www.czasopisma.pan.pl www.journals.pan.pl
Archives of Acoustics Vol. 47, No. 4, pp. 457–467 (2022), doi: 10.24425/aoa.2022.142892

Research Paper

Road, Tram and Aircraft Traffic Noise Annoyance Related to the Number of Noise Events and the Equivalent Sound Level

Jan FELCYN*, Patrycja PTAK

Department of Acoustics, Faculty of Physics, Adam Mickiewicz University
Poznań, Poland

*Corresponding Author e-mail: janaku@amu.edu.pl

(*received January 10, 2022; accepted May 21, 2022*)

Noise mapping is based on long-term noise indicators, such as L_{N} or L_{DEN}. On the other hand, transportation intensity changes during a day (road traffic peak hours) or a year (more flights during holidays) and this variability is not reflected in single sound level values. We wanted to find out whether not only sound level but also the number of noise events is the factor influencing noise annoyance assessment. Ambisonic recordings of real traffic in a city were used. Road, tramway, and aircraft traffic were investigated and two factors were manipulated: the equivalent sound level value and the number of noise events. All stimuli were presented in an anechoic chamber. The results showed that sound level is always a statistically significant parameter while the number of events has an impact only for tramways and airplanes. Moreover, the difference is observed only between one or more subgroups, no matter what the sound level value was. For road traffic this relation was not found to be statistically significant. It was also shown that the existence of tramway bonus or airplane malus is linked with the number of noise events.

Keywords: noise annoyance; noise event; transportation noise; sound level.

1. Introduction

According to World Health Organization (2018) noise pollution poses a high risk for people's health. The influence of noise leads to several effects, such as ischemic heart disease (Dzhambov *et al.*, 2016), depression and anxiety (Beutel *et al.*, 2016), sleep disturbance (Elmenhorst *et al.*, 2014), etc. People also declare that high noise levels are one of reasons for moving outside the city (Beim, Tölle, 2008). Thus, it is important to control the noise and reduce its levels in everyday life.

This controlling procedure is mainly based on the long-term noise metrics, such as L_{N} or L_{DEN}. Their formal usage was introduced in the Environmental Noise Directive in 2002 by the European Commission (European Union, 2002). Those metrics are used to compute noise maps and prepare action plans. However, these characteristics are averaged over the whole year which makes them insensitive to the temporal structure of the sound.

On the other hand, our everyday auditory perception of the world relates to the sounds which are noticeable. This means they stand out from the background noise or are loud enough to mask other sounds and be consciously perceived. Such a single sound is often called a noise event or sound event (Brown, De Coensel, 2018). Regarding transportation noise, this is often the sound produced by a single pass-by of a vehicle (or group of them). Traffic flow intensity can be expressed using number of pass-bys – e.g. as number of light and heavy vehicles per hour on the road or the number of overflights per hour near an airport. All noise events which are noticed by people can influence both their perception of noise and their health, thus it could be important to include this parameter in noise action plans (Torija *et al.*, 2012).

The observation of sound events is crucial, especially in long-term automatic noise monitoring, e.g. near airports. Several methods were introduced to automatically recognize planes' landings and take-offs while ignoring other sounds (Asensio *et al.*, 2012;

Kłaczyński, Pawlik, 2015). This problem is also addressed more generally for the classification of noise events (Kim *et al.*, 2020; Wang *et al.*, 2021).

Obviously, the increase in the number of noise events in a given period of time leads to an increase in the overall sound level – which could sometimes lead to an overestimation of the noise levels in a given area (Gajardo *et al.*, 2014). But what happens when we artificially keep the overall sound level always the same and manipulate the number of events? The relation between number of sound sources and people's reactions to noise was analysed in several studies (Fields, 1984; Sato *et al.*, 1999; Vogt, 2005). The simplest way to express changes in the number of sound events is to provide this number itself. In energy terms, a tenfold increase in the number of events is equal to 10 dB. Thus, when applying linear regression using both sound level values and the number of events, a parameter k was introduced (Vogt, 2005). When k is equal to 10, both predictors – the number of noise events and their level – are equally important. In his paper, Vogt referred to previous papers which addressed the influence of the number of events on annoyance. For example, some experiments did not reveal a relation between the number of overflights and annoyance (Rice, 1977; Vogt *et al.*, 1995). Moreover, Rice (1980) found that when the number of events is lower than 16 per hour, there is no influence on noise annoyance.

However, there are also other approaches: analysis of the quiet-time between events (Morinaga *et al.*, 2018) or the ratio between the duration of noisy events and background noise. This ratio could be defined in different ways, Kaczmarek and Preis (2010) called it the distortion of informational content, while Wunderli *et al.* (2016) introduced the term intermittency ratio (IR).

No matter how we express the changes in the number of noise events, the results of studies are not conclusive. For road traffic noise Sato *et al.* (1999) showed that number of pass-bys is not a statistically significant factor. In contrast, in the studies by Kaczmarek and Preis (2010) and Brink *et al.* (2019) respondents preferred intermittent road traffic noise (lower annoyance ratings) over the fluent noise (higher annoyance ratings). For aircraft noise, Morinaga *et al.* (2018) have found that the correlation between the noisiness perceived by respondents and equivalent sound level increases when the number of overflights (or the duration of quiet-time intervals) were added to the model; this was also showed by Brink *et al.* (2019). Authors also investigated railway noise, and the same conclusions as for aircraft traffic were drawn (the more intermittent a noise is, the higher annoyance ratings).

The problem of the number of noise events, expressed in IR, was also investigated by the author of this article in the preliminary study, limited only to three different sound level values and road traffic solely. The results showed that IR was not a statistically significant factor for the annoyance ratings (Felcyn, 2021). However, this problem needed further research, especially regarding also other types of noise sources. If the number of sound events was the important factor, we would group two independent variables (the number of events and sound levels) and one dependent variable (noise annoyance ratings), and would try to create "equal-annoyance curves" (an analogue of isophonic curves). This would help in the creation of action plans and allow another factor to be included in noise policies.

To find out if such curves can be established, an experiment involving three different noise sources was carried out. In the next sections we will introduce its methodology, present the results, and conclude our findings.

2. Methods

To find out if the number of noise events and equivalent sound level has an influence on people's annoyance assessments, an experiment was prepared and carried out in an anechoic chamber in our laboratory. Three different noise types were presented: road, tramway, and aircraft noise.

2.1. Stimuli

Each type of noise was recorded in the city of Poznań, Poland, in different locations, to preserve the conditions in which the background noise and noise from other sources are minimal. Thus, recordings were made in three various places. All sources were recorded using an ambisonic microphone, giving a 4-channel B-Format file in the output.

Road traffic noise was recorded near a four-lane street with low plants in the middle strip. The noise was recorded during the daytime. Because of the traffic intensity, it was not possible to record single pass-bys, so whole packages (containing several noise events) were registered. The number of light and heavy vehicles was counted. The microphone was placed at a height of 1.2 m and 30 m from the middle of the median strip.

Tramway traffic was recorded behind a shopping mall where the tramway track leads to the near tramway depot. As in Poznań many streets have a tramway in the middle of them, it was the best place to record tramway traffic only. However, the location had two main drawbacks. Firstly, sometimes huge cargo trucks were moving around the mall, generating additional noise. Secondly, near this place there is a railway shunting yard, so there were also trains passing nearby. However, because the recording procedure lasted 2 days, there were enough recordings of good quality.

Again, the noise was recorded with the same ambisonic microphone, 30 m from the middle of a tramway track and at height of 1.2 m.

Aircraft traffic was recorded in Przeźmierowo, the parish adjoining Poznań from the West. This time, the microphone was placed near a quiet street, but directly below the skyway to Ławica Airport (EPPO). The microphone was placed on a (4 m) high stand. Both landings and take-offs were recorded. The recording procedure lasted around one week (to gather all the frequently used machines and both operations).

Raw B-Format data were carefully analysed to detect the best recordings (without other noise sources, high background noise, ambulances, and so on). After the selection process, noise stimuli were prepared.

At first, the duration of each stimulus was set to 5 minutes. This is enough to represent different number of events at the same time. Then, based on the statistics of traffic flow in Poznań, the number of noise events (NNE) represented for each type of noise was chosen. For road traffic noise, having several packages of this noise and taking into account traffic intensity, NNE was set to be: 110, 120, 130, 140, and 168. It was impossible to create stimuli with NNE equal to 150 or 160 because neither of the recording combinations gave this value.

For tramway traffic, NNE was 1, 2, 3, 4, 6, or 8. One tramway per 5 minutes is the flow characterising tramway tracks with only one tramline (the typical frequency of trams in Poznań is one tram per 10 minutes on each tramline). On the other hand, 8 trams per 5 minutes is the heavy traffic observed near important public transportation hubs.

Taking into account that Ławica airport has only one runway, NNE for aircraft traffic was set to 1, 2, or 3. 3 is the maximum number during normal holiday periods (before COVID-19). For both tram and aircraft traffic there was only one recording of a noise event and it was simply multiplied several times across the timeline of a stimulus. This was necessary to eliminate the possible influence of the different types/models of transportation means on the annoyance ratings.

All the created stimuli were presented at five different sound levels: 50, 55, 60, 65, and 70 dBA (5-minute equivalent sound level). To make the stimuli more realistic, all types of noise were presented with the background noise. This noise was recorded in the same place as aircraft noise and included typical city background noise (distant road traffic, steps, sometimes birds, etc.). Background noise was always kept at 40 dBA. Each stimulus was also faded in and out to avoid sudden peaks (2.5 s fades were used at the beginning and in the end).

Summing up, 70 stimuli were created (3 sources presented at 5 sound levels, 5 various NNE for road traffic, 6 for trams and 3 for aircraft noise).

2.2. Participants

70 participants took part in the experiment: 36 men, 32 women, and 2 non-binary people. Their age was between 19 and 58 years with a mean value of 27.4 and SD = 9.6. Participants were recruited through an announcement at the university and information on social media. They filled in a short survey about their experience and attitudes towards different environmental aspects, and noise annoyance. Each participant was also paid for his/her involvement.

2.3. Procedure

2.3.1. Reproduction of stimuli

Each stimulus was presented in an anechoic chamber. A custom-designed reproduction system, consisted of 26 loudspeakers (25x Yamaha NS5 and one subwoofer Velodyne) was used to create the soundscape using the ambisonic technique.

To transform the B-Format recordings into 26 channels a DAW Reaper was used with the plugins from IEM Plug-in Suite. They are an open-source set of audio processors designed for the management of 3D sound. Using them, based on the coordinates of each loudspeaker, the 26-channel master bus was created, taking into account both delay and intensity compensation.

In the middle of the chamber (i.e. in the middle of the whole system) a chair was placed where the participants sat during the listening tests. The whole system was calibrated using the SVAN soundmeter and pink noise, measured in the same place where the participant's head was during the test. Pink noise was emitted from each loudspeaker separately and was measured three times in a row to give the same 1-minute equivalent sound pressure level. A subwoofer was also calibrated regarding its phase coherence with the other loudspeakers. This was done by measuring transfer functions of a subwoofer and the full-band loudspeaker. A measuring microphone was placed above the chair (in the place of a listener's head). Signals from the sub and full-band loudspeakers had to have the same phase, so a very short time delay was applied to the sub to achieve it. The whole system is presented in Fig 1.

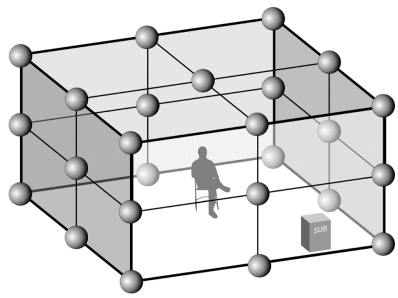

Fig. 1. Ambisonic system used in the experiment.

2.3.2. Listening tests

As has been mentioned before, 70 different stimuli were created based on original transportation noise recordings. They were presented in a random order for each participant. Each participant took part in 3 listening sessions. The listeners were asked to bring a book and read during the session to eliminate effects of concentration on noise. After each stimulus they filled in a short survey in which they chose the type of noise they had heard (cars, trams, planes or other) and then rated the annoyance they had experienced using a numerical ICBEN scale (0-10, according to ISO 15666:2021, International Organization for Standardization, 2021). There was also an open-field question to provide any details which were important for the participant in a given stimulus. After each session a break of at least 24 hours was required before participants listened to the next set of stimuli.

The whole procedure was interrupted several times due to the pandemic situation in Poland and restrictions related to it. Thus, some participants started before lockdown and finished the experiment after that. Therefore, it was not possible to continuously run the experiment during the autumn of 2020 and spring of 2021.

2.4. Acoustical characteristics of stimuli

Apart from subjective data about noise annoyance from participants, we also computed some objective characteristics of the stimuli. To do that, in the chair placed in the anechoic chamber (where participants normally sit) we mounted a dummy head (Neumann KU100) and recorded each stimulus using it. Slightly above the head we also mounted a sound meter to control the overall sound level of the played stimulus.

All the recorded data were then analysed in ArtemiS Suite 11 software. We computed several acoustical and psychoacoustical characteristics – versus time, as single values and percentile values. They were then correlated with annoyance ratings, but this analysis is not a part of this paper.

3. Results

3.1. Participants reliability

Before launching any analysis, all results were inspected in the context of listeners' reliability and consistency. This was done using two reliability tests: Cronbach's alfa and intra-class correlation, both were computed in *R* software using the "psych" package. Each test was also computed in the one-item dropped approach.

The results of these tests were similar: two respondents had significantly lower coefficient values than the others. Thus, they were excluded from further analyses. Moreover, for another respondent many answers were missing. It was decided that also this participant's data would not be analysed. Finally, results were obtained based on data from 67 respondents.

3.2. Repeated measures ANOVA

In the experiment each respondent rated a given stimulus once. As each participant rated all stimuli, such an experimental procedure is known as repeated measures. There were two independent variables – sound level and NNE – and the dependent variable was the annoyance rating. In this case, repeated measures ANOVA was used to determine which factors are statistically significant.

The data were not normally distributed and the assumption of sphericity was not met either. In this case, a non-parametric ANOVA version should be applied. However, a non-parametric two-way design is not trivial, especially regarding possibilities of computing post-hoc tests. Moreover, we wanted to analyse probabilities of our hypotheses conditioned by the gathered data. This problem cannot be checked using the most frequent approach, so the Bayesian version of ANOVA was used. Nevertheless, we also computed non-parametric ANOVA with Conover's post-hoc tests to check if the results are stable and reliable regardless the method used. As the results were the same, we report here only Bayesian approach.

Both Bayesian and non-parametric ANOVA were computed in the JASP software. In the Bayesian method there is no classical p-value, which is somehow replaced by the Bayes Factor (BF). The greater the value of BF is, the stronger the proof that one hypothesis is more probable than the other. According to Jeffreys (1998), a BF value greater than 100 has very strong evidence power. On the other hand, Kass and Raftery (1995) proposed a value of 150 for the same power.

Generally speaking, Bayesian statistics is based on prior beliefs (expressed as a given distribution, for example, binomial) and results from data. When the gathered data are used the distribution changes and we obtain posterior distribution. As we always have two hypotheses – null and alternative – for both of them we have also prior and posterior distributions. To have an idea of how much it is possible that one of the hypotheses is more probable than the other, we compute posterior model odds. BF is a fraction between both posterior model probabilities (for H_0 and H_1, see the Eq. (1)). The Bayes factor quantifies the strength of evidence provided by the data. When BF is denoted as BF_{01}, it is the probability of H_0 over H_1; BF_{10} describes probability of H_1 over H_0:

$$\mathrm{BF}_{01}(d) = \frac{P(H_0|d)}{P(H_1|d)}. \tag{1}$$

The experimental data were divided into three subsets, covering cars, trams, and airplanes, and each subset was analysed separately. At first, the data were analysed using Bayesian repeated measures ANOVA.

3.2.1. Road traffic

Taking into account both independent variables (sound level and NNE), only the former has a high BF value (BF_{10} = 4.328e+164) while the latter is lower than 1 (which means that the greater probability is that this variable is not influential). Model with both factors has also a high BF_{10} value, however, it is slightly lower than that for the sound level only. Finally, a model with both factors and interaction between them has a lower BF_{10} value than both already mentioned models. This suggests that the best model takes into account only the sound level factor. All the results are shown in Table 1.

Table 1. The results of repeated measures Bayesian ANOVA for road traffic.

Models	BF_{10}	Error [%]
Sound level	4.328e+164	0.471
NNE (number of cars)	0.035	1.373
NNE + Sound level	2.759e+164	1.301
NNE + Sound level + NNE * Sound level	2.521e+160	3.583

Post hoc tests for sound level revealed that all groups are different from each other (the smallest BF value was observed between groups of 50 and 55 dBA; BF_{10} = 6.941e+09, while the greatest was between 50 and 70 dBA, BF_{10} = 2.668e+72, see Table 2).

The averaged R^2 of the model was estimated to 65.7%. A visual summary of the analysis is presented in Fig. 2. Each point represents a mean value of the annoyance rating given for a stimulus. Error bars show 95% confidence intervals (CI). In each figure in this manuscript, CI were computed the same way. We used the boot package for R and bootstrapped mean values

Table 2. The results of post hoc tests computed for the sound level factor for road traffic.

Level		Prior odds	Posterior odds	BF_{10}	Error [%]
50	55	0.320	2.820e+09	6.941e+09	7.683e−13
	60	0.320	5.056e+24	3.111e+24	1.331e−28
	65	0.320	4.185e+51	3.060e+51	1.066e−56
	70	0.320	1.573e+74	2.668e+72	3.013e−79
55	60	0.320	339263.473	9.644e+05	7.709e−09
	65	0.320	2.235e+33	2.291e+34	9.349e−38
	70	0.320	6.930e+53	5.294e+55	8.017e−59
60	65	0.320	2.061e+11	2.745e+12	9.478e−15
	70	0.320	3.714e+39	1.714e+43	1.308e−44
65	70	0.320	6.400e+11	1.108e+14	2.965e−15

using 10 000 replications. Then, CI were computed directly from the distributions using the function boot.ci.

3.2.2. Tramway traffic

For this type of noise ANOVA results showed that both the sound level and NNE factors have high BF values (3.764e+133 and 1.044e+12, respectively). A model based on both these factors has BF = 2.97e+152, while a model with additional interaction has BF = 2.228e+150. In this case, the best model for the data is that which includes both NNE and the sound level. Detailed results are presented in Table 3.

Table 3. The results of repeated measures Bayesian ANOVA for tramway traffic.

Models	BF_{10}	Error [%]
Sound level	3.764e+133	0.676
NNE (number of trams)	1.044e+12	0.527
NNE + Sound level	2.970e+152	1.827
NNE + Sound level + NNE * Sound level	2.228e+150	1.153

As both independent variables had large BF values, a post hoc test was run for them. The results of those analyses are shown in Table 4 (sound level) and Table 5 (NNE).

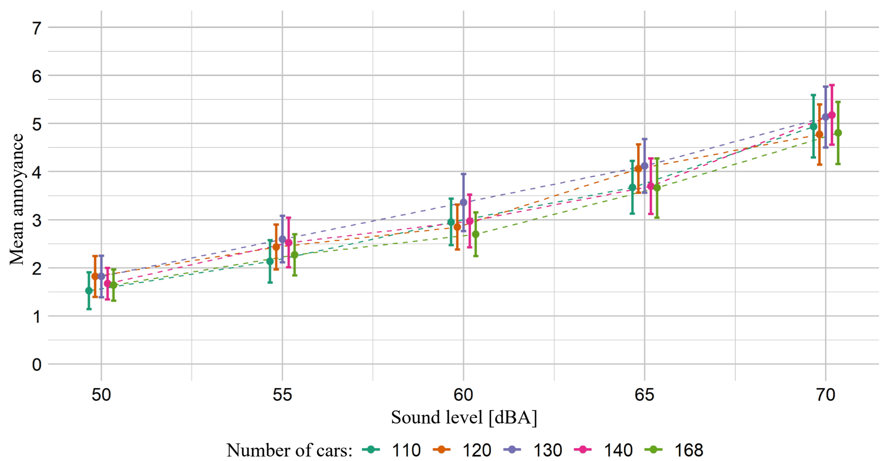

Fig. 2. Mean annoyance ratings with 95% confidence intervals for road traffic noise.

Table 4. The results of post hoc tests computed for the sound level factor for tramway traffic.

Level		Prior odds	Posterior odds	BF_{10}	Error [%]
50	55	0.320	163.055	509.547	1.783e−05
	60	0.320	8.946e+12	2.796e+13	4.071e−16
	65	0.320	1.931e+36	6.034e+36	1.034e−39
	70	0.320	3.561e+70	1.113e+71	4.421e−76
55	60	0.320	40.031	125.097	7.047e−05
	65	0.320	2.048e+22	6.4e+22	1.614e−25
	70	0.320	1.012e+55	3.163e+55	4.556e−59
60	65	0.320	2.025e+13	6.328e+13	1.799e−16
	70	0.320	5.269e+50	1.647e+51	1.366e−54
65	70	0.320	6.768e+18	2.115e+19	5.188e−22

Table 5. The results of post hoc tests computed for the number of pass-bys (NNE) factor for tramway traffic.

NNE		Prior odds	Posterior odds	BF_{10}	Error [%]
1	2	0.260	44134.873	1.697e+05	4.582e−8
	3	0.260	7.485e+7	2.879e+08	2.292e−11
	4	0.260	2.148e+8	8.262e+08	7.783e−12
	6	0.260	7.638e+10	2.938e+11	1.878e−14
	8	0.260	9.229e+12	3.550e+13	1.352e−16
2	3	0.260	0.107	0.412	0.023
	4	0.260	0.582	2.238	0.004
	6	0.260	11.415	43.904	2.058e−4
	8	0.260	133.161	512.158	1.694e−5
3	4	0.260	0.023	0.088	0.110
	6	0.260	0.077	0.296	0.033
	8	0.260	0.240	0.923	0.010
4	6	0.260	0.027	0.104	0.096
	8	0.260	0.045	0.173	0.057
6	8	0.260	0.018	0.069	0.141

As one can see, for sound level each pair has large BF_{10} values – however, the smallest value was computed between 55 and 60 dBA (BF_{10} = 125.097), while the largest was between 50 and 70 dBA subgroups ($BF_{10,U}$ = 1.113e+71).

Another situation was observed for a different number of pass-bys. Generally, large BF_{10} values were found for all combinations between one and more noise events (with BF_{10} values from 1.697e+05 to 3.550e+13). Values of BF_{10} greater than 1 were also found only for three pairs: 2–4 (BF_{10} = 2.238), 2–6 (BF_{10} = 43.904), and 2–8 (BF_{10} = 512.158). All the other combinations had values lower than 1.

All the mean values of noise annoyance ratings with 95% credible intervals are presented in Fig. 3. Also note that for this type of noise R^2 was equal to 58.1%.

3.2.3. Air traffic

For air traffic noise, similarly to tramway noise, both the sound level and NNE factors were found to have large BF_{10} values: sound level – 9.144e+62 and the number of overflights – 9356.771. The model which incorporates both factors has the greatest BF_{10} value, 2.199e+69. Adding the interaction part to the model drops its BF_{10} to 8.938e+65. The statistics for the models are presented in Table 6.

Table 6. The results of repeated measures Bayesian ANOVA for air traffic.

Models	BF_{10}	Error [%]
Sound level	9.144e+62	0.480
NNE (number of planes)	9356.771	0.678
NNE + Sound level	2.199e+69	0.945
NNE + Sound level + NNE * Sound level	8.938e+65	1.170

Regarding the results of post hoc tests, for the sound level factor BF_{10} values were generally high, with two exceptions: a pair of 55 and 60 dBA subgroups had BF_{10} = 12.319, and 50–55 had BF_{10} = 70.890. Detailed results are presented in Table 7.

When analysing the number of overflights, high BF_{10} values were revealed for the relations between 1 and both 2 and 3 noise events (BF_{10} = 5.432e+04 and 1.316e+06, respectively). On the other hand, the BF_{10} value computed between 2 and 3 overflights was very low, 0.072. The results of this analysis are presented in Table 8.

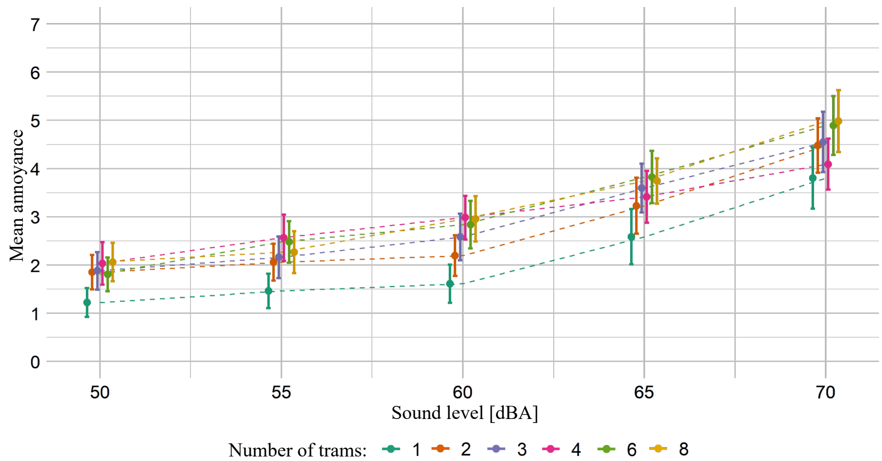

Fig. 3. Mean annoyance ratings with 95% confidence intervals for tramway traffic noise.

Table 7. The results of post hoc tests computed for the sound level factor for air traffic.

Level		Prior odds	Posterior odds	BF_{10}	Error [%]
50	55	0.320	22.685	70.890	2.826e−6
	60	0.320	1.107e+9	3.46e+09	5.381e−14
	65	0.320	1.293e+20	4.04e+20	1.921e−25
	70	0.320	1.252e+32	3.91e+32	1.474e−38
55	60	0.320	3.942	12.319	1.547e−5
	65	0.320	2.387e+11	7.46e+11	2.195e−16
	70	0.320	1.037e+26	3.24e+26	1.072e−31
60	65	0.320	6737.446	21054.52	1.016e−8
	70	0.320	4.325e+19	1.35e+20	6.043e−25
65	70	0.320	254295.424	794673.2	2.655e−10

Table 8. The results of post hoc tests computed for the planes number (NNE) factor for air traffic.

NNE		Prior odds	Posterior odds	BF_{10}	Error [%]
1	2	0.587	31888.232	5.432E+04	1.706e−7
	3	0.587	772557.231	1.316E+06	6.704e−9
2	3	0.587	0.042	0.072	0.152

The mean annoyance ratings for air traffic noise can be seen in Fig. 4 and are provided with 95% credible intervals. For aircraft noise R^2 was 59.5%.

3.3. Linear regression

Based on the results obtained in ANOVA, we decided to compute also linear regression models for each sound source. It can be seen from Figs 2–4 that the relations between mean annoyance ratings and sound level values are linear, especially for cars. For trams and planes the relations are also linear, taking into account credible intervals.

To compute linear regression equations we used the function *lmer* from the *R* package *lmerTest*. This function is robust for violations of assumptions in residuals and the shape of distributions. Moreover, it also takes into account random effects. In this case, the factor "subject" was chosen to be the random effect.

For each type of noise, the regression was computed as the relation between noise annoyance ratings and the NNE and sound level values, "subject" was taken as the random factor. The results for each type of noise are presented in Table 9.

It can be clearly seen that the results of regression analyses are similar to those from Bayesian ANOVA. Once again, sound level is a statistically significant factor in each case, while the NNE is not for road traffic. However, this type of analysis provides some more interesting information.

First of all, it seems that quite a high value of R^2 values from ANOVA is mainly due to the random effect of subjects. In each case, the R^2 value computed only for fixed effects is low (0.21 for road, 0.18 for tramway, and 0.17 for aircraft traffic). Comparing these values to the R^2 of the whole data (0.65, 0.55, and 0.59, respectively) leads to a conclusion that the majority of variance in the data is due to the differences between subjects.

Another fact is that each regression equation seems to have similar values of coefficients for the predictors. The intercept value varies between −6.17 for road and −5.6 for aircraft traffic. On the other hand, the coefficient for the sound level is 0.16 for road, 0.13 for tramway, and 0.14 for aircraft. The biggest differences were found for the coefficient values for NNE. As this predictor is not statistically significant for road traffic, its coefficient in that case is equal to 0. For trams it is 0.13 and for aircraft 0.36 – suggesting that this predictor somehow has a stronger influence for aircraft noise annoyance than for tramways. However, is has to be mentioned that for airplanes only three different numbers of overflights were used, while for tramways it was 6.

For each model also a value of the Bayesian information criterion (BIC) is provided. It has the lowest value for aircraft noise and the highest for trams. However, these are values for three different models, thus this parameter cannot be used to directly compare all the models.

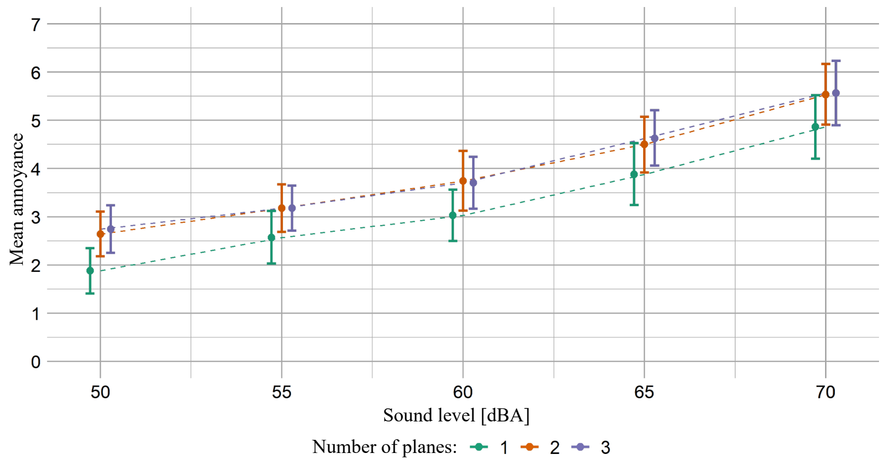

Fig. 4. Mean annoyance ratings with 95% confidence intervals for air traffic noise.

Table 9. The results of robust linear regression analyses computed for each type of noise.

	Road traffic				
	Estimate [95% CI]	Std. Error	df^*	t-value	p
(Intercept)	−6.17 [−7.02, −5.32]	0.43	915.43	−14.20	1.71e−41
Cars' level	0.16 [0.15, 0.17]	0.01	1597.02	31.72	1.14e−171
No. of cars	0 [−0.01, 0.00]	0.00	1597.02	−0.97	0.33
BIC = 6258.93 R^2 (fixed) = 0.21 R^2 (total) = 0.65					
	Tramway traffic				
	Estimate [95% CI]	Std. Error	df^*	t-value	p
(Intercept)	−5.49 [−6.16, −4.82]	0.34	777.95	−16.10	1.51e−50
Trams' level	0.13 [0.12, 0.14]	0.00	1938.01	27.27	7.16e−139
No. of trams	0.13 [0.10, 0.15]	0.01	1938.01	8.77	3.83e−18
BIC = 7653.09 R^2 (fixed) = 0.18 R^2 (total) = 0.55					
	Aircraft traffic				
	Estimate [95% CI]	Std. Error	df^*	t-value	p
(Intercept)	−5.6 [−6.59, −4.60]	0.51	831.69	−11.02	1.78e−26
Planes' level	0.14 [0.13, 0.16]	0.01	935.00	19.41	8.27e−71
No. of planes	0.36 [0.23, 0.48]	0.06	935.00	5.64	2.28e−08
BIC = 4090.26 R^2 (fixed) = 0.17 R^2 (total) = 0.59					

* *df* – effective degrees of freedom estimated using Satterthwaite formula.

4. Discussion

On the basis of the ANOVA and regression results, it can be seen that the relations between annoyance ratings and sound level values are linear for all types of noises. Moreover, no matter which noise we discuss, the values of intercepts and coefficients are quite similar.

From the ANOVA results we know that R^2 values are between 58.1% (for trams) and 65.7% (for road traffic). However, linear regression analyses showed that the majority of variance comes from differences between listeners. R^2 for fixed predictors oscillates around 0.2, which is in line with common findings that not more than 30% of variance in people's answers could be explained by sound level values (Marquis-Favre *et al.*, 2005). However, here R^2 also covers the influence from NNE.

Large differences in participants' responses are not surprising and were found in other research too (Miedema, Oudshoorn, 2001; Schultz, 1978). Thus, there are many non-acoustical factors proposed to explain the diversity of the ratings (Fields, 1993; Job, 1988). They are not covered in this paper, but will be discussed in future manuscripts. At this point we just want to mention that such variability in the answers given by humans is common and should not be treated as an error.

Interestingly enough, the number of pass-bys/acoustical events (NNE) was found to be statistically significant for both tramway and aircraft traffic, but not for road noise. Moreover, for both cases post-hoc analyses revealed that statistically significant differences were observed only between the "one" versus "more than one" subgroups.

To better illustrate this phenomenon, Fig. 5 presents the results in which tramway and aircraft traffic was divided into both those subgroups.

From Fig. 5 we can see that the lowest annoyance ratings were given for the tramway traffic when only one tram was presented. Then, three relations are very

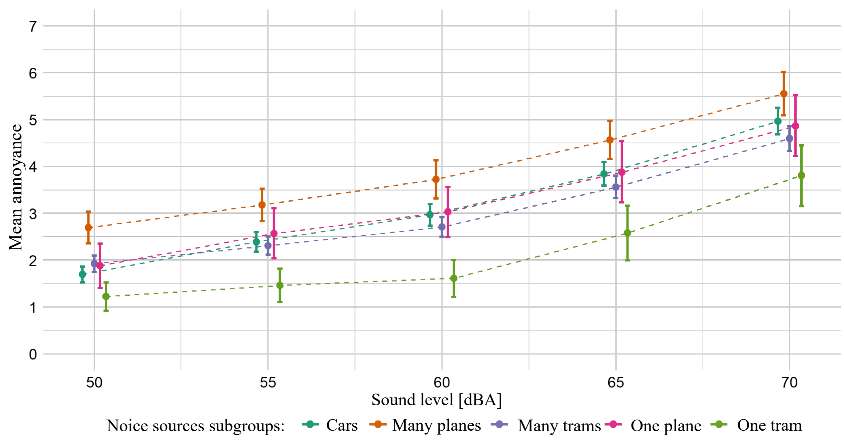

Fig. 5. Mean annoyance ratings of different sound sources while tramway and aircraft traffic is divided into subgroups one and more than one pass-by.

similar to each other: many trams (more than one), road traffic and one plane. Lastly, the highest ratings were obtained for the many planes (more than one) case.

This observation can be interpreted in several ways. Firstly, people were asked to read a book during the experiment. Thus, in some surveys participants reported that they had not heard any noise source (only background noise) – but in reality there was a tram pass-by or plane overflight. In such cases the common rating was zero. We analysed this phenomenon based on answers about recognised noise sources and the values of annoyance ratings.

Globally, cars were recognised as cars in 98.69% of cases. For planes this number was 90.75%, and for trams it was 80.6%. Moreover, trams were mainly misrecognised as cars – in 14.54% of cases. These disproportions are even more apparent when analysing the cases with the minimum NNE. For one pass-by of a tram, people rated it with zero on the ICBEN scale in 25% of observations. For one overflight of a plane it was 14.33%. When all the data are considered, it seems that people tend to treat one noise event as not annoying at all. Especially when we compare it to the stimuli with NNE equal to 2. In these cases, 0 ratings were given for 16.12% of cases for trams, and 4.48% of cases for planes – significantly less frequently than for one pass-by.

Another possible explanation of the low ratings for one noise event cases is that maybe humans have something like a long-term integration mechanism. Single, accidental events are not enough to "excite" the annoyance perception, but more than one (in a period of 5 minutes in this experiment) is enough to evoke higher annoyance. And then, no matter how many pass-bys there would be, the rating is the same. In other words, one noise event is countable but more – not, so no matter how many vehicles were observed, the only difference is between one and more. This assumption could be also the explanation why the number of pass-bys was not statistically influential for road traffic. It is rather impossible that people could count vehicles when their number is so high (110 and more). Thus, each situation was similar to the others, and the only important factor was the sound level. It seems reasonable to conduct further research in which cars would also be represented with only several pass-bys.

On the basis of Fig. 5 one can also draw another interesting conclusion. When neglecting one event cases, there is clearly no tramway bonus (an analogue to railway bonus described by Fastl *et al.* (1996) and Fields, Walker (1982)) or aircraft malus (Gjestland, 2007; Müller *et al.*, 2016). Thus, in contrast to many papers (especially to exposure-response curves established by Miedema and Oudshoorn (2001)), when the different noise sources have the same sound level values, trams are not less annoying and planes are not more annoying than road traffic noise. However, these mechanisms can be clearly observed for the one event stimuli. Roughly estimating, in this case the tramway bonus is between 10 to 15 dB and the aircraft malus is around 10 dB. Thus, it seems that in this experiment both phenomena are present only when one noise event occurs. When there are more of them, differences vanish. This could be the explanation why in some papers a tramway bonus (or rather railway bonus) is observed, while in the others it is not. However, further research is needed to verify this hypothesis.

As was mentioned in the Introduction, there are two different parameters aimed at somehow describing temporal changes in the stimulus. Both of them (DR and IR) were calculated for each stimulus. However, replacing with their values the simple factor of the number of events did not change anything in the analyses of the results. On the other hand, because of the different experimental approach than in (Vogt, 2005), it was impossible to compute the k parameter. Nevertheless, based on linear regression coefficients, some conclusions can be drawn. For road traffic noise, the number of events is not statistically significant, so its coefficient is 0. For tramway noise, both the sound level and the number of events have the same coefficient value, 0.13. Finally, for aircraft noise, the sound level has the coefficient of 0.14 while the number of overflights has a value equal to 0.36. This could suggest that number of events is more influential than sound level values, but it should be noted that for both trams and aircraft statistically significant differences were shown only between one or more noise event subgroups (see ANOVA results). Thus, the values of coefficients are lower when one event observations are excluded from the data. It seems that generally NNE affected annoyance ratings less than the sound level. This finding is in line with results from (Vogt, 2005).

5. Conclusions

In this research we have shown that a sound level is the main influential factor for noise annoyance assessment for all types of stimuli (road, tramway, and aircraft traffic). The number of noise events (NNE) is related to the noise annoyance ratings for tramway and aircraft traffic noise but is not statistically significant for road traffic. However, statistical significance was found only between cases when one noise event occurred versus more than one.

Each type of noise was characterised by similar parameters of linear regression models. Thus, globally tramway bonus and aircraft bonus was not observed. But they are represented in the data when we split tramway and aircraft groups into one event and many events subsets.

The explained variance in annoyance ratings oscillates around 60%. However, only 20% is explained by the relation to computed predictors (mainly equivalent sound level). The rest of the variance has its roots in differences between respondents, so non-acoustical factors should also be taken into account.

Acknowledgments

This research was funded by the National Science Centre with the Preludium grant (2017/25/N/HS6/01164) and by the National Centre for Research and Development (grant NOR/POLNOR/Hetman/0073/2019-00). The authors also declare no conflict of interest.

References

1. Asensio C., Recuero M., Ruizn M. (2012), Aircraft noise-monitoring according to ISO 20906. Evaluation of uncertainty derived from the classification and identification of aircraft noise events, *Applied Acoustics*, **73**(3): 209–217, doi: 10.1016/j.apacoust.2011.09.002.
2. Beim M., Tölle A. (2008), Segregation processes between declining older buildings and suburbanisation – Case study Posen, *disP – The Planning Review*, **174**(3): 51–65, doi: 10.1080/02513625.2008.10557016.
3. Beutel M.E. *et al.* (2016), Noise annoyance is associated with depression and anxiety in the general population – The contribution of aircraft noise, *PLOS ONE*, **11**(5), doi: 10.1371/journal.pone.0155357.
4. Brink M. *et al.* (2019), A survey on exposure-response relationships for road, rail, and aircraft noise annoyance: Differences between continuous and intermittent noise, *Environment International*, **125**: 277–290, doi: 10.1016/j.envint.2019.01.043.
5. Brown A.L., De Coensel B. (2018), A study of the performance of a generalized exceedance algorithm for detecting noise events caused by road traffic, *Applied Acoustics*, **138**: 101–114, doi: 10.1016/j.apacoust.2018.03.031.
6. Dzhambov A.M., Dimitrova D.D. (2016), Association between noise pollution and prevalent ischemic heart disease, *Folia Medica*, **58**(4): 273–281, doi: 10.1515/folmed-2016-0041.
7. Elmenhorst E.-M., Quehl J., Müller U., Basner M. (2014), Nocturnal air, road, and rail traffic noise and daytime cognitive performance and annoyance, *The Journal of the Acoustical Society of America*, **135**(1): 213–222, doi: 10.1121/1.4842475.
8. European Union (2002), *Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise*.
9. Fastl H., Kuwano S., Namba S. (1996), Assessing the railway bonus in laboratory studies, *Journal of the Acoustical Society of Japan (E)*, **17**(3): 139–148, doi: 10.1250/ast.17.139.
10. Felcyn J. (2021), The influence of a signal's time structure on the perceived noise annoyance of road traffic noise, *Journal of Environmental Health Science and Engineering*, **19**: 881–892, doi: 10.1007/s40201-021-00655-4.
11. Fields J.M. (1984), The effect of numbers of noise events on people's reactions to noise: An analysis of existing survey data, *The Journal of the Acoustical Society of America*, **75**(2): 447–467, doi: 10.1121/1.390469.
12. Fields J.M. (1993), Effect of personal and situational variables on noise annoyance in residential areas, *The Journal of the Acoustical Society of America*, **93**(5): 2753–2763, doi: 10.1121/1.405851.
13. Fields J.M., Walker J.G. (1982), Comparing the relationships between noise level and annoyance in different surveys: A railway noise vs. aircraft and road traffic comparison, *Journal of Sound and Vibration*, **81**(1): 51–80, doi: 10.1016/0022-460X(82)90177-8.
14. Gajardo C.P., Morillas J.M.B., Escobar V.G., Vílchez-Gómez R., Gozalo G.R. (2014), Effects of singular noisy events on long-term environmental noise measurements, *Polish Journal of Environmental Studies*, **23**(6): 2007–2017.
15. Gjestland T. (2007), The socio-economic impact of noise: A method for assessing noise annoyance, *Noise and Health*, **9**(35): 42–44, doi: 10.4103/1463-1741.36979.
16. International Organization for Standardization (2021), ISO/TS 15666:2021 – Acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys, https://www.iso.org/standard/74048.html (access: 12.10.2021).
17. Jeffreys H. (1998), *The Theory of Probability*, 3rd ed., Oxford University.
18. Job R.F.S. (1988), Community response to noise: A review of factors influencing the relationship between noise exposure and reaction, *The Journal of the Acoustical Society of America*, **83**(3): 991–1001, doi: 10.1121/1.396524.
19. Kaczmarek T., Preis A. (2010), Annoyance of time-varying road-traffic noise, *Archives of Acoustics*, **35**(3): 383–393.
20. Kass R.E., Raftery A.E. (1995), Bayes factors, *Journal of the American Statistical Association*, **90**(430): 773–795, doi: 10.1080/01621459.1995.10476572.
21. Kim J., Min K., Jung M., Chi S. (2020), Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition, *Building and Environment*, **181**: 107092, doi: 10.1016/j.buildenv.2020.107092.
22. Kłaczyński M., Pawlik P. (2015), Automatic detection system of aircraft noise events during acoustic

climate long-term monitoring near airport, *Vibroengineering PROCEDIA*, **6**: 352–356.

23. Marquis-Favre C., Premat E., Aubrée D. (2005), Noise and its effects – A review on qualitative aspects of sound. Part II: Noise and Annoyance, *Acta Acustica united with Acustica*, **91**(4): 626–642.

24. Miedema H.M., Oudshoorn C.G. (2001), Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals, *Environmental Health Perspectives*, **109**(4): 409–416, doi: 10.1289/ehp.01109409.

25. Morinaga M., Tsukioka H., Kaku J., Kuwano S., Namba S. (2018), A laboratory investigation into the effect of quiet-time interval between aircraft noise events on overall noisiness, *The Journal of the Acoustical Society of America*, **144**(1): 11–22, doi: 10.1121/1.5044403.

26. Müller U., Elmenhorst E.-M., Mendolia F., Basner M., McGuire S., Aeschbach D. (2016), Effects of nocturnal air and rail traffic noise on sleep, [in:] *Proceedings of the 22nd International Congress on Acoustics*.

27. Rice C.G. (1977), Development of cumulative noise measure for the prediction of general annoyance in an average population, *Journal of Sound and Vibration*, **52**(3): 345–364, doi: 10.1016/0022-460X(77)90564-8.

28. Rice C.G. (1980), Trade-off effects of aircraft noise and number of events, [in:] *Proceedings of the 3rd International Congress on Noise as a Public Health Problem*, pp. 495–510.

29. Sato T., Yano T., Björkman M., Rylander R. (1999), Road traffic noise annoyance in relation to average noise level, number of events and maximum noise level, *Journal of Sound and Vibration*, **223**(5): 775–784, doi: 10.1006/jsvi.1999.2153.

30. Schultz T.J. (1978), Synthesis of social surveys on noise annoyance, *The Journal of the Acoustical Society of America*, **64**(2): 377–405, doi: 10.1121/1.382013.

31. Torija A.J., Ruiz D.P., Alba-Fernandez V., Ramos-Ridao Á. (2012), Noticed sound events management as a tool for inclusion in the action plans against noise in medium-sized cities, *Landscape and Urban Planning*, **104**(1): 148–156, doi: 10.1016/j.landurbplan.2011.10.008.

32. Vogt J. (2005), The relative impact of aircraft noise and number in a full-factorial laboratory design, *Journal of Sound and Vibration*, **282**(3–5): 1085–1100, doi: v10.1016/j.jsv.2004.03.059.

33. Vogt J., Lemburg C., Jaencke L., Kalveram K.-T. (1995), Trading level for number: further evidence that experimentally induced annoyance does not increase, *Inter-Noise 95 Proceedings*, pp. 857–860.

34. Wang Y., Zhao G., Xiong K., Shi G., Zhang Y. (2021), Multi-scale and single-scale fully convolutional networks for sound event detection, *Neurocomputing*, **421**: 51–65, doi: 10.1016/j.neucom.2020.09.038.

35. World Health Organization (2018), *Environmental Noise Guidelines for the European Region (2018)*, available at: https://www.who.int/europe/publications/i/item/9789289053563 (access: 25.02.2019).

36. Wunderli J.M. *et al.* (2016), Intermittency ratio: A metric reflecting short-term temporal variations of transportation noise exposure, *Journal of Exposure Science and Environmental Epidemiology*, **26**(6): 575–585, doi: 10.1038/jes.2015.56.