Bachelor-Thesis

Einsatz und Anpassung von Methoden der Audio-Feature-Extraktion am Beispiel von Indoor Soundscapes

Vorgelegt von: Fabian Rosenthal

Angefertigt im Rahmen der Bachelorprüfung
für den Studiengang Ton und Bild am Fachbereich Medien
der Hochschule Düsseldorf
Bearbeitungszeitraum: 5. November 2021 – 7. Februar 2022

Betreuer: Prof. Dr.-Ing. Jochen Steffens
Zweiter Prüfer: Dipl.-Ing. Siegbert Versümer M. Sc.

https://doi.org/10.20385/opus4-4399
Einsatz und Anpassung von Methoden der Audio-Feature-Extraktion am Beispiel von Indoor Soundscapes

KURZFASSUNG

Im häuslichen Umfeld erleben Menschen abhängig von ihrer Wohnsituation unterschiedlichste Geräuschumgebungen (Indoor Soundscapes), die einen hohen Einfluss auf das Wohlbefinden haben. Im Rahmen einer Feldstudie wurde daher die theoretisch und praktisch relevante Frage untersucht, welche Merkmale der wahrgenommenen Geräusch-umgebungen als besonders ereignisreich oder angenehm bewertet werden. Einhundertfünf Teilnehmer*innen berichteten nach der Experience-Sampling-Methode zeitgesteuert über auftretende Geräuschumgebungen des häuslichen Alltags. Sie bewerteten deren subjektive Wirkung gemäß Soundscape-Standard. Zudem fertigten sie in-situ Audioaufnahmen der Geräuschszenarien an. Die 6594 Tonaufnahmen wurden einer Audioinhaltsanalyse unterzogen und im Zuge dessen wurden vier Featuresets verschiedener Berechnungsansätze extrahiert. Mithilfe der Perzentilen LASSO-Regularisierung wurden lineare gemischte Modelle für Angenehmheit und Ereignisreichtum sowie drei Multilevel-Modelle für Angenehmheit aufgestellt. Die besten Modelle erklären 9 % der Varianz von Angenehmheit und 27 % der Varianz von Ereignisreichtum durch feste Effekte selektierter Prädiktoren aller getesteten Featuresets. Angenehmheit sinkt vor allem, wenn lautheitsbasierte Features hohe Werte zeigen. Ereignisreichtum ist am stärksten abhängig von kurzen Spitzen des C-bewerteten Schalldruckpegels und wird im Vergleich zu Angenehmheit stärker von Zeitschwankungen der Features bestimmt. Durch den Vergleich der Featuresets wird deutlich, dass die Modelleffekte bekannter psychoakustischer Größen durch Hinzufügen von MFCC-Features verbessert werden. Darüber hinaus wird anhand von Multilevel-Modellen gezeigt, dass die Angenehmheit des lautesten Viertels der Tonaufnahmen deutlich besser durch feste Effekte erklärbar ist als der leisere Rest des Datensatzes. Kongruent mit aktuellen Befragungsstudien kann durch Audiofeatures eine Kategorienabhängigkeit der Soundscape-Bewertungen belegt werden: Musik und Sprache werden als angenehmer bewertet, technische Geräusche und Anlagenrauschen hingegen als unangenehmer. Hoher Ereignisreichtum hängt mit menschgemachten Geräuschen (z. B. Poltern, Stuhlrücken) zusammen. Diese Ergebnisse stellen eine wichtige Komponente in der umfassenden Beschreibung komplexer Geräuschumgebungen dar und ihnen kommt in Zeiten der SARS-CoV-2-Pandemie, in denen Wohnraum vermehrt heterogen genutzt werden muss, eine erhöhte Bedeutung zu.

INHALTSVERZEICHNIS

Abkürzungen und Konventionen

.	Dezimaltrennzeichen
GTCC(-N)	Gammatone Cepstral Coefficient; mit N = Rang des Koeffizienten
ICC	Intra Class Coefficient (Intraklassenkorrelation)
IQR	Inter Quartile Range (Interquartilsabstand)
LASSO	Least Absolute Shrinkage and Selection Operator
MFCC(-N)	Mel Frequency Cepstral Coefficient; mit N = Rang des Koeffizienten
MIR	Music Information Retrieval
$R^2_{marginal}$	Bestimmtheitsmaß; Anteil der festen Effekte

Anmerkung zur Verwendung verschiedener Perzentil-Definitionen

Für die Berechnung von Perzentilen legen Statistik und Akustik unterschiedliche Definitionen an. Die vorliegende Arbeit folgt bei der Beschreibung von Zeitreihen der Audiofeatures stets der Definition der Akustik, in der das n-te Perzentil angibt, dass n % der Werte eines der Größe nach geordneten Datensatzes oberhalb dieses Wertes liegen. Allen Anwendungen, die nicht Zeitreihenbeschreibungen sind, wird die statistische Definition zu Grunde gelegt, in der das n-te Perzentil den Wert angibt, für den n % der Daten kleiner oder gleich sind. Die statistische Definition wird für alle abgebildeten Boxplots, den Algorithmus der Perzentilen LASSO-Regularisierung sowie bei Trennung der Multilevel-Teilmengen eingesetzt.

EINLEITUNG

Die unterschiedlichen Geräuschumgebungen im häuslichen Umfeld haben einen hohen Einfluss auf das persönliche Wohlbefinden. Die vorliegende Arbeit befasst sich mit der Beschreibung dieser komplexen akustischen Geräuschumgebungen. Ziel ist es, das Portfolio der bekannten akustischen und psychoakustischen Einflussgrößen auf die wahrgenommene Qualität von Geräuschumgebungen um audiospezifische Merkmale zu erweitern und in Bezug auf Besonderheiten der in Wohnräumen auftretenden Geräusche anzupassen.

Es ist vielfach belegt, dass psychologischer Stress eine Reaktion auf Lärm ist (Gunn et al., 1975; Wolsink et al., 1993; Lercher, 1996; Stallen, 1999) und weitreichende Folgen für die körperliche Gesundheit nach sich ziehen kann (Babisch, 2002; World Health Organization, 2011; Beutel et al., 2016). Auch eine Minderung der kognitiven Leistungsfähigkeit kann nachgewiesen werden (Klatte et al., 2017). Wahrgenommene Qualität akustischer Umgebungen ist aber auch dann messbar, wenn keine psychische oder physische Schädigung durch überaus starken Lärm auftritt: So ist z. B. empirisch belegt, dass leise tonhaltige Geräusche besonders störend wahrgenommen werden (Oliva et al., 2017). Andererseits deuten Studien darauf hin, dass als unangenehm oder störend empfundene Geräuschumgebungen durch positiv wahrgenommene Schallereignisse maskiert werden könnten, die gezielt zugelassen oder hinzugefügt werden (Skoda et al., 2014; Torresin et al., 2019). Eine Modellierung umfassender situativer und persönlicher Einflussfaktoren auf die Wahrnehmung von Geräuschumgebungen in Alltagssituationen wird bei Versümer et al. (2020) vorgestellt. Die Autoren zeigen anhand einer retrospektiven Befragung zu Alltagsgeräuschen, dass die persönliche Befindlichkeit, die Zugehörigkeit von Geräuschen zu bestimmten Kategorien und die persönliche Fähigkeit, Geräusche mental auszublenden, starke Faktoren im Zusammenhang mit der Bewertung von ‚Annoyance' (engl. Ärger, Belästigung) sind. Versümer et al. belegen im Einklang mit früheren Studien (Kuwano et al., 2003), dass in Wohnräumen selbst leisere Installationsgeräusche als sehr störend empfunden werden.

Eine umfassende empirische Betrachtung komplexer Geräuschumgebungen ist daher für alle Gewerke sowie Geräte- und Anlagenhersteller, die in die Gestaltung von (Wohn-)Innenräumen involviert sind, von großer Bedeutung: Erkenntnisse können helfen, die (Bau-)Vor-

haben, Produkte und Konzepte im Hinblick auf die Lebensqualität der Nutzer*innen und Verbraucher*innen zu optimieren (Berglund, 2006; Torresin et al., 2020).

Um diese Erkenntnisse zu gewinnen, wurden in der hier präsentierten Feldstudie 6594 Tonaufnahmen der Geräuschumgebungen von 105 Proband*innen einer umfassenden und systematischen Audioinhaltsanalyse unterzogen. Mit Methoden, die in den Fachgebieten (Psycho-)Akustik, Sprachforschung und Music Information Retrieval (MIR) gute Ergebnisse zeigen, wurden Audiomerkmale, sog. Audiofeatures, aus den Tonaufnahmen extrahiert und selektiert. Eine solche empirische Studie zu akustischen, psychoakustischen und klanglichen Einfluss-faktoren auf die wahrgenommene Qualität von Geräuschumgebungen in Wohnräumen liegt gegenwärtig noch nicht vor.

SOUNDSCAPE

Die mehrdimensionale Wirkung komplexer Geräuschumgebungen des Alltags auf den Menschen lässt sich mit dem sog. Soundscape-Ansatz beforschen. Laut Rahmenkonzept der DIN ISO 12913-1 (DIN Deutsches Institut für Normung e.V., 2018) beschreibt ‚Soundscape' ein Wahrnehmungskonstrukt, das in Beziehung zu einem physikalischen Phänomen, nämlich der akustischen Umgebung, gesetzt wird. Soundscapes existieren durch die menschliche Wahrnehmung der akustischen Umgebung und sind angelehnt an das Konzept ‚Landscape', das ebenso auf der menschlichen Wahrnehmung basiert.

Der Begriff wurde in den 1970er Jahren im Kontext des ‚World Soundscape Project' von Schafer (1977) als abstrakte Reaktion auf die zunehmende Lärmbelastung in den Städten etabliert. Truax stellt die entscheidenden Kerngedanken heraus: Schafer gehe weg von einer reinen Haltung gegen Lärm hin zu einem positiveren Ansatz, der Hörer*innen ins Zentrum rücke (Truax, 2019). Soundscape-Forschung geht in ihren methodischen Ansätzen und Untersuchungsformen letztlich auf die im ‚World Soundscape Project' etablierten Ideen zurück (Truax & Barrett, 2011; Brambilla et al., 2013; Jeon & Hong, 2015).

Botteldooren und Coensel (2009) setzen sich ebenfalls mit dem hörer*innenzentrierten Ansatz auseinander und beschreiben, dass das zuvor genannte Interaktionselement in Soundscapes mit der Interaktion bei Wahrnehmung von Musik oder einer Landschaft vergleichbar sei. Ein zentrales Moment dieser Interaktion sei die Aufmerksamkeit der Hörer*innen (Botteldooren & Coensel, 2009). Daher beschäftigen sich viele Studien damit, wie Aufmerksamkeit,

Auffälligkeit von Geräuschen und Schallquellenerkennung zusammenspielen und ein mentales Abbild der Geräuschumgebung erzeugen (Duangudom & Anderson, 2007; Botteldooren & Coensel, 2009; Coensel, 2010; Oldoni et al., 2013; Kaya & Elhilali, 2014, 2017; Selzer et al., 2017).

Das Rahmenkonzept des Soundscape-Standards ISO 12913-1 (DIN Deutsches Institut für Normung e.V., 2018) ordnet die Hörempfindung selbst als Element des Wahrnehmungskonstrukts Soundscape in das Spannungsfeld aus akustischer Umgebung und Kontext ein, der wiederum auch die Schallquellen der akustischen Umgebung, die folgende Interpretation der Hörempfindung und die Reaktion auf selbige mitbestimmt. Die Reaktionen sind ihrerseits wieder mit dem Kontext rückgekoppelt, was den Interaktionsansatz von Soundscape unterstreicht.

Weiterhin kann Soundscape drei besondere Hörstile implizieren, die bei Botteldooren et al. (2011) diskutiert werden: Hörer*innen können demnach 1. in Bereitschaft auf ein (un-)erwartbares, bekanntes oder wichtiges Schallereignis oder 2. in bestimmten Kontexten auf der Suche nach ganz bestimmten Geräuschen sein. Zudem kann 3. „Hören einer Geschichte" stattfinden, d. h. eine Fokussierung auf eine ganz bestimmte Sequenz in einer Vielzahl von Geräuschen der Umgebung. Das Hören würde so zu einer multisensorischen Erfahrung mit entsprechenden Ansprüchen an die Forschung (Botteldooren et al., 2011).

Studien gewinnen mithilfe des Soundscape-Ansatzes Erkenntnisse über die Vielzahl von Emotionen und Affekten, die von Proband*innen im Zusammenhang mit der akustischen Umgebung wahrgenommen oder erlebt werden (Västfjäll et al., 2003; Västfjäll, 2003; Berglund, 2006; Nilsson et al., 2007; Brambilla et al., 2013; Aletta et al., 2016; Versümer et al., 2020). Die gefundenen Affekte lassen sich auf den Dimensionen Arousal und Valence des von Russel vorgestellten ‚Circumplex Model of Affect' auftragen (Russell, 1980; Russell et al., 1981; Västfjäll, 2003; Axelsson et al., 2010). Bezogen auf wahrgenommene Qualität von Soundscapes lässt sich die Dimension Valence mit Angenehmheit und die Dimension Arousal mit Ereignisreichtum übersetzen.

Im Folgenden wird eine Übersicht gegeben, wie sich Soundscape-Studien mit den Dimensionen der wahrgenommenen Qualität auseinandersetzen. Zuerst werden zwei Laborstudien vorgestellt, die sich mit der Beweisbarkeit der Dimensionen und den Zusammenhängen mit

psychoakustischen Messgrößen beschäftigen. Nachfolgend werden verschiedene empirische Studien betrachtet, die Soundscape-Dimensionen für Urban Soundscapes beforschen, und abschließend wird der Bereich der Studien beleuchtet, die sich mit Indoor Soundscapes auseinandersetzen.

Mit Axelsson et al. (2010) wurde eine entscheidende Arbeit vorgelegt, die Fragebögen zur Erhebung der Affekte der wahrgenommenen Qualität der Soundscapes durch Dimensionsreduktion vereinheitlicht und vereinfacht: Die zugrundeliegenden Dimensionen Angenehmheit und Ereignisreichtum wurden durch eine Hauptkomponentenanalyse über 116 Items eines Laborversuchs berechnet. Der benötigte Fragebogen konnte somit reduziert werden. Dieser und die dazugehörigen Berechnungsempfehlungen für Angenehmheit und Ereignisreichtum wurden in den Soundscape-Standard ISO 12913-3 aufgenommen (DIN Deutsches Institut für Normung e.V., 2021). In der oben genannten Studie untersuchten Axelsson et al. (2010) überdies, welche akustischen und psychoakustischen Messgrößen der Soundscapes mit den gefundenen Wahrnehmungsdimensionen korrelieren. Für Angenehmheit wurden negative Korrelationen mit dem A-bewerteten äquivalenten Dauerschalldruckpegel (L_{Aeq}), dem zehnten Perzentil der Lautheit (N_{10}) sowie der zeitlichen Variabilitäten von Schalldruckpegel und Lautheit (L_{A10-90} bzw. N_{10-90}) berichtet, während für Ereignisreichtum positive Korrelationen mit dem äquivalenten Dauerschalldruckpegel (L_{Aeq}) und dem zehnten Perzentil der Lautheit (N_{10}) nachgewiesen werden konnten.

Ein methodisch gänzlich anderer Ansatz für die Aufklärung von in Soundscapes zugrundeliegenden Dimensionen ist bei Aletta et al. (2017) zu finden: Proband*innen sortierten gedruckte Spektrogramme von Soundscapes anhand des Aussehens. Darüber wurde ein Clustering erzeugt und es wurden drei Dimensionen von Soundscapes abgeleitet. Aletta et al. zeigen weiterhin, dass keine über die getesteten Soundscapes gemessene (akustische oder psychoakustische) Messgröße allein die gefundenen Dimensionen hinreichend vorhersagt. Sie ziehen den Schluss, dass komplexe Soundscapes ebenso komplexe Vorhersagemodelle erforderlich machen.

Für den Bereich des urbanen Raums wurde der Soundscape-Ansatz bereits vielfach eingesetzt, um zu verstehen, wie die akustischen Umgebungen der Stadt auf Hörende wirken können. Dabei lassen sich die Studien in reine Befragungsstudien, Simulationsstudien und empirische Studien mit Tonaufzeichnung gruppieren.

Ein Großteil der Studien beschäftigt sich mit der Erhebung und Auswertung von in-situ beantworteten Fragebögen (z. B. durch Passanten) ohne Tonaufzeichnungen jeder individuellen Soundscape pro Observation (Tardieu et al., 2008; Brocollini et al., 2010; Hong & Jeon, 2015; Yilmazer & Acun, 2018; Tarlao et al., 2021). Ebenso gibt es reine Methoden- und Simulationsstudien ohne empirische Stichprobe oder Situationsbezug (Botteldooren et al., 2011; Filipan et al., 2015; Kang et al., 2016). Park et al. (2014) stellen eine urbane Kartierung durch ein freies Netz aus kleinen netzwerkbasierten Stationen vor, die ständig umgebende Soundscapes messen, aufzeichnen und an einen Server senden. Das Projekt ist plattformübergreifend angelegt, sodass sowohl Forschende mit verschiedenen technischen Hintergründen als auch interessierte Laien teilnehmen können. Weitere Studien befassen sich z. B. mit der automatischen Klassifikation von Soundscapes bestehender Datenbanken (Salamon & Bello, 2015).

Eine für die vorliegende Arbeit zentrale Studie ist Ricciardi et al. (2015). Erstmalig zeichneten Proband*innen in-situ Tonaufnahmen von Urban Soundscapes auf Smartphones auf. Dabei haben die Teilnehmer*innen von der Studie vorgegebene Orte über einen längeren Zeitraum abgearbeitet, wodurch über ähnliche Observationen (Tageszeit) des gleichen Orts gemittelt werden konnte. Allerdings verspielt die Studie dadurch die Möglichkeit, dass Proband*innen ihre typischen, alltäglich erlebten Soundscapes berichten. Als Methode, mit der diese typischen Alltagsstichproben erhoben werden könnten, schlagen Steffens et al. (2017) die Experience Sampling Method vor.

Im Bereich der Indoor Soundscapes entscheiden sich Studien zumeist für Räume ganz bestimmter Funktion (Yilmazer & Bora, 2017; Thomas et al., 2019; Wang et al., 2020). Durch Interpretation der Ergebnisse können dann konkrete Verbesserungsvorschläge gemacht werden: Wang et al. (2020) empfehlen anhand ihrer Studienergebnisse von Soundscapes in Bahnhöfen z. B. bauliche Trennung von Warte- und Transitbereichen zur Verbesserung der wahrgenommenen Qualität.

In seiner Masterthesis untersucht Orhan (2019) Indoor Soundscapes in verschiedenen Museen und kombiniert die Befragung mit akustischen Messgrößen zu den relevanten Räumen. Ähnlich wie bei der Simulationsstudie von Jeon et al. (2022) wird aber nur eine sehr geringe Anzahl an akustischen Messgrößen auf Effekte getestet.

Ein interessantes Ergebnis ihrer Laborstudie präsentieren Dedieu, Lavandier et al. (2019): In der Stichprobe gab es zwei Gruppen von Proband*innen: Eine grundsätzlich geräuschempfindlichere Gruppe mit Bevorzugung von Naturgeräuschen und Ablehnung gegenüber Geräuschen aus benachbarten Wohnungen sowie eine zweite etwas weniger lärmempfindliche Gruppe, die gegenüber verschiedenen Geräuschquellen im Mittel toleranter ist.

Während der SARS-CoV-2-Pandemie haben gesetzliche Eindämmungsmaßnahmen in Form der sog. Lockdowns zeitweise die Freizeitgestaltung sowie das Arbeits- und Wohnverhalten der Bevölkerung ganz erheblich beeinflusst. Der Beforschung von Indoor Soundscapes im Wohnraum kommt im Lichte dessen eine neue, größere Bedeutung zu. Torresin et al. stellen eine zweiteilige Onlinestudie vor, die sich detailliert mit dem Zusammenhang von Angenehmheit und Ereignisreichtum der Indoor Soundscapes und dem Wohlbefinden Londoner Teilnehmer*innen während des verhängten Lockdowns auseinandersetzt (Torresin et al., 2021; Torresin et al., 2022). Sie können in den Daten starke negative Effekte unangenehmer Soundscapes auf das Wohlbefinden nachweisen.

Eine große aktuelle Befragungs-Studie berichtet unter Einbeziehung demografischer Faktoren ähnliche Zusammenhänge (Erfanian et al., 2021): Die Ergebnisse zeigen für angenehme Soundscapes positive Effekte und für ereignisreiche Soundscapes negative Effekte auf das Wohlbefinden.

Audioinhaltsanalyse und Feature-Extraktion

Die vorliegende Arbeit befasst sich mit der Aufklärung von Angenehmheit und Ereignisreichtum für den Bereich Indoor Soundscapes durch eine umfassende Audioinhaltsanalyse von empirischen Soundscape-Tonaufnahmen. Sie widmet sich der Frage, welche Effekte bestimmter Geräuscheigenschaften aufgezeichneter Geräuschumgebungen auf die wahrgenommenen und subjektiv bewerteten Dimensionen Angenehmheit und Ereignisreichtum nachgewiesen werden können. Ausgehend von den Grundbegriffen der Audioinhaltsanalyse werden anhand zentraler Studien relevante Merkmale vorgestellt, die deutlich machen, warum die Audioinhaltsanalyse auch in der Soundscape-Forschung eine geeignete Methode ist.

Lerch (2012) gibt eine weit gefasste Definition von Audioinhaltsanalyse, in der sie als Extraktion von Information aus dem Audiosignal wie z. B. der digital gespeicherten Musikaufnahme selbst beschrieben wird. Mitrović et al. (2010) schreiben der Identifikation und Extraktion von

relevanten oder angebrachten inhaltsbasierten Features der Audioinhaltsanalyse eine zentrale Rolle zu. Die extrahierten Metadaten können alle Informationen der Audiodateien enthalten, die für die Repräsentation oder Erklärung des rohen Audiosignals von Bedeutung sind. Mitrovic et al. (2010) bringen zudem den Aspekt einer kompakten, expressiven und leicht durch Maschinen zu verarbeitenden Beschreibung des Signals ein.

Begrifflich ist das Audiofeature der Merkmalsträger an sich. Die sog. Audiofeature-Extraktion, die Erhebung der Features, ist der Audioinhaltsanalyse hierarchisch untergeordnet. Dennoch werden Audioinhaltsanalyse und Audiofeature-Exktraktion in der Literatur oft synonym verwendet.

Besonders seit Aufkommen von Musik-Streaming und Musikempfehlungsalgorithmen ist die Mehrheit der Anwendungsfälle der Audiofeature-Extraktion eher im Bereich der Klassifikation von Musik, insbesondere Musikgenres, zu sehen (Pohle et al., 2009; Seyerlehner et al., 2010; Fu et al., 2011b, 2011a; Schlüter & Osendorfer, 2011; McFee et al., 2012). Aber auch rein geräuschbezogene Studien setzten sich mit der Klassifikation akustischer Szenen auseinander (Gaunard et al., 1998; Povinelli et al., 2006; Giannoulis et al., 2013; Barchiesi et al., 2015; Alías et al., 2016).

Audiofeature-Extraktion ist in allen technischen und wissenschaftlichen Bereichen relevant, die sich für die Ähnlichkeit, Klassifizierung oder Bewertung der akustischen Eigenschaften von Geräuschen, Sprache oder Musik interessieren (Mitrović et al., 2010): Studien der Medizinforschung setzen Audiofeature-Extraktion ein, um bspw. Tonaufnahmen von Schluck- oder Lungengeräuschen zu klassifizieren (Aboofazeli & Moussavi, 2005; Ronzhin et al., 2016, Elsetrønning et al. 2020). In der Biologie werden ähnliche Methoden z. B. für die Klassifikation von Tier-stimmen genutzt (Benko & Perc, 2009; Pieretti et al., 2011; Lee et al., 2013).

Typischerweise werden die Daten vor der Audiofeature-Extraktion in der zeitlichen Auflösung reduziert, indem das Audiomaterial in Zeitblöcke unterteilt wird. Diese Methode wird unter dem Begriff Bag-of-Frames zusammengefasst und findet in einigen musikbezogenen Studien Anwendung (Nam et al., 2012; Wülfing & Riedmiller, 2012). In der Geräuschforschung kam der Bag-of-Frames-Ansatz in der bereits genannten Studie von Gaunard et al. (1998) und auch in mehreren Studien zu Urban Soundscapes zum Einsatz (Aucouturier et al., 2007; Lagrange et al., 2015). Wirksame Einsatzmöglichkeiten von Bag-of-Frames im Zusammenspiel mit ausge-

wählten Klassifikationsalgorithmen stellen Giannloulis et al. (2013) in ihrer Meta-Studie zur Klassifikation von akustischen Szenen vor.

Wichtige grundlegende Informationen über die Audiodaten können mit basalen Funktionen wie Energiegehalt und Frequenz der Nulldurchgänge extrahiert werden (Muhammad & Alghathbar, 2009). Aussagen über das Frequenzspektrum eines vorliegenden Signals lassen sich z. B. durch statistische Lagebeschreibung des Spektrums tätigen. Die statistischen Momente Schiefe und Wölbung können die spektrale Beschaffenheit repräsentieren und angeben, ob Pegel tiefer Frequenzen gegenüber hohen Frequenzen überwiegen oder ob ein-zelne Bänder herausstechen (Lerch, 2012). Andere Repräsentationen des Spektrums müssen erst über eine Verkettung von Funktionen berechnet werden: Für die Feature-Gruppe der Cepstral Coefficients muss das Signal mehrfach transformiert werden (Bogert & Ossanna, 1966; Childers et al., 1977; Oppenheim & Schafer, 2004). Die Anwendung bestimmter Filterbänke, wie z. B. der Mel-Skala oder der Gammatone-Filterbank, erlaubt dann, das Frequenzspektrum in Bändern kompakt darzustellen (Li et al., 2001; McKinney & Breebaart, 2003; Lee et al., 2009; Muhammad & Alghathbar, 2009; Yin et al., 2011; Fan et al., 2015).

Wenn eine inhaltsgetriebene Selektion der zahlreichen Features z. B. aus Komplexitätsgründen nicht möglich ist, muss nach der Feature-Extraktion eine automatisierte oder algorithmische Bewertung darüber erfolgen, welche Audiofeatures im Sinne der Problemstellung relevante Informationsträger sind (Guyon & Elisseeff, 2003). Wenn das grundlegende Ziel der Audioinhaltsanalyse Klassifikation ist, stehen eine Vielzahl an Algorithmen zur Wahl, die in den Feature-Daten z. B. nach Mustern entsprechend einem Codebook suchen (Lee et al., 2007; Rabaoui et al., 2008; Seyerlehner et al., 2008; Weiss & Bello, 2010, Lazaro et al. ,2017). Diese sind aber für Regressionsprobleme ungeeignet (Barchiesi et al., 2015). Relevante Features für die Regressionsmodellierung müssen dann über spezielle Methoden selektiert werden (Guyon & Elisseeff, 2003; Hastie et al., 2009; Bountourakis et al., 2015). Dabei entscheiden sich Studien z. B. zwischen schrittweiser Regression (Aletta & Kang, 2018; Lepa et al., 2020; Erfanian et al., 2021) oder Regularisierung mittels Least Absolute Shrinkage Operator (LASSO) (Gauthier et al., 2017; Rahimi et al., 2017; Versümer et al., 2020). Das letztgenannte Verfahren ist für die Methode der vorliegenden Arbeit relevant, weil es eine höhere Generalisierbarkeit von Modellen erreichen kann (Roberts & Nowak, 2014).

ZIELSETZUNG UND HYPOTHESEN

Die vielfältigen Befunde der Geräusch- und Soundscapeforschung zu Einflüssen von akustischen Messgrößen auf die wahrgenommene Hörempfindung und die dargelegten Möglichkeiten der Audioinhaltsanalyse lassen erwarten, dass eine Verknüpfung des Soundscape-Ansatzes mit in-situ Befragungsstil und Audioinhaltsanalyse das Verständnis von alltäglichen Geräuschumgebungen maßgeblich vertiefen kann (Steffens et al., 2017).

Aufgrund dessen sollen die Indoor Soundscapes durch eine kontrollierte in-situ Befragung mit Einsatz der Experience Sampling Methode beforscht werden. Dabei soll der Soundscape-Ansatz konsequent und umfassend mit den Analysemethoden der Audiofeature-Extraktion und den Möglichkeiten der Selektion verknüpft werden.

Der Schwerpunkt der Arbeit wird auf die explorative Beforschung der Audiofeatures gesetzt. Dennoch sollen ausgehend von den Ergebnissen der ähnlichen Studien folgende Forschungshypothesen überprüft werden:

1. Bewerteter Ereignisreichtum hängt positiv von hoher zeitlicher Schwankung in den Feature-Wertereihen ab. (H1)
2. Die Beschreibungen der Feature-Zeitreihen klären mehr Varianz von bewertetem Ereignisreichtum auf als die Durchschnittswerte der Features. (H2)
3. Die bewertete Angenehmheit hängt stark negativ von der psychoakustischen Lautheit ab. (H3)

Darüber hinaus sollen folgende offene Forschungsfragen explorativ untersucht werden:

1. Welche Audiomerkmale sind in Bezug auf die subjektive Bewertung von Angenehmheit und Ereignisreichtum relevant? (F1)
2. Wie gut sind die auf Einzelgeräusche abgestimmten psychoakustischen Größen geeignet, die komplexen, teils stark überlagerten Geräuschszenen zu beschreiben? (F2)
3. Lassen sich die beiden Dimensionen Angenehmheit und Ereignisreichtum für komplexe Indoor Soundscapes gleichermaßen gut modellieren? (F3)

METHODE

FELDSTUDIE

Die vorliegende Arbeit untersucht die Zusammenhänge von Geräuscheigenschaften der Indoor Soundscapes und der wahrgenommenen Qualität in einer Feldstudie, deren Daten zwischen April 2021 und Oktober 2021 erhoben wurden.

Einhundertfünf Proband*innen haben nach der Experience Sampling Method (Kubey et al., 1996; Hektner et al., 2007; Larson & Csikszentmihalyi, 2014; Steffens et al., 2017; van Berkel et al., 2018) über einen Zeitraum von jeweils zehn Tagen in-situ in ihrem häuslichen Umfeld 15-sekündige, binaurale Tonaufnahmen der Soundscapes durchgeführt und zu jeder getätigten Tonaufnahme Bewertungen der Soundscapes, der persönlichen Befindlichkeit und des situativen Kontextes abgegeben.

Das eingesetzte technische Equipment umfasste baugleiche Smartphones mit der für die vorliegende Studie individualisierten Anwendung *movisens xs* (movisens GmbH, 2021) sowie an der Hochschule Düsseldorf entwickelte Tonaufnahmegeräte mit binauralen Mikrofonen. Vor jeder Ausgabe der Geräte wurden Kalibrierungsaufnahmen angefertigt, sodass Ein-flüsse durch Batterieschwankungen später kompensiert werden konnten.

Die Teilnehmenden bekamen eine persönliche Studieninstruktion und ihre Teilnahme wurde entsprechend der Anzahl vollständig abgegebener Aufnahmen und ESM-Bewertungen nach einem vorher kommunizierten Schema mit mindestens 20 Euro und in einer Maximalhöhe von 100 Euro vergütet.

ESM-BEFRAGUNG UND TONAUFZEICHNUNG

Die Proband*innen konnten das Zeitfenster für die Ausübung der Studie über den individuellen Teilnahmezeitraum hinweg täglich frei in *movisens xs* wählen und tagsüber sowie in bereits angebrochenen Zeiträumen jederzeit korrigieren. Während der frei eingestellten Zeiträume löste die App stündlich mit einer Randomisierung von ± zehn Minuten eine Befragung aus (reguläre ESM-Befragung). Die Proband*innen hatten dann die Möglichkeit, die aktuelle Befragung um einige Minuten zu verschieben oder ganz zu verweigern. Über diese automatisierte ESM-Befragung hinaus bestand die Möglichkeit, Bewertungen initiativ abzugeben. Der

nächste automatische Befragungszeitpunkt wurde dann erst wieder eine Stunde (mit Randomisierung) nach der initiativen Bewertung ausgelöst.

Bei regulärer wie initiativer Befragung wurden die Proband*innen zuerst zur Anfertigung der Tonaufnahmen aufgefordert. Sie waren instruiert, dass sich alle nachfolgenden Fragen nur auf Geräusche beziehen, die (zeitlich) auf der Aufnahme enthalten sein können. Es bestand die Möglichkeit, die zuletzt angefertigte Tonaufnahme zu löschen und zu wiederholen. Die Teilnehmenden waren aufgefordert, sich während der Tonaufnahme möglichst still zu verhalten. Die in die Ohren einzuhängenden Mikrofone haben die akustische Umgebung kopfbezogen und ohne Körperschall durch Greifen oder Halten aufgezeichnet.

Fragebogen

Nach der Tonaufnahme haben die Proband*innen sechs Kurzfragebögen zu persönlicher Befindlichkeit und Situation, acht Fragen zum auffälligsten Geräusch inklusive der Kategorie des auffälligsten Geräusches und zu acht Items des Soundscape-Standards beantwortet (Tabelle 1). Die vorliegende Arbeit nutzt die Antworten zu letztgenanntem Fragebogen und zur Frage, welcher Kategorie das auffälligste Geräusch (im Vordergrund) zugehörig ist.

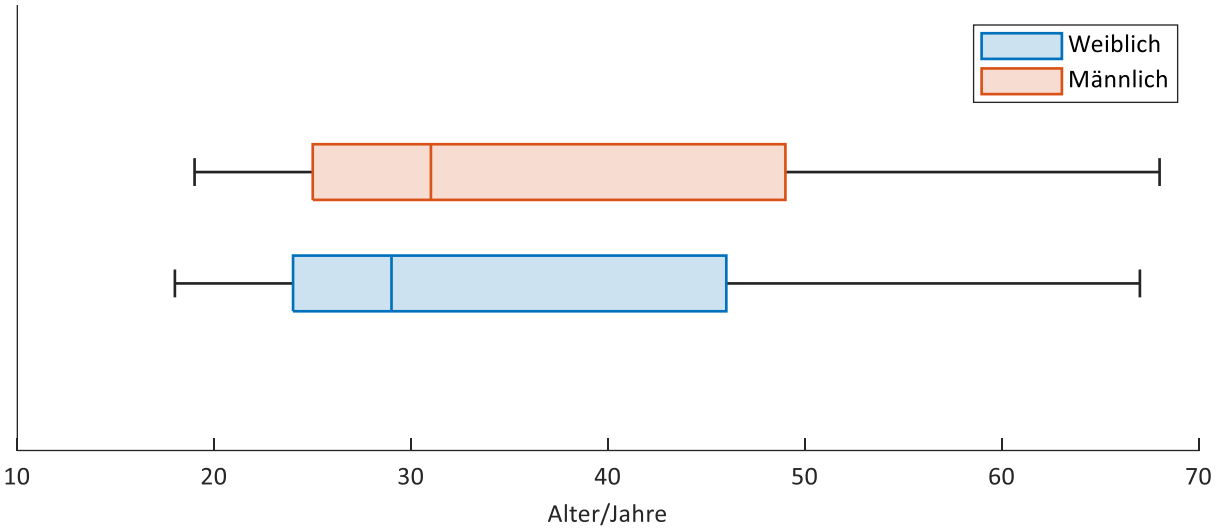
TABELLE 1: ITEMS (A-H) UND SKALA (1-5) ZUR WAHRGENOMMENEN SOUNDSCAPE-QUALITÄT NACH „C.3.1.3. QUESTIONNAIRE PART 2“ DES ISO 12913-2 (ISO INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2018) DER ANTWORT-SKALA WURDEN DIE WERTE 0 BIS 4 ZUGEORDNET.

„Wie sehr stimmen Sie zu, dass die Geräuschumgebung inklusive des auffälligsten Geräusches folgendes ist?“	
a) angenehm	1) Stimme nicht zu
b) chaotisch, hektisch	2) Stimme eher nicht zu
c) lebendig, abwechslungsreich	3) Stimme weder zu noch nicht zu
d) ereignisarm, statisch	4) Stimme eher zu
e) ruhig, erholsam	*5)* Stimme zu
f) störend, lästig	
g) ereignisreich, dynamisch	
h) monoton, eintönig	

Stichprobe

Einhundertfünf Proband*innen konnten für die vorliegende Arbeit berücksichtigt werden. Sie waren im Durschnitt 35.97 Jahre alt (SD = 13.99 J). Siebenundfünfzig Personen waren weiblich (54.29 %) und 48 männlich (45.71 %). Altersunterschiede zwischen den Geschlechtern sind den Boxplots in Abbildung 1 zu entnehmen. Die mittlere Haushaltsgröße betrug zwei Personen (Mean = 2.13; SD = 0.91).

Abbildung 1: Boxplot des Alters der Proband*innen in Jahren gruppiert nach Geschlecht (Innenlinie in Box = Median; Breite der Box = IQR; Box-Außenkanten = 25 % bzw. 75 % Quartil; Whisker = Minima/Maxima mit höchstens 1.5 IQR Abstand zur Box)



Datensatz

Der Datensatz besteht aus zwei Teilen: Der per *movisens xs* erhobene Teil wird im Folgenden ESM-Datensatz genannt, der zweite Teil, die angefertigten Tonaufnahmen, wird als Audio-datensatz bezeichnet.

ESM-Datensatz

Der ESM-Datensatz wurde nach vollständig abgegebenen Bewertungen gefiltert und die Items aus Tabelle 1 gemäß des Soundscape-Standards ISO 12913-3 (DIN Deutsches Institut für Normung e.V., 2021) nach Gleichung 1 und Gleichung 2 zusammengefasst.

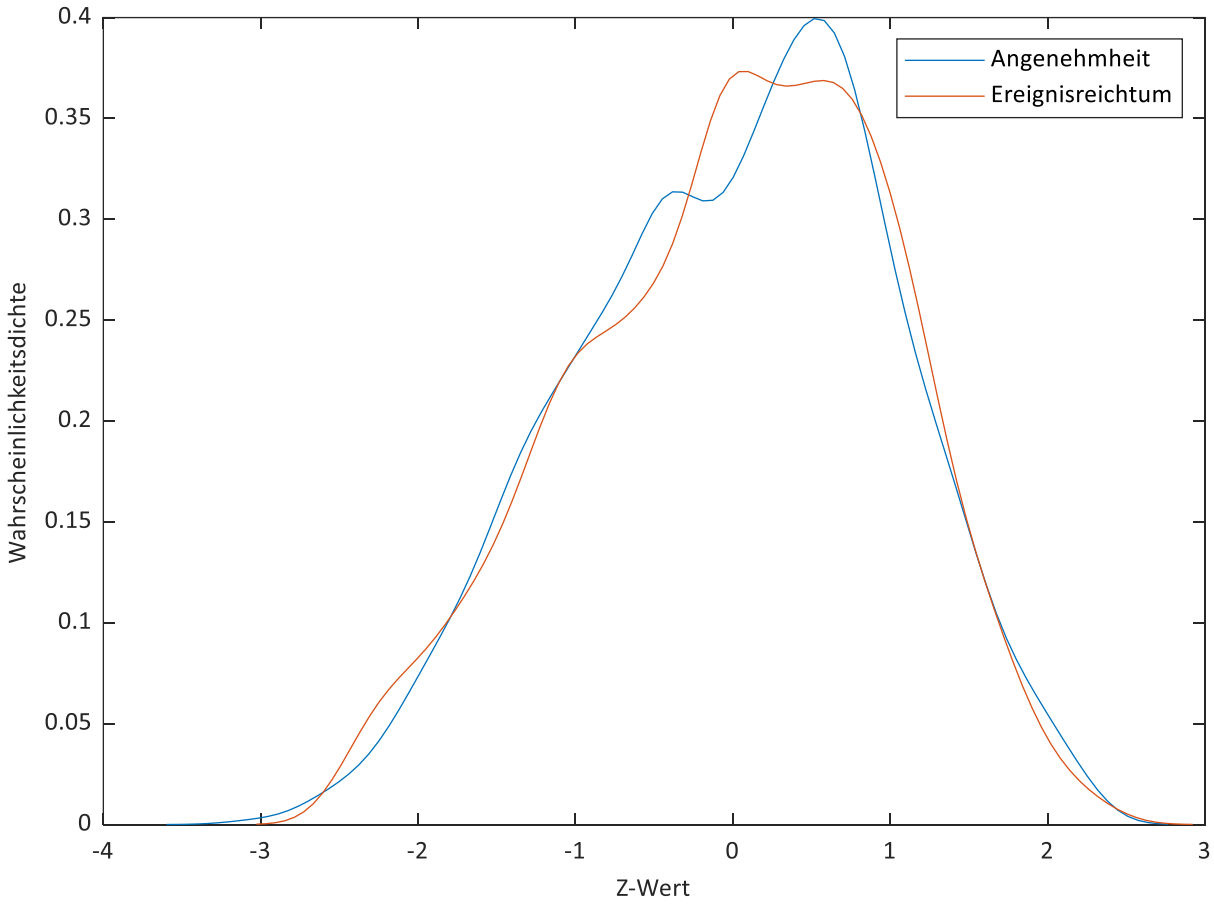
GLEICHUNG 1 $$\boldsymbol{Angenehmheit = (a - s) + \cos 45° \cdot (r - ch) + \cos 45° \cdot (l - m)}$$

GLEICHUNG 2 $$\boldsymbol{Ereignisreichtum = (e - ea) + \cos 45° \cdot (ch - r) + \cos 45° \cdot (l - m)}$$

MIT a = ANGENEHM; s = STÖREND; r = RUHIG; ch = CHAOTISCH; er = EREIGNISREICH; ea = EREIGNISARM; l = LEBHAFT; m = MONOTON

Die resultierenden Werte der Zielgrößen Angenehmheit und Ereignisreichtum wurden Z-standardisiert. Die Verteilung dieser beiden Zielgrößen kann der Abbildung 2 entnommen werden. Darüber hinaus liefert Anhang A weiteres Material zur Verteilung der Zielgrößen (Abbildung A1) und zur Statistik der durchschnittlichen, für jede*n Proband*in gemittelten, Bewertungen von Ereignisreichtum und Angenehmheit gruppiert nach Geschlecht (Abbildung A2).

ABBILDUNG 2: PLOT DER WAHRSCHEINLICHKEITSDICHTEFUNKTIONEN (MITTELS KERNEL-GLÄTTUNG GESCHÄTZTE WERTE) ÜBER DIE SKALA DER ZIELGRÖßEN ANGENEHMHEIT UND EREIGNISREICHTUM IN Z-WERTEN (BANDBREITE DER GLÄTTUNG JEWEILS 0.19; DIE ZIELGRÖßEN HABEN DIE WÖLBUNGEN 2.49 (ANGENEHMHEIT) UND 2.46 (EREIGNISREICHTUM)).



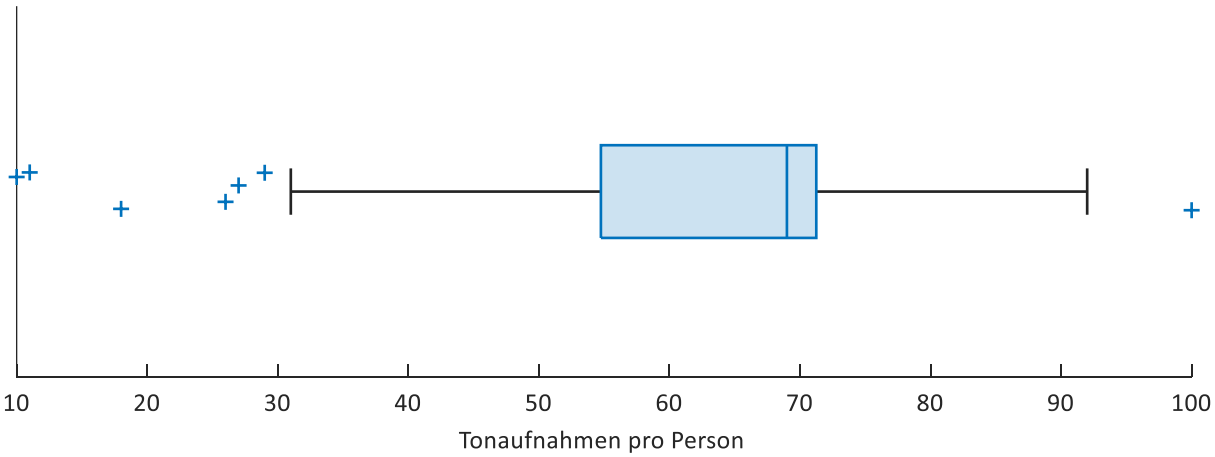
Zuordnung der Datensätze

Da die Aufnahmegeräte von den Smartphones unabhängig waren, mussten die Tonaufnahmen den zugehörigen Observationen im ESM-Datensatz in der Aufbereitung zugeordnet werden. Einem ESM-Zeitpunkt wurden diejenigen Aufnahmen zugeordnet, die sechs Minuten vor der Bewertung sowie bis zum Abschluss der Bewertung angefertigt wurden. Dieses durch Versuch und Irrtum bestimmte Intervall erzeugte die meisten eineindeutigen Zuordnungen. Lagen nicht eineindeutige Zuordnungen vor, z. B. Doppelzuordnungen mehrerer Aufnahmen in demselben Zeitintervall, wurden die Aufnahmen auf etwaige Störungen hin überprüft. Lag keine wahrnehmbare Störung vor, wurde die neuste der angefertigten Tonaufnahmen zugeordnet.

Audiodatensatz (Deskriptive Statistik)

Der Audiodatensatz umfasst 6594 Aufnahmen von je 15 Sekunden Dauer (PCM; Wortbreite = 16 Bit; Abtastfrequenz = 44.1 kHz). Die Audiogesamtdauer beträgt somit 27 Stunden, 28 Minuten und 30 Sekunden. Durchschnittlich fertigte jede*r Proband*in 62.8 Tonaufnahmen an (SD = 16.2). Über die Verteilung der Häufigkeiten von Tonaufnahmen pro Proband*in informiert der Boxplot in Abbildung 3.

Abbildung 3: Boxplot der absoluten Häufigkeiten der Tonaufnahmen pro Proband*in. (Innenlinie in Box = Median; Breite der Box = IQR; Box-Außenkanten = 25 % bzw. 75 % Quartil; Whisker = Minima/Maxima mit höchstens 1.5 IQR Abstand zur Box; Ausreißer (Kreuze) = Datenpunkte mit Abstand > 1.5 IQR zur Box; zur Verdeutlichung werden Ausreißer mit Jitter dargestellt)



FEATURE-EXTRAKTION

Die vorliegende Arbeit extrahiert für die Beantwortung der explorativen Forschungsfrage F1 vier Featuresets aus dem Audiodatensatz. Die Kodierung durch die hier aufgeführten Großbuchstaben wird in den späteren Modellnamen weiter fortgeführt:

1. Das BA-Featureset beinhaltet Features, die in mehreren Schritten an die Besonderheiten der Indoor Soundscapes angepasst wurden.
2. Das MIR-Featureset ist eine unangepasste Extraktion aller Standard-Features der MIRToolbox (Lartillot, 2021).
3. Das PSY-Featureset enthält die in ISO 12913-3 (DIN Deutsches Institut für Normung e.V., 2021) geforderten akustischen und psychoakustischen Messgrößen.
4. Das RA-Featureset extrahiert Werte mittels der Berechnungsmethode Relative Approach in HEAD Acoustics ArtemiS Suite (HEAD Acoustics, 2018).

Vor der Audioinhaltsanalyse wurde für das BA-Featureset und MIR-Featureset zu jeder observierten Audiodatei der linke oder rechte Audiokanal mit dem größten Betragsdigitalwert selektiert. Aus jeder so gewonnenen Monodatei pro Observation des Audiodatensatzes wurden in drei Schritten Audiofeatures extrahiert: Erstens wurde jede Datei gemäß der Bag-of-Frames-Methode mit einem Hanning-Fenster in Frames aufgeteilt, die eine Länge von 50 ms und einen Hop-Faktor von 0.5 (50 % Überlappung) aufwiesen, sodass jede 15-sekündige Datei mit 599 sequentiellen Frames repräsentiert werden konnte. Zweitens wurde auf jedem der Frames ein Set ausgewählter *Grundfunktionen* berechnet, sodass pro Datei 599 Werte pro Grundfunktion gespeichert wurden. Drittens wurden die Frame-Reihen der Grundfunktionen durch statistische, zeitbezogene oder andere, nichtlineare *Aggregationsfunktionen* zu Einzahlwerten zusammengefasst. Die erhaltenen Einzahlwerte und damit die Verknüpfung von Grund- und Aggregationsfunktion sind die eigentlichen Audiofeatures, die den Dateien und somit Observationen zugeordnet werden. Die Gesamtzahl der Features ergibt sich aus dem Produkt der Anzahl der Grundfunktionen mit der Anzahl der Aggregationsfunktionen. Der über die Aggregationsfunktionen angepasste Featuresatz wird als BA-Featureset bezeichnet.

Grundfunktionen des BA-Featuresets

Im Bereich der Grundfunktionen wurde ein umfangreiches Set von Mel Frequency und Gammatone Cepstral Coefficients (MFCC bzw. GTCC) inkl. der Delta-Werte und Delta-Delta-Werte, d. h. der Koeffizientendifferenzen benachbarter Frames, bis Rang dreizehn gewählt (Mitrović et al., 2010). Der per definitionem mit der Signalenergie korrespondierende Rang Null der Koeffizienten wurde nicht berücksichtigt.

Mit der Wahl der Filterbänke sind Anpassungen an das menschliche Hören möglich. So kann die Gammatone-Filterbank als Filterelement einer Cochlea-Simulation betrachtet werden (Patterson et al., 1992; Valero & Alias, 2012). Zwar ähneln die GTCC durch die Mittenfrequenzen der Bänder stark den MFCC, zeigen aber in Studien etwas bessere Ergebnisse bei Signalen mit hohem Störpegel (Yin et al., 2011; Lerch, 2012). Insgesamt scheinen GTCC zudem besser geeignet, die spektralen Komponenten unter 1 kHz zu repräsentieren, denen für viele Geräuscharten eine große Bedeutung zukommt (Valero & Alias, 2012). Aus diesen Gründen fiel die Entscheidung, sowohl MFCC als auch GTCC in das Featureset aufzunehmen.

Diverse Studien zeigen, dass es nützlich ist, die Cepstral Coefficients mit Lagebeschreibungen und Kennzahlen des Spektrums zu kombinieren (Pachet & Zils, 2003; Mörchen et al., 2005; Mörchen et al., 2006; Lee et al., 2013; Alías et al., 2016; Ronzhin et al., 2016). Aus diesem Grund kamen neun Kenngrößen zur Beschreibung der groben Form des Frequenzspektrums der Signale zum Einsatz: Beispielhaft soll die Funktion Spectral Slope hervorgehoben werden, die mit Annäherung einer Regressionsgeraden an Spektrallinien arbeitet. Die Steigung der angenäherten Geraden gibt Aufschluss über Abfall oder Anstieg der Pegel zu den hohen Frequenzen hin (Lerch, 2012).

Darüber hinaus wurden mit Entropy, Spread und Flux drei Funktion zur Abschätzung bzw. Berechnung des Zufallsgehalts, der Bandbreite und der Spektraldifferenzen zu vorherigen Frames eingesetzt (Lartillot, 2021).

Das Set wurde um zwei Funktionen zur Berechnung eines möglichen Grundtons (Pitch) und des Verhältnisses von Energie des harmonischen Signalanteils zur Gesamtenergie (Harmonic Ratio) ergänzt. Die Funktion Predictivity Ratio berechnet, wie gut das Signal mit einem linearen Modell vorhergesagt werden kann. Rauschsignale erzeugen üblicherweise einen hohen Fehler, harmonische Signale einen geringen Fehler (Lerch, 2012). Darüber hinaus wurden die

RMS-Energie und die Zerocrossing-Rate, d. h. die Rate der Nulldurchgänge, ermittelt. Eine Übersicht der eingesetzten Grundfunktionen liefert Tabelle 2.

TABELLE 2: ÜBERSICHT ÜBER GRUNDFUNKTIONEN IM BA-FEATURESET UND ANZAHL DER ERGEBNISVEKTOREN PRO AUDIODATEI. JEDER ERGEBNISVEKTOR PRO FUNKTION HAT DIE GRÖßE 599×1.

Grundfunktion	Anzahl
MFCC, Delta, Delta-Delta (je Rang 1-13)	65
GTCC, Delta, Delta-Delta (je Rang 1-13)	65
Spectral: Centroid, Crest, Decrease, Flatness, Kurtosis, Skewness, Rolloff Point, Slope, Flux, Spread, Entropy	11
Pitch, Harmonic Ratio, RMS Energy, Zerocrossing Rate, Predictivity Ratio	5

AGGREGATIONSFUNKTIONEN DES BA-FEATURESETS

Für die Selektion und spätere Regression muss jeder Observation (also jeder Tonaufnahme) ein Featurewert zugeordnet werden. Die Wertereihen der Grundfunktionen (Jeweils 599 Werte) müssen mit Aggregationsfunktionen zu Einzahlwerten zusammengefasst werden (Mörchen et al., 2006; McDermott et al., 2013).

Auf jeder der hier erzeugten Wertereihen wurde daher eine Schar von Aggregationsfunktionen berechnet. Sie umfassen die ersten vier statistischen Momente (Varianz, Standardabweichung, Schiefe, Wölbung), Minima und Maxima, vier Perzentile (akustische Definition), den Interquartilsabstand von 75. und 25. Perzentil und neben Modalwert und arithm. Mittelwert auch die Berechnung der lokalen Maxima pro Sekunde (Lokale-Maxima-Rate) und die Mittelung ihrer Prominenz. Für letztere wird die Prominenz über alle lokalen Maxima zunächst aufsummiert und dann auf die Anzahl der Maxima normiert. Darüber hinaus wurde mit der Rekonstruktion des Phasenraums ein nichtlineares Verfahren zur Zusammenfassung der Zeitwertreihen herangezogen (Lindgren et al., 2003; Aboofazeli & Moussavi, 2005; Benko & Perc, 2009). Der rekonstruierte Phasenraum wurde einer Hauptkomponentenanalyse unterzogen. Als finaler Featurewert wurde der signifikante Eigenwert der Kovarianzmatrix extrahiert. Wie bei Mörchen et al. (2006) wurde die Zeitverschiebung (Time lag) der Phasenraumrekonstruk-

tion nicht über die Nullstellen der Autokorrelation (Ma & Han, 2006) bestimmt. Stattdessen wurde die Funktion für die ersten fünf Zeitverschiebungen berechnet und als jeweils unabhängiges Feature behandelt.

Aufschluss über die Korrelation des Signals mit sich selbst gibt die Autokorrelationfunktion (ACF). Ein starker Abfall der ersten Werte der Autokorrelationsfunktion entspricht einem Signal, das nur von einer kleinen Anzahl vorausliegender Werte abhängt, also ein hohes Maß an Zufälligkeit hat und tendenziell ein weißes Spektrum besitzen wird. Die ersten zehn Werte der ACF (entspricht einer Verschiebung um zehn Einheiten) wurden mit einer Regressionsgeraden angenähert. Die Steigung der Geraden ist der Featurewert. Dasselbe Prozedere wurde für die ersten zehn Werte der partiellen Autokorrelationsfunktion durchgeführt. Die Arbeit folgt auch mit diesem Ansatz den Vorschlägen der Studie von Mörchen et al. (2006).

Um Konflikte durch lineare Abhängigkeiten in der späteren Selektion und Modellbildung zu vermeiden, wurden Kollinearitäten identifiziert und betroffene Features aus dem BA-Featureset entfernt (Funktion siehe Anhang D). In der resultierenden Matrix mit vollem Spaltenrang umfasste das BA-Featureset 1560 Audiofeatures. Über die Anzahl der angewendeten Aggregationsfunktionen informiert Tabelle 3.

TABELLE 3: ÜBERSICHT DER AGGREGATIONSFUNKTIONEN UND ANZAHL DER ERGEBNISVEKTOREN PRO GRUNDFUNKTION PRO AUDIODATEI IM BA-FEATURESET. RESULTIERENDE VEKTOREN HABEN DIE GRÖßE 1×1. FÜR PERZENTILE GILT DIE AKUSTISCHE DEFINITION.

Aggregationsfunktion	Anzahl
Statistische Momente: Var, SD, Schiefe, Wölbung	4
Min. & Max.; Perzentile 1, 5, 10, 50; Interquartilsabstand (25 % - 75 %)	7
Arithm. Mittelwert, Modalwert	2
Eigenwert nach PCA auf rekonstruiertem Phasenraum (Dimension = 2; Time lag = 2 – 5)	4
Steigung einer Regressionsgeraden auf ACF- und partieller ACF	2
Entropie, Lokale-Maxima-Rate, Gemittelte Prominenz	3

FEATURES DER MIRTOOLBOX (MIR)

Um später die im BA-Set enthaltenen Features mit unangepassten Features zu vergleichen, wurde die MIRToolbox für MATLAB herangezogen (Lartillot et al., 2008; Lartillot, 2021). Da sie

in vielen Arbeiten standardmäßig miteinbezogen wird, ist sie ein geeignetes Vergleichsinstrument zur Beurteilung der angepassten Features des BA-Featuresets. Studien attestieren zudem eine breite Funktionspallette (Moffat et al., 2015).

Die Toolbox kann für einen Ordner mit Audiodateien mithilfe des Bag-of-Frames-Ansatzes ein Set aus den in der Toolbox zur Verfügung stehenden Features berechnen. Die intern genutzten Grundfunktionen (Tabelle 4) werden mit drei bis fünf Aggregationsfunktionen zusammengefasst (Mean, SD, Steigung, Frequenz der Periodizität, Amplitude der Periodizität). Die Auswahl und Anzahl der Aggregationsfunktionen ist abhängig von voreingestellten ‚Default'-Werten der MIRToolbox und diese wurden nicht verändert. Weil die MIRToolbox aber keine parallelisierte Berechnung anbietet, wurde ein Workaround entwickelt, um die Rechendauer zu verringern (siehe Anhang D). Features, die NA-Werte enthalten, mussten für die Selektion entfernt werden. Der Satz der MIRToolbox umfasst nach dieser Bereiniung 305 Features.

TABELLE 4: ÜBERSICHT DER GRUNDFUNKTIONEN DER MIRTOOLBOX UND ANZAHL DER ERGEBNISVEKTOREN PRO AUDIODATEI

Grundfunktion	Anzahl
MFCC, Delta, Delta-Delta bis Rang 13	65
Rms, fluctuation, zerocross, lowenergy, spectralflux	5
Tempo, attack time, attack slope	3
Centroid, brightness, spread, skewness, kurtosis, rollof95, rollof85, entropy, flatness, roughness, irregularity	11
Chromogram (peak), keyclarity, mode, hcdf	4

PSYCHOAKUSTISCHE MESSGRÖßEN (PSY)

Vor Berechnung des PSY-Featuresets wurde jeder Kanal der binauralen Audiodateien mit dem zugehörigen, aus den Kalibrierungsaufnahmen errechneten, Faktor kalibriert. Übersteuerungen durch Betragsdigitalwerte über eins wurden durch adäquate Pegelskalierung unterbunden. Die Berechnungen der akustischen und psychoakustischen Messgrößenwerte fanden in HEAD Acoustics ArtemiS Suite 13.1 statt. Unter Zuhilfenahme der ASX-Entwicklerschnittstelle wurden die Einzahlwerte in MATLAB importiert.

Für das PSY-Featureset wurden alle Messgrößen extrahiert, die im Standard zur Datenanalyse von Soundscapes, ISO 12913-3 (DIN Deutsches Institut für Normung e.V., 2021) gefordert werden. Dieser legt neben der Betrachtung von Schalldruckpegeln mit A- und C-Gewichtung nach ISO 1996-1 (ISO International Organization for Standardization, 1996, 2017a) auch die Erhebung bzw. Berechnung der psychoakustischen Größen Lautheit nach ISO 532-1 (ISO International Organization for Standardization, 2017b; DIN Deutsches Institut für Normung e.V., 2020), Schärfe nach DIN 45692 (DIN Deutsches Institut für Normung e.V., 2009), Rauheit und Schwankungsstärke nach Zwicker & Fastl (1999) sowie der Tonalität nach ECMA 74 (Ecma International, 2021) nahe.

Im Soundscape-Standard ist ein Mindestmaß an Perzentilen und Durchschnittswerten zu finden, das spezifisch auf die jeweilige Messgröße abgestimmt ist (z. B. kubische Mittelung der Lautheit). Über dieses Mindestmaß hinaus wurden angelehnt an Aletta et al. (2017) für jede Größe zusätzliche Perzentile (1, 5, 10, 25, 50, 75, 90, 95, 99; akustische Definition), als Maß für zeitliche Variabilität die Perzentildifferenzen (P1-P99, P5-P95, P10-P90, P25-P75) sowie ein Maß für spektrale Variabilität ($L_{Ceq} – L_{Aeq}$) berechnet. Bei allen Berechnungen wurde wie im ISO-Standard gefordert der jeweils höhere Wert der Messgrößenwerte für den linken und rechten Kanal extrahiert. Insgesamt wurden für das Set akustischer bzw. psychoakustischer Messgrößen 103 Features extrahiert.

ARTEMIS: RELATIVE APPROACH (RA)

Die HEAD Acoustics ArtemiS-Suite bietet neben den standardmäßigen Messgrößen die proprietäre Analysemethode Relative Approach, die durch Einbeziehung vorheriger Zeit- bzw. benachbarter Frequenzwerte ermittelt, inwiefern ein Signal zum Messzeitpunkt von einem berechneten Erwartungswert abweicht (HEAD Acoustics, 2018). Relative Approach wurde in der Vergangenheit bei der akustischen Beschreibung von Turbinengeräuschen oder von Verkehrslärm eingesetzt (A. Fiebig et al., 2009; Sottek & Genuit, 2009).

Werte dieser Methode sind stets Relativwerte, weil der aktuelle Signalwert in Bezug zu seinen umgebenden Werten gesetzt wird. Es stehen dazu verschiedene Basisanalyse-Modelle (z. B. 1/n-Oktavfilter und Lautheit) zur Verfügung, auf denen dann eine bestimmte Variationsanalyse ausgeführt wird. Die Variationsanalyse Regression berechnet eine von Zeit- oder Frequenzmustern abhängige Regression für den aktuellen Signalwert. Die Prominence-Berechnung vergleicht den aktuellen Signalwert mit einem Mittelwert, der über ein

peripheres Rechteck im Spektrogramm gebildet wird. Aus den mannigfaltigen Kombinationsmöglichkeiten wählt die vorliegende Arbeit neun Analysefunktionen im Relative Approach aus (Tabelle 5). Deren Parameter wurden anhand einiger charakteristischer Beispielgeräusche aus dem Datensatz entsprechend des angeratenen Vorgehens der Dokumentation (HEAD Acoustics, 2018) eingestellt, um eine subjektiv große Varianz der Werte für unterschiedliche Geräuschbeispiele zu erzeugen. Als Einzahlwerte wurden neben den Durchschnittswerten und Summen wie bei den psychoakustischen Messgrößen Perzentile (1, 5, 10, 25, 50, 75, 90, 95, 99; akustische Definition) und Perzentildifferenzen (P1-P99, P5-P95, P10-P90, P25-P75) als zeitliche Schwankungsmaße berechnet. Das RA-Featureset umfasst insgesamt 136 Features.

TABELLE 5: ÜBERSICHT DER GRUNDFUNKTIONEN IM RA-FEATURESET. BEI FFT-BERECHNUNGEN IST DIE ÜBERLAPPUNG JEWEILS 50 %. ANGABE DER FFT-LÄNGE IN SAMPLES.

Feature	Basisanalyse	Variationsanalyse
ra_1/12F_RT	1/12 Oktavfilter, A-Bewertung, Zeitbewertung = 5 ms	Zeitmuster
ra_1/12FFT_RT	1/12 Oktav (FFT), A-Bewertung, FFT-Länge = 1024 Samples	Zeitmuster
ra_1/12F_RF	1/12 Oktavfilter, A-Bewertung, Zeitbewertung = 5 ms	Frequenzmuster
ra_1/12FFT_RF	1/12 Oktav (FFT), A-Bewertung, FFT-Länge = 1024 Samples	Frequenzmuster
ra_1/3F_RT	1/3 Oktavfilter, A-Bewertung, Zeitbewertung = 5 ms	Zeitmuster
ra_1/3FFT_RT	1/3 Oktavfilter, A-Bewertung, FFT-Länge = 1024	Zeitmuster
ra_HM	Hörmodell, A-Bewertung	Frequenzmuster
ra_Pr1	Lautheit (FFT/Head), FFT-Länge = 256	Prominence 3D
ra_Pr2	Lautheit (FFT/Head), FFT-Länge = 2048	Prominence 3D

FEATURE-SELEKTION

Für jedes Featureset wurde eine Selektion relevanter Features für Angenehmheit und Ereignisreichtum durchgeführt. Damit alle Features gleichberechtigt in die Selektion eingehen,

wurden sie zuvor Z-standardisiert. Als Selektionsverfahren kam die Perzentile LASSO-Regularisierung zum Einsatz (Roberts & Nowak, 2014).

Perzentile Lasso-Regularisierung

Der sogenannte ‚Least Absolute Shrinkage and Selection Operator' Lambda wird genutzt, um unwichtige bzw. redundante Features aus Modellen zu entfernen.

Die in dieser Arbeit eingesetzte LASSO-Funktion löst in allgemeiner Form das folgende Problem:

$$\min_{\beta_0,\beta}\left(\frac{1}{N}Devianz(\beta_0,\beta)+\lambda\sum_{j=1}^{p}\left|\beta_j\right|\right)$$

MIT β_0 = INTERCEPT (SKALAR), p = ANZAHL PRÄDIKTOREN, β = PRÄDIKTOR-KOEEFIZIENTEN (VEKTOR DER LÄNGE p), N = ANZAHL OBSERVATIONEN, λ = REGULARISIERUNGSPARAMETER. DIE DEVIANZFUNKTION IST ABHÄNGIG VON DER GEWÄHLTEN VERTEILUNG (MATHWORKS, 2012).

Der Algorithmus berechnet ausgehend von demjenigen maximalen Lambda, bei dem gerade alle Prädiktor-Koeffizienten null sind, für einhundert auf ein logarithmisches Netz verteilte Lambda-Werte den durch k-fache Kreuzvalidierung gemittelten Modellfehler. Es wird dann das Lambda ausgewählt, das einen Standardfehler von demjenigen Lambda entfernt ist, für das in der Kreuzvalidierung gemäß der obigen Formel der niedrigste Modellfehler (Devianz) berechnet worden war (1SE-Regel). Das gewählte Lambda heißt 1SE-Lambda.

Die Kreuzvalidierung unterliegt prinzipiellen Ergebnisschwankungen, die von der zufälligen Zuweisung der Teilmengen (Folds) abhängt. Die Ergebnisschwankung beeinflusst im Standard-Verfahren der LASSO-Regularisierung sowohl die errechneten Modellfehler als auch die resultierenden Modellgrößen. Nach Roberts & Nowak (2014) kann das Ergebnis stabilisiert werden, indem ein Perzentil (statistische Definition) über N Ergebnisse für das 1SE-Lambda unabhängiger LASSO-Regularisierungen mit jeweils zufälliger Teilmengenzuordnung für die Kreuzvalidierungen berechnet wird. Roberts & Nowak schlagen anhand zweier Realdatenstudien das 95. Perzentil für das optimale Lambda vor. Zumeist erreicht ein so hoch gewähltes optimales Lamda eine deutlich ‚strengere' Bestrafung als ein optimales Lambda auf Grundlage des Modal- oder Mittelwerts aller 1SE-Lambdas. Das verkleinert die Varianz der Modellgröße und ermöglicht robuste Schlussfolgerungen, die nicht von den zufälligen Teilmengenzuweisungen der Kreuzvalidierung abhängen (Roberts & Nowak, 2014, S. 210).

Die vorliegende Arbeit setzt entsprechend das 95. Perzentil ein (Definition der Statistik: 95 % der 1SE-Lambdas liegen unterhalb oder sind gleich diesem Wert). Das Standard-LASSO-Verfahren wurde innerhalb des Perzentilen LASSO 100-mal mit jeweils zufälligen Teilmengenzuordnungen der Kreuzvalidierung ausgeführt. Die Anzahl der Teilmengen in der Kreuzvalidierung wurde aus Effizienzgründen und im Sinne der Forschung auf fünf ($k_{CV} = 5$) festgesetzt (Krstajic et al., 2014). Nach seiner Berechnung durch das 95. Perzentil wird das Lambda in ein Standard-LASSO eingesetzt, um den zugehörigen Ergebnisvektor zu erhalten. Seine Koeffizienten sind nur für die im Perzentilen LASSO selektierten Features ungleich null. Der danach gefilterte Datensatz wird in einem linearen gemischten Modell zur Berechnung der festen und zufälligen Effekte genutzt.

Die Penaltisierung der Prädiktorkoeffizienten in Abhängigkeit des Regularisierungsparameters kann beispielhaft anhand eines Modells in Abbildung A3 (siehe Anhang A) nachvollzogen werden.

MODELLBILDUNG

In Anschluss an die für jede Kombination aus Featureset und Zielgröße durchgeführte Selektion wurde ein lineares gemischtes Modell berechnet, das sowohl feste Effekte durch die Audiofeatures als auch zufällige Effekte durch die Proband*innen enthält. Die signifikanten Prädiktoren der Einzelmodelle wurden dann für jede Zielgröße erneut einer Perzentilen LASSO-regularisierten Selektion unterzogen und in zwei Modelle zusammengeführt.

Die Modellnamen werden entsprechend dem Schema ‚Zielgröße_Featureset' benannt. Dabei wird für die Zielgröße Angenehmheit ‚PL' und für Ereignisreichtum ‚EV' eingesetzt. Die Kodierung der Featuresets erfolgt mit den genannten Großbuchstaben. Die kombinierten Modelle der Selektionen aus signifikanten Prädiktoren aller Featuresets heißen PL_SEL für Angenehmheit und EV_SEL für Ereignisreichtum.

In den SEL-Modellen wurde noch ein weiteres Feature getestet: die Dateigröße des Formats MPEG-1 Audio Layer 3 (MP3). Jede Audiodatei war mit dem LAME-Encoder (siehe Anhang C) bei variabler Bit-Rate MP3-kodiert und somit datenreduziert worden. Grundsätzlich können MP3-Encoder Audiomaterial, das unterschiedlich hohe spektrale Verdeckung und zeitliche Maskierung enthält, unterschiedlich stark datenreduzieren. Setzt man die Bitrate variabel ein, wird die Audioqualität konstant gehalten und es ändert sich nur die Dateigröße mit dem Grad

der Verdeckung bzw. Maskierung. Somit könnte die MP3-Dateigröße im Zusammenhang mit akustischer Komplexität stehen und feste Effekte auf die Wahrnehmungsdimensionen von Soundscapes enthalten.

Als zentrale Größe für die Güte der Modelle wurde das Bestimmtheitsmaß $R^2_{marginal}$, also der Anteil der Varianzaufklärung durch die festen Effekte, gemäß Nakagawa bestimmt (Nakagawa & Schielzeth, 2013; Nakagawa et al., 2017). Anhand dessen können die Modelle unterschiedlicher Featuresets miteinander verglichen werden.

MULTILEVEL-MODELLIERUNGEN

Zur genaueren Prüfung der in Hypothese H3 vermuteten Zusammenhänge von Angenehmheit und psychoakustischer Lautheit wurden drei Teilmengen-Modelle mit entsprechender Regularisierung durch das Perzentile LASSO aufgestellt.

Um den Einfluss von lauter Musik zu beleuchten, wurden alle 602 Observationen (9.13 %) ausgeschlossen, in denen Proband*innen angegeben hatten, dass das auffälligste Geräusch Musik war. Eine Überprüfung durch Modellbildung für alle Soundscapes, die ausschließlich vordergründige Musik enthalten, war wegen der geringen Anzahl und einhergehender Singularitäten nicht möglich.

Des Weiteren wurden zwei über die Lautheit bestimmte Multilevel-Modelle zweier Teilmengen des Datensatzes berechnet. Dazu wurde das 75. Perzentil (statistische Definition) der Lautheit als willkürlicher Trennungswert gewählt, der deutlich über dem Mittelwert der Lautheit liegt: Observationen wurden entweder der leiseren Teilmenge N_low zugeordnet, wenn ihre durchschnittliche kubische Lautheit kleiner gleich dem 75. Perzentil über die Lautheit aller Observationen war. Observationen, deren Lautheitswert darüber lag, wurden der Teilmenge N_high zugeordnet. Die Multilevel-Modelle wurden nach Perzentiler LASSO-Selektion über das Featureset berechnet, das zuvor das beste Einzelmodell für Angenehmheit geliefert hatte.

Weil das 75. Perzentil willkürlich gewählt wurde, ist der Verlauf der Modellkoeffizienten für feste und zufällige Effekte in Abhängigkeit des Trennpunktes zu betrachten. Dieser kann in einer Näherung durch folgendes Vorgehen berechnet werden: Das Perzentil wird iterativ hochgezählt und in jeder Schleife werden für beide Level (low und high) die Modellstatistiken $R^2_{marginal}$ und $ICC_{conditional}$ berechnet. Dabei wird vereinfachend für jedes Perzentil die gleiche

Modellformel angenommen und nicht neu über LASSO-Selektion ermittelt. Der beispielhafte Verlauf der Modellkoeffizienten über die Perzentile gibt Aufschluss über die relative Modellierbarkeit in Abhängigkeit des gewählten Trennpunkts für die Teilmengen N_low und N_high (siehe Abbildung A4 in Anhang A).

SOFTWARE

Eine Gesamtliste aller in die Selektion eingeflossenen Features, die wichtigsten in der Arbeit eingesetzten Funktionen und Skripte sowie einige beispielhafte Featurefunktionen des BA-Featuresets liegen in Anhang D vor.

Der durch *movisens xs* erzeugte ESM-Datensatz wurde in RStudio (Version siehe Anhang C) aufbereitet und dem Audiodatensatz zugeordnet. Die Vorbereitung der Tonaufnahmen, die Feature-Extraktion des BA- und MIR-Featuresets sowie die Selektion durch Perzentile LASSO-Regularisierung wurde in MathWorks MATLAB mit verschiedenen Toolboxen (Versionen siehe Anhang C; Funktionen und Skripts siehe Anhang D) durchgeführt. Die PSY- und RA-Featuresets wurden in HEAD Acoustics ArtemiS Suite 13.1 berechnet. Die Modellrechnungen fanden in RStudio (Version siehe Anhang C) mit dem lme4-Package statt.

ERGEBNISSE

Im Folgenden werden die zentralen Ergebnisse der linearen gemischten Modelle hervorgehoben und die forschungsrelevanten Befunde berichtet. Zunächst werden die Ergebnisse für die Modelle der Angenehmheit und nachfolgend auch der Multilevel-Modellierung für Angenehmheit vorgestellt. Im Anschluss werden die zentralen Ergebnisse der Modelle für Ereignisreichtum ausgewertet. Die detaillierte Übersicht aller Einzelmodelle ist den Ergebnistabellen in Anhang B zu entnehmen.

Für alle nachfolgend berichteten Ergebnisse wird ein Signifikanzniveau von $\alpha = 0.05$ angenommen. Die berichteten β-Werte beruhen wie bereits dargelegt auf Z-standardisierten Feature-Werten.

ANGENEHMHEIT

PL_BA: ANGENEHMHEIT & BA-FEATURESET

Im Modell PL_BA ($R^2_{marginal} = 0.05$) wurden bei MFCC-12 und Predictivity-Ratio zwei mit Lokaler-Maxima-Rate aggregierte Features selektiert ($\beta = -0.17$; $p < .001$ bzw. $\beta = -0.05$; $p < .001$). Diese Rate kann ausdrücken, ob der Zeitverlauf der Features rauschhaft ist. Das wäre bei vordergründigen Geräuschquellen wie Haushaltsgeräten oder Lüftungen etc. der Fall, denen hier also eine unangenehmere Bewertung zugesprochen werden kann. Die Effekte durch Kurtosis (geringe Wölbung) des Deltas von MFCC-3 ($\beta = -0.06$; $p < .001$) und Modalwert des Deltas von GTCC-1 ($\beta = -0.03$; $p = .01$) zeigen zudem einen Negativeffekt durch hohe Zeitschwankung im tieferen Frequenzbereich an. Wenn der Delta-Wert selten hoch ist (entspricht auch dem kleinen Modalwert), gibt es wenig starke Werteänderungen in den Wertereihen des GTCC-1. Bei starken Trittschallgeräuschen z. B. wäre in den Werten das Gegenteil, also ein häufig hohes Delta, zu erwarten. Die Werte für Skewness des Spectral-Centroids ($\beta = 0.05$; $p = .002$) lassen darauf schließen, dass Soundscapes, deren spektrale Schwerpunkte eher in Richtung tieferer Frequenzen verschoben sind, angenehmer bewertet wurden als Soundscapes mit höherem Schwerpunkt.

PL_MIR: ANGENEHMHEIT & MIR-FEATURESET

PL_MIR leistet mit $R^2_{marginal} = 0.05$ dieselbe Varianzaufklärung wie PL_BA, hat aber mit neun selektierten Features die höhere Modellgröße. Zu berichten ist, dass zwei für Musik und tonale

Signale designte Features in das Modell selektiert worden sind: Die Aggregation der Harmonic Change Detection Function (Harte et al., 2006) durch die Amplitude der maximalen Periodizität (Lartillot, 2021) zeigt einen negativen Effekt ($\beta = -0.08$; $p < .001$). Das Feature zeigt in seinen Extremwerten sehr gut Musik und einzelne Instrumente an (Minima) und reagiert hingegen auf rauschhafte, scharfe Geräusche, die bspw. beim Kochen und Braten vermehrt auftreten, mit Maximalwerten. Darüber hinaus wurde ein Feature selektiert, das mit der Keyclarity-Funktion auf die Erkennbarkeit von Grundtönen in Musik spezialisiert ist ($\beta = 0.04$; $p < .001$). Die Standardabweichung dieser Funktion ist besonders gering, wenn die Erkennbarkeit von Grundtönen im Signal nie schwankt, was auf ein starkes Anlagenbrummen, Fiepen o. ä. hindeuten würde. Das erzeugt im vorliegenden Modell eine negativere Bewertung der Angenehmheit.

PL_PSY: Angenehmheit & PSY-Featureset

Das beste Einzelmodell für Angenehmheit bilden die akustischen und psychoakustischen Messgrößen ($R^2_{marginal} = 0.07$). Hervorzuheben sind die beiden positiven Effekte durch Werte der Schwankungsstärke, die über 95 % der Zeit erreicht werden ($\beta = 0.18$; $p < .001$) und die zeitliche Schwankung der Hörempfindung Schärfe ($\beta = 0.14$; $p < .001$). Die beiden Features sind hochsensitiv für vordergründige Sprache im Raum. Ein negativer Effekt geht von der kubisch gemittelten Lautheit aus ($\beta = -0.14$; $p < .001$). Dieser Befund erlaubt, die Hypothese H3 des negativen Zusammenhangs von Lautheit und Angenehmheit zu bestätigen.

PL_RA: Angenehmheit & RA-Featureset

Die Lautheitsbefunde aus PL_PSY gehen einher mit der Selektion der Relative-Approach-Features in PL_RA ($R^2_{marginal} = 0.05$): Der Prominence-Berechnung (Pr1) liegt eine Basisanalyse der Lautheit zu Grunde. Dabei wird der momentane Signalwert von einem Schätzwert abgezogen, der über die Lautheit der vorherigen Signalwerte vorhergesagt wurde. Dauerhaft hohe Werte der Prominence des Relative Approach bedeuten ständige Änderungen des Signals. Für diese dauerhaften, aber auch für spontane Musterabweichungen in 10 % der Zeit ist ein negativer Effekt festzustellen ($\beta = -0.23$; $p < .001$ bzw. $\beta = -0.10$; $p < .001$). Entgegengesetzt dazu ist der positive Effekt durch hohe Dauerwerte der Basisanalysefunktion Hearing Model (HM) zu nennen ($\beta = 0.29$; $p < .001$). Hearing Model detektiert tieffrequente Signalmuster, wie sie z. B. in Musik vorkommen, sehr zuverlässig. Auch im vorliegenden Datensatz schlägt das Feature

mit hohen Werten deutlich auf Musik und Sprache an. Tonaufnahmen, die hohe Dauerwerte in diesem Feature zeigen, wurden angenehmer bewertet.

PL_SEL: Angenehmheit & signifikante Prädiktoren

Das umfassende LASSO-selektierte Modell der signifikanten Prädiktoren der Einzelmodelle erreicht keine deutlich höhere Varianzaufklärung der festen Effekte und verbessert sich lediglich auf $R^2_{marginal} = 0.09$. Den stärksten Einfluss hat die durchschnittliche Lautheit ($\beta = -0.2$; $p < .001$). Die Selektion dieses Features im Kontext der signifikanten Prädiktoren aller Featuresets unterstreicht das Zutreffen von Hypothese H3. Das 95. Perzentil der Basisanalyse Hearing Model des Relative Approach sowie die Zeitschwankung der Schärfe zeigen gleiche Effektstärken (beide $\beta = 0.13$; $p < .001$). Mit etwas geringerem Effekt gleicher Richtung ist das 95. Perzentil der Schwankungsstärke zu nennen ($\beta = 0.08$; $p < .001$). Diese drei letztgenannten Features sind sensitiv für Sprache und Musik. Maximalwerte werden z. B. für Gespräche mehrerer Sprecher*innen erreicht.

Die Lokale-Maxima-Rate des MFCC-12 ($\beta = -0.04$; $p = .006$) und Amplitude der Periodizität auf MFCC-10 ($\beta = -0.03$; $p = .002$) zeigen Effekte durch Periodizität im Bereich höherer Ränge auf. Die Features schlagen vor allem bei surrenden Haushaltsgeräten, aber auch leiseren rauschhaften Geräuschen an. Sie haben aber relativ zu den lautheitsbasierten psychoakustischen Features deutlich geringere Effektstärken. Einen gegensätzlichen Effekt zeigt der Durchschnittswert des MFCC-7 ($\beta = 0.03$; $p = .002$). Unter den Maximalwerten dieses Features sind mehrere Tonaufnahmen zu finden, die leise Vogelstimmen enthalten.

Die HCDF- und Keyclarity-Features des Modells PL_MIR sind mit gleichen Effektrichtungen auch wieder im Modell PL_SEL vorhanden ($\beta = -0.04$; $p = .008$ bzw. $\beta = 0.04$; $p < .001$). Die zuvor genannten Erläuterungen gelten hier gleichermaßen.

Teilmengen- und Multilevel-Modelle

Um den Zusammenhang von Lautheit und Angenehmheit besser zu verstehen waren Datensatzteilmengen erzeugt worden. Weil das PSY-Featureset bereits im Einzelmodell PL_PSY die beste Varianzaufklärung gezeigt hatte, wurde es für die Multilevel-Modelle den anderen Featuresets vorgezogen.

PL_NoMusic: Angenehmheit & keine Musik

Für das Modell ohne diejenigen Observationen, die Musik im Vordergrund haben, konnte eine Varianzaufklärung durch die festen Effekte von $R^2_{marginal} = 0.08$ festgestellt werden. Es kommt nicht zu einer deutlichen Verbesserung der Modellgüte im Vergleich zu PL_PSY oder PL_SEL. Das lässt die Schlussfolgerung zu, dass bewusst im Raum gespielte oder gehörte Musik, die typischerweise hohe Lautheit hat, aber dennoch sehr angenehm bewertet ist, die Modellierung von Angenehmheit nicht maßgeblich verzerrt.

Multilevel PL_N: Angenehmheit & Lautheit

In der Multilevel-Modellierung sind vier zentrale Befunde festzustellen: Erstens ist die Varianzaufklärung durch feste Effekte für Observationen hoher Lautheit deutlich höher als für alle Observationen ($R^2_{marginal,\ PL\_N\_high} = 0.13$; $R^2_{marginal,\ PL\_SEL} = 0.09$). Zweitens ist für PL_N_low im Vergleich zu PL_PSY ein Abfall der Modellgüte auf das Niveau von PL_BA und PL_PSY zu beobachten ($R^2_{marginal} = 0.05$). Drittens ist die Lautheit selbst nur für den lauteren Teil der Observationen, also PL_N_high als Prädiktor selektiert worden ($\beta_{high} = -0.09$; $p = 0.001$). Im Gegensatz dazu enthält PL_N_low außerdem das 1. Perzentil der Schärfe mit einem geringen Effekt ($\beta_{low} = 0.09$; $p < .001$).

In beiden Modellen PL_N_low und PL_N_high ist wie auch schon in PL_SEL ein Effekt durch das 95. Perzentil der Schwankungsstärke signifikant ($\beta_{low} = 0.13$ bzw. $\beta_{high} = 0.19$). Die Zeitschwankung der Schärfe wurde in beiden Modellen durch LASSO selektiert, jedoch ist der Effekt nur in PL_N_high signifikant ($\beta_{high} = 0.18$; $p < .001$).

Die Auswertung des Plots der Modellstatistiken in Abhängigkeit der gewählten Perzentile zur Erzeugung der Datensatzteilmengen N_low und N_high (Abbildung A4 in Anhang A) zeigt klar auf, dass die Angenehmheit des lauteren Teils des Datensatzes immer deutlich besser durch feste Effekte beschreibbar ist als die des leiseren. $R^2_{marginal}$ der lauteren Teilmenge N_high ist allerdings mit der Anzahl der zugehörigen Observationen sehr stark negativ korreliert ($R = -0.92$; $p < .001$). Es ist an dieser Stelle methodisch nicht auszuschließen, dass auch Kausalität zwischen der besseren Varianzaufklärung lauter Teilmengen und der Reduktion der Anzahl von Observationen besteht. Der Zusammenhang kann in Folgeuntersuchungen aufgeklärt werden, indem bei Iteration über Lautheitsschwellwerte immer auch die Anzahl der Observationen der Teilmengen angeglichen wird.

Davon unabhängig ist aber die Validität des Befundes, dass die lautere Datensatzteilmenge immer besser modelliert werden kann als die leisere. Lautheit stellt sich als wichtige Clustering-Variable heraus.

Darüber hinaus ist die Deselektion von Lautheit als Prädiktor der Angenehmheit in PL_N_low ein wichtiges Ergebnis: Für die Angenehmheit grundsätzlich leiserer Geräusche spielen feine Lautheitsunterschiede dann keine Rolle mehr.

EREIGNISREICHTUM

EV_BA: EREIGNISREICHTUM & BA-FEATURESET

Aus dem BA-Featureset wurde im Perzentilen LASSO ein Modell aus acht statistisch signifikanten Prädiktoren für die Zielgröße Ereignisreichtum selektiert ($R^2_{marginal}$ = 0.14). Darunter befinden sich zwei Features der GTCC-Familie und sechs MFCC-Features. Die beiden größten positiven Effekte können für den Durchschnittswert des GTCC-1 ($\beta = 0.14$; $p < .001$) und das 5. Perzentil des MFCC-2 ($\beta = 0.08$; $p < .001$) berichtet werden, die größten negativen Effekte gehen von den Lokalen-Maxima-Raten der Delta-Werte des MFCC-12 ($\beta = -0.13$; $p < .001$) und der Delta-Delta-Werte des MFCC-10 ($\beta = -0.08$; $p < .001$) aus. Betrachtet man nur die Aggregationsebene, machen Features, die über die Lokale-Maxima-Rate aggregiert wurden, ca. 33 % und Features, die über eine lmSlope-PACF-Funktion aggregiert wurden, ca. 20 % der relativen Effektstärke aus. Beide Features beider Gruppen weisen geringe negative Effekte auf. Eine hohe Lokale-Maxima-Rate auf den höheren Rängen der Cepstral Coefficients deutet eher auf rauschartige Signale hin. Sprache, Musik und z. B. Klappergeräusche zeigen geringe Raten, aber werden ereignisreicher bewertet. Die negativen Effekte der lmSlope-PACF-Funktionen zeigen, dass Geräusche als ereignisreich bewertet wurden, deren Korrelation mit sich selbst schnell bzw. stark abfällt. Das trifft vor allem auf Klappern, Poltern etc. zu. Die festen Effekte erklären 14 % der Varianz von Ereignisreichtum, was dem Niveau der Varianzaufklärung durch die zufälligen Effekte des Nullmodells entspricht (τ_{00} = 0.15). Durch die algorithmische Selektion der Features zur zeitlichen Korrelation und Schwankung sowie der MFCC-Deltas kann die Nullhypothese zu H1, dass kein Zusammenhang zwischen Feature-Zeitschwankungen und Ereignisreichtum besteht, verworfen werden. Der Effekt durch GTCC-1-Mean lässt aber nicht zu, auch für H2 die Nullhypothese zu verwerfen. Das bedeutet, es wird belegt, dass Zeitschwankungen in den Features einen signifikanten positiven Effekt auf Ereignisreichtum haben, aber

es kann nicht nachgewiesen werden, dass der Unterschied im Vergleich zur Beschreibung mit Durchschnittswerten überzufällig ist.

EV_MIR: Ereignisreichtum & MIR-Featureset

Das Modell EV_MIR erklärt über feste Effekte ($R^2_{marginal}$ = 0.12) einen etwas geringeren Anteil der Varianz von Ereignisreichtum als EV_BA. Mit vier Prädiktoren handelt es sich um das kleinste Modell für Ereignisreichtum. Es gibt einen deutlichen negativen Effekt durch den Durchschnittswert des Spectral Spread (β = -0.2; $p < .001$), der als Quasibandbreite des Spektrums vor allem für rauschartige Signale groß wird. Dieser Befund deutet darauf hin, dass die Nullhypothese zu H2 angenommen werden muss. Zwei positive Effekte mit jeweils β = 0.1 ($p < .001$) konnten für die Standardabweichungen von MFCC-12 und MFCC-13 modelliert werden. Sprache zeigt in diesem Bereich der MFCC typischerweise höhere Standardabweichungen als rauschartige, technische Geräusche. Ein ungewöhnlicher Befund ist mit dem positiven Zusammenhang der Standardabweichung der Keyclarity und Ereignisreichtum zu berichten (β = 0.07; $p < .001$): Die Funktion Keyclarity, die insbesondere für Musik programmiert wurde, gibt die Stärke der Erkennbarkeit eines berechneten Grundtons an (Lartillot, 2021). Die Bewertung von Ereignisreichtum ist hoch, wenn die Standardabweichung über eine Wertereihe der Keyclarity groß ist, die Erkennbarkeit von Grundtönen also stark schwankt. Das würde vor allem für Brummgeräusche oder andere dauerhafte Töne zutreffen. Die positiven Modellparameter der MFCC-Standardabweichungen stützen die Hypothese H1.

EV_PSY: Ereignisreichtum & PSY-Featureset

Das Modell EV_PSY zeigt mit $R^2_{marginal}$ = 0.23 zusammen mit EV_RA die beste Varianzaufklärung von Ereignisreichtum durch feste Effekte. Der positive Effekt durch das erste Perzentil des C-bewerteten Schalldruckpegels zeigt dabei mit Abstand die größte Stärke (β = 0.38; $p < .001$). Hohe Schalldruckpegel in 1 % der Zeit werden ereignisreicher bewertet. Die Selektion der C-Bewertung versus der deutlich stärker an die Eigenschaften des Gehörs angepassten A-Bewertung kann auf die Bedeutung tieffrequenter Trittschallgeräusche wie z. B. Poltern hindeuten: Die A-Bewertung dämpft tiefe Frequenzen deutlich stärker (Ca. 19 dB Differenz der Bewertungskurven bei 100 Hz), was dem menschlichen Hören nachempfunden ist. Dass dennoch die C-Bewertung selektiert wurde, deutet daraufhin, dass tieffrequente kurzzeitige Peaks im Schalldruckpegel grundsätzlich ereignisreicher bewertet werden, u. U. auch aufgrund der zugehörigen Geräuschkategorie.

Des Weiteren lässt sich in diesem Modell ein insgesamt großer Einfluss der zeitlichen Variabilitäten der Tonalität ($\beta = 0.15$; $p < .001$), der Lautheit ($\beta = -0.08$; $p < .001$) und der Schärfe ($\beta = 0.08$; $p < .001$) feststellen, sodass die Nullhypothese zu H1, dass kein signifikanter Zusammenhang zwischen Zeitschwankungen und Ereignisreichtum besteht, weiterhin verworfen werden kann. Allerdings scheint das Vorzeichen des Effekts geräuschabhängig zu sein.

Fragen wirft folgender Befund der Schwankungsstärke auf: Das 95. Perzentil zeigt einen deutlichen positiven Effekt ($\beta = 0.14$; $p < .001$), während das 90. Perzentil einen negativen Effekt aufweist ($\beta = -0.08$; $p < .001$). Hier scheinen die beiden Features für Geräuschszenen des vorliegenden Audiodatensatzes über die Differenz von 5 % in der Perzentilberechnung stark unterschiedliche Werte anzunehmen, was zu den gegensätzlichen Richtungen der Effekte führt.

EV_RA: Ereignisreichtum & RA-Featureset

Das EV_RA-Modell ($R^2_{marginal} = 0.23$) weist für das 95. Perzentil der Werte der 1/3-Oktavfilter-Basisanalyse und Zeitmuster-Regression eine Effektstärke von $\beta = 0.68$ ($p < .001$) für Ereignisreichtum auf. Das Feature reagiert sensitiv auf andauernde, gröbere, d. h. vom Terzbandfilter (1/3-Oktave) erfassbare, Frequenzmusteränderungen und liefert Maximalwerte bei im Raum wiedergegebener Musik und lauter Sprache.

Darüber hinaus wurden drei Perzentile sowie drei zeitliche Schwankungen in Relative Approach-Features selektiert. Die RA-Grundfunktion 1/12F_RT (feine Zeitmusterauflösung) wurde mit den Perzentilen P10 und P95 selektiert: Diese Grundfunktion detektiert vor allem zeitliche Muster. Kurzzeitig auftretende Änderungen oder Brüche mit Mustern dieser Basisanalyse dämpfen Ereignisreichtum also gleichermaßen wie die Dauerwerte des Features.

EV_SEL: Ereignisreichtum & signifikante Prädiktoren

Die signifikanten Prädiktoren der Einzelmodelle wurden wiederum in einem Perzentilen LASSO in ein umfassendes Modell selektiert. Durch dieses Vorgehen lässt sich die Varianzaufklärung durch die festen Effekte auf 27 % steigern, während der Betrag durch die zufälligen Effekte weiterhin mit 15 % zu Buche schlägt. Die sieben stärksten Effekte entstammen erwartungskonform dem PSY- und RA-Featureset, deren Prädiktoren bereits die Einzelmodelle mit den höchsten marginalen R^2 geliefert hatten.

Die zwei mit Abstand stärksten Effekte machen das erste Perzentil des C-bewerteten Schalldruckpegels ($\beta = 0.25$; $p < .001$) und das 95. Perzentil des Relative Approach mit terzbandgefilterter Zeitmusterregression ($\beta = 0.21$; $p < .001$) aus. Dies korrespondiert mit besonders dynamischen Soundscapes, die z. B. dynamische oder laute Musik und Sprache oder laute Kurzzeitgeräusche wie z. B. Poltern enthalten.

Effektstärken ähnlicher Größenordnungen können mit positivem Vorzeichen für die zeitlichen Schwankungen der Tonalität ($\beta = 0.12$; $p < .001$) und Schärfe ($\beta = 0.09$; $p < .001$) sowie in negativer Ausprägung für die Schwankung von Prominence im Relative Approach ($\beta = -0.10$; $p < .001$) angegeben werden. Die Zeitschwankungen von Schärfe und Tonalität sind vor allem groß für Soundscapes mit Gesprächen mehrerer Sprecher, lebendigem Wasserplätschern (z. B. am Waschbecken), aber auch für außergewöhnlichere tonale Geräuschereignisse wie chaotisches Vogelgezwitscher oder festinstallierte Sirenen. Der Effekt des Prominence-Features wirft hingegen Fragen auf und soll in der Diskussion gesondert betrachtet werden.

Darüber hinaus wurden das 75. Perzentil der Tonalität ($\beta = -0.06$; $p < .001$) und das 99. Perzentil der Schärfe ($\beta = -0.06$; $p < .001$) selektiert. Die verbleibenden sechs Features der Cepstral Coefficients, sowie das Keyclarity-Feature machen jeweils Effektstärken aus, deren Betrag unter 0.05 liegt.

Die zusätzlich getestete Dateigröße der Audiodateien nach Konvertierung in das MP3-Format wurde zwar im Perzentilen LASSO in EV_SEL selektiert, zeigte aber keinen statistisch signifikanten Effekt ($p = .068$). Für PL_SEL wurde das Feature im Perzentilen LASSO nicht selektiert.

Grundsätzlich ist auffällig, dass die Varianzaufklärung der festen Effekte aller Modelle für Angenehmheit deutlich schlechter ist als die der Modelle für Ereignisreichtum. Sie bleibt im besten Modell der selektierten Prädiktoren für Angenehmheit unter 10 %, während bei Ereignisreichtum 27 % erreicht werden. Im Sinne der Forschungsfrage F3 ist ein deutlicher Rückstand bei Angenehmheit gegenüber Ereignisreichtum festzustellen.

DISKUSSION

Die Ergebnisse dieser Arbeit ermöglichen einen Einblick in den Zusammenhang von signal- bzw. geräuschspezifischen Eigenschaften von Soundscapes und deren Bewertungen durch Proband*innen. Im Folgenden werden zunächst Betrachtungen zum Vergleich des BA- und des MIR-Featuresets angestellt. Anschließend soll die Bedeutung der Cepstral Features in dieser Arbeit bewertet werden. Schließlich werden die zentralen Befunde der Modelle im Hinblick auf ihre inhaltlichen Implikationen für Indoor Soundscapes diskutiert. Anhand dessen soll nachgezeichnet werden, welche klanglichen Soundscape-Eigenschaften zu Bewertungstendenzen von Angenehmheit und Ereignisreichtum führen und welche Schlussfolgerungen auch im Hinblick auf die aktuelle pandemische Situation gezogen werden können.

VERGLEICH BA VS. MIR

Das MIR-Featureset wurde in der vorliegenden Arbeit als „Out of the Box"-Lösung eingesetzt, die im Verlgeich mit dem BA-Featureset als Baseline diesen sollte. Obwohl das BA-Featureset mit 1560 Features, besonders im Hinblick auf die Anzahl der Aggregationsfunktionen, um ein Vielfaches umfangreicher als das MIR-Featureset ist, zeigt es gemessen an den festen Effekten in Modellen keine hervorragende Leistung. Zwar scheint es Tendenzen bezüglich besserer Eignung der BA-Aggregationsfunktionen bei Ereignisreichtum zu geben, bei Angenehmheit ist es jedoch umgekehrt und in PL_SEL sind insgesamt mehr MIR-Features vorhanden. Die festen Effekte der BA- und MIR-Modelle sind für beide Zielgrößen fast gleich groß.

Vor diesem Hintergrund könnten Entscheidungen für eines der beiden Featuresets auch anhand nicht inhaltlicher Faktoren getroffen werden: So kann bei zukünftiger Entscheidungsfindung für die Wahl von Featuresets mit einfließen, dass die MIRToolbox umfangreich und fachlich fundiert dokumentiert ist. Ein großer Nachteil, besonders für große Datenmengen, ist, dass die MIRToolbox aktuell keine direkt nutzbare Parallelisierbarkeit der Rechenprozesse bietet. Auf der an-deren Seite sind im BA-Featureset die Berechnungen zur Rekonstruktion der Phasenräume sehr zeitaufwendig. Die Effizienz wird auch durch die hohe Anzahl der Features verschlechtert, weil sich die Rechendauer des LASSO-Algorithmus stark erhöht. Die Anzahl der Features des Datensatzes geht dabei sogar kubisch in die Rechenkomplexität ein (Efron et al., 2004).

Cepstral Coefficients

Durch das Hinzufügen der Cepstral Features MFCC bzw. GTCC in die Modelle verbessert sich das $R^2_{marginal}$ von PL_SEL bzw. EV_SEL nicht additiv um den Betrag des $R^2_{marginal,}$ der für Pl_BA bzw. EV_BA erreicht wurde. Die Prädiktionsleistung von Cepstral Features fällt insgesamt geringer aus als in anderen Studien (Yin et al., 2011). Eine Erklärung dafür könnte eine Schnittmenge der z. B. durch psychoakustische Größen und Cepstral Features erklärten Varianz sein. Somit verbessern sich die Modelle der selektierten Prädiktoren (SEL) nur um den Betrag der Varianzaufklärung, die nicht schon von den anderen Features mit größeren Effektstärken geleistet wird. Sie bringen also nicht viel mehr neuen Effekt in das Modell ein bzw. erklären fast nur diejenige Varianz, die schon von den anderen Prädiktoren erklärt wird. Allerdings scheinen die selektierten MFCC-Features auch nicht redundant zu sein, weil sie allen Prüfungen der Modellfehler und der 1SE-Regel der LASSO-Regularisierung standhielten und ebenfalls signifikante Effekte haben. Eine Erklärung könnte die Transformation ins Cepstrum sein, die die Features von den effektstärkeren psychoakustischen Messgrößen dekorreliert (Oppenheim & Schafer, 2004). Auch wenn sie den Informationsgehalt des Modells nicht maßgeblich steigern, aber von den effektstärkeren psychoakustischen Messgrößen unabhängige Verteilungen besitzen, können sie dazu beitragen, Rauschen zu unterdrücken und sind dann nicht im eigentlichen Sinne redundant (Guyon & Elisseeff, 2003). Die in Studien berichtete Überlegenheit von GTCC gegenüber MFCC kann jedoch anhand der vorgestellten Modelle nicht bestätigt werden (Valero & Alias, 2012).

Bemerkenswert ist die Selektion der neuen Features mit Aggregation durch Lokale-Maxima-Rate und lmPACF-Funktion in den Modellen für Ereignisreichtum, die in der Aggregation der MFCC-Grundfunktionen sehr gut rauschartige Signale und Geräusche beschreiben können, deren Werte durch Periodizität oder andere Muster stark von vorausliegenden Zeitwerten abhängen. Sie erweisen sich als Prädiktoren, die durch leichte bis moderate Effekte das Gesamtmodell unterstützen und vergrößern. Die Dekorrelation zu unmittelbar vom Signalpegel abhängigen Größen wie dem Schalldruckpegel scheint dabei zentral zu sein.

Angenehmheit

Mit Blick auf die Forschungsfrage F3 zeigt die vorliegende Studie, dass sich Ereignisreichtum besser modellieren lässt als Angenehmheit. Dabei ist auffällig, dass die zufälligen Effekte allein, also jene, die durch die Proband*innen erklärbar sind, selbst im besten Modell (PL_SEL)

mehr als doppelt so viel Varianz wie die Audiofeatures erklären (ICC = 0.22; $R^2_{marginal} = 0.09$). Es wird deutlich, dass die hohen in Laborstudien berichteten Modellstatistiken und Korrelationen in heterogenen Alltagssituationen mit komplexen Soudscape-Zusammensetzungen nicht reproduziert werden können, wenn generalisierbare Verfahren für die Feature-Selektion herangezogen werden. Dieser Befund deckt sich mit aktuellen Studien, die über komplexe Situation des Wohnalltags, wenn auch nicht unter Einbeziehung von Audiofeatures, geforscht haben (Versümer et al., 2020; Erfanian et al., 2021; Torresin et al., 2021). Eine Erklärung für die auffällig niedrigeren Modellwerte der Angenehmheit könnte darin bestehen, dass Angenehmheit, einem Alltagsverständnis folgend, deutlich stärker mit Emotionen verknüpft ist als Ereignisreichtum, der emotionsneutraler ist. Eine Veränderung der Gesamtstimmung durch externe emotionale Einflussfaktoren, die nicht Gegenstand dieser Studie waren, könnte bei der Abgabe der Bewertungen einen Fehler eingebracht haben (Steffens et al., 2017).

Dennoch lassen im Modell PL_SEL signifikante Effekte bei der Betrachtung von Angenehmheit feststellen: Kongruent mit den Erwartungen durch bisherige Studien (Tordini, 2014; Dokmeci Yorukoglu & Kang, 2017; Mohamed & Dokmeci Yorukoglu, 2020) zeigt sich ein Rückgang der Angenehmheit durch steigende durchschnittliche Lautheit der Soundscapes. Das entspricht grundsätzlich der funktionalen Anlage der Messgröße nach ISO 532-2 (ISO International Organization for Standardization, 2017b). Durch den empirischen Beleg in dieser Arbeit kann die Hypothese H3 somit als bestätigt gelten.

Weitere negative Effekte auf Angenehmheit gehen von Geräuschen aus, die starkes Brummen und im höheren Frequenzbereich starke rauschhafte Komponenten enthalten, wie es für haushaltstypische Maschinen, Anlagen, Belüftungen bzw. Geräusche der Anwendung der genannten, zutrifft (z. B. Bratgeräusche in der Küche; vgl. Ergebnisse PL_SEL: Lokale-Maxima-Raten der MFCC). Dies geht einher mit den Ergebnissen zu den besonders störenden Geräuschkategorien der Online-Befragung bei Versümer et al. (2020).

Weiter zeigt sich, dass hohe Dauerwerte in besonders musiksensitiven Features (ra_HM_P95) die bewertete Angenehmheit deutlich steigern. Wenn dieses Feature Musik anzeigt, wird auch die Angenehmheit tendenziell steigen. Dass abgespielte Musik eine wichtige und zwar angenehm bewertete Geräuschkategorie ist, ist zuvor von Versümer et al. (2020) belegt worden. Dieser Effekt ist ebenfalls in den Daten der aufgezeichneten Soundscapes messbar. Da die Daten der vorliegenden Feldstudie während der SARS-CoV-2-Pandemie erhoben

wurden, einer Zeit, in der die Menschen im beruflichen wie im privaten Leben und ganz besonders auch in Home-Office-Situationen außergewöhnlichen psychologischen Belastungen ausgesetzt waren, darf Musik als Regulator in der Stressbewältigung (MacDonald, 2013) keinesfalls unterschätzt werden. Dem berichteten Befund kommt daher, übereinstimmend mit einschlägigen Befragungen (Torresin et al., 2022), besondere Bedeutung zu.

Das Modell PL_NoMusic zeigt aber durch systematischen Ausschluss der Observationen mit vordergründiger Musik, dass die Lautheit dieser vordergründigen Musik die Modellierung von Angenehmheit nicht maßgeblich verzerrt. Das scheint mit ein Grund dafür zu sein, dass die LASSO-Selektion unterschiedliche lautheitsbasierte Features mit nicht identischen Verteilungen in das Modell PL_SEL aufgenommen hat.

Anhand der Multilevel-Modellierung mit den Teilmengen für laute und leise Soundscapes wurde überdies aufgezeigt, dass in den lautesten 25 % der observierten Tonaufnahmen 8 % mehr Varianz als in den leiseren 75 % der Tonaufnahmen erklärt werden kann. Des Weiteren ist die Anzahl der signifikanten selektierten Prädiktoren der beiden Teilmengen-Modelle nicht gleich. Diese Modellunterschiede der beiden Lautheitslevels zeigen, dass über die Lautheit ein sinnvolles Clustering des Datensatzes erreicht werden kann.

Der starke Abfall der festen Effekte für leise Soundscapes deutet auf einen grundsätzlichen Verlust des Aussagegehalts der Audiofeatures für leise und sehr leise Soundscapes hin, während die Varianz zwischen den Probanden für das Nullmodell von PL_N_low signifikant höher ist als für PL_N_high ($\tau_{00,\,low} = 0.26$; $\tau_{00,\,high} = 0.15$). Zu geringer Lautheit hin können die allgemein messbaren Größen weniger Varianz erklären, während der Einfluss von persönlicher Einschätzung durch Proband*innen größer wird.

Vor allem bei der Varianzaufklärung der Angenehmheit für die leise Teilmenge von Indoor Soundscapes ist Anschlussforschung zu leisten. Denkbar wäre, dass in Modellen der leisen Soundscapes bestimmte situative oder persönliche Faktoren hervortreten, welche die höheren τ_{00}-Werte erklären. Zudem könnte die Frage untersucht werden, ob die bei Dedieu et al. (2019) gefundenen Unterschiede einer tendenziell lärmempfindlicheren und einer tendenziell lärmtoleranteren Probandengruppe auch in den vorliegenden Daten eine Rolle spielen.

Ereignisreichtum

Die Varianz von Ereignisreichtum ist durch alle in dieser Arbeit eingesetzten Featuresets hinreichend gut zu beschreiben. Insgesamt ist aber ein klarer Schwerpunkt bei psychoakustischen Features zu sehen: Insbesondere fallen jene Features auf, die über niedrige Perzentile zeitlich kurze Spitzenwerte beschreiben (psy_SPLC_P1) oder im Zusammenhang mit hoher zeitlicher Schwankung stehen (psy_Tonality_P1-P99). Hohe Werte werden vor allem für lautes Klappern, chaotisches Staubsaugen und Sprache mit starken Plosivlauten erreicht. Ebenso korrespondieren hohe Werte dieser Features z. B. mit Signaltönen in Soundscapes, in denen auch weitere Quellen überlagert auftreten. Die Modelle identifizieren auf Grundlage der psychoakustischen Messgrößen sehr zuverlässig die komplexen und lebhaften Soundscapes, sodass mit Bezug auf die explorative Forschungsfrage F2 eine gute Eignung im Bereich der ereignisreichen Soundscapes bestätigt werden kann.

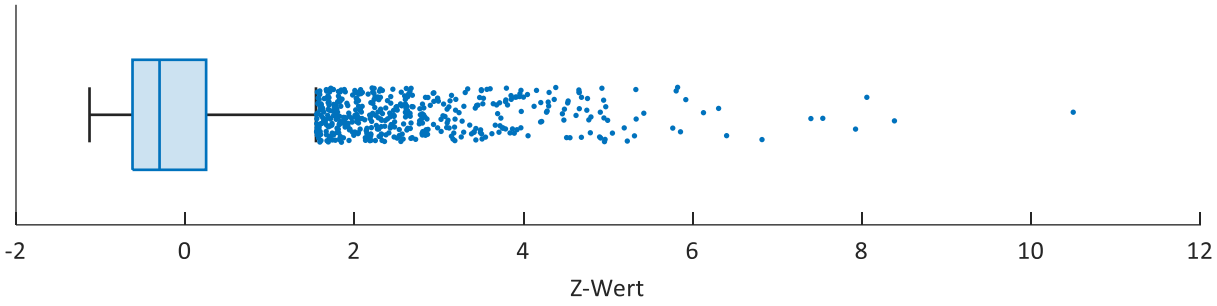
Die Selektion der Zweitschwankungsfeatures deutet darauf hin, dass die Hypothese H2 zum Zusammenhang von Zeitschwankungen in Features und Ereignisreichtum zutrifft. Allerdings kann bei den vorliegenden Häufigkeiten keine Überzufälligkeit nachgewiesen und die Nullhypothese nicht verworfen werden.

Starke Effekte wurden auch für die Features des Relative Approach festgestellt. Die Fähigkeit dieser spezialisierten Berechnungsmethoden, die Abweichung von Mustern im Frequenz- und Zeitbereich messbar zu machen, erweist sich für komplexe Geräuschumgebungen als geeignet. In Kombination mit herkömmlichen psychoakustischen Größen erreichen sie starke Modellstatistiken.

Allerdings wirft ein Feature des Relative Approach, ra_Pr1_P1-P99, im Modell EV_SEL durch die Richtung seines Effekts Fragen auf: Das Feature bildet als Zeitschwankungsmaß die Differenz aus den in 1 % und in 99 % der Zeit erreichten Werten der Prominence-Grundfunktion. Alle anderen vergleichbaren selektierten Zeitschwankungen psychoakustischer Messgrößen haben aber entgegengesetzte Effekte. Das kann anhand der Tonaufnahmen nicht erklärt werden, für die ra_Pr1_P1-P99 Maximalwerte annimmt: Audiodateien mit Soundscapes, die durchaus ereignisreich erscheinen (bspw. lebhafte Atmosphäre an geöffnetem Fenster), erreichen hohe Werte. Allerdings sind die Maximalwerte des Features auch extreme Ausreißer. Über die Verteilung und insbesondere die Ausreißer gibt der Boxplot in Abbildung 4

Aufschluss. Als besonders ist dabei festzustellen, dass zwei von fünf Dateien mit Höchstwerten des Features elektromagnetische Einstreuungen durch den Mobilfunk enthalten. Das Feature scheint gewissermaßen ein Indikator dieser klaren Frequenzmuster zu sein, die nicht Teil der akustischen Umgebung waren und nicht von Proband*innen bewertet werden konnten. Das verzerrt die inhaltliche Aussagekraft bzw. die Interpretierbarkeit der Maxima an dieser Stelle und zeigt, dass die Audiofeatures auch den in den Datensatz eingetragenen Fehler mitrepräsentieren müssen.

ABBILDUNG 4: BOXPLOT DER Z-STANDARDISIERTEN FEATUREWERTE VON RA_PR1_P1-P9. (INNENLINIE IN BOX = MEDIAN; BREITE DER BOX = IQR; BOX-AUßENKANTEN = 25 % BZW. 75 % QUARTIL; WHISKER = MINIMA/MAXIMA MIT HÖCHSTENS 1.5 IQR ABSTAND ZUR BOX; AUSREIßER (PUNKTE) = DATENPUNKTE MIT ABSTAND > 1.5 IQR ZUR BOX; AUSREIßER WERDEN ZUR VERDEUTLICHUNG MIT JITTER DARGESTELLT)



Würde aber das Feature über seinen Großteil der Werte zu keiner guten Vorhersageverbesserung beitragen, hätte es in den Kreuzvalidierungen des Perzentilen LASSO hohe Modellfehler erzeugt und wäre nach der 1SE-Regel nicht selektiert worden.

Übersetzt man die gefundenen Effektstärken von EV_SEL sprachlich zu einer Soundscape, so werden diejenigen alltäglichen Geräuschumgebungen ereignisreich bewertet, die vielschichtige Überlagerungen von dynamischen Geräuschen wie Sprache und impulshafte Geräusche des Wohnens wie z. B. Klappern oder Rücken von Stühlen enthalten. Das zeigen die positiven Effekte durch 1 %-Werte des C-bewerteten Schalldruckpegels sowie die Schwankungen der Schärfe und der Tonhaltigkeit. Weniger ereignisreich bewertet werden dagegen rauschhafte, technische Schallquellen mit geringer zeitlicher Variabilität oder insgesamt geringerer Dynamik, was an den negativen Effekten durch Dauerpegel der Schärfe und durch die mit Lokaler-Maxima-Rate aggregierten Features der höheren MFCC-Ränge deutlich belegt werden kann.

Torresin et al. (2021) berichten für Indoor Soundscapes einen leicht negativen Zusammenhang zwischen Wohlbefinden und Ereignisreichtum, der auch bei Erfanian et al. (2021) bestätigt wird. Daher kann mit EV_SEL auch angedeutet werden, inwiefern die genannten Geräuschquellen der Soundscapes dem Wohlbefinden zu- oder abträglich sind.

Wie schon bei der Diskussion zu den Modellen der Angenehmheit wird übereinstimmend mit dem Forschungsstand deutlich, welche Rolle den Geräuschkategorien zukommt (Versümer et al., 2020). Dieser Befund ist wichtig, weil durch Mitmenschen im Wohnumfeld verursachte Geräusche im Gegensatz zu den durch Haushaltsgeräte verursachten nicht einfach ‚abgeschaltet' werden können. Als Regulator bleibt dann die persönliche Fähigkeit, Störschalle mental auszublenden, was einen der größten Faktoren bei der Bewertung von Soundscapes als störend oder nervig ausmacht (Versümer et al., 2020). Erkenntnisse über die Peak-End-Rule und ihre Übertragung auf das Soundscape-Konzept zeigen weiterhin, wie entscheidend einzelne markante Geräuschereignisse die Aufmerksamkeit selbst in Anwesenheit von dauerhaften Störschallen auf sich ziehen können (Västfjäll, 2004; Steffens & Guastavino, 2015). Die Peak-End-Rule erklärt die Wichtigkeit des Effekts durch hohe Schalldruckpegel, die nur in einem Prozent der Zeit gemessen werden.

Das Modell EV_SEL bestätigt die in früheren Laborversuchen (Axelsson et al., 2010) gefundenen Zusammenhänge von Ereignisreichtum und Zeitvariabilitäten psychoakustischer Messgrößen für komplexe Indoor Soundscapes. Demgegenüber kann EV_SEL durch feste Effekte zwischen fünf und zwanzig Prozentpunkten mehr Varianzaufklärung leisten, als Modelle aktueller empirischer Studien, in denen Ereignisreichtum z. B. anhand situativer, persönlicher und demographischer Faktoren modelliert wird (Erfanian et al., 2021; Torresin et al., 2022).

Offene Befragungen zeigen, dass Soundscapes für Home-Office-Situationen mehr als doppelt so oft mit den Eigenschaften störend oder ablenkend assoziiert werden als für reine Wohnsituationen (Torresin et al., 2021). Die Ergebnisse dieser Arbeit reichen deshalb nicht nur weit in die bauakustische Planung und die Stadtplanung hinein, sondern zeigen auch Konfliktpotentiale auf, die in Zeiten gesetzlicher Maßnahmen zur Eindämmung der SARS-CoV-2-Pandemie und den damit einhergehenden zeitlich oder räumlich unvermeidlichen Konfrontationen in Familie und Nachbarschaft offenbar werden können.

LIMITATIONEN

Mit Bezug auf das Design der vorliegenden Studie muss diskutiert werden, welche Vor- und Nachteile es für die Studienergebnisse gehabt haben könnte, die Möglichkeit einer initiativen Befragungseinleitung zu implementieren, anstatt die Feldstudie auf die automatisierte ESM-Befragung zu beschränken. Um Proband*innen die Möglichkeit zu geben, die Studie um ihre ‚Pflichttermine' des Tages herum durchzuführen, war eine initiative Funktion wichtig. Das wird der Gesamtzahl an Observationen zugutegekommen sein. Proband*innen könnten andererseits zu einem gewissen Anteil Geräusche bewusst ‚gesammelt' haben, anstatt eine für sie zufällige Stichprobe des Tagesgeschehens zuzulassen. Diese Haltung ließe sich nach Botteldooren et al. (2011) mit einem ‚suchenden' Hörstil beschreiben. Dieses trotz klarer Instruktionen unvermeidbare Proband*innenverhalten kann das nicht modellierbare Rauschen des Datensatzes erhöht haben. Zu viele auf diese Weise eingebrachte Observationen (d. h. nicht nur initiativ abgegeben, sondern auch mit der Implikation, besonders unangenehme Geräusche für die Studie zu ‚sammeln') könnten die hohe Varianz des Nullmodells für PL_SEL erklären.

Des Weiteren muss betrachtet werden, dass die hier vorgestellte Feldstudie sich an Proband*innen richtete, die sich zeitlich und organisatorisch dazu in der Lage sahen, regelmäßig Tonaufnahmen am Tag anzufertigen. Diese Voraussetzung war wichtig, um das Versuchsequipment gut auszulasten, effektiv an Proband*innen zu verteilen und eine insgesamt hohe Anzahl an Proband*innen und Observationen in die Feldstudie einbeziehen zu können. Zwar herrschte im Erhebungszeitraum der Feldstudie eine recht hohe Home-Office-Quote, jedoch ist eine Diskrepanz in verschiedenen Berufsständen zu verzeichnen (Corona Datenplattform, 2021). So wurden bestimmte Berufsgruppen eher ausgeschlossen, die bei einem Versuchsdesign mit längerer individueller Erhebungsdauer und geringerer Zahl von Observationen pro Tag durchaus hätten teilnehmen können. Dazu könnten bspw. Betroffene von Schicht- und Bereitschaftsdiensten oder Arbeiter*innen gehören (Corona Datenplattform, 2021).

Eine Limitation im Hinblick auf die Aussagefähigkeit der Audiofeatures entsteht in der vorliegenden Arbeit durch eine fehlende ‚Ground Truth': Zwar ist der vorliegende Datensatz sehr groß, allerdings ist jede Soundscape der Methode entsprechend immer nur von genau einer Proband*in bewertet worden, sodass der Einfluss von beiläufig, unwahrheitsgemäß oder beliebig beantworteten Items immer voll in die statistischen Zusammenhänge jeder Observation

eingegangen sein muss. Solche zufälligen Observationen können nicht vorhergesagt werden und senken den Betrag der insgesamt in Modellen erklärbaren Varianz. Folgeuntersuchungen könnten eine Vergleichsstudie durchführen, in denen Proband*innen die hier aufgezeichneten Tonaufnahmen als Stimuli in einem Hörversuchslabor anhören und bewerten. Das kann aufklären, welcher Zufallseffekt von den Proband*innen dieser Arbeit ausgeht und wieviel Varianz tatsächlich durch Rauschen im Datensatz unerklärbar bleibt.

Offen bleibt weiterhin die Frage, warum die Audiofeatures in den ganz leisen Dateien des Datensatzes deutlich weniger Varianz erklären können. Unter Umständen ist in diesen Dateien kaum Information gespeichert. Für die Forschung zu ganz leisen Soundscapes müssen weitere Versuchsmethoden entwickelt werden. Die Bildung einer ‚Ground Truth' könnte auch hier ein entscheidender Faktor sein.

ZUSAMMENFASSUNG

Die vorliegende Arbeit hat sich zum Ziel gesetzt, das Portfolio bekannter klanglicher und akustischer Einflussgrößen in der Soundscape-Wahrnehmung zu erweitern. Dazu wurde die Eignung von über 2000 Audiofeatures für die Vorhersage von Angenehmheit und Ereignisreichtum an 6594 Indoor Soundscapes getestet. Die Umfragedaten sind mit der zeitgemäßen Experience Sampling Method in einer Feldstudie mit 105 Proband*innen erhoben worden. Die erstmalige Anwendung dieser systematisch kontrollierten Befragungstechnik in Kombination mit Tonaufzeichnung von Indoor Soundscapes erzeugt einen höchst umfangreichen Datensatz, der auch der weiteren Anschlussforschung große Potentiale bietet.

Aus vier großen Featuresets wurden mithilfe der Perzentilen LASSO-Regularisierung relevante, generalisierbare Prädiktoren selektiert und in linearen gemischten Modellen auf ihre Effektstärken, Signifikanz und die Varianzaufklärung der Soundscape-Dimensionen hin überprüft. Die Modelle zeigen klar, dass in komplexen empirischen Soundscapes 27 % der Varianz von Ereignisreichtum durch feste Effekte der Audiofeatures zu erklären sind, wohingegen 9 % der Varianz von Angenehmheit durch feste Effekte erklärt werden konnten. Für die Angenehmheit des lautesten Viertels der Soundscapes ließen sich 13 % der Varianz erklären. Angenehmheit wird grundsätzlich stark von lautheitsbasierten Features beeinflusst, während Ereignisreichtum maßgeblich von kurzzeitig auftretenden Spitzenwerten sowie zeitlichen Schwankungen in den Features bestimmt wird. Die Studie zeigt allein anhand der Audiofeatures, welche große

Bedeutung den Geräuschkategorien bei der Bewertung der wahrgenommenen Qualität von Indoor Soundscapes zukommt.

Es ist deutlich geworden, dass die Modelle für Angenehmheit und Ereignisreichtum durch konsequente Methoden der Audiofeature-Extraktion selbst bei Anwendung strengster Selektionsalgorithmen an Varianzaufklärung gewinnen. Dabei lässt sich trotz komplizierter Funktionsverkettungen in den Features ein plausibles Bild der Soundscapes zeichnen. Es kann nachgewiesen werden, dass Musik neben sozialer Interaktion eine entscheidende Rolle im häuslichen Alltag spielt, die Angenehmheit positiv beeinflusst. In Zeiten der SARS-CoV-2-Pandemie, in der das häusliche Verhalten durch vermehrte Nutzung der Räume als Home-Office teils drastische Änderungen erfahren hat, kommt den Ergebnissen dieser Studie eine besondere Bedeutung zu. Sie fügen sich schlüssig in ein Gesamtbild von Indoor Soundscapes zu Pandemie-Zeiten, in denen bewusst erzeugte angenehme Geräusche Entspannungs- und Stressbewältigungsstrategie sein können, während ereignisreiche Geräusche auch im Hinblick auf geändertes Sozial- und Arbeitsverhalten Potential für Störung haben (Torresin et al., 2022). Bauakustische sowie städtebauliche Planung sollte diesen genannten Aspekten im Sinne einer positiven Wohn- und flexiblen Arbeitsumgebung Rechnung tragen: Wohnräume müssen die Möglichkeit bieten, positive, angenehme Indoor Soundscapes bewusst zu erzeugen, wobei Störungen durch technische und maschinelle Geräusche reduziert werden sollten. Mit Blick auf die pandemischen Herausforderungen kommt in Mehrpersonenhaushalten und Mehrparteienhäusern der Schaffung von Balance zwischen freier Entfaltung im Wohnraum einerseits und dem Gefühl der Ungestörtheit und Unabhängigkeit von nicht kontrollierbaren akustischen Geräuschumgebungen andererseits große Bedeutung zu.

LITERATURVERZEICHNIS

Aboofazeli, M. & Moussavi, Z. (2005). Analysis and Classification of Swallowing Sounds using Reconstructed Phase Space Features. *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, 421–424. https://doi.org/10.1109/ICASSP.2005.1416330

Aletta, F., Axelsson, Ö. & Kang, J. (2017). Dimensions Underlying the Perceived Similarity of Acoustic Environments. *Frontiers in Psychology*, *8*, 1162. https://doi.org/10.3389/fpsyg.2017.01162

Aletta, F. & Kang, J. (2018). Towards an Urban Vibrancy Model: A Soundscape Ap-proach. *International Journal of Environmental Research and Public Health*, *15*(8), 1712. https://doi.org/10.3390/ijerph15081712

Aletta, F., Kang, J. & Axelsson, Ö. (2016). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*, *149*, 65–74. https://doi.org/10.1016/j.landurbplan.2016.02.001

Alías, F., Socoró, J. & Sevillano, X. (2016). A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Applied Sciences*, *6*(5), 143. https://doi.org/10.3390/app6050143

Aucouturier, J.-J., Defreville, B. & Pachet, F. (2007). The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *The Journal of the Acoustical Society of America*, *122*(2), 881–891. https://doi.org/10.1121/1.2750160

Axelsson, Ö., Nilsson, M. E. & Berglund, B. (2010). A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, *128*(5), 2836–2846. https://doi.org/10.1121/1.3493436

Babisch, W. (2002). The Noise/Stress Concept, Risk Assessment and Research Needs. *Noise & health*, *4*(16), 1–11.

Barchiesi, D., Giannoulis, D., Stowell, D. & Plumbley, M. D. (2015). Acoustic Scene Classification: Classifying environments from the sounds they produce. *IEEE Signal Processing Magazine*, *32*(3), 16–34. https://doi.org/10.1109/MSP.2014.2326181

Benko, T. P. & Perc, M. (2009). Nonlinearities in mating sounds of American crocodiles. *Bio Systems*, *97*(3), 154–159. https://doi.org/10.1016/j.biosystems.2009.05.011

Berglund, B. (2006). From the WHO Guidelines for community noise to healthy soundscapes. *Proceedings of the Institute of Acoustics, 2006*, 1–9.

Beutel, M. E., Jünger, C., Klein, E. M., Wild, P., Lackner, K., Blettner, M., Binder, H., Michal, M., Wiltink, J., Brähler, E. & Münzel, T. (2016). Noise Annoyance Is Associated with Depression and Anxiety in the General Population- The Contribution of Aircraft Noise. *PloS one*, *11*(5), e0155357. https://doi.org/10.1371/journal.pone.0155357

Bogert, B. & Ossanna, J. (1966). The heuristics of cepstrum analysis of a stationary complex echoed Gaussian signal in stationary Gaussian noise. *IEEE Transactions on Information Theory*, *12*(3), 373–380. https://doi.org/10.1109/TIT.1966.1053903

Botteldooren, D. & Coensel, B. de (2009). The role of saliency, attention and source identification in soundscape research. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, *2009*(5), 2198–2205.

Botteldooren, D., Lavandier, C., Preis, A., Dubois, D., Aspuru, I., Guastavino, C., Brown, L., Nilsson, M. & Andringa, T. C. (2011). Understanding urban and natural soundscapes. *Forum Acusticum 2011*. 2047–2052. European Accoustics Association (EAA).

Bountourakis, V., Vrysis, L. & Papanikolaou, G. (2015). Machine Learning Algorithms for Environmental Sound Recognition. In G. Kalliris & C. Dimoulas (Hrsg.), *Proceedings of the Audio Mostly 2015 on Interaction With Sound - AM ,15.* 1–7. ACM Press. https://doi.org/10.1145/2814895.2814905

Brambilla, G., Gallo, V., Asdrubali, F. & D'Alessandro, F. (2013). The perceived quality of soundscape in three urban parks in Rome. *The Journal of the Acoustical Society of America*, *134*(1), 832–839. https://doi.org/10.1121/1.4807811

Brocollini, L., Waks, L., Lavandier, C., Marquis-Favre, C., Quoy, M. & Lavandier, M. (2010). Comparison between multiple linear regressions and artificial neural networks to predict urban sound quality. *Proceedings of 20th International Congress on Acoustics, ICA 2010*.

Childers, D. G., Skinner, D. P. & Kemerait, R. C. (1977). The cepstrum: A guide to processing. *Proceedings of the IEEE*, *65*(10), 1428–1443. https://doi.org/10.1109/proc.1977.10747

Coensel, B. de (2010). A model of saliency-based auditory attention to environmental sound. *20th International Congress on Acoustics (ICA-2010)*, 1–8.

Corona Datenplattform (2021). Homeoffice im Verlauf der Corona-Pandemie. *Themenreport 02*(2). Bonn.

Dedieu, R., Lavandier, C., Camier, C. & Berger, S. (2019). Pleasantness of typical acoustic environments inside a living room in a European residential context. *International Congress of Acoustics.* https://hal.archives-ouvertes.fr/hal-02489915/

DIN Deutsches Institut für Normung e.V. (2009). *DIN 45692 - Messtechnische Simulation der Hörempfindung Schärfe* (DIN 45692:2009-08). Berlin. Beuth Verlag GmbH.

DIN Deutsches Institut für Normung e.V. (2018). *ISO 12913-1: Soundscape- Teil 1: Definition und Rahmenkonzept* (ISO 12913-1:2018-02). Berlin. Beuth Verlag GmbH.

DIN Deutsches Institut für Normung e.V. (2020). *DIN ISO 532-1 - Akustik – Verfahren zur Berechnung der Lautheit – Teil 1: Verfahren nach E. Zwicker (ISO 532-1:2017, korrigierte Fassung 2017-11);* (ISO 532-1). Berlin. Beuth Verlag GmbH.

DIN Deutsches Institut für Normung e.V. (2021). *ISO 12913-3: Soundscape - Teil 3: Datenanalyse* (ISO/TS 12913-3:2021-06). Berlin. Beuth Verlag GmbH.

Dokmeci Yorukoglu, P. N. & Kang, J. (2017). Development and testing of Indoor Sound-scape Questionnaire for evaluating contextual experience in public spaces. *Building Acoustics*, *24*(4), 307–324. https://doi.org/10.1177/1351010X17743642

Duangudom, V. & Anderson, D. (2007). Using Auditory Saliency to Understand Complex Auditory Scenes. *2007 15th European Signal Processing Conference*, 1206–1210.

Ecma International (2021). *ECMA 74: Acoustics: Measurement of airborne noise emitted by information technology and telecommunications equipment* (ECMA 74). London. BSI British Standards.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, *32*(2), 407–499.

Elsetrønning, A., Rasheed, A., Bekker, J. & San, O. (2020) On the effectiveness of signal decomposition, feature extraction and selection on lung sound classification. http://arxiv.org/pdf/2012.11759v1

Erfanian, M., Mitchell, A., Aletta, F. & Kang, J (2021). Psychological well-being and demographic factors can mediate soundscape pleasantness and eventfulness: A large sample study. *Journal of Environmental Psychology*, *77*, 101660. https://doi.org/10.1016/j.jenvp.2021.101660

Fan, J., Thorogood, M., Riecke, B. E. & Pasquier, P. (2015). Automatic Recognition of Eventfulness and Pleasantness of Soundscape, 1–6. https://doi.org/10.1145/2814895.2814927

A. Fiebig, S. Guidati & A. Goehrke. (2009). *Psychoacoustic Evaluation of Traffic Noise*. http://pub.dega-akustik.de/NAG_DAGA_2009/data/articles/000368.pdf

Filipan, K., Boes, M., Coensel, B. de, Domitrović, H. & Botteldooren, D. (2015). Identifying and recognizing noticeable sounds from physical measurements and their effect on soundscape. *10th European Congress and Exposition on Noise Control Engineering*, 1559–1564.

Fu, Z., Lu, G., Ting, K. M. & Zhang, D. (2011a). Music classification via the bag-of-features approach. *Pattern Recognition Letters*, *32*(14), 1768–1777. https://doi.org/10.1016/j.patrec.2011.06.026

Fu, Z., Lu, G., Ting, K. M. & Zhang, D. (2011b). A Survey of Audio-Based Music Classification and Annotation. *IEEE Transactions on Multimedia*, *13*(2), 303–319. https://doi.org/10.1109/TMM.2010.2098858

Gaunard, P., Mubikangiey, C. G., Couvreur, C. & Fontaine, V. (1998). Automatic Classification of Environmental Noise Events by Hidden Markov Models. *Applied Acoustics*, *54*(3), 187–206. https://doi.org/10.1016/s0003-682x(97)00105-9

Gauthier, P.-A., Scullion, W. & Berry, A. (2017). Sound quality prediction based on systematic metric selection and shrinkage: Comparison of stepwise, lasso, and elastic-net algorithms and clustering preprocessing. *Journal of Sound and Vibration*, *400*, 134–153. https://doi.org/10.1016/j.jsv.2017.03.025

Giannoulis, D., Benetos, E., Stowell, D., Rossignol, M., Lagrange, M. & Plumbley, M. D. (2013). Detection and Classification of Acoustic Scenes and Events: An IEEE AASP Challenge. *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. https://doi.org/10.1109/WASPAA.2013.6701819

Gunn, W. J., Patterson, H. P., Cornog, J., Klaus, P. & Connor, W. K. (1975). *A model and plan for a longitudinal study of community response to aircraft noise*. NASA Langley Research Center.

Guyon, I. & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of machine learning research*(3), 1157–1182.

Harte, C., Sandler, M. & Gasser, M. (2006). Detecting harmonic change in musical audio. In X. Amatriain, E. Chew & J. Foote (Hrsg.), *Proceedings of the 1st ACM workshop on Audio and music computing multimedia - AMCMM ,06,* 21–26. ACM Press. https://doi.org/10.1145/1178723.1178727

Hastie, T., Tibshirani, R. & Friedman, J. (2009). The Elements of Statistical Learning. Data Mining, Inference, and Prediction. *Cited on*, *33*.

HEAD Acoustics. (2018). *Relative Approach*. https://cdn.head-acoustics.com/fileadmin/data/de/Application-Notes/Relative-Approach-Analyse-02.2018.pdf

Hektner, J. M., Schmidt, J. A. & Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. SAGE.

Hong, J. Y. & Jeon, J. Y. (2015). Influence of urban contexts on soundscape perceptions: A structural equation modeling approach. *Landscape and Urban Planning*, *141*, 78–87. https://doi.org/10.1016/j.landurbplan.2015.05.004

ISO International Organization for Standardization (1996). *ISO 1996-1 - Acoustics: Basic quantities and assessment procedures* (ISO 1996-1). Vernier, Geneva. International Organization for Standardization.

ISO International Organization for Standardization (2017a). *ISO 1996-2:2017: Determination of sound pressure levels* (ISO 1996-2:2017). Vernier, Geneva. International Organization for Standardization.

ISO International Organization for Standardization (2017b). *ISO 532-2 2017 - Loudness nach Moore-Glasberg* (ISO 532-2). Vernier, Geneva. International Organization for Standardization.

ISO International Organization for Standardization (2018). *ISO 12913-2 Part 2: Data collection and reporting requirements* (ISO/TS 12913-2:2018(E). Vernier, Geneva. International Organization for Standardization.

Jeon, J. Y. & Hong, J. Y. (2015). Classification of urban park soundscapes through perceptions of the acoustical environments. *Landscape and Urban Planning*, *141*, 100–111. https://doi.org/10.1016/j.landurbplan.2015.05.005

Jeon, J. Y., Jo, H. in, Santika, B. B. & Lee, H. (2022). Crossed effects of audio-visual environment on indoor soundscape perception for pleasant open-plan office environments. *Building and Environment*, *207*, 108512. https://doi.org/10.1016/j.buildenv.2021.108512

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A., Bento Coelho, J. L., Maffei, L. & Lavia, L. (2016). Ten questions on the soundscapes of the built environment. *Building and Environment*, *108*, 284–294. https://doi.org/10.1016/j.buildenv.2016.08.011

Kaya, E. M. & Elhilali, M. (2014). Investigating bottom-up auditory attention. *Frontiers in Human Neuroscience*, *8.* https://doi.org/10.3389/fnhum.2014.00327

Kaya, E. M. & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160101. https://doi.org/10.1098/rstb.2016.0101

Klatte, M., Spilski, J., Mayerl, J., Möhler, U., Lachmann, T. & Bergström, K. (2017). Effects of Aircraft Noise on Reading and Quality of Life in Primary School Children in Germany: Results From the NORAH Study. *Environment and Behavior*, *49*(4), 390–424. https://doi.org/10.1177/0013916516642580

Krstajic, D., Buturovic, L. J., Leahy, D. E. & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, *6*(1), 10. https://doi.org/10.1186/1758-2946-6-10

Kubey, R., Larson, R. & Csikszentmihalyi, M. (1996). Experience Sampling Method Applications to Communication Research Questions. *Journal of Communication*, *46*(2), 99–120. https://doi.org/10.1111/j.1460-2466.1996.tb01476.x

Kuwano, S., Namba, S., Kato, T. & Hellbrück, J. (2003). Memory of the loudness of sounds in relation to overall impression. *Acoustical Science and Technology*, *24*(4), 194–196. https://doi.org/10.1250/ast.24.194

Lagrange, M., Lafay, G., Défréville, B. & Aucouturier, J.-J. (2015). The bag-of-frames approach: A not so sufficient model for urban soundscapes. *The Journal of the Acoustical Society of America*, *138*(5), EL487-92. https://doi.org/10.1121/1.4935350

Larson, R. & Csikszentmihalyi, M. (2014). The Experience Sampling Method. In M. Csikszentmihalyi (Hrsg.), *Flow and the Foundations of Positive Psychology* (S. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lartillot, O. (2021). MIRToolbox 1.8.1 User's Manual. https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-8-1.pdf/@@download/apatfile/manual1.8.1.pdf

Lartillot, O., Toiviainen, P. & Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Hrsg.), *Studies in Classification, Data Analysis, and Knowledge Organization. Data Analysis, Machine Learning and Applications* (S. 261–268). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_31

Lazaro, A., Sarno, R., R., J. A. & Mahardika, M. N. (2017) Music Tempo Classification Using Audio Spectrum Centroid, Audio Spectrum Flatness, and Audio Spectrum Spread based on MPEG-7 Audio Features. In *2017 3rd international conference on science in information technology (ICSITech),* 41–46. https://doi.org/10.1002/0470093366.fmatter

Lee, C.-H., Hsu, S.-B., Shih, J.-L. & Chou, C.-H. (2013). Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features. *IEEE Transactions on Multimedia*, *15*(2), 454–464. https://doi.org/10.1109/TMM.2012.2229969

Lee, C.-H., Shih, J.-L., Yu, K.-M. & Lin, H.-S. (2009). Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features. *IEEE Transactions on Multimedia*, *11*(4), 670–682. https://doi.org/10.1109/TMM.2009.2017635

Lee, C.-H., Shih, J.-L., Yu, K.-M. & Su, J.-M. (2007). Automatic Music Genre Classification using Modulation Spectral Contrast Feature. In *Multimedia and Expo, 2007 IEEE International Conference on.* IEEE. https://doi.org/10.1109/icme.2007.4284622

Lepa, S., Steffens, J., Herzog, M. & Egermann, H. (2020). Popular Music as Entertainment Communication: How Perceived Semantic Expression Explains Liking of Previously Unknown Music. *Media and Communication*, *8*(3), 191–204. https://doi.org/10.17645/mac.v8i3.3153

Lerch, A. (2012). *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press.

Lercher, P. (1996). Environmental noise and health: An integrated research perspective. *Environment International*, *22*(1), 117–129. https://doi.org/10.1016/0160-4120(95)00109-3

Li, D., Sethi, I. K., Dimitrova, N. & McGee, T. (2001). Classification of general audio data for content-based retrieval. *Pattern Recognition Letters*, *2001*(5), 533–544.

Lindgren, A. C, Johnson, M. T & Povinelli, R. J (2003). Speech recognition using reconstructed phase space features. In *2003 IEEE Inter-national Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ,03)* (I-60-3). IEEE. https://doi.org/10.1109/ICASSP.2003.1198716

Ma, H. & Han, C. (2006). Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction. *Frontiers of Electrical and Electronic Engineering in China*, *1*(1), 111–114. https://doi.org/10.1007/s11460-005-0023-7

MacDonald, R. A. R. (2013). Music, health, and well-being: a review. *International journal of qualitative studies on health and well-being*, *8*, 20635. https://doi.org/10.3402/qhw.v8i0.20635

McDermott, J. H., Schemitsch, M. & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature neuroscience*, *16*(4), 493–498. https://doi.org/10.1038/nn.3347

McFee, B., Barrington, L. & Lanckriet, G. (2012). Learning Content Similarity for Music Recommendation. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(8), 2207–2218. https://doi.org/10.1109/tasl.2012.2199109

McKinney, M. F. & Breebaart, J. (2003). *Features for Audio and Music Classification*. https://jscholarship.library.jhu.edu/handle/1774.2/22

Mitrović, D., Zeppelzauer, M. & Breiteneder, C. (2010). Features for Content-Based Audio Retrieval. In *Advances in Computers. Advances in Computers: Improving the Web, 78*, 71–150. Elsevier. https://doi.org/10.1016/S0065-2458(10)78003-7

Moffat, D., Ronan, D. & Reiss, J. D. (2015). An evaluation of audio feature extraction toolboxes. *Proceedings of the 18th Int. Conference on Digital Audio Effects*. https://www.researchgate.net/publication/282858086_An_Evaluation_of_Audio_Feature_Extraction_Toolboxes

Mohamed, M. A. E. & Dokmeci Yorukoglu, P. N. (2020). Indoor soundscape perception in residential spaces: A cross-cultural analysis in Ankara, Turkey. *Building Acoustics*, *27*(1), 35–46. https://doi.org/10.1177/1351010X19885030

Mörchen, F., Ultsch, A., Thies, M. & Löhken, I. (2006). Modeling timbre distance with temporal statistics from polyphonic music. *IEEE Transactions on Audio, Speech and Language Processing*, *14*(1), 81–90. https://doi.org/10.1109/TSA.2005.860352

Mörchen, F., Ultsch, A., Thies, M., Löhken, I., Nöcker, M., Stamm, C., Efthymiou, N. & Krümmerer, M. (2005). MusicMiner: Visualizing timbre distances of music as topographical maps. Univ..

movisens GmbH. (2021). *Experience Sampling - movisensXS - movisens GmbH*. https://www.movisens.com/en/products/movisensXS/

Muhammad, G. & Alghathbar, K. (2009). Environment Recognition from Audio Using MPEG-7 Features: Jeju, Korea, 10 - 12 December 2009; [including the 2009 International Workshop on Ubiquitous Multimedia Computing and Communications (UMCC 2009). *2009 Fourth International Conference on Embedded and Multimedia Computing (EM-Com 2009)*.

Nakagawa, S., Johnson, P. C. D. & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nam, J., Herrera, J., Slanley, M. & Smith, J. (2012). Learning Sparse Feature Representations for Music Annotation and Retrieval. *ISMIR*, 565–570. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.452.9286&rep=rep1&type=pdf

Nilsson, M. E., Botteldooren, D. & Coensel, B. de (2007). Acoustic Indicators of Soundscape Quality and Noise Annoyance in Outdoor Urban Areas. *Proceedings of the 19th International Congress on Acoustics*.

Oldoni, D., Coensel, B. de, Boes, M., Rademaker, M., Baets, B. de, van Renterghem, T. & Botteldooren, D. (2013). A computational model of auditory attention for use in soundscape research. *The Journal of the Acoustical Society of America*, *134*(1), 852–861. https://doi.org/10.1121/1.4807798

Oliva, D., Hongisto, V. & Haapakangas, A. (2017). Annoyance of low-level tonal sounds – Factors affecting the penalty. *Building and Environment*, *123*, 404–414. https://doi.org/10.1016/j.buildenv.2017.07.017

Oppenheim, A. V. & Schafer, R. W. (2004). From frequency to quefrency: A history of the cepstrum. *IEEE signal processing Magazine*, *21*(5), 95–106.

Orhan, C. (2019). *A comparative study on indoor soundscape in museum environments* [Doktorarbeit, Bilkent University]. repository.bilkent.edu.tr. http://repository.bilkent.edu.tr/handle/11693/52316

Pachet, F. & Zils, A. (2003). Evolving Automatically High-Level Music Descriptors from Acoustic Signals. *International Symposium on Computer Music Modeling and Retrieval*, 42–53.

Park, T. H., Turner, J., Musick, M., Lee, J. H., Jacoby, C., Mydlarz, C. & Salamon, J. (2014). Sensing Urban Soundscapes. In *EDBT/ICDT Workshops*, 375–382.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. & Allerhand, M. (1992). Complex Sounds and Auditory Images. In *Auditory Physiology and Perception,* 429–446. Elsevier. https://doi.org/10.1016/b978-0-08-041847-6.50054-x

Pieretti, N., Farina, A. & Morri, D. (2011). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, *11*(3), 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005

Pohle, T., Schnitzer, D., Schedl, M., Knee, P. & Widmer, G. (2009). *On Rhythm and General Music Similarity*. https://archives.ismir.net/ismir2009/paper/000020.pdf

Povinelli, R. J, Johnson, M. T, Lindgren, A. C, Roberts, F. M. & Ye, J. (2006). Statistical models of reconstructed phase spaces for signal classification. *IEEE Transactions on Signal Processing*, *54*(6), 2178–2186. https://doi.org/10.1109/TSP.2006.873479

Rabaoui, A., Davy, M., Rossignol, S. & Ellouze, N. (2008). Using One-Class SVMs and Wavelets for Audio Surveillance. *IEEE Transactions on Information Forensics and Security*, *3*(4), 763–775. https://doi.org/10.1109/tifs.2008.2008216

Rahimi, S., Andris, C. & Liu, X. (2017). Using yelp to find romance in the city: A case of restaurants in four cities. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, 1–8.

RC Team. (2013). *R: A language and environment for statistical computing*. http://r.meteo.uni.wroc.pl/web/packages/dplr/vignettes/intro-dplr.pdf

Ricciardi, P., Delaitre, P., Lavandier, C., Torchia, F. & Aumond, P. (2015). Sound quality indicators for urban places in Paris cross-validated by Milan data. *The Journal of the Acoustical Society of America*, *138*(4), 2337–2348. https://doi.org/10.1121/1.4929747

Roberts, S. & Nowak, G. (2014). Stabilizing the lasso against cross-validation variability. *Computational Statistics & Data Analysis*, *70*, 198–211. https://doi.org/10.1016/j.csda.2013.09.008

Ronzhin, A., Potapova, R. & Németh, G. (2016). *Fusing Various Audio Feature Sets for Detection of Parkinson's Disease from Sustained Voice and Speech Recordings*, 9811, 328–337. Springer International Publishing. https://doi.org/10.1007/978-3-319-43958-7

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A., Ward, L. M. & Pratt, G. (1981). Affective Quality Attributed to Environments. *Environment and Behavior*, *13*(3), 259–288. https://doi.org/10.1177/0013916581133001

Salamon, J. & Bello, J. P. (2015). Unsupervised Feature Learning for Urban Sound Classification. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 171–175.

Schafer, M. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books.

Schlüter, J. & Osendorfer, C. (2011). Music Similarity Estimation with the Mean-Covariance Restricted Boltzmann Machine. In *2011 IEEE 10th International Conference,* 118–123. https://doi.org/10.1109/ICMLA.2011.102

Selzer, J., Becker-Schweitzer, J., Oehler, M. & Skoda, S. (2017). Salienz von Umweltgeräuschen: Welchen Einfluss haben Intensität, zeitlicher Verlauf und spektraler Kontrast? *Fortschritte der Akustik - DAGA 2017.*

Seyerlehner, K., Widmer, G. & Knees, P. (2008). Frame level audio similarity-a codebook approach. https://www.dafx.de/paper-archive/2008/papers/dafx08_61.pdf

Seyerlehner, K., Widmer, G. & Pohle, T. (2010). Fusing block-level features for music similarity estimation. *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 225–232. http://dafx10.iem.at/papers/seyerlehnerwidmerpohle_dafx10_p31.pdf

Skoda, S., Steffens, J. & Becker-Schweitzer, J. (2014). Road traffic noise annoyance in domestic environments can be reduced by water sounds. https://www.researchgate.net/publication/266970366_Road_traffic_noise_annoyance_in_domestic_environments_can_be_reduced_by_water_sounds

Sottek, R. & Genuit, K. (2009). Sound quality evaluation of fan noise based on advanced hearing-related parameters. *Noise Control Engineering Journal*, *57*(4), 384. https://doi.org/10.3397/1.3159391

Stallen, P. J. M. (1999). A theoretical framework for environmental noise annoyance. *Noise & health*, *1*(3), 69–80.

Steffens, J. & Guastavino, C. (2015). Trend Effects in Momentary and Retrospective Soundscape Judgments. *Acta Acustica united with Acustica*, *101*(4), 713–722. https://doi.org/10.3813/AAA.918867

Steffens, J., Steele, D. & Guastavino, C. (2017). Situational and person-related factors influencing momentary and retrospective soundscape evaluations in day-to-day life. *The Journal of the Acoustical Society of America*, *141*(3), 1414. https://doi.org/10.1121/1.4976627

Tardieu, J., Susini, P., Poisson, F., Lazareff, P. & McAdams, S. (2008). Perceptual study of soundscapes in train stations. *Applied Acoustics*, *69*(12), 1224–1239. https://doi.org/10.1016/j.apacoust.2007.10.001

Tarlao, C., Steffens, J. & Guastavino, C. (2021). Investigating contextual influences on urban soundscape evaluations with structural equation modeling. *Building and Environment*, *188*, 107490. https://doi.org/10.1016/j.buildenv.2020.107490

Thomas, P., Dekoninck, L., van Hove, S., All, A., Conradie, P., Marez, L. de, Huisseune, H., Plets, D. & Botteldooren, D. (2019). Characterization of the indoor kitchen soundscape. *23rd International Congress on Acoustics (ICA 2019)*, 4154–4157. https://doi.org/10.18154/RWTH-CONV-239646

Tordini, F. (2014). Is there more to saliency than loudness? *6th Workshop on Speech in Noise (SPiN): Intelligibility and Quality, Marseille*.

Torresin, S., Albatici, R., Aletta, F., Babich, F. & Kang, J. (2019). Assessment Methods and Factors Determining Positive Indoor Soundscapes in Residential Buildings: A Systematic Review. *Sustainability*, *11*(19), 5290. https://doi.org/10.3390/su11195290

Torresin, S., Albatici, R., Aletta, F., Babich, F., Oberman, T., Stawinoga, A. E. & Kang, J. (2021). Indoor soundscapes at home during the COVID-19 lockdown in London–Part I: Associations between the perception of the acoustic environment, occupantś occupantś activity and well-being. *Applied Acoustics*, *183*, 108305. https://www.sciencedirect.com/science/article/pii/s0003682x21003996

Torresin, S., Albatici, R., Aletta, F., Babich, F., Oberman, T., Stawinoga, A. E. & Kang, J. (2022). Indoor soundscapes at home during the COVID-19 lockdown in London – Part II: A structural equation model for comfort, content, and well-being. *Applied Acoustics*, *185*, 108379. https://doi.org/10.1016/j.apacoust.2021.108379

Torresin, S., Aletta, F., Babich, F., Bourdeau, E., Harvie-Clark, J., Kang, J., Lavia, L., Radicchi, A. & Albatici, R. (2020). Acoustics for Supportive and Healthy Buildings: Emerging Themes on Indoor Soundscape Research. *Sustainability*, *12*(15), 6054. https://doi.org/10.3390/su12156054

Truax, B. (2019). Acoustic Ecology and the World Soundscape Project. In *Sound, Media, Ecology* (S. 21–44). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-16569-7_2

Truax, B. & Barrett, G. W. (2011). Soundscape in a context of acoustic and landscape ecology. *Landscape Ecology*, *26*(9), 1201–1207. https://doi.org/10.1007/s10980-011-9644-9

Valero, X. & Alias, F. (2012). Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. *IEEE Transactions on Multimedia*, *14*(6), 1684–1689. https://doi.org/10.1109/tmm.2012.2199972

van Berkel, N., Ferreira, D. & Kostakos, V. (2018). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, *50*(6), 1–40. https://doi.org/10.1145/3123988

Västfjäll, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, *6*(2), 181–188. https://doi.org/10.1089/109493103321640374

Västfjäll, D. (2004). The „end effect“ in retrospective sound quality evaluation. *Acoustical Science and Technology*, *25*(2), 170–172. https://doi.org/10.1250/ast.25.170

Västfjäll, D., Kleiner, M. & Görling, T. (2003). Affective Reactions to and Preference for Combinations of Interior Aircraft Sound and Vibration. *The International Journal of Aviation Psychology*, *13*(1), 33–47. https://doi.org/10.1207/S15327108IJAP1301_3

Versümer, S., Steffens, J., Blättermann, P. & Becker-Schweitzer, J. (2020). Modeling Evaluations of Low-Level Sounds in Everyday Situations Using Linear Machine Learning for Variable Selection. *Frontiers in psychology*, *11*, 570761. https://doi.org/10.3389/fpsyg.2020.570761

Wang, B., Kang, J. & Zhao, W. (2020). Noise acceptance of acoustic sequences for indoor soundscape in transport hubs. *The Journal of the Acoustical Society of America*, *147*(1), 206. https://doi.org/10.1121/10.0000567

Weiss, R. & Bello, J. P. (2010). Identifying repeated patterns in music using sparse convolutive non-negative matrix factorization. *11th Int Society for Music Information Retrieval Conference*, 123–128.

Wolsink, M., Sprengers, M., Keuper, A., Pedersen, T. H. & Westra, C. A. (1993). *Annoyance from wind turbine noise on sixteen sites in three countries*. https://www.researchgate.net/publication/317167748_Annoyance_from_Wind_Turbine_Noise_on_Sixteen_Sites_in_Three_Countries

World Health Organization. (2011). *Burden of disease from environmental noise: quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/326424

Wülfing, J. & Riedmiller, M. (2012). *Unsupervised Learning of Local Features for Music Classification*. https://archives.ismir.net/ismir2012/paper/000139.pdf

Yilmazer, S. & Acun, V. (2018). A grounded theory approach to assess indoor soundscape in historic religious spaces of Anatolian culture: A case study on Hacı Bayram Mosque. *Building Acoustics*, *25*(2), 137–150. https://doi.org/10.1177/1351010X18763915

Yilmazer, S. & Bora, Z. (2017). Understanding the indoor soundscape in public transport spaces: A case study in Akköprü metro station, Ankara. *Building Acoustics*, *24*(4), 325–339. https://doi.org/10.1177/1351010X17741742

Yin, H., Hohmann, V. & Nadeu, C. (2011). Acoustic features for speech recognition based on Gammatone filterbank and instantaneous frequency. *Speech Communication*, *53*(5), 707–715. https://doi.org/10.1016/j.specom.2010.04.008

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics. Facts and Models*. Springer Science & Business Media.

ANHANG A

ABBILDUNG A1:

BOXPLOTS DER BEWERTETEN ANGENEHMHEIT UND EREIGNISREICHTUM IN Z-WERTEN. (INNENLINIE IN BOX = MEDIAN; BREITE DER BOX = IQR; BOX-AUßENKANTEN = 25 % BZW. 75 % QUARTIL; WHISKER = MINIMA/ MAXIMA MIT HÖCHSTENS 1.5 IQR ABSTAND ZUR BOX; AUSREIßER (KREISE) = DATENPUNKTE MIT ABSTAND > 1.5 IQR ZUR BOX)

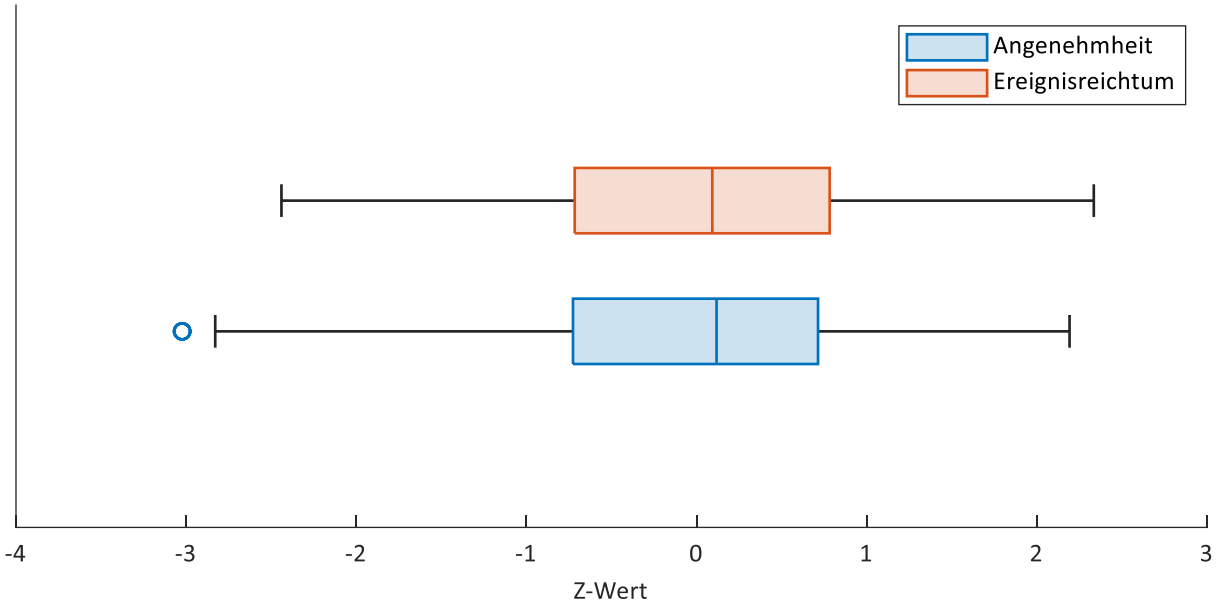


ABBILDUNG A2:

BOXPLOTS DER PERSÖNLICHEN DURCHSCHNITTLICHEN BEWERTUNG VON ANGENEHMHEIT UND EREIGNISREICHTUM PRO PERSON IN Z-WERTEN GRUPPIERT NACH GESCHLECHT (INNENLINIE IN BOX = MEDIAN; HÖHE DER BOX = IQR; BOX-AUßENKANTEN = 25 % BZW. 75 % QUARTIL; WHISKER = MINIMA/ MAXIMA MIT HÖCHSTENS 1.5 IQR ABSTAND ZUR BOX; AUSREIßER (KREISE) = DATENPUNKTE MIT ABSTAND > 1.5 IQR ZUR BOX). ABWEICHUNGEN DER MITTELWERTE DER GESCHLECHTER BEI BEIDEN ZIELGRÖßEN NICHT STATISTISCH SIGNIFIKANT ($t(103)_{ANG.} = -1.7$; $p = 0.09$ BZW. $t(103)_{EREIGN.} = -1.7$; $p = 0.1$).

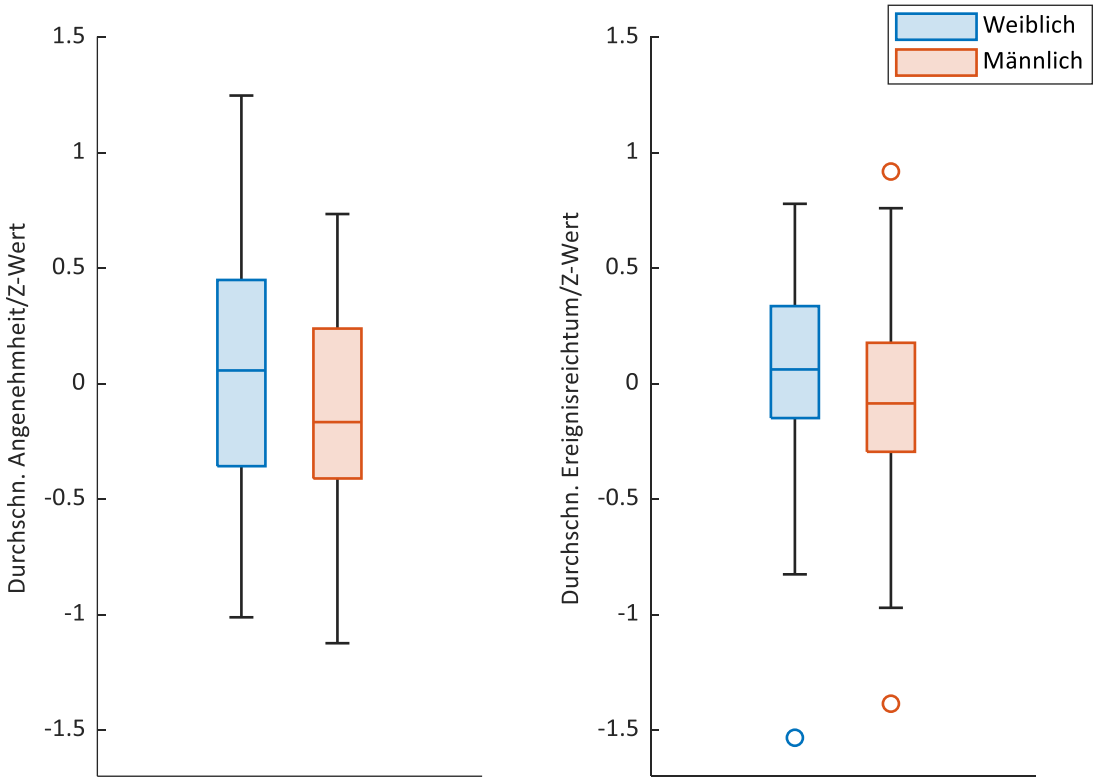


ABBILDUNG A3:

LAMBDA-TRACE-PLOT DER IN DER PERZENTILEN LASSO-REGULARISIERUNG FÜR MODELL EV_BA GEMITTELTEN PRÄDIKTORKOEFFIZIENTEN β ÜBER DEN MODELLPARAMETER LAMBDA (OBEN GANZER WERTEBEREICHR; UNTEN AUSSCHNITTSVERGRÖßERUNG UM DAS PERZENTILE LAMBDA). DIE ROTE X-LINIE ‚PERCENTILE LAMBDA' GIBT DEN WERT AN, DER ALS 95. PERZENTIL DER 100 1SE-LAMBDAS BERECHNET WURDE.

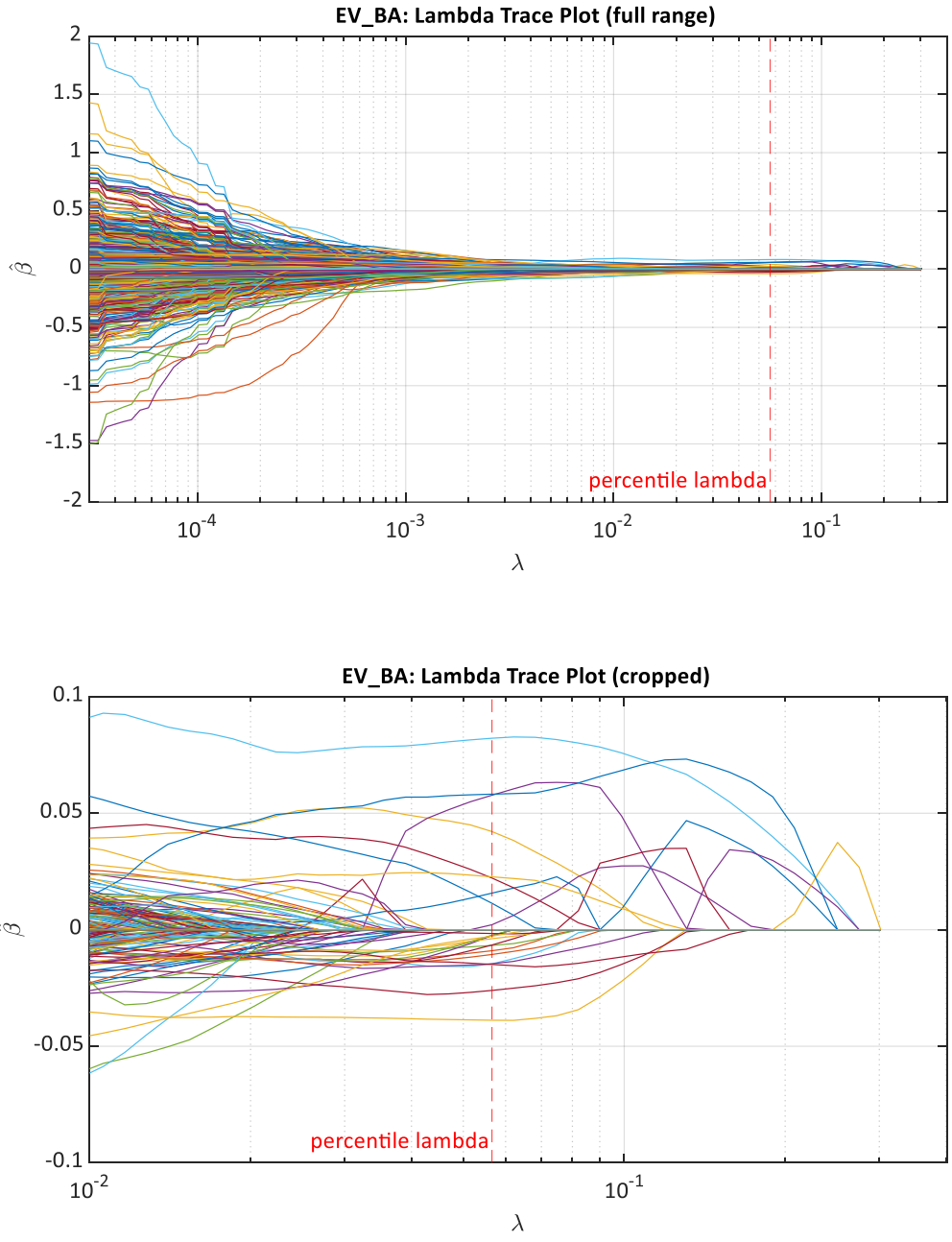
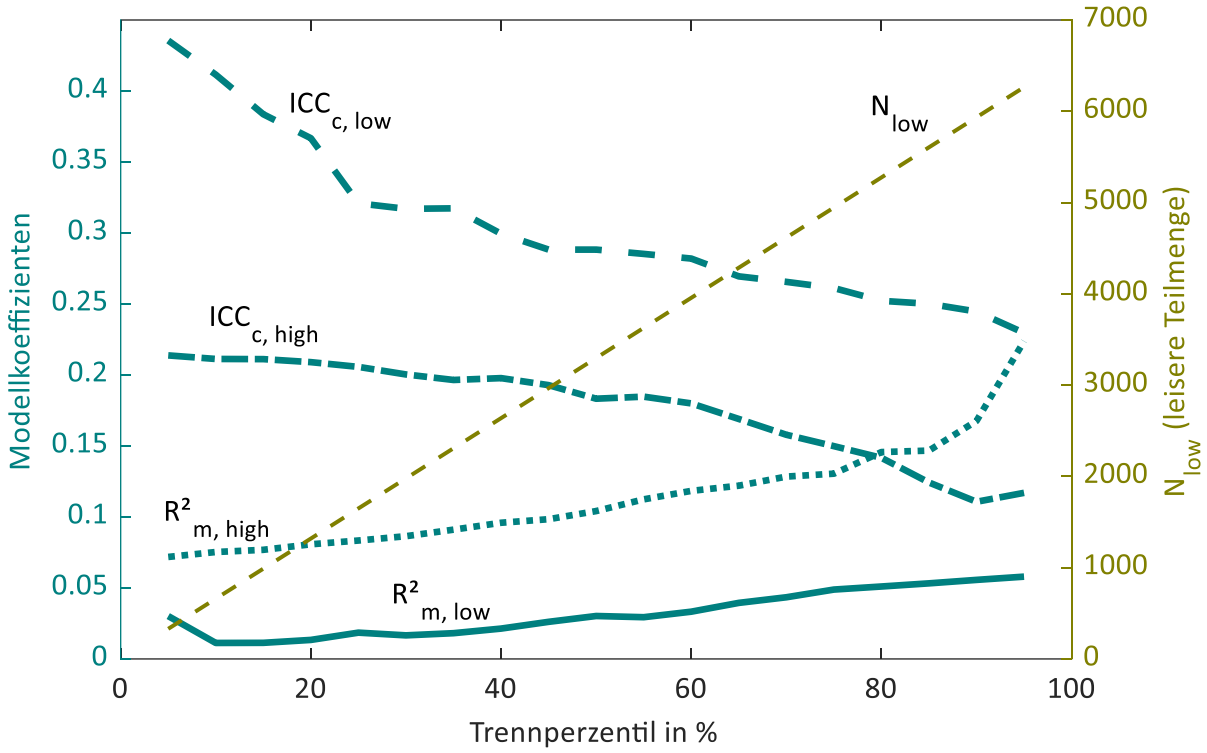


ABBILDUNG A4:

PLOT DER MODELLKOEFFIZIENTEN $R^2_{marginal}$ UND $ICC_{conditional}$ (NORMIERT; Y-ACHSE LINKS) VON MULTILEVEL-MODELLEN FÜR DATENSATZTEILMENGEN N_low (LEISERE TEILMENGE) UND N_high (LAUTERE TEILMENGE) SOWIE ANZAHL DER OBSERVATIONEN FÜR DIE LEISERE TEILMENGE low (ABSOLUTE ANZAHL; Y-ACHSE RECHTS) ÜBER DIE TRENNPERZENTILE 5…95 (IN %; X-ACHSE), DIE ALS LAUTHEITS-SCHWELLWERTE FÜR DIE TEILMENGEN low UND high EINGESETZT WURDEN. VEREINFACHEND WURDE FÜR JEDE ITERATION DIE MODELLFORMEL *'ANGENEHMHEIT ~ 1 + FLUCTUATIONSTRENGTH_P95 + LOUDNESS_P99 + SHARPNESS_P1-P99 + SHARPNESS_P1 + (1 | ID)'* ANGENOMMEN, DIE IN PERZENTILER LASSO-SELEKTION FÜR DAS TRENNPERZENTIL 75 ERMITTELT WURDE (QUELLCODE SIEHE DATEI D13 IN ANHANG D).



ANHANG B

VOLLSTÄNDIGE ERGEBNISTABELLEN DER LINEAREN GEMISCHTEN MODELLE NACH PERZENTILER LASSO-REGULARISIERUNG

Für alle Modelle, in denen der Datensatz nicht in Teilmengen modelliert wurde, gilt N_{ID} = 105 und Observations = 6594.

Legende für feste Effekte	
Predictors	Feature-Bezeichnung
β	Z-standardisierte Regressionskoeffizienten
[CI]	95 % Konfidenzintervalle nach dem Schema [Lower Limit (2.5 %), Upper Limit (97.5 %)]
p	Zufallswahrscheinlichkeit; p-Werte kleiner als 0.001 wurden zur besseren Lesbarkeit durch <.001 ersetzt; Prädiktoren, aber nicht Intercepts, deren p-Werte über dem Signifikanzniveau von α = 5 % liegen, sind in der Schriftfarbe ‚GRAU' dargestellt.
[df]	Anzahl der Freiheitsgrade

Legende für zufällige Effekte (Random Effects)	
σ^2	Intraklassenvarianz
$\tau_{00\ ID}$	Interklassenvarianz (Varianz des Nullmodells)
ICC	Intraklassenkorrelation
$R^2_{marginal}$	Anteil der Varianzaufklärung durch feste Effekte
$R^2_{conditional}$	Gesamte Varianzaufklärung durch feste und zufällige Effekte

ERGEBNISTABELLEN ANGENEHMHEIT

PL_BA

(Angenehmheit & BA-Featureset)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.11, 0.07]	.697	[104]
ba_mfcc_12_LocMax/s	-0.17	[-0.19, -0.15]	<.001	[6547.4]
ba_mfccDelta_3_kurtosis	-0.06	[-0.08, -0.04]	<.001	[6517.9]
ba_predictivityRatio_LocMax/s	-0.05	[-0.08, -0.03]	<.001	[6570.2]
ba_spectralCentroid_skewness	0.04	[0.01, 0.06]	.002	[6522.7]
ba_gtccDelta_1_mode	-0.03	[-0.06, -0.01]	.010	[6575.4]
Random Effects				
σ^2	0.74			
$\tau_{00\ ID}$	0.22			
ICC	0.23			
$R^2_{marginal}$	0.05			
$R^2_{conditional}$	0.26			

PL_MIR

(Angenehmheit & MIR-Featureset)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.11, 0.08]	.706	[104]
mir_mfcc_13_Std	0.12	[0.09, 0.15]	<.001	[6522.2]
mir_hcdf_PeriodAmp	-0.08	[-0.11, -0.05]	<.001	[6523.2]
mir_keyclarity_Std	0.04	[0.02, 0.07]	<.001	[6522]
mir_mfcc_2_Mean	0.04	[0.02, 0.07]	.001	[6552.1]
mir_mfcc_7_Mean	0.04	[0.02, 0.06]	<.001	[6554.6]
mir_kurtosis_Mean	-0.04	[-0.07, -0.01]	.002	[6562.2]
mir_mfcc_6_PeriodAmp	-0.04	[-0.06, -0.01]	.001	[6532.3]
mir_mfcc_10_PeriodAmp	-0.04	[-0.06, -0.01]	.002	[6521.3]
mir_mfcc_13_PeriodAmp	-0.03	[-0.05, -0.01]	.006	[6519.9]

Random Effects	
σ^2	0.74
$\tau_{00\ ID}$	0.22
ICC	0.23
$R^2_{marginal}$	0.05
$R^2_{conditional}$	0.27

PL_PSY

(Angenehmheit & Psychoakustische Größen)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.11, 0.07]	.670	[104.2]
psy_FluctuationStrength_P95	0.18	[0.15, 0.21]	<.001	[6539.2]
psy_Sharpness_P1-P99	0.14	[0.12, 0.17]	<.001	[6554.4]
psy_Loudness_rmc	-0.14	[-0.16, -0.11]	<.001	[6565.2]
Random Effects				
σ^2	0.72			
$\tau_{00\ ID}$	0.22			
ICC	0.23			
$R^2_{marginal}$	0.07			
$R^2_{conditional}$	0.29			

PL_RA

(Angenehmheit & Relative Approach)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.11, 0.07]	.695	[104]
ra_HM_P95	0.29	[0.26, 0.33]	<.001	[6548.1]
ra_Pr1_P99	-0.23	[-0.26, -0.20]	<.001	[6541.7]
ra_Pr1_P10	-0.1	[-0.13, -0.06]	<.001	[6554.2]

Random Effects	
σ^2	0.73
$\tau_{00\ ID}$	0.22
ICC	0.23
$R^2_{marginal}$	0.05
$R^2_{conditional}$	0.27

PL_SEL

(Angenehmheit & selektierte Prädiktoren)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.11, 0.07]	.676	[104]
psy_Loudness_rmc	-0.20	[-0.24, -0.17]	<.001	[6537.9]
ra_HM_P95	0.13	[0.09, 0.17]	<.001	[6511.8]
psy_Sharpness_P1-P99	0.13	[0.10, 0.15]	<.001	[6534]
psy_FluctuationStrength_P95	0.08	[0.04, 0.11]	<.001	[6512.3]
ba_mfcc_12_LocMax/s	-0.04	[-0.07, -0.01]	.006	[6506.5]
mir_ hcdf_PeriodAmp	-0.04	[-0.07, -0.01]	.008	[6514.2]
mir_ keyclarity_Std	0.04	[0.02, 0.06]	<.001	[6515.7]
mir_ mfcc_10_PeriodAmp	-0.03	[-0.06, -0.01]	.002	[6520.6]
mir_ mfcc_7_Mean	0.03	[0.01, 0.06]	.002	[6552.9]
mir_ mfcc_13_Std	0.03	[0.00, 0.06]	.064	[6504.2]
Random Effects				
σ^2	0.70			
$\tau_{00\ ID}$	0.22			
ICC	0.24			
$R^2_{marginal}$	0.09			
$R^2_{conditional}$	0.31			

ERGEBNISTABELLEN EREIGNISREICHTUM

EV_BA

(Ereignisreichtum & BA-Featureset)

Predictors	β	[CI]	p	[df]
(Intercept)	0.00	[-0.08, 0.07]	.942	[102.2]
ba_gtcc_1_mean	0.14	[0.11, 0.18]	<.001	[6583.9]
ba_mfccDelta_12_LocMax/s	-0.13	[-0.16, -0.10]	<.001	[6527.9]
ba_mfcc_2_P5	0.08	[0.05, 0.11]	<.001	[6570.8]
ba_mfccDeltaDelta_10_LocMax/s	-0.08	[-0.11, -0.05]	<.001	[6508.6]
ba_mfcc_13_lmSlopePACF	-0.06	[-0.08, -0.03]	<.001	[6503.4]
ba_mfccDelta_9_skewness	-0.05	[-0.07, -0.02]	<.001	[6557]
ba_mfcc_6_lmSlopePACF	-0.04	[-0.07, -0.01]	.004	[6511.4]
ba_gtcc_3_lmSlopePACF	-0.03	[-0.06, -0.01]	.014	[6524.8]
ba_spectralDecrease_entropy	0.02	[-0.02, 0.05]	.336	[6579.8]
Random Effects				
σ^2	0.72			
$\tau_{00\ ID}$	0.15			
ICC	0.17			
$R^2_{marginal}$	0.14			
$R^2_{conditional}$	0.29			

EV_MIR

(Ereignisreichtum & MIR-Featureset)

Predictors	β	[CI]	p	[df]
(Intercept)	0.00	[-0.08, 0.08]	.962	[102.8]
mir_spread_Mean	-0.20	[-0.22, -0.17]	<.001	[6571.6]
mir_mfcc_12_Std	0.10	[0.05, 0.15]	<.001	[6516.4]
mir_mfcc_13_Std	0.10	[0.05, 0.15]	<.001	[6505.3]
mir_keyclarity_Std	0.07	[0.05, 0.1]	<.001	[6538.6]
Random Effects				
σ^2	0.73			
$\tau_{00\ ID}$	0.15			
ICC	0.17			
$R^2_{marginal}$	0.12			
$R^2_{conditional}$	0.27			

EV_PSY

(Ereignisreichtum & psychoakustische Größen)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.01	[-0.09, 0.07]	.843	[102]
psy_SPLC_P1	0.38	[0.35, 0.41]	<.001	[6567.4]
psy_Tonality_P1-P99	0.15	[0.13, 0.18]	<.001	[6516.4]
psy_Fluctuationtrength_P95	0.14	[0.10, 0.17]	<.001	[6514.3]
psy_Loudness_P5-P95	-0.08	[-0.12, -0.05]	<.001	[6522.2]
psy_Sharpness_P1-P99	0.08	[0.06, 0.11]	<.001	[6570.5]
psy_Fluctuationtrength_P90	-0.08	[-0.11, -0.05]	<.001	[6519.7]
psy_TonalityHMS_P75	-0.07	[-0.09, -0.05]	<.001	[6512.7]
psy_Sharpness_P99	-0.07	[-0.09, -0.04]	<.001	[6567.5]

Random Effects	
σ^2	0.64
$\tau_{00\ ID}$	0.15
ICC	0.20
$R^2_{marginal}$	0.23
$R^2_{conditional}$	0.38

EV_RA

(Ereignisreichtum & Relative Approach)

Predictors	β	[CI]	p	[df]
(Intercept)	0.00	[-0.08, 0.07]	.905	[102.2]
ra_1/3F_RT_P95	0.68	[0.62, 0.75]	<.001	[6584.7]
ra_1/12F_RF_P5-P95	0.52	[0.38, 0.67]	<.001	[6514.4]
ra_1/12F_RT_P10	-0.46	[-0.54, -0.37]	<.001	[6514.1]
ra_1/12F_RT_P95	-0.36	[-0.43, -0.29]	<.001	[6545.3]
ra_HM_P95	0.18	[0.13, 0.23]	<.001	[6526.4]
ra_Pr1_P1-P99	-0.15	[-0.18, -0.11]	<.001	[6541.5]
ra_1/12F_RF_P25-P75	0.13	[0.05, 0.21]	.001	[6504.5]
ra_1/12F_RF_P10-P90	0.03	[-0.15, 0.22]	.715	[6509.5]
Random Effects				
σ^2	0.64			
$\tau_{00\ ID}$	0.16			
ICC	0.19			
$R^2_{marginal}$	0.23			
$R^2_{conditional}$	0.38			

EV_SEL

(Ereignisreichtum & selektierte Prädiktoren)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.01	[-0.09, 0.07]	.852	[101.6]
psy_SPLC_P1	0.25	[0.22, 0.29]	<.001	[6535.3]
ra_1/3F_RT_P95	0.21	[0.18, 0.25]	<.001	[6532.1]
psy_Tonality_P1-P99	0.12	[0.09, 0.15]	<.001	[6503.5]
ra_Pr1_P1-P99	-0.10	[-0.13, -0.07]	<.001	[6539.5]
psy_Sharpness_P1-P99	0.09	[0.06, 0.11]	<.001	[6541.8]
psy_TonalityHMS_P75	-0.06	[-0.09, -0.04]	<.001	[6501.5]
psy_Sharpness_P99	-0.06	[-0.08, -0.03]	<.001	[6557.1]
ba_mfccDeltaDelta_10_LocMax/s	-0.04	[-0.07, -0.01]	.007	[6492.1]
ba_mfcc_6_lmSlopePACF	-0.04	[-0.06, -0.01]	.003	[6496.1]
mir_keyclarity_Std	0.04	[0.02, 0.06]	<.001	[6515.5]
ba_mfcc_13_lmSlopePACF	-0.03	[-0.06, -0.01]	.005	[6490.4]
ba_mfccDelta_12_LocMax/s	-0.03	[-0.06, -0.00]	.033	[6497.3]
ba_gtcc_3_lmSlopePACF	-0.03	[-0.05,-0.01]	.016	[6506.9]
ba_mfccDelta_9_skewness	-0.02	[-0.04, -0.00]	.032	[6534.8]
mp3FileSize	-0.02	[-0.05, 0.00]	.068	[6563.4]
psy_FluctuationStrength_P95	0.02	[-0.01, 0.05]	.221	[6513.5]
ba_spectralDecrease_entropy	0.02	[-0.01, 0.04]	.142	[6568.5]
Random Effects				
σ^2	0.61			
$\tau_{00\ ID}$	0.15			
ICC	0.20			
$R^2_{marginal}$	0.27			
$R^2_{conditional}$	0.41			

ERGEBNISTABELLEN ANGENEHMHEIT (MULTILEVEL)

PL_NoMusic

(Angenehmheit & psychoakustische Größen;

keine vordergründige Musik)

Predictors	β	CI	p	[df]
(Intercept)	-0.02	[-0.11, 0.07]	.671	[104.7]
psy_Loudness_rmc	-0.14	[-0.18, -0.09]	<.001	[5922.3]
psy_Fluctuationtrength_P95	0.12	[0.09, 0.15]	<.001	[5923.9]
psy_Sharpness_P1	0.10	[0.06, 0.15]	<.001	[5949.8]
psy_Sharpness_P1-P99	0.09	[0.04, 0.14]	<.001	[5926.9]
psy_TonalityHMS_P50	-0.07	[-0.09, -0.04]	<.001	[5901.9]
LoudnessPercentile99	0.0	[-0.04, 0.05]	.884	[5904]
Random Effects				
σ^2	0.70			
$\tau_{00\ ID}$	0.22			
ICC	0.24			
$N_{\ ID}$	105			
Observations	5983			
$R^2_{marginal}$	0.08			
$R^2_{conditional}$	0.30			

PL_N_low

(Angenehmheit & psychoakustische Größen; Lautheit ≤ 75. Perzentil)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.13, 0.08]	.649	[104.2]
psy_Fluctuationtrength_P95	0.13	[0.10, 0.16]	<.001	[4898.2]
psy_Sharpness_P1	0.09	[0.04, 0.15]	<.001	[4915.1]
psy_Sharpness_P1-P99	0.05	[-0.01, 0.10]	.094	[4893]
Random Effects				
σ^2	0.69			
$\tau_{00\ ID}$	0.26			
ICC	0.27			
$N_{\ ID}$	105			
Observations	4946			
$R^2_{marginal}$	0.05			
$R^2_{conditional}$	0.31			

PL_N_high

(Angenehmheit & psychoakustische Größen; Lautheit > 75. Perzentil)

Predictors	β	[CI]	p	[df]
(Intercept)	-0.02	[-0.12, 0.07]	.602	[94.8]
psy_Fluctuationtrength_P95	0.19	[0.14, 0.24]	<.001	[1623.3]
psy_Sharpness_P1-P99	0.18	[0.12, 0.24]	<.001	[1640.7]
psy_Loudness_P99	-0.09	[-0.14, -0.04]	.001	[1611.3]
Random Effects				
σ^2	0.72			
$\tau_{00\ ID}$	0.15			
ICC	0.17			
$N_{\ ID}$	102			
Observations	1648			
$R^2_{marginal}$	0.13			
$R^2_{conditional}$	0.28			

ANHANG C

TABELLARISCHE ÜBERSICHT ÜBER DIE NUTZUNG VON SOFTWAREVERSIONEN

Einsatz	Software
ESM-Befragung	Movisens XS (movisens GmbH, 2021)
Vorbereitung des ESM-Datensatzes; Modellbildung	R 4.1.2 (RC Team, 2013) & RStudio 2021.09.0 Build 351 R-Packages: lme4, lmerTest, MuMIn, readxl, writexl, performance, DHARMa, finalfit, dplyr, ggplot2, solitude, readr, tools, utils, xlsx, lattice, tidyr, reshape2, gsubfn, emmeans, clusterBootstrap, fastDummies, forwrassp, stringr, timechange, chron, lubridate, sjPlot
Audio-Verarbeitung, Feature-Extraktion (BA-Featureset), Feature-Selektion	MathWorks MATLAB R2021b Mathworks Toolboxen: Wavelet Toolbox v 6.0, Image Processing Toolbox v 11.4, Statistics and Machine Learning Toolbox v 12.2, Signal Processing Toolbox v 8.7, Predictive Maintenance Toolbox 2.4, Partial Differential Equation Toolbox 3.7, Deep Learning Toolbox v 3.7, System Identification Toolbox v 9.15, Global Optimization Toolbox v 4.6, Econometrics Toolbox v 5.7, DSP System Toolbox v 9.13, Parallel Computing Toolbox v 7.5, Curve Fitting Toolbox v 3.6, Audio Toolbox v 3.1, Phased Array System Toolbox v 4.6, Optimization Toolbox v 9.2
Extraktion MIR-Featureset	MIRToolbox v 1.8.1 (Lartillot et al., 2008; Lartillot, 2021)
Extraktion PSY- und RA-Featureset	HEAD Acoustics ArtemiS Suite 13.1
MP3-Dateigröße	LAME-Encoder: https://lame.sourceforge.io/

ANHANG D

DIGITALER ANHANG

Der digitale Anhang ist dieser Arbeit als USB-Stick beigelegt und umfasst folgende Dateien:

Index	Dateiname	Typ	Einsatz
D2	CompleteFeatureList.xlsx	Liste	Feature Auflistung
D3	Extract_BA_Features.mlx	Skript	BA-Feature-Extraktion
D4	Extract_MIR_Features.m	Skript	MIR-Feature-Extraktion
D5	FeatureAggregation_lmSlopeACF.m	Funktion	BA-Featureset
D6	FeatureAggregation_localMaximaRate.m	Funktion	BA-Featureset
D7	FeatureAggregation_meanProminence.m	Funktion	BA-Featureset
D8	FeatureAggregation_PC1PS2.m	Funktion	BA-Featureset
D9	FeatureAggregatrion_lmSlopePACF.m	Funktion	BA-Featureset
D10	Feature_TimePredictivityRatio.m	Funktion	BA-Featureset
D11	Feature_TimeRms.m	Funktion	BA-Featureset
D12	Feature_TimeZeroCrossingRate.m	Funktion	BA-Featureset
D13	multilevel_iteration_script.R	Skript	Multilevel-Modellierung
D14	Selection_percentileLassoGLMFunction.m	Skript	Feature-Selektion
D15	Utility_audioPreparationScript.m	Skript	Audio-Vorbereitung
D16	Utility_licols.m	Funktion	Vor Selektion
D17	Utility_makeParallelMIR.m	Funktion	MIR-Feature-Extraktion
D18	Utility_myACA.m	Funktion	BA-Featureset

Der Anhang ist unter der DOI 10.5281/zenodo.11060722 veröffentlicht.