Audiovisual Interaction in the Perception and Classification of Urban Soundscapes

Audiovisuele interactie in de perceptie en classificatie van stedelijke geluidslandschappen

Kang Sun

Promotoren: prof. dr. ir. B. De Coensel, prof. dr. ir. D. Botteldooren
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: elektrotechniek

UNIVERSITEIT GENT

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. B. Dhoedt
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2018 - 2019

ISBN 978-94-6355-170-0
NUR 962, 973
Wettelijk depot: D/2018/10.500/88

UNIVERSITEIT GENT

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie

Promotoren:

Prof. dr. ir. Bert De Coensel
Prof. dr. ir. Dick Botteldooren

Examencommissie:

Prof. dr. ir. Daniel De Zutter (voorzitter)	Universiteit Gent
Prof. dr. Mats E. Nilsson	Stockholm University
Prof. dr. Anna Preis	Adam Mickiewicz University
Prof. dr. Frank Witlox	Universiteit Gent
Prof. dr. ir. Annelies Bockstael	Universiteit Gent
Prof. dr. ir. Timothy Van Renterghem (secretaris)	Universiteit Gent
Prof. dr. ir. Bert De Coensel (promotor)	Universiteit Gent
Prof. dr. ir. Dick Botteldooren (promotor)	Universiteit Gent

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie
Tech Lane Ghent Science Park – Campus A (Campus Ardoyen)
iGent, Technologiepark-Zwijnaarde 15, 9052 Gent, Belgie

Tel.: +32 9 264 3316
Fax.: +32 9 264 3593

Proefschrift tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: elektrotechniek
Academiejaar 2018-2019

FACULTEIT INGENIEURSWETENSCHAPPEN EN ARCHITECTUUR

Acknowledgment

Acknowledgment is the last one to be written as it is tremendously arduous to look at own past. For a long time I am even avoiding to do so as the four years spent in Ghent University was... abundant. I only have a small gap between internal defense and public defense, which leaves me limited amount of time to write about feelings and people. With thorough deliberation, what I want to deliver in this acknowledgment is: sincerity. In this acknowledgment you will not find any name in a tedious long list that's almost like my phone book. But you will find your name between the lines if you are touched by one or several words.

First thing first: my gratitude to both supervisors. There are people that inspire you by their actions, surprise you by their keen sense of research directions, impress you by their composure when troubles come, help you with their professionalism and selflessness, and trust you unconditionally. I think my supervisors are quite matching and even beyond these statements.

When the snows fall and the white winds blow, the lone wolf dies, but the pack survives. There is no research without a supporting research group. In the lab when I yell for help for various problems; in the kitchen when we chat a cup of coffee for keep-up; on the way to field test when we sing a carpool karaoke; at those paper submission days when we figure out trivial but troublesome questions; and at those nights when we have unavoidable deadlines. Every now and then, these little moments with colleagues are flashing through my mind when thinking of PhD.

Friends are angels who lift us to our feet when your wings have trouble remembering how to fly. Friendship exists in the fallen leaves we step on in a hiking trip; in the reflection of the sunset in the river where we kayak; in the sea breeze and sandy beach where we spend a lazy afternoon; in the ice cream we take during a walk in Korenmarkt; in the cheers when we have a good shot in the sport court; in the swaying candle lights where we talk through the irritation; in the postcards where all the best wishes are signed; and in the peace of saying goodbye with “call whenever”.

Last but not the least, I thank the care, support and disinterested love from my parents, even from thousands of miles away. From east to west, from day to night, this treasure is always with me, never fades.

Kang Sun

Ghent, October, 2018

Table of Contents

Nederlandse Samenvatting

Als gevolg van verstedelijking neemt de bevolkingsdichtheid, mobiliteit en bijgevolg de aanwezigheid van mechanische geluiden in stedelijke gebieden over de hele wereld toe. Geluidshinder, vooral in en rond de woning, en de relatie ervan met blootstelling aan lawaai, is de afgelopen decennia grondig onderzocht, omdat het een van de meest prominente effecten is van blootstelling aan lawaai, zoals erkend door de Wereldgezondheidsorganisatie. Omgevingsgeluid kan echter ook een positieve invloed hebben: het kan de gemoedstoestand verbeteren, een prettige herinnering oproepen aan een eerdere ervaring, of aanmoedigen om te ontspannen en te herstellen. Omgevingsgeluiden roepen gedachten en emoties op, en kunnen onze stemming beïnvloeden of zelfs ons gedrag sturen. Bijgevolg beschouwt het wetenschappelijk onderzoek naar omgevingsgeluid de stedelijke geluidsomgeving steeds meer als één geheel, inclusief de positieve en de negatieve effecten die geluid kan bieden. De term geluidslandschap wordt door ISO gedefinieerd als een "akoestische omgeving zoals waargenomen of ervaren en/of begrepen door een persoon of mensen, in context". Steden bestaan uit vele soorten openbare ruimten, elk met hun kenmerkend geluidslandschap. Geïnspireerd door de potentiële positieve effecten die een geschikte akoestische omgeving kan hebben op het welzijn van de burgers en de aantrekkelijkheid van de stad, trekt de uitdaging van het ontwerp van de akoestische omgeving van open stedelijke ruimten daarom al decennialang de aandacht.

Architecten en stedenbouwkundigen erkennen steeds meer het belang van het geluidslandschap in de perceptie van de stedelijke openbare ruimte en de identiteit van een stad. Geluid en beeld kunnen echter niet als afzonderlijke entiteiten worden beschouwd; de beoordeling van onze leefomgeving wordt beïnvloed door zowel het landschap als het omgevingsgeluid. Bovendien is de invloed van visuele factoren op de perceptie van geluid nog niet volledig begrepen. In enquêtes rond omgevingsgeluid wordt het effect van visuele elementen, zoals het uitzicht vanuit het raam van de leefruimte in de woning, op de perceptie van het geluid in de eigen leefomgeving regelmatig opgenomen, maar minder vaak dan andere contextuele of demografische factoren. Bovendien wordt de beoordeling van omgevingsgeluid beïnvloed door een interactie tussen horen en zien, maar ook door persoonlijke factoren. Deze laatste weerspiegelen de verschillen in reactie op audiovisuele stimuli, toegeschreven o.a. aan iemands vermogen om de aandacht te focussen. Deze individuele verschillen blijken over het algemeen verder te gaan dan demografische verschillen en gevoeligheid aan geluid, en daarom wordt de perceptie van omgevingsgeluid best op een holistische manier behandeld.

In dit proefschrift worden een aantal laboratoriumexperimenten beschreven die trachten een beter begrip te geven van audiovisuele interactie in de perceptie van stedelijke geluidslandschappen. In hoofdstuk 2 wordt een experiment beschreven, uitgevoerd in een replica van een woonkamer, dat het effect van het zicht vanuit het raam op geluidsoverlast onderzoekt. Dit experiment werd zo realistisch mogelijk ontworpen. Zo kregen de deelnemers de opdracht om tijdens het experiment lichte activiteiten uit te oefenen, om niet op het geluid te focussen, en werd de duur van blootstelling aan elke stimulus hierop ingesteld. Omdat dit experiment erop gericht was om het effect van het uitzicht vanuit het raam te onderzoeken, werd een directe vergelijking tussen verschillende visuele stimuli vermeden door de verschillende delen van het experiment uit te voeren op verschillende dagen. Daarnaast was het experiment ook gericht op het identificeren van verschillen in geluidsgevoeligheid en het vermogen tot concentreren tussen personen. Om meer informatie te verkrijgen dan wat typisch via vragenlijsten kan worden bekomen, was een experiment met goede controle over de stimuli noodzakelijk. Dit vormde echter een uitdaging: beoordelen van geluidshinder op een ecologisch valide manier in experimentele opstelling is niet triviaal, omdat het gevaar bestaat dat in een experiment de belangrijkste verborgen factor die wordt onderzocht, nl. niet-vrijwillig gerichte aandacht, wordt vervangen door gerichte aandacht.

In het experiment beschreven in hoofdstuk 2 werd vastgesteld dat (1) de zichtbaarheid van de geluidsbron meer invloed heeft op zelf-gerapporteerde geluidshinder dan de zichtbaarheid van groene elementen; (2) zelf-gerapporteerde geluidsgevoeligheid de sterkste persoonlijke factor is, waarbij personen die gemakkelijk worden afgeleid door visuele elementen een significant lagere geluidsoverlast bij hetzelfde blootstellingsniveau melden; (3) er twee significante interacties zijn bij de voorspelling van zelf-gerapporteerde geluidshinder: a) tussen geluidsgevoeligheid en zichtbaarheid van de geluidsbron, en (b) tussen visuele dominantie, als een persoonlijke factor, en de zichtbaarheid van groene elementen.

De interactie tussen deze factoren levert aanvullend bewijs om de rol van audiovisuele aandacht in de studie van geluidsoverlast te ondersteunen. In hoofdstuk 3 worden vervolgens de verschillen tussen personen in hoe zien of horen hun perceptie domineert verder onderzocht, en wordt een onderliggend mechanisme met de naam "audiovisuele aanleg" voorgesteld. Hierbij wordt een onderscheid gemaakt tussen nauwkeurige en minder nauwkeurige luisteraars, en tussen proefpersonen die wel of niet kunnen worden afgeleid door incongruente visuele informatie. Om dit voorgestelde mechanisme te valideren, werden de resultaten van twee eerder uitgevoerde experimenten dieper geanalyseerd. Het eerste experiment concentreert zich op zelf-gerapporteerde geluidshinder in de woonomgeving (en werd besproken in hoofdstuk 2); het tweede experiment richt zich op de mate waarin open stedelijke ruimten als aangenaam worden ervaren. In het eerste experiment bleek de invloed van de zichtbaarheid van vegetatie op zelf-gerapporteerde geluidshinder door audiovisuele aanleg te worden gewijzigd. In het tweede experiment werd vastgesteld dat de beoordeling van het lopen over een brug wordt beïnvloed door audiovisuele aanleg, in het bijzonder wanneer een opvallend geluidsscherm wordt gebruikt om de geluidsniveaus van snelwegverkeer te verminderen. Er kan hieruit dus worden geconcludeerd dat audiovisuele vaardigheden van invloed kunnen zijn op de beoordeling van de leefomgeving.

Naast het ontwarren van de interactie-effecten tussen auditieve, visuele, persoonlijke en contextuele factoren in de perceptie van stedelijke geluidslandschappen, blijft ook het beschrijven en classificeren zelf van geluidslandschappen een uitdaging. Net als bij de perceptie, dient het opnemen, reproduceren en classificeren van stedelijke geluidslandschappen ook op een holistische manier te worden uitgevoerd, omdat elk geluidslandschap "in context" dient te worden beschouwd. Tot op heden bestaat er echter geen gestandaardiseerd protocol voor het audiovisueel opnemen van stedelijke geluidslandschappen en voor de immersieve weergave ervan. Voor het classificeren van geluidslandschappen werden in het verleden reeds een aantal holistische methoden voorgesteld, zoals het bekende “affect circumplex” model, dat toelaat om geluidslandschappen te classificeren in een 2D voorstelling. Hoewel het erg populair is, werd dit beoordelings- en classificatiekader ook onderworpen aan enige kritiek, omdat het niet volledig rekening houdt met de context en het doel van de omgeving.

De opkomst van realistische en betaalbare immersieve audiovisuele reproductiesystemen, zoals virtuele realiteit (VR) brillen, ondersteund door steeds efficiëntere modellen voor auralisatie, maakt een immersieve reproductie van geluidslandschappen in laboratoriumomgeving mogelijk. Een dergelijke reproductie kan ook een waardevol instrument vormen voor participatieve evaluatie van het geluid in stadsontwerp. In hoofdstuk 4 wordt een immersieve methodologie voor het opnemen en reproduceren van geluidslandschappen voorgesteld, waarbij spatiale audio wordt gecombineerd met 360-graden video. Uit de resultaten van een eerste experiment blijkt dat deze reproductiemethodologie als ecologisch valide kan worden beschouwd, in termen van realisme en immersiviteit. Vervolgens wordt een hiërarchische methode voor het classificeren van geluidslandschappen voorgesteld, waarbij een onderscheid wordt gemaakt tussen onopvallende en opvallende, tussen storende en activiteits-ondersteunende, en ten slotte tussen kalmerende en stimulerende geluidslandschappen. Een tweede experiment, ontworpen om de voorgestelde classificatiemethode te vergelijken met bestaande methoden, wordt vervolgens besproken. Op basis van de resultaten van dit tweede experiment werd een model geconstrueerd dat gebaseerd is op een beperkt aantal akoestische indicatoren. Dit model maakt het mogelijk om een geluidslandschap te classificeren in een van vier vooropgestelde categorieën, met een nauwkeurigheid van meer dan 88%.

Samengevat, de belangrijkste bevindingen en nieuwe technieken geïntroduceerd in deze thesis zijn:

- Audiovisuele aanleg, een persoonlijkheidskenmerk dat gelijklopend is aan andere psychologische concepten zoals geluidsgevoeligheid, heeft een modererend effect of audiovisuele interacties in zowel binnen- als buitenomgevingen.
- Een nieuwe methode is geïntroduceerd voor immersieve audiovisuele reproductie van buitenomgevingen, gebaseerd op het simultaan presenteren van 360-graden video en spatiale audio.
- Een hiërarchische methode voor het classificeren van stedelijke geluidslandschappen is voorgesteld, dewelke gebaseerd is op de mate waarin het geluidslandschap bijdraagt aan de perceptie van de omgeving als geheel.

English Summary

Due to urban sprawl and rural urbanization, the population density, urban mobility, and, consequently, the abundance of mechanical sounds in urban areas across the world is ever increasing. Noise annoyance, especially in and around the dwelling, and its relation with noise exposure, has been investigated thoroughly in recent decades, as it is one of the most prominent effects of noise exposure, as recognized by the World Health Organization. However, ambient sound may also provide a positive influence, such as enhancing a person's mood, triggering a pleasant memory of a prior experience, or encouraging a person to relax and recover. Ambient sounds may evoke thoughts and emotions, may influence our mood or steer our behavior. As a consequence, scientific research on environmental sound is steadily moving from considering urban noise as a nuisance to considering the urban soundscape as a whole, including the positive as well as the negative effects sound may provide. The urban soundscape is defined by ISO as an "acoustic environment as perceived or experienced and/or understood by a person or people, in context". Cities are comprised of many types of public outdoor spaces, each with their distinctive soundscape. Inspired by the potential positive effects a suitable acoustic environment may have on well-being of citizens and the attractiveness of the city, the challenge of designing the acoustic environment of urban public outdoor spaces has therefore attracted attention since decades.

It is increasingly acknowledged by (landscape) architects and urban planners that the soundscape contributes significantly to the perception of urban public open spaces and the identity of a city. However, landscape and soundscape cannot be considered as separate entities; the appraisal of our living environment is influenced by landscape and soundscape alike. Moreover, the influence of visual factors on sound perception is not yet completely understood. In environmental noise surveys, the effect of visual elements, such as the view from the window, on the perception of the sound within one's living environment has been addressed before, yet less frequently than other contextual factors. Moreover, this appraisal is influenced by an interaction between audition and vision, as well as by personal factors. The latter reflect the differences in reaction to audiovisual stimuli, attributed to attitude, sensory and attention focusing capabilities. These individual differences are commonly found to go beyond demographic information and noise sensitivity, and therefore, environmental sound perception should be treated in a holistic manner.

In this dissertation, a set of experimental studies are described that attempt to achieve a better understanding of audiovisual interaction in the perception of urban soundscapes. In Chapter 2, an experiment performed in a mockup living room is described, that investigates the effect of the view from the window on noise annoyance. This experiment was designed to be ecologically valid as much as possible. Firstly, participants were instructed to engage in some light activity

during the experiment in order not to focus on the sound, and the exposure time for each stimulus was set accordingly. Secondly, since the aim of this experiment was to investigate the effect of the view from the window, direct comparison between different visual stimuli was avoided by showing the visual stimulus in a natural setting, a mockup window, and by presenting the different visual stimuli on different experiment days. In addition, the experiment aimed to identify subjective noise sensitivity and attention focusing capability as personal factors. To be able to go beyond questionnaires for assessing personal factors, a laboratory study using well controlled stimuli was opted for. This presented a challenge: assessing noise annoyance in an ecologically valid way in an experimental setup is rather difficult as the main hidden factor under investigation, i.e. non-voluntary attention, is replaced by focused attention in a listening experiment.

In this experiment it was found that (1) sound source visibility, as a functional parameter of the visual setting, has more impact on self-reported noise annoyance than the visibility of green elements within the visual scene; (2) self-reported noise sensitivity remains the strongest personal factor, yet persons being easily distracted by visual elements report significantly lower noise annoyance at the same exposure level; (3) two significant interactions can be observed in the prediction of self-reported noise annoyance: (a) noise sensitivity interacts with sound source visibility; (b) vision dominance, as a personal factor, interacts with the visibility of green elements.

The interaction between these factors provides additional evidence to support the role of audiovisual attention in the emergence of noise annoyance. Chapter 3 further explores the individual difference in how vision or audition dominates perception, and based on the results of a laboratory experiment, an underlying mechanism labelled as “audiovisual aptitude” is proposed. A deeper analysis allowed to distinguish between accurate and less accurate listeners, and between participants that are easily visually distracted and those that are not. To validate this proposed mechanism, two previously conducted laboratory experiments were re-analyzed. The first experiment focuses on self-reported noise annoyance in a living room context (and was discussed in Chapter 2); the second experiment focuses on the perceived pleasantness of outdoor public spaces. In the first experiment, the influence of visibility of vegetation on self-reported noise annoyance was found to be modified by audiovisual aptitude. In the second experiment, it was found that the overall appraisal of walking across a bridge is influenced by audiovisual aptitude, in particular when a visually intrusive noise barrier is used to reduce highway traffic noise levels. Thus it could be concluded that audiovisual aptitude may affect the appraisal of the living environment.

Next to disentangling the effects of audition, vision, personal factors and context on the perception of the urban soundscape, simply describing and classifying soundscapes by itself remains a challenge. As with perception, the recording and classification of urban soundscapes should also be performed in a holistic manner, as each soundscape has to be considered “in context”. To date, however, no standardized protocol exists for immersive audio-visual recording and playback of urban acoustic environments with soundscape in mind. For classifying soundscapes, a number of holistic methods have been proposed, such as the well-known circumplex model of affect that can be mapped to a two-dimensional plane. Although very popular, this assessment and classification framework has also been subject to some critique, as it does not fully take into account context and the purpose of a space.

The advent of realistic and affordable immersive audio-visual reproduction systems (head-mounted displays), backed by increasingly efficient and realistic acoustic simulation and auralization models, has enabled the immersive reproduction of soundscapes in a laboratory environment. Immersive virtual reality could also become a valuable tool for interactive participatory evaluation of the soundscape in urban planning and design projects. In Chapter 4, an immersive soundscape reproduction methodology that combines spatial audio with 360-degree video, presented through a virtual reality headset, is proposed. An audiovisual experiment is presented, which shows that the reproduction methodology is perceived as ecologically valid in terms of realism and immersion. Subsequently, a hierarchical method for soundscape classification is proposed, which distinguishes between backgrounded and foregrounded, disruptive and supportive, and finally calming and stimulating soundscapes. A second experiment is presented that was designed to compare the proposed classification method with existing methods. On the basis of the results of this experiment, a model based on a limited number of acoustical indicators was constructed that allows to classify a soundscape in each of the four proposed categories, with an accuracy exceeding 88% on an independent dataset.

To conclude, the main findings and novel techniques introduced in this dissertation are:

- Audiovisual aptitude, as a personal factor similar to other well-known psychological concepts such as noise sensitivity, moderates audiovisual interactions in the assessment of both indoor and outdoor environments.
- A novel method for immersive audiovisual reproduction of outdoor environments is introduced, which is based on a simultaneous presentation of 360-degree video and spatial sound recordings.
- A hierarchical classification scheme for urban soundscapes is proposed, based on how the soundscape contributes to the perception of the overall environment.

List of Abbreviation

AAO	Auditory attention Attracting Object
AIC	Akaike information criterion
AM	Amplitude
AO	attention Attracting Object
API	Application Programming Interface
ANOVA	Analysis of variance
BIC	Bayesian information criterion
EEG	Electroencephalogram
FM	Frequency modulation
FPR	False positive rate
GLMM	Generalized linear mixed model
ICBEN	International Commission on Biological Effects of Noise
ISO	International Organization for Standardization
NDVI	Normalized difference vegetation index
NoiSeQ	Noise-Sensitivity-Questionnaire
PCA	Principal component analysis
PTA	Pure tone audiometry
ROC	Receiver operating characteristic
SE	Standard Error of the Estimate
SPL	Sound pressure level
TPR	True positive rate
VAO	Visual attention Attracting Object
VR	Virtual reality
2D	2 dimension

List of Symbols

df	degree of freedom
F	*F*-statistic
G_{RGB}	RGB greenness (R(red)G(green)B(blue))
J	Youden index
L_{Aeq}	A-weighted, equivalent sound pressure level
$L_{Aeq,\ 1\ min}$	A-weighted equivalent sound pressure levels during the one-minute period
L_{AFmax}	The maximum level with A-weighted frequency response and Fast time constant
L_{Ax}	A-weighted noise level just exceeded for x% of the measurement period
L_{den}	Day-evening-night equivalent level
L_{dn}	Day-night equivalent level
N	Loudness
p	significance
R, R^2	The Pearson product-moment correlation coefficient
S	Sharpness
SL	Saliency
S(x)	fuzzy set score
S'(x)	adjusted fuzzy set score, where the AND and NOT operator is implemented as a probabilistic t-norm and fuzzy negation
Δ	difference

List of Publications

Articles in international journals

- Sun K, De Coensel B, Echevarria Sanchez GM, Van Renterghem T, and Botteldooren D. (2018). Effect of interaction between attention focusing capability and visual factors on road traffic noise annoyance. *Applied Acoustics*, 134, 16-24.
- Sun K, Echevarria Sanchez GM, De Coensel B, Van Renterghem T, Talsma D, and Botteldooren D. (2018). Personal audiovisual aptitude influences the interaction between landscape and soundscape appraisal. *Frontiers in Psychology*, 9:780.
- Echevarria Sanchez GM, Van Renterghem T, Sun K, De Coensel B, and Botteldooren D. (2017). Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. *Landscape and Urban Planning*, 167, 98-107.
- Sun K, De Coensel B, Filipan K, Aletta F, Van Renterghem T, De Pessemier T, Joseph W, and Botteldooren D. Classification of soundscapes of urban public open spaces. Submitted to *Landscape and Urban Planning*.

Articles in conference proceedings

- Sun K, Botteldooren D, De Coensel B. (2018). Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation. *Proceedings of the 47th International Congress and Exposition on Noise Control Engineering*. Institute of Noise Control Engineering.
- Sun K, De Coensel B, Echevarria Sanchez GM, Van Renterghem T, and Botteldooren D. (2017). Effects of Human Factors on the Interaction between Visual Scene and Noise Annoyance. *In INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 255(2), 5530-5536. Institute of Noise Control Engineering.
- De Coensel B, Sun K, and Botteldooren D. (2017). Urban Soundscapes of the World: selection and reproduction of urban acoustic environments with soundscape in mind. *In INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 255(2), 5407-5413. Institute of Noise Control Engineering.
- Echevarria Sanchez GM, Van Renterghem T, Sun K, De Coensel B, and Botteldooren D. (2017). Personal factors affecting the audio-visual perception of the urban public space. *In INTER-NOISE and NOISE-CON*

Congress and Conference Proceedings, 255(2), 5714-5723. Institute of Noise Control Engineering.

- Botteldooren D, Sun K, Echevarría Sánchez GM, De Coensel B, and Van Renterghem T. (2017). The influence of audio-visual aptitude on audio-visual interaction in appraisal of the environment. *In 12th International Congress on Noise as a Public Health Problem* (ICBEN 2017).
- Sun K, De Coensel B, Echevarría Sánchez GM, Van Renterghem T, and Botteldooren D. (2016). Effects of sound source visibility on sound perception in living room environment. *In INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 253(2), 6420-6425. Institute of Noise Control Engineering.
- Echevarria Sanchez GM, Van Renterghem T, Sun K, De Coensel B, and Botteldooren D. (2016). The relative importance of visual and sound design in the rehabilitation of a bridge connecting a highly populated area and a park. *In INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 253(2), 6810-6816. Institute of Noise Control Engineering.
- Sun K, Liang X, Botteldooren D, and De Coensel B. (2015). Evaluating the university campus soundscape: The case of Tianjin University. *In 10th European Congress and Exposition on Noise Control Engineering* (Euronoise 2015) (pp. 1517-1521).
- De Coensel B, Sun K, Wei W, Van Renterghem T, Sineau M, Ribeiro C, Can A, Aumond P, Lavandier C, and Botteldooren D. (2015). Dynamic noise mapping based on fixed and mobile sound measurements. *In 10th European Congress and Exposition on Noise Control Engineering* (Euronoise 2015) (pp. 2339-2344).

1

Introduction

1.1 Urban soundscape

1.1.1 Urban life

An urban area is a human settlement with high population density and infrastructure of built environment. Urban areas are created through urbanization and are categorized by urban morphology as cities, towns, conurbations or suburbs (Wikipedia contributors, 2018). Based on the EEA Report: Urban sprawl in Europe – the ignored challenge (EEA, 2006), urban sprawl has accompanied the growth of urban areas across Europe over the past 50 years. Countries or regions with economic activity and high population density such as Belgium, The Netherlands, southern and western Germany, northern Italy and the Paris region are experiencing the most visible impacts of urban sprawl. Urbanization is progressing rapidly in lesser developed regions as well, and the urban population is anticipated to grow an average 2.3% per year in the developing world between 2000 and 2030 (UN, 2000). Almost all of the world's total population growth in this period is expected to be absorbed by urban areas within less developed regions (Brockherhoff, 2000). According to the latest estimate and projection released by the Population Division of the United Nations, the world's urban population continues to grow at a higher rate than the total population of the world, and 3 billion people or approximately 48% of the world population are now city dwellers (UN, 2004). By 2050, about 70% of the World's population will be living in cities (UN, 2014). Thus, urbanization is a challenge and a fact that both developed and developing countries have to face.

According to the general rule of global urbanization, urbanization levels between 30% and 70% are considered to indicate accelerated development (Northam, 1975). During such periods, the required support of money, resources and population transfer may greatly reduce the environmental carrying capacity and lead directly to tremendous pressure on the urban environment (Wang et al., 2008; Chen, 2007). Thus urban areas gather all the resources, enjoy the benefits and unavoidably face all the consequences. Accordingly, it is essential to develop methods of enabling rapid development in a sustainable manner, while maintaining a high quality of life through coordination of urbanization and the environment (Li et al., 2012).

Life inside urban areas is significantly associated with various forms of physical activity and health outcomes (Ewing et al., 2008). It therefore warrants attention to study the drawbacks of urbanization next to its benefits. The environmental impacts of urban sprawl (Johnson, 2001) include (but are not restricted to): loss of environmentally fragile lands, reduced regional open space, higher levels of air pollution, higher energy consumption, decreased aesthetic appeal of landscape (Burchell et al, 1998), loss of farmland, reduced diversity of species, increased runoff of stormwater, increased risk of flooding (Adelmann, 1998; PTCEC, 1998), excessive removal of native vegetation, monotonous (and regionally inappropriate) residential visual environment, absence of mountain views, ecosystem fragmentation (Margules and Meyers, 1992). One invisible aspect that has drawn an increasing amount of attention in recent decades, is the sonic environment. The latter consists of the sounds originating from all the urban elements (Schafer, 1993).

1.1.2 From sonic environment to soundscape

One important aspect of urban life is sound. The urban sonic environment is one of the defining factors of a city. Cities are comprised of a wide variety of outdoor spaces, each with their distinctive sonic environment, which is typically composed of sounds from human activity, mechanical sounds, and sounds from nature. We hear voices, vehicles, birds, wind in trees, machinery, footsteps, raindrops, telephones, the hum and beeps of our electronics, dogs barking, and even more. The sonic environment of a place or space is the sound from all sources that could be heard by someone (Brown, et al., 2015).

Humanity is increasingly urban, but continues to depend on nature for its survival. Natural ecosystems that are located outside or stretch beyond the city limits benefit cities as well (Bolund and Hunhammar, 1999). For instance, they possess the capacity to counterbalance environmental exposure (Van Kamp et al., 2003). Human reaction to sound can be traced back to our biological origins. For example, research has found that chronic and frequent sound stimuli interfere with animals' abilities to detect sounds which may be important for survival, whereas intermittent and unpredictable sound is often perceived as a threat (Francis and Barber, 2013). In a similar manner, long-term exposure to high-level sounds in urban areas affects the well-being of residents (Ewing et al., 2008) and quality of urban life. Historically, this has driven people into negative thinking of sound – therefore the term “noise” is often used. Sound in outdoor environments has traditionally been considered in negative terms as both intrusive and undesirable (Jennings and Cain, 2013). Indeed, various adverse health effects are related to individual noise exposure in residential areas (WHO, 1999). Research on sleep disturbance caused by traffic noise has found that respondents living in noisy areas have significantly more difficulties with falling asleep, and exhibit poor sleep quality, tiredness after sleep and possible increase use of sleeping pills (Jakovljević et al., 2006).

However, sound is essential for mental health (Schlesinger and Meadow-Orlans, 1972) and may provide positive effects as well, such as enhancing a person's mood, triggering a pleasant memory of a prior experience, or encouraging a person to relax and recover (Payne, 2013). Thus, it is obvious that all different kinds of sounds that form the urban sonic environment have to be considered. In general, people tend to save the wanted sounds and eliminate the unwanted, suggesting that it is not purely about the sonic environment, but also about how people perceive and understand it.

The concept of “soundscape” was first used in urban context by Southworth in 1969 (Southworth, 1969). Later on, the Canadian composer Murray Schafer popularized the term (Schafer, 1969) and presented the “World Soundscape project”, in which he introduced soundscape as an acoustic field of study. He suggested that it is less easy to formulate an exact impression of a soundscape, as compared to the case of landscapes (Schafer, 1977). In later work, Schafer commented that “Noise pollution results when man does not listen carefully. Noises are the sounds we have learned to ignore. We must seek a way to make environmental acoustics a positive study program.” (Schafer, 1993). In 2014, the International Organization for Standardization (ISO) has defined soundscape as an “acoustic environment as perceived or experienced and/or understood by a person or people, in context” (ISO, 2014). Note that the term “soundscape” used in this dissertation refers to the ISO definition.

1.2 Soundscape evaluation

1.2.1 Overview

By its definition, soundscape research represents a paradigm shift in the field of sound evaluation. First, it involves human perception in the assessment of sound, and second, it expands on classical physical measurements and makes reference to the use of different investigative measurement methods (Schulte-Fortkamp and Fiebig, 2015). Access to high-quality sonic environments may positively affect well-being, quality of life (WHOQOL Group, 1998), and environmental health through restorative or health and wellbeing promoting mechanisms (Van Kamp et al., 2015). However, in order to get a better understanding of soundscapes, how to improve them and to obtain positive effects on well-being and quality of life, it might be easier to start with the opposite – annoyance.

It is necessary to specify certain phrases used in this chapter. On the one hand, the term "annoyance", in epidemiological research, refers to retrospective yearly averaged reported noise annoyance. It is often measured through interviews at home or questionnaire surveys and highly relies on one's experience rather than one's instant sensory perception. On the other hand, within the paradigm of soundscape, the term "annoyance" is typically used to refer to instantaneous annoyance or, in a broader way, one's short-term perception and understanding of a sonic environment. Short-term annoyance refers to one's displeasure with the environmental sound, which is probably closer to activity disturbance, but does not integrate over different activities. In the remainder of this work, the term annoyance will refer to instantaneous annoyance, unless stated otherwise. Obviously, instantaneous annoyance has an influence on retrospective annoyance (Västfjäll, 2004). Thus, epidemiological annoyance is useful as a guideline for assessing observed differences in short-term annoyance.

With the rapid speed of urbanization, dealing with noise is an unavoidable challenge. The influence of sound exposure on annoyance in public open space (De Coensel et al., 2005; Pedersen and Persson Waye, 2004), as well as in and around dwellings (De Coensel et al., 2007; Sato et al., 1999), has been explored in depth. Noise exposure has a clear impact on human health, on sleep disturbance and on human behavior in general (Ouis, 2001; Öhrström et al., 2007; Douglas and Murphy, 2016; Evans et al., 2001). It is believed that good soundscape quality in suburban green areas and city parks can only be achieved if the traffic noise exposure during daytime is below 50 dB(A) (Nilsson and Berglund, 2006; Nilsson, 2007). The presence of construction noise in combination with road traffic noise, and the level of road traffic noise are strongly related to annoyance rating (Jeon et al., 2010). On the one hand, these studies emphasize the importance of sound in the emergence of noise annoyance. On the other hand, in earlier studies non-acoustic factors, such as landscape, social and behavioral factors, are found to be important modifiers for sound perception (Yu and Kang, 2008; Jeon et al., 2011; Liu et al., 2013). Furthermore, audio-visual interaction, which has first been studied in the field of object recognition (Erber, 1969), also influences the perception of the sonic environment (Carles et al., 1992). In addition, inter-individual differences also arise as an important factor that modifies the perception of the sonic environment (De Coensel and Botteldooren, 2006; Filipan et al., 2017).

In the following sections a closer look is given to the aspects mentioned above, and the underlying mechanisms that combine these factors in an ecological valid way are examined.

1.2.2 Saliency and attention

Environmental sound by definition is not the primary focus of attention of a person submerged in it. Rather, specific sounds that stand out, that are salient, attract attention and become auditory objects as the listener starts paying attention to them (Botteldooren et al., 2015). The key transfer from sonic environment to soundscape is to notice. The process of noticing a sound is influenced by two interchanging processes: top-down and bottom-up attention (Terroir et al., 2013; Kaplan and Kaplan, 1989). On the one hand, top-down attention is voluntary: it assumes active listening to the sounds occurring in the environment. On the other hand, bottom-up attention is involuntary and is influenced by the listeners' general state of mind, the task at hand and very importantly – the sonic environment. To investigate the bottom-up attention to sound, saliency is introduced as a concept. Saliency indicates how much a specific sound or a sound event stands out of its background. As a consequence, the higher the saliency, the higher the probability of a sound being noticed. Although related to perception, it is possible to define the physical characteristics that contribute to saliency (Kaya and Elhilali, 2017).

Not only the composition of the acoustic environment determines what sounds are noticed but also the attentiveness, current activities, and expectations of the listener and its prior knowledge of the sounds that could be heard. Understanding human auditory scene analysis and the important role of auditory attention (Oldoni et al., 2013) allows us to outline better soundscape assessment methods and to come to enhanced methodologies for designing desirable soundscapes within a specific context and for a specific use (Kang et al., 2016). Wood and Cowan (1995) replicated and extended Moray's (1959) investigation of the cocktail party phenomenon, which refers to a situation in which one can attend to only part of a noisy environment, yet highly pertinent stimuli such as one's own name can suddenly capture attention. The findings on working memory capacity reflect individual differences in the ability to control attention and avoid distraction (Conway et al., 2001). It has been shown that high-working memory capacity individuals are less susceptible to the effects of auditory distractors (Beaman, 2004; Sörqvist, 2010). Orienting is fundamentally a multifaceted reaction to an event's significance, engaging sensory-motor processes that support not only passive and active attention, but what is viewed here as its foundation: natural selective attention (Bradley, 2009).

Thus it is important to bear in mind the concept of attention thoroughly. In this dissertation, it is used as a guideline not only in perception of a specific soundscape situation, but also in describing and classifying the soundscape later on.

1.2.3 Audiovisual interaction

1.2.3.1 General concept

In early days, when the concept "audiovisual interaction" was introduced in speech recognition, it was found that observers rely increasingly more on visual cues for speech information as the signal/noise ratio is degraded. Furthermore, audiovisual speech recognition performance was found to be more variable than audition-only performance (Erber, 1969). Audiovisual interaction is the result of the interplay between three main factors: sound, vision and person. These factors are not independent, but interact with each other, working in multiple layers and modifying perception. Clearly, audiovisual speech stimuli are easier to recognize than audition-only or visual-only stimuli in terms of accuracy, speed and understanding. One possible explanation from neural studies is that the multisensory brain areas, playing a role in audiovisual integration of phonemes and graphemes, participate in the neural network supporting the supramodal concept of a "letter of alphabets", having both auditory (phonemic) and visual (graphemic) qualities (Raij et al., 2000). Neurophysiological and behavioral studies in animals also outlined the principles underlying the crossmodal spatial integration between auditory and visual stimuli in space perception, pointing out an enhanced visual perceptual performance with audiovisual stimuli (Bolognini et al., 2005).

In soundscape studies, it is found that the sound and not the visual component dominates the patterns of preference, which is attributed to the more varied nature of the sounds presented, in comparison with the relatively homogenous quality of the visual scenes shown (Carles et al., 1992). However, "bad" visual scenes would contaminate judgments of what we hear (Viollon et al., 2002). By definition, the notion of soundscape puts emphasis not only on the physical characteristics of the sound but also on the perception of the listener, as well as on the relationship between both (Ge et al., 2009). These contradictory findings, from first sight, may due to the degree of matching between visual and sound information. Also, the degree of implication of the perceiver based on the audiovisual stimuli might play a role.

1.2.3.2 Auditory factors

For the European Union's noise indicator, L_{den}, exposure-effect relationships have been derived (Miedema and Oudshoorn, 2001). The relationship between sound exposure and annoyance goes beyond the level (Landström et al., 1995; Raimbault and Dubois, 2005). Even at low noise levels, a small percentage of people are still highly annoyed (Fields, 1993). Not only the average noise level over a particular time period, but also the highest noise level during that time period has a significant correlation with annoyance (Sato et al., 1999), which suggests that noise control measures should target noisy vehicles. Specific for quiet (rural) soundscapes, earlier research suggests that a multi-criteria approach is a good option for soundscape quality assessment, as the restorative and appealing power of the area should be taken into account (De Coensel and Botteldooren, 2006). By only relying on sound pressure levels averaged over long time periods, and by suppressing all aspects of quality, the specific acoustic properties of environmental noise leading to annoyance cannot be fully identified; annoyance caused by environmental noise has a broader linkage with various acoustical properties such as frequency spectrum, duration, impulsiveness, tonal and low-frequency components, etc. than only with sound pressure level (SPL) (Fastl et al., 1996).

In addition to purely acoustical factors, the sound source plays an important role. Specific types of sounds and their associated meanings have been found to be more important in influencing the perceived restorativeness of the soundscape than its overall sound pressure level (Payne, 2013). For instance, the noise from wind turbines, which are increasingly being installed in the USA and in Europe, strongly annoy people living in their vicinity (Pedersen and Persson Waye, 2004). Another example is railway noise, which is found to be less annoying than aircraft and road traffic noise at a similar noise level (Fields and Walker, 1982). Sounds from nature facilitate recovery from sympathetic activation after a psychological stressor (Alvarsson et al., 2010). High proportion of people are favorable to water sound and birdsong, which are sounds that tend to be perceived as more tranquil and less invasive (Yang and Kang, 2005). Sound marks, such as sound signals (footsteps, voices), complex dynamics governing natural sound include the chorus of birds singing or the sound of wind blowing in trees, changes in the murmur of a passing plane or music backgrounds, would subsequently shape territories, which ultimately improve the quality of everyday life (Raimbault and Dubois, 2005; Botteldooren et al., 2006; De Coensel et al., 2003).

Inspired by the importance of the sound source, many researchers attempt to systematically categorize sound sources that potentially appear in the urban environment, as an approach to study the urban soundscape (Brown et al., 2011; Yang and Kang, 2005; Lavandier and Defréville, 2006). As an example, semantic criteria are applied by Schafer to distinguish between road traffic (car–truck–motorcycle), other forms of transportation (railway, aircraft), working machines (street cleaning, working site), music, people’s presence (speech, walking), and nature (wind, animals) (Schafer, 1977; Delage, 1980). Such classification respects the objective of the sound, which leaves a certain impression on people over the long term. This approach helps to diagnose the main components of a new sonic environment at first sight, and allows to employ a strategy of keeping the wanted sounds and eliminating the unwanted sounds. However, this approach doesn’t consider the possible interactions between each category, nor the interaction with other factors involved in perception. For example, earlier research has found that water sounds with relatively greater energy in low-frequency ranges were effective for masking noise caused by road traffic (You et al., 2010). Previous research of using water sounds to mask road traffic noise revealed that urban soundscape preference is affected by the acoustical characteristics of water sounds (sharpness) and visual images of water features (Jeon et al., 2012). This reveals that first, active noise control could consider introducing a more favorable sound; and second, visual information modifies the perception of the auditory scene. This kind of audiovisual interaction provides a way for urban designers and urban planners to optimize urban soundscapes.

Other factors, such as exposure duration, occasion, etc., may also play a role. Earlier research found that longer exposure resulted in increased annoyance but did not alter the differential effect of disruption on annoyance, which might indicate that annoyance cannot be conceived of as a purely perceptual sound property, rather, it is influenced by the degree of interference with the task at hand (Zimmer et al., 2008). Although music is generally considered to be a positive sound, music can also interfere with activities. Earlier research indicates that music could distract, and thus lower task performance, when sound levels are sufficiently high (Wolfe, 1983). Moreover, identifiable music could trigger memory and emotion, and thus makes concentration harder while performing an unrelated task. The above suggests that soundscapes should be analyzed in a holistic manner, rather than being focused only on the auditory factor.

1.2.3.3 Visual factors

Previous research has indicated that more than 80% of the human sensory input is visual (Rock and Harris, 1967). As stated before, a "good" view might increase one's auditory perception and vice versa. Regarding the visual factor, a green view which contains vegetation has been frequently mentioned as being positive for perception. Earlier research suggested that exposure to restorative environments facilitates recovery from mental fatigue (Berto, 2005). Visiting natural environments in urban area (such as urban parks) has been shown to achieve great restorative effects (Hartig et al., 1991). Moderate evidence is found in electroencephalogram (EEG) studies, in which it is shown that the presence of vegetation may reduce the negative perception of noise (Yang et al., 2011).

Some may not agree. For instance, it is not always the case that a green surrounding is perceived as better, when it is combined with different types of sound environment (Brambilla and Maffei, 2006). Also, when using green noise barriers to reduce noise annoyance, the visually attractiveness is important (Hong and Jeon, 2014; Veisten et al., 2012). Though people have a certain preference for suburban green areas and city parks, an earlier study suggested that such areas can only be perceived as having a good soundscape quality if the traffic noise level is below 50dB(A) (Nilsson and Berglund, 2006). On the contrary, other research found that in an at-home situation, the road traffic noise facade insulation, measured in-situ at each dwelling, could not be linked to self-reported noise annoyance (Van Renterghem and Botteldooren, 2016). They further showed that a real view on outdoor vegetation was essential for reducing noise annoyance.

Another visual factor that is considered frequently is sound source visibility. An earlier study pointed out that seeing the sound source would increase subjective annoyance (Zhang et al., 2003). A similar trend was also found in a wind turbine noise study (Pedersen and Larsman, 2008), which suggests that blocking the view to the sound source might ultimately help to reduce annoyance. However, others found that, under the same noise exposure level, average ratings of noisiness were higher when the degree of visual screening was higher (Watts et al., 1999). Previous research also showed that noise annoyance behind transparent barriers (where the sound source can be seen) is lower than noise annoyance behind opaque barriers (Maffei et al., 2013). It is suggested that people tend to be more anxious when a moving sound source cannot be seen. Recent research has therefore attempted to explain this inconsistency by the type of sound source:

adding visual information to a listening experiment tended to reduce annoyance if the sound source was believed to have a positive influence, while annoyance increased for mechanical sound sources (Preis et al., 2016). This again highlights the effect of audiovisual interaction, as the auditory factor (the sound type) modifies visual preference.

1.2.3.4 Individual differences

Epidemiological research has shown that personal factors, such as age, gender, education and noise sensitivity, as well as social variables, modify the influence of sound exposure on retrospective annoyance at home (Guski, 1999). For annoyance from transportation noise, age has an effect (Day-Night-Level equivalent to 5 dB) (Miedema and Vos, 1999). The relationship between age and annoyance forms an inverse U-shaped curve, where the middle-aged group has the highest annoyance (Janssen et al., 2011; Miedema and Vos, 2004). Some research reported that women were more likely to report high noise annoyance (Dratva et al., 2010), while other research found there is no significant relationship between gender and noise annoyance (Miedema and Vos, 1999).

Such demographic information is easy to access and might/might not have an impact on noise annoyance. However, subjective noise sensitivity, which was first introduced by Weinstein (1978) as a quantity measurable with a set of questionnaires, was shown to be a very stable personality trait which is determined both by inheritance and experience (Schreckenberg et al., 2010; Västfjäll, 2002). This personal trait reflects the attitude towards a wide range of sounds, which does not necessarily link to individual demographic information (Stansfeld, 1992; Weinstein, 1978). Since then, a large number of studies have confirmed the positive correlation between noise sensitivity and annoyance. Nevertheless, recent research also showed that one's personality has an independent effect on noise sensitivity (Shepherd et al., 2015), which suggests there is more beyond noise sensitivity when it comes to a person's general attitude towards sound. With the previous section talking about the interaction between visual information and sound, it provokes the question whether there is a personal trait that reflects the reaction towards audiovisual stimuli, including attitude, sensory and attention focusing capabilities.

For audiovisual stimuli, earlier research has shown the benefit of vision in understanding speech (Musacchia et al., 2007). By contrast, it has also been shown that in situations of uncertainty, e.g. in a bimodal-inducer (auditory and visual) situation, when the inducers conflicted temporally, observers tend to follow the more reliable auditory cue (Apthorp et al., 2013). Some research has shown that older and younger persons obtained similar performance with purely auditory stimuli, but older adults have poor performance with audiovisual modality (Sommers et al., 2005). This again confirms the interaction effect and also the question whether an individual difference could be linked with audiovisual interaction.

1.2.4 Perception – a holistic approach

In Section 2.1, the difference between instantaneous annoyance and retrospective annoyance has been briefly discussed. Though Section 2 started with a discussion on annoyance, it went beyond to discuss sensory perception in general, since by definition, the soundscape is perceived or experienced and/or understood (ISO, 2014). The usage of the term perception here refers to the appraisal of the environment. While the discussion on noticing sound, saliency, attention, and audiovisual interaction has looked into each aspect separately, it is essential to combine them in a holistic way.

The auditory and visual senses are the major contributors to obtaining information from the surrounding environment (Liu and Kang, 2018). From the discussion above on audiovisual interaction, it can be concluded that (in)congruence between visual and auditory information strongly affects the appraisal of the sonic environment (Viollon et al., 2002). This might explain the stated inconsistencies in the effects of auditory and visual factors on perception, as a single factor might work in different directions to impact perception. Some may argue that sound is the dominating factor. Nevertheless, a view on an urban green area paired with high-level sound exposure does not provoke the common mindset of such environments. It also suggests that for improving the quality of the urban soundscape, one cannot only take into account one single aspect. The traditional approach for tackling noise issues is focused on reducing the noise level, by blocking streets with a noise barrier, etc. This strategy might work to some degree, but clearly better results could be achieved. Many have attempted to approach this issue beyond addressing only auditory or visual aspects, and explored the best combination of audiovisual measures in specific situations (Hong and Jeon, 2014; Liu et al., 2013; Preis et al., 2016).

Previous research has found that the more sound is congruent with expectation, the less is the evoked annoyance and, conversely, the more is its acceptability (Brambilla and Maffei, 2006). The phrase "expectation" used here refers to what a person expects to encounter in a certain place. A plausible basis for expectation for the soundscape of a location is the concept of soundscape "competence" proposed by Truax (2001), which is related to an individual's experiences. People expect certain types of sound to be present in a particular space. Earlier research found that the perception of the sonic environment, both real and simulated, is affected by expectation in several different ways (Bruce and Davies, 2014). Note that there is a difference between the expectation of particular sound sources and the expectation of the soundscape as a whole; the latter was found to be driven significantly by prior experience of similar spaces and also by perceived loudness (Bruce and Davies, 2014). Furthermore, expectations might also depend on many social and economic factors and are very difficult to predict, especially within a universal model (Botteldooren et al, 2001; Zhang and Kang, 2007).

The congruency of the audio-visual environment and the expectation are reminiscent of the fact that the soundscape definition contains “in context” (ISO, 2014), which suggests that specific methods for soundscape quality optimization apply to specific situations. However, the study on this topic should go deeper, looking for the commonness and individuality, the underlying mechanisms, and expected effects. It should encourage urban designers and urban planners to obtain a full understanding of the situation before taking actions: the components of the sonic environment, the function of the urban space, its users, society backgrounds, etc. (Kang et al., 2016). Conversely, starting from the expected optimal soundscape, what could be done? This echoes the top-down and bottom-up approaches for obtaining solutions to a problem.

As a final note, one should bear in mind that perception is not restricted to saliency, attention and audiovisual interaction. Other factors, such as other sensory context (odor, heat and humidity), weather, climate, etc. should also be considered in the future. Thus, the holistic approach to soundscape should be an evolving concept.

1.3 Soundscape collection and classification

1.3.1 Soundscape collection

Since urban soundscape studies have received more and more attention during the last decades, researchers have encountered a tremendous amount of soundscape examples worldwide. Soundwalks are often used as a methodology for soundscape evaluation, in which participants are physically in a specific location carefully chosen by the researchers (Semidor, 2006). Soundwalks are a practice that was devised by Schafer, when he established the World Soundscape Project at Simon Fraser University during the late 1960s and early 1970s (Schafer, 1969). It is an empirical method for identifying a soundscape and components of a soundscape in various locations (Adams et al., 2008). In a soundwalk procedure, participants are asked to evaluate their subjective perception on a given scale during or after being exposed to the soundscape (Westerkamp, 1974). Often a sematic questionnaire is used, with questions on the sonic or total environment (Kang and Zhang, 2010). Such an approach collects the perception of existing urban spaces to a very high degree, with participants being physically exposed in the environment. Nevertheless, certain drawbacks were also found with this approach. First, organizing a soundwalk is a costly procedure and sometimes the procedure is evenly spread across seasons (Yang and Kang, 2005). Second, though the locations might have been carefully selected, the actual situation during the soundwalk is still unpredictable. Third, attention is explicitly focused on the sonic environment during soundwalks and although the whole physical context is ecologically valid (participants are in the real space), the activity of the person and its natural flow of attention is not ecologically valid. Thus, a lab reproduction method might be a valuable alternative.

In a laboratory listening experiment, participants are typically presented with previously recorded audio stimuli (Jennings and Cain, 2013). Visual stimuli and different audio-visual stimuli combinations are also often presented (Carles et al., 1992), to investigate audiovisual interactions as stated in previous section. The stimuli often contain various combinations of different auditory and visual cues, and their duration varies from seconds (You et al., 2010; Lavandier and Defréville, 2006) to minutes (Payne, 2013), where researchers believe it is sufficient for the purpose of the study at hand. With stimuli of a short duration, attention will be largely focused on the environment and one may need longer exposure and distracting activities to increase ecological validity. Thus, part of the bottom-up mechanisms governed by saliency might not occur in those situations. Nevertheless, in a lab experiment, the visual stimulus is often presented in a two-dimensional form (e.g. on a screen or projected on a wall) with a limited (cropped) view; and participants are in an artificial lab environment after all. Compared to the soundwalk method, this provides less visual information and physical immersion to the participants.

As context is an important part of the soundscape and the visual setting is an important cue for context, examples of acoustic environments should be embedded in accurate 360-degree visualization. Immersive virtual reality could also become a valuable tool for interactive participatory evaluation of the soundscape in urban planning and design projects (Puyana-Romero et al., 2017), as virtual reality reproduction systems are rapidly becoming affordable and widely available. To date, however, no unique protocol or standards exist for immersive audio-visual recording and playback of urban environments with soundscape in mind (Hong et al., 2017). Standardization efforts with regards to spatial audio recording have been started recently by ISO (2018). Hence, developing a database of high-quality immersive recordings of existing spaces and a unified lab playback system are highly valuable, which might then serve as an ecologically valid baseline for studying the perceptual outcome of noise control and soundscape measures.

1.3.2 Soundscape classification

Instead of leaving the many soundscape examples pale and disorganized, it is a challenge to sort them into groups based on shared traits. Soundscape classification based on perception, for instance, has been tried by many researchers. When asked to describe the urban acoustic environment, persons tend to name audible sounds and their sources and may relate the quality of the environment to the meaning given to these sounds (Dubois et al., 2006). In this procedure, sound sources, sound descriptors and soundscape descriptors are undeniably influenced by cognitive effects (Davies et al., 2013). Classification schemes based on urban sound source sorting have been proposed (Brown et al., 2011), which does not capture the influence of the composition as a whole on persons and therefore should be complemented by more holistic indicators.

Holistic descriptors that have been proposed previously and that could be used for classification include: pleasantness, music-likeness, restorativeness, appropriateness. (Aletta et al., 2016a; Botteldooren et al., 2006). A lot of research has focused on the soundscape descriptors inspired by emotion-denoting adjectives (Aletta et al., 2016a). The well-known circumplex model of affect (Russell, 1980) identifies eight affective concepts that can be mapped to a two-dimensional plane. Previous research (Axelsson et al., 2010) translated core affect to the physical environment that causes it and showed that outdoor soundscape quality may be represented by two main orthogonal components: pleasantness and eventfulness. This assessment and classification framework has been applied in many studies and has proven to be rather representative from the perspective of emotion. However, doubts and critiques towards this assessment arise as well. Regarding the core affect model itself, research has identified a main problem with the two-dimensional approach offered by Russell: a variety of overlapping emotional concepts can be placed in the same quadrant of the model (e.g., Ekkekakis, 2008). It has been argued that a representative soundscape for one quadrant label in the 2D core affect model seems rare (Axelsson, 2009). Some may even argue that a perception from an individual shall not be measured by emotion only because multiple effects are included, e.g. noticing the sound environment.

Other classification methods that are not perception dependent have also been implemented. Rychtáriková and Vermeir (2013) sorted the soundscapes in urban public places into 20 categories based on their shape formed by 13 acoustical parameters. Note that this is not in line with the current understanding of soundscape, since soundscape, according to the definition, includes perception. Though it seems rather redundant and restricted from a sonic perspective, it provides an alternative thinking of soundscape classification that is based on objective acoustical parameters. Moreover, Torija et al. (2014) proposed an automatic soundscape classification model based on acoustical as well as perceptual criteria, pushing soundscape classification to the next level.

1.4 Outline of the dissertation

This dissertation is structured into three main parts. An overview of each part is given in the following subsections.

1.4.1 Individual differences modify the effect of visual information on noise annoyance

Noise annoyance, especially in and around the dwelling, has been investigated thoroughly in recent decades as it is one of the most prominent effects of traffic noise exposure (Ouis, 2001; De Coensel et al., 2007). Still, the influence of visual factors on sound perception is not completely understood, especially in the at-home situation. Audiovisual attention focusing and gating are expected to play a role at the perceptual stage. This would also imply the existence of inter-person differences in exposure-effect relationships beyond known factors such as noise sensitivity.

To explore these hypotheses, Chapter 2 describes a noise annoyance experiment conducted in a mockup living room. The noise annoyance experiment involved 16 audiovisual stimuli, which were a combination of 4 window-view video sceneries and 4 sound fragments, to investigate the relative importance of sound source visibility and green elements visibility. In this setting, it was found that (1) sound source visibility, as a functional parameter of the visual setting, has more impact on self-reported noise annoyance than the green element's visibility which describes the quality of the visual; (2) self-reported noise sensitivity remains the strongest personal factor, yet persons being easily distracted by visual elements report significantly lower noise annoyance at the same exposure level; (3) two significant interactions can be observed in the prediction of self-reported noise annoyance: (a) noise sensitivity interacts with sound source visibility; (b) vision dominance, as a personal factor, interacts with the visibility of green elements. The interaction between these factors provides additional evidence to support the role of audiovisual attention in the emergence of noise annoyance.

1.4.2 Individual differences and the concept of audiovisual aptitude

Chapter 3 further explores the individual differences that are found in the experiment discussed in Chapter 2, which might have an effect on annoyance and an interaction with auditory/visual factors. An interaction between audition and vision in the appraisal of the stimuli used in the living environment has been found, and this interaction was found to be influenced by personal factors. In Chapter 3, an auditory deviant detection experiment in an ecologically valid and complex context is described, which allows us to (1) distinguish between accurate and less accurate listeners; and (2) distinguish between participants that are easily visually distracted and those that are not. To conclude, this individual difference reflects the different attitude and reaction towards audiovisual stimuli, which can be labeled as "audiovisual aptitude". Chapter 3 explores the relationship between this factor and demographic information. Also, this individual difference is found to be aligned with many well-known psychology concepts and effects, such as the Colavita effect (Colavita, 1974), inattentional blindness (Simons and Chabris, 1999) and inattentional deafness (Macdonald and Lavie, 2011).

To further analyze the effect of this personal factor, two previously conducted laboratory experiments were re-analyzed. One is focusing on the perceived pleasantness of using outdoor public spaces in a Virtual Reality environment. It was found that the overall appraisal of walking across a bridge is influenced by audiovisual aptitude, in particular when a visually intrusive noise barrier is used to reduce highway traffic noise levels. Another one is the experiment in Chapter 2, where it was found that the influence of visibility of vegetation on self-reported noise annoyance was modified by audiovisual aptitude. Therefore, Chapter 2 and 3 should be viewed together.

1.4.3 Soundscape collection and soundscape classification in (visual) context

It is increasingly acknowledged by (landscape) architects and urban planners that the soundscape contributes significantly to the perception of urban public open spaces. In Chapter 4, first, a soundscape recording and playback system is presented, which combines a 360-degree view camera with ambisonics/binaural recording, and a virtual reality headset and headphone for playback. A first experiment was conducted to validate this method and it was perceived as ecologically valid in terms of realism and immersion. Second, a hierarchical method for soundscape classification that is based on the contribution of soundscape to the perception of the total environment is proposed. This method distinguishes between backgrounded and foregrounded soundscapes, disruptive and supportive soundscapes, and finally calming and stimulating soundscapes. To validate this classification scheme, a second experiment was conducted with a set of immersive audio-visual recordings recorded worldwide as stimuli. This alternative classification method was then compared to the 2D core affect model (Axelsson et al., 2010), and well-separated classes were found. Finally, a set of models based on a limited number of acoustical indicators are constructed that could correctly classify a soundscape in the proposed classification scheme.

This third chapter concerns soundscape collection and classification in a holistic manner. Compared to previous research, immersive stimuli that are more close to the real environment are used. The alternative classification method that is proposed is less dependent on emotion as compared to existing methods, but rather depends on the contribution of the soundscape to the overall perception of the environment. Other researchers are encouraged to perfect this approach, and to contribute to the database of soundscape recordings collected worldwide.

2

Effect of interaction between attention focusing capability and visual factors on road traffic noise annoyance

Sun K, De Coensel B, Echevarria Sanchez GM, Van Renterghem T, and Botteldooren D

Published in *Applied Acoustics*, 134, 16-24, 2018.

This chapter presents two laboratory experiments: an experiment in a mockup living room to explore the effect of view from the window on annoyance at home, and a listening experiment to test attention focusing capability as a personal factor. It is found that this personal factor is comparable to noise sensitivity regarding the size of its effect on perception, and that it interacts with visual factors. Part of this research was presented at the 2016 Internoise conference in Hamburg (Sun et al., 2016).

2.1 Introduction

In recent decades, the relationship between noise exposure and annoyance, especially in and around the dwelling, has been explored in depth (Ouis, 2001; Sato et al.,1999). Hence, noise annoyance has now been recognized by the World Health Organization as the strongest and best proven effect of environmental noise on people. For the European Union's noise indicator, L_{den}, exposure effect relationships have been derived (Miedema and Oudshoorn, 2001). It has also been shown that noise annoyance could be an indicator for effects of noise on health and well-being (Niemann et al., 2006; Ndrepepa and Twardella, 2011; Honold et al., 2012). The determinants of annoyance were investigated in related studies leading to complex models (Fyhri and Klæboe, 2009; Botteldooren et al., 2002). Epidemiological research has indeed shown that not only the average sound level influences annoyance, but also personal factors modify the exposure effect relationship (such as age, gender, education and noise sensitivity, as well as other environmental factors (Fields, 1993; Guski, 1999; Miedema and Vos, 1999)). In particular, subjective noise sensitivity was shown to be a very stable personality trait which is determined both by inheritance and experience (Öhrström et al., 1988; Västfjäll, 2002; Schreckenberg et al., 2010; Van Kamp et al., 2004; Heinonen-Guzejev et al., 2005).

In environmental noise surveys, the effect of visual elements such as the view from the window on long-term noise annoyance have been addressed before (Van Renterghem and Botteldooren, 2016; Li et al., 2010; Pedersen and Persson Waye, 2007; Aletta et al., 2016b), yet less frequently than other contextual factors. Audiovisual interactions in combination with noise annoyance in and around the dwelling is a multifaceted effect that is not easy to grasp. In experimental work related to urban environments, the congruence between visual and sound information was strongly affecting the appraisal of the sonic environment, in terms of visual influence (Viollon et al., 2002). Although congruence may also play a role in occurrence of annoyance in and around the dwelling (Hong and Jeon, 2015), more basic aspects of the audiovisual experience have been suggested, such as visibility of sound source (Pedersen and Larsman, 2008). Some studies pointed out that seeing the sound source would increase subjective annoyance (Zhang et al., 2003), others found that visually screened traffic was perceived as more noisy (Watts et al., 1999; Maffei et al., 2013). In addition, the general quality of the visual setting and more particularly, the visibility of green elements was shown to have a direct influence. Visually attractive and green noise barriers tend to be more efficient in reducing noise annoyance (Hong and Jeon, 2014). Recent research (Preis et al., 2016) has nevertheless confirmed the complexity of the audiovisual interaction: in a lab experiment, adding visual information to a listening experiment tended to reduce annoyance if the sound source was believed to have a positive influence, while annoyance increased for mechanical sound sources.

Psychophysical knowledge may help understanding the complex influence of visual information on perceived noise annoyance in and around the dwelling. Prior research has shown that noticing sounds can be regarded as a precursor for noise annoyance (De Coensel et al., 2009). In this view, sounds that attract more attention would more likely cause annoyance. Audiovisual stimuli, which are irrelevant for the tasks a person is involved in, may capture involuntary attention, a process where sensory modalities interact at different levels in the brain (Koelewijn et al., 2010). This could lead to an increase in annoyance for visible sources. In addition, individual differences in the capability of focusing attention has recently been shown to affect the cocktail party effect (Oberfeld and Klöckner-Nowotny, 2016). Distractibility may be a personality trait that can be defined also in the healthy population (Forster and Lavie, 2016). Hence, it seems useful to study whether distractibility could be a personal factor affecting the influence of the visual scene on noise annoyance or even the emergence of noise annoyance itself.

It should be noted, however, that occasional attention saccades to environmental factors not only cause increased noticing and therefore possible annoyance. Attention restoration theory predicts that such attention switches may enhance restoration and therefore would not be appraised as annoying (Kaplan, 1995; Raanaas et al., 2011). A better understanding of audiovisual interactions in perception of the environment may lead to better urban planning and soundscape design (Hao et al., 2015).

In this article, an experimental study is described that aims at confirming the hypothesis on the mechanisms underlying the effect of the view from the window on noise annoyance. In addition, the experiment aims at identifying subjective noise sensitivity and distractibility as personal factors influencing this effect. To be able to go beyond questionnaires for assessing personal factors, we opted for a lab study using well controlled stimuli. Assessing noise annoyance in an ecologically valid way in an experimental setup is rather difficult as the main hidden factor under investigation, i.e. non-voluntary attention, is replaced by focused attention in a listening experiment. For this reason, two specific requirements were introduced in the experimental design. Firstly, the exposure time for each stimulus was 10 minutes and participants were instructed to engage in some light activity during the experiment in order not to focus on the sound. Earlier studies (De Coensel et al., 2007; Van Renterghem et al., 2013) have shown that this protocol is valid. Secondly, since the target of this study is the effect of the view from the window, direct comparison between different visual stimuli is avoided by showing the visual stimulus in a natural setting, a mockup window, and by presenting the different visual stimuli on different days. The additional distractibility experiment is conducted at the very end not to reveal the focus on visual information.

2.2 Methodology

2.2.1 Overview

The first part of this study is a road traffic noise annoyance experiment conducted in conditions that should resemble the everyday living context as closely as possible. Participants were exposed to 16 audiovisual stimuli (Figure 1) during 4 separate experimental days in the same mockup living room. At each experimental day, the view from the window was fixed and the audio fragments varied. The participants were led to believe this experiment was about rating the perceived annoyance of 16 environmental sound conditions in a living room. Each audiovisual stimulus was played for 10 minutes, in order to give participants enough time to engage in some light activity and to adapt to the living room environment. After the presentation of each audiovisual stimulus, they were asked to rate their perceived noise annoyance during the past 10 minutes on an 11-point scale (from 'Not at all' (0) to 'Very much' (10) annoyed) (ISO, 2003).

Since detecting the effects of visual factors on sound perception was the objective of this study, all other factors were carefully controlled in order to eliminate their impact on sound perception as much as possible. For example, during each experimental day, participants were asked to sit in the same seat in the mockup living room, which gave them the same perspective to all scenes. It was also assured that the room setup, the lighting, and the room ventilation remained unchanged. The acoustic playback level was controlled by measuring the sound level in the center of the room. Participants were also asked to refrain from drinking alcohol or unusual amounts of coffee or taking medical drugs before the experiment. In addition, it was asked not to listen to loud music while waiting to participate in the experiment.

The design of the experiment assumes that the auditory memory of participants was erased in between experimental days. However, there may still be a degree of habituation to the experimental setup. Therefore the order of presentation of the 4 visual settings during 4 days was randomized between participants.

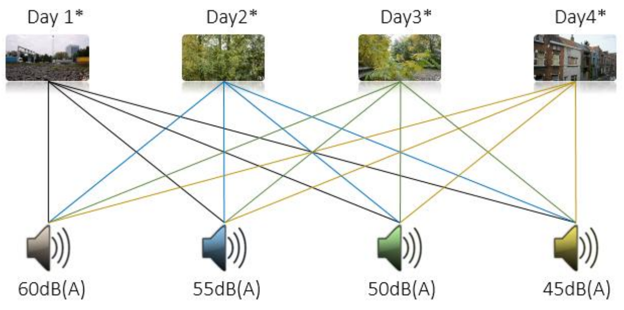


Figure 1 – 16 audiovisual stimuli (combination of 4 sound fragments and 4 window-view sceneries). (*The order of experimental days was randomized).

The second part of the experiment was only conducted the fourth day, after the regular test was completed. It consisted of a listening task focused on detecting deviant auditory scenes. This was to avoid impact on the subsequent days. The second part also included the short version of the noise sensitivity questionnaire proposed by Weinstein (Weinstein, 1978).

2.2.2 Mockup living room

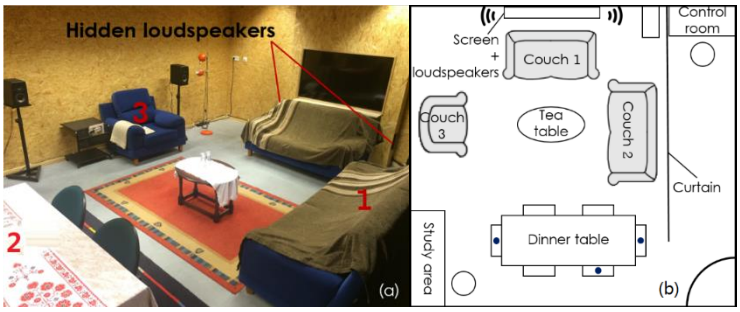


Figure 2 – Layout of the mockup living room: (a) photograph; (b) schematic drawing (not true to scale).

The mockup living room was arranged as shown in Figure 2. A 60-in. television screen, projecting window-view videos, was fixed in a specially-made cabinet integrating it in the wall and making it resemble a window. Two loudspeakers were hidden in the cabinet to make the sound appear to come from the window. Note that the loudspeakers visible in Fig.2a were not used in this experiment. The control room is positioned in the corner, separated from the living room by a large thick curtain. A subwoofer is also positioned next to the control room, which ensures that low frequency sound is reproduced realistically.

As shown in Fig.2a, three sitting positions were marked in this room. Participants were suggested only to sit in these preselected seats, which gives them certain perspectives to the mock-up window (obviously, they are not being told that this was the reason).

2.2.3 Audiovisual stimuli

2.2.3.1 Window-view video sceneries

The four videos contained a mixture of different natural and man-made landscape elements. Four screenshots of the videos (all taken near the city of Ghent, Belgium) are shown in Figure 3. Scene (a) provides an open view of highway traffic and contains very few green elements; (b) allows vision on some parts of the highway through the woods; (c) contains a totally green visual setting; and (d) shows a row of houses along a non-busy street, hiding a highway from sight. The sound source was completely visible in scenery (a) and partly visible in scenery (b), while in (c) and (d) no sound source was visible. On the other hand, scenery (b) and (c) contained dominant natural elements, whereas scenery (a) and (d) contained mostly man-made elements.

Video (a) has been synchronized to the audio, video (b) is not but the highway view is rather limited so that individual – possibly loud – vehicles cannot be detected anyhow. For the last two video’s, synchronization is not relevant.

Figure 3 – The four window-view sceneries used in the experiment.

2.2.3.2 Audio fragments

Four audio fragments with different sound level are created by simulating the effect of a change in the window acoustic insulation. The original traffic noise audio fragment was recorded simultaneously with the video recording at the location of scene (a) (see Figure 3) with a B-field microphone, in a four-channel B-format. This audio recording was then transformed into a two-channel format using VVMic (Visual Virtual Microphone) 3.4. Two channels played back near the left and right of the window can still give a sense of movement of individual cars. By playing the sound from the loudspeakers behind the television screen/window, the sound spatialization of a common living room is achieved. This recording will represent the open-window sound exposure for the participants.

When presenting audiovisual information to the listener, it is important that the auditory and visual cues on source distance are congruent. Hence we opted for noise mitigation through window insulation to mimic sound level variation in this study, as this would keep the spectro-temporal variation of the traffic sound consistent with the visual distance. In addition, this gave a plausible reason to the participants why different noise levels had to be evaluated. According to the work of Tadeu and Mateus (Tadeu and Mateus, 2001), three transmission loss curves were selected to represent a (closed) single glazed, a double glazed and a triple glazed window (specific choices: ‘single layer 8mm’, ‘double 8+4, d=10mm’, ‘triple 8+4+4, d1=100, d2=50’). The original audio recording was filtered accordingly using Sony Soundforge software to mimic the different closed window acoustic insulation spectra as shown in Figure 4.

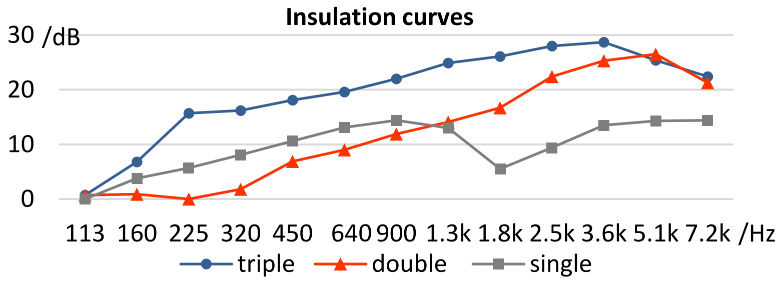


Figure 4 – Frequency attenuation (insulation curve after calculation).

By fixing the volume of the audio card of the playback PC, the media player software and the amplifier of the loudspeakers, the overall exposure sound level of the original audio fragment is settled at an equivalent sound pressure level of 60dB(A) (in the center of the room) for the assumed open window sound exposure. The overall presentation sound level for the single, double, and triple glazed is reduced towards 55dB(A), 50dB(A) and 45dB(A), respectively, to make sure a clear level difference would be detected.

Participants were told that these sounds correspond to four different window insulations. It is assumed that this method of presentation ensures that it does not direct a participant’s attention to differences in the view from the window. As the difference between the sounds is in fact not the main target of the investigation, the above procedure for generating the different sound excerpts only needs to suggest ecological validity so a more advanced calibration of the room response is not essential.

2.2.4 Course of the experiment

It was already mentioned that the order of presentation of the visual context should be randomized to avoid bias by habituation to the experimental conditions during the subsequent sessions. In addition, within one experimental session, the 4 sound environments are also presented in random order to decrease the bias that might be caused by the previous sound experience. There are A_4^4=24 possibilities for the order of video presentation over the four experiment days, and an equal number of 24 possibilities for the order of audio fragment presentation during each experimental day. To prevent large level differences between subsequent tests, the maximum change in sound level between subsequent fragments was limited to 10 dB(A). This reduced the number of possible sound presentation orders to 12. The sound order randomization is applied after the videos have been assigned randomly between experimental days by adhering to the following rules: each scene should be coupled two times with all 12 sound orders, and over all experiment days, all four scenes should have a different audio fragment order. This randomization ensures that all possibilities are covered, and is expected to eliminate any impact of order of presentation on the results.

Participants were told that the experiment is designed to study their disturbance by road traffic noise in a living room environment. All they had to do was relaxing as if they were in their own living room. They were allowed to read a book, browse a magazine, have some drinks, play with their phone to some extent, or even chat with the other participants. However, activities that require a high level of concentration, such as bringing work-related documents, was forbidden. This setting (1) is close to real life; and (2) prevents that participants would focus too much on listening to the sound. Note that although activity disturbance may be a cause of annoyance, this experiment was not designed to assess activity disturbance itself. This would require a more stringent task design and a different range of sound exposures.

In between the 10-minutes lasting exposures, there was a one minute break, during which every participant was asked a single question: ‘Thinking about the last 10 minutes staying in this living room, which number from 0 to 10 best shows how much you were annoyed or not annoyed by the traffic noise?’ (ISO, 2003).

2.2.5 Audiovisual aptitude and noise sensitivity assessment

It is known that the response to a retrospective annoyance question is only partly determined by the equivalent noise level. Individual differences in response have been related to human factors such as gender, age and noise sensitivity. As this research is focusing on the effect of the view from the window on reported noise annoyance, an additional personal factor labeled “audiovisual aptitude” is added. This factor measures how strongly the visual context influences the ability of a person to detect differences in the auditory scene and remember them. Section 3 will elaborate on the possible perceptual and psychological phenomena that could underlay this new factor. To measure “audiovisual aptitude”, at the end of the 4th day of the above-described experiment, a second experiment is conducted. It contains four audiovisual scenarios, in which either the audio or visual parts was altered in a subtle way (Sun et al., 2016). The experimental design consists of a deviant detection task where three alternatives are presented once for each trial. The deviant has to be detected when only sounds are presented and when sounds are presented in the presence of a visual distractor. This ecologically valid alternative to basic psychological stimuli is intended to investigate whether a person is more vision or audition oriented but also measures its sensitivity to inattentional deafness (Macdonald and Lavie, 2011).

On the outcome of this experiment, two classification principles are applied: auditory resolution and visual distractibility. Auditory resolution distinguished between persons that make no errors on the blind listening test, i.e. they detect the deviant in each of the four cases. This allows to distinguish the careful listeners with good auditory memory that are able to detect even the smallest change. Visual distractibility distinguishes between the persons that do well on the blind listening test but get misled by the incongruent visual information and make at least one error in deviant detection in this case. In other words this group gets misled by the visual information. Hereby, two human factors arise: auditory acuity and vision dominance (Giard and Peronnet, 1999). More information on this experiment can be found in (Sun et al., 2016).

Finally, at the end of the complete experiment, after four days, a more elaborate questionnaire was presented to all participants to collect some personal information and more in-depth questions, including age, gender, education level and noise sensitivity, via a widely-used noise sensitivity survey (Weinstein, 1978). In addition, the hearing status of all participants was assessed via pure tone audiometry (PTA) carried out in a quiet but not sound-proof room using a regularly calibrated AC5Clinical Computer Audiometer.

2.3 Results and Analysis

2.3.1 Participants

In total 75 participants conducted this experiment, 6 of them were excluded from the final dataset due to either bad hearing (based on a pure tone audiometric test performed on the 4th day), or not completing the full experiment. Basic demographic information is listed in Table1.

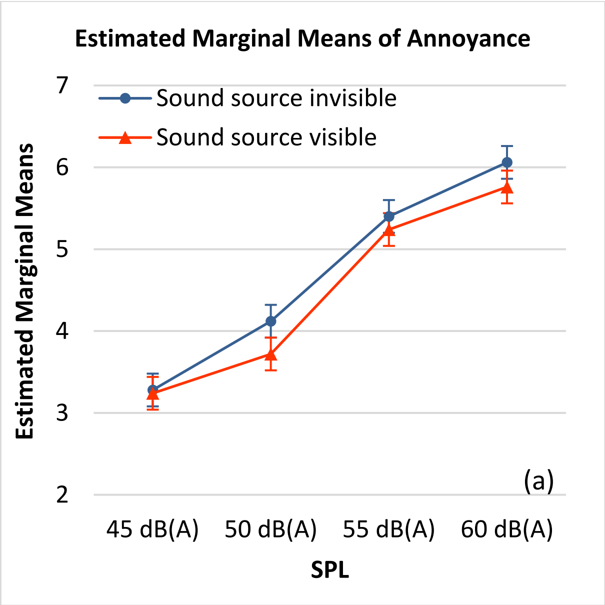
Table 1 – Basic information of 69 participants.

Factors	Categories	Number	Percentage/%
Gender	Female	28	40.6
	Male	41	59.4
Age*	Junior(20~27yrs)	37	53.6
	Senior(28~46yrs)	32	46.4
Education	Below M.S	20	29
	Above M.S	49	71

*The age variation of participants is from 20 to 46 yrs. The average value is 27.9 and the median value is 27.

2.3.2 Visual factors

As described in Section 2.3.1, the content of four window-views can be sorted based on two features: the visibility of sound source and the presence of green elements. In Figure 3, (b) and (c) contain dominating green elements, while (a) and (d) do not. On the other hand, in (a) and (b), the sound source (highway traffic) is visible, while in (c) and (d), it is not. Figure 5 indicates the difference of estimated marginal means of annoyance based on these two features.



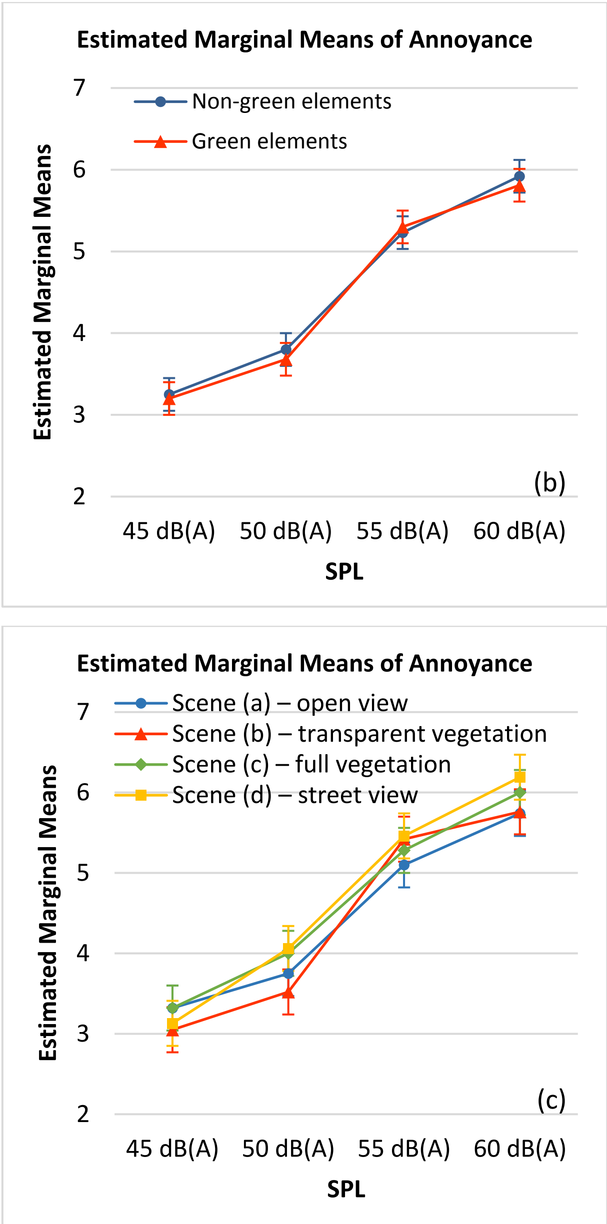


Figure 5 – The annoyance difference of (a) visibility of sound source, (b) visibility of green elements and (c) four window-view scenes.

(The error bars represent the standard errors on the averages: +/- 1SE)

As shown in Fig.5a, average reported noise annoyance increases strongly with the noise level, including the change in spectrum caused by the window insulation. It should be noted that different sound environments were presented during the same day and thus a direct comparison was possible. When looking at the influence of the window view, participants tend to be less annoyed when the sound source was visible (scenery (a) and (b)). Meanwhile, there is also a larger jump between low SPL and high SPL in this category (red line in Fig.5a). When the sound source is visible, people's annoyance tends to be divided into two stages for either low and high levels. At both the low and high levels, the annoyance increases with SPL are not as fast as when the sound source is invisible. Nevertheless, the annoyance-SPL regression tends to be more linear when the sound source is invisible. Visible green elements do not seem to have a large influence (Fig.5b) in this overall analysis.

As all experimental conditions have been assessed by each participant in the study, and personal factors are assumed to have a significant effect on the self-reported annoyance rating, two level statistics treating person as a random variable is appropriate. The different sound environments are characterized by their A-weighted sound level, but also differ in spectral characteristics. Therefore, SPL is treated as an ordinal variable for the exposure condition rather than as a continuous variable.

A mixed factor generalized linear model fit is applied, using participant as a random factor to generalize these results. This model considers only the sound (SPL) and the visual factor(s). For visual factor(s), it is tested with only the 4 views (sceneries) or with green elements visibility and sound source visibility as descriptor of the window view. Besides, it is also tested to add the interaction between the sound and the visual factor(s) and to remove the insignificant factor(s). The best model (with the lowest information criterion) from the above-mentioned ones is listed in Table 2. The effect of sound source visibility on reported noise annoyance is statistically significant while the visibility of green elements is not. Also, none of the interactions between sound and visual factor(s) has statistical significance. However, as shown in Fig.5a, the relatively small difference between lines and the overlapping of standard error bars suggests that the significance of sound source visibility will be less pronounced as stronger factors get involved in the model.

Table 2 – Generalized linear model 1.

Fixed Effects	Target: Annoyance			
Source	F	df1	df2	Sig.
Intercept	178.129	4	1.099	.000
Sound source	7.493	1	1.099	.006
SPL	235.008	3	1.099	.000
'Participant' is used as random factor.				

2.3.3 Human factors

A frequently mentioned personal factor, noise sensitivity, is investigated in this study. The post-processing divided participants into two groups based on the neutral score, i.e. choosing the neutral answer for each single question in Weinstein's questionnaire (Kishikawa et al., 2006). In total, 57 participants obtained a score higher than the neutral score, which leads them to be marked as being highly sensitive to noise, whereas all others are categorized as having low noise sensitivity. As shown in Figure 6, people with high sensitivity are clearly much more annoyed than people with low sensitivity.

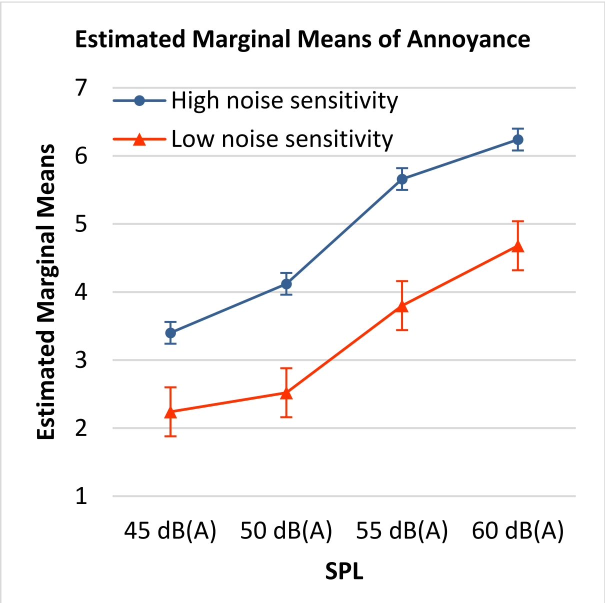
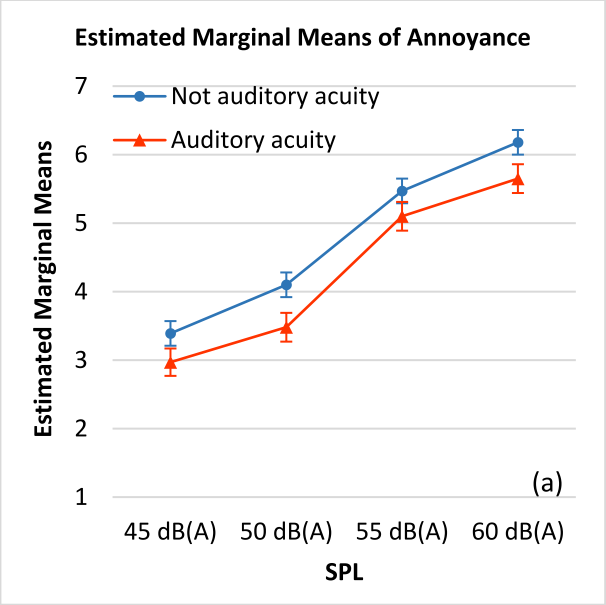


Figure 6 – Dependence of reported noise annoyance on exposure for different sensitivity categories.
(The error bars represent the standard errors on the averages: +/- 1SE)

As mentioned in Section 2.5, participants are clustered according to their audiovisual aptitude along two dimensions: auditory acuity and being vision dominated. Fig.7a shows that participants with good auditory acuity (30 participants) are less annoyed than others. The second factor selects the group labeled vision dominated (13 participants). They have good auditory acuity but are easily distracted by incongruent visual stimuli. These vision dominated participants are notably less annoyed than the other 56 participants, as shown in Fig.7b.



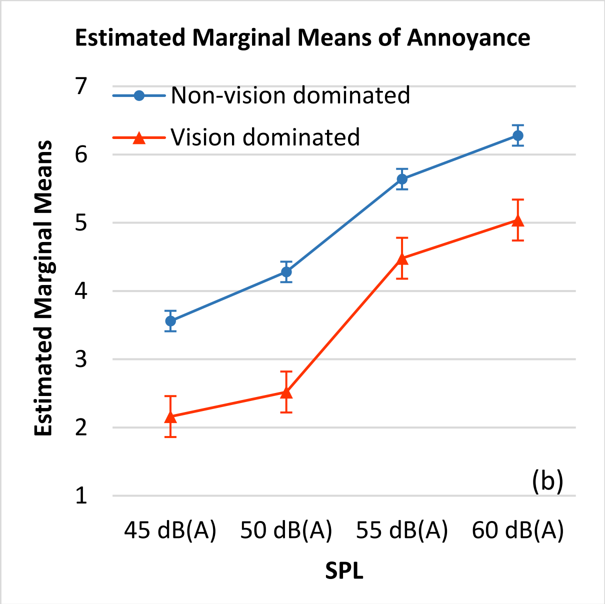


Figure 7 – Reported noise annoyance as a function of exposure differentiated according to (a) auditory acuity and (b) being vision dominated.
(The error bars represent the standard errors on the averages: +/- 1SE)

To test the significance of these human factors, a generalized linear model focusing on the human factors is constructed. Still, participant is used as a random factor to generalize the current results. For visual factor(s) in this model, it is tested with only the 4 views (sceneries) or with green elements visibility and sound source visibility. Similar to model 1, it is also tested to remove the insignificant factors. The best model (with the lowest information criterion) is shown in Table 3. As can be seen, sensitivity and being vision dominated are statistically significant whereas auditory acuity is not. This indicates (1) the importance of noise sensitivity as a human factor; (2) the limitation of auditory acuity by purely focusing on auditory resolution; and (3) the potential influence of being vision dominated on perception.

Table 3 – Generalized linear model 2.

Fixed Effects	Target: Annoyance			
Source	F	df1	df2	Sig.
Intercept	66.779	11	1.091	.000
Gender	2.374	1	1.091	.124
Education level	0.901	1	1.091	.343
Age	2.791	1	1.091	.095
Sensitivity	5.803	1	1.091	.016
Auditory acuity	0.019	1	1.091	.889
Vision dominated	4.021	1	1.091	.045
SPL	234.860	3	1.091	.000
Green	0.349	1	1.091	.555
Sound source	7.488	1	1.091	.006
’Participant’ is used as random factor.				

2.3.4 Interaction between personal factors and window view

In the generalized linear models derived above (Table 2 and Table 3), personal factors and window view are treated as independent factors. The goal of this study is nevertheless to detect the personal factors that can affect the influence of window view on perceived noise annoyance. Therefore, a generalized linear model is fitted that includes interactions, especially interactions between above mentioned human factors and visual factors.

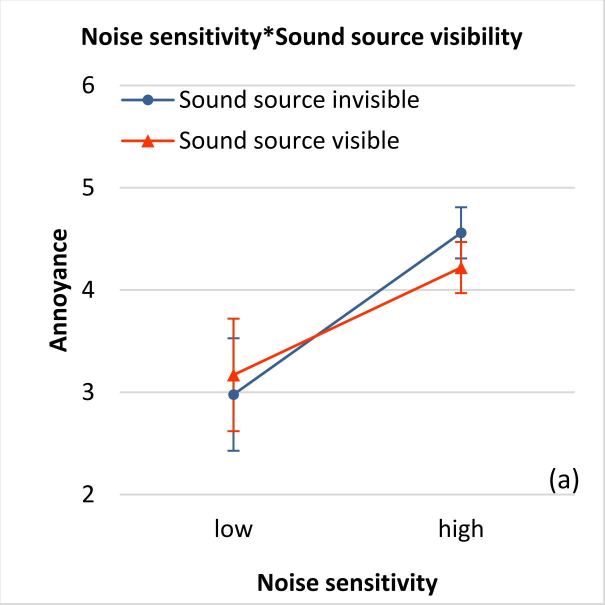
Table 4 shows all the variables mentioned in this study. Individually, many of them showed statistical significance in models for noise annoyance. However, since more variables are involved, some of them are no longer statistically significant due to the strong effect of the interactions. In the human factors category, sensitivity and being vision dominated remain influential factors. On the other hand, descriptors of the view from the window are no longer statistically significant.

Table 4 – Generalized linear model 3.

Fixed Effects	Target: Annoyance			
Source	F	df1	df2	Sig.
Intercept	50.283	15	1.087	.000
Gender	2.438	1	1.087	.119
Education level	0.925	1	1.087	.336
Age	2.866	1	1.087	.091
Sensitivity	5.960	1	1.087	.015
Auditory acuity	0.020	1	1.087	.888
Vision dominated	4.129	1	1.087	.042
SPL	236.894	3	1.087	.000
Green	2.254	1	1.087	.134
Sound source	0.352	1	1.087	.553
Sensitivity*Green	1.610	1	1.087	.205
Sensitivity*Sound source	5.941	1	1.087	.015
Vision dominated *Green	4.894	1	1.087	.027
Vision dominated *Sound source	0.098	1	1.087	.754
’Participant’ is used as random factor.				

The results also involve the interaction between visual factors and two human factors: sensitivity and being vision dominated, which remain statistical significant in the model with interactions. Two out of the four interactions are statistically significant in model 3. The first one is the interaction between noise sensitivity and sound source visibility (Fig.8a). This interaction supports two observations: (1) The dependence of noise annoyance on noise sensitivity increases when the sound source is not visible; (2) For noise sensitive people, sound source visibility decreases annoyance while for noise insensitive people sound source visibility slightly increases annoyance.

The second statistically significant interaction is the one between being vision dominated and green element visibility (Fig.8b). In this study, the visibility of green elements in the window view averaged over all participants does not have a statistically significant influence on reported noise annoyance. For vision dominated persons the visibility of green elements increases noise annoyance. For the remainder of the participants, there is nearly no effect of visibility of green elements in the window view.



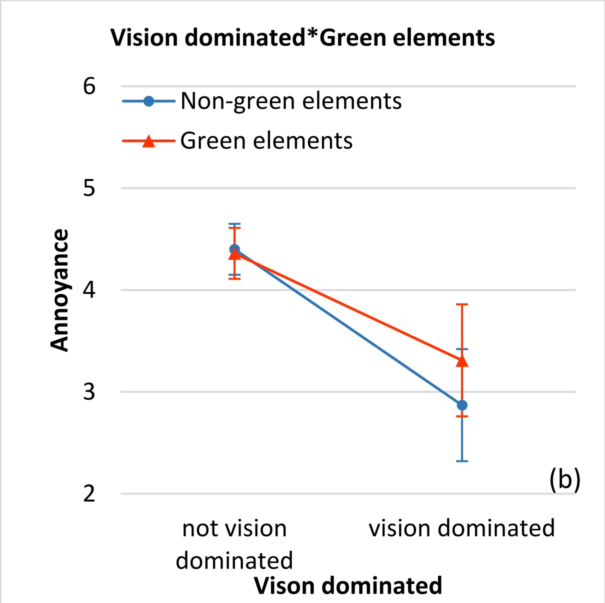


Figure 8 – The interaction between (a) sensitivity and sound source visibility and (b) being vision dominated and green elements visibility.
(The error bars represent the standard errors on the averages: +/- 1SE)

2.4 Discussion

A laboratory experiment was designed to increase our understanding of the mechanisms governing the effect of the window view on perceived noise annoyance. In particular, the experiment aimed at uncovering effects that may occur during processing of audiovisual stimuli. With these goals in mind, the experiment was designed to minimize influences of reasoning and general context setting by the visual elements. Thus, the aim was to avoid that test participants would consider living in a higher quality neighborhood based on the view from the window. Amongst others, the views were therefore chosen not to be particularly attractive gardens or landscapes. As preceding experience and the duration of the tests may influence the annoyance response, auditory stimuli were presented in random order during one test day and visual context was changed in random order between experimental days. The large number of possibilities combined with a limited number of participants resulted in the fact that some particular orders were presented to a single participant only. An ANOVA test checking the influence of stimuli orders showed no statistical significant ($p>0.05$) effect. Therefore, this randomization of the presentation order was shown to have no effect.

This study derived three generalized linear models, considering visual factors, human factors and interactions in addition to sound as independent variables. The information criterion, estimators of the relative quality of statistical models, of these three models are shown in Table 5. A lower information criterion value indicates a better quality of the model. The first model introduces information on the view from the window. Model 2 shows that adding personal information improves the predictability of reported noise annoyance. Finally, model 3 emphasizes that the interaction between these personal factors and the view from the window might explain the inconsistent evidence of the impact of window view on reported noise annoyance.

Table 5 – Comparison between three generalized linear models

		Model 0 (sound only)	Model 1 (visual factors)	Model 2 (person factors)	Model 3 (person-visual interaction)
Information Criterion*	Akaike Corrected	4088	4083	4036	4028
	Bayesian	4103	4098	4051	4043

* Models with smaller information criterion values fit better.

Concerning the direct impact of view from the window (model 1), it was shown that adding the four views separately did not result in any improvement of the model in terms of Akaike information criterion (AIC) or Bayesian information criterion (BIC). Entering the presence of green and the visibility of the source as separate variables resulted in a slight improvement, but only the visibility of the source had an effect. Moreover, adding interaction effects between sound level and window view, which might have been expected on the basis of Figure 5, did not improve the model. Table 2 shows that sound source visibility has statistical significance and thereby confirms previous audiovisual experiments (Preis et al., 2016). Figure 5a further shows that people tend to be less annoyed when the sound source is visible. However, some early research on sound source visibility (Zhang et al., 2003) pointed out that hiding the sound source from sight would reduce annoyance for students in a classroom setting. The current finding is consistent with more recent research (Matsuyama et al., 2014) putting forward the hypothesis that people tend to be more anxious when a moving sound source cannot be seen. Expectation and attention focusing could be a potential explanation for these – at first sight – contradictory findings. In a situation with a sound-irrelevant task requiring high concentration, like for instance following courses in a classroom, the noise distracts attention from the primary task and is against people’s expectations; adding congruent visual information will increase audiovisual saliency and will worsen this situation. In situations where people's attention is mainly led by the noise – as in the current experiment – introducing visual information matches people's expectation and therefore could slightly lower annoyance.

Another conclusion that can be drawn from model 1 is the limited importance of visible green elements (Fig.5b). Yet, visible green typically tends to be positive in many soundscape studies (Li et al., 2010; Gidlöf-Gunnarsson and Öhrström, 2007). Van Renterghem and Botteldooren (2016) pointed out that a green window view significantly reduces self-reported noise annoyance at home, and this effect becomes stronger with an increasing percentage of green elements in the window plane. In real-life settings, a green window view does not only stand on itself, but also delivers information on the general quality of neighborhood or the presence of appealing green areas nearby, both factors that were shown to influence reported noise annoyance. This study, however, was designed not to contain such information, as it is conducted in an underground lab with artificial outside view, and the chosen views accounts for the limited space between the window and a highway. The green scenes in this study essentially hide the source and do not suggest the presence of a park or green area.

Among the human factors introduced in model 2, noise sensitivity has a strong impact, consistent with many studies using the same method of measuring self-reported sensitivity (Okokon et al., 2015). More importantly, the refined assessment of individual audiovisual aptitude gives strong proof of the visual distraction hypothesis. Vision dominated individuals tend to be less annoyed at the same noise level (Figure 7). The personal factor being vision dominated has a high significance in model 2. A small effect of auditory acuity is also seen in Figure 7, but this effect does not statistically significantly contribute to model 2. It is interesting to note that other personal factors like gender, age or education level do not statistically significantly contribute to the model. The effect of these factors may be captured by noise sensitivity and being vision dominated. Additionally, the result also indicates that the methodology of determining these two factors, through audiovisual aptitude investigation, is reliable.

The model with interactions (model 3, Table 4) gives a balanced view on the influence of visual factors, expectations and congruence of audio and visual information. The model improvement caused by adding the interactions exceeds the improvement by adding information on window view without taking personal factors into account. Two interactions are observed. The first statistically significant interaction is between sound source visibility and noise sensitivity (Fig.8a). This interaction indicates, on the one hand, that highly noise sensitive people are notably more annoyed when the sound source is invisible. Scenarios with invisible sound sources do not match the soundscape and this may give highly noise sensitive persons a feeling of insecurity, intensifying noise annoyance. On the other hand, people with low noise sensitivity are less likely to notice the environmental noise. Visible noise sources increase the probability that these persons notice the traffic sound and get annoyed by it. People implicitly express their general attitude towards noise by their sensitivity. High noise sensitivity may also indicate more awareness of the environment in general. They expect the visual to match the audio information. Hence, when the sound source is visible, the satisfaction of getting their expectations fulfilled would decrease annoyance by noise. Finally, it can be noted that this observation also matches the discussion in the previous paragraphs stating that the effect of visibility of the source may depend on the context, where visibility of the source reduces annoyance in a context that stimulates listening. Noise sensitive persons are more likely to be listening.

The second significant interaction is between vision domination and green element visibility (Fig.8b). For non-vision dominated persons, the presence of green in the visual scene does not affect their annoyance rating. Vision dominated persons, however, report higher annoyance when the window view contains the almost impervious green elements as used in the current research. This may imply that these persons are shaping their expectations based on the visual scene rather than to rate noise annoyance based on the noise alone. Interestingly, experimental results involving incongruence of visual and audio information are the direct reasons for these people to be identified as being vision dominated, as described in section 1.4. Furthermore, the larger difference caused by green elements visibility in vision dominated people shows their greater concern about the visual information, compared to non-vision dominated people.

Audiovisual aptitude, the new factor that was shown in these experiments to explain at least partly the variance in effects of window view on self-reported noise annoyance, is a feature that is orthogonal to noise sensitivity. This could be shown by the lack of correlation between these two factors. However, there is also a clear underlying reason for this. According to Soames Job (Job, 1999), noise sensitivity includes factors such as "level of physiological reactivity to stimulation generally; hearing acuity; attitudes to noise in general; beliefs about harmful effects of noise in general; vulnerability caused by stressors other than noise; level of social support and other available coping mechanisms." It is thus a much wider concept than audiovisual aptitude that measures a person's sensory capability of perceiving increasingly subtle elements of the soundscape. Though annoyance is an outcome of many combined mechanisms, the inner willingness to perceive and pay attention to the soundscape seems relatively more important than the capability. The reader should however bear in mind that the similarity between rating scales for sensitivity and annoyance could also reveal an underlying similarity in rating behavior, which is not present in the deviant detection test used to rate audiovisual aptitude.

2.5 Conclusion

In this study an ecologically valid experiment was performed in which a series of audiovisual stimuli were presented in a mock-up living room with the goal to create a better understanding of the influence of window view on reported noise annoyance. Regarding visual factors, sound source visibility was shown to have more impact than green element visibility on self-reported annoyance. Regarding human factors, noise sensitivity was found to have the strongest statistical significant effect on annoyance. A specially designed audiovisual aptitude assessment exposed two reliable human factors, which were shown to explain the large variation in effects of window view on noise annoyance. The results of the experiment validate hypotheses on the role of expectations and multi-sensory attention in perception and appraisal of the sound environment.

Although the noise itself obviously is the dominating factor in the emergence of noise annoyance, it only explains a limited part of the variance. Hence, it is essential to study other factors involved which have the potential for becoming noise mitigation measures. Visibility of the source and a green window view have been mentioned as environmental modifiers of the noise exposure annoyance relationship, yet evidence has been inconclusive. In the present noise annoyance experiment, it was found that the effect of being a vision dominated listener is almost as significant as the effect of noise sensitivity – a known stable personality trait – but more importantly, this personal factor interacts with visual factors. This factor should therefore be considered in future investigations.

A number of limiting factors can be identified with the design of the current experiment. E.g. participants were asked to participate on 4 separate days, with the goal to erase their auditory memory. Still, it is impossible to assure that participants are in the same mood on each of the experimental days. Since this study is on audiovisual perception, one can expect that the mental status and mood of the participants has an effect on the results. Next to this, human factors and visual factors are investigated in this study, yet the acoustical properties of the stimuli are only described in terms of sound pressure level. In many sound quality studies, it has been shown that other features such as frequency and temporal content, sharpness and loudness also change people's preference towards sounds. However, in this study, the precise psychoacoustical characteristics of the sounds were not the essential targets, as the main goal was to study audiovisual interaction.

The visual factors, personal factors and interactions identified in this work help to understand the mechanisms underlying the emergence of noise annoyance. The audiovisual aptitude factor that was introduced in this study could be applied in audiovisual studies as an extended personal factor next to noise sensitivity. The experiment used for assessing audiovisual aptitude is not easily transferrable to field interviews and may benefit from being replaced by more suitable tests or questionnaires for this purpose. The interactions also may have consequences on the design of acoustic and visual elements in urban soundscapes. For this, audiovisual aptitude should be related to demographic variables, lifestyle, and context to allow to identify the most vulnerable groups. Two practical implications of recognizing the existence of a personal factor that affects the influence of visual setting on noise annoyance, could be identified. Firstly, it constitutes a warning that noise annoyance mitigation that would be based on changing visual context may not work for all subpopulations (with different audiovisual aptitude) in the same way. Secondly, urban sound planners may opt for a worst case approach that leads to acceptable perception of the living environment also for the most noise sensitive people and those that are not vision dominated.

3

Personal audiovisual aptitude influences the interaction between landscape and soundscape appraisal

Sun K, Echevarria Sanchez GM, De Coensel B, Van Renterghem T, Talsma D, and Botteldooren D

Published in *Frontiers in Psychology*, 9:780, 2018.

This chapter further explores the personal factor discussed in Chapter 2, and introduces an underlying mechanism labeled "audiovisual aptitude", which distinguishes between persons in terms of listening accuracy and sensitivity to visual distraction. Two previously conducted laboratory experiments are re-analyzed to validate this proposed mechanism. It is concluded that audiovisual aptitude may affect the appraisal of the living environment. This work was carried out in the framework of the SONORUS "Urban Sound Planner" project, supported by the People Programme Marie Curie Actions of the European Union's Seventh Framework Programme. Part of this research was presented at the 2017 Internoise conference in Hong Kong (Sun et al., 2017) and at the 2017 ICBEN conference in Zurich (Botteldooren et al., 2017).

3.1 Introduction

The phrase 'soundscape' used in this study is as defined by International Organization for Standardization (ISO): an "acoustic environment as perceived or experienced and/or understood by a person or people, in context" (ISO, 2014). The subjective appraisal of our living environment is influenced by landscape and soundscape alike. It is well known that these influences are not independent. This interaction partly originates at a low level of auditory and visual perception. In soundscape theory, the importance of visual context on soundscape appraisal has been stressed (Botteldooren et al., 2015; Weinzimmer et al., 2014). Using virtual reality, it was likewise shown that the sonic environment affects overall pleasantness of the public space even when the participants in the experiment focused on visual designs and were kept unaware of the sound (Echevarria Sanchez et al., 2017). In the home environment, it has been shown that vegetation as seen through a window affects the self-reported noise annoyance at home (Van Renterghem and Botteldooren, 2016; Li et al., 2010; Leung et al., 2017). The visibility of a sound source may also affect the awareness of sound. On the one hand, it has been shown that people get more annoyed when the sound source is visible (Zhang et al., 2003), while other studies found that sound is actually less annoying when the source is visible (Maffei et al., 2013). It remains currently unknown what drives these differences. In this paper, we forward the hypothesis that a personal factor or multiple personal factors influence the interaction between landscape and soundscape appraisal. Personal traits and beliefs are known to influence the perception and appraisal of the sonic environment both at home (e.g. noise sensitivity (Miedema and Vos, 2003; Heinonen-Guzejev, 2009)) and in public spaces (e.g. meaning given to tranquility (Filipan et al., 2017) and recreation (Miller et al., 2014; Pilcher et al., 2009)). So it is not unlikely that this additional personal factor would indeed exist.

Previous studies have already shown that considerable individual differences exist in the way humans process audiovisual information, ranging from differences in connectivity between auditory and visual pathways (e.g., (Van den Brink et al., 2013)), to selective preferences in processing auditory or visual material (Giard and Peronnet, 1999). More generally, when engaged in a visual task, participants tend to ignore auditory stimuli, as demonstrated by the well-known Colavita effect (Colavita, 1974). One striking result from many studies on the Colavita effect is that when participants are presented with either auditory or audiovisual stimuli, and have to respond to a change in the auditory stimulus, they usually do so accurately on the auditory-only trials, but fail to detect this change when an audio-visual stimulus is presented to them. A main question is why participants miss such an auditory change.

One possible answer comes from Simons and Chabris, who explored how an unexpected object could go unnoticed during a monitoring task, in a phenomenon they described as inattentional blindness (Simons and Chabris, 1999). Recent research also demonstrates that a single discrete visual distractor can improve the detectability of an unexpected object in an inattentional blindness task (Pammer et al., 2014). Visual distractor processing tends to be more pronounced when the perceptual load of a task is low compared to when it is high (perceptual load theory (Lavie, 1995)). Sandhu and Dyson studied the effect of auditory load on visual distractors and vice versa. They found that in both attend auditory and attend visual conditions, the distractor processing was evident, especially when the distractors were visual (Sandhu and Dyson, 2016). Perceptual load theory has been supported from assessing the impact of perceptual load on the flanker task (Eriksen and Eriksen, 1974), as well as behavioral paradigms, such as negative priming (Lavie and Fox, 2000), implicit learning (Jiang and Chun, 2001) and inattentional blindness (Cartwright-Finch and Lavie, 2007).

A possible explanation for inattentional blindness based on perceptual load theory is that conscious perception of task-irrelevant stimuli critically depends upon the level of task-relevant perceptual load rather than intentions or expectations (Cartwright-Finch and Lavie, 2007). Aging could increase the susceptibility to inattentional blindness (Graham and Burke, 2011). Likewise, individual differences in cognitive ability related to working memory and executive functions affect inattentional blindness (Fougnie and Marois, 2007). Several studies have shown that this phenomenon could be associated with general fluid intelligence (O'Shea and Fieo, 2015) and executive attentional control (Kahneman, 1973). Moreover, an explanation in terms of attention and working memory capacity can explain individual differences in perceiving audiovisual stimuli.

As a counterpart to inattentional blindness, Macdonald and Lavie reported that people could also miss sounds in high-visual-load condition; a phenomenon which they described as "inattentional deafness" (Macdonald and Lavie, 2011). It stands in parallel with inattentional blindness, following the same procedure of reducing perceptual processing of task-irrelevant information in high-load tasks. Therefore, one could expect various forms of "inattentional deafness" resembling the known forms of "inattentional blindness" (Mack and Rock, 1998), ranging from failing to recognize meaningful distractor objects (Lavie et al., 2009) to failing to notice the presence of stimuli (Neisser and Becklen, 1975).

Earlier research has also shown the benefit of vision in speech-reception (Musacchia et al., 2007). By contrast, it has also been shown that in situations of uncertainty, observers tend to follow the more reliable auditory cue (Apthorp et al., 2013). Very mild forms or hearing damage might lead to reduced speech intelligibility (Füllgrabe et al., 2015; Bharadwaj et al., 2014) and thus a stronger reliance on visual cues. But, it was also observed that some persons are simply more auditory dominated while others are more visual dominated (Giard and Peronnet, 1999).

The above discussion indicates that there might be individual differences in the way people perceive audiovisual stimuli that would be more pronounced in a rather complicated audiovisual environment, possibly due to individual differences in distractibility. Individual levels of distractibility can vary from slight facilitation from a noisy background to severe disruption (Ellermeier and Zimmer, 1997). It has been suggested that individual differences in working memory capacity underlie individual differences in susceptibility to auditory distraction in most tasks and contexts (Sörqvist and Rönnberg, 2014). The findings on working memory capacity reflect individual differences in the ability to control attention and avoid distraction (Conway et al., 2001). It has been shown that high-working memory capacity individuals are less susceptible to the effects of auditory distractors (Sörqvist, 2010; Beaman, 2004). A Recent study showed that attention restoration is achieved through increased exposure to natural sounds, while conversely, human-caused sounds reduce attention restoration (Abbott et al., 2016).

Throughout this article, the personal factor which was discussed above and that is expected to influence how persons perceive and appraise a combined auditive and visual stimulus will be labelled *audiovisual aptitude*. The term *aptitude* was chosen to highlight our hypothesis that this personal factor reflects a natural ability to process audiovisual scenes. This ability includes focusing on either (the visual or auditory) part of the scene and its composition in both simple and complex scenes. Its detailed meaning will further be explored in the discussion section.

This paper uses an audiovisual deviant detection experiment, with real-life scenes containing multiple visual and audio elements, to categorize persons according to their auditory acuity and their distractibility by incongruent visual stimuli. Two previously conducted experiments (labeled experiment 2 (Sun et al., 2018b) and experiment 3 (Echevarria Sanchez et al., 2017) in the following sections) have been reanalyzed by including audiovisual aptitude as a personal factor. Audiovisual aptitude is expected to modify the effect of the view from the window on reported noise annoyance in experiment 2. In experiment 3, it modifies the effect of sonic and visual stimuli on pleasantness of walking across a bridge.

The audiovisual deviant detection experiment was designed to focus on the skills and sensitivities that matter for environmental sound perception. Previous research has shown that sounds that can be recognized relate to the overall appraisal of soundscapes in public places such as parks (Axelsson et al., 2010; Pilcher et al., 2009; Miller et al., 2018). Likewise, it was shown that noticing sounds from outside influences annoyance at home (De Coensel et al., 2009). In general, perception is a comprehensive process, in which a single factor sometimes cannot explain the final result (Botteldooren et al., 2006; Brown, 2012). Thus, the first part was designed to test the participant’s ability to analyze complex auditory scenes and identify individual sounds in it. An ecologically valid setting assures that participants can also rely on personal experience and context-related expectation, factors that will also influence the appraisal of the environment in everyday life. A deviant detection task is chosen where the deviant is a complex auditory scene in which one sound is missing. To explore the influence of visual information on sound perception that is explained above, the second part of the test adds the visual context that matches the auditory scene. Congruent visual information on the deviant (missing sound) would be beneficial in general for the deviant detection task. Yet, as people are in general expected to be more visually guided (Colavita effect), participants could then simply detect the visual deviant, which would not be very instructive for identifying their audiovisual aptitude. Hence, the information on the deviant was made incongruent between the visual and the auditory information, making distraction and perceptual load dominant mechanisms.

3.2 Methodology

3.2.1 Overview

This study uses three experiments conducted by the same participants to identify the personal differences in audiovisual aptitude (experiment 1) and to explore how these differences influence perception of the environment (experiment 2&3).

The first experiment explores audiovisual aptitude. It consists of a blind audio test (Part 1) and audiovisual test (Part 2) sharing the same audio track. During both tests, participants were requested to detect the deviant auditory stimulus amongst three fragments. This experiment contained 4 scenarios, in which either the audio or visuals altered. This ecologically valid alternative to simple psychological stimuli is intended to investigate whether a person's visual attention mechanism dominates auditory attention.

Meanwhile, the same participants joined the other two experiments, one focusing on road traffic annoyance at home and the other on the perceived quality of the public space. These have been analyzed in view of the audiovisual aptitude. This setting allows to explore whether the personal audiovisual aptitude identified in experiment 1 can be used to explain differences in response in the other two experiments.

With the criteria of good (peripheral) hearing and completing the whole experiment, this study collected 68 participants (28 Female, M_{age}=27.9, SD=5.05, range: 20-46 yrs, 48 obtained a master degree or higher). In later analysis, participants were classified based on gender, age (divided into two groups by median value 27, group 1: 20-27 yrs, 37 participants, M_{age}=24.2, SD=1.8; group 2: 31 participants, 28-46 yrs, M_{age}=32.5, SD=3.9.) and education. All the principles outlined in the Helsinki Declaration of 1975, as revised in 2000 (World Medical Association, 2001), have been followed in all the experiments involving human subjects. All participants signed an informed consent form before the start of the experiments.

3.2.2 Experiment 1: Audiovisual aptitude

3.2.2.1 Layout of the paired test

As shown in Table 1, the audio test (Part 1) only contains the audio content, while the video test (Part 2) contains both sound and vision. In each part, participants were asked a single question after experiencing the three items: ‘Which of the three items sounds most differently from the other two?’. In Part 1, item 2 was the correct answer, whereas in Part 2 item 5 was the correct answer. During the analysis stage, in Part 1, choosing item 2 will be marked as correct, and consequently, choosing item 1 or 3 will be considered as mistake 1 (M1). In Part 2, item 5 is correct, and 4 and 6 mistakes (M2).

Table 1 – Overview of audio-visual scenarios studied in Experiment 1.

	Item No.	File format	Content		Mistake type
			Auditory	Vision	
Part 1	1	audio	background sound + AAO	black screen	M1
	2	audio	background sound	black screen	
	3	audio	background sound + AAO	black screen	M1
Part 2	4	video	background sound + AAO*	background view + VAO*	M2
	5	video	background sound	background view + VAO	
	6	video	background sound + AAO	background view	M2

*Congruent Visual attention Attracting Object (VAO) and matching Auditory attention Attracting Object (AAO).

3.2.2.2 Scenarios content

This study uses 4 different scenarios. Content details of the videos are listed in Table 2. Figure 1 shows screenshots of the 4 scenarios.

Figure 1 – Snapshots for 4 scenarios, (a): Airport car, (b): Restaurant, (c): Aircraft, (d): City park.

Table 2 – Visual and auditory context for each of the scenarios used in the audiovisual aptitude experiment together with congruent visual attention attracting object (VAO) and matching auditory attention attracting object (AAO).

No.	a	b	c	d
Scenario	Airport car	Restaurant	Aircraft	City park
Main visual context (background view)	terminal window view to parking apron	student restaurant at sitting position	terminal window view to airport runway	a bunch of chicken in the park
Main auditory context (background sound)	broadcasting, people talking, aircraft engine	people talking, eating, forks and plates	airport outside sound, wind, shuttlebus passing	chicken crowing and walking on fallen leaves
VAO	shuttlebus passing	tapping finger	departing aircraft	walking pigeon
AAO	shuttlebus sound	finger tapping sound	aircraft departing sound	pigeon cooing, walking on leaves
Total duration	0:27	0:35	1:00	0:55
AO duration (percentage)	0:12 (44.4%)	0:12 (34.3%)	0:24 (40%)	0:11 (20%)

In Figure 1, the object (VAO) that is absent in one of the videos in each scenario is indicated with a circle, while its path and moving direction are shown with the solid lines and arrows. Scenario (a) shows a view of a tarmac through a terminal window, with several aircrafts and a few shuttle buses far in the scene. The background sound consists of terminal announcements and people talking. Scenario (b) is a crowded student restaurant, with people eating, talking and laughing (forming the background sound). The attention attracting object in scenario (b) is a tapping finger, with its small movement within the range of the solid line circle as shown in Fig.1b. Scenario (c) shows an aircraft runway in front of a terminal window with many shuttle buses and vans moving around. Differently from scenario (a), the background of this scenario is an outdoor site with various mechanical sounds. The attention attracting object, a departing aircraft, occurs in the background of the scene. Scenario (d) shows a small city in a city outskirt, containing chickens on the left side of the screen, as well as a few cars passing by behind the park. The background sound here consists in chicken sounds, park sounds and city background sound. All four scenarios were recorded with a stable camera.

For each scenario, item 6 is the stimulus where the attracting object was removed from the visual. In scenario (a), (c) and (d), the (visually) attracting objects were removed. In scenario (b), the tapping finger was replaced by a stable hand lying on the table.

3.2.2.3 Procedure

This experiment was conducted scenario by scenario. In part 1 of the test, participants were asked to listen to items 1, 2 and 3 presented with audio only (black screen). In part 2, participants were asked to watch items 4, 5 and 6 from the same scenario. Once they finished a particular scenario, they could move on to the next one until all four scenarios were experienced.

The four scenarios were presented in random order and also the order of presenting the items was randomized. Each item could be played only once, and there was no backtrack and alteration once a single scenario was completed. All participant finished this experiment with the same headphones in the same quiet room (with a background noise of about 30 dBA).

In addition, personal information like age, gender and education level, as well as noise sensitivity (via Weinstein's questionnaire (Weinstein, 1978)) were recorded ($M_{sensitivity}$=79.40, SD=10.95, participants were split into two groups with midpoint 73.5 afterwards). The hearing status of all participants was assessed via pure tone audiometry (PTA) carried out in a quiet but not sound-proof room using a regularly calibrated AC5Clinical Computer Audiometer.

3.2.3 Experiment 2: Annoyance in living room

In a mock-up living room (Figure 2), participants were asked to engage in some light activities for 10 minutes while hearing highway traffic sounds. After 10 minutes, the standard ICBEN noise annoyance question was asked using an 11-point answering scale, referring to the past 10 minutes. This experiment was conducted with four sound pressure levels (45 dB(A), 50 dB(A), 55 dB(A) and 60 dB(A), measured in the centre of the living room) corresponding to four different acoustical window insulation cases. The following three days, the same experimental procedure was repeated. However, while participants were led to believe that they simply evaluated again four window types, what actually changed was the video playing in the background to simulate a window view (Table 3). With this experimental design, we aimed to go beyond simple loudness evaluation (as can be expected by playing a short sound fragment only). In addition, we hid the true purpose, especially regarding our interest in the visuals displayed as a window view. More details on this experiment can be found in (Sun et al., 2018b).

Figure 2 – The mock-up living room with hidden loudspeakers indicated next to the mock-up window.

Table 3 – Snapshots from the videos played in the mock-up window.

	Green elements	No green elements
Sound source visible		
Sound source invisible		

3.2.4 Experiment 3: Perception of public space

The third experiment is complementary to the second one in two ways. Firstly, it considers the public space, more specifically the perceived environmental quality of a bridge crossing a ring road giving access to a park. Secondly, four visual designs were evaluated, hiding the fact that our interest is now in the effect of the noise coming from the highway below the bridge on audiovisual quality assessment. To achieve this, on each day of the experiment the participants evaluated a walk across the bridge in a virtual environment displayed to them using oculus rift (Figure 3). A sequence of four rather different visual designs were displayed to them each day (Figure 4), yet the sound coming from the highway under the bridge stayed the same. Participants were asked to rate the pleasantness of the total experience without specifically referring to sound. On the subsequent days, they evaluated visually identical environments yet the sound changed without informing the participants. More details on this experiment can be found in (Echevarria Sanchez et al., 2017).

In this experiment, participants were virtually moving across the bridge following a pre-defined path, but they could freely move their head. An important and interesting aspect that could be analyzed with this setup is the head movement, which is a proxy for their looking behavior, reflecting where people's (visual) attention is directed to (Gibson and Pick, 1963). Recording the looking behavior allows assessing the frequency and total duration of gazing at the highway during the walk. This counting is based on the head movement of the participants and the screen middle point is used as a proxy for the visual focus point. This recording in only performed with the four matching situations (visual designs with the corresponding sonic environments).

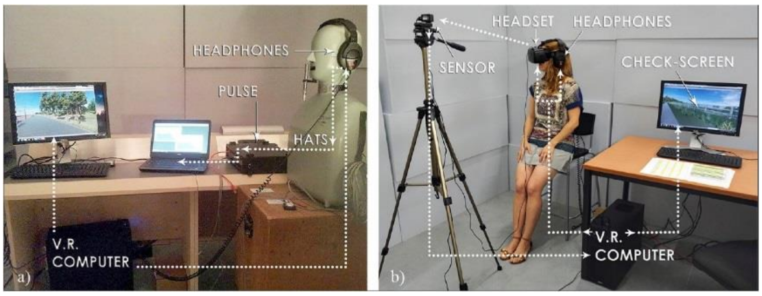


Figure 3 – a) Equipment used for calibration. b) Equipment used for Virtual reality experiment.

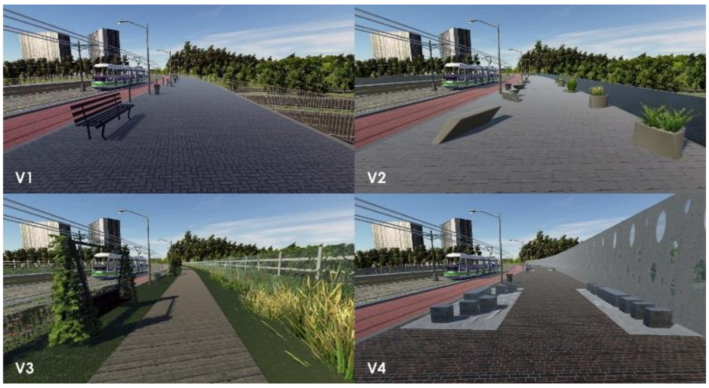


Figure 4 – Snapshot of the virtual reality display of the 4 bridge designs; the barrier seen on the right progressively increases in height when going from V1 to V4, reducing the highway noise level.

3.2.5 Statistical analysis

To test whether the personal factors have an impact on the results of part 1 and 2 in experiment 1, a repeated analysis of variance (anova) test was conducted. To observe the relation between a sound factor (the duration of the attention attracting object) and the overall result of part 1 and disparity between overall results in part 1 and 2, a linear regression was performed. Furthermore, in experiment 2 and 3, first, a generalized linear model is built to find the fittest classification of participants through experiment 1 – that is the classification that results in the best model quality. Then, a mixed-effect generalized linear model targeting at noise annoyance (Exp.2) and pleasantness (Exp.3) is conducted, using 'participant' as a random factor to generalize the results, accounting for various factors including the fittest personal factor via experiment 1. The Akaike Information Criterion (AIC) is used to rate the model quality (models with smaller AIC values fit better). At last, an anova test is conducted to check the impact of personal factors on the gazing time in experiment 3. The statistics analysis in this study was conducted in SPSS statistics (version 25).

3.3 Results and Analysis

3.3.1 Audiovisual aptitude

3.3.1.1 Overview

Figure 5 shows the percentage of the participants that made a mistake in different parts of the audiovisual aptitude experiment. In part 1 (M1), scenario ‘park’ is where people made most mistakes while scenario ‘airport car’ led to the smallest number of mistakes. Despite the scenario differences, task performance in general decreases by adding a visual setting containing incongruent information on the deviant. Comparing the differences between M1 and M2, visual information makes the task performance significantly worse in some scenarios ('airport car' and 'aircraft'), while in other scenarios, it has less effect. Further analysis will focus on personal factors that can be deduced.

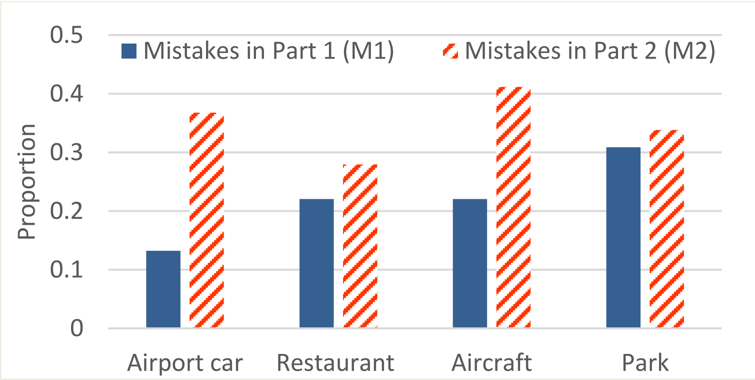


Figure 5 –Proportion of the participants making mistakes in different scenarios of the aptitude experiment.

3.3.1.2 Effect of personal factor

Aiming at M1, an anova test with factor scenario and various personal factors was made. The result shows that the factor education ($F_{1,264}$=2.31; p>.05), gender ($F_{1,264}$=1.25; p>.05), noise sensitivity ($F_{1,264}$=0.052; p>.05) and age ($F_{1,264}$=0.11; p>.05) are not significant. Interestingly, the interaction between the factors scenario and age is significant ($F_{3,264}$=2.97; p<.05), as shown in Figure 6.

On the other hand, the same procedure applied to M2 reveals that the factors education ($F_{1,264}$=1.11; p>.05), gender ($F_{1,264}$=0.46; p>.05) and noise sensitivity ($F_{1,264}$=0.054; p>.05) are not significant, while age ($F_{1,264}$=9.98; p<.01) is a significant factor, as shown in Figure 7.

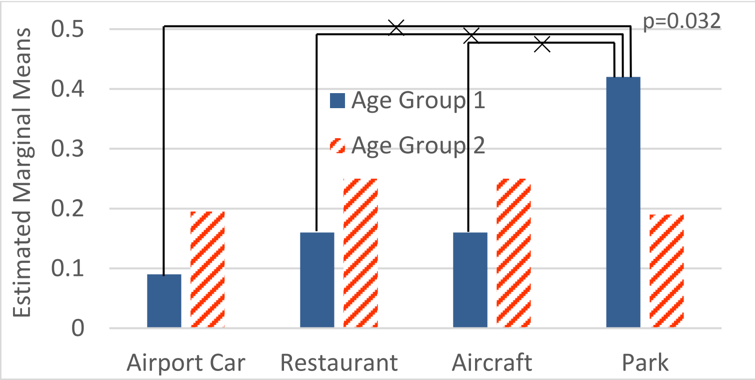


Figure 6 – Interaction between scenario and age on M1 mistakes.
(Age Group 2 is older than Age Group 1;
×: population marginal means significantly different).

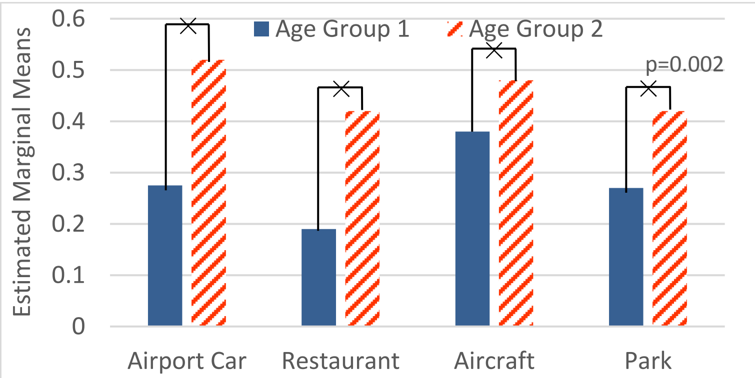


Figure 7 – Age effect on M2 mistakes.
(×: population marginal means significantly different).

As can be seen in part 1, factor age itself has no statistical significance on M1. Still there is a very strong interaction between age and scenario. Younger participants made more errors in scenario 'park' (Figure 6). In part 2 of the experiment, age is a statistically significant factor, namely older participants made more mistakes than younger ones in all scenarios (Figure 7).

Furthermore, Figure 8 shows the difference between results in part 1 and part 2, which suggests the effect of visual distraction on each age group in the four scenarios. A rather smaller variation among all four scenarios occurs in older participants.

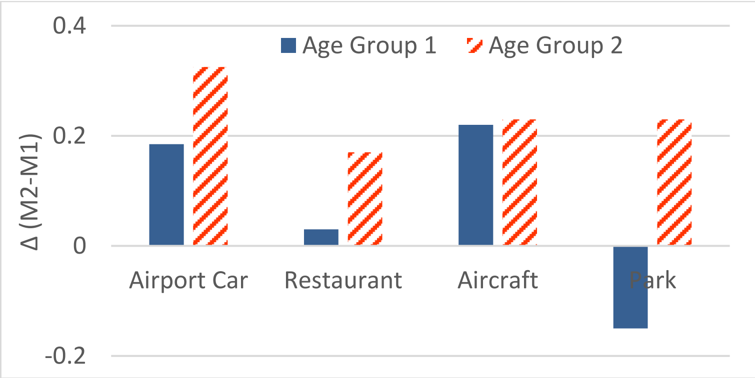


Figure 8 – Disparity of M1 and M2 by age groups.

3.3.1.3 Effect of sound features

The observation task in part 1 could be described as a pure sound deviant detection. The variation of results between each scenario (M1, Figure 5) should be ascribed to the sound itself. One feature that differs between scenarios is the total duration (%) of the attracting object (AO) stimuli, as shown in Table 2. A one-way anova test involving duration (%) as a factor on the results of M1 (on each participant) shows it has statistical significance ($F_{3,264}$=2.54; $p<.05$). In Figure 9, the correlation between AO duration (%) and M1 also supports the hypothesis that longer AO duration (%) decreases the difficulty of the sonic deviant detection task; the chance of making errors increases with decreasing duration.

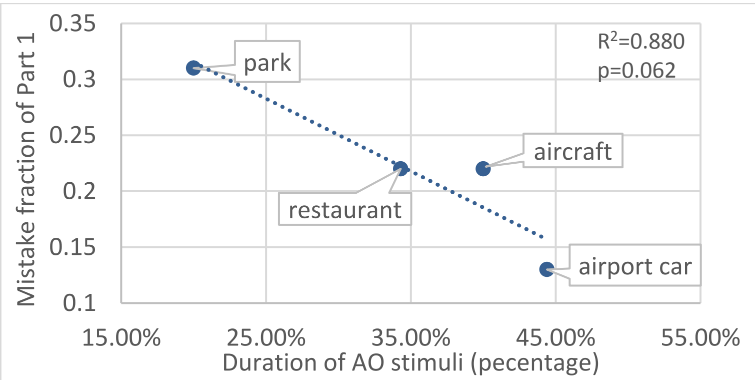


Figure 9 – Correlation between duration (%) of AO stimuli and M1.

In Figure 5, the difference between M1 and M2 suggests that the mistakes caused by the incongruent visual information also span a wide range: scenario ‘airport car’ has the biggest (Δ(M2-M1)=0.24) and scenario ‘park’ has the smallest (Δ=0.03) effect. This trend (Figure 10) also applies to the other two scenarios –

scenario 'aircraft' (duration of AO=40%; Δ=0.19) and scenario 'restaurant' (duration of AO=34.3%; Δ=0.06). Despite the correlation between the duration (%) of AO and M1 (Figure 9), Figure 11 further shows the correlation between M1 and Δ.

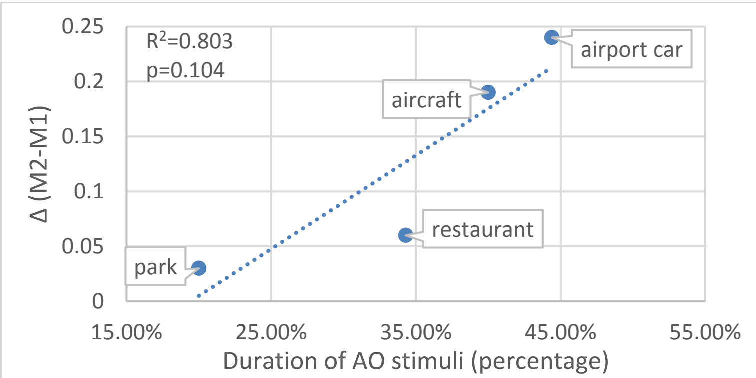


Figure 10 – The correlation between AO duration (%) and Δ (M2-M1) (disparity of M1 and M2).

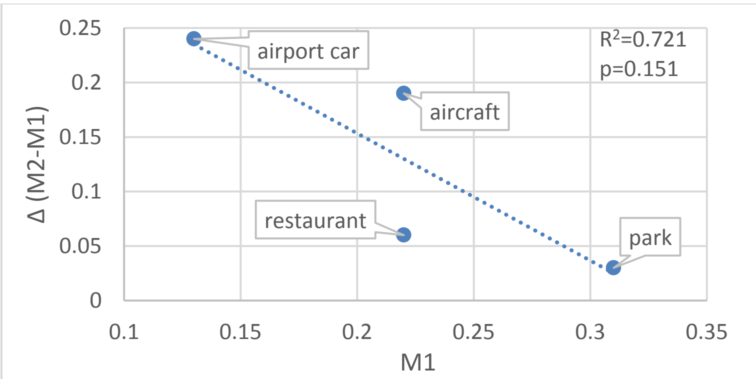


Figure 11 – The correlation between M1 and Δ (M2-M1).

3.3.1.4 Clustering by audiovisual aptitude

Combining the results of part 1 and part 2 in two dimensions (Figure 12) gives a clear view of the distribution of the participants. Participants were categorized into four groups. Group 1 (29.4%) are participants who made no mistakes in Part 1 but made at least one mistake after introducing the visual information (Part 2). Participants in group 2 (44.1%) made at least one mistake in both tests. On the contrary, group 3 (14.7%) are participants who made no mistake in any of the tests. Participants in group 4 (11.8%) made at least one mistake in Part 1, but flawlessly performed after introducing the visual information (Part 2).

These four groups generally represent different reactions towards the audiovisual stimuli, which would affect the perception as in the task performance. In the following analysis of the second and third experiment, this classification of participants will be referred to as audiovisual aptitude.

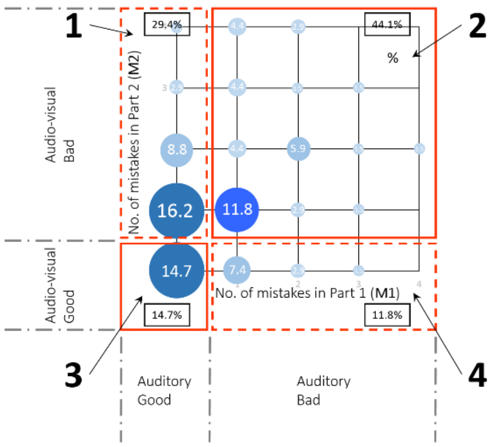


Figure 12 – Participants grouping in the audiovisual aptitude experiment.

3.3.2 Effect of audiovisual aptitude on annoyance at home

Previous analysis of this experiment showed the dominating effect of the sound level on noise annoyance and a smaller influence of the window view (Sun et al., 2018b). To test the effect of audiovisual aptitude, a generalized linear model was built targeting annoyance and involving only sound pressure levels and various ways of categorizing the four groups that were identified before. Table 4 shows the comparison of models with different groupings, aiming at searching for the best model (with lowest information criterion). Model 14 is better than other models, even though it increases the degrees of freedom. More factors and interactions are included to model 14 using a stepwise adding/removing methodology. Statistical significance of model deviance reduction when including an additional variable has been checked by likelihood ratio testing (based on the Chi-square distribution). Table 5 shows details of the best model (model 14+) with all statistically significant factors.

Table 4 – Comparison between models in living room experiment.

Model	Aptitude clustering				df	Information Criterion (Akaike Corrected)
	1	2	3	4		
1	A	B	B	B	4	3961.255
2	B	A	B	B	4	3964.488
3	B	B	A	B	4	3961.430
4	B	B	B	A	4	3989.188
5	A	A	B	B	4	3990.073
6	A	B	A	B	4	3989.473
7	A	B	B	A	4	3988.186
8	A	A	B	C	5	3960.111
9	A	B	A	C	5	3987.032
10	A	B	C	A	5	4014.913
11	A	B	B	C	5	3991.336
12	A	B	C	B	5	3960.627
13	A	B	C	C	5	3991.185
14	A	B	C	D	6	3957.773
14+						3934.948

Table 5 – Details of model 14+ in living room experiment.

``	Target: Annoyance at home			
Source	F	df1	df2	Sig.
Intercept	58.739	13	1.073	.000
Noise sensitivity	6.663	1	1.073	.010
SPL	242.440	3	1.073	.000
Noise sensitivity*Sound source	6.003	2	1.073	.003
Audiovisual aptitude*Green	2.451	7	1.073	.017

*'Participant' is used as random factor.

Even though audiovisual aptitude is not significant as a single effect due to the presence of more important factors (namely SPL and noise sensitivity), there is a strong interaction between audiovisual aptitude and visibility of green elements (see the window scenes of the living room, section 2.3). Details of this interaction are shown in Figure 13. Persons from all aptitude groups are slightly less annoyed when green elements are visible from the windows except in group 1. On the contrary, these persons that score very well on the purely auditory deviant detection task (Part 1, Exp.1), but fail when an incongruent visual element is added (Part 2, Exp.1), are less annoyed when a window scene without green elements is present.

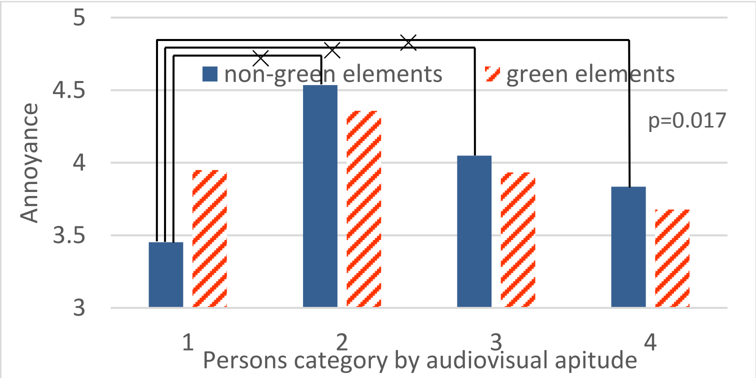


Figure 13 – The interaction between audiovisual aptitude and green elements visibility on annoyance (×: population marginal means significantly different).

3.3.3 Effect of audiovisual aptitude on perceived quality of the public space

3.3.3.1 Models for perceived quality

Analysis of the third experiment showed the strong effect of the visual bridge design and a more moderate effect of highway sound on the pleasantness rating (Echevarria Sanchez et al., 2017). In this it should be noted that sound was only changed in between days to deliberately hide changes. The same procedure as in the previous experiment is applied, using a generalized linear model now targeting pleasantness and involving only sound environment, bridge design, and audiovisual aptitude. As in the previous experiment, statistical significance of model deviance reduction has been checked by likelihood ratio testing. Model 14+ adding more interactions to model 14 using subsequent adding and removing of factors, further improved the model quality. Details are shown in Tables 6 and 7.

Table 6 – Comparison between models in public space experiment.

Model	Aptitude clustering				df	Information Criterion (Akaike Corrected)
	1	2	3	4		
1	A	B	B	B	7	4161.258
2	B	A	B	B	7	4134.640
3	B	B	A	B	7	4160.538
4	B	B	B	A	7	4160.429
5	A	A	B	B	7	4161.331
6	A	B	A	B	7	4161.570
7	A	B	B	A	7	4161.065
8	A	A	B	C	8	4160.176
9	A	B	A	C	8	4164.030
10	A	B	C	A	8	4160.841
11	A	B	B	C	8	4213.013
12	A	B	C	B	8	4160.962
13	A	B	C	C	8	4161.575
14	A	B	C	D	9	4133.550
14+						4123.957

Table 7 – Details of model 14+ in public space experiment.

Fixed Effects	Target: Pleasantness in public space			
Source	F	df1	df2	Sig.
Intercept	12.582	27	1.060	.000
Bridge design	63.038	3	1.060	.000
Sound environment	2.670	3	1.060	.046
Audiovisual aptitude*Bridge design	2.516	9	1.060	.007
Audiovisual aptitude*Sound env.	2.502	9	1.060	.008
*‘Participant’ is used as random factor.				

A strong interaction occurs between audiovisual aptitude and both bridge design and sound environment. In Figure 14, only people from aptitude group 2 have an increasing pleasantness rating with lower contribution of highway sound. Group 1 and 3 have a special preference for the sound environment with the 2nd and 3rd strongest contribution of highway sound, 68.6 dB(A) and 65.3 dB(A), respectively. Oddly, people from group 4 prefer the sound environment with the strongest highway sound more than any others. In Figure 15, people in all aptitude groups show a common high appraisal of bridge design 3 (including vegetation, Figure 4, V3), followed by design 2. Design 1 and 4 lead to relatively low pleasantness ratings, with design 4 being even slightly worse than design 1 for most people. However, the only exception is group 3 (those who performed without errors in the aptitude experiment, in both part 1 and 2): design 4 is much higher rated than design 1. In addition, Figure 16 shows the effect of audiovisual aptitude on pleasantness of the matching audiovisual combinations, namely the bridge design with the corresponding sonic environment. Persons from group 1, 2 and 3 share the similar trend, except for people from group 3 slightly preferring bridge 4 rather than bridge 2. However, for persons in group 4, bridge 4 is clearly the worst and the other three bridges do not differ from each other very much.

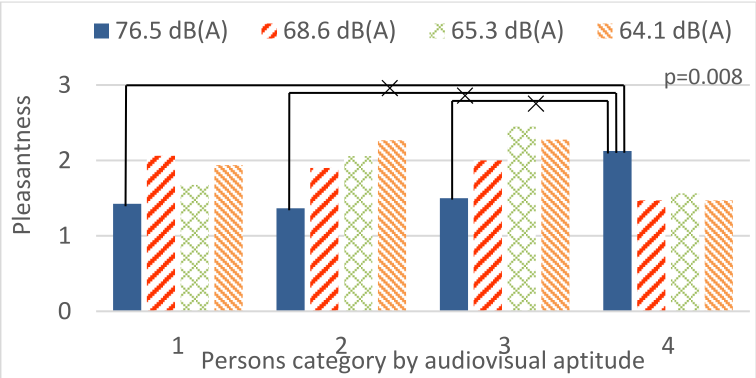


Figure 14 – The interaction between audiovisual aptitude and sound environment (highway SPL is used as a label) on pleasantness.
(×: population marginal means significantly different).

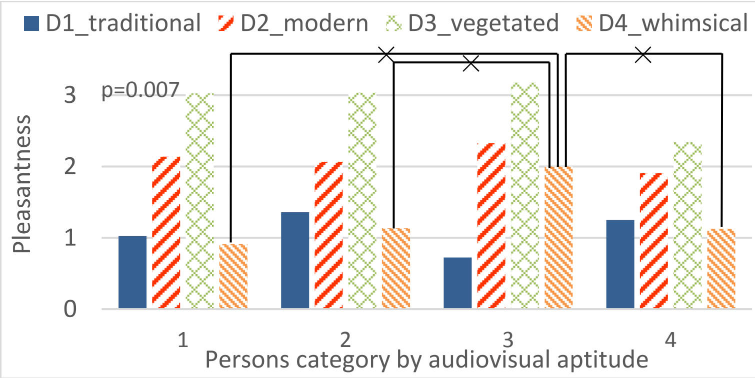


Figure 15 – The interaction between audiovisual aptitude and bridge design on pleasantness.
(×: population marginal means significantly different).

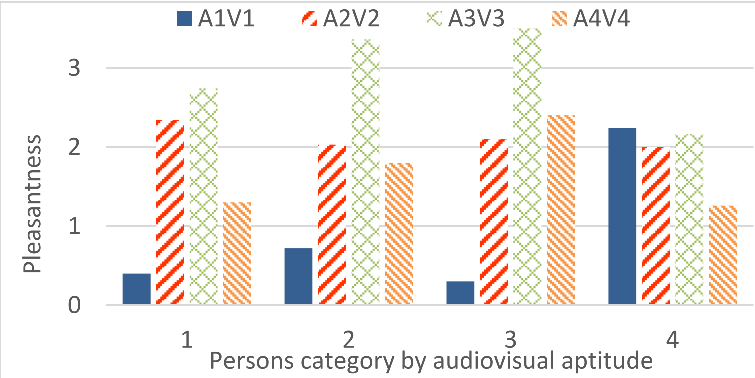


Figure 16 – Effect of audiovisual aptitude on pleasantness of matching audiovisual designs.

3.3.3.2 Looking behavior study: the gazing time

A one-way anova test with factor bridge design and gazing time (total time, Table 8) shows this is a statistical significant factor ($F_{3,224}$=8.84; p<.01). It reveals that at bridges 1 and 2 (Figure 4, V1&V2), people tend to look more often and longer at the highway. These two bridges both contain rather low edge barriers, visually exposing the sound source directly. Also, in all four bridge designs, the average gazing time is longer than the median gazing time, which shows that participants who actually look at the highway traffic do this for a longer time.

An anova test targeting at total gazing time involving the factor bridge design and personal factors shows that education ($F_{1,220}$=3.03; p>.05), gender ($F_{1,220}$=2.50; p>.05), age ($F_{1,220}$=3.77; p>.05) and noise sensitivity ($F_{1,220}$=0.04; p>.05) have no statistical significance, while audiovisual aptitude ($F_{3,212}$=2.73; p<.05) is significant. However, there is no strong interaction between the factors bridge design and audiovisual aptitude ($F_{9,212}$=0.72; p>.05). Moreover, looking back at the overall pleasantness, no clear correlation between total gazing time and pleasantness is found ($F_{113,228}$=0.64; p>.05).

Table 8 – Total gazing time for each bridge design.

Bridge Designs	Gazing time					
	Total time (seconds)		No. of times		Average time (seconds)	
	average	median	average	median	average	median
1	14.58	11.9	2.84	3	4.85	4
2	14.48	11.6	2.88	3	4.50	4.06
3	7.81	4.6	1.72	1	2.97	3.05
4	7.19	5.7	1.53	1	3.83	2.95

Note that in this section, the four bridges not only differ from each other by visual design, but also the sound level from the highway is decreasing from bridge 1 (highest) to bridge 4 (lowest). Figure 17 shows that persons in aptitude group 1 and 3, who made no errors in Part 1 of audiovisual aptitude experiment (Exp.1), look at traffic longer than the other two groups. Figure 18 shows that bridge 1 and 2, which have a rather low barrier and thus higher highway noise levels, result in more gazing time than in case of the other two bridges.

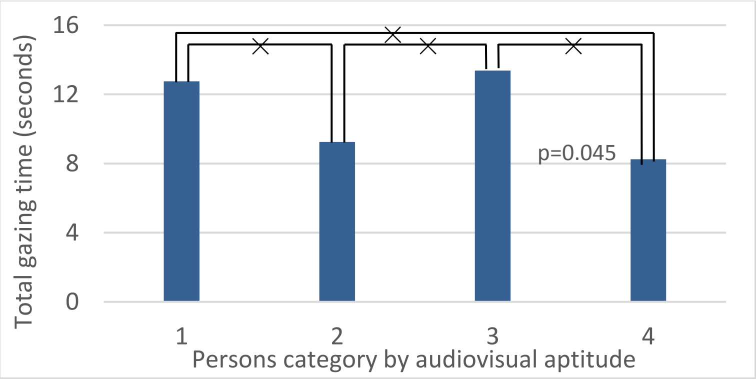


Figure 17 – Effects of audiovisual aptitude on total gazing time. (×: population marginal means significantly different).

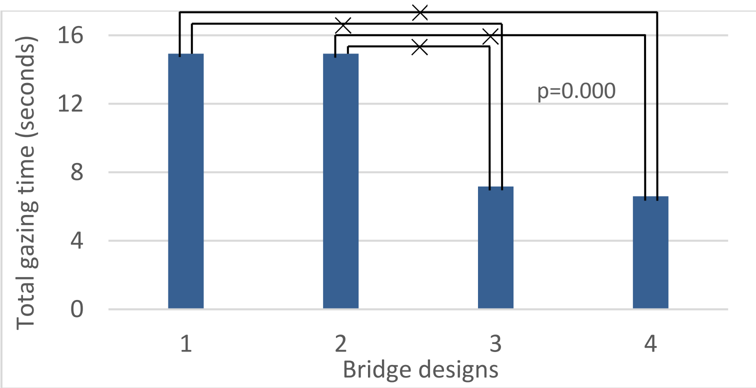


Figure 18 – Effects of bridge designs on total gazing time. (×: population marginal means significantly different).

3.4 Discussion

The goal of current study was to provide evidence for the existence of a personal factor that could influence the perception of landscape and soundscape and their interaction. For this purpose, an experiment (Exp.1) was designed to explore the individual difference in capability for unraveling audiovisual stimuli and its distractibility from auditory acuity. This personal factor was labeled *audiovisual aptitude*. Two other experiments (Exp. 2 and 3) were re-analyzed involving this personal factor. We found that in experiment 2, this individual difference modified the impact of window views on self-report noise annoyance in a living room context. In experiment 3, this individual difference altered the impact of highway sound pressure level and visual bridge design on the pleasantness rating of a public space. It also affected the looking behavior during the perception of the public space.

Our audiovisual aptitude test categorizes people according to their ability to perform the purely auditory test at one hand and the audiovisual test at the other. It is a rather strict way of grouping participants in four groups. For instance, aptitude group 3 does not allow a single mistake. Each of the groups identified in Figure 12 can be characterized in more detail and the underlying reasons for people to belong to this group may be explored. This also makes the definition of the factor audiovisual aptitude more precise.

For persons in aptitude group 1, incongruent visual information interferences the performance on the auditory task for the average person. They perform very well on the blind auditory test but start making mistakes once incongruent visual information is presented to them simultaneously. Macdonald and Lavie highlighted the level of perceptual load in a visual task as a critical determinant of inattentional deafness, an equivalent of inattentional blindness (Macdonald and Lavie, 2011). Persons in this group were successful in the sound deviant task with a low visual perceptual load (black screen, Part 1), but failed when the visual perceptual load increased (Part 2) which could be explained by being more vulnerable to inattentional deafness. Collignon et al. suggested the possibility of visual dominance in emotional processing under incongruent auditory and visual stimuli. However, this visual dominance in affect perception does not occur in a rigid manner, namely the visual dominance will disappear if the reliability of the visual stimuli is diminished (Collignon et al., 2008). The reliability of visual and auditory information influences the cross-modal asymmetry effects in temporal perception (Wada et al., 2003).

Group 2 contains most of the participants in this study. Although they often detect deviant auditory stimuli correctly with or without visual information, they make at least one error in both tasks with a slight tendency of making more errors when visual incongruent information is present (Figure 12). The complexity of the test arises either from the cocktail party effect (Conway et al., 2001) or the visual distraction effect on perception (Simons and Chabris, 1999). Both phenomena have been identified before. Hearing damage, even at a level where people would not report hearing problems or tonal audiometry does not show significant threshold shifts, could still cause reduced auditory scene analysis capacity (Füllgrabe et al., 2015). Auditory neuropathy has recently been identified as one possible cause (Bharadwaj et al., 2014). Although the age of the participants in this study does not warrant expecting a high incidence of hearing damage, some participants could clearly have more difficulties in performing the test. Also at the cognitive level we can expect some groups to perform worse (Edwards, 2016).

Persons in group 3 succeed in detecting the deviant sound in each of the four situations regardless of the presence of incongruent visual information. They could be labeled hearing specialists and are probably auditory dominated. Noise sensitivity was found before to be moderately stable and associated with current psychiatric disorder and a disposition to negative affectivity (Stansfeld, 1992), which is at least partly inherited (Heinonen-Guzejev, 2009). The present study included the Weinstein noise sensitivity survey. Persons in this group do not answer consistently different on this noise sensitivity questionnaire, which seems to indicate that another characteristic is measured by the proposed test. Other authors also noted that despite the fact that noise sensitivity has been established and widely applied in noise-related studies, it reveals only one personality trait. Miedema and Vos questioned the validity of ascribing noise sensitivity to a general negative affectivity among people (Miedema and Vos, 2003). Recent research also showed that the personality had an independent effect on noise sensitivity (Shepherd et al., 2015).

Finally, group 4 contains people that seem to be helped by the incongruent visual information while detecting deviant sound environments. They are the smallest group in this study. For purely visual tasks, it was demonstrated that a single discrete visual distraction can improve the detectability of an unexpected object (Pammer et al., 2014). Yet, it is equally likely that the visual information gives them a clue on what sounds they need to listen for in the auditory deviant detection task. Some people may have acquired the skill to compensate for their inability to form auditory objects in an auditory scene analysis task via top down mechanisms grounded in visual information.

The usefulness of the personality factor identified by the proposed audiovisual test for understanding the perception of the soundscape, and specifically the interaction between the visual and the sonic environment in it, is illustrated with two experiments.

Experiment 2 focused on road traffic noise annoyance in a living room environment. Comparing predictive models showed that keeping the four groups identified above (as separate groups) explained the observations best. Figure 13 further shows that participants belonging to aptitude group 2, 3 and 4 reported less noise annoyance when green elements were visible from the window, which is consistent with many studies (Van Renterghem and Botteldooren, 2016; Maffei et al., 2013). However, persons belonging to group 1 behaved significantly differently. They reported more annoyance at the same noise exposure when green elements were shown in the window pane (Table 3). To explain these observations, it should first be noted that the green views in this case did not provide an appealing and readable green area following the reasoning in (Kaplan and Kaplan, 1989). Instead, it only served as a visual barrier between the window and a highway. For this reason, the positive effect found in other studies may be less pronounced or even reversed. The deviating influence of a green window view on the annoyance response in group 1 may be explained in several ways. Persons in this group were identified as visual dominant and the mediocre quality of the green may have a stronger negative effect on them. Such a green view is also incongruent with the sonic environment. Persons in aptitude group 1, which are easily distracted by incongruent visual information, may value congruence more and experience the expectation gap more strongly. This expectation gap could confuse them and push them to reporting more annoyance by the traffic noise.

The evaluation of the pleasantness of crossing a bridge over the highway using virtual reality (experiment 3) also revealed significant differences between the audiovisual aptitude groups. Figure 16 shows that the most obvious group with deviant pleasantness evaluation is group 4. These participants value the audiovisual design 1 (without barrier) much more than other participants and at the same time they seem to find less pleasure in the green design (A3V3). To investigate further the reasons for this deviant rating, a closer investigation of Figures 14 and 15 reveals that it is not the visibility of the source that makes the original situation (A1V1) more pleasurable but to some extent the higher highway noise level. However, the magnitude of the effect is much more pronounced in the physically matching situation. Thus, congruency of the audiovisual information seems to play a role. In the perceived restorativeness soundscape scale (PRSS) study, Payne pointed out that specific types of sounds and their associated meanings were more important in influencing the perceived restorativeness of the soundscape than its overall sound pressure level (Payne, 2013). Considering the relatively lower pleasantness rating of the green design (A3V3) in group 4 compared to the other groups, the effect in this case seems better explained by the lower pleasure rating of the visual design (D3) as seen in Figure 15. Combining all of these observations leads to the hypothesis that persons belonging to group 4 value congruency of audiovisual information and moreover prefer to see the highway that produces the sound they hear. This matches what could be expected by the description of possible traits within this group 4 given above: these people need visual information to understand the auditory scene. Not having this information leads to a lower pleasantness rating.

Also group 3 shows deviant pleasantness ratings, in particular they value the design including a high noise barrier (A4D4) more than others (Figure 16). Looking at Figures 14 and 15 it becomes clear that this is caused by a significantly higher pleasantness rating of visual design 4 even if averaged over combinations with different highway sound levels. Earlier, this group was identified as hearing specialists, persons that are very skillful in identifying deviant sounds and that do not get misled by incongruent visual information. At first sight, this may contradict the observation that the bridge design 4 is rated more pleasantly even if combined with different highway noise levels. However, the hypothesis is forwarded that seeing the high noise barrier already induces the feeling that highway noise will be mitigated, a fact that is highly appreciated by this group.

In addition, Figure 14 shows that most participants (aptitude groups 1, 2 and 3) are following a trend of higher pleasantness rating with decreasing highway sound pressure level, despite the small difference between them. Even though the experiment was conducted on different days and the level difference can be as low as 1.2 dB(A), such a trend was still obtained. The presence of sounds that can create a frame of reference such as footsteps and a tram pass by could explain this (Echevarria Sanchez et al., 2017).

The virtual reality method used in experiment 3 also allows to monitor the head movement of the participants in the study. Participants belonging to group 1 and 3 turned their head significantly longer towards the cars on the highway. Participants in these groups make no errors on the auditory deviant detection task but may fail in the presence of incongruent visual information. Head movement is helpful in auditory scene analysis (Kondo et al., 2014), yet persons belonging to group 1 and 3 are not expected to need this information as they are performing very well on the purely auditory test. A more plausible explanation for the observed difference between groups might be that it reflects a stronger focus on environmental sound.

Hence experiment 2 and 3 show that the personal factor obtained from the aptitude experiment modifies perception of the audiovisual environment, both in a home setting and in the public space. This consistent and stable personal factor could be a potential modifier in studies on the interaction between visual and auditory information in perception experiments and could affect the way the urban environment is designed.

The core strength of the categorization should be ascribed to the aptitude experiment itself, so this experiment is analyzed in more detail. The test has been designed to assess the aptitude of participants in the auditory scene analysis step in auditory perception and to measure resistance against incongruent visual information. Indirectly it integrates an assessment of peripheral hearing status and attention focusing and gating capabilities of the person. For this reason, the test was based on ecologically valid and complex auditory and visual scenes rather than on more abstract test that are commonly used in psychology. This choice was made to maximize the probability of finding significant associations to the noise annoyance and public space perception. An appropriate test should be sensitive, reproducible, and easy to understand.

To guarantee sensitivity for all persons, the test consisted of four different contexts and deviants that could be more or less easily detected: then scenario 'airport car' would be the easiest one while scenario 'park' the hardest. This range in difficulty is mainly achieved by the duration (%) of AO stimuli as shown in Section 3.1.3. Figure 10 indicates that in scenario 'airport car', the monitoring task is relatively easy (perceptual load of the task is low), the visual distraction is sufficiently working. While vice versa, in scenario 'park', the monitoring task is rather hard (perceptual load of the task is high), the visual distractor processing tends to be less pronounced. This comparison agrees with perceptual load theory (Lavie, 1995). Figure 11 confirms that the more difficult the purely auditory task, the lower the influence of the visual distractor.

Furthermore, the sensitivity of the test for age of the participant reflects the sensitivity of the test. Earlier research suggested that older adults were more affected by irrelevant speech in a monitoring task (Bell et al., 2008). The age deficits occurred in many conditions and increased with the similarity of distractor and target (Scialfa et al., 1998). Cohen and Gordon-Salant also stated that older adults may be more susceptible to irrelevant auditory and visual competition in a real-world environment (Cohen and Gordon-Salant, 2017). Some research has shown that older and younger persons obtained similar performance with purely auditory stimuli, but older adults have poor performance with audiovisual modality (Sommers et al., 2005). These findings are congruent with the presented study, as stated in section 3.1.2. However, in part 1 of the audiovisual aptitude experiment, younger participants made less mistakes in all scenarios except for scenario 'park' (Figure 6). In figure 8, the smaller variation in older participants suggests that the visual distraction tends to have a more equalized effect on them. However, for younger participants, there's a bigger difference between scenarios, which might indicate that the visual distraction process highly depends on the context for younger people. Early research showed the effect of sound familiarity on recognition (Cycowicz and Friedman, 1998), which could suggest a large part of younger participants in this experiment were unfamiliar with a natural sonic environment.

The latter observation could lead to poor reproducibility of the test in another group of persons with different familiarity with the audiovisual scenes that are presented. This could be a plie for choosing a more abstract audiovisual test. The reported experiments were intended to show the existence of a difference in audiovisual aptitude between persons that could affect perception of the sonic and visual environment. It nevertheless has some limitations. An auditory deviant detection test with a limited number of scenarios will not reveal the full truth of above-mentioned hypothesis. The scenarios may not have been optimally chosen to balance familiarity with the environment amongst all participants. In addition to the age influence, other demographic factors may lead to a change in behavior in specific scenarios. For such an experiment, the number of participants matches widespread practice. However, using larger test populations may uncover other and more subtle influences and relationships. Also the verification – experiments 2 and 3 – has certain shortcomings. In section 3.3.2, for instance, the head movement was used as a proxy for eye movement since no eye tracer, compatible with the VR headset, was available at the time of the experiment.

3.5 Conclusion

Our study provides evidence for the existence of a personal factor that influences the effect of the view from a living room window on perceived noise annoyance by highway traffic noise and the effect of both the visual design and the highway noise level on perceived pleasantness of crossing a bridge over a highway. This personal factor, which we labeled audiovisual aptitude, may explain differences in perception of the (audiovisual) environment observed in other studies. It was shown that this personal factor differs from noise sensitivity, a known personality trait. It could become as important as noise sensitivity in understanding differences in perception of the living environment when both landscape and soundscape matter.

In this work, a deviant detection experiment was used to categorize persons according to their audiovisual aptitude. It was shown that categorization in four groups resulted in more performant models for predicting the above-mentioned influences than using less groups. Each group could be linked to personal factors identified previously in literature. Nevertheless, it can be expected that such an extensive test resulting in four groups might not be necessary. Based on the insights gained in this work, an audiovisual aptitude questionnaire may be constructed.

Future research may also focus on finding the neurological basis for the difference in audiovisual aptitude between persons. Recent research shows that high noise sensitivity is associated with altered sound feature encoding and attenuated discrimination of sound noisiness in the auditory cortex (Kliuchko et al., 2016). Audiovisual aptitude is expected to be related to attention moderated auditory scene analysis.

4

Classification of soundscapes of urban public open spaces

Sun K, De Coensel B, Filipan K, Aletta F, Van Renterghem T, De Pessemier T, Joseph W, and Botteldooren D

Submitted to *Landscape and Urban Planning*.

This chapter presents an immersive soundscape reproduction method based on the presentation of spatial audio combined with 360-degree video, and a hierarchical method for soundscape classification. An experiment is conducted to validate this classification. This work was carried out in the framework of the Urban Soundscapes of the World project, supported by the HEAD Genuit Foundation, and of the C3PLACES project, supported by the European Union's H2020 research innovation programme. Part of this research was presented at the 2017 Internoise conference in Hong Kong (De Coensel et al., 2017) and at the 2018 Internoise conference in Chicago (Sun et al., 2018a).

4.1 Introduction

Soundscape, as defined by the International Organization for Standardization (ISO), is an "acoustic environment as perceived or experienced and/or understood by a person or people, in context" (ISO, 2014). The urban soundscape contributes to the perceived quality of the urban environment and the identity of a city. Ambient sounds may evoke thoughts and emotions, may influence our mood or steer our behavior. Cities are comprised of many types of public outdoor spaces, each with their distinctive soundscape. Inspired by the potential positive effects a suitable acoustic environment may have on well-being of citizens and the attractiveness of the city, the challenge of designing the acoustic environment of urban public outdoor spaces has attracted attention since decades (Southworth, 1969; Schafer, 1994).

During the past decades, research on the urban sound environment and soundscape has grown, driven by increased population density and abundance of mechanical sounds in mega-cities across the world. Sound in outdoor environments has traditionally been considered in negative terms as both intrusive and undesirable (Jennings and Cain, 2013). However, sound may provide positive effects as well, such as enhancing a person's mood, triggering a pleasant memory of a prior experience, or encouraging a person to relax and recover (Payne, 2013). Where classical noise control exclusively focusses on reducing levels of unwanted sounds, soundscape design requires new tools. Hence the advent of realistic and affordable immersive audio-visual reproduction systems (head-mounted displays), backed by increasingly efficient and realistic acoustic simulation and auralization models (Vorländer, 2008) has been identified as a key enabling technology. Immersive virtual reality could also become a valuable tool for interactive participatory evaluation of the soundscape in urban planning and design projects (Puyana-Romero et al., 2017; Echevarria Sanchez et al., 2017), as virtual reality reproduction systems are rapidly becoming affordable and widely available.

Design is often inspired by good examples. As context is an important part of the soundscape and the visual setting is a string cue for context, examples of acoustic environments should be embedded in accurate 360-degree visualization. To date, however, no unique protocol or standards exist for immersive audio-visual recording and playback of urban environments with soundscape in mind (Hong et al., 2017). In addition to providing examples, high-quality immersive recordings of existing spaces are highly valuable to serve as an ecologically valid baseline for studying the perceptual outcome of noise control and soundscape measures. Hence, such recordings are now being collected in cities across the globe. To unlock such collections, a suitable classification is needed and best examples of each class need to be identified.

One could consider a purely acoustical categorization (Rychtáriková and Vermeir, 2013). However, according to the soundscape definition (ISO, 2014), soundscape evaluation should not be restricted to acoustical determinations only (Zannin et al., 2003), as the social context (Maris et al., 2007), visual context (Sun et al., 2018b) and individual differences need to be included (Dubois et al., 2006).

When asked to describe the urban acoustic environment, persons tend to name audible sounds and their sources and may relate the quality of the environment to the meaning given to these sounds (Dubois et al., 2006). In view of the importance of audible sounds, classification schemes based on urban sound source sorting have been proposed (Léobon, 1995; Brown et al., 2011). Such classifications can easily be applied to collections of audio-visual recordings through listening experiments conducted by sound specialists, yet one should remain aware that attention plays an important role in the perception of the acoustic environment in a real context (Oldoni et al., 2013). Classification based on audible sources does not capture the influence of the composition as a whole on persons and therefore should be complemented by more holistic indicators.

Holistic descriptors that have been proposed previously and that could be used for classification include: pleasantness, music-likeness, restorativeness, appropriateness. (Aletta et al., 2016a; Botteldooren et al., 2006). A lot of research has focused on the soundscape descriptors inspired by emotion-denoting adjectives (Brown, 2012; Aletta et al., 2016a). The well-known circumplex model of affect (Russell, 1980) identifies eight affective concepts that can be mapped to a two-dimensional plane. Previous research (Berglund and Nilsson, 2006; Axelsson et al., 2010) translated core affect to the physical environment that causes it and showed that outdoor soundscape quality may be represented by two main orthogonal components: pleasantness and eventfulness. In such a 2D model specific directions are labelled : exciting (45˚), chaotic (135˚), monotonous (225˚) and calm (315˚).

Although very popular, this assessment and classification framework has also been subject to some critique. Regarding the core affect model itself, research has identified a main problem with the two-dimensional approach offered by Russell: a variety of overlapping emotional concepts can be placed in the same quadrant of the model (e.g., Ekkekakis, 2008). Based on the 2D core affect model, Latinjak (2012) proposed a three-dimensional model, where a third dimension, namely “time perspective”, was added next to arousal and valence. In addition, the classification of soundscape in the pleasantness – eventfulness plane assumes that the environmental sound is attentively listened to. It assumes that perceiving the sonic environment is a main purpose of an individual visiting a place, which is not often the case. Unawareness of the surroundings (inattentional blindness (Simons and Chabris, 1999) and inattentional deafness (Macdonald and Lavie, 2011)) occurs especially during moments with reduced attention towards the environment. The sonic environment is thus often backgrounded.

Besides the soundscape descriptors and the 2D core affect model, a triangular qualitative urban sound environment mapping technique was recently proposed (Kamenický, 2018). This research used activities, mechanisms and presence to build an objective soundscape map based on composition of sound events. A significant correlation between qualitative cognitive-semantic variables clustering and quantitative acoustic and psychoacoustic parameters agglomerative clustering was proposed.

In an urban environment, the soundscape, the landscape, etc., and its users form an ecological entity. It might therefore be more suitable if the soundscape classification of existing urban sites could be treated within such a holistic context. With the aforementioned discussion in mind, we propose a coarse hierarchical classification that could be used for labelling audiovisual collections or as a first mapping of the city. The proposed classification, shown in Figure 1, was first suggested in De Coensel et al. (2017). In a first stage, soundscapes are classified according to whether they are backgrounded or contain foregrounded sound elements when perceived within context (Botteldooren et al., 2015) – where only visual context has been considered here. Foregrounded sound affects the overall perception of the environment. In a second stage, one could distinguish between sonic environments that are disruptive or supportive for the envisaged use. Disruptive sound environments could lead to annoyance. Finally, the sonic environment could be supportive for the overall experience of the living environment in many different ways. Here, the proposed classification follows the arousal dimension of core affect to distinguish between calming (reducing arousal) and stimulating (increasing arousal). We forward the hypothesis that the proposed classification system is strongly related to the sonic environment itself and less sensitive to differences between people than previous classification systems and therefore more appropriate for classifying the audio-visual representation of a place.

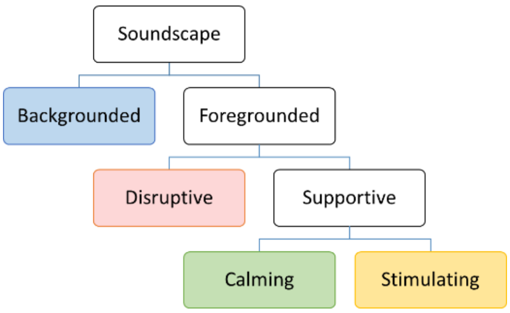


Figure 1 – Proposed hierarchical classification of urban soundscapes.

It is worth noticing that the proposed classification is not crisp; one could potentially mathematically formalize this classification using fuzzy set memberships.

In this article, the proposed classification will for the first time be made operational through a questionnaire that is administered to a panel of volunteers that is experiencing the immersive playback at the laboratory of a collection of audio-visual recordings at an urban site (Section 2.2.3). This will allow to explore the rationality of the proposed soundscape classification, the underlying affiliation between categories and its comparison with the 2D core affect model (Section 3.3). Classification of a collection achieved by questioning persons about the soundscape as experienced in the virtual reality environment has some drawbacks: because of the variability between persons (Sun et al., 2018c), this requires an assessment panel of sufficient size, which results in a large effort and cost for classifying new recordings. Hence this paper also proposes models based on acoustical parameters (Section 3.5).

4.2 Methodology

4.2.1 Collection

4.2.1.1 Site selection protocol

Sampling of urban sites for performing soundscape evaluation studies is most often performed in an *ad hoc* manner. Systematic site selection methods for landscape studies, conservation and planning are often based on objective factors such as land cover (Gillespie et al., 2017), as well as perception, visual preference and emotional attachment of local residents (Longstreth, 2008; Walker and Ryan, 2008). The latter are typically evaluated through surveys or interviews, in order to select a sample of sites covering a wide range of landscapes (Tress et al., 2006).

A similar approach for site selection was also applied at the early stage of this study. An online questionnaire survey was conducted among 30 to 50 inhabitants (depending on the city), in which they were asked to pinpoint outdoor public spaces within their city that they perceive along the soundscape perception dimensions of pleasantness and eventfulness. Locations obtained from the online survey were then spatially clustered using the Google MapClusterer API, which allows extracting a shortlist of prototypical locations. This approach was designed to lead to a range of urban sites with a large variety in soundscapes, more or less uniformly covering each of the four quadrants of the 2D core affect perceptual space (Axelsson et al., 2010; Cain et al., 2013). In each city, participants were recruited among local students, and through calls for participation on relevant Facebook pages and with local guide associations. Details of the site selection protocol can be found in De Coensel et al. (2017).

4.2.1.2 Audio-visual recording

Combined and simultaneous audio and video recordings were performed at the selected locations within each city, using a portable, stationary recording setup. Photographs of this setup are shown in Figure 2. The setup consists of the following components: binaural audio (HEAD acoustics HSU III.2 artificial head with windshield and SQobold 2-channel recording device), first-order ambisonics (Core Sound TetraMic microphone with windshield and Tascam DR-680 MkII 4-channel recording device) and 360-degree video camera (GoPro Omni spherical camera system, consisting of 6 synchronized GoPro HERO 4 Black cameras). The ears of the artificial head, the video camera system and the ambisonics microphone are located at heights of about 1.50m, 1.70m and 1.90m, respectively. It was chosen to stack the audio and video recording devices vertically, such that no horizontal displacement between devices is introduced, which could otherwise result into an angular mismatch for the localization of sound sources in the horizontal plane. A minimal separation distance of about 20cm between the camera and both the binaural and ambisonics microphones is required, such that these do not show up prominently on the recorded video, and can be masked easily using video processing software. All audio was recorded with a sample rate of 48 kHz and a bit depth of 24 bits, and were stored in uncompressed .wav format; moreover, the binaural recordings were performed according to the specifications set forth in ISO TS 12913-2 (ISO, 2018). Note that the recording setup is highly portable: when disassembled, all components can be carried by a single person. Assembling the setup takes about 10 minutes, and batteries and memory of all recording devices allow for about a full day of recording.

At each location, the recording system is oriented towards the most important sound source and/or the most prominent visual scene—this orientation defines the initial frontal viewing direction for the 360-degree video and ambisonics recordings, and the fixed orientation for the binaural recordings. Time synchronization is performed at the start of each recording by clapping hands directly in front of the system; this also allows checking correct 360-degree alignment of all components when post-processing. At each location, at least 10 minutes of continuous recordings were performed, such that 1-minute or 3-minute fragments containing no disturbances can be extracted easily. During recording, the person handling the recording equipment was either hiding (in order not to show up on the 360-degree video) or, in case hiding was not possible, blended in the environment (e.g. performing the same activities as the other people around).

Figure 2 – Recording setup (*Left*: photo on location (Boston); *Right*: position diagrammatic sketch).

4.2.1.3 Post-processing for Virtual Reality

Since the six cameras from GoPro Omni use a parallel program, the six individual videos are automatically synchronized. The stitching work that combines these six videos together as a single 360-degree video is achieved with Autopano Video and Autopano Giga from Kolor software team. It gives the postproduction a stable, color-balanced and sustained 360-degree view. Since the postproduction captures the full surroundings, it is impossible to know what the viewer will eventually be focusing on (within the 360-degree sphere) at any given moment. In this study, only the opening scene of each recording (the coordinates of the image) was fixed, which ensures all the participants receive the same view at the beginning. With this setting, it also sets a reference for the audio-spatial synchronization.

Since the GoPro Omni cameras stand between the tripod stand, the HEAD and the Tascam (Figure 2), the videos will also record these devices, shown in zenith and nadir (top and bottom) in the postproduction, respectively. These were carefully camouflaged with a patch created in Photoshop, ensuring that no recording equipment appears in the final playback. Also, a color equalization has been applied to the postproduction by using ffmpeg (saturation=2), which highlights the color vividness in the video. All videos were exported in 4k quality.

Together with the presentation by an Oculus Virtual Reality device, it gives a visually realistic and immersive experience as if the participants were in the place standing right on the recording position.

These 360-degree video is paired with ambisonics audio recording. The reason why first-order ambisonics audio can be used is explained in Appendix. Video and audio synchronization was conducted by ffmpeg. Google Spatial Media Metadata Injector was used to achieve the spatial audio effect, that the audio field changes following the head rotation.

4.2.2 Experiment: Soundscape classification

4.2.2.1 Material & participants

In total, 50 one-minute recordings were selected from the complete recording in this experiment (e.g.: Figure 3). One minute is very short for assuring that participants are not focusing on the sound, but this time interval was chosen as a compromise that still gave a good impression but would not take too much time from the users of the collection. Table 1 gives the overview of their basic characteristics namely location, time, and $L_{\mathrm{Aeq,\,1\,min}}$ (A-weighted equivalent sound pressure levels during the one-minute period). The L_{Aeq} of each stimulus was calculated on the basis of the binaural signal, applying an independent-of-direction (ID) equalization, and taking the energetic average between both ears.

To allow for completely independent validation of prediction models, the whole experiment was repeated two times. First, 25 soundscapes (Table 1 – classification 1) were chosen for participant group 1 (20 participants, 6 female, $\mathrm{Age_{mean}}$=28.9 yr, standard deviation 2.8 yr, range: 25-35 yr). Five cities (Montreal, Boston, Tianjin, Hongkong and Berlin) were included in the experiment, and each city contributed with 5 soundscapes. The soundscapes were presented city by city to the participants. The city order and the order of soundscapes in each city were randomized.

Another 25 recordings (Table 1 – classification 2) were presented to participant group 2 (20 participants, 5 female, $\mathrm{Age_{mean}}$=30.2 yr, standard deviation 5.6 yr, range: 22-46 yr). The number of soundscapes per city was different now. These 25 recordings were grouped into 5 groups of 5 soundscape each, avoiding e.g. that one group contained only parks. The group order and the order of soundscapes in each group were again fully randomized.

All participants had normal hearing status which was assessed via pure tone audiometry (PTA) carried out in a soundproof room using a regularly calibrated AC5Clinical Computer Audiometer. All participants had normal color vision which was tested by the “Ishihara test for color deficiency” (Ishihara, 1957). The participants performed the perception experiment individually, and were offered a gift voucher as compensation.

Figure 3 – Example: snapshot of stimuli R0001. (more stimuli could be found in Supplement 1).

Table 1 – Overview of stimuli: (upper) classification 1, (lower) classification 2.

Label	City	Date	Time	Location	Longitude	Latitude	$L_{Aeq,1min}$/dB
R0002	Montreal	2017/6/22	8:43	Place d'Armes	45.504683	-73.55715	66.5
R0003	Montreal	2017/6/22	9:43	Tour de l'horloge	45.511973	-73.545911	55
R0007	Montreal	2017/6/22	15:26	Chalet du Mont-Royal	45.503405	-73.587005	54.8
R0010	Montreal	2017/6/22	17:53	Square Phillips	45.503807	-73.568543	67.5
R0011	Montreal	2017/6/22	19:10	Place Jacques Cartier	45.50768	-73.552625	66.1
R0015	Boston	2017/6/28	12:41	Old State House	42.359039	-71.057139	69.5
R0016	Boston	2017/6/28	13:11	Quincy Market	42.35986	-71.055825	74.6
R0017	Boston	2017/6/28	13:47	Post Office Square	42.35623	-71.0556	65.8
R0018	Boston	2017/6/28	14:23	R. F. Kennedy Greenway	42.354721	-71.052073	66.1
R0020	Boston	2017/6/28	16:31	Paul Revere Mall	42.365687	-71.053446	57.4
R0022	Tianjin	2017/8/24	8:54	Peiyang Square (TJU campus)	39.107327	117.170222	62.2
R0026	Tianjin	2017/8/24	11:46	Water Park North	39.090986	117.163317	60.4
R0029	Tianjin	2017/8/24	15:29	Haihe Culture Square	39.130202	117.193256	73.5
R0031	Tianjin	2017/8/24	16:26	Tianjin Railway Station	39.133779	117.203206	65.2
R0033	Tianjin	2017/8/24	17:59	Nanjing Road	39.118566	117.185557	65.3
R0036	Hong Kong	2017/8/29	15:43	Wanchai Tower	22.279705	114.17245	68.7
R0040	Hong Kong	2017/8/30	7:44	Hong Kong Park	22.277824	114.161488	64.1
R0041	Hong Kong	2017/8/30	8:50	Wong Tai Sin Temple	22.342062	114.194042	69.7
R0047	Hong Kong	2017/8/30	13:36	Peking Road	22.296512	114.171813	77
R0048	Hong Kong	2017/8/30	14:30	Ap Lei Chau Waterfront	22.245093	114.155663	62.2
R0050	Berlin	2017/9/9	16:57	Breitscheidplatz	52.504926	13.336556	72.4
R0054	Berlin	2017/9/10	11:32	Gendarmenmarkt	52.513517	13.3929	60.8
R0058	Berlin	2017/9/10	14:18	Lustgarten	52.518604	13.399195	65.2
R0060	Berlin	2017/9/10	15:39	James-Simon Park	52.521787	13.399158	65.9
R0061	Berlin	2017/9/10	16:32	Pariser Platz	52.516145	13.378545	67.7

R0001	Montreal	2017/6/22	8:02	Palais des congrès	45.503457	-73.561461	65.8
R0004	Montreal	2017/6/22	10:39	Place Marguerite-Bourgeoys	45.507368	-73.555006	62.1
R0005	Montreal	2017/6/22	12:21	Parc La Fontaine	45.523279	-73.568341	53.7
R0006	Montreal	2017/6/22	14:22	Monument à Sir George-Étienne Cartier	45.514488	-73.586564	58.7
R0008	Montreal	2017/6/22	16:26	McGill University campus	45.504202	-73.576833	54.7
R0012	Boston	2017/6/28	9:36	Boston Public Garden	42.353478	-71.070151	62.5
R0013	Boston	2017/6/28	10:12	Boston Common	42.353705	-71.065063	62.3
R0023	Tianjin	2017/8/24	9:23	Jingye Lake (TJU campus)	39.107495	117.166476	57.4
R0027	Tianjin	2017/8/24	12:14	Water Park Center	39.087846	117.162092	58.5
R0030	Tianjin	2017/8/24	16:00	Century Clock	39.13262	117.198314	63.2
R0032	Tianjin	2017/8/24	16:55	Jinwan Plaza	39.131835	117.202969	60.7
R0034	Tianjin	2017/8/24	18:44	Drum Tower	39.140833	117.174355	54.5
R0037	Hong Kong	2017/8/29	16:14	Johnston Road	22.277781	114.176621	71.6
R0038	Hong Kong	2017/8/29	17:07	Taikoo Shing	22.286715	114.218385	64.6
R0039	Hong Kong	2017/8/29	17:55	Victoria Park	22.281835	114.187832	57.0
R0042	Hong Kong	2017/8/30	9:44	Nelson Street	22.318352	114.170164	67.2
R0043	Hong Kong	2017/8/30	10:32	Signal Hill Garden	22.296008	114.174859	62.1
R0045	Hong Kong	2017/8/30	12:45	Hong Kong Cultural Centre	22.29343	114.170038	60.7
R0049	Hong Kong	2017/8/30	15:53	The Peak	22.270879	114.150917	55.6
R0052	Berlin	2017/9/10	9:28	Tiergarten	52.512166	13.347172	53.3
R0053	Berlin	2017/9/10	10:48	Leipziger Platz	52.509296	13.37818	68.8
R0055	Berlin	2017/9/10	12:08	Checkpoint Charlie	52.507796	13.390011	66.5
R0057	Berlin	2017/9/10	13:43	Neptunbrunnen	52.519829	13.406623	66.2
R0062	Berlin	2017/9/10	18:06	Sony Center	52.510166	13.373572	66.9
R0063	Berlin	2017/9/10	18:31	Potsdamer Platz	52.509192	13.376332	67.4

4.2.2.2 Experiment setup

Participants joined this experiment inside a soundproof booth (Figure 4), where the process was monitored through a double-glassed window from outside. Stimuli were played back using a PC (placed outside the booth), equipped with the GoPro VR Player 3.0 software, which allowed to play back video with spatial audio. The 360-degree video was presented through an Oculus Rift head-mounted display. The audio was played back through Sennheiser HD 650 headphones, driven by a HEAD acoustics LabP2 calibrated headphone amplifier. The gain of the ambisonics audio has been adjusted such that their level is as close as possible to that of the corresponding binaural audio tracks.

During the experiment, participants remained seated (seat height: 0.50m), which allowed them to freely move their head and look around in all directions but physically remained at a fixed position. The sensor for Oculus Rift was placed on a tripod (height: 1.20m), keeping approximately the same height as the participant's head position. A microphone was mounted on the tripod and was driven by a laptop, which was used to monitor the experiment from outside. When participants needed to answer questions during the experiment, they could do it by (verbal) talking and the experimenter could mark it from outside the booth. By this procedure, a holistic immersed experience was maintained throughout the full experiment.

Figure 4 – Experiment setup (*Left*: inside the booth; *Right*: view from monitoring position).

4.2.2.3 Procedure

Soundscape classification according to Figure 1 was achieved via a questionnaire. The questionnaire was designed to follow the hierarchical nature of the classification and with brevity in mind. To assess foregrounding/backgrounding of the sound within the holistic experience participants were asked: (Q3) *How much did the sound draw your attention?* To frame this question, a more general question (Q1) *In general, how would you categorize the environment you just experienced?* was added. The options for answering this question already focus attention on the more pleasurable evaluation: "*calming/tranquil*" to "*lively/active*" but with a clear option "*neither*" in between. The question distinguishing disruptive from supportive environments relates to possible activities: (Q4) *Would the sound environment prevent you from doing the activities above?* A question that again required some framing by listing possible activities in Q2 (see Figure 5). The answers to Q2 are not used and hence the choice of possible activities is not critical.

Finally, Q5 evaluates the contribution of the sonic environment as being supportive to the perception of the overall environment. This question *defines* the labels *calming* and *stimulating* as sonic environments that contribute to the *calmness/tranquility* and the *liveliness/activeness* of the place respectively.

Participants experienced the one-minute stimuli first, followed by the 5 questions presented in the VR screen with a black background (Figure 5). Participants needed to answer all 5 questions verbally. Hence also the choice for a 5-point answer scale with answering categories equidistantly spaced is in agreement with Fields et al. (2001). Note that question 5 has two versions, only one (5a or 5b) is presented to the participants. This is based on the answer in question 1: participants answering “very calming/tranquil” or “calming/tranquil” received question 5a, while participants answering one of the other choices got question 5b. After answering the questions, the next stimuli were presented. Thus, participants did not have to take off the headset between experiencing each stimulus.

The experiment was divided in 5 sections, each section contained 5 stimuli (in classification 1, one city is one section, while in classification 2, one group is one section, see Section 2.3.1). Between each section, there is a small break where participants could take the headset off. During this break, participants needed to answer additional questions regarding to the 5 stimuli they just experienced. Participants got 5 photos of the opening scenes of the stimuli in the same order as the stimuli play order. Below each photo, participants first needed to put a score on a 11-point scale (from 0: “not at all” to 10: “extremely”) on the following questions: “*How well do you remember the sound environment that goes with this picture?*”, and “*How would you rate the sound environment of this place in terms of "full of life and exciting"/"chaotic and restless"/"calm and tranquil"/"lifeless and boring"?*”(Axelsson, 2015a), respectively. After this break, the next 5 stimuli were presented to the participants with the same procedure until all 25 stimuli (i.e. 5 sections) were evaluated.

After the participants finished the 25 stimuli, two questions regarding the overall reproduction quality were asked, specifically on the realism and immersion, using an 11-point scale. The questions presented during the break and at the end of experiment were answered on paper, thus an 11-point scale could be seen as continues scale.

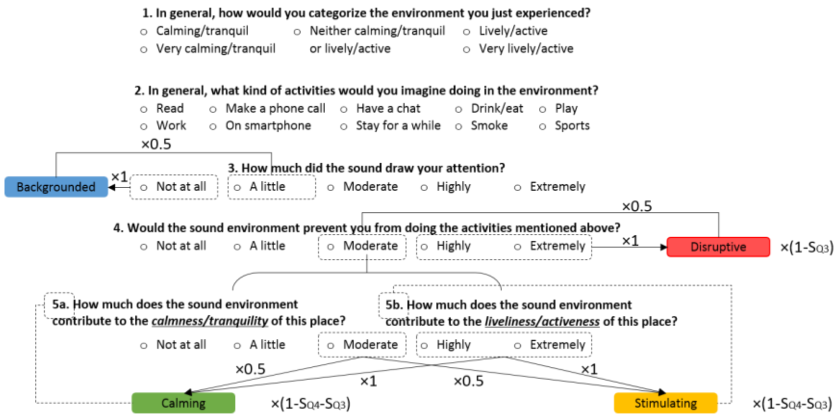


Figure 5 – overview of the questions and flow. (colored parts: fuzzy scoring in proposed classification).

4.2.2.4 Data processing

In this study, the fuzzy membership set of the four proposed classes *backgrounded, disruptive, calming,* and *stimulating* is based on the answers in question 3, 4, 5a and 5b, as marked in Figure 5, where $S_A(x)$ is the membership degree of soundscape x in the fuzzy set A. The fuzzy membership set, i.e. the correspondence between the answer on the question and the degree of belonging to each class, is given in Table 2.

Table 2 – The fuzzy membership set for each class of soundscape.

Question	Answer					Fuzzy set
	Not at all	A little	Moderate	Highly	Extremely	
Q. 3	1	0.5	0	0	0	$S_{backgrounded}(x)$
Q. 4	0	0	0.5	1	1	$S_{disruptive}(x)$
Q. 5a	0	0	0.5	1	1	$S_{calming}(x)$
Q. 5b	0	0	0.5	1	1	$S_{stimulating}(x)$

To account for the hierarchical structure of the proposed classification scheme, exclusion rules should be implemented. For example, a soundscape cannot be disruptive if it is backgrounded or it cannot be supportive if it is disruptive. In mathematical form, this implies a transformation of the membership degree:

$$S'_{backgrounded} = S_{backgrounded}$$
$$S'_{disruptive} = S_{disruptive}(1 - S_{backgrounded})$$
$$S'_{calming} = S_{calming}(1 - S_{disruptive} - S_{backgrounded})$$
$$S'_{stimulating} = S_{stimulating}(1 - S_{disruptive} - S_{backgrounded})$$

where the AND and NOT operator were implemented as a probabilistic t-norm and fuzzy negation.

The above procedure was applied to each soundscape-participant combination. For each soundscape, the average membership over all participants on the four classes was also calculated.

Next to this, participants also evaluated each soundscape in terms of the 2D core affect model ("full of life and exciting", "chaotic and restless", "calm and tranquil" and "lifeless and boring") on an 11-point scale. Similarly, the average score using the 2D core affect model quadrant categories for each soundscape was also calculated.

4.2.2.5 Psychoacoustical indicators and saliency

A preliminary study (Appendix) showed that either ambisonics or binaural recordings could be used for the reproduction. The gain of the ambisonics audio tracks has been adjusted such that their level is as close as possible to that of the corresponding binaural audio tracks. As the binaural tracks were recorded with a fully calibrated setup, the acoustical properties of the recordings are calculated on the basis of the one-minute binaural tracks using HEAD acoustics ArtemiS 8.3. The values for equivalent A-weighted sound pressure level (L_{Aeq}), percentile (L_{Axx}) and maximum sound levels (L_{AFmax}) were calculated as the energetic average of both left and right ears, whereas the values for loudness (N), sharpness (S) and corresponding percentile and maximum values were calculated as the arithmetic average between left and right ears.

Sounds that are noticed have a strong influence on the perception of soundscape (Kang et al., 2016, Terroir et al., 2013, De Coensel et al. 2009). Noticing of the sound is influenced by two interchanging processes: top-down and bottom-up attention. Top-down attention is voluntary: it assumes an active listening for the sounds occurring in the environment. On the other hand, bottom-up attention is involuntary and is influenced by the sonic environment alone.

To investigate the bottom-up attention to sound, saliency as a concept is introduced. Saliency indicates how much the specific sound or a sound event stands out of its background. In consequence, the higher the saliency, the higher the probability of a sound being noticed. Although related to perception, it is possible to define the physical characteristics that contribute to saliency (Kaya and Elhilali, 2017). In this study, we used a computational model (Filipan et al., 2018) which calculates the saliency of the sound by simulating several aspects of the measured physiological response of the brain. This saliency model has two processing stages implemented: auditory periphery and brain processing. Auditory periphery simulates the initial transformation of the sound from the acoustic wave to the firing of neurons. The second stage of the model is related to the sensitivity of the human auditory cortex to spectrotemporal modulations (Santoro et al., 2017; Schönwiesner and Zatorre, 2009) that are frequently encountered in speech and biological vocalizations. This reaction is simulated by mapping the tonotopically spaced output of the periphery to both amplitude (AM) and frequency modulation (FM) space. The mapping is achieved by using resonator filters for the AM and summation of the differently delayed signals across frequency bands for the AM/FM combination space. These signals are then fed through the sensory activation stage, a part of the model that simulates defocusing of the attention (Xue et al., 2014, Krause et al. 2013) by inhibiting the excitatory input.

To summarize the saliency of the sound in a single value indicator, all demodulated signals (spread over the frequency bands and AM/FM frequencies) are summed and saturated using a logarithm function. Finally, one-minute indicators for the time-evolution of the overall saliency are calculated: maximum (SL_max), average (SL_avg), median (SL_median) and 5, 10, 50, 90 and 95 percentile values (SL_xx).

4.2.2.6 Visual factors

The visual factors in each stimulus were also assessed, specifically the percentage of green pixels – a proxy for vegetation – and the number of people. The 50 stimuli were also labelled by the density of people appearing in the video using a qualitative 5-point scale, ranging from none (labelled as "1") to extremely dense (labelled as "5"). The proportion of each person density grade is 22%, 30%, 26%, 14%, 8% of the cases (from 1 to 5), respectively.

The opening scene in each stimulus was used to calculate the green area percentage. The digital pictures consisted of 4096 × 1632 pixels and were saved in .png format. The "RGB greenness" parameter G_{RGB} (Crimmins and Crimmins, 2008; Richardson et al., 2007) is used and calculated as G_{RGB} = (G-R) + (G-B), where G, R and B are the relative intensities of the green, red and blue channels in the RGB picture, respectively. A more robust assessment of green vegetation is the (broadband) normalized difference vegetation index (NDVI), however, requiring a measurement of near-infrared light. RGB greenness was shown to perform quite similar to NDVI in capturing the amount of vegetation as concluded by Richardson et al. (2007).

In a next step, an appropriate threshold was set. Note that all green is included when calculating G_{RGB}; so not only leaves from trees and bushes but also grass zones. Non-green vegetation is missed in this assessment. However, in this study, vegetation is predominantly green colored. Accidental non-vegetation green-colored objects were manually removed, typically accounting for only small zones in the photographs. Such a manual action was needed in less than 10% of the pictures. In Figure 6, examples are shown for a low, a moderate and a high vegetation percentage.

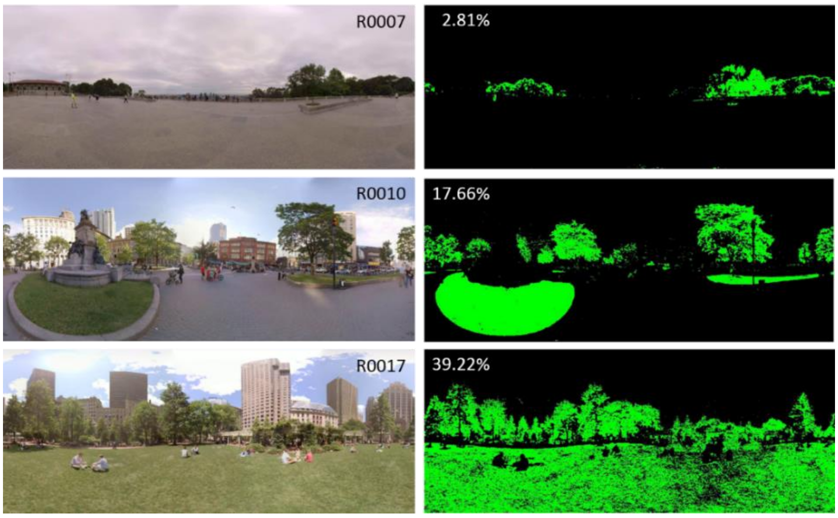


Figure 6 – Examples of opening scene of 360-degree videos, contain a low, a moderate and a high green percentage. (*Left*: the original photographs; *Right*: the corresponding photographs with only the pixels that were identified as green retained).

4.2.3 Statistical analysis

To observe relationships between the proposed soundscape categories, a principal component analysis (PCA) was performed. A PCA was also applied to the quadrant classifications in the 2D core affect model. Moreover, a mixed factor generalized linear model (GLMM) was constructed for the four proposed categories to analyze the contribution of underlying physical parameters to the classification. The fittest model for each soundscape category was looked for, using the Akaike Information Criterion (AIC) as model quality indicator (models with smaller AIC values fit better). Finally, predicting models from classification 1 and 2 were built via linear regression, to predict the scores on four soundscape categories. A receiver operating characteristic (ROC) analysis was made to check the prediction quality. The statistical analysis in this study was conducted using the SPSS statistics software (version 25).

4.3 Results

4.3.1 Audiovisual reproduction quality

Two items were analyzed regarding the quality of the proposed reproduction system: realism and immersion (Section 2.3.3). Earlier research proposed "plausibility" of a virtual acoustic environment, defined as "a simulation in agreement with the listener's expectation towards an equivalent real acoustic event" (Lindau and Weinzierl, 2012). The answers of the immersion and realism questions (see Figure 7), as a holistic measure, reveals the ecological validity of the experiment and the level of plausibility reached by the set-up. This proves that the carefully designed experiment and the VR 360-degree video paired with spatial audio reproduction allows the participants to be virtually present at the recording location.

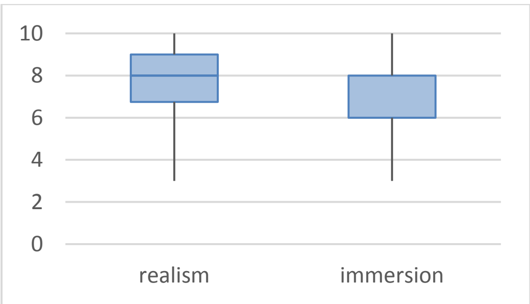


Figure 7– Realism and immersion of the reproduction quality.

4.3.2 Correlation between audiovisual perception and soundscape clustering

A crisp way to categorize the soundscapes is to compare the fuzzy membership to the proposed four classes. If the membership to one specific class is much larger than in the others, this soundscape is sorted in this class. Otherwise, this soundscape categorization remains unclear. Figure 8 shows the distribution of soundscapes that can be categorized into one of the four classes (i.e. 70.1% of cases), over the general audiovisual perception of the environment (answer to question 1). More specifically, *backgrounded* was found in 18% of the case, while *disruptive*, *calming*, *stimulating* was found in 18%, 14.5%, 19.6% of the cases, respectively.

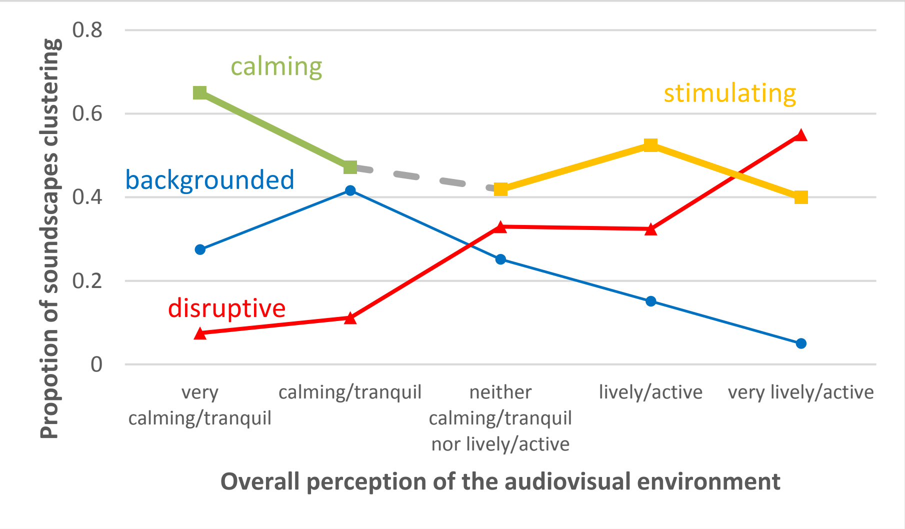


Figure 8 – Proportion of each soundscape category as a function of overall perception.

For the *backgrounded* category, the sound at the location does not lead to awareness of the acoustical environment. The distribution shows that an overall "very lively/active" environment is very unlikely if the soundscape is *backgrounded* but then tends more towards a "calming/tranquil" environment. The *disruptive* category shifts the curve towards the "lively/active" side making a "very calming/tranquil" overall environment very unlikely. The supportive soundscape (*calming* and *stimulating*) pushes the curve towards the extremes in overall perception. A higher proportion of *calming* soundscapes appears in the overall perception cases of "very calming/tranquil". It is striking that for the option "very lively/active", the proportion of *disruptive* soundscapes is higher than the proportion of *stimulating* soundscapes, which might suggest that a relatively larger number of environments with a non-supportive soundscape were selected as stimuli.

4.3.3 Principal component analysis

In Figure 1, soundscapes are divided into *backgrounded* and foregrounded by attention causation. The foregrouded soundscapes consist of three categories, corresponding to the negative and positive effects. A principal component analysis (PCA) is applied to the average score on *disruptive*, *calming* and *stimulating* for 50 stimuli. Figure 9a shows the triangle of three foregrounded soundscape categories in the plane spanned by the two principal components. In particular, component 1 explains 71.06% of variance, while component 2 explains 22.09%.

The average score on the four proposed soundscape classifications forms a 4×50 size matrix, with values varying from 0 to 1. A threshold is set to the matrix for binary results to highlight the most pronounced 25% of the scores in the matrix. The threshold is set at 0.32, and 53 values out of 200 are greater than this threshold. It is found that 29 soundscapes clearly belong to one of the four proposed categories (*backgrounded*: 9, *disruptive*: 7, *calming*: 3, *stimulating*: 10), 12 soundscapes cover two categories and 9 soundscapes cannot be sorted into any of these categories. Figure 9a shows the distribution of 50 soundscapes in the PCA analysis, they are colored based on the binary results of the proposed classification.

As a comparison, the scores on four quadrant categories in the 2D core affect model also forms a 4×50 size matrix. A threshold of 5.79 is set to the matrix to highlight the most pronounced 25% of the scores. 52 values out of 200 are greater than the threshold in the matrix. It is found that 28 soundscapes are determined by one of the four quadrant categories (chaotic: 6, exciting: 6, tranquil: 16, boring: 0), 12 soundscapes cover two categories and 10 soundscapes cannot be sorted into any of these categories. In Figure 9b, 50 soundscapes are colored based on the binary results in the 2D core affect model.

Similarly, a PCA is also applied to the four quadrant categories in the 2D core affect model. In Figure 10a, component 1 explains 55.1% of variance, while component 2 explains 30.9%. Also, Figure 10 shows the distribution of 50 soundscapes in PCA analysis, colored by the 2D core affect model classification and the proposed classification, respectively.

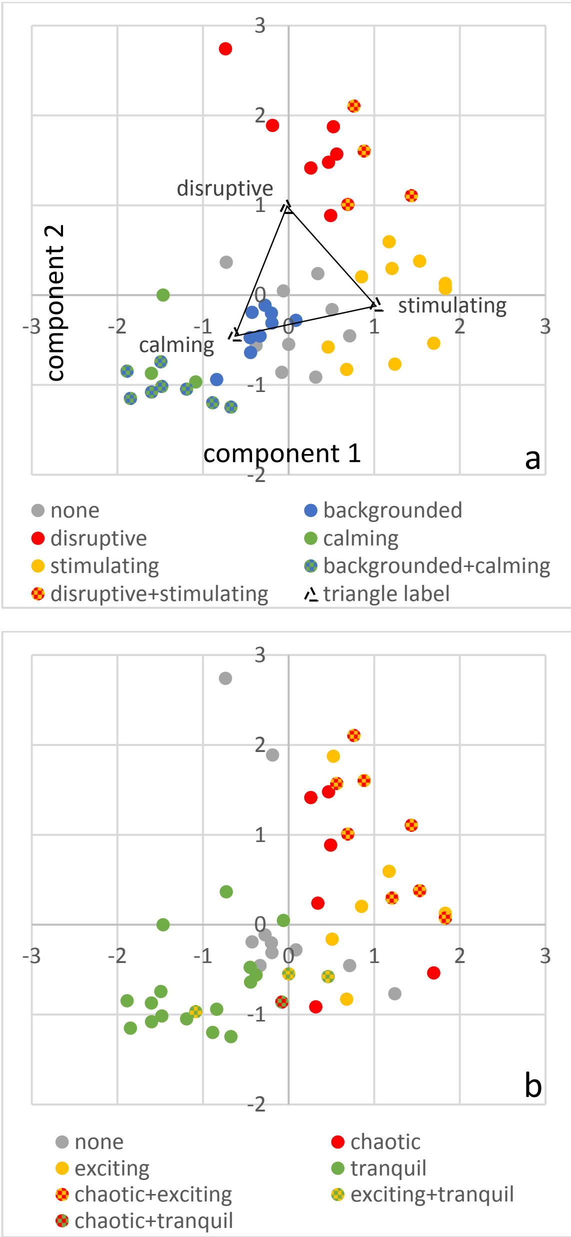


Figure 9 – Component plot based on fuzzy classification in rotated space; a: (triangle label) and 50 soundscapes distribution (colored in proposed classification); b: 50 soundscapes distribution (colored in 2D core affect model classification).

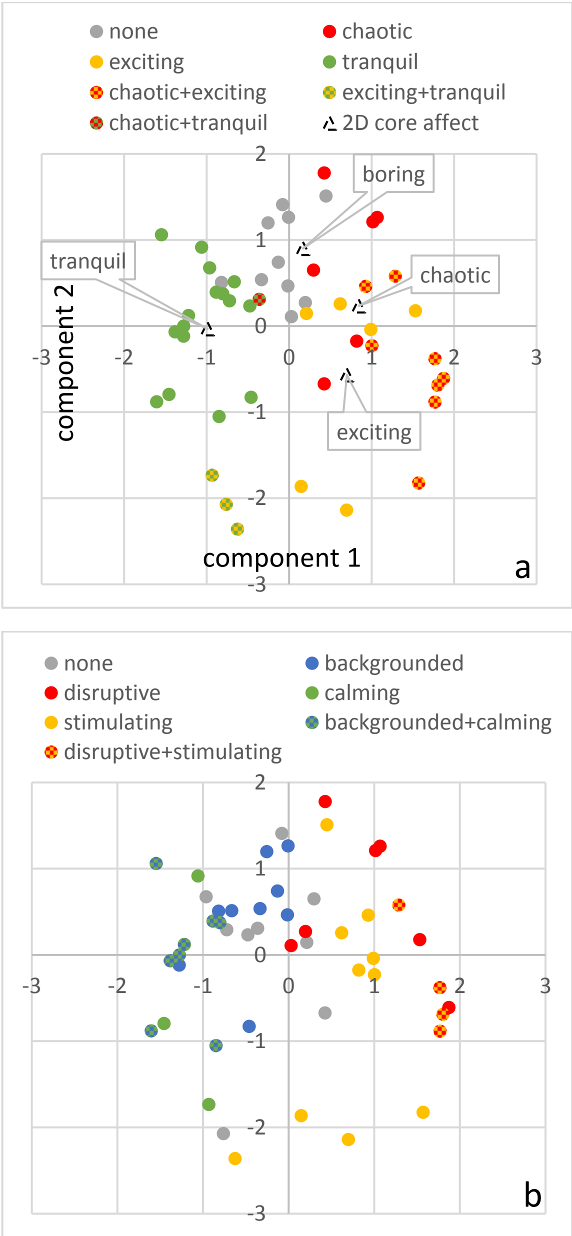


Figure 10 – Component plot based on answers to the core affect model question in rotated space; a: 50 soundscapes distribution (colored by the 2D core affect model classification); b: 50 soundscapes distribution (colored by the proposed classification).

4.3.4 Factor analysis

4.3.4.1 Relationships between soundscape class and memorization

During the small break in between experiencing 5 environments (see Section 2.3.3), a question about the memorization degree of the soundscape was asked, with the corresponding picture presented. To evaluate whether this memorization degree has a correlation with the scores on the proposed four soundscape categories, a mixed factor generalized linear model fit was applied, using participants as random factor. It is found that the memorization has significance in *backgrounded* ($F_{1,498}$=25.626; p<0.001) and *disruptive* ($F_{1,498}$=6.814; p<0.01), but not in *calming* ($F_{1,498}$=2.238; p>0.05) and *stimulating* ($F_{1,498}$=3.745; p>0.05). Naturally, the score of the *backgrounded* category has a negative correlation with memorization, while for the *disruptive* category, it is positively correlated.

4.3.4.2 Physical factors explaining soundscape classification

Taking into account all above-mentioned factors, a mixed factor generalized linear model fit was applied, with a stepwise method and using participant as random factor. Table 3 shows the fittest model results, with the Akaike Information Criterion (AIC) as a model quality indicator. The results suggest that the physical parameters that were tested fit the *backgrounded* category model best. All categories involve both acoustical factors and visual factors, except for the *disruptive* category. This might indicate that in a *disruptive* soundscape, the sound is dominating the perception.

Table 3 – Generalized linear mix model results of proposed soundscape categories.

glmm	AIC		F	df1	df2	coefficient	sig.
backgrounded	319.231	corrected model	48.081	5	994	0.458	0.000
		L_{A05}	55.591	1	994	-0.041	0.000
		N_{05}	30.428	1	994	0.023	0.000
		S_{max}	19.228	1	994	-0.068	0.000
		SL_median	10.011	1	994	-0.037	0.002
		Green pixels	6.827	1	994	-0.116	0.009
disruptive	511.113	corrected model	29.200	8	991	-1.432	0.000
		L_{A95}	45.799	1	991	-0.525	0.000
		L_{A90}	43.224	1	991	0.547	0.000
		SL_95[#]	6.205	1	991	-0.035	0.013
		S_{50}	12.919	1	991	-0.480	0.000
		N_{05}	12.287	1	991	0.040	0.000
		N	5.469	1	991	-0.046	0.020
		S_{95}	6.886	1	991	0.302	0.009
		S_{05}	4.538	1	991	0.145	0.033

calming	591.150	corrected model	40.721	6	993	1.327	0.000
		L_{AFmax}	103.492	1	993	-0.020	0.000
		Person density	12.645	4	993	(=1)0.172 (=2)0.024 (=3)0.003 (=4)-0.057 (=5)0*	0.000
		S_{50}	22.805	1	993	0.106	0.000
stimulating	535.742	corrected model	40.829	5	994	0.755	0.000
		Person density	16.435	4	994	(=1)-0.196 (=2)-0.077 (=3)-0.064 (=4)0.091 (=5)0*	0.000
		SL_median	39.724	1	994	0.067	0.000

*: This coefficient is set to 0 because it is redundant.
#: SL_95: 95% exceed saliency level.

4.3.5 Soundscape classification prediction

The previous section explored the factors that could modify the membership set of the proposed four categories. As stated before, an important challenge is to create models based on acoustical parameters, that predict soundscape classification as accurately as possible within the context of the definition of soundscape. For this purpose, classification 1 and classification 2 (Table 1) that were conducted with two groups of totally different stimuli, and applied to two groups of different participants, will be treated as two independent data sets. As stated in section 2.2.4, each soundscape gets an average membership score for each of the proposed soundscape classes. We will investigate whether a model based on physical parameters that is extracted from one of the classifications can predict this membership score for the other classification.

4.3.5.1 Prediction models from classification 1

A linear regression on 25 stimuli in classification 1 is applied, using a stepwise approach to access all possible acoustical parameters. Table 4 shows the remaining predictors, as well as the detailed model for each class membership.

Table 4 – Results of linear regression for 25 stimuli in classification 1.

label	Soundscape category	R^2	SE	prediction equation – from classification 1	predictors	sig.
1-1	backgrounded	0.546	0.100	y=-0.017x+1.393	$x=L_{A05}$	0.000
1-2	disruptive	0.719	0.095	$y=0.029x_1-0.014x_2-0.922$	$x_1=L_{A05}$, $x_2=L_{A95}$	L_{A05}(0.000) L_{A95}(0.006)
1-3	calming	0.606	0.129	y=-0.023x+1.936	$x=L_{AFmax}$	L_{AFmax}(0.000)
1-4	stimulating	0.667	0.100	y=0.105x+0.722	x=SL_95	SL_95(0.001)

SE: Std. Error of the Estimate.

When applying the equations in Table 4, it is easy to get the predicted scores of proposed soundscape categories for 25 stimuli in classification 2. To compare this prediction with the experimental value in classification 2, a receiver operating characteristic (ROC) analysis is applied. Figure 11 shows the ROC curve of the prediction, referring the experimental binary results of classification 2 as criterion. The parameter in this ROC curve is the threshold for crisp classification. Table 5 further shows the detailed results of the model prediction quality.

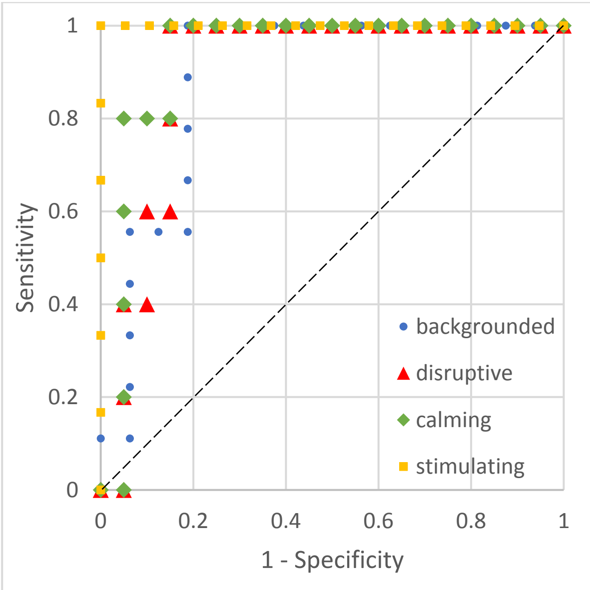


Figure 11 – ROC curve of predictions from classification 1.

Table 5 – The ROC curve area analysis from classification 1.

	Area Under the Curve				
	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
backgrounded	0.889	0.068	0.002	0.755	1.000
disruptive	0.900	0.063	0.007	0.777	1.000
calming	0.930	0.054	0.003	0.824	1.000
stimulating	1.000	0.000	0.000	1.000	1.000

a. Under the nonparametric assumption.

b. Null hypothesis: true area = 0.5.

As shown in Figure 11 and Table 5, the ROC curve shows the numeric results of the predictions. The Youden index (J) is often used as a criterion for selecting the optimum cut-off point (Schisterman et al., 2005). The Youden index is defined as shown in Eq. 1, and it ranges from -1 to 1. A higher value for J represents a lower proportion of totally misclassified results, i.e. a better prediction. Table 6 shows the maximum J value and its corresponding threshold.

$$J = sensitivity + specificity - 1 \qquad \text{(Eq. 1)}$$

Table 6 – Maximum Youden index for predictions (from classification 1) in proposed four category.

label	soundscape category	Highest *J*	Recommended threshold	Accuracy
1-1	backgrounded	0.812	0.3101	0.88
1-2	disruptive	0.85	0.1592	0.88
1-3	calming	0.85	0.4659	0.88
1-4	stimulating	1	0.1916	1

4.3.5.2 Prediction models from classification 2

Vice versa, the same procedure applies to classification 2. Table 7 shows the results of linear regression (stepwise) applied to classification 2 and the model details for each category. The prediction for 25 stimuli in classification 1 is compared with the binary results of the experimental value in classification 1, using ROC analysis (Figure 12). Table 8 further shows the detailed results of the prediction quality. Similarly, Table 9 shows the maximum *J* value and the corresponding threshold for predictions from classification 2.

Table 7 – Results of linear regression for 25 stimuli in classification 2.

label	Soundscape category	R^2	SE	prediction equation – from classification 1	predictors	sig.
2-1	backgrounded	0.603	0.113	y=-0.026x+1.894	x=L_{A05}	0.000
2-2	disruptive	0.360	0.148	y=0.020x-1.111	x=L_{A05}	0.002
2-3	calming	0.512	0.138	y= -0.028x$_1$+1.161x$_2$+1.76	x$_1$=L_{AFmax}, x$_2$=S_{50}	L_{AFmax}(0.000) S_{50}(0.027)
2-4	stimulating	0.663	0.090	y=0.023x-1.221	x=L_{A10}	L_{A10}(0.001)

SE: Std. Error of the Estimate

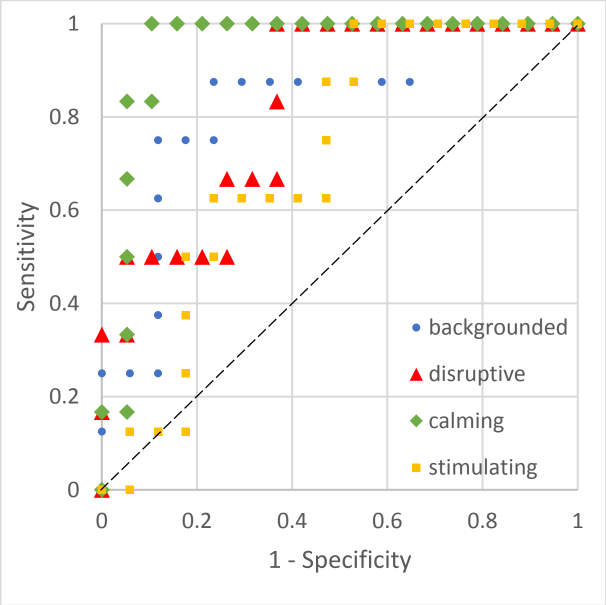


Figure 12 – ROC curve of predictions from classification 2.

Table 8 – The ROC curve area analysis from classification 2.

	Area Under the Curve				
	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
backgrounded	0.831	0.09	0.009	0.655	1.000
disruptive	0.825	0.089	0.019	0.65	0.999
calming	0.947	0.046	0.001	0.857	1.000
stimulating	0.713	0.103	0.091	0.511	0.915

a. Under the nonparametric assumption.

b. Null hypothesis: true area = 0.5.

Table 9 – Maximum Youden index for predictions (from classification 2) in proposed four category.

label	Soundscape category	Highest *J*	Recommended threshold:	Accuracy
2-1	backgrounded	0.64	0.107	0.8
2-2	disruptive	0.632	0.2644	0.72
2-3	calming	0.895	0.1184	0.92
2-4	stimulating	0.471	0.3037	0.64

4.3.5.3 Prediction quality comparison

Taking the recommended threshold, the numeric result is transferred into a dichotomous result. As stated before, the experimental binary results are used as criterion. In the ROC analysis, the accuracy ($\frac{true\ positive+true\ negative}{total\ sample}$) is indicating the proportion of total correctly classified results. Table 6 and 9 show the accuracy of each prediction taking the recommended threshold, respectively. They indicate that it is better to predict *backgrounded* soundscape with 1-1, and for *disruptive* and *stimulating* soundscape, 1-2 and 1-4 predicts better. Whereas for predicting a *calming* soundscape, 2-3 is clearly better. Another way to detect the quality of the predictions is considering the true positive to false positive rate (TPR to FPR). As shown in Figure 13, a smaller distance between prediction dots and point (0,1) indicates a higher prediction quality. The relative distance also indicates that for the proposed four categories, model 1-1, 1-2, 2-3 and 1-4 are optimized choices.

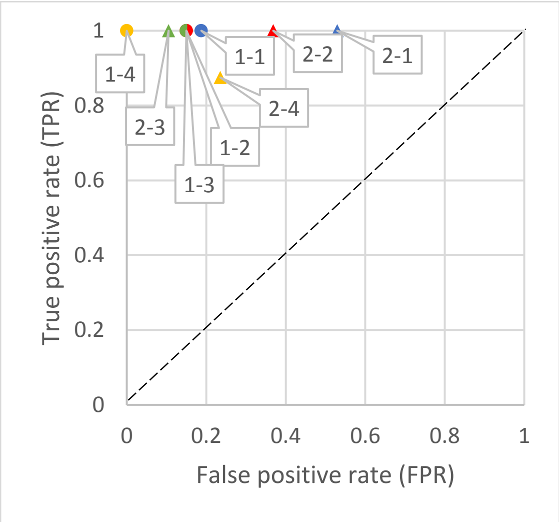


Figure 13– The ROC space and plots of the 8 predictions at recommended thresholds. (labels are referred to Table 6 and Table 9).

4.3.5.4 Models from all 50 stimuli

Based on the above comparison, a better model is selected for each category (model 1-1, 1-2, 2-3, 1-4). Table 10 gives the models that are built on the dataset of all 50 stimuli, with the indicators obtained from the optimized models built on the subgroups that best extrapolated to an independent dataset. Within this study, we cannot test this model with other recordings as verification. However, it can serve as a guideline once the new recordings and new subjective assessment are done.

Table 10 – Model details for all 50 stimuli.

label	Soundscape category	R^2	SE	prediction equation – from classification 1	predictors	sig.
0-1	backgrounded	0.521	0.112	y=-0.018x+1.464	x=L_{A05}	0.000
0-2	disruptive	0.488	0.128	y=0.027x_1-0.015x_2-0.733	x_1=L_{A05}, x_2=L_{A95}	L_{A05}(0.000) L_{A95}(0.006)
0-3	calming	0.426	0.150	y= -0.020x_1+0.079x_2+1.440	x_1=L_{AFmax}, x_2=S_{50}	L_{AFmax}(0.000) S_{50}(0.098)
0-4	stimulating	0.501	0.114	y=0.078x+0.643	x=SL_95	SL_95(0.000)

SE: Std. Error of the Estimate.

4.4 Discussion

4.4.1 Backgrounded soundscapes

Since this experiment was conducted in a soundproof lab and participants wore a headphone, which could inevitably have drawn the participants' attention to the sound. Hence, it can be expected that less soundscapes will be categorized as *backgrounded* in the VR setting than in the real world. Nevertheless, we opted to treat the *backgrounded* category rather strict and limited its membership function to answers “Not at all” (1) and “A little” (0.5). It should be noted that categorizing a soundscape as *backgrounded* excludes any of the foregrounded classes and hence, as soon as the sonic environment has even the smallest influence, it should be considered as foregrounded.

As the *backgrounded* soundscape is defined as a soundscape that does not contribute to the overall experience of the place by the question used to identify it, it is logical that this class of soundscapes does not catch a lot of attention. If not heard, such a soundscape will neither leave an impression in memory which is supported by a significant negative correlation with memorization (Section 3.4.1). This could be the preferred soundscape for private dwellings where inhabitants may prefer to add their own sounds. Earlier research (Axelsson, 2015b) categorized one outdoor space type as "my space", where crowds and mechanical sounds should be inaudible and sounds of nature and individuals should be only moderately audible. This supports the idea that *backgrounded* soundscapes are appropriate for “my space”. The distribution of this soundscape over general perception of environments shown in Figure 8, shows a trend towards an overall “calming/tranquil” perception of the environment. This reveals that a *backgrounded* soundscape is not very likely found in a lively and active environment, nor should it be. Nevertheless, some examples among the 50 stimuli used in this study, which are labelled as *backgrounded* determined based on their binary results (Section 3.3), contain road traffic and people talking (e.g.: R0002, R0017, etc.).

As the *backgrounded* soundscape does not attract attention, it covers a separate dimension and hence it was not included in the PCA (Section 3.3). In Figures 9&10, the stimuli labelled as *backgrounded* in the proposed classification scheme were labelled as “none” in 2D core affect model, i.e. not dominated by any category. This might be explained by the fact that a *backgrounded* soundscape could be allocated by all emotional components. It has been argued that a representative soundscape for the “lifeless and boring” label in the 2D core affect model seems rare (Axelsson, 2009; Bahalı and Tamer-Bayazıt, 2017), which is also the case in this study (Figure 10a). However, some *backgrounded* stimuli are located close to the “lifeless and boring” label in Figure 10b which might suggest that a “lifeless and boring” soundscape does not attract attention. Hence in an experiment that focusses attention on sound, either sonic environments that could lead to such a soundscape are not included or explicit foregrounding changes people’s perception. Note that this does not suggest that the *backgrounded* and “lifeless and boring” are completely overlapping since the two classifications are from different domains.

The generalized linearized model for individual soundscape classification with progressive inclusion of significant physical parameters shows that also visual factors contribute to the soundscape being *backgrounded* (Table 3 and Supplement). Visible green reduced the chance for a soundscape to become labelled as *backgrounded*. This is consistent with previous work highlighting the importance of visual factors in the construct of annoyance at home – the place where *backgrounded* soundscapes may be most appropriate (Gidlöf-Gunnarsson and Öhrström, 2007; Van Renterghem and Botteldooren, 2016). While comparing the fittest model for each soundscape category (Table 3), it seems that physical parameters built the best model for *backgrounded* (with lowest AIC compared to other categories), thus it seems easier to predict on the basis of physics when the sound environment will not be noticed. This is not an unexpected outcome.

The stable model for predicting *backgrounded* soundscapes (see Section 3.5) only retains L_{A5} as an acoustical indicator. To be *backgrounded*, sonic environments should simply not contain any loud sounds whatever their origin and whatever their duration. Note that focusing on the highest level using low percentile statistical indicators (or an equivalent level) is consistent with models for annoyance at home and the above observation that *backgrounded* soundscapes might be most appropriate for the environmental contribution to the private dwelling.

4.4.2 Disruptive soundscapes

Disruptive soundscapes are defined as sonic environments that prevent the users of the space from doing activities they would otherwise engage in. This conceptual soundscape relates very strongly to affordance and activity appropriateness as proposed in Nielbo et al. (2013) and Andringa and Van Den Bosch (2013). It is, to a certain extent, also aligned with the concept of "appropriateness", which has been suggested as key determinant of soundscape evaluation (Axelsson, 2015a).

Among all three foregrounded categories, *disruptive* is the only one that significantly correlates to memorization (Section 3.4.1), which suggests that such a soundscape leaves a strong – albeit negative – impression. The distribution of *disruptive* soundscapes over categories of overall appreciation of the environment shows an increasing trend towards "lively/active" and neutral evaluation (Figure 8). A straightforward interpretation is that *disruptive* soundscapes prevent the overall environment to be "calming/tranquil", yet it could be compatible with an environment that is neither calming nor lively or even with a "lively/active" environment. Soundscapes in this category tend to be loud, accompanied by a high density of people (Supplement 2).

It seems that *disruptive* is close to "chaotic and restless" in the 2D core affect model from the description, as well as certain overlaps in binary results of stimuli (Figure 9&10). In the PCA (Figure 9a), *disruptive* determined soundscapes are concentrated in the upper part of the triangle, while two outliers are slightly deviated to the negative axes of component 1. When analyzing these two outliers (R0013 & R0029), a shared trait was found: both stimuli contains a (visually) peaceful park, there are nearly no human activities and the weather is nice. In R0029, a honk from a boat appears all of a sudden. In R0013, a sustained noise from a lawnmower (not visible) appears in the background. These unexpected occurrences trigger some participants to report a disturbance while others chose to ignore these two stimuli and focus on the calming aspects of the soundscape. These two stimuli were labelled as "none" in the PCA analysis based on the 2D core affect model (Figure 9b).

The generalized linear model combines many non-orthogonal factors to predict the *disruptive* category but does not contain visual factors in the fittest model (Table 3). The dominance of sound in such a case is in line with many studies dealing with the perception of "unpleasant" soundscapes (Guastavino, 2006; Davies et al., 2013). Moreover, *disruptive* leads to the best prediction model among the three foregrounded categories (Table 3, AIC), which supports the use of the disruptive-supportive subdivision as second stage division (Figure 1).

Finally, looking at the predictive models for average soundscape classification (see also Section 3.5), additional insight in this category of soundscape can be obtained. The predictive models contain L_{A5} and L_{A95} as acoustic descriptors, or looking in more detail at the signs and magnitude of the coefficients, L_{A5} and L_{A5}-L_{A95}, both with a positive trend. This indicates that in addition to the sound level – measured here as L_{A5} – that also appears in the classification of *backgrounded*, the fluctuation of the sound – measured here as L_{A5}-L_{A95} – is important for the soundscape to become disruptive. Previous work has suggested the importance of the latter difference or a similar indicator of fluctuation, sometimes referred to as *emergence*, for predicting the pleasantness of public place soundscapes (Nilsson et al., 2007; Liu and Kang, 2015), as well as for annoyance at home (Bockstael et al., 2011), but never found such strong effects.

4.4.3 Calming soundscapes

Supportive soundscapes are expected to contribute to the overall experience of a place. They should match expectations created by the context and purpose of the place. In a design phase the type of support expected could be put forward by the urban designer. In this study the type of support one may expect, *calming* or *stimulating*, is mainly evoked by visual information. Therefore, in the procedure (Figure 5), questions 5a and 5b were only asked based on the answer in question 1 (i.e. when the overall perception is "calming/tranquil", it is assumed the soundscape would support the "calming/tranquil" atmosphere). If a not very "calming/tranquil" soundscape appears in an overall "calming/tranquil" environment, the fuzzy scores will only give a lower score for *calming*, rather than categorizing the soundscape as *stimulating*. Thus, *calming* and *stimulating* are not opposites of each other. Because of this construction, the combined distribution of *calming* and *stimulating* soundscapes over overall perception (Figure 8) is not very informative, but at least shows a somewhat stronger importance of the soundscape in "very calming/tranquil" environments.

Stimuli identified as "calm and tranquil" in the 2D core affect model also appear in the *calming* region of the PCA based on the proposed classification (Figure 9) and vice versa (Figure 10). This is not surprising as the distinction between the *calming* and *stimulating* type of supportive environments is mainly in the arousal dimension of core affect. In addition, the pleasantness dimension seems to bare some resemblance with not being disruptive. It is also found that the *calming* category is close to *backgrounded*, as 8 stimuli out 12 were identified as belonging to these two categories (Figure 9a). One possible explanation, focusing on attention, is that as the stimuli in *calming* soundscapes lead to passive attention fading (Bradley, 2009). This shifts the perception towards *backgrounded*. This vacillates the soundscape perception along the attention causation, which makes it stringent to label a soundscape as *calming*. However, despite the crossover between *calming* and *backgrounded*, these two categories are still different. Firstly, *calming* soundscapes make the overall environment being perceived as "calm and tranquil" and "very calm and tranquil" (Figure 8). Secondly, the percentage of (visual) vegetation is not a significant factor for explaining *calming* soundscapes (Table 3 and Supplement 2).

The *calming* category seems most difficult to predict from physical quantities (Table 3), which is not surprising given the high correlation between *backgrounded* and *calming* regarding physical parameters, and since attention causation in the first stage as division is stronger than arousal in the third stage (Figure 1). As for visual factors, a vegetation-dominated view is not a prerequisite for the soundscape to be classified as *calming* yet the visual presence of people plays a key role: too many people reduce the calmness of the soundscape.

Sharpness (S_{50}) and the absence of strong peaks (L_{AFmax}) appear both in the explorative GLM and the predictive models (see also Section 3.5). Sharpness is typically higher for natural sounds and lower for mechanical ones (Boes et al., 2018). A lot of research confirmed the positive effect of e.g. natural sounds (Payne, 2013, Van Renterghem, 2018) and the negative effect of mechanical sound (Bijsterveld, 2008).

4.4.4 Stimulating soundscapes

Finally, the *simulating* category is defined by the questionnaire as a soundscape that supports the liveliness and activeness of the environment. It is expected to arouse people, to encourage them to get involved. Music or music-like sound, for instance, could achieve such an effect (Botteldooren et al., 2006; Raimbault and Dubois, 2005), which was also found in some stimuli in this study (e.g., R0010, R0058, etc.). This type of soundscape helps the whole environment to be perceived as "lively/active" (Figure 8). However, compared to *disruptive*, a rather lower proportion of *stimulating* appears in an overall "very lively/active" perception. This might suggest that environments with such soundscapes attract people's attention but is slightly more likely to cause activity interference. Given a closer look at the 4 stimuli that are crossing these two categories (Figure 9a), all of them contain a lot of people, so some people may judge this crowd disturbing for their envisaged activities.

When putting *stimulating* soundscapes in the PCA plane of the 2D core affect model, they lay in between "chaotic and restless" and "full of life and exciting" (Figure 10a). As defined in the proposed classification, this category supports the liveliness and activeness of the environment. The GLM suggests that the presence of people is necessary (Table 3). It is consistent with previous research (van den Bosch et al., 2018), which suggests that human sounds add to the eventfulness of a soundscape and the perceived audible safety. It is worth noting that only when the visual person density is high, this category seems to be favored while lower person densities tend to favor *calming* soundscapes.

Finally, both the explanatory GLM and the predictive models (See also Section 3.5) for *stimulating* soundscapes contain the continuous fraction of saliency. Saliency, as defined in the model based on amplitude and frequency modulations, focusses strongly on vocalisations. Hence it is also indicative of the presence of human sounds. Previous work showed that the second order time derivative of the level in the 500 Hz octave band – which is also an indicator for amplitude fluctuations – correlates well with the presence of human voices (Aumond et al., 2017).

4.4.5 The soundscape classification approach

The main goal of this study was to propose and operationalize a coarse, holistic soundscape classification method and propose it as a labeling tool for audio-visual collections. This classification is not expected to be covering all details and further taxonomy could be used. The proposed classification is based on the contribution of the soundscape to the whole environmental perception.

The proposed classification scheme is unique in recognizing that in context, environmental sounds may remain backgrounded and that only sonic environments containing foregrounded elements may significantly contribute to the overall experience of the urban environment. Thus the *backgrounded* class is introduced as an orthogonal dimension.

A good classification of the remaining foregrounded soundscapes: *disruptive*, *calming* and *stimulating* should be minimally overlapping or maximally separated and therefore form a triangle in the principle component space. This was proven to be indeed the case. Moreover, although the classes slightly overlap and soundscapes may have a finite fuzzy membership to multiple classes at the same time, a tendency for good separation is indeed visible (Figure 9a). Recent research (Kamenický, 2018) also uses a triangle (activities, mechanisms and presence) for classification, which suggests a spectrum evolution of soundscapes in between the extremes. The evolution between soundscape categories is also embodied by the stimuli crossing two categories. It suggests that the soundscape perception is fluid and could be modified by time, person and context (Maris et al., 2007; Sun et al., 2018c).

The proposed classification is compared to the popular classification in a 2D core affect plane. There are some obvious similarities between both classifications yet in the plane of the first two principle components classes, the latter seems less separated. This could be because another dimension is sampled and the core affect classification is richer, but as the variance explained by the first two components is even higher than for the proposed classification, this does not seem the case. This might suggest that in a given soundscape (with fixed physical parameters), detecting attention causation is easier than classifying emotion perception. It highlights the importance of involving attention causation in soundscape classification. None of the observed soundscapes is dominantly "boring" as observed above, which argues in favor of eliminating this dimension. It should be noted however that in this study, the data for the proposed classification were collected right after each stimulus, while the data of the 2D core affect model were collected afterwards (Section 2.2.3). This might introduce the deviation of acoustical memory in perception (Darwin and Baddeley, 1974). However, no significant correlation was found between memorization and any of the four categories in the 2D core affect model.

Understanding the soundscape needs to isolate it from the whole environment that contains more than the sonic environment, but it is also important to use the whole environment as a guideline to classify the soundscape. Visual context, specifically two items in this study (Supplement 2), were found significant in both whole environment perception and the crisp clustering, though the latter represents 70.1% of the variance (Section 3.2). This is not the case in some of proposed categories. For example, for *disruptive*, the visual factors do not influence significantly. On the other hand, the soundscape also modifies the overall perception (e.g., two outliers in *disruptive* category).

Although soundscape – by definition – involves perception within context, a classification of sonic environments with soundscape in mind should benefit from capturing common understanding by society rather than personal preferences. Hence the proposed classification avoided the pleasantness dimension in affect which is expected to be more individual than the arousal dimension. If this attempt to remove individual differences from the classification was successful, it should be possible to construct predictive models solely based on physical parameters. This will be shown in the next Section.

4.4.6 Prediction models

The main goal of building prediction models is labelling new audio-visual recordings in the collection without the use of a panel. As the main application of the collection is to provide representative exemplars for each category, the prediction models do not need the refinement to resolve ambiguous situations and therefore could be based on a limited database of 50 samples. Another goal of building a model purely based on acoustical parameters could be to construct "soundscape maps". Also for this application simple models are preferred.

Thus, in this study, models predicting soundscape classification with a limited number of acoustical parameters were considered. The strongest possible model validation was assured by confirming model performance on the outcome of independent experiments. The linear models produce a membership degree for each of the four classes. Model comparison is done on sharp, binary classifications. The choice of threshold allows to balance between the risk of obtaining false positives and false negatives.

For model validation, the recommended threshold is based on the Youden Index which selects an optimal balance between sensitivity and specificity. This results in most crisp classification models combine the highest possible specificity with the highest possible sensitivity and appear in the upper left corner of Figure 13 (7 out of 8 dots). The recommended threshold for each model (Table 6&10), is lower than the value used to crisply classify the experimental results (0.32). This causes more than 25% data to be classified and therefore the model approach is less critical than the experimental approach. This may lead to false classification but it ensures that all possible example in each category are selected. Because it includes some soundscapes into one category unnecessarily, it might need additional panel tests to purify the selected soundscapes.

An alternative way to select the threshold is to push the outcome to maximal specificity (i.e. minimal FPR component). This method ensures that all automatically selected soundscapes are representative exemplars of a certain category, but it faces the fact that some soundscapes that could be a representative of a certain category, will be filtered out. As more audiovisual recordings are thus thrown out of the classification, this increases the work of site recording as a bigger collection is needed to start from. Thus, both methods for selecting the threshold have advantages and drawbacks. The choice depends on whether panel tests costs more than site recording or the other way around.

Besides the comparison between the models built on subgroups, Table 10 gives the models from the data of all 50 stimuli. Based on this study, they cannot be rigorously bilaterally verified. However, model parameter selection from the best models for the two subgroups are used without adding new parameters, which should reduce the risk of overfitting on the pooled data. Coefficients are nevertheless optimized for the pooled data. The models of Table 10 are therefore our suggestions for best available models.

4.4.7 Limitations

The experimental approach used in this work has a few drawbacks. Although using audio-visual reproduction through virtual reality is a huge improvement over older methods to experience sonic environments in context, it still lacks other sensory context: odor, heat and humidity, etc. And, although the 360-degree visual scenery is a very strong que for setting the context, it does not contain all information about a place, its use, its socio-cultural meaning, etc. During the experiment, we also received feedback on the resolution of VR Rift glasses for which, at the moment, there is no significantly better alternative.

The selection procedure for collecting the audio-visual recordings in each city was rather stringent and recordings from cities from different continents were included in the study. Nevertheless, there may be some bias in the database used for constructing the models. The distributions of soundscape with a different person density are not evenly (Section 2.2.6) since the real recording needed to consider the accessibility and operating possibility (i.e.: safety, stability, etc.). It is natural that more recordings in the database were made with less people (e.g.: parks) rather than at crowded places (e.g.: a shopping street).

Regarding the models, we are convinced that additional indicators and alternative machine learning techniques could have been used. E.g. regarding visual factors, it only assessed two items, as many aspects were shown to have an impact on soundscape perception (e.g., sound source visibility, number of vehicles, etc.). The database is open and will be extended in future so we encourage researchers to use it to test their hypotheses.

4.5 Conclusion

This study proposes a hierarchical soundscape classification methodology that is grounded in attention causation and reflects the contribution of the soundscape to the overall perception of the environment. The methodology is made operational through a matching brief questionnaire. The proposed hierarchical classification scheme offers an alternative to the 2D core affect model, and is based on how well the soundscape is noticed, how it interferes with possible activities that could be performed at the site, and includes the overall appreciation of the environment. It (1) accounts for the existence of *backgrounded* soundscapes that do not catch attention; (2) forms a clear triangular construct between *disruptive*, *calming* and *stimulating*, which offers a clear separation of soundscape categories; (3) explores the multiple factors that might modify the four categories, both in terms of acoustics and vision. Finally, a set of models based on acoustical parameters is built to predict the partial membership to the proposed soundscape categories, which might be used to classify soundscapes without involving participants. It has a high proportion of correctly classified soundscapes, validated by verification on a completely independent dataset (other participants and other soundscapes). By using the proposed soundscape classification methodology, it is at least possible to identify the most pronounced examples in each category.

The methodology is developed with the classification of a repository of audiovisual recordings from around the world in mind, yet it could be applied in other application domains. It is tested on an ecologically valid, realistic and immersive soundscape reproduction system to be applied in a laboratory. This holistic method includes soundscape collection, on-site recordings and final playback.

Within the framework of the “Urban Soundscapes of the World” project, more soundscape recordings will gradually be added into the database. It is hoped that, together, this ecologically valid reproduction system and the models that automatically classify soundscapes as the recordings enter the database will allow to build a growing international collection. This will offer urban planners the most interesting exemplars worldwide for each type of soundscape, inspiring and guiding future urban sound planning and design.

Appendix

Preliminary study – Validation of the recording and playback protocol

Overview

With the virtual reality device presents the video, it is expected to pair with corresponding audio recording, that ensures a high quality and spatial effect. Note that the audio recording by GoPro Omni cameras itself was not used in this study. As the recording contains both ambisonics and binaural audio (Figure 2), it is essential to decide which audio recording performs better through headphone playback when combined with virtual reality. A preliminary experiment was designed for this purpose.

Binaural audio recordings, performed using an artificial head, are generally considered to provide the highest degree of realism. Using an artificial head, the sound is recorded as if a human listener is present in the original sound field, preserving all spatial information in the audio recording. The main disadvantage of binaural audio recordings is that the frontal direction, and as such the acoustic viewpoint of the listener, is fixed by the orientation of the artificial head during the recording. This drawback could in theory be solved using ambisonics audio recording (Gerzon, 1985), a multichannel recording technique that allows for unrestricted rotation of the listening direction after recording. In principle, this technique could therefore provide an alternative to binaural recordings in the context of soundscape studies. However, the ambisonics technique has its own disadvantages, such as the more complex process of playback level calibration and equalization as compared to the binaural technique, the necessity of head tracking and real-time HRTF updates in case of playback through headphones, and the limited spatial resolution that can be achieved with lower-order ambisonics recordings—to date, there are no truly portable higher-order ambisonics recording systems available. Nevertheless, (first-order) ambisonics has become the de facto standard for spatial audio in VR games and platforms providing 360 video playback such as YouTube or Facebook.

Material & Experiment setup

Five 1-minute recordings were chosen for experiment 1 (Table I). The stimuli contain a fixed HD video, cut out from the original video in the frontal viewing direction, and padded with black in order to obtain again a 360-degree spherical video that can be viewed through a head-mounted display. This creates a “window” effect, forcing the participant to watch only in the frontal direction (Supplement 3). Furthermore, these stimuli are created in two flavors: with first-order ambisonics spatial audio track (allowing for head rotation) and with binaural audio track (which provides a fixed, i.e. head-locked, listening direction).

Table I – Stimuli for validation experiment.

Label	City Location	Date Time	Longitude Latitude	$L_{Aeq,1min}$
R0001	Montreal Palais des congrès	2017/6/22 8:02	45.503457 -73.561461	65.8
R0012	Boston Boston Public Garden	2017/6/28 9:36	42.353478 -71.070151	62.5
R0030	Tianjin Century Clock	2017/8/24 16:00	39.13262 117.198314	63.2
R0038	Hong Kong Taikoo Shing	2017/8/29 17:07	22.286715 114.218385	64.6
R0055	Berlin Checkpoint Charlie	2017/9/10 12:08	52.507796 13.390011	66.5

The experiment setup is the same as described in Section 2.2.2. During the experiment, participants were seated inside a soundproof booth. Recordings are played back using a PC (placed outside the booth), equipped with the GoPro VR Player 3.0 software, which allows to play back video with spatial audio. The 360-degree video is presented through an Oculus Rift head-mounted display, and the participant could freely move the head and look around in all directions. The audio is played back through Sennheiser HD 650 headphones, driven by a HEAD acoustics LabP2 calibrated headphone amplifier. Stimuli with binaural audio track are automatically played back at the correct level, as the headphone amplifier and headphones are calibrated and equalized for the artificial head that made the recordings. The gain of the ambisonics audio tracks have been adjusted such that their level is as close as possible to that of the corresponding binaural audio tracks.

Procedure & Participants

Since 5 stimuli paired with 2 audio recordings were involved, these 10 videos were played randomly to participants (20 participants, 6 female, Age_{mean}=28.9 yr, standard deviation 2.8 yr, range: 25-35 yr). After each video, 6 questions were shown in the VR screen (Table II, Guastavino et al., 2007). Participants needed to answer each question on a 5-point scale by verbal talking.

Table II – Questions and scale.

Questions:	Answer (5-point scale)
1. The sonic environment sounds __ enveloping.	little – very
2. I feel __ immersed on the sonic environment.	little – very
3. Representation of the sonic environment:	poor – good
4. Readability of this scene:	poor – good
5. Naturalness, true to life:	not truthful – truthful
6. The quality of the reproduction is __.	poor – good

Results

Table III shows the results of the comparison between ambisonics (allowing head rotation) and binaural (head-locked) audio playback. The table shows, on a scale from 1 to 5, the median scores on the questions asked (similar results are obtained with average scores). When there is a difference in median between the binaural and ambisonics playback cases, the higher value is underlined.

Table III – Median score of 5 pairs of soundscapes in the second stage of the experiment (a: ambisonics, b: binaural).

Label	Envelop-ment		Immer-sion		Represen-tation		Read-ability		Realism		Overall quality	
	a	b	a	b	a	b	a	b	a	b	a	b
R0001	4.0	4.0	3.5	4.0	4.0	3.5	4.0	3.0	3.5	4.0	4.0	4.0
R0012	3.5	4.0	3.0	3.5	3.0	3.0	3.0	3.5	3.0	3.0	3.0	3.0
R0030	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
R0038	4.0	3.5	4.0	3.0	4.0	4.0	4.0	3.5	4.0	4.0	4.0	4.0
R0055	4.0	4.0	4.0	3.0	4.0	4.0	4.0	4.0	4.0	3.0	4.0	3.0

Earlier research (Guastavino et al., 2007) showed that ambisonics audio results in a high degree of envelopment and immersion. Intuitively, one would expect that the possibility of rotating one's head during playback would result in a higher degree of envelopment and immersion, as compared to the case when one's listening direction is locked. On the other hand, due to the limited spatial resolution offered by first-order ambisonics, one would expect the binaural reproduction to result in a higher degree of readability and realism. The results shown in Table III do not allow to draw these conclusions; using a two-sample *t*-test with significance level 0.05, no significant difference is found between both sound reproduction methods, for any of the perceptual dimensions considered. Moreover, the difference between soundscapes is found to be larger than between the audio reproduction methods; some differences are significant, e.g. between R0012 and R0030 regarding representation (both ambisonics and binaural) and realism (binaural), or between R0012 and R0055 regarding immersion (ambisonics), readability (ambisonics) and representation (both ambisonics and binaural). This pilot test therefore justifies the use of ambisonics in the first stage of the experiment; either reproduction method could have been used.

Supplement

1. Full list of stimuli can be found in flowing Youtube link:

https://www.youtube.com/playlist?list=PL7YplJbeU4sKnGbO_p3EZwClZnShSkkHY

2. Effect of visual factors

To evaluate the visual factors (Section 2.3.6), a mixed factor generalized linear model was built using the proposed four soundscape categories, with participant as a random factor. Moreover, this model is also applied to the general perception of the audiovisual environment (Figure 5, question 1) and the crisp clustering of stimuli (Section 3.2). As shown in Table A, person density is a significant factor in all above-mentioned outcomes, while green pixel fraction is only significant in *backgrounded*, audiovisual perception and crisp clustering (explained in Section 3.2).

Table A – Generalized linear mix model results on visual factors.

glmm		F	df1	df2	sig.
backgrounded	corrected model	13.260	5	994	0.000
	Person density	16.151	4	994	0.000
	Green pixels	9.524	1	994	0.000
disruptive	corrected model	7.454	5	994	0.000
	Person density	9.234	4	994	0.000
	Green pixels	2.271	1	994	0.099
calming	corrected model	23.877	5	994	0.000
	Person density	20.407	4	994	0.000
	Green pixels	0.549	1	994	0.459
stimulating	corrected model	31.714	5	994	0.000
	Person density	30.769	4	994	0.000
	Green pixels	0.829	1	994	0.363
audiovisual perception	corrected model	13.665	20	976	0.000
	Person density	14.326	16	976	0.000
	Green pixels	2.909	4	976	0.019
crisp clustering	corrected model	4.975	20	976	0.000
	Person density	4.451	16	976	0.000
	Green pixels	3.184	4	976	0.013

‘Participant' is used as random factor.

3. Snapshot of video example (R0001) used in validation experiment (Appendix).

5

Conclusion and future work

5.1 Conclusion

This dissertation has explored a number of aspects of soundscape perception and classification in a holistic way. In Chapters 2 and 3 of this dissertation, a series of audiovisual laboratory experiments into soundscape perception were discussed. Chapter 2 discussed an experiment conducted in a mock-up living room, with the goal to create a better understanding of the influence of window view on reported noise annoyance. Sound source visibility was found to have more impact on self-reported annoyance than green element visibility, and noise sensitivity was found to have the strongest statistical significant effect on annoyance. Chapter 2 then further explored the role of audiovisual interaction and multi-sensory attention in perception and appraisal of the sonic environment. A potential individual difference (termed audition/vision dominated) was discovered, which reflected the differences in reliability on the detection of auditive/visual cues between test persons. Chapter 3 then further explored this individual difference and rephrased it as “audiovisual aptitude”. This personal factor was found to be related to general attitude towards audiovisual stimuli, in reference to a number of other psychological effects. It was further shown that this personal factor differs from noise sensitivity, a known stable personality trait. Through reanalysis of two earlier experiments, audiovisual aptitude was found to modify the influence of visibility of vegetation on self-reported noise annoyance, and to influence the overall appraisal of walking across a bridge in virtually reality, in particular when a visually intrusive noise barrier is used to reduce highway traffic noise levels.

In Chapter 4, a hierarchical soundscape classification methodology was proposed, grounded in auditory attention and reflecting the contribution of the soundscape to the overall perception of the environment. This scheme offers an alternative to the 2D core affect model, and is based on how well the soundscape is noticed, how it interferes with possible activities that could be performed at the site, and how it influences the overall appreciation of the environment. The classification approach first accounts for the existence of *backgrounded* soundscapes that do not catch attention, and then forms a triangular construct between *disruptive*, *calming* and *stimulating*, offering a clear separation of soundscape categories. Subsequently, an ecologically valid, realistic and immersive soundscape reproduction system was presented. This holistic method involves soundscape collection through on-site immersive audiovisual recordings, and playback through a head-mounted display. Chapter 4 then finished with an attempt at automatic soundscape classification, with a set of models based on acoustical parameters, to predict the partial membership to the proposed soundscape categories. The prediction models were found to be accurate to a reasonable degree.

The results of the experiments conducted in this work contribute to the understanding of the perception and classification of urban soundscapes. As its main innovative aspects, this work

- showed that a personal factor labeled audiovisual aptitude modifies the effect of audiovisual interaction on perception, such that this personal trait should be addressed in urban design and urban planning;
- introduced a hierarchical soundscape classification method that is based on the contribution of the soundscape to the overall perception of the environment, taking into account the effect of auditory attention;
- presented an immersive soundscape recording and reproduction method, that combines spatial audio with 360-degree video, and showed its validity in terms of realism and immersion.

5.2 Limitations and future work

Although the results of the experiments carried out in this study are discussed in detail in the previous chapters, there are still a number of limitations related to the methodology that could be discussed. Although the perception experiment discussed in Chapter 2 was specifically designed to minimize the influence of auditory memory, still, a large number of personal factors could not be controlled with the experimental design, e.g. the mental status and the mood stability of the test persons may have varied over the different days over which the test took place. Psychoacoustical characteristics of the sound, such as frequency and temporal content, sharpness and loudness, have also been shown to change sound preference. These characteristics of sound are not explored to the fullest extent in the present work. Similar limitations apply to other visual factors influencing soundscape perception, such as space openness, brightness and color fullness. Moreover, as it should be stressed that perception is to be investigated in a holistic manner considering all contextual factors, it would be interesting to investigate the influence of other sensory factors, such as odor, heat or humidity.

The assessment of audiovisual aptitude discussed in Chapter 3 is based on the performance of participants on a detection task carried out within a laboratory context, which is less susceptible to judgmental biases that may affect self-assessments (e.g. effects of mode of questionnaire administration) (Bowling, 2005). This task was designed to be correlational with regard to personal factors, which are unlikely to be manipulated experimentally. However, it has been debated if personality even could be a causal factor, following the dictum “no causation without manipulation” (Holland, 1986). Besides, it might be difficult to quantify such a strong influencing personal factor with the limited sample size that was used in this study. Thus, audiovisual aptitude remains a hypothesis and definitely needs further investigation. Current experimental results from four scenarios might not be easily transferable to field interviews. However, an extensive test resulting in four categories of respondents might not be necessary in practice. With these thoughts in mind, the following steps are suggested, to establish a better understanding of audiovisual aptitude, needed for its possible future application:

1) To extend the current set of scenarios, applying the same sampling idea but using different scenes (including various attracting objects and deviant appearing durations). This extension should increase the variation within the scenarios, and thus form a broader dataset. Experiments should further be carried out with a more diverse set of participants (regarding cultural background etc.) and experiment material should be randomly chosen from the dataset to avoid bias. Comparison of the results of a series of experiments could verify the rationality of such a personal factor.

2) In recent years, the relation between noise sensitivity and particularities of auditory processing in the central nervous system has been investigated with the use of brain imaging techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) (Fedele et al., 2015). A recent study combined EEG and MEG to measure neural sound feature processing in the central auditory system, and found that high noise sensitivity is associated with altered sound feature encoding and attenuated discrimination of sound noisiness in the auditory cortex (Kliuchko et al., 2016). In this thesis, it was found that audiovisual aptitude is a similar but independent personal trait as compared to noise sensitivity. Thus, inspired by the trend of investigation into the underlying mechanisms of noise sensitivity, a similar procedure could also be applied to the further investigation of audiovisual aptitude. For instance, this approach could be used to investigate the neurological basis (e.g.

auditory cortex and visual cortex) for the differences in audiovisual aptitude between individuals.

3) Based on the insights gained in this work and future investigations, an audiovisual aptitude questionnaire may be constructed to make such process operational and easily adaptable. Following the example of the development of the Noise-Sensitivity-Questionnaire (NoiSeQ) (Schütte et al., 2007), the reliability of such a questionnaire (relative and absolute Generalizability-coefficient) should be above precision level 1 "accurate measurement" as described in ISO (2004). An audiovisual aptitude questionnaire might also differentiate between main domains of daily life (such as leisure, work, habitation, communication, and sleep) and ideally, the ratings should be age and gender independent.

4) Audiovisual aptitude is expected to be related to attention moderated auditory scene analysis. To further simplify the operational procedure for measuring audiovisual aptitude, the relationships between audiovisual aptitude, demographic information and/or other information that is easy to obtain should be investigated.

In the review of Van Renterghem (2018), the positive effect of vegetation on the perception of environmental noise has been shown to occur in many studies. In this thesis, this factor is not strongly pronounced, at least from first sight. In Figures 12 and 13 in Chapter 3, it can be seen that for the majority of participants (70.6%), self-reported noise annoyance is lower with a vegetation window view. However, with group 1 (29.4% of participants) having a strong opposite opinion, this factor of vegetation window view is not statistically significant overall. This actually raises the importance of recognizing the role of personal factors, as well as any interaction such as found in Chapter 2 (Figure 8a) and Chapter 3 (Figures 14 and 15). This finding might help to explain the contradictory results on the same factor in different studies. Specifically, people with different noise sensitivity react to the sound source visibility in an opposite way (Figure 8a); people in group 4 value the highest sound level better than other groups (Figure 14); people in group 3 prefer the last bridge design as compared to other groups (Figure 15).

On a similar note, in Table 3 (Chapter 4), the amount of green pixels (as a proxy of vegetation) only appears to be relevant in the case of *backgrounded* soundscapes but not in the other cases, particularly not in the case of *calming* soundscapes. Watts et al. (2013) found a close relationship between green space, as determined not only by the amount of greenery but also by the presence of natural landmarks in general, and perceived tranquility. However, in this work, the classification is based on the contribution of the soundscape to the calmness of a space, thus, the same strong dependence on the visual scene was not likely to be found. Other research has pointed out that subjective experience is more closely linked to the connectivity state of the auditory cortex than to its basic sensory inputs (Hunter et al., 2010). The latter study nevertheless confirmed that visual context can modulate functional connectivity of the auditory cortex with regions implicated in the generation of subjective states.

Other research indicates that the accessibility and potential use of nearby green areas reduces long-term noise annoyances and prevalence of stress-related psychosocial symptoms (Gidlöf-Gunnarsson and Öhrström, 2007). In this light, the definition of the hierarchical soundscape classification scheme proposed in Chapter 4 needs to be addressed, which is based on the contribution of the soundscape to the perception of the overall environment. In the procedure (Figure 5, Chapter 4), the overall perception (question 1) determines whether the participants will have to answer the *calming*/*stimulating* question (question 5a/b). The visual content most likely determines the overall (first) impression (e.g. a green space might be calming/tranquil), but it does not confirm how strong the soundscape is supporting this overall perception. The questionnaire focuses on the soundscape and thus reflects the limited influence of green on soundscape perception that was found.

Furthermore, even though the amount of green pixels and the person density are found to be significant in terms of audiovisual overall perception and crisp clustering (Table A, Supplement, Chapter 4), they are not included in the prediction model (section 4.3.5). There are several reasons of not doing so. First, the rationality of using green pixels as a proxy of vegetation remains to be discussed. The amount of green pixels does not account for the details of vegetation such as visual quality, the distance, the distribution, etc. (Nilsson et al., 2012). Second, the horizontal range of the visual field in humans is around 150 degrees (Traquair, 1938), whereas the recordings used in the experiment are 360 degrees. The amount of green pixels is analyzed for the whole 360-degree scene, however, participants were free to rotate their head and thus it was hard to track where exactly they were viewing in the video. Third, the density of people is not evenly distributed over the various scenes used in the experiment (none to extremely: 22%, 30%, 26%, 14%, 8%).

The locations for performing the audiovisual recordings discussed in Chapter 4 were selected using an online survey, and as such there might have been some bias in the database used for constructing the models. Although the 360-degree visual scene is a very strong cue for setting the context, it does not contain all information about a place, such as its use, its socio-cultural meaning, etc. Although it is a huge improvement to use virtual reality for playback as compared to presentation on a screen, other sensory context is still lacking. To conclude this, the knowledge of the role of the visual context in soundscape perception is still at an early stage, as well as our understanding of soundscape.

With regards to the soundscape collection, reproduction and classification approaches presented in this dissertation, the following suggestions might be explored in future research:

1) In the procedure of the experiment in Chapter 4, the presentation of the 360-degree video might be overwhelming and might shift the focus of attention from the sound to the video. As the questionnaire reminds the participants about the sound, it might have had an influence on the results of the experiment. It could be an idea to study the differences in perception using systematically manipulated sound environments (e.g. by varying the amount of noise and/or the amount of positive sound components), where the questions in the study only refer to the environment as a whole without mentioning the sound environment (e.g. Echevarria Sanchez et al., 2017). Such studies could be conducted to explore, for instance, at what levels distant road traffic noise in a city park starts to influence the overall appraisal of the place.

2) Table 10 (Chapter 4) presents a model based on 50 recording samples. It is suggested to test the model with new recordings, and thus to verify the rationality of the model. It could be that with more audiovisual recordings, the model will converge to a more stable state. Also, it is expected that other factors might show to be of relevance in the model.

3) In this study, only static (stationary) recordings are used. Within a given area, a moving recording of the public space could be conducted as well (e.g. Aumond et al., 2017). It would be a natural next step to map the soundscape distribution in the given area, using the proposed hierarchical scheme. This approach would result in a "city soundscape map", useful for local residents and other users, as people have a desire for certain environments to fulfill their own purpose. Such an action would also call for public attention to the sound environment in a proper, holistic, and participatory way, instead of only complaining about the traffic noise, for instance.

4) Various ways to make it possible for public space users to "compose" their own sound environments are currently being investigated (e.g. within the framework of the EU project C3Places). Users hereby manipulate the soundscape by temporarily introducing additional sounds with varying properties (e.g. sound type and sound level), controlled through their smartphone that is connected with loudspeakers placed in the public space. In such a way, a desired sound environment can be offered to specific users, which leads to a better satisfaction when using an urban space.

5) With the development of virtual reality and augmented reality, it has become possible to (visually) plan an urban design change before it is executed within an existing environment. In this way, urban designers can investigate the effects of different scenarios. To enhance the quality of such virtual environment designs, it is

suggested to include, next to sound, other contextual factors, such as odor, heat, or humidity. In the foreseen future, it will be possible for people to move (walk) in an existing environment through virtual reality glasses.

5.3 Urban sound planning

Urban sound planning stresses the importance of involving sound in urban planning. Landscape, soundscape, people and environment do not stand on their own, and should be considered in a holistic way. This PhD generated some new insights and foregrounded known factors that could influence the urban sound planning process in future.

The potential effects of the visual scene and in particular its dependence on individual differences between people was highlighted. Quantitative results relating the effect of window view on equivalent noise effect reduction could not be obtained due to the limitations discussed above, but this may not even be the goal given the strong interactions between multiple modalities. Is the sound environment ever perceived on its own? Regarding the work on audiovisual aptitude, some suggestions could be stated for urban planners and designers, even without quantitative results. In urban planning, certain rules should be followed regarding legal basis, ethics, politics, social issues, etc. (Levy, 2016). Essentially, the goal of urban planning and design is to build appropriate environments that fulfill their desired function and provide their desired experience as good as possible, especially with an appropriate soundscape (Aletta et al., 2016a). An important aspect for urban planners is the foreseen user of the space. Pre-investigation of the foreseen users of a public open space should be a prerequisite to plan an environment that suits the given group, especially when the users have special needs. Furthermore, urban planners may opt for a worst-case approach that leads to an acceptable perception of the living environment also for the most noise sensitive people and those that are not vision dominated. For instance, controlling vegetation visibility and sound source visibility should ensure that noise annoyance is acceptable for most people.

The aim of Chapter 4, as well as of the Soundscape of the World project, is to raise awareness of the importance of sounds in urban planning. In urban planning, function, landscape, accessibility, etc. is often first considered, and much less consideration is paid to sound design. Two audiovisual recording examples might be a good reference, as they were found to be outliers for *disruptive* soundscapes (R0013 and R0029, section 4.4.2, Chapter 4). Both contain a (visually) peaceful park, nearly no human activities and nice weather, which should not be perceived as *disruptive*. However, a sustained noise from a lawnmower (not visible) in R0013 and a sudden honk (from a boat) in R0029 totally shift the perception. A badly designed soundscape or an environment polluted with unwanted sounds might make a carefully planned urban space perceived as not tranquil at all. On the other hand, changing the sound environment might also shift the perception of a public space for the better. For instance, R0058 and R0060 both were recorded inside a park with a vegetation dominated view, and both also contain street music performance, which makes the soundscape perceived as *stimulating*. It is hoped that the ecologically valid reproduction system presented in this work, and the models that automatically classify soundscapes as recordings enter the database, will allow to build a growing international collection. The various good examples of urban spaces in this database offer urban planners the inspiration and guidance for future urban sound planning and design. Researchers are also encouraged to test their hypotheses using the database of immersive audiovisual recordings, which is freely available, and to contribute by enhancing the database.

Finally, as a side result, when urban sound designers give demonstrations on optimizing the urban environment, the ecological validity of the design of the experiment, its setup and context, should be respected. Specifically, the environment should sufficiently resemble the target environment and persons experiencing the environment should do this in a natural, unfocussed way. A two-track design (i.e. having participants perform a task while questions are asked with a hidden agenda) should take all relevant psychological effects into account (e.g. auditory memory). Realism and immersion can be increased with the use of new techniques (e.g. virtual reality, augmented reality, spatial audio). These techniques should definitely be included also by practitioners.

References

Abbott LC, Taff BD, Newman P, Benfield JA, and Mowen AJ. (2016). The influence of natural sounds on attention restoration. Journal of Park and Recreation Administration, 34(3), 5–15.

Adams MD, Bruce NS, Davies WJ, Cain R, Jennings P, Carlyle A, Cusack P, Hume K, and Plack C. (2008). Soundwalking as a methodology for understanding soundscapes.

Adelmann GW. (1998). Reworking the landscape, Chicago style. Hastings Center Report, 28(6), s6-s11.

Aletta F, Kang J, and Axelsson Ö. (2016a). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning, 149, 65-74.

Aletta F, Kang J, Fuda S, and Astolfi A. (2016b). The effect of walking sounds from different walked-on materials on the soundscape of urban parks. Journal of Environmental Engineering and Landscape Management, 24(3), 165-175.

Alvarsson JJ. Wiens S, and Nilsson ME. (2010). Stress recovery during exposure to nature sound and environmental noise. International journal of environmental research and public health, 7(3), 1036-1046.

Andringa TC, and Van Den Bosch KA. (2013). Core effect and soundscape assessment: Fore-and background soundscape design for quality of life. In INTER-NOISE and NOISE-CON congress and conference proceedings (Vol. 247, No. 6, pp. 2273-2282). Institute of Noise Control Engineering.

Apthorp D, Alais D, and Boenke LT. (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. Journal of vision, 13(5:3), 1-15.

Aumond P, Can A, De Coensel B, Botteldooren D, Ribeiro C, and Lavandier C. (2017). Modeling soundscape pleasantness using perceptual assessments and acoustic measurements along paths in urban context. Acta Acustica united with Acustica, 103(3), 430-443.

Axelsson Ö, Nilsson ME, and Berglund B. (2010). A principal components model of soundscape perception. The Journal of the Acoustical Society of America, 128(5), 2836-2846.

Axelsson Ö. (2009). May information load be the key dimension underlying soundscape perception?. In Inter-Noise 2009. Institute of Noise Control Engineering of the USA.

Axelsson Ö. (2015a). How to measure soundscape quality. In Proceedings of the Euronoise 2015 conference, Maastricht, The Netherlands. pp. 1477-1481.

Axelsson Ö. (2015b). Towards guidelines for soundscape design. In AESOP Prague Annual Congress 2015: Definite Space–Fuzzy Responsibility, Prague, Czech Republic. pp. 802-808.

Bahalı S, and Tamer-Bayazıt N. (2017). Soundscape research on the Gezi Park–Tunel Square route. Applied Acoustics, 116, 260-270.

Beaman CP. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(5), 1106-1118.

Bell R, Buchner A, and Mund I. (2008). Age-related differences in irrelevant-speech effects. Psychology and aging, 23(2), 377-391.

Berglund B, and Nilsson ME. (2006). On a tool for measuring soundscape quality in urban residential areas. Acta Acustica united with Acustica, 92(6), 938-944.

Berto R. (2005). Exposure to restorative environments helps restore attentional capacity. Journal of environmental psychology, 25(3), 249-259.

Bharadwaj HM, Verhulst S, Shaheen L, Liberman MC, and Shinn-Cunningham BG. (2014). Cochlear neuropathy and the coding of supra-threshold sound. Frontiers in systems neuroscience, 8.

Bijsterveld K. (2008). Mechanical sound: Technology, culture, and public problems of noise in the twentieth century. MIT press.

Bockstael A, De Coensel B, Lercher P, and Botteldooren D. (2011). Influence of temporal structure of the sonic environment on annoyance. In 10th International Congress on Noise as a Public Health Problem (ICBEN-2011) (Vol. 33, pp. 945-952). Institute of Acoustics.

Boes M, Filipan K, De Coensel B, and Botteldooren D. (2018). Machine Listening for Park Soundscape Quality Assessment. Acta Acustica united with Acustica, 104(1), 121-130.

Bolognini N, Frassinetti F, Serino A, and Làdavas E. (2005). “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. Experimental brain research, 160(3), 273-282.

Bolund P, and Hunhammar S. (1999). Ecosystem services in urban areas. Ecological economics, 29(2), 293-301.

Botteldooren D, Andringa T, Aspuru I, Brown AL, Dubois D, Guastavino C, Kang J, Lavandier C, Nilsson M, Preis A, and Schulte-Fortkamp B. (2015). From sonic environment to soundscape. Soundscape and the Built Environment; Kang J, Schulte-Fortkamp B (Eds), 17-42.

Botteldooren D, De Coensel B, and De Muer T. (2006). The temporal structure of urban soundscapes. Journal of sound and vibration, 292(1-2), 105-123.

Botteldooren D, Sun K, Echevarría Sánchez GM, De Coensel B, and Van Renterghem T. (2017). The influence of audio-visual aptitude on audio-visual interaction in appraisal of the environment. In 12th International Congress on Noise as a Public Health Problem (ICBEN 2017).

Botteldooren D, Verkeyn A, and Lercher P. (2001). How can we distinguish exposure and expectation effects in integrated soundscape analyses?. In Proceedings 17th International Congress on Acoustics, published on CD-ROM, Vol. II, Rome, Italy.

Botteldooren D, Verkeyn A, and Lercher P. (2002). Noise annoyance modelling using fuzzy rule based systems. Noise and Health, 4(15), 27-44.

Bowling A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of public health, 27(3), 281-291.

Bradley MM. (2009). Natural selective attention: Orienting and emotion. Psychophysiology, 46(1), 1-11.

Brambilla G, and Maffei L. (2006). Responses to noise in urban parks and in rural quiet areas. Acta Acustica united with Acustica, 92(6), 881-886.

Brockherhoff MP. (2000). An urbanizing world. Population Bulletin, 55(3), 3-44.

Brown AL, Gjestland T, and Dubois D. (2015). Acoustic environments and soundscapes. Soundscape and the Built Environment; Kang J, Schulte-Fortkamp B (Eds), 1-17.

Brown AL, Kang J, and Gjestland T. (2011). Towards standardization in soundscape preference assessment. Applied Acoustics, 72(6), 387-392.

Brown AL. (2012). A review of progress in soundscapes and an approach to soundscape planning. International Journal of Acoustics and Vibration, 17(2),73-81.

Bruce NS, and Davies WJ. (2014). The effects of expectation on the perception of soundscapes. Applied Acoustics, 85, 1-11.

Burchell RW, Shad NA, Listokin D, Phillips H, Downs A, Seskin S, Davis JS, Moore T, Helton D, and Gall M. (1998). The costs of sprawl – Revisited: Transportation Research Board Report 39 (No. Project H-10 FY'95), National Academy Press,Washington, DC.

Cain R, Jennings P, and Poxon J. (2013). The development and application of the emotional dimensions of a soundscape. Applied Acoustics, 74, 232-239.

Carles J, Bernáldez F, and Lucio JD. (1992). Audio-visual interactions and soundscape preferences. Landscape research, 17(2), 52-56.

Cartwright-Finch U, and Lavie N. (2007). The role of perceptual load in inattentional blindness. Cognition, 102(3), 321-340.

Chen J. (2007). Rapid urbanization in China: a real challenge to soil protection and food security Catena, 69(1), 1-15.

Cohen JI, and Gordon-Salant S. (2017). The effect of visual distraction on auditory-visual speech perception by younger and older listeners. The Journal of the Acoustical Society of America, 141(5), EL470-EL476.

Colavita FB. (1974). Human sensory dominance. Attention, Perception, & Psychophysics, 16(2), 409-412.

Collignon O, Girard S, Gosselin F, Roy S, Saint-Amour D, Lassonde M, and Lepore F. (2008). Audio-visual integration of emotion expression. Brain research, 1242, 126-135.

Conway ARA, Cowan N, and Bunting MF. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. Psychonomic bulletin & review, 8(2), 331-335.

Crimmins MA, and Crimmins TM. (2008). Monitoring plant phenology using digital repeat photography. Environmental Management, 41, 949-958.

Cycowicz YM, and Friedman D. (1998). Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. Brain and Cognition, 36(1), 30-51.

Darwin CJ, and Baddeley AD. (1974). Acoustic memory and the perception of speech. Cognitive Psychology, 6(1), 41-60.

Davies WJ, Adams MD, Bruce NS, Cain R, Carlyle A, Cusack P, Hall DA, Hume KI, Irwin A, Jennings P, and Marselle M. (2013). Perception of soundscapes: An interdisciplinary approach. Applied Acoustics, 74(2), 224-231.

De Coensel B, and Botteldooren D. (2006). The quiet rural soundscape and how to characterize it. Acta Acustica united with Acustica, 92(6), 887-897.

De Coensel B, Botteldooren D, and De Muer T. (2003). 1/f noise in rural and urban soundscapes. Acta acustica united with acustica, 89(2), 287-295.

De Coensel B, Botteldooren D, Berglund B, Nilsson ME, De Muer T, and Lercher P. (2007). Experimental investigation of noise annoyance caused by high-speed trains. Acta Acustica united with Acustica, 93(4), 589-601.

De Coensel B, Botteldooren D, De Muer T, Berglund B, Nilsson ME, and Lercher P. (2009). A model for the perception of environmental sound based on notice-events. The Journal of the Acoustical Society of America. 126(2), 656-665.

De Coensel B, De Muer T, Yperman I, and Botteldooren D. (2005). The influence of traffic flow dynamics on urban soundscapes. Applied acoustics, 66(2), 175-194.

De Coensel B, Sun K, and Botteldooren D. (2017). Urban Soundscapes of the World: selection and reproduction of urban acoustic environments with soundscape in mind. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 255(2), 5407-5413. Institute of Noise Control Engineering.

Delage B. (1980). Paysage sonore urbain: recherche n° 79-27, juin 1979. Plan construction.

Douglas O, and Murphy E. (2016). Source-based subjective responses to sleep disturbance from transportation noise. Environment International, 92-93, 450-456.

Dratva J, Zemp E, Dietrich DF, Bridevaux PO, Rochat T, Schindler C, and Gerbase MW. (2010). Impact of road traffic noise annoyance on health-related quality of life: Results from a population-based study. Quality of life research, 19(1), 37-46.

Dubois D, Guastavino C, and Raimbault M. (2006). A cognitive approach to soundscapes: using verbal data to access auditory categories. Acta Acust United Acust, 92(6), 865-874.

Echevarria Sanchez GM, Van Renterghem T, Sun K, De Coensel B, and Botteldooren D. (2017). Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. Landscape and Urban Planning, 167, 98-107.

Edwards B. (2016). A model of auditory-cognitive processing and relevance to clinical applicability. Ear and hearing, 37, 85S-91S.

EEA Report No. 10/2006: Urban Urban sprawl in Europe – the ignored challenge, ISSN: 1725-9177.

Ekkekakis P. (2008). Affect circumplex redux: the discussion on its utility as a measurement framework in exercise psychology continues. International Review of Sport and Exercise Psychology, 1(2), 139-159.

Ellermeier W, and Zimmer K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. The Journal of the Acoustical Society of America, 102(4), 2191-2199.

Erber NP. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech, Language, and Hearing Research, 12(2), 423-425.

Eriksen BA, and Eriksen CW. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Attention, Perception, & Psychophysics, 16(1), 143-149.

Evans GW, Lercher P, Meis M, Ising H, and Kofler WW. (2001). Community noise exposure and stress in children. The Journal of the Acoustical Society of America, 109(3), 1023-1027.

Ewing R, Schmid T, Killingsworth R, Zlot A, and Raudenbush S. (2008). Relationship between urban sprawl and physical activity, obesity, and morbidity. In: Marzluff J.M. et al. (eds) Urban Ecology (pp. 567-582). Springer, Boston, MA.

Fastl H, Kuwano S, Schick A.(eds.). (1996). Recent Trends in Hearing Research. Oldenburg, BIS Verlag.

Fedele T, Scheer HJ, Burghoff M, Curio G, and Körber R. (2015). Ultra-low-noise EEG/MEG systems enable bimodal non-invasive detection of spike-like human somatosensory evoked responses at 1 kHz. Physiological measurement, 36(2), 357-368.

Fields JM, and Walker JG. (1982). Comparing the relationships between noise level and annoyance in different surveys: A railway noise vs. aircraft and road traffic comparison. Journal of Sound and Vibration, 81(1), 51-80.

Fields JM, De Jong RG, Gjestland T, Flindell IH, Job RFS, Kurra S, Lercher P, Vallet M, Yano T, Guski R, and Felscher-Suhr U. (2001). Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. Journal of sound and vibration, 242(4), 641-679.

Fields JM. (1993). Effect of personal and situational variables on noise annoyance in residential areas. The Journal of the Acoustical Society of America, 93(5), 2753-2763.

Filipan K, Boes M, De Coensel B, Lavandier C, Delaitre P, Domitrović H, and Botteldooren D. (2017). The personal viewpoint on the meaning of tranquility affects the appraisal of the urban park soundscape. Applied Sciences, 7(1), 91.

Filipan K, De Coensel B, Aumond P, Can A, Lavandier C, and Botteldooren D. Auditory Saliency Triggers Change in Pleasantness Assessment of the Soundscape Perceived during Walking Trips through Urban Environments. (2018). *Submitted to Special Issue Advances in Environmental Acoustics, Building and Environment.*

Forster S, and Lavie N. (2016). Establishing the attention-distractibility trait. Psychological science, 27(2), 203-212.

Fougnie D, and Marois R. (2007). Executive working memory load induces inattentional blindness. Psychonomic bulletin & review, 14(1), 142-147.

Francis CD, and Barber JR. (2013). A framework for understanding noise impacts on wildlife: an urgent conservation priority. Frontiers in Ecology and the Environment, 11(6), 305-313.

Füllgrabe C, Moore BC, and Stone MA. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Frontiers in Aging Neuroscience, 6, 347.

Fyhri A, and Klæboe R. (2009). Road traffic noise, sensitivity, annoyance and self-reported health—A structural equation model exercise. Environment International, 35(1), 91-97.

Ge J, Lu J, Morotomi K, and Hokao K. (2009). Developing soundscapegraphy for the notation of urban soundscape: its concept, method, analysis and application. Acta Acustica United with Acustica, 95(1), 65-75.

Gerzon MA. (1985). Ambisonics in multichannel broadcasting and video. Journal of the Audio Engineering Society, 33(11), 859-871.

Giard MH, and Peronnet F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of cognitive neuroscience, 11(5), 473-490.

Gibson JJ, and Pick AD. (1963). Perception of another person's looking behavior. The American journal of psychology, 76(3), 386-394.

Gidlöf-Gunnarsson A, and Öhrström E. (2007). Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. Landscape and Urban Planning, 83(2-3), 115-126.

Gillespie MAK, Baude M, Biesmeijer J, Boatman N, Budge GE, Crowe A, Memmott J, Morton DR, Pietravalle S, Potts SG, Senapathi D, Smart SM, and Kunin WE. (2017). A method for the objective selection of landscape-scale study regions and sites at the national level. Methods in Ecology and Evolution, 8(11), 1468-1476.

Graham ER, and Burke DM. (2011). Aging increases inattentional blindness to the gorilla in our midst. Psychology and aging, 26(1), 162-166.

Guastavino C, Larcher V, Catusseau G, and Boussard P. (2007). Spatial audio quality evaluation: comparing transaural, ambisonics and stereo, In Proceedings of the 13th International Conference on Auditory Display (ICAD), Montréal, Canada.

Guastavino C. (2006). The ideal urban soundscape: Investigating the sound quality of French cities. Acta Acustica united with Acustica, 92(6), 945-951.

Guski R. (1999). Personal and social variables as co-determinants of noise annoyance. Noise and health, 1(3), 45-56.

Hao Y, Kang J, and Krijnders JD. (2015). Integrated effects of urban morphology on birdsong loudness and visibility of green areas. Landscape and Urban Planning, 137, 149-162.

Hartig T, Mang M, and Evans GW. (1991). Restorative effects of natural environment experiences. Environment and behavior, 23(1), 3-26.

Heinonen-Guzejev M, Vuorinen HS, Mussalo-Rauhamaa H, Heikkilä K, Koskenvuo M, and Kaprio J. (2005). Genetic component of noise sensitivity. Twin Research and Human Genetics, 8(3), 245-249.

Heinonen-Guzejev M. (2009). Noise sensitivity medical, psychological and genetic aspects. [Dissertation]. Helsinki: Department of Public Health, University of Helsinki.

Holland PW. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Hong JY, and Jeon JY. (2014). The effects of audio–visual factors on perceptions of environmental noise barrier performance. Landscape and Urban Planning, 125, 28-37.

Hong JY, and Jeon JY. (2015). Influence of urban contexts on soundscape perceptions: A structural equation modeling approach. Landscape and Urban Planning, 141, 78-87.

Hong JY, He J, Lam B, Gupta R, and Gan WS. (2017). Spatial Audio for Soundscape Design: Recording and Reproduction. Applied Sciences, 7(6), 627.

Honold J, Beyer R, Lakes T, and Van der Meer E. (2012). Multiple environmental burdens and neighborhood-related health of city residents. Journal of Environmental Psychology, 32(4), 305-317.

Hunter MD, Eickhoff SB, Pheasant RJ, Douglas MJ, Watts GR, Farrow TF, Hyland D, Kang J, Wilkinson ID, Horoshenkov KV, and Woodruff PW. (2010). The state of tranquility: Subjective perception is shaped by contextual modulation of auditory connectivity. Neuroimage, 53(2), 611-618.

Ishihara S. (1957). Test for Colour Deficiency – 24 Plates Edition. Tokyo: Kanehara Shuppan, 24.

ISO (2003). ISO/TS 15666:2003. Acoustics—assessment of noise annoyance by means of social and socio-acoustic surveys. Geneva: International Organization for Standardization.

ISO (2004). ISO 10075-3. Ergonomic principles related to mental workload - Part 3: Measurement and assessment of mental workload. Genf: International Organization for Standardization.

ISO (2014). ISO 12913-1:2014 Acoustics — Soundscape — Part 1: Definition and Conceptual Framework. Geneva: International Organization for Standardization.

ISO (2018). ISO/PRF TS 12913-2. Acoustics—Soundscape—Part 2: Data collection and reporting requirements. Geneva: International Organization for Standardization.

Jakovljević B, Belojević G, Paunović K, and Stojanov V. (2006). Road traffic noise and sleep disturbances in an urban population: cross-sectional study. Croatian medical journal, 47(1), 125-133.

Janssen SA, Vos H, Eisses AR, 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), 3746-3753.

Jennings P, and Cain R. (2013). A framework for improving urban soundscapes. Applied Acoustics, 74(2), 293-299.

Jeon JY, Lee PJ, Hong JY, and Cabrera D. (2011). Non-auditory factors affecting urban soundscape evaluation. The Journal of the Acoustical Society of America, 130(6), 3761-3770.

Jeon JY, Lee PJ, You J, and Kang J. (2010). Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. The Journal of the Acoustical Society of America, 127(3), 1357-1366.

Jeon JY, Lee PJ, You J, and Kang J. (2012). Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces. The Journal of the Acoustical Society of America, 131(3), 2101-2109.

Jiang Y, and Chun MM. (2001). Selective attention modulates implicit learning. The Quarterly Journal of Experimental Psychology: Section A, 54(4), 1105-1124.

Job RFS. (1999). Noise sensitivity as a factor influencing human reaction to noise. Noise and Health, 1(3), 57-68.

Johnson MP. (2001). Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. Environment and planning A, 33(4), 717-735.

Kahneman D. (1973). Attention and effort (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kamenický M. (2018). Enhanced sound source composition methods for qualitative mapping of urban sound environment. In 11th European Congress and Exposition on Noise Control Engineering (Euronoise 2018).

Kang J, Aletta F, Gjestland TT, Brown LA, Botteldooren D, Schulte-Fortkamp B, Lercher P, van Kamp I, Genuit K, Fiebig A, and Coelho JL. (2016). Ten questions on the soundscapes of the built environment. Building and Environment, 108, 284-294.

Kang J, and Zhang M. (2010). Semantic differential analysis of the soundscape in urban open public spaces. Building and environment, 45(1), 150-157.

Kaplan R, and Kaplan S. (1989). The experience of nature: A psychological perspective. CUP Archive.

Kaplan S. (1995). The restorative benefits of nature: Toward an integrative framework. Journal of Environmental Psychology, 15(3), 169-182.

Kaya EM, and Elhilali M. (2017). Modelling auditory attention. Phil. Trans. R. Soc. B, 372(1714), p.20160101.

Kishikawa H, Matsui T, Uchiyama I, Miyakawa M, Hiramatsu K, and Stansfeld SA. (2006). The development of Weinstein's noise sensitivity scale. Noise and Health, 8(33), 154.

Kliuchko M, Heinonen-Guzejev M, Vuust P, Tervaniemi M, and Brattico E. (2016). A window into the brain mechanisms associated with noise sensitivity. Scientific reports, 6, 39236.

Koelewijn T, Bronkhorst A, and Theeuwes J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. Acta Psychologica, 134(3), 372-384.

Kondo HM, Toshima I, Pressnitzer D, and Kashino M. (2014). Probing the time course of head-motion cues integration during auditory scene analysis. Frontiers in neuroscience, 8, 170.

Krause B, Márquez-Ruiz J, and Cohen Kadosh R. (2013). The effect of transcranial direct current stimulation: a role for cortical excitation/inhibition balance?. Frontiers in human neuroscience, 7, p.602.

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

Latinjak AT. (2012). The underlying structure of emotions: A tri-dimensional model of core affect and emotion concepts for sports. Revista Iberoamericana de Psicología del Ejercicio y el Deporte, 7(1), 71-87.

Lavandier C, and Defréville B. (2006). The contribution of sound source characteristics in the assessment of urban soundscapes. Acta acustica united with Acustica, 92(6), 912-921.

Lavie N, and Fox E. (2000). The role of perceptual load in negative priming. Journal of Experimental Psychology: Human Perception and Performance, 26(3), 1038-1052.

Lavie N, Lin Z, Zokaei N, and Thoma V. (2009). The role of perceptual load in object recognition. Journal of Experimental Psychology: Human Perception and Performance, 35(5), 1346-1358.

Lavie N. (1995). Perceptual load as a necessary condition for selective attention. Journal of Experimental Psychology: Human perception and performance, 21(3), 451-468.

Léobon A. (1995). La qualification des ambiance sonores urbaines. Natures Sciences Société, 3(1), 26-41.

Leung TM, Xu JM, Chau CK, and Tang SK. (2017). The effects of neighborhood views containing multiple environmental features on road traffic noise perception at dwellings. The Journal of the Acoustical Society of America, 141(4), 2399-2407.

Levy JM. (2016). Contemporary urban planning. Taylor & Francis.

Li HN, Chau CK, and Tang SK. (2010). Can surrounding greenery reduce noise annoyance at home?. Science of the Total Environment, 408(20), 4376-4384.

Li Y, Li Y, Zhou Y, Shi Y, and Zhu X. (2012). Investigation of a coupling model of coordination between urbanization and the environment. Journal of environmental management, 98, 127-133.
Lindau A, and Weinzierl S. (2012). Assessing the plausibility of virtual acoustic environments. Acta Acustica united with Acustica, 98(5), 804-810.
Liu F, and Kang J. (2018). Relationship between street scale and subjective assessment of audio-visual environment comfort based on 3D virtual reality and dual-channel acoustic tests. Building and Environment, 129, 35-45.
Liu J, and Kang J. (2015). Soundscape design in city parks: exploring the relationships between soundscape composition parameters and physical and psychoacoustic parameters. Journal of Environmental Engineering and Landscape Management, 23(2), 102-112.
Liu J, Kang J, Luo T, and Behm H. (2013). Landscape effects on soundscape experience in city parks. Science of the Total Environment, 454, 474-481.
Longstreth R. (ed.). (2008). Cultural landscapes: balancing nature and heritage in preservation practice. Minneapolis: University of Minnesota Press.
Macdonald JSP, and Lavie N. (2011). Visual perceptual load induces inattentional deafness. Attention, Perception, & Psychophysics, 73(6), 1780-1789.
Mack A, and Rock I. (1998). Inattentional blindness. Cambridge: MIT Press.
Maffei L, Masullo M, Aletta F, and Di Gabriele M. (2013). The influence of visual characteristics of barriers on railway noise perception. Science of the Total Environment, 445-446, 41-47.
Margules CR, and Meyers JA. (1992). Biological diversity and ecosystem fragmentation – an Australian perspective. Ekistics, 59(356-357), 293-300.
Maris E, Stalen PJ, Vermunt R, and Steensma H. (2007). Noise within the social context: annoyance reduction through fair procedures. Journal of the Acoustical Society of America, 121(4), 2000-2010.
Matsuyama M, Nisimura R, Kawahara H, Yamada J, and Irino T. (2014). Development of a Mobile Application for Crowdsourcing the Data Collection of Environmental Sounds. In: Yamamoto S. (eds) Human Interface and the Management of Information. Information and Knowledge Design and Evaluation. HIMI 2014. Lecture Notes in Computer Science, vol 8521. Springer, Cham.
Miedema HM, and Oudshoorn CG. (2001). Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. Environmental health perspectives, 109(4), 409-416.
Miedema HM, and Vos H. (1999). Demographic and attitudinal factors that modify annoyance from transportation noise. The Journal of the Acoustical Society of America, 105(6), 3336-3344.
Miedema HM, and Vos H. (2003). Noise sensitivity and reactions to noise and other environmental conditions. The Journal of the Acoustical Society of America, 113(3), 1492-1504.

Miedema HM, and Vos H. (2004). Noise annoyance from stationary sources: Relationships with exposure metric day–evening–night level (DENL) and their confidence intervals. The Journal of the Acoustical Society of America, 116(1), 334-343.

Miller ZD, Hallo JC, Sharp JL, Powell RB, and Lanham JD. (2014). Birding by Ear: A Study of Recreational Specialization and Soundscape Preference. Human Dimensions of Wildlife, 19(6), 498–511.

Miller ZD, Taff BD, and Newman P. (2018). Visitor experience of wilderness soundscapes in Denali national Park and Preserve. International Journal of Wilderness, 24(2).

Moray N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. Quarterly Journal of Experimental Psychology, 11, 56-60.

Musacchia G, Sams M, Skoe E, and Kraus N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proceedings of the National Academy of Sciences, 104(40), 15894-15898.

Ndrepepa A, and Twardella D. (2011). Relationship between noise annoyance from road traffic noise and cardiovascular diseases: a meta-analysis. Noise and Health, 13(52), 251-259.

Neisser U, and Becklen R. (1975). Selective looking: Attending to visually specified events. Cognitive psychology, 7(4), 480-494.

Nielbo FL, Steele D, and Guastavino C. (2013). Investigating soundscape affordances through activity appropriateness. In Proceedings of Meetings on Acoustics ICA2013 (Vol. 19, No. 1, p. 040059). ASA.

Niemann H, Bonnefoy X, Braubach M, Hecht K, Maschke C, Rodrigues C, and Robbel N. (2006). Noise-induced annoyance and morbidity results from the pan-European LARES study. Noise and Health, 8(31), 63-79.

Nilsson ME, and Berglund B. (2006). Soundscape quality in suburban green areas and city parks. Acta Acustica united with Acustica, 92(6), 903-911.

Nilsson ME, Botteldooren D, and De Coensel B. (2007). Acoustic indicators of soundscape quality and noise annoyance in outdoor urban areas. In Proceedings of the 19th International Congress on Acoustics.

Nilsson ME, Jeon JY, Rådsten-Ekman M, Axelsson Ö, Hong JY, and Jang HS. (2012). A soundwalk study on the relationship between soundscape and overall quality of urban outdoor places. The Journal of the Acoustical Society of America, 131(4), 3474-3474.

Nilsson ME. (2007). Soundscape quality in urban open spaces. Proceedings of Inter-Noise 2007, Institute of Noise Control Engineering, Istanbul, Turkey.

Northam RM. (1975). Urban Geography. John Wiley & Sons, New York.

O'Shea DM, and Fieo RA. (2015). Individual differences in fluid intelligence predicts inattentional blindness in a sample of older adults: a preliminary study. Psychological research, 79(4), 570-578.

Oberfeld D, and Klöckner-Nowotny F. Individual differences in selective attention predict speech identification at a cocktail party. eLife. 2016;5:e16747.

Öhrström E, Barregård L, Andersson E, Skånberg A, Svensson H, and Ängerheim P. (2007). Annoyance due to single and combined sound exposure from railway and road traffic. The Journal of the Acoustical Society of America, 122(5), 2642-2652.

Öhrström E, Björkman M, and Rylander R. (1988). Noise annoyance with regard to neurophysiological sensitivity, subjective noise sensitivity and personality variables. Psychological Medicine, 18(3), 605-613.

Okokon EO, Turunen AW, Ung-Lanki S, Vartiainen A, Tiittanen P, and Lanki T. (2015). Road-Traffic Noise: Annoyance, Risk Perception, and Noise Sensitivity in the Finnish Adult Population. International Journal of Environmental Research and Public Health, 12(6), 5712-5734.

Oldoni D, De Coensel B, Boes M, Rademaker M, De Baets B, Van Renterghem T, and Botteldooren D. (2013). A computational model of auditory attention for use in soundscape research. The Journal of the Acoustical Society of America, 134(1), 852-861.

Ouis D. (2001). Annoyance from road traffic noise: a review. Journal of Environmental Psychology, 21(1), 101-120.

Pammer K, Korrel H, and Bell J. (2014). Visual distraction increases the detection of an unexpected object in inattentional blindness. Visual Cognition, 22(9-10), 1173-1183.

Payne SR. (2013). The production of a perceived restorativeness soundscape scale. Applied Acoustics, 74(2), 255–263.

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), 379-389.

Pedersen E, and Persson Waye K. (2004). Perception and annoyance due to wind turbine noise—a dose–response relationship. The Journal of the Acoustical Society of America, 116(6), 3460-3470.

Pedersen E, and Persson Waye K. (2007). Wind turbine noise, annoyance and self-reported health and wellbeing in different living environments. Occupational & Environmental Medicine, 64, 480-486.

Pilcher EJ, Newman P, and Manning RE. (2009). Understanding and managing experiential aspects of soundscapes at Muir Woods National Monument. Environmental Management, 43(3), 425–435.

Preis A, Hafke-Dyx H, Szychowska M, Kocinski J, and Felcyn J. (2016). Audio-visual interaction of environmental noise. Noise Control Engineering Journal, 64(1), 34-43.

PTCEC, (1998). Report of the Pennsylvania 21st Century Environment Commission. Harrisburg, PA: 21st Century Environment Commission.

Puyana-Romero V, Lopez-Segura LS, Maffei L, Hernández-Molina R, and Masullo M. (2017). Interactive Soundscapes: 360°-Video Based Immersive Virtual Reality in a Tool for the Participatory Acoustic Environment Evaluation of Urban Areas. Acta Acustica united with Acustica, 103(4), 574-588.

Raanaas RK, Evensen KH, Rich D, Sjøstrøm G, and Patil G. (2011). Benefits of indoor plants on attention capacity in an office setting. Journal of Environmental Psychology, 31(1), 99-105.

Raij T, Uutela K, and Hari R. (2000). Audiovisual integration of letters in the human brain. Neuron, 28(2), 617-625.

Raimbault M, and Dubois D. (2005). Urban soundscapes: Experiences and knowledge. Cities, 22(5), 339-350.

Richardson AD, Jenkins JP, Braswell BH, Hollinger DY, Ollinger SV, and Smith M. (2007). Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. Oecologia, 152, 323-334.

Rock I, and Harris CS. (1967). Vision and touch. Scientific American, 216(5), 96-107.

Russell JA. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161-1178.

Rychtáriková M, and Vermeir G. (2013). Soundscape categorization on the basis of objective acoustical parameters. Applied Acoustics, 74(2), 240-247.

Sandhu R, and Dyson BJ. (2016). Cross-modal perceptual load: the impact of modality and individual differences. Experimental brain research, 234(5), 1279-1291.

Santoro R, Moerel M, De Martino F, Valente G, Ugurbil K, Yacoub E, and Formisano E. (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. Proceedings of the National Academy of Sciences, 114(18), 4799-4804.

Sato T, Yano T, Björkman M, and Rylander R. (1999). Road traffic noise annoyance in relation to average noise level, number of events and maximum noise level. Journal of Sound and Vibration, 223(5), 775-784.

Schafer RM. (1969). The new soundscape. Don Mills: BMI Canada Limited.

Schafer RM. (1977). The tuning of the world. Alfred A. Knopf.

Schafer RM. (1993). The soundscape: Our sonic environment and the tuning of the world. Simon and Schuster.

Schafer RM. (1994). The Soundscape: Our Sonic Environment and the Tuning of the World. Destiny Books, Rochester, Vermont.

Schisterman EF, Perkins NJ, Liu A, and Bondell H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology, 16(1), 73–81.

Schlesinger HS, and Meadow-Orlans KP. (1972). Sound and sign: Childhood deafness and mental health (p. 7). Berkeley: University of California Press.

Schönwiesner M, and Zatorre RJ. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proceedings of the National Academy of Sciences, 106(34), 14611-14616.

Schreckenberg D, Griefahn B, and Meis M. (2010). The associations between noise sensitivity, reported physical and mental health, perceived environmental quality, and noise annoyance. Noise and Health, 12(46), 7-16.

Schulte-Fortkamp B, and Fiebig A. (2015). Impact of Soundscape in Terms of Perception. Soundscape and the Built Environment; Kang J, Schulte-Fortkamp B (Eds), 69-88.

Schütte M, Marks A, Wenning E, Griefahn B. (2007). The development of the noise sensitivity questionnaire. Noise and Health, 9(34), 15-24.

Scialfa CT, Esau SP, and Joffe KM. (1998). Age, target-distractor similarity, and visual search. Experimental Aging Research, 24(4), 337-358.

Semidor C. (2006). Listening to a city with the soundwalk method. Acta Acustica united with acustica, 92(6), 959-964.

Shepherd D, Heinonen-Guzejev M, Hautus MJ, and Heikkilä K. (2015). Elucidating the relationship between noise sensitivity and personality. Noise & health, 17(76), 165-171.

Simons DJ, and Chabris CF. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception, 28(9), 1059–1074.

Sommers MS, Tye-Murray N, and Spehar B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. Ear and hearing, 26(3), 263-275.

Sörqvist P, and Rönnberg J. (2014). Individual differences in distractibility: an update and a model. PsyCh journal, 3(1), 42-57.

Sörqvist P. (2010). Effects of aircraft noise and speech on prose memory: What role for working memory capacity?. Journal of Environmental Psychology, 30(1), 112-118.

Southworth M. (1969). The sonic environment of cities. Environment and Behavior, 1(1), 49-70.

Stansfeld SA. (1992). Noise, noise sensitivity and psychiatric disorder: epidemiological and psychophysiological studies. Psychological medicine monograph supplement, 22, 1-44.

Sun K, Botteldooren D, De Coensel B. (2018a). Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation. Proceedings of the 47th International Congress and Exposition on Noise Control Engineering. Institute of Noise Control Engineering.

Sun K, De Coensel B, Echevarria Sanchez GM, Van Renterghem T, and Botteldooren D. (2018b). Effect of interaction between attention focusing capability and visual factors on road traffic noise annoyance. Applied Acoustics, 134, 16-24.

Sun K, De Coensel B, Echevarria Sanchez GM, Van Renterghem T, and Botteldooren D. (2017). Effects of Human Factors on the Interaction between Visual Scene and Noise Annoyance. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 255(2), 5530-5536. Institute of Noise Control Engineering.

Sun K, De Coensel B, Echevarría Sánchez GM, Van Renterghem T, and Botteldooren D. (2016). Effects of sound source visibility on sound perception in living room environment. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 253(2), 6420-6425. Institute of Noise Control Engineering.

Sun K, Echevarria Sanchez GM, De Coensel B, Van Renterghem T, Talsma D, and Botteldooren D. (2018c). Personal audiovisual aptitude influences the interaction between landscape and soundscape appraisal. Frontiers in nPsychology, 9:780.

Tadeu AJB, and Mateus DMR. (2001). Sound transmission through single, double and triple glazing. Experimental evaluation. Applied Acoustics, 62, 307-325.

Terroir J, De Coensel B, Botteldooren D, and Lavandier C. (2013). Activity interference caused by traffic noise: Experimental determination and modeling of the number of noticed sound events. Acta Acustica united with Acustica, 99(3), 389-398.

Torija AJ, Ruiz DP, and Ramos-Ridao ÁF. (2014). A tool for urban soundscape evaluation applying support vector machines for developing a soundscape classification model. Science of the Total Environment, 482, 440-451.

Traquair HM. (1938). An Introduction to Clinical Perimetry, Chpt. 1. London: Henry Kimpton. 4–5.

Tress B, Tress G, Fry G, and Opdam P (eds.). (2006). From Landscape Research to Landscape Planning – Aspects of Integration, Education and Application. Dordrecht, The Netherlands: Springer.

Truax B. (2001). Acoustic communication (Vol. 1). Greenwood Publishing Group.

UN. (2000). World Urbanization Prospects: The 1999 Revision, New York: United Nations-Department of Economic and Social Affairs. Population Division, United Nations.

UN. (2004). World Urbanization Prospects: The 2003 Revision, New York: United Nations-Department of Economic and Social Affairs. Population Division, United Nations.

UN. (2014). World Urbanization Prospects: The 2014 Revision Highlights, New York: United Nations-Department of Economic and Social Affairs. Population Division, United Nations.

van den Bosch KA, Andringa TC, Post WJ, Ruijssenaars WA, and Vlaskamp C. (2018). The relationship between soundscapes and challenging behavior: A small-scale intervention study in a healthcare organization for individuals with severe or profound intellectual disabilities. Building Acoustics, 25(2), 123-135.

Van den Brink RL, Cohen MX, van der Burg E, Talsma D, Vissers ME, and Slagter HA. (2013). Subcortical, modality-specific pathways contribute to multisensory processing in humans. Cerebral cortex, 24(8), 2169-2177.

Van Kamp I, Job RFS, Hatfield J, Haines M, Stellato RK, and Stansfeld SA. (2004). The role of noise sensitivity in the noise–response relation: a comparison of three international airport studies. The Journal of the Acoustical Society of America, 116(6), 3471-3479.

Van Kamp I, Klæboe R, Brown AL, and Lercher P. (2015). Soundscapes, human restoration and quality of life. Soundscape and the Built Environment; Kang J, Schulte-Fortkamp B (Eds), 43-68.

Van Kamp I, Leidelmeijer K, Marsman G, and De Hollander A. (2003). Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study. Landscape and urban planning, 65(1-2), 5-18.

Van Renterghem T, and Botteldooren D. (2016). View on outdoor vegetation reduces noise annoyance for dwellers near busy roads. Landscape and Urban Planning, 148, 203-215.

Van Renterghem T, Bockstael A, De Weirt V, and Botteldooren D. (2013). Annoyance, detection and recognition of wind turbine noise. Science of the Total Environment, 456-457, 333-345.

Van Renterghem T. (2018). Towards explaining the positive effect of vegetation on the perception of environmental noise. Urban Forestry & Urban Greening.

Västfjäll D. (2002). Influences of current mood and noise sensitivity on judgments of noise annoyance. The Journal of psychology, 136(4), 357-370.

Västfjäll D. (2004). The “end effect” in retrospective sound quality evaluation. Acoustical Science and Technology, 25(2), 170-172.

Veisten K, Smyrnova Y, Klæboe R, Hornikx M, Mosslemi M, and Kang J. (2012). Valuation of green walls and green roofs as soundscape measures: Including monetised amenity values together with noise-attenuation values in a cost-benefit analysis of a green wall affecting courtyards. International journal of environmental research and public health, 9(11), 3770-3788.

Viollon S, Lavandier C, and Drake C. (2002). Influence of visual setting on sound ratings in an urban environment. Applied Acoustics, 63(5), 493-511.

Vorländer M. (2008). Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, Springer, Berlin.

Wada Y, Kitagawa N, and Noguchi K. (2003). Audio–visual integration in temporal perception. International journal of psychophysiology, 50(1), 117-124.

Walker AJ, and Ryan RL. (2008). Place attachment and landscape preservation in rural New England: A Maine case study. Landscape and Urban Planning, 86(2):141-152.

Wang JY, Da LJ, and Song K. (2008). Temporal variations of surface water quality in urban, suburban and rural areas during rapid urbanization in Shanghai, China. Environmental Pollution, 152, 387-393.

Watts G, Chinn L, and Godfrey N. (1999). The effects of vegetation on the perception of traffic noise. Applied Acoustics, 56(1), 39-56.

Watts G, Miah A, and Pheasant R. (2013). Tranquillity and soundscapes in urban green spaces—predicted and actual assessments from a questionnaire survey. Environment and Planning B: Planning and Design, 40(1), 170-181.

Weinstein ND. (1978). Individual differences in reactions to noise: a longitudinal study in a college dormitory. Journal of Applied Psychology, 63(4), 458-466.

Weinzimmer D, Newman P, Taff D, Benfield J, Lynch E, and Bell P. (2014). Human Responses to Simulated Motorized Noise in National Parks. Leisure Sciences, 36(3), 251–267.

Westerkamp H. (1974). Soundwalking. Originally published in Sound Heritage, Vol. III(4), Victoria B.C., revised 2001.

WHO. (1999). Guidelines for community noise. Edited by Berglund B, Lindvall T, and Schwela DH. World Health Organization.

WHOQOL Group. (1998). The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties. Social science & medicine, 46(12), 1569-1585.

Wikipedia contributors. (2018, August 11). Urban area. In Wikipedia, The Free Encyclopedia. Retrieved 14:01, August 13, 2018, from https://en.wikipedia.org/w/index.php?title=Urban_area&oldid=854484432

Wolfe DE. (1983). Effects of music loudness on task performance and self-report of college-aged students. Journal of Research in Music Education, 31(3), 191-201.

Wood NL, and Cowan N. (1995). The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953). Journal of Experimental Psychology: Learning, Memory, & Cognition,21, 255-260.

World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bulletin of the World Health Organization, 79(4), 373-374.

Xue M, Atallah BV, and Scanziani M. (2014). Equalizing excitation–inhibition ratios across visual cortical neurons. Nature, 511(7511), p.596.

Yang F, Bao ZY, and Zhu ZJ. (2011). An assessment of psychological noise reduction by landscape plants. International journal of environmental research and public health, 8(4), 1032-1048.

Yang W, and Kang J. (2005). Soundscape and sound preferences in urban squares: a case study in Sheffield. Journal of urban design, 10(1), 61-80.

You J, Lee PJ, and Jeon JY. (2010). Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise. Noise Control Engineering Journal, 58(5), 477-483.

Yu L, and Kang J. (2008). Effects of social, demographical and behavioral factors on the sound level evaluation in urban open spaces. The Journal of the Acoustical Society of America, 123(2), 772-783.

Zannin PHT, Calixto A, Diniz FB, and Ferreira JAC. (2003). A survey of urban noise annoyance in a large Brazilian city: the importance of a subjective analysis in conjunction with an objective analysis. Environmental Impact Assessment Review, 23(2), 245-255.

Zhang B, Shi L, and Di G. (2003). The influence of the visibility of the source on the subjective annoyance due to its noise. Applied Acoustics, 64(12), 1205-1215.

Zhang M, and Kang J. (2007). Towards the evaluation, description, and creation of soundscapes in urban open spaces. Environment and Planning B: Planning and Design, 34(1), 68-86.

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), 1039-1051.