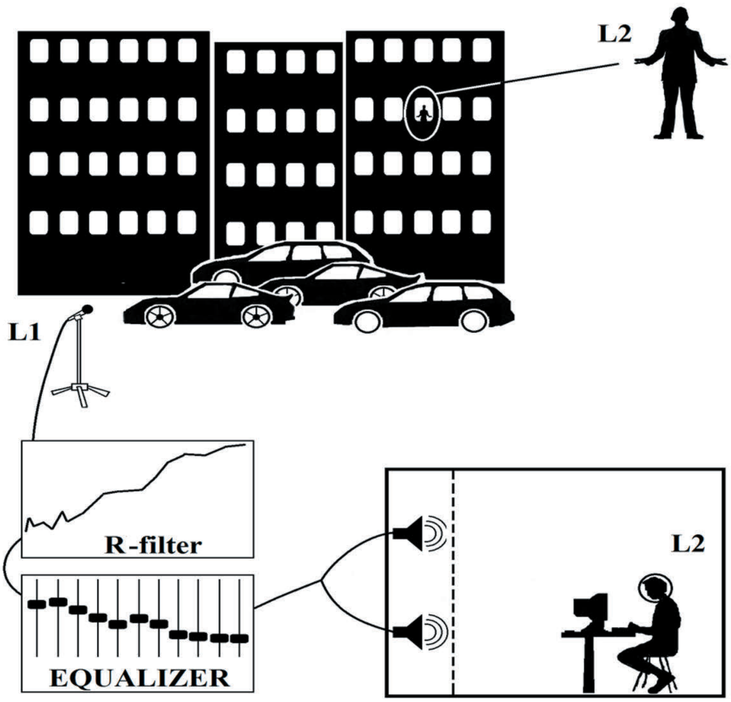
TURUN YLIOPISTO
UNIVERSITY OF TURKU



SUBJECTIVE LISTENING EXPERIMENTS FOR ANNOYANCE INVESTIGATION

David Oliva Elorza

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1650 | MEDICA – ODONTOLOGICA | TURKU 2022

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

SUBJECTIVE LISTENING EXPERIMENTS FOR ANNOYANCE INVESTIGATION

David Oliva Elorza

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1650 | MEDICA – ODONTOLOGICA | TURKU 2022

University of Turku

Faculty of Medicine
Institute of Biomedicine
Medical Physics and Engineering
Doctoral Programme in Clinical Research

Supervised by

Adjunct Professor, Juhani Soini
Turku University of Applied Sciences
Turku, Finland

D.Sc. (Tech.), Panu Maijala
Patria Oyj
Tampere, Finland

Reviewed by

M.D., Docent, Markku Sainio
Helsinki University Hospital
Helsinki, Finland

D.Sc. (Tech), Nick Zacharov
Meta – Reality Labs
Paris, France

Professor, Catherine Lavandier
Cergy-Pontoise University
Cergy-Pontoise, France

Opponent

D.Sc. (Eng), Professor (Associate),
Rodrigo Ordoñez
Aalborg University
Aalborg, Denmark

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover Image: David Oliva

ISBN 978-951-29-8941-6 (PRINT)
ISBN 978-951-29-8942-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2022

To my mum, my wife, and my kids.

UNIVERSITY OF TURKU
Faculty of Medicine
Medical Physics and Engineering
DAVID OLIVA: Subjective Listening Experiments for Annoyance Investigation
Doctoral Dissertation, 151 pp.
Doctoral Programme in Clinical Research
September 2022

ABSTRACT

Noise limits and guidelines that consider only the sound pressure level or the loudness of noises are not efficient in protecting people from all the adverse effects of noise. Other physical characteristics, e.g., tonality, modulation, and frequency content, should also be considered, especially when the noise level is low and it cannot cause hearing risk, but might lead to annoyance and disturbance. Annoying noises have an impact on health and well-being, but this impact and its relationship with the physical properties have not been sufficiently studied. Subjective annoyance caused by noises like those we experience in living spaces and offices should be further investigated via psychoacoustic laboratory experiments. The primary aim of this work was to develop systematic, effective, and reliable methodology to perform this type of psychoacoustic tests. The secondary aim was to investigate the objective metrics that best predict subjective annoyance in four typical noise conditions: ventilation noise in office spaces, traffic noise in homes, neighbors' noise in homes, and noises with tonal components in homes. The main result was the development of the methodology, which in turn enabled us to define our own standards and guidelines. Furthermore, we identified the objective metrics that best correlated with subjective annoyance in each one of the four studied noise situations. In offices, five metrics predicted subjective ratings reasonably well. Noise with sound energy at higher frequencies was less tolerated. Noise with a slope of -7 dB per octave band increment resulted in the highest satisfaction. In dwellings, related to neighbors' living sounds, four metrics of airborne sound insulation performed well to predict annoyance. We demonstrated that 50–80 Hz bands should not be included in the objective rating. In dwellings, related to five types of traffic noise transmitted through façade elements, one metric $R_{\mathrm{w}}+C_{50\text{–}3150}$ performed significantly better than the others. The last experiment proved that tonality is not properly considered in current standards and noise guidelines. The performed psychoacoustic research demonstrated that other physical properties than the sound pressure level should be considered when assessing noise annoyance, and it provided evidence to the objective metrics that would make noise guidelines more efficient with respect to health protection.

KEYWORDS: Noise, annoyance, loudness, perception, psychoacoustics, listening tests, perceptual evaluation, metrics, living spaces, offices, traffic, background noise

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Lääketieteellinen Fysiikka ja Tekniikka
DAVID OLIVA: Subjektiivisia kuuntelukokeita häiritsevyyden tutkimiseksi
Väitöskirja, 151 s.
Turun kliininen tohtoriohjelma
Syyskuu 2022

TIIVISTELMÄ

Melurajat ja ohjeet suojelevat ihmisiä melun haitallisista vaikutuksista, mutta ne ottavat enimmäkseen huomioon vain melun äänenpainetason tai voimakkuuden. Muut fyysiset ominaisuudet, kuten kapeakaistaisuus, modulaatio ja taajuussisältö, joilla on selvä vaikutus subjektiiviseen kokemukseen ja häiritsevyyteen, jätetään usein huomiomatta. Ärsyttävät äänet saattavat noudattaa lakia niiden kielteisistä vaikutuksista huolimatta, koska niiden äänenpainetaso ei ylitä yhtään melurajaa. Asuintilojen ja toimistojen melun aiheuttamaa subjektiivista ärsytystä tulisi tutkia tarkemmin psykoakustisten laboratoriokokeiden avulla. Työn ensisijaisena tavoitteena oli kehittää järjestelmällinen, tehokas ja luotettava menetelmä tämän tyyppisten psykoakustisten testien suorittamiseksi. Lisäksi selvitettiin, mitä muita objektiivisia mittareita, kuin äänenpainetaso tai äänenvoimakkuus, ennustavat parasta subjektiivista ärsytystä ja häiritsevyyttä. Työssä tutkittiin neljää tyypillistä meluolosuhdetta: toimistotilojen ilmanvaihdonääniä, kaupungin liikenteen melua kodeissa, naapurin melua kodeissa, ja kapeakaistaisia komponentteja sisältävää melua. Päätuloksena oli menetelmän kehittäminen, joka mahdollisti omien standardien ja toimintaohjeiden määrittämisen. Lisäksi tunnistettiin objektiiviset mittarit, jotka korreloivat paremmin subjektiivisen häiritsevyyden kanssa kussakin neljästä tutkitusta melutilanteesta. Toimistoissa viisi mittaria ennusti kohtuullisen hyvin subjektiivisia luokituksia. Kohinaa, joka kuului korkeammilla taajuuksilla toimivalla äänenergialla, siedettiin vähemmän. Asunnoissa, kun asumisääniä syntyy naapurin asunnossa, neljä ilmaääneneristysmittaria toimi hyvin ennustamaan asukkaiden subjektiivista ärsytystä. Osoitettiin, että 50–80 Hz: n kaistoja ei pitäisi sisällyttää objektiiviseen luokitukseen. Myös asunnoissa, liittyen viitteen eri liikennemeluun kantautumassa sisätilaan julkisivuelementtien kautta, yksi metrinen R_w+$C_{50–3150}$ toimi huomattavasti paremmin kuin muut. Viimeinen koe osoitti, että tonaalisuutta ei oteta asianmukaisesti huomioon nykyisissä standardeissa ja meluohjeissa. Tämä tutkimus osoitti, että oikein suoritetut psykoakustiset kokeet tarjoavat laadullista ja määrällistä tietoa subjektiivisesta häiritsevyydestä, ja että näiden tietojen perusteella voidaan määrittää objektiiviset mittarit, jotka tekisivät ohjearvoista tehokkaampia melun haitallisilta vaikutuksilta suojauduttaessa.

AVAINSANAT: Melu, häiritsevyys, äänekkyys, kokemus, psykoakustiikka, kuuntelukokeita, havainnollinen arviointi, mittareita, asumistilat, toimistot, liikennemelu, taustamelu

Table of Contents

Abbreviations

AM	Amplitude modulation
dB	Decibel, unit for sound pressure level
Hz	Hertz, international unit for frequency
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
$L_{A,eq}$ or precisely $L_{p,A,eq}$	A-weighted equivalent sound pressure level
$L_{C,eq}$ or precisely $L_{p,C,eq}$	C-weighted equivalent sound pressure level
$L_{N,ISO}$	Loudness according to ISO 532-1:2017
$L_{N,ANSI}$	Loudness according to ANSI S3.4-2007
LFN	Low frequency noise
NCB	Balanced noise criterion
Phon	Unit of measurement of loudness level
RTN	Road traffic noise
R_w	Weighted sound reduction index according to ISO 717-1:2013
SIL	Speech interference level
SNQ	Single number quantity
SPL	Sound pressure level
STC	Sound transmission class according to ASTM E413-10:2010
STI	Speech transmission index
WTN	Wind turbine noise

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Hongisto, V., Oliva, D., and Rekola, L., 2015. Subjective and objective rating of spectrally different pseudorandom noises – Implications for speech masking design. The Journal of the Acoustical Society of America, 137(3), pp.1344–1355.

II Hongisto, V., Oliva, D., and Keränen, J., 2014. Subjective and objective rating of airborne sound insulation–living sounds. Acta Acustica united with Acustica, 100(5), pp.848–863.

III Hongisto, V., Oliva, D., and Rekola, L., 2018. Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. The Journal of the Acoustical Society of America, 144(2), pp.1100–1112.

IV Oliva, D., Hongisto, V., and Haapakangas, A., 2017. Annoyance of low-level tonal sounds–Factors affecting the penalty. Building and Environment, 123, pp.404–414.

1 Introduction

Noise, non-desired sound (Smedley 1845), is physically present around us 24/7 in indoor and outdoor spaces (Berglund 1995). Noise is normally experienced at bearable sound pressure levels, and so typically is not a hearing risk (Sulaiman et al. 2013, Themann 2019). In dwellings and offices spaces, the sources of noise are ventilation systems, devices of any type, and human speech (Alonso et al. 2020, Hongisto 2017). Outdoors, noise is created by traffic, city activities, industry, and other sources of environmental noise (Veitch et al. 2002). Non-desired sound is basically everywhere always, and it affects the health and well-being of millions of people all around the world (WHO 2018, Di Blasio 2019). It is well accepted that environmental noise is an important public health issue. The recent report "Environmental Noise Guidelines for the European Region" (WHO 2018) even stated several non-auditory effects that can be produced by noise as we experience it; ischemic heart disease, hypertension or strokes, cognitive development, sleep disturbance, and other variables related to metabolic diseases and quality of life in general. In Europe, environmental noise is a public health issue, and its negative impacts on human health and well-being are a concern to both the general public and policymakers (European Commission 2002). New guidelines have been updated recently for environmental noise (WHO 2018).

Noise level limits are applied in most countries to protect society from the adverse effects of noise on health (Kim & Berg 2010, WHO 2018). The measurements of the noise are done by following acoustic standards, and the measured values are compared to the corresponding limits (Śliwińska-Kowalska & Zaborowski 2017). From a legal point of view, the law is clear when it describes how to measure and what are the specific limits. However, noise limits are not harmonized between countries (Kylliäinen 2016, Kylliäinen 2017). In Finland, for instance, noise limits to protect health from noise have been defined by the Ministry of the Environment (ME 2017) and by the Ministry of Social Affairs and Health (MSAH 2015). At European level, efforts are continuously made to define and apply harmonized methods and limits, for instance, COST T0901 action (Rasmussen 2010, Rasmussen 2019).

$L_{A,eq}$, the A-weighted equivalent sound pressure level, is the metric most used to quantify noise (IEC 61672-1:2013). At homes and office spaces, $L_{A,eq}$ values are most of the time between 30 and 60 dB, which is about 20 dB under the hearing risk level and 20 dB above the hearing threshold (ISO 226:2003). This metric relates to loudness and not to the annoyance of the sound (Nilsson 2007, McMinn 2013). It can provide identical values to differently experienced noises, and thus often it is not suitable to consider annoyance, where other factors also moderate the experience. Annoyance is the feeling of displeasure associated with any certain agent (Guski et al. 2017), and noise annoyance is one of the major effects of noise (Guski et al. 1999). Noise annoyance is a multi-faceted psychological and subjective concept where individual emotions and attitudes combine with numerous types of noises, situations, and moderators. Hence the difficulty to obtain an objective metric to describe it well and accurately. To understand the complex notion of annoyance better, as we did in this work, looking for other measurable properties and physical quantities of noise, such as spectrum, tonality, amplitude modulation, roughness, and sharpness is needed (Zwicker & Fastl 2013, Kuttruff 2016). Parameters quantifying these aspects could be used in models predicting when annoyance occurs or what the impact on health and its magnitude are. In addition to the physical characteristics, there is a relationship between annoyance and exposure time (Brink et al. 2019, Zimmer et al. 2008). They can also be used to develop better quantitative and objective methods to improve noise guidelines, and thus more studies to verify validity of the limits are justified.

With respect to the importance of studying noise perception, there is a special characteristic of our body that is good to keep in mind. First, hearing is the only one of the five human senses, unless damaged, that would require an external object to block it—so to say. We can close our eyes to avoid seeing, we can breathe through our mouth to prevent us from smelling unpleasant odors, we eat normally only what tastes good, and we easily can keep our hands in our pockets when something is not to be touched. But we cannot close our ears or shut our hearing system voluntarily. For many of us, it is very difficult to escape from a certain noise once it is heard. Before, noise was defined as non-desired sound, and now also as sound that impacts negatively on the well-being and health of humans. It annoys, disturbs, and affects our capacity to rest (Hume et al. 2012) or to concentrate while performing cognitive demanding or other types of tasks (Haka et al. 2009). A priori, noise triggers some reactions in our brain that are related to displeasure or other negative effects. Cavanna (2014) investigated how these reactions work at the clinical level.

We spend a great part of our lives in dwellings and indoor spaces and so it makes sense to investigate how annoyance takes place in these surroundings and how to measure it. The aim is to obtain objective noise metrics that correlate with subjective annoyance and not with subjective loudness. The benefit would be to consider noise risks on health and well-being better.

This research of noise perception may bring other benefits to society, for instance because of its applicability in civil and industrial engineering. First, it could help architects and engineers to determine the performance of building constructions with respect to the protection from noise annoyance. That would allow to classify building constructions, like floors and façades, with respect to their performance in perceptual terms, and enable the optimization of certain layers of the construction according to achieved reduction of annoyance perception. Second, manufacturers of devices, e.g., ventilation outlets or machines and basically any type of hand-held devices, can use that knowledge to optimize the noise emission spectra produced by their machines and to provide better acoustic satisfaction in the spaces where those noises are being heard (Park et al. 2020, Wu et al. 2020).

Our work concentrates on the psychoacoustics research of sounds with $L_{\mathrm{A,eq}}$ between 25 and 50 dB, well below the levels known to cause adverse audiological effects. Psychoacoustics methods were used to understand better the relationships between the physical characteristics of sounds and their perceptual attributes (Moore 2014). The primary aim was to study subjective annoyance perception of noises with sound pressure levels like those typically existing in dwellings and offices. The research sides with environmental medicine and public health science. The methods comprised listening experiments with test subjects, which we hope were optimized in terms of effectiveness and efficiency. Each of the Publications I–IV originates from independent experiments and noise situations. All experiments were short term in duration, under 90 minutes. The subjects were exposed to a certain number of sounds for periods of time shorter than half a minute. Subjects self-reported in rating scales the annoyance experienced in each case. Each experiment was designed to answer at least one research question. We analyzed the physical characteristics of the test sounds producing more and less annoyance and made conclusions about how this type of noise could be quantified better in that context. Our approach during the research was to determine which ones, from all available acoustic metrics or parameters, provided the best correlation with the ratings provided by the test subjects. Perhaps one day the presented data and evidence supports future noise limits and guidelines.

During a decade, our research group investigated in several ways the subjective impact of noise on humans and the way to quantify it with objective metrics. The output of that work includes several scientific publications. Thirteen papers published in peer-review journals were co-written by the author of this thesis (Haapakangas et al. 2011, Hongisto et al. 2014, Hongisto et al. 2015, Hongisto et al. 2016, Kylliäinen et al. 2016, Virjonen et al. 2016, Hongisto et al. 2017a, Hongisto et al. 2017b, Kylliäinen et al. 2017, Oliva et al. 2017, Hongisto et al. 2018, Hongisto et al. 2019, Haapakangas et al. 2020). Publications I–IV are part of this collection of references. Most of the work was performed during two research projects, ÄKK (2012–2014) and Annoyance (2016–2019). The research team acted in the Finnish Institute of Occupational Health until 2016. The team moved to Turku University of Applied Sciences while the second project was still taking place.

2 Review of the Literature

The following literature review focuses on subjective annoyance caused by noises with a level range of $L_{A,eq}$ = 25–50 dB, and with physical characteristics similar to the noises typically present in homes and offices spaces. The review occupies seven sections, and it is divided in two main parts. The first provides a historical overview of the listening experiments related to hearing sensitivity and subjective loudness and annoyance. The second part of the review concentrates explicitly on papers directly related to Publications I–IV. Cross-sectional field surveys were not applied in this work, and so they are only discussed later in Appendix 1.

2.1 Noise and health

The impact of noise on health is studied in environmental medicine as part of public health science. The impact of noise on health depends on many factors, the most important being its physical properties, i.e., sound pressure level, the exposure length, and the time at which the event(s) occurs. Intrinsic factors of the noise, like its nature and several physical properties, and extrinsic factors, like the possibility to control the noise or the occupation during the exposure, have been related to its impact on health. Health effects on people can be divided in physiological and psychological, Figure 1. The figure organizes the physiological and psychological health effects using three dimensions; horizontally the sound pressure level $L_{A,eq}$ and vertically the incidence. Incidence refers to the number of individuals who develop the specific health-related event. A third dimension has been added vertically to the right of the figure to represent the severity, but it remains uncertain how to sort the effects with respect to their seriousness. The figure has been created for this work and has not been taken from any reference so small deviations from other works might be expected.

Hearing impairment is the best documented adverse effect of loud noise exposure. In occupational industrial environments, hearing damage has been shown to occur after long and short-term exposures of $L_{A,eq}$ > 80 dB and $L_{C,eq}$ > 140 dB, respectively (Manninen & Aro 1079). Specific protection codes in most countries use these levels as limits to the allowable noise exposure levels that should not be exceeded in working environments. Non-auditory adverse effects include cardiovascular reactions. Raised blood pressure and hypertension can appear already at lower levels, $L_{A,eq}$ > 70 dB, (Knipschild 1977, Lercher et al. 1993, Muzet & Ehrhart 1980, Stansfeld et al. 2000). Heart rate and depth increments (Knipschild 1977, Parrot et al. 1992), and changes in respiration rate, headache and nausea have also been reported (Crook & Langdon 1974).

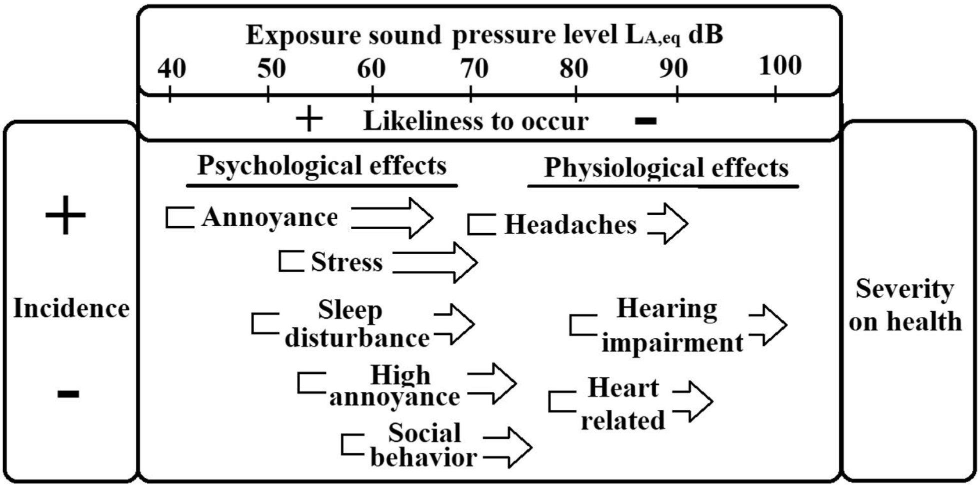


Figure 1. The most common psychological and physiological health effects of noise are ordered according to the sound pressure level $L_{A,eq}$ of exposure, also considering the likeliness of occurrence in society's population and the severity on health in terms of damage. This figure has not been taken from any previous source. The orders of the effects with respect to the three dimensions is based on estimations of the author from the numerous sources referred in this text.

Community noise is usually at levels below $L_{A,eq}$ 70 dB, where most of the psychological effects occur. The level range from 30 to 70 is of interest, as it is what we mostly experience daily. Outdoors, the sound pressure level is often 40 to 70 dB $L_{A,eq}$, while in indoor spaces, like dwellings and offices, it varies mostly between 30 and 60 dB. The most important psychological effects, annoyance, stress and high annoyance, sleep disturbance, and social behavior changes are also ordered in Figure 1, according to dimensions $L_{A,eq}$ and incidence. The order of severity of these effects in the figure is only an estimation of the author.

Annoyance is the most widespread and well-documented psychological response to noise (Stansfeld & Brown 2000). It starts to appear in some individuals at levels as low as 40 dB, and so its incidence is the highest. Most probably annoyance occurs before any other psychological effect appears, and it is related to the habituation and acceptance processes. Sleep disturbance might occur without previous annoyance symptoms. Annoyance includes fear and anger related to the belief of being harmed (Cohen & Weinstein 1981), which may act as a health stressor increasing the annoyance experience. Social behavior changes could occur in some individuals at higher noise levels, after episodes of frustration when communication is strongly impaired. A long-term effect can be unwanted aversive changes in affective state and social behavior (Job 1996). Jones et al. 1981 have discussed the psychological impact in urban communities in relation to willingness to help, increased aggression, and in processing attentional social cues (Jones et al. 1981). These effects are severe, but less likely to occur, and so they have been less studied in the literature. Interference of noise with work actions of cognitive complexity, communication, and other daily activities like sleeping and reading were reported by Szalma & Hancock (2011).

Stress and high annoyance are less likely to occur, but their severity is higher. Moderators, subjective opinions, and uncontrolled continuous exposure to the noise may impair habituation and acceptance.

Sleep disturbance and high sleep disturbance relate very much to aircraft traffic near airports. The impact of this and other traffic noises on sleep quality has been studied in terms of sleep awakenings and sleeping times in several studies (Öhrström & Rylander 1982, Öhrström 1989, Myllyntausta et al. 2020). Sleep disturbance depends on the levels experienced in the sleeping location and on other physical characteristics of the noise, as impulsiveness, nature of the sound, and temporal distribution. It is known that the elderly and young are the most and less susceptible groups to suffer from sleep disturbance and high sleep disturbance, but apart from this, the health impact is most conveniently studied based on subjective and attitude factors, and not age or gender.

The impact on health of annoyance, stress, high annoyance, and social behavior depends very much to each person's subjective feelings associated with the noise, e.g., the degree of interference with activities or previous history (Schreckenberg et al. 2010), and thus it is important to understand how these may impact on individuals. Noise sensitivity is a variable often used to explain the variance between individual psychological responses under the same type of exposure (Stansfeld 1992, Stansfeld et al. 2000, Weinstein 1978). After the work of Weinstein, noise sensitivity has been measured in noise studies with 21 questions. Weinstein said that noise sensitivity represents two aspects: being critically discriminating about the environment and having higher neuroticism scores. Neuroticism is a negative emotional arousal causing poorer ability to manage psychological stress and a tendency to complain (Anderson 1971, Ormel et al. 2012). This and other psychiatric disorders are rare, and they are likely to occur in persons with other mental health symptoms. Van Dijk 1986 mentioned that ill mental health might produce psychiatric disorders, where the noise annoyance may act as feedback of the negative impact, here acting as a health stressor. Cases of phobic disorders and neurotic depression were reported for instance by Wing et al. 2012.

In the context of community noise and health, moderators also have an impact. They are the personal and social aspects of the residents that modify subjective perception and annoyance, including past disturbances, attitudes, and expectations. Personal moderators are sensitivity to noise, fear of harm connected with the source, personal evaluation of the source, and coping capacity with respect to noise. Social moderators are general social thoughts about the source, trust or misfeasance with source authorities, history of noise exposure, and expectations of residents (Guski 1999).

To reduce the negative impact of noise on health and society, it is important to consider the likelihood of noise events in daily activities and the possibility to reduce or control them. Control strategies mainly include improving the sound insulation quality of constructions, placing absorbents to reduce the overall levels and reverberation times, and other strategies to reduce likeliness of appearance. In offices, irrelevant speech has been reported to be the most annoying type of noise, mostly because of its ability to impair cognitive demanding tasks (Haka et al. 2009, Salame & Baddeley 1982). The most effective means of action for this type of spaces concentrate on reducing the level of the irrelevant speech and its intelligibility. By contrast, in dwellings and living spaces, most typical sources of annoyance are traffic coming through the façade constructions and noise created in neighboring spaces. Living noise can be airborne, but also structure-borne like impact noises in floors. From an engineering point of view, the insulation quality of the dividing element can be measured and quantified with respect to any situation, but it is important to ensure that selected metrics suit the specific case well. Each existing building code applies any of the available weighting methods, but most often these are based on loudness and not on annoyance data. The ongoing discussion, see for instance Kylliäinen et al. 2016, Virjonen et al. 2016, and Kylliäinen et al. 2017, supports more studies and development to improve the annoyance-based noise metrics and models.

2.2 Health impact research with listening experiments

Listening experiments have been designed and applied to study sound and noise perception for numerous purposes. It is not possible in this thesis to specify all of them neither to discuss or enumerate their findings. Nevertheless, the following text presents three types of experiments to illustrate their applicability.

Audibility and hearing sensitivity experiments, like Whittle et al. 1972 and Møller & Andresen 1984, have shown the complexity of our hearing system. Audibility studies determined that loudness perception is not constant along the audible frequency range, and that equal changes of sound pressure level do not produce comparable changes in loudness experience. Human hearing is explained in the next section.

The second group is formed by sound quality evaluation experiments, which is the psychological technology that analyzes the sound in quantitative ways according to the feeling of humans. It attempts to objectively qualify and quantify the quality of sound based on human hearing standards. They can be subdivided in further groups. First, sound quality experiments aiming to identify the physical characteristics and the metrics that correlate best with subjective experience in system quality terms. The outcome of these studies are sound quality metrics, like loudness, roughness, sharpness, and many others. The downside of these metrics is that they can be calculated from sound pressure level measurements, but they require more elaborated mathematical methods and calculations (see for instance the work of Fastl 2006 and Zwicker & Fastl 2013). For instance, Jeon et al. 2011 showed that subjective perception of the noise produced by air-condition appliances could be predicted when tonality and fluctuation strength metrics were considered in addition to the sound pressure level. A special group of sound quality experiments focuses on analyzing the quality of a digital signal with respect to the original sound that it aims to mimic. Sound quality experiments were also applied to improve the quality of audio in videogames (Neidhardt et al. 2017, Pouru 2019). From an audiological perspective, Pouru investigated sound perception in videogames. He tested what combinations of directivity patters provided the best immersion experience. Eighteen sound stimuli were created from the combination of three sources (cat's meow, water flushing, typing machine) and six parameters related to the directivity of listener's hearing. The sounds were presented randomly using headphones and virtual reality helmet. In this case, the visual stimulus was a 360 degrees image with mountains but lacking visual details in the close field. In each sound, the test subject had to identify while rotating and using their own hearing sense, both the front and back directions where the sound was coming from, and then rate how realistic on a 1–5 scale the experience was in general. It was found that front and back positions were always identified, but only two of the directivity patterns produced acceptable immersion experience when rotating.

The third type of listening experiments study the impact of noise on health and the different adverse outcomes reviewed in the previous section. Zimmer et al. 2008 demonstrated that annoyance was influenced by the task at hand and by the degree of interference with the task produced by the sound. Irrelevant speech, or background speech, has been reported to be the most bothersome noise source in the office environment (Sundstrom et al. 1994). Haapakangas et al. (2011) studied with 54 subjects the impact of five speech masking conditions on performance in cognitive demanding tasks. Silence was the best situation for acoustic comfort and concentration. When background speech was present, water-based noise was rated as acceptable masker, while music was not. A recent experiment by Dedieu (Dedieu et al. 2020) investigated people's strategies to mask neighbors' noise to improve acoustic pleasantness in living rooms. A group of participants chose to mask neighbors' noise with ventilation noise, while others preferred to disguise the noise from the adjacent apartment with environmental noise. Finally, other studies focused on sleep and stress. The effect of traffic noise on sleep quality was studied by Myllyntausta et al. 2020, and the stress effect of noises while performing cognitive tasks was investigated by Radun et al. 2021 and Haapakangas et al. 2011. Mucci et al. 2020 have recently published a review on urban noise and psychological distress.

2.3 Loudness perception

Our hearing system is the result of an evolutionary process, and this development was greatly influenced by the need to use oral communication and by the sounds around us. The pioneering studies on loudness of Fletcher & Munson in 1933 and Kryter & Pearsons in 1965 provided a good starting point regarding the sensitivity of our hearing system. Subsequent studies using pure steady tones and broadband sounds have helped to determine the relationship between sound pressure level and loudness perception. Normal hearing values are standardized in ISO 226:2003. The last revision of ISO 226 presents the equal-loudness level contours, see Figure 2.

The unit of measurement for loudness levels is the phon, Hz is the international unit for frequency (horizontal axis), and dB is the unit for sound pressure level (vertical axis). Sound pressure level is a logarithmic measure of the effective pressure of a sound relative to the ambient atmospheric pressure. Therefore, loudness is the human perception of sound pressure level. The equal-loudness contours explain the way we hear and how sounds of distinct pitch compare with each other. Our hearing system does not perform linearly, and for instance we are more sensible to frequencies between 1000 and 3000 Hz. A dip representing this effect can be seen in the equal-loudness contours. The curves represent what energies at each frequency produce similar loudness experience when compared to each other. The curves are not either parallel, and thus equal variations of sound pressure level at two separate frequency regions do not lead to the equal change of loudness experience. For instance, a sound at 63 Hz needs to have a level of 40 dB to be heard, but once it is heard, a small change in its level leads to a larger increment in loudness than what the same increment of level produces at higher pitches.

Figure 2 provides more information. The listening frequency range in the horizontal axis is divided in octave bands, and the sound pressure level (SPL) in the vertical axis is expressed in decibels, dB. Both are expressed in logarithmic scales. The SPL in dB solves how our hearing system responds logarithmically and not linearly to the changes of sonic pressure. The audible frequency range is divided in frequency bands called octaves. This filtering method allows to split the audible spectrum into bands, for instance in 1/1 bands which are centered at 31, 63, 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. In Figure 2, the vertical lines represent the center frequency of the 1/3 octave bands, but only every third of them is labelled in the horizontal axis.

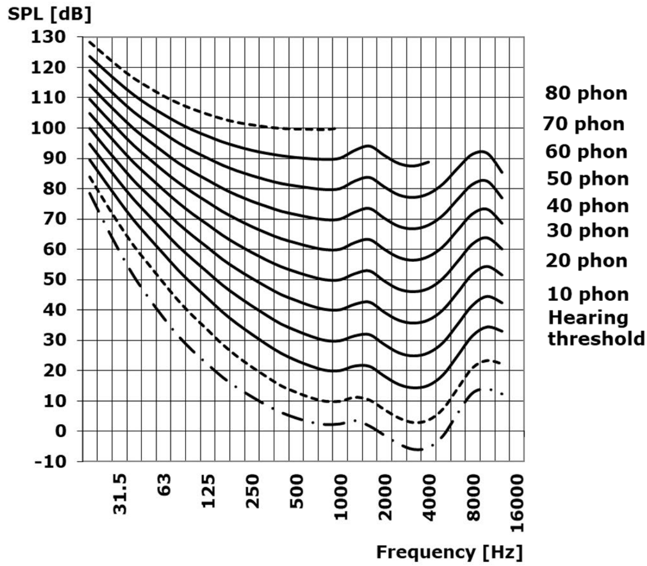


Figure 2. The hearing threshold and the equal-loudness level contours under free-field listening conditions as presented in ISO 226:2003. The horizontal axis is the frequency measured in Hertz (Hz), and the vertical is the sound pressure level measured in dB. The lines express the loudness level in phons at which different frequencies are perceived equally loud. The solid lines represent the loudness level at which different frequencies are perceived equally loud. The lower dashed line represents the threshold of hearing, and the upper dotted line the level of pain to the auditory system.

2.4 Loudness metrics

The most applied metric related to hearing is the A-weighted equivalent sound pressure level, $L_{A,eq}$. It is measured in dB according to IEC 61672-1:2013. The metric compensates the differences in hearing sensitivity at $L_{A,eq}$ = 40 dB, which is the 40-phons curve, calculates the logarithmic sum of the level measured at each frequency, and expresses the sound by a single number quantity. Figure 3 presents the A-weighting network. Its shape resembles the inverse of the 40 phons line shown in Figure 2. Because human hearing is not linear, e.g., explained in the figure by non-parallel curves, using only one curve as reference for both loud and less loud sounds might lead to errors to quantify hearing in terms of loudness. Several studies during the last years, for instance Nilsson 2007, Huang et al. 2008, and Bolin et al. 2014, have studied the suitability of this metric $L_{A,eq}$. It is generally accepted that it works well for loudness with SPLs between 40 and 70 dB (McMinn 2013), but it has also been demonstrated (see next section) how it fails to explain annoyance. Certainly, it is easy to use and available in all sound level meters. Other metrics have been developed, and standardized in some cases, with the aim to quantify loudness experience following other approaches. For instance, ISO includes two methods: ISO 532-1:2017 applies the Zwicker method (Zwicker 1961, Zwicker 1977), and ISO 532-2:2017 applies the Moore-Glasberg method (Glasberg & Moore 2006). ANSI S3.4:2007 is also based on Glasberg & Moore 2006, but it is slightly different than ISO 532-2:2017. ITU-R BS.1770-4 provides a method to determine subjective programmed loudness and true-peak signal level of broadcasted signals.

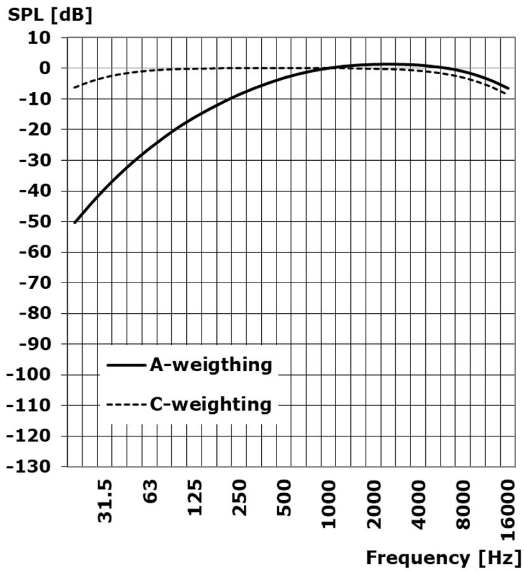


Figure 3. The A-weighting and C-weighting networks as presented in IEC 61672-1:2013. The horizontal axis expresses the frequency in Hz in 1/3 octave bands and the vertical the SPL in dB. The networks enable to calculate the A-weighted and the C-weighted equivalent sound pressure levels.

2.5 Annoyance perception

Annoyance is an unpleasant mental state characterized by irritation, distraction, and discomfort. Related to our topic, noise annoyance is seen as the major effect of noise (Guski et al. 1999) and as a critical health outcome (WHO 2018). At levels below those known to cause physical health effects, noise annoyance, like annoyance in general, has a strong subjective character. Annoyance perception varies between individuals and strong differences are expected. Noise annoyance is multifaceted.

Laird and Coye (1929) considered it like an emotion, because noise was experienced affectively, typically producing displeasure. According to Hall et al. (1985) noise annoyance is the result of disturbance because of the interference of noise with intended activities. Situational and personal factors influence the way we react to noise. The same noise is not perceived equally in unassociated environments or times of the day. Investigations in that field have been done, for instance, by Lee et al. (2021) with construction noise, Benz et al. (2021) with neighbor and living noises, Gille et al. (2016) considering only traffic noise, Zhang and Ma (2021) for railway noise, and by Michaud et al. (2016) for wind turbine noise.

The crucial question is, which parameters and which metrics would help us to predict annoyance? The most common approach is to investigate the correlation between reported annoyance and the physical characteristics of the noise. The characteristic can be any measurable metric, and it is possibly to consider situational factors like location and time of occurrence. Using this approach, new models have appeared recently for: traffic noise by Ascigil-Dincer and Demirkale (2021) and by Bravo-Moncayo et al. (2019), industrial noise by Paszkowski and Loska (2017), and railway noise by Vallin et al. (2018). Other studies developed their models using more elaborate metrics describing special physical characteristics of the noise, like sharpness, roughness, and tonality (Zwicker & Fastl 2013, Kuttruff 2016).

2.6 Discussion of selected literature

Sounds and noises present several physical characteristics other than sound pressure level that impact and affect the way we perceive them (Kuttruff 2016). It has not been possible to find a metric that works well for all types of sounds (McMinn 2013), and hence it would be justified to apply specific metrics for each individual noise situation. The following review presents and discusses psychoacoustic studies with $L_{\mathrm{A,eq}}$ in the 25–50 dB range, because from now on we concentrate only in community noise as experienced in indoor spaces like dwellings and offices. Publications I–IV studied machine noise in offices, living sounds in apartments, traffic noise in dwellings, and machine noise with tonality as special characteristic.

2.6.1 Experiments on *loudness* and *annoyance*

Loudness is an auditory perception while annoyance relates to the opinion or state of mind regarding disturbance or discomfort, see review by Aletta et al. (2018). Both are subjective, and therefore several studies applying psychoacoustics methods have studied how they correlate to each other, and what acoustic metrics predict them best. Acoustic metrics may be single number quantities (SNQ) able to quantify properties of a sound in any way. Works to define the field of range and better metrics than

$L_{A,eq}$ to specific situations have been done. Kjellberg & Goldstein (1985; 12 subjects, 45 sounds, $L_{A,eq}$ = 35–73 dB) showed that $L_{A,eq}$ was not the best metric to predict loudness. Nilsson (2007; 30 subjects, 69 sounds, $L_{A,eq}$ = 47–78 dB) demonstrated that loudness, $L_{N,ISO}$, according to the Zwicker method (ISO 532-1:2017), predicted subjective loudness better than $L_{A,eq}$ for traffic sounds. Torija & Flindell (2015; 33 subjects, 16 sounds, $L_{A,eq}$ = 37–57 dB) investigated the effect on annoyance of traffic noise with different spectra, and presented evidence that $L_{A,eq}$ was not a good predictor of annoyance in traffic noise, and that subjective annoyance and loudness did not correlate well with each other. Park & Bradley (2009) presented linear correlation evidence between subjective loudness and annoyance when investigating neighbors' noises transmitted through wall constructions. Persson et al. (1990; 98 subjects, 20 sounds, $L_{A,eq}$ = 40–70 dB) deducted that $L_{A,eq}$ did not correlate well with annoyance at frequencies below 200 Hz. Poulsen (2003; 18 subjects, 8 sounds, $L_{A,eq}$ = 20–35 dB) did not discuss the correlation between subjective loudness and annoyance, but his data suggested that annoyance did not correlate linearly with A-weighted SPL. In this case, the SPL of the stimuli was low, and the test stimuli had physical characteristics, like impulsiveness and tonality, which demonstrate that $L_{A,eq}$ failed to predict annoyance when these characteristics were present. Ryherd & Wang (2008; 30 subjects, 6 sounds, $L_{A,eq}$ = 46–48 dB) did not find correlation between $L_{A,eq}$ and annoyance in an investigation with tonal sounds, either. Noises with $L_{A,eq}$ = 35–65 dB are not a risk for the damage of the auditory system (Sulaiman et al. 2013, Themann 2019). The studies listed above seem to also indicate that subjective annoyance should not be studied with loudness-only based metrics.

2.6.2 Studies with office noise

Open-plan offices and offices account for over 50% of workspaces in Europe (European Parliament 2020), at least before the Covid-19 pandemic. Research has been done to investigate the stress, tiredness, and the negative impact of noise to work performance and cognitive tasks (Danielsson & Bodin 2009, Jensen et al. 2005). Ventilation noise and the speech from others have been reported to have the highest negative impact on well-being and cognitive tasks in these types of workspaces (Haapakangas et al. 2017, Yadav et al. 2021). Dedicated studies by Haapakangas et al. (2008) and Kaarlela-Tuomaala et al. (2009) demonstrated that in offices the speech by colleagues was the most disturbing and detrimental noise. On the other hand, ventilation noise can be used to reduce the disturbing effect of speech. Veitch et al. (2002; 35 subjects, 15 sounds, $L_{A,eq}$ = 41–44 dB) studied the effect of ventilation noise on cognitive demanding tasks and demonstrated that acoustic satisfaction increased when speech intelligibility decreased. They reported annoyance increments when the masking noise had a spectrum with high frequency content. Ebissou et al. (2015) studied the impact of speech on concentration (57 subjects, 5 sounds, $L_{A,eq}$ = 46 dB). Half of the participants performed identically under all noise situations, i.e. they were insensitive to speech intelligibility, while the other half performed worse when STI (speech transmission level) was 0.45 and above. Hongisto et al. (2016) applied combinations of speech and masking noise (32 subjects, 4 sounds, $L_{A,eq}$ = 33–42 dB), and demonstrated that the ventilation noise in individual offices helps to reduce the problems associated to speech coming from neighboring rooms in cases where the sound insulation properties of the dividing element was weak. A recent work (Lenne et al., 2020) showed that speech masking systems using noises with $L_{A,eq}$ = 45 dB were counter-productive in the long term. Despite the fact that the masking system did significantly reduce the perception of the level of intelligible conversations, it increased, on the other hand, the negative effect on the annoyance caused by equipment noise (the masking sound was perceived by the office workers as ventilation noise). While some studies, like Haapakangas et al. (2011) and Haka et al. (2009), have confirmed the validity and usefulness of speech masking systems, it is not a fast-spreading technology. A long-term in situ experiment to test the validity of water-based sounds for masking was performed by Hongisto et al. (2017a).

2.6.3 Studies with airborne and impact noise from neighbors at homes

We spend, perhaps, over one third of our lives at home, and so it is important to create acoustically comfortable living spaces to secure our rest and well-being protecting us from environmental and living noises. Living noise originates from neighbors' activities, and can be conversation, TV watching, music playing, practice of music instruments, children vividly playing, partying with friends, use of mechanical devices for cleaning or food preparation, etc. (Muellner & Rychtáriková 2013). Living noises can be easily heard if constructions' dividing spaces are weak.

The properties of noises change when they travel from one space to another through the dividing element. The changes in sound pressure level and spectrum can be used to quantify the quality of the construction, but that is not always easy. For instance, if we use the sound reduction index (R_w) as the metric to quantify the quality of a heavy construction (monolithic concrete) and of a light construction (staggered double wall), it could happen that for both we obtain identical values. They would then be equally good in those terms, even though the sound quality of the noise was different. In other words, the metric could block the same number of decibels, but the nature of the transmitted sound differs greatly between the two constructions, as the cement-based construction blocked the low frequencies better while the lighter panel performed better at high ones. The research to verify the performance of sound insulation metrics is abundant. Psychoacoustic experiments related to building acoustics, e.g., Zwicker (1977), Ellermeier et al. (2004), Furihata et al. (2007), Horvart et al. (2012), Fastl (2006), and Zwicker & Fastl (2013), have helped to develop metrics like those applied in ISO 717-1:2013 and ISO 16283-1:2014 to measure airborne sound insulation properties of wall constructions and the impact sound insulation properties of floors. The performance of airborne sound insulation metrics was evaluated further by Park et al. (2008), Park & Bradley (2009), Pedersen et al. (2012), Rychtáriková et al. (2012), and Hongisto et al. (2016). Kylliäinen et al. (2017; 55 subjects, 44 sounds, $L_{A,eq}$ = 16–38 dB) investigated the perception of footsteps from upstairs through nine types of floor constructions. Park and his team published three works devoted to sound insulation metrics. In Park et al. (2008; 15 subjects, 20 sounds, $L_{A,eq}$ = 35 dB) the speech transmission through walls was studied. They found that STC (sound transmission class according to ASTM E413-10:2013) and R_w (weighted sound reduction index according to ISO 717-1:2013) metrics were not good predictors and proposed new ones based on arithmetic average in frequency bands. In the next study by Park & Bradley (2009a), music and speech heard at the other side of dividing walls were rated by participants according to annoyance and loudness. The correlation analysis did not find any SNQ able to predict the annoyance for music and speech at the same time. In the third paper, Park & Bradley (2009b) provided spectral adaptation terms to improve the annoyance prediction of both speech and music noise conditions. Rychtáriková et al. (2012; 40 subjects, 64 sounds) studied two wall constructions, one heavy and one with light construction, which presented the same value of the metric R_{living}. R_{living} is identical to $R_w + C_{50–5000}$ (ISO 717-1:2013). R_{living} and other metrics were also evaluated by Hongisto et al. (2014). The evaluation of the responses proved that R_{living} was not a good SNQ for this type of sounds and situations. Hongisto et al. (2016; 32 subjects, 4 sounds, $L_{A,eq}$ = 33–42 dB) studied the impact on cognitive work and subjective experience of four combinations of sound insulation and sound masking properties in private office rooms. The work revealed that sound insulation guidelines should account better the acoustic satisfaction and distraction caused by speech.

2.6.4 Studies with traffic noise inside homes

Road traffic noise, RTN, is probably the environmental noise most present in our lives, and thus research on this topic is extensive. The recent report by WHO (2018) includes 17 road traffic noise cross-sectional studies describing the relationship between sound pressure levels in dwellings and the percentage of people highly annoyed by it. High annoyance rates are 10% as the noise inside the dwellings exceeds 50 dB $L_{A,eq}$. Listening experiments focusing on RTN include, for instance, Ishiyama & Hashimoto (2000; 29 subjects, 24 sounds, $L_{A,eq}$ = 50–70 dB), Versteld

& Vos (2002; 23 subjects, 4 sounds, $L_{A,eq}$ = 38–50 dB), Torija & Flindell (2014), Torija & Flindel (2015; 33 subjects, 16 sounds, $L_{A,eq}$ = 37–57 dB), and de la Prida et al. (2020; 119 subjects, 30 sounds). These experiments confirmed that sound insulation metrics do not correlate well with perceived annoyance for all combinations of the RTN spectra. Again, the challenge originates in the number of combinations. The spectra indoors depend on the combination of façade sound insulation properties and RTN type. The type, for instance speed, vehicle size, asphalt, and weather conditions, etc. define the spectra of the noise both outdoors and indoors. In turn, the sound insulation quality of façades depends on the applied materials, their thickness, and their placement with respect to each other. Road-traffic spectrum has been standardized in ISO 717-1:2013.

In the experiment of Dincer & Yilmaz (2015; 40 subjects, 25 sounds), the test subjects lived in the same city district in which the RTN stimuli was recorded. This was seen as appropriate to ensure that test subjects were familiar prior the experiment with the test stimuli, and rated it based on their previous experience. The stimuli were created from in situ traffic recordings, which were then filtered with sound insulation values from typical building construction from that district. They investigated how different types of sounds compare to each other in terms of annoyance. Traffic volume and speed had an obvious effect on annoyance, while accelerating was the most annoying characteristic. Ordoñez et al. (2013; 16 subjects, 30 sounds) evaluated the suitability of three psychoacoustic methods to gather subjective data, which was later correlated with objective metrics. Six traffic sounds were filtered to consider the insulation properties of 10 constructions.

2.6.5 Studies regarding wind turbine noise

Wind turbine noise, WTN, is a special case within environmental noise cases, and in the last years it has received a lot of attention from researchers, governments, and habitants (WHO 2018). The most important physical characteristics of WTN are low frequency noise (Møller & Pedersen 2011), infrasonic tones (Leventhall 2006, Bolin et al. 2011), tonality (Søndergaard & Pedersen 2013, Yokoyama & Tachibana 2016), and amplitude modulation (Ioannidou et al. 2016). These physical characteristics seem to contribute to higher ratings of perceived annoyance of WTN in comparison to other environmental noises (Janssen et al. 2011, Schäffer et al. 2016).

Over 300 residents living within 2 km from the nearest wind turbine in three areas of Finland participated in the survey research of Radun (Radun et al. 2019). The concern for health effects was the most important factor related to both WTN annoyance and sleep disturbance. Other factors related to WTN annoyance were area, noise sensitivity and the general attitude towards wind power as a form of energy production. The sound level also explained outdoor annoyance and sleep disturbances. Recent research by Maijala (Maijala et al. 2020, Maijala et al. 2021) describe a set of sub-studies investigating the role of infrasound in health complaints related to wind farms. They performed measurements in people's houses and found no perceptible levels of infrasound despites the complaints. A total of 70 out of 1351 survey respondents (5%) reported symptoms they attributed to infrasound from a wind farm. Sound measurements were performed in two uninhabited dwellings at 1.5 km from a wind farm. Results showed that annoyance was related to the total sound level and amplitude modulation of the WTN. The work did not show that infrasound is not annoying, but that people could complain about it even when it is not measurable.

The effect of amplitude modulated synthetic wide band noise was studied by Virjonen et al. (2019; 40 subjects, 92 sounds, $L_{A,eq}$ = 29–49 dB). Sounds with the same sound pressure level but different modulation frequency and depth led to unequal subjective annoyance ratings. They suggested that depending on those two physical characteristics, a penalty from 4 to 12 dB should be added to the sound pressure level of the noise in question. Schäffer and his team recently performed two concise studies regarding WTN (Schäffer et al. 2016 and Schäffer et al. 2018). In the first study (60 subjects, 30 sounds, $L_{A,eq}$ = 35–60 dB), WTN and RTN were compared to short-term annoyance reactions. WTN was synthetically produced as presented in Manyoky et al. (2014), while RTN was recorded from car pass-by events. Thirty stimuli with a length of 25 seconds each were rated regarding annoyance in an 11-point scale from 0 to 10 by the 60 participants. Visual factors were excluded from the experiments, and so the researchers consider that observed differences in annoyance reactions between WTN and RTN were associated only with the acoustic characteristics of the sounds and not the visual appearance of the sources. The analysis of data revealed that the same $L_{A,eq}$ produced higher annoyance in WTN than in RTN. WTN has modulation as special physical characteristic, which are described by the amplitude modulation and modulation frequency. The study presented a direct link between the acoustic characteristics of the noises and the annoyance reactions and proposed to verify that in future long-term exposure studies. In the second study, Schäffer et al. (2018; 52 subjects, 20 sounds, $L_{A,eq}$ = 37–43 dB) changed the methodology and studied wind turbine noise as it would be perceived in residential areas outside the dwellings. The work concentrated now on the subjectively perceived short-term annoyance reactions. The study was perhaps the first to study technical aspects like spectral shape, depth of periodic amplitude modulation, and occurrence of random amplitude modulation. The level of the sounds was constant at $L_{A,eq}$ = 40 dB, aiming for the approximate level of windmills at distances of one kilometer. The laboratory setup asked 52 participants to rate the annoyance produced by 20 stimuli of 10 seconds' length each. Participants did not get visual cues during the listening, but they were instructed to imagine the noise to come from windmills. The study showed that annoyance increased with increasing energy content of the noise towards the low-frequency range, as well as with the depth of periodic amplitude modulation and its randomness. The results confirm that that other metrics than $L_{A,eq}$ should be used.

3 Aims

The following psychoacoustic research was designed for indoor noise with three aims in mind. First, with a focus on environmental health, we wanted to determine how the physical characteristics impacted on annoyance. The results should provide valuable information to later perform risk assessments and risk prevention in relation to health. Second, and with an engineering standpoint, we aimed to find out what acoustic metrics suited best to predict subjective annoyance experience in specific noise situations. This could help architects and engineers to determine the performance of building constructions with respect to the protection from noise annoyance, instead of loudness as it is still done. Third, we wanted to develop cost-effective psychoacoustic testing methods and facilities, which would enable the correct execution of this research activity and any future experiments.

The secondary aims of this Thesis are related to the main research questions in each of the Publications I–IV. The aims relate mostly to questions of interest for acousticians, the manufacturing industry, and noise regulation parties. The research was performed during two research projects, ÄKK (2012–2014) and Annoyance (2016–2019), funded by Business Finland, Finnish Institute of Occupational Health, Turku University of Applied Sciences, and several companies with research interests in the field. Each Publication considered a specific noise situation and a set of metrics to be investigated. The secondary aims are summarized:

- Determine what is the spectrum of ventilation broad-band noise causing best acoustic satisfaction. Identify which SNQs out of 15 are the best predictors of acoustic satisfaction. Publication I.
- Determine which one out of 12 standardized SNQs of airborne sound insulation predicts best the subjective annoyance ratings of living sounds inside dwellings. Publication II.
- Determine which one out of 25 different SNQs used for sound transmission explains best the subjective annoyance ratings inside dwellings when traffic noises transfer through façade constructions. Publication III.
- Determine how tonal sounds are perceived compared to non-tonal sounds at overall levels comparable to those existing in residential spaces. Determine

the role of tonal frequency, tonal audibility, and overall level in subjective annoyance perception. Publication IV.

4 Materials and Methods

The methods applied in Publications I–IV are described next. The chapter is divided in four main sections, one for each publication, and each section is further divided in seven parts. The seven parts are; research question, experimental setup, laboratory setup, sound stimuli, subjective measures, experimental procedure and communication, and statistical analysis. We avoid as much as possible the repetition of common terms of the experiments.

4.1 Subjective and objective rating of spectrally different pseudorandom noises—Implications for speech masking design. Publication I.

Research questions

Publication I, Hongisto et al. (2015), concentrated on noises created by ventilation machines in office spaces. The research investigated the feelings of office workers in terms of annoyance or acoustic satisfaction with respect to this type of noise. The main research question was what type of noise spectra leads to the best acoustic satisfaction and the least annoyance. The secondary aims were two; Find out which ventilation noise would be better suited for use as a speech masker in open plan offices and calculate what objective single number quantities correlate best with acoustics satisfaction.

In open-plan offices, speech is the most detrimental type of noise, as it impairs concentration and cognitive work (Haapakangas et al. 2008, Kaarlela-Tuomaala et al. 2009). Speech masking systems are often used as control measures for this type of cases. They try to improve the perceived sound environment by reducing the intelligibility of irrelevant speech. Ventilation noise can be used as a masker, but normally the masking noise is played via hidden speakers. Haapakangas mentioned that office workers might not always understand the reason to artificially increment noise levels and might present resistance to it, despite they have demonstrated its positive impact on performance and satisfaction.

Experimental setup

Twenty-three test subjects (15 female and 18 male, mean age 40, age range 25–65) participated in the experiment one at a time. They rated the loudness, the annoyance, and seven other subjective measures for a total of 11 spectrally different pseudorandom noises. The subjective measures were obtained from nine questions of the type; *how disturbing is the noise?*, *how much it impairs your concentration during working?*, and *could you stand that noise for long periods of time?*.

The main research question in this study was what metrics work best to describe the sound quality of ventilation noise in offices. The quality of the metric was evaluated according to how well each metric correlated with the subjective perception. Fifteen SNQs were selected, and they were calculated from the sound pressure levels measured in the area later occupied by the head of the test subject. The complete list of the 15 SNQ and how they are calculated is described in detail in Publication I.

Laboratory setup

The laboratory, 6.7 x 4.6 x 2.7 m, was divided in two parts by a light-weight construction to hide the speakers from the participants. The outlook of the room is presented in Figure 4.

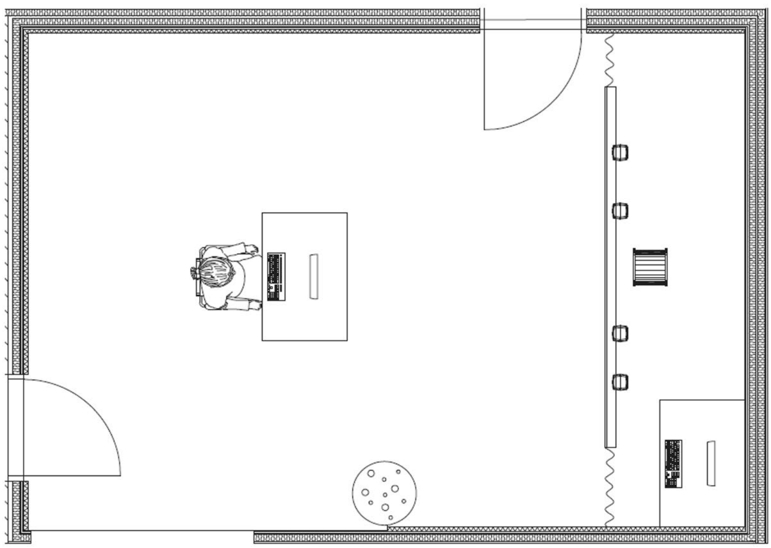


Figure 4. The layout of the laboratory used in Publication I.

The room represented a possible office space and the subject sat in the middle in the room. Light, temperature, and air ventilation were set to guarantee the maximum comfort and avoid them as variables. The reverberation times in the room were 0.37, 0.48, 0.34, 0.26, 0.27, and 0.25 seconds in the octave bands between 63 and 8000 Hz. The background level in the room was $L_{A,eq}$ = 24 dB, and the linear equivalent sound pressure levels in the octave bands 63–8000 Hz were 43, 34, 24, 16, 12, 12, 12 and 12 dB. The background was almost 20 dB lower than the test sounds, and thus it did not affect the subjective ratings of the sounds. The sounds were presented to the subjects via a multi-speaker system with four loudspeakers at heights 150 and 50 cm (Genelec 8010A) and a passive subwoofer (JayHo). The system was hidden from the subject's view behind one gypsum board partition. The attenuation of the partition was not of interest because the sound stimuli were adjusted to match the target spectra in the location occupied by the subject. It was ensured with measurements that the sound field matched the spectra homogeneously within the area occupied by the head of the listener. Because of the placement of the speakers, the sound field was not diffuse in the room, neither was the sound experienced to come from the ceiling where typically ventilation machines are located. We believed this did not have a negative impact in the quality of the research. A software, especially developed for this test, presented the sounds to the subjects, and recorded their responses.

Sound stimuli

The eleven experimental *sounds* were created from pseudorandom pink noise, but their spectra differed. Audio filtering was performed with Sound Quality software 7698 developed by Brüel&Kjaer. All sounds were set to have the same A-weighted equivalent level sound pressure level $L_{A.eq}$ = 42 dB. This level was selected because it is acceptable for masking noise systems in open plan offices (Hongisto et al. 2017a) and because noises louder than 45 dB $L_{A,eq}$ have been previously judged as too loud (Veitch et al. 2002). The spectra of the eleven sounds are presented in Figure 5, which for comparison purposes also presents the background level in the room (red line) and the hearing threshold (green line). The spectra presented in the figure is the average of six measurements taken in the space occupied by the head of the subjects. Speech sounds were not presented during the experiment, e.g., the subjects only listened and rated the masking noises.

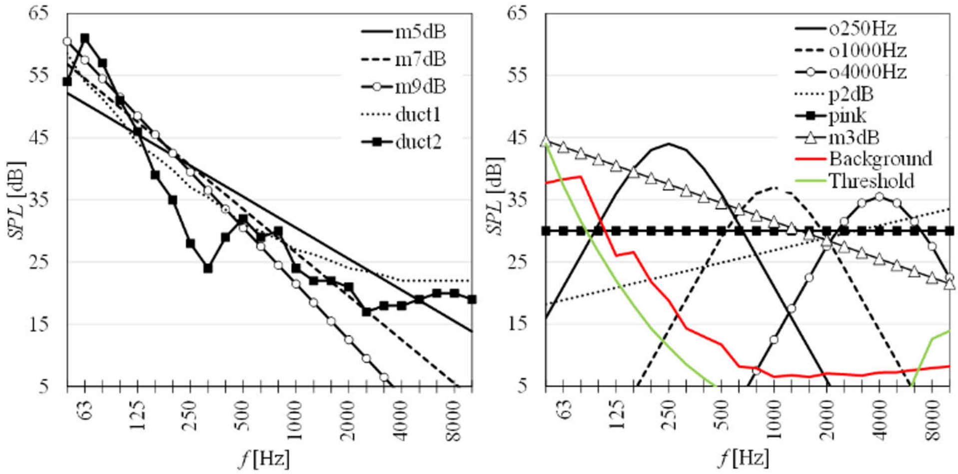


Figure 5. The linear sound pressure levels (vertical axis) of the eleven experimental sounds at each 1/3 octave frequency band from 50 Hz to 10000 Hz (horizontal axis).

Subjective measures

Nine subjective measures were used; *loudness* (*loud*), *rumbling* (*rumb*), *roaring* (*roar*), *hissing* (*hiss*), *pleasantness* (*plea*), *disturbance* (*dist*), *easiness to get used* (*habi*), *concentration impairment* (*conc*), *work efficiency* (*work*). The first four measures describe the noise from a physical or sound quality perspective, while the other five represent attitude towards the noise. The meaning of the measures was explained to the subjects. The attitude measures are based on the research by Haapakangas et al. (2011). The subjects were not working during the test, but they were instructed to imagine being working in a corresponding acoustic environment for a whole working day. Therefore, the five attitude questions inquire about long-term effects. The test subjects had to listen each sound for 90 seconds. Thereafter the questions appeared to the computer screen, and they responded to the nine questions. The rating was done by mouse-clicking over a 20 cm long horizontal line below each question. The line was divided in six equal parts by five cross lines. Verbal scaling was used among the cross lines to facilitate the rating. Rating was then scaled from 0 to 100. The *loudness* and the five attitude measures correlated strongly with each other, and so we created a sum variable *Acoustic Satisfaction* as the main dependent variable. The Cronbach's alpha of this variable was determined for all 11 sounds and the values were high, between 0.89 and 0.95. It was expected that this variable would simplify and improve the quality of the statistical analysis.

$$Acoustic\ Satisfaction = 1/6\,\{plea + habi + work + (100 - loud) + (100 - dist) + (100 - conc)\}$$

Experimental procedure and communication with subjects

The test subjects were recruited from our company, Finnish Institute of Occupational Health, but they were not part of the acoustic group and were naïve with listening experiments. The experiment was very short, compared to those of Publications II–IV. Subjects were not given any compensation for the participation. During a five-minute speech, the subjects were told that they should imagine themselves working in a similar office and noise environments. The use of the software to rate the sounds and run the experiment was explained. They were asked to rate each of the 11 sounds according to the nine subjective measures. The subjective measures were explained to them.

Design and statistical analysis

The statistical analysis considered that the experiment had a repeated measures design. The independent variable was the *sound*. The subjects listened to them in a randomized order. The dependent variables were the nine subjective measures and the sum variable *Acoustic Satisfaction*. Analysis was performed both in Excel 2013 (Microsoft 2014) and SPSS 20 (IBM 2012). Non-parametric Friedman's test was applied to determine the significance of the differences of the dependent variables among the 11 experimental sounds. Pair comparisons were made for all possible 55 combinations of sounds. Non-parametric Wilcoxon signed rank test was used in the pair comparisons. The p values were corrected with the Benjamini-Hockberg correction. The difference of the *Acoustic Satisfaction* value between two sounds was considered statistically significant when $p < 0.05$. The correlation coefficients between the indices and the dependent variables were determined using the individual ratings and not the average. Pearson's correlation coefficients for the 11 indices were calculated, and they achieved a statistically significant level of $p < 0.05$ and $p < 0.001$ when the coefficient was larger than 0.13 and 0.25, respectively.

4.2 Subjective and Objective Rating of Airborne Sound Insulation–Living Sounds. Publication II.

Research questions

Publication II applied an experimental setup with living sounds, i.e., noises like loud talking, TV watching, music, dog barking, etc. produced in people's dwellings. We were interested in how they are perceived at the other side of the constructions used to separate the apartments. The aim of the study was to determine which standardized airborne sound insulation SNQ predicted best the subjective ratings of living sounds.

The number of SNQs in the literature is vast, see e.g., Park & Bradley 2009a, Park & Bradley 2009b, Rasmussen 2009, and Rasmussen 2019. We limited our analysis to standardized SNQs found in ISO 717-1 (ISO 717-1:2013) and ASTM E413 (ASTM E413-10:2013).

Experimental setup

Fifty-nine subjects (40 female, 19 male, age range 20 to 43, mean 27) participated in the experiment. The test subjects were asked to imagine themselves being at home reading while listening to the sounds of the experiment. The subjects rated the *loudness*, *disturbance* and *acceptability* of living sounds, i.e., sounds created in neighbors' apartments and transmitted through the separating walls. The experimental setup is explained in Figure 6. The six sounds were recorded in six indoor spaces, L1. Afterwards, the recordings were filtered with sound reduction index filters of nine walls, SRI, to simulate how these six sounds would be heard in the living room at the other side of the construction, L2. The L2 situation was then simulated in laboratory conditions.

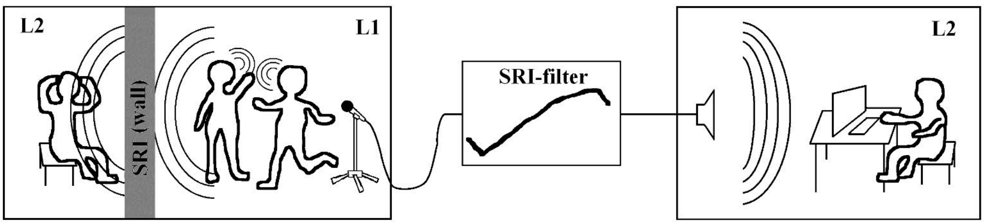


Figure 6. The actual listening experience of neighbor's living sound in L2 was simulated by feeding the sound recorded in six real living rooms, L1, through an audio filter including the SRI of the wall. The test subject experiences in the laboratory, in theory, the same situation L2 than a real neighbor would have regarding the sounds coming from the adjacent apartment.

Laboratory setup

The laboratory room was the same as in Publication I. The room was improved with five paintings and a carpet to give it a home space appearance (Figure 7). The loudspeakers, as before, were located behind the light partition.

Sound stimuli

In this test, we selected six living sounds, which were filtered to reproduce how they would be experienced at the other side of a wall. Nine walls were simulated, and so the total of 54 sound stimuli were created. The six living sounds were Guitar (acoustic guitar playing four chords), Music1 (Together again by Janet Jackson), Music2 (Feuer frei by Rammstein), Baby cry, Loud speech, and Dog bark. These six living sounds represent in our opinion the most typical situations in dwellings with respect to annoyance from neighbors. The walls were simulations of nine typical wall constructions applied in new apartments in Finland, and so they had a sound reduction index R'_w between 50 and 60 dB. The SNQs describing the walls were calculated according to SNQs of ISO 717-1 (ISO 717-1:2013) and ASTM E413 (ASTM E413-10, 2010). The 12 SNQs were: R_w, $R_w + C_{100–3150}$, $R_w + C_{100–5000}$, $R_w + C_{50-3150}$, $R_w + C_{50–5000}$, $R_w + C_{tr,100–3150}$, $R_w + C_{tr,100–5000}$, $R_w + C_{tr,50–3150}$, $R_w + C_{tr,50–5000}$, STC, STCno8, R_{speech}.

Figure 7. The laboratory with the configuration applied in Publication II.

The sound pressure level of the 54 experimental sounds varied from 8 to 40 dB $L_{A,eq}$. This means that the sound pressure level of part of the sound stimuli at some frequencies was lower than the background noise level. Test subjects experienced in those cases higher levels than those of the signals because they heard a combination of the sound stimuli and the background in the room. However, part of the spectrum of the experimental sounds was above the spectrum of the background noise, and so subjects could hear and identify the sounds. During the preparation, up to 30 dB higher playback levels were used to enable accurate spectrum measurements also at those bands where the levels would fall below the background noise level when the actual listening level is used. Publication II includes detailed information regarding the procedure of recording of noises and mixing of sound stimuli. For simplicity, most of that information is not presented here. The reader interested in acoustical information, like the type of constructions, layers, thickness, mass, principal material, sound reduction index R'_w and 11 other SNQs, is invited to read the original paper.

Subjective measures

The subject was asked to imagine the following situation (citing the original work): "*Imagine that you are alone at home in a multi-story building in silence and peace. You are in a relaxed mind set. You are reading a magazine or a book or you are browsing the internet, and you start to hear a sound from neighboring dwelling behind the wall*". The subject had to listen to each sound sample completely before she or he could rate the three subjective measures. While each sound was played for the first time, the following text was displayed on the screen; "*You hear this kind of sound coming from your neighbor*". The dependent variables of the experiment were three subjective measures: *loudness, disturbance* and *acceptability*, which were asked with sentences; "*How loud is the sound?*", "*How disturbing is the sound?*", and "*Would the sound be acceptable if it could be heard in your own home?*". The extreme alternatives were labelled accordingly, e.g., "*0: Not at all disturbing*" and "*10: Extremely disturbing*". Our presumption was that *disturbance* and *acceptability* seemed to be more relevant than *loudness* because the sounds were not especially loud. It should be clarified that our analysis focused on the difference between the ratings given by subjects and not on absolute values. For instance, acceptability and acceptability thresholds might change between persons, but that should not impact too much as long each test subject uses the rating scale consistently. *Acceptability* has not previously been used very much in this type of experiments, and we asked to see its potential as a measure.

Experimental procedure and communication with subjects

The test subjects were recruited via notice boards in nearby buildings and were paid with an EUR 20 gift card. The presumptions were normal hearing ability, Finnish native language and currently residing in a multi-story building. The latter condition secured experience with sounds like the fictional situation.

The communication protocol with the subjects was improved with respect to the previous study. The experimental procedure was now more elaborated, as we wanted to ensure that all subjects would receive an identical set of instructions and that it was clear for them how to proceed and what to expect in each phase of the experiment. The communication lasted about 15 minutes. During that time, the consent form was explained and signed, and the instructions regarding the experiment were given. The most important concepts were explained twice via oral and written instructions.

The experiment took about 75–90 minutes and it was divided in five phases: the Weinstein sensitivity test (Weinstein 1978), hearing sensitivity measurements using the Hughson-Westlake method in frequencies 125, 500, 1000, 2000 and 4000 Hz in both ears (Madsen Electronics OB822 Clinical Audiometer), familiarization phases to become familiar with the future sounds, the rehearsal phase for practicing the use of the subjective rating scale, and the listening experiment itself. The Weinstein sensitivity test included the 21 questions designed to address affection and attitudes to both general noise and daily environmental sounds. We performed the test with the hope of performing further research regarding the relationship between subjects' attitudes to noise and the three subjective measures. That research will perhaps be conducted in the future.

The test presented the 54 experimental sounds in a pre-randomized order. Sounds were presented one at a time. The six living sounds were presented in sets, each set having the nine test sounds obtained from the nine constructions. Subjects needed to complete a rating set to get access to the next one. The presentation order of living sounds and wall filters was randomized to ensure that the presentation order effects were balanced. Each set of nine test sounds was preceded by a dummy sound. The ratings of the dummy sounds were not analyzed. The dummy sounds are similar in terms of information to the test sounds, but they have a different spectrum. They were placed to avoid taking into the analysis the first subjective measure rating given by the test subjects. The importance of dummy sounds is discussed later in Appendix 1.

Design and statistical analysis

The primary task was to determine the linear correlation between the subjective measures and the 12 SNQs selected to represent the physical properties of the nine walls. The 72 data points obtained for each subjective measure (6 sounds x 12 walls) were used here to calculate the Pearson's correlation between the individual ratings of the 59 participants for each wall (531 individual responses) for each one of the living sounds. There are two ways to determine Pearson's correlation coefficient. The most justified method would be to determine the correlation coefficient in the individual level, called $R_{P,I}$. The value was determined by including all 531 individual responses (59 subjects and 9 walls per subjective measure) for each living sound. A scientifically less justified method would be to determine the correlation coefficient in the sample-average level, called $R_{P,A}$, i.e. the average judgment given by the subjects for each wall. As the individual data contains 531 scattered points while the sample-averaged data contains only 9 points, the values of $R_{P,A}$ are systematically larger than the values of $R_{P,I}$. A very consistent difference between $R_{P,A}$ and $R_{P,I}$ could be found with all SNQs and the six living sounds. The difference was 0.16 ± 0.04 for *loudness*, 0.21 ± 0.04 for *disturbance* and 0.28 ± 0.07 for *acceptability*. The main conclusions of this paper, i.e., the nomination of the SNQs best predicting the subjective measures, would be the same whether we base the analysis on $R_{P,A}$ or $R_{P,I}$. We decided to report the squared values of $R_{P,A}$ to enable an easier comparison with the results of Park & Bradley (2009a), who also used the sample-averaged method. We abbreviated the square of $R_{P,A}$ by R^2 which is the coefficient of determination.

One example of the statistical analysis is presented in Figure 8, showing the disturbance rating obtained for two living sounds, Music2 and Dog bark, against the insulation properties of the nine walls according to the SNQ R_w. The vertical axis is the disturbance rating, and the horizontal axis is R_w. For simplicity reasons, the figure shows only the averaged value, but as previously said, the 59 individual answers for each wall were considered in the correlation analysis. This way we could determine which SNQ predicted better, i.e., correlated the most, with the subjective ratings. Overall, when the R^2 value is high, the SNQ predicts the mean subjective rating well. In this example, R_w predicts better the disturbance caused by Music2 than by Dog bark.

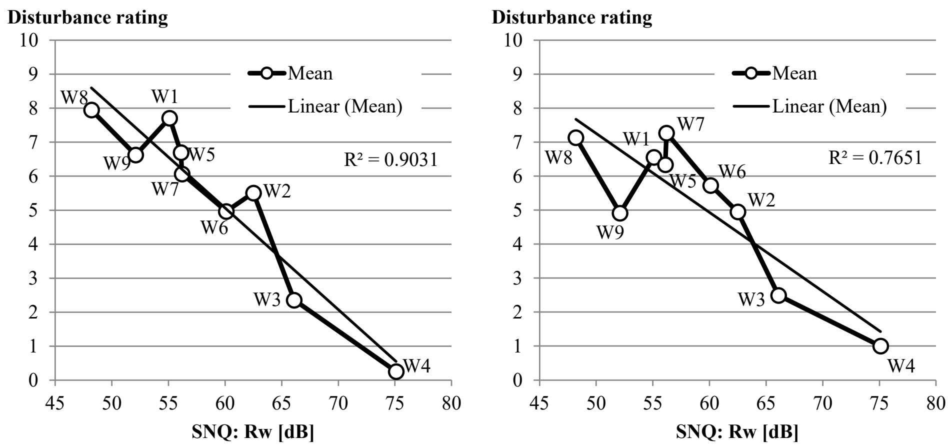


Figure 8. Visual example of the performed correlation analysis, presenting the mean disturbance ratings (vertical axis) for sound types Music2 (left) and Dog bark (right) for the nine walls W1-W9 when they were ordered according to R_w (horizontal axis). A linear trendline has been fitted to both sets of data. The R^2 value determines how well the SNQ correlates (i.e., predicts) the subjective disturbance rating.

4.3 Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. Publication III.

Research questions

Publication III concentrates on people's subjective perception of road traffic noise, RTN, inside their homes. According to the World Health Organization, RTN is the most common source of environmental noise, and sleep disturbance the most adverse health effect related to it (WHO 2018, Hurtley 2009). One problem regarding the control of traffic noise is that different countries use different SNQs to quantify and regulate the noise insulation properties of façades. This has led to the situation in which it is not clear which measurable metric correlates best with sound comfort or annoyance.

The performance of sound insulation metrics depends very much on the spectrum of the analyzed sounds. Our hypothesis was that the performance order of constructions with respect to any objective SNQ would not necessarily match the subjective annoyance rating order. Our aim was to evaluate how 25 SNQs used for sound transmission through façade constructions explain the subjective ratings of various traffic sounds. Sixty experimental sounds were created from five spectrally different road-traffic sound types, which were filtered to sound as they would be heard inside a room after being transmitted through the 12 façade constructions.

Experimental setup

Forty-three subjects (28 female, 15 male, age range 21 to 50, mean 27) participated in the experiment. The subjects rated the *loudness* and the *annoyance* of airborne road-traffic noise transmitted from outside to inside through façade constructions. Five spectrally different road traffic noise were recorded and filtered with 12 filters that represented the airborne sound reduction index of 12 façade elements. The experimental setup, see Figure 9, aimed to reproduce the same sound conditions, L2, that would be experienced inside a dwelling when the traffic noise from outside, L1, is transmitted through the façade. The sound recordings at L1 were filtered and equalized to simulate the 12 façade constructions. The test subjects were told to imagine themselves reading a book or magazine at home while listening to the sounds. Their task was to imagine the situation and rate the *loudness* and *annoyance* of the 60 experimental sounds on a scale from 0 to 10.

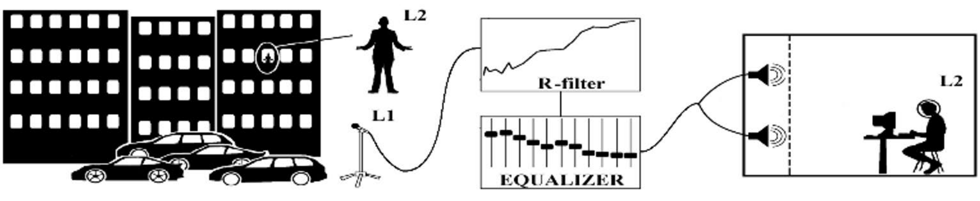


Figure 9. The listening experience of traffic sounds inside dwellings L2 was created by applying filters to real traffic sound recordings L1. The filters simulate the sound reduction properties produced by façade constructions. The test subject in the laboratory experienced the same L2 sounds than the real habitant of the dwelling.

Laboratory setup

A laboratory room of new construction was used in this experiment (Figure 10). The room was fully isolated and specially designed for psychoacoustic experiments. For instance, the room was placed over springs to minimize the occurrence of impact sounds from the building, and the ventilation system was rearranged to minimize background noise. The reverberation time of the space was 0.45, 0.32, 0.30, 0.14, 0.11, 0.12, 0.12 and 0.11 in the octave bands between 63 and 8000 Hz. The background noise level in the room was 20 dB $L_{A,eq}$. The loudspeakers were hidden behind a curtain. The computer was placed outside the room to minimize the background level.

Figure 10. Outside and inside views of the laboratory room used in the listening experiments presented in Publications III and IV.

Sound stimuli

The sound stimuli were 20 seconds long. The recordings of RTN were performed outdoors for several hours in the city of Turku, Finland. Stereo recordings were done in the frequency range 20–20000 Hz with a digital recorder (Fostex FR-2LE, Fostex, Japan), and two 1/2 in. pre-polarized condenser free-field microphones (NTI M2010, NTI Audio, Liechtenstein) separated by 20 cm. Binaural recordings were not used since we applied loudspeaker playback in the experiment and perception of the direction of the sound source was irrelevant.

Five different *sound types* S1–S5 were selected to represent the five conditions. S1 was light vehicles in an urban street at 50 km/h, S2 and S3 were light vehicles in a motorway at 80 and 100 km/h respectively, S4 was heavy vehicles in an urban street at 60 km/h, and S5 included both heavy and light vehicles in an urban street at 60 km/h. The spectra of the five *sound types* differ quite much from each other, see the left side of Figure 11. The 12 façade constructions were typical façade constructions used in Finland. The sound reduction indexes R of the façades are presented in Figure 11, on the right. The 60 test sounds were basically created applying filters to reduce the sound pressure level of the sounds S1–S5 at each frequency band by the sound reduction index R_w of the 12 facades constructions. After the filtering, the 60 test sounds had an equivalent A-weighted sound pressure level ranging from 12 to 46 dB. Publication III included over two pages full of information regarding the façade constructions, the spectra of the 60 experimental sounds, and the 25 SNQs used in the analysis. For simplicity, this information is not included here. The reader interested in all that the acoustical information is invited to read the original publication.

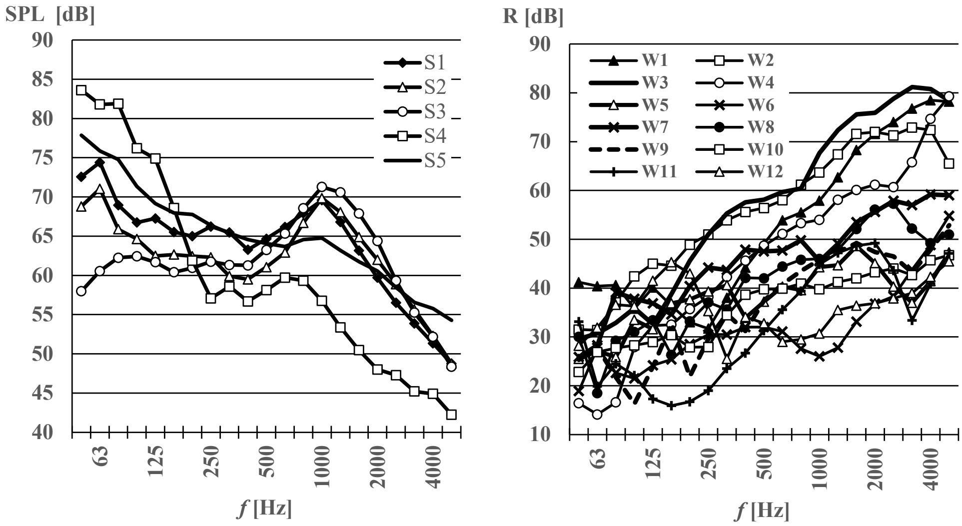


Figure 11. Left) The linear sound pressure levels SPL (vertical axis) of the traffic sound types S1–S5 recorded outdoors at each 1/3 octave band frequency from 50 to 5000 Hz (horizontal axis). Right) The sound reduction index R (vertical axis) of the 12 façade constructions W1–W12 at each 1/3 octave band frequency from 50 to 5000 Hz (horizontal axis).

Subjective measures

To make the judgments closer to the real residential situation, the participants were instructed in the following way (citing the original work): "*During the experiment, imagine that you live in a dwelling in an apartment building close to a street. Imagine that you are alone at home in peace and quiet. You are in a relaxed state of mind. You are reading a magazine or a book, or you are browsing the internet, and you start to hear sounds from the road. Although the sounds are short, imagine that you would hear them continuously.*" The subject had to listen each sample completely before the rating of the two subjective measures was enabled. The dependent variables were the two subjective measures *loudness* and *annoyance*. While listening to the sound for the first time, the text "*You hear this kind of sound in your dwelling*" was shown to the test subject in the screen. After that, the sound started to play again from the beginning, until the subject responded to the two sentences; "*How loud is the sound?*" and "*How annoying is the sound?*". The rating scale went from 0 to 10 and the extreme alternatives were labelled accordingly; "0 *Not at all loud*", "10 *Extremely loud*", "0 *Not at all annoying*" and "10 *Extremely annoying*". In the lower part of the screen, a checkbox was added with the text "*The sound cannot be heard at all*", which could be pressed by the subjects when they could not hear any sound at all. Inaudibility ratings were only obtained in 20 cases out of the 2580 possible data points we had in this test.

Experimental procedure and communication with subjects

The participants were recruited via student organizations and were paid with an EUR 20 gift card. The inclusion criteria were normal hearing ability, Finnish native language and currently residing in a multi-story building. The latter condition aimed to secure experience with the fictional situation. All participants were living in similar types of buildings and have prior experience with the type of sounds used in the test. The experiment consisted of seven phases and lasted for 75–90 minutes. The phases were: Reading and signing the consent form, Weinstein sensitivity test, hearing sensitivity measurements, familiarization phase to become familiar with the future sounds, rehearsal phase for practicing the use of the rating scales, the experiment phase which took about 30 minutes, and the final feedback about the experiment. The experiment phase presented the 60 sound stimuli in random order.

Design and statistical analysis

The analysis started with a first stage evaluation regarding the normality of the rating distributions and the presence of outliers or unrealistic ratings. This analysis was done using Excel 2013 (Microsoft) and SPSS (IBM SPSS 20, IBM Corp, Armonk, NY) simultaneously. The ratings were normally distributed, except for the extreme sounds receiving the lowest loudness and annoyance ratings. None of the participants was classified as an extreme outlier, but there were a few individual outliers. Individual outliers were those ratings 1.5 times the interquartile range above the 75^{th} percentile or below the 25^{th} percentile. Individual outliers did not influence the results of successive analysis.

The recorder road-traffic included frequency-dependent level fluctuations before façade filtering. Because the façade filters W1–W12 changed the spectrum significantly, the variability of the A-weighted SPL did not remain constant for a given sound type. Temporal level variations were measured for all 60 sounds. We used the SNQ $\Delta L_A = L_{A5} - L_{A95}$ to express this, where L_{A5} and L_{A95} [dB] are the 5% and 95% percentile levels determined for the duration of the sound sample (20 s) using fast time sampling. L_{A5} means the A-weighted SPL, which is exceeded for 5% of the samples. Pearson's linear correlation analysis was conducted for each sound type to test whether the mean ratings depend on temporal variability.

The primary task was to determine the linear Pearson's correlation coefficients between the subjective measures and the 25 SNQs selected to measure the acoustical properties of the 12 façades. In this case, and similarly to Kylliäinen et al. 2017, all individual responses from participants were considered. The example in Figure 12 presents the statistical analysis that was performed for any combination of subjective measure and SNQ. In the figure, only the mean of the 43 participants' ratings, the standard deviation, and the most extreme ratings (min and max) are indicated. The correlation coefficient r_P was statistically highly significant (p = 0.0005, one-tailed) when $|r_P^2| > 0.23$. The significance of the difference between two correlation coefficients was evaluated according to the method of Steiger (Steiger 1980). The difference between two r_P^2 values was statistically significant ($p < 0.05$, one-tailed, N = 516) when the values differed by more than 0.05 ($r_P^2 > 0.70$).

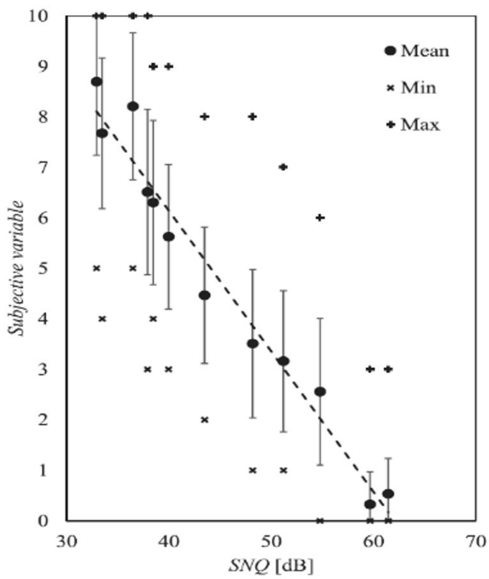


Figure 12. An example representing the idea behind the statistical process to verify which SNQ best correlated with subjective perception. In this example, the sound type is S3, the subjective variable is *loudness* (vertical axis), and the SNQ is $AA_{100\text{-}5000}$ (horizontal axis). The mean of 43 test subjects' ratings, the standard deviation, and the most extreme ratings (Mix, Max) are indicated.

4.4 Annoyance of low-level tonal sounds – Factors affecting the penalty. Publication IV.

Research questions

Existing literature shows that tonal noises are more annoying than non-tonal sounds with the same A-weighted SPL (Kryter & Pearsons 1965, Persson et al. 1990, Ryherd & Wang 2008, Hansen et al. 2011). Most of the existing research works on annoyance have concentrated on situations with $L_{A,eq}$ in the range 50–70 dB (Angerer et al. 1991, More & Davies 2010, Di et al. 2015), and the studies considering lower sound level situations are fewer (Landström et al. 1994, Hünerbein et al. 2010, Poulsen 2003). The research above has not determined what is the impact on annoyance of the two physical characteristics that best describe tonality, e.g., tonal frequency and tonal audibility when the level varies between $L_{A,eq}$ = 25–45 dB. Tonal audibility is the prominence of the tone with respect to the rest of the noise and tonal frequency is the frequency at which the tone, i.e., narrowband, is located. Tonal audibility aims to quantify how much the tone can be discerned or identified, and the rest is considered background. The calculation of tonal audibility has been standardized in ANSI S1.13:2005, ECMA-74:2015, DIN 45681:2005, ISO 1996-2:2007, but each standard proposes its own calculation method and so the results are not comparable. Based on the calculated tonal audibility, the standards apply a penalty, k. The calculated penalty is added to the measured or predicted SPL to counteract the negative effect that tonality might have on annoyance. Currently, several countries apply penalty levels in the corresponding noise legislation when sounds present tonality. In Finland, for example, the penalty is 5 dB in buildings for any type of environmental sounds presenting tonal components (Ministry of the Environment 2017), while it is 3 or 6 dB depending on the tonal audibility when the regulation concerns health protection (Ministry of Social Affairs and Health 2015).

Publication IV, Oliva et al. 2017, concentrated on noise including tonal components. The purpose of the study was to determine how tonal sounds are perceived compared to non-tonal sounds at low levels like those experienced in residential spaces. Our first hypothesis was that tonal sounds produce different subjective loudness and annoyance experiences. If that occurred, penalty values should be updated based only on annoyance ratings. Based on existing literature, our second and third hypotheses were that penalty depends on tonal frequency and tonal audibility. We wanted to evaluate what standards performed better. In this work, we used the methods applied in ANSI S1.13:2005, ECMA-74:2015, DIN 45681:2005, and ISO 1996-2:2007. It should be noted that none consider the overall level of the sound, for instance $L_{A,eq}$, as an input parameter. Our fourth and final hypothesis was that penalty should also consider the overall level of the sound. If that happens, new building codes should also include penalties for low and high sound pressure levels separately. To test this hypothesis, the levels of the sounds were 25 and 35 dB $L_{A,eq}$ for tonal sounds, and 19–45 dB $L_{A,eq}$ for the background.

Experimental setup

The experiment was conducted using a repeated measures design, where all participants rated all experimental sounds. Experimental sounds included tonal and non-tonal sounds. The twenty tonal sounds were built from the combination of five tonal components and four tonal audibility levels. The experiment consisted of four blocks, to ask separately *loudness* and *annoyance* for levels 25 and 35 dB $L_{A,eq}$.

Laboratory setup

The same laboratory room as in Publication III was used, but the audio setup changed. In this experiment, the audio stimuli were given simultaneously via headphones (Sennheiser HD 580), speakers (Genelec 8020A), and subwoofer (Genelec 7050B). This hybrid system was done to minimize the impact of resonance modes. Resonances are constructive and/or destructive effects in the sound field appearing especially at low frequencies between parallel walls. They are explained in Appendix 1.

Each sound was measured separately to check that the playback level and the spectrum were as desired. Measurements were conducted with a head and torso simulator Brüel & Kjær 4100 (Brüel & Kjær 2013), because the experiment was performed with headphones. To ensure that the headphones were correctly mounted during the measurements, pink noise was first played through the playback system. The headphones were mounted so that the spectrum measured from both ear-microphones was identical. The frequency-dependent diffuse-field correction was applied (Brüel & Kjær Pulse Sound Quality 15.1.0). The correction considers the amplification caused by the torso (head, ear pinna, etc.). The amplification is approximately 0 dB, 0 dB, 0.4 dB, 1.3 dB, 3.5 dB, at frequencies 50, 110, 290, 850 and 2100 Hz, respectively (Brüel & Kjær 2013).

Sound stimuli

Two kinds of experimental sounds were used as sound stimuli: tonal sounds and the reference sounds. The reference sounds were broadband sounds resembling comfortable ventilation noise. They were used to compare the experience between tonal and non-tonal sounds. The $L_{A,eq}$ of the non-tonal sounds varied from 19 to 45 dB. The $L_{A,eq}$ for the tonal sounds was either 25 dB or 35 dB. We developed the following calculation method to assist with the estimation of the number of decibels that a non-tonal sound would require to be experienced as equally annoying as a tonal sound. The idea is explained mathematically later. The 20 tonal sounds were combinations of the same broadband sound used as the reference sounds and an added tonal component. Five tonal frequencies and four tonal audibility levels were selected to create the 20 tonal components. The tonal frequencies were set at 50, 110, 290, 850 and 2100 Hz. The frequencies are not harmonic with each other. The tonal audibility defines the prominence of the tone over the broadband component. We used as reference the tonal audibility levels of 5, 10, 17.5 and 25 dB calculated according to standard DIN 45681:2005. Figure 13 presents the spectra of four sounds having the same tonal frequency but different tonal audibility. The left side shows the spectra in 1/3rd octave bands, and on the right, the same but in narrow frequency bands. This figure aims to illustrate, for instance, how the level of masking component decreases when the tonal audibility increases. All sounds have the same overall level 25 dB $L_{A,eq}$. Furthermore, Figure 14 illustrates how the variable tonal frequency compares sounds with the same tonal audibility.

The experiment was separated in four Blocks, to enable the study of hypothesis 4, i.e., the penalty depends on the overall level. The 20 tonal sounds were presented at two overall levels, 25 dB $L_{A,eq}$ for Blocks 1 and 3, and 35 dB $L_{A,eq}$ for Blocks 2 and 4. The rating of *loudness* occurred in Blocks 1 and 2, and the rating of *annoyance* in Blocks 3 and 4. Non-tonal sounds, i.e. the reference sounds, with level in the range $L_{A,eq}$ = 19–45 dB were always played in each Block. The length of all sound stimuli was 9 seconds. The subjects listened to the complete stimuli before the rating scale appeared and rating was enabled. The stimuli kept playing during the rating phase.

Subjective measures

The two dependent variables were the subjective measures *loudness* and *annoyance. Loudness* rating, Blocks 1 and 2, was done after the question "*How loud is the sound*". In Blocks 3 and 4, and to make the subjective rating of *annoyance* more relevant to the real residential situation, the participants were given the following instructions (citing Publication IV): *"Imagine that you are alone at home in silence and peace. You are in a relaxed mind set. You are reading a magazine, or a book, or you are browsing the internet and you start to hear a sound from the environment*". In addition, a picture of a typical living room was shown to give a hint of a domestic environment. The *annoyance* was asked with the following question "*How much the sound bothers, disturbs or annoys you?*" The judgement of *loudness* and *annoyance* was given on a scale from 0 to 10. The extreme alternatives of the scale were verbally labelled with "*Not at all*" and "*Extremely*".

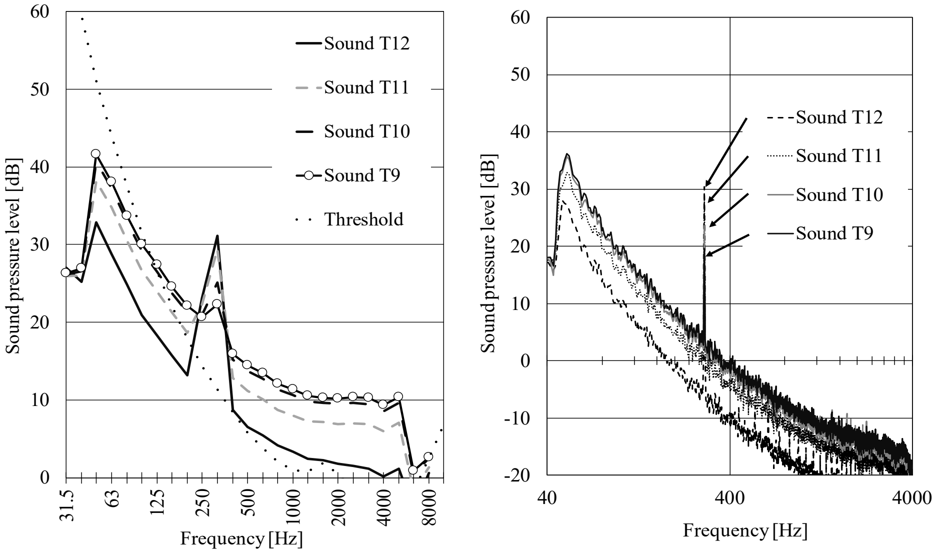


Figure 13. The spectra of four experimental sounds, T9, T10, T11, T12, which have the same tonal frequency, 290 Hz, but different tonal audibility levels. All of them have the same overall level 25 dB $L_{A,eq}$. Left) Spectra in 1/3rd octave bands. The hearing threshold according to ISO 226:2003 is presented also for comparison purposes. Right) Narrow band analysis. The arrows show the position of the peak.

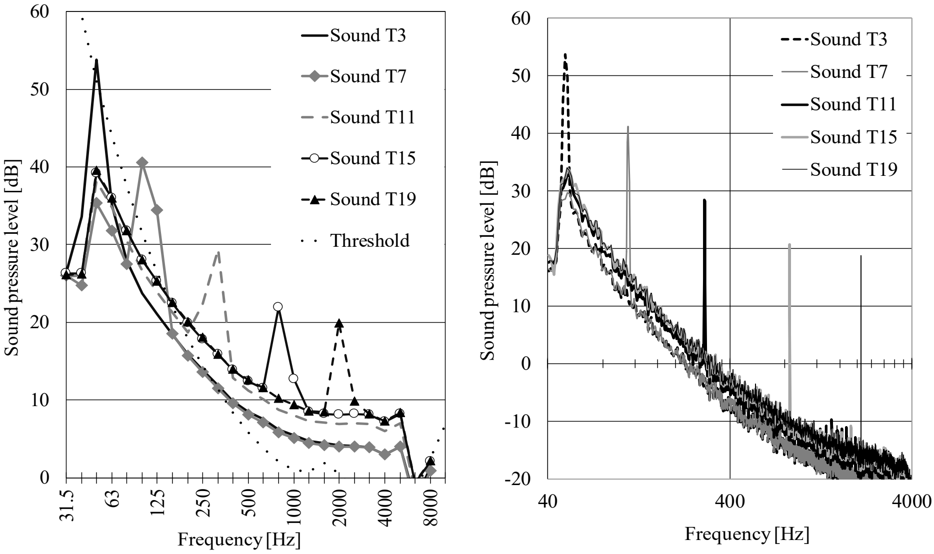


Figure 14. The spectra of five experimental sounds, T3, T7, T11, T15, T19, which have the same tonal audibility and the same overall level 25 dB $L_{A,eq}$, but different tonal frequency. Left) Spectra in 1/3rd octave bands. The hearing threshold according to ISO 226:2003 is presented also for comparison purposes. Right) Narrow band analysis.

Experimental procedure and communication with subjects

The participants were recruited via university email lists, and they received an EUR 20 gift card. None of them was professionally related to our research group nor the field of acoustics. The experiment took about 75–90 minutes and consisted of eleven phases. The phases were: Reading and signing consent form, answering Weinstein's noise sensitivity questionnaire (Weinstein 1978), hearing sensitivity measurements, familiarization phase to become familiar with the future sounds, rehearsal phase for practicing the use of the subjective loudness rating scale, Block 1, Block 2, a 5-minute break, rehearsal phase for practicing the use of the subjective annoyance rating scale, Block 3, and Block 4. In the familiarization phase the subjects listened to 13 of the sounds used later. The presentation order of the sounds within each Block was predefined with 10 different pseudorandom orders. The reference sounds were evenly distributed with the tonal sounds, and successive tonal sounds did not have the same tonal frequency.

Data analysis and determination of penalty

The research focused on the determination of the penalty values for tonal sounds, and our hypothesis was that the penalty depended on three physical characteristics: overall level, tonal frequency, and tonal audibility. The mathematical method behind the calculation of the penalty values is explained in Figure 15. First, the mean ratings of annoyance for the non-tonal reference sounds are plotted against the sound pressure level of each of these sounds (empty circles). Linear interpolation is done. Thereafter, the annoyance of the tonal noise of interest is added to the plot (black square). The penalty value, k, is obtained from the difference between the level of a reference sound receiving the same annoyance rating than the tonal sound and the level of the tonal sound. In other words, k is the number of decibels that the non-tonal sound should be amplified to be perceived equally annoying as the tonal sound under concern. Positive values of k imply that tonal sounds are more annoying than non-tonal sounds of the same level, while negative values of k the other way around. Most of the data analysis was performed using Excel. SPSS (IBM SPSS 20, IBM Corp, Armonk, NY) was used to test the normality of the rating distributions of each sound and the presence of outliers.

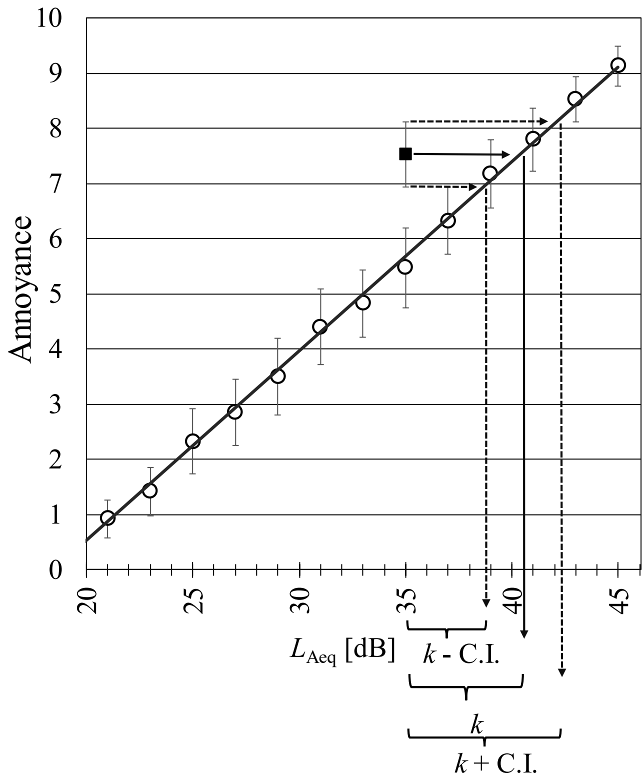


Figure 15. Example of the determination of penalty value *k* for the experimental sound with overall level 35 dB $L_{A,eq}$, tonal frequency 850 Hz, and tonal audibility 17 dB. The penalty (line with arrow) and its uncertainty (dashed line with arrows) determined by the 95% confidence interval were determined by finding the apparent level of the equally annoying non-tonal sound using the fitted line. In this case, the penalty was *k* = 5.3 dB and the confident interval C.I. = 0.9.

5 Results

This chapter presents the results of Publications I–IV. The results highlight the most interesting findings from each original paper. Detailed data about the performance of metrics and single number quantities are only briefly presented here. The data and all details about the applied statistical analyses are better explained in the original publications.

5.1 Results from Publication I

The results were valid thanks to the quality of the experimental design. Friedman's test showed a significant difference between the sounds for all nine subjective measures. This means that the sounds were rated significantly different from each other. The correlation coefficients between the six subjective measures *plea (how pleasant is the sound?)*, *habi (how easy is the noise to get used to?)*, *work (I could work efficiently with this noise for long periods of time)*, *loud (how loud is the noise?)*, *dist (how disturbing is the noise?)*, and *conc (how much the noise would impair your concentration during working?)* was very high. The Cronbach's alpha of the sum variable *Acoustic Satisfaction* (see section Subjective Measures in Chapter 4.1) was very high for every sound, between 0.89 and 0.95. The sum variable was more stable, i.e., less sensitive to small inconsistencies, than the individual measures because it averages the ratings of the six measures. Therefore, it was justified to base our conclusions of the subjects' preference order of the sounds in terms of the sum variable *Acoustic Satisfaction*. The results are shown in Figure 16. The vertical axis presents the names of the 11 investigated sounds and the horizontal axis is the *Acoustic Satisfaction*. The pair comparison analysis enabled to differentiate two groups A and B of sounds. Sounds in A or B group do not significantly differ from the most satisfactory and most dissatisfactory, respectively.

The correlation between the single number quantities and the dependent variables enable also to determine what SNQ correlated best with *Acoustic Satisfaction*, see Figure 17. The metrics (vertical axis) with higher Pearson's correlation coefficients (horizontal axis) predicted *Acoustic Satisfaction* best. The best correlation with *Acoustic Satisfaction* was obtained with *STI1* (R_P=0.47 in Fig. 16, $p<0.001$).

Loudness level $L_{N,ANSI}$ predicted the *Acoustic Satisfaction* almost equally well as L_{Lo-Hi}, *SIL*, and *NCB* (see definitions in Publication I).

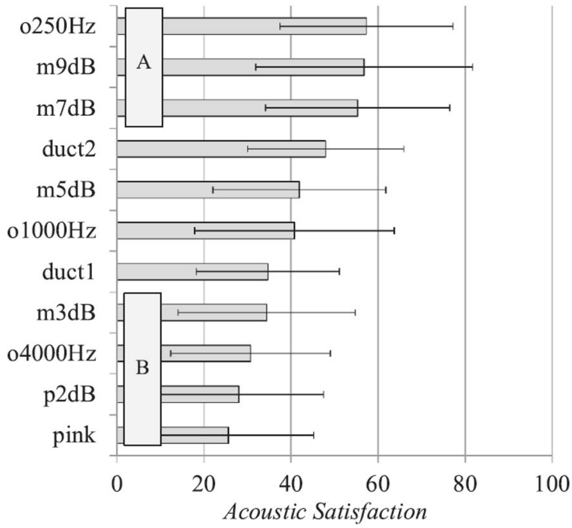


Figure 16. The means and the standard deviations of the sum variable *Acoustic Satisfaction*. The sounds that were not significantly different from the most satisfactory sound (o250Hz) and the most dissatisfactory sounds (pink) are joined to groups A and B, respectively.

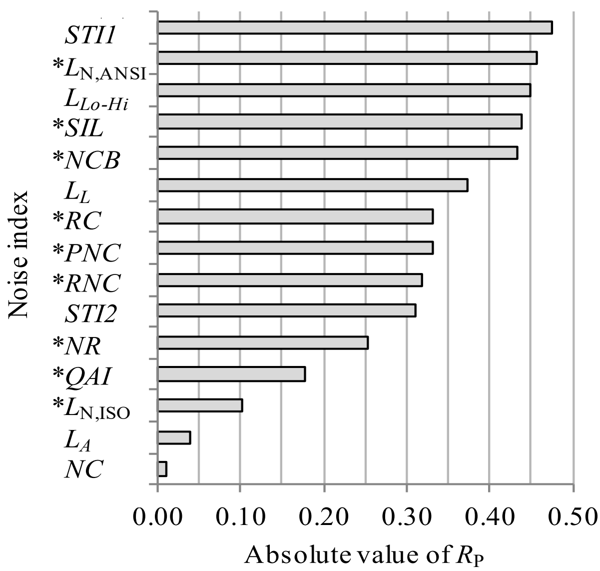


Figure 17. The absolute Pearson's correlation coefficients between the sum variable *Acoustic Satisfaction* and the 15 noise metrics or SNQs (vertical axis). The indices are sorted from best to worst. Absolute values of R_P exceeding 0.25 indicates a significant correlation (p<0.001) with Acoustic Satisfaction. *, the value of RP was negative.

5.2 Results from Publication II

The squared Pearson correlation coefficients, R^2, between the 12 studied metrics and the subjective measures of *loudness*, *disturbance*, and *acceptability* for the six *sound types* are presented in Tables 1, 2, and 3, respectively. In each table, the first column is the metric, and the second column is the frequency range in Hertz which it comprehends. Columns 3–8 are the R^2 values for the six sound types, and the last column, Mean, indicates the average R^2 value over the six sound types. The correlation curves between four selected SNQs (R_w, $R_w+C_{50–3150}$, $R_w+C_{100–3150}$, $R_w+C_{tr,50–3150}$) and *disturbance* are shown in Figure 18.

Table 1. The squared Pearson correlation coefficients (R^2) between the SNQs and the average rating of *loudness.* The limiting values for statistical significance were R^2>0.444 (p<.05), R^2>0.637 (p<0.01, underlined values), and R^2>0.806 (p<0.001, bolded values).

a) Loudness	Range [Hz]	Guitar	Music (T)	Music (L)	Baby	Speech	Dog	Mean
R_w	100–3150	**0.92**	**0.90**	**0.87**	0.63	**0.92**	0.71	**0.83**
$R_w + C_{100–3150}$	100–3150	**0.91**	**0.92**	**0.86**	0.56	**0.88**	0.62	0.79
$R_w + C_{100–5000}$	100–5000	**0.91**	**0.93**	**0.87**	0.57	**0.88**	0.62	0.80
$R_w + C_{50–3150}$	50–3150	**0.83**	**0.93**	**0.82**	0.49	0.74	0.47	0.71
$R_w + C_{50–5000}$	50–5000	**0.83**	**0.93**	**0.82**	0.49	0.74	0.47	0.71
$R_w + C_{tr,100–3150}$	100–3150	**0.83**	**0.89**	0.77	0.41	0.76	0.44	0.68
$R_w + C_{tr,100–5000}$	100–5000	**0.83**	**0.89**	0.77	0.41	0.76	0.44	0.68
$R_w + C_{tr,50–3150}$	50–3150	0.52	0.71	0.54	0.24	0.35	0.14	0.42
$R_w + C_{tr,50–5000}$	50–5000	0.52	0.71	0.54	0.24	0.35	0.14	0.42
STC	125–4000	**0.92**	**0.89**	**0.89**	0.69	**0.92**	0.74	**0.84**
STCno8	125–4000	**0.92**	**0.86**	**0.89**	0.70	**0.95**	0.79	**0.85**
R_{speech}	200–5000	0.80	0.68	0.77	0.77	**0.89**	**0.91**	0.80

Table 2. The squared Pearson correlation coefficients (R^2) between the SNQs and the average rating of *disturbance*. The limiting values for statistical significance were R^2>0.444 (p<.05), R^2>0.637 (p<0.01, underlined values), and R^2>0.806 (p<0.001, bolded values).

b) Disturbance	Range [Hz]	Guitar	Music (T)	Music (L)	Baby	Speech	Dog	Mean
R_w	100–3150	**0.95**	**0.86**	**0.90**	0.62	**0.94**	0.77	**0.84**
$R_w + C_{100–3150}$	100–3150	**0.92**	**0.90**	**0.84**	0.54	**0.86**	0.66	0.79
$R_w + C_{100–5000}$	100–5000	**0.93**	**0.90**	**0.85**	0.54	**0.86**	0.66	0.79
$R_w + C_{50–3150}$	50–3150	**0.85**	**0.96**	0.75	0.46	0.70	0.51	0.71
$R_w + C_{50–5000}$	50–5000	**0.86**	**0.96**	0.75	0.46	0.70	0.51	0.71
$R_w + C_{tr,100–3150}$	100–3150	**0.83**	**0.87**	0.70	0.38	0.70	0.47	0.66
$R_w + C_{tr,100–5000}$	100–5000	**0.83**	**0.87**	0.70	0.38	0.70	0.47	0.66
$R_w + C_{tr,50–3150}$	50–3150	0.54	0.79	0.42	0.22	0.31	0.17	0.41
$R_w + C_{tr,50–5000}$	50–5000	0.54	0.79	0.42	0.22	0.31	0.17	0.41
STC	125–4000	**0.95**	**0.88**	**0.92**	0.68	**0.93**	0.80	**0.86**
STCno8	125–4000	**0.95**	**0.81**	**0.94**	0.69	**0.98**	**0.84**	**0.87**
R_{speech}	200–5000	**0.84**	0.63	**0.90**	0.79	**0.98**	**0.97**	**0.85**

Table 3. The squared Pearson correlation coefficients (R^2) between the SNQs and the average rating of *acceptability*. The limiting values for statistical significance were R^2>0.444 (p<.05), R^2>0.637 (p<0.01, underlined values), and R^2>0.806 (p<0.001, bolded values).

c) Acceptability	Range [Hz]	Guitar	Music (T)	Music (L)	Baby	Speech	Dog	Mean
R_w	100–3150	**0.92**	**0.87**	**0.85**	<u>0.67</u>	**0.88**	<u>0.77</u>	**0.83**
$R_w + C_{100-3150}$	100–3150	**0.87**	**0.87**	<u>0.78</u>	0.58	<u>0.78</u>	<u>0.66</u>	<u>0.76</u>
$R_w + C_{100-5000}$	100–5000	**0.87**	**0.88**	<u>0.78</u>	0.58	<u>0.78</u>	<u>0.66</u>	<u>0.76</u>
$R_w + C_{50-3150}$	50–3150	<u>0.76</u>	**0.91**	<u>0.67</u>	0.49	0.62	0.51	<u>0.66</u>
$R_w + C_{50-5000}$	50–5000	<u>0.76</u>	**0.91**	<u>0.68</u>	0.49	0.62	0.51	<u>0.66</u>
$R_w + C_{tr,100-3150}$	100–3150	<u>0.74</u>	**0.81**	0.62	0.41	0.61	0.47	0.61
$R_w + C_{tr,100-5000}$	100–5000	<u>0.74</u>	**0.81**	0.62	0.41	0.61	0.47	0.61
$R_w + C_{tr,50-3150}$	50–3150	0.43	<u>0.69</u>	0.35	0.21	0.24	0.17	0.35
$R_w + C_{tr,50-5000}$	50–5000	0.43	<u>0.69</u>	0.35	0.21	0.24	0.17	0.35
STC	125–4000	**0.91**	**0.90**	**0.87**	<u>0.74</u>	**0.87**	<u>0.79</u>	**0.85**
STCno8	125–4000	**0.95**	**0.83**	**0.91**	<u>0.74</u>	**0.94**	**0.85**	**0.87**
R_{speech}	200–5000	**0.88**	<u>0.69</u>	**0.91**	**0.85**	**0.99**	**0.98**	**0.88**

5.3 Results from Publication III

The subjective ratings of *loudness* and *annoyance* for each *sound type* for every one of the 12 façades are presented in Figure 19. The values are averages calculated from participants ratings. The analysis to verify the potential of the studied SNQs to predict annoyance or loudness is done in two parts. First, the squared correlation coefficients R^2 between the SNQs and the subjective ratings of *loudness* and *annoyance* are calculated. This data is presented in Table 4. SNQs with high R^2 values predict annoyance or loudness better than those obtaining smaller values. The squared correlation coefficient was statistically significant at a level of p=0.01 (one-tailed) when R^2>0.13. Values smaller than 0.13 are indicated by n.s in the table. Second, the SNQs are sorted from best to worst depending on how well they predicted loudness and annoyance. The SNQ with the largest R^2 receives rank 1 while the SNQ with the smallest R^2 receives the last rank. The rank order of the 25 studied SNQs for *loudness* and *annoyance* for each *sound type* is presented in Table 5.

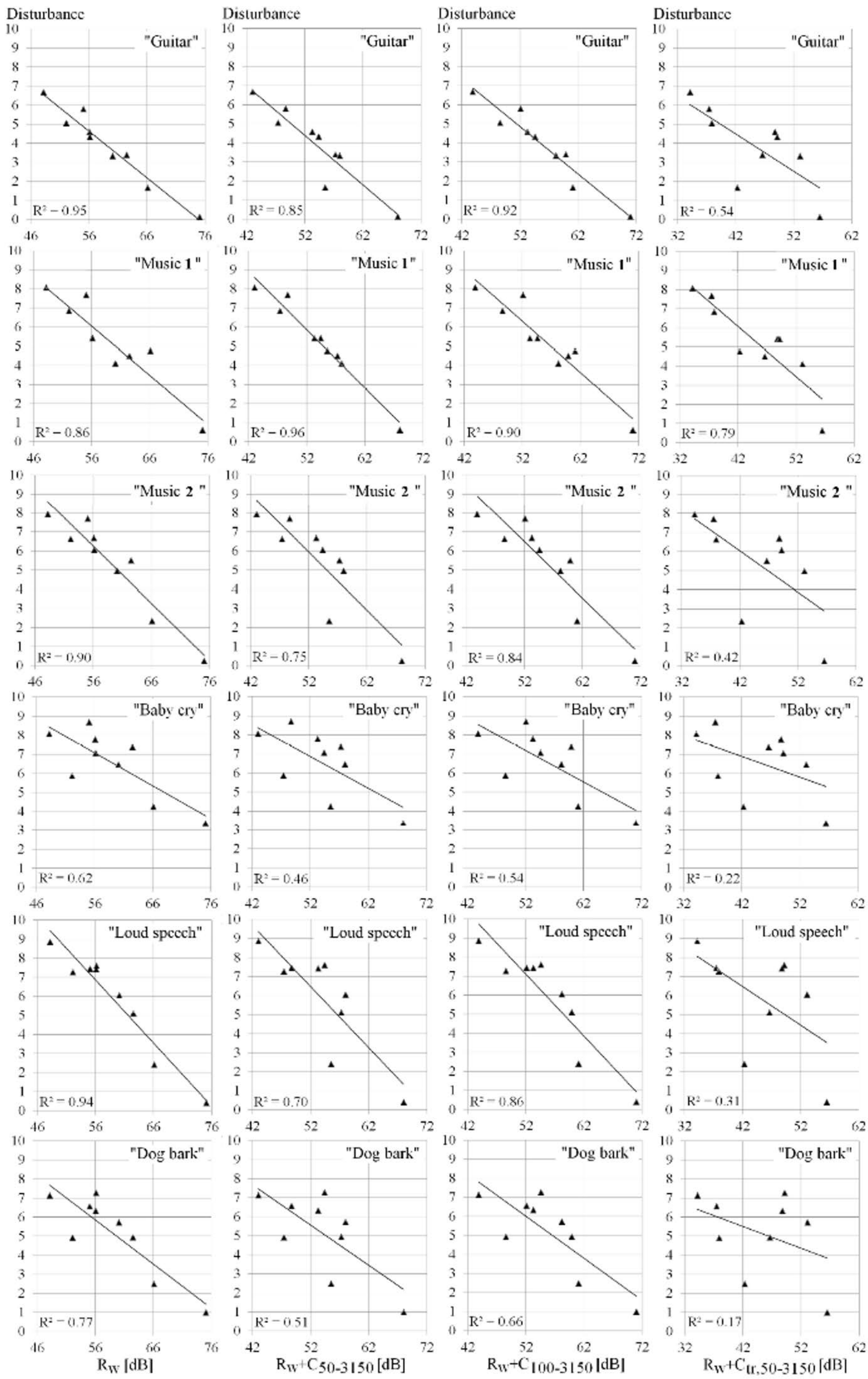


Figure 18. The linear correlations between four standardized sound insulation ratings (R_w, R_w+$C_{50-3150}$, R_w+$C_{100–3150}$, R_w+$C_{tr,50–3150}$) with respect to subjective ratings of *disturbance* for the six *Sound types*.

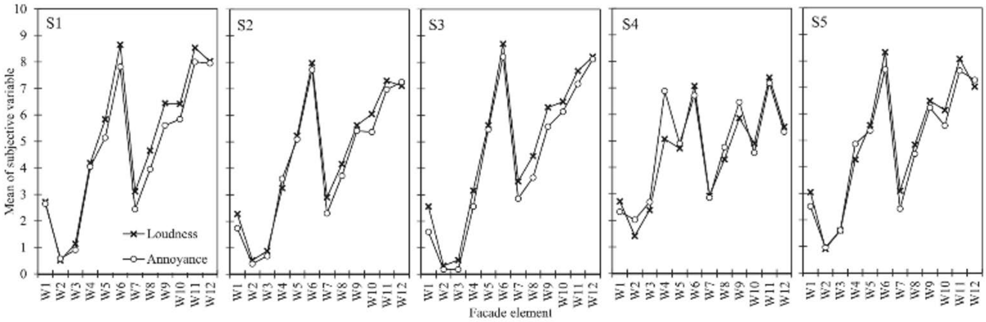


Figure 19. Distribution of the mean of the subjective variables *loudness* and *annoyance* for each of the façade elements W1–W12 and for each of the five *sound types* S1–S5.

Table 4. Squared Pearson's correlation coefficients, R^2, between the 25 analyzed metrics and the subjective *loudness* and *annoyance* for the five *sound types* S1–S5.

	Loudness					*Annoyance*				
SNQ	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
R_w	0.74	0.73	0.78	0.48	0.69	0.55	0.56	0.64	0.21	0.50
$R_w + C_{100–3150}$	0.74	0.72	0.77	0.49	0.69	0.55	0.56	0.62	0.22	0.50
$R_w + C_{100–5000}$	0.74	0.72	0.77	0.49	0.69	0.55	0.56	0.62	0.21	0.50
$R_w + C_{50-3150}$	0.74	0.72	0.76	0.52	0.70	0.56	0.57	0.62	0.25	0.52
$R_w + C_{50–5000}$	0.75	0.72	0.76	0.52	0.70	0.56	0.57	0.62	0.25	0.52
$R_w + C_{tr,100–3150}$	0.71	0.69	0.72	0.51	0.68	0.53	0.53	0.57	0.24	0.49
$R_w + C_{tr,100–5000}$	0.71	0.69	0.72	0.51	0.68	0.53	0.53	0.58	0.24	0.50
$R_w + C_{tr,50–3150}$	0.66	0.62	0.63	0.55	0.64	0.50	0.51	0.50	0.33	0.50
$R_w + C_{tr,50–5000}$	0.66	0.62	0.63	0.55	0.64	0.50	0.51	0.50	0.33	0.50
$STA_{100–5000}$	0.74	0.72	0.77	0.49	0.69	0.55	0.56	0.62	0.21	0.50
$STA_{50–5000}$	0.74	0.72	0.76	0.52	0.70	0.56	0.57	0.62	0.25	0.52
$AA_{100–5000}$	0.71	0.72	0.77	0.46	0.67	0.51	0.55	0.63	0.21	0.48
$AA_{50–5000}$	0.71	0.71	0.76	0.48	0.67	0.51	0.55	0.62	0.24	0.48
STC	0.72	0.71	0.76	0.46	0.67	0.53	0.54	0.61	0.19	0.48
STC_{no8}	0.74	0.73	0.78	0.47	0.68	0.55	0.56	0.64	0.21	0.49
$EA_{100–5000}$	0.62	0.58	0.59	0.51	0.61	0.45	0.45	0.46	0.26	0.45
$EA_{50–5000}$	0.42	0.38	0.36	0.45	0.43	0.30	0.32	0.28	0.34	0.35
EA_{63}	0.16	0.15	0.13	0.23	0.17	n.s.	0.14	n.s.	0.23	0.16
EA_{125}	0.34	0.31	0.30	0.38	0.38	0.22	0.23	0.21	0.23	0.26
EA_{250}	0.57	0.54	0.57	0.42	0.56	0.43	0.40	0.43	0.20	0.40
EA_{500}	0.72	0.70	0.76	0.47	0.67	0.53	0.55	0.62	0.21	0.49
EA_{1000}	0.64	0.66	0.73	0.36	0.57	0.47	0.51	0.61	0.15	0.41
EA_{2000}	0.65	0.67	0.74	0.38	0.59	0.47	0.52	0.62	0.18	0.43
EA_{4000}	0.56	0.57	0.62	0.33	0.51	0.40	0.43	0.53	0.15	0.37
OITC	0.65	0.61	0.62	0.54	0.64	0.49	0.49	0.49	0.30	0.49

Table 5. The rank order of the 25 studied SNQs for *loudness* and *annoyance* based on the R^2 values of Table 4. The last column represents the overall rank of each SNQ according to the ten rank orders obtained for each subjective variable in the five *sound types*.

	Loudness					*Annoyance*					
SNQ	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	Overall Rank
R_w	4	1	1	13	7	4	4	2	18	6	4
$R_w + C_{100-3150}$	7	6	5	10	5	6	7	6	14	7	6
$R_w + C_{100-5000}$	6	3	3	11	4	7	6	4	15	8	5
$R_w + C_{50-3150}$	3	8	9	4	2	3	3	12	6	2	2
$R_w + C_{50-5000}$	1	4	7	5	1	1	1	10	7	1	1
$R_w + C_{tr,100-3150}$	14	14	16	8	10	12	14	16	9	11	14
$R_w + C_{tr,100-5000}$	13	13	15	7	9	11	13	15	10	10	9
$R_w + C_{tr,50-3150}$	16	18	18	2	16	16	18	19	3	5	15
$R_w + C_{tr,50-5000}$	15	17	17	1	15	15	17	18	2	4	12
$STA_{100-5000}$	8	5	4	12	6	8	8	5	16	9	8
$STA_{50-5000}$	2	7	8	6	3	2	2	11	8	3	3
$AA_{100-5000}$	11	9	6	18	14	13	11	3	19	17	13
$AA_{50-5000}$	12	10	12	14	11	14	10	9	11	15	11
STC	10	11	11	17	12	10	12	13	22	16	16
STC_{no8}	5	2	2	15	8	5	5	1	20	12	7
$EA_{100-5000}$	20	20	21	9	18	20	20	21	5	18	19
$EA_{50-5000}$	23	23	23	19	23	23	23	23	1	23	21
EA_{63}	25	25	25	25	25	25	25	25	13	25	25
EA_{125}	24	24	24	22	24	24	24	24	12	24	24
EA_{250}	21	22	22	20	21	21	22	22	21	21	22
EA_{500}	9	12	10	16	13	9	9	8	17	13	10
EA_{1000}	19	16	14	23	20	19	16	14	24	20	20
EA_{2000}	18	15	13	21	19	18	15	7	23	19	18
EA_{4000}	22	21	20	24	22	22	21	17	25	22	23
OITC	17	19	19	3	17	17	19	20	4	14	17

5.4 Results from Publication IV

The subjective *loudness* and *annoyance* ratings for the non-tonal sounds are presented in Figure 20. The $L_{A,eq}$ in Blocks 1 and 3 was 25 dB, while in Blocks 2 and 4 the overall A-weighted level was 35 dB. In Figure 21, the subjective *loudness* and *annoyance* ratings for the tonal sounds are presented. The penalty values k for annoyance are shown in Figure 22, in the left and right for tonal sounds with $L_{A,eq}$ = 25 or 35 dB, respectively.

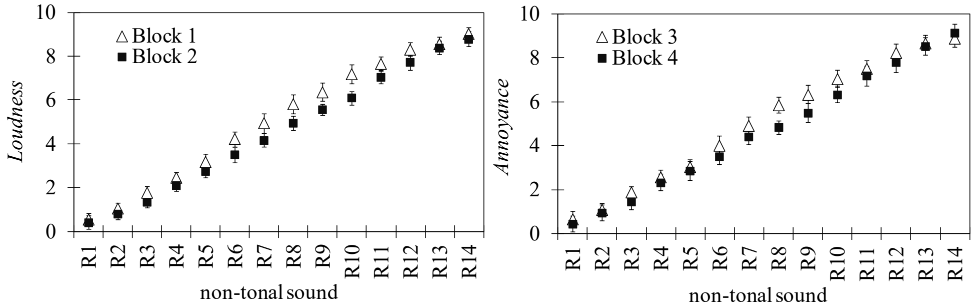


Figure 20. Mean rating values and the 95% confidence intervals for the non-tonal reference sounds R1–R14. Left) *loudness* in Block 1 ($L_{A,eq}$ = 25 dB) and Block 2 ($L_{A,eq}$ = 35 dB). Right) *annoyance* in Block 3 ($L_{A,eq}$ = 25 dB) and Block 4 ($L_{A,eq}$ = 35 dB).

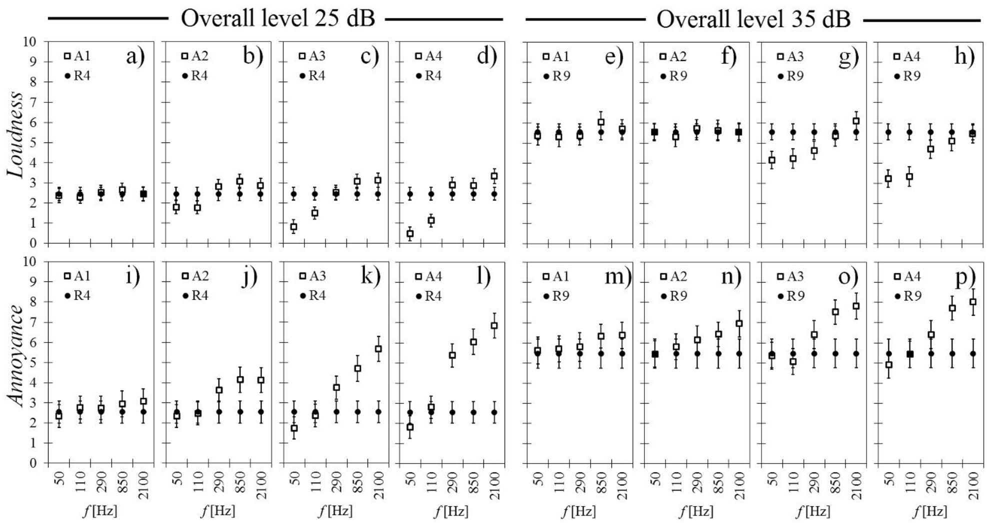


Figure 21. Mean *loudness* ratings (upper figures) and mean *annoyance* ratings (lower figures) of the tonal sounds for each of the four tonal audibility levels A1–A4. The tonal audibility levels, i.e., the prominence of the tones, were 5, 10, 17.5 and 25 dB, respectively. For comparison purposes, the ratings of the non-tonal sound with similar sound pressure level are also shown; R4 with $L_{A,eq}$ = 25 dB in the left-hand figures (figures a, b, c, d, i, j, k and l), and R9 with $L_{A,eq}$ = 35 dB in the right-hand side figures (figures e, f, g, h, m, n, o, p). In all figures, the open square marks represent tonal sounds, and the closed square marks are non-tonal reference sounds.

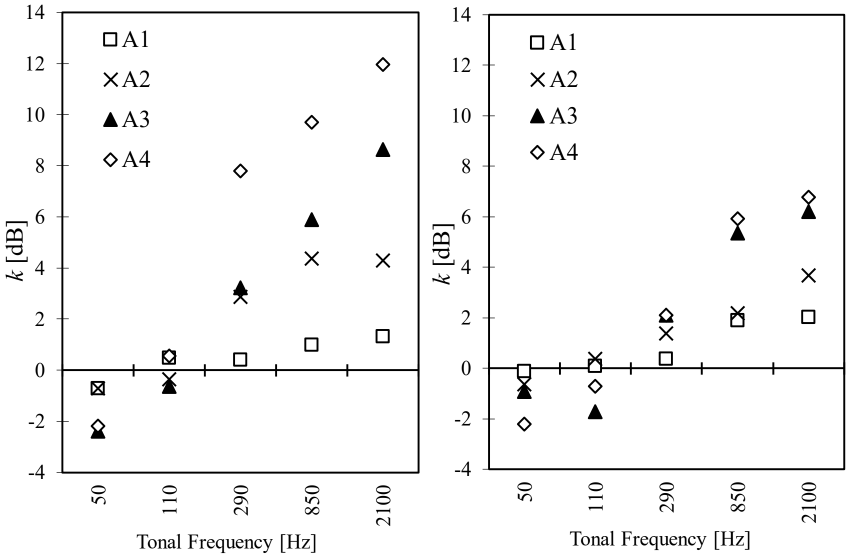


Figure 22. The mean penalty values *k* calculated from *annoyance ratings* as a function of the tonal frequency for tonal audibility levels A1–A4. The tonal audibility levels, i.e., the prominence of the tones, were 5, 10, 17.5 and 25 dB, respectively. Left) for tonal sounds at 25 dB $L_{A,eq}$ Right) for tonal sounds at 35 dB $L_{A,eq}$.

6 Discussion

This section discusses the possible contribution of Publications I–IV and this Research to the field of acoustics. All the work aimed to determine the single number quantities and metrics that correlate best with annoyance perception. This knowledge could be used in future updates of noise codes related to community noise in dwellings, and hence it might interest acousticians and authorities deciding on noise regulations. In our opinion, the new data performs better because it is based on annoyance ratings and not loudness. For the type of noise conditions we investigated, annoyance is the most possible health reaction, and it makes sense to change the current metrics and limits based on that new knowledge and new sets of measurements. Each Publication solved a different type of community noise, and we hope that could be our contribution to the fields of environmental medicine and public health science. It would be convenient to verify the presented findings with new psychoacoustic experiments.

Furthermore, we would like to present and discuss critically some of the methods and the knowhow we developed during this Research. For clarity reasons we have moved that discussion to Appendix 1. We think that should be useful to other researchers preparing new psychoacoustic experiments. Topics include laboratory and field setups, preparation of laboratory, and preparation of sounds.

6.1 Publication I

Publication I focused on the noise field experienced in open-plan offices. The first aim was to identify what spectra of ventilation noise were the most acceptable and the most disturbing for people performing cognitive tasks in this type of environment. The 11 experimental sounds had different spectra but identical sound pressure level $L_{A,eq}$ = 42 dB. This arrangement enabled us to compare the subjective ratings based only on the sound spectra. The results are in principle only valid for sounds with the same $L_{A,eq}$. Based on our experience and on published data, e.g., ISO 226:2003, we presume the range of validity of these results could be extended to at least ±8 dB over that level. However, we do not recommend using noises with higher levels than 42 dB $L_{A,eq}$ to mask speech, as 45 dB has proved to cause discomfort in open plan offices (Lenne et al. 2020).

The contribution of this work is twofold. First, manufacturers of ventilation equipment could use the published data to fine-tune the noise emission spectra of their machines and outlets. We recommend that ventilation equipment producing at the working positions similar spectra as the experimental sounds achieving the worst *acoustic satisfaction* should be avoided or modified. Our results indicated that noises with sound energy mostly at low frequencies, like test sounds o250Hz, m9dB and m7dB, were better tolerated. Sounds with energy at high pitch received the worst ratings. Second, developers of speech masking systems could apply the provided knowledge to evaluate the expected performance of their own devices. For any given masking noise, it is possible to calculate the SNQs which correlated best with *Acoustic Satisfaction*. After that, the performance of the device can be analyzed comparing the new measured values with those of the noises receiving the best ratings of *acoustic satisfaction* in our study. Our results say that the most important SNQs for this noise situation are the four in the top part of Figure 17, specially *STI1* (speech transmission level). This metric was developed to measure speech intelligibility. Humans are more sensitive to the frequency range where speech consonants are located, and perhaps using a metric that focuses in this area makes sense, since irrelevant speech was the most common source of annoyance in offices (Haapakangas et al. 2011). However, if we had to choose one metric it would be *SIL* (speech interference level as defined in ANSI S12.2:2008). This metric worked as well as the other best ones, i.e., *STI1*, $L_{N,ANSI}$, and *NCB*, and it is not so laborious to calculate.

Out of the sounds we studied, our recommendation for background noise in offices is m7dB (sound with slope of -7 dB per octave within octave bands 63 and 8000 Hz). It provides both *acoustic satisfaction* and good speech masking. These conclusions agreed well with the results published by Veitch et al. (2002), who also investigated the acoustic satisfaction produced by background noises of different spectra. For instance, both studies demonstrated the preference of office workers towards background noises with low frequency content but good masking properties.

Acoustic satisfaction was defined in Navai and Veitch (2003) as a dimension of environmental satisfaction, i.e., the state of contentment with physical environmental conditions. In our work, *acoustic satisfaction* was a numerical value obtained averaging the responses of test subjects to the loudness and five other attitude questions, which were answered on a line scaled from 0 to 100 (see end of section 4.1). A similar approach was previously applied by Veitch et al. (2002), who calculated acoustic satisfaction from the responses to 14 questions. However, in the work of Veitch, the questions could be answered using three different rating scales, e.g., 5-point scales, 7-point scale, and one 0–100 sliding scale question concerning the intelligibility of the speech sounds. We did not find information on their published paper about how these ratings were combined to obtain the final acoustic satisfaction value. In our work, the decision to calculate the *acoustic satisfaction* was made once we verified that the loudness and the five attitude measures correlated strongly with each other. Cronbach's alpha of this variable was determined for all 11 sounds and the values were high, between 0.89 and 0.95. Using only one dependent variable, instead of six, simplified and improved the quality of the statistical analysis.

The analysis of SNQs also gives information about the standardized method to measure loudness that performs best. $L_{N,ISO}$ (ISO 532, 1975) is based on equal loudness contours. The experimental sounds had identical $L_{A,eq}$, and therefore their loudness measured according to the ISO standard was mostly constant. By contrast, $L_{N,ANSI}$ (ANSI S3.4, 2007) provided a larger range of numerical values and it correlated well with the subjective measures. Results recommend the use of $L_{N,ANSI}$ over $L_{N,ISO}$ to predict the subjective perception of sounds with similar SPL. However, it should be noted that standard ISO 532:1975 was updated after the publication of this study to ISO 532:1-2017 and ISO 532:2-2017, and thus it is possible that the new methods proposed by the International Organization for Standardization perform better than the older version we used.

Last, in this experiment $L_{A,eq}$ was not a good descriptor, because all our sounds had the same $L_{A,eq}$. Previous field surveys in office workspaces and open field, for instance Keighley (1970), Tang et al. (1996), and Ayr et al. (2003), have shown that A-weighted level correlated well with subjective experience. Not necessarily our results conflict with the previous studies, but we proved that two sounds with equal A-weighted levels could produce significantly different subjective experiences and ratings.

6.2 Publication II

The research investigated how six living noises were perceived in neighbor's apartments when they are transmitted through 12 wall types. The main goal of the research was not to identify the most annoying noises nor the insulation performance of the 12 walls, but the performance of available metrics to describe sound insulation. The performance was based on subjective annoyance ratings. The main conclusions regarding this research are based in the mean R^2 presented in Tables 1, 2, and 3 for *loudness*, *disturbance*, and *acceptability*. The R^2 represents the results of the correlation analysis, and it tells how well each metric can predict any of the three subjective measures. *STCno8*, *STC*, R_w, and R_{speech} where the best predictors of *loudness*, *disturbance*, and *acceptability*. All other SNQs performed significantly worse than these four.

The mean R^2 is a good compromise since all sounds were distinct and had different spectra. The results are not the same when the analysis is performed for each individual *sound type*, because each had different spectra, and therefore, some

SNQs adapted better than others to the weighting terms embedded in the calculation process. For instance, *Music1* had strong low frequency content, and the SNQs that expanded the frequency range of interest to include from 50 to 80 Hz performed well, for instance $R_w + C_{50–3150}$ and $R_w + C_{50–5000}$. A very detailed discussion about the suitability of all SNQs with respect to each individual *sound type* was included in Publication II.

None of the SNQs reached very significant R^2 values for the sound "Baby cry". Considering simultaneously the three subjective measures the best descriptors were *STC* and *STCno8*. "Baby cry" had a considerable amount of sound energy at high frequencies. However, it could occur that poor correlation is the result of opposing psychological emotions between the subjects with respect to that noise. In our research we did not make use of the results from the Weinstein Sensitivity test that subjects performed at the beginning of the experiment. They might be the objective of a future paper. Almost similar results were obtained with "Dog bark", which was predicted best by R_{speech} when *loudness*, *disturbance* and *acceptability* were simultaneously considered.

The sound types used in this experiment belong to the most disturbing sounds in homes according to Hongisto et al. (2013) and Rychtáriková et al. (2012). The sounds were played down to 50 Hz to test the adequacy of few SNQs reaching these frequencies. To calculate the likelihood of people listening such low frequencies, we mostly would need the proportion of people using subwoofers and good speakers to play television and music. We believe playing both "Music1" and "Music2" from 50 Hz was scientifically justified since it represented cases occurring in real dwellings.

The results presented in this study are valid only for the selected sound types, their level range, walls, and frequency band. Different results might appear when other sounds are used, or different walls are simulated. Extrapolation to other types of sounds and constructions should therefore be done with care, especially at levels higher than the ones we used, which were $L_{A,eq}$ = 20–40 dB. The 12 walls are a good range of constructions used in Finland and Europe, and we reproduce a total of 72 laboratory sounds to analyze how they perform with respect to six types of living noises. The process to sort the performance order of the 12 walls based on loudness, annoyance, and acceptability ratings worked well.

Furthermore, we determined the SNQs that predict best the three subjective measures. Because the 72 sounds represent actual cases well, we expect the experiment was valid to determine what are the best SNQs to measure sound insulation. The SNQs that we took in our study represent what is currently used in Europe. The presented engineering data could be used in future updates of building codes in any country. Considering that noise annoyance as a health impact can be produced by many types of noise, we propose the use of metrics that performed well for most noises, like *STCno8*, *STC*, R_w, and R_{speech}. Using metrics that work only for one type of noise might be applied when needed. We limited our analysis to SNQs used in Europe and found in ISO 717-1:2013 and ASTM E413-10:2010. The analysis could be continued with less usual SNQs like EA and AA which have been used in other studies, like in Park & Bradley 2009.

The length of the stimuli was 18 seconds. After that, the sound continued playing until the subjective rating was given. Typical exposure times in real life are longer, and thus the experience of the sound might be different in the real set-up. This however might not severely affect the results presented here. Our results are based on the difference between the ratings given for each sound, and we expect exposure time to have only a small influence on the ratings, not affecting the main results of the research. Laboratory set-ups are not able to reproduce real situations in which it is possible to control the noise, for instance talking to the neighbor or changing the activity or the location to mitigate the impact of the noise event. Long exposure in real life cases might produce different results, since stressors like repetitive music listening or dog barking might increase levels of noise sensitivity. Habituation and noise sensitivity were not investigated in our experiment. None of our sounds was especially annoying, but it could be hard to habituate to any of them if it is experienced repeatedly and with no control. We never asked the subjects, which one of the six sound types would be the less preferred to be repeatedly occurring in their homes.

The age of the subjects, median 26, was below the average in Finland (45). Only one-third of the subjects were male. Age and gender effects were not looked for.

6.3 Publication III

Publication III was designed to verify which internationally standardized single number quantities, SNQs, for sound insulation characterization of façades performed best with respect to the subjective *loudness* and *annoyance*. Five traffic sounds, S1–S5 were used. The façade constructions were ordered with respect to their performance towards the two subjective measures. The squared correlation coefficients, (see Table 4), revealed statistically significant differences between the 11 investigated SNQs. The largest differences and smallest differences were obtained for *sound types* S3 (light vehicles in motorway at 100 km/h) and S5 (both heavy and light vehicles in urban street at 60 km/h following ISO 717-1:2013 spectrum), respectively. Therefore, we considered justified to conduct a rank analysis to find out which SNQs were the best predictors of *loudness* and *annoyance* for each *sound type*, (see Table 5). We did not find in the literature any paper publishing similar rank analysis based in such a large set of data. $R_w + C_{50\text{–}5000}$ was the best descriptor of *sound types* S1 (light vehicles in urban street at 50 km/h), S2 (light vehicles in motorway at 80 km/h), and S5 (light and heavy vehicles in urban street at 60 km/h). It received the largest number of first-order ranks. This is important because S1, S2 and S5 are in our opinion among the most relevant types of road-traffic noises.

The differences between the SNQs to predict the subjective measures were expected, because each SNQ applies a different weighting method over the frequency spectra of the noise. For instance, S3 had a strong proportion of sound energy in the high frequency range, and that explains why R_w and *STCno8*, which ignore 50–80 and 50–100 Hz frequency ranges respectively, were the best descriptors for that specific sound. Similarly, S4 (heavy vehicles in urban street at 60 km/h) had strong low frequency content, and it was best predicted by $R_w + C_{tr,50–5000}$ (*loudness*) and $EA_{50–5000}$ (*annoyance*). It is uncommon that roads are only used by heavy vehicles, except close to harbors, truck depots, or bus stations, and sometimes only at peak times. Therefore, the use of these two metrics could only be reserved for those situations exactly. S5 was probably the most important *sound type* in our study since it contains both light and heavy vehicles at 60 km/h speed. This combination is quite common in cities. The spectrum of S5 matched the road-traffic spectrum of ISO 717-1:2003.

The subjective ratings of *loudness* and *annoyance* were in close agreement, but the standard deviation of *annoyance* was larger (see Figure 19). There can be two reasons for that; first, annoyance is a subjective perception of a sound while loudness is an auditory perception. Second, *annoyance* might have more subjective character than *loudness*, and so it is more susceptible to individual differences.

6.4 Publication IV

The results presented in Figure 21 confirmed our Hypothesis 1; the penalty values are different when calculated from *loudness* and *annoyance*. We believe that test subjects perfectly understood them as separate subjective measures and rate them properly. Tonal sounds were normally rated as less loud than the non-tonal sound with the same $L_{A,eq}$ for tonal frequencies 50, 110 and 290 Hz. The contrary was observed for tonal frequencies 850 and 2100 Hz. Regarding *annoyance*, tonal sounds were rated as more annoying with respect to non-tonal sounds with identical $L_{A,eq}$ when the tonal frequencies were 290, 850 and 2100 Hz. The main results, penalty values k (Figure 22), partly support Hypothesis 2, i.e., the penalty depends on tonal frequency. There is a clear tendency of the penalty to increase with the tonal frequency. The penalty values were statistically significantly larger than 0 dB at the three highest tonal frequencies. The penalty was statistically non-significant at the two lowest tonal frequencies. Furthermore, the results partially support Hypothesis 3, i.e., the penalty depends on tonal audibility. At 290 Hz and above, but not below, there is a clear tendency of the penalty to increase with the tonal audibility.

Different penalty values *k* were obtained at 25 and 35 dB $L_{A,eq}$. The results proved Hypothesis 4, i.e., the penalty value depends on the overall level. It seems the penalty depends on the overall level at which the sounds are objectively experienced. Penalty *k* was higher at 25 dB than at 35 dB $L_{A,eq}$ for tonal frequencies 290, 850 and 2100 Hz. We presented the Blocks in the same order, so there could be an order effect here that we did not control. The order of the Blocks was not counterbalanced because our primary aim was to analyze the perception of each individual Block, $L_{A,eq} = 25$ or 35 dB, but not the differences between them, which in turn we obtained. We predefined with pseudorandom orders the appearance of the experimental sounds to avoid effects related to the harmony of the tonal components. It could be worthwhile to investigate next if the low sound pressure level of the masking explains why the penalty values are different at 25 and 35 dB $L_{A,eq}$. When the level was low, 25 dB, the masking component of the test sound was close and even under the hearing threshold, but that tone was audible. This absence of masking could make the annoyance feeling more intense. Tonal audibility, as defined in ISO 1996-2:2007 and DIN 45681:2005, is the level difference between the tone and the masking. However, the calculation of the penalty values in these standards does not consider cases where masking is inaudible or close to it. This situation occurs in quiet residential environments when appliances produce clearly distinguishable tonal components.

Furthermore, the calculated penalty values obtained in this research are not in agreement with the penalty values provided in standards ISO 1996-2:2007 and DIN 45681:2005, neither in noise codes applied in Finland (Ministry of the Environment 2017, Ministry of Social Affairs and Health 2015). DIN 45681 predicted a 6 dB penalty *k* basically for all tonal sounds with tonal audibility 17 dB and 25 dB. The penalty was approximately 4 dB for tonal audibility 10 dB, and 1–3 dB for tonal audibility 5 dB. Our values disagreed even more with the predictions by ISO 1996-2, which independently of the tonal frequency granted a penalty of 6 dB for all tonal sounds with tonal audibility 10 dB and above. The reason for the disagreement may be that the standards are partially based on studies applying higher overall sound pressure levels than ours, and/or that they do not consider the sound pressure level of the sound at all. These deficiencies could be considered in future revisions of standards related to tonal sounds.

A special case appeared for tonal sound with $L_{A,eq} = 35$ dB, tonal frequency 50 Hz, and tonal audibility level of 17 dB. The penalty was statistically significantly negative. This seems to be in contraposition to typical conceptions regarding the negative impact of low frequency sounds. Previous literature, for instance Pawlaczyk-Luszczynska et al 2003, Angerer et al. 1991, More & Davies 2010, Di et al. 2015, using higher SPLs than ours, showed that sounds with low frequency were perceived as more annoying than sounds of higher pitch at comparable sound pressure levels. We obtained different results in this experiment, where noises with high frequency tones produced more negative annoyance responses. Our results are supported also by the findings in Publication I. They suggest that low frequency tones and noise might be preferred over high frequency cases, at least at the sound pressure level range that we studied. Furthermore, our results confirmed that, contrary to current situation, for instance ISO 1996-2:2007 and DIN 45681:2005, penalty is not needed when the A-weighed SPL is 25 and 35 dB for tonal sounds including tones at low frequencies, i.e., 50 and 110 Hz. The penalty increased with the pitch. Our first explanation here is that our experiment, contrary to others', concentrated on sounds with low SPL. The equal loudness contours (Figure 2) explain how low frequencies require higher sound pressure levels than high frequencies to be heard. Once low frequencies are heard, small increments of level can produce large changes of loudness, and therefore, also the disturbance and annoyance are expected to increase.

Comparison with previous laboratory studies was difficult because not many have systematically analyzed the annoyance effects of tonal audibility and tonal frequency in the level range we did. Poulsen (2003) studied low frequency sounds and the suitability of objective methods to predict subjective ratings. They analyzed the perception of eight low frequency sounds at three $L_{A,eq}$ = 20, 27.5 and 35 dB. Half of the sounds included low frequencies tones, i.e., at 25, 62 and 75 Hz. The tonal audibility was not a systematic variable, and the spectra of the sounds were not documented enough to enable a trustworthy comparison. However, as we did, they found tonal sounds producing lower annoyance than non-tonal sounds. Using 20 participants, Hünerbein et al. (2010) investigated equal annoyance contours of six tones with tonal frequencies 32, 44, 72, 115, 180 and 400 Hz, each of them at two tonal audibility levels, 5 and 10 dB. Audibility levels were calculated according to ISO 1996-2:2007. The tones were embedded in three different masking sounds, played at three levels: 39, 44, and 49 $L_{A,eq}$. They also described the tonal audibility, tonal frequency, and overall level to influence subjective perception. They presented, likewise, non-linear relationships between the annoyance and the tonal audibility when the level of the tone is close to the threshold of audibility. Their and our results prove that penalty values should not be constant, as it would occur with ISO 1996-2. The study of Landström et al. (1994) also proved that at low levels, tones with lower frequencies are preferred. The twenty test subjects had to modify the frequency of the tones and mark the most and least acceptable tonal frequency. The overall level of the sound was kept constant at $L_{A,eq}$ = 40 dB. The averaged most and least acceptable frequencies were 58 and 380 Hz. Our results oppose to those of Waye et al. (1997) and Bengtsson et al. (2004), because our test sounds did not have the same physical characteristics. Their test stimuli included amplitude modulated sounds, e.g., the level of the noise or of the tone changes periodically in small time intervals.

They reported amplitude modulation to have an impact on annoyance. The level of our sounds was always constant, and so the results are not comparable. Furthermore, the test subjects participating in each research were reporting sensitiveness to low frequency noise. Thus, their results might not be valid to represent the entire population, because most people do not report to be sensitive to low frequency noise.

The non-tonal sounds were used in this experiment to enable the comparison with the tonal sounds in terms of subjective annoyance. The selection of the spectra of non-tonal sounds requires discussion. The spectrum was designed to be like the shape of the 40-phon equal loudness contour (Figure 2), and for analogy, like the inverse of the A-weighted curve (Figure 3). Thus, the sound energy was balanced along the frequency range because each frequency band contributed equally to the total loudness. Results from Publication I indicated that this type of spectra should not have a special annoying effect on listeners.

The sounds did not have any temporal variation (steady-state sounds), so it was possible for the participants to imagine how the sound would be perceived in their own dwellings, even though the exposure duration was very short compared to real life exposure times in residential environments. Poulsen (1991) reported that annoyance ratings remained similar for exposure times of 1, 5, 15 and 30 min, when participants were instructed to rate the sound in the laboratory in a similar way as we did. Comparably, short exposure times have been used in other listening experiments (Little & Mabry 1969, Nilsson 2007). Based on the above-mentioned studies, we believe that longer exposure times would not have had a strong effect on the reported penalty values.

Furthermore, not only in Publication IV but also in Publications I–III, the results were not based on absolute ratings, but on the differences between the ratings. In other words, it was not so important for us to reproduce real living conditions, but that the subjects would use the rating scale properly and accordingly to the experienced noises.

Our experiment was unique because we investigated low levels and systematically varied tonal frequency and tonal audibility. The results are in principle only valid for the five tonal frequencies and the four tonal audibility levels that were studied. Nevertheless, the results published in Publication IV led to the development of a penalty model that was published recently, Hongisto et al. 2019.

7 Summary/Conclusions

Publications I–IV aimed to find out which single number quantities should be considered in obligatory building codes or national classification schemes. Many countries, even within the European Union, apply different metrics in their national legislation to evaluate the quality of constructions and of living spaces. The values presented in one country might not be relevant in the neighboring one. Here comes the importance of listening experiments to identify which metrics or single number quantities correlate best with subjective perception and annoyance. It can be claimed that we have studied the noise environments experienced by millions of people, and therefore the work we did was relevant and of general interest. Our studies validated and discredited metrics for combinations of noise types and situations. This knowledge is expected to be of use in the future when new noise regulations or guidelines are prepared.

Subjective annoyance is linked to human well-being and health, and so our experiments apply to environmental health research. In our opinion, psychoacoustic experiments in environmental health research are rare worldwide and only few laboratories and research groups perform them with such diligence and level of detail. To our knowledge, we are the only research group in Finland performing this type of consecutive research. The research group performed almost 20 experiments related to subjective listening perception during the last decade, and the author contributed directly to about half of them. The result of that work has been presented in over 50 scientific references, including both peer-reviewed journals and international congresses. We believe our work has had a positive impact on the scientific community and hopefully one day on the society in general.

This thesis presented also the methodology developed to perform psychoacoustic listening tests where test subjects were exposed to ordinary sounds in typical living environments. The author hopes that all this information helps other researchers to perform successful listening experiments.

Acknowledgements

To be honest, four years ago this work was not on my list of priorities. I had left the acoustic group where this research was done, and I was in the process of reinventing myself in another research group and creating my own gaming company. So, to say, there was not much motivation to talk about decibels anymore. I would like to start thanking my supervisor Adj Prof Juhani Soini for convincing me to get this job done and supporting me in this and few other projects. This paper has been written with the support and the critical comments of other experts. Thanks to supervisors Dr Panu Maijala and Professor Pekka Hänninen, and to reviewers Dr Markku Sainio, Dr Nick Zacharov, and Dr Catherine Lavandier. Today, after spending uncountable hours with this paper, I can look back and be glad about getting the job done and its quality. You all contributed to that, and I really appreciate it. My gratitude goes also to several colleagues from the Finnish Institute of Occupational Health and Turku University of Applied Sciences. Dr Valtteri Hongisto is especially thanked for sharing with me correct scientific practices and pushing me forward. Dr Jukka Keränen and Dr Petra Virjonen are deeply thanked for helping me during several years with so many theoretical and practical issues related to acoustics. Thanks to Jarkko Hakala for making me laugh while building the laboratory spaces and installing the required equipment.

At a personal level, I thank my wife Päivi for her patience and support, my three kids Ukko, Eero and Pablo, for making me feel so special and giving me so many moments of joy, my grandfather for showing me the path, and my mum for helping me to follow it. Sure, I also thank my father, my sisters, my grandmother, my great family and all the friends I have. Thanks all of you for being part of my life.

Publications I, II and III were part of the ÄKK project, funded by Tekes (current Business Finland), the Ministry of Environment, and nine companies of the building industry. Publication IV was part of Anojanssi research project, which was funded by Tekes, Turku University of Applied Sciences, Ministry of Environment, Ministry of Social Affairs and Health, and eight private companies.

Turku, August 2022
David Oliva Elorza

David Oliva Elorza

References

Aletta, F., Oberman, T., and Kang, J. 2018. Associations between positive health-related effects and soundscapes perceptual constructs: A systematic review. International journal of environmental research and public health, 15(11), 2392.

Alonso, A., Patricio, J., Suárez, R. and Escandón, R., 2020. Acoustical retrofit of existing residential buildings: Requirements and recommendations for sound insulation between dwellings in Europe and other countries worldwide. Building and Environment, 174, p.106771.

Anderson, C. M. B., 1971. The measurement of attitude to noise and noises. GB National Physical Laboratory.

Angerer, J.R., McCurdy, D.A. and Erickson, R.A., 1991. Development of an annoyance model based upon elementary auditory sensations for steady-state aircraft interior noise containing tonal components.

ANSI S1.13, 2005 Measurement of Sound Pressure Levels in Air. Acoustical Society of America, New York.

ANSI S3.4, 2007 Procedure for the Computation of Loudness Steady Sounds, Acoustical Society of America, New York.

ANSI S12.2, 2008 Criteria for Evaluating Room Noise, Acoustical Society of America, New York.

ASTM E413-10, 2010. Classification for Rating Sound Insulation. ASTM International: West Conshohocken, PA, USA.

Ayr, U., Cirillo, E., Fato, I. and Martellotta, F., 2003. A new approach to assessing the performance of noise indices in buildings. Applied acoustics, 64(2), pp.129-145.

Bengtsson, J., Waye, K.P. and Kjellberg, A., 2004. Evaluations of effects due to low-frequency noise in a low demanding work situation. Journal of sound and vibration, 278(1-2), pp.83-99.

Berglund, B. and Lindvall, T. eds., 1995. Community noise. Stockholm: Center for Sensory Research, Stockholm University and Karolinska Institute.

Bolin, K., Bluhm, G. and Nilsson, M.E., 2014. Listening test comparing A-weighted and C-weighted sound pressure level as indicator of wind turbine noise annoyance. Acta Acustica united with Acustica, 100(5), pp.842-847.

Bolin, K., Bluhm, G., Eriksson, G. and Nilsson, M.E., 2011. Infrasound and low frequency noise from wind turbines: exposure and health effects. Environmental research letters, 6(3), p.035103.

Brink, M., Schäffer, B., Vienneau, D., Foraster, M., Pieren, R., Eze, I.C., Cajochen, C., Probst-Hensch, N., Röösli, M. and Wunderli, J.M., 2019. A survey on exposure-response relationships for road, rail, and aircraft noise annoyance: Differences between continuous and intermittent noise. Environment international, 125, pp.277-290.

Brüel & Kjær 2013. Product data. Sound quality head and torso simulator types 4100 and 4100-D. https://www.bksv.com/-/media/literature/Product-Data/bp1436.ashx (last accessed 2.11.2020).

Brüel & Kjaer. Pulse Sound quality software type 7698. https://www.bksv.com/media/doc/bp1589.pdf (last accessed 2.11.2020).

Cavanna, A.E., 2014. What is misophonia and how can we treat it?, Expert Review of Neurotherapeutics, 14:4, 357-359.

Cohen, S., & Weinstein, N., 1981. Nonauditory effects of noise on behavior and health. Journal of social issues, 37(1), 36-70.

Crook, M. A., & Langdon, F. J., 1974. The effects of aircraft noise in schools around London Airport. Journal of Sound and Vibration, 34(2), 221-232.

Danielsson, C.B. and Bodin, L., 2009. Difference in satisfaction with office environment among employees in different office types. Journal of Architectural and Planning Research, pp.241-257.

Di, G.Q., Chen, X.W., Song, K., Zhou, B. and Pei, C.M., 2016. Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises. Applied acoustics, 105, pp.164-170.

Di, G.Q., Zhou, X.X. and Chen, X.W., 2015. Annoyance response to low frequency noise with tonal components: A case study on transformer noise. Applied Acoustics, 91, pp.40-46.

Di Blasio, S., Shtrepi, L., Puglisi, G.E. and Astolfi, A., 2019. A cross-sectional survey on the impact of irrelevant speech noise on annoyance, mental health and well-being, performance and occupants' behavior in shared and open-plan offices. International journal of environmental research and public health, 16(2), p.280.

DIN 45681:2005. Acoustics – Determination of tonal components of noise and determination of a tone adjustment for the assessment of noise immissions. German Institute for Standardization, Berlin, Germany.

Ebissou, A., Parizet, E. and Chevret, P., 2015. Use of the speech transmission index for the assessment of sound annoyance in open-plan offices. Applied Acoustics, 88, pp.90-95.

ECMA 74:2015. Measurement of airborne noise emitted by information technology and telecommunications equipment, 13th edition, June 2015, ECMA International, Geneva, Switzerland.

Ellermeier, W., Zeitler, A. and Fastl, H., 2004. Predicting annoyance judgments from psychoacoustic metrics: identificable versus neutralized sounds. In Proc. of 33rd Intern. Congress on Noise Control Engineering INTER-NOISE 2004, Prague, Czech Republic.

European Commission, 2002, European Noise 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

European Parliament 2020, Policy department D, Budgetary affairs. Open spaces at EU institutions versus traditional work spaces: justification, evolution, evaluation and results. https://www.europarl.europa.eu/cmsdata/213064/20201015_Brochure_Open%20spaces.pdf (Last accessed 16.3.2021).

Fastl, H., 2006. Psychoacoustic basis of sound quality evaluation and sound engineering. In Proc. 13th Intern. Congress on Sound and Vibration ICSV13, Vienna, Austria.

Fletcher, H. and Munson, W.A., 1933. Loudness, its definition, measurement and calculation. Bell System Technical Journal, 12(4), pp.377-430.

Furihata, K., Yanagisawa, T., Asano, D.K. and Yamamoto, K., 2007. Development of an experimental noise annoyance meter. Acta acustica united with acustica, 93(1), pp.73-83.

Glasberg, B. R., Moore, B. C. J. Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. J. Acoust. Soc. Am. 2006, 120 pp. 585–588

Guski, R., 1999. Personal and social variables as co-determinants of noise annoyance. Noise and health, 1(3), p.45.

Guski, R., Schreckenberg, D. and Schuemer, R., 2017. WHO environmental noise guidelines for the European region: a systematic review on environmental noise and annoyance. International journal of environmental research and public health, 14(12), p.1539.

Haapakangas, A., Helenius, R., Keskinen, E. and Hongisto, V., 2008, July. Perceived acoustic environment, work performance and well-being–survey results from Finnish offices. In 9th International congress on noise as a public health problem (ICBEN) (Vol. 18, No. 8, pp. 21-25).

Haapakangas, A., Kankkunen, E., Hongisto, V., Virjonen, P., Oliva, D. and Keskinen, E., 2011. Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices. Acta Acustica united with Acustica, 97(4), pp.641-655.

Haapakangas, A., Hongisto, V., Eerola, M. and Kuusisto, T., 2017. Distraction distance and perceived disturbance by noise—An analysis of 21 open-plan offices. The Journal of the Acoustical Society of America, 141(1), pp.127-136.

Haapakangas, A., Hongisto, V. and Oliva, D., 2020. Audio-visual interaction in perception of industrial plants–Effects of sound level and the degree of visual masking by vegetation. Applied Acoustics, 160, p.107121.

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E. and Hongisto, V., 2009. Performance effects and subjective disturbance of speech in acoustically different office types–a laboratory experiment. Indoor air, 19(6), pp.454-467.

Hansen, H., Verhey, J.L. and Weber, R., 2011. The magnitude of tonal content. A review. Acta Acustica united with Acustica, 97(3), pp.355-363.

Hansen, K., Nguyen, P., Lechat, B., Zajamsek, B., Alamir, M., Micic, G., Catch-eside, P., 2019. Comparison between perceptual responses to traffic noise and wind farm noise in a non-focused listening test. In Proceedings of ACOUSTICS 2019.

Hongisto, V., Makila, M., Haapakangas, A., Hakala, J., Hyona, J. and Kylianen, M., 2013, September. Acoustic satisfaction in multi-storey buildings built after 1980-Preliminary results of a field survey. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 247, No. 4, pp. 4125-4132). Institute of Noise Control Engineering.

Hongisto, V., Oliva, D. and Keränen, J., 2014. Subjective and objective rating of airborne sound insulation–living sounds. Acta Acustica united with Acustica, 100(5), pp.848-863.

Hongisto, V., Oliva, D. and Rekola, L., 2015. Subjective and objective rating of spectrally different pseudorandom noises—Implications for speech masking design. The Journal of the Acoustical Society of America, 137(3), pp.1344-1355.

Hongisto, V., Mäkilä, M. and Suokas, M., 2015. Satisfaction with sound insulation in residential dwellings–The effect of wall construction. Building and Environment, 85, pp.309-320.

Hongisto, V., Varjo, J., Leppämäki, H., Oliva, D. and Hyönä, J., 2016. Work performance in private office rooms: The effects of sound insulation and sound masking. Building and Environment, 104, pp.263-274.

Hongisto, V., Varjo, J., Oliva, D., Haapakangas, A. and Benway, E., 2017. Perception of water-based masking sounds—long-Term experiment in an open-plan office. Frontiers in psychology, 8, p.1177.

Hongisto, V., Oliva, D. and Keränen, J., 2017. Indoor noise annoyance due to 3–5 megawatt wind turbines—An exposure–response relationship. The Journal of the Acoustical Society of America, 142(4), pp.2185-2196.

Hongisto, V., Oliva, D. and Rekola, L., 2018. Subjective and objective rating of the sound insulation of residential building façades against road traffic noise. The Journal of the Acoustical Society of America, 144(2), pp.1100-1112.

Hongisto, V., Saarinen, P. and Oliva, D., 2019. Annoyance of low-level tonal sounds–A penalty model. Applied Acoustics, 145, pp.358-361.

Hoofnagle, C.J., van der Sloot, B. and Borgesius, F.Z., 2019. The European Union general data protection regulation: what it is and what it means. Information & Communications Technology Law, 28(1), pp.65-98.

Horvat, M., Domitrović, H. and Jambrošić, K., 2012, January. Suitability of 3D sound reproduction and the influence of background noise on subjective assessment of sound insulation. In Euronoise 2012.

Huang, Y., Di, G., Zhu, Y., Hong, Y. and Zhang, B., 2008. Pair-wise comparison experiment on subjective annoyance rating of noise samples with different frequency spectrums but same A-weighted level. Applied Acoustics, 69(12), pp.1205-1211.

Hume, K.I., Brink, M. and Basner, M., 2012. Effects of environmental noise on sleep. Noise and health, 14(61), p.297.

Hünerbein von, S., King, A., Hargreaves, J., Moorhouse, A. and Plack, C., 2010. Perception of Noise from Large Wind Turbines. University of Salford, Acoustics Research Centre UK University of Manchester, Division of Human Communication and Deafness, UK.

Hurtley, C. ed., 2009. Night noise guidelines for Europe. WHO Regional Office Europe.

IEC 61672-1, 2013. Electroacoustics - Sound level meters - Part 1: Specifications, International Electrotechnical Commission.

Ioannidou, C., Santurette, S. and Jeong, C.H., 2016. Effect of modulation depth, frequency, and intermittence on wind turbine noise annoyance. The Journal of the Acoustical Society of America, 139(3), pp.1241-1251.

ISO 226:2003. Acoustics–Normal equal-loudness-level contours. International Organization for Standardization, Geneva, Switzerland.

ISO 532:1975. Acoustics – Method for calculating loudness level, International Organization for Standardization, Geneva, Switzerland

ISO 532-1:2017. Acoustics — Methods for calculating loudness — Part 1: Zwicker method, International Organization for Standardization, Geneva, Switzerland.

ISO 532-2:2017. Acoustics — Methods for calculating loudness — Part 2: Moore-Glasberg method, International Organization for Standardization, Geneva, Switzerland.

ISO 717-1:2013. Acoustics – Rating of sound insulation in buildings and of building elements – Part 1: Airborne sound insulation, International Organization for Standardization, Geneva, Switzerland.

ISO 1996-2:2007. Acoustics – Description, measurement and assessment of environmental noise -- Part 2: Determination of environmental noise levels. Annex C. International Organization for Standardization, Geneve, Switzerland.

ISO 3740:2019. Acoustics — Determination of sound power levels of noise sources — Guidelines for the use of basic standards. International Organization for Standardization, Geneva, Switzerland.

ISO 9612:2009. Acoustics — Determination of occupational noise exposure — Engineering method. International Organization for Standardization, Geneva, Switzerland.

ISO 15666:2003. Acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization, Geneva, Switzerland.

ISO 16283-1:2014. Acoustics — Field measurement of sound insulation in buildings and of building elements — Part 1: Airborne sound insulation. International Organization for Standardization, Geneva, Switzerland.

Ishiyama, T. and Hashimoto, T., 2000. The impact of sound quality on annoyance caused by road traffic noise: an influence of frequency spectra on annoyance. Jsae Review, 21(2), pp.225-230.

ITU-T, P.58, 2011. Head and torso simulator for telephonometry, International Telecommunication Union, Geneva, Switzerland.

Janssen, S.A., Vos, H., Eisses, A.R. and Pedersen, E., 2011. A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources. The Journal of the Acoustical Society of America, 130(6), pp.3746-3753.

Jensen, K. L., Arens, E., and Zagreus, L,. 2005. Acoustical quality in office workstations as assessed by occupant surveys," in Proceedings of Indoor Air, Tenth International Conference on Indoor Air Quality and Climate, September 4–9, Beijing, China, pp. 2401–2405.

Jeon, J.Y., You, J., Jeong, C.I., Kim, S.Y. and Jho, M.J., 2011. Varying the spectral envelope of air-conditioning sounds to enhance indoor acoustic comfort. Building and Environment, 46(3), pp.739-746.

Job, R. S., 1996. The influence of subjective reactions to noise on health effects of the noise. Environment international, 22(1), 93-104.

Johansson, O., 2000, August. Relationship between psychoacoustic descriptors and annoyance: Regarding sound in home environments. In Proc. Inter-Noise (p. 1669).

Jones, D. M., Chapman, A. J., & Auburn, T. C., 1981. Noise in the environment: a social perspective. Journal of Environmental Psychology, 1(1), 43-59.

Kaarlela-Tuomaala, A., Helenius, R., Keskinen, E. and Hongisto, V., 2009. Effects of acoustic environment on work in private office rooms and open-plan offices–longitudinal study during relocation. Ergonomics, 52(11), pp.1423-1444.

Keighley, E.C., 1970. Acceptability criteria for noise in large offices. Journal of Sound and Vibration, 11(1), pp.83-93.

Kim, J., Lim, C., Hong, J. and Lee, S., 2010. Noise-induced annoyance from transportation noise: Short-term responses to a single noise source in a laboratory. The Journal of the Acoustical Society of America, 127(2), pp.804-814.

Kim, R. and Van den Berg, M., 2010. Summary of night noise guidelines for Europe. Noise and health, 12(47), p.61.

Kjellberg, A. and Goldstein, M., 1985. Loudness assessment of band noise of varying bandwidth and spectral shape. An evaluation of various frequency weighting networks. Journal of Low Frequency Noise, Vibration and Active Control, 4(1), pp.12-26.

Knipschild, P., 1977. V. Medical effects of aircraft noise: community cardiovascular survey. International Archives of Occupational and Environmental Health, 40(3), 185-190.

Kryter, K.D. and Pearsons, K.S., 1965. Judged noisiness of a band of random noise containing an audible pure tone. The Journal of the Acoustical Society of America, 38(1), pp.106-112.

Kuttruff, H., 2016. Room acoustics. Crc Press.

Kylliäinen, M., Takala, J., Oliva, D. and Hongisto, V., 2016. Justification of standardized level differences in rating of airborne sound insulation between dwellings. Applied Acoustics, 102, pp.12-18.

Kylliäinen, M., Hongisto, V., Oliva, D. and Rekola, L., 2017. Subjective and objective rating of impact sound insulation of a concrete floor with various coverings. Acta Acustica united with Acustica, 103(2), pp.236-251.

Landström, U., Kjellberg, A., Söderberg, L. and Nordström, B., 1994. Measures against ventilation noise–which tone frequencies are least and most annoying?. Journal of Low Frequency Noise, Vibration and Active Control, 13(3), pp.81-88.

Landström, U., Åkerlund, E., Kjellberg, A. and Tesarz, M., 1995. Exposure levels, tonal components, and noise annoyance in working environments. Environment International, 21(3), pp.265-275.

Lee, S., Kim, K., Choi, W. and Lee, S., 2011. Annoyance caused by amplitude modulation of wind turbine noise. Noise Control Engineering Journal, 59(1), pp.38-46.

Lee, J., Francis, J.M. and Wang, L.M., 2017. How tonality and loudness of noise relate to annoyance and task performance. Noise Control Engineering Journal, 65(2), pp.71-82.

Lee, J. and Wang, L.M., 2020. Investigating multidimensional characteristics of noise signals with tones from building mechanical systems and their effects on annoyance. The Journal of the Acoustical Society of America, 147(1), pp.108-124.

Legarth, S.V., 2007. Auralization and assessments of annoyance from wind turbines. Proc. 2nd Int. Meeeting on Wind Turbine Noise, Lyon.

Lenne, L., Chevret, P. and Marchand, J., 2020. Long-term effects of the use of a sound masking system in open-plan offices: A field study. Applied Acoustics, 158, p.107049.

Lercher, P., Hörtnagl, J., & Kofler, W. W. (1993). Work noise annoyance and blood pressure: combined effects with stressful working conditions. International archives of occupational and environmental health, 65(1), 23-28.

Leventhall, G., 2006. Infrasound from wind turbines-fact, fiction or deception. Canadian acoustics, 34(2), pp.29-36.

Little, J.W. and Mabry, J.E., 1969. Sound duration and its effect on judged annoyance. Journal of Sound and Vibration, 9(2), pp.247-262.

Maijala, P., Turunen, A., Kurki, I., Vainio, L., Pakarinen, S., Kaukinen, C., Lukander, K., Tiittanen, P., Yli-Tuomi, T., Taimisto, P., et al. Infrasound Does Not Explain Symptoms Related to Wind Turbines; Report of the Prime Minister’s Office: Helsinki, Finland, 2020.

Maijala, P., Kurki, I., Vainio, L., Pakarinen, S., Kuuramo, C., Lukander, K., Tiippana, K., Virkkala, J., Stickler, E. and Sainio, M. Annoyance, perception, and physiological effects of wind turbine infrasound. J. Acoust. Soc. Am. 2021, 149, 2238–2248.

Manninen, O., & Aro, S., 1979. Noise-induced hearing loss and blood pressure. International Archives of Occupational and Environmental Health, 42(3), 251-256.

Manyoky, M., Wissen Hayek, U., Heutschi, K., Pieren, R. and Grêt-Regamey, A., 2014. Developing a GIS-based visual-acoustic 3D simulation for wind farm assessment. ISPRS International Journal of Geo-Information, 3(1), pp.29-48.

McMinn, T., 2013. A-weighting: Is it the metric you think it is?. Acoustics 2013, pp.1-4.

MSAH. Ministry of Social Affairs and Health, 2015, Decree 545-2015 of the Ministry of Social Affairs and Health on Health-related Conditions of Housing and Other Residential Buildings and Qualification Requirements for Third-party Experts, 2015. Translation from Finnish, http://www.finlex.fi/en/laki/kaannokset/2015/en20150545.pdf

ME. Ministry of the Environment, 2017. Decree of the Ministry of the Environment the sound environment of the building (796/2017), 24 November, Helsinki, Finland.

Monteiro, C., Machimbarrena, M., de la Prida, D. and Rychtarikova, M., 2016. Subjective and objective acoustic performance ranking of heavy and light weight walls. Applied Acoustics, 110, pp.268-279.

More, S. and Davies, P., 2010. Human responses to the tonalness of aircraft noise. Noise Control Engineering Journal, 58(4), pp.420-440.

Muzet, A., & Ehrhart, J. (1980). Habituation of heart rate and finger pulse responses to noise in sleep. Noise as a Public Health Problem. Rockville, Maryland: ASHA Reports, 10, 401-404.

Myllyntausta, S., Virkkala, J., Salo, P., Varjo, J., Rekola, L., & Hongisto, V., 2020. Effect of the frequency spectrum of road traffic noise on sleep: A polysomnographic study. The Journal of the Acoustical Society of America, 147(4), 2139-2149.

Møller, H. and Andresen, J., 1984. Loudness of pure tones at low and infrasonic frequencies. Journal of Low Frequency Noise, Vibration and Active Control, 3(2), pp.78-87.

Møller, H. and Pedersen, C.S., 2011. Low-frequency noise from large wind turbines. The Journal of the Acoustical Society of America, 129(6), pp.3727-3744.

National Advisory Board on Research Ethics. (2009). “Ethical principles of research in the humanities and social and behavioral sciences and proposals for ethical review,” Helsinki, Finland.

Navai, M. and Veitch, J.A., 2003. Acoustic satisfaction in open-plan offices: review and recommendations. Institute for Research in Construction, Research Report RR-151.

Neidhardt, A., Liebal, J. and Paasonen, J., 2017. Human echolocation in virtual acoustic environments: Estimating the direction of a close wall. In 4th Int. Conf. on Spatial Audio (ICSA), Graz, Austria.

Nilsson, M.E., 2007. A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound. Journal of Sound and Vibration, 302(1-2), pp.197-207.

Oliva D., Hongisto V., Keränen J., Koskinen V., Measurement of low frequency noise in rooms - Method for constant and intermittent noise, Sisäympäristölaboratorio raporttisarja 2, Työterveyslaitos, Helsinki 2010.

Oliva, D., 2012. Questionnaire about Low Frequency Noise measurements in rooms. Journal of Low Frequency Noise, Vibration and Active Control, 31(1), pp.13-20.

Oliva, D., Hongisto, V. and Haapakangas, A., 2017. Annoyance of low-level tonal sounds–Factors affecting the penalty. Building and Environment, 123, pp.404-414.

Onakpoya, I.J., O'Sullivan, J., Thompson, M.J. and Heneghan, C.J., 2015. The effect of wind turbine noise on sleep and quality of life: A systematic review and meta-analysis of observational studies. Environment international, 82, pp.1-9.

Ordoñez, R., Visentin, C., Markovic, M. and Fausti, P., 2013. Objective and subjective evaluation of façade sound insulation. In 42nd International Congress and Exposition on Noise Control Engineering 2013, INTER-NOISE 2013: Noise Control for Quality of Life (Vol. 2, pp. 1523-1532). OAL-Osterreichischer Arbeitsring fur Larmbekampfung.

Ormel, J., Riese, H., & Rosmalen, J. G., 2012. Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect?. Clinical psychology review, 32(1), 71-79.

Park, H.K., Bradley, J.S. and Gover, B.N., 2008. Evaluating airborne sound insulation in terms of speech intelligibility. The Journal of the Acoustical Society of America, 123(3), pp.1458-1471.

Park, H.K. and Bradley, J.S., 2009. Evaluating standard airborne sound insulation measures in terms of annoyance, loudness, and audibility ratings. The Journal of the Acoustical Society of America, 126(1), pp.208-219.

Park, H.K. and Bradley, J.S., 2009. Evaluating signal-to-noise ratios, loudness, and related measures as indicators of airborne sound insulation. The Journal of the Acoustical Society of America, 126(3), pp.1219-1230.

Park, S.H., Lee, P.J., Lee, B.K., Roskams, M. and Haynes, B.P., 2020. Associations between job satisfaction, job characteristics, and acoustic environment in open-plan offices. Applied Acoustics, 168, p.107425.

Parrot, J., Petiot, J. C., Lobreau, J. P., & Smolik, H. J., 1992. Cardiovascular effects of impulse noise, road traffic noise, and intermittent pink noise at LAeq= 75 dB, as a function of sex, age, and level of anxiety: a comparative study. International archives of occupational and environmental health, 63(7), 477-484.

Patsouras, C., Fastl, H., Widmann, U. and Hölzl, G., 2002. Psychoacoustic evaluation of tonal components in view of sound quality design for high-speed train interior noise. Acoustical Science and Technology, 23(2), pp.113-116.

Pawłaczyk-Luszczyńska, M., Dudarewicz, A., Waszkowska, M. and Sliwińska-Kowalska, M., 2003. Assessment of annoyance from low frequency and broadband noises. International journal of occupational medicine and environmental health, 16(4), pp.337-343.

Pedersen, E. and Larsman, P., 2008. The impact of visual factors on noise annoyance among people living in the vicinity of wind turbines. Journal of Environmental Psychology, 28(4), pp.379-389.

Pedersen, T.H., Antunes, S. and Rasmussen, B., 2012. Online listening tests on sound insulation of walls–A feasibility study. In Proceedings of Euronoise.

Persson, K., Björkman, M. and Rylander, R., 1990. Loudness, annoyance and dBA in evaluating low frequency sounds. Journal of Low Frequency Noise, Vibration and Active Control, 9(1), pp.32-45.

Poulsen, T., 1991. Influence of session length on judged annoyance. Journal of Sound and Vibration, 145(2), pp.217-224.

Poulsen, T., 2003. Comparison of objective methods for assessment of annoyance of low frequency noise with the results of a laboratory listening test. Journal of low frequency noise, vibration and active control, 22(3), pp.117-131.

Pouru, L., 2019. The Parameters of Realistic Spatial Audio: An Experiment with Directivity and Immersion. Turku University of Applied Sciences

de la Prida, D., Pedrero, A., Navacerrada, M.Á. and Díaz-Chyla, A., 2020. An annoyance-related SNQ for the assessment of airborne sound insulation for urban-type sounds. Applied Acoustics, 168, p.107432.

Rasmussen, B., 2010, December. Harmonization of sound insulation descriptors and classification schemes in Europe: COST Action TU0901. In Proceedings of European Symposium on Harmonisation of European Sound Insulation Descriptors and Classification Standards.

Rasmussen, B.,2019. Sound insulation between dwellings–Comparison of national requirements in Europe and interaction with acoustic classification schemes. In 23rd International Congress on Acoustics (pp. 5102-5109). Deutsche Gesellschaft für Akustik (DEGA eV).

Rychtáriková, M., Mülner, H., Stani, M., Chmelík, V. and Glorieux, C., 2012. Does the living noise spectrum adaptation of sound insulation match the subjective perception?. In Proceedings-European Conference on Noise Control (pp. 180-184).

Ryherd, E.E. and Wang, L.M., 2008. Implications of human performance and perception under tonal noise conditions on indoor noise criteria. The Journal of the Acoustical Society of America, 124(1), pp.218-226.
Sakamoto, S., Yokoyama, S., Tsujimura, S. and Tachibana, H., 2013, September. Loudness evaluation of general environmental noise containing low frequency components. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 247, No. 5, pp. 3234-3240). Institute of Noise Control Engineering.
Salame, P., & Baddeley, A., 1982. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. Journal of verbal learning and verbal behavior, 21(2), 150-164.
Schreckenberg, D., Griefahn, B., & Meis, M., 2010. The associations between noise sensitivity, reported physical and mental health, perceived environmental quality, and noise annoyance. Noise and health, 12(46), 7.
Schwarz, D., Lemaitre, G., Aramaki, M. and Kronland-Martinet, R., 2016, Effects of Test Duration in Subjective Listening Tests. International Computer Music Conference (ICMC), Hans Timmermans, Sep 2016, Utrecht, Netherlands. pp.515-519.
Schäffer, B., Schlittmeier, S.J., Pieren, R., Heutschi, K., Brink, M., Graf, R. and Hellbrück, J., 2016. Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: A laboratory study. The Journal of the Acoustical Society of America, 139(5), pp.2949-2963.
Schäffer, B., Pieren, R., Schlittmeier, S.J. and Brink, M., 2018. Effects of different spectral shapes and amplitude modulation of broadband noise on annoyance reactions in a controlled listening experiment. International journal of environmental research and public health, 15(5), p.1029.
Śliwińska-Kowalska, M. and Zaborowski, K., 2017. WHO environmental noise guidelines for the European region: a systematic review on environmental noise and permanent hearing loss and tinnitus. International journal of environmental research and public health, 14(10), p.1139.
Smedley, E., 1845. Encyclopaedia Metropolitana; or, universal dictionary of knowledge, on an original plan: comprising the twofold advantage of a philosophical and an alphabetical arrangement, with appropriate engravings: Edited by Edw. Smedley, Hugh Jam. Rose, and H. John Rose.
Stansfeld, S. A. (1992). Noise, noise sensitivity and psychiatric disorder: epidemiological and psychophysiological studies. Psychological medicine monograph supplement, 22, 1-44.
Stansfeld, S., Haines, M., & Brown, B., 2000. Noise and health in the urban environment. Reviews on environmental health, 15(1-2), 43-82.
Steiger, J.H., 1980. Tests for comparing elements of a correlation matrix. Psychological bulletin, 87(2), p.245.
Sulaiman, A.H., Seluakumaran, K. and Husain, R., 2013. Hearing risk associated with the usage of personal listening devices among urban high school students in Malaysia. public health, 127(8), pp.710-715.
Sundstrom, E., Town, J. P., Rice, R. W., Osborn, D. P., & Brill, M. 1994. Office noise, satisfaction, and performance. Environment and behavior, 26(2), pp.195-222.
Szalma, J. L., & Hancock, P. A., 2011. Noise effects on human performance: a meta-analytic synthesis. Psychological bulletin, 137(4), 682.
Søndergaard, L.S. and Pedersen, T.H., 2013. Tonality in wind turbine noise. IEC 61400-11 ver. 2.1 and 3.0 and the Danish/Joint Nordic method compared with listening tests. Wind Turbine Noise 2013.
Tachibana, H., Yano, H. and Sonoda, Y., 1990. Subjective assessment of indoor noises—Basic experiments with artificial sounds. Applied Acoustics, 31(1-3), pp.173-184.
Tachibana, H., Sonoda, Y., Iwamoto, K., Kuwano, S. and Namba, S., 1993. Validity of arithmetic average of sound pressure levels in octave bands as a loudness index. Journal of the Acoustical Society of Japan (E), 14(3), pp.197-204.
Tang, S.K., Burnett, J. and Poon, C.M., 1996. Aural environment survey in air-conditioned open-plan offices. Building Services Engineering Research and Technology, 17(2), pp.97-100.

Themann, C.L. and Masterson, E.A., 2019. Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. The Journal of the Acoustical Society of America, 146(5), pp.3879-3905.
Torija, A.J. and Flindell, I.H., 2014. Differences in subjective loudness and annoyance depending on the road traffic noise spectrum. The Journal of the Acoustical Society of America, 135(1), pp.1-4.
Torija, A.J. and Flindell, I.H., 2015. The subjective effect of low frequency content in road traffic noise. The Journal of the Acoustical Society of America, 137(1), pp.189-198.
Van Dijk, F. J., 1986. Non-auditory effects of noise in industry. II. A review of the literature. International archives of occupational and environmental health, 58(4), 325-332.
Veitch, J.A., Bradley, J.S., Legault, L.M., Norcross, S. and Svec, J.M., 2002. Masking speech in open-plan offices with simulated ventilation noise: noise level and spectral composition effects on acoustic satisfaction. Institute for Research in Construction, Internal Report IRC-IR-846.
Versfeld, N.J. and Vos, J., 2002. A-weighted equivalent sound level as a predictor of the annoyance caused by road traffic consisting of various proportions of light and heavy vehicles. Journal of sound and vibration, 253(2), pp.389-399.
Viollon, S., Lavandier, C. and Drake, C., 2002. Influence of visual setting on sound ratings in an urban environment. Applied acoustics, 63(5), pp.493-511.
Virjonen, P., Hongisto, V. and Oliva, D., 2016. Optimized single-number quantity for rating the airborne sound insulation of constructions: Living sounds. The Journal of the Acoustical Society of America, 140(6), pp.4428-4436.
Virjonen, P., Hongisto, V. and Radun, J., 2019. Annoyance penalty of periodically amplitude-modulated wide-band sound. The Journal of the Acoustical Society of America, 146(6), pp.4159-4170.
Waye, K.P., Rylander, R., Benton, S. and Leventhall, H.G., 1997. Effects on performance and work quality due to low frequency ventilation noise. Journal of sound and vibration, 205(4), pp.467-474.
Weinstein, N.D., 1978. Individual differences in reactions to noise: a longitudinal study in a college dormitory. Journal of applied psychology, 63(4), p.458.
Whittle, L.S., Collins, S.J. and Robinson, D.W., 1972. The audibility of low-frequency sounds. Journal of sound and vibration, 21(4), pp.431-448.
WHO, 2018. Environmental Noise Guidelines for the European Region. World Health Organization. Regional Office for Europe, Copenhagen.
Wing, J. K., Cooper, J. E., & Sartorius, N., 2012. Measurement and classification of psychiatric symptoms: An instruction manual for the PSE and CATEGO program. Cambridge University Press.
Wu, H., Sun, X. and Wu, Y., 2020. Investigation of the relationships between thermal, acoustic, illuminous environments and human perceptions. Journal of Building Engineering, 32, p.101839.
Yadav, M., Cabrera, D., Kim, J., Fels, J., & de Dear, R., 2021. Sound in occupied open-plan offices: Objective metrics with a review of historical perspectives. Applied Acoustics, 177, 107943.
Yokoyama, S. and Tachibana, H., 2016, August. Perception of tonal components contained in wind turbine noise. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 253, No. 6, pp. 2435-2446). Institute of Noise Control Engineering.
Zimmer, K., Ghani, J. and Ellermeier, W., 2008. The role of task interference and exposure duration in judging noise annoyance. Journal of sound and vibration, 311(3-5), pp.1039-1051.
Zwicker, E. and Fastl, H., 2013. Psychoacoustics: Facts and models (Vol. 22). Springer Science & Business Media.
Zwicker E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. 1961, 33 (2) p. 248
Zwicker, E., 1977. Procedure for calculating loudness of temporally variable sounds. The Journal of the Acoustical Society of America, 62(3), pp.675-682.
Öhrström, E., & Rylander, R., 1982. Sleep disturbance effects of traffic noise—a laboratory study on after effects. Journal of Sound and Vibration, 84(1), 87-103.
Öhrström, E., 1989. Sleep disturbance, psycho-social and medical symptoms—a pilot survey among persons exposed to high levels of road traffic noise. Journal of Sound and Vibration, 133(1), 117-128.

Appendix 1. Psychoacoustic experiments.

A good starting point to design a psychoacoustics experiment is NT ACOU 111 (Nordtest 111:2002). Beyond that, there are no other standards providing orientation about how to perform listening experiments like those conducted during this research. This issue has been mentioned before (Genuit 2010). The author has participated in nearly 20 laboratory experiments, and the experience and knowhow are shared in this Appendix. The aims are to support the validity of the previously presented results, and to complement somehow this work with a critical discussion of the methods we applied. The discussion may be useful in the preparation of future psychoacoustic listening tests.

Differences between laboratory experiments and field surveys

Two methods enable the study of subjective noise perception: socio-acoustic field surveys (ISO 15666:2003) and listening experiments. In field surveys, the respondents are normally asked to rate and comment the noise condition they experience during a defined period. In laboratory experiments, the participants are asked to rate sets of sound stimuli that might not represent previously experienced noise conditions. Each method has its pros and cons. Field surveys are useful when the research relates to the noises that a certain population group experience during their lives. They enable to investigate the responses of many persons in a cost-effective way. The main drawback relates to the verification of the physical characteristics of the noise(s). For instance, it is possible to obtain the opinion of a thousand persons with respect to the noise coming from a certain road or airport, but it is hard to verify what the respondents do actually experience in their homes, because each house is at a different distance from the noise source and has different sound insulation properties. Nassur et al. (2019) solved this challenge recently. They investigated the effects of exposure to aircraft noise on heart rate during sleep in populations living near airports. The study performed sound pressure level measurements in front of the façade of the participants' dwellings and inside their bedrooms, while simultaneously their heart rate was measured. Similar approaches were applied by Landström et al. (1995) and by Hongisto et al. (2005) to study noise annoyance. Landström studied the correlation between the noise measured in 439 working places and the annoyance rating provided by the 439 workers who worked in each of these places. Hongisto analyzed the responses of 159 habitants of multi-story buildings and performed sound insulation and sound pressure level measurements for verification purposes in the dwellings.

Laboratory experiments and field surveys might provide different results. During field surveys, people are not necessarily exposed to the stimuli under investigation while answering to the questionnaire. Field surveys following ISO 15666 (ISO 15666:2003) ask the respondent to rate verbally and numerically the sound experience in defined time periods, for instance the last 12 months, and so the rating could be based on their memory, previous experiences, and other moderators building the opinion. This cannot be controlled, and thus it needs to be accepted in the results. The impact of moderators could be forecasted with a noise sensitivity test. In experiments of Publications II–IV we performed the Weinstein sensitivity test (Weinstein 1978). We did not notice any participant with anomalous responses and the data was not investigated any further.

In laboratory conditions the participant is normally instructed to visualize a certain situation, for instance, being at home or in the garden reading a book or having some rest. They should imagine the stimuli to be part of that visualization. To compensate the lack of connection to the real case, it is possible to provide the participants a visual cue to facilitate the process, but that is not the most common. Visual cues were used for instance in Legarth 2007 and in Di et al. 2016 in studies on annoyance from WTN and electrical elements, respectively. In Publications III and IV, during the communication phase, we provided a picture of a typical living room to the participants to facilitate the visualization of the fictional situation. The effect on subjective perception of visual cues was analyzed by Viollon et al. (2002) and Haapakangas et al. (2020). Haapakangas investigated how the visual background, i.e., the amount of vegetation hiding the factory producing the noise affected the subjects' rating.

Listening experiments in laboratory conditions provide yet an extra benefit. The opinion of subjects regarding the sound conditions are a priori impartial because they can concentrate on the actual listening experience and not on the previous background and opinion regarding the sound under study (Schäffer et al. 2016, Legarth 2007, Maijala et al. 2021). However, if the respondent has previous experience with the noise in question, it is recommended to evaluate the impact of her/his ratings in the results. Noise sensitivity tests are handy in these situations. By contrast, in field surveys it is expected that moderators and previous experience affect the opinions of residents. In a WTN, the results were affected by confounding factors, such as attitudes, sensitization, and existence of economic benefits or devaluation of property because of nearby wind turbines (Janssen et al. 2011). In the field study by Pedersen & Larsman (2008), residents close to wind turbines were more annoyed by wind turbine noise than people who did not have one near their own residence. Links between visual experience of wind turbine generators and self-reported negative health effects were published by Onakpoya et al. (2015).

In our opinion, it can be self-evident that listening experiments are not able to provide an absolute value to how annoying a sound is. However, they enable to investigate the relative differences between sets of noises, as we did, and that was very useful. Laboratory experiments are laborious to prepare and scale up, but once they are ready, they provide more flexibility and control over the sound stimuli and input parameters.

Exposure times and duration of test

Listening experiments enable the detailed study of several sounds at once. Research studies involving over 100 experimental sounds have been reported for instance by Angerer et al. 1991, Park & Bradley 2009, Bolin et al. 2014, Di et al. 2016, and Schwarz et al. 2016, with 240, 120, 135, 117 and 121 sounds, respectively. Such large number of test sounds typically require that the length of the stimuli be reduced to avoid fatigue of the test subjects. Poulsen (1991) presented to the test subjects noises of four different lengths, from 1 to 30 minutes, and found no differences in the rating of the annoyance with respect to the duration of the test. Schwarz et al. (2016) analyzed the change in rating behavior during a long-lasting listening experiment. They observed a small but statistically significant upwards tendency towards the end of the test. The duration of the experiment was not especially long, on average 36 minutes, but the number of the stimuli was large, 121, each lasting seven seconds. Perhaps the repetitiveness of the same rating task had there a higher impact than the total length of the research itself.

The length of the test experiment is the total time by the test subject in the laboratory space. In Publication I it was less than 45 minutes and in Publications II, III and IV it was increased to 75 minutes. The increment is due to the complexity of the later tests. Seventy-five minutes was our best balance between the number of sound stimuli that we needed and the length of the sample. It can be mentioned that subjects are not necessarily in a hurry when they perform the tests and take their own time to rate the sounds accordingly. In Publication IV, the minimum exposure duration of the sound stimuli was 9 seconds, but on average each participant listened to each stimulus for 15 seconds before the rating was given. In Publications I and IV, the sounds did not have any temporal variation, i.e., they were steady-state sounds, but in Publications II and III, the living and traffic sounds varied with time. The participants typically reported afterwards that it was easy to imagine how the test sounds would have been experienced in real offices or dwellings, even for longer time periods despite the short exposure length. Poulsen (1991) reported that annoyance ratings remained similar for exposure times of 1, 5, 15 and 30 minutes. Other listening experiments, for instance Little & Mabry (1969) and Nilsson (2007) have used similar exposure times as ours, and Schäffer et al. (2016) found an optimum length of the stimuli to be 20 seconds when studying 30 sounds. However, in Zimmer et al. (2008) the rating of sounds depended on the exposure time and the task at hand. Veitch et al. (2002) used 18 minutes exposure time before judging was enabled. This length is closer to typical sound exposure times, but long exposure times limits the number of sounds that can be studied in an experiment.

In our opinion, in experiments where the task is to compare sounds and not to look to the absolute ratings of noise, the use of shorter exposure times like the ones we used in our experiments, e.g., 15–25 seconds, seems justified. In dwellings and office spaces, the noise exposure times are usually longer, and the rating of the sound may change. The subjective grounds for the rating are different in field and in laboratory studies. However, the focus of our studies was to collect subjective ratings and compare those, but not to gather absolute ratings. If the exposure time were to have an effect, it would most probably not affect the correlation coefficients, just the absolute levels of the subjective ratings. We believe that longer exposure times than the ones used in these four experiments would not significantly change the results.

Laboratory

Psychoacoustic and acoustic laboratories need to be very silent to secure the quality and reliability of the experiment. The background noise level should be always lower than the level of the investigated test stimuli, and we recommend that this is verified for all frequencies in the audible range. The background noise level in our room was under 25 dB $L_{A,eq}$, and it was below the hearing threshold in few individual one-third octave bands. According to Kylliäinen et al. (2015), such low levels exist only in about 25% of Finnish living rooms. Perhaps such setting did not represent the situation experienced by most persons in their own living rooms. The benefit of having such low background levels was to ensure the relationship between the sound stimuli and the subjective ratings. In Publication III, the most efficient façade constructions reduced the level of traffic sounds to very low levels. If the background level would have been like the level of the stimuli, we would not be sure which one was actually rated by the subject. In Publication II, if our laboratory had 35 dB $L_{A,eq}$ as in Park et al. (2009), instead of the 23 dB we had, about 30% of the experimental sounds would have been inaudible and the results affected accordingly.

Reproduction of the sound stimuli with speakers

Our tests always included sounds from 50 Hz and above. Sounds containing such low frequency content cannot be reproduced with normal speakers, and therefore we added subwoofers in all experiments. This issue has been addressed for instance by Møller & Pedersen (2011) in WTN studies including sound energy in frequencies 20 to 50 Hz. Expanding our experiments to low frequencies forced us to consider and verify the existence of resonance modes. First, we avoided placing the test subjects in locations where wave theory calculations predicted room modes. Second, we performed exhaustive measurements in the area occupied by the test subjects to verify the spectra of the stimuli. Room modes are a collection of resonance effects that can appear in a room when the space is excited by an acoustic source. Resonance effects and standing waves usually appear when the stimuli contain a considerable amount of sound energy at low frequencies and when the laboratory room has dimensions comparable to the wavelength of the low frequency sound. Standing waves are responsible for changes up to 20 dB at low frequencies within different points in the same room, see for instance Oliva et al. 2010 and Oliva 2012. Sharp SPL changes occur mostly in the middle of any dimension of the room dimensions. If we imagine a room three meters high, it has the room mode 001 at 57 Hz. A test subject with the head at 1.5 meters' height would most probably listen nothing at around 57 Hz frequency. The speakers play the sound, but that disappears in certain areas. Analyzing the room modes before the experiment to select the optimal position for the listener is mandatory, as suggested by Poulsen (2002). In our experiments, and especially in Publication IV because of tonal sounds at 50 and 110 Hz, we verified the correctness of the laboratory setup with measurements.

Reproduction of the sound stimuli with headphones

Sometimes, the sound stimuli need to be played via headphones, as was done in Publication IV where a hybrid method to play the sound stimuli was applied. The test subjects heard the sounds played simultaneously via headphones, speakers, and a subwoofer. This arrangement was designed to ensure that resonance mode effects would have a minimum impact, and that low and high frequency tones were played with accuracy and confidence. The SPL was verified with a head and torso simulator Brüel & Kjær 4100 (Brüel & Kjær 2013). Head and torso simulators enable to measure the spectra arriving to the external auditory ear channel when the test subjects are wearing headphones. The simulator has a similar shape and weight than humans and includes two microphones located inside the ears immediately at the beginning of the auditory canal. Measurements with this dummy head require special consideration, and it is needed to follow precise calibration and measurement procedures (ITU 2011). It is important to notice that when using dummy heads, the size and shape of the head and ears affects the sound field. Some frequencies are boosted, and others are attenuated. These effects can be quantified by the so-called head related transfer function, HRTF. In other words, measuring with a dummy and with a single microphone produce different results, and the HRTF is the function characterizing that difference. The HRTF of our dummy head was calculated from the difference of SPL measured with a single microphone and with the dummy head in open field and diffuse field conditions. According to our measurements, the correction applied to our tonal sounds was 0 dB, 0 dB, 0.4 dB, 1.3 dB, 3.5 dB, at frequencies 50, 110, 290, 850 and 2100 Hz, respectively. Our experience is that it is worth to measure the HRTF and compare those results with the ones provided by the manufacturer of the device.

Selection, preparation, and verification of sound stimuli

The correct selection and preparation of the sound stimuli are of crucial importance for the test to succeed; likewise, it is particularly important to verify that the sounds experienced by the test subject match the target spectra. We recommend analyzing the sound pressure level, the spectra, and the information content. The information content, i.e., the type of sound being simulated, needs to be checked because the applied filters might have distorted the signal. Second, the spectra of the audio file and the spectra of the played sound at the listener position do rarely match. As explained above, the room resonances and the emission properties of the used speakers or headphones influence the final sound field, and extra filtering is typically needed. Sound pressure level verification measurements are mandatory to ensure that the sound stimuli and the target matches well. For physical reasons related to sound insulation, the background noise in laboratories might be higher at low frequencies than the target spectra of the sound stimuli. In these cases, the background disturbs the measurements, and it becomes difficult to verify what the spectra of the stimuli are. In certain cases, and to overcome this difficulty, we amplified the level of the stimuli sounds while doing the verification measurements. After that, we reduced the level at all frequency bands with a level limiter. This way we ensured that the spectra of the stimuli was correct along the whole frequency range of interest.

The process to verify the sound stimuli is nevertheless straightforward. Based on our experience we propose the following procedure. 1) Define the target spectra, i.e., the spectra of the sound that the subject should listen to. 2) Create a temporal audio file. 3) Play the signal through the speakers or the headphones. 4) Measure the sound pressure level and the spectra in several points around the area to be occupied by the head of the subject. 5) Calculate the difference between the target value and the measured values. 6) If the difference in any frequency band exceeds ± 1dB, for instance, apply a digital filter in the frequency band to correct this error. 7) If the difference in the overall sound pressure level exceeds ± 1dB, apply a digital filter to correct the discrepancy. 8) Repeat steps 3–7 until the error is smaller than ± 1dB at all frequency bands, and ± 0.5 dB with respect to the overall level. 9) Verify that the signal has not been distorted.

An additional fact to consider when designing sound stimuli is that the rating of the stimuli needs to answer the research question. In Publication II, where we studied annoyance of living sounds transmitted through wall partitions, the six selected sounds represented well the most annoying sounds created by neighbors according to Hongisto et al. (2013). The wall constructions separating the dwellings were also selected from those typically used in Finland. Finally, the level of the sounds was adjusted so that they represented the level in both sending and receiving rooms. The excellent work of Park and Bradley (2009) perhaps failed in this sense since the sound insulation values they used were too low to represent actual constructions between dwellings.

Presentation order of the sound stimuli

In Publications II and IV the presentation order of the stimuli was predefined with pseudo-random orders at our convenience. This enabled us to present the sounds in orders that we thought were appropriate; also, to avoid counter-productive random situations, for instance, quieter sounds first and louder sounds last. In publication II the randomization process applied to the 54 audio files (six sound types by nine walls) was balanced at both dimensions using Latin square orders. In Publication IV we always ensured that two consecutive tonal sounds would not have the same tonal frequency. In Publication I and III we were able to just randomize all the files because the order was not expected to affect to the subjective ratings.

Dummy sounds

Dummy sounds are part of the habituation process needed to properly instruct the subjects about the type of sounds to be rated in the test. They are used in psychoacoustic experiments with two purposes. First, they allow to present to the test subjects a representative collection of the experimental sounds before the actual rating. This action aims to enable the subjects to get to know the sounds and facilitate the creation of the rating scale in their minds prior the actual test. For instance, dummy sounds were used during the practice session in Publication II to practice the rating scale. In this case we created the additional audio files with a pop music piece from singer Jack Johnson. The track was filtered to sound as being transmitted through the nine walls as it was done for the rest of the test sounds. Second, it is expected that the first rating of a new series might not be accurate, and so dummy sounds can be played at the start of any new series of sounds. Those ratings are then ignored. In Publication III and IV we created dummy sounds to start the series, and the respective ratings were omitted from the analysis.

However, during our research, we were interested in evaluating how the presentation order could affect the ratings. We tried to answer two small questions: how tired the subject becomes with the test, or how the rating changes during their time there. In Publication III we replayed few audio files twice and performed a small *repeatability analysis* (unpublished). We did not see a statistical difference between the three sounds that we repeated twice. We considered that the rating remained sufficiently consistent during the experiment.

Number of test subjects

The performed literature review identified several experiments with too small numbers of test subjects, e.g., less than 10. For us, the smallest number of participants was 23 in Publication I. We obtained enough statistically significant differences. In Publications II, III, and IV the number of subjects was doubled because the research questions required more demanding statistical methods.

We were able to perform a listening experiment with over 50 subjects in about two months. Recruitment methods were created to systematically recruit personnel and students from our university and from the buildings nearby. We could argue that other set of participants could lead to different results, and the discussion can be expanded to other aspects like age, gender, hearing sensitivity, noise sensitivity, moderators and so on. In psychoacoustic experiments it should be tried to expand the age distribution and use large enough ranges. Keeping track of all that data is a good idea if any kind of distribution analysis is needed later.

Ethical considerations

The listening experiments of Publications I–IV included experimental sounds with low sound pressure levels, and so there was not a risk for hearing loss. In our cases, the studies were conducted in accordance with the guidelines of the National Advisory Board on Research Ethics 2009. According to Finnish research policy, we did not apply a statement from the ethics board because all of the following criteria were met: (1) The study did not interfere with the physical integrity of participants. (2) The study complied with the principle of informed consent. (3) Participants were over 15 years old. (4) The study did not expose participants to such exceptionally strong stimuli (e.g., violence) that the evaluation of potential harm would have required special expertise. (5) The study did not include a risk of long-term mental harm beyond the risks encountered in normal life. (6) Participation in the study did not entail a security risk for participants. Experiments occurring after 25th May 2018 need to fulfill the General Data Protection Regulation EU 2016/679 (Hoofnagle et al. 2019) with respect to acquisition and handling of personal information of test participants.

Communication with subjects

We considered it especially important to familiarize the participants with the type of sounds and the rating procedures before the actual experiments took place for the effectiveness of our tests. That had a rewarding impact in the quality of the experiments. In Publication IV the 95% confidence intervals of the ratings were satisfactory low. During the experiments, we developed a standard procedure where all test subjects received exactly the same information. To avoid rating bias, we considered important not to tell the participants the research questions of the experiment, neither what type of data we were gathering. At most, those things were explained if someone had some questions after the test. On the other hand, we put a lot of attentions to ensure that all participants understood how to use the rating scale and understand the experiment in the same way. Communication with subjects is rarely described in the literature, even though it is crucial to ensure that the ratings are not biased and that all subjects understand their role in the experiment in a similar way.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2022

ISBN 978-951-29-8941-6 (PRINT)
ISBN 978-951-29-8942-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)