Towards Soundscape Fingerprinting: Development, Analysis and Assessment of Underlying Acoustic Dimensions to Describe Acoustic Environments

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor-Ingenieur
(abgekürzt: Dr.-Ing.)

genehmigte

Dissertation

von

M.Sc. Lutz Jakob Bergner

geboren am 01.03.1987 in Bremen

2023

1. Referent: Prof. Dr. Jürgen Peissig
2. Referent: Prof. Dr. phil. André Fiebig
Tag der Promotion: 27. April 2023

Für meine Eltern.

ACKNOWLEDGMENT

I sincerely thank Prof. Dr. Jürgen Peissig for his supervision and mentoring during my doctorate endeavor and for his confidence in my work.

I am very grateful for the fruitful discussion with Prof. Dr. André Fiebig who reassured me that it is worthwhile to follow the path chosen.

This dissertation would not exist like this without the persistent support of Johanna and Alma. They provided the necessary patience and impatience, each with the right amount at the right time. I humbly thank you.

Further thanks go to people who have accompanied and enriched my scientific path so far as teachers, colleagues, mentors, and role models: Karlheinz Brandenburg, Sandra Brix, Tobias Gehlhaar, Jasmin Hörmeyer, Robert Hupke, Roman Kiyan, Susanne Könecke, Song Li, Yuqing Li, Wolfgang Mauersberger, Marcel Nophut, Nils Poschadel, Stephan Preihs, Roman Schlieper, Daphne Schössow, Frank Schultz, Mario Seideneck, Christoph Sladeczek, Stefan Weinzierl, Stephan Werner.

Thank you, Rebecca Knoop and Frank Trojanowski, for your vibrant proofreading.

Finally, I would like to thank those who contributed to this work in various ways by participating in the listening experiment, by providing valuable advice, encouragement, and support or by giving me the general opportunity to pursue a career in science: Charlotte, Christine, Jörg, Julia, Karin, Katharina, Kathrin, Maren, Michael, Niklas, Peter, Rolf, Sebastian D., Sebastian R., Tjorven, Tom, Ulrike, Wanja, Wolfgang, the KOOP group, the peer mentoring group.

ABSTRACT

Soundscape according to the definition in ISO 12913-1 describes an acoustic environment as perceived by humans in context. In order to be able to assess a soundscape holistically, the components *acoustic environment*, *person* and *context* should be described sufficiently to enable triangulation.

Person-based soundscape assessment has been the subject of extensive research over the past decades to date, leading to a good understanding of the main emotional dimensions. On the acoustic side, e.g., in modeling emotional responses by acoustic features, parameters describing loudness are widely used, also from the point of view of legal regulations. These parameters are often complemented by established psychoacoustic measures. However, it is unknown to what extent these parameters are suitable to adequately describe and compare acoustic environments for hypotheses concerning humans.

The presented dissertation aims to contribute to this field by means of an exploratory, empirical, and data-based approach. First, the general requirements of the aim – the description of acoustic environments – are defined and accompanied with concepts and findings from current research areas. Subsequently a methodology is developed that allows for the identification of underlying acoustic dimensions on the basis of empirical observational data of real world acoustic environments by means of multivariate statistical methods. It contains considerations on the physical sound field, the human auditory system, as well as appropriate signal processing techniques. The methodology is then applied to an exemplary extensive dataset of various Ambisonics soundscape recordings. The resulting expressions of the acoustic dimensions are evaluated and discussed with respect to plausibility and perceptual consistency. Finally, two application examples are presented to further validate the methodology and to test the applicability of acoustic dimensions in concrete research scenarios.

It was found that the presented methodology is suitable to identify dimensions for the description of acoustic environments. Furthermore, the dimensions found form a suitable basis for further soundscape analyses.

Keywords: soundscape, acoustic dimensions, auditory perception, acoustic signal processing, multivariate statistical methods

ZUSAMMENFASSUNG

Soundscape (nach ISO 12913-1) beschreibt eine akustische Umgebung, wie sie von Menschen im Kontext wahrgenommen wird. Eine ganzheitliche Beurteilung einer Soundscape wird demnach durch Triangulation der Aspekte *akustische Umgebung*, *Person* und *Kontext* hergestellt.

Die personenbezogene Bewertung von Soundscapes war und ist bis heute Gegenstand umfangreicher Forschungsarbeiten, die zu einem weitreichendem Verständnis der wichtigsten emotionalen Dimensionen geführt haben. Auf der akustischen Seite sind Parameter weit verbreitet, die die Lautstärke beschreiben. Ergänzt werden diese Parameter oft durch etablierte psychoakustische Größen. Unbekannt ist jedoch, inwieweit diese (psycho-)akustischen Parameter tatsächlich geeignet sind, Soundscapes zu beschreiben und zu vergleichen hinsichtlich den Menschen betreffender Hypothesen.

Hierzu soll diese Dissertation einen Beitrag leisten. Der dabei verfolgte Ansatz ist explorativ, empirisch und datenbasiert. Zunächst werden Anforderungen an das Ziel – die Beschreibung akustischer Umgebungen – definiert und mit Konzepten aus aktuellen Forschungsgebieten ergänzt. Anschließend wird eine Methodik entwickelt, die es erlaubt, fundamentale akustische Dimensionen zu identifizieren auf der Basis empirischer Beobachtungsdaten realer akustischer Umgebungen und mit Hilfe multivariater statistischer Methoden. Sie enthält Überlegungen zum physikalischen Schallfeld, zur menschlichen Hörwahrnehmung sowie zu geeigneten Signalverarbeitungstechniken. Die Methodik wird anschließend auf einen beispielhaften Datensatz von Ambisonics Soundscape-Aufnahmen angewandt. Die resultierenden akustischen Dimensionen werden hinsichtlich ihrer Plausibilität und wahrnehmungsbezogenen Konsistenz diskutiert. Schließlich werden zwei Anwendungsbeispiele vorgestellt, um die Methodik weiter zu validieren und um die Anwendbarkeit der akustischen Dimensionen in konkreten Forschungsszenarien zu testen.

Hierbei kann festgestellt werden, dass die gefundenen Dimensionen einen hohen Grad an Varianz akustischer Umgebungen erklären können und gut interpretierbar sind. Sie bilden somit eine geeignete Grundlage für die hier dargestellte Analyse von Soundscapes. Die Methodik ist dabei variabel erweiterbar, sodass vielfältige Anwendungen und Forschungsarbeiten bzgl. akustischer Umgebungen ermöglicht werden.

Schlagwörter: Soundscape, akustische Dimensionen, Hörwahrnehmung, akustische Signalverarbeitung, multivariate statistische Methoden

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The present work deals with the question of how acoustic environments can be described comprehensively in order to make them comparable and distinguishable. For that, a methodology is developed that allows for the identification of underlying acoustic dimensions of general acoustic environments. The aim of the method is to provide dimensions that are manageable in number and plausible to interpret, and that explain a high degree of variance with respect to the observation of general acoustic environments. If these requirements are met, the dimensions are capable to describe acoustic environments for various applications.

1.1 MOTIVATION

The motivation to engage in the research to describe properties of acoustic environments in the scope of this thesis came up in the course of an interdisciplinary research project "WEA-Akzeptanz" [1]. The project aimed for the investigation and modeling of wind turbine noise (WTN) throughout the entire signal path from the source of origin within the turbine, blades and tower to a common wind turbine sound source model to sound propagation over large distances from the emission to the immission location up to the final perception of human individuals in their everyday surrounding. In examining the far end of this chain, namely the transition from immission to perception, three fundamental questions emerged that contributed to motivate this work:

- What acoustic properties are necessary and appropriate as input parameters to model human perception and response to WTN scenarios?
- What acoustic properties are critical to reproduce when measuring perceptual responses in laboratory experiments in which subjects are exposed to WTN scenarios?
- How can the respective acoustic environments be classified and categorized to make generally valid statements, knowing that WTN scenarios are variable, depending on wind farm size and layout, immission site distance, wind speed and direction, weather, and atmospheric conditions, among other factors?

A potential answer to all of these question would be to have a set of unambiguous parameters that are capable to describe an acoustic

environment in order to make it accessible. Such a set would be able to transfer invisible sound and its characteristics as perceived by humans into a quantifiable and representable domain.

Legal regulations on noise immission, e.g. due to wind farms, usually set thresholds to variants of the sound pressure level (SPL), which are supplemented by consideration of the spectral and temporal composition and the presence of distinct tones and information content, such as in the German regulation on noise immission, the "TA-Lärm" [2]. Of course these regulations are simplified in order to make them applicable and legally compliant but at the same time they provide measurable acoustic properties with respect to human auditory perception. A review of scientific sources revealed another picture, in which a wide variety of acoustic parameters are used to investigate specific research hypotheses on human perception. However, a common agreement on a set of interpretable and unambiguous acoustic parameters was missing there as well.

These reasons motivated the author of this work to dig deeper into the topic, to collect methods and findings and finally to develop a methodology to identify underlying acoustic dimensions for a comprehensive description of acoustic environments.

1.2 AREAS OF APPLICATION

The motivation above already indicates potential fields of application of an agreed set of acoustic dimensions. Four examples are elaborated in the following. First, a set of fundamental acoustic dimensions can be used to compare acoustic environments in general in an exploratory way. This can be employed for example in academia or public administration to give aggregated information on the acoustic properties of environments. In the above example the proposed methodology could be used to compare acoustic properties of WTN scenarios with urban scenarios in order to adopt legal regulations for specific use cases. Second, when causal relationships are studied in relation to acoustic environments, a set of comprehensive descriptors is very useful. This can be the case either when an acoustic environment is the result of individual sound-emitting processes, or – as in the example above – when consequences follow from the acoustic environment, such as human responses. This research direction is also tied to the discipline of auditory scene analysis (ASA), initially developed by Albert S. Bregman [3], which investigates the human perception of complex acoustic scenes by means of principles of the Gestalt psychology, namely segmentation, integration and segregation of acoustic events. Third, computer algorithms that are designed to predict specific outcomes using data mining, machine learning, or artificial intelligence methods can benefit if their input variables are already self-contained, methodologically justified, and characterizing. Examples for that are

disciplines such as the detection and classification of sound events and scenes (AED, AEC, ASC) or the computational analysis of auditory scene (CASA). Fourth, research on soundscapes, that is "the acoustic environment as perceived by humans, in context" [4], can benefit from the presented methodology in such that an objective and general description of the acoustic environment is available for more complex human-centered investigations as indicated in the first two examples above.

1.3 OUTLINE

The proposed methodology is presented within in the following five chapters, including the discussion of fundamental assumptions and requirements, the development of methods, and their subsequent validation. In Chapter 2 the framework of soundscape is introduced and necessary conceptual and physical aspects relevant for the proposed methodology are derived from that. In Chapter 3 the methods for the identification of underlying acoustic dimensions are developed and applied resulting in a set of acoustic descriptors that are used for describing acoustic environments. Subsequently these dimensions are applied to concrete application examples aiming for steps of validation. This consists of a perceptual evaluation (Chapter 4), an investigation on ecological validity of reproduced acoustic environments (Chapter 5) and an exemplary investigation of specific classes of acoustic environments to make statements if the identified dimensions can be generalized (Chapter 6). Finally, in Chapter 7 the results will be summarized and discussed.

2

SOUNDSCAPE

Soundscape as understood nowadays is a multidisciplinary construct that describes an acoustic environment that is perceived by humans under the influence of general and personal context. The term was first mentioned in the work of Southworth in 1969 [5] for describing how the sonic environment influences the overall human experience and perception of cities. The term became more widespread in and after the work of Schafer in 1977 [6]. In it, observations of real soundscapes are described, beginning with sounds of the classical elements of earth, water, air, and fire, through sounds of the animal world, to man-made sounds and complex sounds in acoustic environments of modern urban coexistence, including the disruptive changes brought about by the industrial and electrical revolutions. The observations did not only refer to the pure physical-acoustic processes but always also to the person-related view and effect of the respective acoustic environment. From this compilation Schafer developed a first conceptual framework of nomenclature, classification and assessment of soundscapes. After that, the concept of soundscape entered various disciplines, including studies on acoustical ecology (soundscape studies), urban planning and architecture, music, and noise pollution[7]. The understanding of evaluating acoustic environments holistically in order to be able to make statements about them prevailed. An attempt to provide a framework to assess all relevant aspects of soundscapes can be found in [8]. A quite comprehensive collection of aspects of soundscapes can be found in [9], wherein contributions of many renowned experts in this field are gathered. The collaborative research on soundscape eventually led to standardization, namely on the definition and basic conceptual framework and nomenclature in ISO 12913-1:2014 [4], the data collecting and reporting requirements in ISO/TS 12913-2:2019 [10] and the data analysis in ISO/TS 12913-3:2020 [11]. A fourth part on guidelines for the assessment of soundscape investigation results is currently proposed. These standards together with the underlying scientific work form the basis for the scope of this thesis. Therefore, the assumptions and requirements that are taken into account in this work are discussed in the following sections.

2.1 CONCEPTUAL FRAMEWORK

Soundscape as defined in ISO 12913-1 provides a framework to describe and assess the acoustic, human-related and contextual aspects of acoustic environments. The acoustic environment itself, that is the

physical properties of a sound field, is present in this framework as sound sources positioned in the three-dimensional space that are subject to sound propagation effects as well as room acoustical influences (reflection, absorption, scattering). The acoustic environment is then perceived by humans by means of psychophysical, specifically psychoacoustic processes and effects. The mere perception of an acoustic environment is then fed to cognitive processes of interpretation and comparisons to inner references in the auditory system. From that emotional reactions arise that lead to manifold outcomes such as active or passive consequences. The entire chain from sound emission to emotional reaction is subject to contextual influence. For example, the same acoustic environment can be interpreted differently dependent on the time of the day or the current personal mood. Or similarly, the emitted sound from the same sound source can propagate differently depending on the season. A depiction of this framework is shown in Figure 2.1 as an adaptation of [4]. The multidisciplinarity

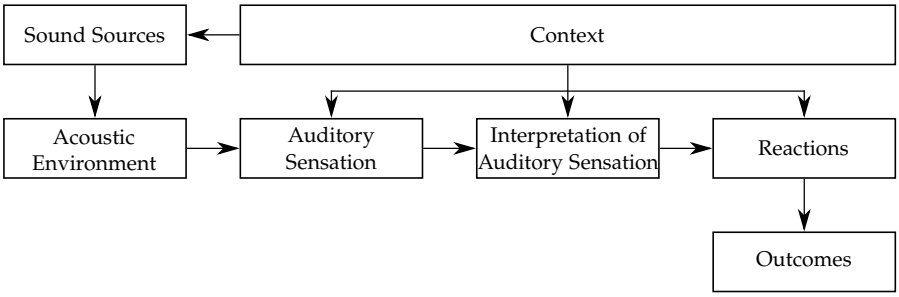


Figure 2.1: Conceptual framework of soundscapes, adapted from ISO 12913-1 [4].

of soundscape is also reflected by the suggestions for description and assessment in ISO/TS 12913-3[11]. In order to increase validity of a specific soundscape description, triangulation between the entities acoustic environment, person and context is proposed. That means, that certainty within one of these entities can be used to validate the influence of another. However, the standard does not (yet) offer concrete measures and procedures to describe all of the entities that can be applied and accepted generally but rather provides methodological guidelines on how to approach this necessity. This work aims to contribute to the valid description of one of these entities, namely the acoustic environment which is explained more in detail in the following.

2.2 ACOUSTIC ENVIRONMENT

As indicated before, an acoustic environment is composed by sound sources that are located at a specific, moving, and/or diffuse location and whose sound emission is altered by propagation effects as well as room acoustic influences until a mixture of sound waves arrives at

a specific receiver position. An unambiguous and physically correct description of this acoustic environment could be performed by capturing the three-dimensional sound field at the receiver position in a specific period of time. The sound field can be represented mathematically as $P(\theta, \varphi, r, \omega)$ where the position of the receiver location is denoted in spherical coordinates for radius r, azimuth φ, and inclination θ at a given frequency $\omega = 2\pi f$. The physical representation of the sound field can be written as a series of spatio-temporal coefficients $p_{nm}(r, \omega)$ of directional basis functions as established in the *spherical harmonic decomposition* (SHD) [12, 13]

$$P(\theta, \varphi, r, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} p_{nm}(r, \omega) Y_n^m(\theta, \varphi) \tag{2.1}$$

$$\text{with} \quad Y_n^m(\theta, \varphi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} \mathrm{P}_n^m(\cos\theta) \mathrm{e}^{\imath m \varphi}, \tag{2.2}$$

where $Y_n^m(\theta, \varphi)$ represents the spherical harmonic basis functions and P_n^m the associated Legendre polynomials. Another representation deduced from that is the superposition of incoming plane waves with coefficients $\tilde{p}_{nm}$ as it is done in *plane wave decomposition* (PWD) [12, 13]

$$P(\theta, \varphi, r, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \tilde{p}_{nm} \left(\frac{\omega}{\mathrm{c}}\right) b_n \left(\frac{\omega}{\mathrm{c}} r\right) Y_n^m(\theta, \varphi), \tag{2.3}$$

where $b_n\left(\frac{\omega}{\mathrm{c}}\right)$ denotes an appropriate radial function at the speed of sound c. Mathematical frameworks, numerical approximations as well as recording techniques of both procedures are well established in the field of sound field analysis and synthesis and/or virtual acoustics as we will see and use later in Section 3.1. Strictly speaking, however, this representation of sound pressure is only valid for the specific receiver position at a given time. Furthermore, it is difficult to derive interpretations from it that reflect a human-centered evaluation of a soundscape and is therefore only conditionally suitable for a general description.

A different approach to describe an acoustic environment is the use of semantic descriptors. ISO/TS 12913-2 provides guidelines on how the acoustic environment of a specific soundscape should be documented both semantically (e.g. sound sources, composition, foreground/background sounds, tonalities; weather/wind conditions, time of year/day; place of reception, location within the place, dwellings) and by means of accompanying acoustic measurements (esp. sound pressure level (SPL) and their statistical quartiles). This approach offers a high interpretability for the subsequent analysis of correlations between properties of the acoustic environment and human (emotional) responses, but on the other hand lacks objectivity and reliability, since the characteristic of such a semantic description is influenced by the respective investigator.

The approach that is followed in this work aims to merge the benefits of both approaches. The *signal-based* SHD or PWD representation of the sound field is further analyzed to derive abstract information that can be *semantically* interpreted. For this purpose, it is investigated which acoustic properties are actually present in different acoustic environments and which (combination of) properties can be made responsible for the discrimination of acoustic environments.

2.3 ACOUSTIC AND NON-ACOUSTIC CONTEXT

Although this thesis is concerned with the identification of dimensions to describe general acoustic environments, a brief discussion of context is given here for completeness. Context in the scope of soundscape describes influences that change physical, psychophysical, and psychological aspects of a soundscape on an intra- and interindividual level, i.e., context can vary between individuals, but also, for example, between times of day.

Physical context refers to influences that alter acoustic aspects and might be also assigned to the general characteristic of the acoustic environment according to Figure 2.1. Examples are wind and weather conditions that influence sound propagation, time of the year regarding background noise and sound absorption due to vegetation and maybe even constructional aspects like opened windows and aspects concerning the composition of a soundscape, e.g. the presence/absence of masking or background sounds such as fountains in public places [14].

Personal context on the other hand describes how the personality and experience of an individual influence their auditory perception and especially interpretation and qualitative evaluation. Examples are

VISUAL CUES e.g. if a/the dominant sound source is visible,

EXPERIENCE with and attitude towards individual sound sources and respective causes as well as entire acoustic environments, e.g. large crowds or specific engine noises,

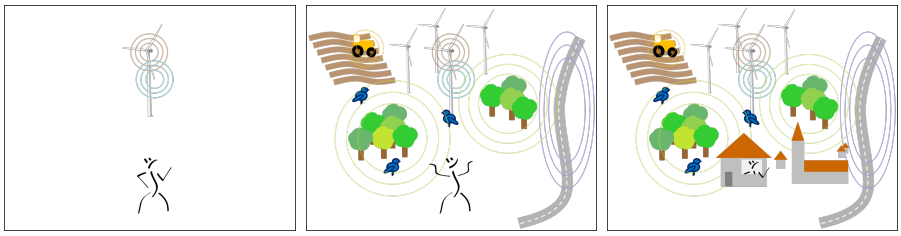
PERSONAL WELL-BEING including chronic impairments and physical complaints as well as current mood, fitness, sleepiness etc.,

ACTIVITY while exposure, e.g. sleep, rest, leisure time, working hours,

PLACE of exposure, e.g. within the own housing space, in the garden/on the balcony, at work, at public places,

EXPECTATION towards the acoustic characteristic.

All these aspects can influence the evaluation of an acoustic environment, for example if certain noise exposure is acceptable or not. A



(a) (b) (c)

Figure 2.2: Context-sensitive soundscape using the example of a wind turbine noise scenario. Perception of an isolated sound source (a), sound source embedded in usual (acoustic) environment (b), change of personal context from visitor to resident (c).

visualization of context-sensitive soundscape using a wind energy noise scenario can be taken from Figure 2.2. The sound emitted from a certain source (here: wind turbine) may isolated be perceived as annoying by an individual due to its modulated and tonal character and therefore causes disapproval (Figure 2.2a). However, if this dominant sound source is embedded into its usual (acoustic) environment with other sound sources present like that from an overland road or agricultural activities as well as masking noise due to wildlife and vegetation, the overall impression may change to a more positive or indifferent reaction. This effect can be enhanced if the scenario is set up in such a way that the subject visits the soundscape only for a foreseeable time and can leave it again at any time (Figure 2.2b). If the context is changed a second time by designing the scenario in such a way that exactly the same acoustic environment is perceived in one's own home, the reaction may again turn into a negative or rejecting direction (Figure 2.2c). It can be seen from this example that the influence of physical and personal context may lead to a very different perception and rating of valence. This finding also underlines the recognition that identification of underlying acoustic dimension, as sought in this work, is not sufficient to model emotional responses to a particular soundscape, but rather can be used to objectively assess the acoustic component of that soundscape.

2.4 CAPTURE & REPRODUCTION

An important aspect of soundscapes in the context of this thesis, but also general to soundscape research, is how to record and store a representation of one. We have learned that a soundscape is a vivid construct that changes over time and between perspectives. Thus, for reasons of documentation, suitable methods to retain the relevant aspects of a soundscape must be applied. This documenting character is even surpassed in its requirements if the recorded soundscape is to be reproduced and, as far as possible, create the same auditory

impression as the real-world reference. This applies in particular if reproducible laboratory studies on perceptual and emotional aspects of soundscapes are to be carried out. Various modern audio recording, processing and reproduction techniques that can be used for this task are discussed in this chapter.

The technical specification ISO/TS 12913-2 [10] recommends the use of binaural recordings for the use of laboratory-based listening experiments, preservation and archiving. Since the human auditory system is trained to deduce all relevant acoustic information from one's left and right ear signals, this seems to be a straightforward approach. The technology involves calibrated recording with an acoustic artificial head that resembles an average human head (and torso), outer ear and ear canal, with microphones placed at the location of the eardrum. The recorded signals can be reproduced with calibrated headphones and resemble a comparable auditory experience with high aural accuracy. The technology of binaural recording, rendering and reproduction has been developed over decades [15, 16, 17, 18] and is subject to current research of various aspects including modeling of binaural hearing [19], measurement [20], modeling [21] and individualization [22] of head-related transfer functions (HRTF), binaural reproduction with differing room acoustic properties [23] and applications in virtual and augmented reality scenarios [24]. However, binaural recordings suffer from certain drawbacks: (1) The recording is directional static, that means a head movement during reproduction with headphones leads to an implausible moving/rotating soundscape which can only be overcome with dynamic rendering of virtual acoustic environments. (2) There is a potential mismatch between the shape of head and pinna of the dummy head and the subject who listens to the reproduction which leads to incorrect reproduction of the HRTF and thus coloration and/or localization errors. (3) Reproduction of binaural recordings must – with few exceptions – be reproduced with headphones which alters at least the personal context of subjects between real-world reference and reproduction.

These drawbacks can be reduced or even overcome when microphone array recordings are conducted. These are usually based on the description of the three-dimensional sound field by means of spherical harmonic decomposition (SHD) and/or plane wave decomposition (PWD) as previously elaborated in Eqs. (2.1) and (2.3). A technical approach to transfer the mathematical fundamentals into a manageable signal representation is called Ambisonics. This technology has also been developed over many decades [25, 26] up to now in order to optimize technological properties towards perceptual validity, including microphone array engineering [27, 28], Ambisonics encoding schemes [29], signal processing of spatial optimization and manipulation [30], and rendering or decoding schemes that can be applied for both, headphones reproduction [31] and multichannel loudspeaker systems [32].

Due to its versatility and commercial availability, Ambisonics provides good possibilities for soundscape recordings for both documentation and reproduction purposes. However, as stated in ISO/TS 12913-2, there does not yet exist standardized procedures for calibrating the recording and reproduction stages for a fully auditory correct signal path, which is why this point is left up to the individual investigator.

Both technologies, binaural and Ambisonics recording and rendering are used in manifold soundscape studies. A review on recording techniques, including stereo, binaural, first- and higher-order Ambisonics (FOA, resp. HOA) can be found in [33]. It also provides an overview on reproduction techniques with loudspeakers and headphones as well as methodological thoughts. The project "Urban Soundscapes of the World" [34] uses a combined method of binaural and FOA recording for investigations on the general technological methodology [35]. Binaural recordings were exemplarily utilized for psychoacoustic analysis in [36] or for the use in *augmented reality* application within the project "I Hear NY4D" [37]. A hybrid capture of Ambisonics, binaural and additional mono signals is proposed in the project "Soundscape Indices (SSID)" [38] which also investigated the difference between binaural and mono reproduction [39]. Ambisonics recording and reproduction was used for example in [40] in order to retreive emotional dimensions that were developed on basis of field studies, so called sound walks. There exist a mentionable body of freely available Ambisonics and binaural recording databases for arbitrary use for other research groups, such as the "Eigenscape" database [41], the "Ambisonics Recordings of Typical Environments" (ARTE) database [42], the bespoken "Urban Soundscapes of the World" database [34] or the binaural "TAU Urban Acoustic Scenes 2020" [43]. A more detailed overview of available datasets is conducted in Section 3.2. An important aspect of reproducing soundscape recordings for the use of listening experiments is the question of ecological validity, i.e. if the reproduction yields acoustic and potentially non-acoustic properties comparable to the real-world reference. Investigations on this were carried out in [44, 45, 46] as well as by the author of this thesis in [47]. Another investigation on the comparability of emotional responses between stereo reproduction in an online listening experiment and Ambisonics reproduction in an respective laboratory experiment was carried out in [48]. General considerations about audio reproduction in uncontrolled internet surveys are given in [49], to which the author contributed. From the above references it can be summarized that the use of both Ambisonics and binaural reproduction are viable methods for investigating human emotional responses on soundscapes if the technological signal chain – from recording to processing to reproduction – is treated with the necessary care including calibration and equalization of headphones and loudspeaker systems.

2.5 PERCEPTION & ASSESSMENT

The assessment of perception and emotional response is the key factor in soundscape research. The framework, concepts, and technological background described above and in the respective ISO/TS 12913-1/2/3 have their main purpose in gaining a better understanding of how acoustic environments affect mood and well-being of individuals. This is also reflected in the fact that there is an extensive body of research literature on this subject which can not be covered in its entirety at this point. However, some major developments, methodologies and findings are to be discussed here.

An important field of research was (and is) the identification of perceptual dimensions when human beings evaluate soundscapes, i.e. what categories or descriptors would they use to describe an acoustic environment. Research on that includes the method to provide a multitude of semantic descriptors to subjects while exposed to soundscapes whose rated results are subsequently aggregated to latent perceptual dimensions as proposed in 1981 by Russell, Ward, and Pratt for general environments in [50] and performed in the scope of soundscapes in [51, 52, 53, 54, 55]. The findings from these and other contributions led to the definition of a two-dimensional model of affective quality in soundscape evaluation in ISO/TS 12913-3 [11], namely pleasantness and eventfulness. These two dimensions on perpendicular axes span a space where human evaluation of soundscapes can be located as seen in Figure 2.3. The standard also provides a methodology and

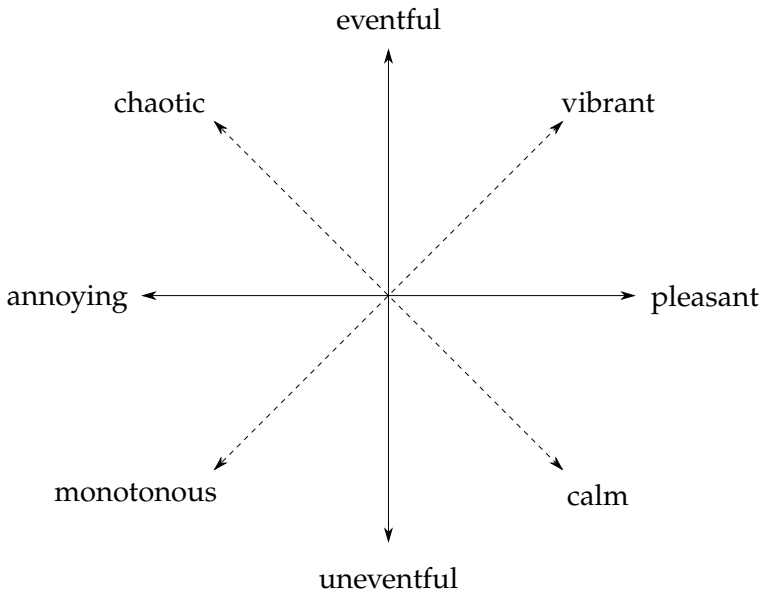


Figure 2.3: Affective qualities of soundscapes according to ISO 12913-3 [11]

questionnaire template to assess these dimensions which will be used later in Section 4.4. A further discussion of how emotions are not only

a result of soundscapes, but are themselves fed back into the chain of perception, interpretation, and response is presented in [56].

The semantic descriptors and dimensions of responses is one part of understanding the effects of soundscapes. The other part is the respective acoustic and/or non-acoustic cause that leads to such judgment. There have been various attempts to model affective qualities with acoustic indicators, without any particular approach being able to remove all ambiguities so far. It seems to be confirmed that – beside the situational context of exposure – loudness plays an important role which is also reflected in many legal regulations. Further psychoacoustic measures can be taken to refine prediction models as suggested e.g. in [36]. A thorough overview of modeling attempts and methods can be found in [57], whereas a review on potential acoustic and psychoacoustic indicators can be found in [58].

Furthermore, in [59, 60] the author has proposed a methodology for the capturing and reproduction of wind turbine noise scenarios in laboratory experiments and contributed to perceptual assessment of these scenarios in the same project scope [61, 62, 63, 64].

3

METHODOLOGY OF SOUNDSCAPE FINGERPRINTING

In this chapter the fundamental methodology of the presented work is rolled out, developed and applied. It consists of the discussion of appropriate acoustic indicators, databases of soundscape recordings, the identification of the underlying acoustic dimensions by means of multivariate statistical methods and consideration how to assess them.

Conceptual parts of the following chapter have already been published in parts by the author in [65, 66].

3.1 SIGNAL-BASED INDICATORS

In the scope of this work, the description of an acoustic environment as part of a soundscape aims for distinction. That means that relevant information are to be provided in such a way that shared and diverging properties of the respective acoustic environments become not only visible but also interpretable from a human-centered perspective. For its development, a wide variety of signal-based acoustic metrics, hereafter referred to as *indicators*, form the basis for this description. To motivate the classification and the selection of appropriate indicators, a naïve semantic description of an acoustic environment will serve as an example:

Example. *I'm sitting in an office at home. I hear the ticking of a clock in the room. The baby in the apartment above me is crying. Outside my window, cars pass by frequently. From far away, I sometimes hear the siren of an emergency vehicle.*

This description is probably quite meaningful and it is not difficult to put yourself in this acoustic situation. However, this description contains several assumptions and background information that are personally learned from previous similar experience but that are critical for the correct classification and interpretation of the descriptive example. Thus, an analogue technical description of the above would not be feasible without further information processing. A technically understandable description of the same environment, on the other hand, could be as follows:

Example. *The position of the receiver is located in a fairly quiet indoor room of small dimensions and low reverberation time. A localizable, steady, quiet, and transient noise can be noticed in a very regular temporal succession. Another, elevated and band-limited sound with tonal and noise-like components modulates in level and spectrum in an irregular temporal slope.*

Loud, low-frequency harmonic sounds with spatial movement and coherent increase and decrease of level occur irregularly on average two times per minute. Spatially diffuse, alternating tonal sounds in the musical interval of a fourth with apparent Doppler effect can be noticed on average once every 20 minutes.

This second description contains signal parameters which may be generated, calculated and analyzed by means of acoustical signal processing but whose assessment and interpretation is not easily human-readable anymore without having the original description at hand. However, it paves the way for quantifiable and objective description and comparison. A detailed analysis of the used technical terms reveal that information is provided that refer to the spectral properties of sound sources (noise-like, band-limited, low-frequency harmonic, tonal), as well as the loudness or level (quiet, increase of level, loud), spatial orientation (localizable, elevated, spatially diffuse), and temporal succession (very regular, irregular, two times per minute, once every 20 minutes) and finally information on the general room acoustical and environmental aspects (quiet, indoor, low reverberation). A description like this can be used as the basis for general a-priori categories for acoustic environments, namely quality (in the following referred to as **Q**), loudness (**L**), spaciousness (**S**) and time (**T**). The category *quality* must not be confused with *valence* but represents characteristics, that helps human beings to identify sound sources, such as information on timbre and spectral composition as well as short time temporal succession. *Loudness* includes information on the total number, distance and level of simultaneous sound sources as well as on the acoustic ambience while *spaciousness* represents general location and envelopment of the sound sources. The category *time* includes information on how often and how long a sound event occurs and if individual properties are changing over time or stay constant. In order to find appropriate signal-based acoustic indicators that are capable to represent these categories, relevant research areas to provide potential metrics were identified such as soundscape studies [10, 58], music information retrieval (MIR)[67], acoustic event detection (AED), acoustic event classification (AEC), acoustic scene classification (ASC)[68], acoustic scene analysis (ASA), or bioacoustics [69]. All of these disciplines provide metrics that were found to be suitable for the respective field of research hypotheses. Research attempts to model sound and noise quality and perception for example often rely on A- and C-weighted sound pressure levels or conventional psychoacoustic measures such as loudness, sharpness, roughness and fluctuation. The computational detection and classification of acoustic events and scenes often include parameters like short-time spectrograms or Mel-frequency cepstral coefficients (MFCCs). In bioacoustics e.g. specific characteristics of the time-frequency spectrogram are used to detect sounds of wild life (e.g. bird singing) or human-generated sounds

which is used to formulate metrics on the bioacoustic complexity or richness.

As mentioned above, each discipline has developed its specific set and use of acoustic indicators which often have a common foundation but differing parametrization and interpretation. In the scope of this thesis these works were considered and indicators were selected that are widely established. The resulting selection for quality indicators can be found in Tables 3.1, loudness indicators in Table3.2, and spaciousness indicators in Table 3.3 respectively each listing the indicator's expected value range, signal basis, scientific or standardized reference and computational implementation.

Table 3.1: Acoustic indicators for a-priori category quality with respective information on value range, signal basis, reference and implementation.

Indicator	**Range**	**Unit**	**Signal**	**Ref.**	**Impl.**
Roughness	[0,1] (lin)	Asper	mean binaural (FB)	[70, 71]	[72]
Sharpness	[0,20] (lin)	Acum	mean binaural	[70]	[72]
Fluctuation Strength	[0,1] (lin)	Vacil	mean binaural (FB)	[70]	[72]
Periodic Modulation Frequency	[0.1,100] (lin)	Hz	mean binaural (FB)	Own	cf. C
Periodic Modulation Depth	[-10,10] (lin)	None	mean binaural (FB)	Own	cf. C
Stochastic Modulation Depth	[-10,10] (lin)	None	mean binaural (FB)	Own	cf. C
Spectral Centroid	[80,12000] (log)	Hz	mean binaural	[73]	[72]
Spectral Spread	[0,1000] (lin)	Hz	mean binaural	[73]	[72]
Spectral Skewness	[0,1] (lin)	None	mean binaural	[73]	[72]
Spectral Kurtosis	[0,10000] (lin)	None	mean binaural	[73]	[72]
Spectral Entropy	[0,1] (lin)	None	mean binaural	[74]	[72]
Spectral Flatness	[0,1 (lin)	None	mean binaural	[75]	[72]
Spectral Crest Factor	[0.1,500] (log)	None	mean binaural	[73]	[72]
Spectral Flux	[$1e^{-12}$,$1e^{-1}$] (log)	None	mean binaural	[76]	[72]

Continued on next page

Table 3.1 – continued from previous page

Indicator	Range	Unit	Signal	Ref.	Impl.
Spectral Slope	[-1e^{-6},1e^{6}] (lin)	None	mean binaural	[67]	[72]
Spectral Derease	[-10,1] (lin)	None	mean binaural	[73]	[72]
Spectral Rolloff Point	[0.1,8000] (log)	Hz	mean binaural	[76]	[72]
Octave Band Energy	[-120,0] (lin)	dB	mean binaural	[77]	[72]
Mel Frequency Cepstral Coefficient	[-100,100] (lin)	None	mean binaural	[78]	[72]
Timbral Booming	[-20,20] (lin)	None	mean binaural	[79]	[80]

Table 3.2: Acoustic indicators for a-priori category loudness with respective information on value range, signal basis, reference and implementation.

Indicator	Range	Unit	Signal	Ref.	Impl.
Loudness (Zwicker)	[0.01,100] (log)	Sone	pressure (FB)	[71]	[72]
Loudness (Moore-Glasberg)	[0.01,100] (log)	Sone	pressure (FB)	[81]	[72]
Loudness units relative to full scale (momentary)	[-80,0] (lin)	LUFS	pressure (FB)	[82, 83]	[72]
Loudness units relative to full scale (short-term)	[-80,0] (lin)	LUFS	pressure (FB)	[82, 83]	[72]
Loudness units relative to full scale (integrated)	[-80,0] (lin)	LUFS	pressure (FB)	[82, 83]	[72]
Loudness units relative to full scale (integrated)	[-10,10] (lin)	LU	pressure (FB)	[82, 83, 84]	[72]
True Peak	[-10,10] (lin)	LU	pressure (FB)	[83]	[72]
L_A	[0,120] (lin)	dB	pressure (FB)	[85]	[72]
$L_{A,eq}$	[0,120] (lin)	dB	pressure (FB)	[85]	[72]
$L_{A,peak}$	[0,120] (lin)	dB	pressure (FB)	[85]	[72]
$L_{A,max}$	[0,120] (lin)	dB	pressure (FB)	[85]	[72]

Table 3.3: Acoustic indicators for a-priori category spaciousness with respective information on value range, signal basis, reference and implementation.

Indicator	Range	Unit	Signal	Ref.	Impl.
Horizontal direction of arrival	[-180,180] (lin)	degree	ambisonic (FB)	[86]	[87]
Longitudinal direction of arrival	[-90,90] (lin)	degree	ambisonic (FB)	[86]	[87]
Diffuseness	[0,1] (lin)		ambisonic (FB)	[86]	[87]
Interaural level difference	[-30,30] (lin)	dB	binaural (FB)	[16]	[19]
Interaural time difference	[-1,1] (lin)	ms	binaural (FB)	[16]	[19]
Interaural cross correlation	[0,1] (lin)		binaural (FB)	[16]	[19]
Spherical Directivity Index	[0,20] (lin)	dB	pressure (FB)	[88], Own	cf. C
Vertical Directivity Index	[0,20] (lin)	dB	pressure (FB)	[88], Own	cf. C
Horizontal Directivity Index	[0,20] (lin)	dB	pressure (FB)	[88], Own	cf. C
Spherical Pressure Ratio	[-40,0] (lin)	dB	ambisonic (FB)	Own	cf. C

For this work, each indicator is calculated as time series for overlapping frames of 100 ms each and hop size of 50 ms to respect both time-integrating behavior of the human auditory model [16] and time-variance of acoustic scenes. It is recognized here that there may be acoustic events and psychoacoustic effects that are difficult to detect with this temporal resolution. At the same time, averaging through large analysis windows contributes to the increased robustness of the results against statistical and measurement noise. Furthermore, the majority of indicators is calculated frequency-dependent. For that, the broadband signals are filtered using ten octave filters with center frequencies given in Table 3.4 and indicators are calculated for each octave band individually. Again, this spectral resolution is not sufficient to separate the filter bands of human hearing or the spectral composition of individual sound sources. However, it offers the possibility to detect a general and interpretable frequency dependence of the acoustic indicators.

The indicators themselves are based on one of three signal representations of the same acoustic environment: The quality and loudness indicators may be calculated either from a monophonic pressure representation or from a binaural signal, while the spaciousness indicators require a binaural and spherical harmonic (Ambisonics) signal

Table 3.4: Frequency limits in Hz of analysis bands.

ID	0	1	2	3	4	5	6	7	8	9
f_c	31	62	125	250	500	1000	2000	4000	8000	16000

representation of the three-dimensional soundfield. The latter two representations incorporate spatial information of an acoustic environment such as the location of sound sources or the envelopment of sound. The signal representations used for calculating the individual indicators can be taken from the column "Signal" in above tables where an additional "(FB)" denote calculation in frequency bands. In order to maintain consistency and to reduce data complexity, all three representations stem from the same recording of a specific acoustic environment. For that, microphone array recordings are necessary that can be transformed into the spherical harmonic domain as it is established in Ambisonics encoding and rendering [12]. The order of the ambisonic recordings generally determine the spatial confidence. However even first-order Ambisonics (FOA) recordings are suitable for the analysis in this work. The binaural representation is derived by convolution with appropriate head-related transfer functions (HRTF) [89] as it is established in [30, 90]. The monophonic sound pressure representation on the other hand is proportional to the oth-order Ambisonics component [86].

3.2 SOUNDSCAPE DATABASES

The presented approach to identify underlying acoustic dimensions of soundscapes is data-based, that means that the potential findings are not theory-driven but stem from (real world) observations in the form of soundscape microphone recordings. Of course, the results are not developed without any prior knowledge and assumptions, as the compilation of indicators has shown previously. Thus, the underlying database for the statistical analysis was also chosen carefully to fulfill certain requirements. First, the database determines the soundscape population in statistical terms and underlying acoustic dimension deduced from that are strictly only valid for samples of this population. From these assumptions comes the requirement that the database itself must cover as wide a range of acoustic environment classes as possible, including indoor and outdoor scenarios, urban and rural, noisy and quiet, etc. The presence of sound source classes according to ISO/TS 12913-2, namely sounds of technology, human and nature, should also be ensured with similar frequency. Technical requirements are Ambisonics recording of first or higher order with sufficient quality (at least 44.1 kHz @ 16 Bit). In order to calculate correct level and loudness indicators, certain information on calibration should be provided, either in the form of microphone sensitivity, gain and A/D conversion

or, preferably, as separate sound pressure level reference measurement e.g. as L_{Aeq} either time-dependent or for an entire recording. There exists a surprisingly large body of soundscape recordings, that fulfill parts or all of the mentioned requirements. Table 3.5 shows a selection of potential databases without claiming to be complete. The first three

Table 3.5: Potential soundscape databases.

Name	Soundscape classes	Domain	Quant.	Total Length
EigenScape [41]	beach, busy street, park, pedestrian zone, quiet street, shopping centre, train station, woodland	HOA	64	640 min
ARTE [91]	library, office, church, living room, café, dinner party, street, train station, food court	MOA	13	29 min
Soundfield by Røde Ambisonic Sound Library [92]	indoor crowd, playground, car, foyer, library, mall, market, metro, street, steam train, traffic, train station	FOA	35 / 237	137 min
Urban Soundscapes of the World [34]	urban	FOA	130	130 min
I Hear NY4D [93, 37]	urban	FOA	6	N/A
Mhacoustics demos [94]	music, wild life, indoor, crowd,	HOA	8	N/A
TAU-NIGENS Spatial Sound Events 2020 [95]	various single sound events	FOA (synth.)	800	800 min
TAU Urban Acoustic Scenes 2020 [43]	airport, indoor shopping mall, metro station, pedestrian street, public square, street with medium level of traffic, urban park	binaural	N/A	2400 min
TUT Database [96]	bus, cafe, car, city, forest, store, home, lakeside, library, station, office, train, tram, park	binaural	1170	585 min

databases, namely the *Eigenscape*, *ARTE*, and *Soundfield Library* are utilized for the development of the acoustic dimensions. The reasons for this is mainly a balanced selection of soundscape classes without any class being overly represented (especially urban soundscapes). The selection also ensures a wide range of Ambisonics order from first-

order (Soundfield), to 4th-order (EigenScape) to 4th-/7th-mixed-order (ARTE).

3.3 DETERMINING UNDERLYING DIMENSIONS

3.3.1 *Concept of Factor Analysis*

The idea pursued in this work is that the multitude of indicators contain information describing the properties of an acoustic environment that are relevant when a human being perceives and contextualises the same environment. Just as humans can classify their environment acoustically on the basis of their two ear signals, a procedure is now to be developed here that provides an abstract construct for the description and identification of acoustic environments on the basis of the indicators presented in the previous section. In other words, it is assumed that the observed indicators above are realizations of certain underlying acoustic dimensions that characterize an acoustic scene or environment. These assumptions allow the application of exploratory factor analysis (EFA; hereafter referred to as FA for convenience) [97] as schematically depicted in Figure 3.1. Similar to the related prin-

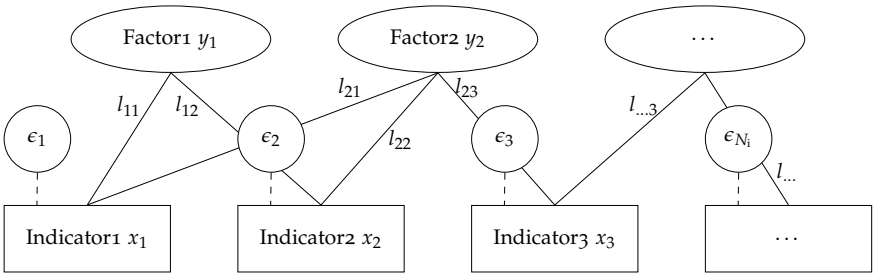


Figure 3.1: Concept of Factor Analysis with loadings l_{ij} and unique variances ϵ_i.

ciple component analysis (PCA), FA can be used here to aggregate data variances (and thus information) by transforming the observed indicator time series from an original space into an optimized space of latent dimensions. The methodological differences between PCA and FA concern the perspective: while PCA assumes that the observed indicators constitute the ground truth, which in turn can be described by principal components, FA implies that the (hidden) latent factors constitute the ground truth and the observed indicators are more or less arbitrary realizations of it. In mathematical terms FA can thus be described in a generative way as

$$\boldsymbol{X} = \boldsymbol{L}\boldsymbol{Y} + \boldsymbol{\epsilon} \quad , \tag{3.1}$$

where $\boldsymbol{X}$ is the matrix of indicator observations of dimension $[N_\mathrm{i} \times N_\mathrm{o}]$ (N_i: number of indicators; N_o: number of observations), $\boldsymbol{L}$ a

specific loading matrix of dimension $[N_\mathrm{f} \times N_\mathrm{i}]$ (N_f: number of factors) and according latent factor scores $\boldsymbol{Y}$ of dimension $[N_\mathrm{f} \times N_\mathrm{o}]$ and $\boldsymbol{\epsilon}$ a diagonal matrix of unique variances.

The conceptual difference between FA and PCA becomes relevant if the amount of explained variance is investigated. It is generally assumed that in FA the variance of each indicator is made up by two types, common and unique variance. It is further assumed that the common part of the variance is also observable in other indicators and thus originates from one factor. The unique variance $\boldsymbol{\epsilon}$ refers directly and exclusively to the indicator itself and consists of the specific variance of the indicator plus an error term. The theoretical background of PCA assumes that the unique variance equals zero and the total variance of the indicator remains in the model. Thus, PCA can be seen as special case of the general factor analysis. The factors in FA on the other hand only express the common part of variance of the observed indicators. That means in practice that each indicator may inhibit portions of specific variance ϵ_s as well as measurement and analysis noise ϵ_n which both are not included in the factors as denoted in Figure 3.1 with $\epsilon_i = \epsilon_{si} + \epsilon_{ni}$. Hence, we allow the indicators to be imperfect realizations of the factors which relaxes the necessary requirements of the indicators.

Except for the input matrix $\boldsymbol{X}$, all terms on the right side of the generative equation of FA in (3.1) are unknown, however, certain assumptions are met. The factors are uncorrelated which leads to $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{I}$ where Cov denotes the covariance matrix and $\boldsymbol{I}$ the identity matrix. Further $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are independent from each other and the mean or expectation of the factor scores is zero $\mathrm{E}(\boldsymbol{Y}) = 0$. With these assumptions it can be shown that

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}) &= \mathrm{Cov}(\boldsymbol{L}\boldsymbol{Y} + \boldsymbol{\epsilon}) && (3.2)\\ &= \boldsymbol{L}\mathrm{Cov}(\boldsymbol{Y})\boldsymbol{L}^T + \mathrm{Cov}(\boldsymbol{\epsilon}) && (3.3)\\ &= \boldsymbol{L}\boldsymbol{L}^T + \mathrm{Cov}(\boldsymbol{\epsilon}) \quad . && (3.4) \end{aligned}$$

With $\boldsymbol{\epsilon} = \boldsymbol{O}$ in PCA, this equation can be solved by means of eigenvector decomposition of the covariance matrix of the observed indicators. In general FA, it is solved by means of an iterative estimation that minimizes the unique variance $\boldsymbol{\epsilon}$ using maximum likelihood as formulated in [98, Ch. 21.2] and implemented in [99].

The loading matrix comprises the individual weights of each indicator into each factor

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1N_\mathrm{i}} \\ l_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & l_{ji} & \vdots \\ l_{N_\mathrm{f}1} & l_{N_\mathrm{f}2} & \cdots & l_{N_\mathrm{f}N_\mathrm{i}} \end{bmatrix} \quad . \tag{3.5}$$

The vertical squared sum over rows, i.e. among factors, yields the communalities h_i^2 of the indicators. This metric represents the amount

of the initial indicator's variance that is explained by the identified factors

$$h_i^2 = \sum_{j=1}^{N_\mathrm{f}} l_{ji}^2 \quad . \tag{3.6}$$

In PCA $h_i^2 = 1$ for all indicators if summed over all principle components. The sum over columns, i.e. among indicators, yields the sum of square loadings (in PCA: eigenvalues of covariance matrix) or explained variance of a certain factor

$$s_j^2 = \sum_{i=1}^{N_\mathrm{i}} l_{ji}^2 \quad . \tag{3.7}$$

This measure indicates the weight of a particular j-th factor, which is important when deciding which factors to retain. Dividing $\boldsymbol{L}$ by the respective explained variances yields the relative Loading $\boldsymbol{L_{rel}}$ (in PCA: eigenvectors of covariance) that includes the assignment of the indicators to the respective factors and represents the direction of the transformation

$$\boldsymbol{L_{rel}} = \boldsymbol{L} \cdot \mathrm{diag}\{\boldsymbol{s}\}^{-1} \quad . \tag{3.8}$$

In order to apply FA to indicators of different scales and units, preprocessing of the initial indicator vectors must be applied. For that, an interval range of expected values was defined for each indicator and scaling was applied accordingly to derive relative values within this interval. Since FA is only capable of identifying linear relationships, non-linear indicators must also be treated accordingly. Ratio-scaled indicators with reference to frequency/Hz are converted to frequency in octaves relative to 10Hz to regard the logarithmic behavior of auditory pitch perception. Conversions and expectation intervals for each indicator can be taken from Tables 3.1, 3.2, and 3.3. Finally, a z-standardization was applied to each indicator, that means removal of the mean and normalization to unit variance.

Pure FA produces mutually independent (uncorrelated) factors where the first factor includes maximum variance. However, this might result in a loading matrix that is difficult to interpret. In these cases, a further rotation of the loading matrix $\boldsymbol{L}$ aims for a simple structure with few high loadings and many low loadings. In this work the orthogonal rotation method *varimax* was chosen to preserve uncorrelated factors while increasing interpretability.

3.3.2 *Identification of Underlying Acoustic Dimensions*

The previously explained multivariate method of factor analysis is subsequently applied to the soundscape recording databases de-

scribed in Section 3.2. The raw indicator time series were scaled to its predetermined value range and further normalized by means of z-standardization. In order to reduce transient effects within the calculation of the indicators, the first and last 5 seconds of each time series were excluded from further processing. Finally, the input matrix $\boldsymbol{X}$ consists of $N_\mathrm{o} = 903,753$ observations of $N_\mathrm{i} = 304$ indicators. Even though the choice of *varimax*-rotated factor analysis was made thoroughly, initially all four potential methods, namely PCA and FA each with and without rotation, were calculated to keep the exploratory characteristic of the overall investigation. Figure 3.2 shows the progress of the cumulative explained variance for these four cases. It shows the maximum amount of the explained common variance in FA of 72.62 % which means that 27.38 % of all indicator's variance is unique (specific variance + error). The influence of the *varimax* rotation

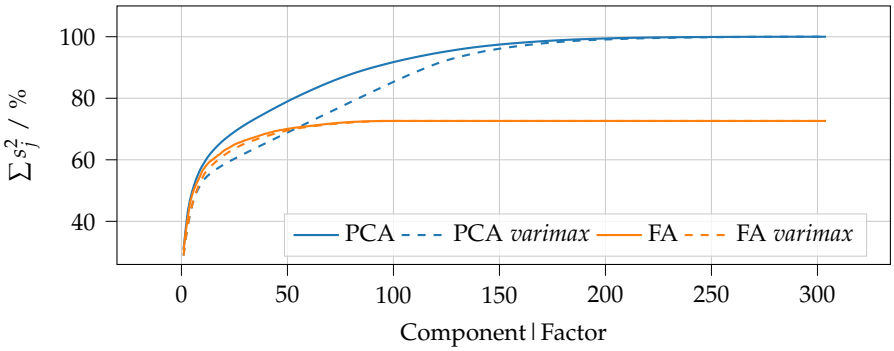


Figure 3.2: Comparison of cumulative explained variance of the pure and *varimax*-rotated PCA and FA.

(dashed lines) becomes visible and especially in PCA the difference to the unrotated method is distinct. An important analysis step is the choice of how many factors or principle components to retain as being relevant. The goal is to maintain a manageable number of factors that allow for meaningful interpretation and can be used to describe a data set at the same time. One approach for this task is to keep all factors with variance larger than the variance of a single indicator. Since we normalized all indicators to unit variance, this means all factors with $s_j \geq 1$ should be kept. This method is called the Kaiser rule [100] and has some drawback since the number of retained factors is high and the distribution of explained variance is neglected. Another method is to visually analyze the scree plot, which depicts the explained variance as line plot. The factor at which a "knee" or an "elbow", i.e. a significant change in slope, is observed is selected as the limit value. Exemplary scree plots are shown at the top of Figures 3.3 and 3.4. The knee is visible at the second factor (factor 4) which would result in a single factor to retain and may be insufficient to characterize the dataset at all. A third method is

the parallel analysis [101] which takes the overall slope of explained variances into account and is based on a Monte-Carlo simulations of random data of the same size. It is assumed to be a more robust method for deciding how many factors to retain and is utilized in this work. Table 3.6 shows the result for the Kaiser criterion next to parallel analysis for all four applied methods. We find that the eight

Table 3.6: Number $N_{\mathrm{f,r}}$ of relevant factors (FA) or principle components (PCA) according to Kaiser criterion and parallel analysis.

	Kaiser criterion		Parallel analysis		
	$N_{\mathrm{f,r}}$	$\sum_{j=1}^{N_{\mathrm{f,r}}} s_j^2$	$N_{\mathrm{f,r}}$	$N_{\mathrm{i,r}}$	$\sum_{j=1}^{N_{\mathrm{f,r}}} s_j^2$
PCA	55	245.20 (80.66 %)	7	108	163.62 (53.82 %)
PCA varimax	85	244.61 (80.46 %)	7	104	149.80 (49.28 %)
FA	21	192.77 (63.41 %)	7	189	158.52 (52.14 %)
FA varimax	28	196.62 (64.68 %)	8	114	156.22 (51.39 %)

most relevant factors in *varimax*-rotated FA explain 51.39 % of the total variance. This is only a moderate result, however given the large number of samples it is a good tradeoff between number of factors and variance explained. The following analysis will then be based on the eight most relevant factors of the *varimax*-rotated factor analysis. This information will be omitted in the following. A direct effect of the *varimax* rotation can be observed if the number of indicators $N_{\mathrm{i,r}}$ is observed, that is necessary to describe >50 % of all relevant factors. In the case of FA this number decreases from 189 to 114, while the explained variance remains almost equal (52.14 % to 51.39 %). From that follows that a smaller number of indicators is capable to explain the same amount of variance and thus makes it easier to interpret a factor. The interpretation of a factor depends largely on the composition of indicators that are loaded from it. A first analysis of the composition can be found in Figure 3.3. At the top plot the explained variance of the relevant factors is depicted. The bottom plot provides an overview of the loading matrix between all 304 indicators and the eight factors. For better visibility, the indicators are ordered in the categories loudness, quality and spaciousness. Certain patterns and regularities can be observed as well as the fact that there are factors that include most indicator loading within a certain indicator category. As a further analysis and processing step, a so called 50 % filter was applied to the factor composition. That means that only those most prominent indicators are kept whose cumulative variance makes up at least half or 50 % of the respective factor's variance. The result of this filtered loading matrix is depicted in Figure 3.4 and stands in comparison to Figure 3.3. Further, a more detailed analysis of the

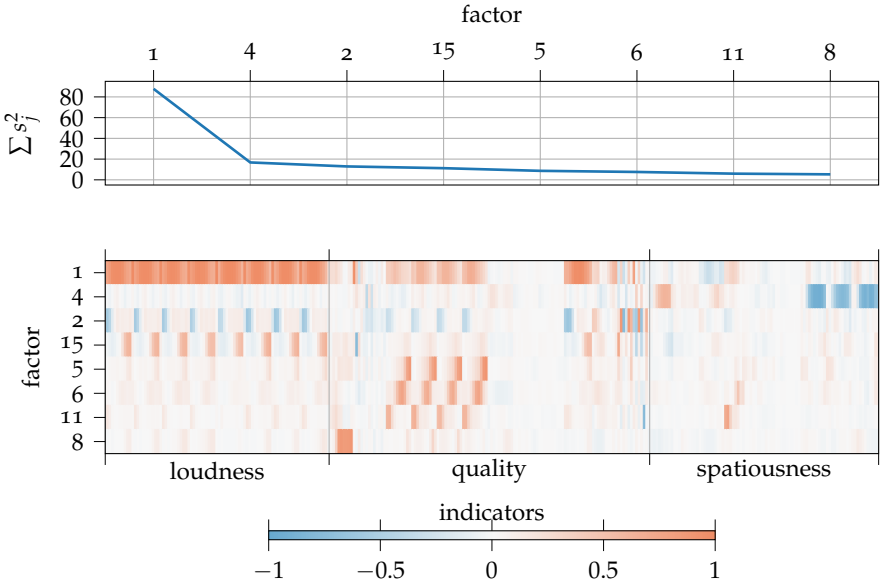


Figure 3.3: Scree test of explained variance (top) and schematic distribution of indicator loadings for the first eight relevant rotated factors (bottom).

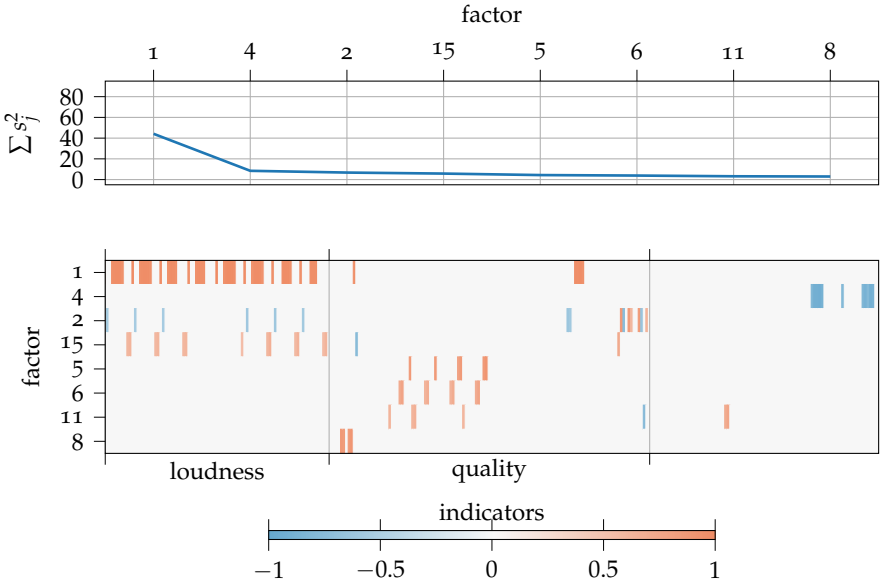


Figure 3.4: Scree test of explained variance and schematic distribution of cleaned indicator loadings for the first eight relevant rotated factors.

factor composition is provided in Table 3.7.

Table 3.7: Indicator composition the first eight relevant rotated factors j with respective explained variance s_j^2 and relative loadings $l_{\mathrm{rel},ij}$ in parentheses. Trailing numbers of the indicators denote the frequency bands. $N_{\mathrm{i},j}$ denotes how many of the total number of indicators account for $\geq$51 % of the factor's explained variance.

Factor j	s_j^2	$N_{\mathrm{i},j}$	**Indicators**
1	87.77 (28.9%)	48	loudnessZwicker(0.106), loudnessZwickerBands04(0.105), loudnessZwickerBands05(0.105), loudnessZwickerBands03(0.105), LA(0.104), LAeq(0.104), LAeqBands03(0.104), LABands03(0.104), LAeqBands05(0.104), LABands05(0.104), lufsMomBands03(0.104), LABands04(0.104), LAeqBands04(0.104), lufsPeakBands03(0.103), lufsMomBands05(0.103), lufsMomBands04(0.103), lufsPeakBands04(0.103), lufsPeakBands05(0.103), LAmaxBands03(0.103), oct06(0.103), LApeakBands03(0.103), oct05(0.103), oct04(0.103), lufsShortBands03(0.103), lufsShortBands04(0.103), lufsShortBands05(0.102), lufsMom(0.102), LAmaxBands04(0.102), loudnessZwickerBands06(0.102), LAmax(0.102), LAmaxBands05(0.102), mfcc00(0.102), LApeakBands04(0.102), lufsShort(0.102), LApeakBands02(0.101), lufsPeak(0.101), LAeqBands02(0.101), LABands02(0.101), LApeakBands05(0.101), LAmaxBands02(0.101), LApeak(0.101), loudnessZwickerBands02(0.100), oct07(0.100), LABands06(0.100), LAeqBands06(0.100), lufsPeakBands06(0.100), lufsMomBands06(0.100), lufsMomBands02(0.098)
4	16.77 (5.5%)	11	sphDI07(-0.226), sphDI06(-0.225), sphDI05(-0.224), sphDIEl07(-0.223), sphDI08(-0.218), sphDIEl05(-0.216), sphDIEl08(-0.215), sphDIEl04(-0.209), sphDI04(-0.207), sphDIEl06(-0.196), sphDIAz06(-0.195)
2	12.92 (4.3%)	15	spectralCentroid(0.242), spectralEntropy(0.238), spectralRolloffPoint(0.236), spectralCrest(-0.210), spectralSkewness(-0.183), oct01(-0.175), lufsPeakBands00(-0.173), spectralSpread(0.173), lufsMomBands00(-0.170), lufsShortBands00(-0.165), spectralFlatness(0.165), oct02(-0.164), LABands00(-0.159), LAmaxBands00(-0.157), LAeqBands00(-0.157)
15	11.23 (3.7%)	15	mfcc01(-0.231), sharp(0.213), lufsMomBands09(0.196), lufsPeakBands09(0.189), LABands09(0.186), lufsShortBands09(0.184), LAeqBands09(0.183), LAmaxBands09(0.183), lufsMomBands08(0.179), LAmaxBands08(0.176), LABands08(0.176), LAeqBands08(0.174), lufsPeakBands08(0.171), lufsShortBands08(0.167), loudnessZwickerBands09(0.163)
5	8.67 (2.9%)	6	modDepthS09(0.305), modDepthP109(0.294), modDepthS08(0.294), modDepthP209(0.290), modDepthP309(0.289), modDepthP308(0.263)

Continued on next page

Table 3.7 – continued from previous page

Factor j	s_j^2	$N_{i,j}$	**Indicators**
6	7.59 (2.5%)	8	modDepthS06(0.276), modDepthP106(0.274), modDepthP105(0.262), modDepthS05(0.255), modDepthP306(0.250), modDepthP206(0.248), modDepthP305(0.228), modDepthP205(0.226)
11	5.96 (2.0%)	7	spectralSlope(-0.318), iacc01(0.306), iacc00(0.281), modDepthP201(0.266), modDepthP200(0.256), modDepthP101(0.254), modDepthS00(0.247)
8	5.29 (1.7%)	4	fluct09(0.388), fluct05(0.374), fluct08(0.369), fluct06(0.367)

It shows the composition of the first eight relevant factors and can be read as following:

Example. *Factor 2 inhibits a variance of* $s_j^2 = 18.64$ *which makes up 6.1 % of the total variance of the dataset. The listed 31 most prominent indicators make up* ≥ 50 *% of the factor's variance, i.e.* ≥ 9.32*. The most important indicator is* `spectralSkewness` *with a loading of* $l_{12} = l_{\text{rel},ij} \cdot \sqrt{s_j^2} = -0.174 \cdot \sqrt{18.64} = -0.751$.

The investigation of the composition of the factors emphasize the relevance of the factor rotation. For completeness, the table of the unrotated FA can be taken from the Appendix Table A.1. Trailing numbers in most listed indicators refer to the frequency bands from 00 ($f_c = 31\,\text{Hz}$) to 09 ($f_c = 16\,\text{kHz}$) (see Table 3.4, p. 20). The indicators that are listed and assigned to the respective factors do have some common background. Either they stem from the same indicator group calculated in different frequency bands (trailing numbers) or are assigned to the same indicator category. Hence, the final analysis step of FA itself is the interpretation of the factors and the attempt to describe it semantically.

FACTOR 1 This factor comprises 48 indicators that mainly measure sound pressure levels (`LA`, `LABands`, `LAeq`, `LAeqBands`) or loudness metrics (`loudnessZwicker`, `loudnessZwickerBands`, `lufsBands`). Beside the broad band indicators, the most occuring frequency bands are (in decreasing order) 04, 05, 03, 02, 06, corresponding to center frequencies $f_{c,04} = 500\,\text{Hz}$, $f_{c,05} = 1\,\text{kHz}$, $f_{c,03} = 250\,\text{Hz}$, $f_{c,02} = 125\,\text{Hz}$, $f_{c,06} = 2\,\text{kHz}$ and covering the (low) mid frequency range. Since this is also the frequency range where most sound energy is located in general soundscapes this spectral focus is comprehensible. The suggestion for a semantic descriptor for this factor would be: **Loudness**.

FACTOR 4 The 11 indicators that characterize this factor stem from the group of spherical directivity indices (DI) of the incoming soundfield. All three variations of it are present, covering either

the full sphere (`sphDI`) (5×), the horizontal plane (`sphDIAz`) (1×) or the vertical plane (`sphDIEl`) (5×). The (high) mid frequency range from $f_{c,04} = 500\,\text{Hz}$ to $f_{c,08} = 4\,\text{kHz}$ is most represented here. The DI provides information, whether the incoming sound energy origins from one or more distinct directions or if it arrives from all directions more or less equally. In the context of soundscape this could be an indication on the number and spatial extent of (dominant) sound sources. An example for a high spherical DI would be a loud car passing by (distinct direction at a time despite movement), low DI could be associated with a forest soundscape with unlocateable vegetation noise and arbitrarily surrounding sound sources. Thus, a semantic suggestion would be: **Sound Source Envelopment**.

FACTOR 2 This factor incorporates 7 of 15 indicators that are singular metrics for the spectral composition of the soundscape, namely spectral centroid, entropy, roll-off point, crest factor, skewness and flatness. The other eight indicators comprise loudness and SPL metrics for the low frequency range ($f_{c,00} = 31\,\text{Hz}$). The composition of the loadings with alternating signs (e.g. `spectralCentroid`: positive; `LABands00`: negative) corresponds well such that it can be interpreted as general timbre, whether the spectral composition can be described as low, mid or high frequency. Suggestion: **Timbre**.

FACTOR 15 An interpretation of this is not immediately obvious. Similar to factor 1, level and loudness metrics dominate the composition (13 out of 15 indicators). However, these metrics are present for the two highest frequency bands with center frequencies $f_{c,08} = 8\,\text{kHz}$ and $f_{c,09} = 16\,\text{kHz}$. This agrees well with the indicator with the second largest (absolute) loading, namely sharpness, which characterizes the high-frequency spectral components. Because of that, the semantic suggestion for this factor is: **High-Frequency Timbre**.

FACTOR 5 The indicators of this factor are variations of the modulation depth. The suffixes are either `Sxx`, where `xx` is a placeholder for the frequncy band and `S` refers to the stochastic part of amplitude modulation. The suffix family `Pixx` refer to the depth of the `i`th periodic modulation frequency in the frequency band `xx`. Thus we see that this factor is composed by the depth of the periodic and stochastic modulations of the high frequency bands with $f_{c,08} = 8\,\text{kHz}$ and $f_{c,09} = 16\,\text{kHz}$. This results in the suggestion for a semantic descriptor: **High-Frequency Modulation**

FACTOR 6 Factor 6 is analogous to factor 5 characterized by modulation depth indicators, however this time for the mid frequency

range $f_{c,05} = 1\,\text{kHz}$ to $f_{c,06} = 2\,\text{kHz}$. Semantic suggestion: **Mid-Frequency Modulation**

FACTOR 11 The indicators of factor 11 are composed of various indicator groups. They have in common to represent the low frequency range. The spectral slope describes, if the magnitude of the spectrum increases with higher frequencies (positive slope) or decreases (negative slope). A negative loading for this indicator agrees with positive loadings for other low frequency indicators describing spectral energy. The inter-aural cross correlation IACC on the other hand, a binaural indicator, denotes the similarity between left and right ear signal regardless of the temporal delay (inter-aural time difference, ITD). A large IACC is reached, if the two ear signal have similar temporal structure at a similar magnitude. This is usually the case in simple and uncluttered soundfields, e.g. with only one sound source (if possible in front of the listener) and/or in conditions with certain room acoustic properties (anechoic or simple and undistorted reflection patterns) [102]. The third group of indicators are again formed by the modulation depth in the low frequency range. The interpretation of this composition might be slightly ambiguous. An attempt would be the presence of a distinct and well defined low frequency sound source, hence, the semantic suggestion for this factor is: **Low-Frequency Sound Source**

FACTOR 8 The last of the identified relevant factors is made up by four indicators of the group fluctuation strength. This indicator models the perception of generally higher fluctuation frequencies in the range betwenn $[0.1, 100]\,\text{Hz}$. Here in factor 8 this indicator occurs for the frequency bands 5, 6, 8 and 9, covering the high mid to high frequency range $f_{c,05} = 1\,\text{kHz}$, $f_{c,06} = 2\,\text{kHz}$, $f_{c,08} = 2\,\text{kHz}$, $f_{c,09} = 16\,\text{kHz}$). Semantic suggestion: **Mid-High-Frequency Fluctuation**.

To summarize this important section on the development and identification of the underlying acoustic dimensions Table 3.8 shows the semantic descriptors that are used in the remainder of this work.

As mentioned above this methodology was previously applied, analyzed, and published by the author in [66]. There, the dimensions "Dynmaic Range" and "Loudness Progression" were identified instead of the dimensions MF-MOD (F) and LF-SOURCES (G). The fact that these dimensions do not appear anymore is their composition from the indicator representing the LUFS range. This indicator was not considered in this work because its calculation procedure is time-integrating and leads to non-independent time observations, which in turn violates the basic requirements for factor analysis. Thus, the dimensions presented here already include the first steps of validation and optimization.

Table 3.8: Semantic descriptors for underlying acoustic dimensions.

Dim.	Fac.	Expl. Var.	Descriptor	Label
A	1	87.77 (28.9 %)	"Loudness"	Loud (A)
B	4	16.77 (5.5 %)	"Sound Source Envelopment"	Envel (B)
C	2	12.92 (4.3 %)	"Timbre"	Timbre (C)
D	15	11.23 (3.7 %)	"High-Frequency Timbre"	HF-Timbre (D)
E	5	8.67 (2.9 %)	"High-Frequency Modulation"	HF-Mod (E)
F	6	7.59 (2.5 %)	"Mid-Frequency Modulation"	MF-Mod (F)
G	11	5.96 (2.0 %)	"Low-Frequency Sound Sources"	LF-Sources (G)
H	8	5.29 (1.7 %)	"Mid-High-Frequency Fluctuation"	MHF-Fluct (H)

3.3.3 *Alternative Methods*

The method used to identify the underlying acoustic dimensions, namely factor analysis (FA), was chosen for several reasons, including statistical simplicity and robustness, widespread use of the method in similar and neighboring disciplines and research questions, and clear and straightforward path of interpretation. Needless to say, there are other methods and approaches to target a similar outcome. Thus, some examples of alternative methods are presented in the following discussing basic principles as well as advantages and disadvantages.

Principle Component Analysis (PCA)

As already mentioned in the development of the acoustic dimensions, PCA has a lot in common with FA. Strictly speaking, PCA is a special case of FA with the assumption, that all indicator's variance is kept and all unique variance is zero $\epsilon_i = 0 \forall i$. This implies all observations of all indicators are assumed to be valid disregarding any erroneous influences such as measurement and analysis noise. Thus, PCA is mainly a dimension reduction method rather than a method to discover latent constructs. Due to the above assumptions the calculation of PCA is even easier than for FA, though. It can be realized by eigenvalue decomposition or singular-value decomposition of the input covariance matrix.

Independent Component Analysis (ICA)

The ICA (cf. [103]) is a method used for blind source separation like the *cocktail party problem*. Considering N simultaneous microphone recordings $x_i(t)$ in a room where M persons talk to each other $s_j(t)$. The individual recordings are then mixtures of the spoken signals

$$\begin{aligned} x_i(t) =& a_{i0}s_0(t) + a_{i1}s_1(t) + \ldots + a_{ij}s_j(t) \\ & \text{for} \quad i = 0 \ldots N-1, \quad j = 0 \ldots M-1 \end{aligned} \tag{3.9}$$

where a_{ij} denotes a real constant factor of the mixing matrix A resembling the problem formulation in matrix notation as $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$. The aim is now to discover the originally spoken signals $\boldsymbol{s}$ and, as a side product the respective mixing matrix $\boldsymbol{A}$ which represents the sound propagation and room reflections in the above example of the cocktail party problem. ICA provides solutions to this generative model by means of an adaptive process. First, a pre-whitening is applied to the observed variables $\boldsymbol{x}$ as $\boldsymbol{y} = \boldsymbol{V}\boldsymbol{x}$, such that $\boldsymbol{y}$ is decorrelated $E\{\boldsymbol{y}\boldsymbol{y}'\} = \boldsymbol{I}$, e.g. by means of a PCA. Finally the basis signals $\hat{\boldsymbol{s}}$ are iteratively estimated by a suitable rotation $\boldsymbol{U}$ that maximizes the non-normality of the signal densities $\hat{\boldsymbol{s}} = \boldsymbol{U}\boldsymbol{y}$. This is because it can be assumed that the sum of non-normal random variables approaches normality with increasing number of observations as in the central limit theorem [104]. For the maximization of the non-normality various methods exist, e.g. by maximization of kurtosis or negentropy or by a fix-point approach [105].
From that follows that the benefits of ICA is the identification of hidden time series or signals. An analogue application to identify acoustic dimension as independent components (IC) would then result in a highly overdetermined model since $N_{\text{ind}} \gg N_{\text{IC}}$ which is likely to introduce unknown artifacts. However, the method is used in this work to compare the time series of the resulting factor scores of acoustic environments that are very similar and share the same temporal context. An application of the ICA is used for the comparison of music reproduction of the same audio content but with different loudspeaker setups in Chapter 6.

Further Methods of Machine Learning

Computational methods that are subsumed under the term *Machine Learning* (ML) incorporate statistical concepts as well as emulations of neural networks in order to predict a desired output. In most cases ML is used to conduct one of two tasks, namely classification or regression. While in classification an observation of specific variables is predicted to belong to one of several predefined and discrete classes, regression predicts a quasi-continuous output value on the basis of the observations of input variables. The respective prediction models are developed in advance on basis of supervised and/or unsupervised learning of a wide range of potential observations. The application of the determination of the underlying acoustic dimension corresponds to the search for suitable input variables for a ML model of any kind. This preprocessing task is also called feature extraction in ML, and various suitable methods are available. In this section a small selection is briefly summarized, a detailed comparison of the presented approaches and others is out of scope of this work.
The previously described methods of FA, PCA, and ICA are usually considered as options of feature extraction. Further methods are

e.g. kernel PCA, autoencoder, or t-distributed stochastic neighbor embedding (t-SNE).

Kernel PCA aims to overcome the linearity constraints in PCA and FA [106]. Since the latter two methods are only capable to aggregate linear dependencies between input variables, certain pre-processing steps must be conducted as described in Section 3.1 which in turn might lead to information loss or corruption. Kernel PCA then introduces an arbitrary mathematical kernel that is applied in between the original data space and the target space. However, this kernel is fixed for all input indicators and its application on linear dependencies may be erroneous.

Autoencoders are realizations of neural networks with the aim to compress data and reduce data dimensionality [107]. The input variables $\boldsymbol{x}$ are first encoded into a representation of lower dimension $\boldsymbol{h}$ by means of any (also non-linear) activation function and subsequently decoded into a reconstruction of the input $\boldsymbol{y} = \hat{\boldsymbol{x}}$ as output with another activation function. The encoding and decoding schemes are optimized by means of a cost function which compares input and reconstruction for example by its mean squared error (MSE) $L(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|$. The intermediate representation $\boldsymbol{h}$ would be of interest in the context of this work and could be an analogy to the underlying acoustic dimensions. However, in the case of autoencoders the number of dimensions of $\boldsymbol{h}$ must be predetermined and its representation is not based on statistical or logical relationships but rather on its ability to reconstruct the input with as low error as possible. Thus, its use is mainly appropriate for mere irrelevance reduction rather than for interpretable dimension aggregation. However, due to its ability to incorporate non-linear activation functions it might be a reasonable method for cross-validating the presented results of FA.

Finally, t-distributed stochastic extraction is a method originally developed for visualization of high-dimensional data in a two- or three-dimensional data plot [108]. This is done by calculating probabilities that two high-dimensional data points are similar. The distribution of these probabilities is then compared to the distribution of the original data and is optimized in a way, that similar data points are located close to each other on a map while dissimilar points are located far from each other. Again, the interpretability of the aggregation is not the main motivation here.

3.4 ASSESSMENT

After developing underlying acoustic dimensions for describing acoustic environments in this chapter, application examples for the assessment of specific research questions on soundscapes are discussed in the three following chapters. For that, appropriate analysis procedures must be defined. In general, the description of acoustic environments

with the above developed methodology benefit if comparisons of two or more environments are conducted. For doing that, two approaches are discussed here, namely statistical analysis of significant differences for gathering quantitative evidence and appropriate visualizations for qualitative assessment.

3.4.1 *Statistical Procedures*

In the following chapters of validating applications statistical methods are applied in order to make statements on the similarity and difference between various acoustic environments. Since the choice of appropriate statistical methods is usually being made on basis of the hypothesis type, data structure and data prerequisites, the methods utilized in the following chapters are being discussed briefly at this point.

First, the data structure is described. The data matrix on which basis significant differences are being investigated has either of two sizes and shapes,

STRUCTURE 1: $[N_{\text{dim}} \times N_{\text{ae}} \times N_{\text{cond}} \times N_{\text{t}}]$

STRUCTURE 2: $[N_{\text{dim}} \times N_{\text{ae}} \times N_{\text{t}}]$,

where N_{dim} denotes the number of dimensions, N_{ae} the number of acoustic environments, N_{cond} the number of conditions of each acoustic environment, and N_{t} the number of time samples per condition. The data structures differ in whether the acoustic environments are present in different conditions or not. In the first case, the hypothesis is usually constructed around the question whether differences exist between the conditions for all acoustic environments. The second case implies hypotheses comparing the acoustic environments themselves. In the case of the investigation on acoustic ecologic validity for example, we compare two conditions of each acoustic environment, namely the original recording and its respective re-recording. With the information above we can construct a data structure type 1. The first decision to be made is whether parametric or non-parametric statistical methods can be applied. Parametric methods, such as ANOVA or t-tests require at least three properties to be met: normality of each group, equal variance in each group (homoscedasticity), and independent samples within each group. A group in our case would be each time vector. Due to the linear combinations of indicators within the dimensions, factor scores can be assumed to be of interval level. Further, normality can be assessed either by visually observe the histogramm of the values with overlay of a normal distribution, or by a specific Q-Q-plot which compares the quantiles of the real data with those of an ideal normal distribution, or by means of a quantitative test such as the Shapiro-Wilk test [100]. Usually, with increasing data size of independent random variables the distributions of regular processes

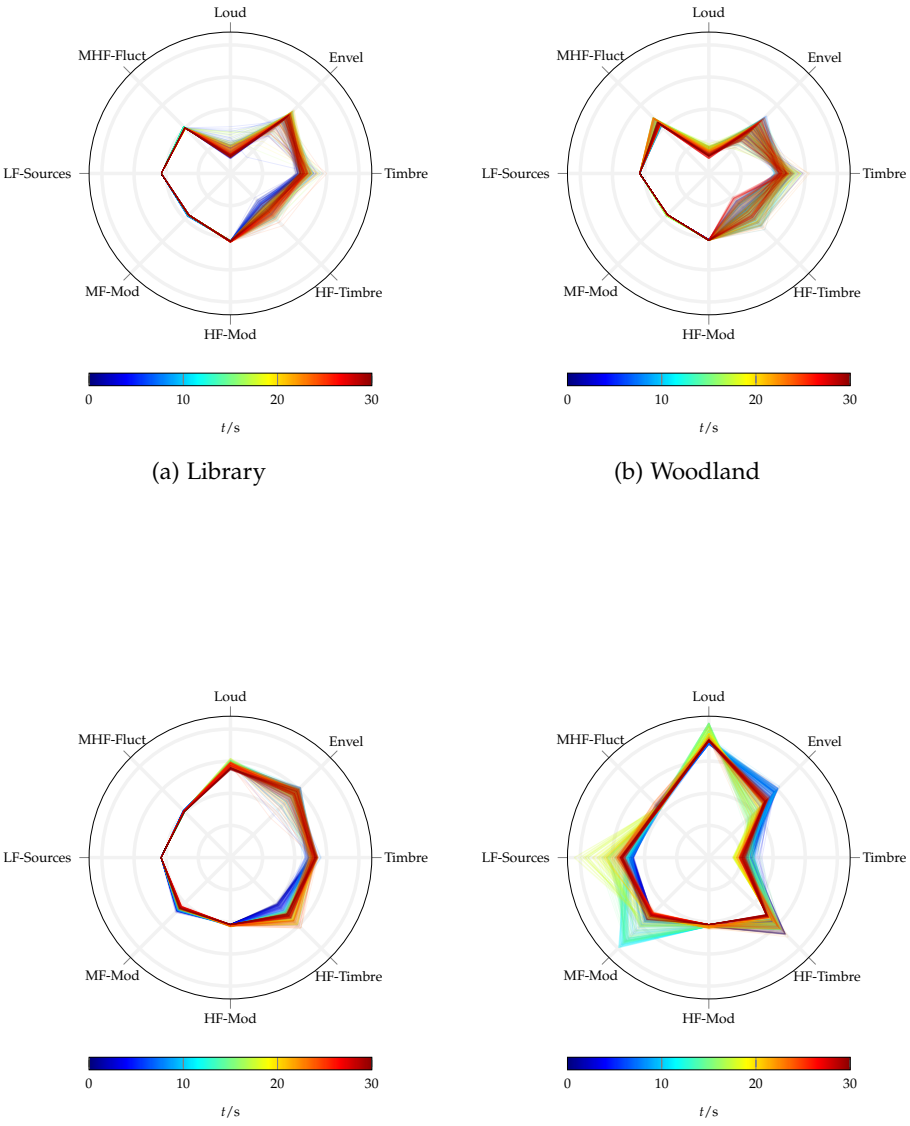
approach normal distribution, which is known as central limit theorem (CLT). However, these tests failed for the dimension scores of acoustic environments. This is not surprising since each time sample of any dimension is dependent on the previous. Acoustic events may occur more or less randomly but inhibit certain deterministic properties. For example a sound source that moves spatially back and forth between point A and point B does not randomly appear at point C which is not located anywhere on the trajectory between A and B. The same is valid for the homogenity of variance in all groups. This property is highly dependent of the characteristic of the acoustic environment and thus can not be asserted at all. Thus, the prerequisites for parametric tests can not be met. With $N_t = 600$ the central limit theorem (CLT) might be reached though [100] where normality can be generally assumed, however the CLT seems not to overcome type II errors (false negative; actual differences are not being detected) [109]. After all, non-parametric methods are utilized for the following analysis.

For the validation parts in the following chapters comparisons of conditions of the same acoustic environment are conducted, thus dealing with data structure type 1 including conditions. Each acoustic dimension is then analyzed separately and isolated, thus, the data structure reduces to $[N_{ae} \times N_{cond} \times N_t]$ for each dimension respectively. The acoustic environments themselves are just random samples covering as wide a range as possible so that the conditions can be focused on. That is why the test design is not two-factorial (acoustic environment AND condition) but a complete block design and since the observations of the conditions are not independent but are collected for the same subject (here: acoustic environment) we assume repeated measures (analogous example: a patient before and after a drug treatment). The above described data structure and hypothesis type leads to the utilization of **Friedman** test for testing the null-hypothesis H_o: *"A difference between conditions could not be found within the analyzed acoustic environments."* This test is based on—as most non-parametric methods—the ranks of the observations rather than on the their absolute or relative values. If the Friedman test exhibits significant differences within the conditions (p-value$< \alpha$ with $\alpha = 0.05$) for a specific acoustic dimension, appropriate post hoc tests are applied. A pairwise application of *Conover* tests [110] reveals the differences between individual conditions (e.g. condition 1 vs. condition 2, cond 1 vs. cond 3,...) which again takes all acoustic environments into account. Beside that, the *Wilcoxon* signed-rank test as non-parametric alternative to the paired t-test may be used in multiple condition comparisons to test on differences between each condition for each acoustic environment. Summarized, the Friedman test is used in this work to test whether there is a difference among conditions at all, the Conover test shows, which condition pair specifically exhibit

differences and the Wilcoxon test reveals the acoustic environments in which the condition differences occur.

3.4.2 *Visualization: The Soundscape Fingerprint*

The basis of soundscape dimension comparisons are the factor scores $\boldsymbol{Y}$. $\boldsymbol{Y}$ is present as time series of the respective dimensions which is why a first straightforward approach is to investigate similarities of acoustic environments on the basis of line plots. A direct comparison of time series in the form $x_2(t) - x_1(t)$ reveals usually not much insight unless both acoustic environments show strong similarities. One step of abstraction is to observe distributions of the factor scores regardless of the temporal behaviour. This can be done, for example, using boxplots showing the median and the 25 % and 75 % quartiles, as well as the whiskers representing the outer ends of the distribution. For symmetric distribution, such as normal distribution, the representation can also be conducted with mean and respective standard deviation. However, as we will find out later, the distribution of factor scores in the scope of this work are not normally distributed for systematic reasons which is why the use of boxplots is qualified. A single boxplot then represents the distribution of a single acoustic dimension for a specific soundscape recording. In order to compare all eight relevant acoustic dimensions, another visualization method is presented here, which resembles a fingerprint and thus serves as the namesake of the presented methodology in this chapter. It consists of a polar plot where each axis represents one of the acoustic dimensions. Each short-time observation of the dimensions is plotted as faint line between these axes. The color-coding represents the temporal succession. This representation provides a summary overview of the acoustic dimensions of a given soundscape recording, as well as the ability to visually compare two or more soundscapes. An example can be found in Figure 3.5.



(a) Library

(b) Woodland

(c) Shopping center

(d) Train station

Figure 3.5: Four exemplary acoustic fingerprints of soundscape recordings.

4

VALIDATION I: PERCEPTUAL EVALUATION

The description of the acoustic properties of soundscapes as proposed in this work relies on physical signal properties. At the same time, all attempts to analyze real-world acoustic environments – not only in soundscape studies, but also in regulatory or commercial settings – target human auditory perception. This fact cannot be emphasized more and therefore will be considered in different facets in the present work. We have already seen that the choice of initial signal indicators was made with perceptual properties in mind. Also the interpretation of the identified acoustic dimension was made from an empirical perspective incorporating human auditory experience. Thus it is straightforward to evaluate the acoustic dimensions that are based on statistical properties of physical signal parameters in terms of perception. This is also an important – if not mandatory – step towards validation of the approach. In order to contribute to this attempt, listening experiments were conducted. The results presented below were published in part in [111].

4.1 METHODOLOGY

The perceptual studies were designed as laboratory experiments with reproduced soundscape recordings.

Stimuli

Suitable excerpts of those soundscape databases that were employed for the dimension development (cf. Table 3.5, p. 21) were produced and processed for reproduction in an laboratory environment. The soundscape stimuli have a length of 30 s each and are faded in and out over 2 s. The selection was made with the following aims:

- Use stimuli of all three data sources with similar share to employ different Ambisonics orders and technologies.
- Cover a wide range of acoustical environments: indoor – outdoor, rural – urban – private, annoying – pleasant, loud – soft, ...
- Assure similar distributions of sound source classes according to ISO 12913-2[10]: sounds of technology, nature, human beings.
- Provide at least one soundscape class that is represented within each database (here: train station).

In total 19 acoustic scenes were selected of which two were additionally used as training stimuli. The excerpts and respective soundscape classes can be taken from Table 4.1.

Table 4.1: Sample draw of soundscape excerpts used for perceptual studies. Additional training stimuli are marked with a $*$.

ID	Database	Name	label	Excerpt
1	ARTE	01_Library	A: Library	01:30-02:00
2 $*$	ARTE	02_Office	A: Office	00:35-01:05
3	ARTE	04_Living_Room	A: LivingRoom	00:58-01:28
4	ARTE	07_Cafe_1	A: Cafe	01:28-01:58
5	ARTE	09_Dinner_Party	A: DinnerParty	01:36-02:06
6	ARTE	11_Train_Station	A: TrainStation	00:18-00:48
7	ARTE	12_Food_Court_1	A: FoodCourt	00:36-01:06
8	Eigenscape	Beach.7	E: Beach	05:38-06:08
9	Eigenscape	Park.5	E: Park	05:28-05:58
10	Eigenscape	PedestrianZone.3	E: Pedestrian A	08:00-08:30
11	Eigenscape	PedestrianZone.5	E: Pedestrian B	06:48-07:18
12	Eigenscape	ShoppingCentre.8	E: Shopping	05:24-05:54
13 $*$	Eigenscape	TrainStation.6	E: TrainStation	03:24-03:54
14	Eigenscape	Woodland.2	E: Woodland	03:16-03:46
15	Soundfield	Kids Playground 1	S: Playground	00:47-01:17
16	Soundfield	Rural Market Busker	S: Busker	01:40-02:10
17	Soundfield	Steamtrain Exterior	S: Steamtrain	02:00-02:30
18	Soundfield	St Kilda Road Traffic	S: Traffic	02:56-03:26
19	Soundfield	Southern Cross Station	S: TrainStation	00:30-01:00

Technical Infrastructure

The three databases use different Ambisonics orders, namely 7th-order MOA with 4th-order spherical and additionally 7th-order horizontal Ambisonics for the ARTE database [91], 4th-order for the Eigenscape database and 1st-order for the Soundfield database. The necessary number of individual Ambisonics channels is calculated as $N_{\mathrm{ch}} = (N_{\mathrm{A,max}} + 1)^2$, where $N_{\mathrm{A,max}}$ denotes the maximum Ambisonics order. The Ambisonics signals were arranged in the digital audio workstation (DAW) Reaper [112] which is able to support tracks with the necessary 64 channels each. As decoder from the Ambisonics domain to loudspeaker signals the widely used AllRADecoder from the *IEM Plug-in Suite* [30] was used. The employed decoder strategy AllRAD utilizes a Ambisonics-to-loudspeaker decoding for a large set of position-optimized virtual loudspeaker setup which is in a second step mapped to the real loudspeaker setup by means of appropriate panning algorithms. This decoder strategy is known to be suitable for unevenly spaced loudspeaker setups and therefore a robust solution in many laboratory setups. The reproduction room itself is an acoustically-optimized media laboratory of approx. 30 m^2, the *Immersive Media Lab* [113] (see Figure 4.1). The laboratory provides in total

Figure 4.1: The *Immersive Media Lab* (IML) that was utilized for perceptual studies.

42 full range loudspeakers of type *Neumann KH120* as well as four subwoofers of type *Neumann KH810* of which 30 speakers and two subwoofers were employed for the reproduction. Each loudspeaker is equalized individually in terms of gain, delay, and flat frequency response in order to fulfill the requirements of ITU-R BS.1116-3 [114]. The equalization procedure for a listening area in which a subject is located is described by the author in [115].

The stimulus audio files were already calibrated for the dimension development part as described in Section 3.2. However, due to the different Ambisonics order and the decoder strategy a calibration of sound pressure level was employed according to the information provided within the recording databases. Since the Soundfield database does not provide calibration or SPL information, the level was adjusted manually and subjectively in order to have plausible differences compared to the other acoustic scenes. In order to analyze the acoustic properties of the reproduced soundscapes as perceived by the subjects in the study, the reproduction was re-recorded with an Eigenmike® EM32. The difference between the original recordings and these re-recordings as well as the influence of the reproduction setup in general will be discussed in detail later in Section 5. For the re-recorded scenes the indicators were calculated and the respective factor scores were deduced. The resulting fingerprints of the soundscape stimuli are collected in Figures B.4 to B.7 of Appendix B. The aggregated distribution of the scores for each dimension and each stimulus scene can be taken from Figure 4.2. A Kruskal-Wallis test on ranks proposes significant differences among the samples within all acoustic dimensions ($H > 9500$; $p < 0.01$). Subsequently, pairwise Dunn's posthoc

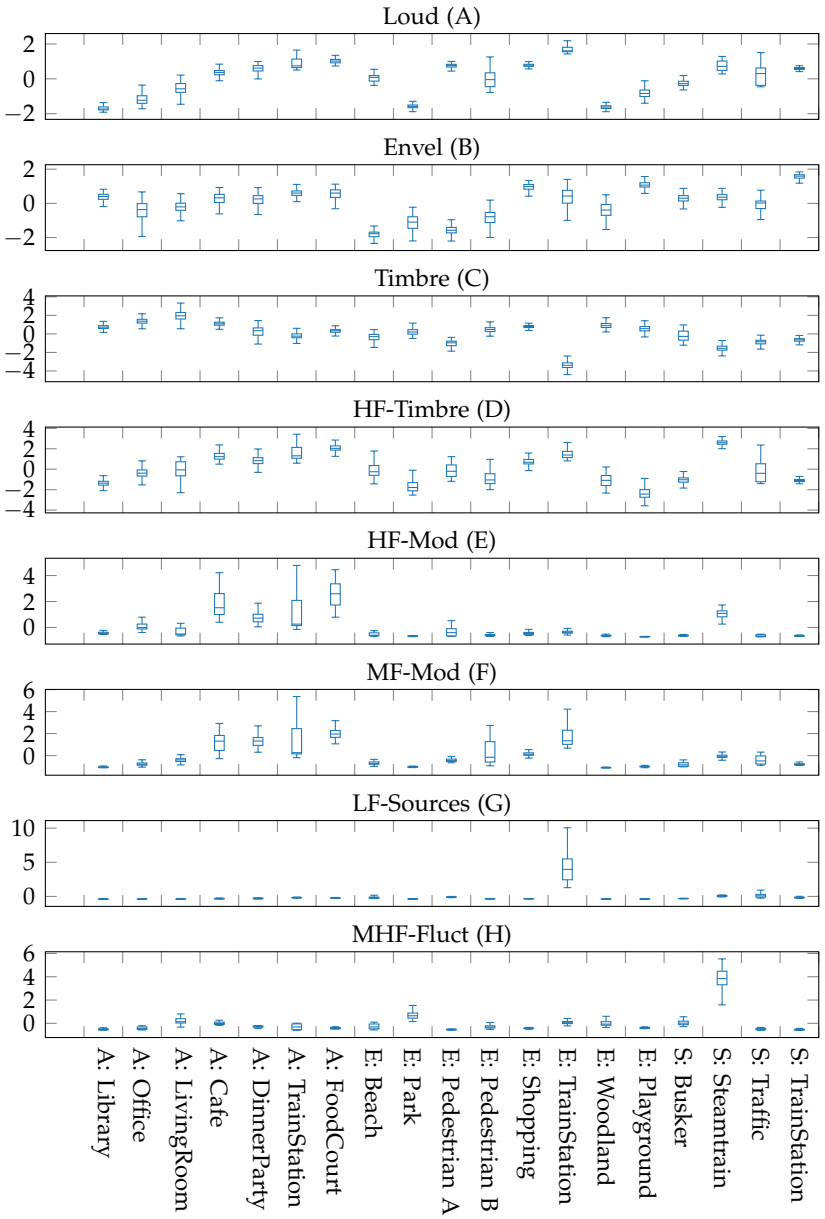


Figure 4.2: Distributions of factor scores of stimuli presented at the listening experiment.

tests with Bonferroni adjustment were performed comparing all 19 samples with each other for each dimension. The result whether each comparison pair differs significantly can be found in Figure 4.3. It is noteworthy that the majority of these comparisons exhibit strong significant differences with $p < 0.01$ (**, red tiles). This result might be influenced by the relatively large number of observations (30 s $\times$ 20 observations/s) and should at this point only describe the difference from a statistical point of view.

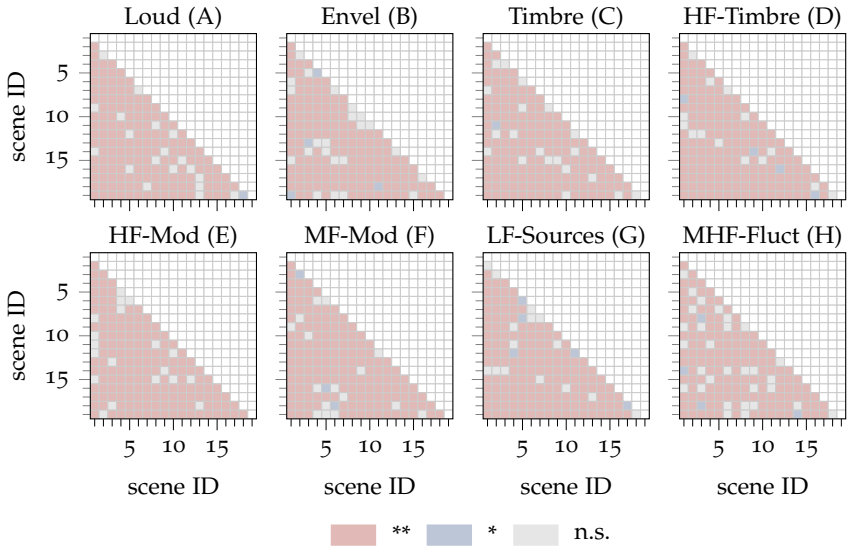


Figure 4.3: Statistical differences between re-recorded sample soundscapes. Red: strong significance ($p < 0.01$), blue: moderate significance ($p < 0.05$), gray: no significance ($p > 0.05$).

Experiment procedure

Participants for the listening experiment were acquired among staff and students of the *Institute of Communications Technology* as well as among individuals from outside the institute, with the aim to obtain as balanced a demographic spectrum as possible. Demographic parameters were collected for all participants, namely place of residence (rural, urban), gender, age and listening experiment experience. In total 20 subjects participated the experiment. The demographic composition can be taken from Table 4.2. The noise sensitivity is assessed

Table 4.2: Statistics of demographic characteristics of all 20 listening experiment participants.

Characteristic	Statistic
Gender	male: **12**, female: **7**, no answer: **1**
Age	range: **[20-69]**, mean: **40.4**, std: **16.7**, median: **35**, P25: **28**, P75: **59**
Habitation	countryside/village: **4**, town: **8**, city: **8**
Hearing ability	very good: **4**, good: **14**, mediocre: **2**, bad: **0**
Participation in listening experiments	0 times: **8**, 1-3: **3**, 4-6: **6**, >6: **3**
Experience (deduced from above)	inexperienced: **11**, experienced: **9**
Mood	very good: **2**, good: **13**, mediocre: **4**, bad: **1**
Very tired	yes: **4**, no: **14**, no answer: **2**
Noise sensitivity	range: **[33-78]** %, mean: **56** %, std: **15** %, median: **56** %, P25: **44** %, P75: **71** %

by a set of nine 4-point items proposed by Zimmer and Ellermeier in [116] which is the short version of a more detailed questionnaire on personal noise sensitivity by the same authors [117]. It is calculated as sum of the answers which results in a value range between 0 and 27 for the respective noise sensitivity. In the above table, the noise sensitivity is scaled with 1/27 and represented as percentage of that ratio.

Also, for testing the hypothesis of whether there are different response patterns between individuals with and without listening experience, the participants were assigned to two categories. Individuals who reported having participated in listening experiments either 0 or 1-3 times were assigned to the group "inexperienced", whereas persons reported 4-6 or >6 participations were assigned to the group "experienced".

The following procedure of the experiment was executed for each participant:

1. **Welcome** Short welcome to the laboratory as well as comforting and calming down participants if necessary.

2. **Procedure information** Information about the experiment procedure was shared (duration, progress etc.), graphical user interface for the questionnaire was presented.

3. **Explanation of nomenclature** The terms and rating parameters were explained with similar wordings and examples throughout the participants but without pre-defined text.

4. **Training 1** The first training stimulus was played back and the questionnaire was filled in jointly by the participant and the examiner.

5. **Training 2** The second training stimulus was played back and the questionnaire was filled in by the participant alone under supervision and support by the investigator.

6. **Start of experiment** Open questions were answered and afterwards the participants were left alone in the laboratory.

7. **Experiment** Each stimulus is presented isolated without A/B comparison or similar. The playback could be started unlimited times and questions/items could be answered throughout the listening experience. Participants were encouraged to take breaks and help themselves with snacks and beverages provided.

8. **End of experiment** After the last stimulus, participants were dismissed from the laboratory room and informally interviewed about the experience. This information was only collected as technical and methodological lessons learned.

9. **Demographic questionnaire** A pen-and-paper questionnaire about demographic details and noise sensitivity was filled in by the participants.

The study consists of three parts, each addressing a different aspect of soundscape perception. The first part investigates what sound sources or sound source classes are perceived by the participants as elaborated in Section 4.2. The second part aims at finding semantic acoustic descriptors for the identified acoustic dimensions. This part is an important step for the validation of the presented methodology and therefore occupies a prominent part (Section 4.3). The third part described in Section 4.4 finally consists of emotional responses of perceived affective quality. It represents the far end of the soundscape framework and serves as one of the main goals of providing descriptive acoustic dimensions as causal background. All three parts were assessed for one stimulus after the other by means of a graphical user interface shown in the appendix Figure B.3 The first and third experiment parts are taken directly from the suggestions of ISO/TS 12913-2. Since they do not directly contribute to the aim of validating the identified acoustic dimensions, they are only reported in a descriptive manner for completeness. The focus in the following will be on the search for semantic counterparts of the acoustic dimensions.

4.2 STUDY I: SOUND SOURCES (ACC. ISO/TS 12913-2)

The ISO/TS 12913-2 suggests three methods for the assessment of soundscapes from a human perspective. All of them are meant to be conducted in-situ, especially during a soundwalk. The most applicable method for laboratory studies with a reproducible experiment design is method A (cf. section C.3.1 in [10]) which provides quantifiable assessment of perceived sound source types, perceived affective quality, overall sound environment quality, and sound environment appropriateness. The first aspect, namely the identification of present sound source classes, is requested in this first part of the experiment.

The participants are asked to what extent they hear the following sound source classes: sounds of technology, sounds of human beings, and sound of nature (see questionnaire in appendix Figure B.3) on an ordinal scale from “not at all” to “a little”, “moderately”, “a lot”, up to “dominates completely”. The resulting distributions are depicted in Figure 4.4 where the ordinal levels are represented by the numbers 1 to 5. It can be seen that the ends of the scale (level 1 and 5), i.e. the absence or complete dominance of a sound source class, are reported quite homogeneously among the 20 participants for example for stimulus E: BEACH or E: WOODLAND. However, this can not be stated for stimuli that are more or less balanced mixtures of sound sources such as E: PARK or S: BUSKER. There, a broader distribution of reported levels of sound source classes can be observed.

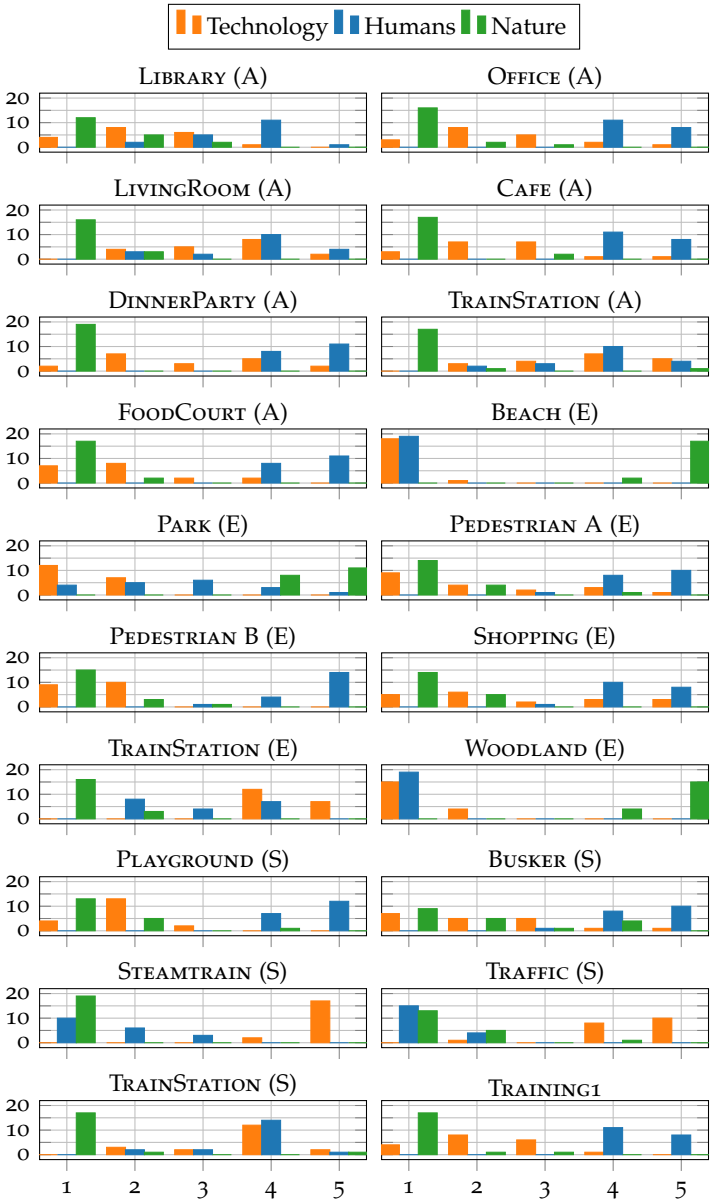


Figure 4.4: Histogram of perceived presence of sound source classes according to ISO/TS 12913-2: sounds of technology (orange), sounds of human beings (blue), sounds of nature (green) from "not at all" (1) to "a little" (2), "moderately" (3), "a lot" (4) to "dominates completely" (5).

If this experiment part is seen as a plausibility check, these results are satisfactory because the unambiguous stimuli were robustly rated by participants in terms of the presence of sound source classes.

4.3 STUDY II: SEMANTIC DESCRIPTION (ACC. SAQI)

An important test within the methodology of acoustic dimensions identification is to find appropriate semantic descriptors that can be

used by everyone to describe the acoustic properties of environments. For that the subjects were asked to rate the selected acoustic scenes by means of semantic differential items according to the *Spatial Audio Quality Inventory* (SAQI). This collection of attributes was developed by expert groups to form a common basis of descriptors for the rating and evaluation of spatial acoustic environments [118, 119]. Although it was originally developed for the technical evaluation of reproduced audio experiences with headphones (binaural synthesis) or multichannel loudspeaker systems (wavefield synthesis, Ambisonics, panning, ...), this inventory can also be used in parts for the evaluation of real soundscapes and, of course, the reproduction of soundscapes in laboratory environments. It consists of 48 descriptors within the eight categories timbre, tonalness, geometry, room, time behavior, dynamics, artifacts and general impression. The descriptors are provided as semantic differential with two opposing attributes. For the listening experiment in this paper, eight of the SAQI items were used as well as two additional attributes that are either defined differently in the SAQI manual (Envelopment) or additionally defined by the author (Fluctuation). These items, listed in Table 4.3, were selected in order to find potential perceptual counterparts for the semantic description of the acoustic dimensions from Table 3.8, p. 32.

Table 4.3: Selected SAQI items used for the experimental study on semantic descriptors. The * denotes additional items introduced by the author.

Quality (Label)	**Semantic Differential**	**Circumscription**
Loudness (Loud)	quieter – louder	Perceived loudness of a sound source. Disappearance of a sound source can be stated by a loudness equaling zero. Example of a loudness contrast: whispering vs. screaming.
Dynamic Range (Dyn)	smaller – larger	Amount of loudness differences between loud and soft passages. In signals with a smaller dynamic range loud and soft passages differ less from the average loudness. Signals with a larger dynamic range contain both very loud and very soft passages.
Fluctuation* (Fluct)	less pronounced – more pronounced	Short-term fluctuations in loudness, e.g. due to speech, knocking, etc.
Tone color (Timbre)	darker – brighter	Timbral impression determined by the ratio of high to low frequency components.
Sharpness (Sharp)	less sharp – sharper	Timbral impression which e.g., is indicative for the force with which a sound source is excited. Example: Hard/soft beating of percussion instruments, hard/soft plucking of string instruments (class. guitar, harp). Emphasized high frequencies may promote a 'sharp' sound impression.

Continued on next page

Table 4.3 – continued from previous page

Quality (Label)	**Semantic Differential**	**Circumscription**
Localizability (Local)	more difficult – easier	If localizability is low, spatial extent and location of a sound source are difficult to estimate, or appear diffuse, resp. If localizability is high, a sound source is clearly delimited. Low/high localizability is often associated with high/low perceived extent of a sound source. Examples: sound sources in a highly diffuse sound field are poorly localizable.
Distance (Dist)	closer – more distant	Perceived distance of a sound source.
Envelopment* (Envelop)	less pronounced – more pronounced	Impression whether the main activity of the acoustic scene is assigned to distinct direction(s) or if the dominant sound source(s) stem from all around the listener.
Naturalness (Natural)	lower – higher	Impression that a signal is in accordance with the expectation/former experience of an equivalent signal.
Presence (Presence)	lower – higher	Perception of "being-in-the-scene", or "spatial presence". Impression of being inside a presented scene or to be spatially integrated into the scene.

All semantic differentials of the items were provided as quasi-continuous Likert scales in the range from -50 to 50 with a stepsize of 1 implemented as slider. An exemplary page of the graphical user interface for a training stimulus can be taken from the appendix, Figure B.3. For the processing of results the values were scaled by the factor $1/50$ to obtain a value interval of $[-1, 1]$.

Beside the raw subjective responses an additional normalization was applied to the ratings which scaled the ratings to the intraindividual absolute maximum rating of all stimuli for a specific item. The following example reflects the effect of this scaling.

Example. *A person rates the item "Dyn" in the range of* $[-0.3, 0.25]$ *among all stimuli and does not use the full capacity of the scale. A scaling factor of* $a = \frac{1}{0.3}$ *is hence applied for all ratings of the item "Dyn" of this specific individual.*

The motivation for scaling item responses is the issue of reference when rating items. The SAQI methodology is based on comparisons either between two conditions in an A/B setup or as a comparison with a subject's internal reference. If all subjects used the same internal reference, no scaling would be required. However, if the reference differs between subjects, the scaling described above helps to reduce this effect. Since it is unknown, if intraindividual scaling to respective inner reference is necessary, both datasets are kept for analysis and are referred to as "unscaled" and "scaled" hereinafter.

Figure 4.5 shows the distributions of the unscaled evaluation responses of the perceptional study on acoustic properties for all subjects as well as discriminated between experienced and inexperienced listeners. Since the distributions seem to be of great variance, the rating

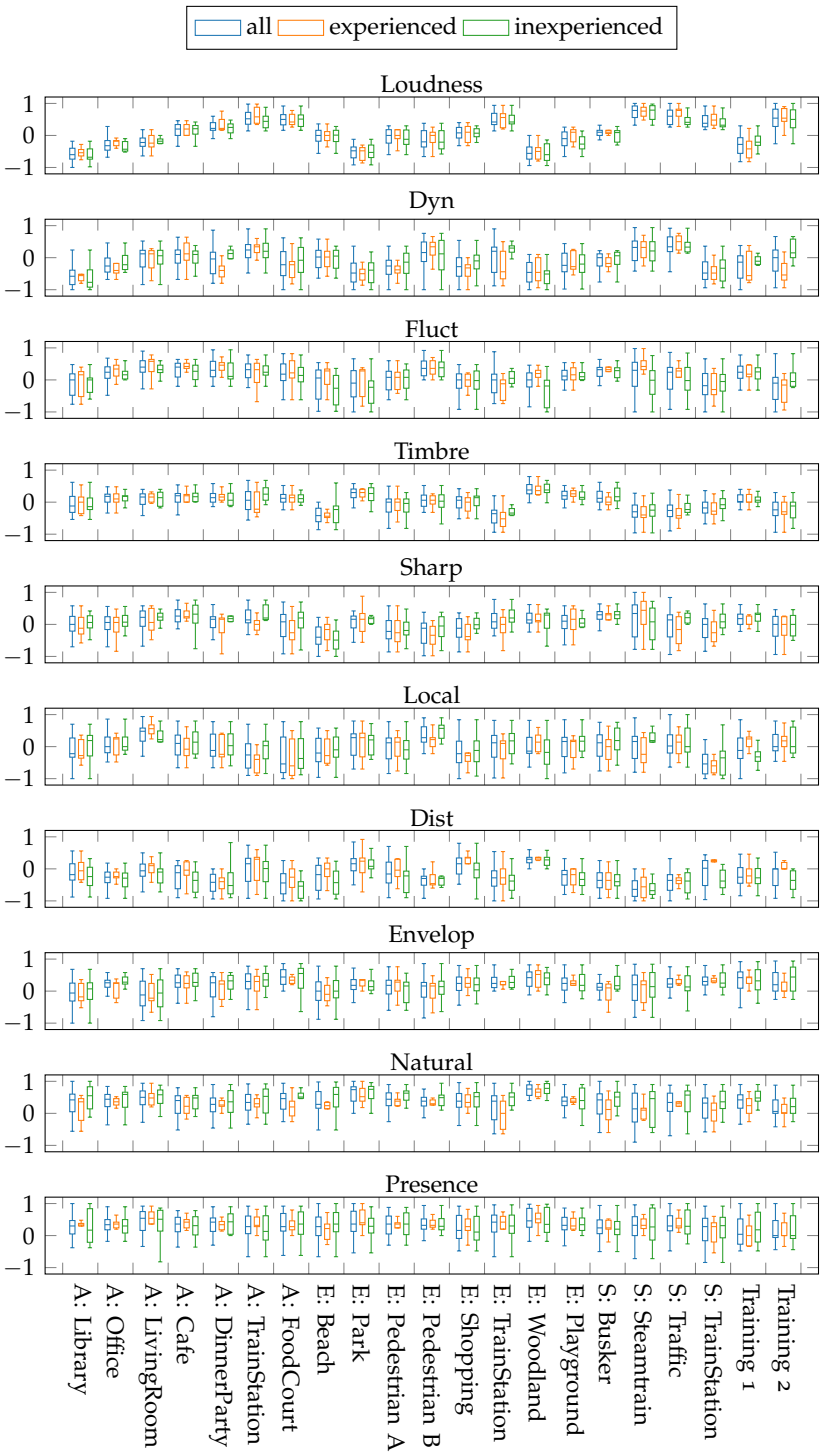


Figure 4.5: Distribution of perceptional item ratings.

pattern of the subjects is analyzed in terms of interquartile range (IQR) to grasp evidence of the consistency or dispersion of answering patterns. The IQR represents the range of values in which the central 50 % of the responses lie. It was calculated for each stimulus and item independently and subsequently averaged for each item. The results for the mean IQR and respective standard deviation in parentheses for each perceptional item is listed in Table 4.4 distinguished between inexperienced, experienced and all subjects as well as between unscaled and scaled responses. Observing the IQRs of the unscaled

Table 4.4: Interquartile range (IQR) and respective standard deviation in parentheses as indicator of consistency within the answering pattern. Left: unscaled ratings, right: scaling to intraindividual minimum and maximum.

	unscaled			scaled		
	all	exp	inexp	all	exp	inexp
Dist	0.51	0.39	0.49	0.63	0.55	0.51
	(0.16)	(0.18)	(0.14)	(0.19)	(0.24)	(0.18)
Dyn	0.56	0.46	0.49	0.71	0.56	0.61
	(0.12)	(0.19)	(0.18)	(0.24)	(0.23)	(0.26)
Envelop	0.45	0.36	0.50	0.60	0.59	0.58
	(0.16)	(0.18)	(0.16)	(0.23)	(0.26)	(0.22)
Fluct	0.53	0.45	0.49	0.73	0.56	0.70
	(0.17)	(0.20)	(0.20)	(0.26)	(0.30)	(0.28)
Local	0.64	0.55	0.60	0.90	0.77	0.80
	(0.14)	(0.21)	(0.17)	(0.24)	(0.33)	(0.25)
Loud	0.36	0.33	0.35	0.40	0.35	0.39
	(0.09)	(0.11)	(0.11)	(0.10)	(0.13)	(0.15)
Natural	0.49	0.36	0.50	0.50	0.46	0.48
	(0.15)	(0.20)	(0.21)	(0.23)	(0.32)	(0.27)
Presence	0.53	0.33	0.66	0.63	0.47	0.72
	(0.12)	(0.13)	(0.19)	(0.21)	(0.22)	(0.31)
Sharp	0.50	0.50	0.41	0.71	0.69	0.60
	(0.17)	(0.18)	(0.19)	(0.24)	(0.28)	(0.28)
Timbre	0.34	0.31	0.33	0.57	0.56	0.54
	(0.09)	(0.12)	(0.10)	(0.18)	(0.25)	(0.17)
mean	0.49	0.40	0.48	0.64	0.56	0.59
	(0.16)	(0.19)	(0.19)	(0.25)	(0.28)	(0.26)
ρ_T	0.834	0.820	0.851	0.840	0.846	0.848

ratings of all subjects reveals that certain perceptual items have explicitly lower IQRs than others. While the items "Timbre" (0.34) and "Loud" (0.36) produce reasonably small IQR, the items "Local" (0.64) and "Dyn" (0.56) inhibit the largest uncertainty. This behavior can be found similarly for the inexperienced and experienced subgroups. The mean IQR range of the experienced group is slightly smaller (0.40) compared to the inexperienced group (0.48) and all subjects together (0.49). However, these differences are within the respective standard deviation (0.16). Beside the IQR, the Cronbach's alpha, also known as tau-equivalent reliability, ρ_T, was calculated. It represents the fit of the

test regarding consistency and ranges (usually) between 0 and 1. A value of > 0.8 represents an appropriate criterion for applied research according to [120] which can be met for all subgroups.

The general behavior of the scaled items with lower and larger IQR remains comparable to the unscaled ratings. At the same time the mean IQR increases distinctly, whereas the reliability remains stable. Since no systematic differences can be observed between unscaled and scaled as well as between experienced, unexperienced and all subjects, the upcoming modeling process incorporates both datasets, scaled and unscaled, as well as all participants without discrimination of experienced and inexperienced subjects.
In order to find interrelations between the identified signal-based acoustic dimensions (cf. Figure 4.2) and the attributes based on subjective perception (cf. Figure 4.5) appropriate statistical analyses were conducted. Generally we are dealing with two distributions that are to be connected: the distribution of short-time factor scores [19 stimuli $\times$ 8 dimensions $\times$ 600 short-time observations] against the distribution of perceptual evaluations among subjects [19 stimuli $\times$ 10 items $\times$ 20 participants]. The shape and type of the data necessitates various methodological assumptions and prerequisites. First, because of their temporal characteristics, the distribution of factor scores cannot be conventionally treated as independent samples, but rather is characteristic of the entire stimulus as a time series. Second, the perceptual items are rated for the entire acoustic scene of 30 seconds. Although subjects were asked to rate the entire listening experience, it may be that the duration immediately preceding the rating has a greater impact than the periods a few tens of seconds before. Since the ratings were allowed to be taken while listening, this effect might be randomized among the subjects. Third, it is not known, whether subjects rated the average of their perception of a specific item or the maximum or something different. If a single dominant acoustic event deviates perceptually from the rest of the stimulus, it can also significantly influence subjects' ratings despite its short-term nature. Keeping these assumptions in mind a correlation process was conducted. It should be reiterated at this point that this work does not attempt to develop entire perceptual models, but rather to contribute to the validation of the identified acoustic dimensions. Hence, correlations between the acoustic factor scores and the item ratings were calculated. To reflect the uncertainty in the previous third assumption, median and maximum of the acoustic factor scores were taken into account for each stimulus and acoustic dimension. Similarly, the median of the subjective rating for each stimulus was utilized since normal distribution and homoscedasticity could not be asserted for all perceptional items within all stimuli. Spearman's correlation of ranks was calculated to reveal general relationships between perceptual items and acoustic dimension rather than exact slopes:

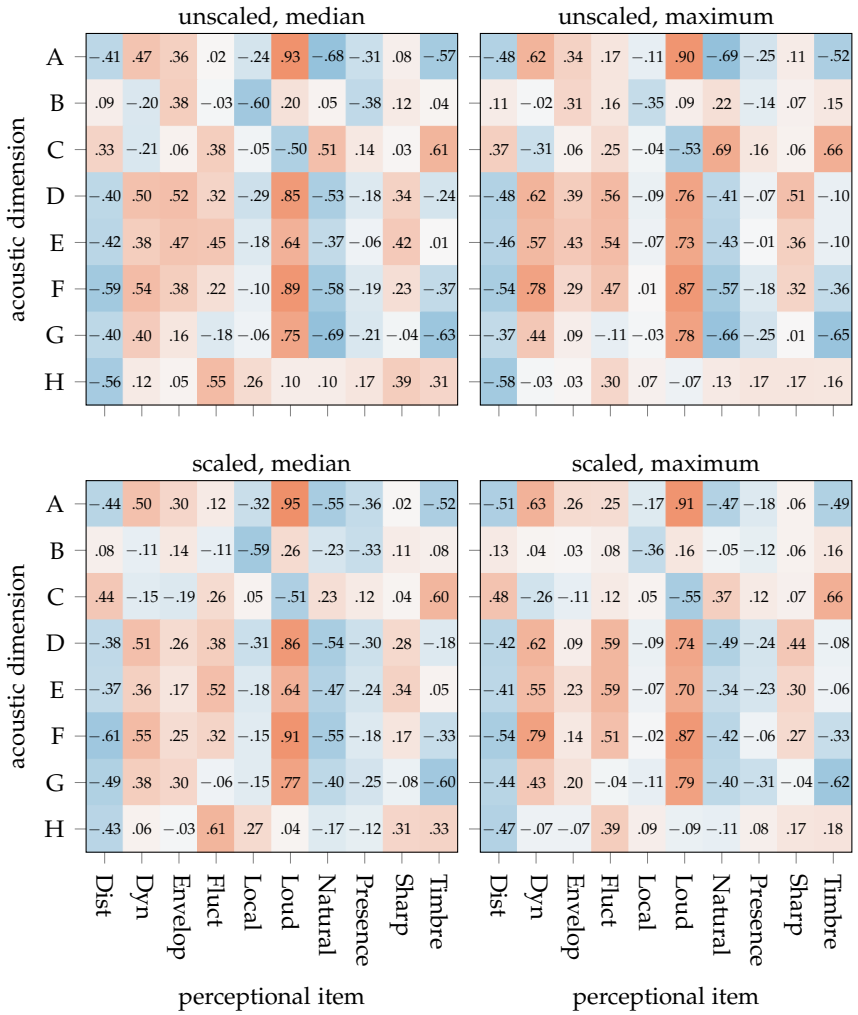


Figure 4.6: Spearman's correlation coefficient matrices between acoustical dimensions and perceptional items. Left: median of acoustic factor scores, right: maximum of acoustic factor scores. Top: unscaled perceptional item responses, bottom: scaling to intraindividual absolute maximum.

Example. *A maximum Spearman's correlation of 1 between the perceptual item "Loud" and the acoustic dimension* Loud *(A) means that the stimulus with the highest median or maximum factor score is also rated with the highest rating of the respective item, the rank order of the stimuli regarding the factor score of* Loud *(A) is the same as regarding the perceptual item of "Loud".*

The resulting correlation factors between acoustic dimension and perceptual item in the range of $[-1, 1]$ are depicted in Figure 4.6. Both, scaled and unscaled item ratings were utilized as well as both location parameters, median and maximum of the acoustic dimensions. Since it can be observed that the choice of scaling (scaled, unscaled) of the perceptive attribute ratings and the location parameter (median or maximum) of the acoustic dimension scores yield different correlations,

Table 4.5: Maximum positive and negative Spearman's correlation R_s for each perceptual item regarding intraindividual scaling and percentile consideration of acoustic dimension score. Moderate and strong correlations $|R_s| > 0.6$ are in bold.

	positive				negative			
Item	Scaling	Pctl.	Dim.	R_s	Scaling	Pctl.	Dim.	R_s
Dist	scaled	P05	C	0.50	unscaled	P90	H	**-0.63**
Dyn	scaled	P85	F	**0.79**	unscaled	P100	C	-0.31
Envelop	unscaled	P05	D	0.58	scaled	P95	C	-0.27
Fluct	scaled	P55	H	**0.68**	unscaled	P60	G	-0.22
Local	scaled	P50	H	0.27	unscaled	P40	B	**-0.62**
Loud	scaled	P50	A	**0.95**	scaled	P90	C	-0.57
Natural	unscaled	P100	C	**0.69**	unscaled	P95	A	**-0.70**
Presence	unscaled	P90	C	0.23	scaled	P10	B	-0.49
Sharp	unscaled	P100	D	0.51	scaled	P00	G	-0.08
Timbre	unscaled	P100	C	**0.66**	unscaled	P100	G	**-0.65**

a systematic analysis on these parameters was conducted. For that, all percentiles from minimum (0 %) to maximum (100 %) in steps of 5 % are calculated for the acoustic dimension scores and for each perceptive attribute the corresponding highest positive and negative correlation are determined as shown in Table 4.5 where correlations of $|R_s| > 0.6$ are set in bold. From this the following findings can be deduced:

- The median of dimension Loudness (A) exhibit strong correlation (0.95) with the perceptional item "Loud" which is an expected result. At the same time distinct negative correlations with the item "Natural" can be observed. Thus, loud soundscape reproductions are rated as unnatural.

- The perceptual item "Local" has its largest negative correlation (-0.62) with Sound Source Envelopment (B) leading to the reasonable explanation that scenes with low acoustic envelopment are rated with higher localizability and vice versa.

- The maximum of dimension Timbre (C) shows reasonable correlation (0.66) with the item "Timbre" which also confirm expectations.

- The dimension High-Frequency Timbre (D) shows moderate positive correlation with the items "Envelop" (0.58) and "Sharp" (0.51) where latter approaches expectations.

- The dimension High-Frequency Modulation (E) shows only low correlation values (cf. Figure 4.6) and thus shows no perceptual counterpart with distinct similarities in this selection.

- The previous finding does not count for the the similar dimension Mid-Frequency Modulation (F) which shows good correlation (0.79) with the item "Dynamic Range".
- The dimension Low-Frequency Sound Sources (G) shows a moderate negative correlation (-0.65) with the item "Timbre" which also seems plausible also in conjunction with the second finding above.
- The P55 of Mid-High-Frequency Fluctuation (H) correlates moderately (0.68) with the item "Fluct" which is plausible but at the same time this dimension shows almost the same amount of negative correlation (-0.63) with the item "Dist" which is a somewhat ambiguous result.

It can be stated that each acoustic dimension except for High-Frequency Modulation (E) can be assigned to one or more perceptual items with distinction. Even though a correlation is not necessarily a result of causality, these findings are promising in such that the identified dimensions as result of statistical signal processing exhibit characteristics that can be reproduced perceptually.

4.4 STUDY III: AFFECTIVE QUALITIES (ACC. ISO/TS 12913-2)

The two dimensions of affective qualities according to ISO/TS 12913-2 between the poles unpleasant and pleasant and between uneventful and eventful respectively span a space to which the main emotional responses can be assigned as depicted in Figure 2.3. The standard suggests an assessment procedure of the two dimensions pleasantness and eventfulness by means of eight descriptors that denote both ends of the main quality axes as well as the respective 45° axes. Each descriptor, namely "pleasant", "calm", "uneventful", "monotonous", "annoying", "chaotic", "eventful", and "vibrant" shall be rated by the participants on an ordinal 5 point Likert scale with the levels "strongly agree", "agree", "neither agree nor disagree", "disagree", and "strongly disagree". This procedure was adopted in the third part of the listening experiment. Subjects were asked to rate each descriptor individually, with possible contradictions, such as agreeing with both eventful and uneventful, allowed. For analysis the ISO/TS 12913-3 proposes to assign the equally spaced values from 5 to 1 to the Likert levels and use the median as measure of central tendency. From that, the assignment of an acoustic environment to the two-dimensional space of affective qualities can be conducted with the following two formulas for the pleasantness P

$$P = (p - a) + \cos 45 \cdot (ca - ch) + \cos 45 \cdot (v - m) \tag{4.1}$$

and the eventfulness E

$$E = (e - u) + \cos 45 \cdot (ch - ca) + \cos 45 \cdot (v - m) \quad , \tag{4.2}$$

where p, a, ca, ch, v, m, e, u denote the first letter of the eight descriptors and their respective median value. The resulting affective quality ratings are depicted in Figure 4.7 for each of the 19 presented soundscape recording excerpts and the two training stimuli. The circle

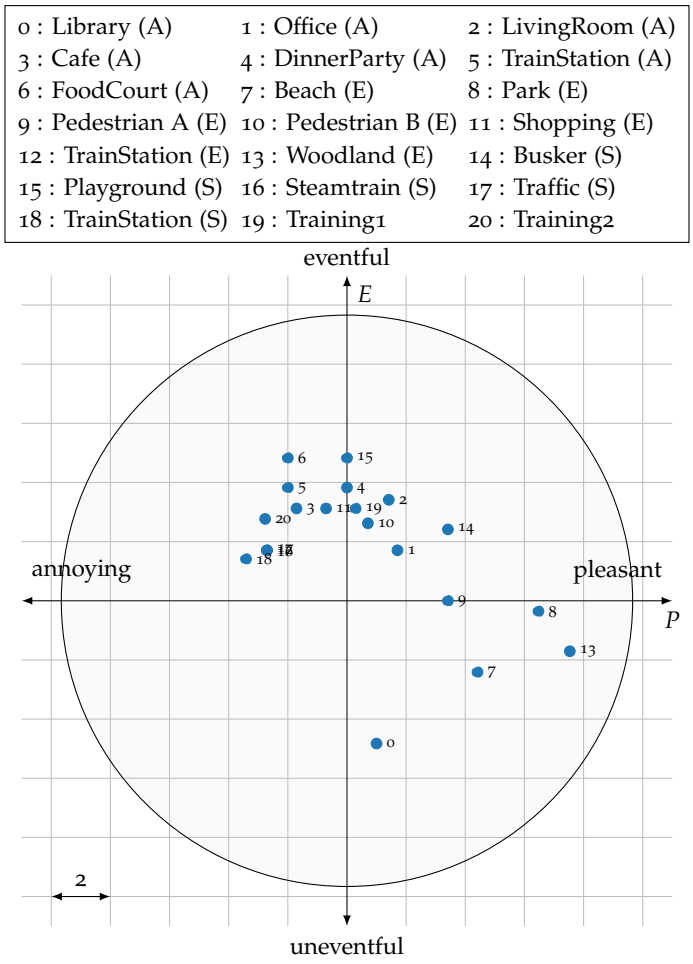
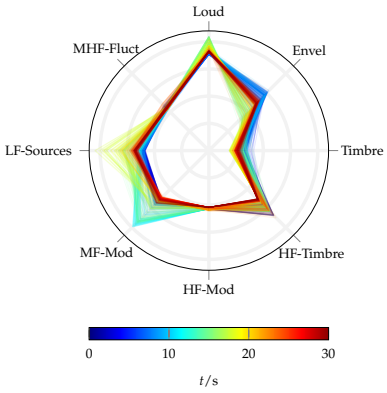


Figure 4.7: Median affective qualities of the 19 stimuli. The shaded circle represents the maximum extent of the dimensions.

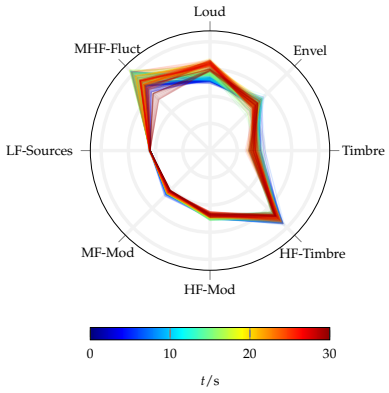
with radius $r = 4 + \sqrt{32} \approx 9.66$ denotes the maximum possible extent of the respective axes. Since the relative locations of the individual soundscape ratings are of interest rather than absolute values, the axis ticks are omitted (however, the grid has a dimension of 2×2). At this point it should be emphasized that this standard-compliant procedure and visualization has its drawbacks since only the median of the ordinal scaled distribution and no information on the dispersion is given. Potential alternatives are discussed in [121].

It can be seen here that three out of four quadrants are occupied and an aggregation of soundscapes exists in the area of moderately eventful and slightly unpleasant. The interpretation of the assignment according to the stimulus name is plausible in most cases. The three stimuli E: TrainStation, S: Steamtrain, and S: Traffic with IDs 12, 16,

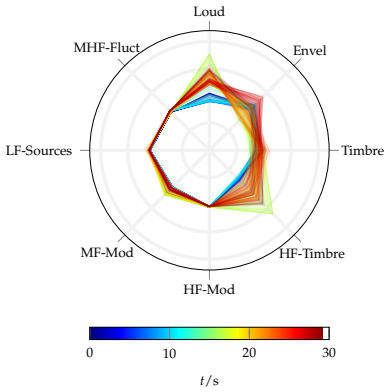
and 17 respectively show the exact same values of $(P|E) = (-2.7|1.7)$ which indicates that their affective qualities are rated equally with respect to their median. All three stimuli are dominated by engine noises. At the same time, the comparison of the acoustic fingerprints representing the time series of the underlying acoustic dimensions in Figure 4.8 shows a very heterogeneous shape (the complete set of fingerprints can be found in Figures B.4 to B.7 of Appendix B).



(a) E: TrainStation



(b) S: Steamtrain



(c) S: Traffic

Figure 4.8: Acoustic fingerprints of three soundscape recordings that show the same median affective qualities.

At this point, the temptation to establish a causal explanation of the perceived affective quality by the acoustic dimensions is resisted. As stated in Section 2.5, attempts to model affective quality using acoustic parameters is an area of research that is attracting much attention. However, most studies show that these efforts have not yet produced satisfactory results due to the lack of physical and personal context. In order to focus on the main aspect of this paper, attempts to establish causal relationships remain as exploratory as presented at this point.

5

VALIDATION II: ECOLOGICAL VALIDITY IN SOUNDSCAPE REPRODUCTION

The development, implementation and validation of fundamental acoustic dimensions allow a range of potential applications. High level applications such as the modeling of perceived sound quality may gain validity because a relevant amount of information on the acoustic properties is provided with this method. Another field of challenging applications is the computer-based identification and recognition of acoustic patterns and events, single sound sources or entire acoustic environments. With a set of acoustic dimensions that cover a significant amount of variance these methods may gain robustness.

In this thesis, two application examples are presented in this and the following chapter that take a step back and serve more as a proof of concept of the proposed method. The examples focus on the comparability of similar acoustic environments and on the identification of diverging acoustic properties. The first example in this chapter investigates the question, if the reproduction of recorded soundscapes by means of Ambisonics rendering in a suitable listening environment is capable to reconstruct the underlying acoustic dimensions that were recorded initially in a real-world acoustic environment.

5.1 METHODOLOGY

The reproduction of acoustic content for assessing any kind of human response is a popular method in various kinds of acoustic and auditory research. Low-level investigations regarding the physiological auditory system as well as more abstract psychoacoustical effects could not be executed without synthetic acoustic reproduction of appropriate stimuli. But also higher-level research that incorporates recordings of real-world sounds such that on sound quality or even emotional response relies in many cases on laboratory studies with reproduced stimuli. Advantages of laboratory studies compared to field studies are in particular the reduced time effort required, the reproducibility or the possibility to present very different acoustic environments quickly one after the other. Thus, within soundscape research, laboratory studies and field studies coexist with various facets, each appropriate for specific hypotheses. Nevertheless, when laboratory studies are conducted, special attention must be paid to the validity of the experimental design. That means that the entire experimental chain, including technical infrastructure, measurement

methods and experiment design, produces results that are consistent with a particular hypothesis. Artifacts introduced to the results by any of the parts of the experiment chain must be known to a certain extent to avoid false conclusions. In experiments where (parts of) real scenarios are reproduced and their results are to be transferred to real scenarios, *ecological validity* must be given. In the case of soundscape research ecological validity can be divided into one part regarding the acoustic environment and another part regarding the non-acoustic context. The aim is that subjects rate a reproduced soundscape similarly compared to an in-situ assessment. An investigation comparing the subjective assessment can be found in [46]. Due to the focus of this work, the validity of the acoustic environment is to be observed by means of the identified acoustic dimensions which was briefly conceptualized in the previous work of the author [59, 115]. The investigation is based on a comparison of two acoustic environments: an original recording (in the following: rec) and a re-recording of the original recording reproduced with an appropriate loudspeaker system (in the following: re-rec). It has to be noted that the original recording does not necessarily represent the pristine acoustic environment without flaw, but is rather itself a (best possible) technical representation of it. The scenarios under test is that of the perceptual study discussed in Chapter 4. 19 excerpts, each of 30 s duration of Ambisonics soundscape recordings were selected, processed, and reproduced with a 30 channel loudspeaker system as described in Section 4.1. This reproduction was then recorded with an Eigenmike® for transformation into the spherical harmonic domain (Ambisonics encoding). The acoustic indicators were calculated as well as the factor scores $\boldsymbol{Y}$ by means of the previously deduced loading matrix $\boldsymbol{L}$ (cf. Section 3.3.2).

5.2 RESULTS

The factor scores of the original recordings and the re-recordings of the respective reproduction were then analyzed. To get an impression, how the time series of the proposed acoustic dimensions actually look like, Figure 5.1 shows an example of the factor scores of four soundscape excerpts, namely A: OFFICE, E: BEACH, E: TRAINSTATION, and S: TRAFFIC. It can be seen that the general slope of the factor scores can be reproduced differently well. In order to quantify this behaviour Spearman's correlation coefficients of ranks were calculated for each dimension and each stimulus which can be taken from Figure 5.2. This shows that for the dimensions LOUD (A), TIMBRE (C), and HF-TIMBRE (D) good correlations can be reached. The dimension ENVEL (B) at least shows positive correlations throughout, while HF-MOD (E), MF-MOD (F), LF-SOURCES (G), and MHF-FLUCT (H) show arbitrary correlation patterns. Thus, it can be stated that the general slope of loudness and timbre can be reproduced. The obvious offset

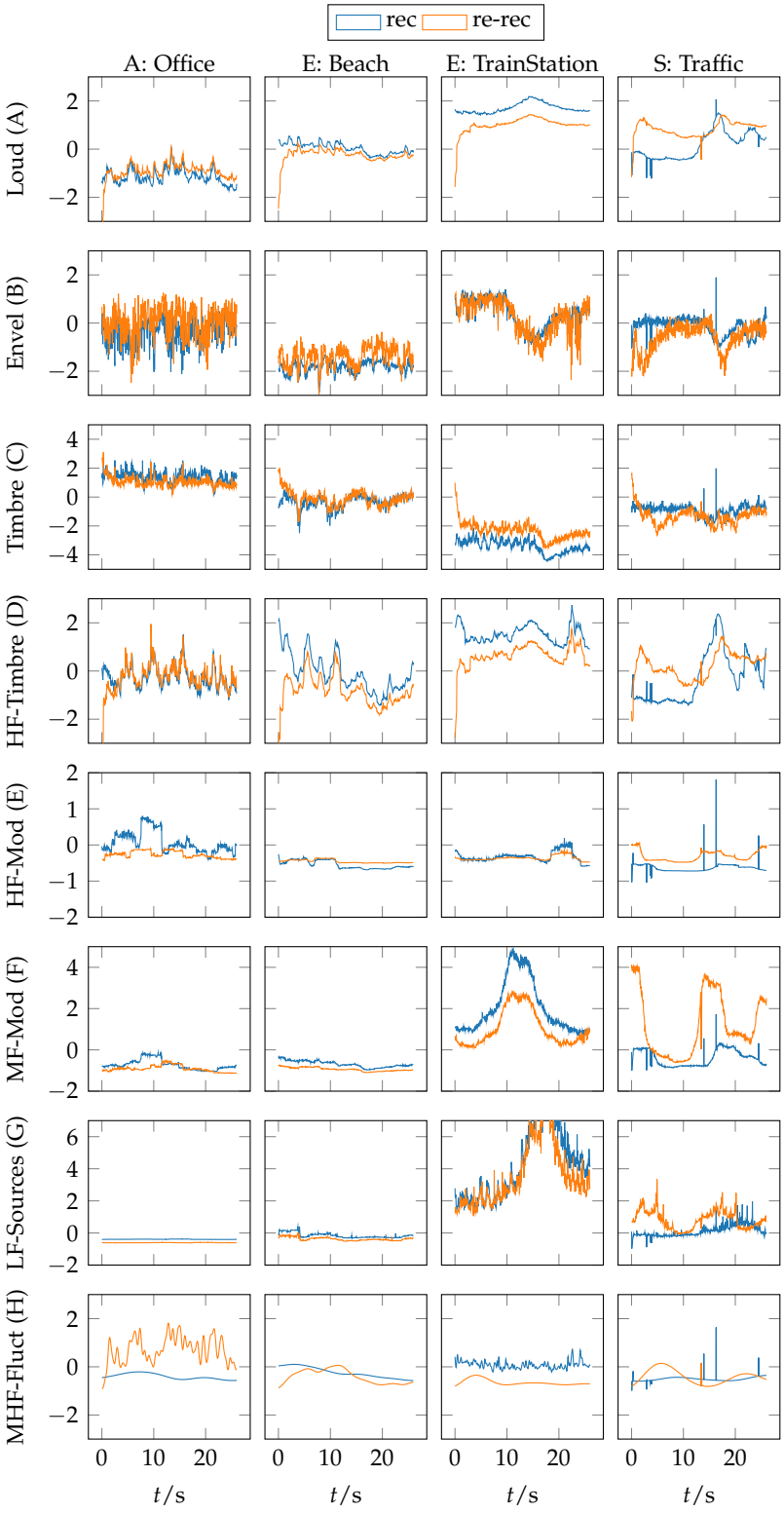


Figure 5.1: Time series of the factor scores of the relevant dimensions for four exemplary acoustic environments comparing recording (blue) and re-recording (orange).

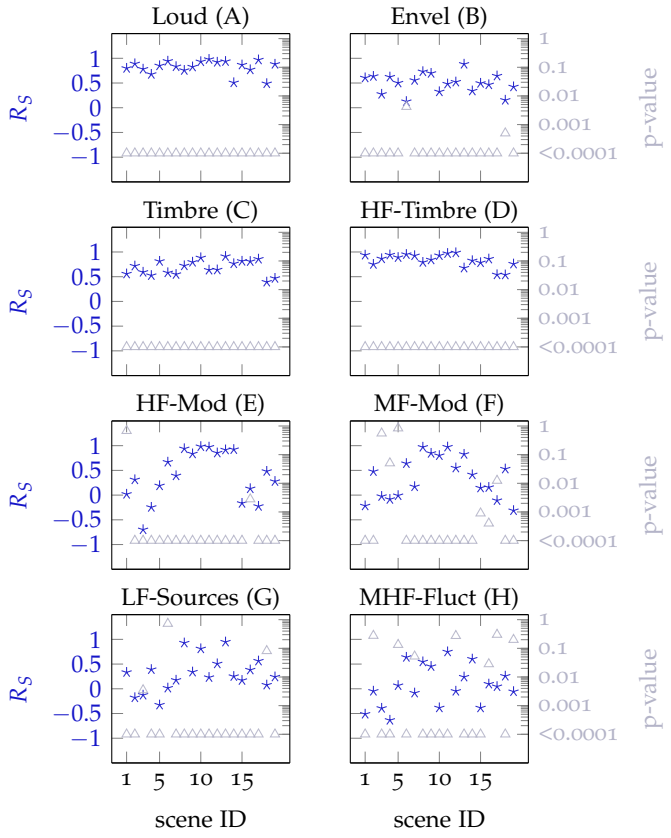


Figure 5.2: Spearman's correlation coefficients R_S between recording and re-recording for each acoustic dimension and corresponding p-values.

from Figure 5.1 is not reflected in Spearman's R_S. However, the correlation pattern of these dimensions indicate an offset that is more or less constant and thus can be treated accordingly. With appropriate manipulations of the audio preprocessing, the reproduction can now be optimized regarding these dimensions. For example, the dimension Loud (A) shows lower values for the re-recording than for the original, which can be targeted with an appropriate reproduction gain factor. Also, the dimension Timbre (C) shows lower values for the re-recordings which might be approached by suitable filtering. Beside the validation step of the methodology, this result also delivers a suitable working basis for the actual application, namely the assessment of the acoustic ecological validity of soundscape reproduction. The question would be if there are systematic differences between recording and re-recording. For that, Figure 5.3 shows the distribution of the factor scores for orginal and re-recorded soundscape representation for all stimuli in detail. It can be seen that most dimensions exhibit a more or less constant offset between the original and re-recorded representation. However, these differences seem mostly not to be significant according to the results of a Friedman test for the null hypothesis H_0:

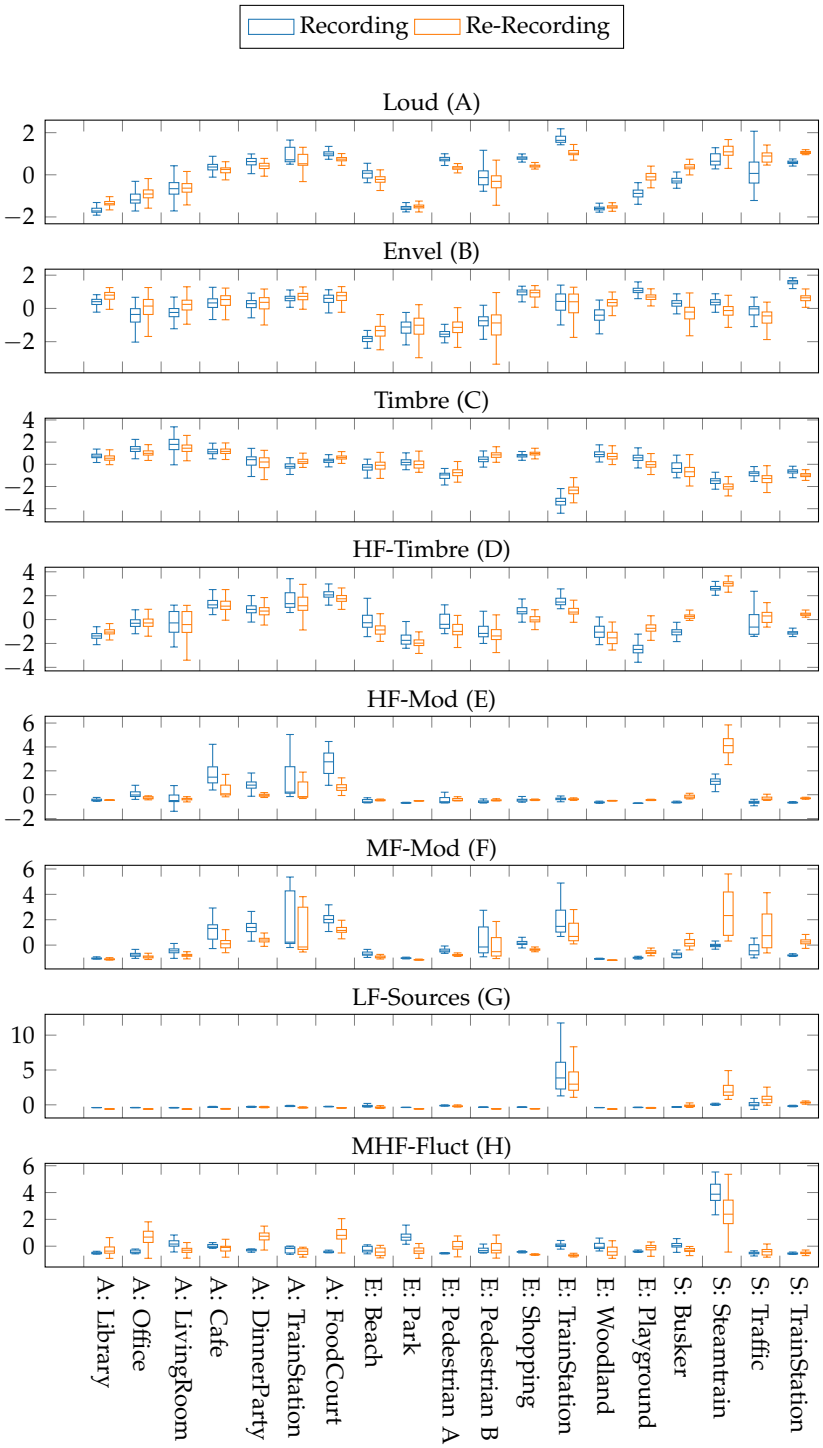


Figure 5.3: Comparison of the distributions of acoustic dimension scores for original recordings and the respective soundfield reproduced in the laboratory.

Table 5.1: Friedman test of significant differences between recordings and re-recordings among the acoustic dimensions with p values and resulting significance as well as the test statistic Kendall's W. Conover p_{con} values denote pairwise posthoc comparison test were applicable.

Dimension	W	p	sig	p_{con}
LOUD (A)	0.002770	0.818546	n/s	N/A
ENVEL (B)	0.024931	0.491297	n/s	N/A
TIMBRE (C)	0.024931	0.491297	n/s	N/A
HF-TIMBRE (D)	0.135734	0.108294	n/s	N/A
HF-MOD (E)	0.069252	0.251349	n/s	N/A
MF-MOD (F)	0.224377	0.038947	*	<0.000001
LF-SOURCES (G)	0.335180	0.011617	*	<0.000001
MHF-FLUCT (H)	0.002770	0.818546	n/s	N/A

"There is no difference in scores of a specific dimension between recording and re-recording." as shown in Table 5.1. Only the dimensions MF-MOD (F) and LF-SOURCES (G) show significant differences here with $p < 0.05$. This was also confirmed with pairwise posthoc Conover tests. However, these statements must be taken with care, since Friedman test on ranks might not discover differences that are not systematic as observable e.g. for the dimension MHF-FLUCT (H) in Figure 5.1, p. 59. By observing the time series it can be said that the first four dimensions LOUD (A), ENVEL (B), TIMBRE (C), and HF-TIMBRE (D) show comprehensible evidence if and how a reproduced sound field represents a original reference. The other dimensions HF-MOD (E), MF-MOD (F), LF-SOURCES (G), and MHF-FLUCT (H) in turn show ambiguous results. The analysis in terms of statistical differences does not necessarily lead to satisfying answers if a reproduction is suitable. It can thus be stated that the identification of statistically relevant acoustic dimensions always requires manual plausibility checks by the investigator, of which this chapter was an example.

6

VALIDATION III: MUSIC REPRODUCTION IN STEREO, SURROUND AND 3D

In this chapter another exemplary application of the proposed methodology of investigating underlying acoustic dimensions is presented that in turn contributes to the validation of the approach. Again, acoustic environments are compared that are expected to have similarities and differences in dimensions between different conditions. This validation step here addresses the question of whether the previously identified dimensions are applicable to all kinds of soundscape recordings or if hypotheses on specific acoustic environments require individual treatment to detect all peculiarities. Parts of the content of this chapter were published by the author in [122] and [123].

This application example is situated in the field of audio engineering and music perception and refers to the development of loudspeaker reproduction systems. With the promise of enhanced spatial imaging, listener envelopment and overall improved listening experience, the number of loudspeakers of commercially available channel-based audio systems and formats increased over the past decades. Starting from stereo reproduction allowing localization of individual music instruments between left and right to quadraphonic sound with four loudspeakers to the commercially very successful 5.1 surround sound with five loudspeakers and an additional subwoofer up to 7.1 surround sound with seven loudspeakers plus subwoofer, the number of loudspeakers increases analogously to the industry's promises of spatial imaging within the listening plane. With the incorporation of additional elevated loudspeakers, marketing terms such as 3D audio or immersive audio become more and more widespread. This development comprises both audio rendering algorithms as well as loudspeaker setups, such as 5.1.4 surround sound which adds four elevated speakers to a standard 5.1 setup (cf. [124, Setup D]) or 22.2 surround sound. The additional height layer of loudspeakers is promised to further increase listener envelopment and spatial plausibility. There exists a mentionable body on research of perceptual effects provoked by these technologies, e.g. in [125, 126, 127]. However, at the same time little information is available on what properties of the reproduced sound field actually change with different reproduction technologies and if the perceptual effects can be explained or modeled with acoustic terms. The methodology of underlying acoustic dimensions is applied here aiming for detection of similarities and differences within the reproduction of music in four exemplary loudspeaker layouts, namely mono, stereo, 5.1 surround sound and

5.1.4 surround sound. Each of these layouts adds spatial direction to the speaker positioning, which is why this selection of formats can serve as an example for other and/or more advanced speaker configurations.

6.1 METHODOLOGY

The stimulus set of this investigation comprises eight excerpts of musical pieces of varying genre, ensemble size and recording/production technique. Each piece of music is available in four versions of different channel-based loudspeaker reproduction formats: *mono* (center loudspeaker), *stereo* (left + right lsp.), *2D* (5.1 surround sound) and *3D* (5.1.4 surround sound). For the production of the stimuli, two audio engineers with experience in multi-channel mixing were engaged to produce three well-sounding mixes (stereo, 2D and 3D) from provided multi-track recordings without any other restrictions. The respective mono version was deduced from the stereo version by averaging left and right channel. An overview of the stimuli can be found in Table 6.1. The loudness of the stimuli within the four playback formats was calibrated to minimize the median deviation of the short-term LUFS (EBU R 128 [82]) time series from the stereo reference. Since this LUFS calibration is conducted in the digital signal domain, the procedure was validated by means of the acoustic loudness measures. Monophonic sound pressure levels L_{Aeq} and loudness according to ISO 532-1 [71] were measured with a Beyerdynamic MM1 microphone at the center of the listening area. Binaural L_{Aeq} and loudness according to ISO 532-2 [81] respectively was captured with a G.R.A.S 45BC-12 KEMAR. The comparison of loudness and level distributions between the formats revealed minor differences in dependence of the respective musical piece, however no systematic and unexpected differences could be found. Another stage of validation of the calibration was performed perceptually by an experienced audio engineer. In order to use the calibrated stimulus set for future listening tests, the overall loudness between the individual pieces of music was adjusted subjectively by the same audio engineer, aiming for plausibility in the reproduction of music with different ensemble sizes and genres.

The reproduction was conducted with the technical infrastructure previously described in Section 4.1. This time nine loudspeakers plus two subwoofers were used that were positioned in accordance with ITU-R BS.2051-2, setup D [124]. The raw stimuli were then re-recorded at the listening point by means of an Eigenmike® EM32 and appropriate Ambisonics processing in the same way as described in Section 5.1.

Table 6.1: Overview of investigated musical pieces.

Label	Piece	Dur. [s]	Genre and orchestration	Production
Laudate	Laudate Dominum (Josep Vila)	33.4	A-cappella choir: 12 singers (SATB)	3D microphone setup + support microphones
Mellow	In a Mellow Tone (Janna Berger)	35.4	Jazz band: ds, db, pf, fem. voice	3D microphone setup + support microphones
Wunderschoen	Im wunderschönen Monat Mai (Robert Schumann)	38.6	Classic song: male voice, pf	3D microphone setup + support microphones
School	School's Out (live; Alice Cooper)	57.5	Full live rock band	single microphones + 3D ambience
Bilder	Pictures of an exhibition (Mussorgsky)	37.3	Large Orchestra	3D microphone setup + support microphones
Walkuere	Ride of the Valkyries (Wagner)	62.5	Opera: Orchestra, fem. voices	manual upmix from commercial 5.1 content
Hantel	Die Hantel (Zweitaktmotor)	61.8	Electropop: synthesizers, male and fem. voices	pure studio production
Rokoko	Rokoko Variations (Tchaikovsky)	68.1	Classic chamber music: cello, woodwind quintet	manual upmix from commercial 5.1 content

Generic and specific acoustic dimensions

As introduced above, the methodology of this validation example targets the question, whether the generic acoustic dimensions developed before are suitable to detect differences and similarities in a sample population that stems from a very specific subarea of acoustic environments, namely the reproduction of music with different loudspeaker systems. In order to do so, the statistical analysis on differences is conducted on two datasets: the factor scores $\mathbf{Y}$ calculated with the generic loading matrix developed in Section 3.3.2 and the factor scores $\mathbf{Y}^*$ that are deduced by means of a loading matrix $\mathbf{L}^*$ that was developed on basis of the specific acoustic environments under test only. This specific loading matrix was calculated with the same assumptions of factor analysis as described in Section 3.3.2. Figure 6.1 shows the schematic diagram of the specific loading matrix that can be compared with the generic one in Figure 3.3 on page 27.

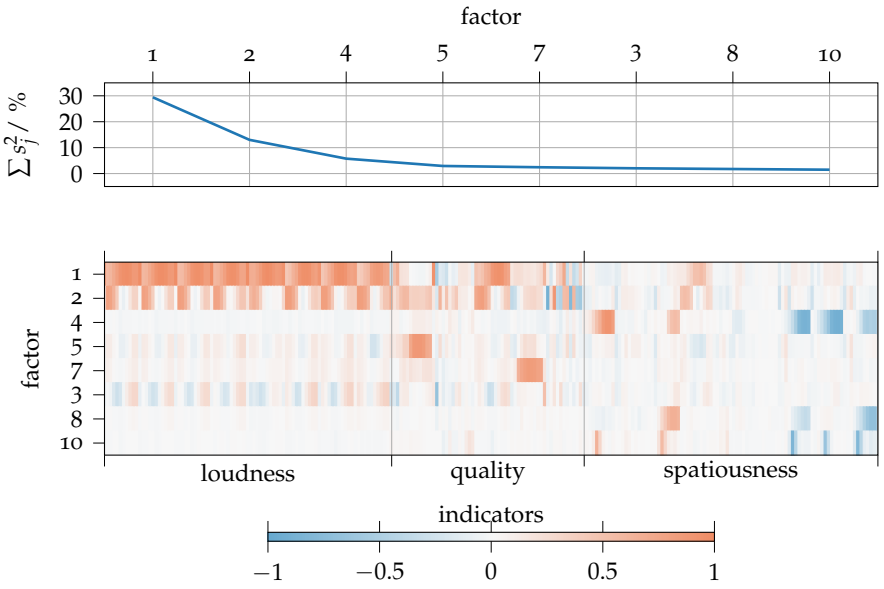


Figure 6.1: Scree test of explained variance (top) and schematic distribution of indicator loadings for the first eight relevant specific factors (bottom).

The respective factor composition is listed in Table 6.2 which in turn may be compared to the generic composition respectively in Table 3.7, p. 28.

Table 6.2: Indicator composition of the first eight relevant rotated factors j of the stimulus specific factor analysis.

Factor j	s_j^2	$N_{i,j}$	**Indicators**
1	69.74 (29.4%)	40	LAeq(0.116), LA(0.116), loudnessZwickerBands05(0.116), LApeak(0.116), LAeqBands06(0.115), loudnessZwicker(0.115), LABands06(0.115), LAeqBands05(0.115), LAmax(0.115), LApeakBands05(0.114), loudnessZwickerBands06(0.114), lufsMomBands06(0.114), LABands05(0.114), LAmaxBands06(0.113), lufsPeakBands05(0.113), lufsPeakBands06(0.113), LAeqBands07(0.113), LApeakBands06(0.113), LABands07(0.113), LAmaxBands05(0.113), lufsMomBands05(0.113), loudnessZwickerBands04(0.112), oct07(0.112), mfcc00(0.112), lufsMomBands07(0.111), oct06(0.111), LApeakBands04(0.111), LAmaxBands07(0.110), oct08(0.110), LAeqBands04(0.110), lufsMom(0.110), LABands04(0.109), lufsPeakBands04(0.109), lufsPeakBands07(0.108), lufsPeak(0.108), LAmaxBands04(0.107), lufsMomBands04(0.107), loudnessZwickerBands07(0.107), LApeakBands07(0.105), oct05(0.104)

Continued on next page

Table 6.2 – continued from previous page

Factor j	s_j^2	$N_{i,j}$	**Indicators**
2	30.81 (13.0%)	26	spectralCentroid(-0.163), spectralDecrease(0.161), lufsMomBands00(0.150), lufsShortBands00(0.149), lufsMomBands01(0.147), lufsPeakBands01(0.147), oct02(0.147), lufsShortBands01(0.146), LAmaxBands00(0.146), lufsPeakBands00(0.145), LABands00(0.143), spectralRolloffPoint(-0.143), oct01(0.142), LAeqBands00(0.142), LAmaxBands01(0.142), LABands01(0.139), LApeakBands00(0.139), LAeqBands01(0.137), LApeakBands01(0.136), lufsShortBands02(0.135), oct00(0.130), lufsMomBands02(0.129), oct03(0.124), lufsPeakBands02(0.123), booming0(0.122), loudnessZwickerBands01(0.118)
4	13.63 (5.8%)	9	sphDIAz08(-0.250), sphDIAz07(-0.248), diff08(0.246), diff07(0.245), sphDIAz09(-0.243), sphDI08(-0.239), sphDI07(-0.235), sphDI09(-0.234), sphDIAz06(-0.232)
5	6.93 (2.9%)	5	fluct06(0.330), fluct05(0.329), fluct04(0.326), fluct07(0.319), fluct03(0.277)
7	5.76 (2.4%)	5	rough05(0.349), rough06(0.347), rough07(0.341), rough04(0.335), rough08(0.319)
3	4.80 (2.0%)	21	mfcc01(-0.306), sharp(0.231), booming2(-0.172), lufsShortBands09(0.153), lufsShortBands08(0.151), lufsMomBands09(0.147), lufsPeakBands09(0.146), oct04(-0.143), lufsPeakBands08(0.141), lufsMomBands08(0.140), LABands08(0.131), LABands09(0.131), lufsPeakBands03(-0.130), LABands03(-0.130), oct05(-0.128), loudnessZwickerBands03(-0.127), booming0(-0.127), LAeqBands03(-0.125), lufsMomBands03(-0.124), LApeakBands08(0.120), LAmaxBands08(0.120)
8	4.15 (1.7%)	5	sphDIEl08(-0.344), sphDIEl09(-0.341), doaEl07(0.320), doaEl08(0.319), sphDIEl07(-0.310)
10	3.52 (1.5%)	3	sphDI04(-0.449), sphDIEl04(-0.431), sphDIAz04(-0.351)

The interpretation with appropriate semantic descriptors of the specific acoustic dimensions can be taken from Table 6.3 that is accompanied by the descriptors of the generic acoustic dimensions for comparability. Both sets of acoustic dimensions show similarities, such as that the dimension Loud (AA) is present dominantly with almost equal amount of explained variance portion. It can also be noted that the timbre is present in both sets with two dimensions, LF-Timbre (BB) and HF-Timbre (FF) for the specific set and Timbre (C) and HF-Timbre (D) for the generic set respectively. The fact that the low-frequency timbre in the specific set explains 13 % of variance indicates that the reproduction of music obviously has more variability in this dimension which may be explained by the use of subwoofers and LFE channel to additionally enhance this frequency range where suitable in an artistic approach. The envelopment of sound is represented in the generic set of dimensions with a single

Table 6.3: Semantic descriptors for generic and specific acoustic dimensions.

(a) specific

Dim.	Fac.	Expl. Var.	Descriptor	Label
AA	1	69.74 (29.4 %)	"Loudness"	Loud (AA)
BB	2	30.81 (13.0 %)	"Low-Frequency Timbre"	LF-Timbre (BB)
CC	4	13.63 (5.8 %)	"High-Frequency Diffusivity"	HF-Diff (CC)
DD	5	6.93 (2.9 %)	"Temporal Fluctuation"	Fluct (DD)
EE	7	5.76 (2.4 %)	"Roughness"	Rough (EE)
FF	3	4.80 (2.0 %)	"High-Frequency Timbre"	HF-Timbre (FF)
GG	8	4.15 (1.7 %)	"Elevational Diffusivity"	El-Diff (GG)
HH	10	3.52 (1.5 %)	"Mid-Frequency Diffusivity"	MF-Diff (HH)

(b) generic

Dim.	Fac.	Expl. Var.	Descriptor	Label
A	1	87.77 (28.9 %)	"Loudness"	Loud (A)
B	4	16.77 (5.5 %)	"Sound Source Envelopment"	Envel (B)
C	2	12.92 (4.3 %)	"Timbre"	Timbre (C)
D	15	11.23 (3.7 %)	"High-Frequency Timbre"	HF-Timbre (D)
E	5	8.67 (2.9 %)	"High-Frequency Modulation"	HF-Mod (E)
F	6	7.59 (2.5 %)	"Mid-Frequency Modulation"	MF-Mod (F)
G	11	5.96 (2.0 %)	"Low-Frequency Sound Sources"	LF-Sources (G)
H	8	5.29 (1.7 %)	"Mid-High-Frequency Fluctuation"	MHF-Fluct (H)

dimension Envel (B) whereas the specific set offers three distinguishable dimensions of diffusivity, namely HF-Diff (CC), El-Diff (GG), and MF-Diff (HH), which also reflects the specific character of the recordings with different spatially positioned loudspeaker setups. The temporal characteristic of the recorded acoustic environments is also present in both sets, namely as Fluct (DD) and Rough (EE) in the specific set and as HF-Mod (E), MF-Mod (F), and MHF-Fluct (H) in the generic set.

6.2 RESULTS

As mentioned above, the results will be taken to discuss the question whether the comparison of such specific acoustic environments is equally successful with help of the generic loading matrix, developed in Section 3.3.2 or if a specifically deduced loading matrix leads to results more useful for discrimination.

Generic Loading Matrix

Figure 6.2 shows the distribution of the factor scores for each stimuli, acoustic dimension, and loudspeaker setup. A first view shows, that certain factors exhibit small differences between loudspeaker

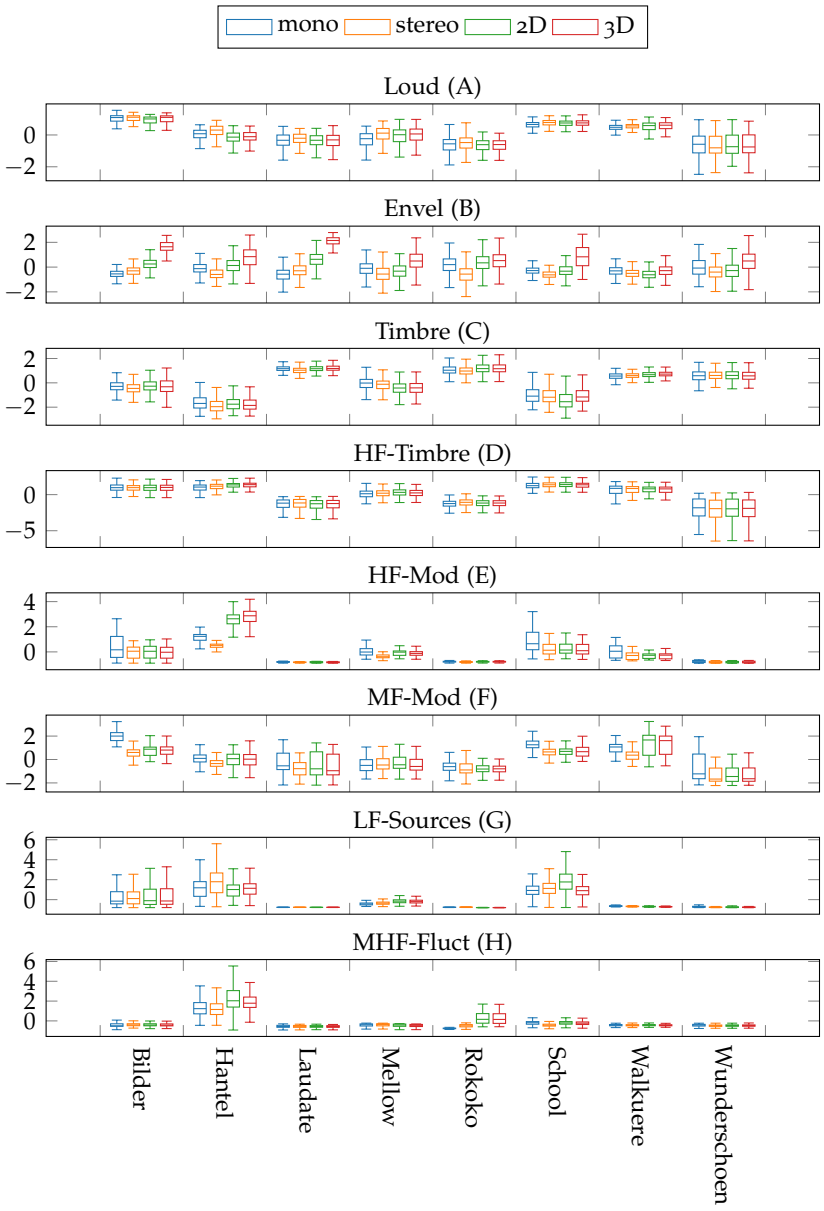


Figure 6.2: Distributions of stimulus factor scores of generic acoustic dimensions. Significant differences between loudspeaker setups can be found in ENVEL (B), HF-MOD (E), and MF-MOD (F).

conditions but larger differences between stimuli, such as LOUD (A), TIMBRE (C), and HF-TIMBRE (D) . Other dimensions show distinct differences between conditions, with distributions appearing to be systematic (e.g. ENVEL (B) increases with increasing number of loudspeakers) or random (e.g. HF-MOD (E) shows differences within the piece Hantel). To quantify the differences, Friedman tests were performed as described in Section 3.4.1 with the null hypothesis H_0: *"There is no difference in scores of a specific factor between mono, stereo, 2D and 3D loudspeaker setups."*. The results of this test are shown in Table 6.4.

Table 6.4: Friedman test of significant differences between loudspeaker setups among the acoustic dimensions with p values and resulting significance as well as the test statistic Kendall's W.

Dimension	W	p	sig
LOUD (A)	0.262500	0.094873	n/s
ENVEL (B)	0.731250	0.000008	**
TIMBRE (C)	0.137500	0.363321	n/s
HF-TIMBRE (D)	0.093750	0.538734	n/s
HF-MOD (E)	0.518750	0.001908	**
MF-MOD (F)	0.500000	0.002724	**
LF-SOURCES (G)	0.081250	0.597847	n/s
MHF-FLUCT (H)	0.293750	0.064542	n/s

It can be seen that the dimension ENVEL (B) shows significant differences between loudspeaker setups which is comprehensible since the number and spatial location of the loudspeaker change leading to a different composition of the spatial sound field. A closer look at the distribution of this dimension in Figure 6.2 shows the tendency that as the number of speakers in a playback setup increases, the scores of this respective dimension increase as well. The Friedman test also shows significant differences between the loudspeaker setups for the dimensions HF-MOD (E) and MF-MOD (F). However, a systematic behavior cannot be asserted from the distribution but rather random deviations can be observed.

Pairwise one-sided Wilcoxon signed-rank tests were performed to determine whether the differences among those three dimensions were systemic or due to specific characteristics. Thus, the comparisons between all loudspeaker setups for each piece of music was conducted one-sided, that is in the way "scores for setup A are greater than for setup B". Table 6.5 shows the resulting p^*-values with Bonferroni correction $p^* = p \cdot N_{comp} = p \cdot (6 \cdot 8)$. It may be read exemplarily as

Example. *According to Table 6.5a the scores within dimension ENVEL (B) for the piece Bilder is significantly greater for the 3D loudspeaker setup than for the 2D setup (**, $p < 0.001$; first row, first column). At the same time, the scores for the piece Wunderschoen are not greater for the stereo setup compared to the mono setup (n/s, $p = 1$; last row, last column)*

Table 6.5: Pairwise one-sided posthoc Wilcoxon test of significant differences between loudspeaker setups of generic acoustic dimensions. Bonferroni adjusted p^* values with resulting significance.

(a) ENVEL (B)

Piece	3D > 2D	3D > Stereo	3D > Mono	2D > Stereo	2D > Mono	Stereo > Mono
Bilder	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Hantel	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Laudate	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Mellow	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Rokoko	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
School	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Walkuere	**	**	n/s	n/s	n/s	n/s
	(<0.001)	**(<0.001)**	(1.000)	(1.000)	(1.000)	(1.000)
Wunderschoen	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(0.001)**	(1.000)	(1.000)

(b) HF-MOD (E)

Bilder	n/s	**	n/s	**	n/s	n/s
	(1.000)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Hantel	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Laudate	n/s	n/s	n/s	n/s	n/s	n/s
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Mellow	n/s	**	n/s	**	n/s	n/s
	(1.000)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Rokoko	**	**	n/s	**	n/s	n/s
	(<0.001)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
School	n/s	n/s	n/s	**	n/s	n/s
	(1.000)	(1.000)	(1.000)	**(<0.001)**	(1.000)	(1.000)
Walkuere	n/s	n/s	n/s	n/s	n/s	n/s
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Wunderschoen	**	**	n/s	n/s	n/s	n/s
	(<0.001)	**(<0.001)**	(1.000)	(1.000)	(1.000)	(1.000)

(c) MF-MOD (F)

Bilder	n/s	**	n/s	**	n/s	n/s
	(1.000)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Hantel	n/s	**	n/s	**	n/s	n/s
	(1.000)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Laudate	n/s	**	n/s	**	n/s	n/s
	(1.000)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Mellow	n/s	n/s	n/s	n/s	**	**
	(1.000)	(1.000)	(1.000)	(1.000)	**(<0.001)**	**(0.002)**
Rokoko	n/s	n/s	n/s	n/s	n/s	n/s
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
School	n/s	**	n/s	**	n/s	n/s
	(0.966)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
Walkuere	n/s	**	**	**	**	n/s
	(1.000)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Wunderschoen	**	n/s	n/s	n/s	n/s	n/s
	(<0.001)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

Most pairwise comparisons within the dimension ENVEL (B) show larger factor scores for the condition with larger number of loudspeakers. This is the case for all comparisons, except the comparison between stereo and mono and except for the piece Walkuere. We can thus assume that this characteristic is systematic. A different behavior can be observed for the dimensions HF-MOD (E) and MF-MOD (F). Here, the statistical difference from the Friedman test can not be operationalized since the pairwise comparisons in Tables 6.5b and 6.5c show no systematic characteristics neither among the comparison conditions nor among the pieces of music. The significant differences here are thus based on other or random reasons.
We can conclude at this point that the generic acoustic dimensions are suitable to detect the very specific differences in a sound field that arise from the use of different loudspeaker setups with the dimension SOUND SOURCE ENVELOPMENT (B). An increase of loudspeakers from mono to stereo, to 2D, up to the 3D setup leads to an increase of the scores of this particular dimensions. When assuming that an increase of spatially distributed sound sources (loudspeakers) amplifies a surrounding or diffuse sound field, this findign confirms the character of ENVEL (B).

Specific Loading Matrix

After gaining satisfactory results with a generic set of acoustic dimensions previously, the following investigation aims towards a more detailed characterization of the specific acoustic environments of music reproduction with different loudspeaker setups. For that, the specific acoustic dimensions developed in Section 6.1 are analyzed in a similar way as the generic dimensions before. Thus, Figure 6.3 shows the distribution of the factor scores. The exploration of the distribution indicates that again certain dimensions show differences between musical pieces and others show differences between loudspeaker setups. In order to quantify latter, a Friedman test with the same null hypothesis H_0: *"There is no difference in scores of a specific factor between mono, stereo, 2D and 3D loudspeaker setups."* was conducted. The dimensions HIGH-FREQUENCY DIFFUSIVITY (CC), ELEVATIONAL DIFFUSIVITY (GG), and MID-FREQUENCY DIFFUSIVITY (HH) show p values that dictate a rejection of H_0 leading to significant differences between the loudspeaker versions. The respective p-values and Kendall's W can be found in Table 6.6. Comparable to the results of the generic set of acoustic dimensions here again the dimensions that describe the spatial composition of the sound field are found to be different which again satisfies reasonable expectations. However, here the spatial characteristics are made up by three dimensions which emphasizes that these characteristics make up a relevant portion of overall variance of the sound field. To ensure that the significant differences are actually

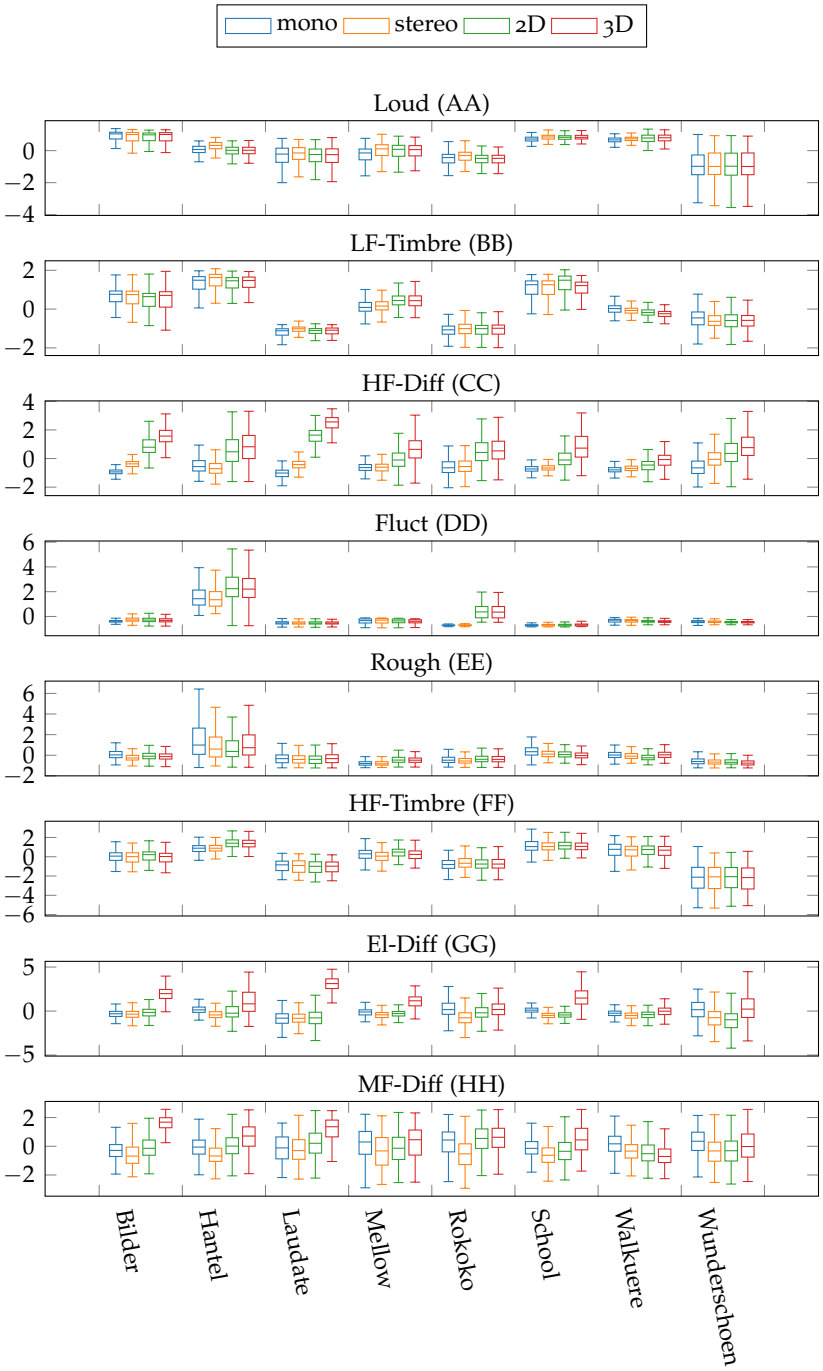


Figure 6.3: Distributions of stimulus factor scores of generic acoustic dimensions. Significant differences between loudspeaker setups can be found in HF-Diff (CC), El-Diff (GG), and MF-Diff (HH).

Table 6.6: Friedman test of significant differences of specific acoustic dimensions between loudspeaker setups.

Dimension	W	p	sig
Loud (AA)	0.281250	0.075483	n/s
LF-Timbre (BB)	0.125000	0.408384	n/s
HF-Diff (CC)	0.925000	0.000000	**
Fluct (DD)	0.106250	0.483549	n/s
Rough (EE)	0.243750	0.118377	n/s
HF-Timbre (FF)	0.231250	0.136654	n/s
El-Diff (GG)	0.731250	0.000008	**
MF-Diff (HH)	0.525000	0.001689	**

due to the different loudspeaker setups paired one-sided Wilcoxon signed-rank tests with the alternative hypothesis H_1: *"Within a specific dimension the scores of loudspeaker setup A is greater than those of setup B."* were conducted. Table 6.7 shows the Bonferroni-adjusted values p^* for all pairwise comparisons.

It can be seen that from all three investigated dimensions a systematic characteristic can be deduced. The dimension HF-Diff (CC) shows significant higher scores for the respective setup with more loudspeakers for almost all comparisons. This meet the expectation that an increase of sound sources (loudspeakers) leads to higher diffusivity in the investigated sound field.

Non-significance can be observed for dimension El-Diff (GG) for the comparison between stereo and mono as well as between 2D and mono setup. This is a plausible outcome, since all speakers are positioned at the same ear level and no differences of elevation are expected. However, this would also apply for the comparison between 2D and stereo, which in turn shows significant differences for six out of eight pieces of music. This behavior is not expected and an explanation can not be delivered at this point. The comparisons of the 3D setup with all other setups without elevated loudspeakers again meets the expectations of higher scores in this dimensions.

The MF-Diff (HH) again shows significantly higher scores for setups with more loudspeakers in 29 out of 48 pairwise comparisons which is an ambiguous outcome. The fact that the piece Walkuere exhibits no difference between the setups for any comparison also indicates a non-systematic but rather content-dependent cause for this dimension.

Summarized we can state that the application of a set of acoustic dimensions that was developed specifically for a certain application turned out to be beneficial in terms of more detailed discrimination. Here, the dimension HF-Diff (CC) could robustly detect sound fields composed of a larger number of sound sources (loudspeakers). This was accompanied and refined by the dimension El-Diff (GG) that was capable of detecting setups with elevated loudspeakers with minor deductions.

Table 6.7: Pairwise one-sided posthoc Wilcoxon test of significant differences between loudspeaker setups of specific acoustic dimensions. Bonferroni adjusted p^* values with resulting significance.

(a) HF-Diff (CC)

Piece	3D > 2D	3D > Stereo	3D > Mono	2D > Stereo	2D > Mono	Stereo > Mono
Bilder	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Hantel	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Laudate	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Mellow	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Rokoko	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
School	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Walkuere	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
Wunderschoen	**	**	**	**	**	**
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**

(b) El-Diff (GG)

Bilder	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Hantel	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Laudate	**	**	**	n/s	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)	(1.000)
Mellow	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Rokoko	**	**	n/s	**	n/s	n/s
	(<0.001)	**(<0.001)**	(1.000)	**(<0.001)**	(1.000)	(1.000)
School	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Walkuere	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Wunderschoen	**	**	n/s	n/s	n/s	n/s
	(<0.001)	**(<0.001)**	(0.083)	(1.000)	(1.000)	(1.000)

(c) MF-Diff (HH)

Bilder	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Hantel	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Laudate	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
Mellow	**	**	n/s	**	n/s	n/s
	(<0.001)	**(<0.001)**	(0.109)	**(<0.001)**	(1.000)	(1.000)
Rokoko	**	**	**	**	**	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)
School	**	**	**	**	n/s	n/s
	(<0.001)	**(<0.001)**	**(<0.001)**	**(<0.001)**	(1.000)	(1.000)
Walkuere	n/s	n/s	n/s	n/s	n/s	n/s
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Wunderschoen	**	**	n/s	n/s	n/s	n/s
	(<0.001)	**(<0.001)**	(1.000)	(0.305)	(1.000)	(1.000)

6.3 TIME SERIES CONSIDERATIONS

The presented methodology is based on similarities and differences of the distribution of factor scores. In principle, this would include the assumption that each short-term observation is independent of any other observation. This of course is not the case since time series are investigated that are tied to a process with both stochastic and deterministic features. Hence, in order to assure that the above made statements are valid not only for the distributions but also for the time series, a further analysis step was conducted. With the help of *independent component analysis* (ICA) [128] as described in Section 3.3.3 it is possible to detect underlying signal bases. The method assumes that observed signals are mixtures of superimposed basis signals. The decomposition of the four signal observations (mono, stereo, 2D, 3D) for each dimension and each piece of music into subcomponents is ought to reveal similarities in the temporal characteristic. Figure 6.4 shows the mixing matrix of the ICA with four basis signal components $s0$, $s1$, $s2$, $s3$ of the piece Laudate, i.e. the respective weights. It can be seen that for the dimensions Loud (AA), LF-Timbre (BB), Rough (EE), and HF-Timbre (FF) a single component is mixed with large weights to the time series of the dimension scores of all loudspeaker setups. This vertical structure means that all four conditions are based on similar time series properties. The dimensions HF-Diff (CC), El-Diff (GG), and MF-Diff (HH) have different characteristics. Here, we cannot identify such structures, which means that the dimensions' time series of the four loudspeaker conditions do differ in a relevant way. These both findings of similarities and differences confirm the assumptions that not only the distributions but also the time series of the identified dimension scores discriminate the four loudspeaker conditions within the dimensions HF-Diff (CC), El-Diff (GG), and MF-Diff (HH).

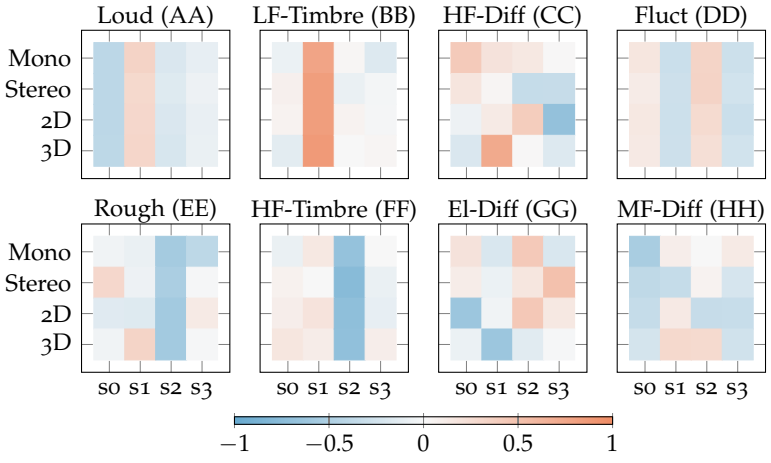


Figure 6.4: ICA mixing matrix for the piece Laudate. Composition of fundamental signal shapes $s0$, $s1$, $s2$, and $s3$ to represent the slopes of the factor scores time series of the respective loudspeaker setups.

7 DISCUSSION

The previous chapters produced decisions, assumptions, results, and consequences that will be summarized and discussed in the following sections. Special focus will be given to the limitations, ambiguities and open questions in order to point out future improvements of the presented methodology.

7.1 SUMMARY

The presented work raised the question of what properties are actually important for distinguishing and comparing acoustic environments from a human perspective.

A satisfying answer to this question would help in various fields of application, e.g. the modeling of the influence of acoustic properties on the perceptual and emotional responses of individuals. Other more technical applications could be general comparisons and classifications of classes of acoustic environments in algorithmic terms or the regulatory assessment of noise immission as elaborated in Chapter 1.

The current state of soundscape research – as standardized in ISO 12913-1 [4] and subsequent technical specifications – seems to have settled on emotional dimensions, namely pleasantness and eventfulness, with which reactions to soundscapes can be broadly represented. To describe a soundscape holistically, a triangulation between the entities acoustic environment, person and context is recommended, as described in Chapter 2. For this reason, and because it can be assumed that there are causal relationships between the entities, there is a need for a comprehensive description of the acoustic properties of soundscapes. The fundamental concepts for describing acoustic environments as part of soundscapes was discussed in the same chapter, including the physical assessment of spatio-temporal sound fields by means of spherical harmonic decomposition, appropriate methods for capturing, representation, and reproduction such as Ambisonics or binaural rendering, the boundary conditions of the acoustic environment representations in terms of acoustic and non-acoustic context as well as general requirements of a representation of the properties of acoustic environments.

From that, the aim of the presented thesis was formulated, namely the identification of underlying acoustic dimensions on the basis of signal-based acoustic indicators that can be used to describe the key properties of acoustic environments in a straightforward way. Chapter 3 then elaborates the methodology to identify these dimensions.

The principal approach is data-based and exploratory, that means that the underlying acoustic dimensions are revealed by observation of a broad range of soundscape recordings. For the observations, a collection of signal-based indicators within the categories quality, loudness, spaciousness and time was selected that was put together on basis of a broad literature review of soundscape studies and neighboring fields such as music information retrieval, sound field analysis and psychoacoustics. A methodological framework of indicators as input and acoustic dimensions as output of factor analysis, a multivariate statistical method, was developed and applied on a dataset containing three different databases of Ambisonics soundscape recordings of approx. 12.5 h in total. The result consists of eight relevant acoustic dimensions that explain 51.4 % of the total variance and whose semantic descriptor is made up by their respective indicator composition, namely Loudness (A), Sound Source Envelopment (B), Timbre (C), High-Frequency Timbre (D), High-Frequency Modulation (E), Mid-Frequency Modulation (F), Low-Frequency Sound Sources (G), and Mid-High-Frequency Fluctuation (H). Finally suggestions were made on how to assess soundscapes by means of these dimensions in terms of statistical analysis and appropriate visualization.

The identified dimensions were then applied to three exemplary cases that serve as steps of validation. Chapter 4 attempts to find correlations between acoustic dimensions and perceptual attributes. For that, a listening experiment was conducted and analyzed that consists of three parts. In the first part general identification of sound source classes was requested from the participants in order to assess the ambiguity of perception of reproduced soundscapes. The second part investigated the consent of perceptual attributes by means of semantic differentials. The participant task was to evaluate a selection of ten attribute pairs for each reproduced soundscape. The perceptual results were then correlated with the eight acoustic dimensions to find appropriate correspondents with plausible success. The third part consisted of an outlook if the acoustic properties and the perceptual evaluations can be taken to interpret emotional responses.

A further validation step is carried out using the example of the investigation of ecological validity in Chapter 5. It was investigated whether the reproduction of a soundscape evokes the same expression of the acoustic dimensions compared to the (original) recording. The differences between recording and re-recording of the reproduction were analyzed and plausible similarities and differences were found with the result that the acoustic dimensions provide relevant information for maintaining ecological validity in terms of acoustic properties.

Finally, in Chapter 6 an investigation was conducted where a single acoustic property – the spatial composition – was varied on purpose,

leaving all other dimensions potentially unchanged. This was done by reproducing music of the same content with varying loudspeaker configurations. It could be shown, that the change in the sound field could be detected in a robust way with the dimension Sound Source Envelopment (B). Further, an adaptation of the acoustic dimensions was applied in order to reflect the peculiar characteristics of the acoustic environments produced by music reproduction with different loudspeaker configurations. This contributed to the question if the set of identified acoustic dimensions can be used to generally describe any acoustic environment. The result showed that in certain cases it is useful to adapt these dimensions in order to identify particular properties of a special subset of acoustic environments. In this case the specific dimensions High-Frequency Diffusivity (CC)and Elevational Diffusivity (GG)were able to differentiate the changes in the sound field in a even more detailed way.

7.2 OUTLOOK

The presented work is a contribution to the research on soundscape and general assessment of acoustic environments that provides novel methods and opens new possibilities in this field. Although parts of the methodology are motivated by neighboring disciplines, to the best of the author's knowledge, similar research has not yet been applied to soundscapes. In this sense, the author does not take the present work and its results as an incontrovertible given, but rather as a new perspective for the study of acoustic environments. Thus, it can be assumed that a constructive challenging of the methodology and results improves the robustness of the approach. Refinements and optimizations of the presented methodology are potentially useful with regard to the following aspects:

- The selection of acoustic indicators may not cover all perceptual relevant aspects of the assessment of acoustic environments. Adding further indicators could increase the interpretability and validity of the derived dimensions. This also counts for the parametrization of indicators such as the time-frequency resolution (window length and number of analysis bands).
- Even though the selection of soundscape recording databases was conducted with care, it may be the case that not all soundscape classes are represented equally. As Chapter 6 has shown, an adapted, dedicated set of acoustic dimensions may be applied for a specific class of soundscape with more explaining power. Thus, guidelines for adaptation of the acoustic dimensions for a specific object of investigation could be developed.
- From the two previous aspects arises the desire to develop new or adapted indicators that increase the variance explanation,

reduce the error and noise susceptibility and at the same time simplify the computational process. This is all the more true since the a priori categories of quality, loudness, spaciousness, and time could be confirmed as relevant and purposeful.

- The development of the underlying acoustic dimensions in this work is based on factor analysis. A further step of cross-validation of the method could be to perform the same task with another (statistical) method as proposed in Section 3.3.3.
- The validation of the identified acoustic dimensions was conducted at the example of three concrete applications in this work. For this purpose, further validation steps are conceivable, including a systematic variation of physical sound field properties as proposed in chapter 6, detailed listening experiments to find more precise perceptual correspondences to the acoustic dimensions, and finally the application to concrete research hypotheses with accompanying soundscape research methods.
- The temporal characteristics of the acoustic dimensions are represented in the presented methodology by means of modulation, fluctuation as well as the time series considerations within the independent component analysis. This aspect could be elaborated in more detail in the future to reflect the time-dependent human auditory perception also from the perspective of Gestalt psychology, thus closing the gap to auditory scene analysis (ASA).

With these proposed future works it can be expected that the methodology is capable to contribute in a relevant way to the description of acoustic environments as part of soundscape as well as neighboring disciplines.

7.3 CONCLUSION

The present work contributes to a comprehensive description of acoustic environments. It can be utilized for soundscape research, i.e. for the investigation on how and why humans perceive and react to their acoustic environment in context. The description of soundscapes by means of triangulation between the entities acoustic environment, person and context gains with the presented results the aspect of a statistical and perceptually relevant description of acoustic properties. Beyond soundscape research the proposed methodology and results can further be used for assessing general acoustic events and environments by means of computer-aided methods of machine learning, such as the detection and classification of acoustic events and scenes. These data-based areas already use methods of dimension reduction and variance aggregation and can benefit from a cross-validated a priori set of acoustic dimensions that are (more or less) independent

from the actual object of investigation.

The source code of the presented framework, including the calculation of indicators, statistical identification and analysis of the acoustic dimensions and visualization routines are available under https://gitlab.com/janywhere. This dissertation is publicly available under http://dx.doi.org/10.15488/13578.

Finally, the author hopes that this work contributes to a better understanding of human perception of acoustic environments and supports the aim to make the world a better sounding place.

REFERENCES

[1] Leibniz Universität Hannover. *WEA-Akzeptanz: Von der Schallquelle zur psychoakustischen Bewertung*. 2017. URL: https://www.wea-akzeptanz.uni-hannover.de/wea-akzeptanz.html (visited on 10/12/2022).

[2] Die Bundesregierung. *Technische Anleitung zum Schutz gegen Lärm – TA Lärm*. Tech. rep. 2017.

[3] A. S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 1990. ISBN: 9780262521956.

[4] ISO. *ISO 12913-1: Acoustics. Soundscape. Part 1. Definition and conceptual framework*. 2018.

[5] M. Southworth. "The Sonic Environment of Cities". In: *Environment and Behavior* 1.1 (June 1969), pp. 49–70. DOI: 10.1177/001391656900100104.

[6] R. M. Schafer. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester: Destiny Books, 1977.

[7] B. Krause. "Anatomy of the soundscape: Evolving perspectives". In: *J. Audio Eng. Soc.* 56.1-2 (2008), pp. 73–80.

[8] M. Zhang and J. Kang. "Towards the evaluation, description, and creation of soundscapes in urban open spaces". In: *Environment and Planning B: Planning and Design* 34.1 (2007), pp. 68–86. DOI: 10.1068/b31162.

[9] J. Kang and B. Schulte-Fortkamp. *Soundscape and the built environment*. Boca Raton, FL: CRC Press, 2016. DOI: 10.1201/b19145.

[10] ISO. *ISO/TS 12913-2: Acoustics. Soundscape. Part 2. Data Collection and reporting requirements*. 2019.

[11] ISO. *ISO/TS 12913-3: Acoustics. Soundscape. Part 3. Data Analysis*. Tech. rep. 2020.

[12] B. Rafaely. *Fundamentals of Spherical Array Processing*. Berlin, Heidelberg: Springer, 2015. DOI: 10.1007/978-3-662-45664-4.

[13] E. G. Williams. *Fourier Acoustics*. London: Elsevier, 1999. DOI: 10.1016/B978-0-12-753960-7.X5000-1.

[14] Ö. Axelsson, M. E. Nilsson, B. Hellström, and P. Lundén. "A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park". In: *Landscape and Urban Planning* 123 (Mar. 2014), pp. 49–60. DOI: 10.1016/j.landurbplan.2013.12.005.

[15] J. Blauert. *Räumliches Hören*. Stuttgart: S. Hirzel Verlag, 1974. ISBN: 9783777622873.

[16] J. Blauert. *Spatial Hearing - The Psychophysics of Human Sound Localization*. 2nd enlarg. Harvard, MA: The MIT Press, 1997. ISBN: 0-262-02413-6.

[17] H. Møller. "Fundamentals of binaural technology". In: *Applied Acoustics* 36.3-4 (1992), pp. 171–218. DOI: 10.1016/0003-682X(92)90046-U.

[18] R. Nicol. *Binaural Technology*. New York, NY: AES Inc., 2010. ISBN: 9780937803721.

[19] P. Majdak, C. Hollomey, and R. Baumgartner. "AMT 1.x: A toolbox for reproducible research in auditory modeling". In: *Acta Acustica* 6.19 (May 2022). DOI: 10.1051/aacus/2022011.

[20] S. Li and J. Peissig. "Measurement of Head-Related Transfer Functions: A Review". In: *Applied Sciences* 10.14 (July 2020), p. 5014. DOI: 10.3390/app10145014.

[21] H. Ziegelwanger, P. Majdak, and W. Kreuzer. "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization". In: *The Journal of the Acoustical Society of America* 138.1 (July 2015), pp. 208–222. DOI: 10.1121/1.4922518.

[22] C. Guezenoc and R. Séguier. "HRTF Individualization: A Survey". In: *AES 145th Convention*. 2018.

[23] A. Neidhardt, C. Schneiderwind, and F. Klein. "Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework". In: *Trends in Hearing* 26 (Jan. 2022). DOI: 10.1177/23312165221092919.

[24] S. Werner, F. Klein, A. Neidhardt, U. Sloma, C. Schneiderwind, and K. Brandenburg. "Creation of Auditory Augmented Reality Using a Position-Dynamic Binaural Synthesis System—Technical Components, Psychoacoustic Needs, and Perceptual Evaluation". In: *Applied Sciences* 11.3 (Jan. 2021), p. 1150. ISSN: 2076-3417. DOI: 10.3390/app11031150.

[25] M. A. Gerzon. "Periphony : With-Height Sound Reproduction". In: *J. Audio Eng. Soc.* 21.1 (1973).

[26] M. A. Gerzon. "Practical Periphony: The Reproduction of Full-Sphere Sound". In: *AES 65th Convention*. 1980.

[27] E. Bates, M. Gorzel, L. Ferguson, H. O'Dwyer, and F. M. Boland. "Comparing Ambisonic microphones - Part 1". In: *AES International Conference on Sound Field Control* (2016).

[28] E. Bates, S. Dooney, M. Gorzel, H. O'Dwyer, L. Ferguson, and F. M. Boland. "Comparing Ambisonic Microphones – Part 2". In: *AES 142nd Convention*. 2017.

[29] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi, and V. Pulkki. "Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis". In: *AES 144th Convention*. Milan, Italy, 2018.

[30] Institute of Electronic Music and Acoustics (IEM) and D. Rudrich. *IEM Plug-in Suite*. 2021. URL: https://plugins.iem.at/.

[31] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich. "A 3D Real Time Rendering Engine for Binaural Sound Reproduction". In: *Proc. of the 2003 International Conference on Auditory Display, ICAD*. Boston, 2003.

[32] F. Zotter and M. Frank. "All-Round Ambisonic Panning and Decoding". In: *J. Audio Eng. Soc.* 60.10 (2012), pp. 807–820.

[33] J. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan. "Spatial Audio for Soundscape Design: Recording and Reproduction". In: *Applied Sciences* 7.6 (June 2017), p. 627. DOI: 10.3390/app7060627.

[34] B. De Coensel, K. Sun, and D. Botteldooren. "Urban Soundscapes of the World: Selection and reproduction of urban acoustic environments with soundscape in mind". In: *Proc. of the 46th INTER-NOISE*. Hong Kong, 2017.

[35] K. Sun, D. Botteldooren, and B. De Coensel. "Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation". In: *Proc. of the 47th INTER-NOISE*. Chicago, USA, 2018.

[36] K. Genuit and A. Fiebig. "Psychoacoustics and its Benefit for the Soundscape Approach". In: *Acta Acustica united with Acustica* 92 (2006), pp. 952–958.

[37] B. B. Boren, M. Musick, J. Grossman, and A. Roginska. "I HEAR NY4D : Hybrid Acoustic and Augmented Auditory Display for Urban Soundscapes". In: *Proc. of the 20th International Conference on Auditory Display, ICAD*. New York, USA, 2014.

[38] A. Mitchell, T. Oberman, F. Aletta, M. Erfanian, M. Kachlicka, M. Lionello, and J. Kang. "The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information". In: *Applied Sciences* 10.7 (Apr. 2020), p. 2397. DOI: 10.3390/app10072397.

[39] C. Xu and J. Kang. "Soundscape evaluation: Binaural or monaural?" In: *The Journal of the Acoustical Society of America* 145.5 (May 2019), pp. 3208–3217. DOI: 10.1121/1.5102164.

[40] W. J. Davies, N. S. Bruce, and J. E. Murphy. "Soundscape Reproduction and Synthesis". In: *Acta Acustica united with Acustica* 100.2 (Mar. 2014), pp. 285–292. DOI: 10.3813/AAA.918708.

[41] M. C. Green and D. Murphy. "EigenScape: A Database of Spatial Acoustic Scene Recordings". In: *Applied Sciences* 7.11 (Nov. 2017), p. 1204. DOI: 10.3390/app7111204.

[42] J. M. Buchholz and A. Weisser. *Ambisonics Recordings of Typical Environments (ARTE) Database*. 2019. DOI: 10.5281/zenodo.2261633. URL: https://doi.org/10.5281/zenodo.2261633.

[43] T. Heittola, A. Mesaros, and T. Virtanen. *TAU Urban Acoustic Scenes 2020*. 2020. DOI: 10.5281/zenodo.3670185. URL: https://zenodo.org/record/3670185 (visited on 10/12/2022).

[44] C. Guastavino and B. F. G. Katz. "Perceptual evaluation of multi-dimensional spatial audio reproduction". In: *The Journal of the Acoustical Society of America* 116.2 (Aug. 2004), pp. 1105–1115. DOI: 10.1121/1.1763973.

[45] C. Guastavino, B. F. G. Katz, J.-d. Polack, D. J. Levitin, and D. Dubois. "Ecological Validity of Soundscape Reproduction". In: *Acta Acustica united with Acustica* 91 (2005), pp. 333–341.

[46] C. Tarlao, D. Steele, and C. Guastavino. "Assessing the ecological validity of soundscape reproduction in different laboratory settings". In: *PLOS ONE* 17.6 (June 2022). Ed. by C. Pegoraro, e0270401. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0270401.

[47] J. Bergner, S. Preihs, and J. Peissig. "Investigation on Ecological Validity within Higher Order Ambisonics Reproductions of Wind Turbine Noisescapes". In: *Proc. of INTER-NOISE*. Madrid, Spain, June 2019.

[48] F. Stevens, D. T. Murphy, and S. L. Smith. "Ecological validity of stereo UHJ soundscape reproduction". In: *AES 142nd Convention* (2017).

[49] Y. Wycisk, R. Kopiez, J. Bergner, K. Sander, S. Preihs, J. Peissig, and F. Platz. "The Headphone and Loudspeaker Test – Part I: Suggestions for controlling characteristics of playback devices in internet experiments". In: *Behavior Research Methods* (May 2022). DOI: 10.3758/s13428-022-01859-8.

[50] J. A. Russell, L. M. Ward, and G. Pratt. "Affective Quality Attributed to Environments - A Factor Analytic Study". In: *Environment and Behavior* 13.3 (1981), pp. 259–288.

[51] Ö. Axelsson, M. E. Nilsson, and B. Berglund. "A principal components model of soundscape perception". In: *The Journal of the Acoustical Society of America* 128.5 (Nov. 2010), pp. 2836–2846. DOI: 10.1121/1.3493436.

[52] J. Kang and M. Zhang. "Semantic differential analysis of the soundscape in urban open public spaces". In: *Building and Environment* 45.1 (Jan. 2010), pp. 150–157. DOI: 10.1016/j.buildenv.2009.05.014.

[53] R. Cain, P. Jennings, and J. Poxon. "The development and application of the emotional dimensions of a soundscape". In: *Applied Acoustics* 74.2 (Feb. 2013), pp. 232–239. DOI: 10.1016/j.apacoust.2011.11.006.

[54] F. Stevens, D. T. Murphy, and S. L. Smith. "Soundscape preference rating using semantic differential pairs and the self-assessment manikin". In: *Proc. of the 13th Sound and Music Computing Conference, SMC*. 2016, pp. 455–462. ISBN: 9783000537004.

[55] F. Aletta, J. Kang, and Ö. Axelsson. "Soundscape descriptors and a conceptual framework for developing predictive soundscape models". In: *Landscape and Urban Planning* 149 (May 2016), pp. 65–74. DOI: 10.1016/j.landurbplan.2016.02.001.

[56] A. Fiebig, P. Jordan, and C. C. Moshona. "Assessments of Acoustic Environments by Emotions – The Application of Emotion Theory in Soundscape". In: *Frontiers in Psychology* 11 (Nov. 2020). DOI: 10.3389/fpsyg.2020.573041.

[57] M. Lionello, F. Aletta, and J. Kang. "A systematic review of prediction models for the experience of urban soundscapes". In: *Applied Acoustics* 170 (Dec. 2020), p. 107479. DOI: 10.1016/j.apacoust.2020.107479.

[58] M. S. Engel, A. Fiebig, C. Pfaffenbach, and J. Fels. "A Review of the Use of Psychoacoustic Indicators on Soundscape Studies". In: *Current Pollution Reports* 7.3 (Sept. 2021), pp. 359–378. DOI: 10.1007/s40726-021-00197-1.

[59] J. Bergner, S. Preihs, and J. Peissig. "On Wind Turbine Noisescape Reproduction for Perceptual Evaluation". In: *Fortschritte der Akustik - DAGA*. Rostock, Germany, Mar. 2019.

[60] J. Bergner, S. Preihs, and J. Peissig. "A Virtual Acoustic Environment for Psychoacoustic Assessment of Wind Turbine Noisescapes". In: *Proc. of the International Conference on Spatial Audio, ICSA*. Ilmenau, Germany, Sept. 2019.

[61] S. Preihs, J. Bergner, and J. Peissig. "Ansätze zur binauralen Erweiterung einer Algorithmik zur lästigkeitsbezogenen Analyse und Synthese der Schallemissionen von Windenergieanlagen". In: *Fortschritte der Akustik - DAGA*. Rostock, Germany, Mar. 2019.

[62] D. Schössow, J. Bergner, S. Preihs, and J. Peissig. "Audiovisuelle Laborstudie zur Lästigkeit von WEA-Schall". In: *Fortschritte der Akustik - DAGA*. Hannover, Germany, Mar. 2020.

[63] S. Preihs, J. Bergner, D. Schössow, and J. Peissig. "On Predicting the Perceived Annoyance of Wind Turbine Sound". In: *Fortschritte der Akustik - DAGA*. Vienna, AUT, Mar. 2021.

[64] S. Preihs, J. Bergner, D. Schössow, and J. Peissig. "Assessing Wind Turbine Noise Perception by means of Contextual Laboratory and Online Studies". In: *Proc. of the 9th Int. Conf. on Wind Turbine Noise*. Apr. 2021.

[65] J. Bergner, S. Preihs, and J. Peissig. "Soundscape Fingerprinting - Methods and Parameters for Acoustic Assessment". In: *Fortschritte der Akustik - DAGA*. Mar. 2021.

[66] J. Bergner and J. Peissig. "On the identification and assessment of underlying acoustic dimensions of soundscapes". In: *Acta Acustica* 6.46 (Oct. 2022). DOI: 10.1051/aacus/2022042.

[67] A. Lerch. *An Introduction to Audio Content Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., July 2012. DOI: 10.1002/9781118393550.

[68] A. Mesaros, T. Heittola, and T. Virtanen. "Acoustic Scene Classification: An Overview of Dcase 2017 Challenge Entries". In: *Proc. of the 16th International Workshop on Acoustic Signal Enhancement, IWAENC*. Sept. 2018, pp. 411–415. DOI: 10.1109/IWAENC.2018.8521242.

[69] J. Sueur, T. Aubin, and C. Simonis. "Seewave, A Free Modular Tool for Sound Analysis and Synthesis". In: *Bioacoustics* 18.2 (Jan. 2008), pp. 213–226. ISSN: 0952-4622. DOI: 10.1080/09524622.2008.9753600.

[70] H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. 3rd Editio. Berlin, Heidelberg: Springer, 2007. DOI: 10.1007/978-3-540-68888-4.

[71] ISO. *ISO 532-1:2017 Acoustics - Methods for calculating loudness - Part 1: Zwicker method*. 2017.

[72] The Mathworks Inc. *Audio Toolbox*. 2022. URL: https://de.mathworks.com/products/audio.html.

[73] G. Peeters. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project". Paris, France, 2004. URL: http://www.citeulike.org/group/1854/article/1562527.

[74] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky. "Spectral entropy based feature for robust ASR". In: *IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, ICASSP*. 2004. DOI: 10.1109/icassp.2004.1325955.

[75] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis. "A Speech/-Music Discriminator of Radio Recordings Based on Dynamic Programming and Bayesian Networks". In: *IEEE Transactions on Multimedia* 10.5 (2008), pp. 846–857. DOI: 10.1109/TMM.2008.922870.

[76] E. Scheirer and M. Slaney. "Construction and evaluation of a robust multifeature speech/music discriminator". In: *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. Vol. 2. 1997, pp. 1331–1334. ISBN: 0-8186-7919-0. DOI: 10.1109/ICASSP.1997.596192.

[77] Acoustical Society of America and American National Standards Institute. "ANSI S1.11: Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters". In: *American National Standards on Acoustics* 1986.734 (1986).

[78] S. B. Davis and P. Mermelstein. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continously Spoken Sentences". In: *Transactions on Acoustics, Speech, and Signal Processing* 28.4 (1980), pp. 357–366.

[79] S. Hatano and T. Hashimoto. "Booming index as a measure for evaluating booming sensation". In: *Proc. of the 29th INTER-NOISE*. 2000, pp. 4332–4336.

[80] AudioCommons and Institute of Sound Recording. *Timbral Models*. 2019. URL: https://www.audiocommons.org/.

[81] ISO. *ISO 532-2:2017 Acoustics - Methods for calculating loudness - Part 2: Moore-Glasberg method*. 2017.

[82] EBU. *EBU - R 128: Loudness normalisation and permitted maximum level of audio signals*. Tech. rep. 2014.

[83] ITU. *ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level*. Tech. rep. 2015.

[84] EBU. *EBU Tech 3342: Loudness Range: A Measure to Supplement EBU R 128 Loudness Normalization*. 2016.

[85] IEC. *IEC 61672-1:2013 Electroacoustics - Sound level meters - Part 1: Specifications*. 2013.

[86] V. Pulkki. "Spatial sound reproduction with directional audio coding". In: *J. Audio Eng. Soc.* 55.6 (2007), pp. 503–516.

[87] A. Politis. "Microphone array procesisng for parametric spatial audio techniques". PhD thesis. Aalto University, Finland, 2016. URL: https://github.com/polarch.

[88] IEC. *IEC 60268-5:2003 Sound system equipment – Part 5:Loudspeakers*. 2003.

[89] B. Bernschütz. "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100". In: *Fortschritte der Akustik - AIA-DAGA*. 2013, pp. 592–595. URL: http://www.audiogroup.web.fh-koeln.de/FILES/AIA-DAGA2013%7B%5C_%7DHRIRs.pdf.

[90] C. Schörkhuber, M. Zaunschirm, and R. Höldrich. "Binaural rendering of Ambisonic signals via magnitude least squares". In: *Fortschritte der Akustik - DAGA*. 2018, pp. 339–342.

[91] A. Weisser, J. M. Buchholz, C. Oreinos, J. Badajoz-Davila, J. Galloway, T. Beechey, and G. Keidser. "The Ambisonic Recordings of Typical Environments (ARTE) Database". In: *Acta Acustica united with Acustica* 105.4 (July 2019), pp. 695–713. ISSN: 1610-1928. DOI: 10.3813/AAA.919349.

[92] Soundfield by Røde. *Ambisonic Sound Library*. URL: https://library.soundfield.com/ (visited on 10/12/2022).

[93] B. B. Boren, A. Andreopoulou, M. Musick, H. Mohanraj, and A. Roginska. "I Hear NY3D: Ambisonic Capture and Reproduction of an Urban Sound Environment". In: *AES 135th Convention*. New York, NY, 2013.

[94] mh acoustics LLC. *em32 Eigenmike Demos*. URL: https://mhacoustics.com/demos (visited on 10/12/2022).

[95] A. Politis, S. Adavanne, and T. Virtanen. *TAU Spatial Sound Events*. 2020. DOI: 10.5281/zenodo.4064792. URL: https://zenodo.org/record/4064792.

[96] A. Mesaros, T. Heittola, and T. Virtanen. "TUT database for acoustic scene classification and sound event detection". In: *Proc. of the 24th European Signal Processing Conference, EUSIPCO*. Budapest, Hungary, Aug. 2016, pp. 1128–1132. DOI: 10.1109/EUSIPCO.2016.7760424.

[97] D. Zelterman. *Applied Multivariate Statistics with R*. Statistics for Biology and Health. Cham: Springer, 2015, pp. 1–393. DOI: 10.1007/978-3-319-14093-3.

[98] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, June 2012. DOI: 10.1017/CBO9780511804779.

[99] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830.

[100] J. Bortz. *Statistik für Human-und Sozialwissenschaftler*. 6th Editio. Berlin, Heidelberg: Springer, 2005. ISBN: 354021271X.

[101] G. J. Lautenschlager, C. E. Lance, and V. L. Flaherty. "Parallel Analysis Criteria: Revised Equations for Estimating the Latent Roots of Random Data Correlation Matrices". In: *Educational and Psychological Measurement* 49.2 (June 1989), pp. 339–345. DOI: 10.1177/0013164489492006.

[102] T. Okano, L. L. Beranek, and T. Hidaka. "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls". In: *The Journal of the Acoustical Society of America* 104.1 (July 1998), pp. 255–265. DOI: 10.1121/1.423955.

[103] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001. DOI: 10.7551/mitpress/3717.003.0014.

[104] J. Bortz and C. Schuster. "Tests zur Überprüfung von Unterschiedshypothesen". In: Springer Berlin Heidelberg, 2010, pp. 117–136. DOI: 10.1007/978-3-642-12770-0_8.

[105] A. Hyvärinen. "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis". In: *IEEE Transactions on Neural Networks* 10.3 (1999), pp. 626–634. DOI: 10.1109/72.761722.

[106] B. Schölkopf, A. Smola, and K.-R. Müller. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation* 10.5 (July 1998), pp. 1299–1319. DOI: 10.1162/089976698300017467.

[107] M. A. Kramer. "Nonlinear principal component analysis using autoassociative neural networks". In: *AIChE Journal* 37.2 (1991), pp. 233–243. ISSN: 15475905. DOI: 10.1002/aic.690370209.

[108] L. van der Maaten and G. Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learing Research* 9 (2008), pp. 2579–2605.

[109] J. L. Devore, K. N. Berk, and M. A. Carlton. *Modern Mathematical Statistics with Applications*. Springer Texts in Statistics. Cham: Springer International Publishing, 2021. DOI: 10.1007/978-3-030-55156-8.

[110] W. J. Conover. *Practical nonparametric statistics*. 3rd Editio. New York, NY: Wiley, 1999. ISBN: 0471160687.

[111] J. Bergner, S. Preihs, and J. Peissig. "Perceptual Correlates of Underlying Acoustic Dimensions in Soundscape Assessment". In: *Fortschritte der Akustik - DAGA*. Hamburg, Germany, 2023.

[112] Cockos Inc. *Reaper Digital Audio Workstation*. 2022. URL: https://www.reaper.fm/index.php.

[113] R. Hupke, J. Ordner, J. Bergner, M. Nophut, S. Preihs, and J. Peissig. "Towards a Virtual Audiovisual Environment for Interactive 3D Audio Productions". In: *AES International Conference on Immersive and Interactive Audio*. Mar. 2019.

[114] ITU. *ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems*. 2015.

[115] J. Bergner, S. Preihs, R. Hupke, and J. Peissig. "A System for Room Response Equalization of Listening Areas Using Parametric Peak Filters". In: *AES International Conference on Immersive and Interactive Audio*. York, UK, Mar. 2019.

[116] K. Zimmer and W. Ellermeier. "Ein Kurzfragebogen zur Erfassung der Lärmempfindlichkeit". In: *Umweltpsychologie* 2.2 (1998), pp. 54–63.

[117] K. Zimmer and W. Ellermeier. "Konstruktion und Evaluation eines Fragebogens zur Erfassugn der individuellen Lärmempfindlichkeit". In: *Diagnostica* 44 (1998), pp. 11–20.

[118] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl. "A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI)". In: *Acta Acustica united with Acustica* 100.5 (Sept. 2014), pp. 984–994. ISSN: 16101928. DOI: 10.3813/AAA.918778.

[119] A. Lindau. *Spatial Audio Quality Inventory (SAQI): Test Manual 1.2*. 2015.

[120] J. C. Nunnally. *Psychometric theory*. 2nd Editio. McGraw-Hill, 1978.

[121] A. Mitchell. "Predictive Modelling of Complex Urban Soundscapes : Enabling an engineering approach to soundscape design". PhD thesis. University College London (UCL), 2022. DOI: 10.13140/RG.2.2.15590.50245.

[122] J. Bergner, D. Schössow, S. Preihs, Y. Wycisk, K. Sander, R. Kopiez, and F. Platz. "Analyzing the Degree of Immersion of Music Reproduction by means of Acoustic Fingerprinting". In: *Fortschritte der Akustik - DAGA*. Stuttgart, Germany, Mar. 2022.

[123] J. Bergner, D. Schössow, S. Preihs, and J. Peissig. "Identification of Discriminative Acoustic Dimensions in Stereo, Surround and 3D Music Reproduction". In: *to be published in J. Audio Eng. Soc.* (Oct. 2022). DOI: 10.17743/jaes.2022.0071.

[124] ITU. *ITU-R BS.2051-2: Advanced sound system for programme production*. Tech. rep. 2018.

[125] C. Eaton and H. Lee. "Subjective evaluations of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings". In: *Acoustical Science and Technology* 43.2 (Mar. 2022), E2169. DOI: 10.1250/ast.43.149.

[126] M. Schoeffler, A. Silzle, and J. Herre. "Evaluation of Spatial / 3D Audio : Basic Audio Quality Versus Quality of Experience". In: *IEEE Journal of Selected Topics in Signal Processing* 11.1 (2017), pp. 75–88.

[127] E. Hahn. "Musical emotions evoked by 3D audio". In: *AES International Conference on Spatial Reproduction* (2018), pp. 380–387.

[128] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". In: *Neural Networks* 13.4-5 (June 2000), pp. 411–430. DOI: 10.1016/S0893-6080(00)00026-5.

[129] M. Schaab, T. Clauss, J. Bergner, and C. Sladeczek. "Personal Sound Zones: Study on the Threshold of Acceptability in an Automotive Environment". In: *Fortschritte der Akustik - DAGA*. Munich, Germany, Mar. 2018.

[130] J. Bergner, C. Sladeczek, and J. Redlich. "Perception-based Investigations on the Monopole Synthesis for Reproduction of Directional Sound Sources". In: *Fortschritte der Akustik - DAGA*. Kiel, Germany, Mar. 2017.

[131] M. Seideneck, J. Bergner, and C. Sladeczek. "Object-based audio in large scale live sound reinforcement controlled by motion tracking". In: *AES 142nd Convention*. Berlin, Germany, May 2017.

[132] J. Bergner, T. Clauss, A. Zhykhar, C. Sladeczek, and S. Brix. "Application of Wave Field Synthesis in Virtual Acoustic Engineering". In: *Proc. of INTER-NOISE*. Hamburg, Germany, Aug. 2016.

[133] C. Sladeczek, D. Beer, J. Bergner, A. Zhykhar, M. Wolf, and A. Franck. "High-Directional Beamforming with a Miniature Loudspeaker Array". In: *Fortschritte der Akustik - DAGA*. Aachen, Germany, Mar. 2016.

[134] J. Bergner, A. Gasull-Ruiz, C. Sladeczek, and S. Brix. "VISTA4F - Development of an Audiovisual Virtual Reality Test Environment for Automotive". In: *Proc. of the 3rd International Conference on Spatial Audio, ICSA*. Graz, Austria, Sept. 2015.

[135] T. Klouche, T. Samulewicz, and J. Bergner. "Measuring the accuracy of microtonal synthesizers: Pianoteq & Vogue". In: *Fortschritte der Akustik - AIA-DAGA*. Mar. 2013.

[136] T. Klouche, T. Samulewicz, and J. Bergner. "Validation of computational tuning systems". In: *Fortschritte der Akustik - DAGA*. Darmstadt, Germany, Mar. 2012.

A

ADDITIONAL TABLES

Table A.1: Indicator composition the first 8 relevant unrotated factors j with respective explained variance s_j^2 and relative loadings $l_{\text{rel},ij}$ in parentheses. Trailing numbers of the indicators denote the frequency bands. $N_{\text{i},j}$ denotes how many of the total number of indicators account for $\geq 51\,\%$ of the factor's explained variance.

Factor j	s_j^2	$N_{\text{i},j}$	**Indicators**
1	92.44 (30.4%)	51	loudnessZwicker(-0.103), LA(-0.102), LAeq(-0.102), loudnessZwickerBands05(-0.101), loudnessZwickerBands04(-0.101), LAeqBands05(-0.101), lufsMom(-0.100), LABands05(-0.100), lufsMomBands05(-0.100), loudnessZwickerBands03(-0.100), LAmax(-0.100), lufsPeakBands05(-0.100), loudnessZwickerBands06(-0.100), LAeqBands03(-0.100), oct06(-0.100), LAeqBands04(-0.100), LABands03(-0.100), LABands04(-0.100), lufsShort(-0.100), mfcc00(-0.100), lufsPeakBands03(-0.100), lufsMomBands03(-0.100), lufsMomBands04(-0.100), lufsPeak(-0.099), lufsShortBands05(-0.099), lufsPeakBands04(-0.099), LApeakBands03(-0.099), LAmaxBands03(-0.099), oct05(-0.099), oct04(-0.099), lufsShortBands03(-0.099), LApeak(-0.099), LAmaxBands05(-0.099), lufsShortBands04(-0.099), LABands06(-0.099), LAeqBands06(-0.099), oct07(-0.099), LAmaxBands04(-0.099), LApeakBands04(-0.098), lufsPeakBands06(-0.098), lufsMomBands06(-0.098), LApeakBands05(-0.098), LApeakBands02(-0.098), LAeqBands02(-0.098), LABands02(-0.097), LAmaxBands02(-0.097), lufsShortBands06(-0.097), loudnessZwickerBands02(-0.096), LAmaxBands06(-0.096), loudnessZwickerBands07(-0.095), lufsPeakBands02(-0.095)
2	18.64 (6.1%)	31	spectralSkewness(-0.174), spectralCentroid(0.166), spectralSpread(0.164), spectralRolloffPoint(0.159), spectralKurtosis(-0.159), spectralFlatness(0.157), sharp(0.154), spectralEntropy(0.144), lufsPeakBands00(-0.126), lufsMomBands00(-0.124), oct01(-0.123), lufsShortBands00(-0.122), oct00(-0.121), oct02(-0.119), mfcc01(-0.117), lufsPeakBands01(-0.116), spectralCrest(-0.116), LABands00(-0.116), lufsMomBands01(-0.115), LAeqBands00(-0.115), LAmaxBands00(-0.114), lufsMomBands09(0.114), lufsShortBands01(-0.112), LAmaxBands09(0.112), lufsPeakBands09(0.111), LABands09(0.110), LAeqBands09(0.110), lufsShortBands09(0.109), LApeakBands00(-0.107), LABands01(-0.103), LAeqBands01(-0.102)

Continued on next page

Table A.1 – continued from previous page

Factor j	s_j^2	$N_{\mathrm{i},j}$	**Indicators**
3	13.98 (4.6%)	25	sphDIo7(0.178), sphDIo9(0.174), sphDIo8(0.173), sphDIAz07(0.171), sphDIAz08(0.159), mfcc01(-0.157), sphDIo5(0.154), sphDIo6(0.153), sphDIEl07(0.151), sphDIEl09(0.147), sphDIEl08(0.145), sphDIEl04(0.144), sphDIAz06(0.143), sphDIo4(0.142), sphDIEl05(0.138), diff07(-0.133), sharp(0.128), sphDIAz09(0.128), sphDIAz05(0.126), diff08(-0.124), diff06(-0.124), spectralEntropy(-0.118), sphDIEl06(0.110), sphDIAz04(0.108), spectralCrest(0.107)
4	14.09 (4.6%)	21	sphDIEl05(-0.187), sphDIo6(-0.186), sphDIo5(-0.185), sphDIEl07(-0.181), sphDIEl06(-0.179), sphDIo4(-0.174), sphDIEl04(-0.173), sphDIEl08(-0.173), sphDIo7(-0.166), sphDIAz05(-0.159), sphDIo8(-0.158), sphDIAz06(-0.155), sphDIAz04(-0.151), sphDIEl03(-0.151), sphDIo3(-0.150), sphDIAz03(-0.122), sphDIAz07(-0.122), diff05(0.120), diff06(0.117), sphDIEl09(-0.115), sphDIAz08(-0.108)
5	8.26 (2.7%)	17	modDepthP109(0.197), modDepthS08(0.196), modDepthS09(0.191), modDepthP108(0.185), modDepthP209(0.184), modDepthP309(0.184), modDepthS07(0.181), modDepthP308(0.175), modDepthP208(0.174), modDepthP207(0.171), modDepthP107(0.170), modDepthP307(0.169), modDepthS06(0.163), modDepthS05(0.162), modDepthP105(0.160), modDepthP206(0.155), modDepthP306(0.153)
6	6.12 (2.0%)	27	mfcc02(-0.200), fluct08(-0.189), fluct09(-0.160), fluct07(-0.160), fluct06(-0.152), modDepthS06(0.152), modDepthS05(0.143), modDepthP106(0.141), modDepthP105(0.138), modDepthP306(0.137), modDepthP206(0.136), mfcc04(-0.134), fluct05(-0.132), modDepthP305(0.128), rough02(0.127), modDepthP205(0.127), sphDIo2(-0.126), rough03(0.123), diff02(0.119), sphDIEl02(-0.119), fluct04(-0.116), modDepthP208(-0.116), spectralSkewness(-0.112), spectralKurtosis(-0.109), doaEl05(0.109), sphDIo1(-0.108), iacc05(0.108)
7	4.99 (1.6%)	17	modDepthS03(0.197), modDepthP203(0.195), modDepthP303(0.194), modDepthP103(0.192), modDepthS02(0.185), modDepthP302(0.181), modDepthP202(0.180), modDepthS09(-0.180), iacc03(0.180), modDepthP102(0.169), spectralSlope(-0.162), modDepthP309(-0.158), iacc02(0.157), modDepthS08(-0.155), modDepthP209(-0.155), modDepthP109(-0.154), modDepthP201(0.147)

B

ADDITIONAL FIGURES

LISTENING EXPERIMENT

Significant differences in the distributions of the factor scores of the stimuli for the listening experiment can be taken from Figures B.1 for the original recordings and B.2 for the re-recorded stimuli.

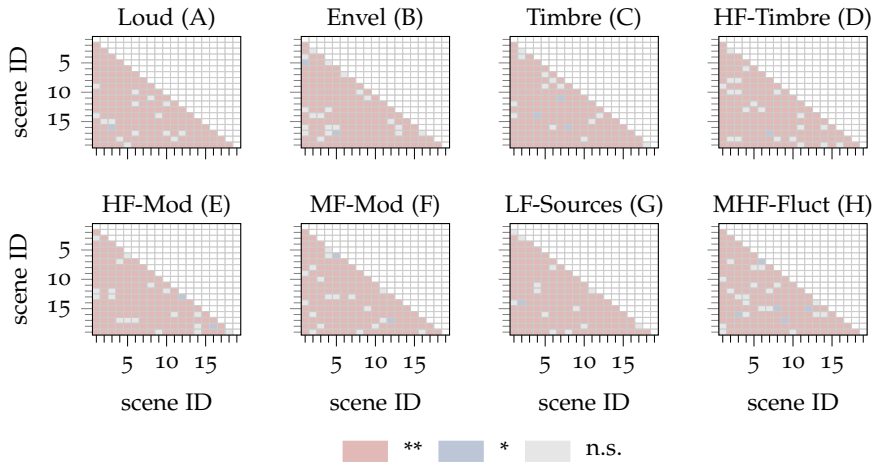


Figure B.1: Statistical differences between recorded sample soundscapes. Red: strong significance ($p < 0.01$), blue: moderate significance ($p < 0.05$), gray: no significance ($p > 0.05$).

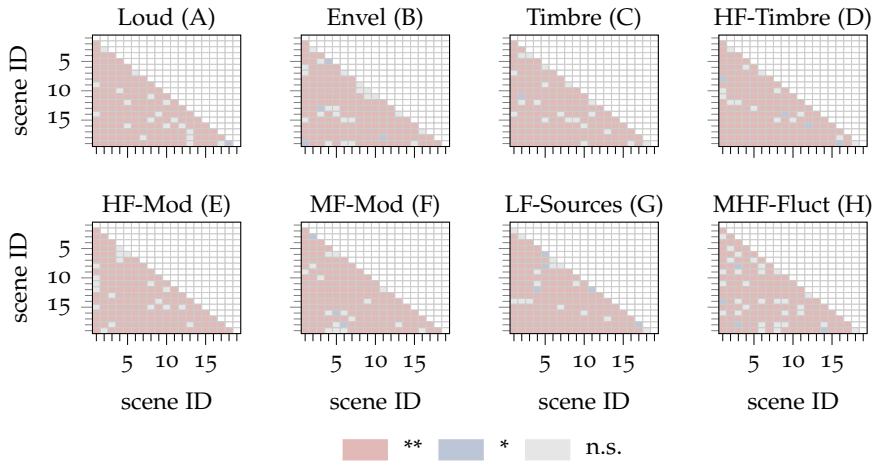


Figure B.2: Statistical differences between re-recorded sample soundscapes. Red: strong significance ($p < 0.01$), blue: moderate significance ($p < 0.05$), gray: no significance ($p > 0.05$).

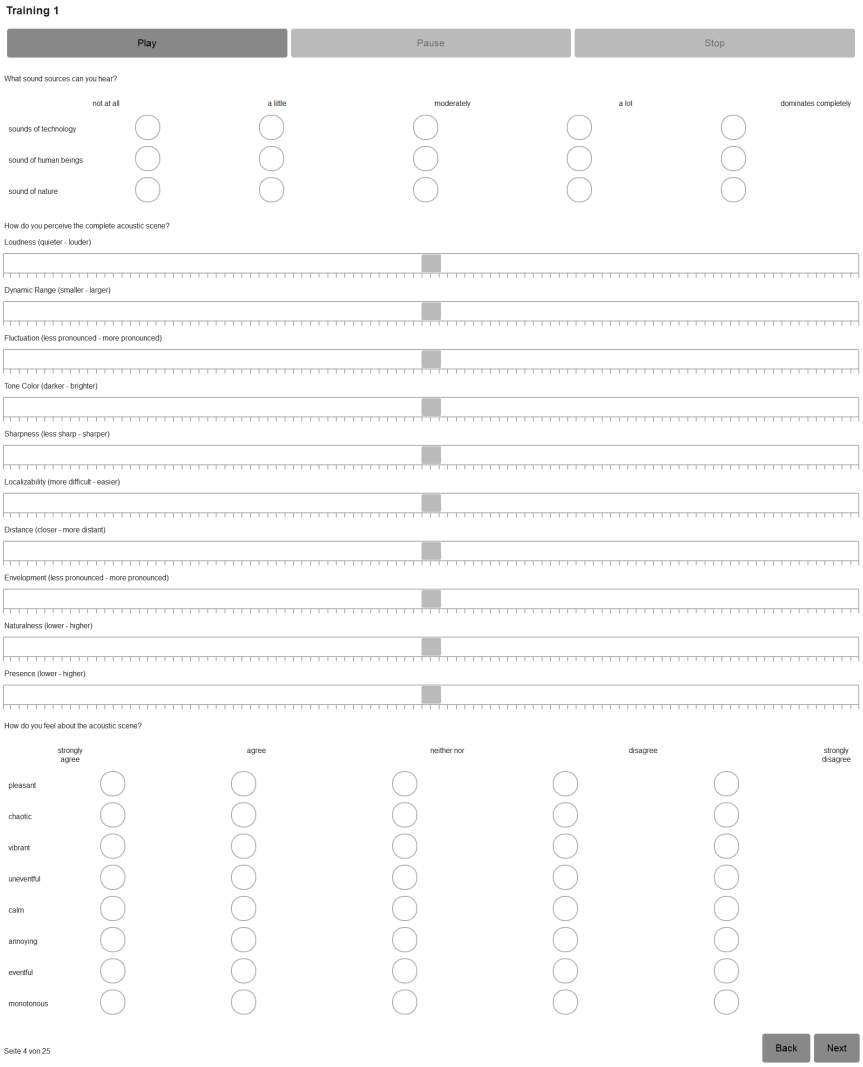
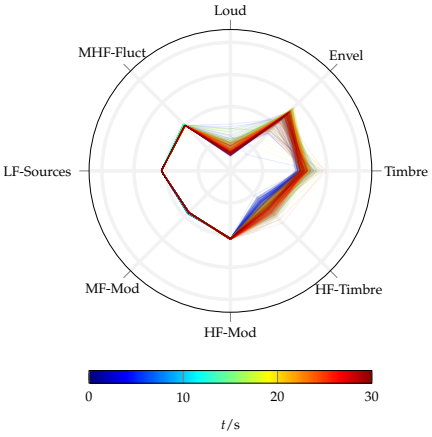
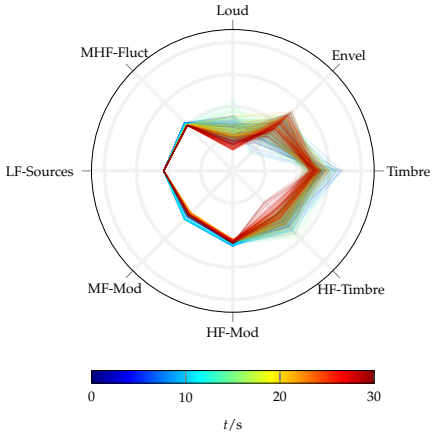


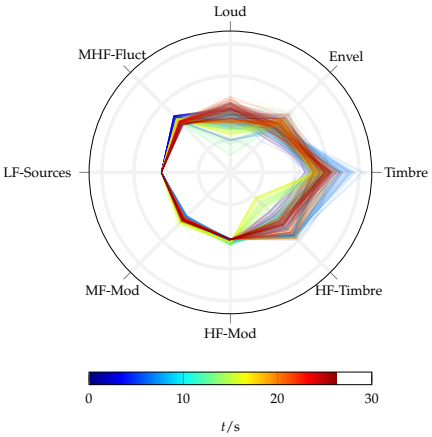
Figure B.3: Graphical user interface of exemplary stimulus of listening experiment.



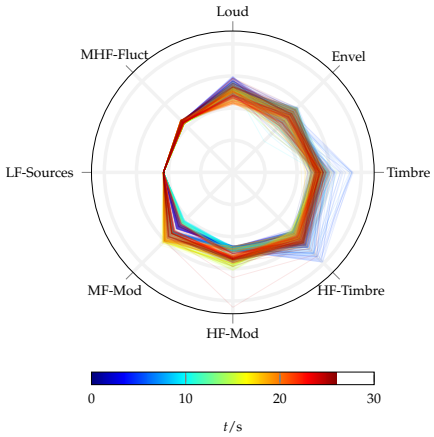
(a) A: Library



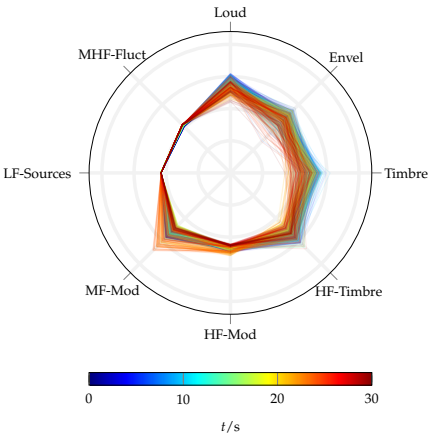
(b) A: Office



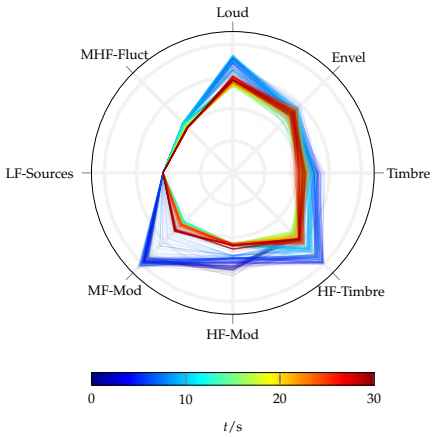
(c) A: LivingRoom



(d) A: Cafe

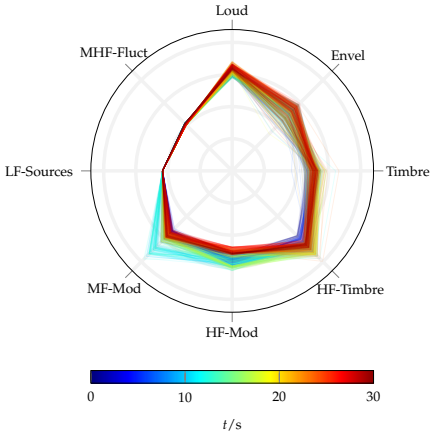


(e) A: DinnerParty

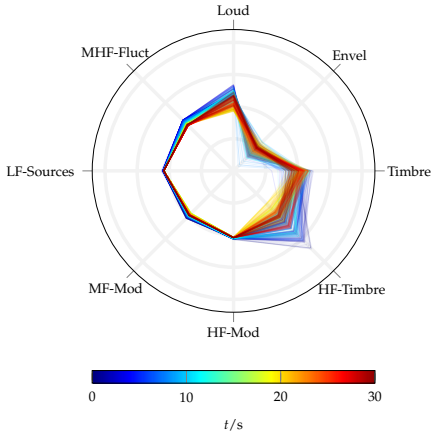


(f) A: TrainStation

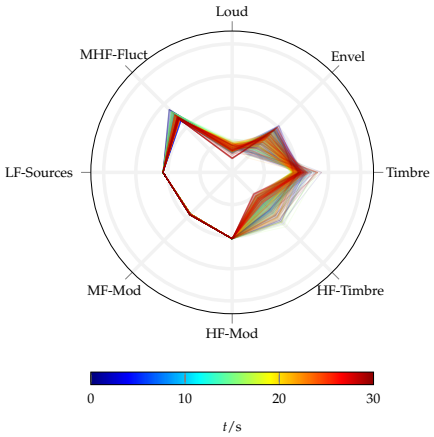
Figure B.4: Fingerprints of soundscape excerpts A: Library, A: Office, A: LivingRoom, A: Cafe, A: DinnerParty, and A: TrainStation.



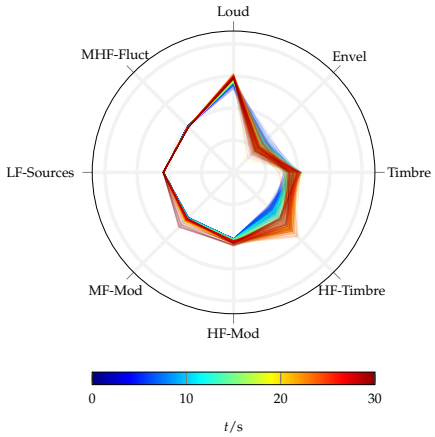
(a) A: FoodCcourt



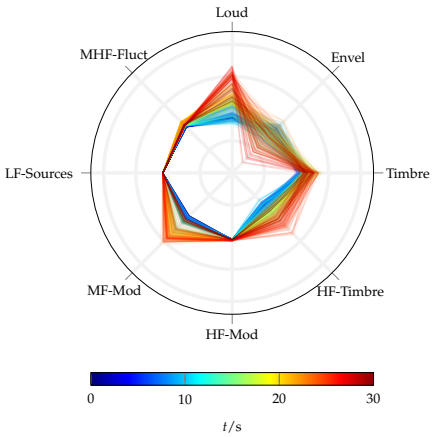
(b) E: Beach



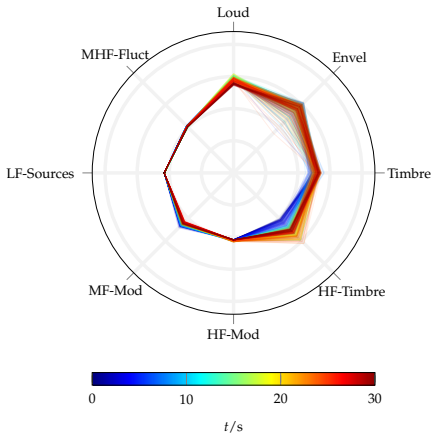
(c) E: Park



(d) E: Pedestrian A

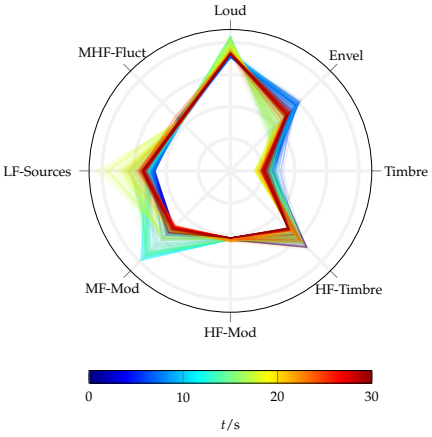


(e) E: Pedestrian B

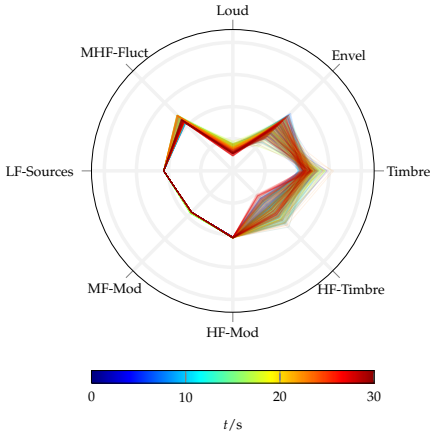


(f) E: Shopping

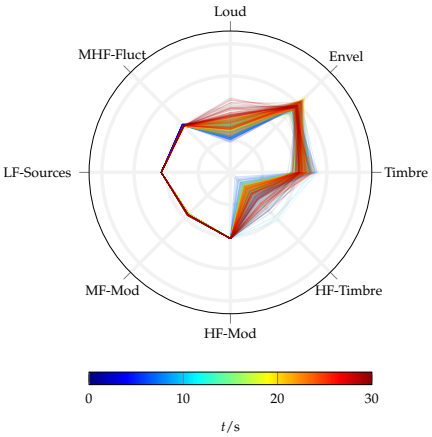
Figure B.5: Fingerprints of soundscape excerpts A: FoodCourt, E: Beach, E: Park, E: Pedestrian A, E: Pedestrian B, and E: Shopping.



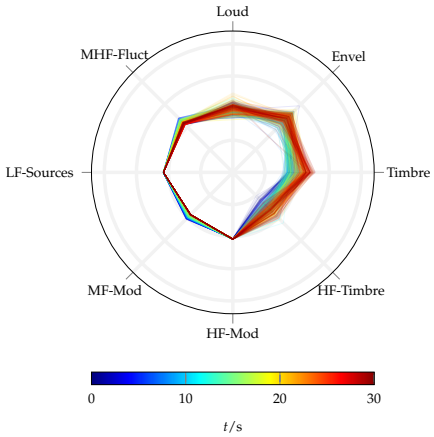
(a) E: TrainStation



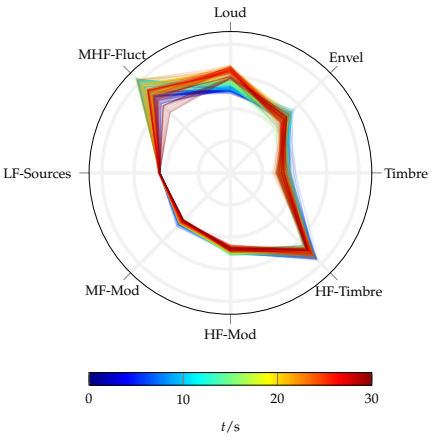
(b) E: Woodland



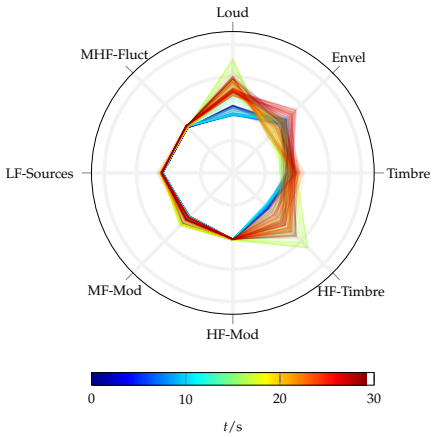
(c) E: Playground



(d) S: Busker

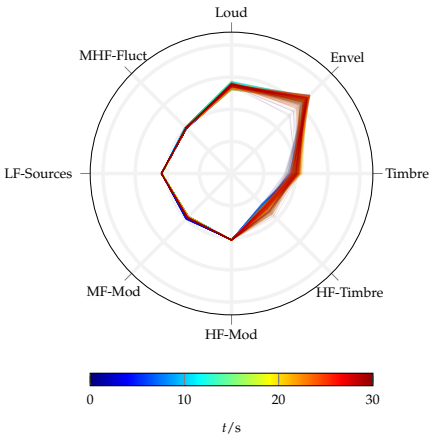


(e) S: Steamtrain



(f) S: Traffic

Figure B.6: Fingerprints of soundscape excerpts E: TrainStation, E: Woodland, S: Playground, S: Busker, S: Steamtrain, and S: Traffic.



(a) S: TrainStation

Figure B.7: Fingerprints of soundscape excerpt S: TrainStation.

C

FURMULAE

INDICATOR CALCULATION

Periodic Modulation Frequency

The periodic amplitude modulation is calculated by means of a peak-finding algorithm that is applied to the spectrum X_{am} that consists of the Fourier transformation of the signal envelope (Hilbert transformation) as in

$$X_{am} = 20 \cdot \log_{10}(\mathcal{F}\{\mathcal{H}\{x\}\}). \tag{C.1}$$

The Fourier spectrum was logarithmically resampled with 96 taps per octave in a frequency range between 0.05 Hz and 20 Hz and the peak-finding algorithm was configured with minimum peak prominence of 6 dB, minimum peak height of $(\max\{X_{am}\} - 24\,\text{dB})$ and minimum peak distance of $1/3$ octaves.

Periodic Modulation Depth

The modulation depth that is associated with the periodic modulation frequency is calculated as described before of the periodic modulation frequency and is scaled such that a depth of 1 corresponds to equal modulation and signal amplitude.

Stochastic Modulation Depth

The stochastic modulation is calculated by substracting up to three periodic modulation envelopes from the original time signal and the modulation depth represents the signal amplitude of this differenced signal.

Spherical Directivity Index

The calculation of the soundfield's directivity index follows Equation C.2

$$DI = 10 \cdot \log_{10}\left(\frac{|p_{ref}|^2}{\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi}|p(\varphi,\theta)|^2 \sin\theta \mathrm{d}\theta \mathrm{d}\varphi}\right), \tag{C.2}$$

where $p_{ref} = \max(|p(\varphi,\theta)|^2)$. The directional sound pressure $p(\varphi,\theta)$ is calculated as plane wave decomposition of the incoming sound with 1° angular resolution.

Vertical Directivity Index

The calculation of the soundfield's vertical directivity index follows Equation C.3

$$DI_v = 10 \cdot \log_{10} \left(\frac{|p_{v,ref}|^2}{\int_{\theta=0}^{\pi} |p_v(\theta)|^2 \sin\theta \mathrm{d}\theta} \right), \tag{C.3}$$

where $p_{v,ref} = \max(|p_v(\theta)|^2)$ and $p_v(\theta) = \int_{\varphi=0}^{2\pi} p(\varphi,\theta)\mathrm{d}\varphi$. The directional sound pressure $p(\varphi,\theta)$ is calculated as plane wave decomposition of the incoming sound with 1° angular resolution.

Horizontal Directivity Index

The calculation of the soundfield's horizontal directivity index follows Equation C.4

$$DI_v = 10 \cdot \log_{10} \left(\frac{|p_{v,ref}|^2}{\int_{\varphi=0}^{2\pi} |p_h(\varphi)|^2 \mathrm{d}\varphi} \right), \tag{C.4}$$

where $p_{h,ref} = \max(|p_h(\varphi)|^2)$ and $p_h(\varphi) = \int_{\theta=0}^{\pi} p(\varphi,\theta)\sin\theta \mathrm{d}\theta$. The directional sound pressure $p(\varphi,\theta)$ is calculated as plane wave decomposition of the incoming sound with 1°angular resolution.

Spherical Pressure Ratio

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation C.5

$$s_p = 20 \cdot \log_{10} \left(\frac{|x_0|^2}{\frac{1}{(N+1)^2} \sum_{n=1}^{N} |x_n|^2} \right) \tag{C.5}$$

where N denotes the maximum order of the Ambisonics signal representation.

Spherical Gradient Ratio

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation C.6

$$s_g = 20 \cdot \log_{10} \left(\frac{\frac{1}{(N+1)^2-1} \sum_{n=1}^{N} |x_n|^2}{\frac{1}{(N+1)^2} \sum_{n=0}^{N} |x_n|^2} \right) \tag{C.6}$$

where N denotes the maximum order of the Ambisonics signal representation.

Spherical Gradient/Pressure Ratio

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation C.7

$$s_{gp} = 20 \cdot \log_{10} \left(\frac{\frac{1}{(N+1)^2-1} \sum_{n=1}^{N} |x_n|^2}{|x_0|^2} \right) \tag{C.7}$$

where N denotes the maximum order of the Ambisonics signal representation.

D

PUBLICATIONS

The following list contains publications for which the author of this thesis is mainly responsible or to which he has contributed. It is sorted by year and divided into contributions directly related to this thesis and publications with other, unrelated topics.

RELATED (FULL PEER REVIEW)

[66] J. Bergner and J. Peissig. "On the identification and assessment of underlying acoustic dimensions of soundscapes". In: *Acta Acustica* 6.46 (Oct. 2022). DOI: 10.1051/aacus/2022042.

[123] J. Bergner, D. Schössow, S. Preihs, and J. Peissig. "Identification of Discriminative Acoustic Dimensions in Stereo, Surround and 3D Music Reproduction". In: *to be published in J. Audio Eng. Soc.* (Oct. 2022). DOI: 10.17743/jaes.2022.0071.

[49] Y. Wycisk, R. Kopiez, J. Bergner, K. Sander, S. Preihs, J. Peissig, and F. Platz. "The Headphone and Loudspeaker Test – Part I: Suggestions for controlling characteristics of playback devices in internet experiments". In: *Behavior Research Methods* (May 2022). DOI: 10.3758/s13428-022-01859-8.

[115] J. Bergner, S. Preihs, R. Hupke, and J. Peissig. "A System for Room Response Equalization of Listening Areas Using Parametric Peak Filters". In: *AES International Conference on Immersive and Interactive Audio*. York, UK, Mar. 2019.

[113] R. Hupke, J. Ordner, J. Bergner, M. Nophut, S. Preihs, and J. Peissig. "Towards a Virtual Audiovisual Environment for Interactive 3D Audio Productions". In: *AES International Conference on Immersive and Interactive Audio*. Mar. 2019.

RELATED (ABSTRACT/NO REVIEW)

[111] J. Bergner, S. Preihs, and J. Peissig. "Perceptual Correlates of Underlying Acoustic Dimensions in Soundscape Assessment". In: *Fortschritte der Akustik - DAGA*. Hamburg, Germany, 2023.

[122] J. Bergner, D. Schössow, S. Preihs, Y. Wycisk, K. Sander, R. Kopiez, and F. Platz. "Analyzing the Degree of Immersion of Music Reproduction by means of Acoustic Fingerprinting". In: *Fortschritte der Akustik - DAGA*. Stuttgart, Germany, Mar. 2022.

[65] J. Bergner, S. Preihs, and J. Peissig. "Soundscape Fingerprinting - Methods and Parameters for Acoustic Assessment". In: *Fortschritte der Akustik - DAGA*. Mar. 2021.

[64] S. Preihs, J. Bergner, D. Schössow, and J. Peissig. "Assessing Wind Turbine Noise Perception by means of Contextual Laboratory and Online Studies". In: *Proc. of the 9th Int. Conf. on Wind Turbine Noise*. Apr. 2021.

[63] S. Preihs, J. Bergner, D. Schössow, and J. Peissig. "On Predicting the Perceived Annoyance of Wind Turbine Sound". In: *Fortschritte der Akustik - DAGA*. Vienna, AUT, Mar. 2021.

[62] D. Schössow, J. Bergner, S. Preihs, and J. Peissig. "Audiovisuelle Laborstudie zur Lästigkeit von WEA-Schall". In: *Fortschritte der Akustik - DAGA*. Hannover, Germany, Mar. 2020.

[60] J. Bergner, S. Preihs, and J. Peissig. "A Virtual Acoustic Environment for Psychoacoustic Assessment of Wind Turbine Noisescapes". In: *Proc. of the International Conference on Spatial Audio, ICSA*. Ilmenau, Germany, Sept. 2019.

[47] J. Bergner, S. Preihs, and J. Peissig. "Investigation on Ecological Validity within Higher Order Ambisonics Reproductions of Wind Turbine Noisescapes". In: *Proc. of INTER-NOISE*. Madrid, Spain, June 2019.

[59] J. Bergner, S. Preihs, and J. Peissig. "On Wind Turbine Noisescape Reproduction for Perceptual Evaluation". In: *Fortschritte der Akustik - DAGA*. Rostock, Germany, Mar. 2019.

[61] S. Preihs, J. Bergner, and J. Peissig. "Ansätze zur binauralen Erweiterung einer Algorithmik zur lästigkeitsbezogenen Analyse und Synthese der Schallemissionen von Windenergieanlagen". In: *Fortschritte der Akustik - DAGA*. Rostock, Germany, Mar. 2019.

UNRELATED

[129] M. Schaab, T. Clauss, J. Bergner, and C. Sladeczek. "Personal Sound Zones: Study on the Threshold of Acceptability in an Automotive Environment". In: *Fortschritte der Akustik - DAGA*. Munich, Germany, Mar. 2018.

[130] J. Bergner, C. Sladeczek, and J. Redlich. "Perception-based Investigations on the Monopole Synthesis for Reproduction of Directional Sound Sources". In: *Fortschritte der Akustik - DAGA*. Kiel, Germany, Mar. 2017.

[131] M. Seideneck, J. Bergner, and C. Sladeczek. "Object-based audio in large scale live sound reinforcement controlled by motion tracking". In: *AES 142nd Convention*. Berlin, Germany, May 2017.

[132] J. Bergner, T. Clauss, A. Zhykhar, C. Sladeczek, and S. Brix. "Application of Wave Field Synthesis in Virtual Acoustic Engineering". In: *Proc. of INTER-NOISE*. Hamburg, Germany, Aug. 2016.

[133] C. Sladeczek, D. Beer, J. Bergner, A. Zhykhar, M. Wolf, and A. Franck. "High-Directional Beamforming with a Miniature Loudspeaker Array". In: *Fortschritte der Akustik - DAGA*. Aachen, Germany, Mar. 2016.

[134] J. Bergner, A. Gasull-Ruiz, C. Sladeczek, and S. Brix. "VISTA4F - Development of an Audiovisual Virtual Reality Test Environment for Automotive". In: *Proc. of the 3rd International Conference on Spatial Audio, ICSA*. Graz, Austria, Sept. 2015.

[135] T. Klouche, T. Samulewicz, and J. Bergner. "Measuring the accuracy of microtonal synthesizers: Pianoteq & Vogue". In: *Fortschritte der Akustik - AIA-DAGA*. Mar. 2013.

[136] T. Klouche, T. Samulewicz, and J. Bergner. "Validation of computational tuning systems". In: *Fortschritte der Akustik - DAGA*. Darmstadt, Germany, Mar. 2012.

E

CURRICULUM VITAE

PERSONAL DETAILS

Name	Bergner
Given Names	Lutz Jakob
Date of birth	01.03.1987
Place of birth	Bremen, Germany

EDUCATION

04/2018 - 04/2023 Dr.-Ing.
Electrical Engineering and Information Technology
Leibniz University Hannover
Thesis: *Towards Soundscape Fingerprinting: Development, Analysis and Assessment of Underlying Acoustic Dimensions to Describe Acoustic Environments*

10/2010 - 06/2014 M.Sc.
Audio Communications and Technology
Technische Universität Berlin
Thesis: *Zur modalen Zerlegung der Richtcharakteristik von Line Array Systemen*

09/2007 - 06/2010 B.Eng.
Media Technology
University of Applied Sciences Emden / Leer
Thesis: *An Opera DVD production: The Audio Production Workflow Based on Pro Tools*

WORK EXPERIENCE

12/2017 - 12/2022 Leibniz University Hannover
Institute for Communications Technology
Hannover, Germany
Research Associate

01/2015 - 11/2017 Technische Universität Ilmenau
Ilmenau, Germany
Research Associate

08/2014 - 11/2017 Fraunhofer Institute for Digital Media Technology (IDMT)
Ilmenau, Germany
Researcher, project manager

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Version submitted for publication as of May 4, 2023.