Model of total annoyance due to combined transportation sound sources in simulated noise scenarios

Jan Felcyn

Manuscript submitted to obtain the doctoral degree in Physics

Supervisor: Anna Preis

Poznań, 2019

This thesis is dedicated to my grandfather Eugeniusz,

professor-emeritus of the Poznań University of Technology, my personal tutor who showed me the beauty of science (and mountains too).

Abstract

Noise is one of the most popular and important factors affecting public health. According to the World Health Organization, more than one million healthy life years in Europe are lost annually due to noise. Since 2002, the European Union has imposed an obligation on Member States to conduct policies focused on monitoring and reducing noise. National governments have enacted local legislation on environmental noise limits. Their values are associated with so-called dose-response curves. These curves describe the reaction of people – expressed as an assessment of annoyance – to noise at different doses. The dose is expressed in sound level values. The dose-response curves are described separately for each type of noise source (road, rail, and aircraft traffic). Meanwhile people, especially city dwellers, are most often exposed to noise from multiple sources occurring at the same time. The issue of the impact of such noise on people is still under discussion and it is not known how to treat the different sound sources that occur at the same time together.

Among the many health effects caused by noise, annoyance is one of the most frequently studied. Noise annoyance is assessed by humans using a standardized procedure and specific scales (standard annoyance scales proposed by ICBEN - International Commission on Biological Effects of Noise). Annoyance can be assessed for both individual types of noise sources and their simultaneous combination - in this case it is called total annoyance. In recent years several models of total annoyance have been proposed, but the results are ambiguous. Moreover, in the laboratory conditions the most often presented recordings do not reflect the real situations encountered in cities every day.

The main purpose of this thesis is to propose new models of total annoyance based on combinations of noise sources simulated in laboratory conditions. At the same time, the existing models of total annoyance were also verified. The combinations were created based on the actual situation in Poznań. Two experiments were carried out in which the listener was "surrounded by sound" thanks to the use of the ambisony technique. The first experiment concerned the most popular type of noise, which is road noise. The aim of the experiment was to find a relationship between the assessment of annoyance of this type of noise and the objective characteristics of the recorded sound. The statistical significance of several relationships was demonstrated, and some of them explained about 70% of variance in the listeners' answers. The second experiment concerned a combination of different types of sound sources (road, tram and aircraft noise) occurring simultaneously. Based on the results obtained, two new models of total annoyance were proposed: Partial Annoyances-Based Model (PABaM) and Personal Characteristics-Based Model (PeCBaM). On the other hand, among the existing models the best suited to the data turned out to be two: the strongest component model and the combined noise source paradox.

Keywords: noise annoyance, total annoyance, urban noise, multiple noise sources, mixed noise sources, road traffic, aircraft, trams, perception of noise, noise effects, public health

Abstrakt

Hałas jest jednym z najpopularniejszych i najważniejszych czynników wpływających na zdrowie publiczne. Zgodnie z danymi Światowej Organizacji Zdrowia, ponad milion zdrowych lat życia jest rocznie traconych z powodu hałasu. Od 2002 roku Unia Europejska nakłada na kraje członkowskie obowiązek prowadzenia polityki skoncentrowanej na monitorowaniu i ograniczaniu hałasu. Rządy państw uchwaliły lokalne akty prawne regulujące dopuszczalne poziomy hałasu w środowisku. Ich wartości są powiązane z tzw. krzywymi dose-response. Krzywe te opisują reakcję ludzi, wyrażoną jako ocenę dokuczliwości hałasu, na hałas o różnej dawce. Dawka wyrażana jest w wartościach poziomu dźwięku. Krzywe dose-response zostały opisane osobno dla poszczególnych rodzajów źródeł hałasu (drogowy, szynowy, lotniczy). Tymczasem ludzie, zwłaszcza mieszkańcy miast, są najczęściej narażeni na hałas wielu źródeł występujących równocześnie. Kwestia wpływu takiego hałasu na ludzi wciąż podlega dyskusji i nie wiadomo jak należy traktować łącznie, jednocześnie występujące, różne źródła dźwięku.

Wśród wielu efektów zdrowotnych wywołanych przez hałas jednym z najczęściej badanych jest dokuczliwość. Dokuczliwość hałasu jest oceniana przez ludzi przy użyciu ustandaryzowanej procedury i konkretnych skal (standardowe skale dokuczliwości zaproponowane przez ICBEN - *International Commission on Biological Effects of Noise*.) Dokuczliwość hałasu może być oceniana zarówno dla pojedynczych typów źródeł hałasu jak i ich jednoczesnej kombinacji – wówczas nazywana jest całkowitą dokuczliwością. W ostatnich latach zaproponowano kilka modeli całkowitej dokuczliwości, jednak wyniki badań są niejednoznaczne. Co więcej, w warunkach laboratoryjnych prezentuje się najczęściej nagrania, które nie odzwierciedlają rzeczywistych sytuacji spotykanych na co dzień w miastach.

Głównym celem niniejszej pracy jest zaproponowanie nowych modeli całkowitej dokuczliwości hałasu w oparciu o symulowane w warunkach laboratoryjnych kombinacje trzech różnych źródeł hałasu. Zostały one stworzone w oparciu o rzeczywistą sytuację występującą w Poznaniu. Jednocześnie, dokonano również weryfikacji istniejących modeli całkowitej dokuczliwości. W pracy przeprowadzono dwa eksperymenty, w których słuchacz był „otoczony dźwiękiem” dzięki użyciu techniki ambisonii. Pierwszy eksperyment poprzedzający główną część pracy dotyczył najpopularniejszego rodzaju hałasu jakim jest hałas drogowy. Celem tego eksperymentu było znalezienie zależności pomiędzy ocenami dokuczliwości hałasu drogowego i obiektywnymi charakterystykami zarejestrowanych nagrań. Wykazano istotność statystyczną kilku zależności, a niektóre z nich wyjaśniały ok. 70% wariancji w odpowiedziach słuchaczy. Drugi eksperyment dotyczył kombinacji trzech różnych rodzajów źródeł dźwięku – hałas drogowy, tramwajowy i lotniczy – występujących równocześnie. Bazując na otrzymanych wynikach (również z eksperymentu pierwszego) zaproponowano dwa nowe modele całkowitej dokuczliwości: model bazujący na cząstkowych ocenach dokuczliwości (PABaM) oraz na indywidualnych charakterystykach słuchaczy (PeCBaM). Weryfikacja istniejących modeli pokazała, że najlepiej pasujące do danych okazały się być dwa: strongest component model oraz combined noise source paradox.

Słowa kluczowe: dokuczliwość hałasu, całkowita dokuczliwość, hałas miejski, źródła hałasu występujące równocześnie, kombinacje źródeł hałasu, hałas drogowy, samoloty, tramwaje, percepcja hałasu, efekty hałasu, zdrowie publiczne

Table of Contents

List of figures

List of tables

1 Review of relevant research

1.1 Meaning of a term “annoyance”

While considering noise and its annoyance, one can ask what actually ‘annoyance’ is. This problem is not trivial as some different annoyance’s definitions were established during last decades. Laird & Coye (1929) referred to Titchener’s theory of emotion. Based on it, people perceive a sound affectively, thus it produces feelings of pleasure or displeasure. But the term ‘annoyance’ could be understood in more ways – and this fact is described by (Guski, 1997). As he pointed out, annoyance could be also described as an attitude, knowledge or even as a consequence of rational decisions.

As papers about noise annoyance multiplied, more annoyance definitions were proposed. However, there is still no scientific agreement on how to define it. One of the best known proposal can be found in the norm ISO/TS 15666. With regards to it, annoyance due to noise is “one person's individual adverse reaction to noise”. This terse and very wide definition does not make anything more clear, even worse, it provokes more questions. We all agree that people differ from each other in many ways so can we be more specific while defining annoyance?

One can conclude that of course, people are different, but the main concept of noise annoyance should be the same in Australia, Japan or Europe. But, as Guski (1997) showed, it is not true – people from various cultures have the same rough concept of annoyance, but they differ in details. Guski showed an example in which *“the Japanese would specify the sound characteristics as "clamorous", and the Chinese would add "dull" sound characteristics”.*

Another problem is that in literature there is a distinction between short-term and long-term annoyance (Bartels et al., 2015a; Schreckenberg and Schuemer, 2010). The first one is related to the situation when people are shortly exposed to noise, especially in laboratory conditions. The second one reflects people’s global opinion about noise, including all experiences they have – this is the annoyance which is investigated mainly during in situ research while people fulfill surveys. In this approach, many other factors could influent people’s judgments, not only noise itself but also social or psychological indicators.

There are also some attempts to introduce a completely new approach to the problem of annoyance – like the one by (Schreckenberg et al., 2018). The authors suggest in their work that annoyance is a multidimensional construct and cannot be reliably tested using only ICBEN scales. They propose to use a new measure called a Multiple-Item Annoyance Scale (MIAS). Based on aircraft noise annoyance data from HYENA study (Babisch et al., 2009a) as well as railway and road traffic noise, they confirmed high usability of a new index.

Nevertheless, noise annoyance rated in standardized scales is still the most common approach in research concerning noise sources. It is frequently related to sound levels observed in the environment. The next subsection contains the most commonly used noise indicators and their definitions.

1.2 Definitions of various noise indicators

1.2.1 Sound levels in the environment

Existing noise indicators, such as the yearly averaged day-evening-night A-weighted sound pressure level, L_{DEN} (or, in North America, averaged day-night level, L_{DN}), the equivalent sound level (in a given time window) $L_{Aeq,\,T}$, and the A-weighted sound exposure level L_{AE} (or SEL) are thought to be good predictors of noise annoyance caused by different sound sources. In other words, the higher the value of the noise indicator, the higher the value of the noise annoyance rating.

From these three indicators, only L_{AE} describes the sound level of a single sound event (such as, e.g., one pass-by of a car), see Eq.1.

$$L_{AE} = 10\log\left(\frac{E_A}{p_0^2 t_0}\right), \quad E_A = \int_{-\infty}^{+\infty} p_A^2(t)dt, \tag{1}$$

where p_0 is reference sound pressure ($p_0 = 2 * 10^{-5}\ Pa$), $t_0 = 1s$ and $p_A\ (t)$ is A-weighted sound pressure changing during time *t.*

$L_{Aeq,\,T}$ is averaged over the time *T* value of sound level (see Eq.2).

$$L_{AEq,T} = 10\log\left(\frac{1}{T}\int_0^T 10^{0,1L_{pA}(t)dt}\right) \tag{2}$$

where $L_{pA}(t)$ is the A-weighted sound pressure level changing during time *t*.

There is a relationship between $L_{Aeq,\,T}$ and L_{AE}. If $L_{Aeq,\,T}$ is calculated for the same sound events, this relationship has the form presented in Eq. 3. If there are different sound events, Eq. 4 represents this relationship.

$$L_{AEq,T} = L_{AE} + 10\log\left(\frac{Nt_0}{T}\right) \tag{3}$$

$$L_{AEq,T} = 10\log\left(\frac{t_0}{T}\sum_{i=1}^{n} 10^{0,1L_{AEi}}\right) \tag{4}$$

The most complex noise indicator, averaged over the longest duration (yearly), is L_{DEN}. It takes into account the whole year, dividing each day into three specific periods and is defined in a European Union Directive 2002/49/EC “relating to the assessment and management of environmental noise” (European Union, 2002):

- **Day**: the time from 7 AM to 7 PM. The averaged sound level of all the days throughout a year is called day level L_D,
- **Evening**: the time from 7 PM to 11 PM. The averaged sound level of all the evenings throughout a year is called evening level L_E. This level is increased by a penalty of 5 dB(A), which reflects the fact that people are more sensitive to noise during the evenings than during the days,
- **Night:** the time from 11 PM to 7 AM. The averaged sound level of all the nights throughout a year is called night level L_N. This level is increased by a penalty of 10 dB(A), which reflects the fact that people are the most sensitive to noise at night as they sleep and noise can disturb their rest.

According to (European Union, 2002), these time periods could be changed by national governments – thus, in Poland, they are defined as: day, from 6 AM to 6 PM, evening, from 6 PM to 10 PM and night, from 10 PM to 6 AM.

In Poland, L_{DEN} is defined in Eq. 5.

$$L_{DEN} = 10\log\left[\frac{1}{24}\left(12 * 10^{0,1L_D} + 4 * 10^{0,1(L_E+5)} + 8 * 10^{0,1(L_N+10)}\right)\right] \tag{5}$$

An indicator similar to L_{DEN} is called L_{DN} and it is used mainly in North America. A penalty used for night time is the same as in L_{DEN} (+10 dB). However, it does not include an ‘evening’ period and time gaps are defined as follows:

- **Day**: the time from 7 AM to 10 PM
- **Night:** the time from 10 PM to 7 AM. The averaged sound level of all the nights throughout a year is called night level L_N. This level is increased by a penalty of 10 dB(A), which reflects the fact that people are the most sensitive to noise at night as they sleep and noise can disturb their rest.

L_{DN} is defined in Eq. 6.

$$L_{DN} = 10\log\left[\frac{1}{24}\left(15 * 10^{0,1L_D} + 9 * 10^{0,1(L_N+10)}\right)\right] \tag{6}$$

All of these indicators are frequently used to measure noise in the environment or during lab studies which investigate a problem of noise annoyance. More detailed information about such studies and their development over years can be found in the next subsection.

1.2.2 Other sound characteristics and their influence on noise annoyance

As only 30% of variance in noise annoyance assessments can be explained by the relation with sound level, other predictors were also investigated. One of the most obvious is loudness which bases on perceptual mechanisms lying behind the process of hearing. There are several models of loudness but each of them is aimed to predict the subjective loudness perceived by people.

Loudness was found to be a good predictor of noise annoyance ratings for urban road traffic (Freitas et al., 2012; Gille et al., 2016c). The other class of predictors could be psychoacoustical characteristics, introduced by Fastl and Zwicker, (2007). Definitions of them can be found below:

- Fluctuation strength (FS), this characteristic is aimed to measure amplitude-modulated sounds and is expressed in vacils. 1 vacil is defined for a 60dB 1 kHz tone modulated at 100% with a tone of 4Hz (Fastl and Zwicker, 2007). Fluctuation strength was found to be a good predictor of road traffic noise annoyance (Kaczmarek and Preis, 2010),

- Roughness, *"roughness is created by the relatively quick changes produced by modulation frequencies in the region between about 15 to 300 Hz. (…) To define the roughness of 1 asper, we have chosen the 60 dB 1 kHz tone that is 100% modulated in amplitude at a modulation frequency of 70 Hz"* (Fastl and Zwicker, 2007). Roughness is a characteristic which can predict noise annoyance of powered two wheelers (Paviotti and Vogiatzis, 2012). Ambiguous results were found for road traffic noise annoyance: Kaczmarek and Preis, (2010) reported a significant correlation while Freitas et al., (2012) did not reveal it,

- sharpness, *"the most important parameters influencing sharpnes are the spectral content and the centre frequency of narrow-band sounds. (…) The reference sound producing 1 acum is a narrow-band noise one critical-band wide at a centre frequency of 1 kHz having a level of 60 dB"* (Fastl and Zwicker, 2007). Sharpness was investigated for road traffic noise by Freitas et al., (2012); Kaczmarek and Preis, (2010), however no proof for a significant correlation was found

Most of psychoacoustical characteristics were incorporated into a new characteristic, psychoacoustical annoyance (PAnn), defined by (Fastl & Zwicker, 2007) as an objective characteristic aimed to correlate with subjective ratings of noise annoyance (*"psychoacoustic annoyance can quantitatively describe annoyance ratings obtained in psychoacoustic* *experiments. Basically, psychoacoustic annoyance depends on the loudness, the tone colour, and the temporal structure of sounds”*). It is related to loudness N_5, fluctuation strength, roughness and sharpness according to following equations:

$$PAnn = N_5\left(1+\sqrt{w_S^2+w_{FR}^2}\right) \tag{7}$$

$$w_S = \left(\frac{S}{acum} - 1.75\right) * 0.25 \log_{10}\left(\frac{N_5}{sone} + 10\right)\ for\ S > 1.75\ acum \tag{8}$$

$$w_{FR} = \frac{2.18}{\left(\frac{N_5}{sone}\right)^{0.4}}\left(0.4 * \frac{F}{vacil} + 0.6 * \frac{R}{asper}\right), \tag{9}$$

where N_5 is a percentile loudness in sone, *S* is sharpness, *F* – fluctuation strength and *R* – roughness. *PAnn* is a dimensionless value. N_5 is *“the loudness which is reached or exceeded in 5% of the measurement time. This means that N_5 represents a loudness value close to the maxima of the loudness-time function of the noise immission”* (Fastl and Zwicker, 2007). Psychoacoustical annoyance was also investigated by Kaczmarek and Preis (2010) but no evidence for its significant relation to people’s noise annoyance assessments was found.

The other class of characteristics which can be taken into account are subjective characteristics of individuals. The most commonly used is people’s sensitivity to noise which is assessed by themselves using dedicated surveys – like e.g. NoiSeQ (Griefahn et al., 2007). Moreover, in literature there are evidence that even people’s fears or attitudes towards the environment and different noise sources can influence their noise annoyance ratings (Bartels et al., 2015b; Marquis-Favre et al., 2005; Méline et al., 2013; Okokon et al., 2015a; van den Berg et al., 2015). All of these factors can be described as subjective as they vary regarding different individuals.

Both psychoaoustical and subjective characteristics can influence peple’s noise annoyance judgments. Nevertheless, there is still no strong evidence if the such relations exist and what is their nature.

1.3 Noise annoyance research over years. History of dose-response curves and ways of rating noise annoyance

The problem of noise annoyance is a field investigated by researchers since many years. One of the first papers concerning the relation between sound and annoyance is the research by Laird & Coye, (1929). In their work, the authors described annoyance mainly as an emotion-like reaction and they were focused on it from workers' point of view. The seven different octave tones, from 64 to 8192Hz were presented in pairs (all possible combinations) to listeners who pointed out which tone had been more annoying. Tones were presented at a constant level of 50 TU (transmission unit, TU was an 'earlier version' of decibel and very similar to it). Laird & Coye found that low and high tones were the most annoying while medium ones (256, 512 and 1024 Hz) were perceived as less annoying.

What might be interesting is that as early as in 1929 the author suggested that *"there is a close relationship between loudness and annoyance, but whether it is a straight line or a curvilinear relationship can only be conjectured."*

Another interesting paper was published in 1946 by Berrien. In this review work author mentioned previous research concerned the problem of noise annoyance, mainly in workplaces. Berrien argued that a degree of annoyance should be described 'from a theoretical zero to some maximum point'. The main conclusion was that some acoustical improvements in work areas (like installing sound-absorbing materials on the walls) could reduce faults and errors made by employees during their duties.

As transportation intensity grew after the 2nd World War, more eyes turned to the problem of noise annoyance – especially regarding the aircraft and road traffic noise. In 60-s and 70-s papers started to concern with noise annoyance induced by big airports – like Heathrow or Los Angeles International Airport (LAX) – or busy cities' streets – like these in London or Paris. Many of these papers were included in one of the most important review research written by Schultz, (1978).

In his work, Schultz synthesized data from a dozen or so field studies, where people living near noisy areas assessed noise annoyance. His main aim was to correlate annoyance's rates with noise exposure expressed as L_{DN}. This work was, somehow, a turning point in the problem of noise annoyance as it was the first paper in which a curve, named later as a 'dose-response curve', appeared.

The main problem in Shultz's research was related to the connection between noise annoyance and long-term noise indicators. As the author noticed, *"it was observed that the correlation between the noise exposure and the individual subjective reactions was poor; typical correlation coefficients ran around 0.3 to 0.4"*. Moreover, in 70-s there was no standardized scale for assessment of noise annoyance, so many scales were used, depending on authors' preferences or assumptions.

However, Schultz noticed that the data is scattered wider when noise levels are lower. He stated that when noise is loud enough, it starts disturbing people and then, they are more consistent in their judgments and deviation in answers is getting narrower. Basing on this observation, Schultz proposed to use, in correlation analysis, only data from people who are "highly annoyed".

But the new question arose: how to define people who are 'highly annoyed' by a given noise source? Roughly speaking, Schultz assumed that all answers expressed as 'very', 'highly', 'extremely' etc. could be described as 'highly'. Based on those assumptions, he concluded that 27-29% from the top of every scale should be taken into account. Of course, the whole process was not straightforward, more details can be found directly in Schultz's work.

Taking all these elements together, Schultz used data from 11 different papers and established a curve reproduced in Figure 1. It was described as follows: *"The mean of the "clustering surveys" data, shown here, is proposed as the best currently available estimate of public annoyance due to transportation noise of all kinds. It may also be applicable to community noise of other kinds".*

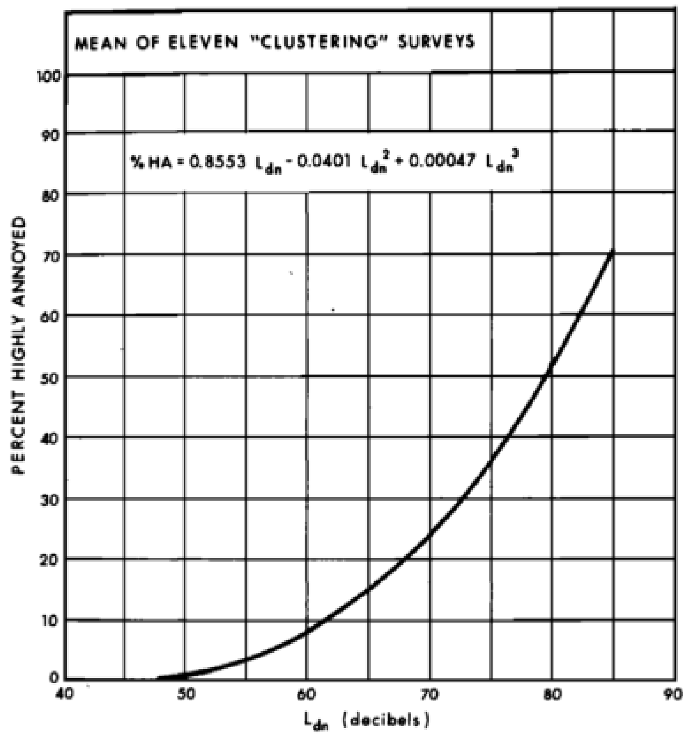


Figure 1. A dose-response curve, established by Schultz (1978) from eleven 'clustering' surveys (all types of noise were treated together).

This type of curve was later named "dose-response curve" because it represents noise dose (expressed in L_{DN} indicator) on the x-axis and percent of people who are highly annoyed (it represents their response, reaction to noise) on the y-axis.

Schultz's work was both significant for the problem of predicting noise influence on people and (at the same time) controversial. Just several years later Hall et al., (1981) pointed out that *"for the same value of L_{DN}, a greater percentage of the sample is highly annoyed by aircraft noise than by road traffic noise"* what was the contrary to Schultz's findings. In 1982 another paper regarding noise annoyance was published by Kryter (Kryter, 1982). In this research author pointed out, that Schultz made some mistakes – and these mistakes were severe, as *"it was prepared to serve as a guide for noise control purposes and has been incorporated in a report of guidelines for environmental impact statement"*. Kryter provided a thorough analysis of aircraft and road traffic noise. He showed that aircraft noise is much more annoying than generated by cars. In that case, he stated, Schultz's work underestimated the percentage of people annoyed by aircraft noise and overestimated those influenced by road traffic. Moreover, Kryter computed the difference between these two kinds of noise to be about 10dB: *"Ten dB or so should be subtracted from L_{dn}'s for street and road traffic noise measured at the front of houses in order to compare the annoyance impact therefrom to the annoyance from aircraft flyover noise in L_{dn}. The result is called the effective L_{dn}, relative to L_{dn} from aircraft noise"*. This approach of comparing two types of noises by artificially adjusting them to give the same annoyance will be then improved by (Vos, 1992).

As time was going, more and more papers concerning noise annoyance were published and they became more detailed. Researchers started to focus on one type of noise annoyance and deeply analyzed its nature. E.g. Fidell et al. (1985) showed that annoyance induced by aircraft noise could be different around large airports compared to small ones.

The data gathered in new papers were then used by Fidell, Barber, & Schultz, (1991) to update already existing Schultz's dose-response curve. Authors stated that even the data almost tripled (comparing to original work), an old curve still gave reasonable results. However, the data was again scattered – it could be observed in Figure 2 where data points and two dose-response curves (original one and updated one) are presented.

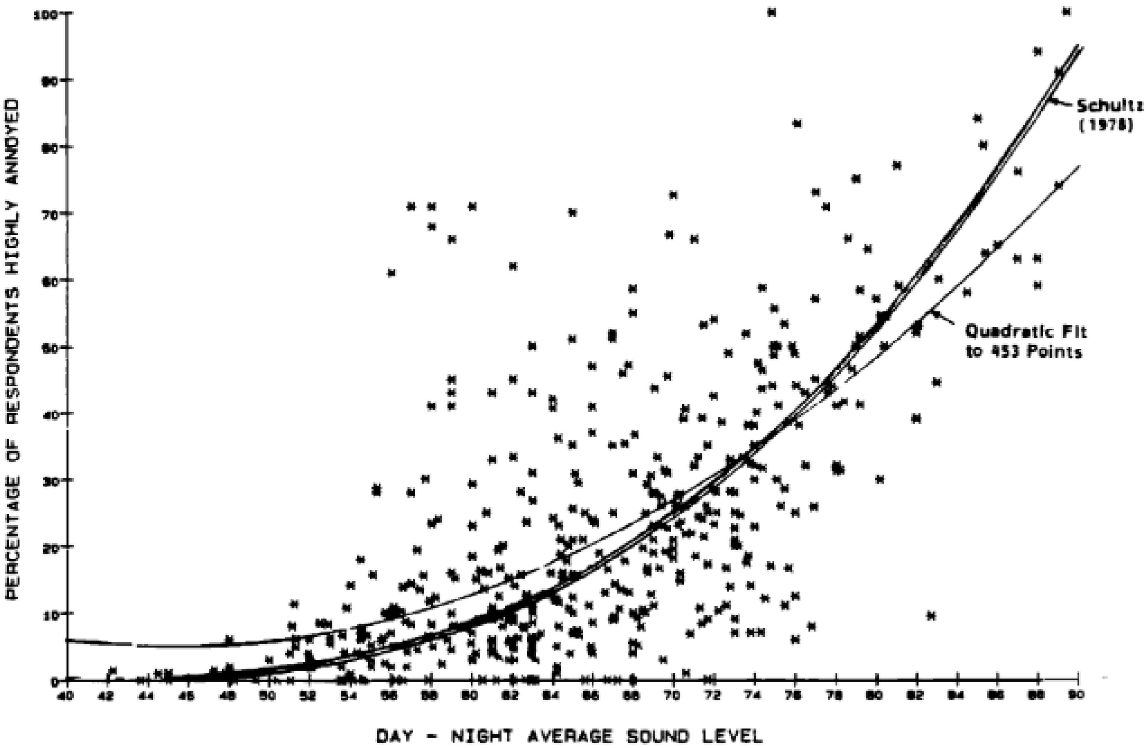


Figure 2. The original dose-response curve presented by Schultz and a new quadratic fit basing on new data. Source: Fidell et al., (1991)

The data grew and more detailed knowledge about relations between noise annoyance and sound level was provided. However, there was still one big obstacle while comparing different findings: researchers tended to use different scales to estimate noise annoyance. In 1993 International Commission on Biological Effects of Noise (ICBEN) started a project aimed to establish a standardized noise annoyance scale. Discussion and research lasted 7 years, finally in 2001, a new scale was proposed (we will refer to that publication later).

This problem of divergent scales was also spotted by James M. Fields, (1994) in his report for NASA. He pointed out 7 problems of Schultz's approach, however the scale seemed to be the greatest obstacle. Interestingly enough, Fields was the very man who led the ICBEN project to find a new standardized noise annoyance scale.

Not only Kryter and Fields realized that the problem of noise annoyance is not straightforward. It was Guski, (1997) who pointed out that, apart from the methodology and type of noise source, the question about annoyance itself can generate bias and answers would be affected by individuals' judgments and experiences.

Only one year later another very important work was published. H. M. E. Miedema & Vos (1998) summarized all the pros and cons of Schultz's approach. Then, based on data used by Schultz (1978) and Fidell (1991) they created three new dose-response curves (%HA related to L_{DN}) – for aircraft, road traffic and railway noise separately. They proved that while equal values of sound levels are presented, the most annoying source is aircraft, followed by road traffic and least annoying railway. They summarized this finding in conclusions section: *"To treat different transportation sources equally with respect to the amount of noise annoyance tolerated, a noise limit in terms of DNL at the most exposed facade must be lower for aircraft noise than for road traffic noise, and the limit for road traffic must be lower than for railway noise. Which DNL values correspond to an equal %HA can be read from the curves presented in this paper"*.

Two years later H. M. Miedema & Oudshoorn, (2001) published more complex research, which presented relations between %LA (low-annoyed), %A (annoyed) and %HA related not only to L_{DN} but also L_{DEN}. They established three cut-off points for these three 'levels of annoyance':

- 28/100 points in a noise annoyance scale for people low-annoyed (%LA)
- 50/100 points in a noise annoyance scale for people annoyed (%A)
- 72/100 points in a noise annoyance scale for people highly-annoyed (%HA)

Miedema and Oudshoorn explained step by step their way of thinking and reasons for using some mathematical approaches. Main problems they had to face were different ranges of annoyance scales and two different noise metrics – i.e. L_{DEN} and L_{DN}. Authors chose to use only L_{DEN} noise metric (values of L_{DN} were transformed into L_{DEN} values using some equations).

In the same year, the special group from ICBEN published finally their findings concerning the problem of using a standardized noise annoyance scale (Fields et al., 2001). Six goals for a new scale were established, among which this scale should *"permit valid international comparisons of survey results within and between languages"* and *"yield an interval-level measurement scale (i.e., the response scale answers are equally spaced) meeting the assumptions for regression and many other analysis techniques"*.

ICBEN members carefully analyzed many different types of scales, including bipolar with or without a middle alternative. They also studied problems of formulating questions, even order of words in it. Finally, many words describing the degree of annoyance were discussed to form a discrete scale with equally-spaced points.

Two questions with slightly different construction (one semantic, with 5 points, and one numerical, with 11 points) were proposed. The semantic question should be asked in the following way:

"Thinking about the last (…12 months or so...), when you are here at home, how much does noise from (…noise source...) bother, disturb, or annoy you; Extremely, Very, Moderately, Slightly or Not at all?"

The second, numerical question, was formulated as follows:

"Next is a zero to ten opinion scale for how much (…source...) noise bothers, disturbs or annoys you when you are here at home. If you are not at all annoyed choose zero, if you are extremely annoyed choose ten, if you are somewhere in between choose a number between zero and ten. Thinking about the last (...12 months or so...), what number from zero to ten best shows how much you are bothered, disturbed, or annoyed by (...source...) noise?"

In addition, authors provided some instructions on how to construct a survey:

- *"Ask all respondents both questions"*
- *"Present the full scale, exactly as worded, to all respondents"*
- *"Place the questions early in the questionnaire, unless this conflicts with other survey objectives"*
- *"If pretests indicate that the questions are perceived as repetitious, include appropriate instructions"*
- *"Prepare written instructions for interviewers"*

Originally, questions and answers were formulated by authors in the following languages: English, Dutch (Flemish), French, German, Hungarian, Japanese, Norwegian, Spanish and Turkish. In 2003, the Polish version of standardized noise reaction questions for community noise surveys was published (Preis et al., 2003). This version of ICBEN questions regarding noise annoyance will be used in this research.

Thanks to these two important findings – three different dose-response curves and standardized scales for assessment of noise annoyance – research about noise annoyance became more consistent and easier to compare between different locations. Also, as a need to establish some law regulations regarding noise annoyance and sound levels spread around the world, the European Union proposed a special directive (European Union, 2002). This directive, called Environmental Noise Directive (END), proposed not only limits of sound levels for different sound sources but also requested country members to monitor noise in big cities and alongside busy transportation infrastructure and also to protect silent areas in the cities (called 'quiet urban areas'). From that moment, EU started policy to limit noise affecting people living in EU countries.

Having the same tools, scientists started to conduct more research, sometimes focused not only on one place, but also different cities or even different countries (Babisch et al., 2009b; Sato et al., 2002). Nowadays we can say that the problem of noise annoyance induced by a single noise source is well-described with many papers about different types of sources:

- **cars/road traffic** (Brown et al., 2015; de Kluizenaar et al., 2011, 2013; Héritier et al., 2014; Okokon et al., 2015a; Sato et al., 2002; Shimoyama et al., 2014; Sung et al., 2017; Torija and Flindell, 2014)

- **aircraft** (Babisch et al., 2009a; Bartels et al., 2013, 2015a; Brink et al., 2008; Gille et al., 2017a; Guoqing et al., 2012; Janssen et al., 2011b; Kroesen et al., 2013; Kroesen and Schreckenberg, 2011; Nguyen et al., 2013)
- **railway** (De Coensel et al., 2007; Di et al., 2014; Gidlöf-Gunnarsson et al., 2012; Kasess et al., 2013; Sato et al., 2004; Trombetta Zannin and Bunn, 2014; Yano et al., 2005; Yokoshima et al., 2008)
- **industrial sound sources** (Alayrac et al., 2010; Axelsson et al., 2013; Morel et al., 2012)

In the early 2000s another type of noise source appeared and attracted scientists – a wind turbine. Big wind turbines located near houses generate a strange, modulated sound which is difficult to adapt to and very annoying. That is why many papers so far try to estimate a depth of modulation and nature of sound propagation from turbines' hubs (Pedersen and Persson Waye, 2004; Bockstael et al., 2012; Ioannidou et al., 2016; Michaud et al., 2016; Seong et al., 2013; Van Renterghem et al., 2013).

More detailed analyses also led to the discovery of two new important tendencies which then became widely accepted and described. First of all, many scientists indicated that a number of flights rapidly grew. The more aircraft in the air, the more noise can be measured around airports. That problem was shown in several papers (Babisch et al., 2009a; Brink et al., 2008; Janssen et al., 2011b) and it was said that original dose-response curves could underestimate the impact of aircraft noise on people's annoyance assessment.

Secondly, scientists have turned their eyes on the problem of high-speed trains (De Coensel et al., 2007; Di et al., 2014; Yokoshima et al., 2008). It was shown that, e.g., Shinkansen trains are more annoying than 'traditional' ones. Some researchers also mentioned the significant role of high vibrations during pass-bys (Gidlöf-Gunnarsson et al., 2012; Sato et al., 2004; Yano et al., 2005).

These two problems finally led to a revision of already existing dose-response curves – those proposed by H. M. Miedema & Oudshoorn, 2001. Gille, Marquis-Favre, & Morel (2016) analyzed data using the original approach from 2001, showing that for aircraft and railway noise original dose-response curves underestimated the number of people who were highly annoyed: "*The results showed that Miedema and Oudshoorn's exposure-response relationships (Miedema and Oudshoorn, 2001) enabled to partially predict the annoyance due to road traffic noise, whereas they underestimated the annoyance due to railway and aircraft noises. New exposure-response relationships were therefore computed from these survey data and by following the whole procedure suggested by Miedema and Oudshoorn (2001). These new exposure-response relationships enabled to improve the calculation of the annoyance due to the different transportation noises*".

However, not only annoyance assessment itself but also other effects (e.g. on health) or different aspects were taken into consideration in research about noise annoyance. Below we provide several sections which enter deeper into certain branches of the whole research.

1.4 Towards understanding of noise annoyance. Different research approaches

1.4.1 Investigating new sources of noise

Analyzing noise sources does not have to be limited to a finite, well-known number of them. In some cases, more detailed analyses can show some interesting findings, unseen at the first glimpse.

One of the most interesting noise sources which are, somehow, similar to trains, are trams. It is difficult to find in literature papers focusing only on them. Tramway noise is limited to cities which have some trams network, however, squeals produced by them (especially on curves) can annoy people. Sometimes trams are analyzed as a single noise source (Pallas et al., 2009; Panulinová, 2017; Trollé et al., 2014), sometimes together with other noise sources, e.g. buses – as part of a public transportation system (Sandrock et al., 2008). Many interesting findings about trams, when more than one noise source is present, were made by Catherine Marquis-Favre's team from Lyon – I will refer to it later, in a section about total annoyance.

The team from Lyon also analyzed powered-two-wheelers (Gille et al., 2016a) revealing that their annoyance is high and can be influenced by the structure of time spaces between single pass-bys. Motorbikes and scooters were also analyzed by (Paviotti and Vogiatzis, 2012).

There was also some research concerning electromobility, i.e. changes in road traffic noise while comparing 'traditional' cars with electric ones. Campello-Vicente, Peral-Orts, Campillo-Davo, & Velasco-Sanchez, (2017) showed that the difference is not large and it does not exceed the maximum of 2dB in favor of electric cars.

1.4.2 Effects of noise on health

One of the most important effects of noise is its influence on humans' life and health. There are many findings which confirm that noise has an impact on people's health – however, the effects are not immediate. Among different drawbacks and diseases, most common ones are:

- sleep disturbance, i.e. awakening as a consequence of noise (Elmenhorst et al., 2014; World Health Organization, 2018)
- *"fatigue, changes in mood, impairment of performance"* (Heimann et al., 2007)

- depression, anxiety, augmenting stress (Beutel et al., 2016; World Health Organization, 2011)
- ischaemic heart disease (IHD), cognitive impairment in children, tinnitus and annoyance (Dzhambov et al., 2016; World Health Organization, 2018)

There was also research considering a link between noise and general mental health (Dzhambov et al., 2017).

Noise effects on health are also described by health and environment organizations, such as World Health Organization (WHO) or European Environment Agency (EEA). WHO measures an impact of noise by terms of disability-adjusted life-years (DALYs). According to WHO report from 2011, every year more than one million healthy years are lost due to noise and its influence on people. Five main contributors to this were mentioned in (Basner et al., 2014) and are presented in Figure 3.

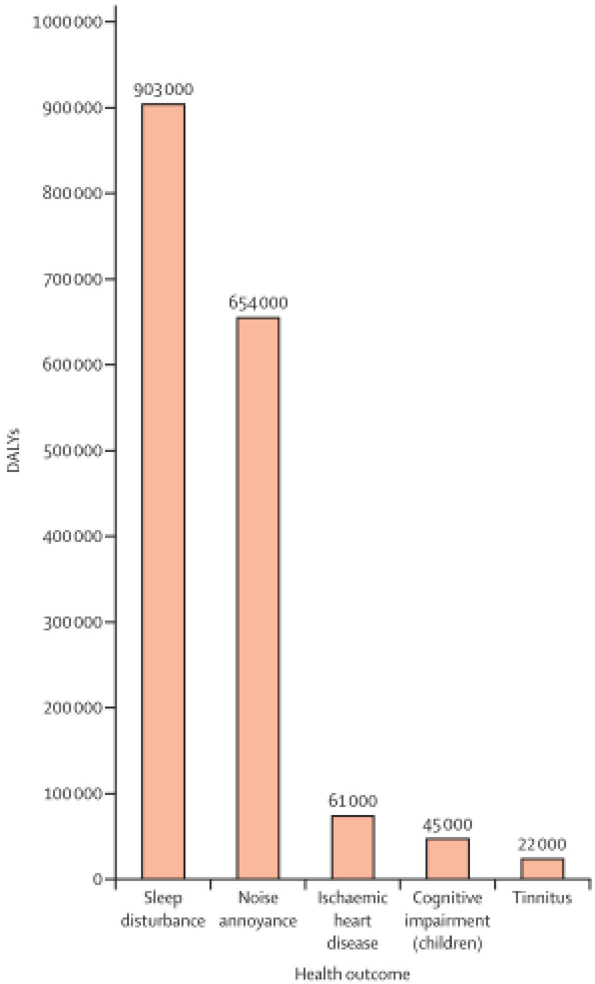


Figure 3. DALYs and five most influent factors regarding noise and its health effects. Taken from (Basner et al., 2014).

Moreover, EEA in its briefing from the beginning of 2017 shows that *"at least 100 million people are exposed to levels of traffic noise that exceed the European Union's (EU) indicator of* *noise annoyance"*. The most dominant noise source is road traffic – all of these sources are shown in Figure 4.

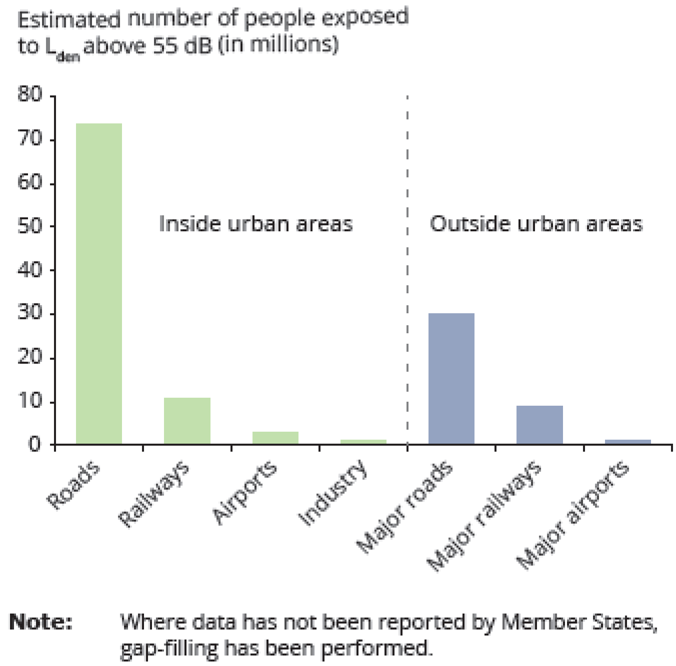


Figure 4. The number of people exposed to different noise sources in the European Union. Source: EEA report from 2011.

All these numbers are available thanks to END directive which imposes on every EU country an obligation to monitor noise in big cities (above 100 000 inhabitants) and alongside busy transportation trails (routes and railways). END also requires to present results of monitoring not only to EU organs but also to make them public and accessible to every citizen.

END also points out that in every noisy area (where noise limits are exceeded) there should be a noise action plan (NAP) implemented – i.e. all necessary actions to limit the noise level to acceptable values. However, nowadays many scientists show that procedures of establishing NAPs are not consistent; also a comparison of noise maps from different cities and countries is difficult (D'Alessandro and Schiavoni, 2015; King and Murphy, 2016).

1.4.3 Noise mapping in the real environment. Soundscape analysis using modern technologies

Measuring noise in a long-term (yearly) approach is not trivial. The most common solution is to monitor noise in dozens of points within a city/alongside a road and then, using some models, extrapolate data to the whole area. These models are different from country to country and there is no agreement if their prediction is reliable (Caschili et al., 2014). That is why European Union proposed to every EU member usage of a standardized procedure (*"common noise assessment methods"*) which is called CNOSSOS-EU and was proposed in a special EU Directive (European Comission, 2015). From 31.12.2018 this method is obligatory for every EU country.

Nevertheless, scientists propose some other solutions to resolve the problem of noise maps' verification and establishing NAPs. Most of them are focused on how to measure people's exposure to multiple noise sources (Licitra et al., 2011) and how methodologically the whole procedure could be improved (Borchi et al., 2016; Felcyn et al., 2018; Martín et al., 2011). Some researchers try to improve noise maps by updating them in real-time (Cai et al., 2017; Wei et al., 2016) while the others see the opportunity in wide usage of smartphones. Several articles propose to use smartphone-based applications to assess acoustical environment (Aspuru et al., 2016; Guillaume et al., 2016; Herranz-Pascual et al., 2016; Murphy and King, 2016; Zuo et al., 2016) or, generally speaking, encourage citizens to take part actively in a process of assessing and protecting acoustics of local area (Vogiatzis and Remy, 2017).

In the last years, not only a sound itself but also its context attracts more attention. This environmental point of view underlines the complexity of human perception and put stress on a term of 'soundscape' (Gidlöf-Gunnarsson and Öhrström, 2007; Preis et al., 2015a; Raimbault, 2006). This term, similar to 'landscape', aims to describe all sounds present in a given area. Moreover, those sounds are not perceived only as waste (like noises) but also as resources which can be enhanced to improve the acoustical experience. 'Soundscape' is also related to, let's say, the complex design of a given place, including, e.g., fountains (Velardi et al., 2017). This way of complex thinking about public space can get more attention in the upcoming years.

1.4.4 Improvements in methodology and spatial audio techniques

The END directive and establishing ICBEN scales for noise annoyance assessment were just a starting point to more complex analyses. The main problem of the relation between noise indicators values (L_{DEN}, $L_{Aeq,T}$, L_{AE}) and people's answers about noise annoyance is the poor correlation between them. As it was shown (Marquis-Favre et al., 2005), not more than 30% of the variance in data could be explained by this relation. That is why many studies try to spread the problem to other acoustical and non-acoustical factors. On the other hand, some scientists propose new variables aimed to replace annoyance scales. One of them is GNR – general noise reaction (Kroesen and Schreckenberg, 2011). As the authors said, in GNR *"all negative feelings and emotions in response to (aircraft) noise are integrated"*. Another parameter is the CTL – community tolerance level (Gille et al., 2017a; Gjestland and Gelderblom, 2017; Schomer et al., 2012). Without analyzing details, CTL shows value in [dBA] (taken from L_{DEN}) which is the limit of noise still acceptable by a community.

Another tendency is to go further in objective parameters of sound and find new ones which could explain better people's noise assessments. This could be for example:

- **for road traffic:** pavement material or vehicle speed (Freitas et al., 2012), having a peaceful room with quiet façade (de Kluizenaar et al., 2013), noise spectrum (Torija and Flindell, 2014), type of a crossroad (Covaciu et al., 2015), different street geometry (Camusso and Pronello, 2016) or number and loudness of noise events/pass-bys (Gille et al., 2016a)
- **for aircraft:** number of complaints about noise (Fidell et al., 2012) or the number of flight operations (Gille et al., 2017a)
- **for railway:** type of brakes (Kasess et al., 2013) or vibrations during pass-bys (Gidlöf-Gunnarsson et al., 2012)

However, as was mentioned above, not only objective characteristics can explain people's judgments of noise. Many subjective ones, related to respondents' experience, feelings, fears and emotions can also influence their perception of noise. One of the most popular is noise sensitivity – i.e. individual 'endurance' of noise. People who are more sensitive tend to give higher values of annoyance ratings to noise. By adding this parameter to noise annoyance models, more variance can be explained than without it (Gille et al., 2016b, 2016c, 2017a). But it is not the only interesting subjective aspect of noise perception. Another one could be the noise perceived in a neighborhood (Méline et al., 2013) or individual fears about influence of noise on one's health (Okokon et al., 2015b; van den Berg et al., 2015). Many non-acoustical factors were mentioned in (Bartels et al., 2013, 2015b). All these findings support a theory that noise annoyance is a complex experience and could not be predicted by just one variable (commonly a sound level).

Not only parameters can change in noise annoyance models. The measuring procedure itself can also be interesting to investigate. Brink et al. (2016) analyzed the structure and order of questions in a questionnaire; he found that location of a question about noise annoyance can influence people's answers. On the other hand, Hermida Cadena, Lobo Soares, Pavón, & Coelho (2017) explained the differences between laboratory and in situ noise annoyance research. While presenting similar stimuli in both conditions, people tend to give higher rates in laboratory conditions.

As knowledge about noise annoyance grew, it became more and more clear that a simple dose-response curve is not the best model to predict noise annoyance. That is why some scientists tried to incorporate more sophisticated methods to this problem.

One of the first trials was to use fuzzy logic models, based on many predictors, not only acoustical ones (Botteldooren and Verkeyn, 2002). Fuzzy modeling was also used to predict aircraft noise annoyance based on certain noise indicators, such as L_{AeqD} or L_{AeqN} (Heleno and Slama, 2013) or to establish NAPs (Ruiz-Padillo et al., 2016).

Another way of thinking is to use multilinear regression in noise annoyance models – like in (Gille et al., 2016c, 2017b; Klein et al., 2017). In general more predictors explain more variance in data.

In noise annoyance research, particularly two ways of presenting sound to participants are used:

- the sound is presented via headphones
- the sound is played from loudspeakers, mainly in a stereo configuration. In this approach, some modifications are sometimes made, e.g. putting loudspeakers in the living room or hiding them behind windows.

However, a sound which is presented, is recorded with one or two microphones (in stereo configuration). This technique preserves information about the horizontal location of sound sources but does not let to do the same with vertical information. Is it possible to keep information about both directions?

The answer to this question is ambisony. It is a technique first described by Gerzon (1973) and aimed to preserve 'a sphere of sound' around the recording point. Because information about sound is spherical, it allows us to keep both horizontal and vertical coordinates of a source. Gerzon introduced an ambisony of the first order which means that information is coded using four channels – in so-called, B-Format. This can be achieved by using channels marked as W (omnidirectional), X, Y and Z (all three directional). They are characterized by the following formulas:

$$W = \frac{1}{k}\sum_{i=1}^{k} s_i \left[\frac{1}{\sqrt{2}}\right] \tag{10}$$

$$X = \frac{1}{k}\sum_{i=1}^{k} s_i [\cos\varphi_i \cos\theta_i] \tag{11}$$

$$Y = \frac{1}{k}\sum_{i=1}^{k} s_i [\sin\varphi_i \cos\theta_i] \tag{12}$$

$$Z = \frac{1}{k}\sum_{i=1}^{k} s_i [\sin\varphi_i], \tag{13}$$

where s_i denotes monophonic signals which should be coded in precise directions using angles ϕ_i (horizontal) and Θ_i (vertical). At this point should be mentioned that spherical coordinates used in Eqs. 10-13 agreed with the convention used in ambisony. It means that:

- coordinates are left-side oriented
- x-axis points out 0° for both azimuth and elevation
- values of angle φ increase counterclockwise towards positive part of y-axis
- values of elevation are positive for points located above XY plane (positive z-axis values)

Presented formulas show that the acoustical field can be synthesized by multiplication of original signals and functions towards specific directions. To do so, we have to use microphones with specific directional characteristics which are shown in Figure 5.

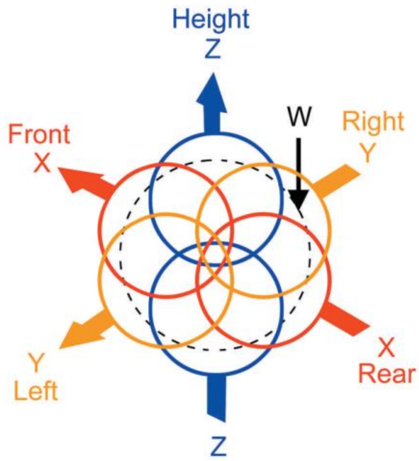


Figure 5. Directional characteristics of a four-capsule microphone aimed to record ambisonic data in B-Format (taken from a manual of Soundfield ST450).

From Fig. 5 one could see that needed data can be recorded using one omnidirectional microphone (W signal, a black dotted line) and three bi-directional/figure of 8 microphones (X, Y, Z signals, red, orange and blue solid circles respectively). As this configuration allows to record all information about sound sources' localization, we can use B-format recordings to decode sound for any number of loudspeakers (but equal or higher than 4). More loudspeakers, better localization of sound. The decoding process can be achieved in two possible ways: projection or quasi-inversion.

1.4.4.1 Projection

A technique of projection means a simple weighted sum of all B-Format canals. This sum is also weighted by a function describing the position of a loudspeaker. Thus, the signal reproduced by a given loudspeaker would be computed using an equation provided below:

$$p_j = \frac{1}{N}\left[W\left(\frac{1}{\sqrt{2}}\right) + X\left(\cos\varphi_j \cos\Theta_j\right) + Y\left(\sin\varphi_j \cos\Theta_j\right) + Z\left(\sin\varphi_j\right)\right], \tag{14}$$

where angles ϕ_j and Θ_j describes a position of j loudspeaker; N is the number of loudspeakers in a whole array. This approach requires a regular structure of an array; for three dimensions the most common are cubic or tetrahedron.

1.4.4.2 Quasi-inversion

This is a matrix version of a decoding process. Assuming that we have a vector of B-Format canals: $B = [W\ X\ Y\ Z]^T$, column-wise vector of all loudspeakers signals denoted as p and matrix of recoding C, elements of the latter will contain values of a function describing the position of a given loudspeaker. The C matrix will have N rows and L columns. In this case, a decoding function could be described as:

$$B = C * p, \tag{15}$$

where the vector of signals, p for loudspeakers is:

$$p = C^{-1} * B, \tag{16}$$

C^{-1} is the reverse version of a matrix C. C is reversible only if is quadratic and this can be achieved when $L = N$ – meaning that the loudspeakers array is regular. If $L \neq N$, a quasi-inversion can be applied, but its more detailed description is not a part of this work.

For an ideally regular array of loudspeakers, projection and inversion give the same results. More information can be found in (Hollerweger, 2008).

Gerzon's research was somehow too innovative for his times and ambisony was not popular. However, it gets more attention nowadays as it seems to fit perfectly into the need to reproduce 3D sound in augmented or virtual reality. Ambisony has also beed used in several works on environmental noise. Alvarsson, Nordström, Lundén, & Nilsson (2014) investigated speech intelligibility in an outdoor living space. They reproduced sound using 8 loudspeakers in a cubic configuration. On the other hand, an ambisonic configuration was used in experiments about audio-visual interactions conducted in Institute of Acoustics at Adam Mickiewicz University in Poznan (Preis et al., 2015b, 2016; Szychowska et al., 2018; Wojaczek, 2016). Nevertheless, to our knowledge, there are no papers concerning the problem of total annoyance in the ambisonic environment, this thesis seems to be the first.

1.5 Total Annoyance – annoyance from multiple noise sources. Review of existing total annoyance models

Road traffic, aircraft, trains, industries or wind turbines – all of these sound sources were investigated in last years. However, in common life, especially in large cities, people are exposed to more than one noise source at the same time. In this case, we say that ‘mixed’ or ‘combined’ noise sources are present. In this work, we will use the term ‘mixed noise sources’.

‘Mixed’ simply means that not only one noise source can be heard, distinguished and named separately by the listener. All of these sources have to be loud enough to be perceived easily in the context of common background noise. When talking about annoyance, more detailed terms has to be provided to avoid ambiguity (Klein, 2015; Miedema, 2004):

- **total annoyance:** sometimes named ‘global annoyance’ – the annoyance evoked by all co-existing noise sources
- **specific annoyance:** it is annoyance only from one noise source, presented separately (solo). It means, that in every case when we investigate only one noise source we actually measure specific annoyance of it
- **partial annoyance:** the term was first introduced by Berglund & Nilsson (1997); it describes annoyance from one noise source which is presented with other noise sources at the same time. It means that every time when one presents many sources to a listener but asks only about annoyance from one of them – asks of partial annoyance

Measuring specific and partial annoyance is sometimes difficult, particularly when we do survey research *in situ*. In this case, it is rather impossible to ask people about specific annoyance, so only partial annoyance can be estimated.

If we want to understand how total annoyance can be measured and predicted, first we should pose a question: how people perceive combined noise sources? Because the problem is related to humans’ perception, the answer is not simple and different mechanisms can be revealed.

The most obvious one seems to be a sort of summation of partial annoyances. This effect, when total annoyance is greater than any partial annoyance, is called ‘synergy’ or ‘synergistic effect’. However, it does not mean that total annoyance has to be a simple sum of partial annoyances.

A synergistic effect was described, e.g., by (Öhrström et al., 2007). In this work combined railway and road traffic noise was assessed in *in situ* studies. Interestingly enough, when only one noise source was loud (comparing to another one), this source determined a total annoyance rate. However, when both sources were equally loud with high sound levels, total annoyance was greater than their partial annoyances: “*as the total traffic sound exposure increased, the prevalence of total annoyance gradually became higher for dwellings where railway and road traffic contributed equally. Thus, there was an interaction between total traffic sound exposure and type of dwelling and this interaction was statistically significant (p=0.003)*”. This interaction can be also observed in Figure 6.

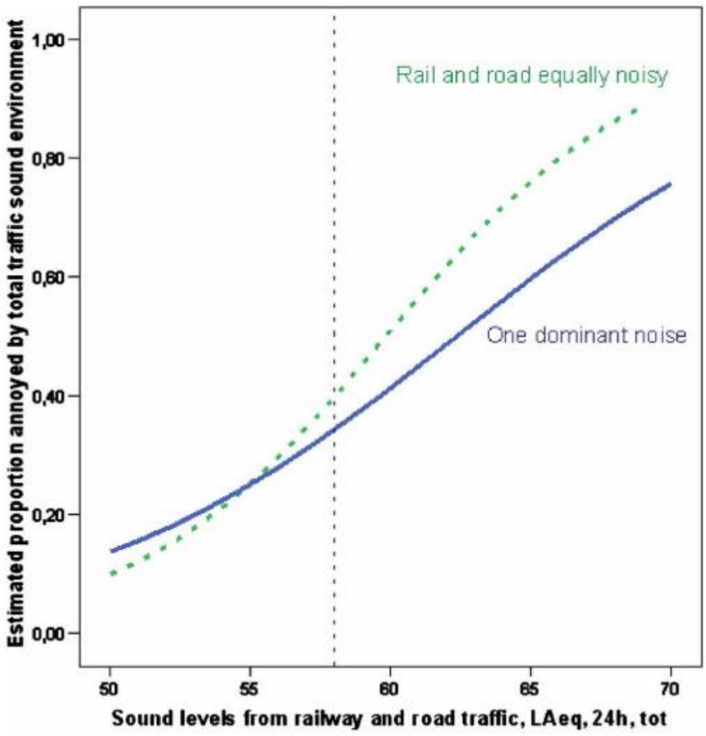


Figure 6. A synergistic effect which occurs when more than one noise source are presented simultaneously (Öhrström et al., 2007).

This interaction, mentioned by Öhrström et al., can be seen as a synergistic effect – because the total annoyance rate is higher than any partial annoyance.

Öhrström et al. (2007) also observed another interesting effect lying under the estimation of total annoyance. When one noise source is louder enough than the other one, the annoyance of the first one (partial or specific) determines also total annoyance (they are equal). This mechanism is called 'dominant' or 'strongest component' model.

A strongest component model could be found in many papers, however its reliability is ambiguous. There are cases when it functions well (Botteldooren and Verkeyn, 2002; Nguyen et al., 2012; Ota et al., 2008; Pierrette et al., 2012) but there are also situations when it does not work or works poorly (Marquis-Favre and Morel, 2015; Morel et al., 2012; Morel and Marquis-Favre, 2013). Such dichotomy makes this model difficult to widely use, nevertheless it is frequently investigated as its implementation is trivial.

On the other hand, there are some situations when annoyance of a given noise source is reduced while presenting the other source. For example, road traffic noise presented separately on a specific sound level could be annoying but when another noise source is added (like aircraft noise) the same road traffic noise would be rated as less annoying. This mechanism is known as 'inhibition'. This phenomenon was primarily investigated by Powell and his colleagues (Powell, 1979; Powell and Rice, 1975) who showed that aircraft noise has smaller annoyance ratings while presented with loud road traffic noise. Powell assumed that the main reason for this phenomenon is the fact that the latter noise masked the former one. Berglund & Nilsson(1998) concluded the same in their review-type paper aimed to describe all already-existing total annoyance models. This spectral masking effect led Morel et al. (2012) to propose a new measure, called 'power of inhibition': this measure explains how likely is for one noise source to dominate over the other – i.e. what is the probability to become 'strongest' and be the dominant part in the strongest component model. However, existence of inhibition is not so obvious – it was again… Morel (with Marquis-Favre, 2013) who showed that inhibition does not always work.

But what about the situation when total annoyance is smaller than the maximum of partial annoyances? This effect is called 'combined noise sources paradox' and its mechanism is not clear for researchers, particularly because we still do not know how humans' brain integrate information about various noise sources. However, this effect definitely exists and sometimes is explained by the position of the noise annoyance question in the questionnaire or understanding of it (Hatfield et al., 2006). Cognitive sciences can provide some help in this area. In many papers it was proved that natural sounds (i.e. birds, waterfalls/fountains, sea waves etc.) are mostly calm and decreases overall annoyance of the whole environment (A. L. Brown, 2010; Preis, Hafke-Dys, Szychowska, Kocinski, & Felcyn, 2016; Szychowska et al., 2018; Velardi et al., 2017). This can lead us to the conclusion, that every time when some natural sounds are present in the environment, a total annoyance assessment can decrease, even if the sound levels of noises are high.

Powell (1979) was interested in annoyance from road traffic and aircraft noises. Based on his findings, he proposed a model which is a simple implementation of an inhibition effect:

$$\Psi_{ij} = \Psi_i' + \Psi_j' \tag{17}$$

Ψ_{ij} is the total annoyance of sources i and j while presented simultaneously. Ψ_i and Ψ_j are the masked annoyance of both sources while they are combined. Inhibition effect was also found by (Izumi, 1988); *"road traffic annoyance is inhibited by train annoyance when the train noise level is higher".* As was mentioned above, Morel et al., (2012) improved somehow this model by establishing the 'power of inhibition'.

Birgitta Berglund, Berglund, Goldstein, & Lindvall, (1981) analyzed three different noise sources: road traffic, jackhammer and pile driver in pairwise comparisons. They proposed to use a 'vector summation model' which had been previously used (with success) to the perception of odors or brightness. The formula of this model is:

$$\Psi_{ij} = \sqrt{\Psi_i^2 + \Psi_j^2 + 2\Psi_i\Psi_j cos\alpha_{ij}}, \tag{18}$$

where indices *i* and *j* represent different noise sources, α_{ij} is a constant. In this research the most variance of data was explained while α was equal to 90°. The same model was used again in a study with aircraft, railway and highway traffic (Berglund and Nilsson, 1998a). On the other hand, Botteldooren & Verkeyn, (2002) pointed out some problems with it, particularly while α>90°. Another limitation is that this model works only for two noise sources combination.

Models based on the summation of energies of different noise sources create separate class. The simplest version of them states that annoyance is a function of the total energy emitted by all sources and summed up:

$$A = f(L_T), \quad L_T = 10\log\sum_{i=1}^{n} 10^{0.1L_i} \tag{19}$$

This model, as well as its more detailed versions, were described by Taylor, (1982). Nevertheless, one can see that this formula simply sums up energies without weighting coefficients – but we know that, for example, railway noise should be less annoying than aircraft while both presented at the same level. Thus, energy summation models are not perceived as reliable what was also proved by H. M. E. Miedema, (2004). Morel et al. (2012) showed that not only energy summation model but also energy difference model provides a poor fit to the data.

The two most commonly used total annoyance models are the strongest component model (based on the effect mentioned above) and weighted summation model.

The strongest component model simply states that a total annoyance is equal to the maximum value of partial/specific annoyances. In (Klein, 2015) a formula which describes it is as follows:

$$A_T = \max_{i=1,n}(A_i) \tag{20}$$

This model works fine for a wide range of situations (Botteldooren and Verkeyn, 2002; Nguyen et al., 2012; Ota et al., 2008; Pierrette et al., 2012), however is limited mainly to situations when different noise sources are clearly separable and the differences in their sound levels are relatively large. In other situations, this model cannot function well, as (Klein, 2015) pointed out: *"the formulation of the strongest component model cannot anticipate synergistic effects or the combined noise sources paradox"*.

The second widely-used model is a weighted-summation model. The procedure for using it can be divided into four steps:

- first, the reference source has to be chosen. A reference source is the source to which all other noise sources will be, let's say, transformed
- secondly, we measure partial annoyances of each noise source (including the reference one)
- then, knowing them, we translate those noise sources to such sound level of a reference source which gives equal annoyance to the measured one for a given noise source

- finally, we just sum up all sound levels to get the final total sound level

That model was introduced by (Vos, 1992). The most common reference source is road traffic – as it is the most frequent noise which people in EU are exposed to. A weighted summation model was found to be a good predictor of total annoyance as it also provides a method to direct comparison between different noise sources (while they are all translated into reference source's sound levels). In his work, Vos used three different sources: road traffic, gunfire and aircraft. While keeping one of the noise sources at the same value of sound level and manipulating levels of the second one, it was possible to present curves representing total annoyance changes with an increasing sound level of the second source. An example of such relations is presented in Figure 7.

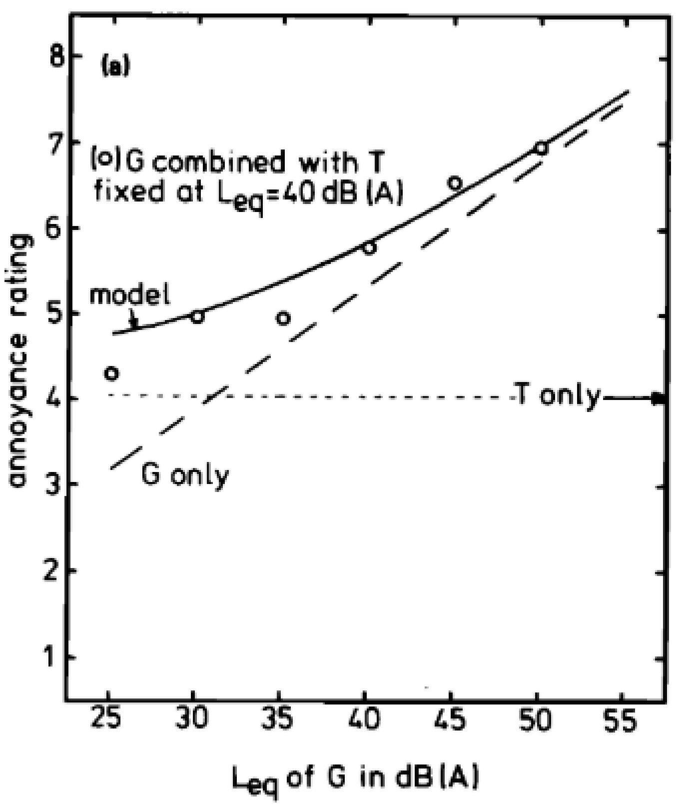


Figure 7. Dose-response curve for combined noises from gunfire (G) and road traffic (T) with the sound level of road traffic fixed at 40 dB (Vos, 1992).

A weighted summation model gave high goodness of fit in some papers (Miedema, 2004; Morel and Marquis-Favre, 2013). It is also worth to mention that curves from Fig. 7 clearly show that when one of the noise sources is much louder than the other one, tendency becomes more like in the strongest component model. Thus, it depicts another popular mechanism of integrating several annoyance ratings. This model bases only on sound levels but we know that the variance explained by the relation between them and annoyance (or %HA) is not high (not more than ~30%).

Of course, there are also some variants of above mentioned models. Those who are interested in it can analyze review-type works like (Janssen et al., 2011a; Laszlo et al., 2012; Marquis-Favre et al., 2005).

All these findings and ambiguous results of experiments show that there is a need for a better understanding of human assessment of combined noise sources and creation of a model which will be stable and reliable in a wide variety of circumstances. This problem was one of the main goals of this thesis.

1.6 Research hypotheses

All research hypotheses presented here will be tested based on results of two experiments. Experiment I will concern road traffic only, Experiment II will include mixed noise sources: road traffic, trams and aircraft.

The main goal of this thesis is to create a new TA model based on most common noise sources which can be met in urban conditions – road traffic, trams and aircraft.

In the previous sections we have shown that only 30% of noise annoyance ratings variance can be explained by noise indicators based on sound level values. We have also mentioned other predictors which can be used, both objective (psychoacoustical features such as roughness, sharpness or fluctuation strength) as well as subjective (individual factors related to noise sensitivity, anxiety or attitudes towards various noise sources). Already existing TA models were presented and described, showing that their results are ambiguous in various ways.

Since the single predictor based on sound level does not explain even half of the variance in noise annoyance assessments, we state that the model could be better with more objective factors incorporated to it i.e. psychoacoustical values described by Fastl and Zwicker (2007). Moreover, the time structure of road traffic may be an important factor – as already shown by Kaczmarek and Preis (2010). They presented four different noise scenarios – characterized by the same $L_{Aeq,10m}$ values, but with different time patterns. In the first scenario, road traffic noise was always presented while in the others, there were some silent periods between pas-bys, when only a background noise was heard. To keep the same $L_{Aeq,10m}$ values, fewer pass-bys were presented, louder they had to be. This different time pattern was found to be a statistically significant factor influencing annoyance assessments. In addition, some relations between psychoacoustical characteristics of recordings (like roughness) and annoyance ratings were found.

However, Kaczmarek et al. presented to people artificially created noise scenarios. Although the pass-bys were recorded in the field, they were then manually put together to create different time patterns. Thus, noise scenarios did not reflect real situations from the environment. Moreover, only one equivalent sound level was observed for all scenarios, equal to 55 dBA. In our opinion it is needed to check if the same tendencies would be observed while presenting the real traffic flow to people with different $L_{Aeq,T}$ values.

Since there is no strong evidence that the psychoacoustical characteristics or time pattern affect the assessment of noise annoyance (the results are inconclusive as shown in section 1.2.2), we wanted to analyze some trends in an initial study which would only consider road traffic noise. It is the most common noise source, so it is a natural choice with only one type of noise in mind. Additionally, if certain relationships could be revealed, their existence would also have an impact on the ratings of the TA.

In this way two hypotheses will be tested in Experiment I:

1) **Road traffic noise annoyance ratings relate to psychoacoustical characteristics of stimuli.** Some years ago, Fastl and Zwicker proposed a parameter called 'Psychoacoustical Annoyance' (PAnn). It bases on other psychoacoustical characteristics, like roughness, sharpness, fluctuation strength and loudness. As PAnn is aimed to be an approximation of a 'real', subjective annoyance, also some correlation between noise annoyance ratings and psychoacoustical values should be observed, especially when stimuli are presented at different sound levels.
2) **For the same LAeq,T, different noise annoyance ratings are observed regarding time patterns and proportions of presented road traffic packages.** The same mechanism of creating silent gaps between road traffic pass-bys will be used here (comparing it to the research of Kaczmarek et al.). There will be different proportions between busy road traffic and periods of silence, and the factor associated with it will be called 'proportion'. On the other hand, if we always preserve the same sound level of each pass-by, the overall sound level will be different – because it will depend on different lengths of silent periods. This factor will be called 'time pattern'. Both these factors should have an impact on the human assessment of the road traffic noise annoyance.

Of the existing TA models, three seem to be the most commonly used. There are two models based on perceptual phenomena, i.e. strongest component model (SCM) or combined noise source paradox (CNSP), however their performances are ambiguous (what was described in the section 1.5). The most popular sound level-based model is the weighted summation model (Vos, 1992). However, it cannot be used in this thesis as the limited experimental procedure requires that the same sound level of road traffic noise be maintained in scenarios of mixed noise sources (see the section 2.2.1.2 for more details). Thus, only SCM and CNSP models will be investigated in this thesis.

However, the most important aspect of this work is to develop a new TA model. As already mentioned, noise annoyance can be related to the objective characteristics of sound sources, PA ratings of each sound source presented in a mix or to subjective characteristics of individuals. Thus, we propose to construct three different TA models, each related to different class of factors.

We state that the best goodness of fit should be obtained for a model using PA ratings of different noise sources. This model will be named as **PABaM** – Partial Annoyance Based Model. As performances of both SCM and CNSP are sometimes poor, we want to establish a new model which would give more explained variance and work well regardless of both SCM and CNSP. The research hypothesis related to this problem is:

3) **TA ratings could be explained by PA ratings of different noise sources.**

PABaM will take into account only these noise sources which can be distinguished by listeners. It means that if one noise source is not identified, it will not have a statistically significant influence impact on the whole model. It was already shown by Klein et al., (2017) that trams do not contribute more than 9% into the TA assessments variance. On the other hand, according to WHO or EEA, road traffic seems to be the most common noise source in Europe. We assume that, comparing to road traffic or aircraft, trams will not influence people's judgements of TA. The next research hypotheses is:

4) **Trams' PA do not contribute to the TA ratings.**

Unlike trams, many papers mention that of the three most common transportation noise sources, i.e. road traffic, trains and airplanes, the latter are the most annoying. Moreover, both trains and cars are on-the-ground vehicles while airplanes fly over our heads. Both these observations suggest that adding aircraft to mixed noise sources scenarios would increase TA ratings given by respondents. The fifth research hypothesis is formulated as:

5) **TA ratings increase when aircraft noise is presented in mixed noise sources scenarios.**

On the other hand, since the results published in many papers suggest the importance of subjective characteristics, we also want to establish a model based solely on their values. We will refer to it using a name **PeCBaM** – People's Characteristics Based Model. PeCBaM will be based on subjective characteristics which will be provided by respondents in a preliminary survey. It will cover the problem of noise sensitivity, fears and attitudes towards different noise sources and their effects. Taking all these factors into account, we believe that the performance of such a model could be better than a model based on objective characteristics of stimuli. Thus, the next research hypothesis is:

6) **Non-acoustical factors (such as noise sensitivity) are better predictors of TA than objective characteristics of stimuli.**

Finally, the model based on objective characteristics (**ReCBaM**, Recordings Characteristics Based Model) will be also computed. Nevertheless, we conclude that its performance will be poor, especially considering that the differences in sound levels between various mixed noise sources are minimal (mixed noise scenarios are dominated by road traffic noise). The last research hypothesis is:

7) **From all three new TA models, the best goodness of fit should be observed for PABaM, followed by PeCBaM and the poorest performance would be observed for ReCBaM.**

Models will be established using a multilinear regression equations, computed using a training subsets and checked using the rest of the data.

2 Method

2.1 Recordings

In this thesis we wanted to simulate the actual traffic flow in Poznań. Poznań has about 500,000 inhabitants and is a capital of the Greater Poland voivodeship. It has tramway network and an airport Ławica (EPPO). Thus, three main transportation noise sources can be distinguished:

- Aircraft; the airport is located in the western side of a city, airplanes take off or land using a skyway over the city, from east to west.
- Tramways; the network is most developed in the center, however there are also lines to biggest estates. Important: there are 7 different types of trams in Poznań
- Road traffic (divided into light and heavy vehicles)

There is also a railway network, however, noise maps do not show any important influence of it on people. This was also proved in our publication where special noise surveys were carried out among Poznań inhabitants (Felcyn et al., 2018). That is why we did not incorporate trains into this study.

2.1.1 Apparatus

As it was mentioned above, noise sources were recorded using an ambisonic microphone. This type of microphone allows to keep all spatial information about a source's movement and its sound. A SoundField ST450 microphone was used. It contains four capsules, one omnidirectional and three eight-shaped. The overall view can be seen in Figure 8.

Figure 8. Construction of a four-capsule ambisonic microphone Soundfield ST450 (picture from the ST450 manual).

The microphone has a dedicated processor that allows to get a B-Format on the output (W, X, Y and Z channels). Every channel was recorded separately by connecting a processor with a Head Acoustics Squadriga II recorder (Figure 9).

Figure 9. Squadriga II recorder (picture from producent's resources).

The sound was processed with 16 bit resolution and 48 kHz of sampling frequency. The output files were in .hdf format and were then converted into .wav files using a dedicated extension from Head Acoustics. Finally, each recording was a 4-channel .wav file.

As we wanted to mix all noise sources in different proportions, every source had to be recorded separately. It was not trivial, especially in the case of trams, because in Poznań tram tracks are commonly located in the middle of the roads. That is why we carefully analyzed the topography and transportation network in a city to finally select four best places to record different noise sources.

2.1.2 Selection of places for recording different noise sources

2.1.2.1 Road traffic noise

The first concern was about number of traffic lanes. In Poznań there are both two and four-lane roads. We decided to record road traffic in both cases.

We wanted to record only road traffic, without additional noise from other sources. Ultimately, two locations were selected:

- For two-lane road we chose Droga Dębińska. It is a straight street, close to allotments. No tram tracks nearby, it is also aside from Ławica's skyway. Traffic is flowing because there are no pedestrian crossings. But there were also two obvious drawbacks of this location. First, a narrow road, dedicated to access allotments, which was sometimes used by drivers. They had to slow down and change a direction of movement, so every time this situation occurred, the procedure was stopped. Secondly, there is a bus stop just few dozen meters away. The bus slowing down and accelerating also made us to stop recordings. However, both these situations did not occur often. A map of the location can be seen in Figure 10.

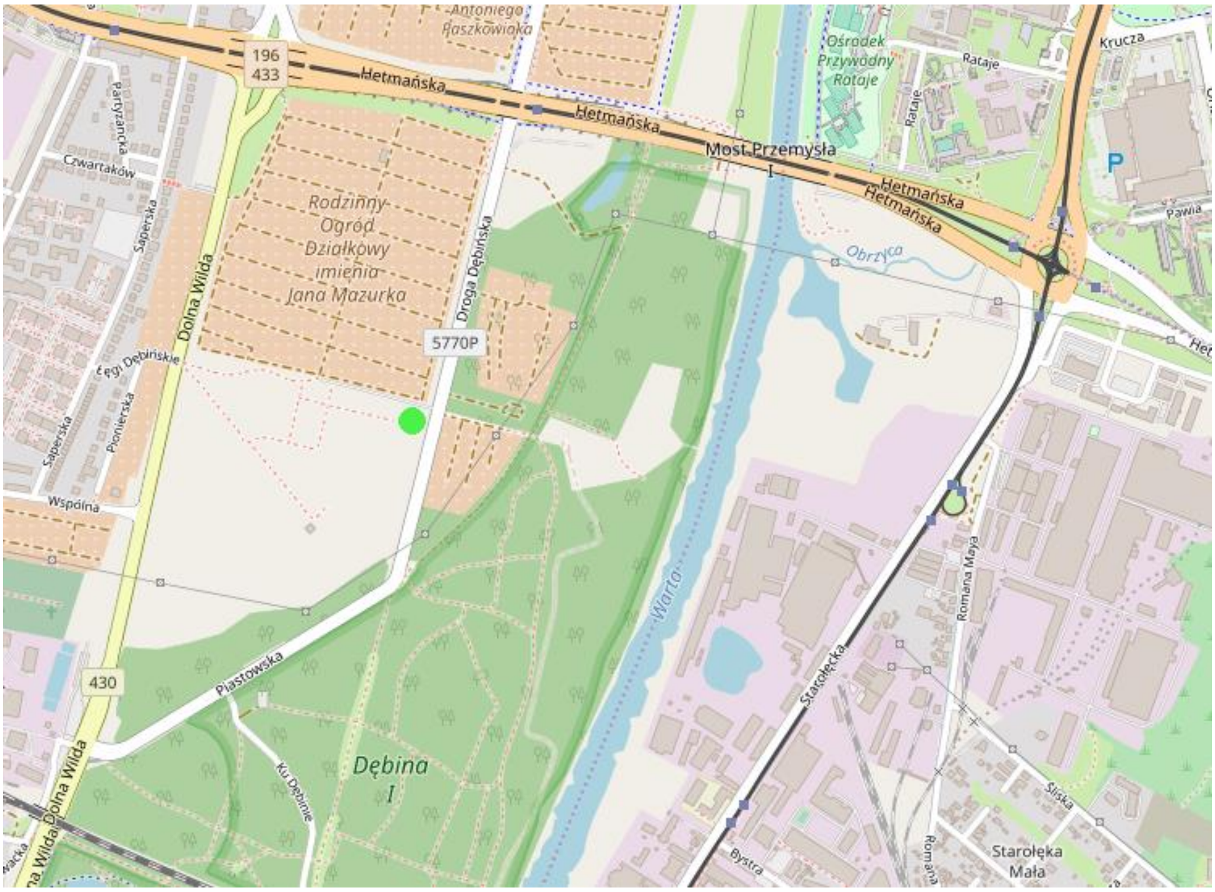


Figure 10. The location of a measuring point of two-lane road with road traffic only. Note: all maps presented in this thesis are taken from open source OpenStreetMap website.

- For four-lane road a location close to Malta Lake was chosen. In the southern side of it there is a Baraniaka street – busy road without trams and far away from railways. However, there is a skyway to Ławica airport over the street, so every airplane interrupted a process of recording. The place was chosen to be also far enough from the traffic lights which were situated in the nearest intersection (Figure 11).

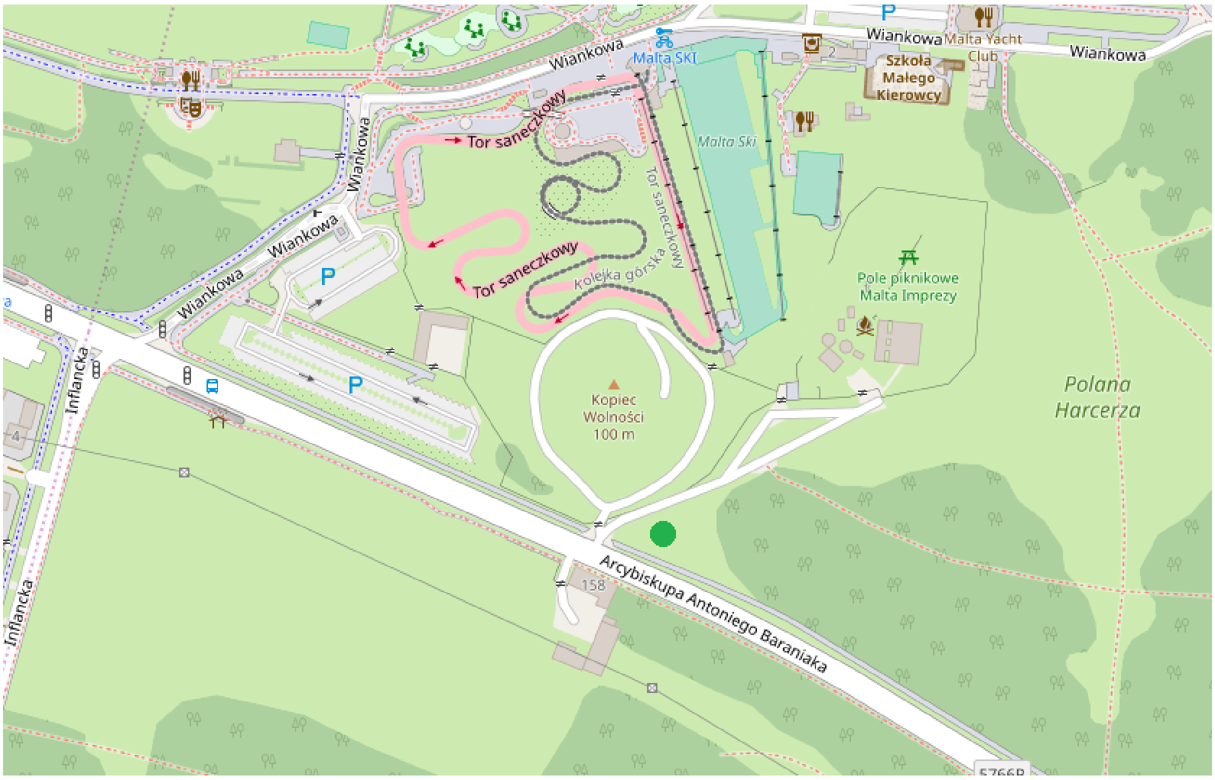


Figure 11. The location of a measuring point of four-lane road with road traffic only.

2.1.2.2 Trams

There are no so many places in Poznań where tramways are far away from road traffic. What is more, as was already mentioned above, there are seven different models of trams in the city. We wanted to record all of them. The best place seemed to be a straight tramway track located near trams' depot – so every tram model could be captured there. Drawbacks? The place is not far from one of the biggest cargo train station in Poland (Franowo) as well as close to the back of the big store (shopping centre, Figure 12). That is why sometimes trains' horns or big trucks interrupted recording.

Figure 12. The location of a measuring point of tramways' noise when trams are only present.

2.1.2.3 Aircraft

We wanted to record airplanes just under a skyway. The skyway above Poznań is located from east to west. Eastern side of Poznań would be a difficult choice as many other noise sources could be observed in the center. That is why we decided to record aircraft in the west side of the Ławica airport. All airplanes fly over Przeźmierowo, a parish neighbouring with Poznań. We located our microphone in a peaceful Wiosny Ludów street there (Figure 13).

Figure 13. The location of a measuring point of aircraft noise. Only incidental road traffic was present there.

2.1.3 Technical aspects of the recording procedure

For road traffic, recordings were made during one day and lasted about 6-7 hours. For trams, as their flow is less fluent, recordings were made during two days, again about 6-7 hours per day. The least frequent were airplanes, so recordings were made during three days, lasting about 8 hours per day. As wind direction changes, sometimes airplanes took off and approached the landings directly above the recording point.

Trams and road traffic were recorded at both 10 and 30 meters from the middle of a track/road. 10m corresponds with the distance between sources and pedestrians on sidewalks. 30m was chosen to mimic the distance between the middle of a transportation trail and a façade of buildings located nearby. Sound meters (SVAN 945A) were also placed at these distances to measure $L_{AEq,T}/L_{AE}$ and other noise indicators. The microphone and sound meters were always positioned 1.2m above the ground (see Figure 14 for the example). Please note that finally only 10m recordings were used in both experiments.

Figure 14. The ambisonic microphone with the sound meter placed nearby during the procedure of recording trams from the distance of 10m.

Road traffic was recorded to keep its natural flow. That is, no single pass-bys were recorded but the whole 'packages', i.e. all vehicles which established a consistent group, mainly controlled by traffic lights in the intersections nearby. For both two and four-lane road we established 'control points' which were located ~50m from the point of measurement. Every time a vehicle passed a control point, the recording was started and lasted until no vehicle was in the 'recording area'. Sometimes a single package contains only one or several vehicles and lasts seconds, sometimes there are dozens of them and several minutes took to take one recording. In every package light and heavy vehicles were counted and average speed of them estimated.

On the other hand, each tramway was recorded separately, unless several of them passed the point of the measurement in the same time. Again, the speed of each tram was also estimated. We also described which model of a tram was recorded.

Aircraft was recorded 4m above the ground, again a sound meter was placed at the same height to measure noise indicators values. All places of measurements are also shown in Figure 15 which presents an overall view of the city of Poznań. The local airport provided information about model of each aircraft.

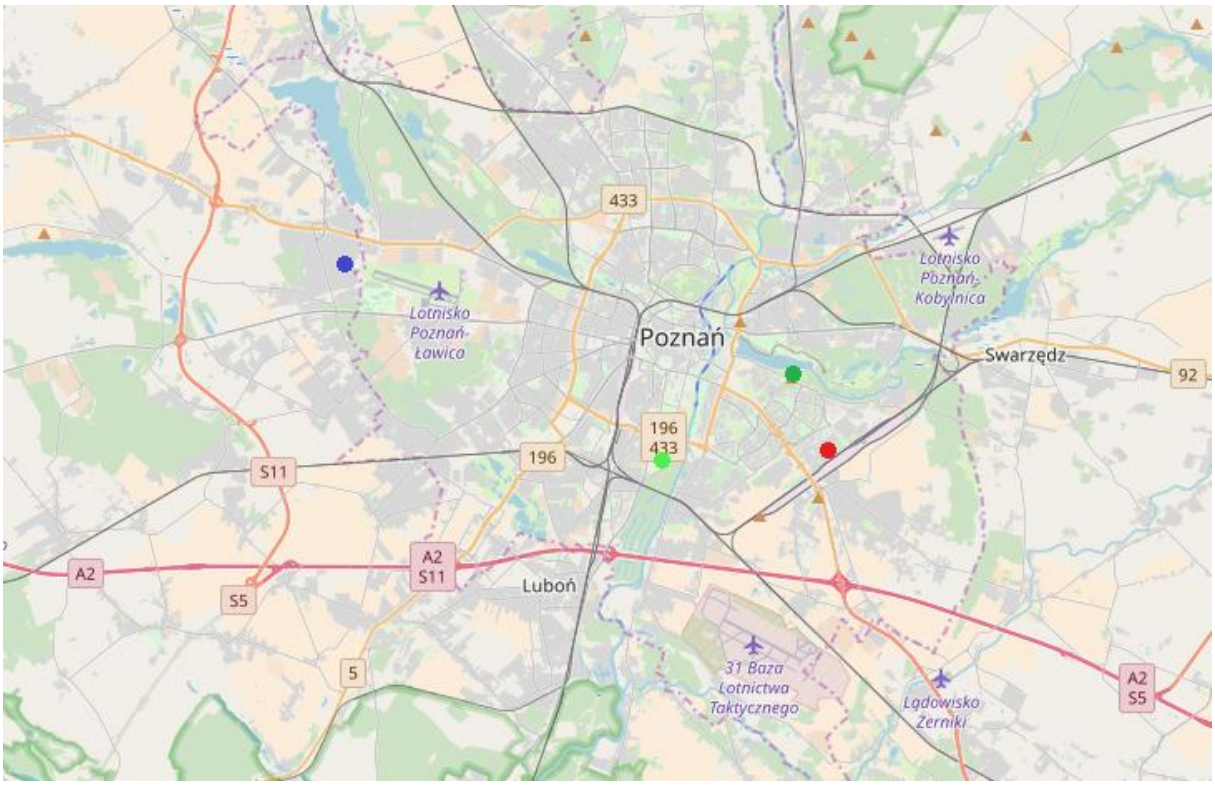


Figure 15. An overview of a city of Poznań with all points of measurements.

2.2 Stimuli

2.2.1 Construction of stimuli

In this research two experiments were conducted. The first related only to road traffic and the second – to mix of cars, trams and airplanes in different proportions.

As already mentioned, our aim was to design the stimuli as close as possible to the actual traffic flow. So, we wanted to simulate road traffic flow, aircraft operations and trams' pass-bys based on information from the city.

2.2.1.1 Experiment I

Traffic information is collected in Poznań by means of inductive loops located at each intersection with lighting. As we wanted to simulate busy road traffic, which is quite common in the center of Poznań, we decided to take data from four-lane road. Przybyszewskiego Street was chosen because it is also a road with a tram line and over it there is a skyway to the Ławica airport – so every source of sound we wanted to use is present there.

The data about traffic flow is expressed in number of light and heavy vehicles per hour. We took numbers related to the peak of traffic and divided them by six to obtain a characteristics of 10-minutes period. The numbers are: 1930 light vehicles/h and 122 heavy vehicles per hour. It means, that heavy vehicles account for ~5.85% of the whole traffic flow. After dividing, it is 322 and 20 respectively.

Based on those numbers, we carefully analyzed all road traffic packages from Baraniaka Street. We had to select best of them, without any interruptions, ambulances, too much windy etc. Finally, ten different packages were chosen, with 331 light vehicles and 20 heavy ones. All recordings lasted about 590 seconds. After this procedure, the order of presentation of the different recordings was combined randomly.

All recordings were faded in and out to limit a problem of a sudden peak in loudness at the beginning of the recording. So, two seconds fade in and fade out were applied. We have also added background noise – because there is still noise in the city, even if there are no noise sources nearby. To do that, we used the same recording as in Kaczmarek & Preis, (2010). It was the quasi-stationary noise recorded at a large distance from a city road infrastructure.

As mentioned in the research hypotheses, we wanted to analyze the different number of pass-bys and time patterns and their impact on human judgment of noise annoyance. Thus, this situation with 331 light and 20 heavy vehicles represented a case, when a traffic flow is dense (peak hours). One could also mention that the time at which cars are observed is 100%, i.e. there are no quiet periods between packages. However, when road traffic is less dense, we can easily think about situation when there are some silent parts between cars. That is why we decided to simulate also a situation with 75% and 50% of proportion between pass-bys and silence.

To create new scenarios with more quiet periods, we carefully selected fewer packages. They were taken from the same set which was used in a '100% case'. We wanted to keep the proportion between light and heavy vehicles the same. It resulted in following scenarios:

- For a '75% case', 7 recordings lasting 424 seconds, with 230 light vehicles and 14 heavy ones; recordings were distributed into a timeline evenly, with the same 'silence gap' between them (each gap lasted ~29 seconds)
- For a '50% case', 5 recordings lasting 280 seconds, with 158 light vehicles and 10 heavy ones; again put evenly, each gap lasted 80 seconds)

Examples of all three cases (100%, 75% and 50%) are presented in Figure 16 – Figure 18.

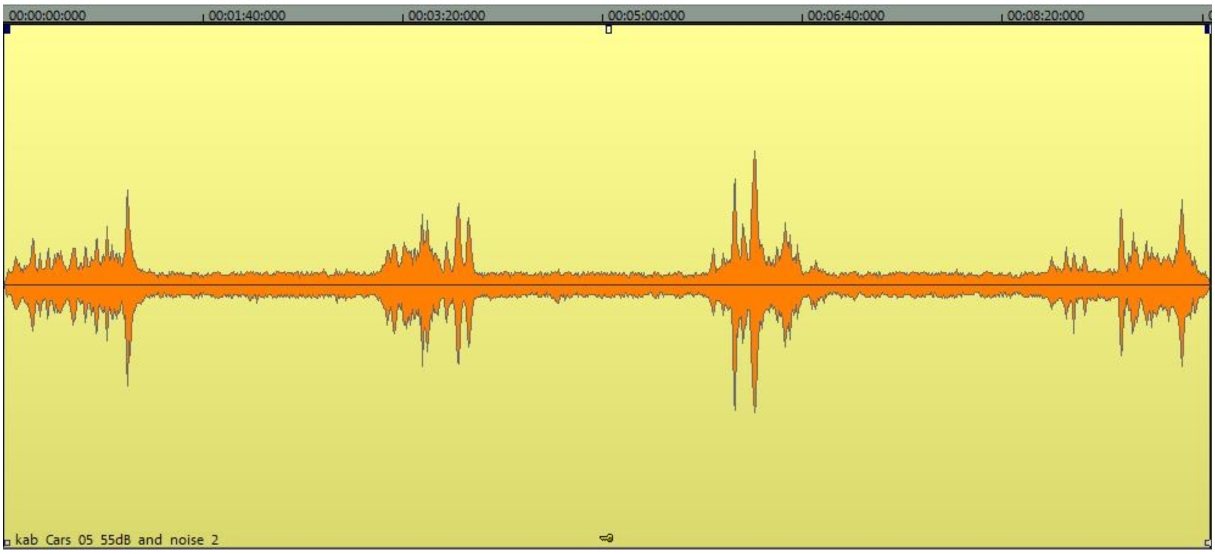


Figure 16. Road traffic noise scenario with 50% of proportion between pass-bys and quiet periods. Noise was presented with a background noise.

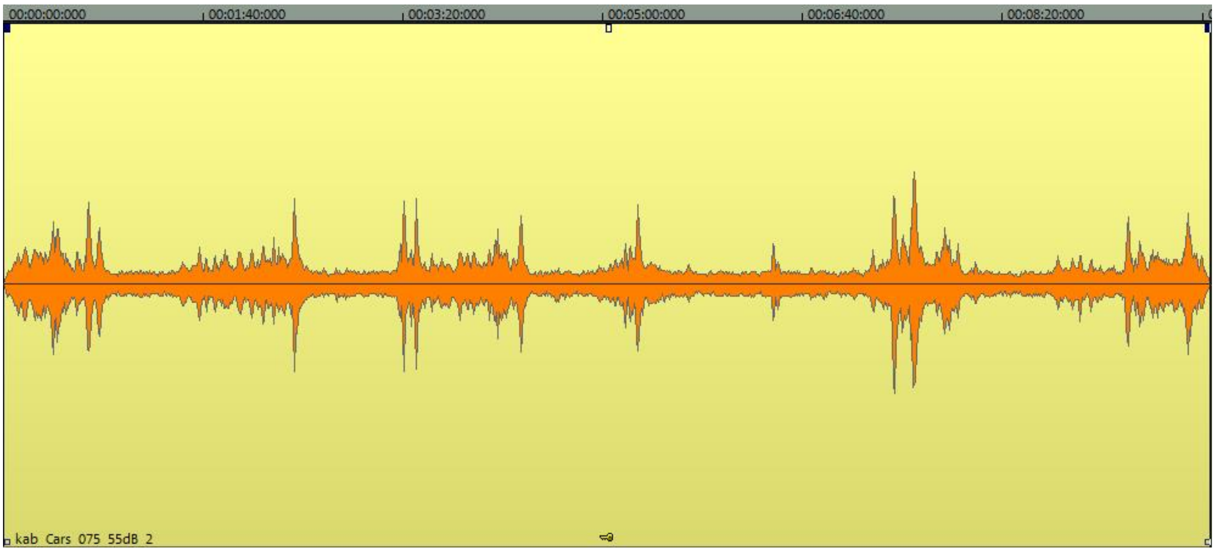


Figure 17. Road traffic noise scenario with 75% of proportion between pass-bys and quiet periods. Noise was presented with a background noise.

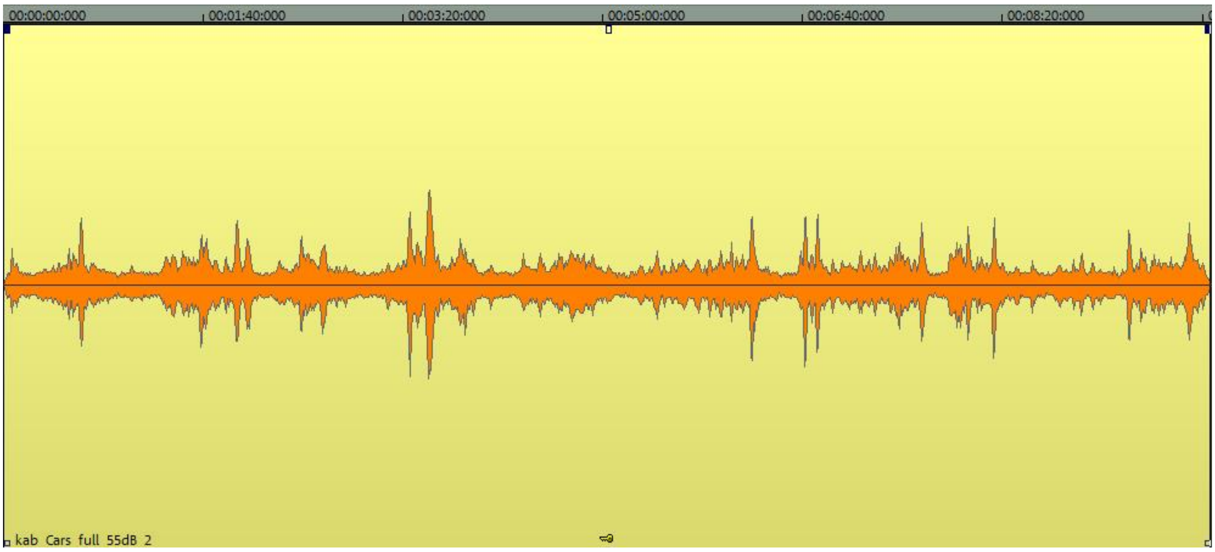


Figure 18. Road traffic noise scenario with 100% of proportion between pass-bys and quiet periods. Noise was presented with a background noise.

Now, after preparation of 3 different scenarios we had to choose noise levels to present them at.

To do so, we analyzed data from a noise map determined for Poznań in 2017 in the area of Przybyszewskiego Street. As can be seen in Figure 19, L_{DEN} values are about 70 dBA around the street. Interestingly enough, we measured similar values during our recording procedure near the Baraniaka Street.

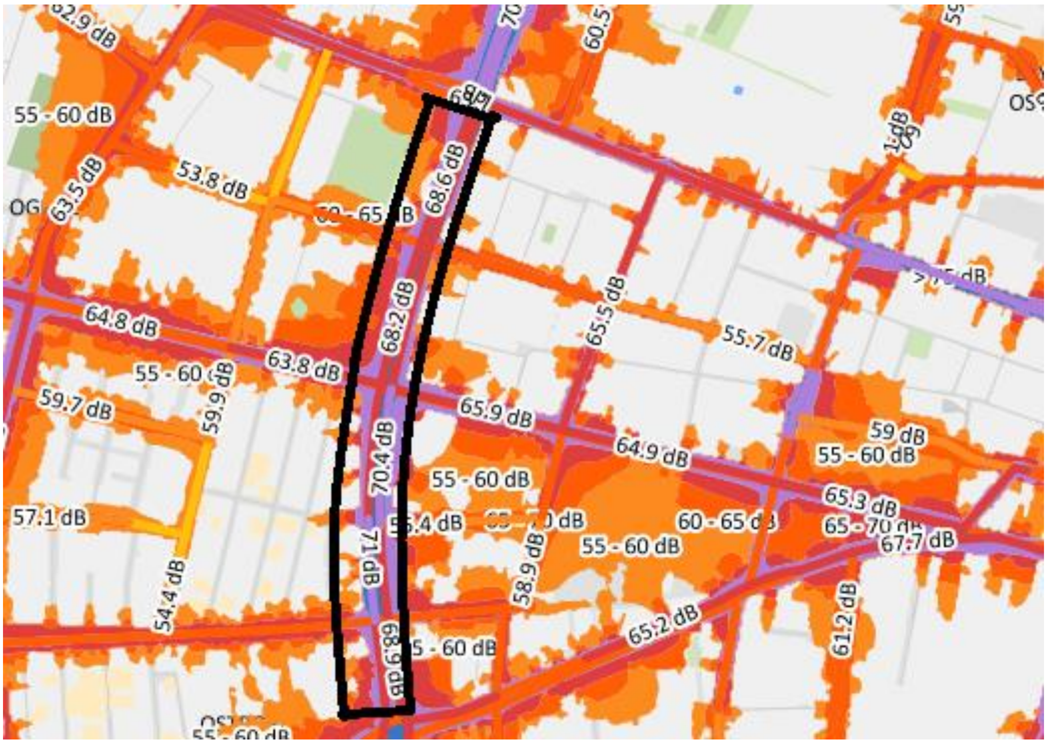


Figure 19. The area of the Przybyszewskiego Street in Poznań which we wanted to simulate in our experiment. This image is a part of the noise map of Poznań.

We decided to use closer recordings (from 10m) as they had better quality comparing with the ones from 30m (lower background noise, less distractors etc.).

The worst case was simulated, i.e. no filtration of the window and façade transfer function was applied. This approach is commonly used by prof. Catherine Marquis-Favre and her team. For example, in (Gille et al., 2016a) they stated that it is difficult to simulate a 'universal' transfer function, it is therefore better to assume the worst case when people spend their time outdoors. In the same way it was assumed that people spend their time in the garden located near a busy street. However, we applied a 10 dBA reduction in the road traffic sound levels – because prolonged exposure to high sound levels can tire our respondents and impair their hearing ability; another reason was the assumption that the garden and street are separated by hedgerow.

$L_{AEq,10m}$ = 60 dBA was the highest value used in the experiment. We also presented stimuli at 55 and 50 dBA. We do know, that this is a strongly limited approach, however it was necessary to keep the experiment short enough, without long-lasting experimental procedures and too many combinations of noises.

All three proportions (mentioned already above) were presented at all three levels. The procedure was as follows:

- put all recordings into a timeline with appropriate silence gaps,
- add a background noise,
- manipulate an overall volume of all recordings to get $L_{Aeq,10m}$ = 60, 55, or 50 dBA. It means that we kept the original volume of the recordings and then changed it for each recording by the same value – so the proportions between the different volumes of the recordings were the same,
- manipulate a volume of a background noise to be around 40 dBA.

Volume manipulation requires more detailed explanation. We have already mentioned that we have maintained the original volume proportions between the different road traffic packages. It can be seen in Figure 16 – the maximum volume of the third package is clearly higher than of the second one. It means that each package had a different sound level, but the value of $L_{Aeq,10m}$ was set to 50, 55 or 60 dBA.

On the other hand, we wondered how the perception of noise could change when artificially set all packages to have the same sound level. In this approach, every road traffic package was set to have the same sound level value – i.e. 50, 55 or 60 dBA. It means that, e.g., for a 50% proportion noise scenario presented at 50 dBA, all four sound packages were changed in volume to emit a sound level of 50 dBA each. Obviously, such a modification influences an overall sound level, i.e. $L_{Aeq,10m}$. For 50% and 75% proportion cases, $L_{Aeq,10m}$ values were slightly lower in this case than in the previous one (see Table 3 for more details).

What was the purpose of using these two methods of setting sound level values? Basically, when the original volume proportions between sound packages are maintained, the relationship between their sound level is 'natural' – i.e. a heavy vehicle pass-by is louder than a passenger car and so on. On the other hand, the dynamic range between the most silent and the loudest package are high – meaning that short-time sound level values vary during the whole 10-minute period.

When each sound package is set to have the same short-time sound level, the differences are smaller, thus the traffic flow seems more 'stable' and the heavy vehicles and passenger cars give similar sound levels. Of course, when the proportion is set to 50% or 75%, the second method leads to lower $L_{Aeq,10m}$ values than in the first case.

However, when scenarios are presented with 100% proportion, each method leads to give the same (or, almost the same) sound level. But the fluctuations of short-time sound levels differ between both approaches. Thus, a question arises: are the people more annoyed by a given noise when it is loud, but quite stable in a short-time sound level, or they prefer 'natural differences', meaning that short-time sound level varies highly? It is commonly assumed that stationary (so, stable in a common sense) sound sources are less annoying as people adapt quickly to the constant sound level. On the other hand, when sound level varies, they should be more annoyed. We tried to simulate this problem with these two approaches. In general, the first situation reflects the stability of short-term sound level values while the second mirrors some fluctuations in these values. Finally, it should be mentioned that on noise maps the values of noise indicators are given in 5-decibel intervals, i.e. their values are represented as 45-50 dBA, 50-55 dBA etc. In our case, both methods gave scenarios which could be represented in such groups. Thus, from 'noise mapping' point of view, all were the same (from the same $L_{Aeq,10m}$ range). But was it so for the respondents?

In the following parts of this thesis, the case where all the proportions of the sound packages have been preserved will be called a 'scenario'. The case when all sound packages were adjusted to give the same sound level values, will be referred to as 'events'.

To illustrate the difference between the two cases, we present two scenarios 75% 60 dBA. In Figure 20 one can see a scenario where $L_{Aeq,10m}$ value was set to 60 dBA. The following figure (Figure 21) shows a scenario in which each sound package has been set to provide a sound level of 60 dBA. It can be clearly seen that in the first case the differences in volume between packages are higher than in the second case. Moreover, two middle sound packages in Figure 21 are louder than in Figure 20 while the rest is quieter than in the first case.

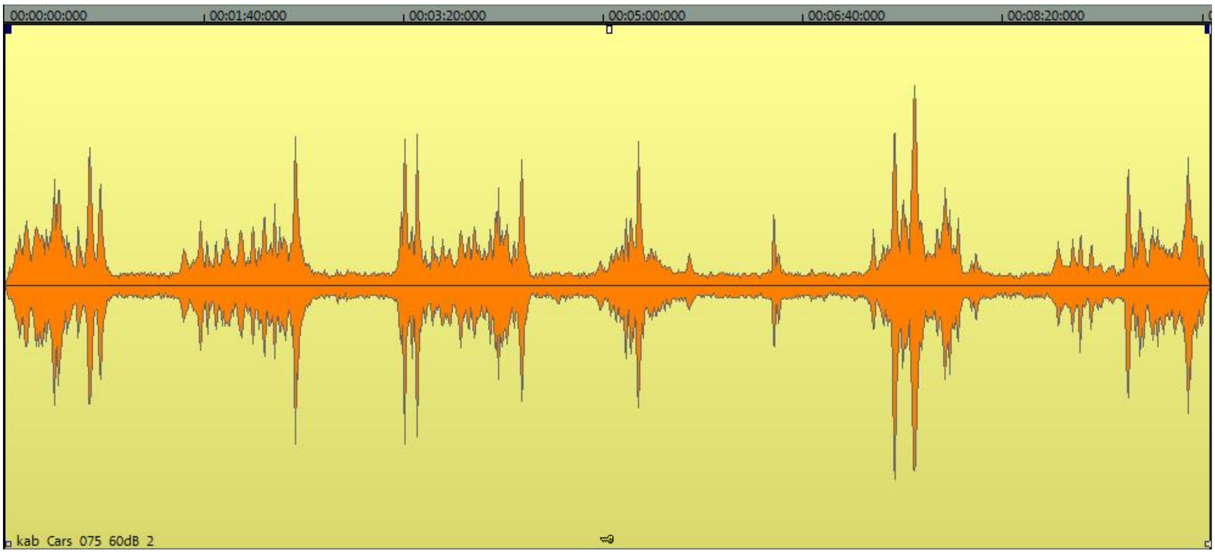


Figure 20. A 'scenario' case, in which all sound packages, presented with 75% proportion, are kept with the original proportions between volumes, but set globally to give $L_{Aeq,10m}$ = 60 dBA .

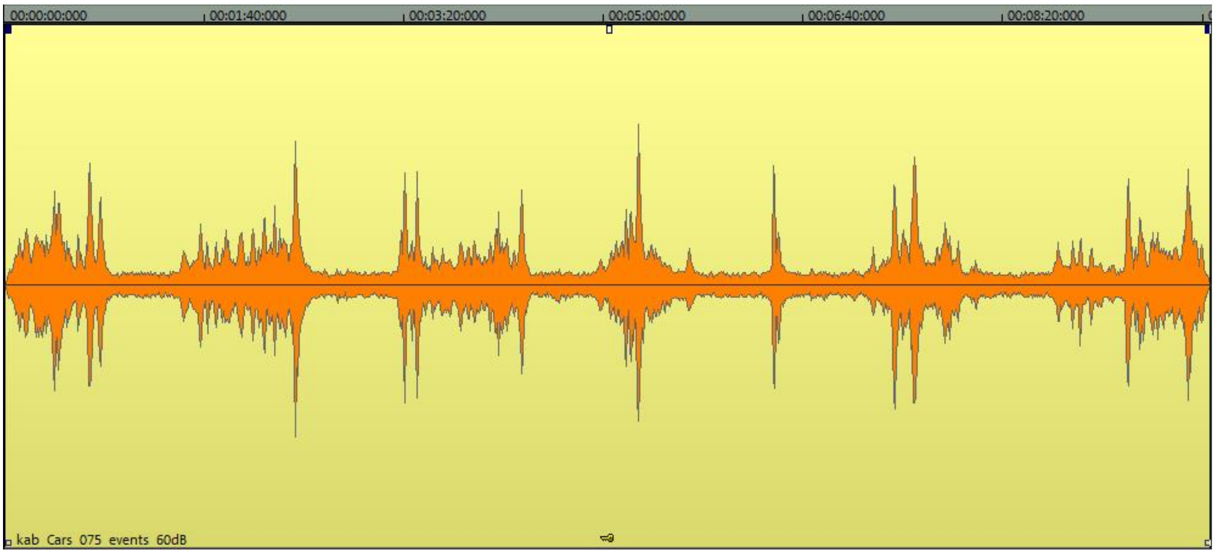


Figure 21. An 'events' case, when all sound packages, presented with 75% proportion, were artificially set to give sound levels of 60 dBA for each.

In summary, different time structures of noise scenarios have been developed. In the following sections we will refer to them using the variable ‘time pattern’ or ‘proportion’. ‘Events’ means that every road traffic package was set to have the same $L_{Aeq,T}$ value. ‘Scenario’ means, that all noise packages were changed by the same decibel step to maintain original proportions of sound level between them, but to give 50, 55 or 60 dBA of $L_{Aeq,10m}$.

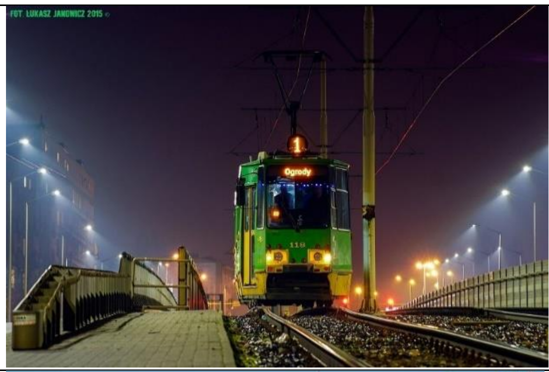
Finally, 18 different road traffic noise scenarios (each of 10 minutes duration) were prepared: 3 different proportions x 3 different sound levels ranges x 2 different time patterns. All scenarios were prepared using B-format of recordings, i.e. based on 4 channels (w, x, y and z).

2.2.1.2 Experiment II

In this experiment, we presented to our participants different mixes of road traffic noise with trams and aircraft noise. The road traffic noise was kept always the same: we used the scenario from Experiment I, so 100% of road traffic flow and sound level of 60 dBA. It was that scenario where original differences between various road traffic packages were kept.

To simulate the noise situation in different places within Poznań area (with more or less busy tramlines), we decided to prepare scenarios with road traffic and different trams’ pass-bys (1, 2, 4, or 8 tramways’ pass-bys were added to the road traffic). First, we analyzed the structure of tramway fleet in Poznań. As it was already mentioned, 7 different types of tramways are used in the city. The structure of the whole fleet is presented in Table 1 (source of data: Municipal Transport Authority, www.mpk.poznan.pl).

Table 1. Models of trams used in the city of Poznań with their number in the whole fleet. All information and photos taken directly from Municipal Transport Authority website, www.mpk.poznan.pl

Model of a tram	Number in the fleet	Proportion of a whole fleet	Photo
Konstal 105 Na/NaDK	42	18.9%	
Moderus Alfa	20	9.0%	
Moderus Beta	48	21.6%	
Siemens Combino	14	6.3%	

Solaris Tramino	46	20.7%	
GT8/GT0	38	6.3%	
Tatra RT6	14	17.1%	

Data about average noise exposure levels of different tramways were provided by AkustiX Company. Near a tramline, they oscillated around L_{AE} = 78-82 dBA. However, with regard to road traffic noise, we applied a value 10dB lower and we did the same with regard to tram noise.

As it has been mentioned before, we created scenarios with 1, 2, 4 or 8 different trams' pass-bys. To simulate real proportion between different trams' models, we used following combinations (in brackets we provide abbreviations then used in a Table 2):

- For 1 tram: 1x Moderus Beta (MB1)
- For 2 trams: 1x Moderus Beta, 1x Solaris Tramino (ST1)
- For 4 trams: 1x Moderus Beta, 1x Solaris Tramino, 1x Konstal 105Na (105Na), 1x GT8 (GT8)
- For 8 trams: 2x Moderus Beta (different recordings; MB1 & MB2), 2x Solaris Tramino (different recordings; ST1 & ST2), 1x Konstal 105Na, 1x GT8, 1x Moderus Alfa (MA), 1x Siemens Combino (SC)

Everytime when a given tram was presented in the scenario, it had the same sound exposure level. However, in a given scenario the moment the tram appears was chosen randomly – with the limitation that different trams' pass-bys do not overlap. L_{AE} values used in the experiment were the same as in the Table 2.

Table 2. Values of L_{AE} characterizing each different type of a tram.

	Model of a tram							
	MB1	**ST1**	**105Na**	**GT8**	**MB2**	**ST2**	**MA**	**SC**
L_{AE} [dBA]	70.7	70.8	71	73.8	71.5	71.5	71.2	69.5

The most complex scenarios consisted of road traffic, trams and aircraft noise. Poznań-Ławica airport has only one runway, so up to 2 aircraft can operate during 10 minutes at most.

As we wanted to simulate real aircraft flow over Poznań, we analyzed data provided by the airport. Our analysis revealed, that most of the aircraft operating there are: Boeing B738 and Airbus A320. Again, AkustiX Company provided information about average sound exposure levels of these planes near Przybyszewskiego Street. For B738 it was 86 dBA and for A320 – 84 dBA. Again, 10 dBA reduction was applied to aircraft noise events.

New scenarios were created as a combination of three sources: every case of mixed road and trams traffic (road traffic + 1,2,4 or 8 trams) with 1 (B738) or 2 (B738 and A320) recordings of aircraft landings added to them. We chose landings because in Poznań (and generally in Poland) the wind most often blows from the west, so aircraft touch down in the direction from east to west. The airport is located to the west from Przybyszewskiego Street.

Finally, 8 new scenarios were constructed: with the same road traffic as for tramways, 1, 2, 4 or 8 trams' pass-bys and 1 or 2 overflights. All trams' and aircraft noise events were put into a timeline randomly, again with the limitation that they do not overlap. In Figure 22 we present the most complex noise scenario, i.e. with 8 trams and 2 airplanes. In Figure 23 we present only trams and airplanes pass-bys (with a background noise), to facilitate an identification of them in the whole scenario. The biggest peaks are airplanes, the rest are trams.

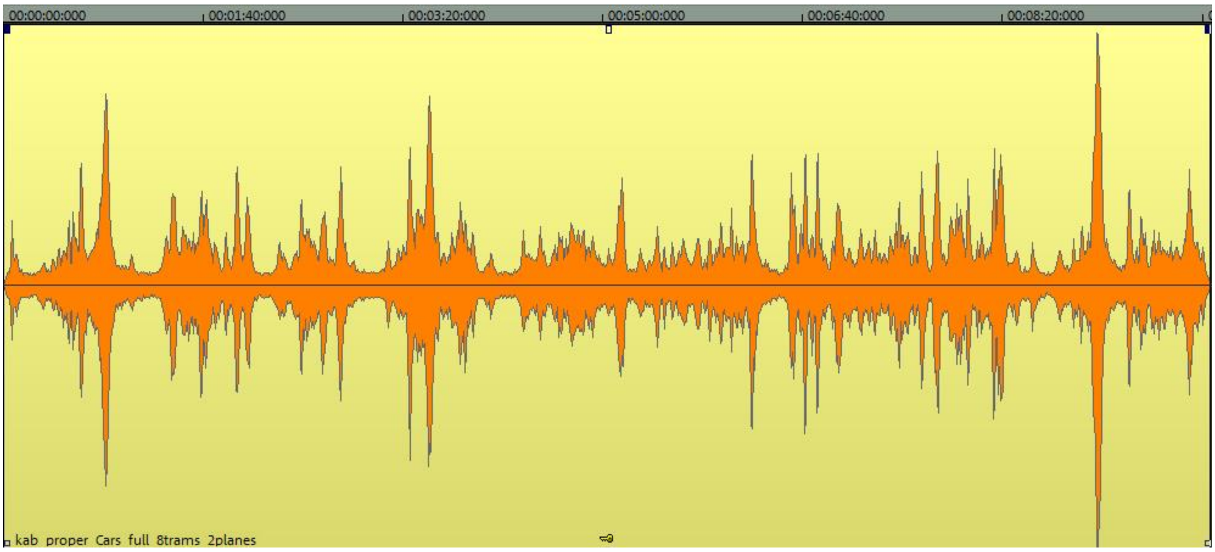


Figure 22. A noise scenario with road traffic, 8 trams and 2 airplanes noises.

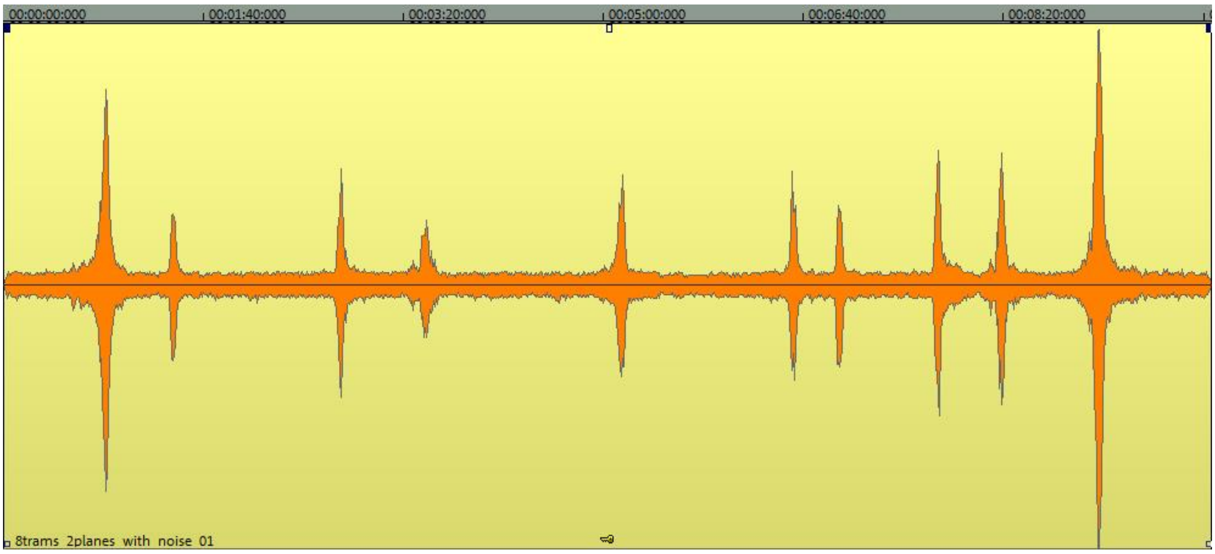


Figure 23. 8 trams and 2 aircraft pass-bys with a background noise.

2.2.2 Calibration of recordings in an ambisonic loudspeakers configuration

In an anechoic chamber located at Institute of Acoustics, a part of the Faculty of Physics of Adam Mickiewicz University in Poznań, 26 loudspeakers (1 was a subbase) were placed in an ambisonic configuration. It is shown in Figure 24.

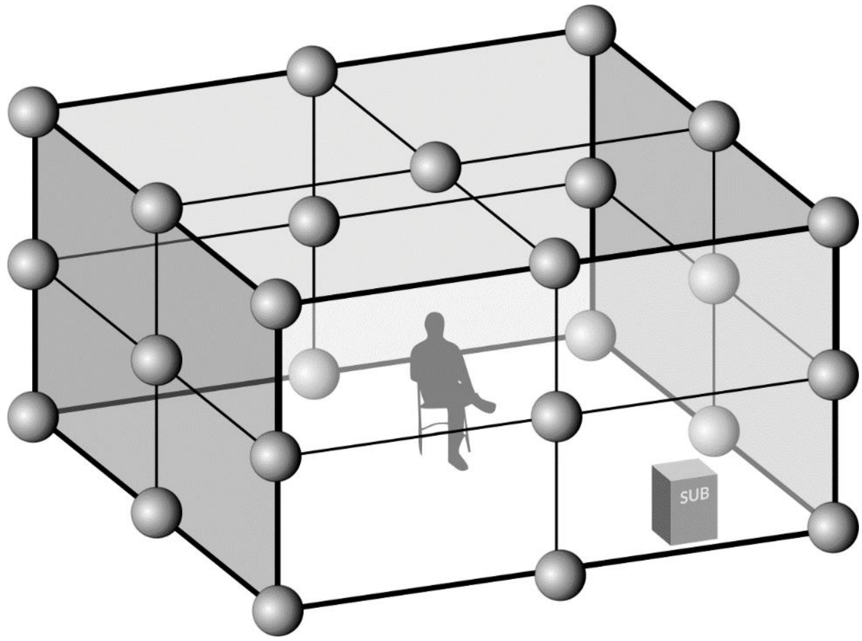


Figure 24. An ambisonic configuration in an anechoic chamber made with 25 Yamaha HS5M loudspeakers. 26 channel is a subbas.

In the lower level of a chamber, in the middle of it, a comfortable chair was placed. It was a place where every listener took part in experiments. In the sitting position, the head of the participant was placed about 1.2m above the ground level. At the same height we placed a calibrated microphone SVANTEK SV11 which was connected to a SVAN 945A sound meter placed outside the chamber.

That point was used as a zero point of a 3D Cartesian coordinates. Positions of all other loudspeakers were measured and described in the same frame of reference. Then, all these coordinates were put into a custom-designed program, written in C#, which transformed B-format recordings into 26-channel .wav files. Compensation of delays (resultant from different distances between zero point and each loudspeaker) was also applied.

This compensation was also used to calibrate the entire system. First, all Yamaha HS10 speakers have been set to the same value using the volume control knob. Secondly, we created an audio session in Samplitude Pro X DAW. 26 different channels with a 10-minute pink noise sample were set. Then, 26 additional buses were created, each representing one speaker. Each track with a pink noise was then connected to the corresponding bus. Next, a noise was emitted from every loudspeaker separately and, using a fader of a corresponding bus in the audio session, a volume was changed to give from every speaker the same sound level ($L_{Aeq,1m}$ = 75 dBA). Sound level was measured 5 times in 1-minute periods for every loudspeaker.

Ultimately, all recordings were put into audio sessions made in Samplitude Pro X (different sessions for different noise scenarios). Since we wanted to carefully control both experiments, the sound sources of all scenarios were placed on different channels. It means that first 26 channels were dedicated to road traffic noise, next 26 – to trams' recordings, and so on, ending with 26 channels for a background noise. Then, every recording which was meant to be played from a specific loudspeaker was sent to the corresponding bus. So for example road traffic from the channel 1, trams from the channel 27 (trams 1), aircraft from the channel 53 (aircraft 1) and background noise from the channel 79 (background 1) were all linked to the bus no. 1. The schematic view of all connections can be found in Figure 25.

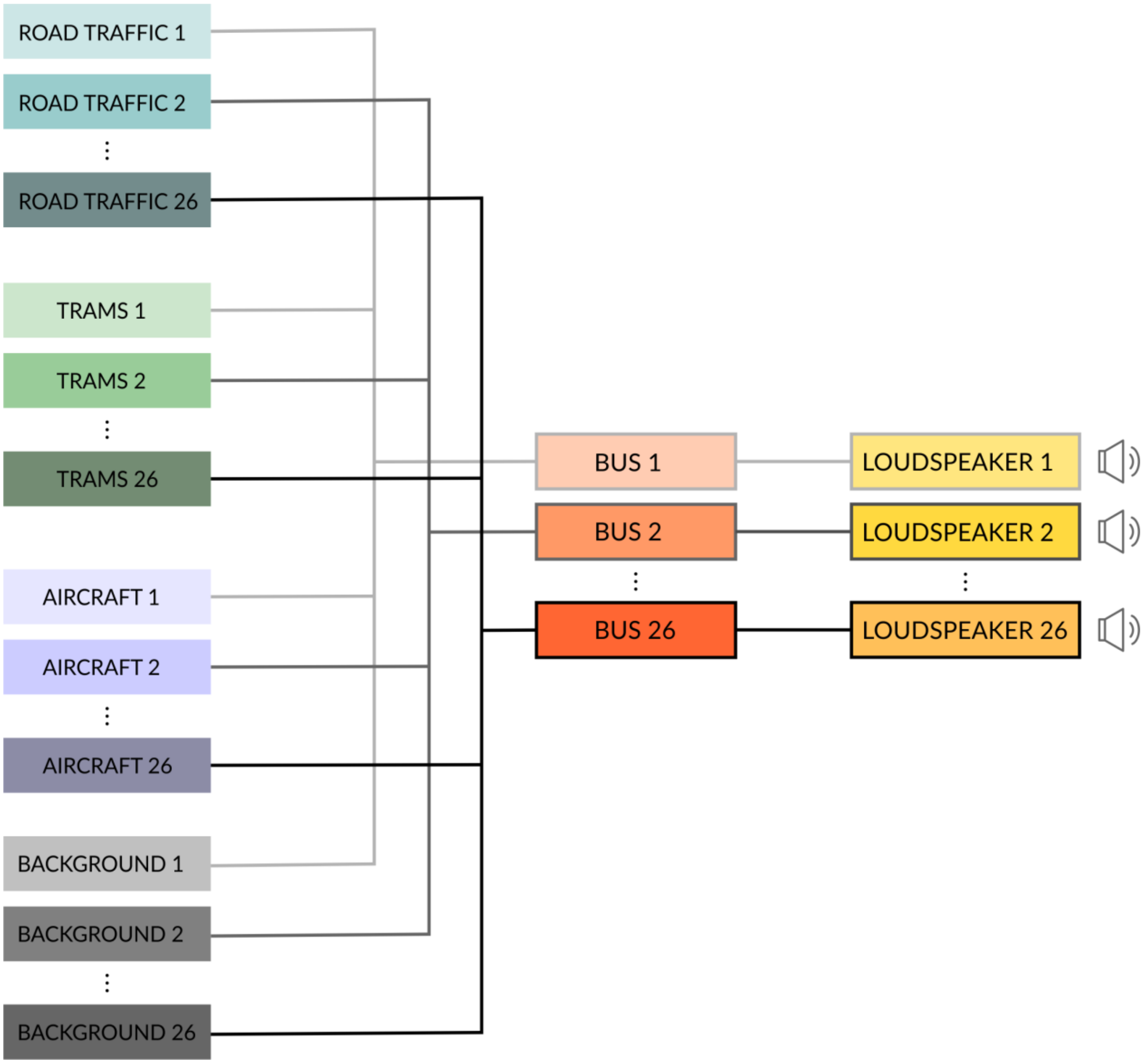


Figure 25. Diagram representing connections in virtual mixer of DAW used to present noise scenarios.

Before calibrating noise scenarios, we calibrated a background noise to give an equivalent sound level around 39-40 dBA during the whole time (i.e. 10 minutes). All noise scenarios were then played from the whole system and $L_{Aeq,10m}$ was measured in the point of participants' head to give the necessary value three times in a row with the tolerance of +/-1 dB. The volume of scenarios was manipulated in the DAW by means of appropriate sliders and amplitude envelopes.

Finally, we received 30 different noise scenarios (including both Experiment I and II) which were then randomly mixed – for each participant in a different way. All scenarios and their characteristics are presented in Table 3.

Table 3. Basic characteristics of all noise scenarios presented in both experiments.

Scenario	Road traffic proportion	Road traffic sound level range ($L_{Aeq,T}$)	Time pattern	Number of trams	Number of planes	$L_{AEq,10m}$ [dBA]
1	.50	45-50	events	0	0	47.1
2	.50	45-50	scenario	0	0	50.0
3	.50	50-55	events	0	0	51.8
4	.50	50-55	scenario	0	0	55.0
5	.50	55-60	events	0	0	56.9
6	.50	55-60	scenario	0	0	60.0
7	.75	45-50	events	0	0	48.5
8	.75	45-50	scenario	0	0	50.0
9	.75	50-55	events	0	0	53.9
10	.75	50-55	scenario	0	0	55.0
11	.75	55-60	events	0	0	58.1
12	.75	55-60	scenario	0	0	60.0
13	1.00	45-50	events	0	0	50.0
14	1.00	45-50	scenario	0	0	50.0
15	1.00	50-55	events	0	0	55.0
16	1.00	50-55	scenario	0	0	55.0
17	1.00	55-60	events	0	0	60.0
18	1.00	60	scenario	0	0	60.0
19	1.00	60	scenario	1	0	60.3
20	1.00	60	scenario	2	0	60.0
21	1.00	60	scenario	4	0	60.1
22	1.00	60	scenario	8	0	60.6
23	1.00	60	scenario	1	1	60.9
24	1.00	60	scenario	2	1	60.2
25	1.00	60	scenario	4	1	60.2
26	1.00	60	scenario	8	1	60.8
27	1.00	60	scenario	1	2	60.7
28	1.00	60	scenario	2	2	60.2
29	1.00	60	scenario	4	2	60.8
30	1.00	60	scenario	8	2	61.0

2.3 Experimental procedure

All scenarios lasted a total of 300 minutes. As each of the participants took part in both experiments, three 100-minutes experimental sessions were prepared for all of them. Order of presenting stimuli was always random, combined before the first session.

First, people were asked to fill in a short survey about their experience with noise in their place of residence, their sensitivity to noise and their concerns/attitudes about different sources and effects of noise. To estimate people's noise sensitivity we used a short version of NoiSeq survey (Sandrock et al., 2007). On the basis of existing documents on other factors that may affect people's judgment about noise annoyance (Bartels et al., 2015b; Marquis-Favre et al., 2005; Méline et al., 2013; Okokon et al., 2015a; van den Berg et al., 2015), we created a 10-question survey to assess their feelings about different types of noise sources and their impact on them. A survey was presented in Polish, however, in Appendix 1 we provide an English version. In short, in the survey the listeners provided the following information:

- Gender, age, postal code
- Rating of the noise annoyance (from all possible noise sources) perceived at home. The standard ICBEN question was posed here with the numerical scale (0-10)
- Types of annoying noise sources which are perceived at home. The following options could be selected (more than one answer could be marked): road traffic, trams, aircraft, trains, industry, open-air events/sports, neighbors or other (with the field to specify them)
- The short NoiSeQ survey. This survey allowed to estimate the global noise sensitivity of individuals as well as three specific types: habitation sensitivity (at home), work sensitivity (during work, while performing certain mental duties) and sleep sensitivity,
- Ten statements on concerns and attitudes towards various noise sources. The answers were given on a 5-point bipolar Likert scale. This question was numbered as 7 and all statements were numbered from 1 to 10. Thus, in the following sections we will refer to the specific phrase using notation *Q7_X*, where *X* is the number of the statement. The statements were as follows:
    - Neighborhood of my place of resience is noisy
    - I am convinced that noise deteriorates my health
    - I am annoyed by the noise in the workplace
    - I do not like travelling by airplanes
    - I do not like using trams
    - I prefer to use a car instead of other means of transport
    - I consider myself to be a proecological person

- My house/apartment is well-isolated from outside noise sources
- Noise produced by my neighbors / coming from a staircase annoys me at home
- I am afraid that noise sources devalue my property

Please note that respondents filled in the surveys as paper questionnaires. Each person also signed a confirmation of voluntary participation in the experiments.

All sessions took place in our anechoic chamber with abovementioned ambisonic system. There was a voice connection between a chamber and a control room – with the microphone and a loudspeaker placed in the chamber (both apart from the ambisonic configuration, had no influence on it) and a microphone and a loudspeaker in a control room. It allowed us to communicate with a participant. Before the main part of each session, participants were presented with short (2 minutes and 30 seconds) examples to show the loudest and least noisy scenarios. Please note that the examples did not use any recording from the main part of the experiment.

We called that short presentation as the 'anchoring procedure'. One might ask why we did it? There are several reasons. First of all, each person participated in three sessions – different days, different hours and maybe different moods, all of which could affect people's judgments. We wanted to limit that possibility by showing before each session what is the 'noise range' in our experiments. Secondly, rating noise annoyance is not a common activity for most people, so, especially at the beginning of an experiment, the biased response may be high as people learn how to assess noise annoyance. By showing them examples we also made it easier to adapt to the whole procedure. Finally, since everyone was exposed to the same examples, potential differences in noise annoyance ratings of the same recording should be the result of individual differences (such as noise sensitivity and so on).

The least loud example in the 'anchoring procedure' was made from several road traffic packages, giving $L_{AEq,T}$ of 50 dBA. They were presented with the same background noise as was used in experiments. The loudest example included one tram pass-by and one aircraft landing operation presented with road traffic (sound level of 60 dBA) and, again, background noise. The order of presentation was also random. Before each example the following instruction was given to a participant: *'You will hear a short example of noise which you can encounter in this experiment. This is an example of the most/least loud noise which could be then presented to you'.*

After the initial presentation, the main part of experiment has started. Participants were instructed to imagine they are sitting in a chair and try to relax. Each person was initially asked to bring a book, to read it during the experiment and not to analyze the structure of the noise presented to them. After each noise scenario, a short paper survey was carried out to assess how difficult it was to read with the presented noise (on a scale similar to an ICBEN scale with 0 as ‘not difficult at all’ and 10 ‘extremely difficult’; this scale will be then called ‘disturbance’) and how annoying it was (‘annoyance’ scale).

As it was shown in the section 1.1, the term ‘annoyance’ is complex, so it can be sometimes difficult for people to understand it always in the same way. In this research we propose to use also a scale of ‘disturbance’. This scale is directly related to annoyance, as in the paper of J.M. Fields et al., (2001) the question about annoyance has a part: ‘How much does noise from (noise source) bother, disturb or annoy you?’. Lam, Chan, Chan, Au, & Hui (2009) shown that both scales correlate – however it was done for a study in Hong Kong. It is why we tried to check the similarity of both scales for the Polish language. The disturbance question was formulated in relation to the reading activity: *“I present you a numerical scale from 0 to 10 to express your opinion how difficult it was to focus on reading during this part of the experiment”*. The annoyance question was formulated according to remarks from J.M. Fields et al., (2001), thus it did not precise the activity. It is therefore possible that the question of disturbance may be understood by respondents as more specific than the question of noise annoyance. More specific question may result in a higher precision of the answer – meaning that disturbance ratings can be slightly higher than those given for the annoyance scale. Such a mechanism was found in Preis et al. (2013).

If only road traffic was presented, a single annoyance question was posed. If more noise sources occurred, there were several questions about noise annoyance (partial annoyances) of each of them (cars, trams and aircraft), Then we presented one question to rate total annoyance (we asked people to rate annoyance *‘together, from all noise sources you have heard’*). Every question about annoyance was formulated accordingly to guidelines made by Fields et al., (2001) in their Polish version (Preis et al., 2003) and used a numerical ICBEN scale to rate noise annoyance. There was also one open-field question in the questionnaire aimed to provide *‘any other interesting thing/detail you noticed in the recording’*. When the participant declared him/herself ready for the next stimulus, another noise scenario was started and a participant was asked to continue reading. After 5 noise scenarios in each session, a short break (lasting about 15 minutes) was taken to go outside the chamber and rest.

2.4 Participants

Participants were recruited by an announcement at the University, as well as information published in social media. Finally, 31 participants took part. They rated all noise scenarios, i.e. those from Experiment I and these from Experiment II. Before the experimental part, a tonal audiogram of both ears was made for each participant. The condition for participation was to have a normal hearing in accordance with the WHO 1997 standard (not more than 25 dB HL for 0.5, 1, 2 and 4 kHz). Fortunately, all people who came forward met these requirements. All participants who started experiments also finished them and were paid for their participation. Finally, 31 participants took part in both experiments. There were 19 women and 12 men among them. The mean age was 28.2 with SD of 7.4 years old.

2.5 Objective characteristics of recordings

The source of subjective data were surveys filled in by the participants. At the same time we wanted to have objective characteristics calculated from acoustical stimuli directly. Nevertheless it was not trivial to analyze as many as 26 channels in the same time.

We decided to use an artificial head. As people, who were sitting in the chair, had only two ears, it seemed to be the best way of 'transformation' from 26 channels to a stereo pair.

To record all stimuli with a head, we put it on a microphone stand and then installed on a chair – to simulate the same situation when people sit and read a book. Then, to calibrate the system, we installed a microphone above the head (halfway between the ears) to measure a sound level and emitted pink noise from all loudspeakers.

We used a head of the Neumann Company KU100. It was connected to inputs of an RME Babyface sound card and a signal was recorded in Samplitude Pro X DAW. We recorded all noise scenarios using it, as well as single noise sources (i.e. only airplanes, presented solely). The artificial head was placed on the same armchair which was used by participants, at the same height (Figure 26).

Figure 26. Configuration in an anechoic chamber while recording noise scenarios to stereo signals.

After making all the recordings, we then put them in Head Acoustics ArtemiS software to get many objective characteristics. We used a recorded pink noise to calibrate the system and obtain reliable results of sound level/loudness analysis. The following characteristics were computed:

- Sound level versus time [dBA] with 250ms time-window; loudness versus time [sone]; specific loudness [sone]
- Psychoacoustic characteristics: fluctuation strength versus time [vacil]; roughness versus time [asper]; sharpness versus time [acum]; tonality versus time [tu]; psychoacoustic annoyance

All of abovementioned parameters were also computed as percentile statistics, i.e. for 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 95%.

3 Results

This chapter is divided into two sections. The first one contains results obtained in the Experiment I – when only road traffic noise was presented to participants. Respectively, the second section presents results from the Experiment II – in which different noise sources were mixed together.

There are inter-individual differences among people while judging noise annoyance, thus both sections start with an inter-rater reliability analysis – respondents with the lowest cohesion with the others were excluded from further analyses.

Next, both sections present results of the first and second experiment respectively. Please note that every time we use a term 'mean' we refer to the arithmetical mean.

3.1 Experiment I. Road traffic noise annoyance related to different time structure and the equivalent sound level

Since the road traffic is the most common noise source in the European cities, the first experiment focuses only on it. Eighteen different noise scenarios were included in this experiment and two research hypotheses (numbered as 1 and 2 in the section 1.6) were tested. 31 participants rated both annoyance and disturbance on two separate numerical ICBEN scales.

3.1.1 Analysis of respondents' reliability

After collecting all data, we started our analyses with some basic statistical tests. First, we wanted to find out if our respondents were coherent in their judgments. To check it, we computed two reliability tests: Cronbach's Alfa and intra-class correlation (ICC) – both for all participants, as well as in the case 'when an item dropped', for annoyance and disturbance scales separately. Results of both analyses are shown in Table 4 and Table 5.

Table 4. Item statistics computed using Cronbach's alfa test of reliability, for both numerical scales. 'R.cor' is 'Item whole correlation corrected for item overlap and scale reliability' (description from 'psych' package in R), mean and SD values are computed from all answers given by a subject.

Subject	Annoyance			Disturbance		
	r.cor	mean	sd	r.cor	mean	sd
Ksz	.2599	2.3333	.7670	.5063	1.9444	.8726
MH	.6474	1.5000	1.4246	.5756	1.7778	1.4371
LS	.5663	5.1667	1.1504	.5028	3.8889	.4714
AP	.5545	3.0556	1.1100	.1674	3.0000	1.0290
MS	.2441	1.6667	1.0290	.3257	1.6667	1.0290
DM	.4102	2.7778	1.8005	.2968	1.3889	1.4200
RF	.6157	5.0556	.8024	.0132	4.6111	.6978
MB	.3765	3.2222	1.6997	.3928	2.7778	1.5551
JW	.6280	2.3889	1.1950	.3467	2.2222	1.1144
NB	.5949	2.1667	2.2029	.6609	1.7778	1.6997
ON-A	.9068	2.6111	1.8830	.8866	2.7222	1.9943
Asch	.5917	1.2222	1.4371	.7395	.7778	1.0603
Mbu	.0678	3.3333	1.6803	**-.0141**	3.2222	2.1020
JLo	.5396	5.4444	2.3319	.2777	5.3889	2.7038
AC	.4645	5.6111	.8498	.3401	4.8889	1.4907
MW	.4973	1.8333	1.2948	.5639	1.6111	1.2433
JL	**-.0661**	4.6667	1.8787	**-.2241**	4.1111	2.0260
Abi	.3616	3.5556	.7048	**-.0472**	3.5556	.7048
AB	.6975	5.7222	1.0178	.6603	5.7222	1.0178
Jsz	.4052	2.5000	1.7573	.1595	1.7222	1.3636
EZ	.7291	.9444	.9376	.6973	.8333	.7859
MG	.0979	1.0556	1.6260	.1826	2.3333	1.9097
PZ	.3198	.7778	.8782	.4959	.6667	.8402
HP	.5819	1.2778	.4609	.6509	1.2778	.4609
MBei	.6812	2.3333	2.7865	.5965	1.9444	2.6892
MKK	.2917	1.3889	.9785	.2657	1.1667	.7071
MP	.5371	1.0556	1.2590	.6747	1.3889	1.1950
SZ	.2681	.5000	.9235	.1585	.1111	.3234
APa	.6454	5.2222	1.1144	.6969	5.3333	.8402
MHu	.5977	2.8333	.9852	.5277	2.2778	.9583
PP	.5957	3.1667	1.5049	.5749	1.8889	.9634

As it can be seen from Table 4, for the scale of annoyance, only the listener 'JL' was negatively correlated with the rest. For the disturbance scale it was 'JL' with 'RF', 'MBu' and 'ABi' too.

Table 5. Results of analysis of intra-class correlation between participants in the 'when an item dropped' case for both scales. 'ICC' is the value of intra-class correlation for the data when a given subject is removed from analysis. Lower and upper bound are 95% CI around ICC value.

Subject	Annoyance			Disturbance		
	ICC	lower bound	upper bound	ICC	lower bound	upper bound
Ksz	.1770	.0945	.3460	.1024	.0462	.2310
MH	.1669	.0878	.3314	.0997	.0445	.2265
LS	.1707	.0903	.3370	.1058	.0483	.2367
AP	.1711	.0906	.3375	.1105	.0513	.2444
MS	.1784	.0955	.3480	.1061	.0485	.2372
DM	.1751	.0933	.3433	.1055	.0481	.2362
RF	.1712	.0906	.3376	.1113	.0518	.2457
MB	.1785	.0956	.3482	.1069	.0490	.2384
JW.	.1687	.0890	.3340	.1068	.0489	.2383
NB	.1680	.0885	.3330	.0978	.0433	.2233
ON-A	.1542	.0794	.3128	.0875	.0369	.2059
Asch	.1677	.0883	.3326	.0985	.0437	.2245
Mbu	.1871	.1013	.3603	.1194	.0569	.2588
JLo	.1699	.0898	.3358	.1119	.0521	.2466
AC	.1738	.0924	.3414	.1039	.0471	.2335
MW	.1716	.0909	.3382	.1007	.0451	.2282
JL	.1953	.1069	.3717	.1276	.0621	.2719
Abi	.1761	.0939	.3447	.1120	.0522	.2468
AB	.1682	.0886	.3333	.0995	.0444	.2262
Jsz	.1752	.0933	.3435	.1105	.0513	.2444
EZ	.1677	.0883	.3326	.1004	.0450	.2278
MG	.1863	.1008	.3592	.1112	.0517	.2455
PZ	.1759	.0938	.3445	.1036	.0469	.2330
HP	.1741	.0926	.3418	.1040	.0471	.2336
MBei	.1658	.0870	.3298	.0973	.0430	.2226
MKK	.1764	.0942	.3452	.1074	.0493	.2392
MP	.1702	.0900	.3362	.0995	.0444	.2262
SZ	.1778	.0950	.3471	.1084	.0500	.2410
APa	.1682	.0887	.3333	.1008	.0452	.2284
MHu	.1706	.0903	.3369	.1038	.0470	.2333
PP	.1698	.0897	.3357	.1027	.0463	.2315

Analysis of results presented in Table 5 also shows that when an item 'JL' is dropped, ICC increases the most (to .1953 for annoyance and .1276 for disturbance). As only 'JL' was negatively correlated in both scales, it was the only participant who was excluded from analyses in Experiment I.

3.1.2 Similarities between annoyance and disturbance scales

As it has been already mentioned, we asked our participants to assess both annoyance and disturbance caused by road traffic noise. However, it seems logical to treat both these terms as very similar. Several tests were computed to check it.

We started by computing basic descriptive statistics which are presented in Table 6. As it can be seen, both annoyance and disturbance scales have the same median and same quartiles. However, mean value of disturbance is slightly lower than for annoyance.

Table 6. Descriptive statistics of both disturbance and annoyance scales used in the Experiment I.

	Annoyance	Disturbance
Valid	540	540
Missing	0	0
Arithmetic Mean	2.7240	2.4630
Median	2.0000	2.0000
Std. Deviation	2.0741	1.9832
Variance	4.3000	3.9340
Skewness	.5511	0.8534
Std. Error of Skewness	.1051	0.1051
Range	9.0000	10.0000
Minimum	0.0000	0.0000
Maximum	9.0000	10.0000
25th percentile	1.0000	1.0000
50th percentile	2.0000	2.0000
75th percentile	4.0000	4.0000

Distribution diagrams reveal more information (Figure 27 and Figure 28). Annoyance has its peak in the value of 2 (the most common answer) while disturbance had the most answers at 1. Annoyance has more responses in the upper range of the scale (from 5) than it has the disturbance. Nevertheless, it is disturbance that has been assessed once in 10, which was not observed for annoyance. This is somehow in accordance with findings from (Preis et al., 2013) where difficulty in speech comprehension was assessed slightly higher than noise annoyance.

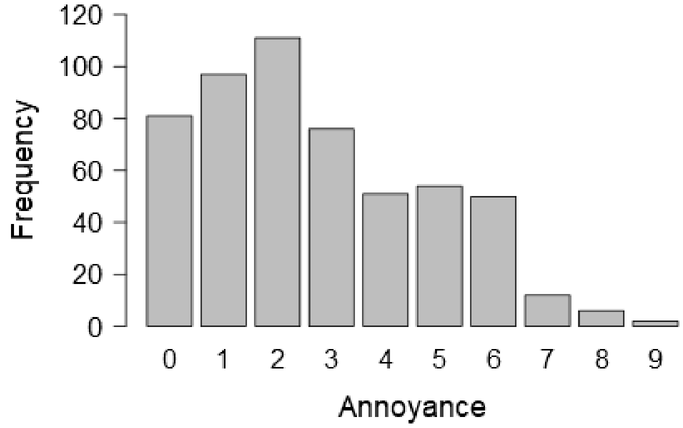


Figure 27. Histogram of annoyance ratings on the numerical ICBEN scale evoked by road traffic noise.

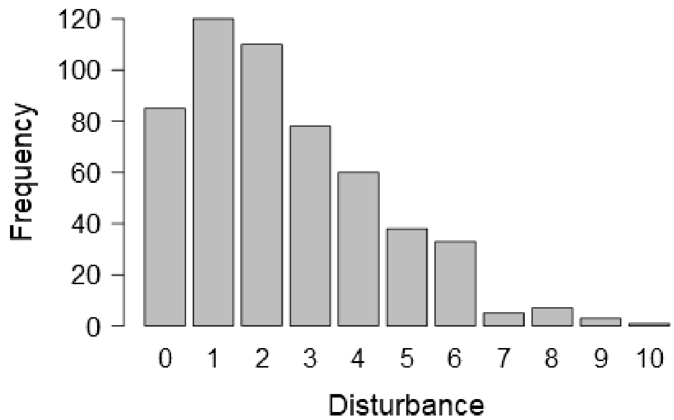


Figure 28. Histogram of disturbance ratings on the numerical ICBEN scale evoked by road traffic noise.

As one can point out that *annoyance* and *disturbance* have both similar meaning, we compared both scales using a Bland-Altman plot (Bland & Altman, 2010). This plot shows differences between both scales in relation to the mean of both ratings. Results can be seen in Figure 29.

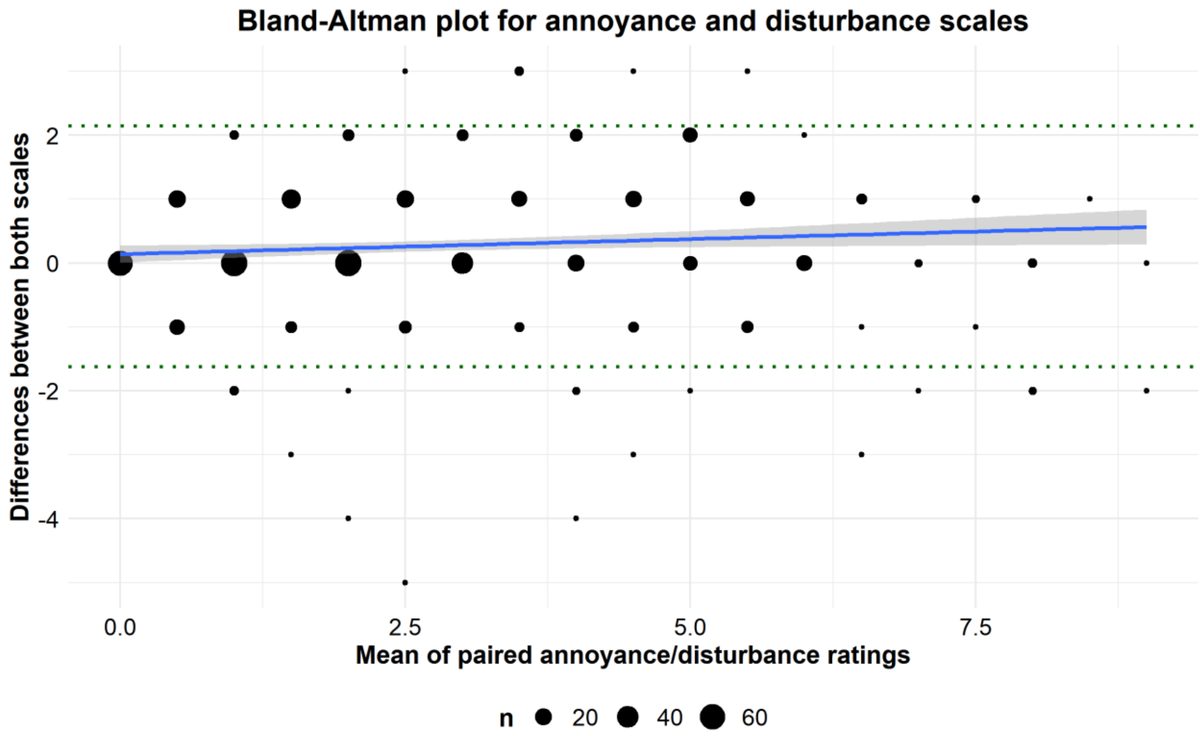


Figure 29. A Bland-Altman plot for both annoyance and disturbance scales used by participants.

The size of points is related to the number of cases with specific values. Moreover, we added a line of regression to show what the trend is as the mean of both scales increases (noise becomes more annoying). As one can see, the line lies around zero and is almost flat – meaning that both scales are comparable in a wide range of ratings. Horizontal green dotted lines represent *2*SD* confidence intervals around differences' mean value. As Bland & Altman suggested, it is recommended that no more than 5% of all observations lie outside those lines. In our case, 540 observations were used and 23 of them did not met this requirement, which represents 4.26% of all cases. Based on this value and the plot, we can say that both scales are interchangeable. So in the following sections we will only use the scale of annoyance – knowing that similar results can also be observed for the scale of disturbance – with the limitation that the disturbance was assessed in relation to reading activity.

3.1.3 Normality and variance of annoyance ratings

A normality test (lillie.test function from the 'nortest' package) showed that for almost every scenario, a requirement of normal distribution was not met. Results of analyses are presented in Table 7. Thus, non-parametric tests based on medians should be applied.

Table 7. Results of normality tests computed for annoyance ratings of every noise scenario presented in the Experiment I.

Sound level range	Proportion	Time pattern	$L_{Aeq,10m}$	D	p	Mean	Median
45-50	0.50	events	47.1	.2709	<.0001	1.8000	1.0
45-50	0.50	scenario	50.0	.1651	.0361	2.3000	2.0
45-50	0.75	events	48.5	.2157	.0010	2.2667	2.0
45-50	0.75	scenario	50.0	.2567	<.0001	2.1000	2.0
45-50	1.00	events	50.0	.1716	.0244	1.7667	2.0
45-50	1.00	scenario	50.0	.1969	.0044	2.1333	2.0
50-55	0.50	events	51.8	.2476	.0001	2.4000	2.0
50-55	0.50	scenario	55.0	.1375	.1570	2.7667	3.0
50-55	0.75	events	53.9	.2083	.0018	2.3667	2.0
50-55	0.75	scenario	55.0	.1880	.0083	2.5667	2.5
50-55	1.00	events	55.0	.1381	.1523	2.8333	3.0
50-55	1.00	scenario	55.0	.1359	.1685	2.6667	2.5
55-60	0.50	events	56.9	.1501	.0828	2.9667	3.0
55-60	0.50	scenario	60.0	.1937	.0056	3.8667	3.0
55-60	0.75	events	58.1	.1469	.0974	3.1667	3.0
55-60	0.75	scenario	60.0	.1540	.0672	3.6000	3.0
55-60	1.00	events	60.0	.2078	.0019	3.5667	4.0
55-60	1.00	scenario	60.0	.1139	.4109	3.9000	4.0

In addition, for each noise scenario and for all participants, the boxplot is presented in Figure 30. It is clear that in almost every case the variance of the response is large and covers most of the range of the ICBEN numerical scale. It means that people are different in terms of the same noise scenario.

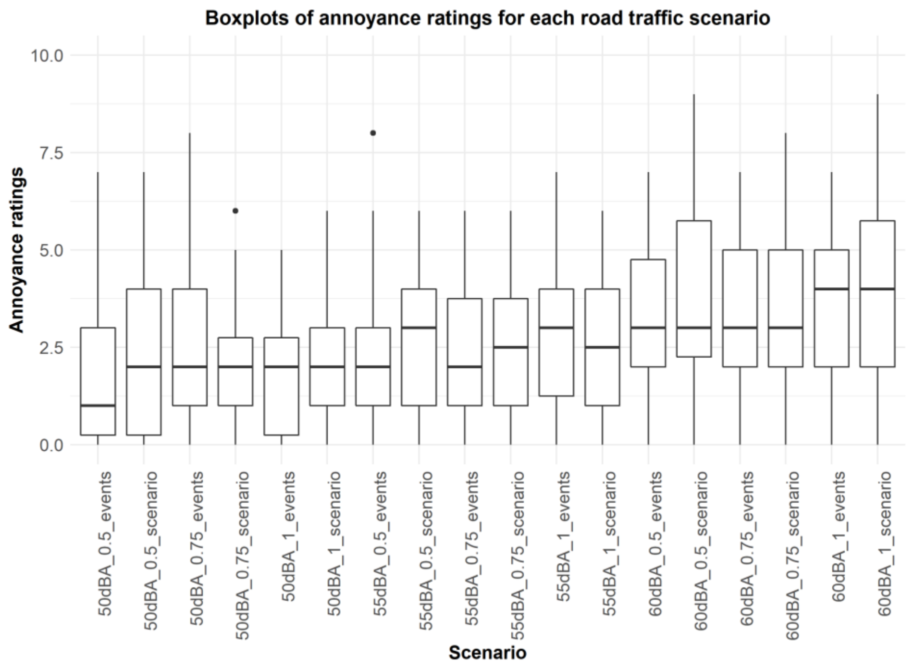


Figure 30. Boxplots of annoyance ratings (over all participants) for each road traffic noise scenario separately.

At this point it is important to clarify some of the problems. The wide variety of people's answers to the question of noise annoyance is not surprising – same problems were mentioned in many other papers on noise annoyance (e.g. H. M. Miedema & Oudshoorn, 2001; Schultz, 1978). This variance is difficult to explain, many scientists assume that non-acoustical factors can significantly affect it. Therefore, when analyzing human noise assessments, it seems reasonable to suggest that this difference is natural and is due to human differences. What is more, we want to emphasize that we are interested in the great picture – global tendencies and relations. From that point of view, a careful analysis of each person individually is pointless. It is obvious that people are different and tend to assess the same noise differently. But when we think about it as a global problem and want to find a good solution (like dose-response curves) in order to predict noise annoyance in cities, we need to generalize and forget about detailed individual analyses.

Another issue is the way of computing summary statistics. Should we take into account only a median? Or can we take a mean, even when normality assumption is not met? Bootstrap methods (Singh & Xie, 2008) seem to be a good solution in such situations. With sufficient number of replications, we can artificially create a distribution that allows us to compute mean and confidence intervals directly from it – without assuming normality.

Using this approach (and a 'boot' package from R), we decided to take a mean of noise annoyance assessment over all participants for each noise scenario separately – creating a vector of 18 means, computed for all 18 noise scenarios. Thus, it was possible to have, e.g., a vector of means correlated with equivalent sound levels. Each time 95% confidence intervals were calculated and displayed.

3.1.4 Relation between objective characteristics and mean annoyance ratings of road traffic noise

For each noise scenario many objective (including psychoacoustical) characteristics were calculated using an ArtemiS software by Head Acoustics. Every recording was described by a separate vector of features. These objective values were correlated with means of road traffic noise annoyance assessments (averaged over all participants).

At first, we have analyzed all possible correlations between means of annoyance and objective characteristics (including all percentile variants). It should be noted that every time we talk about 'correlation' we refer to a Spearman correlation. Next, we limited results only to those which were statistically significant. Then, we manually selected these percentile variants which had the highest values of rho. It turned out that three out of five psychoacoustical characteristics (excluding fluctuation strength and tonality) gave highest values for their 5th percentile variants. Results are presented in Table 8.

Table 8. Results of correlations computed between objective characteristics of noise scenarios and means of annoyance ratings.

Variable	p	rho
$L_{Aeq,10m}$	<.0001	.9629
Level vs time (250ms) 5%	<.0001	.9401
Fluctuation Strength vs time 20%	<.0001	.8822
Loudness vs time 5% (N_5)	<.0001	.9514
PAnn	<.0001	.9463
Roughness vs time 5%	<.0001	.9504
Sharpness vs time 5%	<.0001	.9267
Tonality vs time 80%	.0373	.4938

Description of each objective parameter can be found below:

- $L_{Aeq,10m}$, the A-weighted equivalent sound pressure level measured during 10 minutes of each scenario,
- Level [dBA] versus time, it is a sound level measured in 250ms time windows; 5% refers to the level which was reached or exceeded in 5% of the measurement time (i.e. 10 minutes),
- Loudness versus time 5% (N_5),
- Fluctuation strength versus time 20%; '20%' means that the value obtained is reached or exceeded within 20% of the measurement time,
- Roughness versus time 5%; '5%' means the same as in definition of loudness versus time,
- sharpness versus time 5%,
- psychoacoustical annoyance (PAnn),
- tonality versus time 80%: a characteristic described by Zwicker et al. which expresses amount of energy in tonal components of a signal. We computed tonality using the Hearing Model implemented in ArtemiS; tonality is presented in tonal units [tu].

Since the distribution of annoyance ratings was not found to be normal our main goal was to use a bootstrap approach to be fully independent of normality assumptions. This data was entered into the bootstrap regression analysis. The procedure was as follows:

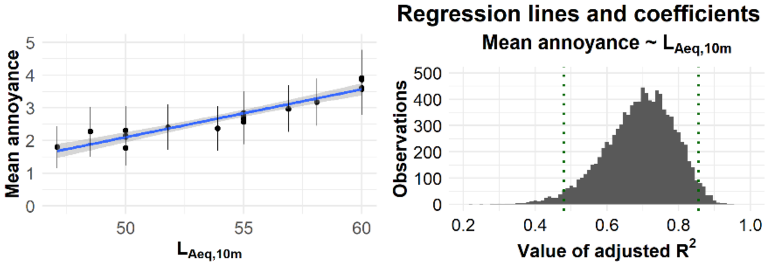
- all eighteen means of annoyance assessments (every mean representing different noise scenario) created one 18-element vector. This vector was bootstrapped with 10 000 replications, giving finally 10 000 different 18-elemented vectors
- these vectors were used in linear regression analysis as a dependent variable. Independent variable was always one of measured objective parameters

- 10 000 different linear regressions were computed for every pair of variables. Coefficients, intercepts, R^2, root mean square errors (RMSE) and p-values were computed as average values with 95% confidence intervals.

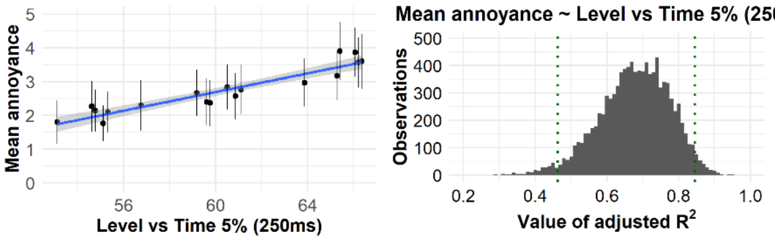
For one pair of variables the mean R^2 was less than 50%. We decided to set a threshold of 50% – assuming that it is justified to exclude all parameters which explained less than 50% of noise annoyance ratings variance. Thus, we have excluded tonality from further analyses. Finally, 7 different characteristics were analyzed: $L_{AEq,10m}$, Level versus time, Loudness versus time 5%, Fluctuation strength versus time 20%, Roughness versus time 5%, sharpness versus time 5% and psychoacoustical annoyance (PAnn).

Plots of all regression lines and bootstrap analyses are presented in Figure 31. On the left side of the diagram mean annoyance ratings with bootstrapped 95% confidence intervals are presented on the y axis while on the x axis different objective characteristics values are shown. Blue lines represent linear regression with the grey background for confidence intervals. In the center are distribution plots of R^2 values for all 10 000 replications with vertical lines representing 2.5% and 97.5% cut-off points (95% confidence intervals). The right side contains tables representing all computed values with their confidence intervals.

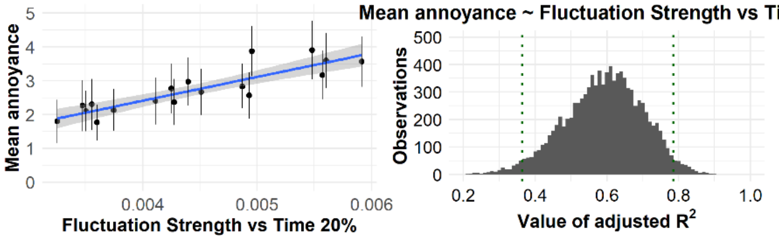
As one can see, the highest R^2 values (around .7) can be observed for: $L_{Aeq,10m}$ (.69), level vs time 5% (.68), loudness vs time 5% (.71), PAnn (.7) and roughness vs time 5% (.7). Another values are: .59 for fluctuation strength vs time 20% and .54 for sharpness vs time 5%.



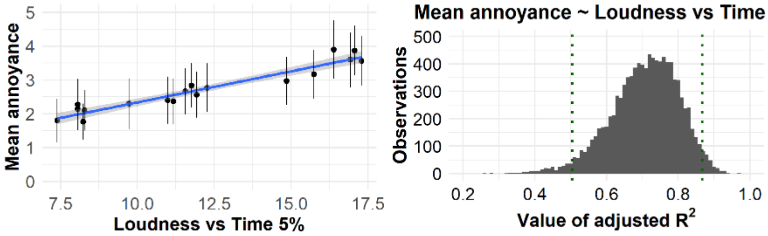
Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.6907	0.4737	0.8543
p	0.0001	0.0000	0.0010
RMSE	0.3890	0.2604	0.5297
Coefficient	0.1465	0.1056	0.1871
Intercept	-5.2233	-7.4068	-3.0284



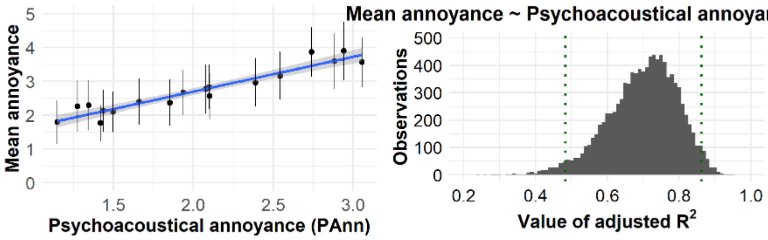
Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.6766	0.4585	0.8445
p	0.0001	0.0000	0.0012
RMSE	0.3984	0.2655	0.5434
Coefficient	0.1385	0.1001	0.1763
Intercept	-5.6183	-7.8665	-3.3196



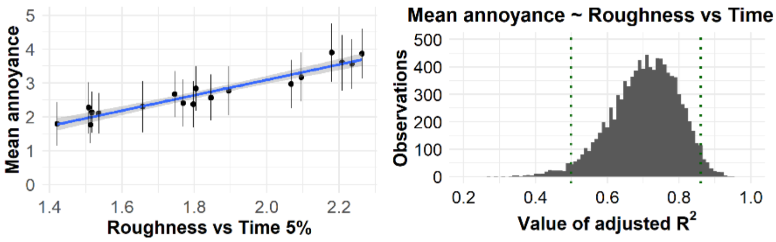
Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.5901	0.3649	0.7909
p	0.0007	0.0000	0.0047
RMSE	0.4499	0.3076	0.5992
Coefficient	703.8326	493.4329	909.3854
Intercept	-0.4063	-1.3167	0.5320



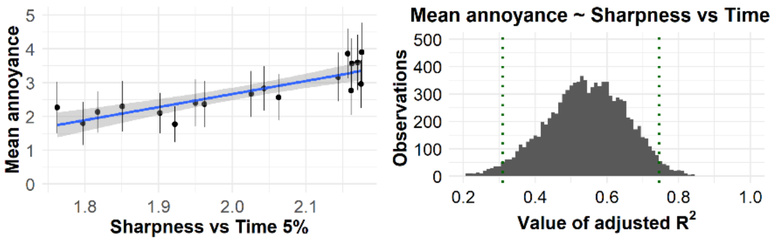
Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.7063	0.4949	0.8651
p	0.0001	0.0000	0.0007
RMSE	0.3789	0.2523	0.5186
Coefficient	0.1839	0.1335	0.2341
Intercept	0.4979	-0.1050	1.1037



Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.6984	0.4811	0.8625
p	0.0001	0.0000	0.0008
RMSE	0.3835	0.2559	0.5219
Coefficient	1.0305	0.7398	1.3154
Intercept	0.6378	0.0582	1.2268



Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.7024	0.4905	0.8612
p	0.0001	0.0000	0.0007
RMSE	0.3816	0.2558	0.5229
Coefficient	2.2715	1.6527	2.8904
Intercept	-1.4500	-2.5752	-0.3242



Parameter	Mean	CI_{Low}	CI_{High}
R^2	0.5418	0.3050	0.7501
p	0.0014	0.0000	0.0103
RMSE	0.4765	0.3328	0.6289
Coefficient	3.8747	2.6704	5.0546
Intercept	-5.0786	-7.4569	-2.6645

Figure 31. Results of bootstrapped regression analyses computed for all significant objective characteristics and mean of noise annoyance assessments for each road traffic noise scenario.

However, linear regression has a number of assumptions that need to be met in order for the model to be credible and stable. Several additional tests have therefore been carried out to check these assumptions:

- Harvey-Collier test, it checks if the use of a linear model for data is justified
- Durbin-Watson test to find out if the residuals are not correlated
- Goldfeldt-Quandt test checks the assumption about homoscedasticity of residuals. Homoscedasticity assumes that all observations in a given vector has the same, finite variance
- Normality test, Kolmogorov-Smirnov test to check the normal distribution of residuals

For all tests we are interested in p-values greater than .05 – which means that the assumptions have been met. All results are provided in Table 9.

Table 9. Results of tests checking linear regression assumptions of different sound characteristics in the Experiment I.

Characteristic	**Harvey-Collier test**		**Durbin-Watson test**		**Goldfeldt-Quandt test**		**Normality test**	
	HC	**p**	**DW**	**p**	**GQ**	**p**	**D**	**p**
$L_{Aeq,10m}$	.8701	.4852	1.9538	.3964	1.4457	.4829	.1406	.5012
Level vs time 5%	.9785	.4409	2.1065	.4965	2.0395	.3621	.1405	.5029
Fluctuation Strength vs time 20%	.7173	.5449	1.9926	.4152	2.7900	.2445	.1434	.4791
Loudness vs time 5%	.8191	.5051	2.1784	.5412	1.9085	.3874	.1400	.5068
PAnn	.7157	.5488	2.0878	.4798	1.6597	.4325	.1406	.5022
Roughness vs time 5%	.9482	.4544	2.1109	.5009	1.9475	.3806	.1401	.5062
Sharpness vs time 5%	2.2734	.1027	1.9065	.3753	1.9334	.3636	.1388	.5170

As one can see, all tests gave results which confirmed the assumptions – which means that the obtained linear regression models are reliable.

Below are the equations for all simple linear regressions that were analyzed. Annoyance is denoted as *A* and each characteristic is coded with the following abbreviations: $LvsT_{250ms}$ is level vs time with 250ms time frame, $FSvsT_{20\%}$ is 20th percentile fluctuation strength vs time, $LoudvsT_{5\%}$ is 5th percentile loudness vs time, *PAnn* is psychoacoustical annoyance, $RvsT_{5\%}$ is 5th percentile roughness vs time and $SvsT_{5\%}$ is 5th percentile sharpness vs time; other parameters should be understandable.

$$A = 0.1466 * L_{Aeq,10m} - 5.226 \quad (21)$$

$$A = 0.1386 * LvsT_{250ms} - 5.6226 \quad (22)$$

$$A = 704.6129 * FSvsT_{20\%} - 0.408 \quad (23)$$

$$A = 0.1841 * LoudvsT_{5\%} + 0.4976 \quad (24)$$

$$A = 1.3120 * PAnn + 0.6379 \quad (25)$$

$$A = 2.2732 * RvsT_{5\%} - 1.4515 \quad (26)$$

$$A = 3.8769 * SvsT_{5\%} - 5.0814 \quad (27)$$

3.1.5 Multilinear regression (MLR) to predict road traffic noise annoyance

Since scientists use MLR in many studies (Gille et al., 2016c, 2017b), we also wanted to check whether this approach is appropriate for our data. To do that, we used the function *step()* from 'stats' package in R. This function allows to compile all independent variables together with the dependent one and calculate which combination of predictors gives the best goodness-of-fit. At this point, however, certain issues need to be clarified.

All objective characteristics which were finally used in our analyses can be divided into one of three groups. Each group of parameters corresponds with the similar feature of sound:

- first group is related to the amplitude of a given signal. Loudness, level versus time and $L_{AEq,10m}$ belong to it
- second group is related to time domain of sound and its modulation – slow changes are represented in fluctuation strength and fast – in roughness
- third group relates to frequency domain and energy in bands – this class is represented by sharpness

Moreover, psychoacoustical annoyance is computed based on loudness, roughness, sharpness and fluctuation strength values, so it is also dependent on them.

Based on these observations, it seems reasonable to select only one representative from each group. Otherwise, some predictors would be responsible for the same variance, thus results of analyses would not be reliable.

We analyzed all possible combinations of predictors from these groups. Nevertheless, none of them was considered statistically significant. In this case, we limit our findings only to simple linear regressions.

3.1.6 Proportion, time pattern, sound level and their influence on road traffic noise annoyance ratings

As it has been already mentioned, road traffic noise scenarios were constructed in such a way, that $L_{Aeq,10m}$ was constant while the time structure of a whole recording varied. This variation was made in two ways: by manipulating the equivalent sound levels of each package in a recording or by manipulating the proportions between 'quiet' periods (when only background noise was present) and 'busy' periods (when cars's pass-bys were presented). The first manipulation in our analyses is called 'time pattern' and the second is named 'proportion'.

Because, as it was said before, an assumption of normality was not met, we took all respondents' ratings of all 18 road traffic scenarios and put them into a robust analysis of variance from a package 'walrus'. 'Robust' in this approach means ANOVA computed for trimmed means – as it was impossible for a three-way design to use bootstrap method or medians. Results of the analysis can be found in Table 10.

Table 10. Results of robust ANOVA analysis with noise annoyance ratings as the dependent variable and cars' sound level, time pattern and proportion as factors.

Factors	Q	p
Proportion	.6792	.7200
Sound Level	46.4578	.0001
Time pattern	3.2949	.0710
Proportion:Sound Level	1.4549	.8370
Proportion:Time pattern	1.4777	.4800
Sound Level:Time pattern	.6558	.7220
Proportion:Sound Level:Time pattern	.7456	.9470

As can be seen from Table 10, a statistical significance was observed only for road traffic sound level, neither for cars proportion, nor for time pattern. Moreover, none of the interactions were considered statistically significant. Post hoc analysis revealed that sound level range was statistically significant for all possible combinations (with p around .001 and .005).

More suggestive information is presented in Figure 32. Higher sound levels implicate higher means of noise annoyance. However, there is one more interesting thing to note: in almost all cases (except two scenarios: 45-50 dBA of sound level with 75% proportion of cars and 50-55 dBA with 100% proportion of cars), people gave higher ratings to globally averaged scenarios ('scenario', triangles) than to scenarios with each package set separately ('events', circles). A relation between road traffic annoyance ratings and $L_{Aeq,10m}$ (see Figure 31) will be discussed later. Nevertheless, it seems that only the sound level has a significant impact on people's judgments about the annoyance caused by road traffic noise.

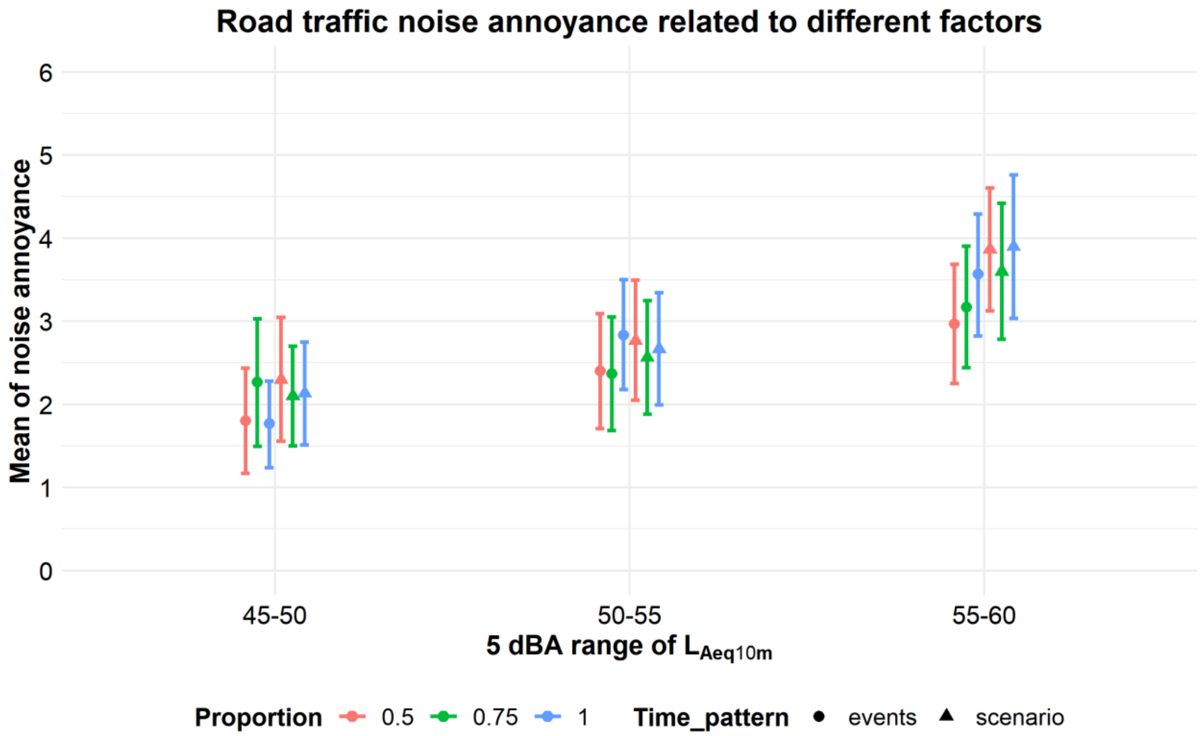


Figure 32. Road traffic noise annoyance related to different factors. Bootstrapped means with 95% confidence intervals are shown for different averaging procedure and different proportion of cars.

3.2 Experiment II: Road traffic noise mixed with trams and aircraft noise

In the Experiment II we presented to our participants combinations of three different noise sources. The main, always the same, was road traffic noise presented at 100% proportion at 60 dBA. This noise was mixed with different number of trams pass-bys (1, 2, 4 or 8) and aircraft overflights (0, 1 or 2). More details about construction of stimuli in this experiment were presented in the section 2.2.1.2 (Construction of stimuli – Experiment II).

In this experiment listeners were presented with mixed noise sources but were asked to assess the annoyance of each noise source separately (partial annoyances, **not** specific annoyances) as well as the general annoyance (total annoyance) resulting from their perception of all noise sources together. Each person was also asked to choose the most annoying noise source for them (only one noise source could be selected).

3.2.1 Reliability analysis

Analogically to the Experiment I, we started with an analysis of the listeners' reliability. Again, we computed both Cronbach's alfa and intra-class correlation (ICC). However, since people rated road traffic, trams' and aircraft noise separately, as well as the general disturbance and annoyance (total annoyance), we conducted both tests separately for each scale, because then we would use them separately in models of total annoyance. So, every test was launched five times to give five different tables of values. Then, we averaged results and based on means decided who should be excluded from further research.

When analyzing Cronbach's alfa test, we focused on corrected r values which describe how well the listener correlates with the rest of the sample group. We were particularly interested in cases where given respondents were negatively correlated with the others. On the other hand, in the analysis of ICC values we focused on 'when an item dropped' option. The higher value of ICC for a listener, the less reliable he/she is.

Table 11 shows the mean values calculated for each listener in relation to the results of both tests. Please note that the initials of respondents are ranked in terms of the value of the corrected r ('r.cor' in the table).

It can be clearly seen that two respondents were negatively correlated with the rest – so excluding them from further analysis was quite obvious. However, no matter which test we choose, after sorting out the initials the same four respondents would always be at the top: 'SZ', 'AB', 'KSz' and 'MG'. These four people have also the lowest values of corrected r, below .1. That is why we decided to exclude all of them from further analyses. Finally, results of 27 respondents were analyzed.

Table 11. Mean values of corrected r and ICC computed for five scales separately for respondents in the Experiment II. Respondents are ordered in relation to values of corrected r.

Respondent	**SZ**	**AB**	**Ksz**	**MG**	**Abi**	**DM**	
r.cor	-.1013	-.0486	.0698	.0724	.1180	.1202	
ICC	.0533	.0516	.0524	.0538	.0487	.0503	
Respondent	**PP**	**AC**	**JLo**	**MHu**	**ON-A**	**HP**	
r.cor	.1253	.1359	.1409	.1539	.1683	.1933	
ICC	.0513	.0503	.0493	.0489	.0512	.0479	
Respondent	**AP**	**Jsz**	**LS**	**APa**	**JW**	**MS**	
r.cor	.2099	.2115	.2121	.2967	.3037	.3186	
ICC	.0493	.0516	.0478	.0479	.0477	.0495	
Respondent	**EZ**	**MBei**	**RF**	**MKK**	**NB**	**PZ**	
r.cor	.3553	.3762	.3838	.3908	.3972	.3973	
ICC	.0478	.0437	.0458	.0466	.0463	.0458	
Respondent	**JL**	**MB**	**Mbu**	**Asch**	**MH**	**MW**	**MP**
r.cor	.4169	.4393	.4597	.5229	.5525	.5698	.6911
ICC	.0423	.0433	.0427	.0438	.0409	.0416	.0373

3.2.2 Correlation between total annoyance ratings and other parameters

The approach used here was the same as in the Experiment I. First, we focused on finding correlations between means of total annoyance ratings (or partial annoyances ratings) and mean values of other parameters. As there was 12 different recordings, each vector had 12 values (values were averaged for each recording, over all participants).

Statistically significant results of Spearman correlation tests can be found in Table 12. As one can see the only one statistically significant correlation for total annoyance ratings was found between them and road traffic annoyance ratings (p = .0068 and rho = .7316).

For trams' annoyance ratings there are four statistically significant correlations. Two of them are related to sound level / loudness (level vs time 95% with p = .0001 and rho = .8932, loudness vs time 20% with p = .0004 and rho = .8581). Also two psychoacoustical characteristics correlate well with trams' annoyance: roughness vs time 95% (p = .0053, rho = .7461) and fluctuation strength vs time (averaged over the whole recording duration, p = .0276 and rho = .6316). Definitions of those parameters were also described on pages 65-66.

Finally, there are also two statistically significant correlations between aircraft annoyance ratings and objective parameters: fluctuation strength vs time 30% (p = .0009 and rho = .9286) and sharpness vs time 5% (p = .0102 and rho = .8333).

Table 12. Results of Spearman and polyserial correlation tests between means of partial and total annoyance ratings and different mean values of objective parameters. A value for polyserial rho for the relation between TA and road traffic PA was not computed, as both variables are continuous.

PA/TA ratings	Parameter	p	rho
Aircraft	**Fluctuation Strength vs time 30%**	.0009	.9286
Trams	**Level vs time 95%**	.0001	.8932
Trams	**Loudness vs time 20%**	.0004	.8581
Aircraft	**Sharpness vs time 5%**	.0102	.8333
Trams	**Roughness vs time 95%**	.0053	.7461
Total	**Road traffic PA**	.0068	.7316
Trams	**Fluctuation Strength vs time**	.0276	.6316

However, as the Spearman correlation test base on ranks, we do not know if these correlations are linear or not. To find it out we created point plots (with errorbars) showing all statistically significant correlations. They are presented in relation to two different PA ratings (trams and aircraft, as no statistically significant correlations were found in the case of road traffic) in Figure 33 and Figure 34. They will be discussed in the 'Analysis' section.

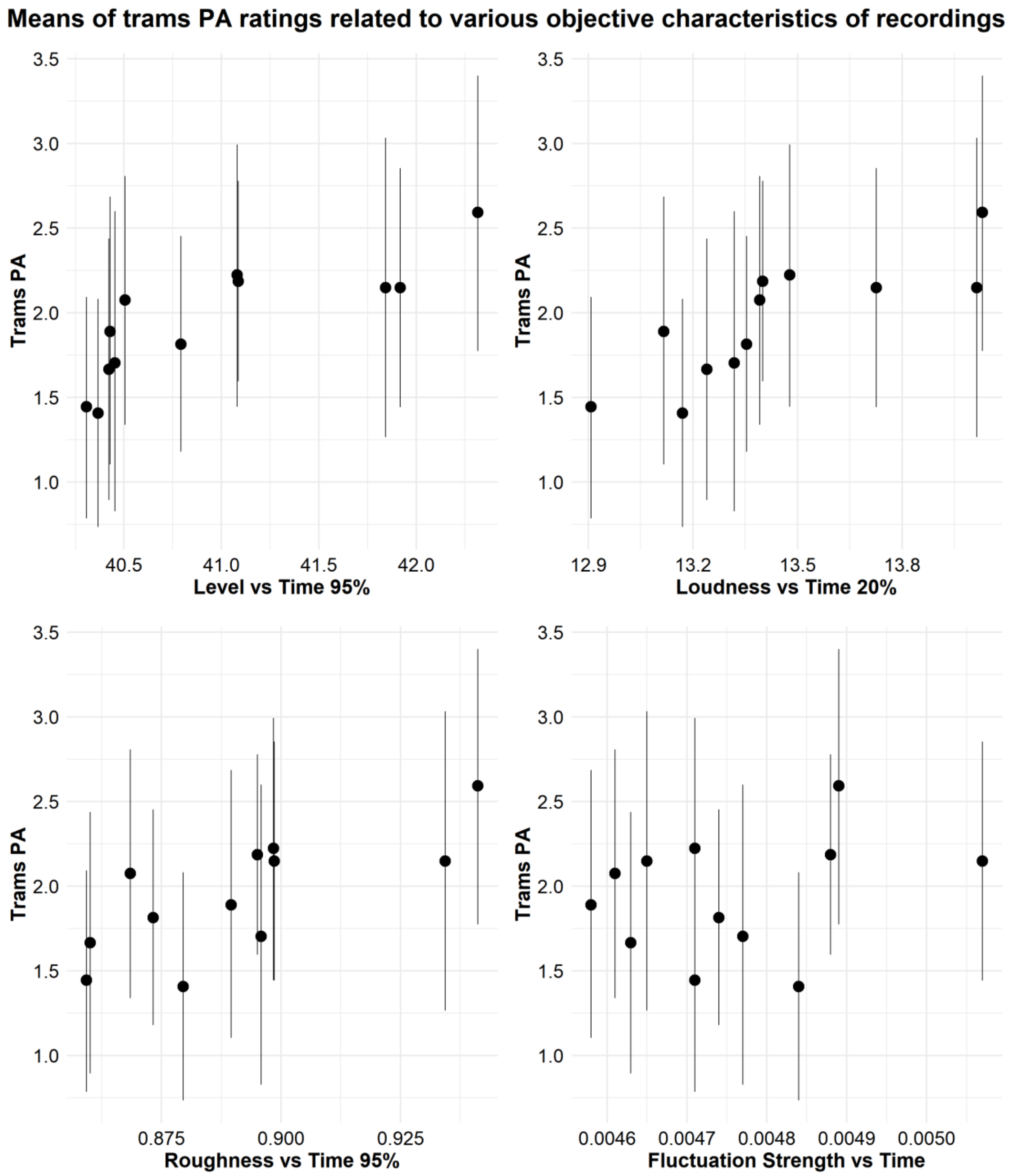


Figure 33. Correlation plots for mean of trams PA (with 95% CI) and objective parameters which were found to be statistically significant in the Spearman correlation test: level versus time 95%, loudness versus time 20%, roughness vs time 95% and fluctuation strength versus time.

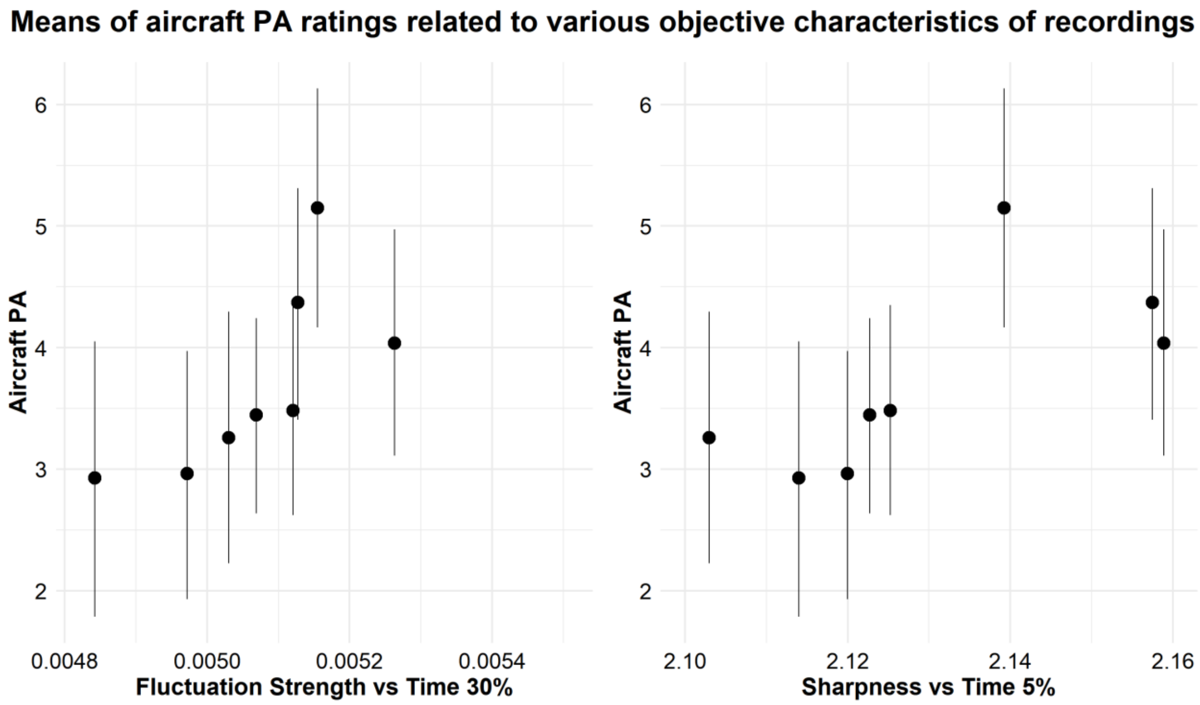


Figure 34. Correlation plots for mean of aircraft PA (with 95% CI) and objective parameters which were found to be statistically significant in a Spearman correlation test.

3.2.3 Verification of existing total annoyance models

Before creation of any new model for prediction of total annoyance (TA), it is worth to analyze already existing models. As it has been mentioned in the section 1.5, several TA models were already established. However, not all of them could be applied to this research. We could not use the vector summation model as it allows only to predict TA based on two different noise sources. It was also impossible to use Vos weighted summation model because sound level of road traffic noise in our scenarios were constant.

Finally, we checked participants' answers with regards to three different TA models: strongest component model (SCM), synergistic model (SM) and combined noise source paradox (CNSP). These three models cover all possible relations between TA ratings and partial annoyances. Strongest component model states that TA is equal to maximum partial annoyance. Synergistic model states that TA is greater than maximum partial annoyance and combined noise source paradox assumes reversely: TA is lower than the maximum partial annoyance.

In our raw data we had 324 records. Of these, only 17 (~5.25%) met the conditions of SM. 160 cases (~49.38%) represented SCM and 147 cases (~45.37%) fulfilled requirements of CNSP. Since 17 cases are very few, the analysis on their basis could be burdened with large errors. Since both SCM and CNSP clearly dominate our data, we then used them for further analysis and comparisons.

3.2.4 Influence of trams and aircraft PA ratings on TA assessment

Our first concern was to check whether the number of trams (1, 2, 4 or 8) and/or aircraft (0, 1 or 2) is a statistically significant factor influencing the assessment of TA. Because the distribution of TA answers was not normal (Liliefors test gave D = .125 with p = 2.791 * 10^{-13}), we again used robust ANOVA procedure from the package 'walrus'. This time we computed it for the whole data, as well as separately for SCM and CNSP cases. In Table 13 and Table 14 main ANOVA results as well as post-hoc tests for a significant factor are shown.

It can be clearly seen that number of trams has no statistically significant influence on total annoyance ratings (F = 1.6089 and p = .662). The interaction between both factors (number of trams and number of planes) also does not have any statistical significance (F = 4.685, p = .602). The only one statistically significant factor is number of planes (F=8.6573, p = .015).

Results of post-hoc tests (Table 14) revealed that only relations between '0' and other cases ('1' or '2') are statistically significant (for 0-1 ψ = -2.72, p = .0215 and for 0-2 ψ = -3.16, p = .0058). Cases in which 1 aircraft was presented do not differ statistically significant from cases with 2 aircraft (ψ = -.44, p = .6871). Please note that *ψ* is the corrected difference between groups and it will be shown also in the other post-hoc tests results.

Table 13. Results of robust 2-way ANOVA computed on the data with number of trams and planes as factors and total annoyance ratings as dependent variable.

Factor	F	p
Trams number	1.6089	.662
Aircraft number	8.6573	.015
Trams number:Aircraft number	4.6850	.602

Table 14. Results of post-hoc tests for number of planes as a factor in the combined noise sources data.

Number of planes		ψ	p	CI low	CI high
0	1	-2.72	.0215	-5.5265	.0865
0	2	-3.16	.0058	-5.8684	-.4516
1	2	-0.44	.6871	-3.0487	2.1687

After dividing the whole data into two subsets – according to the working TA model – we get another interesting results.

Results for SCM are presented in Table 15. As one can see neither number of trams, nor aircraft had a statistically significant impact on people's TA judgments when they rated it accordingly to the SCM model.

Table 15. Results of robust 2-way ANOVA computed on the SCM 'mixed noise sources' data with number of trams and planes as factors and total annoyance ratings as dependent variable.

Factor	F	p
Trams number	.1925	.980
Planes number	4.4290	.119
Trams number:Planes number	3.1366	.810

Finally, we present results of ANOVA analysis for CNSP data (Table 16 and Table 17). Again, as it was shown for the whole data, only number of planes is a statistically significant factor regarding its influence on TA ratings.

Post hoc tests (Table 17) revealed statistically significant difference only between groups with 0 and 2 planes (ψ = -1, p = .001). However, for the 0 vs 1 case, results are very close to the significance threshold of .05 (ψ = -.6875, p = .0621).

All these findings suggest that trams do not contribute to TA ratings while aircraft do.

Table 16. Results of robust 2-way ANOVA computed on the CNSP 'mixed noise sources' data with number of trams and planes as factors and total annoyance ratings as dependent variable.

Factor	F	p
Trams number	3.1555	.388
Aircraft number	11.2736	.006
Trams number:Aircraft number	2.1760	.913

Table 17. Results of post-hoc tests for number of planes as a factor in CNSP data.

Number of planes		ψ	p	CI low	CI high
0	1	-0.6875	.0621	-1.5669	0.1919
0	2	-1.0000	.0010	-1.7133	-0.2867
1	2	-0.3125	.3240	-1.0765	0.4515

3.2.5 Differences between SCM and CNSP data subsets

As the data gathered in the Experiment II can be clearly divided into two groups regarding working TA models (CSM and CNSP), we also wanted to compare both subsets and see if there were any statistically significant differences between them regarding other parameters. We computed bootstrapped robust version of independent samples t-test using a function *t1waybt* from the WRS2 package. The grouping variable was ‘Model’ (dichotomous, ‘SCM’ or ‘CNSP’) and dependent variables were different, because we calculated such tests for each parameter that could describe people (results of surveys filled in by respondents) or recordings. Please note, that parameters with nominal or dichotomous nature were not taken into account in this analysis.

Results of these tests – including only those parameters which gave results with p < .05 – are shown in Table 18. Both trams and aircraft PA ratings were statistically different (t = 13.8452 with p = .0003 and t = 33.9763 with p < .0001 respectively). 6 different parameters describing people involved in the experiment were also statistically significant:

- Age with t = 14.0608 and p = .0003
- Habitation noise sensitivity with t = 4.3617 and p = .0375
- Work noise sensitivity with t = 5.1642 and p = .0241
- Answers about attitude towards noise and its source: Q7_1 (statement about noisiness of the neighborhood) with t = 9.0129 and p = .0041; Q7_2 (statement about carrying for health effects of noise) with t = 5.2648 and p = .02) and Q7_10 (statement about carrying for value of an apartment/house in the noisy environment) with t = 10.2181 and p = .002
- One physical parameter of sound: sharpness vs time 5% with t = 4.92 and p = .0259

In Table 18 we also provide information about means computed for SCM and CNSP data separately as well as variance explained by each parameter and its effect size (ES). It can be clearly seen that the biggest ES has planes' annoyance (.4499). Values above .2 are also given by: age (.2988), trams' annoyance (.2946), Q7_10 (.2533) and Q7_1 (.2365). While taking into account % of a variance explained, the order would be the same.

Table 18. Results of independent bootstrapped t-tests computed for all 'non-annoyance' parameters characterizing people or recordings between subsets of data selected accordingly to the working TA model (SCM or CNSP).

Parameter	Mean value		t value	p value	variance explained	effect size
	SCM	CNSP				
Trams PA	1.5563	2.4218	13.8452	.0003	.0868	.2946
Planes PA	1.6250	3.4422	33.9763	<.0001	.2024	.4499
Age	29.3313	26.2993	14.0608	.0003	.0893	.2988
Habitation	1.5016	1.4303	4.3617	.0375	.0286	.1692
Work	2.1547	1.9932	5.1642	.0241	.0323	.1798
Q7_1	2.8063	2.3742	9.0129	.0041	.0559	.2365
Q7_2	3.7125	3.4286	5.2648	.0202	.0326	.1807
Q7_10	3.4438	3.0476	10.2181	.0020	.0642	.2533
Sharpness vs time 5%	2.2753	2.2811	4.9200	.0259	.0318	.1783

3.2.6 Establishing new models to predict TA

The main objective of this Experiment was to create a reliable TA model that would allow us to predict it based on other parameters.

Before both experiments we had asked people about their experience and attitudes related to noise. We had also computed many objective characteristics of recordings. Having so much data, we decided to establish several models based on three different groups of parameters:

- First group consists of all partial annoyance ratings, i.e. road traffic, trams' and aircraft noise annoyance ratings; model named **PABaM**
- Second group is based solely on individuals' parameters such as age, noise sensitivity or answers to the statements from the question 10; model named **PeCBaM**
- Third group covers all objective parameters such those related to loudness/sound level, spectral/temporal characteristics of recordings and so on; model named **ReCBaM**

In section 1.5 we pointed out that some attempts to establish TA models rely on multilinear regression. That is why this choice was also used by us. However, as one can argue whether the relationship between TA and other characteristics is linear, we also focused on more sophisticated methods. Finally, we decided to use random forest method (Breiman, 2001).

When we want to establish a reliable model, it is very important to remember to divide the data into two subsets: one for training and one for checking. If all data is used as an input to the model, there is a high risk that the model would ideally replicate itself, but powered by new data, its performance would be poor. Therefore, we divide our data into these two subsets in the proportion of 50% of observations for each one.

First, we set up a random forest (RF) model. RF consists of many decision trees which are aimed to split the data into classes (classification procedure) or to predict value at the output (regression). RF was created and tuned using a *caret* package in R. The model was fed with a training set and checked using another data subset. The predicted value was TA rating and predictors were all numerical variables which described both individuals and recordings.

The output of RF gave us a highly sophisticated structure. As we wanted to confirm the stability of the created forest, we did the analysis couple of times. Unfortunately, every time the structure of the forest was completely different – meaning that the output of the function is not stable and cannot be treated as a reliable model. That problem appeared even when a model was fed with the largest possible dataset. Reasons of this failure will be discussed in the 'Analysis' section.

Then, we paid our attention to the MLR models. Again, the data was split into two subsets. However, to avoid instability of output models, we used multiple replication of models. It means that the original data was 10,000 times divided randomly into training and checking subset and they were used to create and then check of 10,000 MLR models. Then, models were assessed using averaged values of assumptions' checking tests and models' parameters.

MLR models were computed separately for three abovementioned groups of parameters for the whole data. The data was also separated into SCM and CNSP variants, however both subsets were only used to check the RMSE of the new TA model.

Before multiple replication of MLR, for every type of relation we used *step* function (including all parameters from a given group) to find the best possible set of parameters which could best explain the variance of the TA ratings. Then, best models were used as formulas in the *lm()* function (which computes **l**inear **m**odels provided with their statistics) and replicated 10 000 times.

This procedure of optimization revealed that prediction of TA based on objective parameters (ReCBaM) is useless. For the whole data, this model explained only 2.46% of variance. For SCM and CNSP it was even impossible to find a statistically significant model – the function returned on the output only a constant relation TA~1. That is why we did not incorporate such models in the replication phase. More details about this could be found in the "Analysis" section.

Our aim was to establish MLR models for the whole data – focusing on partial annoyance ratings (PABaM) and subjective parameters (PeCBaM). The analysis revealed that only road traffic and aircraft partial annoyance had statistically significant influence on TA ratings. Results of PABaM model can be found in Table 19.

As one can see, for both road and aircraft traffic PAs, p is <.0001. Coefficients' mean values are .7907 and .1517 respectively. Intercept value is .4714 with p = .0108. The last provided parameter is variance inflation factor (VIF). VIF provides information about an increase of explained variance which is the effect of collinearity between several predictors. A rule of thumb says that VIF greater than 5 implies high collinearity and the model cannot be classified as reliable (Sheather, 2009). In the case of PABaM, VIF is equal to 1.1284, meaning that there is no collinearity between two predictors.

Table 19. Mean values of basic statistics (including coefficients) computed using 10,000 replications of the PABaM model.

Statistic	Location Statistics	Predictors		
		Intercept	Road Traffic PA	Aircraft PA
Coefficient's value	95% CI Low	.2500	.7267	.1112
	Mean value	.4714	.7907	.1517
	95% CI High	.6801	.8542	.1926
P-value	95% CI Low	<.0001	<.0001	<.0001
	Mean value	.0108	<.0001	<.0001
	95% CI High	.0848	<.0001	<.0001
VIF	95% CI Low	NA	1.0546	1.0546
	Mean value	NA	1.1284	1.1284
	95% CI High	NA	1.2260	1.2260

Analyzing subjective characteristics, the analysis revealed 9 statistically significant predictors (including intercept). From ten statements which were included in the preliminary survey for respondents – regarding their opinions and attitude towards different aspects of noise sources – six were found to be statistically significant. An important variable was also annoyance perceived by listeners in their places of residence as well as (what could be quite surprising) the age of a person. Results of this analysis are shown in Table 20.

As one can see from a Table 20, both age and home annoyance are statistically significant: age with .0495 and $p = 0.265$, and home annoyance with .3125 and $p = .0002$. Intercept has also a significant influence on TA ratings, with a value of -8.8606 and $p < .0001$. As it was already mentioned, six statements from a preliminary survey are statistically significant:

- Statement 1 (Q7_1, "Neighborhood of my place of residence is noisy") with .6893 and $p < .0001$
- Statement 4 (Q7_4, "I do not like travelling by airplanes") with .2529 and $p = .0137$
- Statement 6 (Q7_6, "I prefer to use a car instead of other means of transport") with .4572 and $p = .0001$
- Statement 7 (Q7_7, "I consider myself to be a proecological person") with 1.0514 and $p < .0001$
- Statement 8 (Q7_8, "My house/apartment is well-isolated from outside noise sources") with .7561 and $p < .0001$
- Statement 9 (Q7_9, "Noise produced by my neighbors / coming from a staircase annoys me at home) with .3426 and $p = .0006$

It is also worth to mention that none of the predictors has VIF greater than 3 (even when considering upper 95% limit of confidence intervals).

Table 20. Mean values of basic statistics (including coefficients) computed using 10,000 replications of the MLR model on the whole data between TA and subjective characteristics.

Statistic	Location Statistics	Predictors				
		Intercept	Age	Home annoyance	Q7_1	Q7_4
Coefficient's value	95% CI Low	-10.4784	.0219	.1972	.4776	.1216
	Mean value	-8.8606	.0495	.3125	.6893	.2529
	95% CI High	-7.2713	.0769	.4262	.9110	.3857
P-value	95% CI Low	<.0001	<.0001	<.0001	<.0001	<.0001
	Mean value	<.0001	.0265	.0002	<.0000	.0137
	95% CI High	<.0001	.2180	.0010	.0001	.1133
VIF	95% CI Low	NA	1.4550	1.5535	2.1358	1.0863
	Mean value	NA	1.7117	1.7444	2.4872	1.1691
	95% CI High	NA	2.0646	1.9903	2.9243	1.2870
Statistic	**Location Statistics**	**Predictors**				
		Q7_6	**Q7_7**	**Q7_8**	**Q7_9**	
Coefficient's value	95% CI Low	.2979	.7898	.5692	.2217	
	Mean value	.4572	1.0514	.7561	.3426	
	95% CI High	.6126	1.3119	.9505	.4597	
P-value	95% CI Low	.<0001	<.0001	<.0001	<.0001	
	Mean value	.0001	<.0001	<.0001	.0006	
	95% CI High	.0004	<.0001	<.0001	.0047	
VIF	95% CI Low	1.2034	1.7396	1.3732	1.1077	
	Mean value	1.3800	2.0677	1.5578	1.2140	
	95% CI High	1.6123	2.4765	1.8074	1.3707	

The regression analysis has many assumptions to be met in order for a given model to be reliable and stable. A group of tests to check them were already described in the section 3.1. Results of analyses made for both MLR models can be found in a Table 21.

Based on values of adjusted R^2 one can see that the best fit is for MLR model computed using PA ratings, R^2 = .8316). Lower values can be found for model based on subjective characteristics, but still being around .65, R^2 = .6499. Similar results were obtained regarding Bayesian Information Criterion (BIC) and Aikake Information Criteri (AIC). Both BIC and AIC are measures used in the model selection process. AIC represents the information lost by a model (comparing it to the real data). BIC works similarly, as it is derived from AIC, however it also computes a penalty which comes from multiple predictors used in a model (a risk of overfitting; AIC also incorporates this error, but its values are lower). BIC and AIC values show that the best model was MLR using PA ratings (BIC = 391.4236, AIC = 379.2758). Root mean square error (RMSE) computed between real values of TA from checking data subsets and values predicted by models are also reported – showing again that the best model was MLR on PA ratings (RMSE = .8336).

When analyzing results of assumptions checking tests, only p-values are reported in Table 21. As one can see, in all cases three tests (Goldfeldt-Quandt, Durbin-Watson, Harvey-Collier) gave p values greater than .05 – meaning that conditions were fulfilled. However, when analyzing normality of residuals (using Lilliefors test) there is the case when p is lower than .05: for the MLR model computed using PA rating as predictors (p = .0313).

Table 21. Results of test describing goodness of fit of both MLR models as well as results of tests checking different MLR assumptions.

Statistic	Location Statistics	PABaM	PeCBaM
P-value of a model	95% CI low	<.0001	<.0001
	Mean value	<.0001	<.0001
	95% C high	<.0001	<.0001
adjusted R^2	95% CI low	.7718	.5697
	Mean value	.8316	.6499
	95% C high	.8905	.7247
AIC	95% CI low	317.4251	416.2437
	Mean value	379.2758	456.1832
	95% C high	426.6936	491.0495
BIC	95% CI low	329.5729	445.8920
	Mean value	391.4236	485.9619
	95% C high	438.8414	520.9104
RMSE	95% CI low	.6904	1.2777
	Mean value	.8336	1.4264
	95% C high	.9728	1.5773
P-value of a Goldfeldt-Quandt test	95% CI low	<.0001	.0321
	Mean value	.4993	.4988
	95% C high	1.0000	.9711
P-value of a Durbin-Watson test	95% CI low	.0240	.0254
	Mean value	.5037	.4978
	95% C high	.9693	.9754
P-value of a Harvey-Collier test	95% CI low	.0173	.0208
	Mean value	.5050	.4984
	95% C high	.9739	.9810
P-value of a normality test of residuals	95% CI low	<.0001	.0010
	Mean value	.0313	.1534
	95% C high	.2512	.6216

To have a better view into goodness of fit of each MLR model, histograms of all replicated adjusted R^2 values were drawn. They are presented in Figure 35. Also confidence intervals are marked with dark green dotted vertical lines, mean values are provided too (red solid line).

Figure 35 shows that in both cases values of R^2 are distributed similarly to the normal distribution.

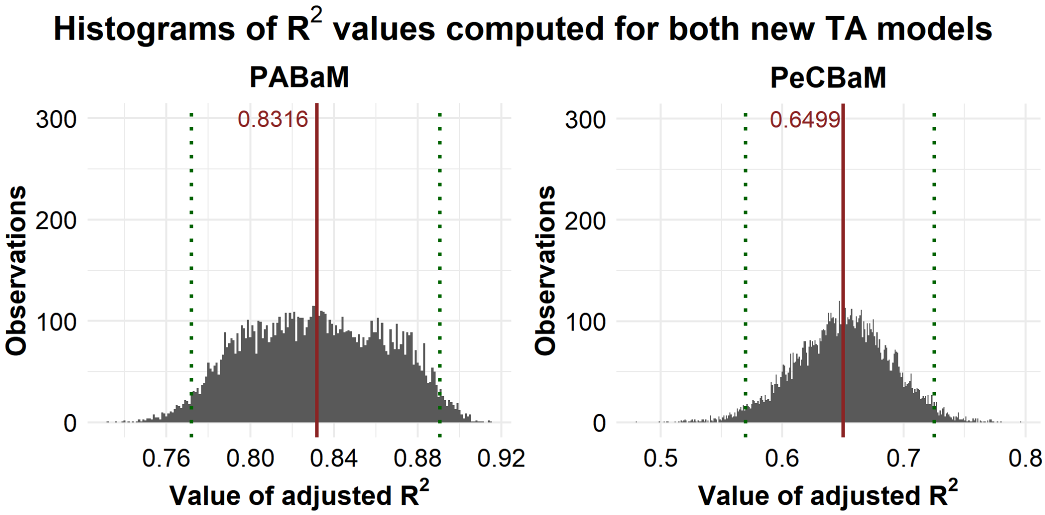


Figure 35. Histograms of adjusted R^2 computed for 10 000 replications for each MLR model. Dotted green lines show upper and lower CI limits; red solid line indicates mean value of R^2 (exact value rounded to the 4 decimal digits is also provided).

To sum up, below we provide equations for both established MLR models of TA. Please note, that PA_{RT} mean PA of road traffic noise and PA_A means PA of aircraft noise. Every statement from question 10 in the preliminary survey is denoted as $Q7_x$, where x is the number of a statement. *HomeAnn* is the annoyance of noise perceived by respondents in their places of living. Eq. 28 describes PABaM and Eq. 29 PeCBaM.

$$TA = 0.7907 * PA_{RT} + 0.1517 * PA_A + 0.4714 \quad (28)$$

$$TA = 0.0495 * Age + 0.3125 * HomeAnn + 0.6893 * Q7_1 + 0.2529 * Q7_4 + \quad (29)$$
$$+ 0.4572 * Q7_6 + 1.0514 * Q7_7 + 0.7561 * Q7_8 + 0.3426 * Q7_9 - 8.8606$$

4 Analysis

4.1 Experiment I

4.1.1 Correlation between road traffic noise annoyance and objective characteristics. Testing research hypothesis no. 1

Analysis of regression lines provide several remarks. First of all, in the case of road traffic noise only, equivalent sound level seems to be a good predictor of people's annoyance assessments. All coefficients and their values were already presented in Figure 31. We start the analysis with consideration of the regression between $L_{AEq,10m}$ and road traffic noise annoyance. Assuming that annoyance is equal to 0, we get a value of $L_{AEq,10m}$ to be ~35.7 dBA. It seems reasonable, because this is a common value of background noise in urban areas (between 30 and 40 dBA). Larger value is obtained when analyzing relation between Level versus time and noise annoyance, it is ~40.56 dBA.

Another finding is that psychoacoustical annoyance (PAnn), which is based on psychoacoustical parameters, correlates well with noise annoyance assessments. However, PAnn tends to give slightly smaller values than people actually perceive – and the mean difference is about 0.63. It seems that PAnn underestimates people's ratings of noise annoyance. As the ICBEN scale consists only of whole numbers, the error can be rounded to 1.

While talking about psychoacoustical parameters, the best fit is achieved for 5% roughness. As roughness corresponds to both amplitude and frequency modulation of sounds (with modulation frequency between 20 and 300 Hz), it can be interpreted that fast changes in sound has a significant influence on people's noise annoyance judgments. Roughness is generally associated with the impression of 'rough sound', which means that the changes occurring in it are too fast to be recognized directly, but still cause a certain feeling of change in amplitude or frequency.

Summing up, we can say that from all interesting parameters we can choose two and each of them corresponds to a different 'dimension' of a noise annoyance caused by road traffic noise. $L_{AEq,10m}$ represents the relation between noise annoyance and, generally speaking, feature of a sound associated with its volume. The second dimension is associated with changes in time-domain and is represented by roughness. All these relations are statistically significant and give high values of fit. As no MLR model was found to be statistically significant, there is no justification of application of multilinear regression models. When talking about road traffic noise annoyance, only one parameter is sufficient. Which of them should be used depends on the tools we have. It seems that the easiest to measure is sound level, i.e. $L_{AEq,10m}$. Both $L_{AEq,10m}$ and roughness give almost the same goodness of fit (~.7).

At this point it is worth to discuss the high percentage of explained variance by $L_{Aeq,10m}$. It was described in many papers that sound level can explain ~30% of annoyance ratings variance at maximum. Thus, there are opinions that A-weighted sound pressure level is not the best choice to describe people's reactions to noise. However, this lack of fit is observed mainly for *in situ* research. People are then asked to rate noise annoyance in relation to the last 12 months. It is difficult to do that, as humans tend to remember only several most influential situations. Many non-acoustical factors can also be significant but they cannot be controlled. On the other hand, in laboratory conditions people are exposed to sounds which are strongly limited in time. Moreover, the same conditions are preserved for all respondents. As the number of influential factors are limited, better results of correlation between sound level values and noise annoyance assessments can be found. In our research the variance in noise annoyance ratings is explained by sound level values at the level of ~70%. What is a reason?

In this research both sound level ($L_{Aeq,10m}$) and loudness give high and similar R^2 values, around .7. It was proved (Berglund et al., 1990) that when stimuli have different loudness, this characteristic determines the noise annoyance ratings – however, people are capable to distinguish annoyance from loudness (Dittrich and Oberfeld, 2009). The other influential factors can be revealed when the loudness of stimuli would be artificially adjusted to be the same. On the other hand, it is assumed that the A-weighted sound pressure level is a good predictor of loudness – but this is only true if spectra of stimuli are similar. Thus, it is not surprising that both loudness and $L_{Aeq,10m}$ appeared to be significant predictors of road traffic noise annoyance ratings.

However, another fact should be mentioned: the regression was established using only 18 data points and some of sound level values repeated several times. It means that from statistical point of view, this approach is limited and further research is needed to better understand the nature of this relationship.

As many objective psychoacoustical characteristics were found to be good predictors of noise annoyance (roughness, sharpness, fluctuation strength, loudness) we can confirm the first research hypothesis: it is true that road traffic noise annoyance ratings relate to psychoacoustical characteristics of stimuli.

4.1.2 Time pattern, proportion and sound level of stimuli related to the road traffic noise annoyance. Research hypothesis no. 2

In the Experiment I one of our goals was to find out whether different proportions between background noise and road traffic, while keeping the same $L_{AEq,10m}$, or different time pattern have an impact on people's judgments of noise annoyance. Results clearly showed that only road traffic sound level has a significant influence. It is also clear from Figure 32. For time pattern generally all cases when a whole scenario was adjusted to a given sound level are rated higher (with two exceptions) than those in which every road traffic package ('events') was adjusted separately. It should not be surprising as when e.g. there are 5 different road traffic packages, each with the same sound level of 60 dBA and the rest time in the scenario is filled with a background noise at 40 dBA, globally the equivalent sound level measured for 10 minutes would be lower than 60 dBA. However, since only the sound level is a significant factor, we can say that research hypothesis no. 6 was not confirmed: in this thesis there are no evidence that for the same $L_{Aeq,T}$ different noise annoyance ratings are observed in relation to different time structure of stimuli (**for cases when the proportion was set to 100%**).

Moreover, there are also no statistical differences between various noise scenarios which are classified to one of three possible 5-decibel intervals (i.e. 45-50, 50-55 or 55-60 dBA). It seems that these intervals, used in noise maps of a given area, are justified and it is not necessary to provide more detailed information as it does not impact people's noise annoyance ratings.

Another fact is that rising proportion (from .5 to 1) between road traffic noise and background noise does not always relate to the rising tendency of noise annoyance ratings. For lower sound levels ranges (45-50 and 50-55 dBA) we can observe something like a bias – mean ratings oscillate around the same value with small differences, sometimes even giving highest ratings for proportion of .5, not for 1. The situation becomes more 'organized' for the highest sound level of 55-60 dBA. In this case, .5 proportion has lower rating than .75 and .75 has the assessment lower than 1. Of course, those tendencies or differences are not so large to have any statistical significance. But can we find a reason why it is so?

To answer that question, we analyzed comments provided in an open question by our respondents. In this question we asked them to note some comments about their feelings, emotions or perception if they felt it was important. As it was an open question not many comments we found (70 for all 540 noise scenarios, giving ~13%). However, we manually segregated them into some groups and then those groups were classified in the way to represent either positive or negative attitude towards a given noise scenario. Both groups are described below:

- positive attitude: ‘almost unheard’, ‘far from the road’, ‘mainly peaceful’, ‘makes me sleeping’, ‘no annoying sources’, ‘peaceful traffic flow’
- negative attitude: ‘annoying steady noise’, ‘ music from a car’, ‘heavy vehicles’, ‘impossible to read’, ‘interrupting sudden noise’, ‘many different noise levels’, ‘road traffic intensity was high’, ‘speeding cars’

It has to be mentioned that one recorded car had a loud music turned on, so it was possible to notice it. Another thing is that when proportion was not 1, people had two opposite attitudes towards fragments with background noise. Several respondents tended to describe them as peaceful or like a far road while the others said that it had been annoying steady noise.

Thus, as two different attitudes could be observed, we decided to run a simple t test using this limited data (only observations with comments qualified to one of both groups) to find out if there is any statistically significant difference. Because groups were not equinumerous (49 negative comments and 21 positive), we used a *t1waybt()* function from ‘WRS2’ package which enables to compute bootstrapped version of t-test. 10,000 replications were used.

Results showed a clear difference between both groups in relation to values of noise annoyance ratings: $t = 33.87$, $p = 0$, variance explained at about 63% and effect size of 0.79. As the mean value of annoyance assessment for positive comments was 1.62 and for negative ones 4.04, it is clear that people with negative attitude gave higher values of annoyance than those with positive feelings.

It is even more interesting when we analyze number of negative and positive comments related to different proportions and sound levels values.

Table 22. Number of positive and negative comments about different noise scenarios in relation to road traffic sound level and proportion between road traffic noise and background noise.

		Sound level							
		Positive				**Negative**			
		45-50	**50-55**	**55-60**	**Sum**	**45-50**	**50-55**	**55-60**	**Sum**
Proportion	**.50**	6	5	1	12	5	10	12	27
	.75	3	0	0	3	3	4	6	13
	1.00	4	1	1	6	0	4	5	9
	Sum	13	6	2	21	8	18	23	49

It can be seen in Table 22 that if sound level increases, more negative comments occurs and fewer positive. On the other hand, the higher the proportion, fewer negative comments. This is probably related to the problem of intermittent noise, it is easier to adapt to loud, but steady noise, than to situation when huge differences in sound levels occur. In our opinion, this is also related to the Figure 32. When sound levels ranges are low (45-50, 50-55 dBA) people are mixed in their opinions, but when sound level is high enough, the proportion starts to influence their perception and it is also reflected in Figure 32.

It is obvious that negative comments should be more frequent while sound level increases. Thus, in our opinion, information from Table 22 are interesting mainly regarding number of positive/negative comments related to proportions between road traffic noise and background noise. This shows that people tend to have more negative attitude towards noise sources when they are presented in different ‘noise packages’ split by ‘quiet’ periods. Nevertheless, this tendency was not statistically significant (as has been shown in ANOVA results in Table 10).

The comparison of the results obtained in this research and in the work of Kaczmarek et al. (2010) shows some ambiguity. The authors found that the proportion was statistically significant while we did not. Possible reasons of it can be discussed. At first, Kaczmarek et al. presented all their stimuli with the same $L_{Aeq,T}$ value, so the sound level was not a factor in his work. In our research, we provided three different sound levels – possibly, this factor was the most important one and ‘covered’ influence of the other variables. Finally, Kaczmarek et al. used artificially prepared noise scenarios while we focused on a real environmental condition. It means that our scenarios were constructed from road traffic packages, not single pass-bys. This constant flow of road traffic may be a reason of no differences between different proportions and time patterns. It is possible that both these factors would have higher impact while the noise was perceived at night. However, this problem was not a part of this thesis.

All results discussed in this section do not let us to confirm the 2nd research hypothesis. Neither proportion nor the time pattern influence road traffic noise annoyance assessments.

4.2 Experiment II

4.2.1 Results of Spearman correlation tests between PA ratings and objective characteristics

In Table 12 some statistically significant Spearman correlations were found between different means of PA ratings’ and means of objective characteristics of recordings. Nevertheless, we stress that all objective parameters in Table 12 are computed for the whole recordings – i.e. not only road traffic but also trams and/or aircraft influence these measurements.

Trams PA ratings were found to correlate with parameters of two natures: related to amplitude (level vs time 95% and loudness vs time 20%) and to temporal structure of sound (roughness vs time 95% and fluctuation strength vs time). It can be interpreted that while presenting mixed noise sources to listeners, their perception of trams PA is influenced by the lower values of sound level (95% means sound level which is presented during at least 95% of the stimulus duration). On the other hand, temporal structure of the stimulus is also very important, regarding both fast and slow changes in amplitude and frequency.

Slow modulations are characterized by fluctuation strength. Fast modulations are expressed in roughness values. High correlation between roughness and annoyance ratings was also discovered by Klein in his PhD thesis (Klein, 2015). Moreover, he stated that best correlation values were observed for 95% roughness – similarly to the results presented in this study. Although correlation analysis is not directional (does not show which variable is dependent and which is a predictor), it seems reasonable to think that both slow and fast temporal changes in a sound structure can impact people's judgments of trams PA.

Aircraft PA correlated with two parameters: fluctuation strength vs time 30%, and sharpness vs time 5%. A fact, that 30% and 5% values of psychoacoustical parameters are enlisted here can point out that higher values of both characteristics have impact on people's judgments of aircraft PA. It seems logical as only 1 or 2 airplanes were presented to listeners, meaning that not more than during 2 minutes (out of ten) they were possible to be perceived. The value of sharpness is related to the amount of energy in the higher frequency bands. It can mean that when aircraft is present, energy in higher frequency bands increases and that defines annoyance of aircraft noise.

On the other hand, some drawbacks of this analysis can be found. From both Figure 33 and Figure 34 it can be seen that 95% confidence intervals of mean PA ratings are wide, meaning that variance of answers was also high. Additionally, points are scattered and it is difficult to fit them to a line or a curve. Moreover, there are only 12 (for trams, Figure 33) or 8 (for aircraft, Figure 34) points representing PA and some clear outliers can be seen. Thus, correlations were found to be statistically significant, but further interpretation of them is difficult. We recommend to analyze these relationships deeper in the other research.

4.2.2 Trams and aircraft PAs and their influence on TA ratings. Hypothesis no. 3, 4 and 5

One of our goals was to find out whether the number of pass-bys of trams or overflights of aircraft had statistically significant impact on TA ratings; appropriate results were shown in Table 13 and Table 14.

Number of aircraft was a statistically significant factor while number of trams was not. Since number of trams varied between 1 and 8, the result could be seen as surprising. We wanted to be absolutely sure that presence of tramways does not change TA ratings. Thus, we did also another analysis in which we incorporated additionally only road traffic noise used in stimuli

– so the ‘0 trams’ condition was also taken into account. This analysis, similarly to the previous one, showed no statistical significance of the ‘number of trams pass-bys’ factor. It means that not only there are no statistical differences between different number of trams but it even does not matter whether the trams are present or not.

It is worth to ask why there is no effect of number of trams on TA ratings? In our opinion, several reasons can be found. First, trams, as a railbound mean of transportation, can be described as less annoying than road traffic or aircraft. This phenomenon is well-described in literature for trains noise and is commonly called ‘railway bonus’. When various noise sources (among them trains) had the same sound level values, it is the railroad noise which would effect lower noise annoyance ratings in comparison to other noise sources. However, this relationship was established for ‘traditional’ railways – last findings show that this phenomenon disappears for high-speed trains. Nevertheless, as trams are slower than trains, it is possible that this mechanism works also in their case.

The second possible reason has, let’s say, spatial nature. As it has been already mentioned, one of the innovative aspects of this thesis is an application of the ambisonic technique. This method let us to keep all spatial information about sound source and allow to recreate it in the anechoic chamber. It means that both road and trams traffic – as they are ‘on the ground’ means of transportation – are presented in the same horizontal plane. As road traffic is far more intense than trams pass-bys, it is possible that trams are masked by road traffic and then they are less perceived or even not perceived at all. Actually, there were eleven observations when in the ‘comments’ section respondents answered that they had not perceived trams – however they had been presented to them. Please note that in the survey filled after each presented noise scenario there was no question about which noise source a respondent heard. Nevertheless, people were asked to provide all their feelings in the comments section – thus when we discuss sources which were distinguishable or not, we refer to the data from this field. Several times people mentioned that they had not heard trams – however, this type of noise was presented to them.

To analyze further a problem of masking trams by road traffic, we computed a 1/3 octave spectrum of each trams pass-bys. Additionally, aircraft and road traffic (one road traffic package) were also used. All spectra (both linear and a-weighted) are shown in Figure 36.

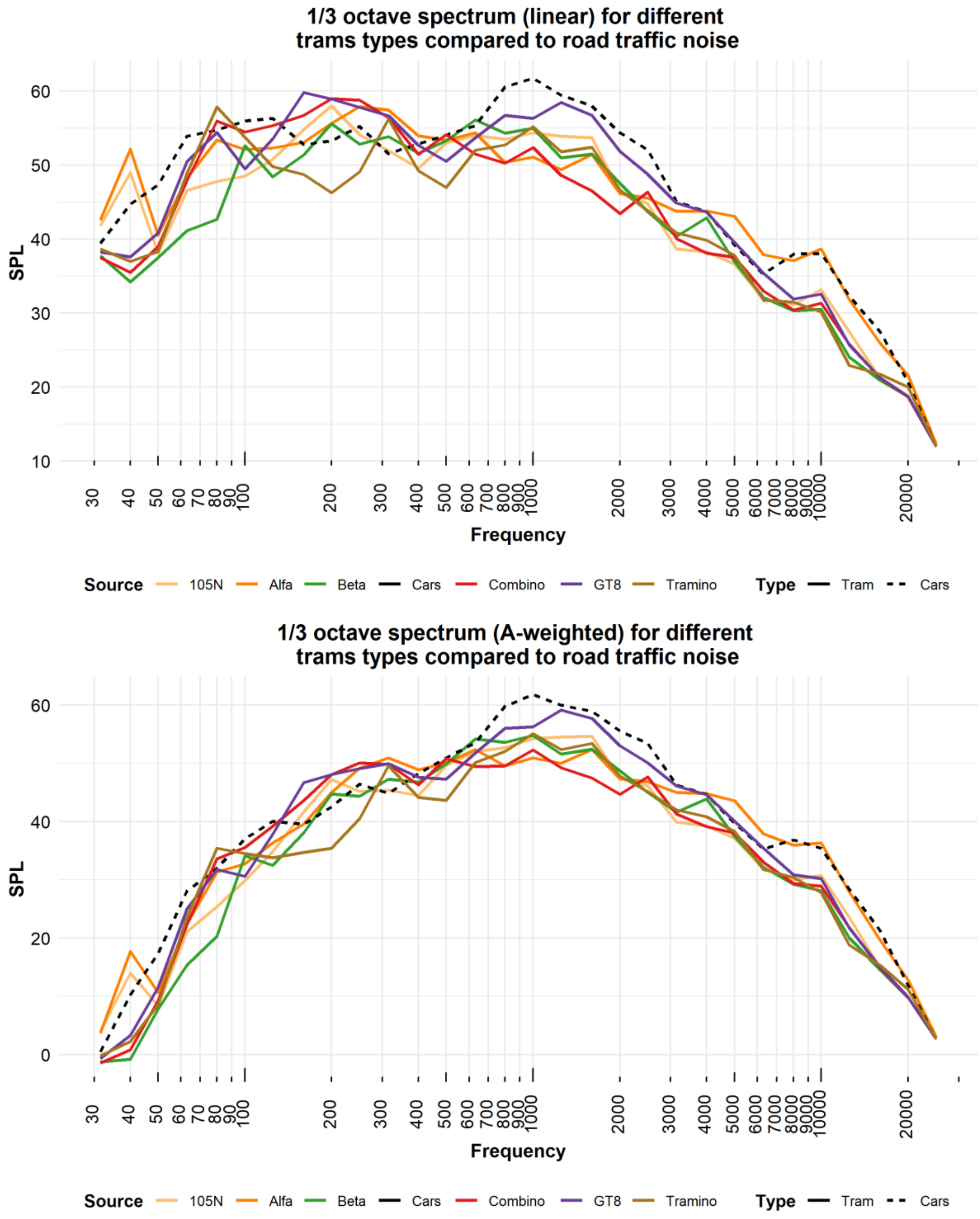


Figure 36. 1/3 octave spectrum (linear – the upper chart and A-weighted – the lower chart) computed for single pass-by of various trams types. For comparison purposes, road traffic spectrum is also provided.

As one can see, trams spectra are very similar to road traffic, especially in higher bands (from about 600 Hz and above). Their energy is slightly higher than for road traffic only between 120 and 400 Hz. No clear peaks were detected for high frequencies – meaning that there are no high-pitched tonal components (squeals). On the other hand one can observe, especially for a linear spectrum, some peaks around 40 and 80 Hz. This is probably related to some resonance in the anechoic chamber, as its suppress of echo starts roughly around 80 Hz.

All this findings can be interpreted that trams noise was simply masked by road traffic or mismatched with the road traffic. It can be also related to more psychological problem. People tend to have their own imaginations and expectations towards various noise sources. It was shown for example by Trollé et al., (2013) that people tend to perceive trams noise not only by sound level values, but also by their tonal components. As there was no squeals in our recordings, trams started to disappear from people's perception as they became similar to road traffic. Moreover, in the city of Poznań trams are considered to be generally efficient mean of public transportation – even when compared to buses. Since its reception is positive, it can also have an impact on people's judgments. However, it is difficult to check it in a quantitatively way. Eventually, it is not only the results of this study where PA of trams do not contribute do TA ratings – similar results were found by Klein et al., (2017). Nevertheless, our findings support a research hypothesis no. 4 that annoyance generates by trams do not contribute to TA ratings.

On the other hand, aircraft PA was found to be statistically significant (Table 13). Post-hoc tests revealed that both combinations between 0 and 1 or 2 were statistically significant.

These both findings – that trams PA does not influence TA ratings while aircraft PA does – can be deeper analyzed with the help of the answers gave by respondents to the question about the most annoying noise source. In the survey which was filled in by respondents after every mixed noise source scenario, there was a question about the most annoying type of noise. Only one type could be chosen. Results in relation to this question can be found in Table 23. Not only the whole data was analyzed, but also two subsets – regarding the perceptual model which was observed in people's ratings (SCM or CNSP).

Because road and trams traffic were presented in all mixed noise scenarios (12 different stimuli) and aircraft only in 8, we also provide the information about percent values, i.e. the absolute numbers of answers were related to the number of occurrences of a given noise source.

Table 23. Number of answers given for the question about most annoying noise source.

Data subset	**Most annoying noise source**		
	Road Traffic	**Tramways**	**Aircraft**
Whole data	199 / 62.4%	36 / 11.3%	84 / 38.0%
SCM	122 / 78.2%	14 / 8.9%	20 / 10.1%
CNSP	65 / 44.5%	20 / 13.7%	61 / 57.0%

For the whole data 199 times (62.4%) road traffic was most annoying, 84 (38%) – aircraft and only 36 times (11.3%) – trams. Thus, it seems that road traffic is the dominant source in most cases. Some people also found aircraft to be the most annoying but only 36 observations reported trams to be the worst source. It seems that trams were chosen in this question too rarely to provide any statistical significance as the factor in TA ratings.

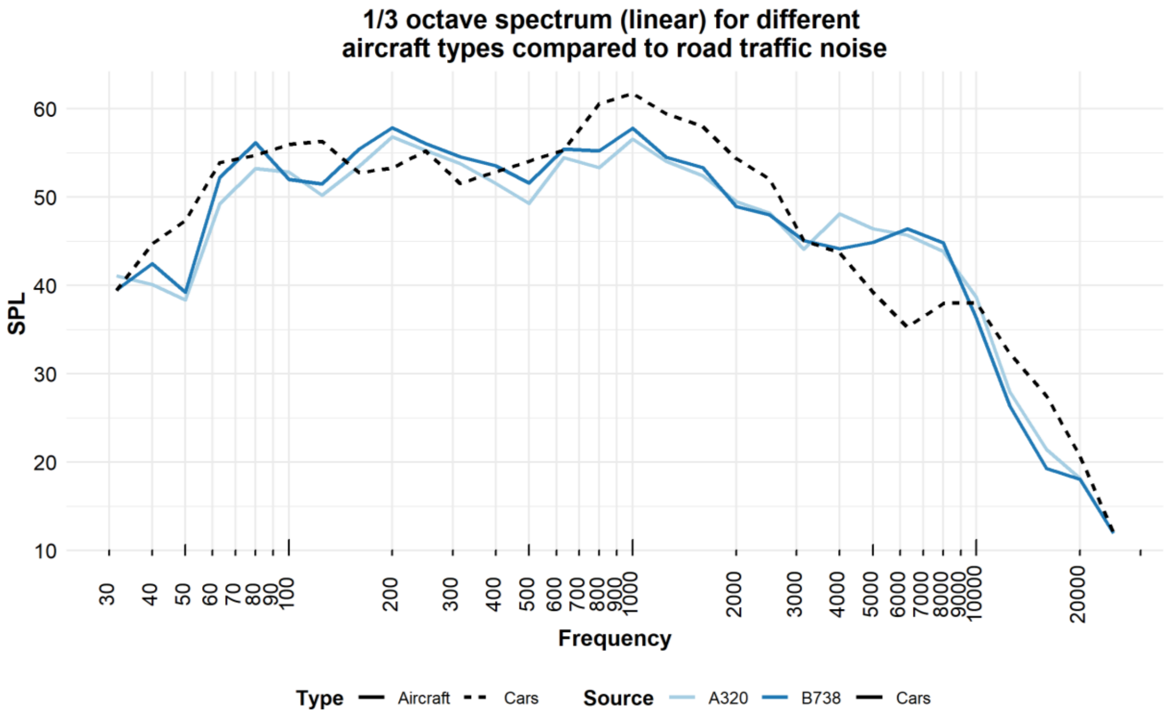
Additionally, we analyze the data in this case separately for both SCM and CNSP subsets. For the SCM data number of planes is not a statistically significant factor. Reasons of that can be possibly found in Table 23: for this subset 122 times (78.2%) road traffic was most annoying, only 20 times aircraft (10.1%) and 14 times trams (8.9%). We assume that the road traffic 'defines' TA ratings in most cases, number of trams or aircraft is not statistically significant as also their PA ratings are lower than road traffic PA assessments.

Finally, number of planes was also statistically significant for CNSP data. However, at threshold of $p<.05$, only relation between 0 and 2 is significant, not between 0 and 1 or 1 and 2. Again, this mechanism can be explained with results of number of most annoying noise sources in relation to number of aircraft. As one can see from Table 24, when number of planes increases, aircraft become more frequently assessed as most annoying (again we also provide percentile information for aircraft). Quite ambiguous situation is for 1 plane: both road traffic and aircraft were found to be the most annoying sources in 23 observations. This could be the explanation why only relation between 0 and 2 planes is statistically significant. Thus, the 5th research hypothesis can be confirmed – TA ratings increase when aircraft noise is presented in mixed noise sources scenarios.

Table 24. Most annoying noise sources regarding number of aircraft in the CNSP data.

Number of planes	**Most annoying noise source**		
	Road Traffic	**Tramways**	**Aircraft**
0	30	9	0
1	23	5	23 / 45.1%
2	12	6	38 / 67.9%

To confirm that finding, we present 1/3 octave spectra for both types of aircraft, i.e. A320 and B738 used in our experiment (compared with the road traffic noise, in both linear and a-weighted versions, Figure 37).



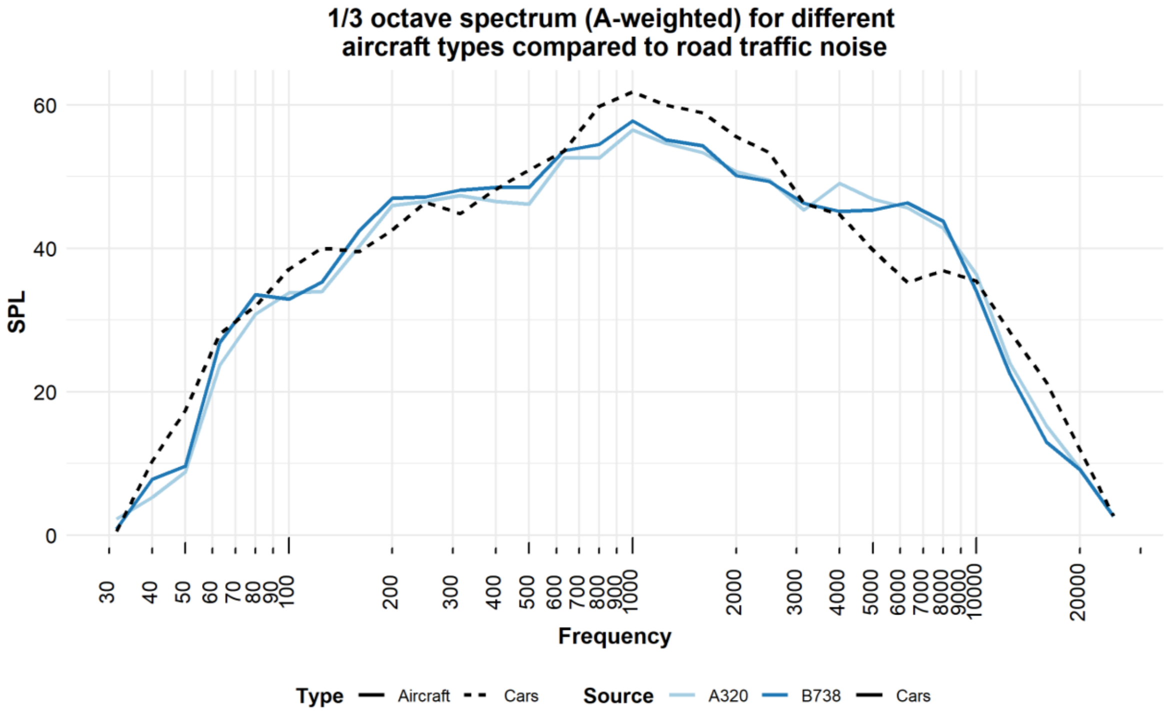


Figure 37. 1/3 octave spectrum (linear and A-weighted) computed for single overflights of two aircraft types. Road traffic spectrum is also provided.

It can be seen that aircraft and road traffic spectra are similar, but with one important exception: energy in bands between ~3000 and 10000 Hz is higher for both aircraft types than for road traffic – and the difference reaches its maximum around 6-8 dB for 6000 Hz. It can explain why aircraft is perceived by people even when a busy road traffic flow is simulated.

We have mentioned before that our assumption is that TA ratings are influenced only by these sources which are distinguishable and recognized by people. Thus, aircraft PA is a significant factor while trams is not. To find another proof for this mechanism we have also computed six basic objective characteristics for each trams' pass-bys and aircraft overflights – in different percentile variants, from 5% to 95%. The analysis revealed that biggest differences can be observed for two parameters: sharpness and tonality. Their 5% and 10% values, for all trams and aircraft compared to road traffic can be seen in Figure 38. Because sharpness and tonality have different scales, we standardized their values to facilitate the possibility of comparison between different noise sources.

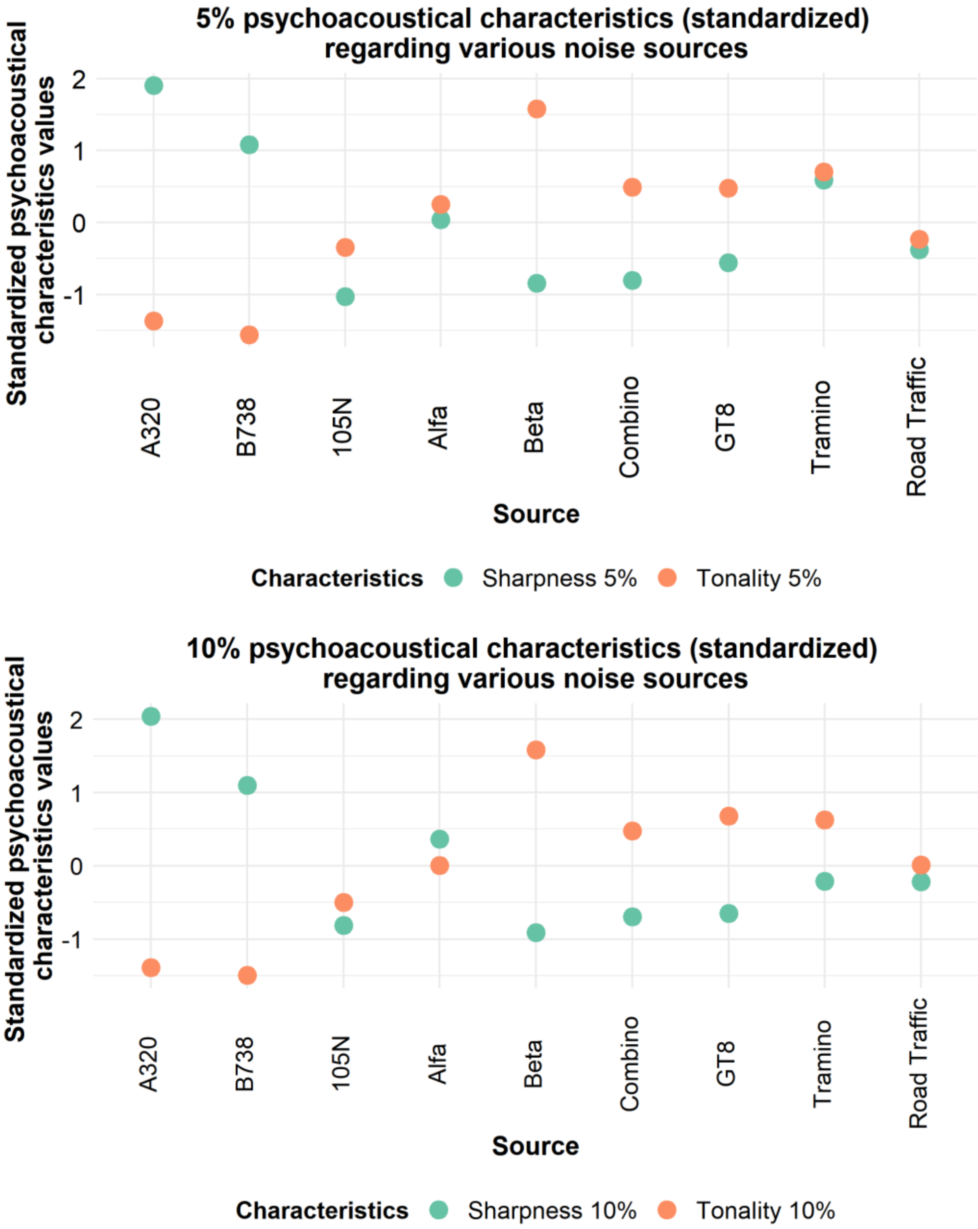


Figure 38. Standardized values of 5% and 10% sharpness and tonality for all trams, aircraft and road traffic.

For 5% and 10% variants of sharpness we can observe high values for both aircraft (A320 and B738) and lower for the rest of noise sources (with the exception for Alfa tram). On the other hand, tonality for aircraft is lower than for the other noise sources (this time, the exception can be observed for Beta tram, with very high value).

Based on these findings we can say that values of both sharpness and tonality differ between aircraft and the other noise sources. It is possible that the differences are so large that could be perceived by respondents in the other psychophysics experiment. As 5% and 10% variants of characteristics differ, it seems that mainly high values of sharpness and tonality determine people's perception of noise sources.

To sum up this section we can confirm research hypotheses no. 3, 4 and 5. It is true that TA ratings could be explained by PA ratings of different noise sources, however in this research only road traffic and aircraft PA were significant predictors while trams PA was not. Thus, it is also confirmed that trams PA do not contribute to TA ratings while aircraft PA do.

4.2.3 Differences between both SCM and CNSP TA models

The main goal of this thesis was to create a new TA model. However, as we have found that two existing models – SCM and CNSP – were observed in most of our respondents ratings, it is worth to analyze the data in the context of their performances.

Before interpreting results of the robust t test (Table 18), we want to direct readers' attention to the 'effect size' column in it. As one can see, in many cases values of effect sizes are low (below .2; this threshold was proposed by Cohen in 1988) meaning that even if the statistical significance was observed, the differences between mean values for given parameters are not so large.

In this case, we will focus only on these parameters which show effect size greater than .2. Having this criterion in mind, five different characteristics can be interpreted. Each of them will be discussed below:

- The biggest value of effect size was observed for aircraft PA. Mean value of it is 1.63 and 3.44 for SCM and CNSP data respectively. Based on it and results already shown in tables above (Table 23 and Table 24), this difference seems easy to explain. For SCM data most annoying noise source is road traffic (in majority of cases) – meaning that aircraft's impact on TA ratings is not large. However, for CNSP data there is clearly some ambiguity – that data could be split for almost equally-numbered groups where road traffic or aircraft is most annoying. That means that in CNSP PA of aircraft is far more important than in SCM data – and it is also reflected in PA ratings. Besides, mean of all PA ratings (for road traffic, tramways and aircraft taken together) is higher for CNSP (3.39) than for SCM (2.4). It can be interpreted that when people are confused which noise source is more annoying (because their PA are high and almost the same), CNSP occurred. However, when they are sure that only one noise source is dominant, PA of other noise sources is low (SCM works).

- Similar mechanism which was observed for aircraft (higher mean PA for CNSP dataset than for SCM dataset) is also observed for trams PA – but with lower effect size values. In our opinion, this phenomenon can be explained similarly as it was for aircraft case.
- A fact that 'age' factor is also significant can be surprising. In many papers it was showed, that age should not have significant effect on TA judgments. However, in this research a difference between mean age of respondents classified into both subsets is about 3 years (29.33 for SCM and 26.3 for CNSP). Why it is so? The simplest explanation seems to be the number of respondents: as only 27 people were included in the analysis of Experiment II, any difference in age could be statistically significant; particularly, when the range of age was large (between 21 to 53). In this case, more respondents should be included in the experiment to get more reliable results related to the age factor.
- Statistical significance was also revealed for the differences in mean value of Q7_10 factor. It is the statement from the question 10 (in the preliminary survey) which relates to the fear about negative influence of noise on the value of a property. As it can be seen from Table 18, mean value of answers to this statement is slightly higher for SCM data (3.44) than for CNSP (3.05) – meaning that people who were classified into SCM subset were more afraid than respondents who established CNSP data. In our opinion, explanation of this finding is not straightforward. It is worth to relate it also to the age factor. Thus, people in SCM data are also (regarding means) older than in CNSP data. It is reasonable to think, that elder people tend to be more frequently owners of houses/apartments than younger population (economic reasons play here a crucial role). When someone is just a tenant, it seems obvious he/she would be less afraid about the effect of noise on the value of a property than owners of it. However, unfortunately, we did not included in a preliminary survey a question about owning of a house/apartment. It seems that such a question should be asked, so in our future work we will include it in a questionnaire.
- The last factor which has an effect size value above the threshold of .2 is Q7_1. This statement is about noisiness of the neighborhood. It seems that people who rated TA according to SCM tend to more agree with the text 'My neighborhood is noisy' than people in CNSP group. This fact can be interpreted in many ways and, in our opinion, it is hard to find a simple solution. As an example of a possible explanation we provide a following consideration. It was already shown that people classified into SCM data

tend to give higher PA ratings to road traffic noise than those in CNSP dataset. Thus, there is a possibility that SCM people are exposed to higher road traffic sound levels in their neighborhood than CNSP respondents are. As we have already shown (Felcyn et al., 2018), percent of people who report to be annoyed by a given noise source correlate positively with the mean annoyance evoked by it. We checked how many times people reported in the preliminary survey in this research that they are annoyed by road traffic noise in their place of residence. For SCM data 102 times out of 160 this option was marked (63.75%) while for CNSP data it was 80 out of 147 (54.42%). Thus, it seems reasonable to think that higher percent values could mean higher noisiness of the neighborhood. However, this explanation is indirect and should be checked more carefully in other research.

All these findings were found to be statistically significant and have effect sizes above .2. On the other hand, it should be observed that variance explained by them are low – the largest, about 20%, was for aircraft PA, but for the other factors it is less than 10%: 8.93% for age, 8.68% for tramways PA, 6.42% for Q7_10 and 5.59% for Q7_1. It means that roughly only 50% of variance in the differences between both subsets can be explained by mentioned factors. It seems there are still other parameters which remain unseen for us.

4.2.4 MLR models (PaBAM and PeCBaM) for prediction of TA. Hypotheses no. 6 and 7

In this thesis an attempt to establish reliable models of TA was made. We decided to split the data into training and checking subsets, in the proportion of 50%-50%.

Since the data about noise annoyance and characteristics of both individuals and recordings are complex, we decided to use random forest to calculate a model of TA. We conducted several attempts using randomForest (RF model) package in R. Nevertheless, results were unstable and each time the structure of the forest was completely different. To be sure that this lack of stability existed, we recomputed the model while setting a seed (random numbers generator) to have the same value. This time a model was always the same – meaning that its structure depends on random numbers generator. Thus, it is not possible using our pooled data to generate a reliable RF model.

Possibly, the main problem is the relatively small dataset. Similarly to many machine learning techniques, RF needs much more data to construct a stable model. This time, since our data had no more than 400 observations (they were also split into learning and checking subsets), it was not enough to feed the algorithm. That is why applying such techniques in this research was impossible.

Before analyzing results of PABaM and PeCBaM, it is worth to consider really poor performance of ReCBaM, i.e. the model which bases on objective characteristics values. In our opinion, it can be explained in two possible ways.

First, all mixed-noise sources scenarios based on the same road traffic noise (presented with the sound level of 60 dBA in the proportion of 100%). It was presented in Table 3 that the differences of observed L_{AEq10m} values between all mixed-noise source scenarios were not higher than 1 dB. Such a low difference could be not observed by respondents – and is even within the range of a measurement error of the used sound meter. It means that all characteristics which base somehow on amplitude or volume cannot be significant in this experiment. The problem of low difference in the range of sound level is related to the structure of our noise scenarios. The mostly influential noise was road traffic noise. Trams and aircraft occurred only sometimes, meaning that even if their L_{AE} values were high (see Table 2), their impact on L_{Aeq10m} values was marginal. Of course, it does not mean that they were not perceived by listeners. However, road traffic noise is the most common in the city of Poznań and our goal was to simulate real urban conditions which can be met in Poznań.

On the other hand, it was shown in Table 12 that some objective characteristics related to sound level (level vs time) or loudness (loudness) correlate well with PA of tramways. However, they do not correlate with TA ratings. It can be interpreted that some changes in these characteristics have some impact on PA ratings, but then these are PA ratings which can predict TA (as it was shown in PABaM). Nonetheless, trams PA ratings were not statistically significant predictors of TA. Thus, eventually, it seems that in these experimental conditions neither sound level nor loudness has an influence on TA ratings. The low variability of sound level values (~1 dB) can be also a source of inconsistent relations between objective predictors and PA ratings which were revealed in Table 12.

4.2.4.1 Analysis of PABaM and PeCBaM performances

For PABaM both road and aircraft traffic PA ratings are significant factors which influence the assessment of TA. However, road traffic has bigger influence (coefficient value is .79) than aircraft with the coefficient of .15. Such results are not surprising as road traffic was the most dominant noise source in all mixed-noise sources scenarios. It is also the type of noise source which is the most frequently met in everyday life in urban areas. Thus, its PA ratings have the biggest impact on predicted TA values. We have also already shown that number of aircraft can influence TA ratings while number of trams does not. The latter observation can be surprising. However, we have shown in the section 4.2.2 that spectra of road traffic and trams noise are similar while the aircraft noise spectrum differs from them. It is the possible reason of the observed relations.

On the other hand, as it can be seen from Table 20, 9 different variables were found to be statistically significant (including intercept) for the PeCBaM. We discuss all characteristics in the following list. The statements which were included in the question 7 of the preliminary survey were described in the section 2.3. The following factors were found to be statistically significant in the PeCBaM:

- **Age:** this predictor has been already discussed in the section 4.2.3 and similar conclusions can be drawn here. However, it is worth to mention that the value of a coefficient of that factor is very small, .05 – meaning that with increase in the age also TA rating increases, but this influence is very weak (~.5 increase in TA for every 10 years increase in age). This is another reason to treat age as an influential factor with some criticism and this statistical significance should be revised in another experiments.
- **Home annoyance:** it is the characteristics which was measured in the preliminary survey using the question about noise annoyance perceived in the place of residence. That question was formulated using ICBEN numerical scale and their recommendations regarding content of the instruction to it. A value of the coefficient .31 suggests that while the perceived noise annoyance at home increases, TA rating also rises. It is rather not surprising as there are some research which confirms this tendency (not only in the relation to noise annoyance at home but also at a place of work: Bartels et al., 2015a; C. Marquis-Favre et al., 2005; Okokon et al., 2015b). It seems that people who are exposed to noise in their place of residence, give higher annoyance ratings to stimuli presented to them in laboratory conditions than those, who are not exposed daily to noise.
- **Q7_1:** this is the first statement from the question 7 from the preliminary survey and it was found to be statistically significant also in the previous section 4.2.3. Q7_1 is related to the same problem as the question about home annoyance – however, different scales were used (11-point numerical ICBEN scale for 'home annoyance' and 5-point bipolar Likert scale for Q7_1). Another thing is that meanings of questions are not the same: question about home annoyance was restricted to the annoyance perceived at home/apartment and Q7_1 relates to the noisiness of the whole neighboring area (also outside). However, to check out if both questions are correlated, we performed a Spearman correlation test which revealed some statistically significant correlation (rho = .21, p = .0001). Nevertheless, as the nature of both questions is different and the VIF values of all predictors in this model are low, we propose to include them both in the MLR model. The value of a coefficient for Q7_1 is .69.

- **Q7_4:** this statement reflects people's fears about travelling by aircraft. The higher rating in this question, the more afraid are people of flying. A positive value of this coefficient (.25) means that people who are more afraid by aircraft also give higher TA ratings. However, answers to this question do not correlate with PA ratings of aircraft noise (rho = -.04, p = .4324). This ambiguity may be surprising because it seems that people are transferring their fear of flying to the assessment of TA. Thus, the nature of this effect can be more complex as there is no direct relation between Q7_4 answers and aircraft PA
- **Q7_6:** statement no. 6 is related to the preference of using a car instead of other means of transportation in a city. People who 'more agree' with this opinion also give higher TA ratings – as the value of a coefficient is positive, .46. This could be not understandable with the first glimpse, as we tend to think that using a private car should be rather associated with higher comfort and then, maybe lower annoyance ratings. On the other hand, especially in Poland, people have many private cars, which lead to many traffic jams during rush hours. According to the General Statistical Department of Poland, in 2016 in Poznań 660 individual passenger cars were registered per 1000 inhabitants (Główny Urząd Statystyczny). Moreover, according to the TomTom Company (TomTom) in the rush hours an average time lost for staying at congestions in Poznań is 39 minutes per day, giving annually 139 hours. Our assumption is that these drawbacks of using a private cars can somehow influence people's judgments of noise annoyance. However, it cannot be simply proved and potential relationships between these phenomena go far beyond the scope of this experiment.
- **Q7_7:** that statement relates to the proecological attitude towards environment. It was proved (Okokon et al., 2015b) that people who care more about the state of the environment are also more annoyed by noise. This relation was confirmed in this research with the high value of coefficient for that factor (1.05). Actually, this is the highest value of all coefficients in this model – meaning that among all other Q7 statements in PeCBaM model it have the biggest impact on TA ratings.
- **Q7_8:** Another relation which can be seen as ambiguous. The eighth statement from the question 7 relates to the quality of acoustical isolation of a house/apartment from outside noise sources. The coefficient's value of .76 suggest that when people are more agreed with the statement that their building is well-insulated, they also rate given stimuli with higher TA ratings. It seems that better insulation should goes with lower TA ratings. Actually, it could be quite opposite. As one lives in the place which is silent, every exposition to the noise can be more annoying – because he/she is not used to such situations. On the other hand, people who are frequently surrounded be

noise adapt to some extent to it and other noises could be less annoying for them. From this point of view, this relation seems to be natural.

- **Q7_9:** The last statistically significant predictor is the statement about how much respondents agree that noise coming from neighbors/staircase annoys them at home. Again, the relation is positive with the coefficient's value of .34. However, comparing it with the previous question lead to another confusion: from one point of view, the better the acoustical insulation, the higher TA ratings. On the other hand, the more agreement that neighboring noise sources annoy inhabitants, the higher TA ratings. Why it is so? In our opinion there is one important difference between both cases. In the question Q7_8 people rate insulation from noises coming from outside. It means these noises are impersonal, although they can be assigned to some categories (e.g. ambulance, aircraft, buses etc.). Noise produced by neighbors are definitely personal, they can be assigned to specific people who could be even known to respondents. This leads to another differentiation: outside noises are independent from our will, we accept them as they are because we do not have any influence on them. On the other hand, noises coming through the wall are voluntary in that way, that they are dependent on a person who produces them. From this point of view, this noise could be eliminated or, at least, diminished – and lack of improvements can be interpreted as a malevolence of neighbors. In our opinion, this 'personal' character of noises rated in Q7_9 can also influence TA ratings.

All these findings prove that non-acoustical factors are also very important variables which have impact on people's noise annoyance assessments. On the other hand, please note that none of NoiSeQ specific noise sensitivities (Global, Habitation, Work or Sleep) is correlated with TA ratings. This can be interpreted that construction of the NoiSeQ survey does not reflect sophisticated nature of people's attitudes, fears and experiences related to noise. Maybe it should be revised somehow? A poor performance of this method was also found by us in (Felcyn et al., 2018). We will refer to that problem in the conclusions section.

The model based on PA ratings (PABaM) is much better than this which uses subjective parameters (PeCBaM, difference in R^2 value is around 20 percentile points), yet both groups of predictors give adjusted R^2 values higher than 60%. For the PA ratings, distribution of adjusted R^2 is wide, meaning that the variance is high (see Figure 35).

Finally, we want to mention one thing: all statements in the question 7 in the preliminary survey had only 5 possible answers from a 5-point bipolar Likert scale. A question designed this way has of course its limitations. We do know that analyzing such short discrete scales numerically may be biased. However, this research was designed as a preliminary study. Our first goal was to find out whether these questions can be crucial factors regarding TA. That is why we think that using this technique is justified in these circumstances.

4.2.4.2 Verification of assumptions required for linear regression. Coefficients values between different models

From the statistical point of view, linear regression is one of most demanding analyses. Many assumptions should be checked and met to treat a model as reliable and not biased. Results of all required analyses are presented in Table 21.

As one can see, all tests give p-values above the threshold of .05, meaning that needed assumptions were met. However, there is one exception. The last row of a table consists of p-values of Liliefors normality test computed on residuals of each model. One of the fundamental assumptions regarding linear regression is that its residuals should be distributed normally. If not, it is possible that there is another predictor, not taken into account, which could explain their atypical variance. P-value of the Liliefors test is lower than .05 for PABaM, however the difference is not large (p = .0313).

On the other hand, adjusted R^2 values are high. We also tested normality of replicated R^2 for each model. For every R^2 vector we computed a Liliefors normality test. As a size of vectors is large (10,000 replications), it influences p-values of test's results. In all cases, p-value was below the significance threshold of .05, puts at range from $<2.2 * 10^{-16}$ to .0003184. As lack of normality means that mean value is not the best position statistic, in Table 25 we provide also other characteristics: median and mode (mode was computed using data rounded to 4 decimal places).

Table 25. Different position statistics computed for adjusted R^2 vectors for all MLR models

Location Statistics	Whole Data	
	PA ratings	**subj. parameters**
Mean	0.8316	0.6499
Median	0.8312	0.6510
Mode	0.8269	0.6424

As one can see, other statistics' values are very similar to means. The differences are not larger than ~.01. It means that mean values is not different from using medians and modes. That is why we will still use them in next sections.

This analysis of two MLR models supports both first and second research hypotheses. It is true that TA ratings can be explained by PA ratings of different noise sources. It is also true, that non-acoustical factors are better predictors of TA than objective characteristics of stimuli. Again, we provide below equations of both PABaM and PeCBaM.

$$TA = 0.7907 * PA_{RT} + 0.1517 * PA_A + 0.4714 \quad (28)$$

$$TA = 0.0495 * Age + 0.3125 * HomeAnn + 0.6893 * Q7_1 + 0.2529 * Q7_4 + + 0.4572 * Q7_6 + 1.0514 * Q7_7 + 0.7561 * Q7_8 + 0.3426 * Q7_9 - 8.8606 \quad (29)$$

4.2.4.3 Checking PABaM and PeCBaM on data subsets

We have already mentioned that our respondents rated noise annoyance in agreement with two already existing TA models: strongest component model (SCM) or combined noise source paradox (CNSP). As our model is aimed to work in both conditions (i.e. no matter which perceptual phenomenon is behind noise annoyance assessments), we decided to check its performance on subsets created according to both models. It means that the whole data was divided in two sets – one consisted of all SCM observations, the other of all CNSP. Thus, the SCM data had 160 observations and CNSP data – 147.

As both PABaM and PeCBaM were computed using larger dataset, it is not possible to compute their performances on the other data using coefficients like R^2. However, to rate the quality of our models, we computed root mean square errors. Again, we used bootstrap method with 10,000 replications to get reliable results. They are provided in the Table 26. The results of RMSE should be interpreted as the mean error between real TA ratings and predictions of a given model. The lower RMSE values, the better goodness of fit of the model.

Table 26. RMSE values with 95% confidence intervals for both PABaM and PeCBaM checked on SCM and CNSP datasets.

RMSE	PABaM		PeCBaM	
	SCM	CNSP	SCM	CNSP
CI Low	.4749	.7735	1.1383	1.2391
Mean	.5774	1.0248	1.3147	1.4656
CI High	.6666	1.2421	1.4730	1.6666

As one can see, PABaM model gives better fit than PeCBaM for both subsets. It can be also clearly seen that it works best for the SCM data with the RMSE value of .5774 – meaning that while someone rates noise annoyance in agreement with SCM, PABaM can predict his/her ratings with the +/- .5 error in TA ratings. On the other hand, PeCBaM always gives errors larger than 1.3. Thus, its performance is burdened with higher errors. It seems that PABaM, which bases on PA ratings, gives better fit and can predict people's reactions to noise more precisely than PeCBaM.

Finally, we want to underline that both models are established based on a simulation of a specific location in Poznań. They are related to transportation situations which can be met in a modern European city. It means that their performances should be similar also in other real environment conditions.

In this section we have shown that research hypotheses no. 6 and 7 can be confirmed. Model based on subjective characteristics of individuals (PeCBaM) is better than this which bases on objective parameters (ReCBaM). And it is true, that the best goodness of fit was observed for PABaM, than for PeCBaM and the worst model was ReCBaM.

5 Conclusions

This research consisted of two Experiments. Both focused on the problem of noise annoyance of transportation noise sources and its assessments in real urban conditions. Collected data and the analyses calculated on the basis of it brought many interesting conclusions and relations. First, we refer to the research hypotheses (section 1.6):

1. **Road traffic noise annoyance ratings relate to psychoacoustical characteristics of stimuli.** Among the psychoacoustical predictors, roughness vs time 5% (with adjusted R^2 value of .7) and loudness N5 (R^2 = .71) proved to be the best, although fluctuation strength vs time 20% (R^2 = .59) and sharpness vs time 5% (R^2 = .53) also proved to be statistically significant.

   Psychoacoustical annoyance (PAnn, which bases on characteristics mentioned above) seems to be a good predictor of road traffic PA as its R^2 value is .7. However, we have shown that this characteristic slightly underestimates people's ratings – comparing computed values with real road traffic noise annoyance assessments.

2. **For the same LAeq,T, different noise annoyance ratings are observed regarding time patterns and proportions of presented road traffic packages.** Results of Experiment I did not confirm this hypothesis. Although some tendencies can be observed, none statistical significance of both factors was found.

3. **TA ratings could be explained by PA ratings of different noise sources**. This hypothesis was confirmed, although not every noise source was found to be statistically significant predictor of TA ratings. Road traffic and aircraft PA can predict TA ratings while trams PA is useless from the statistical point of view. However, it was road traffic PA which was the most influential factor.

4. **Trams PA do not contribute to the TA ratings.** This statement was confirmed. As it was shown in the analysis of the Experiment II, the trams PA factor was not statistically significant. Moreover, there was no statistical difference between a situation with no trams and any number of trams pass-bys. It seems that trams were sometimes difficult to perceive for respondents.

5. **TA ratings increase when aircraft noise is presented in mixed noise sources scenarios.** This mechanism was confirmed because at the time of an aircraft flying over the overall TA rating increased. However, as it can be seen from MLR models' equations, the influence of this factor into TA is not high. Moreover, only the dichotomic fact if the

aircraft was presented or not is statistically significant – meaning that both 0 – 1 and 0 – 2 relations are statistically significant while 1-2 is not.

6. **Non-acoustical factors (such as noise sensitivity) are better predictors of TA than objective characteristics of stimuli.** This hypothesis was confirmed. It was revealed in the Experiment II that various non-acoustical factors, taken together, could explain the variance in TA ratings at the level of 65%.

7. **From all three new TA models, the best goodness of fit should be observed for PABaM, followed by PeCBaM and the poorest performance would be observed for ReCBaM.** This hypothesis was confirmed as PABaM has the R^2 value around 83%, PeCBaM – 65% and ReCBaM only 2.46%.

However, other interesting conclusions can be drawn for both experiments. **In the case of Experiment I we also found that:**

- There are three 'dimensions' responsible for the variance in road traffic noise annoyance assessment. The most important is related to sound amplitude and it can be expressed by $L_{AEq,10m}$ (R^2 = .69), level vs time 5% (R^2 = .68) or loudness vs time 5% (R^2 = .71). The second dimension refers to the changes in time structure and can be described by roughness vs time 5% or fluctuation vs time 20%. The last one is linked to the energy in various frequency bands and is expressed by sharpness vs time 5%.
- Road traffic noise annoyance variance can be explained by only one predictor using a simple linear regression. We attempted to establish multilinear regression for this annoyance, but none of various combinations of characteristics was found to be statistically significant.
- Unlike to other papers where only ~30% of variance in annoyance ratings can be explained by sound level, in our experiment almost 70% of variance of people's judgments can be predicted by an overall sound level, $L_{AEq,10m}$. However, this 30% is met mainly in field studies, where there are many factors which are impossible to control; they can influence people's noise annoyance ratings. In laboratory conditions – which were used in this thesis – many distractors can be reduced and the conditions are the same. Thus, this high percentage of explained variance is understandable.

In the Experiment II the following findings were revealed:

- Several objective parameters were found to be statistically significantly correlated with PA ratings of different noise sources (trams and aircraft, see Figure 33 and Figure 34). However, these relations did not seem to be linear and they are rather difficult to be approximated by any function. Thus, we can say that some relations exist, but we cannot say what is their nature. Moreover, they were established using only 12 (for trams) or 8 (for aircraft) points. As trams PA was found not to be statistically significant in MLR TA models, predicting it is not useful. Regarding aircraft, further research is needed to better understand or even confirm the nature of the revealed relations.

- Majority of respondents rated TA according to the one of two perceptual models SCM or CNSP.
- Trams noise seems to be similar to road traffic – based on the similarity between their 1/3 octave spectra (see Figure 36) – and it can be the possible reason of lack of significance regarding trams PA influence on TA ratings.
- On the other hand, there is a clear peak in the aircraft noise spectrum around 6 kHz, the energy in that band is higher than for road traffic noise. As the difference is around 5-6 dB, it can be perceived and then be responsible for the perception of aircraft noise. Moreover, percentile values of sharpness and tonality were also found to be different for aircraft noise comparing them to the other noise sources (see Figure 38). Thus, aircraft PA is a significant predictor of TA ratings.
- NoiSeq survey aimed to establish individual's noise sensitivity seems to be unreliable, as this thesis is the second paper that found a poor performance of it (and no correlation with noise annoyance assessments). On the other hand, many new individual factors, which were included in the question 7 in the preliminary survey, were found to be statistically significant in MLR models. We propose to construct in the future a new survey using these statements.

Regarding the last conclusion from the list above, we want also to mention, that, based on ambiguous and sometimes strange results, not every statement from the question 7 should be taken into account. In our opinion the most reliable (and already confirmed in other papers) are statements: 1, 7 and 8. Also rating of annoyance perceived at home (on the numerical ICBEN scale) seems to be a reliable factor. Statements 4, 6 and 9, as well as the influence of age, were found to be statistically significant, however we have doubts if they can be treated as reliable and stable predictors. In our opinion, more research is needed to revise their significance.

Finally, it is worth to explain some limits and drawbacks of this thesis.

First, the range of sound levels for all mixed noise source scenarios was really small. ~1dB is the value which can be explained by the measurement error of a sound meter. Thus, it is not surprising that none of factors related to amplitude/volume were found to be statistically significant in the Experiment II. In future research it is needed to construct mixed-noises scenarios which will have much wider range of $L_{Aeq,T}$.

Second, the number of respondents who took part in both experiments is very limited. The duration of the whole experimental procedure and limited funding for respondents' salaries did not let us to create a more sophisticated procedure. Fortunately, as the author of this thesis got a grant for further research, results of this work will be used to plan better and more diverse experiments in the near future.

The next thing is the discrete and short Likert scale used to rate statements from the question 7 in the preliminary survey. It is possible that some other interesting relations were not found due to the limited variability of people's answers. Moreover, as Likert type scale is commonly perceived as ordinal one, one can argue if it is a good practice to use that scale in statistical tests based on variances and mean values. Maybe some statements could be incorporated in surveys and the scale should be changed or widen.

These studies revealed several new findings concerning the assessment of human annoyance and its predictors. New TA models were proposed which can be used in situations where road traffic noise is dominant and other noise sources appear incidentally (thus, the variability of an overall sound level is minimal). Such situation can be observed in the city of Poznań – and real traffic flow characteristics and sound levels were used in both experiments to simulate an urban environment of that city. For the first time in such research an ambisonic technique was used to keep maximum spatial information about all presented noise sources. It was also shown that prediction of TA is the best when is based on PA ratings, rather than subjective or objective parameters.

6 Acknowledgments

This thesis was supervised by professor Anna Preis who always supported and encouraged me in my work – both scientific and educational. It is also impossible to exaggerate help received in all statistical analyses from dr Paweł Kleka. Ambisonic recordings of different noise sources were done with great help from our students. My big thanks go to colleagues from Akustix Company – Michał Gałuszka, Piotr Kokowski, Tomasz Kaczmarek – who helped me in planning experimental procedure and provided information about sound levels of different noise sources in Poznań. Special thanks also go to my father who helped me with English in this thesis. Last, my biggest thanks go to my family. They supported me at every step of my scientific career and I know that without them this thesis would not be created.

This research was partially funded by the Faculty of Physics of the Adam Mickiewicz University and by the National Research Centre with the grant Preludium (UMO-2017/25/N/HS6/01164).

7 Literature

Alayrac, M., Marquis-Favre, C., Viollon, S., Morel, J., and Le Nost, G. (2010). Annoyance from industrial noise: Indicators for a wide variety of industrial sources. *J. Acoust. Soc. Am.* 128, 1128. doi:10.1121/1.3466855.

Alvarsson, J. J., Nordström, H., Lundén, P., and Nilsson, M. E. (2014). Aircraft noise and speech intelligibility in an outdoor living space. *J. Acoust. Soc. Am.* 135, 3455–3462. Available at: http://scitation.aip.org/content/asa/journal/jasa/135/6/10.1121/1.4874625 [Accessed May 27, 2016].

Aspuru, I., García, I., Herranz, K., and Santander, A. (2016). CITI-SENSE: methods and tools for empowering citizens to observe acoustic comfort in outdoor public spaces. *Noise Mapp.* 3. doi:10.1515/noise-2016-0003.

Axelsson, G., Stockfelt, L., Andersson, E., Gidlof-Gunnarsson, A., Sallsten, G., and Barregard, L. (2013). Annoyance and Worry in a Petrochemical Industrial Area—Prevalence, Time Trends and Risk Indicators. *Int. J. Environ. Res. Public Health* 10, 1418–1438. doi:10.3390/ijerph10041418.

Babisch, W., Houthuijs, D., Pershagen, G., Cadum, E., Katsouyanni, K., Velonakis, M., et al. (2009a). Annoyance due to aircraft noise has increased over the years—Results of the HYENA study. *Environ. Int.* 35, 1169–1176. doi:10.1016/j.envint.2009.07.012.

Babisch, W., Houthuijs, D., Pershagen, G., Cadum, E., Katsouyanni, K., Velonakis, M., et al. (2009b). Annoyance due to aircraft noise has increased over the years—Results of the HYENA study. *Environ. Int.* 35, 1169–1176. doi:10.1016/j.envint.2009.07.012.

Bartels, S., Márki, F., and Müller, U. (2015a). The influence of acoustical and non-acoustical factors on short-term annoyance due to aircraft noise in the field — The COSMA study. *Sci. Total Environ.* 538, 834–843. doi:10.1016/j.scitotenv.2015.08.064.

Bartels, S., Márki, F., and Müller, U. (2015b). The influence of acoustical and non-acoustical factors on short-term annoyance due to aircraft noise in the field — The COSMA study. *Sci. Total Environ.* 538, 834–843. doi:10.1016/j.scitotenv.2015.08.064.

Bartels, S., Müller, U., and Vogt, J. (2013). Predictors of aircraft noise annoyance: results of a telephone study. in *Inter-Noise 2013*, 1–10. Available at: file:///C:/Users/Jachu/Documents/IA/conference proceedings/Internoise 2013/pdf/0249.pdf [Accessed April 26, 2017].

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., et al. (2014). Auditory and non-auditory effects of noise on health. *Lancet* 383, 1325–1332. doi:10.1016/S0140-6736(13)61613-X.

Berglund, B., Berglund, U., Goldstein, M., and Lindvall, T. (1981). Loudness (or annoyance) summation of combined community noises. *J. Acoust. Soc. Am.* 70, 1628–1634. Available at: http://scitation.aip.org/content/asa/journal/jasa/70/6/10.1121/1.387229 [Accessed February 24, 2016].

Berglund, B., and Nilsson, M. E. (1997). Empirical issues concerning annoyance models for combined community noises. in *Proceedings of Inter-Noise '97* (Budapest).

Berglund, B., and Nilsson, M. E. (1998a). LOUDNESS OF COMBINED NOISES DERIVED FROM SINGULAR AND CONCURRENT COMMUNITY NOISES. in *Proceedings of International Congress on Acoustics*

1998, 1129–1130.

Berglund, B., and Nilsson, M. E. (1998b). Total annoyance models for community noises explicated. in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Christchurch: Institute of Noise Control Engineering), 1013–1016. Available at: http://www-1ingentaconnect-1com-1813rafvyba33.han.amu.edu.pl/search/article?option1=tka&value1=total+annoyance+models+for+community+noises+explicated&pageSize=10&index=1 [Accessed July 24, 2018].

Berglund, B., Preis, A., and Rankin, K. (1990). Relationship between loudness and annoyance for ten community sounds. *Environ. Int.* 16, 523–531. doi:10.1016/0160-4120(90)90021-W.

Berrien, F. K. (1946). The effects of noise. *Psychol. Bull.* 43, 141–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21018316 [Accessed October 9, 2017].

Beutel, M. E., Jünger, C., Klein, E. M., Wild, P., Lackner, K., Blettner, M., et al. (2016). Noise Annoyance Is Associated with Depression and Anxiety in the General Population- The Contribution of Aircraft Noise. *PLoS One* 11, e0155357. doi:10.1371/journal.pone.0155357.

Bockstael, A., Dekoninck, L., Can, A., Oldoni, D., De Coensel, B., and Botteldooren, D. (2012). Reduction of wind turbine noise annoyance: an operational approach. *Acta Acust. United with Acust.* 98, 392–401. Available at: http://www.ingentaconnect.com/content/dav/aaua/2012/00000098/00000003/art00005 [Accessed May 28, 2016].

Borchi, F., Carfagni, M., Governi, L., Curcuruto, S., Silvaggio, R., Bellomini, R., et al. (2016). LIFE+2008 HUSH project results: a new methodology and a new platform for implementing an integrated and harmonized noise Action Plan and proposals for updating Italian legislation and Environmental Noise Directive. *Noise Mapp.* 3. doi:10.1515/noise-2016-0006.

Botteldooren, D., and Verkeyn, A. (2002). Fuzzy models for accumulation of reported community noise annoyance from combined sources. *J. Acoust. Soc. Am.* 112, 1496. doi:10.1121/1.1506366.

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324.

Brink, M., Schreckenberg, D., Vienneau, D., Cajochen, C., Wunderli, J.-M., Probst-Hensch, N., et al. (2016). Effects of Scale, Question Location, Order of Response Alternatives, and Season on Self-Reported Noise Annoyance Using ICBEN Scales: A Field Experiment. *Int. J. Environ. Res. Public Health* 13, 1163. doi:10.3390/ijerph13111163.

Brink, M., Wirth, K. E., Schierz, C., Thomann, G., and Bauer, G. (2008). Annoyance responses to stable and changing aircraft noise exposure. *J. Acoust. Soc. Am.* 124, 2930–2941. doi:10.1121/1.2977680.

Brown, A. L. (2010). Soundscapes and environmental noise management. *Noise Control Eng. J.* 58, 493. doi:10.3397/1.3484178.

Brown, A. L., Lam, K. C., and Van Kamp, I. (2015). Quantification of the exposure and effects of road traffic noise in a dense Asian city: a comparison with western cities. *Environ. Heal.* 14. doi:10.1186/s12940-015-0009-8.

Cai, M., Yao, Y., and Wang, H. (2017). A traffic-noise-map update method based on monitoring data. *J. Acoust. Soc. Am. Acoust. Soc. Am. J. Acoust. Soc. Am.* 141, 2604–2418. doi:10.1121/1.4979336.

Campello-Vicente, H., Peral-Orts, R., Campillo-Davo, N., and Velasco-Sanchez, E. (2017). The effect of electric vehicles on urban noise maps. doi:10.1016/j.apacoust.2016.09.018.

Camusso, C., and Pronello, C. (2016). A study of relationships between traffic noise and annoyance for different urban site typologies. *Transp. Res. Part D Transp. Environ.* 44, 122–133. doi:10.1016/j.trd.2016.02.007.

Caschili, S., De Montis, A., Ganciu, A., Ledda, A., and Barra, M. (2014). The Strategic Environment Assessment bibliographic network: A quantitative literature review analysis. *Environ. Impact Assess. Rev.* 47, 14–28. doi:10.1016/j.eiar.2014.03.003.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge doi:10.4324/9780203771587.

Covaciu, D., Florea, D., and Timar, J. (2015). Estimation of the noise level produced by road traffic in roundabouts. *Appl. Acoust.* 98, 43–51. doi:10.1016/j.apacoust.2015.04.017.

D'Alessandro, F., and Schiavoni, S. (2015). A review and comparative analysis of European priority indices for noise action plans. *Sci. Total Environ.* 518–519, 290–301. doi:10.1016/j.scitotenv.2015.02.102.

De Coensel, B., Botteldooren, D., Berglund, B., Nilsson, M. E., De Muer, T., and Lercher, P. (2007). Experimental investigation of noise annoyance caused by high-speed trains. *Acta Acust. united with Acust.* 93, 589–601. Available at: http://www.ingentaconnect.com/content/dav/aaua/2007/00000093/00000004/art00007 [Accessed February 24, 2016].

de Kluizenaar, Y., Janssen, S., Vos, H., Salomons, E., Zhou, H., and van den Berg, F. (2013). Road Traffic Noise and Annoyance: A Quantification of the Effect of Quiet Side Exposure at Dwellings. *Int. J. Environ. Res. Public Health* 10, 2258–2270. doi:10.3390/ijerph10062258.

de Kluizenaar, Y., Salomons, E. M., Janssen, S. A., van Lenthe, F. J., Vos, H., Zhou, H., et al. (2011). Urban road traffic noise and annoyance: The effect of a quiet façade. *J. Acoust. Soc. Am.* 130, 1936–1942. Available at: http://scitation.aip.org/content/asa/journal/jasa/130/4/10.1121/1.3621180 [Accessed May 27, 2016].

Di, G.-Q., Lin, Q.-L., Li, Z.-G., and Kang, J. (2014). Annoyance and activity disturbance induced by high-speed railway and conventional railway noise: a contrastive case study. *Environ. Heal.* 13, 12. doi:10.1186/1476-069X-13-12.

Dittrich, K., and Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *J. Acoust. Soc. Am.* 126, 3168–3178. doi:10.1121/1.3238233.

Dzhambov, A. M., Dimitrova, D. D., and Dzhambov, A. M. (2016). Association between Noise Pollution and Prevalent Ischemic Heart Disease. *Folia Medica I Folia Medica* 5858, 273–281. doi:10.1515/folmed-2016-0041.

Dzhambov, A., Tilov, B., Markevych, I., and Dimitrova, D. (2017). Residential road traffic noise and general mental health in youth: The role of noise annoyance, neighborhood restorative quality, physical activity, and social cohesion as potential mediators. doi:10.1016/j.envint.2017.09.009.

Elmenhorst, E.-M., Quehl, J., Müller, U., and Basner, M. (2014). Nocturnal air, road, and rail traffic noise and daytime cognitive performance and annoyance. *J. Acoust. Soc. Am.* 135, 213–222. doi:10.1121/1.4842475.

European Comission (2015). Commission Directive (EU) 2015/996 of 19 May 2015 establishing common noise assessment methods according to Directive 2002/49/EC of the European Parliament and of the Council (Text with EEA relevance). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L0996 [Accessed July 20, 2018].

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. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0049&from=en [Accessed May 31, 2017].

Fastl, H. (Author), and Zwicker, E. (Author) (2007). *Psychoacoustics: Facts and models*. Berlin: Springer Available at: http://search.ebscohost.com/login.aspx?direct=true&db=rih&AN=1992-10206&site=ehost-live&scope=cite.

Felcyn, J., Preis, A., Kokowski, P., and Gałuszka, M. (2018). A comparison of noise mapping data and people's assessment of annoyance: How can noise action plans be improved? *Transp. Res. Part D Transp. Environ.* 63, 72–120. doi:10.1016/J.TRD.2018.04.019.

Fidell, S., Barber, D. S., and Schultz, T. J. (1991). Updating a dosage–effect relationship for the prevalence of annoyance due to general transportation noise. *J. Acoust. Soc. Am.* 89, 221–233. doi:10.1121/1.400504.

Fidell, S., Horonjeff, R., Mills, J., Baldwin, E., Teffeteller, S., and Pearsons, K. (1985). Aircraft noise annoyance at three joint air carrier and general aviation airports. *J. Acoust. Soc. Am.* 77, 1054–1068. doi:10.1121/1.392223.

Fidell, S., Mestre, V., and Sneddon, M. (2012). A potential role for noise complaints as a predictor of the prevalence of annoyance with aircraft noise. *Noise Control Eng. J.* 60, 62. doi:10.3397/1.3677792.

Fields, J. M. (1994). A review of an updated synthesis of noise/annoyance relationships. *NASA Rep. 194950, Giorgia Inst. Technol.* Available at: https://ntrs.nasa.gov/search.jsp?R=19940029797 [Accessed January 15, 2018].

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., et al. (2001). STANDARDIZED GENERAL-PURPOSE NOISE REACTION QUESTIONS FOR COMMUNITY NOISE SURVEYS: RESEARCH AND A RECOMMENDATION. *J. Sound Vib.* 242, 641–679. doi:10.1006/jsvi.2000.3384.

Freitas, E., Mendonça, C., Santos, J. A., Murteira, C., and Ferreira, J. P. (2012). Traffic noise abatement: How different pavements, vehicle speeds and traffic densities affect annoyance levels. *Transp. Res. Part D Transp. Environ.* 17, 321–326. doi:10.1016/j.trd.2012.02.001.

Gerzon, M. A. (1973). Periphony: With-Height Sound Reproduction. *J. Audio Eng. Soc.* 21, 2–10. Available at: http://www.aes.org/e-lib/browse.cfm?elib=2012 [Accessed June 1, 2017].

Gidlöf-Gunnarsson, A., Jerson, T., Ögren, M., and Öhrström, E. (2012). Railway noise annoyance and the importance of number of trains, ground vibration, and building situational factors. *Noise Heal.* 14, 190. doi:10.4103/1463-1741.99895.

Gidlöf-Gunnarsson, A., and Öhrström, E. (2007). Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. *Landsc. Urban Plan.* 83, 115–126. doi:10.1016/j.landurbplan.2007.03.003.

Gille, L.-A., Marquis-Favre, C., and Klein, A. (2016a). Noise Annoyance Due To Urban Road Traffic with Powered-Two-Wheelers: Quiet Periods, Order and Number of Vehicles. *Acta Acust. united with Acust.* 102, 474–487. doi:10.3813/AAA.918966.

Gille, L.-A., Marquis-Favre, C., and Morel, J. (2016b). Testing of the European Union exposure-response relationships and annoyance equivalents model for annoyance due to transportation noises: The need of revised exposure-response relationships and annoyance equivalents model. *Environ. Int.* 94, 83–94. doi:10.1016/j.envint.2016.04.027.

Gille, L.-A., Marquis-Favre, C., and Weber, R. (2016c). Noise sensitivity and loudness derivative index for urban road traffic noise annoyance computation. *J. Acoust. Soc. Am.* 140, 4307–4317. doi:10.1121/1.4971329.

Gille, L.-A., Marquis-Favre, C., and Weber, R. (2017a). Aircraft noise annoyance modeling: Consideration of noise sensitivity and of different annoying acoustical characteristics. *Appl. Acoust.* 115, 139–149. doi:10.1016/j.apacoust.2016.08.022.

Gille, L.-A., Marquis-Favre, C., and Weber, R. (2017b). Aircraft noise annoyance modeling: Consideration of noise sensitivity and of different annoying acoustical characteristics. *Appl. Acoust.* 115, 139–149. doi:10.1016/j.apacoust.2016.08.022.

Gjestland, T., and Gelderblom, F. (2017). Prevalence of Noise Induced Annoyance and Its Dependency on Number of Aircraft Movements. *Acta Acust. united with Acust.* 103, 28–33. doi:10.3813/AAA.919030.

Główny Urząd Statystyczny Statystyczne Vademecum Samorządowca. Available at: http://poznan.stat.gov.pl/statystyczne-vademecum-samorzadowca/ [Accessed November 19, 2018].

Griefahn, B., Marks, A., Gjestland, T., and Preis, A. (2007). ANNOYANCE AND NOISE SENSITIVITY IN URBAN AREAS. in *Paper Presented at the 19th International Congress on Acoustics, Madrid, Spain (September)* (Madrid). Available at: http://www.sea-acustica.es/WEB_ICA_07/fchrs/papers/env-04-004.pdf [Accessed May 26, 2016].

Guillaume, G., Can, A., Petit, G., Fortin, N., Palominos, S., Gauvreau, B., et al. (2016). Noise mapping based on participative measurements. *Noise Mapp.* 3. doi:10.1515/noise-2016-0011.

Guoqing, D., Xiaoyi, L., Xiang, S., Zhengguang, L., and Qili, L. (2012). Investigation of the relationship between aircraft noise and community annoyance in China. *Noise Heal.* 14, 52. doi:10.4103/1463-1741.95132.

Guski, R. (1997). CONCEPTUAL, METHODOLOGICAL, AND DOSE-RESPONSE PROBLEMS RELATED TO ANNOYANCE AND DISTURBANCE. in *InterNoise 97* Available at: https://www.researchgate.net/profile/Rainer_Guski/publication/284771893_Conceptual_methodological_and_dose-response_problems_related_to_annoyance_and_disturbance/links/56b4d9ce08ae3c1b79aaf8f6/Conceptual-methodological-and-dose-response-problems-related-t [Accessed September 1, 2017].

Hall, F. L., Birnie, S. E., Taylor, S. M., and Palmer, J. E. (1981). Direct comparison of community response to road traffic noise and to aircraft noise. *J. Acoust. Soc. Am.* 70, 1690–1698. doi:10.1121/1.387234.

Hatfield, J., van Kamp, I., and Job, R. F. S. (2006). Clarifying “Soundscape”: Effects of Question Format on Reaction to Noise from Combined Sources. *Acta Acust. united with Acust.* 92, 922–928.

Heimann, D., Clemente, M., Olny, X., Defrance, J., Suppan, P., Castelli, S. T., et al. (2007). Air Pollution, Traffic Noise and Related Health Effects in the Alpine Space A Guide for Authorities and Consulters ALPNAP Project Steering Group Editorial Team:: Air Pollution, Traffic Noise and Related Health Effects in the Alpine Space − A Guide for Aut. Available at: http://www.ing.unitn.it/dica/tools/download/Quaderni/ALPNAP_CR_2007_Part_1.pdf [Accessed April 25, 2017].

Heleno, T., and Slama, J. G. (2013). Fuzzy Modeling of Annoyance Caused by Aircraft Noise Using LAeqD and LAeqN MetricsD and LAeqN metrics. *J. Aerosp. Technol. Manag.* 5, 103–110.

doi:10.5028/jatm.v5i1.232.

Héritier, H., Vienneau, D., Frei, P., Eze, I., Brink, M., Probst-Hensch, N., et al. (2014). The Association between Road Traffic Noise Exposure, Annoyance and Health-Related Quality of Life (HRQOL). *Int. J. Environ. Res. Public Health* 11, 12652–12667. doi:10.3390/ijerph111212652.

Hermida Cadena, L. F., Lobo Soares, A. C., Pavón, I., and Coelho, L. B. (2017). Assessing soundscape: Comparison between in situ and laboratory methodologies. *Noise Mapp.* 4, 57–66. doi:10.1515/noise-2017-0004.

Herranz-Pascual, K., García, I., Aspuru, I., Díez, I., and Santander, Á. (2016). Progress in the understanding of soundscape: objective variables and objectifiable criteria that predict acoustic comfort in urban places. *Noise Mapp.* 3. doi:10.1515/noise-2016-0017.

Hollerweger, F. (2008). An Introduction to Higher-Order Ambisonic. Available at: http://flo.mur.at/writings/HOA-intro.pdf/view [Accessed July 26, 2018].

Ioannidou, C., Santurette, S., and Jeong, C.-H. (2016). Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyancea). *J. Acoust. Soc. Am.* 139, 1241–1251. Available at: http://scitation.aip.org/content/asa/journal/jasa/139/3/10.1121/1.4944570 [Accessed May 28, 2016].

Izumi, K. (1988). Annoyance due to mixed source noises— A laboratory study and field surve on the annoyance of road traffic and railroad noise. *J. Sound Vib.* 127, 485–489. doi:10.1016/0022-460X(88)90373-2.

Janssen, S. A., Vos, H., Eisses, A. R., and Pedersen, E. (2011a). A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources. *J. Acoust. Soc. Am.* 130, 3746–3753. Available at: http://scitation.aip.org/content/asa/journal/jasa/130/6/10.1121/1.3653984 [Accessed May 26, 2016].

Janssen, S. A., Vos, H., van Kempen, E. E., Breugelmans, O. R., and Miedema, H. M. (2011b). Trends in aircraft noise annoyance: The role of study and sample characteristics. *J. Acoust. Soc. Am.* 129, 1953–1962. Available at: http://scitation.aip.org/content/asa/journal/jasa/129/4/10.1121/1.3533739 [Accessed May 27, 2016].

Kaczmarek, T., and Preis, A. (2010). Annoyance of time-varying road-traffic noise. *Arch. Acoust.* 35, 383–393. Available at: http://acoustics.ippt.gov.pl/index.php/aa/article/view/232 [Accessed July 27, 2018].

Kasess, C. H., Noll, A., Majdak, P., and Waubke, H. (2013). Effect of train type on annoyance and acoustic features of the rolling noise. *J. Acoust. Soc. Am.* 134, 1071–1081. Available at: http://scitation.aip.org/content/asa/journal/jasa/134/2/10.1121/1.4812771 [Accessed May 27, 2016].

King, E. A., and Murphy, E. (2016). Environmental noise – 'Forgotten' or 'Ignored' pollutant? *Appl. Acoust.* 112, 211–215. doi:10.1016/J.APACOUST.2016.05.023.

Klein, A. (2015). Annoyance indicators for various urban road vehicle pass-by noises and urban road traffic noise combined with tramway noise. Available at: https://www.researchgate.net/publication/323163636_Annoyance_indicators_for_various_urban_road_vehicle_pass-by_noises_and_urban_road_traffic_noise_combined_with_tramway_noise [Accessed July 23, 2018].

Klein, A., Marquis-Favre, C., and Champelovier, P. (2017). Assessment of annoyance due to urban road traffic noise combined with tramway noise. *J. Acoust. Soc. Am.* 141, 231–242. doi:10.1121/1.4973518.

Kroesen, M., Molin, E. J., and van Wee, B. (2013). Measuring subjective response to aircraft noise: The effects of survey context. *J. Acoust. Soc. Am.* 133, 238–246. Available at: http://scitation.aip.org/content/asa/journal/jasa/133/1/10.1121/1.4768795 [Accessed April 27, 2016].

Kroesen, M., and Schreckenberg, D. (2011). A measurement model for general noise reaction in response to aircraft noise. *J. Acoust. Soc. Am.* 129, 200–210. Available at: http://scitation.aip.org/content/asa/journal/jasa/129/1/10.1121/1.3514542 [Accessed May 28, 2016].

Kryter, K. D. (1982). Community annoyance from aircraft and ground vehicle noise. *J. Acoust. Soc. Am.* 72, 1222–1242. doi:10.1121/1.388332.

Laird, D. A., and Coye, K. (1929). PSYCHOLOGICAL MEASUREMENTS OF ANNOYANCE AS RELATED TO PITCH AND LOUDNESS. *J. Acoust. Soc. Am.* 1, 158–163. doi:10.1121/1.1901476.

Lam, K.-C., Chan, P.-K., Chan, T.-C., Au, W.-H., and Hui, W.-C. (2009). Annoyance response to mixed transportation noise in Hong Kong. *Appl. Acoust.* 70, 1–10. doi:10.1016/j.apacoust.2008.02.005.

Laszlo, H. E., McRobie, E. S., Stansfeld, S. A., and Hansell, A. L. (2012). Annoyance and other reaction measures to changes in noise exposure — A review. *Sci. Total Environ.* 435–436, 551–562. doi:10.1016/j.scitotenv.2012.06.112.

Licitra, G., Gallo, P., Rossi, E., and Brambilla, G. (2011). A novel method to determine multiexposure priority indices tested for Pisa action plan. *Appl. Acoust.* 72, 505–510. doi:10.1016/J.APACOUST.2010.07.007.

Marquis-Favre, C., and Morel, J. (2015). A Simulated Environment Experiment on Annoyance Due to Combined Road Traffic and Industrial Noises. *Int. J. Environ. Res. Public Health* 12, 8413–8433. doi:10.3390/ijerph120708413.

Marquis-Favre, C., Premat, E., and Aubrée, D. (2005). Noise and its Effects – A Review on Qualitative Aspects of Sound. Part II: Noise and Annoyance. *Acta Acust. united with Acust.* 91, 626–642. Available at: http://docserver.ingentaconnect.com.ingentaconnecthirzelverlag-actaacusticaunitedwithac2.bu-169.bu.amu.edu.pl/deliver/connect/dav/16101928/v91n4/s2.pdf?expires=1456312391&id=86150596&titleid=75000347&accname=Uniwersytet+im.+Adama+Mickiewicza+-+Biblioteka+.

Martín, M. Á., Tarrero, A. I., Machimbarrena, M., González, J., and Garibay, V. G. de (2011). A methodology to study noise annoyance and to perform Action Plans follow up using as input an existing survey and noise map: Application to the city of Málaga (Spain). *Appl. Acoust.* 72, 495–504. doi:10.1016/J.APACOUST.2011.02.005.

Méline, J., Van Hulst, A., Thomas, F., Karusisi, N., and Chaix, B. (2013). Transportation noise and annoyance related to road traffic in the French RECORD study. *Int. J. Health Geogr.* 12, 44. doi:10.1186/1476-072X-12-44.

Michaud, D. S., Keith, S. E., Feder, K., Voicescu, S. A., Marro, L., Than, J., et al. (2016). Personal and situational variables associated with wind turbine noise annoyance. *J. Acoust. Soc. Am.* 139, 1455–1466. doi:10.1121/1.4942390.

Miedema, H. M. E. (2004). Relationship between exposure to multiple noise sources and noise annoyance. *J. Acoust. Soc. Am.* 116, 949. doi:10.1121/1.1766305.

Miedema, H. M. E., and Vos, H. (1998). Exposure-response relationships for transportation noise. *J. Acoust. Soc. Am.* 104. Available at: https://doi.org/10.1121/1.423927 [Accessed November 15, 2017].

Miedema, H. M., and Oudshoorn, C. G. (2001). Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. *Environ. Health Perspect.* 109, 409. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240282/ [Accessed September 16, 2015].

Morel, J., and Marquis-Favre, C. (2013). Annoyance due to combined road traffic and industrial noises: a simulated environment experiment. in.

Morel, J., Marquis-Favre, C., Viollon, S., and Alayrac, M. (2012). A Laboratory Study on Total Noise Annoyance Due To Combined Industrial Noises. *Acta Acust. united with Acust.* 98, 286–300. doi:10.3813/AAA.918512.

Murphy, E., and King, E. A. (2016). Smartphone-based noise mapping: Integrating sound level meter app data into the strategic noise mapping process. *Sci. Total Environ.* 562, 852–859. doi:10.1016/j.scitotenv.2016.04.076.

Nguyen, L. T., Yano, T., Nguyen, Q. H., Nishimura, T., Sato, T., and Morihara, T. (2013). Community response to aircraft noise around three airports in Vietnam. in, 040003–040003. doi:10.1121/1.4789400.

Nguyen, T. L., Nguyen, H. Q., Yano, T., Nishimura, T., Sato, T., Morihara, T., et al. (2012). Comparison of models to predict annoyance from combined noise in Ho Chi Minh City and Hanoi. *Appl. Acoust.* 73, 952–959. doi:10.1016/j.apacoust.2012.04.005.

Öhrström, E., Barregård, L., Andersson, E., Skånberg, A., Svensson, H., and Ängerheim, P. (2007). Annoyance due to single and combined sound exposure from railway and road traffic. *J. Acoust. Soc. Am.* 122, 2642. doi:10.1121/1.2785809.

Okokon, E., Turunen, A., Ung-Lanki, S., Vartiainen, A.-K., Tiittanen, P., and Lanki, T. (2015a). Road-Traffic Noise: Annoyance, Risk Perception, and Noise Sensitivity in the Finnish Adult Population. *Int. J. Environ. Res. Public Health* 12, 5712–5734. doi:10.3390/ijerph120605712.

Okokon, E., Turunen, A., Ung-Lanki, S., Vartiainen, A.-K., Tiittanen, P., and Lanki, T. (2015b). Road-Traffic Noise: Annoyance, Risk Perception, and Noise Sensitivity in the Finnish Adult Population. *Int. J. Environ. Res. Public Health* 12, 5712–5734. doi:10.3390/ijerph120605712.

Ota, O., Yokoshima, S., and Tamura, S. (2008). The metrics of mixed traffic noise: Results of simulated environment experiments. in, 8. Available at: http://www.icben.org/2008/PDFs/Ota_et_al.pdf [Accessed February 24, 2016].

Pallas, M.-A., Philipps-Bertin, C., Maldonado, M., Legouis, T., Champelovier, P., Chiello, O., et al. (2009). *Bruit et vibrations dus aux tramways : émission et perception*. Available at: http://www.ifsttar.fr/fileadmin/user_upload/editions/inrets/Recherches/Rapport_INRETS_R279.pdf [Accessed April 26, 2017].

Panulinová, E. (2017). Input Data for Tram Noise Analysis. *Procedia Eng.* 190, 371–376. doi:10.1016/j.proeng.2017.05.351.

Paviotti, M., and Vogiatzis, K. (2012). On the outdoor annoyance from scooter and motorbike noise in the urban environment. *Sci. Total Environ.* 430, 223–230. doi:10.1016/J.SCITOTENV.2012.05.010.

Pedersen, E., and Persson Waye, K. (2004). Perception and annoyance due to wind turbine noise—a dose–response relationship. *J. Acoust. Soc. Am.* 116, 3460. doi:10.1121/1.1815091.

Pierrette, M., Marquis-Favre, C., Morel, J., Rioux, L., Vallet, M., Viollon, S., et al. (2012). Noise annoyance from industrial and road traffic combined noises: A survey and a total annoyance model comparison. *J. Environ. Psychol.* 32, 178–186. doi:10.1016/j.jenvp.2012.01.006.

Powell, C. A. (1979). A summation and inhibition model of annoyance response to multiple community noise sources. Available at: http://ntrs.nasa.gov/search.jsp?R=19790024573 [Accessed May 13, 2016].

Powell, C. A., and Rice, C. G. (1975). Judgments of aircraft noise in a traffic noise background. *J. Sound Vib.* 38, 39–50. doi:10.1016/S0022-460X(75)80019-8.

Preis, A., Hafke-Dys, H., Kaczmarek, T., Gjestland, T., and Kleka, P. (2013). The relationship between speech intelligibility and the assessment of noise annoyance. *Noise Control Eng. J.* 61, 255–264. doi:10.3397/1/3761022.

Preis, A., Hafke-Dys, H., Szychowska, M., Kocinski, J., and Felcyn, J. (2016). Audio-visual interaction of environmental noise. *Noise Control Eng. J.* 64, 34–43. doi:10.3397/1/376357.

Preis, A., Kaczmarek, T., Wojciechowska, H., Żera, J., and Fields, J. M. (2003). POLISH VERSION OF STANDARDIZED NOISE REACTION QUESTIONS FOR COMMUNITY NOISE SURVEYS. *IJOMEH Int. J. Occup. Med. Environ. Heal.* 16, 155–159. Available at: http://www.imp.lodz.pl/upload/oficyna/artykuly/pdf/full/Pre6-02-03.pdf [Accessed June 2, 2017].

Preis, A., Kociński, J., Hafke-Dys, H., and Wrzosek, M. (2015a). Audio-visual interactions in environment assessment. *Sci. Total Environ.* 523, 191–200. doi:10.1016/j.scitotenv.2015.03.128.

Preis, A., Kociński, J., Hafke-Dys, H., and Wrzosek, M. (2015b). Audio-visual interactions in environment assessment. *Sci. Total Environ.* 523, 191–200. Available at: http://www.sciencedirect.com/science/article/pii/S004896971500412X [Accessed November 13, 2015].

Raimbault, M. (2006). Qualitative Judgements of Urban Soundscapes: Questionning Questionnaires and Semantic Scales. *Acta Acust. united with Acust.* 92, 929–937. Available at: http://www-1ingentaconnect-1com-1zuuhg50q0e5a.han.amu.edu.pl/content/dav/aaua/2006/00000092/00000006/art00011# [Accessed November 27, 2017].

Ruiz-Padillo, A., Torija, A. J., Ramos-Ridao, A. F., and Ruiz, D. P. (2016). Application of the fuzzy analytic hierarchy process in multi-criteria decision in noise action plans: Prioritizing road stretches. *Environ. Model. Softw.* 81, 45–55. doi:10.1016/j.envsoft.2016.03.009.

Sandrock, S., Griefahn, B., Kaczmarek, T., Hafke, H., Preis, A., and Gjestland, T. (2008). Experimental studies on annoyance caused by noises from trams and buses. *J. Sound Vib.* 313, 908–919. doi:10.1016/j.jsv.2007.12.003.

Sandrock, S., Schutte, M., and Griefahn, B. (2007). The reliability of the noise sensitivity questionnaire in a cross-national analysis. *Noise Heal.* 9, 8. doi:10.4103/1463-1741.34699.

Sato, T., Yano, T., Bjog Rkman, M., and Rylander, R. (2002). COMPARISON OF COMMUNITY RESPONSE TO ROAD TRAFFIC NOISE IN JAPAN AND SWEDEN—PART I: OUTLINE OF SURVEYS AND DOSE-RESPONSE RELATIONSHIPS. *J. Sound Vib.* 250, 161–167. doi:10.1006.

Sato, T., Yano, T., and Morihara, T. (2004). Community Response to Noise from Shinkansen in

Comparison with Ordinary Railways: A Survey in Kyushu, Japan. in *ICA 2004*, 2865–2868. Available at: file:///C:/Users/Jachu/Documents/IA/conference proceedings/ICA-2004/pdf/Th4.X1.3.pdf [Accessed April 26, 2017].

Schomer, P., Mestre, V., Fidell, S., Berry, B., Gjestland, T., Vallet, M., et al. (2012). Role of community tolerance level (CTL) in predicting the prevalence of the annoyance of road and rail noise. *J. Acoust. Soc. Am.* 131, 2772–2786. doi:10.1121/1.3688762.

Schreckenberg, D., Belke, C., and Spilski, J. (2018). The Development of a Multiple-Item Annoyance Scale (MIAS) for Transportation Noise Annoyance. *Int. J. Environ. Res. Public Health* 15, 971. doi:10.3390/ijerph15050971.

Schreckenberg, D., and Schuemer, R. (2010). The impact of acoustical, operational and non-auditory factors on short-term annoyance due to aircraft noise. in *Proceedings - Inter-Noise 2010, 39th International Congress on Noise Control Engineering : 13 - 16 June 2010, Lisbon, Portugal* (Lisboa: Sociedade Portuguesa de Acústica). Available at: https://www.tib.eu/en/search/id/TIBKAT%3A630004412/Proceedings-Inter-Noise-2010-39th-International/ [Accessed December 3, 2018].

Schultz, T. J. (1978). Synthesis of social surveys on noise annoyance. *J. Acoust. Soc. Am.* 64. doi:10.1121/1.382013.

Seong, Y., Lee, S., Gwak, D. Y., Cho, Y., Hong, J., and Lee, S. (2013). An experimental study on annoyance scale for assessment of wind turbine noise. *J. Renew. Sustain. Energy* 5, 052008. Available at: http://scitation.aip.org/content/aip/journal/jrse/5/5/10.1063/1.4821811 [Accessed May 28, 2016].

Sheather, S. (2009). *A Modern Approach to Regression with R*. New York, NY: Springer New York doi:10.1007/978-0-387-09608-7.

Shimoyama, K., Thu, ;, Nguyen, L., Yano, T., and Morihara, T. (2014). Social surveys on community response to road traffic in five cities in Vietnam. in *Inter-Noise*, 1–8. Available at: file:///C:/Users/Jachu/Documents/IA/conference proceedings/Internoise 2014/papers/p109.pdf [Accessed April 26, 2017].

Sung, J. H., Lee, J., Jeong, K. S., Lee, S., Lee, C., Jo, M.-W., et al. (2017). Influence of Transportation Noise and Noise Sensitivity on Annoyance: A Cross-Sectional Study in South Korea. *Int. J. Environ. Res. Public Health* 14. doi:10.3390/ijerph14030322.

Szychowska, M., Hafke-Dys, H., Preis, A., Drzej, J. ˛, Ski, K., and Kleka, P. (2018). The influence of audio-visual interactions on the annoyance ratings for wind turbines. doi:10.1016/j.apacoust.2017.08.003.

Taylor, S. M. (1982). A comparison of models to predict annoyance reactions to noise from mixed sources. *J. Sound Vib.* 81, 123–138. doi:10.1016/0022-460X(82)90181-X.

TomTom Poznań traffic congestion statistics | TomTom Traffic Index. Available at: https://www.tomtom.com/en_gb/trafficindex/city/poznan [Accessed November 19, 2018].

Torija, A. J., and Flindell, I. H. (2014). Differences in subjective loudness and annoyance depending on the road traffic noise spectrum. *J. Acoust. Soc. Am.* 135, 1–4. Available at: http://scitation.aip.org/content/asa/journal/jasa/135/1/10.1121/1.4842456 [Accessed May 27, 2016].

Trollé, A., Marquis-Favre, C., and Klein, A. (2013). Acoustical indicator of noise annoyance due to tramway in in-curve operating configurations. in, 040023–040023. doi:10.1121/1.4799656.

Trollé, A., Marquis-Favre, C., and Klein, A. (2014). Short-Term Annoyance Due to Tramway Noise: Determination of an Acoustical Indicator of Annoyance Via Multilevel Regression Analysis. *Acta Acust. united with Acust.* 100, 34–45. doi:10.3813/AAA.918684.

Trombetta Zannin, P. H., and Bunn, F. (2014). Noise annoyance through railway traffic - a case study. *J. Environ. Heal. Sci. Eng.* 12, 14–14. doi:10.1186/2052-336X-12-14.

van den Berg, F., Verhagen, C., and Uitenbroek, D. (2015). The relation between self-reported worry and annoyance from air and road traffic. *Int. J. Environ. Res. Public Health* 12, 2486–500. doi:10.3390/ijerph120302486.

Van Renterghem, T., Bockstael, A., De Weirt, V., and Botteldooren, D. (2013). Annoyance, detection and recognition of wind turbine noise. *Sci. Total Environ.* 456–457, 333–345. doi:10.1016/j.scitotenv.2013.03.095.

Velardi, L., Hermand, J.-P., and D'Autilia, R. (2017). On timbre in urban soundscapes: The role of fountains. *J. Acoust. Soc. Am.* 141, 4017–4017. doi:10.1121/1.4989238.

Vogiatzis, K., and Remy, N. (2017). Soundscape design guidelines through noise mapping methodologies: An application to medium urban agglomerations. *Noise Mapp.* 4, 1–19. doi:10.1515/noise-2017-0001.

Vos, J. (1992). Annoyance caused by simultaneous impulse, road-traffic, and aircraft sounds: a quantitative model. *J. Acoust. Soc. Am.* 91, 3330–3345. Available at: http://scitation.aip.org.aipjournalsamericaninstituteofphysics-thejournal97.bu-169.bu.amu.edu.pl/docserver/fulltext/asa/journal/jasa/91/6/1.402823.pdf?expires=1456313287&id=id&accname=2104510&checksum=87A28D4E07E6A2B473E8A4C26FA2CEF6.

Wei, W., Renterghem, T. Van, De Coensel, B., and Botteldooren, D. (2016). Dynamic noise mapping: A map-based interpolation between noise measurements with high temporal resolution. *Appl. Acoust.* 101, 127–140. doi:10.1016/j.apacoust.2015.08.005.

Wojaczek, K. (2016). Słuchawkowa i ambisoniczna subiektywna ocena bodźców środowiskowych. Oprogramowanie i analiza wyników.

World Health Organization (2011). WHO | Burden of disease from environmental noise - Quantification of healthy life years lost in Europe. World Health Organization Available at: http://www.who.int/quantifying_ehimpacts/publications/e94888/en/ [Accessed July 19, 2018].

World Health Organization (2018). Environmental Noise Guidelines for the European Region (2018). World Health Organization.

Yano, T., Morihara, T., and Sato, T. (2005). Community response to Shinkansen noise and vibration: a survey in areas along the Sanyo Shinkansen Line. in *Forum Acusticum*, 1837–1841. Available at: http://webistem.com/acoustics2008/acoustics2008/cd1/data/fa2005-budapest/paper/353-0.pdf [Accessed April 26, 2017].

Yokoshima, S., Morihara, T., Ota, A., and Tamura, A. (2008). Reanalysis of dose-response curves of Shinkansen railway noise. Available at: http://www.icben.org/2008/pdfs/yokoshima_et_al.pdf [Accessed April 26, 2017].

Zuo, J., Xia, H., Liu, S., and Qiao, Y. (2016). Mapping Urban Environmental Noise Using Smartphones. *Sensors* 16, 1692. doi:10.3390/s16101692.

Appendix A: The preliminary survey filled by respondents at the beginning of the experimental procedure

Dear Sir or Madam,

I would like to thank you for your interest in this research. Its main objective is to gather data about annoyance caused by different noise sources met in cities. This experiment is anonymous and the information about gender, age or postal code will be used exclusively in statistical and comparative purposes.

At the beginning please fill in the short survey which consists of both metrical and detailed questions – the latter concerning the attitude towards various noise sources. The main part of the experiment, i.e. listening to the sources of noise in different conditions, will then be carried out.

Important: During the experiment please assume that you are at home and want to rest/relax. When the noise will be played, please take and read a book. After each noise scenario you will be asked to fill in a short questionnaire about it.

The survey uses on a numerical scale (from 0 to 10) are used. When answering them, please mark only the **whole numbers** (which are marked on the scale with small vertical lines). Do not select points between them.

1. Gender (W/M)…………………………………………………..

2. Age………………………………………………......................

3. Postal code of the place of residence…………………..

4. Please think about the last 12 months, when you are here at home. Next is a zero to ten opinion scale for how much noise bothers, disturbs or annoys you when you are here at home. If you are not at all annoyed choose zero, if you are extremely annoyed choose ten, if you are somewhere in between choose a number between zero and ten. What number from zero to ten best shows how much you are bothered, disturbed, or annoyed by noise in the place of living?

0 1 2 3 4 5 6 7 8 9 10

5. Please indicate those noise sources (with the ‘x’ sign) which you consider to be annoying in your place of residence:

Road traffic	
Trams	
Aircraft	
Trains	
Industry	
Sports/Entertainment	
Neighbors	
Others (please describe):…………………………………………………………….	

6. The following table contains 13 statements. For each of them, please spontaneously select an answer assessing how much it is in line with your feelings and mark it in the appropriate column with the ‘x’ sign. Possible responses are indicated by digits:

1 – Strongly disagree
2 – Rather disagree
3 – Rather agree
4 – Strongly agree

Statement	**1**	**2**	**3**	**4**
I need an absolutely quiet environment to get a good night's sleep.				
I need quiet surroundings to be able to work on new tasks.				
When I am at home, I habituate to noise quickly.				
I become very agitated if I can hear someone talking while I am trying to fall asleep.				
I am very sensitive to neighbourhood noise.				
When people around me are noisy I don’t get on with my work.				
I am sensitive to noise.				
My performance is much worse in noisy places.				
I do not feel well rested if there has been a lot of noise the night before.				
It would not bother me to live in a noisy street.				
For a quiet place to live I would accept other disadvantages.				
I need peace and quiet to do difficult work.				
I can fall asleep even when it is noisy.				

7. The following table contains 10 statements. For each of them, please spontaneously select an answer assessing how much it is in line with your feelings and mark it in the appropriate column with the ‘x’ sign. **Note: Unlike the previous question, you can choose from five (rather than four) options.** Possible responses are indicated by digits:

1 – Strongly disagree
2 – Rather disagree
3 – Neither disagree not agree
4 – Rather agree
5 – Strongly agree

Statement	**1**	**2**	**3**	**4**	**5**
Neighborhood of my place of residence is noisy					
I am convinced that noise deteriorates my health					
I am annoyed by the noise in the workplace					
I do not like flying					
I do not like using trams					
I prefer to use a car instead of other means of transport					
I consider myself to be a proecological person					
My house/apartment is well-insulated from the outside noise sources					
Noise produced by my neighbors/coming from a staircase annoys me at home					
I am afraid that noise sources devalue my property					