

INDIVIDUAL DIFFERENCES IN MUSICAL SCENE ANALYSIS ABILITIES ACROSS NORMAL-HEARING AND HEARING- IMPAIRED INDIVIDUALS

von der Fakultät für Medizin und Gesundheitswissenschaften der Carl von Ossietzky
Universität Oldenburg zur Erlangung des Grades und Titels eines Doktors der
Philosophie (PhD) eingereichte Dissertation

von

Robin Hake

geboren am 20. März 1993

in Vechta, Deutschland

Robin Hake: *Modelling individual differences in musical scene analysis abilities for normal-hearing and hearing-impaired individuals*

BETREUER:

Prof. Dr. Kai Siedenburg, *Technische Universität Graz, Austria; Carl von Ossietzky Universität Oldenburg, Germany*

Prof. Dr. Steven van de Par, *Carl von Ossietzky Universität Oldenburg, Germany*

ERSTGUTACHTER:

Prof. Dr. Kai Siedenburg, *Technische Universität Graz, Austria; Carl von Ossietzky Universität Oldenburg, Germany*

WEITERE GUTACHTER:

Prof. Dr. Frank Russo, *Toronto Metropolitan University, Canada*

ACKNOWLEDGMENTS

This thesis has been written during my time as an Ph.D. student at the Music Perception and Processing lab in the Department of Medical Physics and Acoustics of the Carl von Ossietzky Universität Oldenburg in Oldenburg, Germany. This dissertation would not have been possible without the support, guidance, and encouragement of many individuals, to whom I wish to express my deepest gratitude.

First and foremost, I would like to thank Dr. Steven van de Par for referring me to the newly established Music Perception and Processing Lab, led by my current supervisor, Prof. Dr. Kai Siedenburg. His recommendation opened the door to an inspiring research environment that has profoundly shaped my academic journey.

My deepest appreciation goes to Prof. Dr. Kai Siedenburg, whose outstanding supervision has been instrumental in my development as a researcher. I greatly value his unique style of guidance, which masterfully balances autonomy with well-timed 'encouragement', fostering an aspiration to refine each piece of work to its highest potential. His dedication to research and mentoring is truly remarkable, and I am profoundly grateful to have had such a role model to inspire my future career.

I would also like to extend my heartfelt thanks to Prof. Dr. Daniel Müllensiefen, who has been an invaluable source of support throughout the years. The efficiency and clarity with which Daniel and Kai have provided feedback are unparalleled, and I have benefited greatly from their expertise and responsiveness. I am also especially thankful to Dr. Klaus Frieler, who reliably came to the rescue whenever my code 'refused' to cooperate. Furthermore, I am deeply grateful to my former colleagues, Aravindan Joseph Benjamin, Simon Jacobsen, Iris Mencke, and especially Michel Bürgel, whose unwavering helpfulness and support have been indispensable throughout this journey. Without Mic's generosity and assistance in times of need, the completion of this dissertation would have taken twice as long.

Finally, my profound thanks go to my wife, Svenja, for her relentless support and for consistently steering me in the right direction. Her encouragement and motivation have been invaluable throughout this process. Additionally, I am deeply thankful to my children, Fritz, Rosemarie and Leve Nori, for teaching me the art of perseverance - even in the absence of sleep.

This work has been shaped by the contributions, patience, and encouragement of many, and I am sincerely grateful to all who have been part of this journey.

Hude, April 2025
Robin Hake

ABSTRACT

This dissertation investigates individual differences in Musical Scene Analysis (MSA) abilities and explores how these abilities are shaped by hearing loss, working memory capacities, musical training, and age. MSA is conceptualised as the musical equivalent to Auditory Scene Analysis (ASA), the foundational auditory process that enables listeners to organise sound sources in complex acoustic scenes (Bregman, 1990).

The dissertation is structured around four empirical studies (Chapters 3 to 6), each targeting a specific aspect of individual variability in MSA ability. Before presenting the empirical studies, Chapter 2 provides a selective literature review examining the key determinants of individual variability in MSA. Drawing from adjacent fields such as ASA in speech perception, hearing impairment research, and music cognition, this review identifies four major contributors to individual differences: peripheral auditory functioning, cognitive abilities (especially attention and working memory), musical experience (i.e., musical training and expertise), and age-related changes. In addition, further modulating influences - such as sex and gender, socioeconomic and cultural background, and multisensory integration - are discussed.

Chapter 3 presents a large-scale survey on hearing health among professional and amateur orchestral musicians ($n = 770$), investigating the prevalence and impact of self-reported hearing problems along several other dimensions of hearing health. Results revealed that while professional musicians reported significantly higher lifetime exposure to music-related noise and demonstrated greater awareness and use of hearing protection than amateurs, both groups showed similarly high rates of hearing-related challenges, including hearing loss, tinnitus, pitch perception challenges, and difficulty identifying instruments in ensembles. These findings suggest that hearing issues in musicians cannot be fully explained by noise exposure or age, highlighting the need for behavioural assessments to better understand underlying perceptual mechanisms.

Chapter 4 addresses this methodological gap by introducing the development and validation of an adaptive MSA test. Designed for use with both normal-hearing and hearing-impaired listeners, the test employs realistic musical mixtures from various genres and uses a simple yes/no paradigm to assess whether a target instrument is perceptually detectable within a polyphonic musical scene. The test was calibrated in large-scale online studies ($N = 656$) and later tested under controlled lab conditions. Across both formats, performance was influenced by stimulus characteristics such as the choice of the target instrument, the target-to-mixture level ratio and number of instruments, and was robustly linked to individuals hearing thresholds. The test showed an acceptable test-retest reliability. Furthermore, correlations with other music perception tests and cognitive measures demonstrated good convergent

validity. Furthermore, substantial intercorrelations between determinants of MSA abilities have been found; prompting for a more sophisticated analysis.

Chapter 5 further explores the cognitive and experiential factors contributing to individual variability in MSA. Using a dual-model statistical approach (ridge regression and gradient boosting), this study examined MSA and speech-in-noise performance in three age and hearing-defined groups (N = 92). The results revealed that while hearing loss was a dominant predictor in both speech and music-based ASA, solely MSA performance was strongly influenced by musical training. Working memory capacity, while initially correlated with better performance, had only a minimal effect after accounting for hearing and cognitive variables. Ageing, however, still exerted a practically meaningful influence across both domains when considered over the entire lifespan. These findings reinforce the idea that MSA depends on a multifactorial interplay of peripheral, cognitive, and experiential factors and that musical training offers a domain-specific enhancement of auditory perception.

Chapter 6 applies the MSA test in among a sample of hearing aid (HA) users (N = 33), examining how HA signal processing, specifically dynamic range compression (DRC), affects MSA performance and subjective musical sound quality. The results showed that HA use improved performance across tasks, but effects of DRC speed varied by outcome: fast DRC enhanced MSA performance, while slow DRC yielded more favourable sound quality ratings. Notably, participants' subjective impressions of instrument clarity often contradicted behavioural results, suggesting that perceptual comfort and scene transparency may diverge in aided listening. Additionally, individual differences in benefit from specific HA settings were not predicted by age or musical background but were linked to duration of HA use and individual hearing thresholds. This study illustrates the importance of personalising HA fittings, especially for music perception, and demonstrates the practical utility of the MSA test in clinical assessment.

The integrated discussion in Chapter 7 synthesised findings from the preceding empirical studies within the broader literature, emphasizing MSA's dependence on converging peripheral auditory, cognitive, and experiential domains. Finally, implications for future research and practical applications are outlined.

ZUSAMMENFASSUNG

Diese Dissertation untersucht individuelle Unterschiede in der musikalischen Szenenanalyse (MSA) und erforscht, wie diese Fähigkeiten durch Hörverlust, Arbeitsgedächtniskapazitäten, musikalische Ausbildung und Alter beeinflusst werden. MSA wird hier als das musikalische Äquivalent zur Auditorischen Szenenanalyse (auch Auditive Szenenanalyse genannt; ASA) konzeptualisiert, dem grundlegenden auditiven Prozess, der es Zuhörern ermöglicht, Schallquellen in komplexen perzeptuelle Szenen zu organisieren (Bregman, 1990).

Die Dissertation ist in vier empirische Studien (Kapitel 3 bis 6) unterteilt, die jeweils einen spezifischen Aspekt der individuellen Variabilität in der MSA-Fähigkeit untersuchen. Vor der Präsentation der empirischen Studien bietet Kapitel 2 ein selektives Review, die die wichtigsten Determinanten der individuellen Variabilität in der MSA untersucht. Anhand angrenzender Bereiche wie der ASA in der Sprachwahrnehmung und der Hörschädigungsforschung identifiziert diese Übersicht vier Hauptfaktoren für individuelle Unterschiede: periphere auditorische Funktion, kognitive Fähigkeiten (insbesondere Aufmerksamkeit und Arbeitsgedächtnis), musikalische Erfahrung (d.h. musikalisches Training und Expertise) und altersbedingte Veränderungen. Darüber hinaus werden weitere modulierende Einflüsse wie Geschlecht, sozioökonomischer und kultureller Hintergrund sowie multisensorische Integration diskutiert.

Kapitel 3 beginnt mit einer groß angelegten Umfrage zur Hörgesundheit unter professionellen und Amateur-Orchestermusikern ($n = 770$), die die Prävalenz und die Auswirkungen von selbstberichteten Hörproblemen entlang verschiedener Dimensionen der Hörgesundheit untersucht. Die Ergebnisse zeigten, dass professionelle Musiker zwar eine signifikant höhere lebenslange Exposition gegenüber musikbedingtem Lärm berichteten und ein größeres Bewusstsein und eine stärkere Nutzung von Gehörschutz als Amateure zeigten, beide Gruppen jedoch ähnlich hohe Raten an hörbezogenen Problemen aufwiesen, einschließlich Hörverlust, Tinnitus, Herausforderungen bei der Tonhöhenwahrnehmung und Schwierigkeiten bei der Identifizierung von Instrumenten in Ensembles. Diese Ergebnisse deuten darauf hin, dass Hörprobleme bei Musikern nicht vollständig durch Lärmexposition erklärt werden können, was die Notwendigkeit behavioraler Messungen unterstreicht, um die zugrunde liegenden Wahrnehmungsmechanismen besser zu verstehen.

Kapitel 4 befasst sich mit dieser methodischen Lücke, indem es die Entwicklung und Validierung eines adaptiven MSA-Tests vorstellt. Der Test wurde für die Verwendung mit normalhörenden und hörgeschädigten Zuhörern entwickelt und verwendet realistische musikalische Mischungen aus verschiedenen Genres. Im Rahmen eines einfachen Ja/Nein-Paradigmas soll bewertet werden, ob ein Zielinstrument in einer polyphonen musikalischen Szene wahrnehmbar ist. Der Test wurde in groß angelegten Online-Studien ($N = 656$) kalibriert und später unter kontrollierten Laborbedingungen validiert. In beiden Formaten

wurde die Leistung durch Stimulusmerkmale wie die Wahl des Zielinstruments, das Ziel-zu-Mischungsverhältnis und die Anzahl der Instrumente beeinflusst und war robust mit den individuellen Hörschwellen verbunden. Eine akzeptable Test-Retest-Korrelation wurde nachgewiesen. Außerdem zeigen Korrelationen mit anderen Musikwahrnehmungstests und kognitiven Messungen zeigten eine gute konvergente Validität. Darüber hinaus wurden signifikante Interkorrelationen zwischen den Determinanten der MSA-Fähigkeiten gefunden, was zu einer anspruchsvolleren Analyse führte.

Kapitel 5 untersucht weiter die kognitiven und erfahrungsbedingten Faktoren, die zur individuellen Variabilität in der MSA beitragen. Unter Verwendung eines dualen statistischen Modellansatzes (Ridge-Regression und Gradient Boosting) untersuchte diese Studie die MSA- und Sprach-im-Rauschen-Leistung in drei alters- und hördefinierten Gruppen (N = 92). Die Ergebnisse zeigten, dass Hörverlust zwar ein dominanter Prädiktor sowohl für die sprach- als auch für die musikbasierte ASA war, die MSA-Leistung jedoch stark von dem Grade des musikalischen Trainings beeinflusst wurde. Die Arbeitsgedächtniskapazität korrelierte zwar zunächst mit einer besseren Leistung, hatte jedoch nach Berücksichtigung von Hör- und kognitiven Variablen nur einen minimalen bis keinen Effekt. Das Alter hatte jedoch über die gesamte Lebensspanne hinweg einen praktisch bedeutsamen Einfluss in beiden Bereichen. Diese Ergebnisse untermauern die Vorstellung, dass die MSA von einem multifaktoriellen Zusammenspiel peripherer, kognitiver und erfahrungsbedingter Faktoren abhängt und dass die musikalische Ausbildung eine relativ domänenspezifische Verbesserung der auditiven Wahrnehmung darstellt.

Kapitel 6 wendet den MSA-Test unter einer Gruppe von Hörgerätebenutzern (N = 33) und untersucht, wie die Signalverarbeitung von Hörgeräten, insbesondere die dynamische Bereichskompression (DRC), die MSA-Leistung und die subjektive musikalische Klangqualität beeinflusst. Die Ergebnisse zeigen, dass die Hörgerätenutzung die Leistung in allen Aufgaben verbesserte, die Auswirkungen der DRC-Geschwindigkeit jedoch je nach Ergebnis variierten: schnelle DRC verbesserte die MSA-Leistung, während langsame DRC günstigere Klangqualitätsbewertungen ergab. Bemerkenswerterweise widersprachen die subjektiven Eindrücke der Teilnehmer hinsichtlich der Instrumentenklarheit oft den Verhaltensergebnissen, was darauf hindeutet, dass das Wahrnehmungswohlbeinden und die Szenentransparenz beim verstärkten Hören auseinandergehen können. Darüber hinaus wurden individuelle Unterschiede im Nutzen bestimmter Hörgeräte-Einstellungen nicht durch Alter oder musikalischen Hintergrund vorhergesagt, sondern waren mit der Dauer der Hörgerätenutzung und den individuellen Hörschwellen verbunden.

Die integrierte Diskussion in Kapitel 7 fasst die Ergebnisse der vorangegangenen empirischen Studien im Rahmen der breiteren Literatur zusammen und betont die Abhängigkeit der MSA von zusammenlaufenden peripheren auditiven, kognitiven und erfahrungsbedingten Bereichen. Abschließend werden Implikationen für zukünftige Forschung und praktische Anwendungen skizziert.

GLOSSARY

Acronyms and abbreviations

ASA	Auditory Scene Analysis
B-GLMM	Bayesian generalised logistic mixed-effects model
BDS	Backwards Digit Span Memory Test
BF	Bayes Factor
CA-BAT	Computerised Adaptive Beat Alignment Test
CFA	Confirmatory Factor Analysis
CI	Confidence Interval
dB	decibel
DRC	Dynamic range compression
e.g.	exempli gratia (for example)
FDT	Frequency Discrimination Task
GMSI	Goldsmiths Musical Sophistication Index
HAD	Hearing Aids Disapproval Scale
HDDA	Hearing-Dependent Daily Activities Scale
HHA	Hearing Health Awareness Scale
HI	Hearing impaired
HL	Hearing loss
i.e.	id est (that is)
ICC	Intraclass correlation
ILD	Inter-aural level differences
IRT	Item response model
kHz	kilohertz
LOOIC	Leave-one-out cross-validation information criterion
MDT	Melodic Discrimination Test
MINT	Music-In-Noise Task
MPT	Mistuning Perception Test
MRHP	Music-related Hearing Problems Scale
MSA	Musical Scene Analysis
NH	Normal hearing
OLSA	Oldenburger Satztest
ORN	Object-related negativity
PTA	Pure tone average
SE	Standard error
SEM	Standard error of measurement
SNR	Signal-to-noise ratio
SPL	Sound pressure level

Acronyms and abbreviations *continued*

SQR	Sound quality rating
TPT	Timbre Perception Test
UNE	Sound level exposure
VR-12	Health-related quality of life
WM	Working memory
WMC	Working memory capacity

Mathematical notation

α	Level of significance
β	Beta coefficient
α	Level of significance
β	Beta coefficient
d	Cohen's d
η^2	Partial eta squared
χ^2	Chi-squared

CONTENTS

1 INTRODUCTION	16
1.1. Musical Scene Analysis	16
1.2. Aim and relevance of this dissertation	17
1.3. Structure of this dissertation	19
2 INDIVIDUAL DIFFERENCES IN MUSICAL SCENE ANALYSIS	20
2.1. Peripheral auditory factors	20
2.2. Cognitive factors	22
2.3. Musical sophistication and training	24
2.4. Ageing	26
2.5. Others	27
3 A SURVEY ON HEARING HEALTH OF MUSICIANS IN PROFESSIONAL AND AMATEUR ORCHESTRAS .	29
3.1. Abstract	31
3.2. Introduction	32
3.2.1. The current study	35
3.3. Methods	35
3.3.1. Participants	35
3.4. Procedure	35
3.5. Materials	36
3.6. Analysis	38
3.7. Results	39
3.7.1. Musician group matching	39
3.7.2. Hearing-related problems	41
3.7.3. Use of hearing protection	44
3.7.4. Noise Exposure	46
3.7.5. Hearing health awareness	47
3.7.6. Hearing aid disapproval	48
3.7.7. Hearing aid usage	49
3.7.8. Health-related quality of life	49
3.7.9. Inter-Scale Correlation Analysis	50
3.8. Discussion	51
3.8.1. Noise exposure and musicians hearing problems	51
3.8.2. Preventive measures and hearing health awareness	54
3.8.3. Hearing aid usage	55
3.8.4. Musicians' overall well-being	56
3.8.5. Limitations	57
3.9. Conclusion	58
3.10. Declarations	59
3.11. Literature	60
4 DEVELOPMENT OF AN ADAPTIVE TEST OF MUSICAL SCENE ANALYSIS ABILITIES FOR NORMAL-HEARING AND HEARING-IMPAIRED LISTENERS	65
4.1. Abstract	66
4.2. Introduction	68
4.2.1. The present study	70
4.3. Experiment 1: Calibration Phase - Part 1	72
4.3.1. Methods	72
4.3.2. Procedure	74
4.3.3. Participants	75
4.3.4. Data analysis	76
4.3.5. Results	79
4.3.6. Discussion	82
4.4. Experiment 2: Calibration Phase - Part 2	83
4.4.1. Methods	83

4.4.2.	Procedure	84
4.4.3.	Participants.....	84
4.4.4.	Data analysis	84
4.4.5.	Results.....	85
4.4.6.	Discussion.....	89
4.5.	Integrated Discussion: Calibration Phase	90
4.6.	Experiment 3: Test Validation	91
4.6.1.	Methods.....	92
4.6.2.	Procedure	95
4.6.3.	Participants.....	96
4.6.4.	Results.....	96
4.6.5.	Discussion.....	98
4.7.	General Discussion	100
4.8.	Declarations	103
4.9.	References.....	104
5	AUDITORY SCENE ANALYSIS IN MUSIC AND SPEECH: INDIVIDUAL DIFFERENCES IN LISTENING ABILITIES	109
5.1.	Abstract.....	110
5.2.	Introduction	112
5.3.	Results	115
5.3.1.	Modelling individual differences.....	116
5.3.2.	Overall factor strength.....	126
5.4.	Discussion	128
5.4.1.	Effect of hearing loss	128
5.4.2.	Effect of working memory capacities.....	129
5.4.3.	Effect of musical training.....	129
5.4.4.	Effect of ageing.....	130
5.5.	Conclusion.....	130
5.6.	Materials and Methods	131
5.6.1.	Participants.....	131
5.6.2.	Test battery.....	131
5.6.3.	Procedure	132
5.6.4.	Data Analyses.....	133
5.7.	Declarations	136
5.8.	References.....	137
6	MUSIC PERCEPTION WITH HEARING AIDS: COMPRESSION DIFFERENTIALLY AFFECTS MUSICAL SCENE ANALYSIS AND MUSICAL SOUND QUALITY	144
6.1.	Abstract.....	145
6.2.	Introduction	147
6.2.1.	Sound quality with hearing aids.....	147
6.2.2.	Auditory scene analysis in music.....	149
6.2.3.	The current study	150
6.3.	Methods.....	151
6.3.1.	Participants.....	151
6.3.2.	Materials	152
6.4.	Procedure	153
6.4.1.	Hearing aid fitting.....	154
6.5.	Analysis	154
6.6.	Results	155
6.6.1.	Effect of amplification and compression across listening tasks.....	155
6.6.2.	Amplification benefits across individuals and listening tasks	158
6.6.3.	Individual differences in hearing aid amplification benefit	161
6.6.4.	Simulated hearing aid output target-to-masker ratio of the MSA task.....	162
6.7.	Discussion	163
6.7.1.	Limitations	167
6.8.	Conclusion.....	168
6.9.	Declarations	169
6.10.	Literature.....	170

7	INTEGRATED DISCUSSION	175
7.1.	Discussion	175
7.1.1.	Subjective hearing problems	175
7.1.2.	Development of a behavioural measure of MSA	177
7.1.3.	Hearing loss, cognition, ageing, and musical training in MSA.....	179
7.1.4.	MSA abilities among hearing aid users	183
7.2.	Implications for future research and possible applications	186
7.3.	Conclusion.....	187
8	APPENDICES	188
8.1.	Figures	188
8.2.	Tables.....	204
9	LITERATURE.....	225
10	LIST OF PUBLICATIONS	241
11	AUTHOR'S DECLARATION.....	242

LIST OF FIGURES

Fig. 3.1 ^a	Demographic profiles of amateur and professional musicians as a function of age, gender, and instrument group distributions before and after the matching procedure.	41 ^b
Fig. 3.2	Distribution of reported hearing loss among amateurs and professional musicians by age group.	42
Fig. 3.3	Amateur and professional musicians scores for Music-Related Hearing Problems and Hearing-Dependent Daily Activities.	44
Fig. 3.4	Musical sound exposure per year and the cumulative music sound exposure for amateur and professional musicians.	46
Fig. 3.5	Scores of for the scales Hearing Health Awareness and Hearing Aid Disapproval as a function of group, hearing loss, and age.	48
Fig. 3.6	Health-related quality of life Veterans RAND 12-Item Health Survey as a function of group, hearing loss, and age.	50
Fig. 4.1	Exemplary schematic illustration of the procedure.	72
Fig. 4.2	Results of Experiment 1.	80
Fig. 4.3	Schematic illustration of the stereo width conditions in frontal azimuth in Experiment 2.	83
Fig. 4.4	Results of Experiment 2.	88
Fig. 4.5	Test-retest-reliability of the first and second set of the MSA as a function of test length is shown on the left and the standard error of measurement as a function of test length on the right.	97
Fig. 5.1	Participant distribution by age and hearing thresholds.	115
Fig. 5.2	Auditory scene analysis scores for music and speech by groups.	116
Fig. 5.3	Effect of hearing thresholds on ASA abilities in music and speech.	119
Fig. 5.4	Effect of working memory capacities (WMC) on ASA abilities in music and speech.	121
Fig. 5.5	Effect of musical training (Gold-MSI subscale) on ASA abilities in music and speech.	123
Fig. 5.6	Effect of age on ASA abilities in music and speech.	125
Fig. 5.7	SHAP-based model predictions and ridge regression estimates for ASA abilities in music and speech.	127
Fig. 5.8	SHAP-based feature contributions for ASA abilities in music and speech.	135
Fig. 6.1	Average pure-tone audiometric threshold (PTA) measurements across participants for both ears.	151
Fig. 6.2	Differences in mean musical scene analysis (MSA) scores, overall sound quality rating (SQR) scores, and (C) speech reception thresholds (SRT in dB SNR) across different conditions: slow, unaided, fast.	156

^a Figures (and tables) are numbered by chapter and sequence. 'Fig. 3.1' denotes the first figure in the third chapter, while 'Fig. A3.1' indicates the first supplementary figure in the appendix of that chapter.

^b Page number.

Fig. 6.3	Effects of hearing aid compression settings on musical scene analysis (MSA), overall sound quality rating (SQR), and speech reception threshold (SRT) scores.	159
Fig. 6.4	Individual preferences (heatmap) for all three conditions (unaided vs. aided fast DRC vs. aided slow DRC) across the three tasks: Musical Scene Analysis test (MSA), overall Sound Quality Ratings (SQR), and Speech Reception Thresholds (SRT).	161
Fig. 6.5	Scatterplots illustrating the relationship between individual characteristics and combined amplification benefit, a composite score reflecting the summed standardised benefit across all tasks (MSA, SQR, and SRT) under both DRC conditions.	162
Fig. 6.6	Difference in TMR between fast and slow DRC versus real ear TMR as reference for various musical instruments.	163
Fig. A3.1	Density plots, along with boxplots and individual data points, illustrate the distribution of Goldsmiths Musical Sophistication Index scores.	188
Fig. A3.2	Distribution of primarily played orchestral category/repertoire among respondents, shown post-matching procedure.	189
Fig. A3.3	This multi-figure depicts the distribution of hearing-related measures and music exposure across age groups in the matched sample of amateur and professional musicians.	190
Fig. A3.4	This multi-figure depicts the distribution of hearing-related measures and music exposure across hearing loss groups in the matched sample of amateur and professional musicians.	191
Fig. A4.1	Graphical illustration of the pure-tone audiometric threshold (PTA) measurements of all participants for the respective better listening ear.	192
Fig. A4.2	Correlation matrix of the adaptive MSA in relation to other psychoacoustic and musical-related tests.	193
Fig. A5.1	Pure-tone audiometric thresholds (PTA) for all participants' better listening ear.	194
Fig. A6.1	Cook's distance analysis for identifying influential data points.	195
Fig. A6.2	Estimated means from Model A1 for musical scene analysis (MSA) scores, Model B2 for sound quality rating factor (SQR) scores, and Model C1 for speech reception thresholds (SRT) across three conditions: slow, unaided, fast.	196
Fig. A6.3	Relationship between overall sound quality ratings (SQR) and individual sound quality dimensions: clearness, fullness, loudness, and sharpness.	197
Fig. A6.4	Comparative assessment of sound quality attributes under different listening conditions.	198
Fig. A6.5	Mean overall sound quality rating (SQR) factor scores across three listening conditions (slow, unaided, and fast) for classical and pop excerpts.	199
Fig. A6.6	Preferred listening condition for each sound quality dimension.	200
Fig. A6.7	Scatterplots showing the relationships between total benefit scores across three measures: SQR (overall Sound Quality Rating) and MSA (Musical Scene Analysis abilities), SRT (Speech Reception Threshold in dB SNR) and MSA, and SRT and SQR.	201
Fig. A6.8	Scatterplots illustrating the relationships between benefit scores across tasks and DRC settings (fast and slow).	202

LIST OF TABLES

Table 3.1	Descriptive Statistics of hearing protection usage by professional and amateur musicians.	45
Table 3.2	The correlation matrix of selected tests from the test battery.	51
Table 4.1	Model comparison of all Bayesian (GLM) Models of Experiment 1.	78
Table 4.2	Model comparison of all Bayesian (GLM) Models of Experiment 2.	85
Table 4.3	The correlation of the adaptive MSA with selected tests from the test battery.	98
Table 5.1	Correlation matrix of test measures.	117
Table A3.1	Self-assessment of individual' hearing impairment.	204
Table A3.2	Self-assessment of individual' hearing impairment in German language.	205
Table A3.3	Musicians average sound level exposure per instrument.	206
Table A3.4	Musicians average sound level exposure per instrument group.	208
Table A3.5	Types of hearing protection: Average attenuation levels and usage.	209
Table A3.6	Types of hearing protection: Average attenuation levels and usage per group.	210
Table A3.7	Results from CFA for items indicating music-related hearing problems.	211
Table A3.8	Results from CFA for items indicating hearing health awareness (HHA).	212
Table A4.1	Model summary of the final Bayesian Model (1F) of Experiment 1.	213
Table A4.2	Model summary of the final Bayesian Model (2F) of Experiment 2.	214
Table A4.3	Model summary of the simple Bayesian Model (2C) of Experiment 2.	215
Table A4.4	Self-assessment of individual' hearing impairment.	216
Table A4.5	Self-assessment of individual' hearing impairment for German participants.	217
Table A6.1	Sound excerpts used within the sound quality rating task.	218
Table A6.2	Sound quality dimensions of the sound quality rating task in the original language (German).	219
Table A6.3	Descriptive Statistics and Correlations for Study Variables.	220
Table A6.4	Final model (A1) summary for the musical scene analysis (MSA) ability scores.	221
Table A6.5	The correlation matrix for the test battery.	222
Table A6.6	Final model (B2) summary for the sound quality rating (SQR) factor scores.	223
Table A6.7	Final model (C1) summary for the speech reception thresholds (SRT) from the Göttinger Sentence Test.	224

INTRODUCTION

Listening to music is often described as a universal experience, yet the way individuals perceive and organize sound can differ – at times considerably. Consider a conductor standing in front a large orchestra, attuned to every detail of the performance. As the strings play in unison, a barely perceptible dissonance catches their attention - one of the seven violins plays slightly out of tune. Meanwhile, in a different setting, another listener is struggling to follow the bassline in a ‘simple’ pop song, unable to distinguish it from the surrounding instruments. What allows one listener to disentangle the multifaceted layers of an orchestral performance while another finds it difficult to separate even a few instruments in a relatively sparse arrangement? Is it a matter of expertise, training, or something more fundamental about auditory perception? While music itself remains unchanged, the way it is experienced can vary profoundly from one individual to another. This dissertation seeks to explore these differences in music perception – more specifically, differences in *Musical Scene Analysis* (MSA) abilities.

1.1. Musical Scene Analysis

Musical perception in complex auditory environments represents a formidable challenge for the human auditory system. When multiple sound sources are present, their acoustic energy often overlaps in time and frequency, creating dense auditory mixtures. To derive meaning from these mixtures, the auditory system must assign sensory elements to individual sound sources and separate them from competing signals. MSA refers to the process by which listeners organize these ‘chaotic’ sound mixtures into distinct perceptual auditory objects, called ‘streams’. Albert Bregman’s seminal work on Auditory Scene Analysis (ASA; 1990) laid the foundation for this concept, describing how the auditory system groups and segregates sounds so that each stream is analogous to a visual object – a coherent percept that can be followed over time. Within a rich musical passage (e.g. an orchestral performance or polyphonic composition), individual streams may represent individual instruments, melodic lines, or vocal parts, each perceived as a separate stream amidst the musical mixture.

Listeners rely on a range of perceptual mechanisms - ‘Gestalt’ principles of perceptual organization - to make sense of these complex musical scenes, allowing them to extract meaningful auditory objects from a dense and overlapping acoustic mixture. For instance, just as visual objects are grouped based on similarity and proximity, the auditory system applies analogous principles: One key factor is temporal proximity, where sounds occurring in rapid succession are more likely to be perceived as belonging to the same auditory object, whereas longer silent gaps facilitate segregation (Bregman & Campbell, 1971; van Noorden, 1975). Proximity with respect to frequency and loudness also play a role, as tones that are close in frequency or intensity are more likely to be grouped together (McAdams & Bregman, 1979). Harmonicity and spectral regularity further contribute to auditory organization; when multiple frequency components share a common fundamental frequency, they tend to be

fused into a single auditory object (Sandell & Darwin, 1996; Darwin, 2005). Another critical factor in MSA is timbre similarity, which refers to the unique spectro-temporal composition of a sound, commonly associated with different musical instruments (McAdams et al., 2004). Despite this, when different frequency components start and change in the same way at the same time, they also have greater probabilities of being grouped as they probably arise from the same source (principle of common onset and common fate; Bregman & Pinker, 1978).

These principles underpin the stimulus-driven (bottom-up) mechanisms of auditory processing that allow the auditory system to segregate complex acoustic scenes. They form the basis for auditory streaming processes described by Bregman (1990) as ‘primitive grouping’. In contrast, ‘schema-based’ grouping (top-down processing) involves higher-level knowledge, learned patterns, and expectations (i.e., the schemas) to organize sound. Here, the listener’s experience and musical knowledge shape perception. A familiar melody, for example, can be perceptually restored even when partially masked by competing sounds, as listeners mentally “fill in” expected patterns (Bey & McAdams, 2002). Bregman’s framework (1990) suggests that primitive grouping establishes the initial perceptual organization, while schema-based processes refine and enhance auditory stream segregation using memory, learned knowledge, and selective attention. This interplay between stimulus-driven organization and cognitive modulation is central to understanding individual differences in MSA. MSA extends these general auditory streaming principles to the specific context of music perception, allowing listeners to form auditory streams from the “messy” acoustic input, turning a cacophony of simultaneous notes into an organized musical experience while also integrating them into meaningful patterns even in the context of music, *“which arguably provides some of the most complex acoustic scenes that a human listener will ever encounter.”* (Pressnitzer et al., 2011, p.1).

1.2. Aim and relevance of this dissertation

As illustrated in the opening paragraph of this dissertation, for some individuals, the ability to follow a single instrument amidst the aggregated layers of an orchestral performance seems almost intuitive (e.g., conductors), while others struggle to discern even the most prominent instruments in rather simple musical arrangements. This disparity illustrates the significant role of individual differences - the unique characteristics and traits that distinguish individuals from each other (Williamson, 2018) - in shaping MSA. Recent studies employing both realistic musical recordings and controlled paradigms using artificially constructed stimuli underscore the diversity of MSA abilities among listeners. For instance, Janata et al. (2002) asked individuals to focus on a specific instrument within a polyphonic musical piece and rate the difficulty of the task. The results showed substantial variability between listeners, with some participants finding the task effortless, while others rated it as highly challenging. Behavioural experiments further demonstrate differences. In a study by Siedenburg et al. (2020), listeners had to discriminate melody or timbre differences within a musical accompaniment, revealing individual performance differences of up to 25 dB in signal-to-masker ratios (SMR) for NH listeners. Variability was even more pronounced among older hearing-impaired (HI) participants, where SMR differences reached 35 dB, with one in

four listeners being unable to perform the task and exhibiting chance-level performance. Substantial interindividual variability in performance was also reported in other studies employing an instrument detection task with realistic polyphonic music (Bürgel et al., 2021, 2023). Similar findings have emerged across various MSA studies, where even some NH individuals fail to complete the task successfully (e.g., Kirchberger & Russo, 2015; Siedenburg et al., 2021). Accordingly, MSA can neither be a uniform ability nor one that can be fully predicted based on conventional measures such as audiometric hearing thresholds alone. Barbara Shinn-Cunningham (2017) suitably described this variability in her review about individual differences:

“Still, in my own lab, we find that even young, healthy adults with NHTs [normal hearing thresholds] have smaller, but still consistent, significant and behaviourally relevant individual differences in hearing ability.”

(Shinn-Cunningham, 2017, p. 2980)

Given this unexplained variability, this dissertation seeks to answer the central research question: *What factors explain individual differences in Musical Scene Analysis abilities and how do they shape musical listening experiences?* Typical research design often oversimplifies the complexity of relationships arising from these differences and thus does not unambiguously allow for inferences about potential factor contributions. Rather, these designs tend to obscure underlying processes by shrouding rich individual data with group data aggregation procedures that mask the rich variability found at the individual level, thereby limiting our understanding of the underlying mechanisms driving perceptual and cognitive performance (O'Connor, 1990). Individual differences in a complex latent construct such as MSA abilities cannot be adequately represent by a single measure or aspect abilities either (Nesselroade & Ford, 1987; Schwartz, 1986). Instead, such constructs require a multidimensional representation - capturing a broader constellation of factors rather than isolated indicators (Cattell, 1966). In response, this dissertation adopts a multivariate approach that foregrounds individual differences by examining how multiple interacting factors uniquely shape listener's perceptual experience.

By adopting this approach, the dissertation intends to advance theoretical models of MSA, elucidate its underlying mechanisms, and inform the design of interventions that enhance music perception for both NH and HI individuals. The implications of these findings extend beyond laboratory measurements, offering practical applications ranging from optimizing hearing aid and cochlear implant technology and auditory training programs to improving accessibility in music education and concert experiences, ultimately improving music accessibility and enjoyment for both NH and HI individuals.

1.3. Structure of this dissertation

The remainder of this dissertation is structured as follows. Chapter 2 provides a selective literature review on individual differences in MSA, drawing from research on ASA, speech perception, music cognition, and research on hearing impairments. Chapters 3 through 6 each present an individual study examining different aspects of MSA variability: Chapter 3 explores the impact of musical sound exposure on professional and amateur musicians' hearing abilities, utilizing a survey-based approach. Chapter 4 introduces a novel experimental test designed to efficiently quantify MSA abilities behaviourally in both NH and HI individuals. Chapter 5 applies this new test to a diverse sample of listeners, disentangling the effects of hearing impairment and cognitive factors on MSA. Chapter 6 extends this investigation by focusing on hearing aid users with moderate-to-severe hearing loss, examining how hearing aid amplification and compression affects individual musical scene perception. Finally, Chapter 7, the integrated discussion, synthesizes findings from the empirical studies, integrating them into a broader framework of music perception. Through this structure, the dissertation aims to provide a comprehensive account of the mechanisms underlying MSA, with implications for both music cognition and clinical audiology.

INDIVIDUAL DIFFERENCES IN MUSICAL SCENE ANALYSIS

Given the acoustic complexity of musical scenes, variability in listeners' ability to parse and interpret music arises from multiple factors. According to Bregman (1990), these individual differences can largely be attributed to two principal sources: (1) bottom-up, stimulus-driven processes governed by the auditory periphery, and (2) top-down cognitive mechanisms, that shape perception through attention, memory, and prior knowledge. The following section reviews these factors in detail. Given the limited number of studies that directly employ MSA paradigms, this selective review draws upon adjacent domains of auditory perception research, including auditory neuroscience, speech-based ASA, hearing impairment research, and music cognition.

2.1. Peripheral auditory factors

MSA fundamentally depends on the ability to perceive the elementary acoustic information in the environment. This may involve frequency components, tempo, spatial location, and harmonic structure (Bernstein & Oxenham, 2006; Bregman, 1990; Carlyon et al., 2001; Ciocca, 2008; Cusack et al., 2004; Marozeau et al., 2013; Moore & Gockel, 2002). Any variation in, or damage to, the auditory pathways, such as that observed in sensorineural hearing loss, may alter or disrupt the processing of essential acoustic cues required for grouping and segregation of sounds within an auditory scene. Indeed, research consistently demonstrates that hearing impairments compromise auditory perception, including frequency resolution, temporal processing, and spatial hearing (Akeroyd et al., 2007; Bronkhorst & Plomp, 1990, 1992; Middlebrooks et al., 2017; Phillips et al., 2000; Shinn-Cunningham & Best, 2008). These limitations ultimately have, depending on the severity of the impairment, pronounced implications for the formation of auditory objects and the ability to segregate overlapping sounds in intricate acoustic scenes. For example, streaming based on frequency and harmonicity cues is notably impaired among (older) adults with hearing loss (Grimault et al., 2001; Oxenham, 2008; Shinn-Cunningham & Best, 2008) and such deficits appear already early in development: In a study comparing HI children with NH controls, Mehrkian et al. (2022) observed attenuated and delayed N2 amplitudes in response to tuned stimuli as well as reduced object-related negativity (ORN) to mistuned stimuli, indicating diminished capacity to detect and segregate sounds based on harmonic cues. Other studies report that adult HI individuals are less able to rely on interaural timing and phase cues, which typically aid in the segregation of target signals from competing sound sources (Füllgrabe et al., 2014; Jerger et al., 1984; Koehnke et al., 1995; Olsen & Noffsinger, 1976). Broader auditory filters and reduced sensitivity to temporal fine structure have also been documented in HI listeners, further compromising their ability to segregate sounds based on subtle spectral and temporal differences (Hopkins et al., 2008; Moore et al., 2006; Shinn-Cunningham & Best, 2008).

To directly assess these influences on auditory streaming, many researchers have employed the classic ABA paradigm (van Noorden, 1975), in which triplet tone sequences are systematically manipulated to vary pitch differences. In this task ‘A’ and ‘B’ represent tones that differ in core auditory attributes like frequency and are presented within silent gaps between triplets. When the A and B tones differ by a small amount in, e.g., frequency, a ‘galloping’ rhythm is heard. As the separation between the A and B tones increases, the likelihood of hearing two simultaneous streams increases. There are variations of this paradigm to specific test the influence of fundamental frequencies and harmonicity, inter-stimulus intervals, intensity differences, spatial separation, timbre, rhythm, pitch modulation (i.e., pitch glide), but also attention (e.g., Boehnke & Phillips, 2005; Carlyon et al., 2001; Haywood et al., 2024; Sussman, 2017). And indeed, HI listeners generally require larger pitch separations (Bayat et al., 2013; Bernstein & Oxenham, 2006), greater mistuned harmonics (Grimault et al., 2001), and greater spatial separation of sources, especially with asymmetrical hearing loss (Moore & Gockel, 2002) to perceive stream segregation, with attentional dysfunctions impairing stream formation even irrespective to the severity of hearing deficits (Carlyon et al., 2001). Tinnitus, a condition characterised by the perception of sound without external stimuli (Møller, 2006), can further impair ASA abilities beyond those caused by hearing loss. In some cases, the salience, persistent magnitude, and psychoacoustic characteristics of the tinnitus signal (in terms of frequency, intensity, bandwidth, and timing) make it resistant to masking and draw attention, ultimately disrupting the ASA process (Durai et al., 2018, 2019; Searchfield et al., 2007, 2012). Indeed, Shinn-Cunningham and Best (2008) emphasize that peripheral hearing deficits like tinnitus increase the cognitive demands associated with auditory object formation and selective attention, leading to slower, less efficient scene analysis. Other evidence suggests largely intact segregation abilities (speech-in-noise), but suggest particular problems with informational masking for tinnitus sufferers (Wang et al., 2022). Yet, even listeners with clinically normal hearing thresholds may exhibit subtle deficits in the fidelity of auditory sensory coding - such as reduced ability to encode fine temporal details crucial for separating auditory sources. This “hidden hearing loss,” or cochlear synaptopathy, impairs the processing of important auditory cues like pitch, spatial location, and timbre (Bharadwaj et al., 2015; Shinn-Cunningham, 2017).

Together, these impairments fundamentally impact everyday auditory perception, such as speech intelligibility, particularly in complex acoustic environments that involve similar cognitive and perceptual mechanisms (Bayat et al., 2013; Johannesen et al., 2016; Ruggles et al., 2012; Shinn-Cunningham & Best, 2008). In contrast, research on the effects of hearing loss in music perception is only beginning to gain momentum. Recent findings suggest that these challenges indeed extend to realistic musical contexts. For instance, in a study using spatially separated excerpts from *Bach’s Art of the Fugue*, older HI adults exhibited reduced ability to follow individual musical voices compared to NH listeners (Siedenburg et al., 2021). Similarly, in melody and timbre discrimination tasks involving musical masking, HI listeners required signal-to-masker ratios (SMRs) approximately 10 dB greater than those of young NH participants (Siedenburg et al., 2020). Importantly, it has been suggested that HI individuals failed to benefit from momentary drops in masker level (known as “listening in the dips”; see also Bronkhorst & Plomp, 1992) – a strategy that is particularly crucial to make effective use

of target object information when interfering sources are present (Rosen, 1992). Extending this line of research, Benjamin and Siedenburg (2024, 2025) employed multi-track music mixes to assess MSA performance among older HI and younger NH listeners. In these tasks, participants had to detect target instruments in complex musical mixtures (similar to the task used in Bürgel et al., 2021, 2023). Results revealed that the HI listener group performed worse than their NH counterparts. Notably, MSA performance declined systematically with increasing hearing loss, with each additional decibel of hearing threshold associated with a 3% reduction in the probability of correctly detecting the target. Furthermore, MSA ability was correlated with subjective audio quality ratings among HI participants (Benjamin & Siedenburg, 2024), providing first indication that the ability to perceptually segregate musical streams contributes to the overall subjective experience of musical mixtures. And while certainly representing a more extreme scenario of impairment, both hearing aid and cochlear implant users have been found to have particular difficulty segregating musical melodies and voices (Limb & Roy, 2014; Paredes-Gallardo et al., 2018; Uys & van Dijk, 2011; Uys & Latzel, 2015, 2016) - with many complain about a lack of musical sound quality and clarity (“smeared/fuzzy/washed out” percept) when listening to music (Madsen & Moore, 2014). All together, these results emphasize that hearing loss not only limits access to acoustic detail but also impairs the ability to perceptually organize complex musical scenes.

2.2. Cognitive factors

The process of ASA is, though grounded in low-level auditory mechanisms, not merely a primitive perceptual process; it is rather deeply influenced by a complex interplay of cognitive factors that shape how listeners perceive and interpret auditory scenes (Bregman, 1990; Bey & McAdams, 2002). Accordingly, a substantial body of research highlights the pivotal role of higher-order cognitive processes in shaping the organisation of complex auditory mixtures (e.g., Wingfield, 1996). Central among these is attention, which acts as like a spotlight within auditory scenes, directing cognitive resources to relevant auditory objects (Auerbach & Gritton, 2022; Kraus et al., 2012). Although early segregation occurs automatically and sets the basis for auditory event formation, auditory stream formation is significantly influenced by attentional engagement and cognitive load (Carlyon et al., 2001; Elhilali et al., 2009; Thompson et al., 2011). For instance, attended auditory objects - such as a violin line within a symphony - have been shown to have enhanced spectro-temporal representations in the auditory cortex (Bey & Zatorre, 2003; King & Walker, 2020; Snyder et al., 2006). Moreover, attention accelerates the perceptual segregation of competing sound streams (Garami et al., 2020), and individual differences in attentional capacity have been shown to predict stream segregation performance in complex listening tasks (Oberfeld & Klöckner-Nowotny, 2016). Similarly, instruments or lead-vocals that naturally draw attention (i.e., salient sources) are more readily detected in polyphonic musical mixtures, even during holistic listening (Bürgel et al., 2021, 2023).

Variations in attention-related cognitive processes, such as working memory (WM) and processing speed, are equally crucial (Wingfield, 1996; Deary, 1994): those individuals with reduced WM capacity often face greater difficulty in navigating in acoustically adverse

(e.g., noisy and reverberant) or informationally complex (e.g., multitalker) auditory scenes (Pichora-Fuller et al., 2016; Zekveld et al., 2009); while individuals with greater WM resources are better able to adapt to difficult listening situations (Lentz et al., 2022; Nitsan et al., 2022; Rudner, 2016). The role of WM has been most extensively studied in speech-based scene analysis. Numerous studies have found moderate to strong correlations between WM capacity and performance on speech perception tasks (see Akeroyd, 2008; Besser et al., 2013; Heinrich et al., 2016). For example, Escobar et al. (2020) showed that even among young NH adults, those with higher WM capacity demonstrated better performance in speech stream segregation tasks and a comprehensive review by Dryden et al. (2017) identified WM, along with processing speed, inhibitory control, crystallized IQ, as major predictors for speech-based stream segregation performance (see also Davidson & Souza, 2024). These findings are consistent with theoretical models, such as the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013), which proposes that WM resources are particularly important for processing of speech; a process that plausibly extends to the perception of lead-vocals in complex musical mixtures. However, the association between WM and speech perception is not consistently observed across all contexts. Emerging evidence suggests that this relationship is modulated by hearing status. Specifically, several studies indicate that the link between WM and speech intelligibility is particularly robust in individuals with HI, but less consistent among NH listeners (Besser et al., 2012; Heinrich et al., 2016; Janse & Andringa, 2021). This pattern supports the interpretation that WM plays a more prominent role in speech segregation when bottom-up acoustic cues are degraded, whereas it may be less critical when auditory inputs are clear and fully intelligible (see review by Füllgrabe & Rosen, 2016).

The role of WM has also been explored within the musical domain. For example, WM has been linked to the sensitivity to harmonic differences, the ability to recognise melodies within musical mixtures, and rhythm processing (Bailey & Penhune, 2010; Eitel et al., 2024; Harrison et al., 2017; Silas et al., 2020). Although notoriously underexplored, the role of WM has also been supported by recent findings using music-specific stream segregation tasks. For instance, Liu et al. (2025) demonstrated a significant relationship between auditory WM abilities and performance on a Music-In-Noise Task (MINT; see Coffey et al., 2019), in which participants discriminated between melodic patterns embedded in musical masking noise. Participants with higher auditory WM performance showed superior abilities to identify melodies under noisy conditions. These findings underline that auditory WM is a key cognitive mechanism facilitating stream segregation in music. Neurophysiological evidence further supports the role of cognitive mechanisms in MSA. Janata et al. (2002) demonstrated that tracking a specific instrument within a realistic polyphonic musical piece was associated with increased activity in frontal and parietal regions, areas implicated in WM and cognitive control (see also Ragert et al., 2014). Moreover, difficulty ratings varied across individuals, likely reflecting differences in WM and attentional capacity.

In summary, while existing research points to a meaningful role of WM in musical stream segregation, the strength and consistency of this relationship, particularly in NH individuals, remains insufficiently understood. Mixed findings in the broader literature, along

with evidence suggesting interaction effects with factors such as hearing impairment, task complexity, and stimulus characteristics, highlight the need for further investigation.

2.3. Musical sophistication and training

Alongside attentional and cognitive processes, listeners naturally acquire a rich understanding of sound structures, such as transition probabilities between sound elements and correlations between sound properties (Pearce et al., 2010; Rohrmeier & Rebuschat, 2012). Prior knowledge of musical structure (also referred to as ‘schemas’) - an aspect of top-down cognitive processing - plays a critical role in MSA and has been shown to facilitate the recognition and analysis of familiar auditory patterns (Bey & McAdams, 2002; Kothinti & Elhilali, 2023; Woods & McDermott, 2018). For example, familiar melodies are more easily extracted from complex mixtures (Szalárdy et al., 2014), an effect also observed with synthetic stimuli (Woods & McDermott, 2018). Structural knowledge of this kind is especially prominent among those individuals with explicit musical training and formal musical education: professional musicians. Indeed, musicians, with their extensive experience in musical structure, are more sensitive to stimuli based on pitch, time and loudness (Brown et al., 2017; Dubinsky et al., 2019; Marques et al., 2007; Kannyo & DeLong, 2011). Moreover, professional musical training has been associated with enhanced encoding of temporal fine structure cues (Mishra et al., 2015) and improved fundamental frequency discrimination in both young and older listeners, regardless of hearing impairments (Bianchi et al., 2019; Sandell & Darwin, 1996). Beyond domain-specific perceptual enhancements, musical training has been further linked to broader cognitive benefits such as faster information processing (Bugos et al., 2024), enhanced attentional resolution (Nandi et al., 2023), and improved WM (Bugos et al., 2007; Herholz & Zatorre, 2012; Talamini et al., 2017; Zhao, 2024), including auditory WM (Liu et al., 2025) – all of which are linked to auditory perceptual performance (see Dryden et al., 2017).

Musicians are thus assumed to be particularly adept at MSA. Supporting this view, Zendel and Alain (2009) showed that musicians outperformed non-musicians in a concurrent sound segregation task based on harmonicity, with neural evidence (object related negativity and P400 components) paralleling behavioural performance. Susini et al. (2023) investigated how musicianship influences interleaved melody recognition, a task requiring listeners to extract a target melody from a stream of overlapping, frequency-interleaved notes. Their results revealed that expert musicians exhibited superior performance, even when frequency separations between target and distractors were minimal, suggesting a heightened capacity for frequency-based stream segregation. These findings replicate and extend earlier work by Bey and McAdams (2002), who also reported musician-related advantages in MSA tasks (see also Slater & Marozeau, 2016). Notably, in contrast to Bey and McAdams, Susini et al. showed that amateur musicians, despite expressing some enhancements in temporal processing, did not exhibit the same streaming advantage as expert musicians in their study, indicating a continuum-based relationship of musical training and MSA abilities. Furthermore, the study explored the potential mediating role of WM, observing that both expert and amateur musicians outperformed non-musicians in a backward digit recall task, reinforcing the

interplay between auditory cognition and stream segregation abilities. Interestingly, Wenhart et al., (2019) used a similar task and showed that among professional musicians, those with absolute pitch are further outperforming those musicians with relative pitch. Consistent with these findings, Sauv   et al. (2022) reported that individuals with higher levels of musical training were more successful at detecting pitch deviants in a four-note melody interleaved with distractors (see also Marozeau et al., 2010).

More recently, Liu et al., (2025) replicated findings of a study by Coffey et al., (2019), who introduced the MINT, showing superior performance for musicians with at least 10 years of experience; an effect that was particularly consistent over varying SNRs which further underscores the robustness of this finding. Importantly, in the original study, Coffey et al. (2019) could demonstrate this effect irrespective of a binary classicisation system for musicianship, in which they demonstrated a direct relationship between the total cumulated practice hours and MINT performance – together, a much more compelling evidence in favour of a causal effect (see Pearl & Mackenzie, 2018). By also considering musical training as a continuum, another study conducted by Siedenburg et al. (2020) showed that musically training is associated with lower signal-to-masker thresholds when detecting tones within a musical accompaniment. Moreover, even individuals without formal musical training but with extensive informal musical experience demonstrate enhanced music perception abilities comparable to formally trained musicians (Correia et al., 2023; see also Micheyl et al., 2006). In fact, the benefits associated with musical experience are so multifaceted that several researchers advocate musical practice as a viable cognitive training intervention to support older adults in compensating for sensory ageing and hearing loss (Dubinsky et al., 2019 Reuter-Lorenz & Park, 2014; Zendel & Alain, 2012), and potentially even mitigating neuropsychiatric symptoms associated with dementia (see Bugos et al., 2024, for a meta-analytic review).

In contrast, when auditory scenes are dominated by speech, such as in conversational or multitalker environments, the benefits of musical training appear less consistent. While some evidence suggests a benefit for speech-in-noise perception, this tends to emerge only for younger listeners and under particularly challenging listening conditions, with minimal or no benefit observed in less demanding scenarios (see Maillard et al., 2023 for meta-analysis). Moreover, the effects are less reliably observed for HI individuals (see McKay, 2021 for systematic review). Certainly, the influence of musical training on speech perception remains a subject of ongoing debate (see also Coffey et al., 2017; Madsen et al., 2019; Schellenberg & Lima, 2024). Similarly, the relationship between musical training and MSA is also not universally observed across all contexts (e.g., Carey et al., 2015; Crawley et al., 2002). Siedenburg et al., (2023) also found no significant correlation between self-reported musical training and the ability to detect target instruments in realistic musical mixtures (see also B  rgel et al., 2021; Benjamin & Siedenburg, 2024, 2025). Likewise, Ba  kent et al. (2018) found no group differences between musicians and non-musicians among young adolescents in a task requiring identification in identifying five-note organ melodies masked by simultaneous piano chords.

2.4. Ageing

Ageing is afflicted to both peripheral auditory and central cognitive changes, affecting MSA in complex, interconnected ways. As people age (beginning typically in later adulthood), cochlear damage (often due to cumulative noise exposure), can result in a degraded auditory input, leading to a general reduction in the density of connections within the brainstem and cortical structures (Cunningham & Tucci, 2017; Ouda et al., 2015; Windle et al., 2023). This degeneration of the auditory system includes dead regions, the loss of synapses and auditory nerve fibres (Liberman, 2017; Lopez-Poveda, 2014; Vinay et al., 2020), as well as reduced neural synchrony in the auditory periphery (Kujawa & Liberman, 2009). These physiological changes impair crucial auditory functions, such as timing information, gap detection, sound localization, and spectral processing (Chen et al., 2010; Lentz et al., 2022; Profant et al., 2019; Syka, 2020; Walton, 2010). Cognitive declines associated with ageing further exacerbate difficulties in streaming abilities (Garami et al., 2020; Goossens et al., 2017). These cognitive declines include a decrease in fluid intelligence, which peaks in the 20s and subsequently declines, affecting the speed and adaptability with which individuals can solve novel problems (Fisher et al., 2019). Additionally, cognitive domains such as processing speed, inhibition, memory, and episodic memory deteriorate throughout adulthood (Park & Bischof, 2013; Salthouse, 1996) and deficits in WM and selective attention have also been observed with ageing (Gow, 2015; Henry et al., 2017; Tucker-Drob et al., 2019). As a result to these peripheral auditory and central cognitive deterioration, older adults experience declines in their ability to process temporal (Füllgrabe et al., 2014; Gordon-Salant et al., 2011; Gordon-Salant & Fitzgibbons, 1993; Souza et al., 2011; Walton, 2010), spatial (Eddins et al., 2018; Moore, 2021), and spectral cues (Moore, 1985; Regev et al., 2023; Relaño-Iborra & Dau, 2022), and their overall effects on scene analysis abilities may be cumulative (Windle et al., 2023).

Several studies report older adults to require higher discrimination thresholds to distinguish a target sound from the background, a difficulty that persists even when amplification is provided to compensate for elevated thresholds (Humes, 2007; Johannesen et al., 2016). Additionally, older adults face challenges with separating competing sound sources and orienting attention to specific auditory items in short-term memory (Alain et al., 2018). Compared to younger adults, they are also less likely to demonstrate automatic stream segregation, requiring active attention to achieve similar perceptual outcomes (Dinces & Sussman, 2017) - a likely contributor to the increased listening effort reported by many older individuals. Furthermore, older adults are more susceptible to informational masking, particularly when background noise is meaningful (e.g., intelligible speech) as opposed to meaningless (e.g., reversed or foreign speech), further emphasising central auditory processing deterioration during ageing (Larsby et al., 2008; Rossi-Katz & Arehart, 2009). Interestingly, while concurrent stream segregation appears to decline with age, sequential stream segregation seems to remain relatively intact, even in older adults with hearing loss (Reuter-Lorenz & Park, 2014).

In spite of the growing attention in recent years, research on ASA in musical context with aging remains limited, and the available findings are inconclusive. For instance, Golden et al. (2017) found no significant differences in MSA performance between patients with major

dementia syndromes and healthy age-matched controls (see also Golden et al., 2015) - even though individuals with dementia typically exhibit substantial cognitive impairments that would be expected to affect auditory perception (Hugo & Ganguli, 2014). In contrast, other studies have demonstrated that older adults tend to perform worse on tasks requiring the detection of a target instrument or vocal line within complex musical mixtures (Siedenburg et al., 2020, 2021; Benjamin & Siedenburg, 2024, 2025). However, these studies are often confounded by the presence of hearing impairments, making it difficult to isolate the specific impact of age-related effects on MSA. Given the high prevalence of age-related hearing loss, disentangling the contributions of peripheral and central aging processes remains a relevant challenge.

2.5. Others

Beyond peripheral auditory functioning, cognition, musical training, and ageing, numerous additional factors potentially contribute to individual differences in MSA. These influences add further complexity to explain the substantial interindividual variability observed. Multilingualism, for instance, has been linked to superior MINT performance (Coffey et al., 2019). Furthermore, multisensory integration can enhance ASA, particularly in speech contexts in which visual cues like lip reading have been shown to facilitate speech segregation (Helfer & Freyman, 2011). Similarly, tactile input has been found to support melody segregation in complex auditory mixtures, together indicating that cross-modal processing may also aid MSA (Slater & Marozeau, 2016).

Genetic, health and lifestyle factors might also influence MSA by moderating or even mediating any peripheral auditory and cognitive changes. For example, sex and gender factors, including their genetic, hormonal, and behavioural underpinnings, might be of importance in contributing to differences in auditory perception (Cahill, 2006). In that, differences between males and females have been documented in various cognitive domains, such as visuospatial ability, verbal fluency, memory, and perceptual speed (Kimura, 1992), as well as in speech-based ASA performance (Zündorf et al., 2011; Lewald & Hausmann, 2013). Environmental exposures associated with gender roles, such as occupational noise, may indirectly impact MSA, e.g., because females tend to experience less severe hearing loss across the lifespan (e.g., Cruickshanks et al., 1998; Hoffman et al., 2017). Nonetheless, a recent massive investigation on over 500.000 participants reported that differences are generally restricted in scope and inconsistent in direction, concluding that sex differences in music perception to be negligible overall (Bertolo et al., 2023).

Socioeconomic status further shapes perceptual and cognitive development by modulating access to musical training and exposure to complex auditory environments. That is, higher socioeconomic backgrounds often afford increased access to formal music education and diverse musical genres, thereby enhancing auditory structural knowledge and fostering stronger music perception abilities (Corrigall & Trainor, 2011; Swaminathan & Schellenberg, 2018). In addition, personality traits and individual preferences can influence the degree of engagement with music and particular genres, shaping auditory schemas and top-down

expectations that directly affect MSA (Chamorro-Premuzic & Furnham, 2007; Rentfrow & Gosling, 2003).

Cultural background and musical enculturation also play a substantial role in shaping perceptual frameworks. Higher-order processes involved in musical expectation, structure recognition, and source segregation are strongly influenced by exposure to specific musical traditions (Anglada-Tort et al., 2023; Bey & McAdams, 2002; Woods & McDermott, 2018; Jacoby et al., 2020). For example, listeners enculturated in Western tonal systems may develop different perceptual sensitivities than those familiar with microtonal or non-Western music traditions (Stevens, 2012). Such cultural exposure shapes the ability to parse pitch, rhythm, and harmonic structures, directly influencing MSA. Nevertheless, some fundamental auditory processes appear to be universal and innate (see Calcus, 2024, and Leibold, 2011, for a review). While still inconclusive overall, some studies even suggest that both Western participants and native Amazonians, despite vastly different musical experiences, tend to perceptually fuse tone combinations related by simple integer ratios (e.g., octaves, fifths), suggesting that basic acoustic constraints may underlie universal patterns in musical perception (McPherson et al., 2020; Mehr et al., 2019).

Taken together, these additional influences, including genetic predispositions, health status, cultural environment, personality, and developmental factors, underscore the complexity of individual variability in MSA. Given the breadth of potential contributing factors, this dissertation focuses on a selected subset of auditory, cognitive, and experiential variables. A more comprehensive investigation of the wider set of influences remains an important task for future research seeking to develop a deeper account of auditory scene processing.

A SURVEY ON HEARING HEALTH OF MUSICIANS IN PROFESSIONAL AND AMATEUR ORCHESTRAS

Submitted to *Trends in Hearing*^c on 28. February 2024

First published online 9. December 2024

by

Robin Hake¹, Gunter Kreutz², Ulrike Frischen³, Merle Schlender¹, Esther Rois-Merz⁴, Markus Meis⁵, Kirsten C. Wagener⁶, Kai Siedenburg^{1, 7}

¹ Dept. of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Germany

² Institute of Music, Carl von Ossietzky Universität Oldenburg, Germany

³ University of Bremen, Germany

⁴ Audienz - Musikalische Hörberatung e.U., Austria

⁵ Cochlear Deutschland GmbH Co. KG, Germany

⁶ Hörzentrum Oldenburg gGmbH, Germany

⁷ Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

^c The original is accessible via: <https://doi.org/10.1177/23312165241293762>

Disclaimer

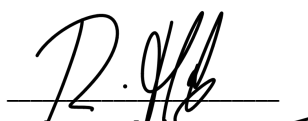
The content of this chapter is identical to that of the original publication.

The citation for the following publication is:

Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburger, K. (2024). A Survey on Hearing Health of Musicians in Professional and Amateur Orchestras. *Trends in Hearing*, 28, 1–20. <https://doi.org/10.1177/23312165241293762>

Author contributions

RH, GH, UF, MS, ER, MM, KW, KS designed the study. RH and KS analysed the data. RH wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.



Robin Hake



Kai Siedenburger

3.1. Abstract

Hearing health, a cornerstone for musical performance and appreciation, often stands at odds with the unique acoustical challenges that musicians face. Utilizing a cross-sectional design, this survey-based study presents an in-depth examination of self-rated hearing health and its contributing factors in 370 professional and 401 amateur musicians recruited from German-speaking orchestras. To probe the nuanced differences between these groups, a balanced subsample of 200 professionals and 200 amateurs was curated, matched based on age, gender, and instrument family. The findings revealed that two-thirds of respondents reported hearing-related issues, prevalent in both professional and amateur musicians and affecting music-related activities as well as social interactions. The comparative analysis indicates that professionals experienced nearly four times more lifetime music noise exposure compared to amateurs and faced more hearing challenges in social contexts, but not in musical settings. Professionals exhibited greater awareness about hearing health and were more proactive in using hearing protection devices compared to their amateur counterparts. Notably, only 9% of professional musicians' playing hours and a mere 1% of amateurs' playing hours were fully protected. However, with respect to their attitudes towards hearing aids, professional musicians exhibited a noticeable aversion. In general, an increase in music-related problems (alongside hearing difficulties in daily life) was associated with a decrease in mental health-related quality of life. This research highlights the importance of proactive hearing health measures among both professional and amateur musicians and underscores the need for targeted interventions that address musicians' specific hearing health challenges and stigmatisation concerns about hearing aids.

3.2. Introduction

Playing musical instruments is associated with long-term health benefits (see review by Dingle et al., 2021). Yet, music-making, whether pursued at an amateur or professional level, may also pose inherent risks, including the potential for developing noise-induced hearing impairments (Schink et al., 2014; Musgrave, 2022). The likelihood of developing such impairments is influenced by the duration and intensity of noise exposure: both short but intense noise levels of 137 dB_{peak} C, as well as comparatively low but prolonged exposure at an average of 85 dB A over several years, have been reported to result in permanent noise-induced hearing loss (Liedtke, 2010; 'Königsteiner' and 'NIOSH' recommendations; Deutsche Gesetzliche Unfallversicherung, 2013, 2020; Chan, 1998). Indeed, musicians regularly confront sound intensity levels that not only challenge but also endanger their hearing (O'Brien et al., 2013). Depending on their instrument, position, and repertoire, they often face peak sound levels of up to 135 dB C and practice durations extending to six hours daily (Rodrigues et al., 2014; Halevi-Katz et al., 2015; Schmidt et al., 2011). Consequently, bound to their profession, musicians are inherently at risk for developing noise-induced hearing impairments (Gong et al., 2018; Schink et al., 2014).

A body of research highlights the risks musicians face in terms of hearing health. Phillips & Mace (2008) report that 48% of music students are exposed to harmful noise levels during daily practice. Similarly, O'Brien et al. (2013) found that 53% of musicians surpass the permissible daily noise exposure limits (see also Tufts & Skoe, 2018), a salient finding, as it does not include time spent in orchestral rehearsals and concert performances (i.e., only solitary practice was considered). It is therefore not surprising that Halevi-Katz et al. (2015) linked greater musical experience (years of persistent music playing) to higher hearing thresholds in the frequency range of 3-6 kHz. Furthermore, Hasson et al. (2009) reported that 40% of musicians experience some form of hearing disorder, with tinnitus (19%) being the most common, followed by 14% suffering from hyperacusis (one form of auditory hypersensitivity), and 6% experiencing difficulties in noisy environments or sound distortions - a finding that is consistent with Harper (2002), who recorded hearing problems among 38% of British and 37% of German musicians. Notably, a systematic review by Di Stadio et al. (2018) revealed that professional musicians in pop/rock genres are at an even higher risk of developing hearing loss and other audiological symptoms compared to classical musicians. Gembris et al. (2018) further indicated that the likelihood of hearing problems of orchestra musicians escalates with age. The impact of hearing loss extends beyond listening abilities, affecting mental health and social well-being. Bigelow et al. (2020) demonstrated that hearing loss correlates with increased psychological distress, greater use of antidepressants, and higher dependency on mental health services among the general population.

However, the relationship between noise exposure and hearing impairment is not always straightforward. For example, Prendergast et al. (2017) had difficulties relating noise exposure to perceptual deficits in young listeners (mostly non-musicians) with normal hearing audiometry. Couth et al. (2020) investigated this in a sample of musicians but could only find sub-clinical effects of noise exposure on hearing. Moreover, Russo et al. (2013)

found no significant differences in hearing issues between instrument groups, even though brass, woodwinds, and percussionists have been reported to be exposed to particularly high sound levels (e.g., Rodrigues et al., 2014; Schmidt et al., 2011).

In summary, musicians commonly face hearing health challenges such as hearing loss and tinnitus, impacting their musical performance and overall quality of life (Halevi-Katz et al., 2015; Hébert & Lupien, 2009; Richter, 2011). Furthermore, hearing health may directly and indirectly affect the ability to perceive, appreciate, and produce music, alongside impacting mental, physical, and social well-being (e.g., Gembris et al., 2018; Musgrave, 2022, Scherer, 2004; Zentner et al., 2008, Bigelow et al., 2020).

Recognizing the risks to hearing health in musicians, it becomes crucial to focus on their habits and preventive strategies. Certainly, adopting behaviours that reduce exposure to loud noise is essential for preventing hearing problems (e.g., Hébert & Lupien, 2009; Olsen et al., 2016; Richter, 2011). Intervention strategies for public health thus often include the promotion of hearing protection (e.g., Occupational Safety and Health Administration, 2002). Still, the actual usage of hearing protection among musicians and the barriers to its consistent application remain key concerns. Laitinen (2005), for example, found that among classical musicians in Finland, the use of hearing protection is sporadic, often commencing only after the onset of auditory symptoms. Specifically, a mere 6% consistently use hearing protection, while 35% seldom do - despite the fact that 31% reported some form of hearing loss and 37% reported experiencing tinnitus. However, these are not uniformly distributed across regions. Among Australian musicians, 40% reported using hearing protection during practice, and 64% claimed occasional earplug use (O'Brien et al., 2013, 2014). This contrasts with orchestras in Germany, where only 38% reported occasional hearing protection use (Zander et al., 2008). Interestingly, there is some indication that younger musicians are more inclined towards protective measures: a study by Couth et al. (2021) found that among 80 British early-career musicians, 77% used hearing protection at least once a week.

The inconsistent use of hearing protection among musicians, despite the well-documented risks, demands an examination of the motivations behind these unhealthy choices. This exploration must consider the significant role of awareness in shaping health-related behaviour. While awareness is often a precursor to behavioural change (e.g., Ajzen, 1991), one might assume that greater awareness of hearing health risks among musicians would naturally lead to increased use of hearing protection. For instance, studies by Keppler et al. (2015a, 2015b, 2015c) suggest that increasing awareness about hearing health substantially contributes to the adoption of protective measures against recreational noise. However, this inference is complicated by findings from Laitinen (2005), where a very high percentage (94%) of musicians reported concern for their auditory health, yet the actual use of hearing protection was notably low (6% consistent usage). This disparity suggests the presence of other barriers within the musical context, which may be intrinsic to the professional demands and social dynamics of musical performance. For example, musicians frequently cite issues with sound quality and difficulties hearing other musicians as significant obstacles to consistent hearing protection use (Laitinen, 2005; Zander et al., 2008; O'Brien et al., 2014). Additional factors such as varying loudness levels during performances or hampered self-monitoring could further discourage musicians from using hearing

protection. Moreover, there is a less explored but equally compelling aspect that may hinder the consistent use of hearing protection among musicians - concerns about one's professional image. The use of hearing protection could potentially be perceived as a sign of auditory weakness, thereby affecting a musician's reputation or audience perception (see Couth et al., 2021).

A similar dynamic may also contribute to the reluctance toward hearing aid adoption. Despite their proven utility in enhancing auditory experiences, hearing aids remain underutilised (e.g., Chern et al., 2023). While the literature does not specifically address the potential impact on musicians' reputation as a factor in hearing aid adoption, musicians repeatedly complain that devices often are not effectively equipped or fitted to account for the full dynamics essential in music (e.g., Madsen & Moore, 2014). Recently, Greasley et al. (2020) extensively explored the complex dynamics of hearing aid usage among musicians by considering both audiologists' and patients' perspectives in the UK. Their study revealed another factor: a significant misalignment in expectations and understanding between these groups. Among musicians who use hearing aids, only 42% report an enhanced enjoyment of music while wearing them, with half experiencing distortions and over two-thirds encountering problems occasionally. A primary concern is the gap in audiologists' training and understanding regarding the specific demands of hearing aids in musical environments. Only a small proportion (37%) of audiologists have received training focused on music, and despite 74% seeking additional information on this subject, many still express only moderate confidence in advising patients about music and hearing aids. Indeed, 58% of patients reported that music listening was never discussed. Consequently, many musicians feel their unique needs for high-quality music listening are not being met, leading to dissatisfaction and low adoption rates. These insights underscore a need for more specialised knowledge and tailor-made solutions in hearing aids for musicians. While awareness is a critical factor, it does not operate in isolation. The limited adoption of hearing protection and hearing aids is a complex interplay of personal, professional, and societal factors, warranting a multifaceted approach to intervention.

In sum, the literature suggests a link between musical noise and an increased risk of hearing problems in musicians, yet generalizing these findings across different groups and levels of exposure remains challenging. Existing studies predominantly focus on professional musicians, disregarding the large proportion of recreational (amateur) musicians - which amount to 14.3 million musicians in Germany alone (Deutsches Musikinformationszentrum, 2021). This necessitates a more detailed investigation into how music-related noise exposure affects musicians, particularly considering the varied intensity levels associated with different instrument groups, the range of environments in which music is performed (from individual practice at home to full orchestral performances), and the frequency and duration of musical activities (that is, considering both professional and amateur musicians). Despite this, the literature provides limited insight into the specific requirements necessary to implement preventive measures effectively. This gap extends to understanding how awareness and concerns about reputation influence musicians' decisions, particularly regarding the use of hearing protection and the adoption of hearing aids.

3.2.1. The current study

Addressing this gap, the current study aimed to provide a comprehensive investigation of hearing health among orchestral musicians, both professional and amateur. This included examining the utilization of and attitudes towards hearing protection and hearing aids, the prevalence and types of hearing problems, and their overall impact on musicians' well-being. While the analysis was exploratory in nature, it was underpinned by two primary hypotheses: Professional musicians were expected to a) report more hearing problems and b) demonstrate a higher level of awareness towards hearing health compared to their amateur counterparts. Additionally, this research aimed to delve into the usage, experiences, and effectiveness of various strategies employed to protect hearing health. An important facet of this exploration was to identify and analyse the key factors that influenced the adoption rates of hearing protection and hearing aids among musicians. Subsidiary to these main goals, the study also explored the influence of demographic factors, specifically age as a critical factor in hearing loss, on these relationships.

3.3. Methods

3.3.1. Participants

In total, 582 amateur musicians and 646 professional musicians initiated the survey. The recruitment strategy for this cross-sectional survey was multifaceted, targeting not only orchestral musicians but also choral musicians, conductors, and the orchestral management. However, this article focuses exclusively on orchestral musicians. Accordingly, non-orchestral musicians, participants who completed less than 50% of the survey, and those who took less than 7 minutes to complete the survey (which was identified as a minimal duration for any serious response strategy) were excluded from the analysis ($n = 617$). To further qualify musicianship among the amateur sample, a criterion based on the question '*How many years have you practised an instrument regularly and daily?*' was applied, using a cut-off of six years for regular practice (see Zhang & Schubert, 2019). This resulted in the exclusion of additional 16 participants.

3.3.1.1. Final sample

After implementing the aforementioned exclusion criteria, the final sample comprised 282 professional musicians ($\text{Age}_{\text{mean}} = 49.2$, $\text{SD} = 11.3$, range = 25-65 years; 130 females, one non-binary, 12 not specified) and 313 amateurs ($\text{Age}_{\text{mean}} = 49.5$, $\text{SD} = 18.8$, range = 25-85 years; 184 females, 7 not specified). For a more nuanced description of the level of musical sophistication among the amateur musicians, see Figure A3.1 in Supplementary Materials (hereafter referred to as 'Supp. Mat.').

3.4. Procedure

The survey was administered from February 2022 to February 2023. Recruitment for this cross-sectional survey was conducted in collaboration with the German Orchestra Association through an internal newsletter, complemented by email invitations distributed to amateur

orchestras within Germany and musicians in Austria. The questionnaire itself was provided by an online web application for creating online surveys called '*SoSci Survey*' (v.3.2.44; Leiner, 2019). Participants were explicitly asked to choose a survey version depending on whether they play music on an amateur or professional level. To accommodate linguistic diversity in orchestras, the questionnaire was available in both English and German, allowing for language selection at any point during the survey. Prior to commencing the survey, participants provided informed consent, acknowledging that participation was both anonymous and voluntary. Notably, no monetary or other incentives were offered for participation. All participants were instructed to complete the survey in their own time, and it was explicitly stated that participants could selectively skip questions if desired. On average, the survey took about 20 minutes to complete. This study was approved by the Carl von Ossietzky University Oldenburg ethics review board (Drs.EK/2021/114).

3.5. Materials

Data were collected on demographic variables and musical background details, including age (categorised in ten-year intervals from “up to 30 years” to “80+ years”), gender, primary (contract) instrument, orchestra category, and repertoire. To assess hearing loss, participants were asked to report their latest audiological reports if available. Those without prior audiological assessment completed an adapted version of the Hearing Aid for Music (HAfM) questionnaire (Greasley, 2022; see Table A1 and Table A2 in Supp. Mat.). Hearing loss was then categorised using the classification system of the World Health Organization (2001) and was then adapted from Martini (1996) to ensure the inclusion of milder forms of hearing loss. This adaptation provides a more sensitive measure of hearing loss, which is particularly important for assessing musicians who might experience subtle, yet impactful, changes in their hearing (see also Humes, 2019). Accordingly, in this study, the following adapted classification was used: mild hearing loss: 20-40 dB HL (i.e., adapted from Martini, 1996), moderate: 41-60 dB HL (WHO), severe: 61-80 dB HL (WHO), profound: > 81 dB HL (WHO).

Additional items targeted hearing protection usage. Specifically, participants were asked to report the number of hours they spend per week practising alone, rehearsing in groups, and performing at concerts – while also including measures on the frequency and type of hearing protection used during these activities. To further assess hearing health-related issues, the questionnaire incorporated several metrics: scales for estimating musical sound levels, measuring musical sophistication, evaluating hearing-related quality of life, scales to identify music-related and daily life hearing challenges, and a hearing aid disapproval scale: **Sound level exposure (UNE)**. The Noise Exposure Structured Interview (NESI; Guest et al., 2018) was adapted into a survey to measure musicians' yearly noise exposure units (UNE). UNE represents a typical yearly exposure to 90 dB A noise over 2080 hours (52 weeks of 8-hour workdays). Accordingly, if a musician accrues one UNE over the period of one year, it is analogous to, for example, a construction worker, being exposed to 90 dB A work-related noise for 52 weeks, 8 hours each working day.

The UNE was calculated using the following equation:

$$UNE = \frac{W * 44_{(weeks)}}{2080_{(hours)}} * \left(P * 10^{\frac{L-A-90_{(dB)}}{10}} + (1 - P) * 10^{\frac{L-90_{(dB)}}{10}} \right)$$

where, W is hours of instrumental noise exposure per week, 2080 corresponds to the number of hours in a working year, P is proportion of time that hearing protection was worn (0 to 1), L denotes the noise level in dB HL, and A representing the level of attenuation provided by hearing protection in dB.

Hours of exposure were based on self-reported averages for practice, group rehearsals, and concert performances. Noise levels (L) were estimated from literature for different instruments and settings (see Table A3-A4 in Supp. Mat.). Hearing protection usage frequency was converted to probability estimates (e.g., "*always*" = 100%). Attenuation values (A) were based on data from Guest et al. (2018; see Table A5-A6 in Supp. Mat.). Lifetime music noise exposure (L-UNE) was derived by multiplying years of exposure by UNE and was log-transformed for analysis. Years of exposure for amateur musicians were calculated based on the reported years of playing an instrument. Notably, the exact years of exposure for professional musicians was not available. Instead, years of exposure were projected by assuming professional musicians would commence their musical education no later than the age of 10 years and begin to perform within orchestra rehearsals and concerts no later than the age of 20 years. This approach likely disregards a significant portion of years of instrumental noise exposure (average starting age of professional musicians have been reported to lie between 6-8 years, see Wesseldijk et al., 2020), yet also implements a rather conservative strategy in the estimation of lifetime music sound exposure. As participants reported both the type of hearing protection used and their usage frequency in various musical settings, UNE was calculated for individual practice, rehearsals and concerts separately and was then combined into a final aggregated score.

Hearing-Dependent Daily Activities Scale (HDDA; Hidalgo et al., 2008). The HDDA is a 12-item scale that assesses the impact of hearing loss on daily life. This scale includes subscales for the ability to perceive basic sounds in everyday life and a subscale assessing the impact of hearing loss on social interactions. Participants responded to questions regarding their hearing abilities in daily life on a 3-point Likert scale (with the options "*No, I can't*"; "*With some difficulty*"; "*Yes, without difficulty*"). The HDDA subscore for basic sound perception yields a final score ranging from 0 to 8, and social interactions spanning from 0 to 16. In both instances, a score of 0 denotes severely impaired hearing abilities. The HDDA is typically used for rapid hearing loss screening, employing a cut-off score of >18 to indicate normal hearing (Jacob et al., 2017). Here it is also used as a variable to estimate hearing difficulties in daily life settings.

Goldsmith Musical Sophistication Index (GMSI; Müllensiefen et al., 2014). Amateur musicians were asked to complete the short version of the Goldsmith Musical Sophistication Index. The short GMSI is a brief, 18-item self-report questionnaire that assesses several aspects of musical expertise. It was designed to capture respondents' overall level of musical sophistication by using subscales for active engagement, emotions, musical training,

perceptual abilities, and singing abilities. Participants were asked to respond to a 7-Point-Likert-Scale (1 = "*Completely Disagree*" to 7 = "*Completely Agree*"). The final composite score ranges from 18 to 126, with higher scores indicating higher level of general musical sophistication.

Health-related quality of life (VR-12; Buchholz et al., 2017). Participant health-related quality of life was assessed through the application of the Veterans RAND 12 Item Health survey. This tool is widely recognized and validated for use across various demographic groups. The VR-12 questionnaire invites responses to statements such as "*How would you describe your state of health in general*", which participants rate on a five-point Likert scale (1 signifying "*poor*" and 5 denoting "*excellent*"). The survey culminates in two distinct subscales, the VR12-PCS and VR12-MCS, each of which are dedicated to the evaluation of physical and mental health, respectively. Both subscales are scored within a range of 0 to 100, with higher scores indicative of a higher reported quality of life.

Music-related Hearing Problems Scale (MRHP). Participants rated six items (e.g., "*I have the impression that the sound is washed out/ blurred.*") on a 5-Point-Likert-Scale, ranging from 1 = "*always*" to 5 = "*never*". The latent one-factor structure was confirmed through Confirmatory Factor Analysis (CFA), showing an internal consistency of $\alpha = .82$ and $R^2 = .44$. Final scores on the scale span from 0 to 10, where higher values are indicative of more severe music-related hearing problems (see Table A7 in Supp. Mat. for details).

Hearing Health Awareness Scale (HHA). The HHA scale is composed of seven items (e.g., "*Hearing health is extremely important to me.*" on a 4-Point-Likert-Scale; 1 = "*true / applies*" to 4 = "*not true / does not apply*". The scale encompasses aspects such as the perceived strength of sound level exposure while playing music ($n = 3$), general attitudes towards hearing health ($n = 3$), and the attitude about wearing hearing protection while performing music ($n = 1$). The CFA achieved a good model fit with a R^2 value of .33 and a Cronbach's alpha of .74. The final scores span from 0 to 10, with higher values being indicative of higher hearing health awareness (see Table A8 in Supp. Mat. for details).

Hearing Aids Disapproval Scale (HAD). The HAD utilised five items (e.g., "*It irritates me, when colleagues with hearing impairments wear hearing aids.*"), encompassing aspects such as underlying apprehensions, societal perceptions, and potential stigma associated with the use of hearing aids among musicians (4-point Likert scale, 1 = "*true / applies*" to 4 = "*not true / does not apply*"). The CFA yields an internal consistency of $\alpha = .79$ with an $R^2 = .44$. The scores derived from this scale range from 0 to 10, where higher values indicate a greater disapproval towards hearing aids (see Table A9 in Supp. Mat. for details).

3.6. Analysis

All statistical analyses were conducted using R statistical software (version 2022.07.2+576; RStudio, 2020). Given the non-normally distributed nature of the data set, the Wilcoxon rank-sum test (also known as the Mann-Whitney U test, abbreviated as $W[\text{sample } n]$) was utilised for comparing two groups. For comparisons involving more than two groups, the Kruskal-Wallis test was employed, abbreviated as $H[\text{degree of freedom } 1, \text{degree of freedom } 2]$. Post-hoc comparisons involved pairwise Wilcoxon rank-sum tests with a Bonferroni correction to

control Type I error rates. Effect sizes and confidence intervals, in addition to p-values, were reported to provide a more comprehensive understanding of the results. The rank-biserial correlation (r_{rb} ; see Kerby, 2014) was used as the measure of effect size for the Wilcoxon tests, with r_{rb} values ranging from 0 to |1|. Values approaching 0 indicate a negligible effect, while an r_{rb} greater than .1 suggests a small effect. An r_{rb} exceeding .3 is indicative of a medium effect, and those above .5 signify a large effect. To supplement the Kruskal-Wallis test results, the Epsilon square measure (ϵ^2) was reported to quantify effect size. Epsilon square values also range from 0 to 1, with higher values indicating stronger effects. The median (Mdn) and the Median Absolute Deviation (MAD) were reported to address the non-parametric nature of the data. MAD , as a measure of statistical dispersion, indicates variability, with higher MAD values suggesting greater variability. Spearman's rank-order correlation was used to assess relationships between continuous and ordinal data. The Chi-squared test was applied for categorical variables. In cases of missing data, listwise deletion was employed, leading to varying sample sizes for each analysis. The investigation centers on a global perspective on the data by mainly focusing on the results of the developed scales which summarise multiple items of the survey. To establish reliable scales for three latent constructs, Confirmatory Factor Analysis (CFA) was applied via the *lavaan* package (v06-15; Rosseel, 2012) in R (RStudio, 2020), utilising all the data prior to any exclusions. The internal consistency of the scale was assessed by calculating Cronbach's alpha, and various model fit indices and percentage of explained variances were reported to ensure the robustness of the established scales. For all three scales, Weighted Least Squares Mean and Variance Adjusted (WLSMV) estimator was used. Additionally, the interested reader may access the detailed data of every individual item in the corresponding open access dataset (see Hake et al., 2024).

3.7. Results

3.7.1. Musician group matching

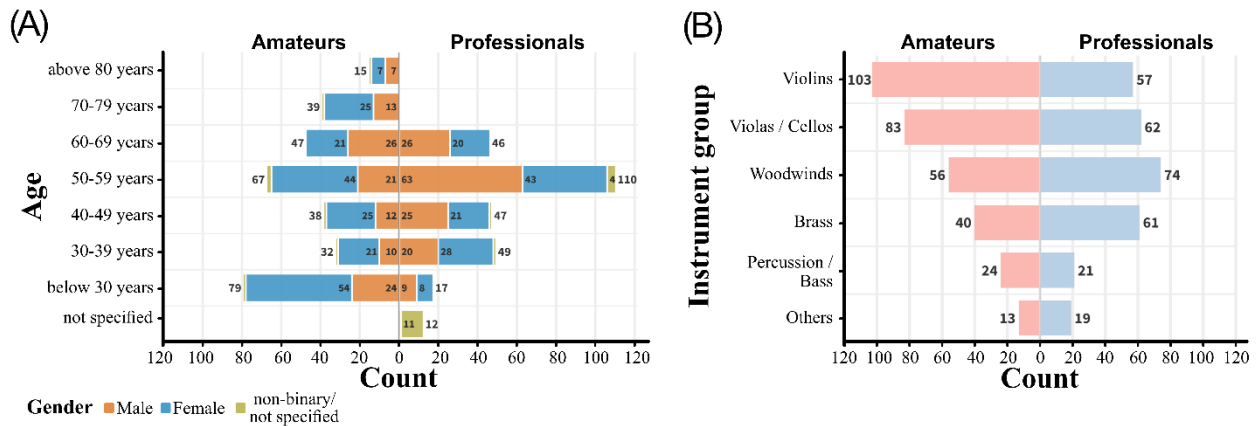
Unmatched groups in comparative analyses can introduce significant bias due to confounding variables, thereby undermining the internal validity of the analysis. To probe for preexisting differences between professional and amateur musicians, a Chi-square test of independence was conducted, showing a significant association between group membership (amateur vs. professional) for age $\chi^2(7) = 133.6, p < .001$. Upon further inspection it became apparent that comparably few young (up to 30 years) and senior professional musicians (70 years and above) participated in the survey compared to the amateurs (see Fig. A3.1). Similarly, there were significant differences between group membership for gender ($\chi^2(3) = 8.3, p = .004$) and for participants instrument family, $\chi^2(5) = 23.5, p < .001$ (see Table A3.3 in Supp. Mat. for the grouping of instruments). Thus, participants' musical group membership (professionals vs. amateurs) demonstrated imbalance across age, gender, and instruments family.

In order to mitigate confounding effects, a matching algorithm was applied, utilising the *MatchIt* package (v4.5.3; Ho et al., 2007). This package serves as a statistical tool designed to ensure comparability between professional musicians and amateurs by balancing the participant groups on several key demographics: age, gender, and instrument family of their primary (contract) instrument. A combination of the 'optimal' and 'exact' matching

procedures was applied. The most critical aspect of this approach was the exact matching on age, since it significantly influences hearing. This method ensured that each pair of participants – comprising one professional and one amateur musician - was matched 'exactly' within the same age group. For the variables of gender and instrument family, the 'optimal' matching method was selected, coupled with the 'euclidean' distance metric. Rather than matching each pair precisely, this strategy aimed to minimise the overall distance across pairs in these two dimensions. Essentially, the result is a subsample in which each professional musician was balanced strictly in terms of age, with a more flexible distribution for gender and instrument family. This method allowed for a more nuanced matching, taking into account the combined effects of age, gender, and instrument family, while ensuring that a sufficient number of participants are retained for subsequent analysis.

After matching, 82 unmatched subjects from the professional musician group, and 113 from the amateur group were excluded. Accordingly, a total of 400 participants remained for the analysis. The distributions for age, gender, and instrument group before and after matching are represented in Figure 3.1. In conclusion, the matching process successfully balanced the key variables of age, gender, and instrument family between the professional and amateur musicians. That is, there are no significant differences between groups neither in age ($\chi^2(4) = 0, p = 1$), gender ($\chi^2(3) = 1.3, p = .73$), and instrument group ($\chi^2(5) = 1.27, p = .94$) after matching. It is critical to note that despite the improvements in balance across these variables, other imbalances remained post-matching, e.g., in repertoire played by the orchestra (see Fig. A3.2 in Supp. Mat.). The demographic distribution of the final sample closely mirrors that of professional musicians in Germany, with deviations from the reference populations remaining within a 5% proportional margin. Thus, the sample presented here can be regarded as representative (see Versorgungsanstalt der deutschen Kultuorchester, 2021).

Pre-matching



Post-matching

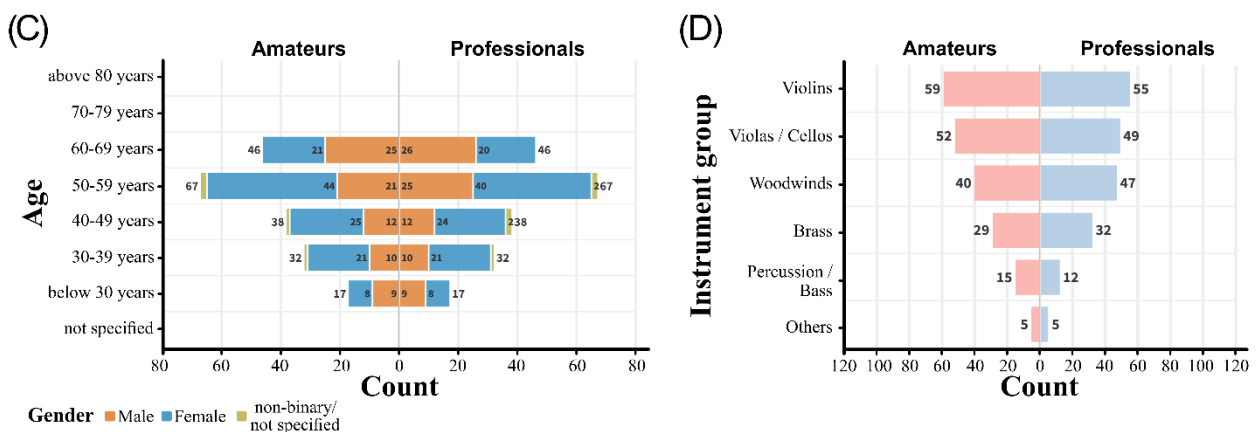


Figure 3.1 Demographic profiles of amateur (pale red) and professional musicians (light blue) as a function of age, gender (A, C), and instrument group distributions (B, D) before (upper row) and after the matching procedure (lower row).

3.7.2. Hearing-related problems

Participants who had received professional audiometric evaluations were instructed to report the final results of those assessments, specifically indicating whether they had been diagnosed with mild, moderate, severe, or profound hearing loss (HL). For participants who had not undergone professional audiometric assessments in the past and were thus unaware of their specific degree of hearing loss, they were instructed to self-estimate their hearing loss. Almost all professional musicians stated to have their hearing checked (97%), whereas only 72% of amateurs had their hearing checked by a professional at least once before. On average the last appointment was 2.7 years ago ($SD = 3.8$) for professionals and 5.2 years ($SD = 6$) for amateurs.

Both self-estimated and self-reported results of professionally diagnosed hearing loss were included in the analysis. Overall, 22% of professionals ($n = 44$) and 23% of amateurs ($n = 46$) reported having at least a mild form of hearing loss (see Fig. 3.2). In addition, 24% of musicians (23.3% amateurs, 24.7% professionals) reported having tinnitus, 49.1% (44.8% amateurs, 53.3% professionals) reported being oversensitive to high sound levels, and 15.9% (11.9% amateurs, 18% professionals) reported other hearing problems. There were no significant differences of reported hearing problems across groups (professionals vs. amateurs) for the degree of hearing loss ($\chi^2(1) = 4.1, p = .26$), tinnitus ($\chi^2(1) = 1.84, p = .22$),

oversensitivity ($\chi^2(1) = 2.87, p = .1$), or reporting other problems ($\chi^2(1) = 2.8, p = .12$). Taken together, two-thirds (66.75%) of the respondents reported experiencing at least one of these kinds of hearing problems - a pattern that is indifferent to the musician's group membership, with 67.5% of amateurs and 66% of professionals affected. Similarly, employing the hearing-dependent daily activities (HDDA) hearing loss screening cut-off (<18 indicating mild impairment; Jacob et al., 2017), a total of 107 individuals (48 amateurs, 59 professionals) were classified as hearing impaired. A chi-squared test found no significant difference in this classification between the groups ($\chi^2(1, 385) = 1.37, p = .242$), further suggesting that the proportion of individuals with hearing difficulties may be similar across both groups.

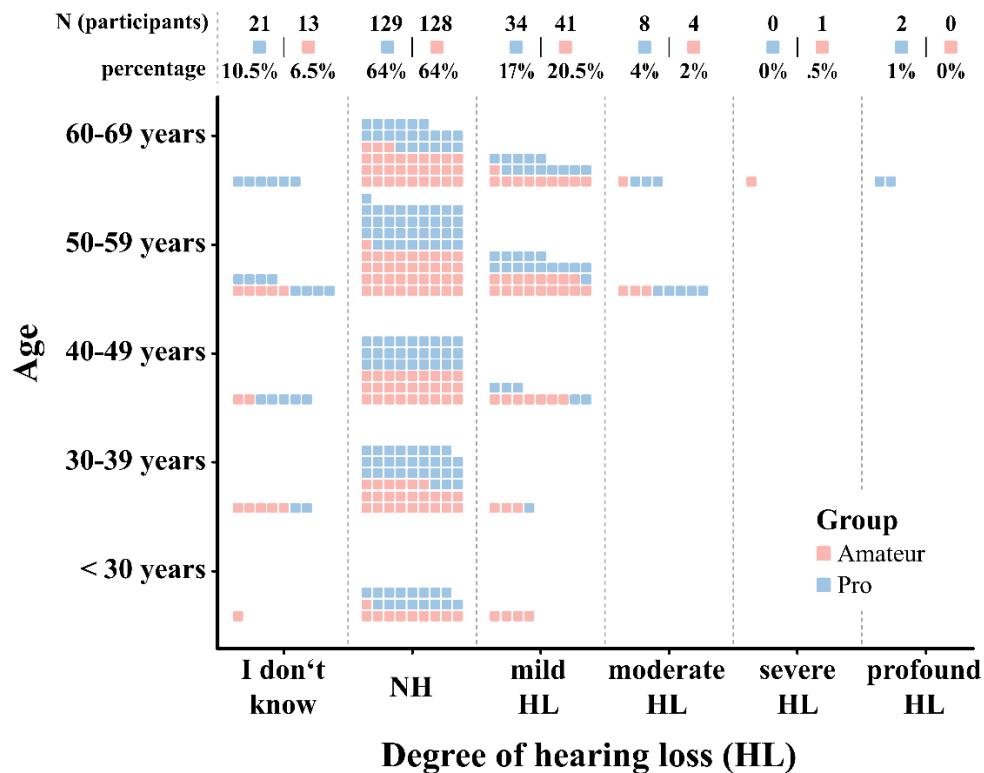


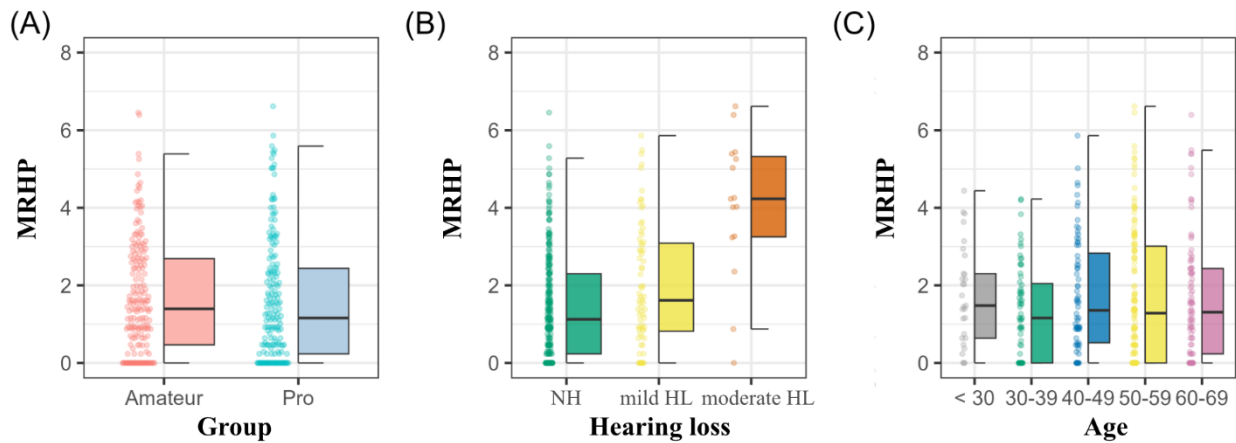
Figure 3.2 Distribution of reported hearing loss (HL) among amateurs (pale red) and professional musicians (light blue) by age group. Both the self-reported audiological diagnoses and self-estimated hearing loss are considered. Accordingly, 5 responses from professionals and 56 from amateurs are based on self-estimated hearing loss rather than self-reported diagnosis from professional hearing screening. Note: NA values have been omitted, so percentages do not accumulate to 100%.

Furthermore, the analysis did not suggest any significant difference in the music-related hearing problems scale (MRHP) between the groups of musicians ($Mdn_{diff} = .24; W = 18554, p = .3, r_{rb}$ (rank biserial) = .06, 95% CI[-.06, .18]; see Fig. 3.3). Differences were observed, however, in the hearing-dependent daily activities (HDDA) scale between the musician groups. Professional musicians reported more problems in social interaction ($Mdn = 12, MAD = 4.45$), compared to amateur musicians ($Mdn = 13, MAD = 2.97$). Lower scores indicate more severe hearing problems. The Wilcoxon Rank-Sum Test revealed this difference to be significant, $W(391) = 21519, p = .03$. The effect size was $r_{rb} = .13$, indicating a small effect. There were no significant differences between the groups in the HDDA subscale for basic sound perception ($W(391) = 14300, p = .59$).

Influence of hearing loss and age. When conducted irrespective of musicians' group membership (also excluding 'I don't know' responses), a Kruskal-Wallis test indicated significant differences in MRHP scores across the participants' degree in hearing-impairment ($H(2, 374) = 27.4, p < .001$), with a subtle effect size ($\epsilon^2 = .07$; see Fig. 3.3). The median MRHP factor score was 1.14 ($MAD = 1.7$) for musicians with 'normal hearing' (NH), 1.61 ($MAD = 1.8$) for the 'mild hearing loss' (mild HL), and 4.23 ($MAD = 1.5$) for the 'at least moderate hearing loss' (moderate HL). A post-hoc pairwise Wilcoxon test revealed that the differences were significant between the 'NH' and the 'mild HL' ($p = .02$), between the 'NH' and the 'at least moderate HL' ($p < .001$), as well as between the 'mild HL' and the 'at least moderate HL' group ($p = .001$). Likewise, HDDA scores showed significant effects with the reported degree of HL. The Kruskal-Wallis test revealed significant variance across the HL types ($H(2, 391) = 100.4, p < .001$), with a moderate effect size ($\epsilon^2 = .27$). The Wilcoxon post-hoc test confirmed significant differences between all groups, i.e., 'NH' ($Mdn = 14, MAD = 3$) and both 'the mild HL' ($Mdn = 9, MAD = 3$) and 'at least moderate HL' ($Mdn = 6, MAD = 1.5$), and also between the latter two (all comparisons reach $p < .001$).

No significant differences were found for participants' MRHP scores between age groups ($H(4, 359) = 3.23, p = .52$). In contrast, significant age-related differences were observed in the HDDA subscore for social interactions ($H(4, 400) = 23.2, p < .001$), although the overall effect size was marginal ($\epsilon^2 = .06$). Post-hoc analysis revealed significant differences between the 'below <30' ($Mdn = 14, MAD = 3$) and '50-59' age groups ($Mdn = 11, MAD = 3; p = .03$), as well as between the '30-39' ($Mdn = 14, MAD = 3$) and both the '50-59' ($p = .001$) and '60-69' ($Mdn = 12, MAD = 4.5, p = .007$) age groups. No other age group comparisons yielded statistically significant results.

Music-related hearing problems (MRHP)



Hearing-dependent daily activities (HDDA) in social situations

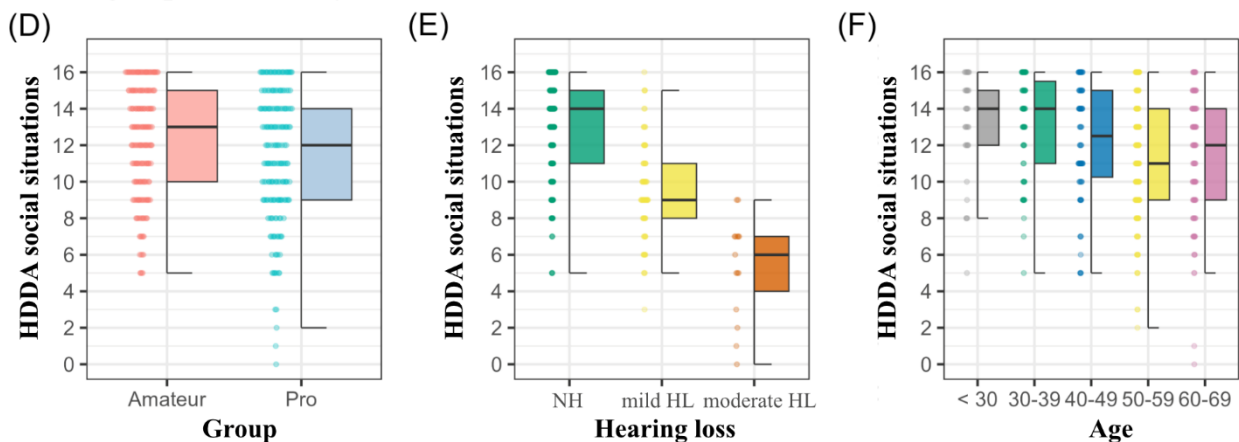


Figure 3.3 Amateur (pale red) and professional musicians (light blue) scores for Music-Related Hearing Problems (MRHP; higher scores equals more severe problems) and Hearing-Dependent Daily Activities (HDDA; lower scores equals more severe problems). Panel A and D compare MRHP and HDDA in social situations scores between amateur and professional musicians. Panels B and E categorise MRHP and HDDA in social interactions by hearing loss severity (normal hearing, mild hearing loss, at least moderate hearing loss). Panels C and F illustrate MRHP and HDDA in social interactions across age cohorts. Here and in the following, boxplots represent the median and interquartile ranges of scores on the right side and individual responses on the left and NA values are excluded.

3.7.3. Use of hearing protection

Overall, a majority of participants (59.4%, $n = 240$) acknowledged that playing music poses a risk to their hearing health due to the exposure to high sound levels; however, merely 52% of those who recognized the risk chose to wear any hearing protection. More strikingly, over 35% of respondents who affirmed the statement (respond with 'true') - "The sound levels in the orchestra are so high at times that I should wear ear protection" - and 77% of those who concurred with 'rather true', abstained from using hearing protection all together. In group comparisons, professionals exhibit a markedly higher tendency to use hearing protection compared to their amateur counterparts (53.5% of professionals vs. 12% of amateurs). However, the majority of musicians reported rarely or never using hearing protection, with percentages ranging from 72.5% to 82.4% for amateurs and from 14.5% to 70.6% for professionals, depending on the playing context (Detailed statistics are provided in Table 3.1).

Notably, only 4% of professional musicians reported always using hearing protection during individual practice, compared to .5% of amateurs. This discrepancy widens during rehearsals, with 9.9% of professionals vs. 0% of amateurs always using hearing protection. The trend persists during concerts, where 10.4% of professionals and only 1.5% of amateurs consistently use hearing protection. By combining reported hearing protection usage with individuals' total instrumental playing time (similar to the method used for estimating the probability of attenuation in the UNE), it was found that amateurs spent 79.5% of their playing time unprotected, 19.5% partially protected (e.g., using protection in one ear or occasionally removing it), and 1% fully protected. In contrast, professional musicians spent 34% of their playing time unprotected, 57% partially protected, and 9% fully protected.

When examining the consistent use of hearing protection in conjunction with the degree of hearing loss, a significant trend was observed ($H(2,364) = 13.25, p = .001$). Specifically, analysed irrespective of musicians group membership, 3% ($SD = .15$) of respondents with normal hearing always used hearing protection. This percentage increased to 5% ($SD = .19$) among those with mild hearing loss and surged to 22% ($SD = .36$) among individuals with at least moderate hearing loss. Post-hoc pairwise Wilcoxon rank-sum tests revealed significant differences between those with at least moderate HL and mild HL ($p < .001$), as well as between those with at least moderate HL and normal hearing ($p = .03$).

Table 3.1

Descriptive Statistics of hearing protection usage by professional and amateur musicians

Use of hearing protection	Never	Rarely	Sometimes	Mostly	Always	Not applicable
Individual practice time						
Amateur	160 (82.4%)	15 (7.8%)	7 (3.6%)	3 (1.5%)	1 (.5%)	8 (4%)
Professionals	137 (70.6%)	27 (13.9%)	13 (3.7%)	4 (2%)	8 (4%)	5 (2.6%)
During rehearsals						
Amateur	140 (72.5%)	17 (8.8%)	22 (11.4%)	7 (3.6%)	0	7 (3.6%)
Professionals	28 (14.5%)	62 (32.3%)	59 (30.7%)	22 (11.5%)	19 (9.9%)	2 (1%)
During concerts						
Amateur	148 (77.1%)	20 (10.4%)	9 (4.7%)	3 (1.5%)	3 (1.5%)	9 (4.7%)
Professionals	58 (30.2%)	46 (24%)	52 (27%)	14 (7.3%)	20 (10.4%)	2 (1%)

Note. N = 400; 'NA' values are omitted.

3.7.4. Noise Exposure

Professional musicians reported significantly higher practice and performance durations than their amateur counterparts. On average, professionals engaged in 7.1 hours ($SD = 4.2$) of weekly practice for concerts, whereas amateurs reported an average of .8 hours ($SD = 1.4$). Rehearsal times followed a similar pattern, with professionals practising 15.2 hours weekly ($SD = 5.9$), in contrast to 3 hours ($SD = 2.3$) for amateurs. Besides group performances, professionals also practised more (i.e., individual practice), averaging 8.6 hours per week ($SD = 5.1$) compared to amateurs' 3.6 hours ($SD = 3.7$).

Based on these practice habits, instrument groups, and hearing protection usage, participants' self-reported musical noise exposure unit (UNE) was derived. Following the classification by Prendergast et al. (2017), individuals with a $\log_{10}(\text{L-UNE})$ of one or higher were categorised as having high noise exposure - equivalent to a decade of exposure at 90 dB A. According to this criterion, a proportion of 50% of professional musicians were classified within the high noise exposure group, a figure that contrasts with the 9.9% of amateur musicians, as illustrated in Figure 3.4.

In addition, the Wilcoxon rank sum test revealed significant group differences in UNE and L-UNE: Professional musicians registered almost three times higher median levels of musical sound exposure ($\text{UNE}_{\text{Mdn}} = .34$, $\text{MAD} = .25$) and almost four times higher lifetime music noise exposure ($\text{L-UNE}_{\text{Mdn}} = 10$, $\text{MAD} = 8.1$), relative to their amateur counterparts, who reported lower levels of $\text{UNE}_{\text{Mdn}} = .09$ ($\text{MAD} = .07$) and $\text{L-UNE}_{\text{Mdn}} = 2.1$ ($\text{MAD} = 1.91$). These differences were statistically significant for both UNE, $W(364) = 5828$, $p < .001$, and L-UNE, $W(364) = 5291$, $p < .001$. The effect sizes were considerable, with $r_{\text{rb}} = -.65$, 95% CI $[-.71, -.57]$ for UNE and $r_{\text{rb}} = -.68$, 95% CI $[-.74, -.61]$ for L-UNE. Notably, 57% of amateurs' lifetime music noise exposure have been experienced during individual practice, 33% during rehearsals, and 10% during concerts. In contrast, professionals experienced 49.4% during practice, 34.2% during rehearsals, and 16.4% during concerts.

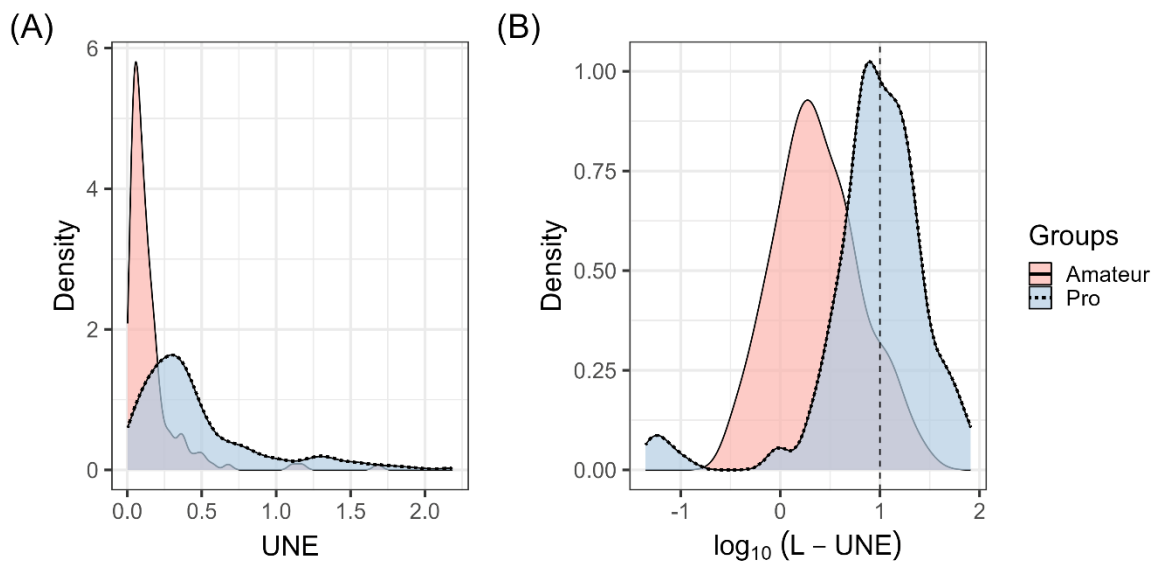


Figure 3.4 Musical sound exposure per year (UNE) and the cumulative (lifetime) music sound exposure (L-UNE) for amateur (pale red) and professional musicians (light blue). 'Density' refers to the empirically estimated probability density function.

Influence of hearing loss and age. In examining the association between UNE and age, the following pattern emerged: UNE was found to be highest among the youngest participants, with those under 30 years showing a median UNE score of .35 ($MAD = .34$). Overall, the UNE score decreased with age, reaching its lowest values in the 40-49 age group, with a median score of .14 ($MAD = .16$), and remaining that low for the rest of the musician's career (see Fig. A3.3 in Supp. Mat.). This effect was statistically significant ($H(2, 364) = 23.2, p < .001$), demonstrating a marginal effect size ($\epsilon^2 = .04$). Post-hoc analyses utilising pairwise t-tests revealed that the 'below 30' age group is significantly higher when compared with the '40-49' ($p < .01$), the '50-59' ($Mdn = .16, MAD = .17, p < .01$), and '60-69' ($Mdn = .18, MAD = .19, p = .04$) age groups. With respect to the degree of hearing loss, no significant differences were observed – neither for the UNE ($H(2, 364) = .58, p = .75$), nor the L-UNE ($H(2, 364) = 1.02, p = .6$).

3.7.5. Hearing health awareness

In the present study, musicians generally perceived hearing health as a topic that does not receive enough discussion, with 71.5% affirming this view as either 'rather true' or 'true'. This sentiment was slightly more common among amateurs (78.5%) compared to professionals (64%). Furthermore, a significant majority (74.75%) expressed a desire for regular counselling on the topic of hearing health and hearing impairment (at least 'rather true'), with a higher interest demonstrated by professional musicians (81.5%) compared to amateurs (68%). However, a quarter of respondents (24%) expressed reluctance, responding with 'not true' or 'rather not true' when asked if they would discuss their hearing loss openly. These rather negative attitudes towards hearing health were further reflected in the Hearing Health Awareness (HHA) factor scores. Overall, amateur musicians exhibited lower average awareness scores ($HHA_{Mdn} = 4.1, MAD = 3$) compared to professional musicians ($HHA_{Mdn} = 8.1, MAD = 1.3$). This difference was statistically significant ($W(359) = 5711, p < .001$) with a very strong effect size ($r_{rb} = -.65, 95\% CI[-.71, -.6]$), suggesting a substantial difference in hearing health awareness between amateur and professional musicians (see also Fig. 3.5).

Hearing health awareness (HHA)

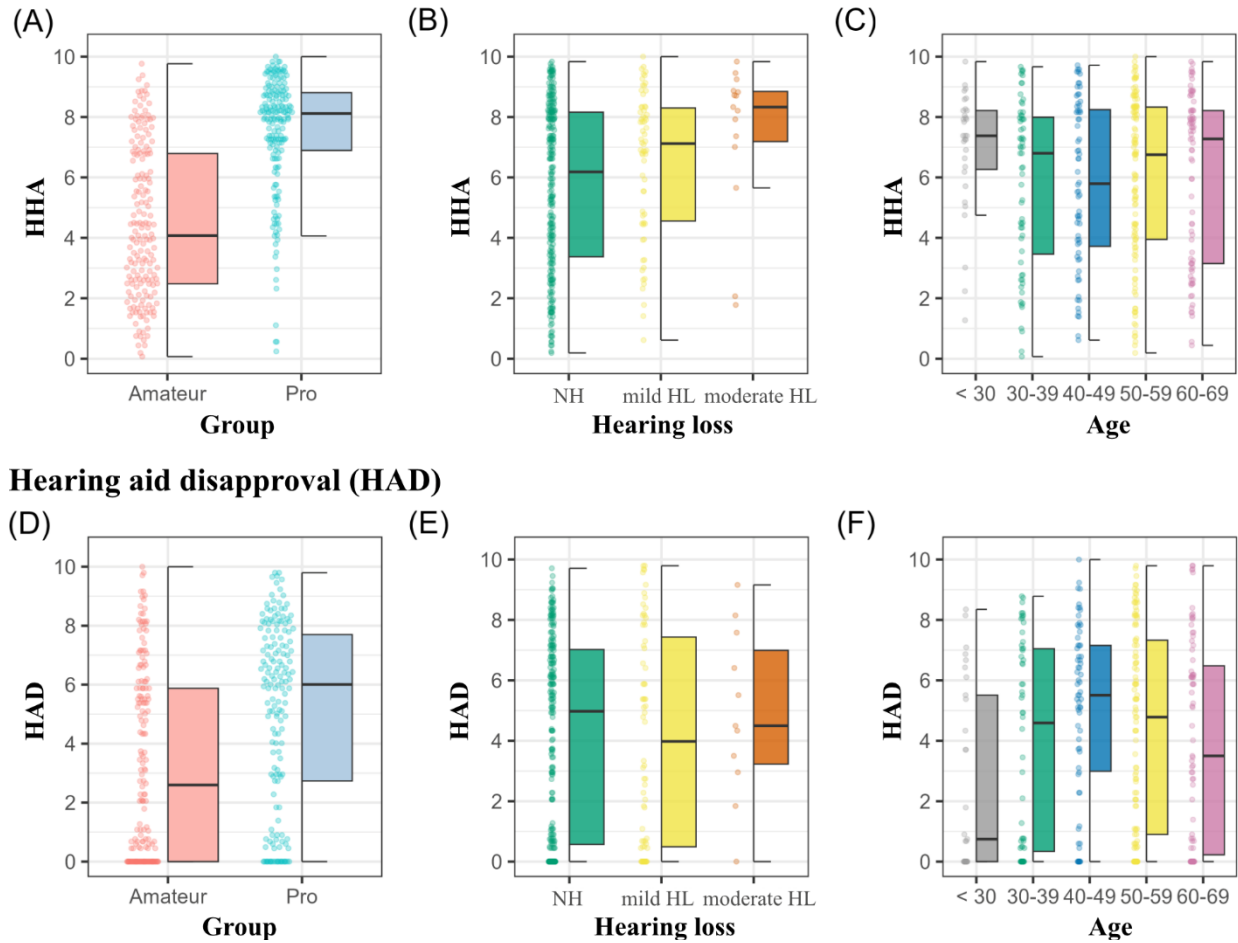


Figure 3.5. Scores of for the scales Hearing Health Awareness (HHA, top row) and Hearing Aid Disapproval (HAD, bottom row) as a function of group (panels A, D), hearing loss (B, E), and age (C, F).

Influence of hearing loss and age. HHA factor scores also revealed significant differences across hearing loss categories ($H(2, 359) = 9.4, p = .01$; see Fig. 3.5B). The NH group had a median HHA score of 6.2 ($MAD = 3.3$), the 'mild HL' group 7.1 ($MAD = 2.4$), and the 'at least moderate HL' group 8.3 ($MAD = 1.4$). However, the effect size was rather small ($\epsilon^2 = .03$). The corresponding post hoc test found only the difference between the 'NH' and 'at least moderate HL' group in HHA scores to be significant ($p = .03$). In terms of age, no differences in HHA factor scores could be observed ($H(4, 359) = 3.2, p = .52$).

3.7.6. Hearing aid disapproval

The analysis of the Hearing Aid Disapproval scale (HAD) revealed significant differences in approval ratings between amateur and professional musicians. Professional musicians exhibited higher levels of disapproval towards hearing aids ($HAD_{Mdn} = 6, MAD = 3.2$) compared to their amateur counterparts ($HAD_{Mdn} = 2.6, MAD = 3.9$; see Fig. 3.5). This difference was statistically significant ($W(335) = 9584.5, p < .001$) and demonstrated a moderate effect size ($r_{rb} = -.32, 95\% CI[-.42, -.2]$).

Influence of hearing loss and age. Moreover, the data indicated a relationship between age and HAD scores, with the Kruskal-Wallis test showing significant differences and a small effect size ($H(4, 335) = 11.44, p = .02, \epsilon^2 = .03$). Pairwise post-hoc tests indicated that the youngest age group 'below 30 years' exhibited lower disapproval scores ($Mdn = .75, MAD = 1.1$) compared to the 40-49 years age group ($Mdn = 5.51, MAD = 2.8$), which showed the highest disapproval ratings, $p = .015$ (see Fig. 3.5F). Despite these differences, an overall trend related to age was not observed, as Spearman's correlation revealed a non-significant relationship between participants' age and their hearing aid disapproval scores ($r(333) = .05, p = .396$). Conversely, HAD scores did not significantly vary across musicians with different types of hearing loss ($H(2, 335) = .44, p = .8$; see Fig. 3.5E).

3.7.7. Hearing aid usage

The proportion of respondents using hearing aids was low. Only 7.8% of respondents ($N_{total} = 31$) reported using hearing aids or having tried them in the past. This usage rate was slightly lower among amateurs (6.5%, $n = 13$) compared to professionals (9%, $n = 18$). The average duration of hearing aid use among these individuals was approximately 9 years, with a slightly longer average usage reported among amateurs (9.8 years) versus professionals (8.6 years). Among those participants who reported wearing hearing aids, a large majority (79.7%) reported improvements in their hearing ('agree' or 'totally agree'). However, the benefits from hearing aids appeared to be uneven across different auditory tasks. Those respondents who reported to use hearing aids regularly ($n_{total} = 15$) found hearing aids to be at least somewhat helpful for speech (100%). Similarly, 73% ($n = 11$) reported benefits for listening with ambient noise in the background. However, the effectiveness of hearing aids for music listening was deemed inadequate by many, with only 60% ($n = 9$) finding them useful for this purpose. This issue was further underscored by 54% ($n = 8$) of respondents who reported the quality of hearing aids as 'rather not sufficient' for making music. Consequently, many refrained from using them, with 33% ($n = 5$) reporting they never used hearing aids for music listening.

3.7.8. Health-related quality of life

In the evaluation of veterans health-related quality of life scale (VR-12), no significant differences were observed between amateur and professional musicians with respect to the mental health component score (VR12-MCS; $W(353) = 16465, p = .35$, see Fig. 3.6) or the physical health component score (VR12-PCS; $W(353) = 16435, p = .37$). Furthermore, no differences could be observed between different degrees of hearing loss and the VR-12 ($H(4,364) = .58, p = .75$). However, the mental health component score correlated with age ($r(364) = .29, p < .001$), with musicians below 30 years reporting the worst mean scores ($M = 38.6, SD = 12$), and the oldest age group of 60+ years the least inclination of mental quality of life issues ($M = 48.9, SD = 9.5$).

Health-related quality of life (VR-12)

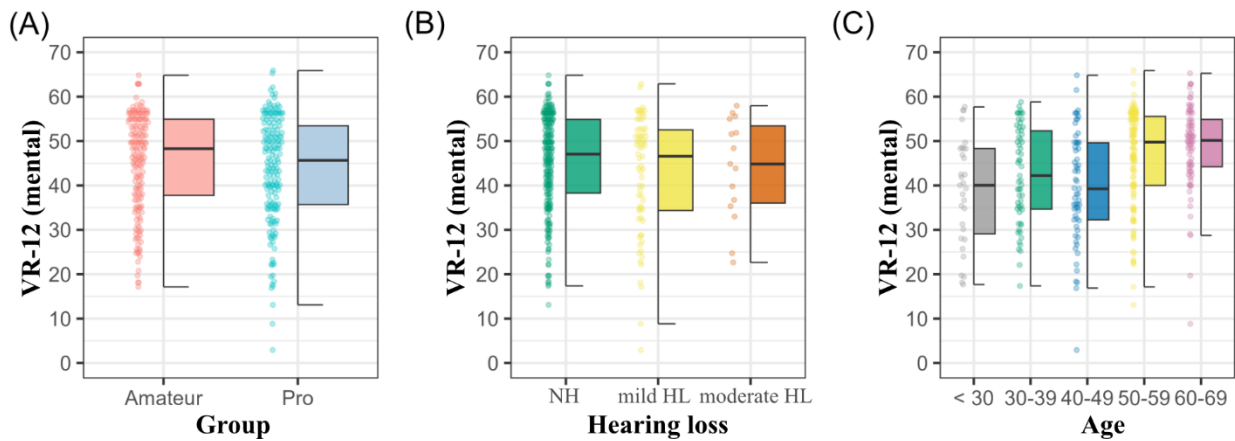


Figure 3.6 Health-related quality of life Veterans RAND 12-Item Health Survey (VR-12) as a function of group (A), hearing loss (B), and age (C).

3.7.9. Inter-Scale Correlation Analysis

In the inter-scale correlation analysis, significant correlations among variables related to age, music-related hearing problems, hearing in daily life, hearing health awareness, and health related quality of life indicators were found (for detailed correlation matrices, see Table 3.2). Specifically, a positive relationship was found between UNE and hearing health awareness ($r(330) = .24, p < .001$), indicating that individuals with higher noise exposure report greater awareness of hearing health related issues. Further, a relationship was observed between HHA and the HDDA subscale for social interactions ($r(350) = -.32, p < .001$), and with MRHP ($r(345) = .20, p < .001$). Besides that, MRHP was moderately correlated with the HDDA subscale for social interactions ($r(365) = -.39, p < .001$). These findings imply that difficulties in hearing, both in social and musical contexts, often co-occur and are associated with heightened awareness of hearing health. Intriguingly, however, the unit of noise exposure (UNE) did not significantly correlate with either the HDDA subscale for social interactions ($r(355) = 0, p = .75$) or MRHP ($r(342) = -.08, p = .152$). Lastly, there was a notable negative correlation between MRHP and the mental health component of quality of life ($r = -.23, p < .001$).

Table 3.2*The correlation matrix of selected tests from the test battery*

Variable	Age	UNE	LUNE	MRHP	HDDA (social)	HAD	HHA	VR12 (mental)
Age	-	-.08	.24***	-.01	-.22***	.05	-.02	-.3***
UNE	364	-	.91***	-.08	.02	.10	.24***	.01
LUNE	362	362	-	-.08	-.06	.14*	.27***	.12*
MRHP	374	344	342	-	-.39***	-.17**	.20**	-.24***
HDDA (social)	391	357	355	367	-	-.14*	-.32***	.20***
HAD	335	312	310	326	330	-	.31***	-.13*
HHA	359	330	328	347	352	317	-	-.13*
VR12 (mental)	353	325	323	340	346	314	332	-

Note. In the upper right half of the correlation matrix, the results are presented as Spearman's correlation coefficients (rho). The lower half of the matrix displays the respective sample sizes (N) used in each analysis. Significance levels are indicated as follows: * $p < .05$, ** $p < .01$, *** $p < .001$. Sig. values are highlighted in bold.

3.8. Discussion

A focal point of this investigation was to determine whether professional musicians would (a) report a greater incidence of hearing problems and (b) exhibit a heightened awareness towards hearing health compared to their amateur counterparts. To allow for a nuanced comparative analysis between these musicians, a carefully curated and balanced subsample of 200 professional and 200 amateur musicians, matched based on age, gender, and instrument family parameters was established.

3.8.1. Noise exposure and musicians hearing problems

The comparative analysis highlighted notable discrepancies in weekly music exposure, with professionals averaging 30.9 hours and amateurs 7.4 hours of active playing. These figures

align with prior findings, stating average practice times of 24-30 hours per week for professional musicians (e.g., Laitinen et al., 2003, Schmidt et al., 2011). The study considered estimates of sound level exposure across various playing settings, from solitary practice to full orchestral rehearsals, alongside the attenuation of music noise exposure via hearing protection usage, thereby establishing a measure for lifetime music noise exposure. The results revealed that professional musicians experience almost three times the yearly music sound exposure and nearly four times the lifetime exposure compared to their amateur counterparts. This significant disparity persists despite a higher occasional usage of hearing protection among professionals (53.5%) compared to amateurs (12%). Examining lifetime exposure, 50% of professional musicians, as opposed to ~10% of amateurs, reached an exposure level equivalent to experiencing 90 dB A for 8 hours daily, 5 days a week, over 52 weeks annually, for a decade. Such levels of exposure, accumulated solely through music playing, are alarming, as they mirror conditions known to cause permanent hearing loss, as documented in various studies and guidelines (e.g., Imam & Hannan, 2017; Liedtke, 2010; see also 'Königsteiner' and 'NIOSH' recommendations; Deutsche Gesetzliche Unfallversicherung, 2013; 2020; Chan, 1998).

As a consequence of these high levels of music sound exposure, it was hypothesised that professional musicians would experience more severe hearing problems - both music-related and in daily life contexts - compared to amateurs. Indeed, two thirds of the musicians reported some kind of hearing problems. But despite the reported differences in musical engagement and its influence on music noise exposure levels, almost no group differences in reported hearing problems could be observed. Specifically, professional and amateur musicians reported comparable rates of self-reported hearing loss (22% amateurs vs. 23% of professional), experiences with tinnitus (23.3% amateurs vs. 24.7% of professional), and oversensitivity to loud sound levels (44.8% amateurs vs. 53.3% of professional). Also scores from the music-related hearing problems scale (MRHP) and the subscore of the hearing-dependent daily activities (HDDA) subscale for basic sounds did not differ significantly between professionals and amateurs. Furthermore, we could neither find an association of noise exposure with HDDA or MRHP, nor could we predict differences between the degree of hearing loss with the music sound exposure metric (UNE). There was only one domain in which professional musicians reported significantly more severe hearing problems compared to amateurs: in the HDDA subscore for social interactions – with a rather small effect size.

The reported prevalence of hearing disorders among musicians, as found in this study, aligns with prior surveys (e.g., Hasson, 2009; Harper, 2002), yet the lack of a significant difference between professional and amateur musicians in terms of hearing impairments, despite differing levels of music noise exposure, is notable. This observation resonates with challenges faced by other researchers, such as Prendergast et al. (2017) and Russo et al. (2013), who also reported difficulties in establishing a straightforward link between levels of noise exposure and hearing impairments among musicians (see also Maghiar et al., 2023). Several reasons may account for these findings.

One possible explanation is that perceptual differences in hearing impairments that result from prolonged noise exposure may manifest gradually and too subtle for detection through self-report methods. Furthermore, musicianship has been linked to enhanced

auditory abilities such as improved rhythm perception, pitch discrimination, and timbre identification (e.g., Nave-Blodgett et al., 2021; Grahn et al., 2009; Micheyl et al., 2006; Tužnik et al., 2018; Zendel & Alain, 2012). There is even some evidence suggesting musical training may also benefit speech perception (Coffey et al., 2017), which is, however, still under debate (e.g., Madsen et al., 2019; McKay, 2021). Accordingly, musical training might be conceived to compensate for the negative effects of hearing impairments. Considering that two-thirds of the participants in this study are under 60 years of age, the subtle nature of noise-induced hearing changes, coupled with potential compensatory mechanisms developed through auditory training, could make early detection challenging. Thus, broad survey methods may not effectively capture these nuanced individual variations and contexts. Future studies should consider integrating hearing assessment methods, such as high-frequency extended pure-tone audiometry (e.g., Couth et al. 2020, 2024), to validate self-reported data and ensure more robust findings.

Another potential explanation of these findings could relate to the operationalization of the unit of music noise exposure (UNE) score. This metric, designed to provide a coarse estimate of prolonged noise exposure, is based on several assumptions that may affect its precision in identifying correlations with hearing problems. A key factor is the specific instrument and the repertoire played by musicians, as these can significantly impact sound exposure. For example, Schmidt et al. (2011) noted a substantial difference of 9.6 dB between repertoires, underscoring the repertoire as a significant factor in musicians' exposure (also see Rodrigues et al., 2014; Laitinen et al., 2003). Moreover, the model adopts a conservative approach, considering a constant noise exposure during individual practice from the age of 10 and orchestral/rehearsal noise exposure from age 20. This cross-sectional method does not account for variations in playing times throughout a musician's career and life, instead it presumes a constant exposure over time. Furthermore, the NESI examination instrument (Guest et al., 2018) used to calculate the musicians' sound level exposure was originally designed to capture both occupational and recreational activities. However, in this study, it was only applied to sound exposure from instrumental music encountered during active music-making activities. Recreational music listening and other everyday sources of noise exposure were not considered. Couth et al. (2020) utilized the same instrument but included recreational activities to calculate the total lifetime noise exposure between musicians and non-musicians (not specifically excluding amateur musicians), finding similar levels of exposure for both groups when recreational activities were included, as the majority of noise exposure could be attributed to recreational activities. Accordingly, in our sample, the musicians' total lifetime noise exposure may be more similar when recreational activities are taken into account, which could explain the lack of differences in reported hearing problems between professional and amateur musicians. It is conceivable that employing more detailed estimates, which account for different career phases and other noise exposure contexts, might uncover effects that the current methodology does not detect. Still, recent work by Couth et al. (2024) also shows minimal effects on hearing from noise exposure utilising a longitudinal design, suggesting that the effects observed in this study may remain stable over time.

Certainly, establishing a direct link between musical sound exposure and hearing impairments proves challenging (see also Couth et al., 2020; Elmazoska et al., 2024; Kähäri et

al., 2001a; 2001b), which suggests that the risk of noise-induced hearing loss in musicians might have been overestimated in previous research. Additionally, there is evidence that occupational noise, such as that encountered on construction sites, may be more detrimental to hearing than recreational sound, including music, possibly due to the dynamic nature of musical sounds (Neitzel & Fligor, 2019). Nonetheless, it is crucial to recognize that many musicians, both amateurs and professionals, are exposed to exceptionally high noise levels during their practice and performances. This exposure is particularly pronounced during individual practice sessions, where long uninterrupted playing and infrequent use of hearing protection are common. Thus, in sum, these findings underscore the importance of preventive measures to preserve musicians' hearing health.

3.8.2. Preventive measures and hearing health awareness

While hearing protection devices offer great potential for improving musicians' hearing health, their utilisation remains alarmingly low. The analysis showed greater (occasional) usage of hearing protection among professional musicians (53.5%) compared to amateurs (12%). This is evident despite 73% of professionals reporting that hearing protection impedes their performance; yet, musicians continue its usage, demonstrating a complex balance between auditory health and musical quality. The majority of musicians who use hearing protection opt for customised solutions (professionals: 84%; amateurs: 65%), indicating an understanding of the importance of personalised fit and comfort. However, a concerning low 1% of amateur musicians' playing hours and only 9% of professional musicians' playing hours are fully protected. In contrast, 80% of amateurs' and 36% of professionals' playing hours remain completely unprotected. Given the potential benefits of hearing protection (e.g., Kwak & Han, 2021), these numbers underscore the need for action to promote greater hearing protection adoption within the music community. Intriguingly, musicians demonstrate a pattern of increasing hearing health awareness as their hearing problems worsen (as also noted by Laitinen, 2005). This negative relationship exists in both music-related settings and social interactions and is mirrored in hearing protection usage: only 3% of those with normal hearing use full protection, increasing to 5% for those with mild HL and a notable 22% for those with at least moderate HL. These findings support the 'reactive action' phenomenon, where proactive measures are often taken only after issues arise.

The low utilisation rate of hearing protection resonates with findings from nearly two decades ago, as reported by Laitinen (2005), where 6% of Finnish musicians were noted to always use hearing protection devices. The stagnation in the adoption of protective measures, even considering variations among different countries, underscores the necessity for effective strategies and interventions that are specifically tailored to the unique needs of the music community. Several factors influence the notably low uptake of hearing protection across the musician community. On the one hand, legislative measures play a crucial role; an overwhelming majority of professionals (97%) regularly undergo hearing checks, typically every three years, as contrasted with 70% of amateurs, who tend to seek audiologist consultations only once every seven years on average. Accordingly, such regulatory mandates, as outlined in the German Noise and Vibrations - Occupational Safety and Health Ordinance

(2007), are likely central in promoting higher rates of hearing protection use among professional musicians. Performance-related anxieties, on the other hand, contribute to the reluctance in adopting hearing protection; 80% of professionals and 32% of amateurs who refrained from using hearing protection said they refrained from using hearing protection because they are concerned their performance might suffer. However, the primary factor is the lack of awareness about hearing health issues. On average, amateur musicians exhibited considerably lower average awareness scores compared to professional musicians. This discrepancy is most evident among younger musicians who, despite facing higher music noise exposure, are least inclined to use hearing protection. This finding contrasts with the report by Couth et al. (2021), which indicated that 77% of young musicians employ hearing protection at least once a week. Intriguingly, young musicians (<30 years) in this study had the highest awareness scores, prompting a re-evaluation of the direct association between awareness and behavioural change as proposed by Ajzen (1991) and underscores the complexity of translating awareness into action.

In conclusion, while establishing a clear link between heightened awareness of hearing health and the consistent use of hearing protection proves challenging, expanding regulatory measures to encompass amateur orchestras could offer substantial benefits. Furthermore, addressing performance-related anxieties and providing information on adaptation processes to mitigate performance impacts could be pivotal components of awareness programs (see also Couth et al., 2021 for model strategies).

3.8.3. *Hearing aid usage*

A mere fraction of musicians across both professional and amateur settings (N = 15) reported continuing usage of hearing aids, highlighting a noteworthy yet anticipated low utilisation rate. Furthermore, 54% of these users expressed that their hearing aids are not sufficient for making music, and only 80% reported that hearing aids improve their hearing at all. This finding aligns with existing literature indicating that current hearing aid technologies may not sufficiently cater to the specialised auditory needs of musicians: Previous survey studies have demonstrated that although hearing aids were perceived as useful in listening to both live and recorded music, their efficacy was deemed less satisfactory for live musical settings - a context of paramount importance for musicians. Common issues reported include distortion, acoustic feedback, imbalanced gain, and compromised tone quality (Madson & Moore, 2014). However, it is essential to acknowledge more recent studies presenting a more favourable view. These suggest that hearing aid usage can indeed improve music enjoyment and appreciation for both speech and music (Chern et al., 2023).

One major concern in this research centers on the potential underrepresentation of individuals with severe hearing impairments who have found hearing aids to be ineffective for musical tasks. Such individuals may have withdrawn from musical activities altogether, thereby becoming statistically invisible within the current study sample. Indeed, more than half of the respondents who reported having tried hearing aids in the past (16 out of 31) refrained from using them. This phenomenon may correspond to an instance of 'survivorship bias,' an analytical error characterised by an undue focus on subjects that 'survive' a given process while overlooking those who do not. Those subjects who report finding hearing aids

effective are likely those who have persisted in musical activities despite hearing challenges. This could inadvertently exclude the experiences of those who have ceased musical involvement due to the inadequacy of hearing aids for music-related tasks. The recognition of survivorship bias and its potential impact on the study's findings has significant implications for this research's validity (that is, the reported utility ratings are likely an overestimation of the usefulness of hearing aids), but also has implications for future investigative directions: Subsequent studies could adopt sampling techniques specifically designed to capture the experiences of individuals who have withdrawn from musical activities due to the limitations of hearing aids and hearing impairments. By doing so, a more comprehensive and accurate understanding of the true effectiveness and limitations of hearing aids may be attained, especially in the domain of music perception and performance.

Our findings further underscore the pivotal role of attitudes towards hearing aids as a moderator in the adoption and sustained utilisation of hearing aids among musicians. In our study, professional musicians manifested a markedly higher degree of disapproval towards hearing aids (HAD) compared to their amateur counterparts. Moreover, one out of four musicians indicated hesitancy in openly discussing their hearing loss in public. This differential disapproval may be anchored in both the stigmatisation of hearing aids and their perceived functional limitations within professional musical settings. The constructed scale for hearing aid disapproval encapsulates multiple facets that resonate with this notion, including irritation caused by colleagues wearing hearing aids, reluctance towards non-discrete devices, apprehensions about diminished peer respect, adverse audience perceptions, and the presumed incongruence between professional musicianship and hearing aid use. These elements collectively converge to form a pervasive stigma surrounding hearing aids, particularly within the realm of professional musicianship. The implications of this aversion are twofold. First, it elucidates an unexplored barrier to hearing aid adoption among a specialised population. Second, it offers a fertile ground for targeted interventions designed to mitigate these attitudes, thereby potentially augmenting hearing aid uptake and sustained usage among professional musicians. Additional empirical inquiries are imperative for a conclusive understanding of the dimensions and implications of hearing aid aversion in this particular cohort.

Importantly, the distribution of HAD among musicians presents a bimodal pattern. While a considerable majority exhibits minimal disapproval, a notable subgroup demonstrates substantial disapproval of hearing aids. There is some indication that older individuals tend to express higher disapproval towards hearing aids, yet a straightforward relationship could not be observed. Certainly, age-dependent inclination towards hearing aid aversion merits further empirical scrutiny, as it could have broader implications for targeted interventions aimed at enhancing hearing aid adoption among ageing musicians.

3.8.4. Musicians' overall well-being

No differences in terms of the health-related quality of life scale could be observed between professional and amateur musicians. The overall state of both physical and mental health in our sample of amateur and professional musicians can be characterised as healthy. This claim is substantiated when comparing published benchmark values from populations undergoing

healthcare challenges. For instance, Buchholz et al. (2017) reported mean scores of 33 on the physical sum scale and 41 on the mental sum scale in a population of patients undergoing orthopedic rehabilitation. Similarly, a study by Hüppe et al. (2022) involving patients with chronic pain reflected comparable results with average scores of 28.1 and 37.6 on the physical and mental scales respectively. In comparison, musicians from this study's sample demonstrates a substantially healthier profile, which aligns more closely with scores anticipated from a generally healthy population. However, there is considerable variability in the subscale for mental health among participants, with 113 musicians scoring below 40 (comparable to orthopaedic rehabilitation; see Buchholtz et al., 2017) and 48 scoring below 30 (comparable to chronic pain; see Hüppe et al., 2022). Notably, a significant proportion of those with lower scores are in the youngest age group, indicating that younger individuals tend to have poorer scores on mental health. This is consistent with previous research into musicians' psychological well-being; for example, Kenny et al. (2014) found that the youngest (female) musicians often report the highest levels of performance-related anxiety. Despite this, an increase in hearing issues related to music and in daily-life was associated with a decline in mental health-related quality of life, suggesting a significant psychological toll from these impairments.

3.8.5. Limitations

Several limitations must be acknowledged to contextualise the present findings. First, despite the efficacy of the matching procedure in equalising pivotal variables such as age, gender, and instrument type, additional imbalances - such as differences in repertoire played - could not be completely controlled for. Second, the process of matching inevitably led to the exclusion of certain participants, thus reducing the sample size. Nevertheless, the remaining sample (N = 400) was sufficiently robust to yield statistically meaningful insights. Importantly, the matching strategy served to better isolate the effects of the variables of interest by minimising potential bias due to confounding variables, thereby justifying the concomitant reduction in statistical power.

A potential self-selection bias poses a concern for the generalizability of the results. The voluntary nature of participation suggests that the sample might disproportionately represent musicians with a particular interest in hearing health, possibly due to personal experiences with hearing impairments. Consequently, the findings may exhibit a bias towards greater hearing health awareness, or a higher incidence of hearing issues compared to the general population of musicians, thereby limiting the generalizability of the results. However, it should be noted that many of the findings align with prior research on hearing health among musicians, not just in Germany but also in countries such as Australia and Finland, yielding corroborative evidence to the present study.

Additionally, the study's reliance on self-reported data introduces another layer of complexity. For example, existing literature suggests that participants in survey research often overstate socially desirable behaviors (e.g., Tourangeau & Yan, 2007), thereby introducing measurement error. This remains uncontrolled and might induce bias, particularly among musicians who are acutely aware of their auditory capabilities and may report subtle issues irrespective of actual noise exposure. Nevertheless, this propensity for

'over-reporting' is more pronounced in interviewer-administered surveys than in self-administered questionnaires, especially when the latter are conducted anonymously and voluntarily. Hence, while the present study has attempted to mitigate this bias through its design, the possibility of misrepresentation of certain attitudes still warrants cautious interpretation of the findings.

3.9. Conclusion

This study sheds light on the intricate relationship between musicians' unique hearing health challenges, attitudes, and preventive behaviours in German-speaking orchestras. Contrary to initial assumptions, professional musicians did not universally experience higher levels of hearing-related issues, despite encountering significantly greater levels of music-related noise exposure. However, these musicians were more proactive in adopting preventative measures such as hearing protection devices. This proactivity did not extend to hearing aid adoption, as professionals displayed noticeable aversion to these devices - an aversion that appears to be partly age-dependent. In this regard, amateur musicians exhibited considerably lower aversion towards hearing aids while showing lower general awareness of hearing health. By highlighting the differences in exposure and preventive actions between professional and amateur musicians, the present study underscores the need for tailored hearing health strategies by specifically suggesting a more proactive and informed approach to the preservation of hearing health within the musical community.

3.10. Declarations

Author contributions. RH, GH, UF, MS, ER, MM, KW, KS designed the study. RH and KS analysed the data. RH wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.

Ethics approval. Ethical approval was provided by the Ethics Committee at University of Oldenburg (Drs.EK/2019/092)

Consent to participate. Informed consent was obtained from all participants in this study, permitting the collection of anonymized data and its potential publication in journal articles.

Acknowledgements. This work has been funded by the Freigeist Fellowship of the Volkswagen Foundation to KS and was supported by the DFG (German Research Foundation): Project Number 352015383 SFB 1330 A6 and C4. The authors wish to thank Moe Touizrar for advice on the manuscript.

Conflict of interest. The authors declare that there is no conflict of interest.

Copyright. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Data Availability Statement. The data from this study has been made publicly accessible via the GESIS research data management platform. It can be accessed online at the provided URL [<https://doi.org/10.7802/2695>] (see Hake et al., 2024).

3.11. Literature

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Bigelow, R. T., Reed, N., Brewster, K. K., Huang, A., Rebok, G. W., Rutherford, B. R., & Lin, F. R. (2020). Association of hearing loss with psychological distress and utilization of mental health services among adults in the United States. *JAMA Network Open*, 3(7), e2010986. <https://doi.org/10.1001/jamanetworkopen.2020.10986>
- Buchholz, I., Kohlmann, T., & Buchholz, M. (2017). Vergleichende Untersuchung der psychometrischen Eigenschaften des SF-36/SF-12 vs. VR-36/VR-12. *Abschlussbericht*.
- Chan, Henry S. (1998). Occupational noise exposure; criteria for a recommended standard.
- Chern, A., Denham, M. W., Leiderman, A. S., Sharma, R. K., Su, I. W., Ucci, A. J., Jones, J. M., Mancuso, D., Cellum, I. P., Galatioto, J. A., & Lalwani, A. K. (2023). The Association of Hearing Loss With Active Music Enjoyment in Hearing Aid Users. *Otolaryngology–Head and Neck Surgery*. <https://doi.org/10.1002/ohn.473>
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- Couth, S., Prendergast, G., Guest, H., Munro, K. J., Moore, D. R., Plack, C. J., Ginsborg, J., & Dawes, P. (2020). Investigating the effects of noise exposure on self-report, behavioral and electrophysiological indices of hearing damage in musicians with normal audiometric thresholds. *Hearing Research*, 395, 108021. <https://doi.org/10.1016/j.heares.2020.108021>
- Couth, S., Loughran, M. T., Plack, C. J., Moore, D. R., Munro, K. J., Ginsborg, J., Dawes, P., & Armitage, C. J. (2021). Identifying barriers and facilitators of hearing protection use in early-career musicians: a basis for designing interventions to promote uptake and sustained use. *International Journal of Audiology*, 61(6), 463–472. <https://doi.org/10.1080/14992027.2021.1951852>
- Couth, S., Prendergast, G., Guest, H., Munro, K. J., Moore, D. R., Plack, C. J., Ginsborg, J., & Dawes, P. (2024). A longitudinal study investigating the effects of noise exposure on behavioural, electrophysiological and self-report measures of hearing in musicians with normal audiometric thresholds. *Hearing Research*, 451, 109077. <https://doi.org/10.1016/j.heares.2024.109077>
- Deutsche Gesetzliche Unfallversicherung. (2013). Präventionsleitlinie „Gehörschutz für Musiker“. Retrieved from https://www.dguv.de/medien/fb-psa/de/sachgebiet/sg_gehoerschutz/praeveleit_musik.pdf
- Deutsche Gesetzliche Unfallversicherung. (2020). *Empfehlung für die Begutachtung der Lärmschwerhörigkeit (BK-Nr. 2301) – Königsteiner Empfehlung – Update 2020*. Dr. Ulrike Wolf & Stefanie Palfner (Redaktion). <https://www.dguv.de/publikationen>
- Deutsches Musikinformationszentrum (2021). *Amateurmusizieren in Deutschland: Ergebnisse einer Repräsentativbefragung in der Bevölkerung ab 6 Jahre*. Institut für Demoskopie Allensbach, Deutscher Musikrat, Bonn. <https://bundesmusikverband.de/zahlen/>
- Di Stadio, A., Dipietro, L., Ricci, G., Della Volpe, A., Minni, A., Greco, A., de Vincentiis, M., & Ralli, M. (2018). Hearing loss, tinnitus, hyperacusis, and diplacusis in professional musicians: A systematic review. *International Journal of Environmental Research and Public Health*, 15(10). <https://doi.org/10.3390/ijerph15102120>
- Dingle, G. A., Sharman, L. S., Bauer, Z., Beckman, E., Broughton, M., Bunzli, E., Davidson, R., Draper, G., Fairley, S., Farrell, C., Flynn, L. M., Gomersall, S., Hong, M., Larwood, J., Lee, C., Lee, J., Nitschinsk, L., Peluso, N., Reedman, S. E., . . . Wright, O. R. L. (2021). How Do Music Activities Affect Health and Well-Being? A Scoping Review of Studies Examining

- Psychosocial Mechanisms. *Frontiers in Psychology*, 12, 713818.
<https://doi.org/10.3389/fpsyg.2021.713818>
- Elmazoska, I., Mäki-Torkko, E., Granberg, S., & Widén, S. (2024). Associations Between Recreational Noise Exposure and Hearing Function in Adolescents and Young Adults: A Systematic Review. *Journal of Speech, Language, and Hearing Research : JSLHR*, 67(2), 688–710. https://doi.org/10.1044/2023_JSLHR-23-00397
- Gembris, H., Heye, A., & Seifert, A. (2018). Health problems of orchestral musicians from a life-span perspective: Results of a large-scale study. *Music and Science*, 1, 1–20.
<https://doi.org/10.1177/2059204317739801>
- Gong, R., Hu, X., Gong, C., Long, M., Han, R., Zhou, L., Wang, F., & Zheng, X. (2018). Hearing loss prevalence and risk factors among older adults in China. *International Journal of Audiology*, 57(5), 354–359. <https://doi.org/10.1080/14992027.2017.1423404>
- Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(23), 7540–7548.
<https://doi.org/10.1523/JNEUROSCI.2018-08.2009>
- Greasley, A. E., Fulford, R. J., Pickard, M., & Hamilton, N. (2020). Help Musicians UK hearing survey: Musicians' hearing and hearing protection. *Psychology of Music*, 48(4), 529–546.
<https://doi.org/10.1177/0305735618812238>
- Greasley, A. E. (2022). Diverse Music Listening Experiences: Insights from the Hearing Aids for Music project. In J. Drever & A. Hugill. *Aural Diversity* Routledge (pp. 134–142).
- Guest, H., Dewey, R. S., Plack, C. J., Couth, S., Prendergast, G., Bakay, W., & Hall, D. A. (2018). The Noise Exposure Structured Interview (NESI): An Instrument for the Comprehensive Estimation of Lifetime Noise Exposure. *Trends in Hearing*, 22.
<https://doi.org/10.1177/2331216518803213>
- Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenbuck, K. (2024). Umfrage zur Hörgesundheits von Ensembleleitung, Berufs- und Amateurmusikerinnen in deutschsprachigen Orchestern und Chören [Data]. *GESIS, Cologne. Data File Version 1.0.0*. <https://doi.org/10.7802/2695>
- Halevi-Katz, D., Yaakobi, E., & Putter-Katz, H. (2015). Exposure to music and noise-induced hearing loss (NIHL) among professional pop/rock/jazz musicians. *Noise and Health*, 17(76), 158–164. <https://doi.org/10.4103/1463-1741.155848>
- Harper, B. S. (2002). Workplace and health: A survey of classical orchestral musicians in the United Kingdom and Germany. *Medical Problems of Performing Artists*, 17, 83–92.
- Hasson, D., Theorell, T., Liljeholm-Johansson, Y., & Canlon, B. (2009). Psychosocial and physiological correlates of self-reported hearing problems in male and female musicians in symphony orchestras. *International Journal of Psychophysiology*, 74, 93–100. doi:10.1016/j.ijpsycho.2009.07.009
- Hébert, S., & Lupien, S. J. (2009). Salivary cortisol levels, subjective stress, and tinnitus intensity in tinnitus sufferers during noise exposure in the laboratory. *International Journal of Hygiene and Environmental Health*, 212(1), 37–44.
<https://doi.org/10.1016/j.ijheh.2007.11.005>
- Hidalgo, J. L. T., Gras, C. B., Lapeira, J. M. T., Martínez, I. P., Verdejo, M. Á. L., Rabadán, F. E., & Puime, Á. O. (2008). The hearing-dependent daily activities scale to evaluate impact of hearing loss in older people. *Annals of Family Medicine*, 6(5), 441–447.
<https://doi.org/10.1370/afm.890>
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199–236. <https://doi.org/10.1093/pan/mpl013>
- Hüppe, M., Schneider, K., Casser, H. R., Knille, A., Kohlmann, T., Lindena, G., Nagel, B., Nelles, J., Pfingsten, M., & Petzke, F. (2022). Characteristic values and test statistical goodness of

- the Veterans RAND 12-Item Health Survey (VR-12) in patients with chronic pain: An evaluation based on the KEDOQ pain dataset. *Schmerz*, 36(2), 109–120.
<https://doi.org/10.1007/s00482-021-00570-5>
- Imam, L., & Hannan, S. A. (2017). Noise-induced hearing loss: A modern epidemic? *British Journal of Hospital Medicine*, 78(5), 286–290.
<https://doi.org/10.12968/hmed.2017.78.5.286>
- Jacob, R., Stelzig, Y., Buschermöhle, M., Berg, D., & Meis, M. (2017). Questionnaires as screening tools to identify persons with hearing deficiencies. *Z Audiol*, 56(3), 103–108.
- Kerby, D. S. (2014). The simple difference formula: an approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, 1.
- Kenny, D., Driscoll, T., & Ackermann, B. (2014). Psychological well-being in professional orchestral musicians in Australia: A descriptive population study. *Psychology of Music*, 42(2), 210–232.
- Keppler, H., Ingeborg, D., Sofie, D., & Bart, V. (2015a). The effects of a hearing education program on recreational noise exposure, attitudes and beliefs toward noise, hearing loss, and hearing protector devices in young adults. *Noise & health*, 17(78), 253–262.
<https://doi.org/10.4103/1463-1741.165028>
- Keppler, H., Dhooge, I., & Vinck, B. (2015b). Hearing in young adults. Part I: The effects of attitudes and beliefs toward noise, hearing loss, and hearing protector devices. *Noise & health*, 17(78), 237–244. <https://doi.org/10.4103/1463-1741.165024>
- Keppler, H., Dhooge, I., & Vinck, B. (2015c). Hearing in young adults. Part II: The effects of recreational noise exposure. *Noise & health*, 17(78), 245–252.
<https://doi.org/10.4103/1463-1741.165026>
- Kusy, A., & Châtillon, J. (2012). Real-world attenuation of custom-moulded earplugs: Results from industrial in situ F-MIRE measurements. *Applied Acoustics*, 73(6–7), 639–647.
<https://doi.org/10.1016/j.apacoust.2012.02.001>
- Kwak, C., & Han, W. (2021). The Effectiveness of Hearing Protection Devices: A Systematic Review and Meta-Analysis. *International Journal of Environmental Research and Public Health*, 18(21). <https://doi.org/10.3390/ijerph182111693>
- Kähäri, K. R., Axelsson, A., Hellström, P.-A., & Zachau, G. (2001a). Hearing assessment of classical orchestral musicians. *Scandinavian Audiology*, 30, 13–23.
doi:10.1080/01050390.1750069536
- Kähäri, K. R., Axelsson, A., Hellström, P.-A., & Zachau, G. (2001b). Hearing development in classical orchestral musicians: A follow-up study. *Scandinavian Audiology*, 30, 141–149.
<https://doi.org/10.1080/010503901316914511>
- Laitinen, H. M., Toppila, E. M., Olkinuora, P. S., & Kuisma, K. (2003). Sound exposure among the Finnish National Opera personnel. *Applied Occupational and Environmental Hygiene*, 18(3), 177–182. <https://doi.org/10.1080/10473220301356>
- Laitinen, H. (2005). Factors affecting the use of hearing protectors among classical music players. *Noise & Health*, 7, 21–29.
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at <https://www.soscisurvey.de>
- Liedtke, M. (2010). Akute Gehörschäden durch extrem hohe Schalldruckpegel. *HNO*, 58(2), 106–109. <https://doi.org/10.1007/s00106-009-2059-0>
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18.
<https://doi.org/10.1177/2331216514558271>
- Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, 9(1), 1–10. <https://doi.org/10.1038/s41598-019-46728-1>

- Maghiar, M. J., Lawrence, B. J., Mulders, W. H. A. M., Moyle, T. C., Livings, I., & Jayakody, D. M. P. (2023). Hearing loss and mental health issues in amateur and professional musicians. *Psychology of Music*, 1–14. <https://doi.org/10.1177/03057356231155970>
- Martini, A. E. (1996). European Working Group on genetics of hearing impairment. European Commission Directorate, Biomedical and Health Research Programme (HEAR). *Infoletter*, 2, 8.
- McKay, C. M. (2021). No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review. *Trends in Hearing*, 25, 233121652098567. <https://doi.org/10.1177/2331216520985678>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Musgrave, G. (2022). Music and wellbeing vs. musicians' wellbeing: examining the paradox of music-making positively impacting wellbeing, but musicians suffering from poor mental health. *Cultural Trends*, 32(3), 280–295. <https://doi.org/10.1080/09548963.2022.2058354>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2). <https://doi.org/10.1371/journal.pone.0089642>
- Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology. General*, 150(2), 314–339. <https://doi.org/10.1037/xge0000903>
- Neitzel, R. L., & Fligor, B. J. (2019). Risk of noise-induced hearing loss due to recreational sound: Review and recommendations. *The Journal of the Acoustical Society of America*, 146(5), 3911. <https://doi.org/10.1121/1.5132287>
- Noise and vibrations - Occupational Safety and Health Ordinance (2007). *Federal Law Gazette I from 6 March 2007 (p. 261), last amended by Article 3 of the Ordinance of 19 July 2010 (Federal Law Gazette I p. 960)*. Retrieved from: https://www.gesetze-im-internet.de/l_rmvibrationsarbschv/
- O'Brien, I., Driscoll, T., & Ackermann, B. (2013). Sound exposure of professional orchestral musicians during solitary practice. *The Journal of the Acoustical Society of America*, 134(4), 2748–2754. <https://doi.org/10.1121/1.4820900>
- O'Brien, I., Ackermann, B. J., & Driscoll, T. (2014). Hearing and hearing conservation practices among Australia's professional orchestral musicians. *Noise and Health*, 16(70), 189–195. <https://doi.org/10.4103/1463-1741.134920>
- Occupational Safety and Health Administration. (2002). *Hearing conservation [OSHA 3074 2002 (Revised)]*. U.S. Department of Labor. <https://www.osha.gov/sites/default/files/publications/osha3074.pdf>
- Olsen, A. D., Gooding, L., R., Shikoh, F., & Graf, J. (2016). Hearing health in college instrumental musicians and prevention of hearing loss. *Medical Problems of Performing Artists*, 31(1), 29.
- Phillips, S. L., & Mace, S. (2008). Sound level measurements in music practice rooms. *Music Performance Research*, 2(1993), 36–47.
- Prendergast, G., Millman, R. E., Guest, H., Munro, K. J., Kluk, K., Dewey, R. S., Hall, D. A., Heinz, M. G., & Plack, C. J. (2017). Effects of noise exposure on young adults with normal audiograms II: Behavioral measures. *Hearing Research*, 356, 74–86. <https://doi.org/10.1016/j.heares.2017.10.007>
- Richter, B. (2011). Hals-Nasen-Ohren-Heilkunde. In C. Spahn, B. Richter, & E. Altenmüller (Eds.), *MusikerMedizin. Diagnostik, Therapie und Prävention von musikerspezifischen Erkrankungen* (pp. 271–290). Stuttgart, Germany: Schattauer.

- Rodrigues, M. A., Freitas, M. A., Neves, M. P., & Silva, M. V. (2014). Evaluation of the noise exposure of symphonic orchestra musicians. *Noise & health*, 16(68), 40–46. <https://doi.org/10.4103/1463-1741.127854>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Russo, F. A., Behar, A., Chasin, M., & Mosher, S. (2013). Noise exposure and hearing loss in classical orchestra musicians. *International Journal of Industrial Ergonomics*, 43(6), 474–478. <https://doi.org/10.1016/j.ergon.2012.11.001>
- Safety Research Corporation of America (2020). *Noise and hearing protection fact sheet*. Retrieved July 5, 2023, from <https://www.srca.net/Blog/NoiseAndHearingProtectionFactSheet.aspx#:~:text=Why%20Can%E2%80%99t%20I,as%20adequate%20protection>
- Scherer, K. R. (2004). Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them? *Journal of New Music Research*, 33(3), 239–251. <https://doi.org/10.1080/0929821042000317822>
- Schink, T., Kreutz, G., Busch, V., Pigeot, I., & Ahrens, W. (2014). Incidence and relative risk of hearing disorders in professional musicians. *Occupational and Environmental Medicine*, 71(7), 472–476. <https://doi.org/10.1136/oemed-2014-102172>
- Schmidt, J. H., Pedersen, E. R., Juhl, P. M., Christensen-Dalsgaard, J., Andersen, T. D., Poulsen, T., & Bælum, J. (2011). Sound exposure of symphony orchestra musicians. *Annals of Occupational Hygiene*, 55(8), 893–905. <https://doi.org/10.1093/annhyg/mer055>
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- Tufts, J. B., & Skoe, E. (2018). Examining the noisy life of the college musician: Weeklong noise dosimetry of music and non-music activities. *International Journal of Audiology*, 57(sup1), S20–S27. <https://doi.org/10.1080/14992027.2017.1405289>
- Tużnik, P., Augustynowicz, P., & Francuz, P. (2018). Electrophysiological correlates of timbre imagery and perception. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 129, 9–17. <https://doi.org/10.1016/j.ijpsycho.2018.05.004>
- Versorgungsanstalt der deutschen Kulturorchester. (2021). *Bericht über das Geschäftsjahr 2021*. Retrieved from <https://www.orchesterversorgung.de>
- Wesseldijk, L. W., Mosing, M. A., & Ullén, F. (2020). Why is an early start of training related to musical skills in adulthood? a genetically informative study. *Psychological Science*, 32(1), 3–13. <https://doi.org/10.1177/0956797620959014>
- World Health Organization (2001). Grades of hearing impairment. *Hearing Network News*, 1: 144–157.
- Zander, M. F., Spahn, C., & Richter, B. (2008). Employment and acceptance of hearing protectors in classical symphony and opera orchestras. *Noise & Health*, 10, 14–26. <https://doi.org/10.4103/1463-1741.39004>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and aging*, 27(2), 410–417. <https://doi.org/10.1037/a0024816>
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494–521. <https://doi.org/10.1037/1528-3542.8.4.494>
- Zhang, J. D., & Schubert, E. (2019). A Single Item Measure for Identifying Musician and Nonmusician Categories Based on Measures of Musical Sophistication. *Music Perception*, 36(5), 457–467. <https://doi.org/10.1525/mp.2019.36.5.457>

DEVELOPMENT OF AN ADAPTIVE TEST OF MUSICAL SCENE ANALYSIS ABILITIES FOR NORMAL-HEARING AND HEARING-IMPAIRED LISTENERS

Submitted to *Behavior Research Methods*^d on 19. September 2023

First published online 13. November 2023

by

Robin Hake¹, Michel Bürgel¹, Ninh K. Nguyen¹, Alinka Greasley², Daniel Müllensiefen^{3,4} & Kai Siedenburg^{1,7}

¹ Dept. of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Germany

² School of Music, University of Leeds, United Kingdom

³ Department of Psychology, Goldsmiths, University of London, United Kingdom

⁴ Hanover Music Lab, Hochschule für Musik, Theater und Medien, Hannover, Germany

^d The original is accessible via: <https://doi.org/10.3758/s13428-023-02279-y>

Disclaimer

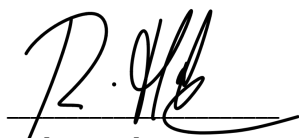
The content of this chapter is identical to that of the original publication.

The citation for the following publication is:

Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., & Siedenburg, K. (2024). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods*(56), 5456–5481.
<https://doi.org/10.3758/s13428-023-02279-y>

Author contributions

RH, MB, DM, and KS designed the study. MB and RH compiled the stimuli. RH programmed the experimental task, collected the data with assistance from AG, and wrote the first manuscript draft. Experiment 2 was part of the BSc thesis of NKN. RH and DM analysed the data. All co-authors edited and approved the final version of the manuscript.


Robin Hake
Kai Siedenburg

4.1. Abstract

Auditory scene analysis (ASA) is the process by which the auditory system makes sense of complex acoustic environments through organising sound mixtures into meaningful events and streams. Although music psychology has acknowledged the fundamental role of ASA in shaping music perception, no efficient test to quantify listeners' ASA abilities in realistic musical scenarios has been published yet. This study presents a new tool for testing ASA abilities in the context of music, suitable for both normal-hearing (NH) as well as hearing-impaired (HI) individuals: The adaptive Musical Scene Analysis (MSA) test. The test uses a simple 'yes-no' task paradigm to determine if the sound from a single target instrument is heard in a mixture of popular music. During the online calibration phase, 525 NH and 131 HI listeners were recruited. The level-ratio between the target instrument and the mixture, choice of target instrument, and number of instruments in the mixture were found to be important factors affecting item difficulty, whereas the influence of the stereo width (induced by inter-aural level differences) only had a minor effect. Based on Bayesian logistic mixed-effects model, an adaptive version of the MSA test was developed. In a subsequent validation experiment with 74 listeners (20 HI), MSA scores showed acceptable test-retest reliability and moderate correlations with other music-related tests, pure-tone-average audiograms, age, musical sophistication, and working memory capacities. The MSA test is a user-friendly and efficient open-source tool for evaluating musical ASA abilities and is suitable for profiling the effects of hearing impairment on music perception.

4.2. Introduction

A necessary foundation for developing an advanced understanding of music perception is auditory scene analysis (ASA) - the process by which the auditory system organizes the acoustic environment into separate coherent events and streams (Bregman, 1990). This process is essential for allowing listeners to make sense of complex sounds and to distinguish between different sources or elements within an auditory scene. ASA is critical for normal hearing in natural environments, but it is also critical for music perception, because disentangling simultaneous streams of sound (e.g., an oboe within an orchestra, or a tenor voice within a choir) is a key part of music appreciation and can be difficult, particularly for hearing-impaired individuals (Siedenburg et al., 2021; Greasley et al., 2020; Madsen & Moore, 2014). Even though music psychology has long acknowledged the fundamental role of ASA in shaping music perception, at the same time, no efficient and ecologically valid test to precisely quantify listeners' ASA ability in realistic musical scenarios has been published yet.

Prior research on ASA has largely employed atomistic approaches, which involve methods to examine auditory perception by isolating and analysing individual components or features within auditory scenes, simplifying complex stimuli to understand the underlying perceptual mechanisms (e.g., Bregman & Campbell, 1971; Micheyl et al., 2013). For example, one study by Bey and McAdams (2002) explored the role of schema-based processes in streaming using a melody recognition task with two unfamiliar six-tone sequences. Their findings revealed that performance improved with increasing frequency difference between target and distractor tones, and listeners performed better when the target alone was played first, but only when there was a difference in mean frequency between the target and distractor tones. Although approaches such as these have significantly contributed to our understanding of ASA, they may not fully capture the intricacies of real-world listening experiences, particularly in the context of music perception.

In a more recent study, Kirchberger and Russo (2015) developed the Adaptive Music Perception (AMP) test, which includes subtests for metre, harmony, melody, and timbre. Additionally, they introduced the melody-to-chord subtest, which tapped into the realm of ASA by asking participants to identify a target melody that is presented simultaneously with a chordal accompaniment. The task requires participants to segregate the target melody from the background chords, which is an essential aspect of ASA. The AMP incorporates an adaptive testing method that dynamically adjusts the difficulty of test items presented to an individual based on their performance in real-time. Nonetheless, the melody-to-chord subtest appeared to be particularly difficult for many participants, with roughly a quarter of normal-hearing (NH) participants and a third of hearing-impaired (HI) participants not being able to complete the task. Moreover, the AMP also uses artificial sound stimuli. Other tests also highlighted difficulties with ASA tasks. Siedenburg et al. (2020) adaptively measured signal-to-masker ratio thresholds of NH and HI listeners in a melody and timbre discrimination task, but also needed to discard data from HI listeners who yielded uninterpretable results. Another non-adaptive study, in which participants were asked to track target instruments in a classical piece while also attending to other instruments playing simultaneously, needed to discard data from HI participants as well due to chance performance (Siedenburg et al., 2021).

These documented challenges in measuring ASA abilities motivated us to develop an adaptive, computer-driven measurement instrument suitable for assessing ASA abilities in the context of music for individuals with a broad range of listening abilities (that is, suitable for both NH and HI listeners). By employing an ecologically valid methodology that incorporates authentic, recorded music, our aim is to capture the intricate and dynamic nature of auditory scenes, integrating multiple auditory features and cognitive processes for incorporating the overall experience, and providing meaningful measurements for both NH and HI individuals.

HI listeners are known to perform poorly compared to NH listeners on music related perceptual tasks such as timbre identification (e.g., Emiroglu & Kollmeier, 2008; Siedenburg et al., 2020), rhythm perception, pitch discrimination (Uys & van Dijk, 2011), melodic intonation (Kirchberger & Russo, 2015; Siedenburg et al., 2020), and auditory scene separation (e.g., Bayat et al., 2013). Older HI listeners also experience degraded spatial auditory processing (Akeroyd et al., 2007). While these effects are thought to result from damage to the ear, the auditory nerve or the nervous system (e.g., Cai et al., 2013), some studies suggest that in fact some of these low performance levels can be explained by the problems associated with aging and associated cognitive declines (Garami et al., 2020; Goossens et al., 2017; Gordon-Salant & Cole, 2016; Vinay et al., 2020). Yet, others argue that neither ageing nor HI can fully explain the observed individual differences, but a combination of these factors (Lentz et al., 2022). Furthermore, a growing body of research deals with the influence of musical sophistication and musical training on auditory perception skills. Even though there is still an ongoing debate as to whether musical training has a beneficial effect on speech-perception (e.g., Parbery-Clark et al., 2009; Bidelman & Yoo, 2020; McKay, 2021), several studies have demonstrated a positive link between musical training and music perception and other acoustical abilities (e.g., Madsen et al., 2019; Siedenburg et al., 2020; von Berg et al., 2021; Zendel & Alain, 2012). Moreover, musicians have been reported to outperform nonmusicians in basic cognitive tasks, such as those related to working memory (Talamini et al., 2017).

Although there is no clear picture of the causal factors underlying individual differences in ASA, it may be difficult to accurately understand ASA in music without taking into account these differences in the ability to process complex auditory scenes. One method which accounts for the large variability in listeners' abilities is adaptive testing. In contrast to standard testing procedures, where a fixed set of items is presented to all test takers regardless of their abilities or performance, in adaptive testing the difficulty level of test items is adjusted based on the responses provided by the test taker (for an overview see van der Linden & Glas, 2000). This allows the difficulty of the administered items to be tailored to the test taker's ability level, rather than presenting a fixed set of items that may be too difficult or too easy for some test takers. There are further benefits of adaptive testing. The comparatively shorter testing times can reduce inattention effects and test bias, making it less likely that test takers will gain advantage by guessing answers, resulting in more precise measurement estimates compared to standard procedures. This results in greater reliability, even when the testing time is reduced by 50-80% compared to non-adaptive tests (de Ayala, 2009; van der Linden & Glass, 2007; Weiss & Kingsbury, 1984). There are several examples of music-related tests that rely on adaptive testing procedures. Modern tests include measures of beat perception (Harrison & Müllensiefen, 2018), melody discrimination (Harrison et al.,

2017), and mistuning perception (Larrouy-Maestri et al., 2019). These tests generally apply Item Response Theory (IRT) models, which is a flexible adaptation approach that can be applied to a wide range of testing situations. However, IRT models require a calibrated item bank with a known difficulty level for all items. We consequently explored factors suitable for manipulating the item difficulty in a task that entails the detection of target sounds in mixtures of popular music.

A common approach to identify factors that affect the underlying construct of interest is to examine the underlying cognitive processes involved in a task (Embretson, 1983). One key aspect of ASA is the ability to segregate individual sound sources from background signals in a complex auditory mixture. This process involves identifying and grouping together sounds that share similar acoustical characteristics, similarities along perceptual dimensions such as pitch and timbre, and which exhibit similar temporal patterning or common onsets (that are principles of common fate, continuity, similarity, and proximity; see Bregman, 1990). Accordingly, instruments (or vocals) with distinctive timbral qualities are often easier to distinguish in an ensemble because their unique acoustic properties enable them to stand out from the mixture. Bürgel et al. (2021) found that participants' performance in a detection task was generally dependent on the target's instrument category, with lead vocals showing a particularly robust attentional salience regardless of low-level acoustic cues. A second relevant component is the acoustical complexity of a mixture. The greater the number of instruments contributing to a musical mixture, the higher the probability of energetic masking (one sound spectrally masks or obscures a quieter sound, making it difficult or impossible to hear) and informational masking (one sound interferes with or disrupts the perception of another sound). Thus, by providing a more complex musical scene, the segregation of individual instruments should become more difficult. Due to the same masking processes, it is also expected that the relative level of the instruments can provide (or hide) important cues for segregating the target sound from the background. Sounds that are louder are typically perceived as more salient than softer sounds. Thus, if the target instrument is presented at a lower level compared to the background mixture, it may be more difficult to segregate it from the mixture and detect it. Another primitive cue that has been identified to be important for encoding perceptual features includes signals' spatial location, as documented by a large range of studies on spatial release from masking (e.g., Litovsky et al., 2021). The spatial separation between a target sound source and interfering sound sources leads to an improvement in target signal detection. The effect of a-priori knowledge about the target location has been studied as well. For instance, in a study by Kidd et al. (2005) participants were asked to identify keywords from a target talker in the presence of two distractors in a spatially separated loudspeakers setting. Conditions in which a-priori knowledge about the target location was provided yielded higher performance compared to conditions in which no cue was provided.

4.2.1. *The present study*

To account for these processes with respect to individual differences among test takers, we designed a straightforward and simple 'yes-no' task (also known as the 'A-Not A' task; see Düvel & Kopiez, 2022) that required participants to decide whether a single target instrument

(or lead singing-vocals) was part of a two-second mixture of instruments. In a calibration phase, two online experiments were conducted in order to establish item characteristics that can be used as predictors in an explanatory Item Response Theory (IRT) model. This phase is necessary to fine-tune the test items for the adaptive test version, ensuring they accurately measure the intended construct and provide meaningful results across a broad range of ability levels. Experiment 1 of the calibration phase investigated the influence of the target-to-mixture level-ratio (labelled as **LEVEL**), the choice of the target instrument (**TARGET**), and the number of instruments in the mixture (**NUM**) on the test results. Experiment 2 focused on the effect of spatial separation in azimuth (at a stereo width of 0, 90°, 180°) introduced by inter-aural level differences (**ILD**).

Cognitive model of the MSA. In the present task, the full cognitive process model includes the following stages: (1) participants perceive the target instrument as a distinct auditory object; (2) participants store the mental representation of the sound of the target instrument in working memory; (3) participants use bottom-up and top-down processing to separate the target instrument from the background mixture based on its acoustic features and prior knowledge. Within this process, stream segregation is guided according to common principles such as similarity, proximity, continuity, and common fate; (4) participants selectively attend to the target instrument's timbre within the mixture (if present) based on their working memory representation while disregarding other sounds in the mixture. This includes a comparison of all segregated auditory streams to an internal mental representation or template of the target instrument's sound; (5) based on the similarity between the segregated auditory stream and the internal template, the listener decides whether the target instrument is present in the mixture or not. Accordingly, the test imposes demands on various cognitive processes, such as perception, working memory, segregation processes, attention, and decision-making. The task's difficulty is presumably influenced by a combination of factors, including the number and relative prominence of instruments within the mixture, as well as the nature of the target instrument.

Based on the proposed cognitive model of the MSA task and the reviewed literature, we formulated four hypotheses for the calibration phase:

1. Decreasing the target-to-mixture level-ratio will make it more difficult for listeners to accurately identify and separate the target sound from the mixture, resulting in lower accuracy.
2. The listener's ability to detect the target instrument within the mixture will decrease as the number of musical instruments in the excerpt increases.
3. There will be differences in detection accuracies for various target instruments. Although the literature provides limited guidance on the direction of these differences, we expect that lead vocals will be the easiest and bass the most difficult to detect, as indicated by Bürgel et al. (2021).
4. An increase in stereo width (induced by inter-aural level differences) will make it easier for listeners to localise and segregate the target sound within the mixture, leading to improved accuracy.

By examining these factors (i.e., **LEVEL**, **TARGET**, **NUM**, **ILD**), the calibration phase aims to optimise the MSA test items for effectively assessing individual differences in auditory scene analysis abilities in a musical context (irrespective of prior musical experience and hearing impairments). In addition to the online calibration phase, we conducted a validation experiment (Experiment 3) in a laboratory context to verify results under controlled conditions. This enabled us to assess the consistency of the MSA test through test-retest reliability analysis and to compare individuals' scene analysis abilities with those in a range of other psychoacoustic and music listening tests.

4.3. Experiment 1: Calibration Phase - Part 1

4.3.1. Methods

4.3.1.1. Test battery

Musical Scene Analysis Test (MSA). The MSA is a 'yes-no' test that resembles a 2-alternative-forced-choice (2-AFC) testing paradigm. The MSA assesses participants' ASA abilities in realistic musical scenarios by asking participants to detect a single target instrument (or lead vocals) in a mixture of instruments. Each trial consisted of a two-second audio excerpt of a single instrument or voice (the target), followed by a one-second silence, and a two-second excerpt with multiple instruments (the mixture). Participants were then asked to decide whether the target was part of the mixture or not (see Figure 4.1 for a schematic illustration).



Figure 4.1 Exemplary schematic illustration of the procedure. Each trial consisted of a two-second excerpt of a single instrument or lead vocals (the target), followed by a one-second gap, and a two-second excerpt with multiple instruments (the mixture). The listener's task was then to decide whether the target instrument was embedded in the mix or not.

All excerpts were drawn from an open-source music-database ("MedleyDB", Bittner et al., 2014; 2016), which consists of real-world multitrack music recordings representing a wide range of musical genres (e.g., pop, rock, world/folk, fusion, jazz, rap, classical). Prior to the extraction, a professional musician with a background in music production meticulously adjusted and post-processed each mix to improve overall audio quality. This process involved refining the balance among individual instruments, fine-tuning volume levels, and minimising

signal leakage. The excerpts were generated using the programming environment MATLAB (Mathworks, 2020). In order to identify a set of suitable candidate tracks, sound levels of each individual instrument within each song were analysed and calculated based on the root-mean-square average over 500 ms time windows for the full duration of the song. If one instrument in the target category and two to six additional instruments had sound levels above -20 dB relative to the instrument's maximum sound level in the song, the song qualified as a candidate base track. By setting a minimum sound level threshold of -20 dB relative to the instrument's maximum sound level in the song, we aim to include only those songs where all chosen instruments are clearly audible.

The candidate list comprised 12126 potential excerpts, each extracted from a distinct 2-second time window within one of the 117 eligible songs in the database. From this list, excerpts were selected pseudo-randomly, with a deliberate effort to minimise duplications of the base song. The selection protocol also ensured an equal distribution of excerpts in terms of the designated target instrument (lead vocals, guitar, bass, or piano) and the number of instruments in the mixture (either three or six). The specific target instruments were selected due to their diverse and widespread accuracy reported in Bürgel et al. (2021), which employed a similar detection task in one of their experimental conditions. The composition of instruments within each mixture was preserved in its original configuration, meaning that it could include a diverse array of instruments such as lead vocals, backing vocals, bass, drums, guitars, keys, piano, percussion, strings, or winds – depending on the base songs used. In half of the mixes, the target instrument did not play in the mixture. In such instances, only excerpts featuring an additional instrument were selected to guarantee the preservation of three or six instrument signals within the musical mixtures for all items. For example, we utilised excerpts originally containing four instruments when a mixture with three instruments was required. Detailed information regarding the specific composition of instruments within each excerpt can be found in the MSA GitHub repository (<https://github.com/rhake14/MSA>).

Overall, this yielded a 4 (*target instrument categories*) $\times$ 2 (*number of instruments in the instrument mixture*) $\times$ 2 (*presence of the target in the mixture*) design, for which 160 different excerpts from 98 base songs were compiled. In addition to the experimental factors target instrument (1) and number of instruments in the mixture (2), the first calibration experiment explicitly examined the influences of the target's level-ratio in comparison to the mixture (3). To this end, only for those excerpts, in which the mix contained the target instrument, four versions were created that varied in their target-to-mixture level-ratio (that is 0, -5, -10, -15 dB). Overall, a total of 400 items were created for the experimental task. Apart from the manipulation of the target instrument, the musical material in the excerpts was left unchanged (i.e., only excerpts were chosen in which the number of instruments corresponded to the desired condition). A logarithmic fade-in and fade-out with a duration of 200ms was applied to the beginning and end of the audio signals. To allow for usage with an online testing platform, all stimuli were converted from WAV format to MP3 with a bit rate of 320 kbit/s stereo (i.e., perceptually lossless compression). All resources, including the MSA test, task description, stimulus details, and example excerpts, are available on the project's GitHub repository (<https://github.com/rhake14/MSA>).

Degree of hearing impairment. Participants were asked to fill out an adaptation of the HAFM (Hearing Aids for Music) National Survey on hearing impairment (e.g., Greasley, 2022). These questions aimed at assessing the type and degree of hearing impairment. Participants were asked “Do you feel you have a hearing loss?” and were able to respond with 5 options ranging from: “No, I do not feel that I have a hearing loss.” to “Yes, I have the feeling of being profoundly hearing impaired”. For each option, a short description was given (e.g., “Yes, I have the feeling of being mildly hearing impaired: When I am talking to a person in a quiet room, I can usually understand a conversation. In noisy situations (e.g., in a pub) and in group conversations, I sometimes have problems understanding speech”). See Tables A4.4 and A4.5 for the complete self-assessment survey.

Goldsmith Musical Sophistication Index (GMS; Müllensiefen, Gingras, Musil, & Stewart, 2014). The Gold-MSI is a brief, 39-item self-report questionnaire that assesses several aspects of musical expertise. It was designed to capture subscales for active engagement, emotions, musical training, perceptual abilities, and singing abilities. Participants were asked to respond to a 7-Point-Likert-Scale (1 = Completely Disagree; 4 = Neither Agree Nor Disagree; 7 = Completely Agree). For both calibration experiments the two subscores for musical training (7 items, for example “I engaged in regular, daily practice of a musical instrument (including voice) for ___ years.”) and for perceptual abilities (9 items, for example “I can tell when people sing or play out of time with the beat.”) were used. The final composite score ranging from 1 to 7, with 7 being the highest possible score, was generated for each subscale. Both the used validated English and German version and other relevant materials are freely available on the Gold-MSI home page (<https://gold-msi.org>).

Huggins Headphone Screening (Milne et al., 2020). This 3-AFC task probes for headphone usage and makes use of a perceptual illusory pitch phenomenon, called the Huggins Pitch. The procedure involves presenting a white noise stimulus to one ear and the same white noise stimulus to the other ear, but 180° phase shifted over a narrow frequency band at about 600 Hz. A faint tone can then be detected, but only when the stimuli are presented dichotically over headphones. Importantly, when the stimuli are presented to one ear alone or over loudspeakers, the sound is very weak or absent. In order to pass the test, listeners needed to properly identify the tone five out of six times. Participants with severe HI struggled with this task and since the test was originally calibrated only among NH individuals, only NH participants need to pass the headphone screening in order not to be excluded from the data analysis. A free demo implementation of the task can be found on GitHub repository (https://github.com/ChaitLabUCL/HeadphoneCheck_Test).

Demographics questionnaire. The demographics questionnaire consisted of several items designed to gather participants' background information. Participants were asked to provide information on their age, gender, and educational level. This demographic data helped to characterise the study sample and provided context for interpreting the results of the main experimental measures.

4.3.2. Procedure

Ethical approval for the study was obtained from the Ethics Committee at the University of Oldenburg and the University of Leeds. Informed consent was obtained from all participants

tested. For Sample 1, the experiment was conducted using *testable.org*, a web-browser-based application for creating behavioural experiments and surveys online (e.g., Rezlescu et al., 2020). The study was conducted in a single online session, with an average completion time of about 35 minutes for participants in Sample 1. For those in Sample 2, the average completion time was 10 minutes. The study was available in both English and German languages. All participants provided digital consent by signing an electronic form and explicitly agreed, through a checkbox format, to remain in a quiet and distraction-free environment for the duration of the experiment. Participants, who usually wear hearing aids, were instructed to remove them for the study. All participants were financially compensated based on an hourly rate of €10; participants in Sample 1 received €5, while those in Sample 2 received €2 (or its equivalent in British pounds for UK residents), reflecting the different administration times for each sample group. Prior to the main experiment, a calibration sound was presented, and participants were instructed to adjust the volume of their playback device to a loud but comfortable level. Then the headphone screening was applied. Prior to the main experiment, participants underwent an MSA training session featuring five unique excerpts from the candidate list, which were not included in the main test set. Immediate feedback was provided after each response, and participants had the option to repeat the training phase as often as desired. The training phase was followed by the main experiment, that is, the MSA, where no feedback was given. A total of 160 trials, each presenting a single target-to-mixture level-ratio version of the 160 excerpts, were administered. The order in which the 160 items were presented and the selection of the target-to-mixture level-ratio version was randomised. Participants were allowed to pause at any time, with a recommended pause after half of the trials. After completing the experimental part of the study, a questionnaire regarding personal information including degree of hearing impairment, age, and gender, as well as the two subscales of the GOLD-MSI for musical training and perceptual abilities, was administered. The same procedure was followed for Sample 2, but participants were only presented 32 items, with each combination of parameters being given no more than once – effectively reducing the duration of the experiment to approximately 10 minutes. The rationale for this approach was to obtain a diverse participant sample with individuals exhibiting varied profiles, such as differing hearing abilities, ages, and genders. Accordingly, in Sample 2, we prioritised the diversity of participants over measurement accuracy. For Sample 2, the experiment was conducted using *psychTestR* (Harrison, 2020), an R package for creating web-browser-based behavioural experiments.

4.3.3. Participants

Two different samples of participants were recruited for Experiment 1: The first sample was recruited through a newspaper article, mailing lists, and a call for participation posted at the online job board of the University of Oldenburg and 126 participants (69 females) took part. Among these, 47 self-reported having at least a mild hearing-impairment (14 females; $M = 61.6$ years, $SD = 16.6$), whereas 79 reported having no hearing-impairment (55 females; $M = 26.9$ years, $SD = 9.3$). The second sample of participants was recruited via the online market research company SoundOut, located in the United Kingdoms, from which 1078 individuals completed the experiment ($M = 30.19$ years, $SD = 11.72$, 598 female). To ensure some control

over the playback conditions during the experiment, participants were instructed to wear two-channel headphones, which was checked with a screening test (Milne et al. 2020). Of the initial sample, 548 individuals failed the headphone screening and were thus excluded. Consequently, the analysis for Experiment 1 included a total of 525 NH participants with ages ranging from 18 - 72 years ($M = 28.6$ years, $SD = 11.02$, 274 female) and 131 HI participants with ages ranging from 23 - 82 years ($M = 40.5$ years, $SD = 20.6$, 67 female). The geographical location of participants was as follows: Australia (1), Canada (10), United Kingdom (135), Ireland (1), New Zealand (2), United States (322), Germany (126), and not specified (59). For a detailed overview with respect to the individual's degree of hearing impairment (i.e., degree of individuals self-rated hearing loss) see Figure 4.2E.

4.3.4. Data analysis

Bayesian generalised logistic mixed effects models (B-GLMM) were used for the analysis. B-GLMM are powerful and flexible alternatives to more commonly used frequentist approaches. In particular, they are able to account for uncertainty in parameter estimation and can provide stable estimates for categorical variables with many levels and smaller sample sizes with the help of informative prior distributions (e.g., Stegmueller, 2013; Dienes & Mclatchie, 2018). Using (Bayesian) mixed effect models with a binary dependent variable (correct/incorrect participant response), we can measure how different aspects of a musical excerpt affect its perceptual processing difficulty. We report the median estimates and 95% confidence intervals of the conditional effects for the final Bayesian logistic mixed-effects model. These values were obtained by averaging the conditional effects estimates across the respective factor of interest. By examining credible intervals and comparing the posterior probabilities of different hypotheses, we evaluate the extent to which detection accuracies differ among the respective conditions. The observed descriptive statistics, including the overall results averaged across items, can be found in the accompanying plots, which offer a visual representation of the data and highlight the main patterns observed in the study (see Figure 4.2 and Figure 4.4). Before conducting the mixed effects analysis, the data were inspected for unexpected response patterns. As a result, a total of 29 items were removed from the final analysis. An individual inspection of these problematic items showed that in some cases the low target-to-mixture levels-ratio made the target inaudible. In other cases, backing vocals were so highly similar to the lead vocals that the experimental task was in fact ill-defined. Eight items with lead vocals as the target instrument were also excluded as they showed a success rate of 100% and thus did not provide valuable discriminative information. It is noteworthy to mention that despite some participants scoring below the 50% chance performance level (occurring only in Sample 2), all participant scores were retained in the analysis. The rationale for this decision is twofold: First, the task inherently varies in item difficulty, and excluding low-performing participants could introduce selection bias, potentially underestimating the true challenge posed by certain items and compromising the generalizability of the results. Furthermore, the sample had already been subjected to a rigorous screening process; a significant cohort ($N = 548$) was excluded for failing the headphone screening test, presumably filtering out participants not adequately committed to the task. Retaining all participant scores was thus deemed crucial for preserving the integrity

and representativeness of the data. All analyses were executed in *R* (v2022.07.2+576; RStudio, 2020) and the Stan modelling language (v2.21.7; Carpenter et al., 2017), using the package *brms* as an interface from *R* to *Stan* (v2.18.0; Bürkner, 2017).

4.3.4.1. Bayesian GLM Model fitting

We fitted several B-GLMM (Bernoulli family with identity link; estimated using Markow-Chain-Monte-Carlo (MCMC) sampling with 35643 observations, 4 chains of 6000 iterations and a warmup of 3000) to predict participants' performance **SCORE** at the level of each individual trial (binary item responses with 0 = incorrect and 1 = correct). The model was built step by step in a hierarchical way by adding one parameter at a time. This allowed us to evaluate the individual impact of each variable. We first added each parameter to the same model structure separately, to get a first understanding of the general predictive performance of each parameter in isolation (as shown in Table 4.1).

We compared three models, each with a single fixed effect, using Bayes Factors (BF) in contrast to a null model that included only random effects and no fixed effects. BF is a measure to quantify the evidence for one model over another. A BF of 1 indicates that both the null and alternative models are equally likely, while a BF greater than 1 suggests that the data better supports the alternative model. Because the specifications for Models 1B, 2B, and 3B differ only in the fixed effect used, we interpret the BF as an indicator of which predictor is most strongly supported by the data for explaining the dependent variable, namely, MSA scores.

Model 1C, which incorporated the **LEVEL** as a fixed effect, and Model 1A (**TARGET**) provided the strongest evidence compared to the null model ($BF_{LEVEL} = \text{Inf}$; $BF_{TARGET} = 1.84\text{e}+48$). This was followed by Model 1B, which included the **NUM** as a fixed effect ($BF = 7.39\text{e}+07$). These results suggest that incorporating the **TARGET** and especially **LEVEL** as fixed effects better explains the observed data than the number of instruments in the mixture (**NUM**), highlighting the potential importance of the choice of the target instrument and the target-to-mixture level-ratio for the MSA task. However, it is important to interpret these results with caution, as the Bayes factor only provides relative evidence between models and does not directly quantify the effect size of each fixed effect. Further investigation, including effect size estimation and comparison, is needed to draw more definitive conclusions about the influence of these factors on the outcome variable (as indicated by the medians of the conditional effects and its accompanying density intervals, see Makowski et al., 2019a; 2019b).

According to this analysis, we planned four different versions of the model, each one adding a parameter in the order of their hypothesised importance to predict the MSA test scores. We then used the leave-one-out cross-validation information criterion (LOOIC) and Bayes Factor (BF) to provide a statistical measure of the model's predictive performance. 'LOOIC' is a model comparison metric derived from the concept of cross-validation. It estimates the predictive accuracy of a model by leaving out one observation at a time, fitting the model to the remaining data, and then predicting the left-out observation. A lower LOOIC value indicates better predictive performance, and the model with the lowest LOOIC value is considered the best (Vehtari et al., 2017). The final model was chosen based on a combination

of these measures, to consider the model's relative fit, its complexity and predictive performance. Model 1G (see Table 4.1) was excluded from the pool of candidate models, as the Pareto k estimate indicates potential issues with the model's convergence or MCMC sampling efficiency. The high BF-value suggests that Model 1F is likely to best explain the observed data when compared to the other models. Additionally, the low LOOIC-value indicates that Model 1F offers the optimal balance between its predictive ability and complexity (see Table 4.1). The final model used (1) **LEVEL**, (2) **TARGET**, and (3) **NUM** as fixed factors. For each of the factors, an interaction effect with the presence of the target in the mixture (**PRESENCE**) was included. Both excerpt and participant were added as random effects, which allowed the intercept to vary across participants and excerpts. Priors over the guessing and inattention parameters were set as beta distribution ($\alpha = 1, \beta = 1$), with an expected lower bound for possible values of .4 and an upper bound of .6 for the guessing parameter (expected success rate if the participant were to answer randomly), and 0 to .1 for the inattention parameter (expected probability of a participant not paying attention) respectively.

Table 4.1

Model comparison of all Bayesian (GLM) Models of Experiment 1

Model	Fixed effects	Random effects	Pareto k check	Model evaluation	
				LOOIC	BF
1A	SCORE ~ TARGET	random intercept for excerpt and participant	< .7	29191.5	1.84e+48*
1B	SCORE ~ NUM	same as A	< .7	29188.9	7.39e+07*
1C	SCORE ~ LEVEL	same as A	< .7	27300	Inf*
1D	SCORE ~ LEVEL + TARGET	same as A	< .7	27302.4	2.84e-10**
1E	SCORE ~ LEVEL + TARGET + NUM	same as A	< .7	27295.2	7.12e-15**
1F	SCORE ~ LEVEL + TARGET:PRESENCE + NUM:PRESENCE	same as A	< .7	27290.45	(reference for 1D, E)**
1G	SCORE ~ LEVEL + TARGET:PRESENCE + NUM:PRESENCE	random slope and random intercept for excerpt and participant	> .7***	25590.3	(excluded)***

Note. Model summary (fixed effects) shows how the model that was passed to the brms package in R [SCORE = binary-coded MSA score; NUM = number of instruments in the mix; LEVEL = level-ratio between target instrument and the mixture; TARGET = choice of the target instrument; PRESENCE = target instrument is part of the mixture. ‘.’ indicates an interaction effect; LOOIC = leave-one-out cross-validation information criterion; BF = Bayes Factor]. Given the structural design of the task, the LEVEL parameter always includes an interaction effect with the presence of the target in the mix (PRESENCE). *Model 1A, 1B, and 1C were tested against a null model, which includes only random

Note (Table 4.1) continued. effects but no fixed effects. **Model 1D and 1E were tested against Model F, to improve the interpretability of the comparison and to guide model selection. ***Pareto k values greater than .7 indicates potential issues with the model's convergence or sampling efficiency, that is the model has not captured the underlying structure of the data accurately or that there might be issues with the MCMC chains (Vehtari et al., 2019).

4.3.5. Results

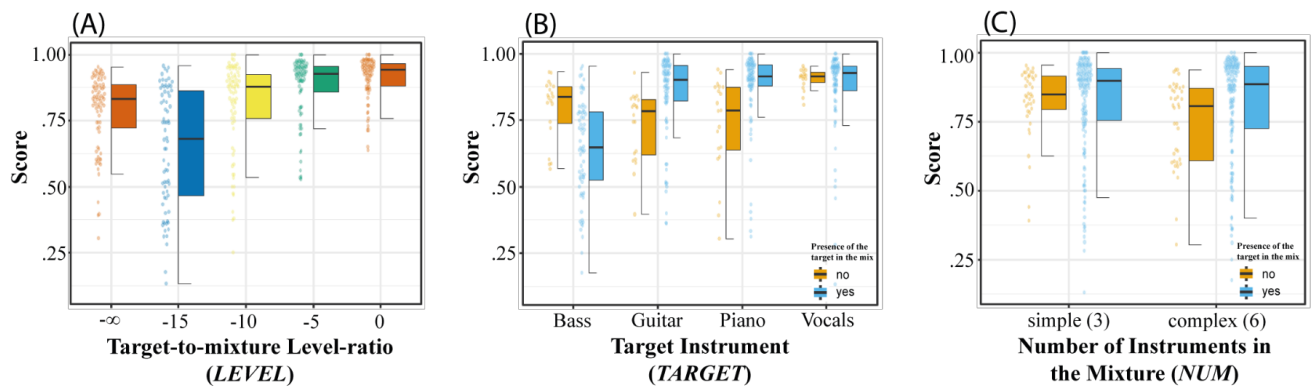
4.3.5.1. Model fit

The final model (1F) convergence and stability of the Bayesian sampling was assessed using $\hat{R}$, which was below 1.01 (Vehtari et al., 2019), whereas Effective Sample Size (ESS) was above 1000 (Burkner, 2017). All Pareto-k estimates were good ($k < .5$; Vehtari et al., 2019). For assessing the model's explanatory power we used the classification accuracy of our model, which represents the proportion of correct predictions out of the total number of predictions. Although classification accuracy does not provide a direct measure of explanatory power, it does offer insights into how well the model is performing in predicting binary outcomes. The final model was able to correctly identify 78.3% of the observed response data. When excluding the random effect information, the three employed fixed-effect factors alone accounted for a reasonable proportion of the model accuracy at 70.7%. The estimates obtained for this model are summarised in Table A4.1. Overall, the model assumed the guessing parameter to be .43, which lies within a reasonable scope of the theoretically assumed .5. According to the model, the inattention parameter was below .01. This suggests inattention effects to be negligible for the short test durations employed in the present research. A detailed overview about the estimated conditional effects of the model (effects of the employed parameter corresponding to all reference conditions) can be found in Figure 4.2D (see also Table A4.2). The results and statistical effects with regard to the individual experimental factors are described in the following.

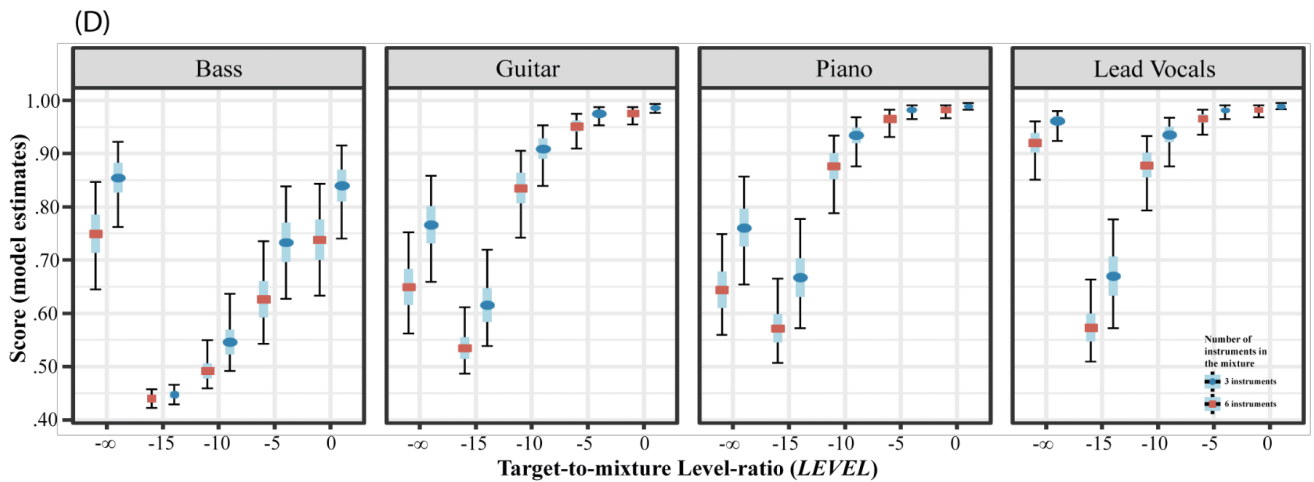
4.3.5.2. Level-ratio between target and mixture

Descriptive statistics and overall results are presented in Figure 4.2A. The model estimates that as the target-to-mixture level-ratio was decreased, the MSA task became increasingly difficult. The median difference in accuracy when changing the level-ratio from 0 dB ($M = 93.6\%$; $CI = [90\%; 96.3\%]$) to -5 dB ($M = 89.8\%$; $CI = [85.4\%; 93.5\%]$) was 3.8 percentage points. Based on the B-GLM model, non-linear one-sided hypothesis testing was performed, indicating the posterior probability (PP) negative difference between 0 dB and -5 dB was above .95. This can be interpreted as evidence for a difference between conditions (for uniform priors the posterior probabilities will exactly correspond to frequentist one-sided p-values; see e.g., Marsman & Wagenmakers, 2017). A substantial difference in performance was also observed as the level-ratio between the target and mixture was further decreased, with the detection rate dropping from a median percentage correct of 80.1% ($CI = [73.3\%, 85.5\%]$) at -10 dB to 56.6% ($CI = [50.5\%, 64.2\%]$) at -15 dB. When there was no target in the mix, the median detection accuracy was 79% ($CI = [70.2\%, 86.6\%]$). In short, the model yielded strong evidence that this difference was meaningful. The congruence between the observed data's descriptives and the model estimates highlights the strong model fit and underscores the robustness of the observed effect.

Item-based Boxplots for Experiment 1



B-GLMM Estimates for Experiment 1



Individual differences for Experiment 1

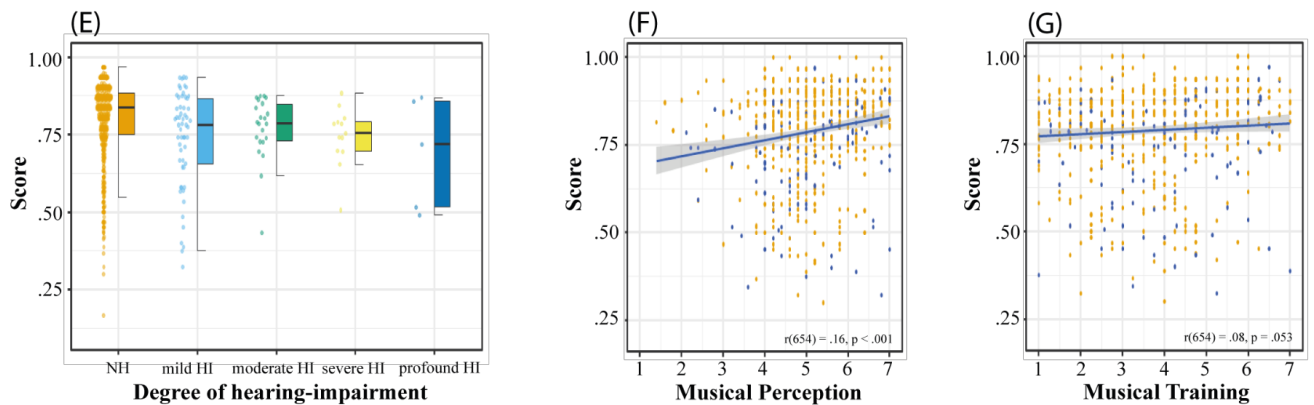


Figure 4.2 Results of Experiment 1. Figures (A), (B), and (C) show the proportion of correct scores of individual test items for each of the employed parameters. Individual dots (left to the boxplots) represent the observed accuracy averaged over participants for each item. Light-blue boxplots correspond to items in which the target was present in the mixture (orange when the target was not present in the mixture). ‘ $-\infty$ ’ corresponds to the items, in which no target was presented in the mixture. Figure (D) shows the conditional effects plot of the final B-GLMM (see Table A2.1).
Figure caption is continued on the following page.

Figure 4.2
continued

Median estimates with inner 50% (the length of the blue bars around the median) and outer 95% (the length of the black bars around the median) high density intervals (HDI) of the estimated parameters are shown. HDIs are uncertainty intervals and resemble the probability of the parameter estimate being within the given intervals. Figure (E) shows the boxplots of participants' average accuracy as a function of the degree of hearing impairment. Figures (F, G) show scatterplots illustrating the relationship between the participants' observed average detection scores and the two subscales of the Gold-MSI (musical training and musical perception). The individual dots represent the score for each participant averaged over all items.

4.3.5.3. *Choice of the target instrument*

The model indicates strong interaction effects of both the number of instruments in the mix and the target category with the presence of the target in the mix (**PRESENCE**). Thus, differences in correct detection rates must be interpreted according to this interaction. In line with previous research (e.g., see Bürgel et al., 2021), lead vocals yielded outstanding accuracy. Even though several items were excluded from this condition due to their perfect detection rates (100% of participant answers correctly), lead vocals still demonstrated the highest detection rates both when the target was presented in the mix ($M = 87.2\%$; $CI = [82.5\%; 91.2\%]$), and when the target was not presented in the mix ($M = 94.1\%$; $CI = [88.7\%; 97.0\%]$). The bass, on the other hand, was the most difficult to detect and also showed the greatest variation: for items with target in mix the median was 60.8% ($CI = [54.3\%; 68.0\%]$), and 80.2% ($CI = [70.3\%; 88.5\%]$) when the target instrument was not included. Both guitar ($M = 84.8\%$; $CI = [80.0\%; 89.1\%]$) and piano ($M = 87.1\%$; $CI = [82.4\%; 91.3\%]$) remained in an easy difficulty range for items in which the target was present, but became moderately difficult when the target was not present (i.e., $M_{\text{guitar}} = 70.8\%$; $CI = [61.1\%; 80.5\%]$ & $M_{\text{piano}} = 70.2\%$; $CI = [60.7\%; 80.3\%]$). Similar to the **LEVEL** factor, non-linear hypothesis testing was performed. Non-negligible differences were found between the bass and all other target instruments, both when the items included the target in the mixture and when the target was missing. When the target was not part of the mixture, the lead vocals also showed substantial differences in detection rates compared to all other instruments. For items in which the target played in the mixture, the differences in performance for lead vocals were apparent only in comparison to the bass. Guitar and the piano had comparable detection rates.

4.3.5.4. *Number of instruments in the mixture*

The B-GLM model indicated a relevant interaction effect for the number of the instruments and the presence of the target in the mixture. Overall, when the target did not play in the mixture, more complex musical excerpts with six instruments in the mixture showed lower detection rates ($M = 74.1\%$; $CI = [65.4\%; 82.7\%]$) compared to simpler mixtures with three instruments ($M = 83.6\%$; $CI = [75.0\%; 90.4\%]$). When the target was present in the mixture, the median detection accuracy was slightly different for both groups: 77.4% ($CI = [72.2\%; 82.5\%]$) for the 6-instrument mix and 82.5% ($CI = [77.4\%; 87.3\%]$) for the 3-instrument mix. Based on the model, this difference was found to be meaningful.

4.3.5.5. *Individual differences factors*

We assessed the relationship between the model-derived MSA scores, which include participant random intercepts adjusted for guessing and inattention parameters, and the GMS perceptual and training scores using Pearson's product-moment correlation. We found a marginal correlation between participants' model-based MSA scores and the musical training subscale of the GMS ($r = .103$, $p = .008$), and a slightly larger correlation between MSA scores and the musical perception subscale ($r = .188$, $p < .001$). Even though the inter-individual variation was quite large within each hearing-group of subjects, we observed a decrease in performance with the self-reported degree of hearing impairment. Averaged across all items, the mean instrument detection accuracy decreased from 79.8% (CI = [78.7%; 80.9%]) for participants with no HI, to 73.9% (CI = [69.8%; 77.9%]) for individuals with mild HI, 77.1% (CI = [72.8%; 81.4%]) for those with moderate HI, 74.5% (CI = [69.3%; 79.6%]) for participants with severe HI, and 69.0% (CI = [53.2%; 84.9%]) for individuals with profound hearing impairment. Participants observed responses are displayed in Figure 4.2 (E, F, G).

4.3.6. *Discussion*

As hypothesised, the model estimates suggested all three employed parameters to have a robust influence on the participants' accuracy. Experiment 1 showed that the accuracy in the MSA test depends on the choice of the target instrument, the number of instruments in the mixture, and the level-ratio between the target and mixture. Lead vocals had the highest detection rates, while the bass was the most difficult to process. The detection rates were lower for more complex musical excerpts with 6 instruments in the mixture compared to simpler mixtures with 3 instruments. As indicated by the sequential model comparison (see BF in Table 4.1), the effect of the level-ratio between the target and mixture had the most substantial effect on the difficulty of the task, with the task becoming increasingly difficult as the target-to-mixture level-ratio decreased. When included in the model, both the number of instruments in the mixture and the choice of the target instruments added comparatively smaller, but non-negligible improvements to the predictive performance of the model. The degree of hearing impairment also showed a relationship to the MSA test, with a decrease in performance as the degree of hearing impairment increased. There was a weak correlation between simple MSA sum scores and musical training and a slightly larger correlation between MSA scores and musical perception. Even among the most highly musically trained participants in this sample, no ceiling effects could be observed, whereas some of the more severely hearing-impaired individuals performed at chance level (see Figure 4.2E, G). This suggests that the test was challenging enough to provide meaningful results for individuals with prior musical expertise and is also likely to be sufficient to measure ASA performance in the context of complex multi-source music for individuals with normal-hearing as well as severe hearing-impairments.

Experiment 1 showed the strongest effects for the factor of the level-ratio between target and mixture. In Experiment 2, we explored whether another important factor of music production, spatialization in terms of a stereo image, proved to be similarly powerful for adjusting item difficulty.

4.4. Experiment 2: Calibration Phase - Part 2

4.4.1. Methods

4.4.1.1. Stimuli

In the second experiment, the same stimuli as in Experiment 1 were used. For this experiment, however, instruments were presented spatially separated along the azimuth. To create the stereo image, inter-aural level differences (ILD) were adjusted. ILDs are critical binaural cues that contribute to sound source localization (e.g., Stecker & Gallun, 2012). Here, we used the equal-power-panning method (e.g., Blauert & Braasch, 2008). Four conditions were thus generated, each characterised by varying panning widths, leading to stereo images with angular widths of (A) 0°, (B) 90°, and (C and D) 180°, yielding a total of 640 items (see Figure 4.3 for a schematic illustration). Specifically, Condition A served as a reference, where all instruments and the target were centrally localised. In Condition B, both the target and the other instruments were randomly allocated to one of the positions at -45°, -27°, -9°, 9°, 27°, or 45° (i.e., 90° of the frontal azimuth) within the stereo field. In Conditions C and D, the positions were expanded to -90°, -45°, -18°, 18°, 45°, or 90° (i.e., 180° of the frontal azimuth). In contrast to the fixed spatial positions of the target instrument in the preceding conditions, Condition D introduced a modification: the spatial location of the target changed between its isolated presentation and its occurrence within the mixture. This alteration effectively removed the informative cue provided by a consistent target location. To avoid ceiling effects, the target-to-mixture level-ratio was kept constant at -10 dB in all conditions.

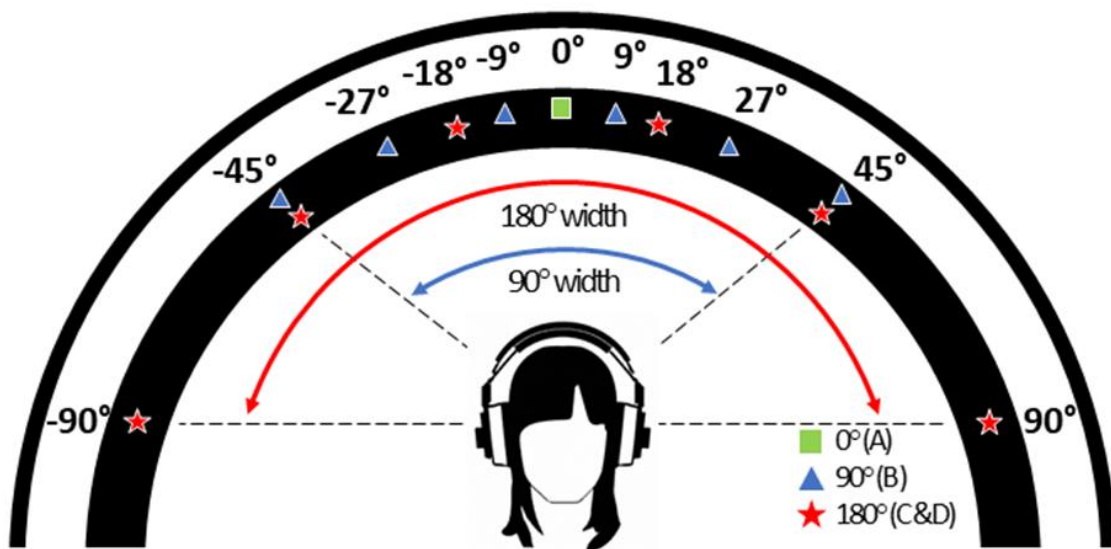


Figure 4.3 Schematic illustration of the stereo width conditions in frontal azimuth in Experiment 2. (A) All instruments are presented at 0°, i.e. the monaural reference; (B) the instruments are distributed evenly at a stereo width of 90° (that is, each instrument is randomly allocated to either one of the positions at -45°, -27°, -9°, 9°, 27°, 45°); (C) all instruments were distributed evenly across the full stereo width of 180°; (D) all instruments were distributed evenly across the full stereo width of 180°, but the position of the target instrument was changed when it was presented during the target presentation phase compared to when it was presented within the mixture (during the mixture presentation phase).

4.4.2. Procedure

The procedure and materials used in this experiment was similar to Experiment 1 and progressed in the following order: sound level calibration, Huggins headphone screening task, training phase of the MSA, data collection phase of the MSA, demographics questionnaire, and the two subscales for musical training and perceptual abilities of the german version of the Gold-MSI self-report questionnaire. Participants listened to all 160 audio excerpts, and for each excerpt, they were presented with one of the four different stereo width conditions. The order in which the excerpts were presented and the selection of the stereo width conditions was randomised across trials.

4.4.3. Participants

The second experiment included a total of 81 participants (53% female), of whom 40 were self-reported NH listeners and 41 self-reported HI listeners. One HI listener was removed from the analysis for stating that he did not use headphones. The remaining 80 participants were on average 43.2 years old ($SD = 21.45$) whereas NH individuals were predominantly younger with a mean age of 26.4 years ($SD = 8.4$, range: 20 – 62 years) and the HI predominantly older with a mean age of 60.2 years ($SD = 17$, range: 17 – 84 years). For a detailed overview with respect to the degree of hearing impairment see Figure 4.4F.

4.4.4. Data analysis

4.4.4.1. Bayesian GLM Model fitting

Similar to the first experiment, we used a Bayesian GLMM (Bernoulli family with an identity link; estimated using MCMC sampling with 12800 observations, 4 chains of 6000 iterations and a warmup of 3000) to predict participant performance. The strategy for building the random and fixed effects model structure was very similar to Experiment 1. However, instead of using the **LEVEL** parameter, the stereo width condition **ILD** (i.e., inter-aural level differences) with an interaction to the presence of the target in the mixture (**PRESENCE**), was included as fixed effect. First, we compared three models which each included one fixed effect only using Bayes factors (see Table 4.2). Model 2B, which included the **TARGET** as a fixed effect, provided the strongest evidence compared to the null model ($BF = 1.84e+06$), followed by Model 2A, which included **NUM** as a fixed effect ($BF = 1.67e+06$), and Model 2C, with **ILD** as fixed effects ($BF = 1.21e+05$). Similar to Experiment 1, these results emphasise the potential importance of the choice of the target instrument. In comparison, the effect size for **ILD** was the lowest.

For the model selection process, the same sequential comparisons approach as in Experiment 1 was used. Based on this approach, Model 2F was selected as the best model, due to its balance between fitting performance and model complexity. Model 2G was not considered, as the Effective Sample Size (ESS) was too low, indicating that the model did not fully converge. Model 2F included the factors **ILD**, **TARGET**, and **NUM**, each with an interaction effect with **PRESENCE**. Furthermore, a random intercept was included to account for the variation in the intercept across participants and excerpts and the prior was defined as in Experiment 1.

Table 4.2*Model comparison of all Bayesian (GLM) Models of Experiment 2*

Model	Fixed effects	Random effects	<i>Pareto</i> <i>k check</i>	Model evaluation	
				<i>LOOIC</i>	<i>BF</i>
2A	<i>SCORE ~ NUM</i>	random intercept for excerpt and participant	< .7	9763.5	1.67e+06*
2B	<i>SCORE ~ TARGET</i>	same as 2A	< .7	9761.3	1.84e+06*
2C	<i>SCORE ~</i> <i>ILD:PRESENCE***</i>	same as 2A	< .7	9739.6	1.21e+05*
2D	<i>SCORE ~ ILD:PRESENCE +</i> <i>TARGET</i>	same as 2A	< .7	9738.8	4.03e+08*
2E	<i>SCORE ~ ILD:PRESENCE +</i> <i>TARGET + NUM</i>	same as 2A	< .7	9739.2	5.03e+11*
2F	<i>SCORE ~ ILD:PRESENCE +</i> <i>NUM:PRESENCE</i> <i>+ TARGET:PRESENCE</i>	same as 2A	< .7	9734.5	1.54e+17*
2G**	<i>SCORE ~ ILD:PRESENCE +</i> <i>NUM:PRESENCE</i> <i>+ TARGET:PRESENCE</i>	random slope and random intercept for excerpt and participant	< .7	9619	(excluded)***

Note. Model summary (fixed effects) shows how the model that was passed to the brms package in R [SCORE = binary-coded MSA score; NUM = number of instruments in the mix; ILD = inter-aural level difference condition; TARGET = choice of the target instrument; PRESENCE = target instrument is part of the mixture. ‘.’ indicates an interaction effect; LOOIC = leave-one-out cross-validation information criterion; BF = Bayes Factor]. *Here we tested against a null model, which includes only random effects but no fixed effects. **Model 2G indicated that the Effective Sample Size (ESS) is too low, indicating that the MCMC algorithm used to estimate the posterior distributions of the model parameters may not have fully converged. A low ESS value can lead to biased or imprecise estimates. ***Given the design of the task, the ILD parameter cannot be interpreted without an interaction effect with the presence of the target in the mix (PRESENCE).

4.4.5. Results

4.4.5.1. Model fit

The model (2F) converged ($\hat{R} < 1.01$, ESS > 1000) and all Pareto-k estimates were in an acceptable range (i.e., $k < .5$). Overall, the model was able to predict 79% of the trial-level observations. When excluding the random effects (random intercepts for excerpts and participants), the model could correctly identify 70.4% of the observed data. See Table A4.2 for a detailed model summary of the estimates and model fit. In line with the procedure of the first experiment, all items with 100% detection rates were excluded from the analysis (4 items with bass as target, 14 with guitar, 20 with piano, and 41 with lead vocals).

4.4.5.2. Stereo width

In this experiment, we investigated the impact of *ILD* on accuracy across four distinct stereo conditions. The B-GLMM model identified a relevant interaction between stereo conditions and target presence within the mixture, emphasizing the need to account for this interaction when interpreting the results. With the target absent from the mixture, median accuracy was as follows: 77.5% (CI = [67.8%; 86.0%]) for condition A (0°), 78.8% (CI = [68.9%; 87.1%]) for condition B (90°), 81.8% (CI = [72.3%; 89.4%]) for condition C (180°), and 77.4% (CI = [67.5%; 86.0%]) for condition D (180° + R). Accordingly, when the target was absent in the mix, a trend was present only from condition A to C (see Figure 4.4E). Conversely, when the target was present in the mixture, a consistent trend in the median accuracy was apparent: Median accuracies were 79.5% (CI = [70.7%; 86.4%]) for condition A (0°), 82.1% (CI = [73.9%; 88.6%]) for condition B (90°), 82.8% (CI = [74.8%; 89.1%]) for condition C (180°), and 86.9% (CI = [79.9%; 92.2%]) for condition D (180° + R). In line with Experiment 1, non-linear hypothesis tests were conducted, affirming a trend of improved accuracy. This improvement was most pronounced in the full stereo width conditions of 180° (C and D) in comparison to the monaural condition A (refer to Figure 4.4D).

4.4.5.3. Choice of the target instrument

In Experiment 2, we observed a similar pattern of interaction effects between the target instrument category and the presence of the target in the mix as in Experiment 1. Consistent with our previous findings, lead vocals demonstrated exceptional accuracy (again, despite the exclusion of several items in this condition due to perfect detection rates). When the target was present in the mix, lead vocals showed a median detection rate of 92.3% (CI = [84.9%; 96.4%]), and when the target was absent, the rate was 88.3% (CI = [79.6%; 93.9%]). In contrast, the bass was the most challenging instrument to detect, with the largest degree of variability: the median was 58.1% (CI = [50.0%; 69.3%]) when the target was in the mix and 80.4% (CI = [70.3%; 88.5%]) when it was not. The guitar and piano demonstrated relatively low difficulty levels for items where the target was present, with median detection rates of 77.2% (CI = [67.1%; 86.2%]) for the guitar and 69.6% (CI = [59.5%; 79.8%]) for the piano. However, their difficulty increased moderately when the target was present (i.e., $M_{\text{guitar}} = 88.2\%$, CI = [78.6%; 94.2%]; $M_{\text{piano}} = 92.8\%$, CI = [85.7%; 96.6%]).

4.4.5.4. Number of instruments in the mixture

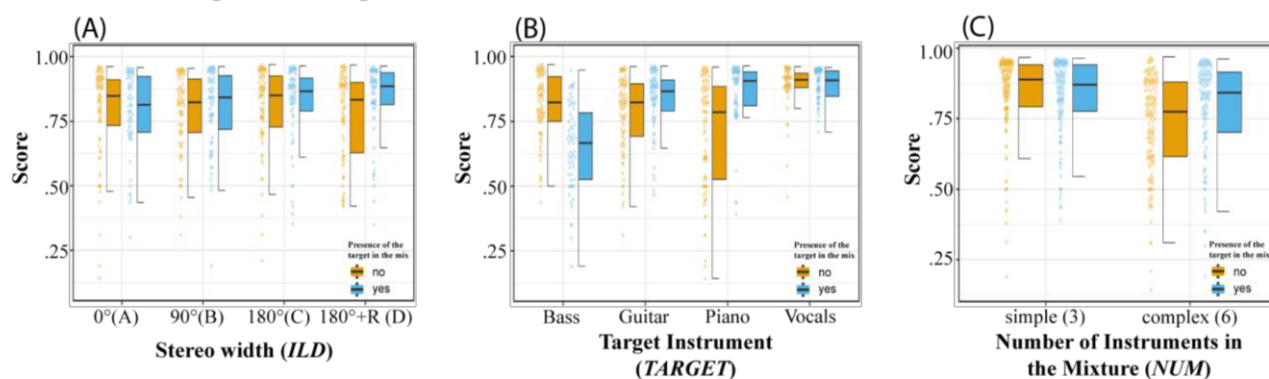
The results with respect to the number of instruments in the mix in Experiment 2 were in line with those from Experiment 1. More complex musical excerpts with six instruments in the mixture showed lower detection rates compared to simpler mixtures with only three instruments. When the target was not present in the mixture, the median for simpler mixtures with three instruments was 89.1% (CI = [80.4%; 94.5%]), while for more complex mixtures with six instruments, it was 68.6% (CI = [57.9%; 79.7%]). According to the model, this difference was considered considerable and meaningful. A similar trend was observed when the target was present in the mixture, albeit less pronounced: The median correct detection

rate was 85.5% (CI = [78.1%; 91.1%]) for the 3-instrument mix and 80.2% (CI = [71.5%; 87.1%]) for the 6-instrument mix.

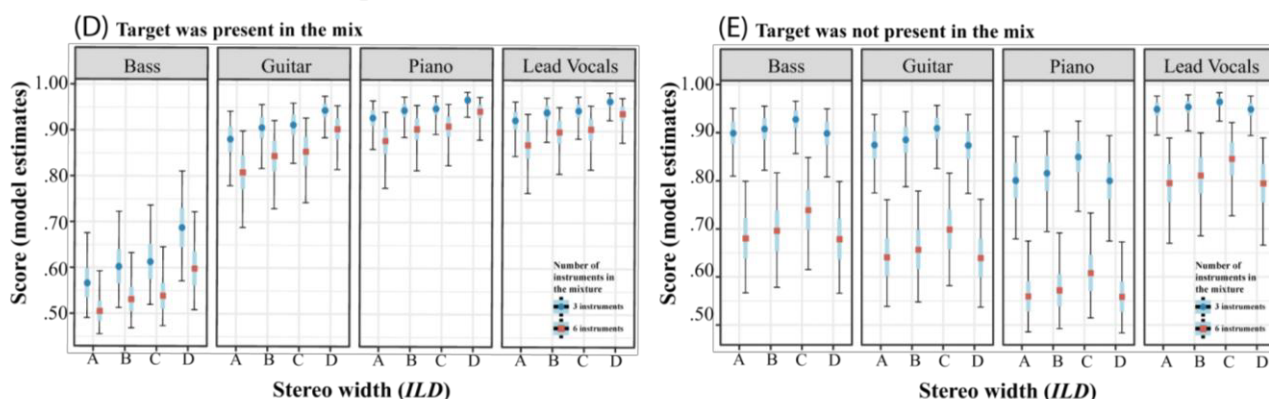
4.4.5.5. Individual differences factors

The model indicated a decrease in accuracy performance with an increase in the degree of hearing impairment. On average, individuals with NH had a mean accuracy of 84.7% (CI = [82.5%; 86.9%]). The mean accuracy for individuals with mild HI was 80.5% (CI = [76.1%; 84.9%]), while for those with moderate HI it was 80% (CI = [77.1%; 82.9%]). For individuals with severe HI, the mean accuracy was 70.2% (CI = [64%; 76.4%]). The lowest mean accuracy was seen in individuals with profound HI, at 54.2% (CI = [44.8%; 63.5%]). The correlation between model-based MSA scores and the GMS musical training scores remained small ($r = -.09$, $p = .45$), while the correlation with musical perception abilities was notably stronger ($r = .20$, $p = .073$).

Item-based Boxplots for Experiment 2



B-GLMM Estimates for Experiment 2



Individual differences factors for Experiment 2

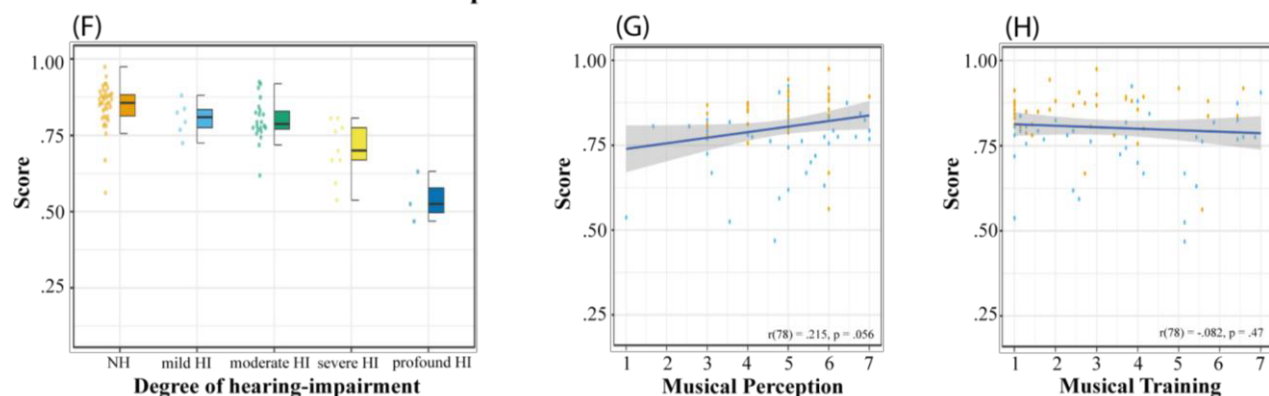


Figure 4.4

Results of Experiment 2. Figures (A), (B), and (C) show the proportion of correct scores of individual test items for each of the employed parameters. Individual dots (left to the boxplots) represent the observed accuracy averaged over participants for each item. Light-blue boxplots correspond to items in which the target was present in the mixture. Figure (D) and (E) shows the conditional effects plot for all conditions of the final B-GLMM (see Table 4.2). Median estimates with inner 50% (the length of the blue bars around the median) and outer 95% (the length of the black bars around the median) high density intervals (HDI) of the estimated parameters are shown. HDIs are uncertainty intervals and resemble the probability of the parameter estimate being within the given intervals.

Figure caption is continued on the following page.

Figure 4.4 Figure (E) shows the boxplots of participants' average accuracy as a function of the degree of hearing-impairment. Figures (F, G) show scatterplots illustrating the relationship between the participants' observed average detection scores and the two subscales of the Gold-MSI (musical training and musical perception). The individual dots represent the score for each participant averaged over all items.

continued

4.4.6. Discussion

Experiment 1 demonstrated that accuracy depended on the target instrument, with lead vocals yielding the highest and bass the lowest overall performance. Furthermore, mixture complexity influenced detection rates, as simpler three-instrument mixtures resulted in higher detection rates compared to more complex six-instrument mixtures. The results also suggested that individuals who reported better musical perceptual abilities tended to perform better on the MSA test, while the role of musical training remained inconclusive. Additionally, listeners with more severe hearing impairments exhibited lower MSA scores. Experiment 2 corroborated these findings.

We hypothesised that increased stereo width, as induced by ILDs, would facilitate target sound localization and segregation within the mixture, leading to improved accuracy irrespective of the presence of the target. However, this was not the case when the target was absent from the mixture. One possible explanation is that *ILD* cues alone are not sufficiently strong to improve accuracy based on the stereo percept. This is unlikely, though, since we observed improved accuracy when the target was present in the mixture, particularly in stereo conditions C (180°) and D ($180^\circ + R$) compared to monaural condition A. Interestingly, condition D displayed even higher accuracy, despite the target position changing between the presentation of the target alone and the mixture. This outcome was unexpected, and we initially assumed that changing the target position would cause confusion, leading to a decrease in detection rate in comparison to condition C. There are two possible explanations for this behaviour. First, participants were not informed about the cue, and the selection of conditions was fully randomised. Accordingly, participants might not have learned to rely on the positional cue due to the fully randomised and disclosed study design and thus never got confused by the change in position within condition D. However, this explanation cannot account for the improvement effect observed between conditions C and D. A second explanation might be the unexpectedness of the positional change of the target's position, which might have led to an acoustical novelty effect, which shifted the cognitive locus of attention towards the unexpected stimuli (i.e., the target instrument at the unexpected position). Research has repeatedly shown that infrequent auditory changes in a series of otherwise repeating sounds trigger an automatic response to the novel or deviant stimulus (e.g., Parmentier, 2014), which supports this assumption. Further research is necessary to elucidate the underlying mechanisms responsible for these performance differences. Overall, the results indicate that increasing stereo width leads to comparably small improvements in accuracy (see BF for Model 2C), but only for trials where the target is present in the mix.

4.5. Integrated Discussion: Calibration Phase

Experiments 1 and 2 were designed to find suitable parameters to manipulate the test items' difficulty to precisely probe MSA abilities of both NH and HI listeners, irrespective of age, musical sophistication, and musical training. We hypothesised (1) the target-to-mix level-ratio, (2) the choice of target instrument, (3) the number of instruments in the mixture, and (4) sound localisation cues to have a substantial influence on auditory scene analysis abilities in the context of music. We demonstrated that all four parameters exert an influence on scene analysis performance. Notably, despite the diversity in participants' musical perceptual abilities, prior musical training, and degree of hearing impairment, test scores still showed a reasonable degree of variability within groups, and we observed no strong ceiling or floor effects. Conclusively, the results of the two calibration experiments established evidence that all four parameters are suitable to adequately manipulate the test-items difficulty.

Although the results identified four important parameters influencing the MSA test to a certain degree, we opted to include only the first three parameters in our test implementation. Our findings indicated only a modest effect of stereo width, i.e., ILDs and only for those items that contained the target. By excluding the ILD parameter, we aimed to enhance the adaptability of our approach for online experimental setups utilising mono audio and to circumvent potential complications for participants with asymmetric hearing impairments. Moreover, as the perception of stereo width could solely contribute to make test items less difficult, and the other parameters were even more efficient for generating less difficult items, we chose to employ a more parsimonious model comprising only the three primary factors.

Based on the findings from Experiment 1 and Experiment 2, we can infer that individual differences in abilities play a notable role in the performance of the MSA task. The predictive accuracy of the final model, including both fixed and random effects, was consistently higher than the model with fixed effects alone. In Experiments 1 and 2, the inclusion of random effects led to an increase in predictive accuracy of 7 percentage points. This suggests that accounting for individual differences is crucial for understanding and predicting MSA performance. These results have significant implications for the development of adaptive MSA testing based on Item Response Theory (IRT). By incorporating individual differences into the model, adaptive testing procedures can more efficiently estimate the underlying ability of participants and tailor the test items to their specific needs. This enables more precise and efficient measurement of MSA performance, while also reducing the likelihood of floor and ceiling effects that may be present in a one-size-fits-all approach. Furthermore, the model's strong predictive performance indicates that the selected factors adequately represent the underlying cognitive processes in an ASA task.

It is also important to acknowledge a few limitations of Experiments 1 and 2. The assessment of participants' degree of hearing impairment relied on self-report, and thus lacked accuracy. Another limitation is the online setting for testing: conducting listening tests remotely poses inherent challenges, such as potential variations in headphone quality and uncontrollable ambient noise, which can affect the reliability and accuracy of the assessments. Notably, the headphone screening test was not administered to participants with hearing impairments, leaving the presentation conditions uncontrolled. For instance, both the types of

playback devices used and compliance with instructions to remove hearing aids remain unknown, thereby further complicating the evaluation of playback conditions. Overall, the challenges associated with conducting online listening tests highlight the importance of carefully considering the limitations and potential problems of these approaches when evaluating individuals with hearing impairments. In the validation Experiment 3, we sought to address these limitations.

4.6. Experiment 3: Test Validation

The primary aim of Experiment 3 was to validate an adaptive version of the MSA test under controlled laboratory conditions. A flexible method for creating an adaptive test is the use of Item Response Theory (IRT) models. These models predict the probability of a correct response based on specific item parameters that capture different ways in which items might vary, such as item difficulty, item discrimination, and parameters for guessing and inattention. In this study, the B-GLMM used in Experiment 1 can be considered an explanatory Item Response Theory (IRT) model (Wilson & de Boek, 2004), as it describes the relationship between a person's latent trait level (i.e., the ability to identify musical instruments in a mixture) and their probability of responding correctly to test items. By incorporating item characteristics (e.g. target instrument, number of instruments in the mix, target-to-mixture level-ratio, presence of the target in the mix) and person-specific random effects, the B-GLMM offers a flexible and robust approach to modelling response data. This makes it suitable for IRT applications in adaptive testing scenarios to assess the participants' ability to identify musical instruments within complex auditory scenes. Our IRT model then allows us to establish validity estimates of the adaptive MSA by examining its correlation with other measures of related constructs, such as speech-in-noise perception, melody discrimination, and mistuning perception. In addition to establishing validity, we also sought to assess the test-retest reliability of the MSA test. One way to assess test-retest reliability within the IRT framework is to estimate the reliability using both the standard error of measurements (SEM) derived from the model, as well as by comparing scores empirically from multiple adaptive MSA measurements within a single participant. We measured participants' MSA ability twice on the same day to investigate test-retest reliability in controlled laboratory conditions, and also compared these measurements with a third set of MSA measurements, obtained through online testing on a different day in the participant's home environment.

To identify potentially important factors that might explain individual differences between test takers, we also assessed working memory, musical sophistication, and individuals' hearing thresholds as indicators of the degree of hearing impairment. This approach enables us to examine the test-retest reliability of the MSA and to pinpoint potential factors contributing to variations in participants' abilities across different musical perceptual domains. Subsequent analyses will scrutinise correlations among these diverse tests to elucidate the latent factors influencing musical perception.

4.6.1. Methods

4.6.1.1. Test battery

Adaptive Musical Scene Analysis Test (MSA). Based on the established B-GLMM of the first calibration experiment (Model F), an adaptive version of the MSA was developed. Parameter estimates for the final model are given in Table A4.1. The final IRT model resembles a classical four-parametric logistic model, in which the discrimination parameter, as well as the guessing and inattention parameter are constrained to be equal across all items. The item difficulty was estimated by using the B-GLMM model (F), which applied the random intercept mixed effects structure for participant and excerpt and by considering the fixed effects for (1) the choice of the target instrument, (2) the number of instruments in the mixture, and (3) the level-ratio between the target and the mixture - each included into the model with an interaction with (4) the presence of the target in the mixture. From these results an estimate of the level of difficulty for each of the 40 parameters combinations could be derived (see Figure 4.2D). The B-GLMM fixed effects were converted to the metric of the desired item response model. By also incorporating the random effect structure of the excerpts, we are also able to get a unique item difficulty estimate for each item. To set up the MSA task, we utilised the *psychTestR* (v2.23.0), an R package that provides the underlying testing mechanisms to create individual test packages that can further be employed in web-browser-based behavioural experiments. The adaptive version of the test was constructed using the *psychTestRCAT* (v1.6.0) package, which uses weighted-likelihood ability estimations of participant ability that range approximately from -3 to +3. The discrimination parameter was set to the standard deviation of the participant intercept (that is, the estimated median of the person random effect structure). The adaptive item selection procedure of the MSA is constructed according to Urry's rule (as cited in Magis & Gilles, 2012). There, moderately difficult items are presented first to estimate participants' ability level using item response theory (de Ayala, 2009). The first items eligible for selection have a difficulty level close to the item bank's mean difficulty, falling within a range of one standard deviation. Subsequently, more difficult items are presented to participants with higher abilities and easier items to participants with lower abilities. After each response, the participant's ability estimates are recalculated, and the next item selected is one that is closely aligned with their updated ability level. The same item is never presented twice. In this experiment, the test-retest reliability was established from the results of the adaptive test with a length of 30 items. For this experiment the MSA-version 2.4 was used.

Oldenburger Satztest (OLSA; Kollmeier et al., 2015; Wagener et al., 1999). The OLSA is an adaptive speech-in-noise perception test. The participants' task is to verbally repeat five-word sentences, spoken by a male speaker, embedded in speech-shaped white noise. The sentences were randomly generated from a predefined word class structure (name-verb-numeral-adjective-object), with each word class drawn from a pool of 10 possible word alternatives - resulting in sentences that were unpredictable for the listeners. While the masker level remained constant, the speech level was adaptively adjusted to determine the individual 50% speech-reception thresholds (SRT). To increase reliability, the composite score of two test lists of 20 sentences each were used. The OLSA is essential for our experiment as it provides insight into the participant's speech perception abilities in a

challenging listening environment, reflecting their ASA skills in the speech domain – measured by a well-established adaptive test. A relevant relationship between OLSA and MSA performance would support the notion that both tasks tap into similar auditory scene analysis abilities.

Frequency discrimination task (FDT). In this test, participants are presented with a series of three tones, one of which differs in frequency, and asked to identify the odd-one out. Feedback on the correct response is provided on every trial. Using an adaptive 2-down-1-up procedure, the test score is calculated from two test-sets of 40 items each, using the geometric mean of the last six reversal points within each set. In the context of the MSA task, the ability to discriminate frequencies allows participants to identify and segregate auditory streams based on their frequency content and thus helps to perceive and distinguish individual instruments within a complex musical scene. It is expected therefore that there would be a relationship between the FDT and MSA scores.

Melodic discrimination test (MDT; Harrison et al., 2017). For each question of this adaptive melodic working memory test, participants hear three versions of the same unfamiliar melody. Each successive version is transposed to a semitone higher in pitch and in one of these versions, a note has been altered. The task is to detect the 'odd one out', while ignoring the transposition in pitch. By utilising weighted-likelihood ability estimations, the MDT uses a similar adaptive item selection procedure to the MSA. The default number of 20 items was used in this experiment. The MDT is important, because it evaluates a participant's working memory and their ability to discriminate auditory information, which relates to stages 2 and 3 of the MSA cognitive model. A strong correlation between MDT and MSA performance would indicate that working memory plays a role in the MSA task.

Mistuning perception test (MPT; Larrouy-Maestri et al., 2019). This adaptive test assesses the ability to perceive mistuning in pieces of music. The task is to decide whether a vocalist is in tune or out of tune with the background music. The main output from the MPT is an ability score, corresponding to the ability estimate for the participant similar to the procedure mentioned for the MDT and MSA. The default number of 30 items was used in this experiment. As the MPT assesses a participant's ability to identify pitch deviations, it can inform their ability to perceive and segregate auditory streams (similar to the FDT), highlighting participants' sensitivity to tuning discrepancies. A relevant relationship between MPT and MSA performance would suggest that sensitivity to mistuning, that is the sensitivity to frequency differences in the context of music, may contribute to the MSA task.

Timbre perception test (TPT, Lee & Müllensiefen, 2020). This test examines timbre perception abilities. Participants use a slider to best reproduce presented stimuli that vary along three important dimensions of timbre: temporal envelope, spectral flux, and spectral centroid. For each dimension a total of six items was presented. The aggregated final estimate represents the average score of all three blocks, ranging from 1 – 100 (higher scores indicate better performance). The TPT directly relates to the participants' ability to perceive and segregate auditory streams based on timbre, a key aspect of the MSA task. A strong correlation between TPT and MSA performance would provide evidence for the importance of timbre perception in the MSA task.

Computerised Adaptive Beat Alignment Test (CA-BAT; Harrison & Müllensiefen, 2018). The CA-BAT is an adaptive test of beat perception ability. This 2-AFC test assesses the ability to recognise the beat in a piece of music. Participants are presented with excerpts where they hear a piece of music together with a click track. The task is to decide whether the click track is on or off the beat in the music. The main output from the CA-BAT is an ability score, computed from the underlying item response model and corresponds to the ability estimate for the participant (similarly to the MDT). The full 25-item test with identical psychometric parameters and adaptive procedure to the original study was used. By making sense of the temporal structure of musical pieces, beat perception can provide the auditory system with prior knowledge about upcoming musical events, thereby guiding stream segregation. In this sense, the ability to perceive the beat helps listeners to organise auditory streams, making it easier to understand and follow the musical structure. A substantial relationship between CA-BAT and MSA performance would support the role of beat perception in the MSA task.

Backwards Digit Span Memory Test (BDS; e.g., Weiss et al., 2016, Talamini et al., 2017). In this classical working memory test, participants remember sequences of digits. A test person is visually presented with digits one after the other and is then asked to recall the digits in the reverse order (e.g., for the sequence 1 2 3 4, the correct answer is 4 3 2 1). In total, two sequences with four digits (i.e., 2x4), 2x5 digits, 4x6 digits, and 4x7 digits were used. The final score represents the proportion of correctly recalled sequences. A strong correlation between BDS and MSA performance would further emphasise the importance of working memory in the MSA task.

Goldsmith Musical Sophistication Index (GMS; Müllensiefen et al., 2014). Instead of employing only the musical training and perception scale as in the calibration phase, the complete index (in German language) was used during the validation experiment. The final composite score ranging from 1 to 7, with 7 being the highest possible score, was generated for each subscale. A robust association between MSA performance and GMS scores would highlight the critical role of listeners' musical sophistication in ASA, whereas a relationship to the subscales that assess musical training and musical perceptual abilities would establish evidence in favour of the predictive validity of the test.

Pure-tone-average audiometry (PTA). For assessing the individuals' degree of hearing-impairment, pure-tone audiometric thresholds were measured with an Interacoustics AD528 portable audiometer. The audiometer was used to measure thresholds at .125, .25, .5, 1, 2, 4, and 8 kHz using the standard clinical ascending-descending procedure in 5 dB steps. Considering that elevated hearing thresholds in frequencies above 4kHz have been shown to be meaningful in domains important for complex ASA listening scenarios (e.g., amplitude modulation detection thresholds and sound localisation; see e.g., Moore, 2020; Narne et al., 2023) the PTA score was calculated as an average over all measured frequencies. When the PTA of the better ear exceeded 20 dB HL, participants were classified as HI (Humes, 2019). A relevant relationship between PTA and MSA performance would imply that hearing sensitivity plays a role in musical scene analysis.

4.6.2. Procedure

Ethical approval was provided by the Ethics Committee at University of Oldenburg. All participants provided written informed consent. The test was administered in two separate parts. The first part was conducted in a controlled laboratory setting, where participants were seated in a soundproof booth. The calibrated equipment consisted of a computer, an RME Babyface soundcard, and Sennheiser HD650 headphones. The long-term sound level was set to 75 dB SPL (A), measured with a Norsonic Nor140 sound-level metre using music-shaped noise as the excitation signal.

The first part of the test battery proceeded in the following order:

Block 1:

- a. Pure-tone-audiometry
- b. Speech-in-noise (OLSA)
- c. Frequency discrimination task

Block 2:

- a. Demographics questionnaire
- b. Huggins headphone screening
- c. Mandatory pause of at least five minutes
- d. Training phase of the MSA
- e. First set of the adaptive MSA with 30 items

Block 3:

- a. Melodic discrimination test
- b. Mistuning perception task
- c. Beat perception task
- d. Second set of the adaptive MSA with 30 items.

In between each of the assessment blocks, participants were given brief breaks that included verbal interactions and instructions from the experimenter. In addition, participants were encouraged to take as many breaks as necessary in-between tests to ensure that they were able to complete the tasks to the best of their ability. The order of the tests was carefully designed to maximise the effectiveness of the testing, taking into account aspects like test difficulty, potential learning effects, and cognitive fatigue. Each task was administered according to established protocols.

The second part of the test battery was performed online by the same group of participants, at least 24 hours after the first part. After adjusting their volume to a loud but comfortable level, participants completed the following tests:

- a. Full Gold-MSI self-report questionnaire
- b. Third set of the adaptive MSA with 30 items
- c. Backwards digit span test
- d. Timbre perception test

This online component of the test battery allowed us to assess participants' MSA abilities in an online setting. The laboratory part of the study took approximately 80-120 minutes, whereas the online part took between 25-40 minutes. Because of missing data, some

of the following analysis was conducted based on a subgroup of participants, which will be specified accordingly.

4.6.3. Participants

The experiment included a total of 74 participants (32 males and 42 females). The participants were divided into four groups: 30 older adults with NH ($M = 63.2$; $SD = 7.2$), 19 older adults with HI ($M = 70.5$; $SD = 7.2$), 24 younger adults with NH ($M = 25.4$; $SD = 4.2$), and one young adult with HI (Age = 26). A graphical illustration of individuals' pure-tone audiometric thresholds can be found in Figure A4.1. Participants were recruited via a call for participation in the local newspaper in Oldenburg, Germany and were compensated for their time (10€ per hour).

4.6.4. Results

4.6.4.1. Test-retest reliability

The two-part, two-day administration design of the study allowed us to assess the consistency of MSA performance over time and compare it in different playback environments. The test-retest reliability of the adaptive MSA was assessed using the intraclass correlation coefficient (ICC; see Koo & Li, 2016), using a two-way mixed-effects model with absolute agreement. For single measurements, which represent the reliability when the adaptive MSA was administered once, a moderate ICC(A,1) of .633 was observed (95% CI [.46; .757], $F(71,54.9) = 4.78$, $p < .001$; Pearson's $r(70) = .67$, $p < .001$). In contrast, the ICC for the mean of multiple measurements (that is the average MSA score when the test was administered twice), improved to a substantial level with an ICC(A,2) of .775 (95% CI [.625; .863], $F(71,51.2) = 4.78$, $p < .001$). Accordingly, the ICC indicates moderate to good consistency across the two sets of 30 items when administered under controlled conditions on the same day (see Figure 4.5A). The results further suggest that the MSA shows moderately accurate and stable reliability values, even when comparing the combined MSA score of the controlled laboratory and the uncontrolled online environment (ICC(A,1) = .60, CI = [.42; .73], $p < .001$; Pearson's $r(66) = .65$, $p < .001$). As the length of the test increased, the model indicated a consistent decrease in the mean estimated standard error of measurement (SEM) for the MSA ability. This trend converged in a final SEM of .29 for the full test of 30 items, as illustrated in Figure 4.5B. Averaged across all participants for whom complete data were available ($n = 65$), the median scores for the first, second, and third MSA item sets were .21 (95% CI = [.17; .26]), .28 (95% CI = [.22; .33]), and .33 (95% CI = [.24; .38]), respectively. A repeated measures ANOVA was conducted to assess the impact of training on test scores across three MSA test sets. The results suggest a small improvement in test scores due to training effects ($F(2, 128) = 3.26$, $p = .042$, $\eta^2 = .014$).

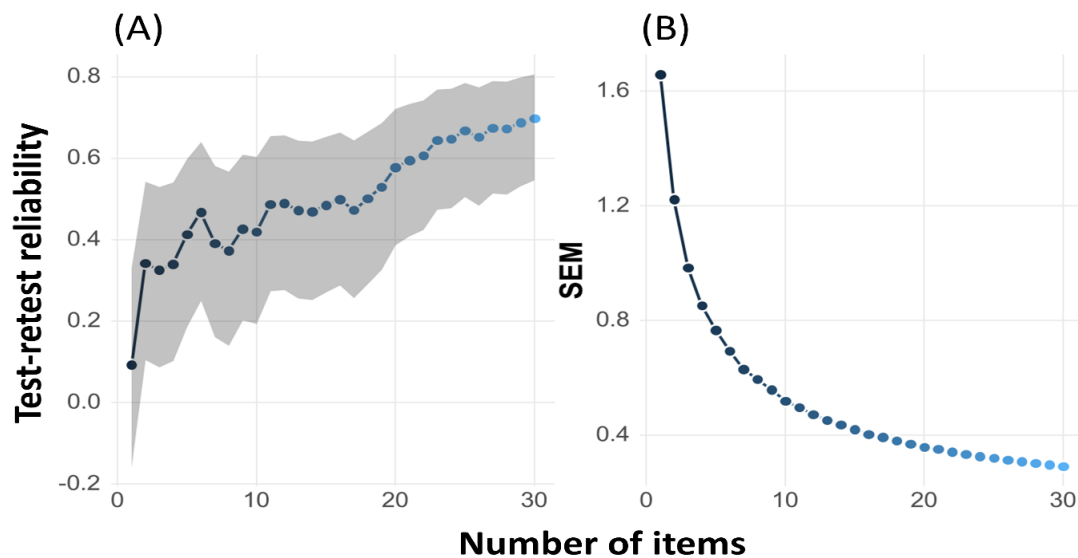


Figure 4.5 Test-retest-reliability of the first and second set of the MSA as a function of test length is shown on the left and the standard error of measurement (SEM) as a function of test length on the right. The test-retest score (on the left) was assessed empirically and is based on the Pearson's product-moment correlation of all participants. The SEM (on the right) was derived from the model ability estimate.

4.6.4.2. Music-related and psychoacoustic tests

In the presented experiment, we found moderate correlations between the MSA test and timbre ($r = .37$), mistuning ($r = .52$), and beat perception ($r = .46$) scores (see Table 4.3). Notably, the strongest correlation was identified between the MSA and melody discrimination ($r = .52$). In addition, a moderate correlation was found between the MSA and speech-in-noise measures (i.e., SRT of the OLSA, $r(63) = -.39$, $p < 0.001$) and after removing two severe outliers (that were minus three standard deviations from the median), the frequency discrimination task revealed a moderate association with the adaptive MSA ($r(61) = -.43$, $p < .001$).

4.6.4.3. Individual differences factors

Similar to the calibration experiments, we found a negative relationship between the degree of hearing impairment (measured by the individuals' better ear elevated hearing thresholds) and MSA scores ($r = -.37$). We also found a negative correlation between age and MSA scores ($r = -.40$), however, both effects were confounded, as indicated by the correlation between age and hearing impairment ($r(63) = .66$, $p < .001$). Interestingly, MSA scores showed the highest correlations with age and pure-tone audiometric scores compared to all other employed music-related perceptual tests in the battery (see Figure A4.2 for a complete correlation matrix). This relationship gives first indication of a potential usage of the MSA as a relevant tool for diagnosing and profiling hearing-impairment in a musical context. Nevertheless, the highest relationship with PTA scores could be observed with the OLSA speech-reception thresholds ($r = .59$; here a positive correlation implies a higher degree of hearing impairment leads to worse performance in the OLSA task). Participants' performance on the MSA tests demonstrated a notable correlation with BDS ($r = .39$), suggesting a moderate association with general working memory capabilities. Our findings further indicate a positive correlation

between musical training and musical perception abilities with ASA abilities. Specifically, the correlation between the MSA test and musical training score was .46, and .37 for musical perception. The strongest correlations with the general musical sophistication score was found with mistuning perception scores ($r = .65, p < .001$), melody discrimination abilities ($r = .53, p < .001$), followed closely by the MSA ($r = .48$).

Table 4.3

The correlation of the adaptive MSA with selected tests from the test battery

	MSA online	SRT	FDT	MDT	CABAT	MPT	TPT
N	66	72	72	72	72	72	66
Correlation	.65**	-.45**	-.39**	.52**	.46**	.52**	.37*
95% CI	[.49, .77]	[-.61,-.24]	[-.57,-.18]	[.33,.67]	[.26,.62]	[.34,.67]	[.14,.56]

Table 4.3 continued

	BDS	GMS general	GMS perceptual	GMS training	Age	PTA
N	66	66	66	66	72	72
Correlation	.36*	.46**	.37*	.46**	-.4**	-.37*
95% CI	[.13,.5 5]	[.24,.63]	[.15,.56]	[.25,.63]	[-.57,-.18]	[-.55,-.15]

Note. Sig. is denoted as *at $p < .01$, ** $p < .001$ (Pearson's correlation). SRT corresponds to the speech-reception thresholds of the OLSA. For a complete correlation matrix see Figure A4.2.

4.6.5. Discussion

The aim of Experiment 3 was to assess the psychometric properties of the adaptive version of the MSA test. Two reliability measures were calculated: the standard error of measurement (SEM) based on the ability estimate of the IRT model and the intra-class test-retest reliability. Both measures help to establish a comprehensive picture of reliability, as the SEM does not require participants to take the test twice, whereas the test-retest reliability does not rely on any model assumptions. Ideally, the optimal test length would be determined by the point at which both test-retest reliability and the standard error of measurement (SEM) estimates plateau. However, in our study, this was not achieved with a test length of 30 items. The final test-retest reliability was acceptable and comparable to other previous measures of scene analysis ability (e.g., Kirchberger & Russo, 2015; Siedenburg et al., 2020). To further improve

the test-retest reliability estimate, it is likely that increasing the test length would be necessary, although this remains to be seen in further empirical research.

To establish validity estimates, it is important to compare the MSA to other established tests that measure similar constructs. Unfortunately, there are no published tests specifically designed to assess ASA abilities with realistic musical excerpts. As a result, we must rely on comparisons with hypothetically associated psychoacoustic and music-related tests. The presented experiment found moderate correlations between the MSA test and various other music-related tests, such as timbre, mistuning, and beat perception, with the strongest correlation found with melody discrimination. The relationship between the MSA and these tests are likely due to shared underlying cognitive processes and sensory representations. For example, within the mistuning perception task, listeners need to have the ability to separate the vocal line from the accompaniment through auditory stream segregation, and then assess the extent to which its pitch content conforms to the prototypical pitch distributions of the relevant musical style (Larrouy-Maestri et al., 2019). Just as in mistuning perception, the MSA task requires the separation of streams in the mixture. This similarity is also reflected in the comparable strengths of the relationship of the participants' frequency discrimination scores to the MSA and those with MPT scores (see Figure A4.2). The moderate correlation between the MSA task and the frequency discrimination task also highlights that the MSA task is reliant on rather fundamental acoustical cues like pitch. Nonetheless, it also requires the ability to integrate multiple auditory cues and is affected by higher-level cognitive processes such as attention and memory. Accordingly, MSA and CABAT (beat perception) share the requirement to process and analyse temporal aspects of musical sound, specifically the rhythms and patterns within a piece.

Another important result is the substantial association between the MSA and the MDT, which assess the ability to discriminate between melodies. Melody discrimination abilities are often interpreted as reflecting more general cognitive traits and involve processes that are similar to those needed in the process of auditory scene analysis: in both tasks musical streams have to be perceptually encoded, stored in memory (including melodic and timbral memory), then compared to each other, followed by a decision-making process. Timbre perception, on the other hand, is a multidimensional attribute which characterises the ability to discriminate between musical sounds, even when sounds are equal in loudness, tempo, and pitch. Accordingly, timbre plays a key role in the recognition of sound sources. Both the TPT and the MSA both require the ability to identify and differentiate musical sounds based on their unique attributes.

In essence, the observed moderate relationships between the MSA test and specific music-related perceptual tasks - namely timbre identification, beat perception, melody discrimination, and mistuning perception - suggest that the MSA test measures similar but not the same cognitive and acoustic processing abilities. This provides compelling evidence for the task's convergent validity. Besides that, the moderate correlation between the MSA task and a speech-in-noise measure further supports its construct validity as both concepts measures the ability to process and parse complex auditory information (speech vs. music), separate relevant sounds from competing sounds (white noise vs. a mixture of instruments), and identify and interpret different elements within an acoustic scene. To further establish

the validity of the MSA task, future research could focus on demonstrating its predictive validity. For instance, researchers could compare the scores of highly trained musicians who are assumed to possess excellent auditory scene analysis skills (such as conductors) with those of musically untrained listeners. This would provide additional support for the MSA task as a valid measure of auditory scene analysis in music.

4.7. General Discussion

The objective of the current study was to establish a new test for ASA abilities in the context of music. Previous efforts to develop tests to assess musical perceptual abilities (Kirchberger & Russo, 2015; Siedenburg et al., 2020) were limited by relying on artificial stimuli and/or not being able to account for individuals with a broad range of perceptual abilities and diverse auditory profiles (e.g., hearing abilities, musical training, age, and cognitive abilities). In light of these limitations, our study aimed to develop an efficient and accessible test for ASA abilities in music. To accomplish this goal, we conducted three experiments.

In the first two experiments, we found that the target instrument, number of instruments in the mixture, target-to-mix level-ratio, and presence of localization cues (perception of stereo) had a relevant impact on an individual's ability to analyse and understand musical scenes. These results indicate that the specific characteristics of stimuli used in ASA tasks can greatly influence an individual's ability to analyse and understand musical scenes. To standardise the MSA test, we conducted a calibration (Experiment 1 and Experiment 2) with a sample of participants with varying degrees of hearing impairment and different musical backgrounds. In Experiment 3, we built an IRT model based on the first three parameters of the already established B-GLMM. This allowed us to develop a reliable assessment tool for evaluating individuals' musical auditory scene analysis abilities. Our findings showed consistent MSA results when administering the test to the same individuals under different presentation conditions or to different individuals under the same conditions. We observed a moderate correlation between MSA and various aspects of musical perception, such as melody discrimination, timbre perception, mistuning perception, and beat perception. These findings emphasise the validity of the assessment and highlight the importance of considering ASA in understanding the interplay of the different facets of musical perception.

We also wanted to explore factors that might explain individual differences among test-takers, including musical background, working memory capacity, age, and degree of hearing impairment. Numerous studies have demonstrated that musical training or musical sophistication not only enhances basic perceptual abilities, such as pitch, timbre, and rhythm detection (Kannyo & DeLong, 2011), but also contributes to higher-order perceptual skills. These include musical imagery (Gelding et al., 2021), mistuning perception (Larrouy-Maestri et al., 2019), auditory tracking (Madsen et al., 2019), and auditory streaming segregation (Zendel & Alain, 2009). The results from Experiment 3 align with this literature by showing a moderate relationship with musical scene analysis abilities. On the contrary, however, Bürgel et al. (2021) found no systematic difference between musicians and non-musicians using a similar instrument detection paradigm, which they attributed to the limited representation of musicians in their sample. Here, we recruited participants with a wide range of musical ability

levels. We nonetheless observed only small correlations of musical training levels with MSA abilities in Experiments 1 and 2, but moderate correlations in Experiment 3, which remains intriguing. A potential explanation for these incompatible findings is the way in which the musical training score of the Gold-MSI is operationalized: the musical training score is more likely assessing a 'lifetime' training score, with the reported musical training potentially dating back many decades (e.g., "I have had formal training in music theory for __ years"). This could mean that someone who has intensively trained in music in the past 10 years may receive the same score as someone who played music extensively for 10 years, but has not played for the last 30 years - reducing the effect of prior musical training on musical perception performances due to decay. Recent musical training, in contrast, might have a stronger relationship with auditory scene analysis abilities than 'lifetime' musical training. Micheyl et al. (2006), for instance, showed that several hours of active training in a pitch discrimination task led non-musicians to achieve similar accuracies as musicians. This suggests that recent musical training might have a comparatively greater impact on ASA abilities. Accordingly, Experiment 1 and 2 might have included 'lifetime' musicians, who have not played considerably in the recent past, while a higher proportion of actively performing musicians participated in the validation experiment. Certainly, this discrepancy must be addressed in future research.

Furthermore, our findings revealed a notable negative correlation between age and MSA scores and a negative association with the degree of hearing impairment. However, similar to Kirchberger and Russo's (2015) study, the effect of the degree of hearing impairment was confounded by the factor of age. This makes it difficult to determine whether the observed correlation between the degree of hearing impairment and MSA performance is due to the hearing impairment itself or factors associated with ageing. Goossens et al. (2017) used a matched-pair design to disentangle the effects of age and hearing impairment on speech perception. Participants were matched on age and then divided into groups based on the degree of hearing impairment, which allowed for a direct comparison of speech-perception in individuals with similar levels of age but different levels of hearing impairment. Goossens et al. (2017) also excluded participants with potential cognitive impairments. The results indicated that even when audiometric thresholds are within normal limits and individuals showed no indication of even mild cognitive impairment, masked speech perception declines by middle age and decreases further with increasing age. However, since the prevalence of hearing-impaired participants is particularly low among the young age group, we were not able to achieve a fully matched-pair design. This is an area for future research. Overall, our findings converge with the notion that even fundamental aspects of music perception such as ASA appears to be affected by working memory, age, and the degree of hearing impairment (cf., Micheyl et al., 2013; Susini et al., 2023).

A notable limitation of the current study is its cultural specificity. The MSA test and its underlying theoretical frameworks are largely rooted in Western-centric music perception research. The chosen stimuli (base tracks and genre), target instruments, and the sample of participants mainly cater to Western audiences. This inherent cultural bias raises questions about the test's cross-cultural applicability and limits the generalizability of our findings. While it may be tempting to aim for a universally applicable MSA test, due to the complexity

and diversity of musical experiences and traditions across cultures, it is more realistic and valid to develop culture-specific versions. For instance, a culture-specific version for Indian music could be developed by sourcing a multi-track database of Indian musical compositions and calibrating the test with Indian listeners. Although these culture-specific tests would feature different stimuli and item banks, they would remain conceptually comparable at the level of the latent variable - MSA ability. Such endeavours would make significant contributions to the field by broadening the scope of music perception research to be more inclusive and representative of global musical experiences (e.g., Jacoby et al., 2020).

Apart from cultural aspects, the test's ecological validity also merits acknowledgment. While the MSA test employs realistic musical stimuli, it should be noted that the listening conditions are nonetheless constrained by the experimental context. Specifically, the stimuli are presented over headphones in a controlled setting (and in a monaural configuration). This may not fully capture the complexities of auditory scene analysis as experienced in more naturalistic settings, such as live concerts, where various acoustic and attentional factors come into play. Therefore, the term 'ecologically valid' is used here with the understanding that the test conditions are a simplification of real-world musical experiences.

In summary, the newly developed adaptive MSA test can serve as a valuable contribution to music perception research by efficiently measuring ASA abilities in a realistic musical context for individuals with a broad range of dispositions. The MSA test is sensitive to the effects of the degree of hearing impairment (i.e., individuals with elevated hearing thresholds) and musical background (musical sophistication and training) but is not limited by it. Future extensions of the current item bank could incorporate excerpts of classical music or excerpts from other musical genres and cultures. The MSA test is an easy-to-implement, user-friendly, enjoyable, and efficient tool for evaluating ASA performances in the context of music. With a training and test time of fewer than 8 minutes, it is quick to administer and allows for flexibility with adjustable test length (alongside known measurement accuracies) to meet the specific needs of the study or application. Additionally, it is open-source and freely available. Detailed information on the operational R package, including an online demo, can be found in our GitHub repository (<https://github.com/rhake14/MSA>). By studying ASA in the context of music, researchers can gain insights that can inform the development of new technologies and techniques for improving music perception in individuals with hearing impairment.

4.8. Declarations

Author contributions. RH, MB, DM, and KS designed the study. MB and RH compiled the stimuli. RH programmed the experimental task, collected the data with assistance from AG, and wrote the first manuscript draft. Experiment 2 was part of the BSc thesis of NKN. RH and DM analysed the data. All co-authors edited and approved the final version of the manuscript.

Ethics approval. Ethical approval was provided by the Ethics Committee at University of Oldenburg (Drs.EK/2019/092) and University of Leeds (FAHC 21-030).

Consent to participate. Informed consent was obtained from all participants in this study, permitting the collection of anonymized data and its potential publication in journal articles.

Acknowledgements. This work has been funded by the Freigeist Fellowship of the Volkswagen Foundation to KS. Special acknowledgement is extended to Klaus Frieler (MPI for Empirical Aesthetics) for his invaluable assistance with the implementation of the MSA as an R package and for his general technical support.

Conflict of interest. The authors declare that there is no conflict of interest.

Copyright. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Open Practices Statements. All program code of the final test (open-source version of the adaptive MSA) is available online at the reported URL (<https://github.com/rhake14/MSA>). All further data and program code will be made available on request.

4.9. References

- Akeroyd, M. A., Gatehouse, S., & Blaschke, J. (2007). The detection of differences in the cues to distance by elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 121(2), 1077–1089.
- Bayat, A., Farhadi, M., Pourbakht, A., Sadjedi, H., Emamdjomeh, H., Kamali, M., & Mirmomeni, G. (2013). A comparison of auditory perception in hearing-impaired and normal-hearing listeners: an auditory scene analysis study. *Iranian Red Crescent medical journal*, 15(11), e9477. <https://doi.org/10.5812/ircmj.9477>
- Bey C., & McAdams S. (2002). Schema-based processing in auditory scene analysis. *Percept. Psychophys.* 64, 844–854.
- Bidelman, G. M., & Yoo, J. (2020). Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. *Frontiers in Psychology*, 11(August), 1–11. <https://doi.org/10.3389/fpsyg.2020.01927>
- Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., & Bello, J. P. (2014). MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research. In *15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan.
- Bittner, R., Wilkins, J., Yip, H., & Bello, J. (2016). MedleyDB 2.0: New Data and a System for Sustainable Data Collection. *International Conference on Music Information Retrieval (ISMIR-16)*, New York, NY, USA.
- Blauert, J., & Braasch, J. (2008). Räumliches Hören. In: Weinzierl, S. (eds) *Handbuch der Audiotechnik*. VDI-Buch. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-34301-1_3
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244–249. <https://doi.org/10.1037/h0031163>
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press).
- Bürgel, M., Picinali, L., & Siedenburger, K. (2021). Listening in the Mix: Lead Vocals Robustly Attract Auditory Attention in Popular Music. *Frontiers in Psychology*, 12(December). <https://doi.org/10.3389/fpsyg.2021.769663>
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Cai, Y., Zhao, F., & Zheng, Y. (2013). Mechanisms of music perception and its changes in hearing impaired people. *Hearing, Balance and Communication*, 11(4), 168–175. <https://doi.org/10.3109/21695717.2013.839356>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guilford Press.
- Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer bayesian analyses over significance testing. *Psychon. Bull. Rev.*, 25, 207–218. doi: 10.3758/s13423-017-1266-z
- Düvel, N., & Kopiez, R. (2022). The paired A–Not A design within signal detection theory: Description, differentiation, power analysis and application. *Behavior Research Methods*, 54(5), 2334–2350.. <https://doi.org/10.3758/s13428-021-01728-w>
- Embretson, S. E. Construct validity: Construct representation versus nomothetic span. *Psychol. Bull.* 93, 179–197 (1983).
- Emiroglu, S., & Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain Research*, 1220, 199–207. <https://doi.org/10.1016/j.brainres.2007.08.067>

- Garami, L., Chow, R., Fakuade, A., Swaminathan, S., & Alain, C. (2020). Orienting Attention to Auditory and Visual Short-term Memory: The Roles of Age, Hearing Loss, and Cognitive Status. *Experimental Aging Research*, 46(1), 22–38.
<https://doi.org/10.1080/0361073X.2019.1693008>
- Gelding, R. W., Harrison, P. M. C., Silas, S., Johnson, B. W., Thompson, W. F., & Müllensiefen, D. (2021). An efficient and adaptive test of auditory mental imagery. *Psychological Research*, 83(3), 1201–1220. DOI: <https://doi.org/10.1007/s00426-020-01322-3>.
- Gordon-Salant, S., & Cole, S. S., (2016). Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. *Ear Hear*, 37(5):593-602. doi: 10.1097/AUD.0000000000000316
- Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, 344, 109–124. <https://doi.org/10.1016/j.heares.2016.11.004>
- Greasley, A. E., Crook, H., & Fulford, R. J. (2020) Music listening and hearing aids: perspectives from audiologists and their patients. *International Journal of Audiology*, 59(9), 694–706.
- Greasley, A. E. (2022). Diverse Music Listening Experiences: Insights from the Hearing Aids for Music project. In J. Drever & A. Hugill. *Aural Diversity* Routledge (pp. 134-142).
- Harrison, P. M. C. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*.
<https://doi.org/10.21105/joss.02088>
- Harrison, P. M. C., & Müllensiefen, D. (2018). Development and validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). *Scientific Reports*, 8(12395), 1–19.
<https://doi.org/10.1038/s41598-018-30318-8>
- Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7, 1–18.
<https://doi.org/10.1038/s41598-017-03586-z>
- Humes, L. E. (2019). The World Health Organization’s hearing-impairment grading system: an evaluation for unaided communication in age-related hearing loss. *International Journal of Audiology*, 58(1), 12–20. <https://doi.org/10.1080/14992027.2018.1518598>
- Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravignani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-Cultural Work in Music Cognition: Challenges, Insights, and Recommendations. *Music perception*, 37(3), 185–195. <https://doi.org/10.1525/mp.2020.37.3.185>
- Kannyo, I., & DeLong, C. M. (2011). The effect of musical training on auditory perception. *Proceedings of Meetings on Acoustics*, 14(May 1994).
<https://doi.org/10.1121/1.4733850>
- Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, 36(2), 217–228. <https://doi.org/10.1097/AUD.0000000000000112>
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, 5, 3–16.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, 51(2), 663–675.
<https://doi.org/10.3758/s13428-019-01225-1>

- Lee, H., & Müllensiefen, D. (2020). The Timbre Perception Test (TPT): A new interactive musical assessment tool to measure timbre perception ability. *Attention, Perception, & Psychophysics*, 82(7), 3658–3675. <https://doi.org/10.3758/s13414-020-02058-3>
- Lentz, J. J., Humes, L. E., & Kidd, G. R. (2022). Differences in Auditory Perception Between Young and Older Adults When Controlling for Differences in Hearing Loss and Cognition. *Trends in Hearing*, 26, 233121652110661. <https://doi.org/10.1177/23312165211066180>
- Litovsky, R. Y., Goupell, M. J., Fay, R. R., & Popper, A. N. (2021). *Binaural hearing*. Springer. Chicago
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing AIDS. *Trends in Hearing*, 18(0), 1–29. <https://doi.org/10.1177/2331216514558271>
- Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, 9(1), 1–10. <https://doi.org/10.1038/s41598-019-46728-1>
- Magis, D., & Gilles, R. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 48(8), 1–31.
- Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. (2019a). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, 4(40), 1541. doi:10.21105/joss.01541
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdtke, D. (2019b) Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, 10, 2767. doi:10.3389/fpsyg.2019.02767
- Marsman, M., & Wagenmakers, E. J. (2017). Three Insights from a Bayesian Interpretation of the One-Sided P Value. *Educational and Psychological Measurement*, 77(3), 529–539. <https://doi.org/10.1177/0013164416669201>
- MathWorks Inc. (2020). *MATLAB* (Version 7.10.0, R2020a) [Computer software]. Natick, Massachusetts.
- McKay, C. M. (2021). No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review. *Trends in Hearing*, 25, 233121652098567. <https://doi.org/10.1177/2331216520985678>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1–2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Micheyl, C., Hanson, C., Demany, L., Shamma, S., & Oxenham, A. J. (2013). Auditory stream segregation for alternating and synchronous tones. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1568–1580. <https://doi.org/10.1037/a0032241>
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Billig, A. J., Chait, M., & Milne, A. (2020). An online headphone screening test based on dichotic pitch. *BioRxiv*, 1–31. <https://doi.org/10.1101/2020.07.21.214395>
- Moore, B. C. (2020). The importance of high-frequency hearing in a medico-legal context. *Journal of the Acoustical Society of America*, 148, 2542–2542. <https://doi.org/10.1121/1.5147057>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9, e89642.
- Narne, V. K., Jain, S., Bharani, Ravi, S. K., Almudhi, A., Krishna, Y., & Moore, B. C. J. (2023). The effect of recreational noise exposure on amplitude-modulation detection, hearing sensitivity at frequencies above 8 kHz, and perception of speech in noise. *The Journal of the Acoustical Society of America*, 153(5), 2562. <https://doi.org/10.1121/10.0017973>

- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.*, 30, 653–661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, 78(3), 321–338. <https://doi.org/10.1007/s00426-013-0534-4>
- Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. In B. Parkin (Ed.), *Real-World Applications in Cognitive Neuroscience* (1st ed.). 10.1016/bs.pbr.2020.06.005
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Siedenburg, K., Röttges, S., Wagener, K. C., & Hohmann, V. (2020). Can You Hear Out the Melody? Testing Musical Scene Perception in Young Normal-Hearing and Older Hearing-Impaired Listeners. *Trends in Hearing*, 24. <https://doi.org/10.1177/2331216520945826>
- Siedenburg, K., Goldmann, K., & van de Par, S. (2021). Tracking Musical Voices in Bach's The Art of the Fugue: Timbral Heterogeneity Differentially Affects Younger Normal-Hearing Listeners and Older Hearing-Aid Users. *Frontiers in Psychology*, 12(608684). doi.org/10.3389/fpsyg.2021.608684.
- Stecker, G. C., & Gallun, F. (2012). Binaural hearing, sound localization, and spatial hearing. In K. K. Tremblay & R. F. Burkard (Eds.), *Translational perspectives in auditory neuroscience: Normal aspects of hearing* (pp. 383-433). Plural Publishing.
- Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *Am. J. Polit. Sci.*, 57, 748–761. doi: 10.1111/ajps.12001
- Susini, P., Wenzel, N., Houix, O., & Ponsot, E. (2023). Psychophysical characterization of auditory temporal and frequency streaming capacities for listeners with different levels of musical expertise. *JASA Express Letters*, 3(8).
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE*, 12(10), 1–21. <https://doi.org/10.1371/journal.pone.0186773>
- Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *The South African journal of communication disorders = Die Suid-Afrikaanse tydskrif vir Kommunikasieafwykings*, 58, 19–47. <https://doi.org/10.4102/sajcd.v58i1.38>
- van der Linden, W. J., & Glas, G. A. W. (2000). Computerized adaptive testing: Theory and practice. *Kluwer Academic*, Dordrecht, Netherlands.
- van der Linden, W. J., & Glas, C. A. W. (2007). Statistical aspects of adaptive testing. *Handb. Stat*, 26, 801–838.
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413-1432. doi:10.1007/s11222-016-9696-4
- Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2019). Pareto smoothed importance sampling. arXiv preprint arXiv:1507.02646.
- Vinay, Sandhya, & Moore, B. C. J. (2020). Effect of age, test frequency and level on thresholds for the TEN(HL) test for people with normal hearing. *International Journal of Audiology*, 59(12), 915–920. <https://doi.org/10.1080/14992027.2020.1783584>
- von Berg, M., Steffens, J., Weinzierl, S., & Müllensiefen, D. (2021). Assessing room acoustic listening expertise. *The Journal of the Acoustical Society of America* 150(4), 2539–2548. <https://doi.org/10.1121/10.0006574>.

- Wagener, K. C., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38(1), 5–14.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361–375.
doi:10.1111/j.1745-3984.1984.tb01040.x.
- Weiss, G. L., Saklofske, H. D., Holdnack, A. J., & Prifitera, A. (2016). WISC-V Assessment and Interpretation. Elsevier. <https://doi.org/10.1016/C2012-0-00415-2>
- Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Response Models: Statistics for Social Science and Public Policy* (pp. 43-74). Springer. https://doi.org/10.1007/978-1-4757-3990-9_2
- Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8), 1488–1498.
<https://doi.org/10.1162/jocn.2009.21140>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and aging*, 27(2), 410–417.
<https://doi.org/10.1037/a0024816>

AUDITORY SCENE ANALYSIS IN MUSIC AND SPEECH: INDIVIDUAL DIFFERENCES IN LISTENING ABILITIES

Submitted to *Scientific Reports* on 12. February 2025

by

Robin Hake¹, Daniel Müllensiefen^{2,3} & Kai Siedenburg^{1,4}

¹ Dept. of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Germany

² Institute for Systematic Musicology, University of Hamburg, Germany

³ Department of Psychology, Goldsmiths, University of London, UK

⁴ Signal Processing and Speech Communication Laboratory Graz University of Technology, Austria

Disclaimer

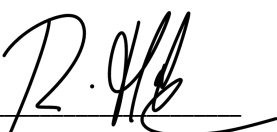
The content of this chapter is identical to the manuscript currently under review for publication at *Scientific Reports*.

The citation for the following publication is:

Hake, R., Müllensiefen, D., & Siedenburg, K. (2025). Auditory scene analysis in music and speech: Individual differences in listening abilities. Manuscript submitted for publication in *Scientific Reports*.

Author contributions

RH, DM and KS designed the study and analysed the data. RH wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.


Robin Hake
Kai Siedenburg

5.1. Abstract

Auditory scene analysis (ASA) is the ability to organize complex auditory mixtures into meaningful events and streams and is fundamental for auditory perception of both speech and music. Individual differences in ASA are recognized in the literature, yet the factors driving this variability remain poorly understood. This study employs a novel music-based ASA task, the Musical Scene Analysis (MSA) test, alongside a speech-in-noise test, to examine the influence of hearing loss, age, working memory capacity (WMC), and musical training. Ninety-two participants were categorised into four groups: 31 older normal-hearing, 34 older hearing-impaired, 26 younger normal-hearing, and one younger hearing-impaired individual. Results reveal a moderate correlation between ASA performance in speech and music ($r = -.5$), suggesting shared underlying perceptual processes, yet the factors influencing individual differences varied across domains. A dual modelling approach using ridge regression and gradient-boosted decision trees identified hearing loss as the strongest predictor of speech-based ASA, with a weaker effect of age, while musical training and WMC had no impact. In contrast, musical training showed a substantial effect on musical ASA, alongside moderate effects of hearing loss and age, while WMC exhibited only a marginal, non-robust effect. These findings highlight both shared and domain-specific factors influencing ASA abilities in speech and music.

5.2. Introduction

Auditory Scene Analysis (ASA) is the fundamental process by which the auditory system organizes complex acoustic environments into meaningful auditory objects and streams, enabling listeners to group, segregate, and track individual sound sources that overlap in time and frequency (Bregman, 1990). In speech, ASA allows individuals to disentangle concurrent sounds, such as separating a single voice from background noise or interfering speakers (commonly referred to as the ‘cocktail party problem’). In the context of multi-source music, this ability - termed Musical Scene Analysis (MSA) - is essential for parsing individual instruments within a mixture (e.g., an oboe within an orchestra) and forms the basis for understanding the rich interplay of instruments, melodies, and harmonies. Individuals are known to vary in their ASA abilities, both in musical and speech dominated acoustic scenes (Alain & Bernstein, 2015; Hake et al., 2024; Siedenburg et al., 2020), yet, the factors driving this variability are not fully understood. This study aims to identify the key contributors to individual differences in ASA abilities and determine their relative importance.

The literature identifies several potential factors that might contribute to individual differences in ASA abilities. Among these, the influence of hearing impairments (such as the effect of elevated audiometric thresholds) has received considerable attention (Anderson & Karawani, 2020; Auerbach & Gritton, 2022). In fact, the negative effects of hearing impairment on speech perception first tend to become obvious in complex acoustic scenes that place considerable demands on ASA such as having a conversation in a noisy restaurant. Specifically, hearing loss, even in mild forms, has been shown to disrupt the processing of acoustic cues that are essential for separating auditory objects within complex scenes. Such impairments are characterized by compromised frequency resolution, attenuated temporal processing, and diminished spatial hearing capabilities (Akeroyd et al., 2007; Middlebrooks, 2017; Phillips et al., 2000; Rimmele et al., 2015). These deficits impair auditory object formation (Shinn-Cunningham & Best, 2008) and the ability to segregate overlapping acoustic streams (Bayat et al., 2013), resulting in notable difficulties in speech-in-speech and speech-in-noise perception (Bayat et al., 2013; Johannesen et al., 2016). Compared to speech, the role of hearing loss in ASA for music has received less attention. However, hearing-impaired individuals have been shown to experience difficulties in tasks that are associated with successful MSA, such as timbre (Choi et al., 2018; Emiroglu & Kollmeier, 2008; Siedenburg et al., 2020), pitch (Uys & van Dijk, 2011), and melody perception (Kirchberger & Russo, 2015; Siedenburg et al., 2020). Siedenburg et al. (2021) directly examined MSA abilities, demonstrating poorer performance among older hearing-impaired (HI) individuals compared to their younger normal-hearing (NH) counterparts in tracking individual instruments within complex musical excerpts. Survey studies further corroborate these findings, with HI individuals frequently reporting difficulties in discriminating instruments (Madsen & Moore, 2014; Vaisberg et al., 2019). A notable limitation of this research, however, is the confounding of age and hearing loss, underscoring the need for studies that isolate the specific effects of hearing impairment on ASA abilities.

Hearing difficulties and impaired speech perception often persist in the absence of overt hearing loss (i.e., elevated hearing thresholds; Goossens et al., 2017; Johannesen et al.,

2016). This indicates the importance of other factors beyond hearing thresholds in shaping auditory perception, with cognitive processes - particularly working memory capacity (WMC) - emerging as a potential contributor to the observed performance variability among individuals. Working memory, the cognitive system responsible for temporarily storing and manipulating information, allows to hold and process transient auditory details, a capability demonstrably associated with auditory tasks outcomes (Akeroyd, 2008). Indeed, research has consistently linked WMC to various aspects of auditory perception (Davidson & Souza, 2024; Humes, 2021; Lentz et al., 2022). For example, individuals with better WMC often excel in speech-related tasks (Dryden et al., 2017; Heinrich et al., 2015; Janse & Andringa, 2021) and are better able to adapt to difficult listening situations (Lentz et al., 2022; Nitsan et al., 2022; Windle et al., 2023). Other research linked low-level concurrent sound segregation to attention and WMC (Garami et al., 2020). Moreover, WMC is closely associated with listening effort, wherein cognitive resources are allocated to overcome auditory obstacles - a challenge that is significantly greater for individuals with lower working memory capacities (Pichora-Fuller et al., 2016; Zekveld et al., 2009). Yet, for ASA-related tasks, the link between WMC and performance were not consistent across all listening conditions. For instance, multitalker scenarios versus single speech streams may differentially engage WMC, and factors such as age and hearing status further modulate this relationship (Akeroyd, 2008; Füllgrabe & Rosen, 2016; Schoof & Rosen, 2014). Furthermore, the extent to which variability in WMC accounts for MSA abilities remains unclear.

Beyond fundamental cognitive abilities, refined listening abilities and knowledge of musical structure, like those acquired by deliberate musical practice may further shape individual perceptual abilities in complex auditory scenarios (Conzelmann & Süß, 2015; Deutsch, 2013; Herholz & Zatorre, 2012; Woods & McDermott, 2018). This reasoning aligns with research demonstrating superior performance of musicians in music-specific tasks like rhythm and beat perception, pitch discrimination, and timbre recognition (Bianchi et al., 2016; Grahn & Rowe, 2009; Micheyl et al., 2006; Nave-Blodgett et al., 2021; Siedenburg et al., 2020; Whittom et al., 2024). Furthermore, musical training has been linked to a cascade of cognitive benefits, including enhanced information processing speed (Bugos et al., 2024), enhanced attentional resolution (Nandi et al., 2023), and even overall improvements in fundamental cognitive regiments like memory (Herholz & Zatorre, 2012; Talamini et al., 2017). This also tends to be associated with enhanced fundamental auditory abilities such as improved frequency discrimination, interaural time differences perception and attentive tracking (Berg et al., 2021; Madsen et al., 2019; Nikjeh et al., 2009; Roden et al., 2014). Musicians also have been shown to outperform non-musicians in auditory stream segregation (Benocci & Calcus, 2024; Elena Benocci & Calcus, 2024; Johnson et al., 2021; Marozeau et al., 2010; Sauvé et al., 2022; Zendel & Alain, 2009), with some evidence suggesting that musical training may even offer some compensation against age-related decline in auditory sensitivity (Dubinsky et al., 2019; Parbery-Clark et al., 2012; Parbery-Clark et al., 2011; Reuter-Lorenz & Park, 2014; Zendel & Alain, 2012). However, it's also worth noting that the influence of musical training on speech perception remains under debate (Coffey, Mogilever, & Zatorre, 2017; Madsen et al., 2019; McKay, 2021; Schellenberg & Lima, 2024). Furthermore, its specific role in MSA processing also remains a subject of active research: While studies like Siedenburg et al. (2020) and Marozeau et al. (2010) demonstrate an advantage for musically trained individuals in MSA tasks, other investigations failed to establish a link between

musical expertise or musical training with MSA performance (Bürgel et al., 2021; Bürgel & Siedenburg, 2023), suggesting a more nuanced relationship or a relatively weak effect that is only observed with sufficient statistical power.

While investigation of peripheral auditory and cognitive factors might offer a unique lens for understanding individual differences in ASA, previous research underscores ageing as a key contributor. Age has been consistently linked to destructive alterations along the auditory pathways and within the cochlea (Cunningham & Tucci, 2017; Liberman, 2017; Windle et al., 2023), with age-related hearing loss emerging as the most prevalent sensory health condition among adults (Bowl & Dawson, 2019; Yueh et al., 2003). The effect of age extends beyond the auditory periphery, though. Age-related neural degeneration within the central auditory system also affects higher-level auditory processing, compromising timing information, gap detection, sound localisation, and spectral cue processing (Chen et al., 2010; Lentz et al., 2022; Profant et al., 2019; Syka, 2020; Walton, 2010) – all of which are essential components for effective ASA (Alain & Bernstein, 2015; Gallun & Best, 2020). These age-related auditory alterations frequently manifest themselves as challenges in filtering out irrelevant competing auditory stimuli (Tun et al., 2002) and difficulties with understanding speech in complex listening environments, even when standard audiometric assessments fall within the normal range (Anderson & Karawani, 2020; Tremblay et al., 2015). Alongside these changes, age also brings a decline in cognitive domains crucial for auditory processing. Processing speed, working memory, and selective attention all diminish with age (Park & Bischof, 2013; Windle et al., 2023), possibly with a negative impact on the ability to detect and track auditory information central to ASA. There is also some evidence that certain age-related auditory effects persist independently of hearing loss and cognitive decline. For example, Lentz et al. (2022) showed that older adults performed worse on several auditory tasks involving temporal processing, even after accounting for hearing loss and working memory. Interestingly, their study also found that stream segregation – a key component of ASA – was the only task that could not be predicted by working memory, hearing loss, or age itself. Certainly, the combined burden of age-related hearing and cognitive changes poses a significant challenge for ASA and warrants careful investigation.

Traditionally, research in auditory perception has primarily concentrated on a limited number of individual factors, often explored in isolation within experiments. However, no single factor can fully explain the variability in ASA performance observed across individuals. Instead, growing evidence highlights the complex interplay between hearing thresholds, cognitive abilities (e.g., working memory), musical training and expertise, and ageing (Fisher et al., 2019; Jayakody et al., 2018; Lin et al., 2013; Loughrey et al., 2018; Madsen et al., 2019; Pelofi et al., 2017). To address the complexities inherent in ASA performance, a multifactorial approach is essential - one that examines the relative contributions of these interrelated factors. Such an approach can provide deeper insights into how individuals organise and interpret complex auditory scenes.

The present study investigates multiple key factors known to shape auditory perception and examines their influence on ASA abilities within a single experimental framework: hearing thresholds, working memory capacity, musical training, and age. To evaluate these factors, the study employs the recently developed Musical Scene Analysis (MSA) test (Hake et al., 2024), which assesses the ability to hear out a target instrument within realistic musical mixtures. Additionally, a speech-in-noise task is employed to examine

ASA abilities within the domain of speech, as it requires auditory stream formation and tracking.

We hypothesise that younger individuals, those with better hearing sensitivity (i.e. lower hearing thresholds), stronger working memory capacity, and greater musical training background will demonstrate superior ASA performance. Furthermore, we predict that the factors influencing musical ASA performance will similarly affect speech-in-noise perception, which could be indicative of shared processes for ASA across the music and speech domains.

5.3. Results

Ninety-two participants were classified into four groups based on age and hearing thresholds (see Figure 5.1). Participants between 18 and 38 years were assigned to the 'younger' groups and participants with pure-tone average (PTA) thresholds of the better ear exceeding 20 dB HL (averaged across octave-spaced frequencies from .125 to 8 kHz) were classified as 'hearing-impaired' (HI).

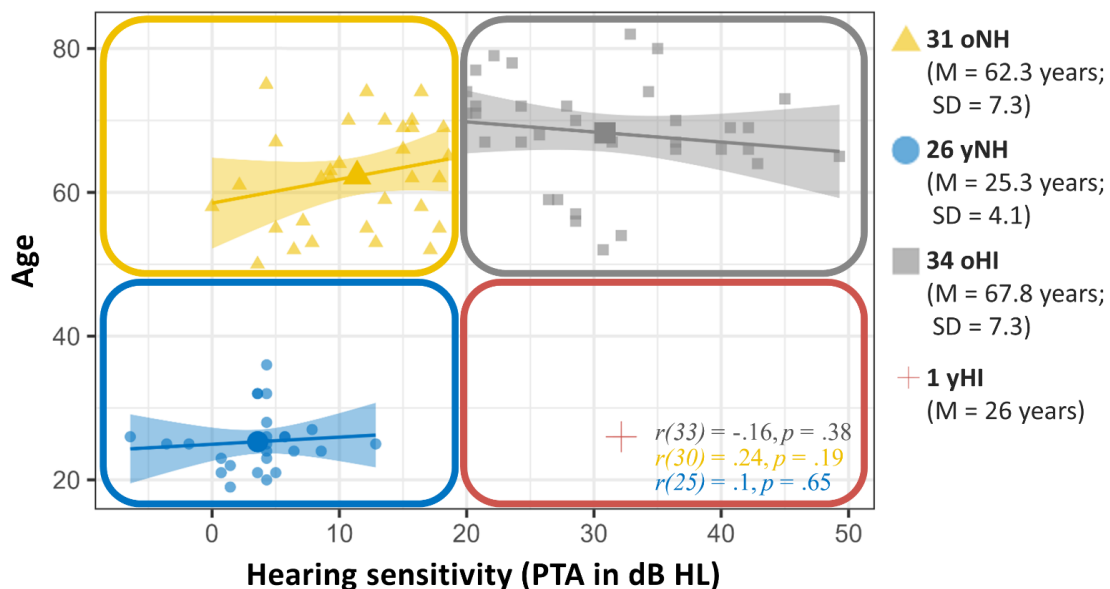


Figure 5.1 Participant distribution by age and hearing thresholds. Colours indicate thresholds for group categorization (age: 50 years; PTA: 20 dB HL). Scatterplots in each quadrant show group-specific correlations between age and PTA, with regression lines and 95% CIs shaded accordingly. Pearson's correlation and significance are noted in the lower right corner. Participant groups are as follows: older adults with normal hearing (oNH, yellow, $n = 31$), older adults with hearing impairment (oHI, grey, $n = 34$), younger adults with normal hearing (yNH, blue, $n = 26$), and one young adult with hearing impairment (yHI, red cross).

Overall, yNH participants demonstrated significantly better performance compared to both older oNH and OHI individuals in ASA related tasks for both speech and music (see Figure 5.2). In the speech domain, yNH participants achieved superior speech reception thresholds (SRT) as measured by the Oldenburg Sentence Test (Wagener, 1999), with a significant effect of group ($F(2, 85) = 20.13, p < .001, \eta^2 = .32$). Similarly, in the music domain, performance on the MSA test revealed a significant main group effect ($F(2, 85) = 11.77, p < .001, \eta^2 = .22$).

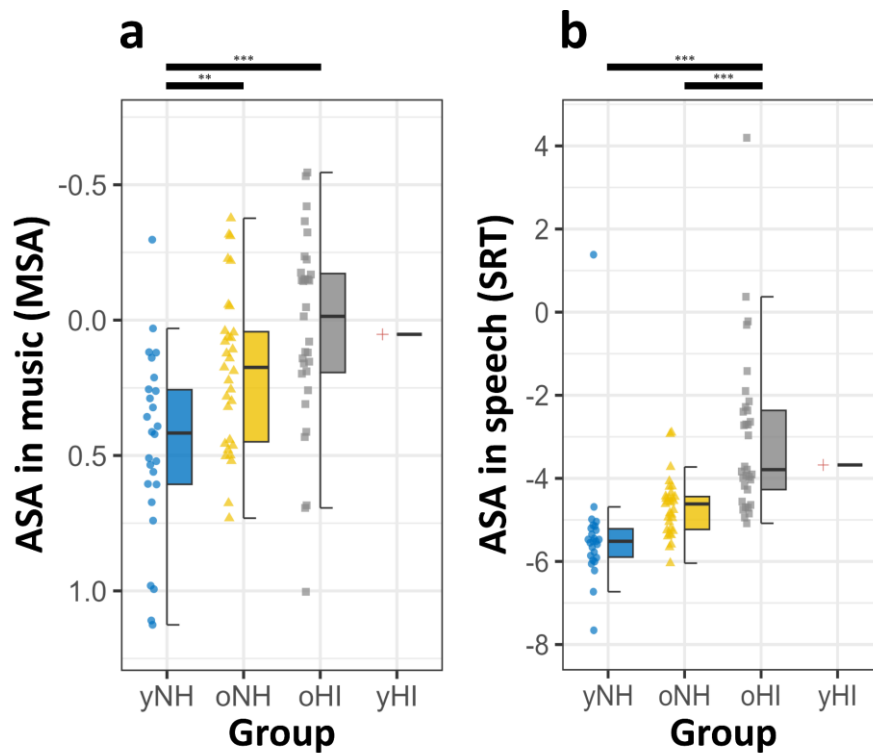


Figure 5.2 Auditory scene analysis scores for music and speech by groups. (a) Musical scene analysis (MSA) scores and (b) Speech-reception thresholds (SRT) by groups. Younger normal-hearing (yNH, blue dots), older normal-hearing (oNH, yellow triangle), older hearing-impaired (oHI, grey squares), younger hearing-impaired (yHI, red cross). Individual scores are plotted as jittered dots on the left to the boxplots. Higher values indicate better performance for MSA, whereas lower values indicate better performance for SRT. To facilitate visual comparison between domains, the y-axis in (a) was inverted, ensuring that better performance is consistently represented by lower values across both panels. Note significance at * $p < .05$, ** $p < .01$, *** $p < .001$

5.3.1. Modelling individual differences

Despite attempts to match age across the groups of oNH and oHI participants and hearing thresholds across the groups of yNH and oNH during recruiting, small but significant differences remained. Specifically, oHI were significantly older than oNH ($M_{\text{difference}} = 5.56$ years, $p = .004$). Additionally, yNH demonstrated significantly better hearing thresholds than the oNH group ($p < .001$), with mean pure-tone average thresholds values of the better ear (PTA) of 3.58 ($SD = 3.79$) and 11.39 ($SD = 5.30$), respectively. Furthermore, all examined factors were significantly intercorrelated (see Table 5.1). These factor intercorrelations complicate isolating the unique contribution of each factor to ASA abilities. For example, the observed strong negative association between PTA and MSA scores ($r = .49$, $p < .001$) might reflect a confound by age-related decline of both hearing and cognition that might influence MSA independently, instead of representing a direct effect of hearing loss.

Table 5.1*The correlation matrix for the test battery*

Variable	MSA	SRT	Age	PTA	Musical training	WMC
MSA	-	-.50***	-.41***	-.49***	.44***	.35***
SRT	84	-	.47***	.64***	-.18	-.27*
Age	88	90	-	.63***	-.23*	-.42***
PTA	87	89	90	-	-.28**	-.42***
Musical training	82	84	85	84	-	.37***
WMC	82	84	85	84	85	-

Note. Above the diagonal are the Pearson correlations calculated from a complete pair of values, which are significant at * $p < .05$, ** $p < .01$, *** $p < .001$. Below the diagonal is the total N for the respective analysis (differences due to missing data). MSA = Musical Scene Analysis; SRT = Speech Reception Thresholds; PTA = Pure Tone Average; WMC = Working Memory Capacity (see Methods for more details).

To account for intercorrelations among predictors, two complementary statistical approaches were employed: Ridge-penalized linear regression and a gradient-boosted decision tree model ('*XGBoost*'; Chen & Guestrin, 2016). By incorporating an L2-norm regularization penalty, ridge regression addresses intercorrelations by shrinking the coefficients of correlated predictors, resulting in more robust and stable model estimates (Gruber, 1998). However, given the plausibility of possible non-linear effects and complex interactions among factors, '*XGBoost*' was employed in addition to the ridge-penalized linear regression model. Importantly, predictor interactions are implicitly considered in '*XGBoost*' models through sequential splits on different factors, allowing the model to capture how the influence of one predictor can vary depending on the value of another.

Both models were used to predict music-related MSA and speech-related SRT scores (see Methods for details on model construction). For MSA, the ridge regression model yielded a root mean square error (RMSE) of .30 and an R^2 of .36. The '*XGBoost*' model demonstrated comparable performance, with an RMSE of .30 for the training set and an RMSE of .31 for the test set, resulting in R^2 values of .34 and .31, respectively. For SRT, ridge regression produced an RMSE of 1.4 and an R^2 of .41, while '*XGBoost*' achieved an RMSE of 1.19 (MAE = .64) for the training set and 1.21 (MAE = .84) for the test set, with corresponding R^2 values of .58 and .55. The small differences in RMSE values between the training and test sets indicate that the '*XGBoost*' models generalize appropriately to new data, indicating a relatively low risk of overfitting.

5.3.1.1. Hearing loss

In both statistical models, poorer hearing thresholds (higher PTA values) were associated with worse ASA performance in both music (MSA) and speech (SRT). Specifically, higher PTA correlated with higher (i.e., worse performance) SRT scores ($r = .64, p < .001$, see Table 1) and lower (worse) MSA scores ($r = -.49, p < .001$). The ridge regression model identified PTA as the strongest predictor of ASA performance across both domains, yielding standardized coefficients of .09 [.04, .13], $p < .001$ for MSA and .96 [.68, 1.26], $p < .001$ for SRT. However, while the overall trend was linear, the strength of this relationship varied across groups; that is, PTA had little influence on ASA performance within the normal-hearing range (< 20 dB HL) but became increasingly pronounced as hearing thresholds deteriorated (Figure 5.3b, d). The SHAP analysis (SHapley Additive exPlanations) provided by the 'XGBoost' model corroborated these group-based findings by quantifying the predictor's marginal contribution to ASA performance. SHAP values reflect the extent to which PTA influences model predictions, independent of other predictors, and are expressed directly in the test score scale (see Methods for details). The SHAP-based trajectory - captured by using a locally estimated scatterplot smoothing (LOESS) regression line - revealed a nonlinear but monotonic trend: PTA had minimal impact below 20 dB HL, followed by a steady increase in its effect up to ~ 40 dB HL, where it plateaued (Figure 5.3c, e). For instance, individuals with ~ 40 dB HL exhibited an expected SRT performance 1.6 dB worse than those with a PTA of ~ 23 dB HL, with the latter representing an SHAP score of 0.

Hearing loss (PTA in dB HL)

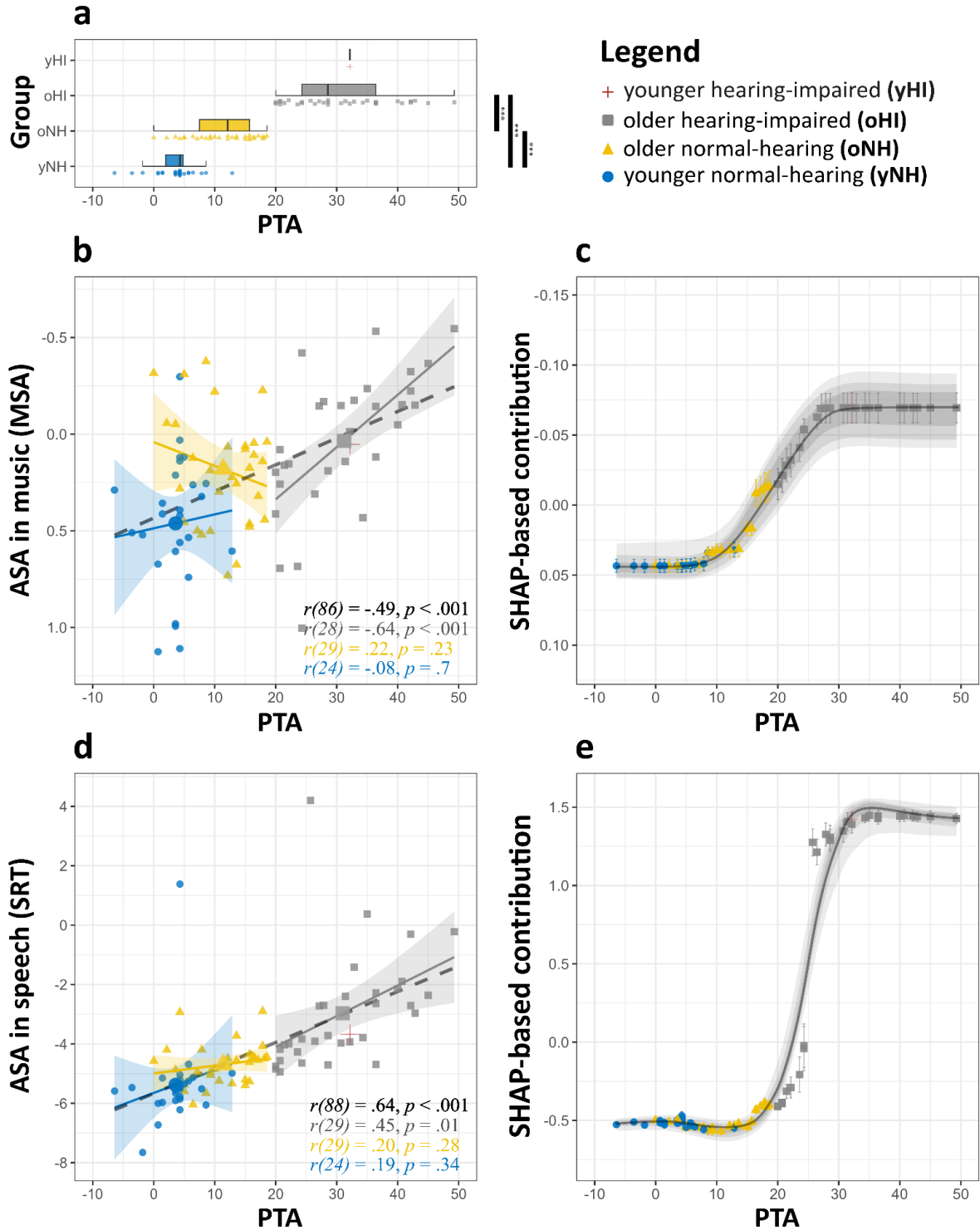


Figure 5.3

Effect of hearing thresholds on ASA abilities in music and speech. (a) Distribution of individual pure-tone average hearing thresholds (PTA) across groups. (b, d) ASA scores for (b) music and (d) speech plotted against PTA. Regression lines are fitted for each group, with shaded areas representing 95% CI. (c, e) SHAP feature contributions for (c) music and (e) speech plotted against PTA, with positive and negative values indicating their respective impact on model predictions. The LOESS curve (span = .35) illustrates the trend, while shaded grey areas show the 95% (light), 80% (medium), and 50% (dark) ranges of LOESS trajectories from the top 1% of models. For interpretability, the y-axis in (b) and (c) is inverted so that better performance visually aligns across domains. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

5.3.1.2. *Working memory capacity*

Working memory capacity (WMC), assessed using the Backwards Digit Span task (Vock & Holling, 2008), correlated significantly with ASA performance in both speech ($r = -.27, p < .05$) and music ($r = .35, p < .001$; see Figure 5.4d). However, these correlations were largely attributable to group membership, defined by age and PTA: Within-group analyses showed no significant correlations between WMC and either SRT or MSA (Figure 5.4b, d). The ridge regression model further indicated that WMC did not significantly predict ASA performance in either domain, yielding a coefficient of $-.03$ $[-.08, .01], p = .16$ for MSA and $.03$ $[-.18, .17], p = .90$ for SRT. Yet, SHAP-based feature analysis suggested a different interpretation: For SRT, SHAP values indicated no discernible effect of WMC, with a near-flat trajectory suggesting negligible predictive relevance across the observed WMC range (Figure 5.4e). In contrast, for MSA, WMC scores revealed a modest but consistent positive effect on performance (Figure 5.4e).

Working memory capacity (WMC)

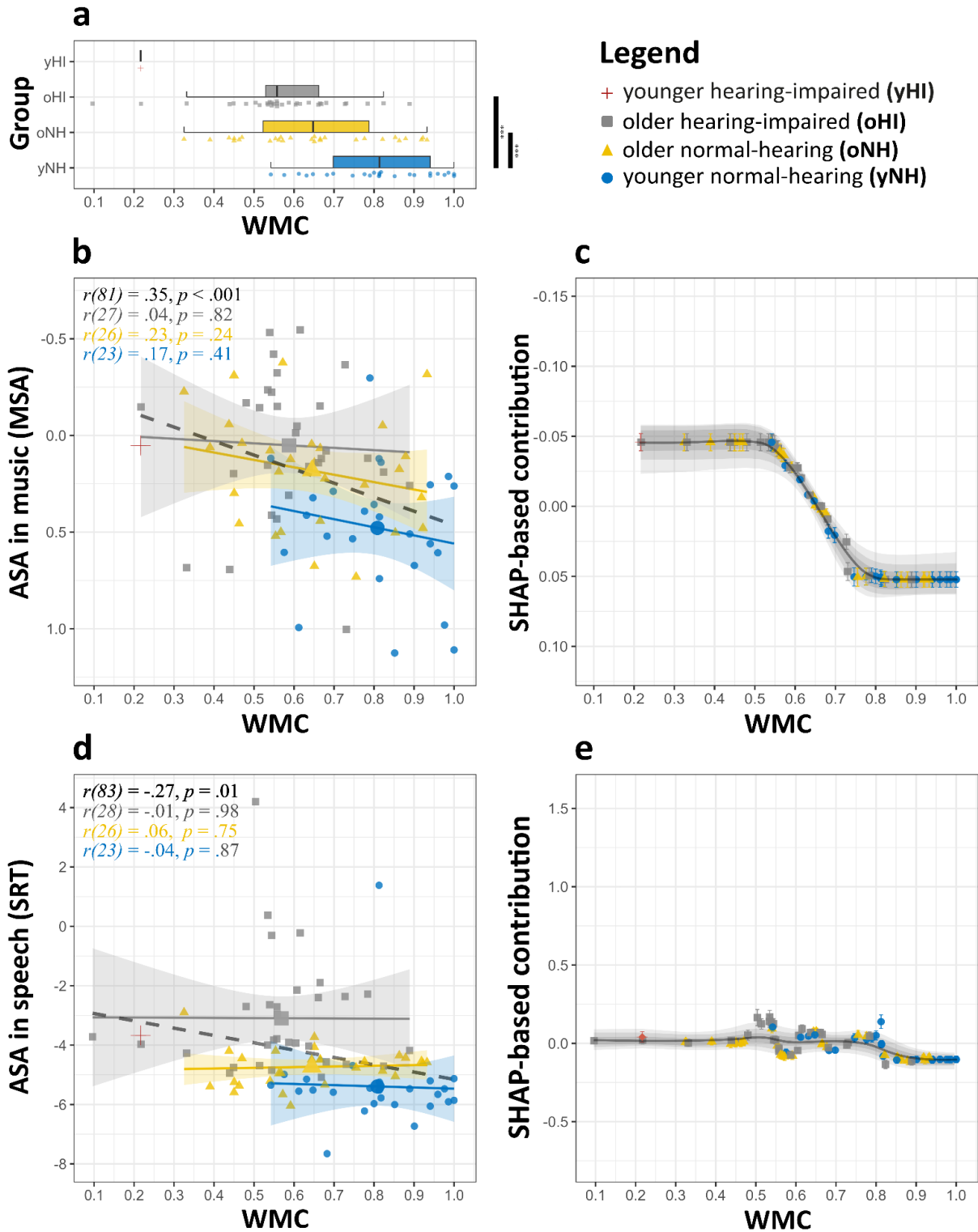


Figure 5.4

Effect of working memory capacities (WMC) on ASA abilities in music and speech. (a) Distribution of WMC across groups. (b, d) ASA scores for (b) music and (d) speech plotted against WMC. Regression lines are fitted for each group, with shaded areas representing 95% CI. (c, e) SHAP feature contributions for (c) music and (e) speech plotted against WMC. The LOESS curve (span = .35) illustrates the trend, while shaded grey areas show the 95% (light), 80% (medium), and 50% (dark) ranges of LOESS trajectories from the top 1% of models. For interpretability, the y-axis in (b) and (c) is inverted so that better performance visually aligns across domains. Significance levels: $*p < .05$, $**p < .01$, $***p < .001$.

5.3.1.3. Musical training

Musical training, as measured by the Gold-MSI subscale (Müllensiefen et al., 2014), exhibited distinct relationships with ASA performance in speech and music. For speech, no significant correlation was observed ($r = -.18$, $p = .11$). In contrast, for music, a moderate positive correlation emerged ($r = .44$, $p < .001$), with these trends remaining generally consistent within groups (Figure 5.5b, d). The ridge regression model further substantiated these findings, indicating that musical training was a significant predictor of MSA performance while showing minimal influence on SRT. For MSA, musical training yielded a standardized coefficient of .08 [$.14$, $.03$], $p = .002$), which was comparable in magnitude to PTA (Figure 5.5c). In contrast, for SRT, musical training did not emerge as a significant predictor ($\beta = -.07$ [$-.30$, $.17$], $p = .53$; Figure 5.5e). SHAP-based feature analysis further supported these findings. For MSA, SHAP values for musical training exhibited a steady increase across training levels. Conversely, in SRT, SHAP values remained near zero across the observed range, suggesting the absence of a meaningful contribution on ASA performance (Figure 5.5c).

Musical training

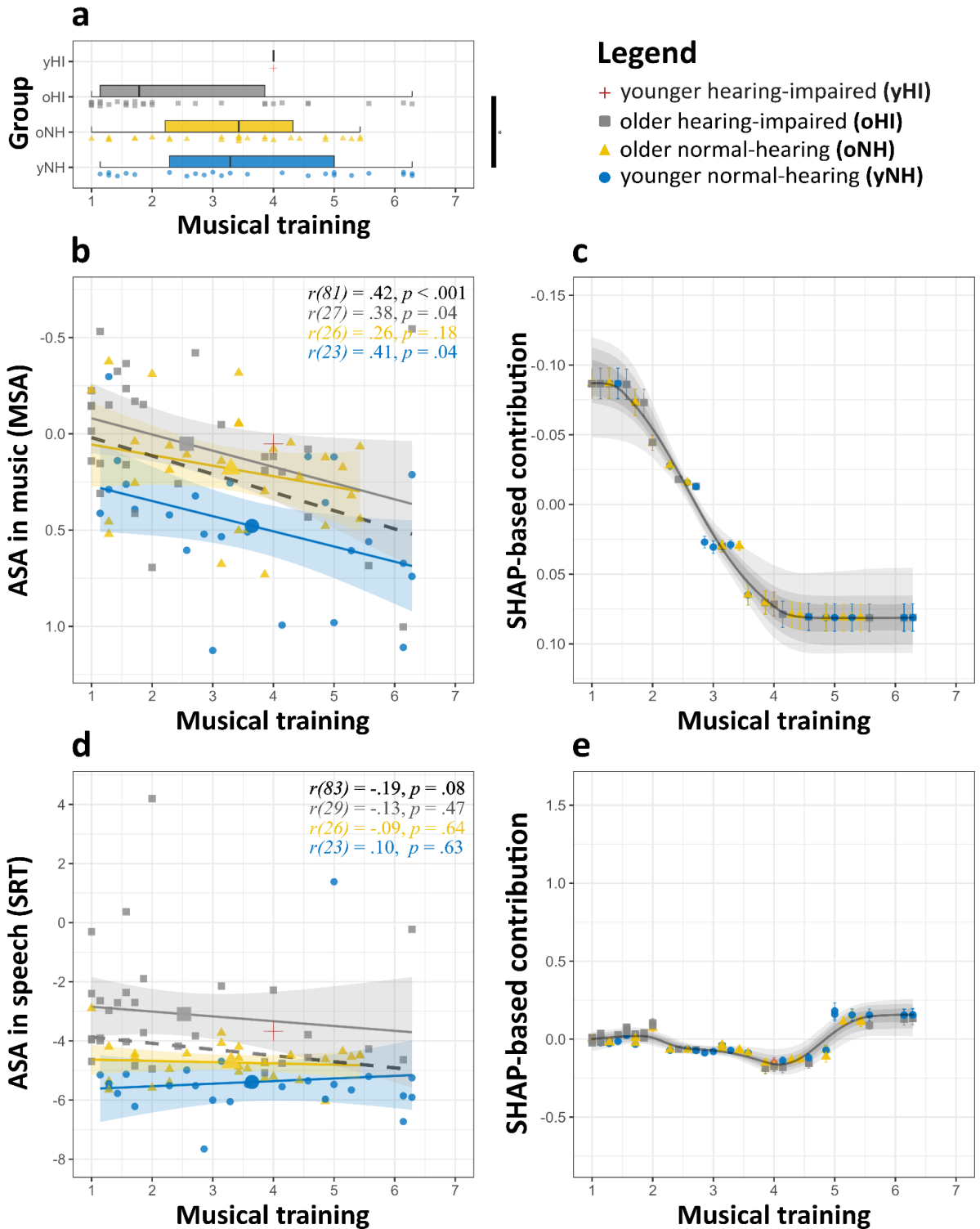


Figure 5.5

Effect of musical training (Gold-MSI subscale) on ASA abilities in music and speech. (a) Distribution of musical training scores across groups. (b, d) ASA scores for (b) music and (d) speech plotted against musical training. Regression lines are fitted for each group, with shaded areas representing 95% CI. (c, e) SHAP feature contributions for (c) music and (d) speech plotted against musical training. The LOESS curve (span = .35) illustrates the trend, while shaded grey areas show the 95% (light), 80% (medium), and 50% (dark) ranges of LOESS trajectories from the top 1% of models. For interpretability, the y-axis in (b) and (c) is inverted so that better performance visually aligns across domains. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

5.3.1.4. Age

Age was significantly correlated with ASA performance in both speech ($r = .47, p < .001$) and music ($r = -.41, p < .001$), suggesting a general decline with increasing age. However, within groups, which are categorised based on PTA and age itself, these correlations disappeared. The only significant within-group correlation, observed for SRT in the yNH group ($n = 25; r = .48, p = .01$; Figure 5.6b, d), was driven by a single extreme outlier. Excluding this participant eliminated the effect ($r = -.03, p = .87$). Despite the absence of age effects within specific groups, SHAP-based analysis revealed a monotonic increase in its contributions across the full lifespan (Figure 5.6c, e) in both domains. Ridge regression results aligned with this pattern, showing a significant effect of age on MSA ($\beta = .05, [.01, .10], p = .02$), while the effect on SRT did not reach significance ($\beta = .17, [-.02, .38], p = .08$). This suggests that, while age effects are less evident within specific subgroups, they manifest across the entire age range.

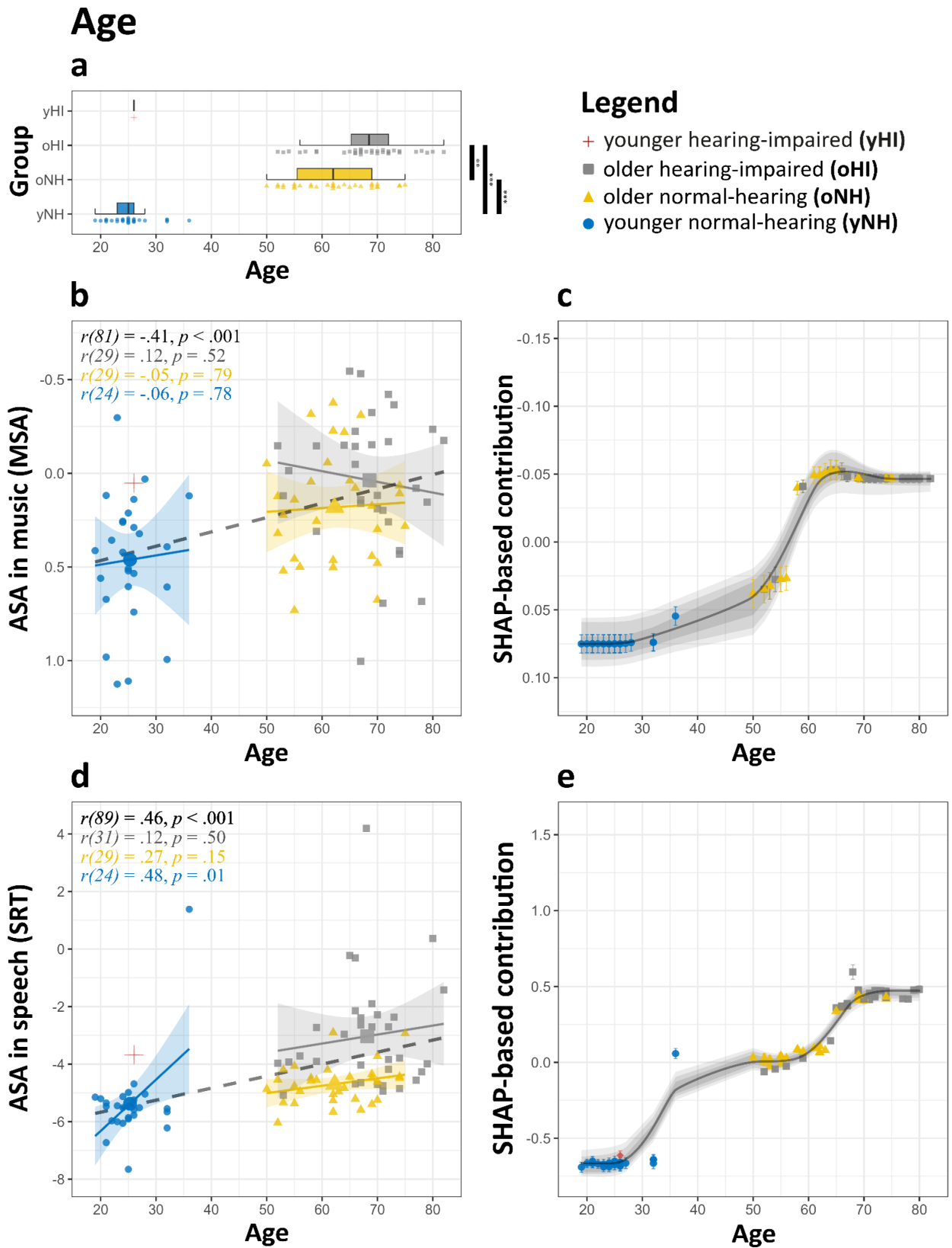


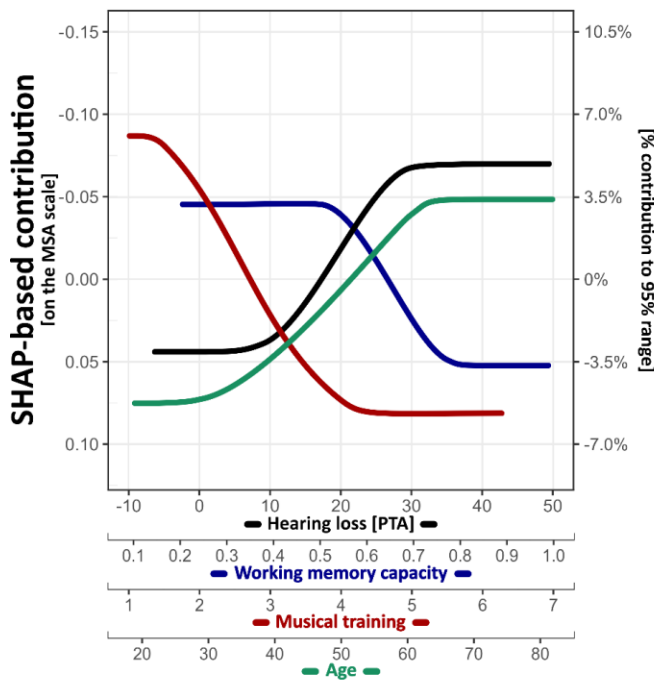
Figure 5.6 Effect of age on ASA abilities in music and speech. (a) Distribution of age across groups. (b, d) ASA scores for (b) music and (d) speech plotted against age. Regression lines are fitted for each group, with shaded areas representing 95% CI. (c, e) SHAP feature contributions for (c) music and (d) speech plotted against age. The LOESS curve (span = .35) illustrates the trend, while shaded grey areas show the 95% (light), 80% (medium), and 50% (dark) ranges of LOESS trajectories from the top 1% of models. For interpretability, the y-axis in (b) and (c) is inverted so that better performance visually aligns across domains. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

5.3.2. Overall factor strength

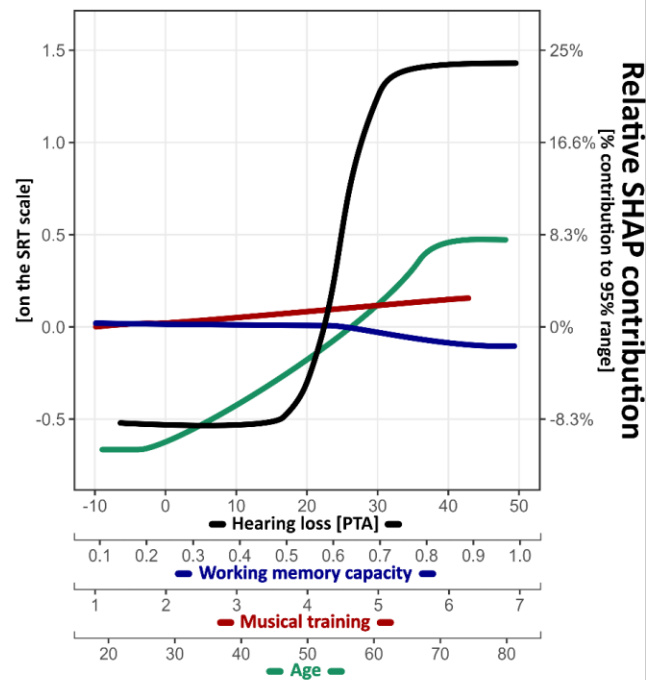
The 'XGBoost' and ridge regression models largely align in identifying key predictors of ASA performance, though slight differences emerge in the relative magnitudes of associations, particularly for MSA. To quantify these relationships, the SHAP contribution range was evaluated in relation to the listeners performance range on the respective test score scale. Specifically, instead of considering the full observed range of MSA and SRT scores, the top and bottom 2.5% of performers were excluded to remove extreme values (i.e., retain only the central 95% of the distribution). This allowed for an estimate of each predictor's relative contribution to ASA performance across domains. For instance, a SHAP value of .5 for SRT corresponded to approximately 8.3% of the performance range within this adjusted 95% distribution (Figure 5.7a, c). Applying this method, musical training accounted for the largest share of variance in MSA (12%), followed by age (10%) and hearing loss (9%), while WMC covered 7%. Ridge regression results showed a similar pattern, ranking hearing loss and musical training as the strongest predictors, with age exerting a slightly weaker effect. WMC showed a small but non-significant association. For speech, both models consistently identify hearing loss as the strongest predictor, with SHAP contributions covering 33% of the observed range. In contrast, neither model detected a meaningful association for WMC or musical training, both of which remained negligible across the observed data range. Age exhibited a weaker association with SRT compared to hearing loss, covering 19% of the SHAP contribution range, but did not reach significance in the ridge model (Figure 5.7b, d).

Gradient boosted decision tree (XGBoost) model

a ASA in music (MSA)

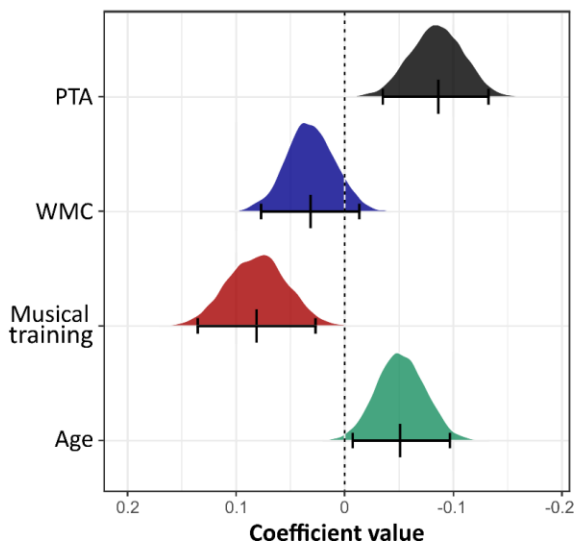


b ASA in speech (SRT)



Ridge regression model

c ASA in music (MSA)



d ASA in speech (SRT)

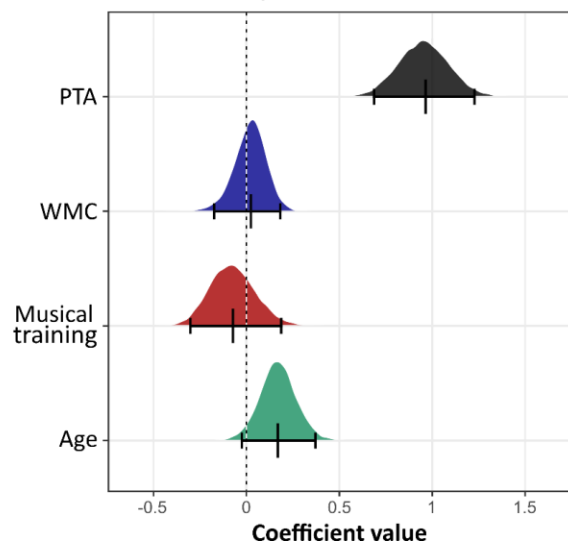


Figure 5.7

SHAP-based model predictions and ridge regression estimates for ASA abilities in music and speech. (a, b) SHAP-based feature contributions from the XGBoost model for MSA (a) and SRT (b), presented as a simplified (straightened) representation of each predictor's association with ASA performance. The left axis denotes the SHAP contribution in the original test score scale, while the right axis represents the relative percentage contribution within the central 95% range of observed scores. (c, d) Ridge regression estimates for MSA (c) and SRT (d), showing coefficient distributions, that are established by bootstrap resampling with 1000 iterations. Black markers indicate the 95% CI.

5.4. Discussion

This study investigated a diverse sample of listeners with varying hearing thresholds, WMC, musical training levels, and ages to evaluate the unique contributions of these factors to ASA performance in music and speech. Consistent with prior research, younger NH individuals outperformed all other groups in both ASA tasks, followed by older participants with NH, with older HI individuals performing the worst. Importantly, however, the observed and well-documented interrelationship among the examined factors complicate the interpretation of individual isolated factor contributions (Davidson & Souza, 2024; Fisher et al., 2019; Jayakody et al., 2018). To address this, two approaches have been employed, one that implies linearity (ridge regression) and another one that does not (called '*XGBoost*'). Notably, individual ASA performance in speech and music are moderately correlated ($r = -.50$), suggesting shared underlying mechanisms across both domains (in line with Bregman, 1990)). Conversely, listener-specific factors exhibited domain-dependent associations with ASA performance, aligning with earlier suggestions of additional and distinct mechanisms for speech and non-speech processing (Bailey, 2010).

5.4.1. *Effect of hearing loss*

We identified a strong relationship between hearing loss (PTA in dB HL) and poorer ASA performance in speech, alongside a moderate relationship with MSA. Both the '*XGBoost*' model estimations and group-level analyses indicated that the adverse effects of hearing loss became apparent even in individuals with mild hearing loss (PTA > 20 dB HL), with ridge regression identifying it as the strongest predictor of ASA abilities in both domains. While some research has reported no significant relationship between hearing loss and stream segregation (Lentz et al., 2022), the present findings align with broader evidence highlighting the critical role of intact auditory pathways for effective ASA across both music and speech (Akeroyd et al., 2007; Hake et al., 2024; Middlebrooks, 2017; Shinn-Cunningham & Best, 2008). Discrepancies in prior findings may be a consequence of differences in task design, as many studies have relied on artificial paradigms (e.g., pure-tone 'ABA' designs) rather than ecologically valid stimuli. Additionally, the extent to which high-frequency hearing loss is considered may be relevant, as ASA deficits may be more pronounced when thresholds above 4 kHz are taken into account (Mishra et al., 2022), which are often disregarded in conventional assessments.

Interestingly, MSA appeared more robust to hearing loss compared to speech-based ASA, with the relative impact of hearing loss on speech being approximately three times greater than in music. This difference may be partially attributed to task demands. In the speech task, participants were required to recall exact words embedded in the auditory scene, whereas in the music task, listeners could potentially take advantage of 'listening in the dips' - a strategy that allows extraction of target sounds during moments of reduced background interference. If this ability remains at least partially preserved, hearing loss may exert a weaker influence on MSA. However, dip listening may also be impaired in individuals with hearing loss (Siedenburg et al., 2020), which could limit the effectiveness of this strategy.

5.4.2. *Effect of working memory capacities*

As listening conditions become more challenging (e.g., increased background noise or increasing the number of instruments in a mixture), auditory processing becomes more effortful, relying increasingly on attentional and WM resources (Nitsan et al., 2022; Pichora-Fuller, 2003). Consequently, we hypothesized that individuals with greater WMC would be better equipped to manage the complexity of musical mixtures, facilitating the retention, rehearsal, and recall of individual instruments. Indeed, while an overall relationship between WMC and ASA performance in both speech and music domains was observed, this relationship was absent within individual groups. Both models further support this, identifying WMC as the least predictive factor of ASA performance overall. Specifically, WMC showed no effect on speech-based ASA. For MSA, model estimates diverged: while ‘*XGBoost*’ suggested a small positive association, the ridge regression model did not yield a statistically significant effect. These results suggest that any potential effect of WMC on ASA is likely minimal, and the current data do not offer strong support for its practical significance.

Nonetheless, this finding is somewhat unexpected, given the well-documented role of WMC in auditory perception (Davidson & Souza, 2024; Humes, 2021) and prior evidence linking WMC to stream segregation and speech-in-noise listening performance (Lotfi et al., 2016), particularly among older hearing-impaired listeners (Dryden et al., 2017). Several factors may explain this discrepancy. One possible explanation is that previously reported associations between WMC and ASA may primarily reflect confounding influences of hearing impairment or aging, rather than a direct causal relationship. For instance, Golden et al. (2017) found no significant differences in ASA abilities in music between Alzheimer’s patients and healthy age-matched controls, despite the former group exhibiting pronounced cognitive impairments, including deficits in WMC (Stopford et al., 2012). Other evidence suggests the WMC association with ASA to be age-dependent, with no effects being reported among NH individuals (Füllgrabe & Rosen, 2016) - a pattern consistent with our group-based findings. Alternatively, task-specific factors may account for the weak or inconsistent findings. The speech-in-noise task employed in this study may impose minimal WMC demands compared to other ASA tasks requiring more demanding scenarios, such as multiple talkers embedded in background sounds. Similarly, the weak but visible trend for an effect of WMC on MSA scores may be attributed to the design of the MSA task, which involves a single-item recall paradigm with a short (1-second) delay, thereby relying to a small degree on a working memory mechanism. While the current study assessed WMC in terms of capacity (i.e., the measurable cognitive limit to store items), future work should consider a broader conceptualization of working memory, e.g. including attentional control, or broader working memory capabilities, as these factors may better explain variability in ASA performance across speech and music domains. Additionally, future studies could incorporate more demanding ASA paradigms, such as those requiring sustained stream tracking or complex auditory scene recall, to better evaluate the role of WMC in advanced ASA tasks.

5.4.3. *Effect of musical training*

Musical practice often involves the identification and discrimination of subtle acoustic details and advance integration of complex auditory streams on a regular basis. Thus, musical training cultivates a unique skill set, potentially offering musicians a distinct advantage within complex auditory scenes like those encountered in ASA tasks (Johnson et al., 2021). To

investigate this, we abstained from a general cut-off score for musicianship, instead considering musical training as a continuum. Our findings revealed a moderate positive relationship between musical training and MSA abilities, independent of hearing loss or age. Specifically, the model suggests that the strength of association of having no musical training compared to extensive musical training on MSA performance is comparable to the impact of mild-to-moderate hearing loss or ageing by approximately 40 years. This finding aligns with research demonstrating the potential of musical training to mitigate age-related auditory decline or declines associated with reduced hearing sensitivity (Dubinsky et al., 2019; Parbery-Clark et al., 2012; Parbery-Clark et al., 2011; Reuter-Lorenz & Park, 2014; Zendel & Alain, 2012). Notably, however, this effect was observed exclusively for musical and not for speech-based ASA. The lack of a clear transfer effect highlights the domain-specific nature of auditory processing, suggesting that skills cultivated through musical education and training may not readily generalise to the unique cognitive and perceptual demands of speech perception in noise. Our findings thus challenge the assumption that musical training universally enhances auditory abilities, raising the question of whether and how its benefits might extend to speech perception - a debate that remains unresolved and polarised in the current literature (Coffey, Chepesiuk, et al., 2017; Coffey, Mogilever, & Zatorre, 2017; Madsen et al., 2019; McKay, 2021; Schellenberg & Lima, 2024).

5.4.4. *Effect of ageing*

Age, with its associated changes in hearing sensitivity, neural degeneration, and general cognitive functioning (Auerbach & Gritton, 2022) often serves as a proxy for a multitude of underlying physiological and auditory influences. These factors, however, complicate the isolation of chronological age's distinct contribution to ASA. To disentangle these effects, we examined a broad age range while controlling for variations in hearing thresholds within groups. At the sample level, we observed a general decline in ASA performance across both speech and music, with increased age and hearing loss. However, within the three groups defined by age range and hearing loss severity, the age effect was largely diminished. Even among older participants (50s-80s), a period typically associated with hearing impairment onset, age was not a strong predictor of ASA performance when hearing loss was accounted for. This suggests that the initial correlations between ASA and age are primarily driven by variability in individuals' hearing thresholds rather than chronological age itself. Nonetheless, considering the full lifespan, aging manifests a practically relevant negative impact on ASA performance in both MSA and speech-based ASA. However, incorporating additional factors - such as attentional control or processing speed - as well as factors capturing hearing impairments beyond overt hearing loss may further account for the variance in performance that is currently attributed to aging.

5.5. Conclusion

This study highlights the complex interplay of factors contributing to individual differences in ASA processes across speech and music domains. While performance in speech and music-related ASA tasks was generally correlated, distinct predictors emerged to be practically impactful for each domain. In speech, hearing loss and age were the most influential factors,

whereas in music, musical training played a pivotal role in predicting MSA abilities. Aging and hearing loss had a less pronounced but still notable impact on the performance in the musical ASA task. Working memory capacity showed no effect on speech-based ASA performance and had only a limited influence on the musical ASA task. These findings highlight the critical role of hearing sensitivity in processing auditory scenes across domains, but particularly in speech-related ASA. Conversely, the observed influence of musical training suggests that extensive musical experience may partially mitigate the effects of hearing loss or aging in domain-specific (i.e. music-related) tasks, highlighting its compensatory potential.

5.6. Materials and Methods

5.6.1. Participants

The experiment included a total of 92 participants (47 males and 45 females). Participants were divided into four groups: 31 older adults with NH ($M = 62.3$, $SD = 7.3$; 18 female), 34 older adults with HI ($M = 67.8$, $SD = 7.7$; 16 female), 26 younger adults with NH ($M = 25.3$; $SD = 4.1$, 11 female), and one young adult with HI (Age = 26, male). A graphical illustration of individuals' pure-tone audiometric thresholds can be found in Figure A5.1 (in Supp. Mat.). If the average pure-tone audiometric measurement for the better ear in a quiet setting surpassed 20 dB HL, subjects were categorised as hearing-impaired (e.g., Humes, 2019).

5.6.2. Test battery

Musical Scene Analysis task (MSA; Hake et al., 2024). The adaptive Musical Scene Analysis test (MSA) is a listening task developed to assess participants' ASA abilities in the context of realistic musical stimuli. The MSA adopts a 'yes-no' paradigm that resembles a 2-alternative-forced-choice task (2-AFC). In each trial, participants hear a two-second excerpt of a single instrument or voice (the target), a one-second silence, and a two-second multi-instrument excerpt (the mixture). Participants are then asked to decide whether the target was part of the mixture or not. The target instrument varies across trials and includes four different instrument types: bass, guitar, lead, and piano. Stimuli are sourced from an open-source music-database ("MedleyDB"), which consists of real-world multitrack music recordings representing a wide range of musical genres (e.g., pop, rock, world/folk, fusion, jazz, rap, classical). All stimuli were presented monaurally. Participant ability is estimated using weighted-likelihood estimation and is projected on a scale of approximately -3 to +3. A score of zero corresponds to the median performance of the reference population from the calibration experiment (Hake et al., 2024), which predominantly included NH and mild HI individuals. The study employed two test sets of the adaptive MSA (version 2.4) with 30 items each.

Backwards Digit Span Memory Test (WMC; Vock & Holling, 2008)). This established working memory assessment requires participants to recall digit sequences in reverse order (e.g., for the sequence 1 2 3 4, the correct response is 4 3 2 1). Digits are presented visually, one at a time. The test comprised sequences of varying lengths (2 trials each of 4-digit and 5-digit sequences; 4 trials each of 6-digit and 7-digit sequences). Performance is measured by the proportion of correctly recalled digits in each sequence. Average results over all sequences are reported.

Oldenburger Satztest (SRT; Wagener, 1999). This adaptive speech-in-noise perception test assesses an individual's ability to comprehend spoken sentences within a background of speech-shaped white noise. Participants have to verbally repeat five-word sentences presented by a male speaker. These sentences were randomly generated from a predefined word class structure (name-verb-numeral-adjective-object), with each word class drawn from a pool of 10 possible word alternatives. The speech level is adaptively varied against a constant noise level to determine the individual's 50% speech reception threshold (SRT in dB HL). The composite score of two test lists of 20 sentences each were used. Lower scores indicate better performance.

Pure-tone-average audiometry (PTA). Hearing thresholds were assessed using standard clinical procedures with an Interacoustics AD528 portable audiometer. Pure-tone thresholds were measured at .125, .25, .5, 1, 2, 4, and 8 kHz. Given the established impact of high-frequency hearing loss on complex auditory processing (Moore, 2020; Narne et al., 2023), the PTA was calculated as an average across all measured frequencies. Participants with a PTA exceeding 20 dB HL in the better ear were classified as hearing-impaired (Humes, 2019). A lower PTA indicates better hearing sensitivity.

Musical training subscale of the Goldsmith Musical Sophistication Index (GMSI; Müllensiefen et al., 2014). The GMSI is a self-report questionnaire that assesses several aspects of musical expertise and experience, including subscales for musical training. Participants responded to questions such as "*At the peak of my interest, I practised my primary instrument for _ hours per day*" and rated their agreement with statements like "*I would not consider myself a musician*" (negatively coded). Responses were given either numerically or on a 7-point Likert scale, ranging from 1 ('*Completely Disagree*') to 7 ('*Completely Agree*'). The final musical training score is calculated from a 7-item set, with scores ranging from 1 to 7, where higher scores indicate greater levels of musical training.

5.6.3. Procedure

Ethical approval for the study was obtained from the Ethics Committee at the Carl von Ossietzky University Oldenburg (Drs.EK/2019/092). All participants provided informed consent prior to participation and received compensation at a rate of €10 per hour. Participants wearing hearing aids were instructed to remove their devices for the duration of the study ($n = 2$). Testing was conducted in two phases. The first phase took place in a controlled laboratory setting. Participants were seated in a soundproof booth and stimuli were presented via calibrated equipment consisting of a computer, an RME Babyface sound card, and Sennheiser HD650 headphones. Long-term sound level was set to 75 dB SPL (A), measured with a Norsonic Nor140 sound-level metre using music-shaped noise. Following a demographics questionnaire, participants completed a pure-tone audiometric test and the speech-in-noise assessment. Participants then undertook the first set of a 30-item adaptive MSA test, followed by additional tests, which will not be addressed in this manuscript. The first phase concluded with a second set of the adaptive MSA. The second testing phase was administered online at least 24 hours after the initial session among the same participants. This phase included the full GMSI self-report questionnaire and the 'Backwards Digit Span' test. Each of the employed tests incorporated a brief training session with practice items. Participants received immediate feedback and could repeat the training as needed. Overall,

the laboratory phase required approximately 80-120 minutes, while the online phase took 25-40 minutes.

5.6.4. Data Analyses

All analyses were executed in R (v2023.10.31; RStudio Team, 2020). The significance level was set at $p < .05$ for all statistical tests. Pearson's correlation coefficients were calculated to examine the relationships between SRT and MSA scores and other auditory and cognitive tests, including Oldenburger Satztest, Backwards Digit Span Memory Test, and the subtest for musical training of the Goldsmith Musical Sophistication Index. Analysis of Variance (ANOVA) was conducted to compare the mean MSA and SRT scores across three different participant groups (excluding the one yHI). Post-hoc comparisons were performed using Bonferroni corrections to control for Type I error. Analyses were conducted on available data, with sample size variations explicitly reported. To predict SRT and MSA scores, both ridge regression and gradient boosted decision tree models were employed. Ridge regression applies L2 regularization to mitigate multicollinearity while assuming linear relationships among predictors. In contrast, the decision tree model accounts for both multicollinearity and complex non-linear interactions without assuming linearity. Predictors included the subtest of the GMSI for musical training, age, PTA, and WMC.

5.6.4.1. Ridge Regression Model Construction

Prior to analysis, all predictor variables were standardized, and only complete data were included (MSA: $n = 82$; SRT: $n = 84$). The optimal regularization parameter (L2) lambda parameter was determined via 10-fold cross-validation. The final ridge regression model was then fitted using this optimal lambda ($\sim .11$ for SRT and $\sim .16$ for MSA). Bootstrap resampling (1,000 iterations) was applied to estimate the variability of the regression coefficients. Confidence intervals (95%) and two-sided bootstrapped p-values were computed for each coefficient, providing robust inferential statistics. This approach allowed for the assessment of the stability and significance of predictor effects across resampled datasets. The ridge regression model was applied to the full dataset, without holding out a separate test set for validation.

5.6.4.2. Gradient Boosted Decision Tree Model (XGBoost)

For the 'XGBoost' model (Chen & Guestrin, 2016), the 'gbtree' booster method was implemented using the 'xgboost' package in R (v1.7.7.1). Complete datasets from participants with available MSA ($n = 89$) and SRT ($n = 91$) scores were included. Model construction employed a random hyperparameter search, an efficient approach for exploring a broad hyperparameter space. Random search is often preferred for hyperparameter optimization as it tests diverse parameter combinations, increasing the likelihood of identifying configurations that minimize validation error (Bergstra & Bengio, 2012). A total of 2,000 random iterations for both MSA and SRT were conducted, with parameter bounds set to balance model complexity and mitigate the risk of overfitting. The specified bounds were as follows: learning rate (*eta*: .05-.15), maximum tree depth (*max_depth*: 3-5), number of boosting rounds (*nrounds*: 50-100), data subsample ratio per boosting round (*subsample*: .5-.9), feature subsample ratio per tree (*colsample_bytree*: .5-.8), minimum child weight (*min_child_weight*: 1-20), L2 regularisation (*lambda*: 0-1.5), L1 regularisation (*alpha*: 0-1.5),

and minimum loss reduction required to make a further split (*gamma*: 0–1.5). The training and test set split was also optimised as part of the tuning process, with bounds set between .5 and .9. Model performance was evaluated using a combination of Root Mean Squared Error (RMSE) and differences in RMSE and R^2 between the training and test sets as an indicator of model fit and overfitting risk. To ensure robust RMSE estimation, repeated random sub-sampling validation (also known as Monte Carlo cross-validation) with 500 iterations was implemented, where each iteration involved a different random split of the data into training and test sets. To further ensure hyperparameter suitability, only models with an RMSE and R^2 difference between the training and test sets below .1 were considered. Instead of selecting a single best-performing model, the average of the top 1% ($n = 20$) of models with the lowest RMSE was then used for the analysis. This approach mitigates the influence of hyperparameter variability and enhances the robustness of the final model estimates. This procedure has been done predicting MSA and SRT scores separately.

For MSA, the final hyperparameter estimates were as follows (mean $\pm$ 95% CI): learning rate ($\eta = .079$ [.069, .090]), maximum tree depth (3.7 [3.36, 4.04]), subsample ratio (.557 [.542, .573]), feature subsample ratio per tree (.684 [.660, .708]), minimum child weight (16.6 [15.7, 17.5]), L2 regularization ($\lambda = .776$ [.629, .923]), L1 regularization ($\alpha = .646$ [.484, .807]), and minimum loss reduction ($\gamma = .254$ [.175, .334]). The average number of boosting rounds was 75.7 [71.0, 80.4], with a data split of 88.3% for training and 11.7% for testing. For SRT, the final hyperparameter estimates were: learning rate ($\eta = .082$ [.07, .093]), maximum tree depth (4.05 [3.63, 4.47]), subsample ratio (.746 [.7, .792]), feature subsample ratio per tree (.693 [.670, .717]), minimum child weight (14.5 [13.1, 15.9]), L2 regularization ($\lambda = .926$ [.745, 1.11]), L1 regularization ($\alpha = .831$ [.615, 1.05]), and minimum loss reduction ($\gamma = .803$ [.611, .995]). The average number of boosting rounds was 72. [66.1, 78.], with 88.7% of data used for training and 11.3% for testing.

5.6.4.3. SHAP-based model estimates for individual feature contribution

Importantly, 'XGBoost' model allows the calculation of SHAP values (SHapley Additive exPlanations), a method for quantifying individual-level feature contributions to model predictions (Lundberg & Lee, 2017). SHAP values assess each predictor's marginal contribution by considering all possible combinations of features being present or absent, providing a nuanced interpretation of how individual differences influence model outcomes. Positive SHAP values indicate that a feature increases the predicted score, while negative values suggest a decreasing effect. Since SHAP values are presented on the original test scale, their interpretation follows the scoring direction of each measure. For SRT, lower values indicate better performance, meaning higher SHAP values correspond to poorer outcomes. In contrast, for MSA, higher values denote better performance, so higher SHAP values reflect improved outcomes. The sum of all SHAP values for a participant, combined with the model's baseline prediction, determines the final predicted score (see Figure 5.8 for an illustration). SHAP value estimates and their corresponding LOESS trajectory are constrained to the observed data range and should not be extrapolated beyond it. For instance, PTA values exceeding 50 dB HL fall outside the estimated range and cannot be reliably inferred from the model.

Final SHAP estimates were derived from the top 1% of models ($n = 20$) to account for variability in hyperparameter configurations. Confidence intervals (99%) for these estimates were generated through 500 bootstrap iterations, enhancing the robustness of feature contribution interpretations.

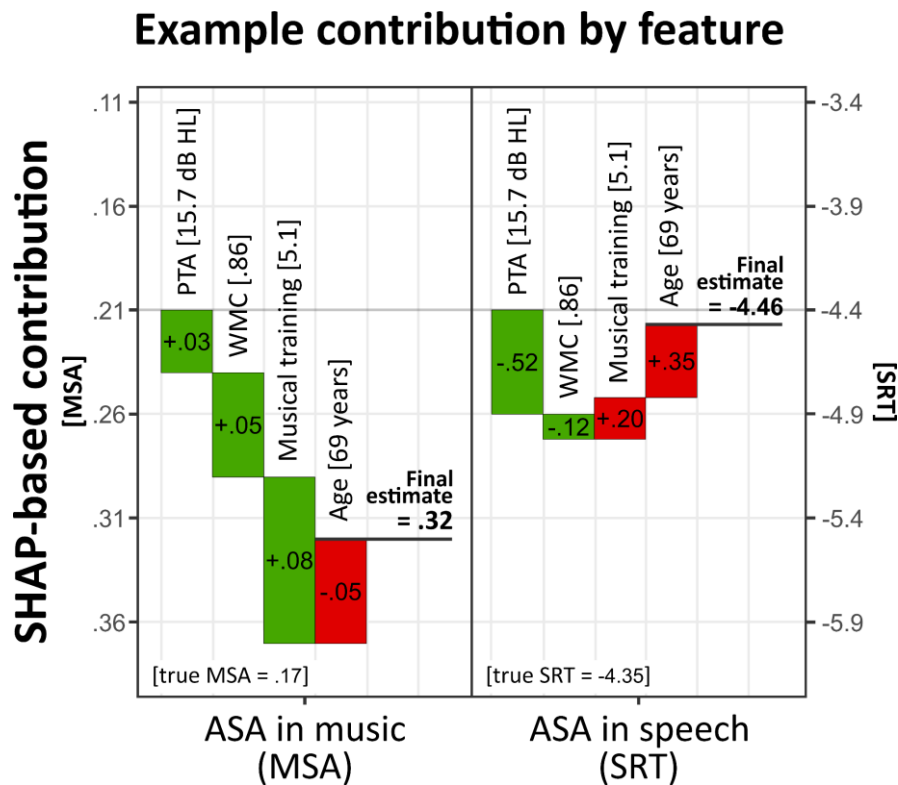


Figure 5.8 SHAP-based feature contributions for ASA abilities in music and speech. Example SHAP-feature contributions for a single participant (Participant #42) for both SRT and MSA. Positive contributions (green) increase the prediction relative to the baseline value (overall mean), while negative contributions (red) decrease it. Lower scores indicate better performance for SRT, whereas higher scores reflect better performance for MSA.

5.7. Declarations

Acknowledgements. This work has been funded by the Freigeist Fellowship of the Volkswagen Foundation to KS. This study was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330 A6.

Author contributions. RH, DM and KS designed the study and analysed the data. RH wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.

Data Availability Statement. The dataset, analysis scripts, and MSA test used in this study are available in the author's GitHub repository (<https://github.com/rhake14>).

Ethics approval. Ethical approval was provided by the Ethics Committee at University of Oldenburg (Drs.EK/2019/092).

Consent to participate. Informed consent was obtained from all participants in this study, permitting the collection of anonymized data and its potential publication in journal articles.

Conflict of interest. The authors declare that there is no conflict of interest.

Copyright. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

5.8. References

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47 Suppl 2, S53-71. <https://doi.org/10.1080/14992020802301142>
- Akeroyd, M. A., Gatehouse, S., & Blaschke, J. (2007). The detection of differences in the cues to distance by elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 121(2), 1077–1089. <https://doi.org/10.1121/1.2404927>
- Alain, C., & Bernstein, L. J. (2015). Auditory Scene Analysis. *Music Perception*, 33(1), 70–82. <https://doi.org/10.1525/mp.2015.33.1.70>
- Anderson, S., & Karawani, H. (2020). Objective evidence of temporal processing deficits in older adults. *Hearing Research*, 397, 108053. <https://doi.org/10.1016/j.heares.2020.108053>
- Auerbach, B. D., & Gritton, H. J. (2022). Hearing in Complex Environments: Auditory Gain Control, Attention, and Hearing Loss. *Frontiers in Neuroscience*, 16, 799787. <https://doi.org/10.3389/fnins.2022.799787>
- Bailey, T. (2010). Auditory pathways and processes: Implications for neuropsychological assessment and diagnosis of children and adolescents. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 16(6), 521–548. <https://doi.org/10.1080/09297041003783310>
- Bayat, A., Farhadi, M., Pourbakht, A., Sadjedi, H., Emamdjomeh, H., Kamali, M., & Mirmomeni, G. (2013). A comparison of auditory perception in hearing-impaired and normal-hearing listeners: An auditory scene analysis study. *Iranian Red Crescent Medical Journal*, 15(11), e9477. <https://doi.org/10.5812/ircmj.9477>
- Benocci, E [E.], & Calcus, A [A.] (2024). Development of Speech Perception in Noise: Effect of Auditory Scene Analysis and Musical Abilities. In *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023* (pp. 3893–3899). European Acoustics Association. <https://doi.org/10.61782/fa.2023.0156>
- Benocci, E [Elena], & Calcus, A [Axelle] (2024). Stream segregation, musical abilities, and the development of speech perception in noise. *JASA Express Letters*, 4(12). <https://doi.org/10.1121/10.0034543>
- Berg, M. von, Steffens, J., Weinzierl, S., & Müllensiefen, D. (2021). Assessing room acoustic listening expertise. *The Journal of the Acoustical Society of America*, 150(4), 2539. <https://doi.org/10.1121/10.0006574>
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(1). JMLR.org
- Bianchi, F., Santurette, S., Wendt, D., & Dau, T. (2016). Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort. *Journal of the Association for Research in Otolaryngology : JARO*, 17(1), 69–79. <https://doi.org/10.1007/s10162-015-0548-2>
- Bowl, M. R., & Dawson, S. J. (2019). Age-Related Hearing Loss. *Cold Spring Harbor Perspectives in Medicine*, 9(8). <https://doi.org/10.1101/cshperspect.a033217>
- Bregman, A. S. (1990). *Auditory Scene Analysis*. The MIT Press. <https://doi.org/10.7551/mitpress/1486.001.0001>
- Bugos, J. A., Gbadamosi, A., Laesker, D., Chow, R., Sirocchi, S., Norgaard, M., Ghent, J., & Alain, C. (2024). Jazz Piano Training Modulates Neural Oscillations and Executive Functions in Older Adults. *Music Perception*, 41(5), 378–392. <https://doi.org/10.1525/mp.2024.41.5.378>

- Bürgel, M., Picinali, L., & Siedenburg, K. (2021). Listening in the Mix: Lead Vocals Robustly Attract Auditory Attention in Popular Music. *Frontiers in Psychology*, 12, 769663. <https://doi.org/10.3389/fpsyg.2021.769663>
- Bürgel, M., & Siedenburg, K. (2023). Salience of Frequency Micro-modulations in Popular Music. *Music Perception: An Interdisciplinary Journal*, 41(1), 1–14. <https://doi.org/10.1525/mp.2023.41.1.1>
- Chen, B., Zhong, Y., Peng, W., Sun, Y., & Kong, W.-J. (2010). Age-related changes in the central auditory system: Comparison of D-galactose-induced aging rats and naturally aging rats. *Brain Research*, 1344, 43–53. <https://doi.org/10.1016/j.brainres.2010.04.082>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System, 11(34), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Choi, J. E., Won, J. H., Kim, C. H., Cho, Y.-S., Hong, S. H., & Moon, I. J. (2018). Relationship between spectrotemporal modulation detection and music perception in normal-hearing, hearing-impaired, and cochlear implant listeners. *Scientific Reports*, 8(1), 800. <https://doi.org/10.1038/s41598-017-17350-w>
- Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017). Neural Correlates of Early Sound Encoding and their Relationship to Speech-in-Noise Perception. *Frontiers in Neuroscience*, 11, 479. <https://doi.org/10.3389/fnins.2017.00479>
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- Conzelmann, K., & Süß, H.-M. (2015). Auditory intelligence: Theoretical considerations and empirical findings. *Learning and Individual Differences*, 40, 27–40. <https://doi.org/10.1016/j.lindif.2015.03.029>
- Cunningham, L. L., & Tucci, D. L. (2017). Hearing Loss in Adults. *The New England Journal of Medicine*, 377(25), 2465–2473. <https://doi.org/10.1056/NEJMra1616601>
- Davidson, A., & Souza, P. (2024). Relationships Between Auditory Processing and Cognitive Abilities in Adults: A Systematic Review. *Journal of Speech, Language, and Hearing Research : JSLHR*, 67(1), 296–345. https://doi.org/10.1044/2023_JSLHR-22-00716
- Deutsch, D. (Ed.). (2013). *The psychology of music* (Third edition). Academic Press.
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, 21, 2331216517744675. <https://doi.org/10.1177/2331216517744675>
- Dubinsky, E., Wood, E. A., Nespoli, G., & Russo, F. A. (2019). Short-Term Choir Singing Supports Speech-in-Noise Perception and Neural Pitch Strength in Older Adults With Age-Related Hearing Loss. *Frontiers in Neuroscience*, 13, 1153. <https://doi.org/10.3389/fnins.2019.01153>
- Emiroglu, S., & Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain Research*, 1220, 199–207. <https://doi.org/10.1016/j.brainres.2007.08.067>
- Fisher, G. G., Chacon, M., & Chaffee, D. S. (2019). Theories of Cognitive Aging and Work. In *Work Across the Lifespan* (pp. 17–45). Elsevier. <https://doi.org/10.1016/B978-0-12-812756-8.00002-5>
- Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology*, 7, 1268. <https://doi.org/10.3389/fpsyg.2016.01268>
- Gallun, F. J., & Best, V. (2020). Age-Related Changes in Segregation of Sound Sources. In K. S. Helfer, E. L. Bartlett, A. N. Popper, & R. R. Fay (Eds.), *Springer Handbook of Auditory*

- Research. Aging and Hearing* (Vol. 72, pp. 143–171). Springer International Publishing. https://doi.org/10.1007/978-3-030-49367-7_7
- Garami, L., Chow, R., Fakuade, A., Swaminathan, S., & Alain, C. (2020). Orienting Attention to Auditory and Visual Short-term Memory: The Roles of Age, Hearing Loss, and Cognitive Status. *Experimental Aging Research*, 46(1), 22–38. <https://doi.org/10.1080/0361073X.2019.1693008>
- Golden, H. L., Clark, C. N., Nicholas, J. M., Cohen, M. H., Slattey, C. F., Paterson, R. W., Foulkes, A. J. M., Schott, J. M., Mummery, C. J., Crutch, S. J., & Warren, J. D. (2017). Music Perception in Dementia. *Journal of Alzheimer's Disease : JAD*, 55(3), 933–949. <https://doi.org/10.3233/JAD-160359>
- Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, 344, 109–124. <https://doi.org/10.1016/j.heares.2016.11.004>
- Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(23), 7540–7548. <https://doi.org/10.1523/JNEUROSCI.2018-08.2009>
- Gruber, M. (1998). *Improving Efficiency by Shrinkage: The James—Stein and Ridge Regression Estimators* (1st ed.). *Statistics: v.156*. CRC Press. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5118495>
- Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., & Siedenburg, K. (2024). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods*(56), 5456–5481. <https://doi.org/10.3758/s13428-023-02279-y>
- Heinrich, A., Henshaw, H., & Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers in Psychology*, 6, 782. <https://doi.org/10.3389/fpsyg.2015.00782>
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, 76(3), 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>
- Humes, L. E. (2019). The World Health Organization's hearing-impairment grading system: An evaluation for unaided communication in age-related hearing loss. *International Journal of Audiology*, 58(1), 12–20. <https://doi.org/10.1080/14992027.2018.1518598>
- Humes, L. E. (2021). Factors Underlying Individual Differences in Speech-Recognition Threshold (SRT) in Noise Among Older Adults. *Frontiers in Aging Neuroscience*, 13, 702739. <https://doi.org/10.3389/fnagi.2021.702739>
- Janse, E., & Andringa, S. J. (2021). The roles of cognitive abilities and hearing acuity in older adults' recognition of words taken from fast and spectrally reduced speech. *Applied Psycholinguistics*, 42(3), 763–790. <https://doi.org/10.1017/S0142716421000047>
- Jayakody, D. M. P., Friedland, P. L., Martins, R. N., & Sohrabi, H. R. (2018). Impact of Aging on the Auditory System and Related Cognitive Functions: A Narrative Review. *Frontiers in Neuroscience*, 12, 125. <https://doi.org/10.3389/fnins.2018.00125>
- Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., & Lopez-Poveda, E. A. (2016). The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216516641055>
- Johnson, N., Shiju, A. M., Parmar, A., & Prabhu, P. (2021). Evaluation of Auditory Stream Segregation in Musicians and Nonmusicians. *International Archives of Otorhinolaryngology*, 25(1), e77–e80. <https://doi.org/10.1055/s-0040-1709116>

- Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, 36(2), 217–228. <https://doi.org/10.1097/AUD.0000000000000112>
- Lentz, J. J., Humes, L. E., & Kidd, G. R. (2022). Differences in Auditory Perception Between Young and Older Adults When Controlling for Differences in Hearing Loss and Cognition. *Trends in Hearing*, 26, 23312165211066180. <https://doi.org/10.1177/23312165211066180>
- Liberman, M. C. (2017). Noise-induced and age-related hearing loss: New perspectives and potential therapies. *F1000Research*, 6, 927. <https://doi.org/10.12688/f1000research.11310.1>
- Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., & Simonsick, E. M. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, 173(4), 293–299. <https://doi.org/10.1001/jamainternmed.2013.1868>
- Lotfi, Y., Mehrkian, S., Moossavi, A., Zadeh, S. F., & Sadjedi, H. (2016). Relation between Working Memory Capacity and Auditory Stream Segregation in Children with Auditory Processing Disorder. *Iranian Journal of Medical Sciences*, 41(2), 110–117.
- Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., & Lawlor, B. A. (2018). Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis. *JAMA Otolaryngology—Head & Neck Surgery*, 144(2), 115–126. <https://doi.org/10.1001/jamaoto.2017.2513>
- Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. <https://doi.org/10.48550/arXiv.1705.07874>
- Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, 9(1), 10404. <https://doi.org/10.1038/s41598-019-46728-1>
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18. <https://doi.org/10.1177/2331216514558271>
- Marozeau, J., Innes-Brown, H., Grayden, D. B., Burkitt, A. N., & Blamey, P. J. (2010). The effect of visual cues on auditory stream segregation in musicians and non-musicians. *PloS One*, 5(6), e11297. <https://doi.org/10.1371/journal.pone.0011297>
- McKay, C. M. (2021). No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review. *Trends in Hearing*, 25, 2331216520985678. <https://doi.org/10.1177/2331216520985678>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Middlebrooks, J. C. (2017). Spatial Stream Segregation. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *Springer Handbook of Auditory Research. The Auditory System at the Cocktail Party* (Vol. 60, pp. 137–168). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_6
- Mishra, S. K., Saxena, U., & Rodrigo, H. (2022). Extended High-frequency Hearing Impairment Despite a Normal Audiogram: Relation to Early Aging, Speech-in-noise Perception, Cochlear Function, and Routine Earphone Use. *Ear and Hearing*, 43(3), 822–835. <https://doi.org/10.1097/AUD.0000000000001140>
- Moore, B. C. (2020). The importance of high-frequency hearing in a medico-legal context. *The Journal of the Acoustical Society of America*, 148(4_Supplement), 2542. <https://doi.org/10.1121/1.5147057>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS One*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>

- Nandi, B., Ostrand, A., Johnson, V., Ford, T. J., Gazzaley, A., & Zanto, T. P. (2023). Musical Training Facilitates Exogenous Temporal Attention via Delta Phase Entrainment within a Sensorimotor Network. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 43(18), 3365–3378. <https://doi.org/10.1523/JNEUROSCI.0220-22.2023>
- Narne, V. K., Jain, S., Bharani, Ravi, S. K., Almudhi, A., Krishna, Y., & Moore, B. C. J. (2023). The effect of recreational noise exposure on amplitude-modulation detection, hearing sensitivity at frequencies above 8 kHz, and perception of speech in noise. *The Journal of the Acoustical Society of America*, 153(5), 2562. <https://doi.org/10.1121/10.0017973>
- Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology. General*, 150(2), 314–339. <https://doi.org/10.1037/xge0000903>
- Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: Influence of music training. *Ear and Hearing*, 30(4), 432–446. <https://doi.org/10.1097/AUD.0b013e3181a61bf2>
- Nitsan, G., Banai, K., & Ben-David, B. M. (2022). One Size Does Not Fit All: Examining the Effects of Working Memory Capacity on Spoken Word Recognition in Older Adults Using Eye Tracking. *Frontiers in Psychology*, 13, 841466. <https://doi.org/10.3389/fpsyg.2022.841466>
- Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiology of Aging*, 33(7), 1483.e1-4. <https://doi.org/10.1016/j.neurobiolaging.2011.12.015>
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PloS One*, 6(5), e18082. <https://doi.org/10.1371/journal.pone.0018082>
- Park, D. C., & Bischof, G. N. (2013). The aging mind: Neuroplasticity in response to cognitive training. *Dialogues in Clinical Neuroscience*, 15(1), 109–119. <https://doi.org/10.31887/DCNS.2013.15.1/dpark>
- Pelofi, C., Gardelle, V. de, Egré, P., & Pressnitzer, D. (2017). Interindividual variability in auditory scene analysis revealed by confidence judgements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 20160107. <https://doi.org/10.1098/rstb.2016.0107>
- Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. (2000). Frequency and temporal resolution in elderly listeners with good and poor word recognition. *Journal of Speech, Language, and Hearing Research : JSLHR*, 43(1), 217–228. <https://doi.org/10.1044/jslhr.4301.217>
- Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, 42 Suppl 2, 2S26-32.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37 Suppl 1, 5S-27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Profant, O., Jilek, M., Bures, Z., Vencovsky, V., Kucharova, D., Svobodova, V., Korynta, J., & Syka, J. (2019). Functional Age-Related Changes Within the Human Auditory System Studied by Audiometric Examination. *Frontiers in Aging Neuroscience*, 11, 26. <https://doi.org/10.3389/fnagi.2019.00026>

- Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review*, 24(3), 355–370.
<https://doi.org/10.1007/s11065-014-9270-9>
- Rimmele, J. M., Sussman, E., & Poeppel, D. (2015). The role of temporal structure in the investigation of sensory memory, auditory scene analysis, and speech perception: A healthy-aging perspective. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 95(2), 175–183.
<https://doi.org/10.1016/j.ijpsycho.2014.06.010>
- Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music*, 42(2), 284–298. <https://doi.org/10.1177/0305735612471239>
- RStudio Team. (2020). *RStudio: Integrated Development for R* (Version v2023.10.31) [Computer software]. <http://www.rstudio.com/>
- Sauvé, S. A., Marozeau, J., & Rich Zendel, B. (2022). The effects of aging and musicianship on the use of auditory streaming cues. *PloS One*, 17(9), e0274631.
<https://doi.org/10.1371/journal.pone.0274631>
- Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, 75, 87–128. <https://doi.org/10.1146/annurev-psych-032323-051354>
- Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience*, 6, 307. <https://doi.org/10.3389/fnagi.2014.00307>
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4), 283–299.
<https://doi.org/10.1177/1084713808325306>
- Siedenburg, K., Goldmann, K., & van de Par, S. (2021). Tracking Musical Voices in Bach's The Art of the Fugue: Timbral Heterogeneity Differentially Affects Younger Normal-Hearing Listeners and Older Hearing-Aid Users. *Frontiers in Psychology*, 12, 608684.
<https://doi.org/10.3389/fpsyg.2021.608684>
- Siedenburg, K., Röttges, S., Wagener, K. C., & Hohmann, V. (2020). Can You Hear Out the Melody? Testing Musical Scene Perception in Young Normal-Hearing and Older Hearing-Impaired Listeners. *Trends in Hearing*, 24, 2331216520945826.
<https://doi.org/10.1177/2331216520945826>
- Stopford, C. L., Thompson, J. C., Neary, D., Richardson, A. M. T., & Snowden, J. S. (2012). Working memory, attention, and executive function in Alzheimer's disease and frontotemporal dementia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 48(4), 429–446. <https://doi.org/10.1016/j.cortex.2010.12.002>
- Syka, J. (2020). Age-Related Changes in the Auditory Brainstem and Inferior Colliculus. In K. S. Helfer, E. L. Bartlett, A. N. Popper, & R. R. Fay (Eds.), *Springer Handbook of Auditory Research. Aging and Hearing* (Vol. 72, pp. 67–96). Springer International Publishing.
https://doi.org/10.1007/978-3-030-49367-7_4
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, 12(10), e0186773.
<https://doi.org/10.1371/journal.pone.0186773>
- Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., & Cruickshanks, K. J. (2015). Self-Reported Hearing Difficulties Among Adults With Normal Audiograms: The Beaver Dam Offspring Study. *Ear and Hearing*, 36(6), e290-9.
<https://doi.org/10.1097/AUD.0000000000000195>
- Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17(3), 453–467.
<https://doi.org/10.1037/0882-7974.17.3.453>

- Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *The South African Journal of Communication Disorders = Die Suid-Afrikaanse Tydskrif Vir Kommunikasieafwykings*, 58, 19–47.
<https://doi.org/10.4102/sajcd.v58i1.38>
- Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2019). A Qualitative Study of the Effects of Hearing Loss and Hearing Aid Use on Music Perception in Performing Musicians. *Journal of the American Academy of Audiology*, 30(10), 856–870.
<https://doi.org/10.3766/jaaa.17019>
- Vock, M., & Holling, H. (2008). The measurement of visuo-spatial and verbal-numerical working memory: Development of IRT-based scales. *Intelligence*, 36(2), 161–182.
<https://doi.org/10.1016/j.intell.2007.02.004>
- Wagener, K. C. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil 1: Design des Oldenburger Satztests. *Zeitschrift Für Audiologie*, 38(1), 4–15.
- Walton, J. P. (2010). Timing is everything: Temporal processing deficits in the aged auditory brainstem. *Hearing Research*, 264(1-2), 63–69.
<https://doi.org/10.1016/j.heares.2010.03.002>
- Whittom, A., Chauvette, L., Bégin, A., Blanchette, I., Tremblay, P., & Sharp, A. (2024). Music Perception in Older Adults With Hearing Loss: Protective Effect of Musical Experience. *Ear and Hearing*. Advance online publication.
<https://doi.org/10.1097/AUD.0000000000001615>
- Windle, R., Dillon, H., & Heinrich, A. (2023). A review of auditory processing and cognitive change during normal ageing, and the implications for setting hearing aids for older adults. *Frontiers in Neurology*, 14, 1122420.
<https://doi.org/10.3389/fneur.2023.1122420>
- Woods, K. J. P., & McDermott, J. H. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of Sciences of the United States of America*, 115(14), E3313–E3322. <https://doi.org/10.1073/pnas.1801614115>
- Yueh, B., Shapiro, N., MacLean, C. H., & Shekelle, P. G. (2003). Screening and management of adult hearing loss in primary care: Scientific review. *JAMA*, 289(15), 1976–1985.
<https://doi.org/10.1001/jama.289.15.1976>
- Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S. M. G., & Houtgast, T. (2009). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognition system. *Ear and Hearing*, 30(2), 262–272. <https://doi.org/10.1097/AUD.0b013e3181987063>
- Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8), 1488–1498.
<https://doi.org/10.1162/jocn.2009.21140>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, 27(2), 410–417.
<https://doi.org/10.1037/a0024816>

MUSIC PERCEPTION WITH HEARING AIDS: COMPRESSION DIFFERENTIALLY AFFECTS MUSICAL SCENE ANALYSIS AND MUSICAL SOUND QUALITY

Submitted to *Trends in Hearing* on 5. December 2024

Accepted 20. February 2025

by

Robin Hake^{1§}, Michel Bürgel^{1§}, Christophe Lesimple², Matthias Vormann³,
Kirsten C. Wagener³, Volker Kühnel², Kai Siedenburg^{1,4}

¹ Dept. of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Germany

² Sonova AG, Switzerland

³ Hörzentrum Oldenburg gGmbH, Germany

⁴ Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

§ Shared first authorship

Disclaimer

The content of this chapter is identical to the manuscript currently under review for publication at *Trends in Hearing*.

The citation for the following publication is:

Hake, R., Bürgel, M., Lesimple, C., Vormann, M., Wagener, K. C., Kühnel, V., & Siedenburg, K. (2025). Music perception with hearing aids: Compression differentially affects musical scene analysis and musical sound quality. *Trends in Hearing* (accepted manuscript).

Author contributions

RH, DM and KS designed the study and analysed the data. RH wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.



Robin Hake



Kai Siedenburg

6.1. Abstract

Hearing aids have traditionally been designed to primarily facilitate speech perception. With regards to music perception, previous work indicates that hearing aid users frequently complain about music sound quality. Yet, the effects of hearing aid amplification on musical perception abilities are largely unknown. This study aimed to investigate the effects of hearing aid amplification and dynamic range compression (DRC) settings on musical scene analysis (MSA) abilities and sound quality ratings (SQR). Additionally, speech reception thresholds in noise (SRT) were measured. Thirty-three hearing aid users with moderate to severe hearing loss participated in three conditions: unaided, and aided with either slow or fast DRC settings. Overall, MSA abilities, SQR and SRT significantly improved with the use of hearing aids compared to the unaided condition. Yet, notable differences were observed regarding the choice of compression settings. Fast DRC led to better MSA performance, reflecting enhanced selective listening in musical mixtures, while slow DRC elicited more favourable SQR. Despite these improvements, variability in amplification benefit across DRC settings and tasks remained considerable, with some individuals showing limited or no improvement. These findings highlight a trade-off between scene transparency (indexed by MSA) and perceived sound quality, with individual differences emerging as a key factor in shaping amplification outcomes. Our results underscore the potential benefits of hearing aids for music perception and indicate the need for personalised fitting strategies tailored to task-specific demands.

6.2. Introduction

Hearing aids have traditionally been designed to facilitate speech perception, aiming to improve clarity and intelligibility (e.g., Popelka et al., 2016). This emphasis is rooted in the critical importance of speech communication in daily life, which has driven technological advancements and design priorities in hearing aid development. However, music perception presents a set of challenges and requirements that are not always adequately addressed by hearing aids optimised for speech. This has led to frequent complaints from hearing aid users about the poor sound quality for music, often reporting that amplified music sounds unnatural or distorted (e.g., Greasley et al 2020; Vaisberg et al., 2019). These complaints suggest that current hearing aid technologies may not fully accommodate the different set of acoustic demands, such as the wide frequency or dynamic range inherent to many classes of musical sounds (Chasin & Russo, 2004; Chasin, 2012; Lesimple et al., 2024). To address this gap, the present study aims to systematically investigate how different hearing aid settings impact perceived sound quality and behavioural measures of music perception. That is, by complementing sound quality ratings with selective listening tasks, this study seeks to provide a deeper understanding of the auditory experience of music among hearing aid users.

6.2.1. *Sound quality with hearing aids*

While surveys have revealed some positive impacts - with 41% of respondents in one study reporting that hearing aids improve music enjoyment (Leek et al., 2008) and another showing that hearing aids helped with both live and reproduced music (Madsen & Moore, 2014) - these findings are not universal. A survey conducted among professional and amateur musicians in German orchestras (Hake et al., 2024) found that 54% of regular hearing aid users rated the quality of their devices as 'rather not sufficient' for making music, and 40% found them inadequate for music listening. Further, Looi et al. (2019) found that those with moderate to severe hearing loss tended to report less enjoyment and perceived music as less melodic when using their devices. Another survey conducted by Greasley et al. (2020) underscores these challenges, reporting that hearing aid users frequently experience feedback, distortion, and difficulties in live music settings - with 67% hearing aid users expressed to have difficulty listening to music. The cumulative impact of these issues has led many hearing-impaired musicians to prefer the sound quality and dynamics experienced without hearing aids, often removing their devices during musical performances (Vaisberg et al., 2019). This tendency of avoidance is further exacerbated by recommendations from some audiologists to forgo these devices for music listening altogether in cases of mild to moderate hearing loss (Greasley et al., 2020; Vaisberg, 2019). This range of experiences, from improved enjoyment to active avoidance, illustrates the need for a deeper understanding of the factors influencing music perception with hearing aids.

A critical component of modern hearing aids that may influence music perception is dynamic range compression (DRC). DRC provides level-dependent amplification tailored to the individual's hearing loss and compensates for the reduced dynamic range associated with sensorineural hearing loss—a phenomenon known as loudness recruitment (Steinberg & Gardner, 1937). The resulting adjustment of amplification enables soft sounds to become

audible, while loud sounds are prevented from becoming uncomfortably loud. Slow DRC systems adjust the gain gradually, maintaining the natural temporal envelope of the input sounds. This helps preserve signal clarity and reduces harmonic and intermodulation distortion. Additionally, slow DRC preserves short-term level changes, aiding sound localization based on interaural level differences (Moore, 2008). However, these systems have drawbacks. For example, slow DRC may struggle with alternating voices of different levels and has a delayed gain adjustment (1-2 sec), which may be problematic in situations with sudden large level changes. Therefore, it offers limited benefit in "listening in the dips" (perceiving "glimpses" of a target; Duquesnoy, 1983), because the gain does not increase significantly during brief dips in the input signal; a situation particularly problematic among hearing-impaired individuals, who already have a reduced ability to perceive sounds during brief dips in background noise (Duquesnoy, 1983; Peters et al., 1998, Siedenburg et al., 2020). Slow DRC are usually implemented in optional music modes for hearing devices (Moore, 2016).

Fast DRC systems, on the other hand, adjust the gain rapidly, which can more closely approximate normal loudness perception and effectively compensate for frequency-dependent changes in loudness recruitment. Fast DRC can also better restore the audibility of weak sounds following intense sounds, facilitating the potential for listening in the dips. These systems perform well when two sources alternate with markedly different levels. In speech recognition it has been shown that intelligibility at low speech levels is improved with fast DRC compared to slow DRC (Kowalewski et al., 2018). However, fast DRC can alter the natural shape of the temporal envelope of sounds and may distort sounds gliding in frequency if the DRC channels are formed using sharp, non-overlapping filters (Moore, 2008). Additionally, it can make background noises sound noisier, which can be annoying for new hearing aid users, and may introduce cross-modulation between voices in a mixture, reducing the ability to perceptually segregate streams (Madsen et al., 2015).

Previous research has provided insights into how dynamic range compression (DRC) settings influence perceived music quality among hearing aid users. Croghan et al. (2014) reported that listeners generally preferred linear amplification or slow-acting DRC over fast-acting settings, particularly highlighting that fast compression adversely affected music preference ratings due to distortions in both temporal and spectral features. Similarly, Madsen et al. (2015) found that clarity ratings for individual musical instruments decreased under DRC conditions compared to linear amplification, although no significant overall difference between fast and slow compression speeds was observed. However, considerable individual variability was evident, with some participants favouring slow-acting compression, potentially due to less alteration of temporal envelope shapes and spectral contrasts. These findings suggest that slow-acting compression may offer perceptual advantages for music listening, though listener preferences appear to be influenced by individual auditory profiles and the acoustic characteristics of the music.

Despite these contributions, several limitations remain. For instance, Madsen et al. (2015) evaluated clarity ratings for isolated instruments rather than full musical mixtures, and presented these under controlled yet artificial playback conditions with limited genre variation. Similarly, Croghan et al. (2014) employed short, pre-selected excerpts rather than continuous or multitrack recordings that more accurately reflect typical music listening

environments. Furthermore, they relied on simulated hearing aids, limiting the ability to assess how individual listeners respond to different DRC settings across aided and unaided conditions. Neither study examined the extent to which fast versus slow DRC settings compare directly to unaided listening, a comparison that is particularly relevant in light of reports that some hearing aid users prefer to listen to music without amplification (e.g., Greasley et al., 2020; Vaisberg et al., 2019). Moreover, both studies focused exclusively on subjective preference ratings, which provide valuable insight into listener impressions but offer limited information about the functional consequences of DRC settings for selective auditory processing. A follow-up investigation should include ecologically valid, genre-diverse music stimuli, adopt within-subjects comparisons across aided and unaided listening, and integrate both subjective and behavioural outcome measures.

6.2.2. Auditory scene analysis in music

A major component in behavioural hearing aid research relates to the ‘cocktail party problem’, which refers to the ability to disentangle concurrent sounds that overlap in time and frequency, for example, separating a single voice from background noise and interfering speakers (McDermott, 2009). This ability is rooted in the process by which listeners organize complex acoustic environments into meaningful auditory objects and streams, a process known as ‘Auditory Scene Analysis’ (ASA; Bregman, 1990). In the context of multi-source music, ASA allows listeners to segregate simultaneously occurring musical sounds, such as a solo oboe melody within an orchestra or a tenor voice in a choir. This application of ASA in the context of music is known as Musical Scene Analysis (MSA; Bregman, 1990). Crucially, this perceptual process is not merely a supportive mechanism but enables the emergence of core musical structures such as melody and harmony in the first place (Huron, 2001). Accordingly, MSA underpins the very organisation and perception of musical structure and is fundamental for the experience of musical textural layering, forms a perceptual foundation for musical clarity, and the ability to selectively attend to specific musical elements within a musical piece (Bey & McAdams, 2002; Huron, 2001).

For normal-hearing individuals, MSA functions rather effortlessly, allowing rich musical experiences where distinct instruments and melodies can be appreciated individually and as a harmonious whole (still substantial interindividual variability in MSA ability has been observed even among normal-hearing populations, see e.g., Siedenburg et al., 2020). Though, this process heavily relies on the integrity of fundamental auditory mechanisms to perceive and discriminate elementary acoustic information in the environment. Hearing impairments, such as sensorineural hearing loss, disrupt this processing by compromising abilities like frequency resolution, temporal processing, and spatial hearing (Akeroyd, 2007; Phillips, 2000; Middlebrooks, 2017), which are all crucial for MSA. As a result, the rich, layered structure of music may collapse into a rather undifferentiated auditory percept, diminishing clarity and blurring timbral contrasts. This may negatively impact both perceived sound quality and musical enjoyment. Tinnitus, a condition characterised by the perception of sound without external stimuli (Møller, 2006), can further impair ASA abilities beyond those caused by hearing loss (Durai et al., 2018, 2019), amplifying this problem even further.

Recent empirical work supports the role of MSA in shaping the perceptual experience of music. For example, Benjamin and Siedenburg (2025) demonstrated that hearing-impaired listeners with higher MSA performance tend to rate the sound quality of musical excerpts more favourably, suggesting that the capacity to resolve individual streams may enhance the aesthetic evaluation of the music and potentially listener satisfaction. Nonetheless, the relationship between MSA and perceived sound quality may not be entirely linear. In a survey of hearing aid users, Madsen and Moore (2014) found that although approximately one-third of respondents reported improved clarity of musical instruments when using dedicated music programs in their hearing aids, no consistent differences in overall tonal quality were reported. This suggests a more complex and potentially listener-specific interaction between ASA processes and subjective sound quality evaluations. Taken together, these findings underscore the need to assess MSA abilities behaviourally, alongside subjective ratings, to gain a more comprehensive understanding of how hearing aid processing strategies affect music perception in ecologically valid listening contexts, particularly those involving complex, dynamically interacting sound sources.

Numerous tests have been developed to quantify (hearing-impaired) listeners' ASA abilities in the speech domain, including the Oldenburg sentence test by Wagener et al. (1999) and the Göttingen sentence test (Kollmeier & Wesselkamp, 1997). In contrast, while several perceptual tasks exist for assessing specific aspects of music perception, such as mistuning detection (Larrouy-Maestri et al., 2019), beat perception (Harrison & Müllensiefen, 2018), or melody discrimination (Harrison et al., 2017), very few probed MSA processes specifically. Even those that do address MSA processes have typically been developed and validated exclusively with normal-hearing populations only, thereby leaving the perceptual challenges faced by hearing-impaired individuals largely underexplored. The 'Adaptive Music Perception' test by Kirchberger and Russo (2015) constitutes a notable exception; however, the subtest related to MSA yielded inconclusive results, with some hearing-impaired participants unable to complete the task successfully. To address this gap, Hake et al. (2024) introduced a new adaptive MSA paradigm designed to efficiently assess musical scene analysis abilities - including among hearing-impaired individuals - by requiring detection of target instruments embedded in recorded polyphonic mixtures from diverse musical genres. Nevertheless, this paradigm has not yet been applied to individuals using hearing aids, leaving open important questions about the influence of amplification and signal processing strategies on MSA performance.

6.2.3. *The current study*

The present study investigates the impact of hearing aid amplification and DRC settings on both MSA abilities and perceived sound quality among hearing aid users. Specifically, this research aims to evaluate how different hearing aid conditions - fast DRC settings and slow DRC settings - affect the ability to analyse musical scenes and the subjective quality of musical excerpts compared to that of not wearing hearing aids. To supplement these measures, speech reception thresholds (SRT) were also assessed to determine the relationship between individual speech intelligibility in noise and MSA ability scores. Overall, amplification by hearing aids is expected to outperform the unaided condition for SRT and music sound quality

assessments. Although hearing aid users often report mixed experiences with music listening, a similar improvement is anticipated for MSA abilities. We further hypothesise that hearing aids with fast DRC settings yield the best performance in MSA and SRT, while slow compression, in contrast, result in higher sound quality ratings compared to fast DRC settings (see Moore, 2008).

6.3. Methods

6.3.1. Participants

This study involved 33 participants with moderate to severe sensorineural hearing loss (mean age = 73.9 years, SD = 10.7 years; 31-78 years range; 17 female). The average duration of hearing aid experience among participants was 18 years and 10 months (SD = 10.8 years; 5-54 years range). To ensure a rather homogeneous sample, participants were required to have used their hearing aids consistently (at least more than 4-8 hours a day) for at least six months for and to have a hearing loss classified within the N3-N4 range according to the Bisgaard profile (Bisgaard et al., 2010). Additionally, asymmetrical hearing loss should not exceed 20 dB for the audiometric frequencies up to 8 kHz and an air-bone gap not exceeding 15 dB for frequencies up to 4 kHz (with a single exception permitted outside this range). Individuals with active ear disease in one or both ears, contraindications to traditional hearing devices (i.e., anatomical anomalies, or other health-related factors that render the use of conventional air-conduction hearing aids inadvisable, inappropriate, or potentially harmful for a patient), or medical conditions requiring clearance for hearing aid fitting (e.g., mastoid cavity) were excluded. Individual hearing losses are illustrated in Figure 1.

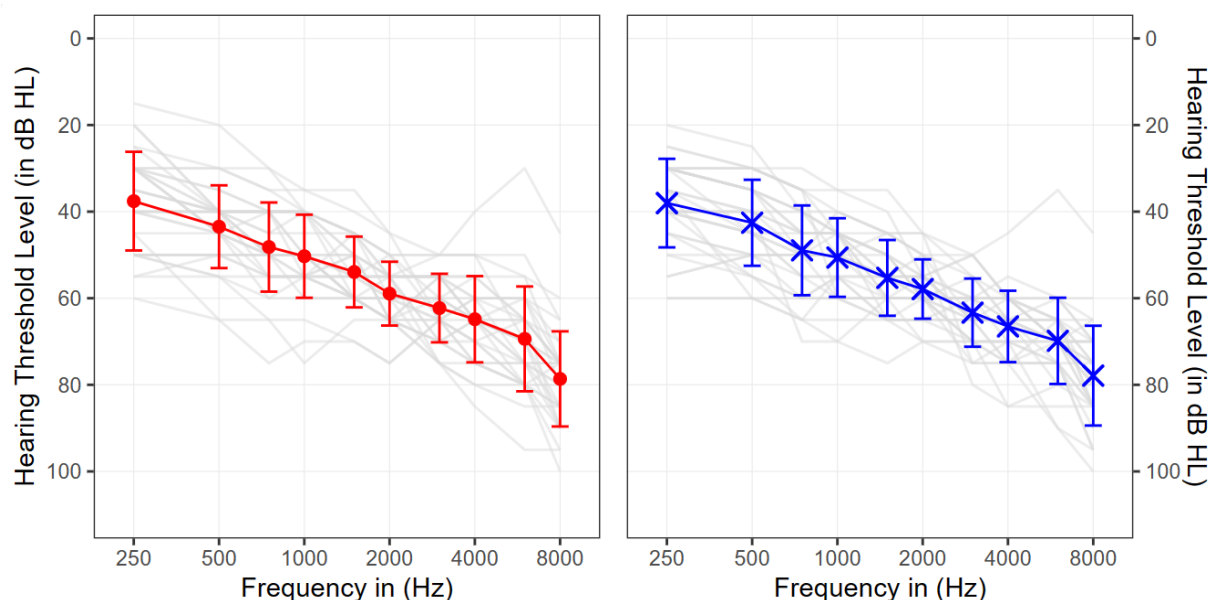


Figure 6.1 Average pure-tone audiometric threshold (PTA) measurements across participants for both ears (red = right ear; blue = left ear). The grey lines show individual participants' threshold. The bold solid lines are representing the pure-tone average over all participants for each listening group.

6.3.2. *Materials*

Audiometry. Hearing sensitivity was assessed in 5 dB HL steps using a PC-controlled clinical Madsen Astera audiometer (Firmware version 1.5.18). Pure-tone thresholds were measured at .125, .25, .5, 1, 2, 4, and 8 kHz. Given the significance of elevated hearing thresholds at frequencies above 4 kHz for complex ASA listening scenarios (Moore, 2020; Narne et al., 2023) pure-tone-average (PTA) was calculated across all measured frequencies.

Musical Scene Analysis test (MSA). The adaptive Musical Scene Analysis test (Hake et al., 2023) is a listening task developed to assess participants' selective listening abilities for realistic music stimuli. The MSA adopts a 'yes-no' paradigm that resembles a 2-alternative-forced-choice task. In each trial, participants listen to an 8-second excerpt of the full musical piece, containing all the instruments and voices included in the original recording. This is followed immediately by a 2-second excerpt of a single instrument or voice (the target), a 1-second silence, and finally a 2-second multi-instrument excerpt (the mixture). Participants had then to decide whether the target was part of the mixture or not. The 8-second presentation of the full musical piece provided sufficient time for the hearing aid's compression system to adapt to the acoustic environment. The task was varied in difficulty by manipulating three conditions: (1) the number of instruments in the final mixture (three or six instruments), (2) the choice of the target instrument (bass, guitar, lead-vocals, and piano), and (3) the target-to-mixture level ratio (ranged from -15 to 0 dB). All MSA stimuli were presented monaurally and without spatial separation of sources. The test initiates with a randomly selected excerpt of medium difficulty. Subsequently, the test adapts to the listener's performance, becoming progressively more challenging or easier as appropriate. This approach enables a reliable estimation of the participant's ability level with a short testing time. The standard test length of 30 items was applied. Participant ability is estimated using weighted-likelihood methods on a scale of -4 to +4; higher scores signify better performance. A score of 0 indicates performance equal to the median of the reference sample ($SD = 1$), which predominantly incorporated normal-hearing individuals (see Hake et al., 2023). Stimuli are sourced from an open-source music-database ("MedleyDB", Bittner et al., 2014; 2016), which consists of real-world multitrack music recordings representing a wide range of musical genres (e.g., pop, rock, world/folk, fusion, jazz, rap, classical). The musical pieces contained in this database are, in general, not widely known to the public. A complete list of the excerpts used, along with the source code of the test, is publicly available (see github.com/rhake14/MSA). The current study employed MSA (long) version 2.8.3.

Sound quality rating task (SQR). Participants were asked to evaluate the quality of thirty 30-second musical excerpts drawn from five pieces of both pop and classical genres. To ensure comparability with the stimuli used in the MSA task, excerpts from the pop genre were also sourced from the MedleyDB. In contrast, classical excerpts were drawn from commercial recordings and included works by Beethoven, Haydn, Mendelssohn, Mozart, and Salieri. These compositions were selected to represent well-known repertoire with diverse orchestration, providing a comprehensive basis for evaluating hearing aid processing strategies across a range of acoustic and structural features. A complete list of all musical excerpts used in both tasks is provided in Table A1 in the Supplementary Materials (hereafter referred to as 'Supp. Mat.'). Each excerpt's sound level was normalised in root-mean-square (RMS). Quality ratings

were conducted across five dimensions, adapted from the 'Dimensions of Quality Ratings for Music' questionnaire by Davies-Venn et al. (2007; see also Gabrielsson et al., 1988), using a 10-point Likert scale with no neutral option; higher values indicated a more pronounced intensity of the respective dimension. The dimensions evaluated were loudness (*'How loud was the sound?'*, with very weak/faint to very loud), sharpness (*'How sharp/shrill or soft was the sound?'*, very soft to very sharp), fullness (*'How full was the sound?'*, very thin to very full), overall quality (*'How good was the overall sound quality of the music?'*, bad to good), and clearness (*'How clearly were you able to distinguish between the individual instruments?'*, unclear to very clear). These quality rating dimensions have been validated for the use with elderly hearing aid wearers (Narendran & Humes, 2003). Each dimension was presented in German, translations are provided in Table A6.2 in 'Supp. Mat.'.

Speech reception thresholds (SRT). Participants' speech reception thresholds were measured using the Göttingen Sentence Test (GOESA; Kollmeier & Wesselkamp, 1997). The GOESA is a speech intelligibility in noise test that employs a standard adaptive procedure, where the masker level remains constant and the speech level is adjusted to determine the 50% speech reception threshold. The test includes 20 everyday sentences per list selected from a pool of 200 sentences, each containing three to seven words, spoken by a male speaker and embedded in speech-shaped white noise. Lower SRT scores (expressed in dB) indicate better performance.

Musical training. To evaluate the individual's level of musical training, the German version of the Goldsmith Musical Sophistication Index (GMSI; Müllensiefen et al., 2014) was employed. The GMSI is a brief self-report questionnaire designed to assess various aspects of musical expertise, including subscales for active engagement and perceptual abilities. For this experiment, only the subscore for musical training was used, which comprises 7 items. Participants responded to questions on a 7-point Likert scale (e.g., *'I engaged in regular, daily practice of a musical instrument (including voice) for_ years.'*, with 0 / 1 / 2 / 3 / 4-5 / 6-9 / 10 or more; *'I have never been complimented for my talents as a musical performer'* with 1 = Completely Disagree to 7 = Completely Agree [reverse coded]). The final composite score ranges from 1 to 7, with higher scores indicating higher level of musical training.

6.4. Procedure

This study was conducted between February and April 2024. Participants were recruited by Hörzentrum gGmbH. The experiment was administered using MATLAB and psychtestR (Harrison, 2020), an R package for creating web-browser-based behavioural experiments. All instructions were presented in German. Participants had prior experience with auditory experiments and were compensated on an hourly basis. Following an initial audiogram measurement using a standard clinical ascending-descending procedure, a total of three hearing test were conducted: First, three sets of the adaptive Musical Scene Analysis test (MSA; Hake et al., 2023) were performed (30 items per set), each presented under a different hearing aid condition: (1) without hearing aids, (2) with hearing aids utilizing fast DRC settings, and (3) with hearing aids utilizing slow DRC settings. The second assessment involved sound quality ratings (SQR), evaluating the perceived sound quality of 30-second

excerpts stemming from five pop and five classical music pieces. A total of 30 items was presented, with 3 items extracted from each music piece and assigned to one of the hearing aid conditions. The third test measured speech intelligibility in noise (SRT) - likewise conducted three times, in each hearing aid condition respectively. All tests were presented via a single frontally positioned loudspeaker at a long-term sound level of 70 dB SPL(A). The order of conditions and the specific stimuli presented in each condition was counterbalanced across participants, and the aided conditions were blind to the listeners. Importantly, prior to data collection, all participants completed structured training sessions for each of the three listening tasks to ensure full comprehension of the task procedures. Specifically, for the sound quality assessment, participants were provided with definitions of each quality attribute, accompanied by example stimuli that were artificially manipulated to represent exaggerated instances of each perceptual dimension (clearness, sharpness, etc.). This training procedure was implemented to ensure that participants correctly understood the intended interpretation of the rating scales, minimised variability arising from inconsistent internal standards, and were thereby able to make appropriate and informed judgements during the rating task. Finally, participants completed the standard demographics questions and the musical training subscale of the Goldsmith-Musical Sophistication Index (Müllensiefen et al., 2014). Participants were encouraged to take breaks between tests, and additional breaks were offered as needed. In total, the data collection took about 2 hours. This study was approved by the Carl von Ossietzky University Oldenburg ethics review board (Drs.EK/2021/031-05), and all standard practices, including informed consent, were followed prior to data collection.

6.4.1. Hearing aid fitting

Participants were binaurally fitted with target fitting software for Audéo L90-R receiver in the canal hearing aids from Phonak. Amplification was defined based on a NAL-NL2 fitting rationale and individual hearing thresholds. The choice of NAL-NL2 was motivated by the amount of prescribed compression necessary to provide effective differences between the tested conditions for the given range of hearing losses. Acoustical coupling was made with standard power domes to maximise the ratio between amplified and direct sound. Frequency lowering was disabled as well as all the noise cleaning algorithms as they are mainly designed for clean speech or speech-in-noise listening situations. Microphone modes were set to 'Real Ear Sound', a beamformer pattern close to the natural pinna effect. Critical gain was measured individually, and feedback canceller was left on the default value. The DRC modes were aligned with those outlined by Chen et al. (2021), where the fast-acting compressor operated with an attack time of 10 ms and a release time of 60 ms, and the slow-acting compressor used an attack time of 1.2 s and a release time of 7.1 s.

6.5. Analysis

Statistical analyses were conducted using both MATLAB and R statistical software (v.2022.07.2+576; RStudio, 2020). Mixed effects models, including random effects for participants, were employed, and contrasts were used to mitigate the risk of p-value inflation

(alpha is set to .05). Effect sizes and confidence intervals, in addition to p-values, were reported to provide a comprehensive understanding of the results. Correlational analyses were conducted using Pearson's correlation for continuous data and Spearman's rank-order correlation for relationships between continuous and ordinal data. The Chi-squared test was applied to categorical variables to ascertain statistical significance. Data point exclusion was based on Cook's distance. Specifically, three data points, which were spatially separated from the others (see Figure A1 in 'Supp Mat'), showed exceptionally poor performances close to chance level. Consequently, these data points were replaced using multiple imputation by chained equations with the *mice* package (v.3.16.0) in R. Specifically, pooled estimates using Bayesian linear regression method with PTA, SRTs, and quality ratings were used to predict MSA scores across five datasets (over 50 iterations). This procedure accounts for variability and uncertainty in test scores, ensuring robust imputation estimates.

6.6. Results

6.6.1. Effect of amplification and compression across listening tasks

6.6.1.1. Musical Scene Analysis (MSA) abilities

MSA abilities were found to be lowest in the unaided condition, which served as the reference ($M = -.68$, $SD = .60$; see Figure 2A). When comparing the compression approaches, the fast DRC condition ($M = -.17$, $SD = .61$) outperformed the slow DRC condition ($M = -.50$, $SD = .69$). Notably, when listeners were equipped with their individual preferred DRC settings, the average MSA ability estimate reached near-normal levels ($M = .003$, $SD = .7$; a). A mixed effects model (Model A1) was fitted to predict MSA scores based on listening conditions (unaided, fast DRC, and slow DRC). The counterbalancing of listening condition presentations across participants was not entirely successful, with one order being represented five times more frequently than the others. To address potential order effects, two additional models were fitted, incorporating the order of presentation and its interaction with listening conditions (Models A2 and A3, respectively). All models included random effects for participants and fixed effects for listening conditions, in which the intercept were allowed to vary. The effect of order was found to be non-significant ($\beta = -.07$, $p = .36$). A likelihood ratio test further indicated that including order did not significantly improve model fit, $\chi^2(1) = .88$, $p = .35$, nor did including the interaction of order (see Table A6.3 in Supp. Mat.). Accordingly, Model A1 was used for the analysis.

The effect for amplification was estimated by comparing the unaided condition with the average of the aided conditions under fast and slow DRC settings, while compression effects were evaluated by directly comparing the fast and slow DRC settings. The results indicated significant effects for amplification ($\beta = -.43$, $SE = .16$, $p = .01$, $d = .63$), indicating that the aided conditions showed a significant improvement over the unaided conditions, with a moderate effect size. Furthermore an effect for compression was found ($\beta = -.43$, $SE = .19$, $p = .03$, $d = .51$), with fast DRC showing improved performance over slow DRC settings, with an effect size approaching moderate significance. Overall, there is considerable individual variability in MSA scores, as indicated by the random effects, with a variance of .10 and a standard deviation of .32.

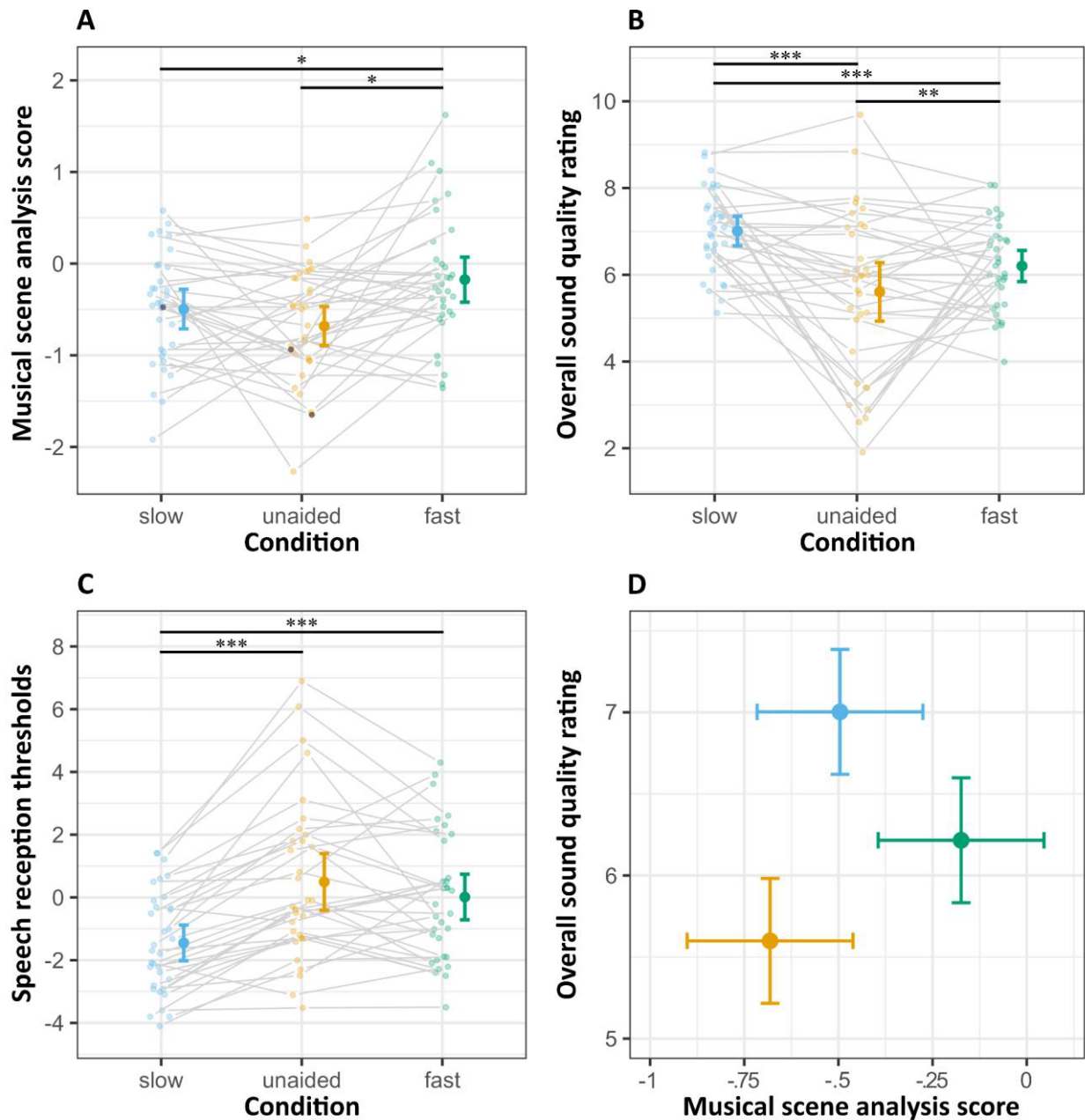


Figure 6.2 Differences in mean (A) musical scene analysis (MSA) scores, (B) overall sound quality rating (SQR) scores, and (C) speech reception thresholds (SRT in dB SNR) across different conditions: slow (blue), unaided (yellow), fast (green). Error bars represent 95% confidence intervals. Individual participant data are displayed as shaded dots. Panel D shows the estimated marginal means and the relationship between MSA scores and SQR scores for each condition (comparisons between SQR and SRT and SRT and MSA can be found in Figure A2 in Sup. Mat.). Model summaries significance between conditions can be found in Table A6.4, A6.6 and A6.7 in Sup. Mat. Note significance at $*p < .05$, $**p < .01$, $***p < .001$.

6.6.1.2. Sound quality ratings

To evaluate overall sound quality, the ‘overall quality’ item was used as the primary measure. Overall quality was highly correlated with clearness ($r(988) = .69, p < .001$), and fullness ($r(988) = .59, p < .001$). Moderate correlations were observed between overall quality and

loudness, ($r(988) = .44, p < .001$). In contrast, sharpness showed a lower yet still significant correlation with overall quality with a pearson correlation ($r(988) = .2, p < .001$). These results are presented in detail in the correlation matrix Table A6.5 and Figure A6.3 (see Supp. Mat.).

Mixed-effects models were fitted to predict the overall sound quality scores based on listening condition (unaided, slow, fast) and musical genre (classical vs. pop) to assess whether adding musical genre, as well as the interaction between condition and genre, revealed meaningful effects or improved the model fit. Four models were compared to determine the most suitable fixed and random effects structure for the data. The first model (B1) included only the listening condition as a fixed effect, while the second model (B2) included both the listening condition and musical genre as fixed effects. The third model (B3) further incorporated the interaction between listening condition and musical genre as a fixed effect. All three models (B1–B3) included random intercepts for both participants and stimuli. In the fourth model (B4), the random slope for listening condition was allowed to vary by stimulus; however, this model encountered boundary issues, leading to a singular fit, and was consequently discarded. Model comparison indicated that the second model (B2), which included listening condition and musical genre as fixed effects, provided the best fit. Detailed model comparison is available in Table A6.3 (Supp. Mat.).

Overall, the mean factor rating for the slow condition was the highest at 7 ($SD = 1.45$), followed by the fast condition at 6.22 ($SD = 1.67$), and the unaided condition had the lowest mean rating at 5.6 ($SD = 2.24$; see Figure 2B). Using contrast specifications similar to the analysis conducted for the MSA scores, an effect for hearing aid amplification was found, indicating that the average of both aided conditions was rated higher than the unaided condition, with ($\beta = 1.01, SE = .1, p < .001, d = .66$). This effect size reflects a substantial improvement in perceived sound quality. Further contrast analysis revealed a significant compression effect (slow vs. fast DRC), with the slow condition rated higher than the fast condition ($\beta = .514, SE = .1, p < .001, d = .51$), indicating a clear preference for the slow condition with a moderate effect size. Similar to the MSA, SQR scores exhibit substantial individual variability, with the random effects model indicating a variance of .87 ($SD = .93$) across participants.

There were also differences in variability across excerpts; however, compared to the subject-specific individual differences, this variability was relatively low, with a variance of only .09 ($SD = .29$). When examining stimulus-specific effects, aided listening conditions generally showed improvements in sound quality perceptions, with slow DRC often outperforming fast DRC, yet some stimuli received similar ratings across all conditions, indicating minimal impact of processing (see Figure 3). Notably, in certain cases, unaided listening was even preferred over fast DRC, suggesting that the benefits of DRC may be stimulus-dependent. Despite this, the model identified a significant main effect of genre, with pop excerpts receiving less favourable ratings ($M = 6.11, SD = 1.97$) compared to classical excerpts ($M = 6.44, SD = 1.82$). This main effect was statistically significant ($\beta = -.33, SE = .15, p = .023$) but had a small effect size ($d = .22$). For a detailed descriptive comparison between genres and sound quality attributes, see Figure A4 and A5 (in Supp. Mat.).

6.6.1.3. *Speech reception thresholds*

The results showed that the unaided condition had the poorest speech-in-noise performance, with the speech reception thresholds (SRT) scoring highest ($M = .49$, $SD = 2.55$; see Figure 2C), followed by the fast DRC condition ($M = .01$, $SD = 2.05$) and the slow DRC condition ($M = -1.45$, $SD = 1.61$). Following the procedure used for MSA and SQR, a mixed effects model was fitted to predict SRT scores and examine the effects of hearing aid amplification and compression. Listening conditions were included as a fixed effect, and random effects allowed the intercept to vary by participants. The contrasts revealed significant effects for amplification, with aided conditions (fast and slow DRC) showing improved SRT scores compared to the unaided condition ($\beta = -1.21$, $SE = .24$, $p < .0001$, $d = 1.08$), reflecting a large effect size. A significant compression effect was also observed, with the slow DRC condition outperforming the fast DRC condition ($\beta = 1.46$, $p < .0001$, $d = .43$), indicating a moderate effect size.

6.6.2. *Amplification benefits across individuals and listening tasks*

The impact of DRC settings on different listening tasks was further explored by examining individual participant performance across tasks. As shown in Figure 3 the effects of DRC on MSA, SQR, and SRT varied across individuals. For SQR and SRT, a strong positive correlation was observed between the benefits from slow and fast compression settings, $r(31) = .93$, $p < .001$, and $r(31) = .72$, $p < .001$, respectively. This pattern suggests that individuals who benefited from one DRC condition tended to benefit from the other, albeit to varying degrees (see Model B2 and C1). More specifically, 23 individuals rated the overall sound quality (i.e., SQR) higher with slow DRC (i.e., individual dots falling below the line of equivalence in Figure 3B), while only three rated fast DRC higher (for an overview over preference for the other quality dimensions see Figure A6 in Sup. Mat.). Seven individuals showed only minor differences between the two conditions, that is within a $\pm 10\%$ margin of one standard deviation benefit distribution, indicating no clear preference (see Figure 3C for individual trajectories). A similar pattern emerged for SRT, with 25 individuals performing better with slow DRC, one performing better with fast DRC, and seven showing indistinct differences (Figure 3E). In contrast, for MSA, the benefits from slow and fast DRC were less consistent across individuals. Nine participants performed better with slow DRC, 18 with fast DRC, and six showed indistinct differences (Figure 3A) and the correlation between slow and fast compression settings for MSA was weak and non-significant, $r(31) = .28$, $p = .11$ (Figure 3B). This highlights that the impact of DRC on MSA varies considerably across individuals compared to SQR and SRT. Although, on average, fast DRC yielded higher scores for MSA, several individuals demonstrated better performance with slow DRC, underscoring the variability in how DRC settings affect MSA across participants.

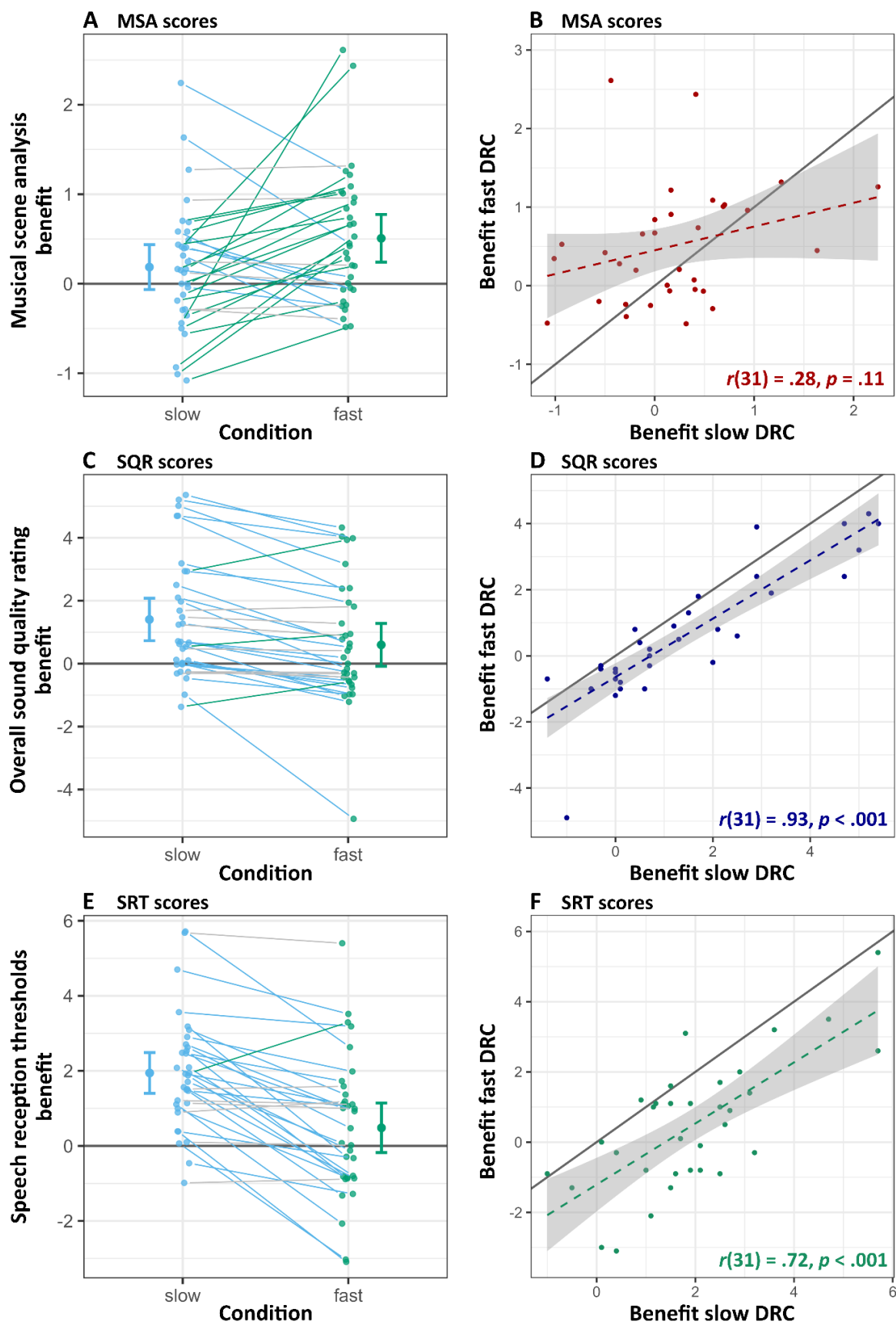


Figure 6.3 Effects of hearing aid compression settings on musical scene analysis (MSA), overall sound quality rating (SQR), and speech reception threshold (SRT) scores. Panel A, C, and E shows individuals benefit scores (that is the unaided condition was either subtracted from the slow or fast DRC condition) of the MSA, SQR, and SRT, respectively. *Figure caption is continued on the following page.*

Figure 6.3 Solid lines connect individual participant scores under DRC conditions, with green lines representing participants who performed better with fast DRC, blue lines for better performance with slow DRC, and grey lines indicating indistinct differences in performance (within $\pm 10\%$ of one standard deviation of the respective benefit distribution). Panels B, D, and F show scatterplots for benefit scores from slow and fast DRC for all tasks. The dashed lines are linear regression fits with 95% confidence intervals shaded in grey, indicating the trend and variability for each test. The solid black diagonal line in each of these panels represents the line of equivalence, where points along this line would indicate equal benefit from fast and slow DRC settings. Pearson correlation coefficients are reported in the lower right corner.

continued

In addition to the differences observed in DRC effects within individual tasks, there was considerable variability in how participants benefited from amplification across tasks. Among all listeners, only four demonstrated consistent preferences for a single hearing aid setting across all tasks, favouring the slow DRC condition (see Figure 4). The remaining participants exhibited varying preferences depending on the specific task. Notably, the strongest alignment was observed between sound quality ratings and speech reception thresholds, with 18 participants showing the highest scores in the slow DRC condition for both tasks.

To further examine the amplification benefit patterns across the conditions, one total benefit score for each task (MSA, SQR, and SRT) was calculated. Specifically, for each task, the benefit of slow and fast DRC was calculated by subtracting the unaided score from the score in each DRC condition. These two values were then summed to yield a total benefit score for each task. Analysis of these total benefit scores revealed no significant correlations between benefits across tasks. The correlation between MSA total benefit and SQR total benefit was weak and non-significant, $r(31) = .12$, $p = .51$ (see Figures A7 in Supp. Mat.). Similarly, no significant relationship was found between MSA total benefit and SRT total benefit, $r(31) = .13$, $p = .48$, nor did the correlation between SQR and SRT total benefit scores reach significance, $r(31) = .24$, $p = .19$. When analysed separately for each DRC setting, no significant correlations were identified across tasks in either the fast or slow DRC conditions (see Figure A8 in Supp. Mat.). However, a marginally significant relationship was observed between SRT and SQR benefits under the slow DRC condition, $r(31) = .33$, $p = .06$. This finding is consistent with the pattern observed in the heatmap (Figure 4), which similarly indicated the highest agreement in preferences in favour of the slow DRC condition between SRT and SQR tasks.

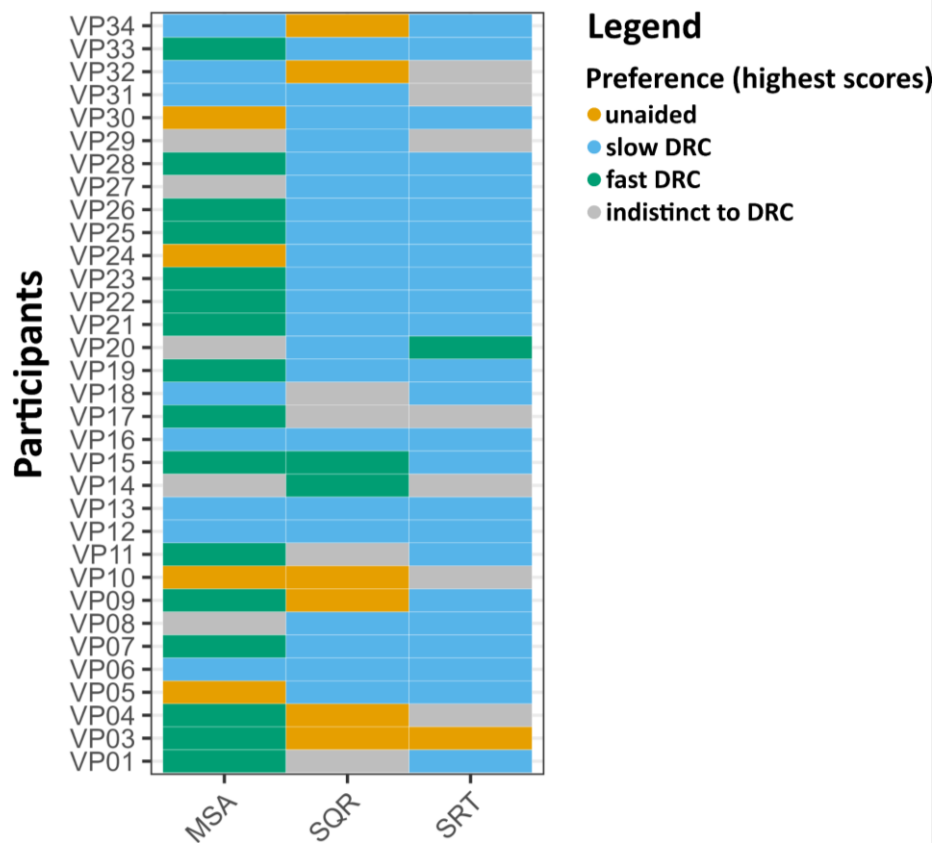


Figure 6.4 Individual preferences (heatmap) for all three conditions (unaided vs. aided fast DRC vs. aided slow DRC) across the three tasks: Musical Scene Analysis test (MSA), overall Sound Quality Ratings (SQR), and Speech Reception Thresholds (SRT). Each row represents a participant, and the columns correspond to the respective tasks. Colours indicate the hearing aid setting associated with the highest scores for each task: pale red (unaided), blue (slow dynamic range compression [DRC]), and green (fast DRC). Grey represents indistinct preferences, where scores fell within a $\pm 10\%$ margin of one standard deviation of the respective benefit distribution, indicating no clear preference for a specific condition.

6.6.3. Individual differences in hearing aid amplification benefit

To explore potential sources of individual variability in amplification benefit, associations were examined between a composite total amplification benefit score and individual factors, including musical training, age, and hearing aid use history. Each task-specific benefit score was first standardised, then summed across tasks to create a total amplification benefit score that reflects overall benefit from amplification. Total amplification benefit did not correlate with musical training ($r(31) = -.19, p = .3$) or age ($r(31) = -.13, p = .47$). Similarly, no significant group differences between males and females were found ($F(1, 31) = 2.18, p = .15$). However, there was a moderate positive relationship between years of hearing aid use and total benefit ($r(31) = .45, p = .001$; see Figure 5), indicating that participants with longer durations of hearing aid use tended to experience greater benefits. There was a moderate positive relationship between years of hearing aid use and total benefit ($r(31) = .45, p = .001$), with greater benefit also associated with higher hearing thresholds (PTA) in the better ear ($r(31) = .63, p < .001$), suggesting that participants with longer hearing aid use and more severe hearing loss derived greater amplification benefits (see Figure 6.5).

These findings raise the question of whether the increased benefit is due to adaptation over time or simply reflects further hearing deterioration (i.e., higher PTA), which increases the contrast between aided and unaided conditions. To clarify these relationships, a linear regression model was conducted to assess the impact of years of hearing aid use and PTA on total benefit. The Variance Inflation Factor ($VIF = 1.15$) indicates that multicollinearity was not severe in this model. The regression analysis revealed that both years of hearing aid use ($\beta = .12, p = .02$) and PTA ($\beta = .28, p < .001$) were significant predictors of total amplification benefit. This suggests that both the duration of hearing aid use and the severity of hearing impairment contribute to the perceived benefit of amplification. The model accounted for 42% of the variance in total benefit ($R^2 = .42, p < .001$), underscoring the importance of both factors. No interaction effect was observed between the two predictors ($p = .89$), indicating that the relationship between the duration of hearing aid use and total benefit does not depend on hearing threshold severity.

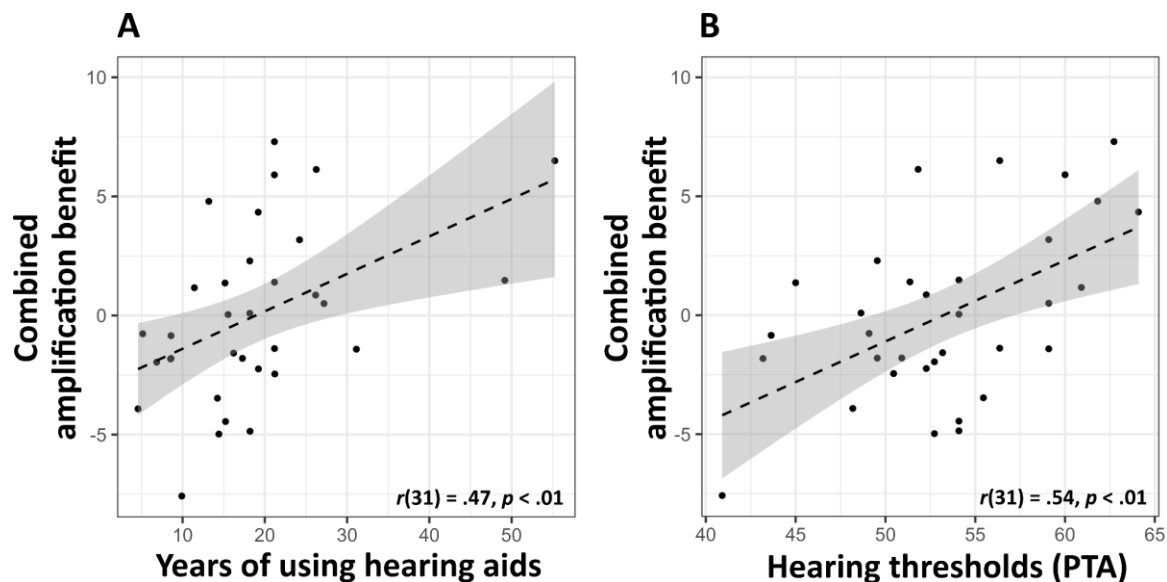


Figure 6.5 Scatterplots illustrating the relationship between individual characteristics and combined amplification benefit, a composite score reflecting the summed standardised benefit across all tasks (MSA, SQR, and SRT) under both DRC conditions. Panel (A) displays combined amplification benefit in relation to years of hearing aid use, and Panel (B) shows combined amplification benefit in relation to hearing thresholds (PTA). Pearson correlation coefficient is provided; the blue line represents the linear regression with a shaded 95% confidence interval.

6.6.4. Simulated hearing aid output target-to-masker ratio of the MSA task

To better understand why fast-acting DRC outperformed the slow-acting DRC settings in the MSA task, the hearing aid output target-to-masker ratio (TMR) of the MSA test material (i.e., the excerpts) was computed by using the inversion technique (Hagerman & Olofsson, 2004). This procedure allows the separation of the hearing aid's output signal, allowing analysis of how each is processed by the device and to compute their respective levels. The principle is to record the same test signal twice: first with the target plus the mixture and then the target minus the mixture. Target and mixture at the output of the hearing aid are then extracted by adding and subtracting the recorded signals, respectively. This was done for all target

instruments from the original pool of stimuli (that is, bass, guitar, lead-vocals, and piano). The reference hearing aid output TMRs were computed for a fitting with 0 dB insertion gain to have a comparable bandwidth with the aided conditions. Hearing aids were fitted with a NAL-NL2 fitting formula for a N4 hearing loss with the fast-acting and the slow-acting DRC.

A consistent pattern was observed (see Figure 6): across all instruments, there is a negative correlation between reference TMR and the difference between fast and slow DRC TMR. The downward trend in each panel indicates that at negative reference TMRs (when the target instrument is more masked within the mixture), fast DRC provides greater benefit, yet the difference between fast and slow DRC settings gradually diminishes, meaning there's little benefit in using fast DRC over slow DRC. At higher reference TMRs (when the target instrument is more prominent in the mixture), slow DRC settings begin to provide more benefit. The average difference favoured fast DRC for bass (.7 dB, 95% CI = [.4, 1]), guitar (.3 dB, 95% CI = [1, .6]), and piano (.3 dB, 95% CI = [.1, .5]), but slow DRC was more beneficial for lead vocals (-.5 dB, 95% CI = [-.8, -.2]).

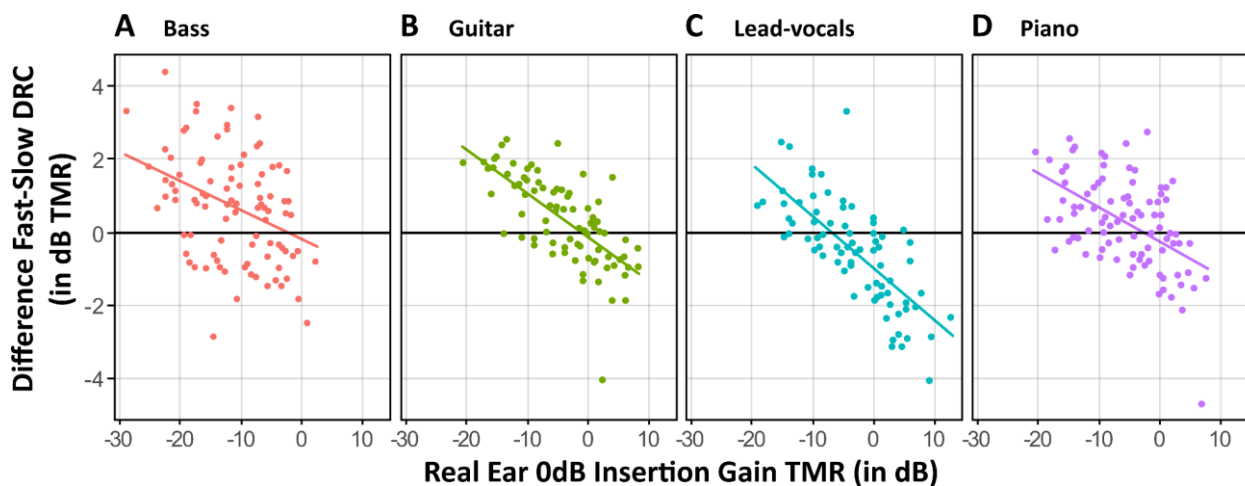


Figure 6.6 Difference in TMR between fast and slow DRC versus real ear TMR as reference for various musical instruments. The graphs show how the effect of compression varies with the reference TMR. Each subplot represents a different instrument: bass, guitar, lead-vocals, and piano. The vertical axis indicates the difference in TMR (dB) between fast and slow DRC where positive values are in favour of fast DRC, while the horizontal axis shows the real ear TMR (dB). The lines represent the linear regression fits for each instrument.

6.7. Discussion

The present study investigated the effects of hearing aid amplification on music and speech perception among individuals with moderate to severe hearing loss. Specifically, we aimed to explore how dynamic range compression (DRC) settings influence musical scene analysis (MSA), perceived musical sound quality (SQR), and speech intelligibility in noise (SRT). The results demonstrated significant improvements in all three listening tasks when participants used hearing aids compared to unaided listening. However, the effects of DRC speed varied across tasks, revealing a nuanced interplay between compression settings and auditory outcomes.

Our primary hypothesis was that slow DRC would lead to superior sound quality ratings due to its ability to preserve natural temporal envelopes and the dynamic range of sounds, thus reducing distortions and enhancing the listening experience (Moore, 2008). This expectation was supported, as slow DRC resulted in the highest overall music sound quality ratings, which were closely associated with enhanced ratings for sharpness, fullness, and clearness of the audio signal. Furthermore, fast compression was expected to improve the audibility of low-level sounds and assist in 'listening in the dips', thereby benefiting auditory scene analysis (ASA) abilities in both speech perception (in noise; e.g., Kowalewski et al., 2018) and music (Siedenburg et al., 2020). Interestingly, however, while fast DRC indeed yielded better results in music listening (as indicated by MSA scores), slow DRC outperformed fast DRC in the speech intelligibility task. This is noteworthy because we had anticipated that fast DRC would enhance both speech reception thresholds and MSA performance. This expectation was based on the shared effects of hearing loss on both speech and music perception (e.g., Hake et al., 2023) and their reliance on common ASA mechanisms, such as object formation, temporal cue processing, and stream segregation. Several potential explanations exist for this outcome.

First, the differential benefit of slow versus fast DRC settings between speech and music outcomes may be due to the distinct, that is the non-overlapping aspects of auditory processing that are uniquely engaged by speech and music. More specifically, music listening may engage different cognitive processing strategies compared to speech (in noise), potentially placing higher demands on attentional resources. However, there is substantial evidence of shared underlying cognitive processes involved in auditory perception across both speech and music domains (e.g., Barros et al., 2017; Shinn-Cunningham & Best, 2008). This overlap is so pronounced that Sallat and Jentschke (2015) have proposed using musical material as a tool for early diagnosis of speech-related impairments. Accordingly, the differing outcomes for MSA and SRT tasks may be attributed to the specific acoustic characteristics of each task instead. For example, compared to speech in noise, music - with its high dynamic range and rapid fluctuations - more often contains short temporal dips that can benefit from rapid gain adjustments to maintain the audibility of all elements (Chasin, 2012; Siedenburg et al., 2020). Fast compression may capture these rapid fluctuations more effectively and thus better restore audibility of musical elements. At the same time, however, fast DRC can amplify noise components in speech gaps, making 'listening in the dips' more difficult in speech-in-noise tasks (May et al., 2020). Furthermore, fast compression can introduce temporal distortions either by altering the natural modulations of the signal or by creating spurious changes in the shape of the temporal envelope (Stone & Moore, 2008; Madsen & Moore, 2014). This trade-off likely contributed to the lower music sound quality ratings observed for fast DRC settings (see also Croghan et al., 2014; Moore, 2008). Conversely, slow DRC adjusts the gain gradually, preserving signal clarity and avoiding abrupt gain changes that could introduce artefacts, effectively maintaining the acoustic integrity of the sound. This contributes to a more natural and pleasant listening experience. Importantly, slow DRC also retains the natural dynamics of the speech envelope, enhancing speech intelligibility, particularly in steady-state noise environments where glimpses are less frequent. Notably, in environments with multiple sound sources, fast DRC has been shown to outperform slow DRC

in enhancing speech intelligibility, likely by capturing auditory glimpses ('dip listening' opportunities), as demonstrated by Salorio-Corbetto et al. (2020) in scenarios with two distractor talkers. Similar results have been observed with negative test SNRs when applying fast-acting compression to speech in the presence of fluctuating, rather than with steady, background noise (Rhebergen et al., 2017; Desloge et al., 2017). This may suggest that DRC benefits for MSA and speech tasks may converge as the acoustic characteristics of the environment match the different tasks' conditions. Other acoustic cues, such as localisation cues, also play a role and may complicate the relationship (Hassager et al., 2017). Future studies should explore more approximate acoustic contexts (e.g., multi-talker environments) between music and speech ASA processes to determine whether comparable DRC effects emerge.

The conflicting evidence regarding the relative benefits of slow and fast compression is well-documented in the speech domain (e.g., see Kates, 2010). Previous studies have suggested that the benefit of slow versus fast DRC further depends on the target-to-mixture ratio (TMR), with fast-acting compression improving audibility at low signal-to-noise ratios (SNRs), but potentially distorting the speech envelope at higher SNRs, leading to reduced intelligibility (Jenstad & Souza, 2005; Stone & Moore, 2008). Our data indicate that for music, the benefit of slow versus fast DRC also depends on the TMR, analogous to SNR in speech tasks (e.g., Rhebergen et al., 2009). Analysis of the TMR hearing aid output revealed that fast DRC was more beneficial than slow DRC when the target instrument was less audible within the mixture (i.e., at low TMRs). In these challenging listening conditions, potentially, the rapid gain adjustments enabled by the fast DRC enhanced the audibility of softer sounds following louder ones, facilitating better separation and identification of individual instruments in complex musical passages. As the target became more prominent in the mix (high TMRs), the difference between fast and slow DRC settings diminished, and slow DRC even became superior due to its ability to preserve sound quality. These results suggest that fast DRC improves audibility when the target signal is softer than the mixture for non-speech instruments, corroborating the behavioural findings from the MSA test.

Individual variabilities in amplification benefits derived from compression settings within the same task further complicate the picture. That is, the preferences for DRC effects does not only depend on the TMR of the stimuli, but also on individual differences. A strong positive relationship was found between the benefit scores in slow and fast DRC conditions for the sound quality ratings, with notable consistency across participants. The same could be observed for the speech-in-noise task. This suggests that, for these tasks, individuals who benefit from one DRC setting also tend to benefit from the other, albeit to differing magnitudes. In contrast, however, the relationship for MSA was weak and insignificant: while fast DRC generally yielded better results on average, many listeners exhibited pronounced benefits with one DRC setting but little to no benefit with the other. Some participants even showed no improvements in MSA performance with amplification under either DRC setting. Accordingly, even within the same individual in the same task, improvements from amplification in one DRC setting do not necessarily translate to improvements in another DRC setting. This might indicate that individual differences factors play a greater role in benefiting from hearing aid usage in the MSA task, possibly due to cognitive factors (e.g., prerequisites of

'listening in the dip' capabilities). This finding highlights the variability in individual responses to different hearing aid settings and the potential task-specific influences on listener preferences.

However, individual differences are not restricted to variability in responses to DRC effects within the same task, but are also observed across tasks. Examination of individual-level outcomes revealed that the amplification benefit in one task does not necessarily correlate with the benefit to another, indicating considerable variability in how participants benefited from amplification across tasks. Specifically, only four participants demonstrated consistent preferences for a single DRC setting across all tasks, with all favouring the slow DRC condition. Most participants, however, exhibited task-dependent preferences, underscoring that the optimal DRC setting varied depending on the specific demands of each task (i.e., cognitively and perceptually). Notably, the highest agreement was observed between sound quality ratings and speech-in-noise tasks, with 18 participants showing the greatest benefit in the slow DRC condition for both tasks. This finding aligns with the results of the correlational analysis, where only the slow DRC condition showed a (marginal) relationship between SQR and SRT benefits. Such patterns suggest that while amplification benefits can differ markedly across tasks, slow DRC may provide more consistent advantages in tasks requiring speech intelligibility and perceived sound quality.

Together, these findings highlight both the importance of individual variability and task-specificity in understanding DRC effects. Importantly, they underscore that DRC influences on listeners' ASA abilities are not only highly task-dependent but also vary significantly across individuals, even within the same task (see also Lunner et al., 1997; Ricketts et al., 2008; Moore et al., 2011). The potential of individualised hearing aid settings is further demonstrated by a comparison of MSA performance between hearing aid users and a reference population from the calibration study (Hake et al., 2023). In the unaided condition, hearing aid users performed at the lower end of the distribution for MSA, with a mean score of -.68, placing them below 75.2% in performance in comparison to the reference population. However, when equipped with their preferred DRC settings, their performance improved substantially, reaching a mean score of .003. This improvement elevated them to a level where only 49.9% of the reference population scored better, indicating that tailored hearing aid settings can restore MSA abilities to near-'normal' levels for many users. Nonetheless, the variability in benefit remains substantial, with some users experiencing no significant improvement or even showing better performance in the unaided condition. This heterogeneity emphasises the critical role of personalization in hearing aid design for enhancing music perception, pointing to the need for deeper investigation into the individual factors that drive these differences.

Indeed, individual differences have been shown to affect compression settings outcomes. For example, despite the well-established dependency of the compression ratio and gain as a function of hearing loss (Scollie et al., 2005), previous research has shown that individuals with higher cognitive abilities are better able to benefit from fast compression, likely due to their ability to take advantage of temporal dips in the background sound (Moore, 2008). In this study, we observed that participants with higher hearing thresholds (i.e., more severe hearing loss) experienced greater benefits from hearing aids across all tasks.

Additionally, previous experience with hearing aids (that is longer history of hearing aid usage) was associated with improved performance, likely reflecting adaptation processes over time, a finding widely supported in the literature. Previous research has shown that long-term adaptation to DRC compression improves not only speech intelligibility but also performance in music-related tasks, such as melody perception (e.g., Uys & Latzel, 2015, 2016). However, it was not clear whether that extends to authentic music. In this study, we were able to demonstrate that these benefits indeed extend to realistic musical sounds.

Importantly, our results showed no effect of age, gender, or musical training on total amplification benefit (see also Knudsen et al., 2010). Additional factors, such as hearing aid usage history (including DRC type) and subclinical auditory processing deficits, such as hidden hearing loss, may also contribute to variability in ASA performance. Future research should explore these individual factors in more detail, including assessments of cognitive abilities, auditory processing skills, and detailed hearing aid experiences. Such investigations could clarify the variability in MSA benefit and inform personalised hearing aid fitting strategies that balance sound quality and selective listening needs across different auditory environments.

6.7.1. Limitations

Several limitations of this study should be acknowledged. First, the participants' degree of familiarity with different DRC settings was not assessed. Differences in, or lack of, acclimatisation to the various DRC modes could have affected performance, as familiarity with certain compression speeds might impact auditory perception. Future studies should incorporate prior exposure to each DRC setting to control for potential adaptation effects. Second, the playback level used during testing was fixed and not varied across conditions. As compression is inherently level-dependent, the use of a single playback level may not capture the full range of effects that DRC settings have on auditory perception. Future studies should include multiple playback levels to better understand how DRC settings interact with sound level to influence performance and subjective experiences. Additionally, this study's operationalisation of musical sound quality was limited. While recent research has identified seven key perceptual attributes of music audio quality for hearing aid users - clarity, harshness, distortion, spaciousness, treble strength, middle strength, and bass strength (Bannister et al., 2024) - we only assessed clarity, sharpness (corresponding to 'harshness'), and fullness, along with an overall quality rating. Incorporating all of these perceptual attributes in future research could provide a more comprehensive understanding of how DRC settings influence music perception. Using these attributes in subjective questionnaires could also complement prior exposure to DRC modes, offering a deeper insight into listeners' perceptual experiences in music listening.

6.8. Conclusion

Our results indicate that a one-size-fits-all approach to hearing aid dynamic range compression settings may not be optimal for music processing. Instead, different auditory tasks may benefit from different compression strategies, and hearing aid settings should be optimised depending on the specific needs and benefits of the listener. Additionally, the observed variability in amplification benefits for music stimuli across participants may reflect individual differences in auditory processing or cognitive abilities, underscoring the importance of personalised hearing aid fitting and tuning. Understanding the interaction between individual differences and hearing aid settings will be critical for optimising auditory outcomes for hearing aid users, particularly in rich musical scenes.

6.9. Declarations

Author contributions. All authors contributed to designing the study. RH, MB, and CL analysed the data. RH and MB wrote the first manuscript draft. All co-authors edited and approved the final version of the manuscript.

Ethics approval. Ethical approval was provided by the Ethics Committee at University of Oldenburg (Drs.EK/2021/031-05).

Consent to participate. Informed consent was obtained from all participants in this study, permitting the collection of anonymized data and its potential publication in journal articles.

Acknowledgements. This work has been funded by the Freigeist Fellowship of the Volkswagen Foundation to KS and was supported by Sonova AG. This work was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330 A6 and A4.

Conflict of interest. This study was financially supported by Sonova AG, which provided prototypes of hearing aids for use in the research. We affirm that the design, execution, and analysis of the study were conducted independently, adhering to rigorous scientific standards, and the findings are presented objectively.

Copyright. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Data Availability Statement. The data supporting the findings of this study are available on GitHub at github.com/rhake14.

6.10. Literature

- Albert, M., Hurrelmann, K., & Quenzel, G. (Eds.). (2019). Shell-Jugendstudie: Band 18. Jugend 2019: *Eine Generation meldet sich zu Wort* (1. Auflage). Beltz.
- Barros, C. G., Swardfager, W., Moreno, S., Bortz, G., Ilari, B., Jackowski, A. P., Ploubidis, G., Little, T. D., Lamont, A., & Cogo-Moreira, H. (2017). Assessing Music Perception in Young Children: Evidence for and Psychometric Features of the M-Factor. *Frontiers in Neuroscience*, 11, 18. <https://doi.org/10.3389/fnins.2017.00018>
- Bannister, S., Greasley, A. E., Cox, T. J., Akeroyd, M. A., Barker, J., Fazenda, B., Firth, J., Graetzer, S. N., Roa Dabike, G., Vos, R. R., & Whitmer, W. M. (2024). Muddy, muddled, or muffled? Understanding the perception of audio quality in music by hearing aid users. *Frontiers in Psychology*, 15, 1310176. <https://doi.org/10.3389/fpsyg.2024.1310176>
- Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, 64(5), 844–854. <https://doi.org/10.3758/bf03194750>
- Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, 14(2), 113–120. <https://doi.org/10.1177/1084713810379609>
- Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., & Bello, J. P. (2014). MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (pp. 155–160). Taipei, Taiwan: International Society for Music Information Retrieval.
- Bittner, R., Wilkins, J., Yip, H., & Bello, J. (2016). MedleyDB 2.0: New data and a system for sustainable data collection. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR-16)*. New York, NY, USA.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- Chasin, M. (2012). Music and hearing aids - an introduction. *Trends in Amplification*, 16(3), 136–139. <https://doi.org/10.1177/1084713812468512>
- Chasin, M., & Russo, F. A. (2004). Hearing AIDS and music. *Trends in Amplification*, 8(2), 35–47. <https://doi.org/10.1177/108471380400800202>
- Chen, Y., Wong, L. L. N., Kuehnel, V., Qian, J., Voss, S. C., & Shangqiguo, W. (2021). Can Dual Compression Offer Better Mandarin Speech Intelligibility and Sound Quality Than Fast-Acting Compression? *Trends in Hearing*, 25, 1-13. <https://doi.org/10.1177/2331216521997610>
- Clement, M., & Kandziora, M. (2021). Tonträgermarkt und Musiknutzung in Deutschland. *Deutscher Musikrat - Deutsches Musikinformationszentrum*(4), 2–15. <https://miz.org/de/beitraege/tontraegermarkt-und-musiknutzung>
- Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, 35(5), e170-84. <https://doi.org/10.1097/AUD.0000000000000056>
- Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, 18(8), 688–699. <https://doi.org/10.3766/jaaa.18.8.6>
- Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & D'Aquila, L. A. (2017). Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability. *The Journal of the Acoustical Society of America*, 141(6), 4452–4465. <https://doi.org/10.1121/1.4985186>
- Dingle, G. A., Sharman, L. S., Bauer, Z., Beckman, E., Broughton, M., Bunzli, E., Davidson, R., Draper, G., Fairley, S., Farrell, C., Flynn, L. M., Gomersall, S., Hong, M., Larwood, J., Lee, C., Lee, J., Nitschinsk, L., Peluso, N., Reedman, S. E., . . . Wright, O. R. L. (2021). How Do Music Activities Affect Health and Well-Being? A Scoping Review of Studies Examining

- Psychosocial Mechanisms. *Frontiers in Psychology*, 12, 713818. <https://doi.org/10.3389/fpsyg.2021.713818>
- Durai, M., O'Keeffe, M. G., & Searchfield, G. D. (2018). A Review of Auditory Prediction and Its Potential Role in Tinnitus Perception. *Journal of the American Academy of Audiology*, 29(6), 533–547. <https://doi.org/10.3766/jaaa.17025>
- Durai, M., Sanders, M., Kobayashi, K., & Searchfield, G. D. (2019). Auditory Streaming and Prediction in Tinnitus Sufferers. *Ear and Hearing*, 40(2), 345–357. <https://doi.org/10.1097/AUD.0000000000000620>
- Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, 74(3), 739–743. <https://doi.org/10.1121/1.389859>
- Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgments and speech intelligibility. *Journal of Speech and Hearing Research*, 31(2), 166–177. <https://doi.org/10.1044/jshr.3102.166>
- Greasley, A., Crook, H., & Fulford, R. (2020). Music listening and hearing aids: Perspectives from audiologists and their patients. *International Journal of Audiology*, 59(9), 694–706. <https://doi.org/10.1080/14992027.2020.1762126>
- Greasley, A., & Crook, H. (2023). Evaluating the impact of Hearing Aids for Music (HafM) resources on practitioners and patients. *BAA Audacity*, 22–27. <https://musicandhearingaids.org/wp-content/uploads/sites/35/2023/04/Evaluating-the-impact-of-Hearing-Aids-for-Music-HafM-resources-on-practitioners-and-patients-BAA-Winter-2023.pdf>
- Hagerman B., & Olofsson, A. (2004). A Method to Measure the Effect of Noise Reduction Algorithms Using Simultaneous Speech and Noise. *Acta Acustica*(90), 356–361.
- Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., & Siedenburg, K. (2023). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods*(56), 5456–5481. <https://doi.org/10.3758/s13428-023-02279-y>
- Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburg, K. (2024). A survey on hearing health of musicians in professional and amateur orchestras. *Trends in Hearing*. <https://doi.org/10.1177/23312165241293762>
- Harrison, P. M. C. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.02088>
- Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7, 1–18. <https://doi.org/10.1038/s41598-017-03586-z>
- Harrison, P. M. C., & Müllensiefen, D. (2018). Development and validation of the computerised adaptive beat alignment test (CA-BAT). *Scientific Reports*, 8(12395), 1–19. <https://doi.org/10.1038/s41598-018-30318-8>
- Hassager, H. G., Wiinberg, A., & Dau, T. (2017). Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. *The Journal of the Acoustical Society of America*, 141(4), 2556. <https://doi.org/10.1121/1.4979783>
- Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences*, 930, 43–61. <https://doi.org/10.1111/j.1749-6632.2001.tb05724.x>
- Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language, and Hearing Research*, 48(3), 651–667. [https://doi.org/10.1044/1092-4388\(2005/045\)](https://doi.org/10.1044/1092-4388(2005/045))
- Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M., & Bullock, R. (2004). Demtect: A new, sensitive cognitive screening test to support the diagnosis of mild

- cognitive impairment and early dementia. *International Journal of Geriatric Psychiatry*, 19(2), 136–143. <https://doi.org/10.1002/gps.1042>
- Kates, J. M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *International Journal of Audiology*, 49(6), 395–409. <https://doi.org/10.3109/14992020903426256>
- Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, 36(2), 217–228. <https://doi.org/10.1097/AUD.0000000000000112>
- Knudsen, L. V., Oberg, M., Nielsen, C., Naylor, G., & Kramer, S. E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature. *Trends in Amplification*, 14(3), 127–154. <https://doi.org/10.1177/1084713810385712>
- Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *The Journal of the Acoustical Society of America*, 102(4), 2412–2421. <https://doi.org/10.1121/1.419624>
- Kowalewski, B., Zaar, J., Fereczkowski, M., MacDonald, E. N., Strelcyk, O., May, T., & Dau, T. (2018). Effects of Slow- and Fast-Acting Compression on Hearing-Impaired Listeners' Consonant-Vowel Identification in Interrupted Noise. *Trends in Hearing*, 22, 2331216518800870. <https://doi.org/10.1177/2331216518800870>
- Krause, A. E., North, A. C., & Hewitt, L. Y. (2015). Music-listening in everyday life: Devices and choice. *Psychology of Music*, 43(2), 155–170. <https://doi.org/10.1177/0305735613496860>
- Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, 51(2), 663–675. <https://doi.org/10.3758/s13428-019-01225-1>
- Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, 19(6), 519–526. <https://doi.org/10.3766/jaaa.19.6.7>
- Lesimple, C., Kuehnel, V., & Siedenbuck, K. (2024). Hearing aid evaluation for music: Accounting for acoustical variability of music stimuli. *JASA Express Letters*, 4(9). <https://doi.org/10.1121/10.0028397>
- Looi, V., Rutledge, K., & Prvan, T. (2019). Music Appreciation of Adult Hearing Aid Users and the Impact of Different Levels of Hearing Loss. *Ear and Hearing*, 40(3), 529–544. <https://doi.org/10.1097/AUD.0000000000000632>
- Lunner, T., Hellgren, J., Arlinger, S., & Elberling, C. (1997). A digital filterbank hearing aid: Three digital signal processing algorithms - user preference and performance. *Ear and Hearing*, 18(5), 373–387. <https://doi.org/10.1097/00003446-199710000-00003>
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18. <https://doi.org/10.1177/2331216514558271>
- Madsen, S. M. K., Stone, M. A., McKinney, M. F., Fitz, K., & Moore, B. C. J. (2015). Effects of wide dynamic-range compression on the perceived clarity of individual musical instruments. *The Journal of the Acoustical Society of America*, 137(4), 1867–1876. <https://doi.org/10.1121/1.4914988>
- May, T., Kowalewski, B., & Dau, T. (2020). Scene-Aware Dynamic-Range Compression in Hearing Aids. In J. Blauert & J. Braasch (Eds.), *Modern Acoustics and Signal Processing. The Technology of Binaural Understanding* (pp. 763–799). Springer International Publishing. https://doi.org/10.1007/978-3-030-00386-9_25
- McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, 19(22), R1024–7. <https://doi.org/10.1016/j.cub.2009.09.005>
- Moore, B. C. J. (2008). The choice of compression speed in hearing AIDs: Theoretical and practical considerations and the role of individual differences. *Trends in Amplification*, 12(2), 103–112. <https://doi.org/10.1177/1084713808317819>

- Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing AIDS and paired comparisons. *Ear and Hearing*, 32(5), 556–568.
<https://doi.org/10.1097/AUD.0b013e31820b5f4c>
- Moore, B. C. J. (2016). Effects of Sound-Induced Hearing Loss and Hearing Aids on the Perception of Music. *Journal of the Audio Engineering Society*, 64(3), 112–123.
<https://doi.org/10.17743/jaes.2015.0081>
- Moore, B. C. J. (2020). The importance of high-frequency hearing in a medico-legal context. *Journal of the Acoustical Society of America*, 148, 2542–2542.
<https://doi.org/10.1121/1.5147057>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS One*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Møller, A. R. (2006). Neural plasticity in tinnitus. *Progress in Brain Research*, 157, 365–372.
[https://doi.org/10.1016/S0079-6123\(06\)57022-0](https://doi.org/10.1016/S0079-6123(06)57022-0)
- Narendran, M. M., & Humes, L. E. (2003). Reliability and validity of judgments of sound quality in elderly hearing aid wearers. *Ear and Hearing*, 24(1), 4–11.
<https://doi.org/10.1097/01.AUD.0000051745.69182.14>
- Narne, V. K., Jain, S., Bharani, Ravi, S. K., Almudhi, A., Krishna, Y., & Moore, B. C. J. (2023). The effect of recreational noise exposure on amplitude-modulation detection, hearing sensitivity at frequencies above 8 kHz, and perception of speech in noise. *The Journal of the Acoustical Society of America*, 153(5), 2562. <https://doi.org/10.1121/10.0017973>
- Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, 103(1), 577–587.
<https://doi.org/10.1121/1.421128>
- Popelka, G. R., Moore, B. C. J., Fay, R. R., & Popper, A. N. (2016). *Hearing Aids* (Vol. 56). Springer International Publishing. <https://doi.org/10.1007/978-3-319-33036-5>
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2009). The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *The Journal of the Acoustical Society of America*, 126(6), 3236–3245.
<https://doi.org/10.1121/1.3257225>
- Rhebergen, K. S., Maalderink, T. H., & Dreschler, W. A. (2017). Characterizing Speech Intelligibility in Noise After Wide Dynamic Range Compression. *Ear and Hearing*, 38(2), 194–204. <https://doi.org/10.1097/AUD.0000000000000369>
- Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, 51(1), 160–172.
[https://doi.org/10.1044/1092-4388\(2008/012\)](https://doi.org/10.1044/1092-4388(2008/012))
- Sallat, S., & Jentschke, S. (2015). Music Perception Influences Language Acquisition: Melodic and Rhythmic-Melodic Perception in Children with Specific Language Impairment. *Behavioural Neurology*, 2015(1), 606470. <https://doi.org/10.1155/2015/606470>
- Salorio-Corbetto, M., Baer, T., Stone, M. A., & Moore, B. C. J. (2020). Effect of the number of amplitude-compression channels and compression speed on speech recognition by listeners with mild to moderate sensorineural hearing loss. *The Journal of the Acoustical Society of America*, 147(3), 1344–1358.
<https://doi.org/10.1121/10.0000804>
- Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Lurnagaray, D., Beaulac, S., & Pumford, J. (2005). The Desired Sensation Level multistage input/output algorithm.

- Trends in Amplification*, 9(4), 159–197.
<https://doi.org/10.1177/108471380500900403>
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4), 283–299.
<https://doi.org/10.1177/1084713808325306>
- Siedenburg, K., Röttges, S., Wagener, K. C., & Hohmann, V. (2020). Can You Hear Out the Melody? Testing Musical Scene Perception in Young Normal-Hearing and Older Hearing-Impaired Listeners. *Trends in Hearing*, 24, 2331216520945826.
<https://doi.org/10.1177/2331216520945826>
- Steinberg, J. C., & Gardner, M. B. (1937). The dependence of hearing impairment on sound intensity. *Journal of the Acoustical Society of America*, 9(1), 11–23.
- Stone, M. A., & Moore, B. C. (2008). Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America*, 123(2), 1063–1076.
<https://doi.org/10.1121/1.2821969>
- Uys, M., & Latzel, M. (2015). Long-Term Effects of Non-Linear Frequency Compression on Performance of Music and Speech Perception. *Journal of Communication Disorders, Deaf Studies & Hearing Aids*, 3(3). <https://doi.org/10.4172/2375-4427.1000139>
- Uys, M., Latzel, M., & van Dyk, M. (2016). The Performance of Music and Speech with Different Frequency Lowering Devices. *Journal of Communication Disorders, Deaf Studies & Hearing Aids*, 4(3). <https://doi.org/10.4172/2375-4427.1000163>
- Vaisberg, J. M. (2019). Understanding hearing aid sound quality for music-listening. *Electronic Thesis and Dissertation Repository*(6145). <https://ir.lib.uwo.ca/etd/6145>
- Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2019). A Qualitative Study of the Effects of Hearing Loss and Hearing Aid Use on Music Perception in Performing Musicians. *Journal of the American Academy of Audiology*, 30(10), 856–870.
<https://doi.org/10.3766/jaaa.17019>
- Wagener, K., Kühnel, V., & Kollmeier, B. (1999). Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. *Zeitschrift für Audiologie*, 38, 4–15.
- Wigan, M., Blamey, P. J., & Zakis, J. (2017). Hearing aids and music: experimental tests for better algorithm selection and design. *Researchgate*.
<https://doi.org/10.13140/RG.2.2.11457.15200>

INTEGRATED DISCUSSION

In the following, the key findings from the four empirical studies presented in this dissertation (Chapters 3 to 7) are synthesised. Findings across studies are compared, contrasted and interpreted in relation to each other, and are contextualised through comparison with findings from previous research. Consideration is also given to the limitations and generalisability of the results. Finally, implications for future research and practical applications are outlined.

7.1. Discussion

Traditional research that explores MSA perception often disregards individual variations by focusing on group-level data only, limiting insights into individual perceptual mechanisms. In contrast, this work adopts an individual differences perspective, integrating subjective reports with behavioural measures to explore MSA's multifaceted nature, particularly within HI populations. To that end, MSA is recognized as a complex process that cannot be adequately explained by a single measure or aspect, but emerges from a dynamic interplay of auditory, cognitive, and experiential factors.

7.1.1. *Subjective hearing problems*

Fundamental peripheral auditory processing abilities constitute an essential foundation for successful MSA. These auditory capabilities, however, are vulnerable to damage through noise-induced hearing impairments. The likelihood of such impairments increases with the intensity and duration of noise exposure, placing musicians at particular risk (Musgrave, 2022; Schink et al., 2014). Addressing this concern, the first study (Chapter 3) provided a comprehensive evaluation of hearing health among orchestral musicians, comprising a balanced subsample of 200 professional and 200 amateur musicians, matched by age, gender, and instrument family. This study systematically examined hearing difficulties encounters by musicians, along exploring the adoption rates of hearing protection, attitudes toward hearing aids, general awareness of hearing health, and the impact these factors on health-related quality of life.

Key findings demonstrated that professional musicians exhibited significantly greater awareness of hearing health issues and were more proactive in using hearing protection devices (reporting protection use for approximately 9% of exposure time) compared to their amateur counterparts (approximately 1% protection use). However, professionals notably displayed greater reluctance toward hearing aid adoption. Furthermore, increased reporting of music-related hearing difficulties, in conjunction with hearing challenges experienced in daily life, correlated negatively with mental health-related quality of life. The survey also quantified cumulative lifetime exposure to musical noise by estimating sound exposure levels across various rehearsal and performance settings. Although no direct dose-response relationship between noise exposure and hearing impairments was evident, findings

underscored substantial exposure among musicians: remarkably, about 50% of professional musicians, compared to approximately 10% of amateurs, encountered music noise exposure levels equivalent to 90 dB(A) for 8 hours daily, 5 days per week, 52 weeks per year, over ten years. These findings highlight the critical need for targeted interventions addressing musicians' unique hearing health challenges and combating stigmatization around hearing aid use.

A focal point of this examination, however, was the prevalence of hearing problems among musicians: two-thirds of musicians reported hearing-related issues, dominant in both professional and amateur musicians affecting music-related activities as well as social interactions. Specifically, about one out of five specified having at least a mild hearing loss, another one out of five reported tinnitus. Reported issues included also difficulty hearing other musicians' performances (one-fourth), challenges in hearing one's own instrument (one-fifth), pitch discrimination difficulties (one-sixth), difficulty recognizing instruments (one-tenth), and experiencing a blurred auditory impression (one-tenth; see also Schurig et al., 2024 for details). Crucially, however, hearing problems could not reliably be related to music noise exposure, ageing, and differences between professionals and amateurs were only minor. Several factors could explain the lack of direct associations observed. First, there might be a compensatory effect that ameliorates the effect of peripheral hearing loss for musicians, as observed across many speech and music-related domains (Alain et al., 2013; Dubinsky et al., 2019; Lo et al., 2025; Parbery-Clark et al., 2012; Parbery-Clark et al., 2011; Reuter-Lorenz & Park, 2014; Zendel & Alain, 2012). Second, methodological limitations of the survey-based approach complicate establishing the link: the survey exclusively reached musicians whose hearing problems were insufficiently severe to prevent continuing on a professional performance level, thereby omitting those most severely affected by hearing problems. The survey is thus afflicted with the 'survivorship bias'. Alternatively, the survey-based methods may inadequately capture subtle auditory deficits related to aging or hearing loss, necessitating more sensitive behavioural or physiological assessments to establish these relationships conclusively.

In summary, while this survey-based approach did not identify a direct relationship between reported music-related hearing difficulties and factors like noise exposure, age, or professional status, it provided critical subjective evidence that musicians - who explicitly train and rely on their auditory capabilities - are susceptible to auditory difficulties directly relevant to MSA. Imminent research should employ comprehensive, sensitive (behavioural) measures to further elucidate these relationships; limitations sought to be addressed by the second study of this dissertation.

7.1.2. Development of a behavioural measure of MSA

Historically, quantifying MSA abilities has proven challenging. Many earlier approaches were atomistic in nature, isolating individual acoustic features from auditory scenes to examine the isolated factor with great precision, yet at the cost of reduced ecological validity. Other approaches often excluded participants at the extremes of ability, such as highly trained musicians or older individuals with hearing impairments, due to uninterpretable performance outcomes (Bregman & Campbell, 1971; Bey & McAdams, 2002; Kirchberger & Russo, 2015; Siedenburg et al., 2020, 2021). In response to these issues, the second study (Chapter 4) aimed to develop an adaptive, ecologically valid tool - the Musical Scene Analysis (MSA) test - to precisely measure listeners' ASA abilities within realistic musical scenarios. The test was developed, calibrated, and validated through two large-scale online experiments involving 525 normal-hearing (NH) and 131 hearing-impaired (HI) participants with diverse hearing profiles, ages, and musical backgrounds.

During the calibration phase, all manipulated stimulus characteristics, including the target-to-mix level ratio, the choice of target instrument, and the number of instruments in the mixture, significantly influenced MSA performance, with each factor contributing in decreasing order of effect size. These findings underscore that the acoustic properties of ASA tasks can substantially shape an individual's ability to analyse and understand complex musical scenes. This pattern is consistent with prior research demonstrating that timbre processing, energetic and informational masking, and target-to-masker level differences critically modulate ASA performance (e.g., Ihlefeld & Shinn-Cunningham, 2008; Rajasingam et al., 2021).

Notably, we observed substantial individual differences in MSA performance across participants, suggesting that listeners engage differently with MSA-related stimulus characteristics. Both NH and HI listeners demonstrating wide-ranging performance from chance-level to near-perfect accuracy. Pivotal for this was the relationship between MSA performance and the degree of hearing loss, as measured by self-report: mean detection accuracy decreased from 80% among NH participants to 69% among those with self-reported profound HI ($n = 5$, with two listeners exhibiting chance performance). Contrary to initial expectations, only a marginal correlation was found between individual MSA scores and the musical training subscale of the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014; $r = .10$), with a slightly stronger correlation for the musical perception subscale ($r = .19$). To further examine these effects, a second experiment was conducted using the same item set but introducing sound localisation cues (via interaural level differences) in a sample of 40 NH and 40 HI participants. This experiment replicated the effects related to the choice of target instrument and the number of instruments in the mixture. However, the addition of localisation cues affected MSA processing only minimally. This was somewhat surprising given that prior knowledge about a target's location has been previously shown to be advantageous for stream segregation (Kidd et al., 2005). Irrespective to that, not all previous findings could be replicated. While a marginal correlation between musical perception and MSA performance was again observed ($r = .22$), the link to musical training disappeared entirely ($r = .00$). In contrast, the impact of hearing loss on MSA performance was robust, with

performance decreasing from 85% in NH individuals to 54% in those with profound HI (chance performance) - further highlighting the central role of peripheral hearing in MSA.

In the same study, an adaptive version of the MSA test was validated under controlled laboratory conditions. With as few as 30 items, this version demonstrated acceptable test-retest reliability, both tested under controlled condition in the acoustic lab, but also in online experimental context, corroborating the reliability of findings from the online-calibration phase. Additionally, the MSA test correlated moderately with measures of frequency discrimination, melody perception, beat perception, mistuning perception, and timbre processing. This pattern aligns with the concept of the *positive manifold*, wherein cognitive and perceptual abilities consistently intercorrelate across various domains due to shared underlying cognitive processes (Woodcock, 1990). Indeed, such manifold effects have also been observed in the auditory domain beforehand (e.g., Pausch et al., 2022). The current findings extend this line of research by showing that MSA itself is part of this manifold. This observation is perhaps unsurprising, given that ASA fundamentally underpins a wide range of musical attributes - including the perception of melody, harmony, timbre, and rhythm (Russo, 2019) - and thus constitutes a core component of musical perception more broadly. Considering the robust interconnections among these perceptual abilities, the longstanding neglect of MSA in music cognition research is particularly striking. The development of the MSA test presented here represents a critical step forward, providing a foundational tool for systematically exploring ASA within musical contexts.

Still, despite stratifying participants by age and hearing status, considerable interindividual variability persisted; variability consistent with both subjective reports (Chapter 3) and behavioural findings from calibration experiments. Moreover, such interindividual variability is well-documented in the broader literature, with studies reporting both subjective descriptions (e.g., "fuzzy" or "washed-out" percept; see Greasley et al., 2020) and behavioural evidence of large performance differences (e.g., Bürgel et al., 2021, 2023; Kirchberger & Russo, 2015; Siedenburg et al., 2020, 2021). To better account for these differences, several explanatory variables were considered in the third study, essentially defining the listeners personal profile to explain individual differences in MSA performance: musical background, WMC, chronological age, and degree of hearing impairment (as determined behaviourally by audiometric thresholds).

In brief, the MSA test showed moderate correlations with all examined predictors: working memory capacity ($r = .36$), musical training ($r = .46$), age ($r = -.40$), and hearing sensitivity ($r = -.37$). Notably, these listener profile variables also exhibited moderate intercorrelations; a pattern that echoes a central theme in the literature: individual differences emerge from a complex interplay of auditory peripheral, cognitive, and experience related factors (e.g., Loughrey et al., 2018). While these person-level characteristics are not themselves part of the positive manifold of auditory-perceptual abilities, they may be regarded as sources of variance that contribute to or shape the observed manifolds underlying processing architecture. To that end, they may serve as underlying determinants influencing performance across multiple auditory domains, potentially explaining why certain individuals perform consistently better across related perceptual tasks. Several hypotheses have been proposed regarding the underlying mechanisms linking these factors. One

prominent account suggests a shared neurobiological vulnerability, whereby age-related degeneration in auditory structures parallels declines in other physiological systems (Thomson et al., 2017); another suggests that the association between hearing loss and cognitive decline is mediated through a decrease in social contact (i.e., social isolation). Concurrently, ageing is linked to cognitive declines in working memory, executive function, and processing speed, which are critical for stream segregation (Dryden et al., 2017; Fisher et al., 2019). These domains are further affected by hearing loss, which increases cognitive load and is associated with accelerated cognitive decline and dementia risk (Bharadwaj et al., 2015; Gallacher et al., 2012; Lin et al., 2011; Pichora-Fuller et al., 2015; Shinn-Cunningham, 2017). Musical training, on the other hand, further complicates the full picture, acting as a potential 'buffer' against these effects by enhancing sensitivity to acoustic cues and supporting relevant cognitive faculties like attention and memory (Clayton et al., 2016; Liu et al., 2025), with benefits persisting into older age (Zendel & Alain, 2009, 2012; Susini et al., 2023; Whittom et al., 2024). These effects, however, appear to be rather task-dependent and do not consistently extend to speech processing (Maillard et al., 2023; McKay, 2021).

Nevertheless, despite differences in methodology, study design, and protocols, many studies investigating these factors continue to confound key influences, for instance, by failing to include younger HI or older NH participants, or by not adequately accounting for musical background and general cognitive abilities (e.g., Humes et al., 2012; Shinn-Cunningham & Best, 2008). The current study improved upon these limitations by including participants with a wider range of hearing sensitivity, age, and musical expertise. However, notable gaps remain: older NH and younger HI individuals were still underrepresented, which restricts the ability to clearly separate the effects of age and hearing loss on MSA performance.

7.1.3. *Hearing loss, working memory, ageing, and musical training in MSA*

The third study (Chapter 5) was aimed to overcome previous methodological limitations by increasing the representation of participant groups that are often underrepresented in MSA-related research, thereby enabling a more comprehensive and robust analysis of the factors contributing to MSA performance. To this end, a dual analytical strategy was employed, combining Ridge-penalized linear regression, which assumes linear relationships, with a gradient-boosted decision tree model (XGBoost; Chen & Guestrin, 2016), which accommodates non-linear relationships and complex interactions. This approach allowed for the robust estimation of each predictor's contribution while accounting for multicollinearity within the final sample, comprised of 31 older NH, 34 older HI, 26 younger NH, and one younger adult with HI. The investigation focused on two central objectives: (1) examining potentially shared underlying processing mechanisms between speech-based and music-based ASA, and (2) exploring how key individual difference factors - such as hearing loss, WMC, musical training, and age - differentially affect ASA performance across these two distinct auditory domains.

In line with the hypothesis, the third study demonstrated a moderate correlation ($r = .49$) between speech-based and music-based ASA performance, supporting the notion of shared underlying processing mechanisms between music and speech. This finding suggests the presence of domain-general ASA mechanisms and aligns with previous research indicating

common perceptual and cognitive foundations: the abilities to perceptually separate pitch patterns and the ability to separate sentences spoken simultaneously by different talkers, for instance, has been shown to rely on the same underlying perceptual and cognitive processes (Mackersie et al., 2001). Additionally, converging evidence from neuroimaging studies suggests considerable overlap in the neural networks underlying music and speech perception, including shared processing pathways in the auditory cortex and associated regions (Abrams et al., 2011; Angulo-Perkins et al., 2014; Leaver & Rauschecker, 2010; Sallat & Jentschke, 2015; Schön et al., 2010). These shared mechanisms provide an empirical foundation for cross-domain transfer effects - for example, the potential for musical training to improve speech perception abilities (e.g., Maillard et al., 2023). At the same time, the non-trivial differences observed between domains point to the continued relevance of domain-specific influences.

Despite the presence of shared processing mechanisms, notable differences emerged in how individual difference factors influenced speech- and music-based auditory scenes. Most prominently, MSA performance appeared more robust to hearing loss compared to speech-based ASA, with hearing loss exerting a substantially greater impact on speech perception. This finding is consistent with results from the calibration phase in Chapter 4, where hearing loss negatively affected ASA performance in both domains, but explained considerably more variance in the speech-based context. These results underscore the importance of intact auditory pathways for effective ASA across both music and speech (Akeroyd et al., 2007; Middlebrooks, 2017; Shinn-Cunningham & Best, 2008). One plausible explanation for differences between domains might be attributed to task-specific listening strategies: in music-based tasks, listeners were likely able to exploit momentary reductions in background energy, so-called "dips" (also referred to as glimpses), to identify target sounds, whereas such a strategy is less viable in steady-state maskers like the steady speech-shaped noise used in the Oldenburg Sentence Test. Future research employing more ecologically valid speech scenarios, such as multi-talker babble, may yield more comparable performance across speech and music-based ASA, as these contexts rely more heavily on similar top-down processing mechanisms.

A central objective of the study was to disentangle the effects of age and hearing loss. It was expected that, after controlling for hearing loss, musical training, and WMC, the direct contribution of chronological age to ASA performance would be substantially reduced. This assumption was based on the recognition that age often serves as a proxy for a wide range of underlying, yet oftentimes unknown, auditory, cognitive, and physiological changes. The results supported this assumption: while an initial decline in ASA performance was observed with increasing age, most of this effect disappeared when accounting for other factors and especially hearing loss, suggesting that age per se may be a less central explanatory factor than frequently assumed. These findings highlight the necessity of methodological rigour in studies attributing performance declines to ageing without adequately controlling for hearing loss and other interrelated factors. Nonetheless, the cumulative effects of ageing over the lifespan remain practically significant, as evidenced by a residual negative impact on ASA performance in both speech and music domains, even after controlling for other variables.

Despite the targeted efforts to balance the sample by age and hearing status, substantial interindividual variability remained within these groups. Numerous studies have pointed to musical experience as key factors underlying this variance (e.g., Zendel & Alain, 2009, 2012; Whittom et al., 2024; Zhao, 2024). Here musical training was operationalized not by comparing musicians versus non-musicians, but by considering musical training as a continuum. Indeed, while contrasting with the results obtained by the calibration experiments of the second study (Chapter 4), the extended and (though not fully) balanced sample used in the third study (Chapter 5) provided robust evidence for a solid relationship between MSA performance and musical training. Specifically, while the validation experiment in Chapter 4 indicated moderate correlations between MSA performance and the musical training subscale of the GMSI ($r = .46$), musical perceptual abilities ($r = .37$), and the general factor of musical sophistication (reflecting a broad spectrum of musical expertise, behaviour, and achievements; $r = .46$), the expanded and more diverse sample of Chapter 5 advocated a similarly strong correlation ($r = .44$). Importantly, the statistical models employed further clarified that this association persisted even taking hearing loss and age into account. Collectively, these results align closely with prior literature documenting enhanced auditory perceptual abilities resulting from musical expertise (Bey & McAdams, 2002; Kothinti & Elhilali, 2023; Woods & McDermott, 2018), the established perceptual advantages of musicians over non-musicians (Brown et al., 2017; Correia et al., 2023; Marques et al., 2007; Micheyl et al., 2006; Mishra et al., 2015), studies explicitly linking musical training to improved auditory perception in musical contexts generally (Dubinsky et al., 2019; Kannyo & DeLong, 2011), and to MSA-related tasks specifically (Zendel & Alain, 2009; Susini et al., 2023; Slater & Marozeau, 2016; Liu et al., 2025).

Notably, the effect sizes observed here are particularly striking because they emerged within a sample largely comprising amateur musicians rather than a group explicitly comparing professional musicians with complete non-musicians. This implies that extensive, professional-level musical practice may not be a necessary prerequisite for observing perceptual benefits associated with musical engagement. It is worth noting that the present findings are consistent with previous research indicating that formal musical training is not strictly required for individuals to reach musician-like performance on tasks involving musical and cognitive skills (Correia et al., 2022). A potential influential factor here may be the frequency of recent musical practice. For example, Micheyl et al. (2006) showed that non-musically trained individuals could achieve performance comparable to classically trained musicians after just 4 to 8 hours of targeted auditory training. Similarly, Laffere et al. (2020) demonstrated that only two hours of explicit auditory training led to improvements in auditory selective attention, enhancing subsequent performance on a working memory task within a distracting auditory environment.

The difference in comparison to the calibration experiments, as well as in the related null effects reported by Bürgel et al. (2021, 2023), is nonetheless unexpected. Potentially, the weaker correlations in the online experiments result from a combination of greater measurement noise, reduced experimental control, and confounding of musical training with age and hearing status, which dampen the likelihood to identify true relationships. However, while differences in measurement quality and environmental control between lab and online

settings may have played some role, the primary explanation may lie in the more balanced sampling. Specifically, the inclusion of participants with broader ranges of individual differences factors, particularly those with varied levels of musical training, may have facilitated a clearer and more accurate representation of the relationship between musical training and MSA abilities. Other null effects, such as those reported by Başkent et al. (2018) among young adolescents, may be explained by developmental factors; adolescents may still be in a critical developmental stage for auditory streaming (Calcutt, 2024; Leibold, 2011) and may not yet have reached a threshold of musical training sufficient for observable auditory benefits. A dedicated large-scale study aimed at testing the robustness and generalisability of this effect may prove valuable.

Interestingly, while the formerly observed relationship between speech and music-based ASA ($r = .49$) indicated domain-general mechanism for ASA processing, the positive effects of musical training were confined to MSA performance only and did not extend to the speech-based ASA task, where no improvement was observed; thus, emphasizing the domain-specific nature of these auditory processing advantages. Generally, this specificity would reinforce ongoing debates regarding the limitations of far-transfer effects, such as the lack of improvements in speech perception resulting from musical training (Coffey et al., 2017; Madsen et al., 2019; McKay, 2021; Schellenberg & Lima, 2024). However, the mixed results might indicate that transfer effect of musical training with ASA tasks might depend on the specific scene itself; giving some accounts on why the benefits of musical training have not been consistently observed across diverse auditory contexts (Crawley et al., 2002; Carey et al., 2015). The present results suggest that auditory processing advantages, particularly those related to MSA, may emerge without extensive or professional-level musical sophistication. These findings challenge traditional binary distinctions between musicians and non-musicians and highlight the value of conceptualizing musical training as a continuous, dose-sensitive variable. They further emphasize the plasticity of auditory and attentional systems and may further suggest that the recency and ongoing frequency of musical engagement may play roles as influential - potentially even more so - than cumulative hours of practice alone. Future research should thus move beyond simplistic dichotomous groupings (i.e., musicians versus non-musicians) and explicitly incorporate these nuanced factors into research designs.

Regarding the relationship between WMC and ASA abilities (both speech-based and music-based), initial expectations anticipated robust links. While an overall relationship between WMC and ASA performance emerged initially across both domains, this association was not significant within individual participant groups once accounting for other predictor variables. After accounting for age, hearing loss, and musical training, the employed statistical models confirmed these findings: WMC exerted no significant effect on speech-based ASA and only a small effect on MSA performance. Thus, the weak or inconsistent associations between WMC and ASA observed here and in previous studies (Füllgrabe & Rosen, 2016; Golden et al., 2017) suggest that apparent WMC effects may predominantly reflect confounding influences that come along with ageing and peripheral HI rather than direct causal relationships. Nevertheless, more demanding ASA tasks with higher cognitive load or energetic and informational complexity might place greater demands on WM systems, an important

consideration for future research given that the tasks employed in this study required minimal memory demands.

Taken together, these findings align with the broader conceptual framework proposed earlier, which characterizes age, hearing loss, cognitive functioning, and musical training as interrelated rather than isolated determinants of MSA performance. The observed attenuation of individual factor effects, particularly for age and WMC, after controlling for other predictors reinforces the perspective that the explored factor variables often operate in concert rather than in isolation. To that end, apparent age-related declines may largely reflect deteriorating peripheral auditory input rather than intrinsic cognitive degeneration. Similarly, the limited explanatory power of WMC might be attributed to its strong entanglement with age and peripheral hearing, both of which increase cognitive load in ASA tasks. In contrast, musical training exhibited a relatively robust and independent association with MSA performance, highlighting its potential as a protective factor that supports auditory skills and buffers against sensory and cognitive declines. Ultimately, these results underscore the necessity of employing multifactorial approaches that account for shared variance and interactive effects when investigating the factors underlying ASA performance across speech and music contexts. Importantly, however, the absence of younger HI participants remains a significant limitation. Future research incorporating this critical group could enable a more nuanced understanding of how HI affects MSA across different stages of development.

7.1.4. *MSA abilities among hearing aid users*

Building on the prior findings from Chapters 3 to 5 - where large interindividual differences in MSA performance were observed among both NH and HI individuals - the final study presented in Chapter 6 aimed to investigate how MSA processing behaves under aided conditions. Specifically, the goal was to examine how hearing aid (HA) signal processing, particularly dynamic range compression (DRC), affects both behavioural and subjective aspects of MSA in a clinical population. In doing so, this study represents a first applied use case of the adaptive MSA test. In addition, musical sound quality and speech-in-noise measures were employed.

Consistent with earlier findings (Chapter 5), a moderate correlation ($r = .55$) between music- and speech-based ASA was observed in unaided listening conditions. Notably, a different speech-in-noise test was used in this experiment, the Göttingen Sentence Test, corroborating the presence of shared auditory processing mechanisms (e.g., Shinn-Cunningham & Best, 2008) as observed in the previous experiment. Interestingly, this relationship was no longer present under aided conditions. Without a more sophisticated, dedicated investigation of this phenomenon, one can only speculate about the underlying reasons. One possibility is that DRC settings in HAs differentially affect the acoustic structure of speech and music, thereby disrupting shared perceptual mechanisms or unevenly altering task demands across domains. Moreover, aided listening introduces additional listener-specific variables - such as prior experience with hearing aids, personal preferences for compression speed, and degrees of acclimatization - that may further modulate perception in idiosyncratic ways. This added layer of inter-individual variability could further weaken the

consistency of domain-general patterns, effectively eroding the correlation between musical and speech-related ASA performance in the aided condition.

HA amplification also showed other interesting outcomes. Indeed, for many listeners, the HA amplification could restore MSA ability comparable to the median performance exhibited by participants from the calibration population of study two (Chapter 4), which predominantly included NH listeners. Moreover, the results showed significant improvements across all three listening tasks (MSA, musical sound quality ratings, speech-in-noise) when participants used hearing aids compared to unaided listening. However, the effects of DRC speed varied across tasks, revealing a nuanced interaction between compression settings and aided outcomes. Specifically, compression differentially affected the behavioural MSA performance and subjective musical sound quality: fast DRC yielded the highest performance for MSA, whereas slow DRC received the best ratings for musical sound quality. Interestingly, this pattern also extended to the question “*How clearly were you able to distinguish between the individual instruments?*” - a question that essentially probes MSA ability subjectively. This apparent contradiction between subjective impression and behavioural outcome suggests that listeners favour the more natural sound of slow DRC, even when it does not support optimal auditory stream segregation. These findings underscore a critical dissociation between behavioural performance and subjective experience in aided listening and more generally highlight the importance of including both objective and subjective assessments when evaluating hearing aid performance, particularly in complex listening domains like music. This might have implications for clinical practice too. The results indicate that optimal HA settings for musical perception may not align with those that yield the best measurable outcomes, which raises questions about which aspects of auditory experience users prioritize in daily life - accuracy or comfort - and suggests that listener preferences may guide fitting decisions as much as, or even more than, raw perceptual ability.

Interestingly, these findings resonate with those from Chapter 3, which examined subjective hearing health and attitudes toward hearing aids among orchestral musicians. While professional musicians demonstrated significantly greater awareness of hearing risks and were more proactive in using hearing protection compared to their amateur counterparts, they also expressed notably stronger disapproval toward hearing aids. Adoption rates were low overall, with only 9% of professionals and 6.5% of amateurs reporting current or past hearing aid use. Among those using hearing aids, only about 60% found hearing aids helpful for music listening, with even fewer deeming the quality sufficient for music-making. A third of the hearing aid users never wore their devices while playing music. This discrepancy points to a critical perceptual barrier: while some respondents attributed their reluctance to concerns about the image projected to colleagues or audiences - suggesting internalised stigma - many cited anticipated auditory distortions as their primary concern, highlighting a deep-rooted scepticism about the suitability of hearing aids for music.

The practical implications emerging from Chapter 6 offer a promising counter-narrative that could potentially address musicians’ scepticism and encourage greater hearing aid adoption among this population. Specifically, results showed not only measurable improvements in MSA abilities under aided conditions but also higher subjective ratings for musical sound quality when employing slower DRC settings - settings that largely preserve

natural dynamic structures and temporal envelopes. While this suggests some form of performance vs. quality duality, clinical recommendations emerging from these combined insights remain valuable: audiologists and clinicians should explicitly communicate to musicians the nuanced performance and quality benefits of different DRC settings, underscoring that hearing aids, when optimally configured, can offer demonstrable benefits in complex musical contexts; importantly, while also highlighting the caveats of the respective settings. Targeted informational and experiential interventions that directly address musicians' specific auditory quality concerns and stigma could facilitate greater acceptance and more effective use of hearing aids among professional musicians. It is important to note, however, that the study in Chapter 6 did not specifically target professional musicians; thus, the demands of music listening and music-making may differ between the general HA-using population studied and musicians. This limits the direct applicability of these results to professional music performance contexts and calls for future work to validate these findings in musician-specific samples.

Clinical considerations, along with their limitations, may also extend to the domain-specific effects observed in the behavioural outcomes presented in Chapter 6. While fast DRC benefitted MSA performance among most listeners, the opposite pattern was found for speech-in-noise, where slow DRC proved more effective. Apparently, unlike to music, preserving natural envelope cues was more critical for speech parsing. These divergent patterns imply a trade-off in HA processing: compression strategies that improve perception in one domain (e.g., in speech dominated scenes), do not necessarily improve perception or may even hinder another (e.g., music dominated scenes). Consequently, one-size-fits-all HA settings may not serve individuals with diverse listening needs. These findings strengthen the case for context-aware or user-adaptive HA algorithms that adjust parameters dynamically depending on listening goals (e.g., speech vs. music).

Once more consistent with earlier findings (Chapters 3–5) and the literature, substantial variability in performance was observed - so pronounced that some moderately to severely HI individuals performed on par with NH participants from previous experiments. Given the relative prominent sample homogeneity in terms of age and audiometric profile, this finding is particularly notable, reinforcing the notion that audiometric thresholds alone are insufficient to predict MSA ability in complex auditory environments (e.g., Siedenburg et al., 2020). Considerable individual variability also emerged regarding the benefits (and preference) associated with different DRC settings across auditory tasks. While some participants demonstrated consistent preferences or performance advantages across tasks, many exhibited context-specific benefits; that is, performing best with fast DRC in the MSA task, preferring unaided listening for musical sound quality, and achieving the highest speech-in-noise scores with slow DRC. These individual patterns were not systematically related to musical training nor chronological age. Instead, they were strongly associated with hearing aid usage duration and individual hearing thresholds, in line with prior findings that highlight the influence of auditory experience and adaptation on amplification outcomes (see also Lunner et al., 1997; Ricketts et al., 2008; Moore et al., 2011).

Altogether, our results support a multifaceted approach to hearing aid optimization, challenging the feasibility of a "one-size-fits-all" solution. They demonstrate that ideal HA

configurations vary not only across auditory environments, but vary among individuals and within individuals across different auditory environments. These eminent individual differences reinforce the necessity for personalized fitting protocols that consider both the acoustic characteristics of the listening scenario and the auditory profiles of users (Scollie et al., 2005; Moore et al., 2011). Finally, the successful employment of the adaptive MSA test in a clinical HA context substantiates its value for future clinical and research applications, providing an ecologically valid measure, sensitive to auditory perception complexities inherent to real-world musical scenarios. Given the expected growing incidence of HA users (Popelka et al., 2016) and the documented dissatisfaction with music perception among this population (Greasley et al., 2020; Looi et al., 2019; Vaisberg et al., 2019), such tools are vital for improving auditory rehabilitation outcomes and user satisfaction.

7.2. Implications for future research and possible applications

The development of the adaptive MSA test represents a significant methodological advancement with broad implications for future auditory perception research. Firstly, the establishment of the adaptive MSA test provides researchers and clinicians with an efficient, robust, reliable, and ecologically valid measurement for assessing ASA abilities specifically tailored to musical context. The initial validation and calibration across diverse populations, including NH and HI individuals, confirmed its sensitivity and applicability for assessing individual differences comprehensively. The employment of the MSA test in HA research (Chapter 6) serves as a practical blueprint, demonstrating the tool's potential to systematically evaluate the effectiveness of various HA settings in real-world listening scenarios. Future studies can readily leverage this methodology to optimize and personalize auditory rehabilitation strategies, particularly for HI listeners and HA users.

Moreover, insights from the individual differences research conducted throughout this dissertation highlight several implications. The consistent finding that individual performance variability persists despite controlling for hearing thresholds and age underscores the necessity for adopting a multifactorial approach to auditory perception research. Such an approach is essential to fully capture the complexity inherent in music perception and ASA abilities. Specifically, the nuanced interrelations among hearing impairment, age, WMC, and musical training suggest these factors do not function independently but interact dynamically. Future studies should, therefore, extend beyond examining isolated effects, focusing instead on exploring interactions through advanced statistical modelling techniques, which effectively account for the interconnectedness of peripheral auditory and cognitive variables.

Practically, these insights advocate for personalized auditory training and rehabilitation approaches. Findings indicate that musical training correlates with enhanced ASA abilities, suggesting its potential role as a cognitive intervention or compensatory mechanism against auditory and cognitive decline. Future applications could involve structured musical training programs designed explicitly for older adults or individuals at risk of developing hearing impairment, potentially serving as preventive or rehabilitative tools to mitigate age-related auditory deficits. The adaptive nature of the MSA test itself could serve as

a diagnostic tool, enabling the targeted identification of specific deficits in ASA / MSA and facilitating individualized training regimens tailored to listener-specific needs.

Additionally, the differential impact of HA dynamic range compression settings on music versus speech perception underscores the complexity of auditory amplification. The MSA test, therefore, could become a standard component in clinical audiology practice, supplementing traditional speech-based tests to provide a holistic assessment of ASA skills in diverse acoustic contexts. Despite this, the considerable variability observed across listeners highlights the importance of considering individual listener profiles in future research and applications. By systematically incorporating listener-specific factors such as musical experience, cognitive capacities, and detailed audiological profiles into assessments and interventions, clinicians and researchers could substantially enhance the effectiveness of auditory rehabilitation and fitting strategies. These expansions would significantly enhance the clinical utility of auditory assessments, enabling clinicians to fine-tune hearing aids more precisely according not only to patients' personal preferences, but also to everyday listening environments like music listening. Ultimately, the insights gained from this dissertation offer promising directions for future research and tangible applications, emphasizing the critical role of personalized, multifaceted approaches to auditory health and music perception capabilities.

7.3. Conclusion

This dissertation set out to deepen our understanding of individual differences in MSA across the lifespan and hearing profiles. Through a series of empirical studies, it established a novel, ecologically valid tool to measure MSA, demonstrated its reliability and practical utility, including for clinical populations, and provided compelling evidence for the multifactorial nature of MSA performance. Hearing loss, cognitive function, and musical training emerged as interrelated yet distinct contributors, with musical training showing particular promise as a protective or compensatory factor. By integrating subjective and behavioural approaches, and systematically exploring the roles of hearing, age, and experience, this work not only fills important gaps in music perception research but also lays the groundwork for future clinical applications and personalised auditory assessments.

APPENDICES

8.1. Figures

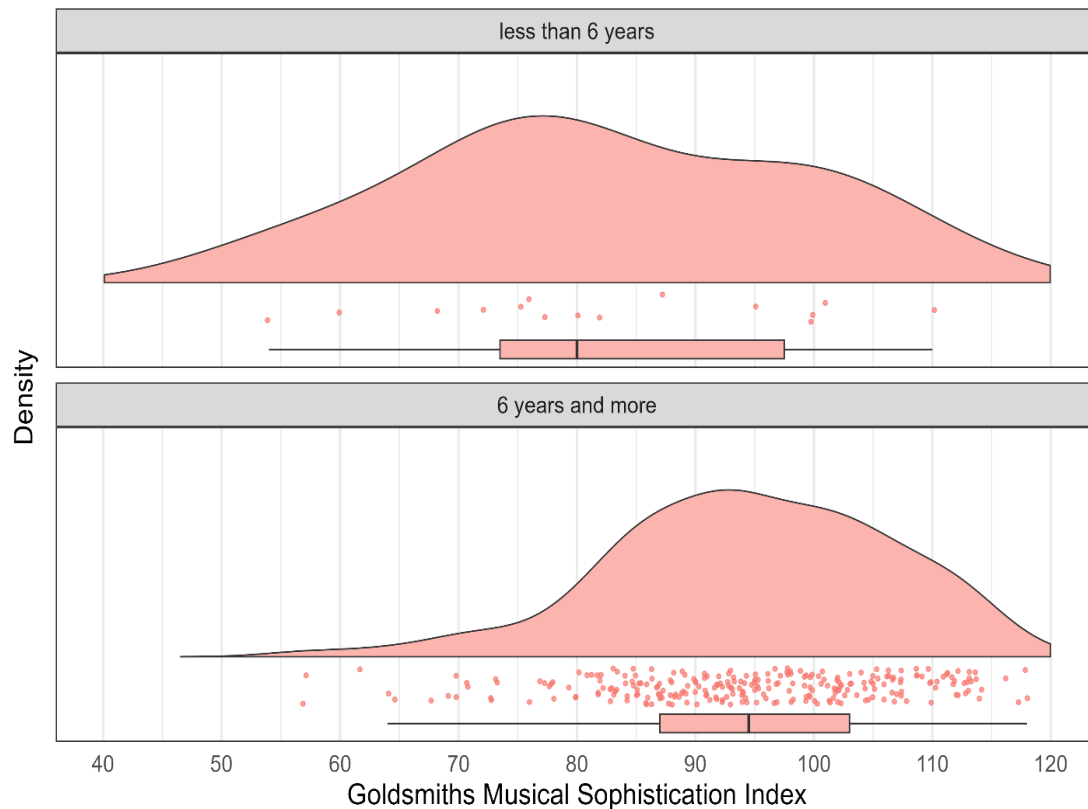


Figure A3.1 Density plots, along with boxplots and individual data points (pale red dots), illustrate the distribution of Goldsmiths Musical Sophistication Index (GMSI; Müllensiefen et al, 2014) scores. These plots compare two groups within the amateur musician sample: participants with <6 years of daily practice (top panel) and the main sample after exclusion of the <6 years practice group (bottom panel). Higher GMSI scores (range: 18-126) indicate higher musical sophistication. The top panel median GMSI ($M = 80$) is comparable to the general population ($M = 77-82$; see Müllensiefen et al, 2014). These participants have been excluded, following the criteria by Zhang & Schubert (2019). The main sample median GMSI (bottom panel) is 95 ($SD = 12.01$). Results shown are before matching.

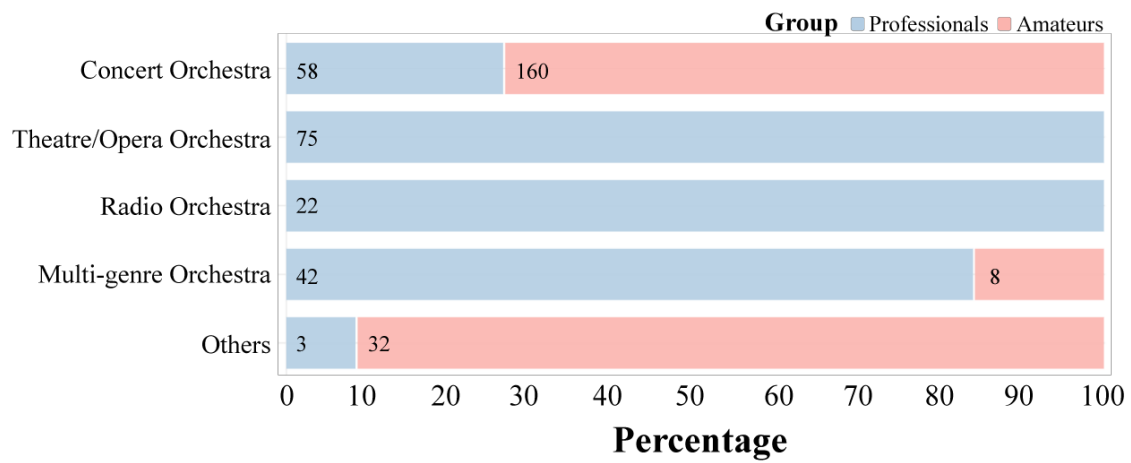


Figure A3.2 Distribution of primarily played orchestral category/repertoire among respondents, shown post-matching procedure. Light blue bars represent professional musicians, and pale bars represent amateurs. Numbers inside each bar indicate the count within each respective category.

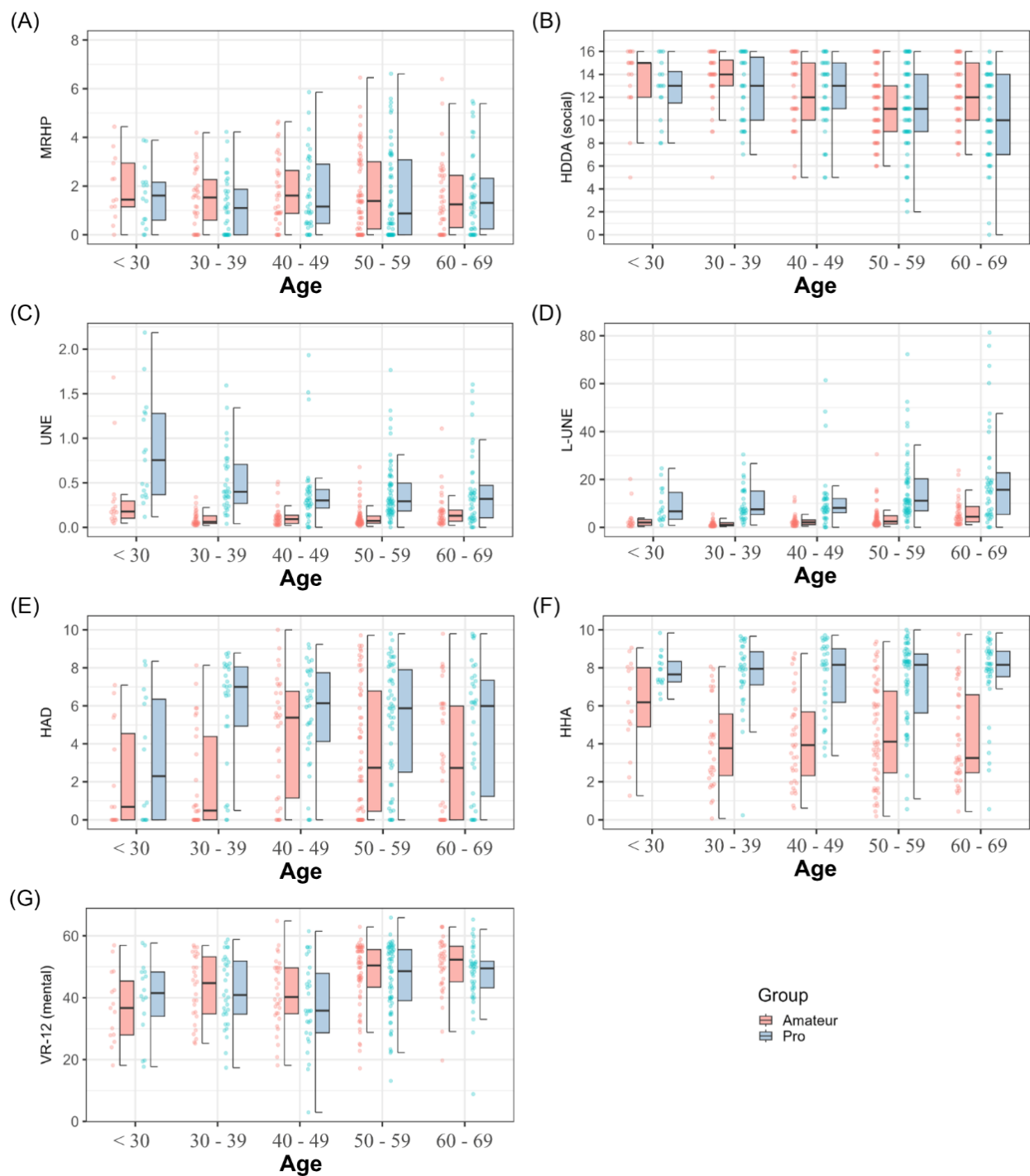


Figure A3.3 This multi-figure depicts the distribution of hearing-related measures and music exposure across age groups in the matched sample of amateur and professional musicians (N = 400). Boxplots show grouped distributions, with dots representing individual responses (amateurs: pale dots, professionals: light blue dots). Panels depict: (A) Music-Related Hearing Problems (MRHP), (B) Hearing-Dependent Daily Activities Subscale for social interactions (HDDA), (C) Unit of Noise Exposure (UNE), (D) Lifetime UNE (L-UNE), (E) Hearing Aid Disapproval (HAD), (F) Hearing Health Awareness (HHA), (G) Veterans Health-related quality of life (VR-12) subscale for mental health.

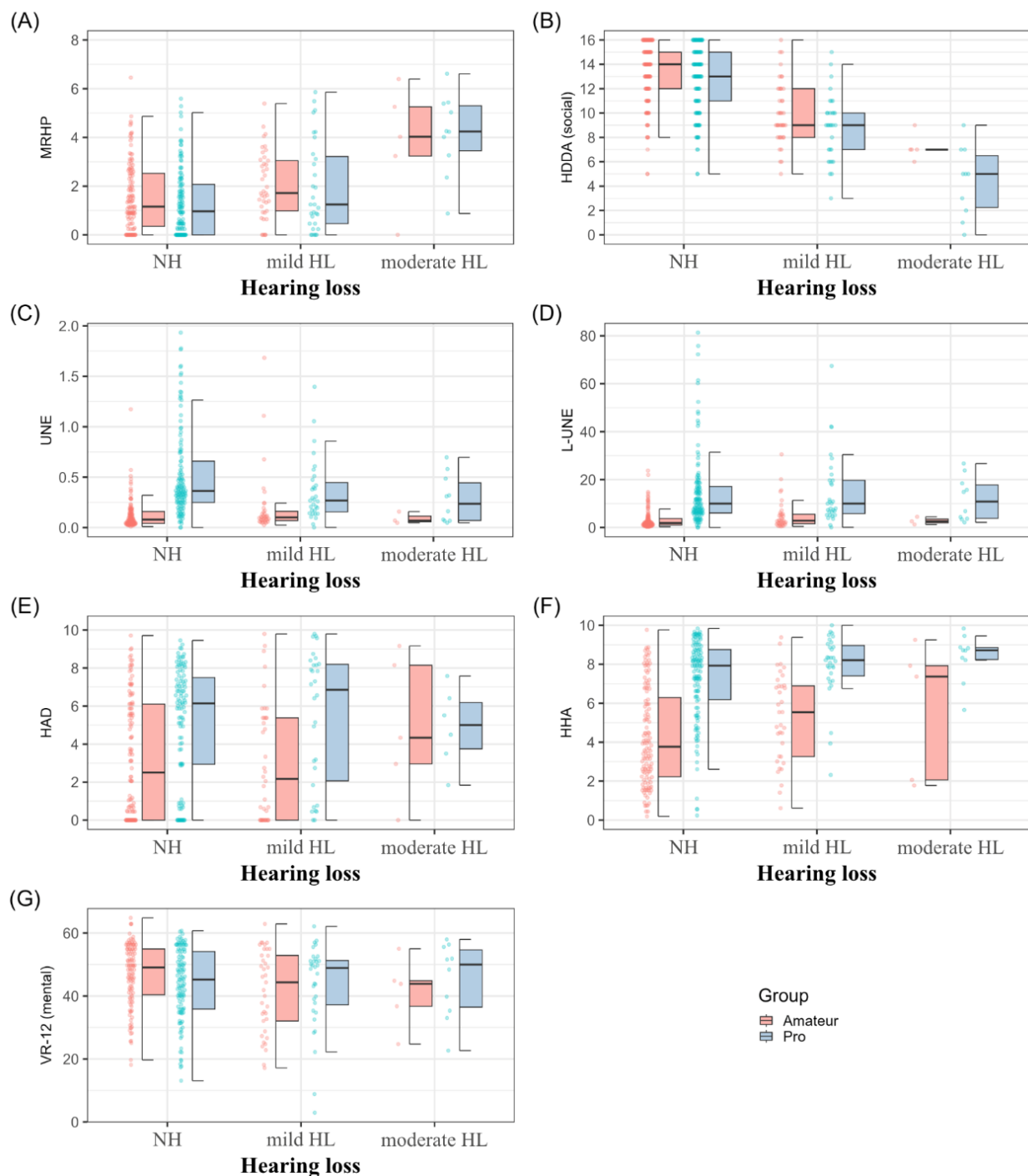


Figure A3.4 This multi-figure depicts the distribution of hearing-related measures and music exposure across hearing loss groups in the matched sample of amateur and professional musicians (N = 400). Boxplots show grouped distributions, with dots representing individual responses (amateurs: pale dots, professionals: light blue dots). Panels depict: (A) Music-Related Hearing Problems (MRHP), (B) Hearing-Dependent Daily Activities Scale (HDDA), (C) Unit of Noise Exposure (UNE), (D) Lifetime UNE (L-UNE), (E) Hearing Aid Disapproval (HAD), (F) Hearing Health Awareness (HHA), (G) Veterans Health-related quality of life (VR-12).

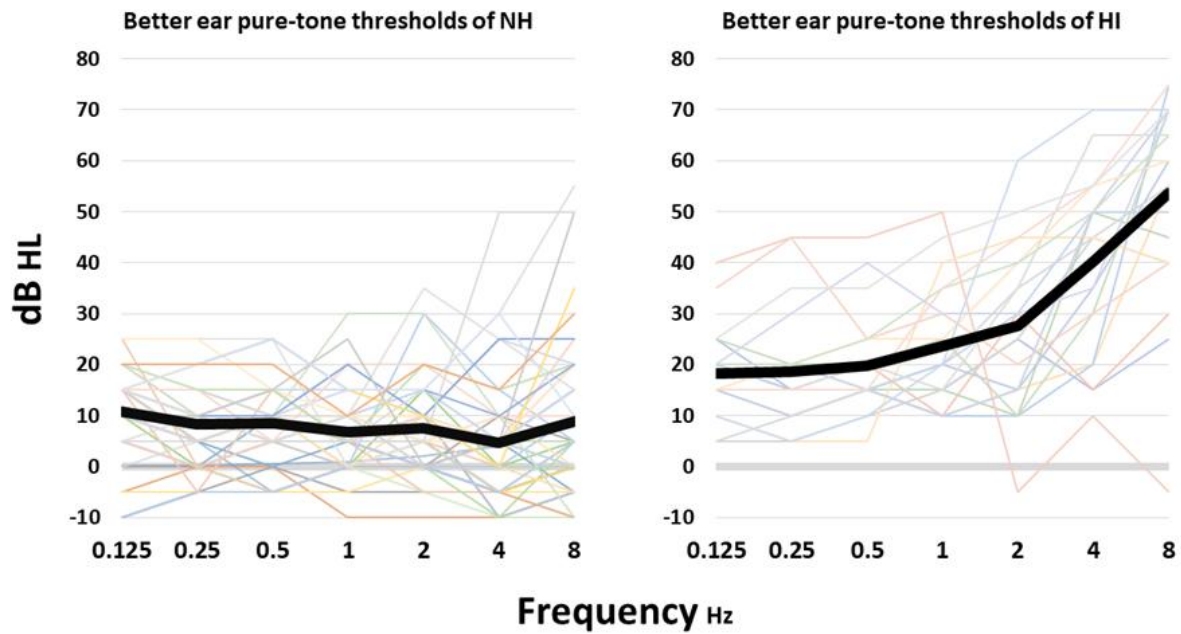


Figure A4.1 Graphical illustration of the pure-tone audiometric threshold (PTA) measurements of all participants for the respective better listening ear. Thresholds for normal-hearing individuals are plotted on the left and for hearing-impaired on the right. Standard clinical ascending-descending procedure in 5 dB HL steps were used. Participants were classified as hearing-impaired, when PTA over all frequency bins for the better ear exceeded 20 dB HL. The greyed lines show individual participants' threshold. The bold solid lines are representing the pure-tone average over all participants for each listening group.

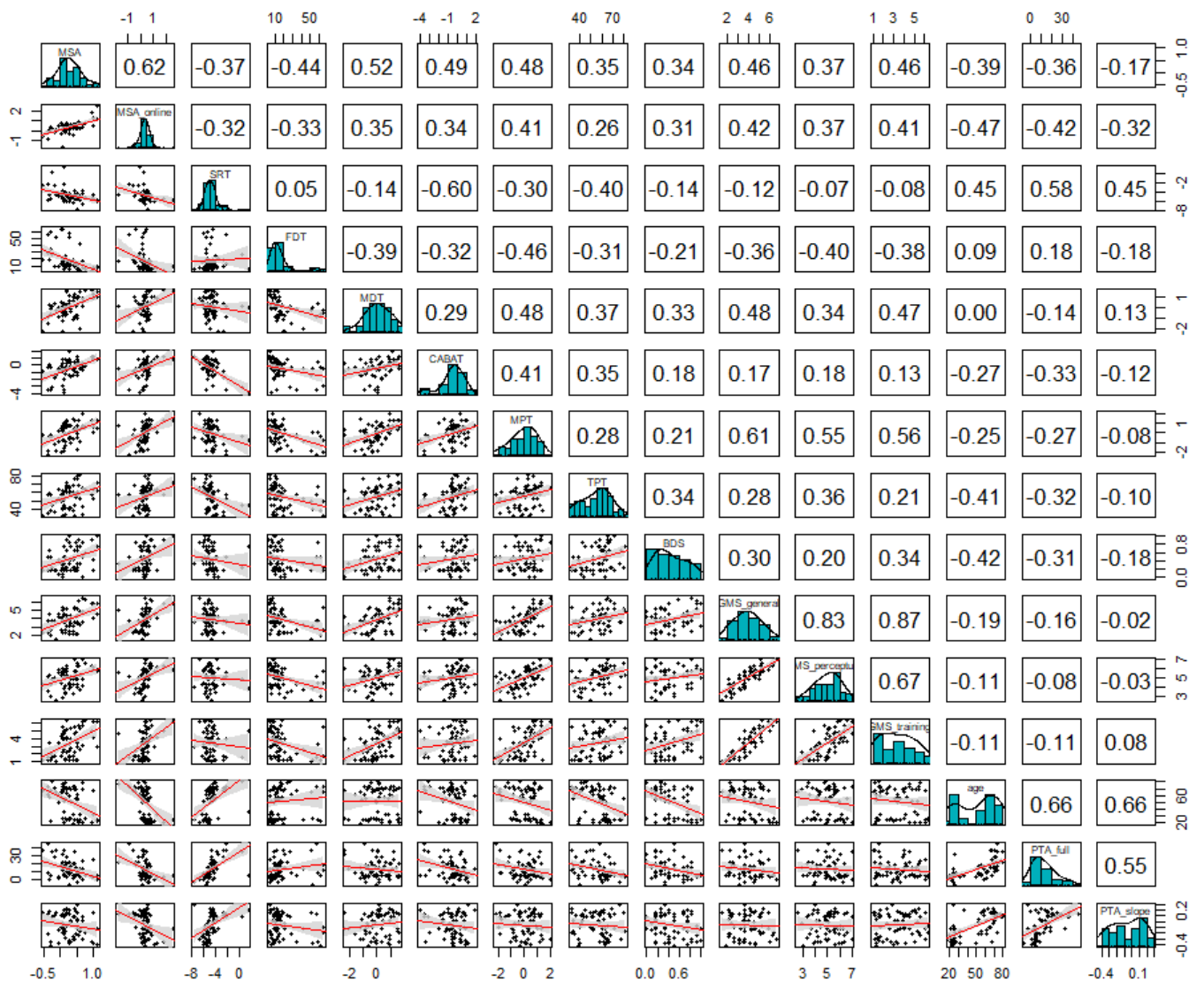


Figure A4.2 Correlation matrix of the adaptive MSA in relation to other psychoacoustic and musical-related tests. MSA is composite score of the first and second item set from the laboratory part of the experiment. MSA3 corresponds to a third test-set that was employed within the online testing environment. Subscales of the Gold-MSI (GMS), age and hearing impairment (PTA) are also included. Speech-reception thresholds (SRT) are the test results of the OLSA speech-in-noise test. For this matrix, extreme outliers and participants with missing data have been excluded (N = 63).

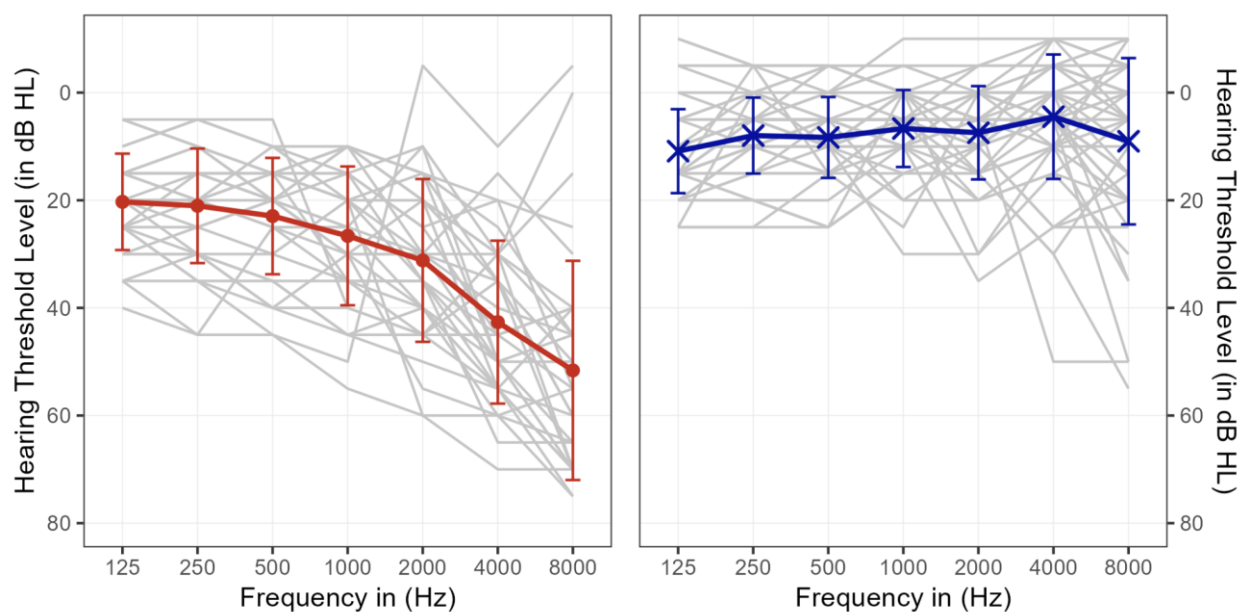


Figure A5.1 Pure-tone audiometric thresholds (PTA) for all participants' better listening ear. Thresholds for individuals with hearing impairment are plotted on the left (red) and for those with normal hearing on the right (blue). A standard clinical ascending-descending procedure in 5 dB HL steps was used. Participants were classified as hearing-impaired if their better ear PTA exceeded 20 dB HL across all frequency bins. The grey lines represent individual participants' thresholds. The bold lines (along SD as error bars) represent the average PTA for each listening group.

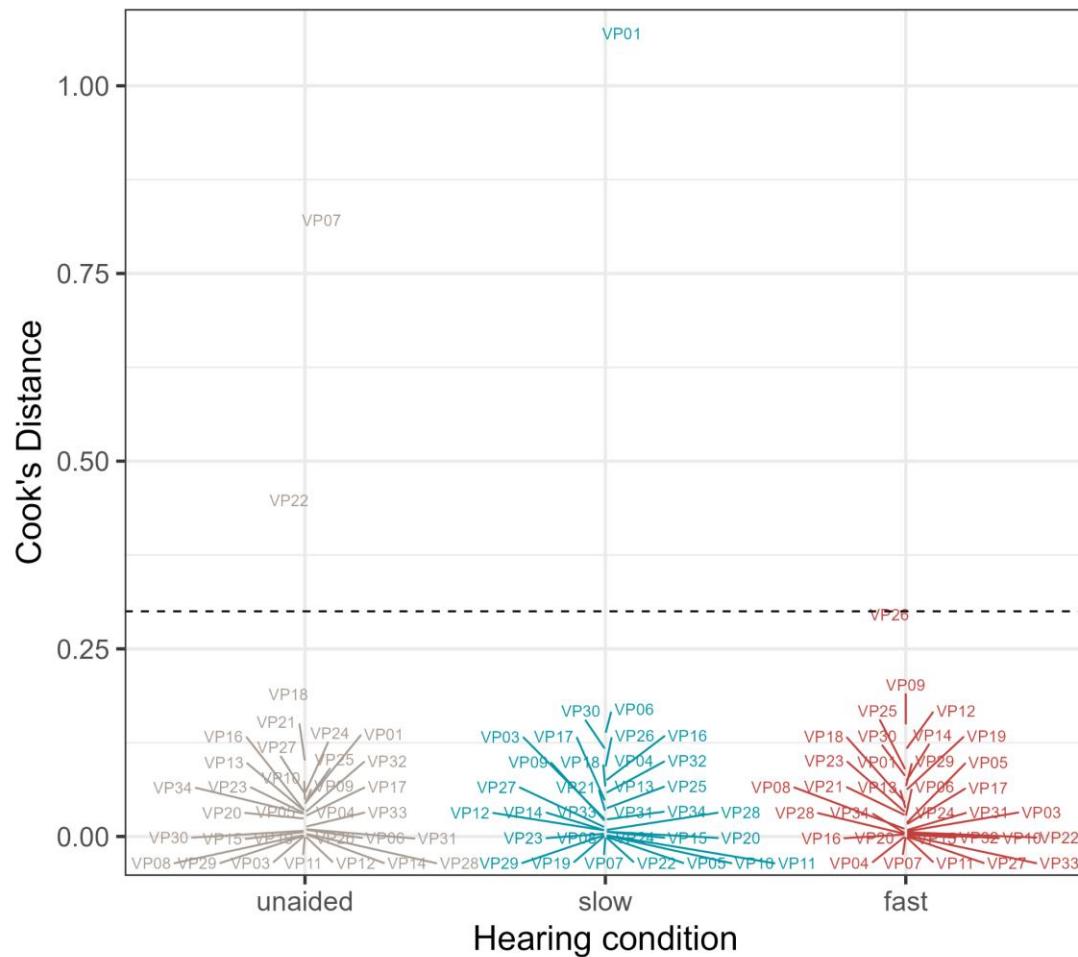


Figure A6.1 Cook's distance analysis for identifying influential data points. The dashed horizontal line indicates the threshold for exclusion, determined through visual inspection. Three data points exceeding this threshold (VP01 in the slow compression condition, VP07 and VP22 in the unaided condition) were subsequently excluded due to chance or near-chance performance on the primary task. These data points were replaced by imputation (see 'Analysis' section).

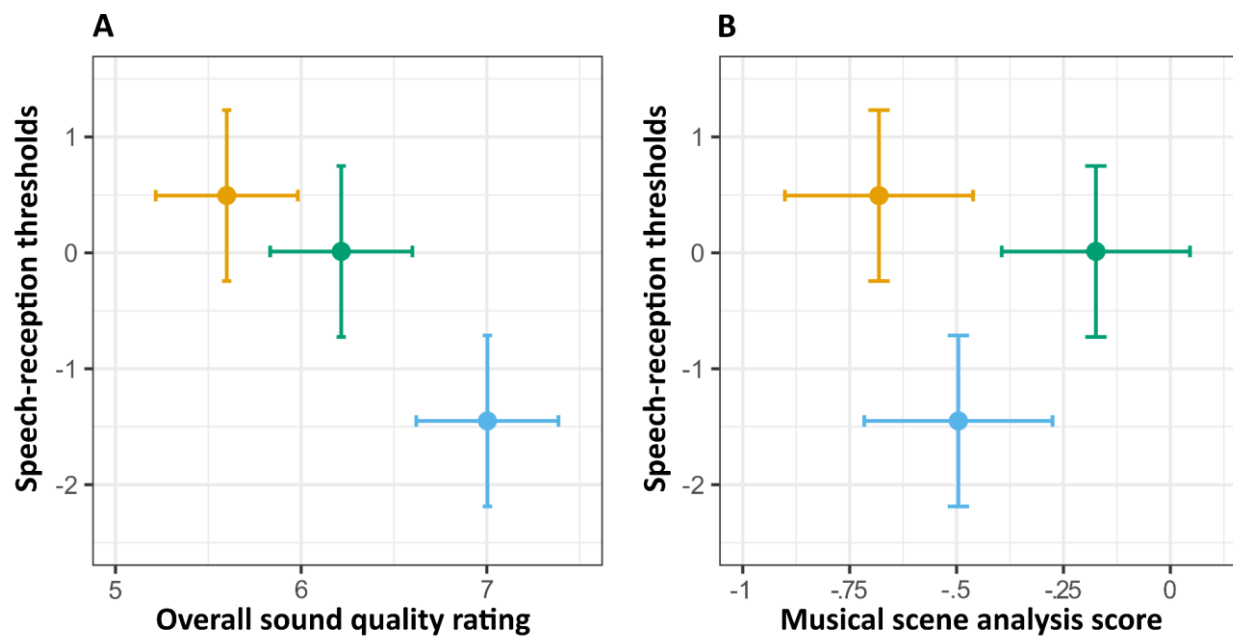


Figure A6.2 Estimated means from Model A1 for musical scene analysis (MSA) scores, Model B2 for sound quality rating factor (SQR) scores, and Model C1 for speech reception thresholds (SRT) across three conditions: slow (blue), unaided (yellow), fast (green). Panel A shows the relationship between SRT and SQR scores, while Panel B displays the relationship between SRT and MSA scores. Error bars represent 95% confidence intervals.

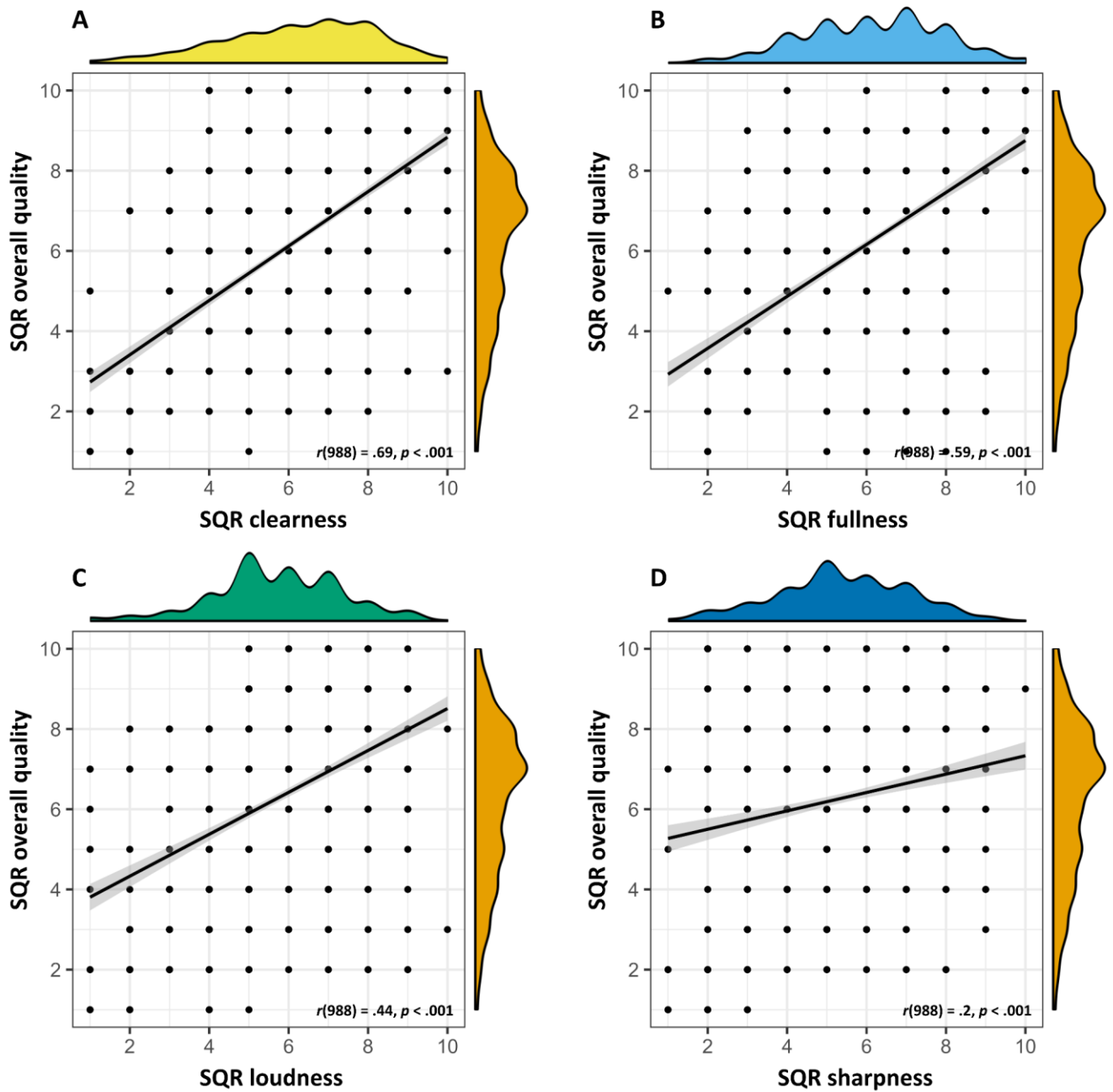


Figure A6.3 Relationship between overall sound quality ratings (SQR) and individual sound quality dimensions: (A) clarity, (B) fullness, (C) loudness, and (D) sharpness. Each plot displays a scatterplot with a linear regression line (solid black) and a 95% confidence interval (shaded area) for each dimension in relation to overall quality. Density plots for each dimension are displayed along the top and right margins to illustrate the distribution of ratings. Pearson correlation coefficients are shown within each panel, indicating significant positive correlations between overall quality and each dimension.

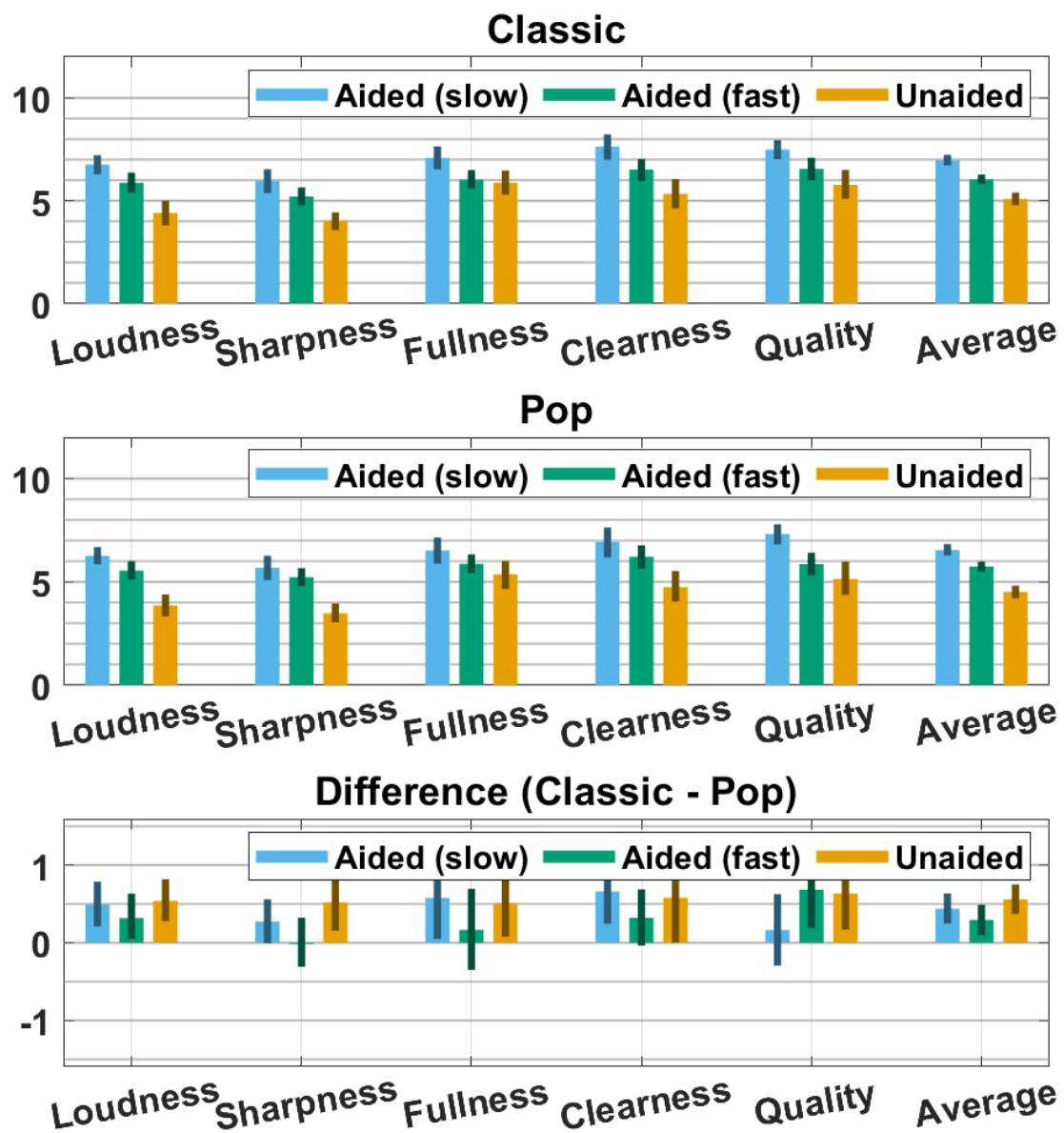


Figure A6.4 Comparative assessment of sound quality attributes under different listening conditions. The bar graphs display ratings for various sound quality attributes (Loudness, Sharpness, Fullness, Clearness, and Overall Quality) across two musical genres: Classical and Pop. The ratings were evaluated under three conditions: unaided, aided with slow compression (Aided slow), and aided with fast compression (Aided fast). The top two panels show the mean ratings for each attribute for Classical and Pop music, respectively. The bottom panel displays the mean differences in ratings between Classical and Pop music for each attribute and condition. Error bars represent standard deviations. Higher values indicate more positive evaluations of the sound quality attributes.

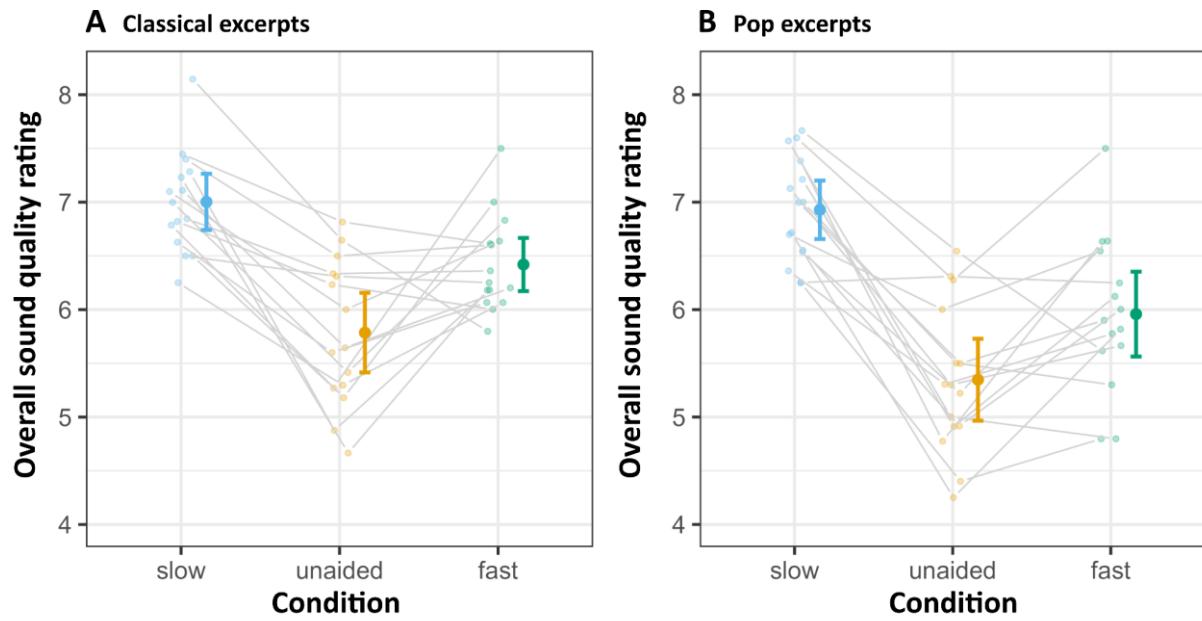


Figure A6.5 Mean overall sound quality rating (SQR) factor scores across three listening conditions (slow, unaided, and fast) for classical (Panel A) and pop excerpts (Panel B). Individual shaded dots represent SQR scores averaged over stimuli. Accordingly, average results of the same stimuli are displayed for the three listening conditions. Error bars represent 95% confidence intervals. Note significance at $*p < .05$, $**p < .01$, $***p < .001$.

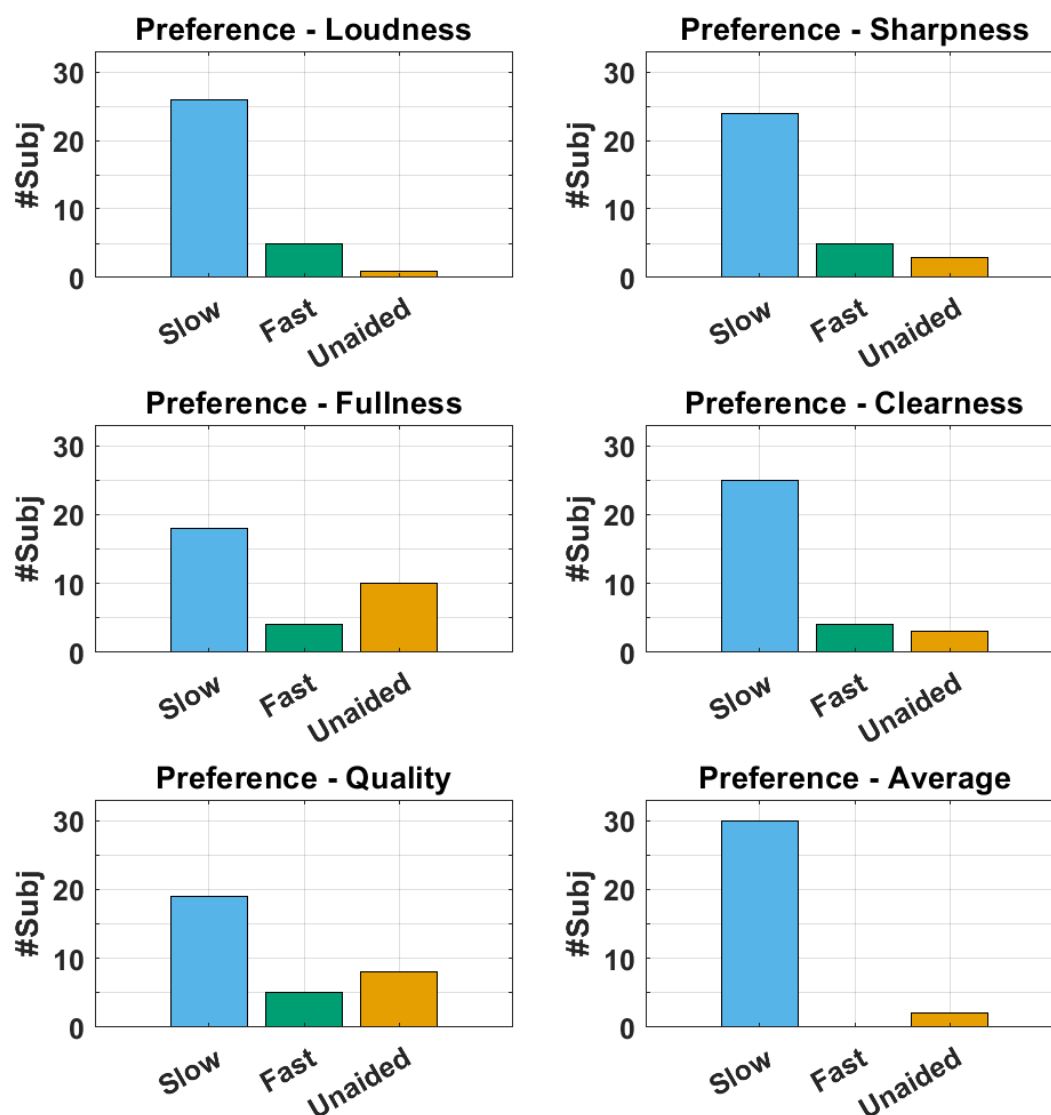


Figure A6.6 Preferred listening condition for each sound quality dimension. The bar graphs display the preferred condition across various sound quality dimensions (Loudness, Sharpness, Fullness, Clearness, and Overall Quality), averaged across two musical genres: Classical and Pop. Ratings were evaluated under three conditions: unaided, aided with slow compression (slow), and aided with fast compression (fast). Each panel represents a different sound quality dimension. The bars indicate the number of subjects who, on average, rated each condition as the highest. Higher values reflect a greater number of subjects preferring that condition. Please note that for loudness and sharpness, the bipolar scales used (e.g., 'very soft' to 'very loud') indicate extreme scores and may thus primarily reflect perceived intensity rather than an explicit preference.

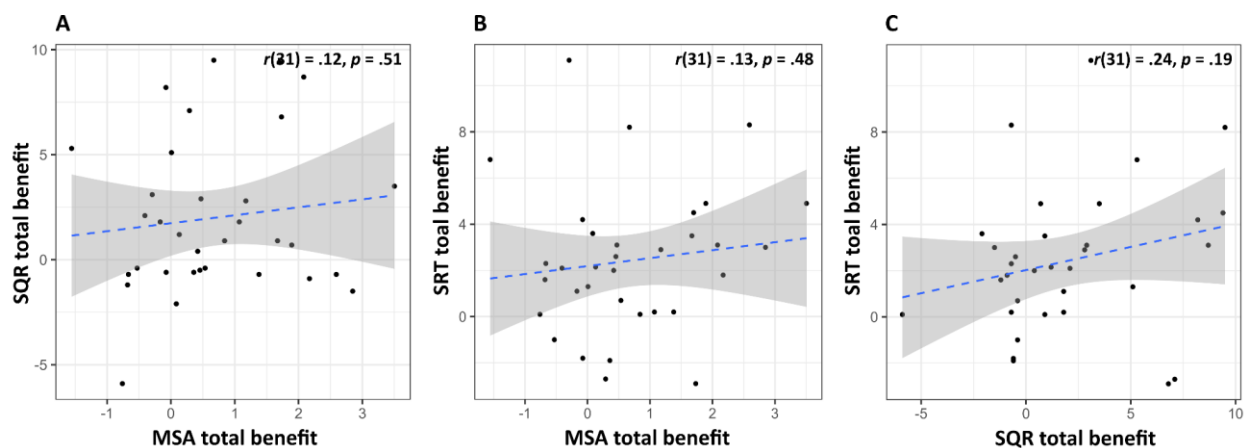


Figure A6.7 Scatterplots showing the relationships between total benefit scores across three measures: (A) SQR (overall Sound Quality Rating) and MSA (Musical Scene Analysis abilities), (B) SRT (Speech Reception Threshold in dB SNR) and MSA, and (C) SRT and SQR. Total benefit scores were calculated by subtracting the unaided condition scores from both the fast DRC and slow DRC condition scores, then summing these differences to create an aggregate do this analysis with the individuals best DRC in task 1 vs. best DRC setting in benefit measure. Each plot displays individual data points along with a linear regression line (dashed blue) and 95% confidence interval (shaded area). Pearson correlation coefficients and p-values are provided in each panel.

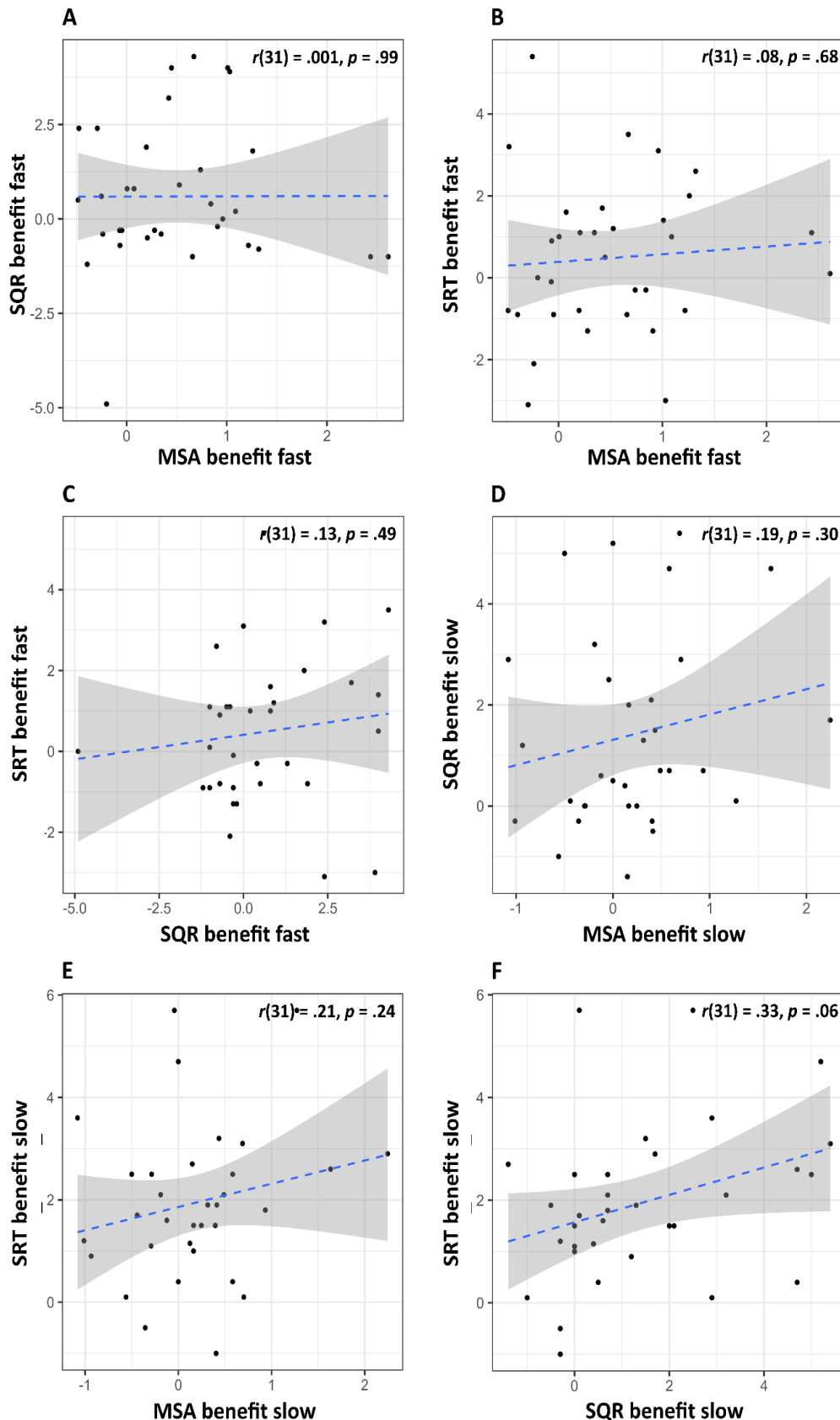


Figure A6.8 Scatterplots illustrating the relationships between benefit scores across tasks and DRC settings (fast and slow). Panels (A), (B), and (C) depict the correlations between benefit scores under the fast DRC condition: (A) overall sound quality rating (SQR) and Musical scene analysis abilities (MSA), (B) Speech reception threshold (SRT) and MSA, and (C) SRT and SQR.

Figure caption is continued on the following page.

Figure A6.8 Panels (D), (E), and (F) depict correlations for the slow DRC condition: (D) SQR and MSA, (E) SRT and MSA, and (F) SRT and SQR. Each plot includes a dashed blue regression line with a 95% confidence interval shaded in gray. Pearson correlation coefficients and p-values are provided in each panel.
continued

8.2. Tables

Table A3.1

Self-assessment of individual' hearing impairment in German language

Question	
Do you feel that you have a hearing loss?	
<i>(While this may not map onto your experience exactly, which of these best describes your hearing without hearing aids.)</i>	
Options	
☐	No, I do not feel that I have a hearing loss.
☐	Yes, I have the feeling of being mildly hearing impaired (hearing loss of 20-40 dB): When I am talking to a person in a quiet room, I can usually understand a conversation. In noisy situations (e.g. in a pub) and in group conversations, I sometimes have problems understanding speech.
☐	Yes, I have the feeling of being moderately hearing impaired (hearing loss of 41-60 dB): I have difficulty understanding speech and often have to ask others to repeat what they have said, both in face-to-face conversations and on the phone. I find it difficult to follow conversations in noisy situations. I have to turn up the TV and radio to hear them clearly.
☐	Yes, I have the feeling of being severely hearing impaired (hearing loss of 61-80 dB): I can barely understand speech without hearing aids and use lip-reading/sign language to assist communication.
☐	Yes, I have the feeling of being profoundly hearing impaired (hearing loss bordering on deafness; 81-95 dB): Ordinary conversations are unintelligible. I cannot hear shouted words and have to rely on lip reading and/or sign language to communicate.
☐	I don't know

Note. Individuals choosing 'I don't know' has been categorized as normal hearing. The version presented to German speaking participants is presented in Table A3.2.

Table A3.2

Self-assessment of individual' hearing impairment in German language

Question	
Haben Sie das Gefühl, dass Sie einen Hörverlust haben?	
<i>(Auch wenn dies nicht genau auf Ihre Erfahrung zutrifft, sollten Sie sich überlegen, welche dieser Angaben Ihr Gehör ohne Hörgeräte am besten beschreibt.)</i>	
Options	
☐	Nein, ich habe nicht das Gefühl einen Hörverlust zu haben.
☐	Ja, ich habe das Gefühl, leichtgradig schwerhörig zu sein (Hörverlust von 20–40 dB): Wenn ich mit einer Person in einem ruhigen Raum spreche, kann ich normalerweise ein Gespräch verstehen. In lauten Situationen (z. B. in einer Kneipe) und bei Gruppengesprächen habe ich manchmal Probleme, Sprache zu verstehen.
☐	Ja, ich habe das Gefühl, mittelgradig schwerhörig zu sein (Hörverlust von 41–60 dB): Ich habe Schwierigkeiten, Sprache zu verstehen, und muss häufig andere bitten, das Gesagte zu wiederholen, sowohl im persönlichen Gespräch als auch am Telefon. In lauten Situationen fällt es mir schwer, Gesprächen zu folgen. Ich muss den Fernseher und das Radio lauter stellen, um sie deutlich zu hören.
☐	Ja, ich habe das Gefühl, hochgradig schwerhörig zu sein (Hörverlust von 61–80 dB): Ich kann Sprache ohne Hörgeräte kaum verstehen und benutze Lippenlesen/Gebärdensprache zur Unterstützung der Kommunikation.
☐	Ja, ich habe das Gefühl, an Taubheit grenzend schwerhörig zu sein (Hörverlust von 81–95 dB): Gewöhnliche Gespräche sind unverständlich. Ich kann geschriebene Wörter nicht hören und muss mich auf das Lippenlesen und/oder Gebärdensprache verlassen, um mich zu verständigen.
☐	Ich weiß nicht.

Note. Individuals choosing 'Ich weiß nicht' has been categorized as normal hearing. The version presented to English speaking participants is presented in Table A3.1.

Table A3.3*Musicians average sound level exposure per instrument*

Instrument Group	Instruments	sound level estimates (dB A)							
		Schmidt et al., 2011 ^a		O'Brien et al., 2013 ^b	Phillips & Mace, 2008 ^c	Laitinen et al., 2003 ^d		Overall	
		<i>IP</i>	<i>GP</i>	<i>IP</i>	<i>IP</i>	<i>IP</i>	<i>GP</i>	<i>IP</i>	<i>GP</i>
Violins	1 st Violin	89 ^e	86	87.5 ^e	87 ^e	86 ^e	88.3 ^e	87.4	87.2
	2 nd Violin	89 ^e	86.2	87.5 ^e	87 ^e	86 ^e	88.3 ^e	87.4	87.3
Violas/ Cellos	Violoncello	-	81.8	80	87 ^e	88	89.8	85.0	85.8
	Viola	90.5	87.7	85.5	87 ^e	86	90.3	87.3	89
Woodwinds	Flute/ Piccolo	101.2	90.4	94	90.4 ^e	96	90.7 ^e	95.4	90.6
	Oboe	-	86.7	85	90.4 ^e	91 ^e	90.7 ^e	88.8	88.7
	Bassoon	88.4	88	88	90.4 ^e	91 ^e	90.7 ^e	89.5	89.4
	Clarinet	-	88.6	95	90.4 ^e	91 ^e	90.7 ^e	92.1	89.7
Brass	Trombone	97.6	89.7	96	95.2 ^e	95	93 ^e	96.0	91.4
	Trumpet	90.9	92	95.5	95.2 ^e	97	92.3	94.7	92.2
	Horn	-	-	93.5	95.2 ^e	95	93 ^e	94.6	93
	Tuba	-	87.8	93.5	95.2 ^e	95	93 ^e	94.6	90.4
Percussion/ Bass	Contrabass	75.6	82.7	75	90.1 ^e	79	87	79.9	84.9
	Kettledrum/ Timbals/ Side drums	-	87.2	97.5	90.1 ^e	99	91.3 ^e	95.5	89.3
Others	Harp	-	-	83	-	89	87	86.0	87
	Celeste	-	90	-	-	-	-	90.0	90

Note. Hours of noise exposure per week was estimated by aggregating musicians' reported average playing durations across various settings - namely, individual practice and group performances (IP = individual practice; GP = group practice). While it's reasonable to assume consistent playing habits throughout the year, these averages were projected over a standard 44-week year (equivalent to 1,780 working hours instead of 2080 hours annually, as is typical in Germany) to calculate annual noise exposure. Sound exposure levels differ notably across instruments, musicians' repertoires, and between solo practice and group performances (e.g., Rodrigues et al., 2014).

Table note is continued on the following page.

Note (Table A3.3) continued. To accommodate these variances, the metric derived sound level estimates from average decibel level data for various instruments and ensembles, playing in different acoustical settings and repertoires, as sourced from existing literature (Laitinen et al., 2003; O'Brien et al., 2013; Phillips & Mace, 2008; Schmidt et al., 2011). Furthermore, sound level estimates that are experienced by musicians were calculated by considering individual practice sessions and group performances separately.

^a Estimates reflect measurements from two orchestras during group rehearsals and actual concerts featuring varied repertoires, including Niels W. Gade's "Ossian Overture" and Dmitri Shostakovich's "Violin Concerto No. 1." Average measurements from both the left and right ears are reported.

^b Measurements of a daily practice session of duration (2.1 h) according to the average from both the left and right ears in a 4.8m x 4m room. Fortissimo passages, which can increase levels by up to 8 dB, are not included in these measurements. ^c Average estimates per instrument group were acquired by measuring students practicing in approx. 3m x 3.66m. ^d Average sound level for performances and orchestra rehearsals of one orchestra playing 'Die Walküre', 'Insect Life', 'Don Giovanni', and 'Swan Lake'.

^e Measurements are averaged over instrument groups.

Table A3.4*Musicians average sound level exposure per instrument group*

Instrument Group	Instruments in group	sound level exposure estimate (dB A)	
		individual practice	rehearsals / concerts
Violins	1st Violin, 2nd Violin	87.4	87.2
Violas/Cellos	Violoncello, Viola	86.1	87.4
Woodwinds	Flute/Piccolo, Oboe, Bassoon, Clarinet	91.4	89.6
Brass	Trombone, Trumpet, Horn, Tuba	94.9	91.7
Percussion/Bass	Contrabass, Kettledrum/Timbals	87.7	87.1
Others	Harp, Celeste	88	88.5

Note. Estimates are based on averages per instrument group from several sound level measurements provided by the literature as reported in Table A3 (see Supp. Mat.; Laitinen et al., 2003; Phillips & Mace, 2008; O'Brien et al., 2013; Schmidt et al., 2011) calculated for individual practice and rehearsals / orchestral concert performances separately. Notably, estimated exposure levels for individual practice sessions are reported to be higher, compared to group performances. The main contributing factor to this higher exposure is the continuous play during individual practice sessions compared to group rehearsals (e.g., see Schmidt et al., 2011). The instrument classification was from Russo et al. (2013).

Table A3.5*Types of hearing protection: Average attenuation levels and usage*

Hearing protection	<i>attenuation dB^d</i>
DIY [e.g., cotton wool] ^a	7
Standard ear plugs ^b	31
Plastic ear plugs ^b	26
Balanced sound-reducing plugs ^b	16
Customised ear plugs ^c	22
None	0

Note. To establish the probability estimate for the attenuation within the unit of noise exposure, reported frequency of hearing protection usage was converted into probability estimates: "Never" equals 0% probability of wearing hearing protection, "rarely" = 25%, "sometimes" = 50%, "mostly" = 75%, and "always" = 100%. Attenuation values (A) of the respective hearing protection devices (ear plugs) were estimated based on the hearing protection guide reported in the literature; ^a Safety Research Corporation of America (2020). ^b Guest et al. (2018). ^c Kusy & Châtillon (2012) ^d Overall estimates are provided.

Table A3.6*Types of hearing protection: Average attenuation levels and usage per group*

Hearing protection	attenuation dB	N (prop) ^d	
		<i>amateurs</i>	<i>professional</i>
DIY [e.g., cotton wool] ^a	7	0	3 (.015)
Standard ear plugs ^b	31	5 (.025)	5 (.025)
Plastic ear plugs ^b	26	0	4 (.02)
Balanced sound-reducing plugs ^b	16	3 (.015)	5 (.025)
Customised ear plugs ^c	22	15 (.075)	89 (.445)
None	0	177 (.885)	94 (.47)

Note. The attenuation depends on several factors, including the frequency of the signal; here overall measures are provided. ^a Safety Research Corporation of America (2020). ^b Guest et al. (2018). ^c Kusy & Châtillon (2012). ^d N (responses) for both groups of musicians after the matching procedure. Proportions are provided in brackets.

Table A3.7*Results from CFA for items indicating music-related hearing problems*

Item		Mean* (SD)	Corrected Item-Total Correlation	Cronbach's α if Item Deleted	Factor Loadings (lambda)
No	When making music and listening to music...				
1	...I have the impression that the sound is washed out/ blurred.	4.6 (.74)	.62	.80	.61
2	...I cannot hear well what I am playing.	4.2 (1.02)	.63	.80	.63
3	...I cannot hear well what other orchestra members	4.0 (.94)	.77	.76	.8
4	...I have problems perceiving intonation or pitch differences clearly.	4.3 (.86)	.70	.78	.71
5	...I have difficulty recognising other instruments.	4.5 (.69)	.66	.79	.65
6	...the music is often too soft or too loud.	4.0 (1.00)	.57	.81	.57
Model Fit Indices				Value	References
Robust Comparative Fit Index				.994	> .95
Robust Tucker-Lewis Index				.990	> .95
Robust RMSEA				.050	< .06
SRMR				.045	< .08
R ²				.44	
Cronbach's α				.82	

Note. N = 575. Raw Cronbach's α = .82, 95% CI[0.8, 0.84]. The kurtosis values are within acceptable ranges. Adequacy of the analysis was assessed by the Kaiser-Meyer-Olkin index (.84) and Bartlett's test of sphericity ($\chi^2(21) = 1261$, $p < .001$). Participants rated items on a 5-Point-Likert-Scale, ranging from 1 = always to 5 = never. *Low scores represent participants consistently providing the most extreme responses on all the assessed items, signifying intense experiences with hearing issues in the context of music.

Table A3.8*Results from CFA for items indicating hearing health awareness (HHA)*

Item		Mean (SD)	Corrected Item-Total Correlation	Cronbach's α if Item Deleted	Factor Loadings (lambda)
No					
1	Hearing health is extremely important to me.	1.3 (.5)	.23	.78	.18
2	I have my hearing checked regularly.	2.62 (1.1)	.40	.77	.34
3	The sound levels in the orchestra are so high at times that I should wear hearing protection.	2.1 (1.1)	.80	.69	.86
4	Do you perceive the sound level exposure in the orchestra profession as a risk to your hearing health?	2.19 (1.5)	.78	.69	.81
5	How do you feel about the sound level exposure during the individual training periods?	3.43 (.7)	.30	.70	.29
6	How do you feel about the sound level exposure during the rehearsals?	2.3 (.8)	.74	.68	.76
7	How do you feel about the sound level exposure during concerts?	2.2 (.8)	.66	.69	.66
Model Fit Indices				Value	References
Robust Comparative Fit Index				.973	> .95
Robust Tucker-Lewis Index				.959	> .95
Robust RMSEA				.081*	< .06
SRMR				.062	< .08
R ²				.37	
Cronbach's α				.76	

Note. N = 553. Raw Cronbach's α = .76, 95% CI [.73, .79]. The kurtosis values are within acceptable ranges. Adequacy of the analysis was assessed by the Kaiser-Meyer-Olkin index (.79) and Bartlett's test of sphericity ($\chi^2(21) = 454, p < .001$). Participants rated items on a 4-Point-Likert-Scale, ranging from 1 = "true/applies" to 4 = "not true/does not apply". *RMSEA is .081 90% CI [.061, .101] with RMSEA values less than .05 indicate a close fit, values up to .08 indicate a reasonable error of approximation, and values greater than .10 suggest a poor fit. This model still falls into the category of a reasonable fit. All items were used to compute the final factor score, as excluding any items did not noticeably enhance the reliability. The final scores were transformed to span from 0 to 10, with higher values indicating a higher hearing health awareness.

Table A4.1

Model summary of the final Bayesian Model (1F) of Experiment 1

Random Effects:			Median	95% HDI	pd	Rhat	ESS
~ excerpt (levels: 156)			1.56	[1.35, 1.79]	1	1	1703
~ participants (levels: 657)			2.08	[1.88, 2.31]	1	1	2003
Fixed Effects:							
LEVEL	PRESENCE YES	<i>Level -15</i>	-2.43	[-3.20, -1.63]	1	1.002	2549
		<i>Level -10</i>	-.00039	[-.75, .77]	.5002	1.002	2511
		<i>Level -5</i>	1.52	[.77, 2.29]	1	1.003	2536
		<i>Level 0</i>	2.33	[1.58, 3.11]	1	1.002	2561
	PRESENCE NO	<i>Level -inf (no target in mix)</i>	.21	[-.78, 1.15]	.6642	1.001	2865
TARGET	PRESENCE YES	<i>BASS</i>	-2.17	[-3.14, -1.21]	.9998	1.002	2420
		<i>GUITAR</i>	.91	[-.02, 1.82]	.9726	1.001	1870
		<i>PIANO</i>	1.32	[.37, 2.23]	.9971	1.002	2119
		<i>LEAD-VOCALS</i>	1.32	[.39, 2.23]	.9962	1	2219
	PRESENCE NO	<i>BASS</i>	.02	[-1.01, 1.10]	.5192	1	2598
		<i>GUITAR</i>	-.7	[-1.76, .37]	.9001	1.001	2408
		<i>PIANO</i>	-.72	[-1.76, .35]	.9066	1.001	2465
		<i>LEAD-VOCALS</i>	1.62	[.58, 2.69]	.9991	1	2834
NUM	PRESENCE YES	<i>3 instruments in the mix</i>	.85	[.20, 1.53]	.9952	1.004	1397
	PRESENCE NO	<i>3 instruments in the mix</i>	.79	[.24, 1.35]	.997	1.005	2006
		<i>guessing intercept</i>	.43	[.41, .45]	1	1.001	4238
		<i>inattention intercept</i>	.00249	[.00, .01]	1	1	5570
Model fit:			Estimate				
ELPD			-13645.2				
LOOIC			27290.45				
WAIC			27282.69				
Sigma			1				
Log_loss			0.36				

Note. Parameter column (most left) shows the interaction of the presence of the target (PRESENCE YES = target instrument is part of the mixture) and the three employed parameter: level-ratio between target and the mixture (LEVEL), choice of the target instrument (TARGET), and the number of instruments in the mix (NUM 3 = three instruments in the mix). The estimates are presented in log-scale format.

Table A4.2

Model summary of the final Bayesian Model (2F) of Experiment 2

Random Effects:			Median	95% HDI	pd	Rhat	ESS
~ excerpt (levels: 160)			1.63	[1.40, 1.89]	1	1	3267
~ participants (levels: 80)			1.56	[1.27, 1.90]	1	1	2272
Fixed Effects:							
<i>ILD</i>	<i>PRESENCE YES</i>	<i>ILD B (90°)</i>	.11	[.17, .44]	.76	1	9398
		<i>ILD C (180°)</i>	.41	[.08, .72]	.99	1	8991
		<i>ILD D (180° + R)</i>	-.008	[-.33, .30]	.52	1	8502
	<i>PRESENCE NO</i>	<i>ILD B (90°)</i>	.33	[.00, .62]	.97	1.001	8614
		<i>ILD C (180°)</i>	.38	[.05, .70]	.98	1	9091
		<i>ILD D (180° + R)</i>	.91	[.51, 1.25]	1	1	8561
<i>TARGET</i>	<i>PRESENCE YES</i>	<i>BASS</i>	.71	[-.07, 1.45]	.96	1.003	2149
		<i>GUITAR</i>	.43	[-.03, 1.19]	.86	1.001	1724
		<i>PIANO</i>	-.22	[-1.01, .57]	.71	1.004	1987
		<i>LEAD-VOCALS</i>	1.51	[.72, 2.30]	1	1	2582
	<i>PRESENCE NO</i>	<i>BASS</i>	-1.37	[-2.16, -.58]	1	1	2010
		<i>GUITAR</i>	1.05	[.29, 1.84]	1	1.001	2084
		<i>PIANO</i>	1.67	[.89, 2.43]	1	1	2438
		<i>LEAD-VOCALS</i>	1.59	[.80, 2.39]	1	1	2834
<i>NUM</i>	<i>PRESENCE YES</i>	<i>3 instruments in the mix</i>	-1.26	[-1.80, -.74]	1	1.001	2233
	<i>PRESENCE NO</i>	<i>3 instruments in the mix</i>	-.46	[-.96, .06]	.96	1.001	2484
		<i>guessing intercept</i>	.41	[.40, .44]	1	1	10165
		<i>inattention intercept</i>	.00199	[.00, .007]	1	1	8578
Model fit:			Estimate				
ELPD			-4867				
LOOIC			9734				
WAIC			9733.1				
Sigma			1				
Log_loss			0.363				

Note. Fixed effects column (most left) shows the interaction of the presence of the target (PRESENCE YES = target instrument is part of the mixture) and the three employed parameter: choice of the target instrument (TARGET), the number of instruments in the mix (NUM 3 = three instruments in the mix), and the stereo width condition, elicited by inter-aural level differences (ILD): (A) All instruments are presented at 0°, i.e. the monaural reference; (B) the instruments are distributed evenly at a stereo width of 90°; (C) the instruments are distributed at full stereo width of 180°; (D) similar stereo panorama as in condition C, but the target is presented at a different position when the target instrument is presented in isolation compared to the presentation phase when the whole instrument mixture is playing. The estimates are presented in log-scale format.

Table A4.3*Model summary of the simple Bayesian Model (2C) of Experiment 2*

Random Effects:		Median	95% HDI	pd	Rhat	ESS
~excerpt (number of levels: 156)		2.12	[1.84, 2.47]	1	1	2273
~participant (number of levels: 567)		1.55	[1.27, 1.90]	1	1	1720
Fixed Effects:						
PRESENCE YES	<i>ILD A (0°)</i>	.61	[.06, 1.16]	.98	1.002	1041
	<i>ILD B (90°)</i>	.89	[.33, 1.43]	1	1.002	1075
	<i>ILD C (180°)</i>	.98	[.42, 1.51]	1	1.002	1081
	<i>ILD D (180° + R)</i>	1.51	[.96, 2.06]	1	1.001	1117
PRESENCE NO	<i>ILD A (0°)</i>	.53	[-.02, 1.05]	.97	1.006	1010
	<i>ILD B (90°)</i>	.63	[.06, 1.16]	.98	1.005	1044
	<i>ILD C (180°)</i>	.91	[.36, 1.44]	1	1.006	1055
	<i>ILD D (180° + R)</i>	.49	[-.07, 1.02]	.96	1.004	1047
<i>guessing intercept</i>		.41	[.40, .44]	1	1	7772
<i>inattention intercept</i>		.00212	[.00, .01]	1	1	7131
Model fit:		Estimate				
ELPD		-4869.90				
LOOIC		9739.80				
WAIC		9738.17				
Sigma		1				
Log_loss		.36				

Note. Fixed effects column (most left) shows the interaction of the presence of the target (PRESENCE YES = target is part of the mixture) and the stereo condition, elicited by inter-aural level differences (ILD): (A) All instruments are presented at 0°, i.e. the monaural reference; (B) the instruments are distributed evenly at a stereo width of 90°; (C) the instruments are distributed at full stereo width of 180°; (D) similar stereo panorama as in condition C, but the target is presented at a different position when the target instrument is presented in isolation compared to the presentation phase when the whole instrument mixture is playing. The estimates are presented in log-scale format. The results illustrate the main effect for the ILD conditions.

Table A4.4*Self-assessment of individual' hearing impairment*

Question	
Do you feel that you have a hearing loss?	
<i>(While this may not map onto your experience exactly, which of these best describes your hearing without hearing aids.)</i>	
Options	
☐	No, I do not feel that I have a hearing loss.
☐	Yes, I have the feeling of being mildly hearing impaired (hearing loss of 20-40 dB): When I am talking to a person in a quiet room, I can usually understand a conversation. In noisy situations (e.g. in a pub) and in group conversations, I sometimes have problems understanding speech.
☐	Yes, I have the feeling of being moderately hearing impaired (hearing loss of 41-60 dB): I have difficulty understanding speech and often have to ask others to repeat what they have said, both in face-to-face conversations and on the phone. I find it difficult to follow conversations in noisy situations. I have to turn up the TV and radio to hear them clearly.
☐	Yes, I have the feeling of being severely hearing impaired (hearing loss of 61-80 dB): I can barely understand speech without hearing aids and use lip-reading/sign language to assist communication.
☐	Yes, I have the feeling of being profoundly hearing impaired (hearing loss bordering on deafness; 81-95 dB): Ordinary conversations are unintelligible. I cannot hear shouted words and have to rely on lip reading and/or sign language to communicate.
☐	I don't know
Note. Individuals choosing 'I don't know' has been categorized as normal hearing. The version presented to German speaking participants is presented in Table A4.5.	

Table A4.5

Self-assessment of individual' hearing impairment in German language

Question	
Haben Sie das Gefühl, dass Sie einen Hörverlust haben?	
<i>(Auch wenn dies nicht genau auf Ihre Erfahrung zutrifft, sollten Sie sich überlegen, welche dieser Angaben Ihr Gehör ohne Hörgeräte am besten beschreibt.)</i>	
Options	
☐	Nein, ich habe nicht das Gefühl einen Hörverlust zu haben.
☐	Ja, ich habe das Gefühl, leichtgradig schwerhörig zu sein (Hörverlust von 20–40 dB): Wenn ich mit einer Person in einem ruhigen Raum spreche, kann ich normalerweise ein Gespräch verstehen. In lauten Situationen (z. B. in einer Kneipe) und bei Gruppengesprächen habe ich manchmal Probleme, Sprache zu verstehen.
☐	Ja, ich habe das Gefühl, mittelgradig schwerhörig zu sein (Hörverlust von 41–60 dB): Ich habe Schwierigkeiten, Sprache zu verstehen, und muss häufig andere bitten, das Gesagte zu wiederholen, sowohl im persönlichen Gespräch als auch am Telefon. In lauten Situationen fällt es mir schwer, Gesprächen zu folgen. Ich muss den Fernseher und das Radio lauter stellen, um sie deutlich zu hören.
☐	Ja, ich habe das Gefühl, hochgradig schwerhörig zu sein (Hörverlust von 61–80 dB): Ich kann Sprache ohne Hörgeräte kaum verstehen und benutze Lippenlesen/Gebärdensprache zur Unterstützung der Kommunikation.
☐	Ja, ich habe das Gefühl, an Taubheit grenzend schwerhörig zu sein (Hörverlust von 81–95 dB): Gewöhnliche Gespräche sind unverständlich. Ich kann geschriebene Wörter nicht hören und muss mich auf das Lippenlesen und/oder Gebärdensprache verlassen, um mich zu verständigen.
☐	Ich weiß nicht.

Note. Individuals choosing 'Ich weiß nicht' has been categorized as normal hearing. The version presented to English speaking participants is presented in Table A4.4.

Table A6.1*Sound excerpts used within the sound quality rating task*

Dimension	Prompt	Lowest rating	Highest rating
Classic	Ludwig van Beethoven	Overture “Coriolan”, Op. 62	https://www.youtube.com/watch?v=DzINFjNPmqI
	Joseph Haydn	Symphony No. 31 in D major	https://www.youtube.com/watch?v=kcVZQel5NqQ
		Hob. I 31 – “Hornsignal”: I. Allegro	
	Felix Mendelssohn	Symphony n3 in A minor, Op.56	https://www.youtube.com/watch?v=RnkYSNA0-tA
	Wolfgang Amadeus Mozart	Fagottkonzert B-Dur K191 Allegro	https://www.youtube.com/watch?v=kQgLCsbaDk8
	Antonio Salieri	Sinfonia in D Major, 'Veneziana' - 1. Allegro assai	https://www.youtube.com/watch?v=Cayi95DvRNI
Pop	Dead Milkmen	Prisoners Cinema	Medley DB
	Hop Along	Sister Cities	Medley DB
	Invisible Familiars	Disturbing Wildlife	Medley DB
	Patrick Talbot	Fool	Medley DB
	Strand of Oaks	Spacestation	Medley DB

Note. Online videos (youtube.com) last accessed on: 29.10.2024.

Table A6.2

Sound quality dimensions of the sound quality rating task in the original language (German)

Dimension	Prompt	Lowest rating	Highest rating
Lautheit	Wie laut war der Klang?	Leise	Laut
Schärfe	Wie scharf/schrill oder weich war der Klang?	Weich	Scharf
Sattheit/Völle	Wie voll war der Klang?	Dünn	Voll
Heraushörbarkeit	Wie deutlich konnten Sie die einzelnen Instrumente voneinander unterscheiden?	Undeutlich	Sehr klar
Gesamtqualität	Wie gut war die allgemeine Klangqualität der Musik?	Schlecht	Gut

Note. -

Table A6.3*Descriptive Statistics and Correlations for Study Variables*

Model	Variable	AIC	BIC	R ² (cond.)	R ² (marg.)	ICC	Sigma
Musical scene analysis							
A1	MSA ~ condition + (1 ID)	192.8	205.8	.322	.1	.247	.552
A2 ^a	MSA ~ condition + order + (1 ID)	193.7	209.2	.327	.11	.248	.551
A3 ^b	MSA ~ condition * order + (1 ID)	197.2	217.9	.341	.11	.261	.552
Sound quality ratings models							
B1	SQR ~ condition + (1 ID) + (1 stim)	3779	3808	.357	.09	.294	1.53
B2 ^c	SQR ~ condition + genre + (1 ID) + (1 stim)	3778	3779	.358	.1	.289	1.53
B3 ^d	SQR ~ condition * genre + (1 ID) + (1 stim)	3781	3825	.36	.1	.289	1.53
B4 ^e	SQR ~ condition + genre + (1 ID) + (1 + condition stim)	3787	3787	NA	.13	NA	1.53
Speech reception thresholds							
C1	SRT ~ condition + (1 ID)	383.7	386.6	.755	.135	.717	1.12

Note. Total $N = 99$. CI = confidence interval; LL = lower limit; UL = upper limit. Compared to the simpler model, the model fit did not significantly improve the model fit (^a $\chi^2 = 1.12$, $p = .29$; ^b $\chi^2 = .52$, $p = .77$, ^d $\chi^2 = .75$, $p = .69$). Compared to the simpler model, the model fit did significantly improve the model fit (^c $\chi^2 = 7.85$, $p = .005$). ^eBoundary issue or singular fit, meaning the model may be overfitted or too complex

Table A6.4*Final model (A1) summary for the musical scene analysis (MSA) ability scores*

Effect	Estimate	SE	95% CI		p
			LL	UL	
Model A1: MSA ~ condition + (1 subject)					
Intercept	-.68	.11	-.9	-.46	< .001
Slow DRC	.19	.14	-.08	.46	.175
Fast DRC	.51	.14	.25	.78	< .001
Random effects					
Subject (Intercept)	.1	.32 (SD)			
Residual	.3	.55 (SD)			

Note. Total Observations = 99. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table A6.5*The correlation matrix for the test battery*

Variable	Overall quality	Clearness	Fullness	Loudness	Sharpness
Overall quality	-	.69***	.59***	.44***	.20***
Clearness		-	.53***	.48***	.26***
Fullness			-	.43***	.10**
Loudness				-	.65***
Sharpness					-

Note. Above the diagonal are the Pearson correlations calculated from a complete pair of values, which are significant at * $p < .05$, ** $p < .01$, *** $p < .001$. N = 988.

Table A6.6*Final model (B2) summary for the sound quality rating (SQR) factor scores*

Effect	Estimate	SE	95% CI		p
			LL	UL	
Model B2: SQR ~ condition + genre + (1 subject) + (1 stimuli)					
Intercept (Unaided)	5.76	.2	5.36	6.16	.37
Slow DRC	1.4	.12	1.17	1.64	<.001
Fast DRC	.62	.12	.38	.85	<.001
Genre: Pop	-.33	.15	-.62	-.05	.023
Random effects					
Subject (Intercept)	.87	.93 (SD)			
Stimuli (Intercept)	.09	.29 (SD)			
Residual	2.35	1.53 (SD)			

Note. Total Observations = 990. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table A6.7

Final model (C1) summary for the speech reception thresholds (SRT) from the Göttinger Sentence Test

Effect	Estimate	SE	95% CI		p
			LL	UL	
Model B2: SRT ~ condition + (1 subject)					
Intercept (Unaided)	.49	.36	-.23	1.22	.18
Slow DRC	-1.94	.28	-2.49	-1.4	<.001
Fast DRC	-.48	.28	-1.03	0.07	.08
Random effects					
Subject (Intercept)	3.18	1.8 (SD)			
Residual	1.3	1.12 (SD)			

Note. Total Observations = 990. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

LITERATURE

- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex (New York, N.Y. : 1991)*, 21(7), 1507–1518. <https://doi.org/10.1093/cercor/bhq198>
- Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., & Taylor, M. J. (2003). Developmental changes in distinguishing concurrent auditory objects. *Brain Research. Cognitive Brain Research*, 16(2), 210–218. [https://doi.org/10.1016/S0926-6410\(02\)00275-6](https://doi.org/10.1016/S0926-6410(02)00275-6)
- Alain, C., Cusimano, M., Garami, L., Backer, K. C., Habelt, B., Chan, V., & Hasher, L. (2018). Age-related differences in orienting attention to sound object representations. *Neurobiology of Aging*, 66, 1–11. <https://doi.org/10.1016/j.neurobiolaging.2018.01.026>
- Albert, M., Hurrelmann, K., & Quenzel, G. (Eds.). (2019). *Shell-Jugendstudie: 18 (2019). Jugend 2019: Eine Generation meldet sich zu Wort* (1. Auflage). Beltz.
- Anglada-Tort, M., Harrison, P. M. C., Lee, H., & Jacoby, N. (2023). Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Current Biology : CB*, 33(8), 1472–1486.e12. <https://doi.org/10.1016/j.cub.2023.02.070>
- Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 59, 126–137. <https://doi.org/10.1016/j.cortex.2014.07.013>
- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47 Suppl 2, S53–71. <https://doi.org/10.1080/14992020802301142>
- Auerbach, B. D., & Gritton, H. J. (2022). Hearing in Complex Environments: Auditory Gain Control, Attention, and Hearing Loss. *Frontiers in Neuroscience*, 16, 799787. <https://doi.org/10.3389/fnins.2022.799787>
- Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, 204(1), 91–101. <https://doi.org/10.1007/s00221-010-2299-y>
- Bayat, A., Farhadi, M., Pourbakht, A., Sadjedi, H., Emamdjomeh, H., Kamali, M., & Mirmomeni, G. (2013). A comparison of auditory perception in hearing-impaired and normal-hearing listeners: An auditory scene analysis study. *Iranian Red Crescent Medical Journal*, 15(11), e9477. <https://doi.org/10.5812/ircmj.9477>
- Benjamin, A. J., & Siedenburg, K. (2024). Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music. *JASA Express Letters*, 4(11), Article 113202. <https://doi.org/10.1121/10.0032474>
- Benjamin, A. J., & Siedenburg, K. (2025). Effects of spectral manipulations of music mixes on musical scene analysis abilities of hearing-impaired listeners. *PloS One*, 20(1), e0316442. <https://doi.org/10.1371/journal.pone.0316442>
- Bernstein, J. G., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *Journal of the Acoustical Society of America*, 120(6), 3929–3945. <https://doi.org/10.1121/1.2372452>

- Bertolo, M., Müllensiefen, D., Peretz, I., Woolley, S. C., Sakata, J. T., & Mehr, S. A. (2023). *Sex differences in music perception are negligible*. <https://doi.org/10.1101/2023.05.23.541970>
- Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise--a review. *Trends in Amplification*, 17(2), 75–93. <https://doi.org/10.1177/1084713813495459>
- Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, 64(5), 844–854. <https://doi.org/10.3758/bf03194750>
- Bey, C. E., & Zatorre, R. J. (2003). Recognition of interleaved melodies. An fMRI study. *Annals of the New York Academy of Sciences*, 999, 152–154. <https://doi.org/10.1196/annals.1284.017>
- Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., & Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *Journal of Neuroscience*, 35(5), 2161–2172. <https://doi.org/10.1523/JNEUROSCI.3915-14.2015>
- Bianchi, F., Carney, L. H., Dau, T., & Santurette, S. (2019). Effects of Musical Training and Hearing Loss on Fundamental Frequency Discrimination and Temporal Fine Structure Processing: Psychophysics and Modeling. *Journal of the Association for Research in Otolaryngology : JARO*, 20(3), 263–277. <https://doi.org/10.1007/s10162-018-00710-2>
- Birtwistle, E., Chernikova, O., Wünsch, M., & Niklas, F. (2025). Training of Executive Functions in Children: A meta-analysis of cognitive training interventions. *Sage Open*, 15(1), Article 21582440241311060. <https://doi.org/10.1177/21582440241311060>
- Boehnke, S. E., & Phillips, D. P. (2005). The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Perception & Psychophysics*, 67(6), 1088–1101. <https://doi.org/10.3758/BF03193634>
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press).
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244–249. <https://doi.org/10.1037/h0031163>
- Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32(1), 19–31. <https://doi.org/10.1037/h0081664>
- Bronkhorst, A. W., & Plomp, R. (1990). A clinical test for the assessment of binaural speech perception in noise. *Audiology*, 29(5), 275–285. <https://doi.org/10.3109/00206099009072858>
- Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America*, 92(6), 3132–3139. <https://doi.org/10.1121/1.404209>
- Brown, C. J., Jeon, E.-K., Driscoll, V., Mussoi, B., Deshpande, S. B., Gfeller, K., & Abbas, P. J. (2017). Effects of Long-Term Musical Training on Cortical Auditory Evoked Potentials. *Ear and Hearing*, 38(2), e74-e84. <https://doi.org/10.1097/AUD.0000000000000375>
- Bugos, J. A., Gbadamosi, A., Laesker, D., Chow, R., Sirocchi, S., Norgaard, M., Ghent, J., & Alain, C. (2024). Jazz Piano Training Modulates Neural Oscillations and Executive Functions in Older Adults. *Music Perception*, 41(5), 378–392. <https://doi.org/10.1525/mp.2024.41.5.378>
- Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health*, 11(4), 464–471. <https://doi.org/10.1080/13607860601086504>

- Bürge, M., Picinali, L., & Siedenburg, K. (2021). Listening in the Mix: Lead Vocals Robustly Attract Auditory Attention in Popular Music. *Frontiers in Psychology*, 12(December). <https://doi.org/10.3389/fpsyg.2021.769663>
- Bürge, M., & Siedenburg, K. (2023). Salience of Frequency Micro-modulations in Popular Music. *Music Perception: An Interdisciplinary Journal*, 41(1), 1–14. <https://doi.org/10.1525/mp.2023.41.1.1>
- Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews. Neuroscience*, 7(6), 477–484. <https://doi.org/10.1038/nrn1909>
- Calcutt, A. (2024). Development of auditory scene analysis: A mini-review. *Frontiers in Human Neuroscience*, 18, 1352247. <https://doi.org/10.3389/fnhum.2024.1352247>
- Cattell, R. B. (1966). Multivariate behavioral research and the integrative challenge. *Multivariate Behavioral Research*, 1(1), 4–23.
- Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 115–127. <https://doi.org/10.1037/0096-1523.27.1.115>
- Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, 137, 81–105. <https://doi.org/10.1016/j.cognition.2014.12.005>
- Crawley, E. J., Acker-Mills, B. E., Pastore, R. E., & Weil, S. (2002). Change detection in multi-voice music: The role of musical structure, musical training, and task demands. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 367–378. <https://doi.org/10.1037//0096-1523.28.2.367>
- Chen, B., Zhong, Y., Peng, W., Sun, Y., & Kong, W.-J. (2010). Age-related changes in the central auditory system: Comparison of D-galactose-induced aging rats and naturally aging rats. *Brain Research*, 1344, 43–53. <https://doi.org/10.1016/j.brainres.2010.04.082>
- Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience : A Journal and Virtual Library*, 13, 148–169. <https://doi.org/10.2741/2666>
- Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians. *PloS One*, 11(7), e0157638. <https://doi.org/10.1371/journal.pone.0157638>
- Clement, M., & Kandziora, M. (2021). Tonträgermarkt und Musiknutzung in Deutschland. *Deutscher Musikrat - Deutsches Musikinformationszentrum*(4), 2–15. <https://miz.org/de/beitraege/tontraegermarkt-und-musiknutzung>
- Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., & Zatorre, R. J. (2019). The Music-In-Noise Task (MINT): A Tool for Dissecting Complex Auditory Perception. *Frontiers in Neuroscience*, 13, 199. <https://doi.org/10.3389/fnins.2019.00199>
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Schellenberg, E. G., & Lima, C. F. (2023). Individual differences in musical ability among adults with no music training. *Quarterly Journal of Experimental Psychology (2006)*, 76(7), 1585–1598. <https://doi.org/10.1177/17470218221128557>
- Corrigall, K. A., & Trainor, L. J. (2011). Associations Between Length of Music Training and Reading Skills in Children. *Music Perception*, 29(2), 147–155. <https://doi.org/10.1525/mp.2011.29.2.147>
- Coutinho, E., & Dikken, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, 27(4), 658–684. <https://doi.org/10.1080/02699931.2012.732559>

- Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. *American Journal of Epidemiology*, 148(9), 879–886. <https://doi.org/10.1093/oxfordjournals.aje.a009713>
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 643–656. <https://doi.org/10.1037/0096-1523.30.4.643>
- Cunningham, L. L., & Tucci, D. L. (2017). Hearing Loss in Adults. *The New England Journal of Medicine*, 377(25), 2465–2473. <https://doi.org/10.1056/NEJMra1616601>
- Darwin, C. J. (2005). Pitch and Auditory Grouping. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research. Pitch* (Vol. 24, pp. 278–305). Springer-Verlag. https://doi.org/10.1007/0-387-28958-5_8
- Davidson, A., & Souza, P. (2024). Relationships Between Auditory Processing and Cognitive Abilities in Adults: A Systematic Review. *Journal of Speech, Language, and Hearing Research : JSLHR*, 67(1), 296–345. https://doi.org/10.1044/2023_JSLHR-22-00716
- Davidson, P., & Winocur, G. (2017). Aging and Cognition☆. In Reference Module in Neuroscience and Biobehavioral Psychology. Elsevier. <https://doi.org/10.1016/B978-0-12-809324-5.00252-2>
- Deary, I. J. (1994). Intelligence and auditory discrimination: Separating processing speed and fidelity of stimulus representation. *Intelligence*, 18(2), 189–213. [https://doi.org/10.1016/0160-2896\(94\)90027-2](https://doi.org/10.1016/0160-2896(94)90027-2)
- Dinces, E., & Sussman, E. S. (2017). Attentional Resources Are Needed for Auditory Stream Segregation in Aging. *Frontiers in Aging Neuroscience*, 9, 414. <https://doi.org/10.3389/fnagi.2017.00414>
- Dritsakis, G., van Besouw, R. M., & O' Meara, A. (2017a). Impact of music on the quality of life of cochlear implant users: A focus group study. *Cochlear Implants International*, 18(4), 207–215. <https://doi.org/10.1080/14670100.2017.1303892>
- Dritsakis, G., van Besouw, R. M., Kitterick, P., & Verschuur, C. A. (2017b). A Music-Related Quality of Life Measure to Guide Music Rehabilitation for Adult Cochlear Implant Users. *American Journal of Audiology*, 26(3), 268–282. https://doi.org/10.1044/2017_AJA-16-0120
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, 21, 2331216517744675. <https://doi.org/10.1177/2331216517744675>
- Dubinsky, E., Wood, E. A., Nespoli, G., & Russo, F. A. (2019). Short-Term Choir Singing Supports Speech-in-Noise Perception and Neural Pitch Strength in Older Adults With Age-Related Hearing Loss. *Frontiers in Neuroscience*, 13, 1153. <https://doi.org/10.3389/fnins.2019.01153>
- Durai, M., O'Keeffe, M. G., & Searchfield, G. D. (2018). A Review of Auditory Prediction and Its Potential Role in Tinnitus Perception. *Journal of the American Academy of Audiology*, 29(6), 533–547. <https://doi.org/10.3766/jaaa.17025>
- Durai, M., Sanders, M., Kobayashi, K., & Searchfield, G. D. (2019). Auditory Streaming and Prediction in Tinnitus Sufferers. *Ear and Hearing*, 40(2), 345–357. <https://doi.org/10.1097/AUD.0000000000000620>
- Eddins, A. C., Ozmeral, E. J., & Eddins, D. A. (2018). How aging impacts the encoding of binaural cues and the perception of auditory space. *Hearing Research*, 369, 79–89. <https://doi.org/10.1016/j.heares.2018.05.001>
- Eitel, M., Ruth, N., Harrison, P., Frieler, K., & Müllensiefen, D. (2024). Perception of Chord Sequences Modeled with Prediction by Partial Matching, Voice-Leading Distance, and

- Spectral Pitch-Class Similarity: A New Approach for Testing Individual Differences in Harmony Perception. *Music & Science*, 7, Article 20592043241257654.
<https://doi.org/10.1177/20592043241257654>
- Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology*, 7(6), e1000129.
<https://doi.org/10.1371/journal.pbio.1000129>
- Escobar, J., Mussoi, B. S., & Silberer, A. B. (2020). The Effect of Musical Training and Working Memory in Adverse Listening Situations. *Ear and Hearing*, 41(2), 278–288.
<https://doi.org/10.1097/AUD.0000000000000754>
- Fisher, G. G., Chacon, M., & Chaffee, D. S. (2019). Theories of Cognitive Aging and Work. In *Work Across the Lifespan* (pp. 17–45). Elsevier. <https://doi.org/10.1016/B978-0-12-812756-8.00002-5>
- Folland, N. A., Butler, B. E., Smith, N. A., & Trainor, L. J. (2012). Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes. *The Journal of the Acoustical Society of America*, 131(1), 993–997.
<https://doi.org/10.1121/1.3651254>
- Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2014). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, 6, 347.
<https://doi.org/10.3389/fnagi.2014.00347>
- Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology*, 7, 1268. <https://doi.org/10.3389/fpsyg.2016.01268>
- Gallacher, J., Ilubaera, V., Ben-Shlomo, Y., Bayer, A., Fish, M., Babisch, W., & Elwood, P. (2012). Auditory threshold, phonologic demand, and incident dementia. *Neurology*, 79(15), 1583–1590. <https://doi.org/10.1212/WNL.0b013e31826e263d>
- Garami, L., Chow, R., Fakuade, A., Swaminathan, S., & Alain, C. (2020). Orienting Attention to Auditory and Visual Short-term Memory: The Roles of Age, Hearing Loss, and Cognitive Status. *Experimental Aging Research*, 46(1), 22–38.
<https://doi.org/10.1080/0361073X.2019.1693008>
- Gfeller, K., Mallalieu, R. M., Mansouri, A., McCormick, G., O'Connell, R. B., Spinowitz, J., & Gellinek Turner, B. (2019). Practices and Attitudes That Enhance Music Engagement of Adult Cochlear Implant Users. *Frontiers in Neuroscience*, 13, 1368.
<https://doi.org/10.3389/fnins.2019.01368>
- Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, 344, 109–124. <https://doi.org/10.1016/j.heares.2016.11.004>
- Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research*, 36(6), 1276–1285. <https://doi.org/10.1044/jshr.3606.1276>
- Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (2011). Auditory Temporal Processing and Aging: Implications for Speech Understanding of Older People. *Audiology Research*, 1(1), e4. <https://doi.org/10.4081/audiores.2011.e4>
- Golden, H. L., Clark, C. N., Nicholas, J. M., Cohen, M. H., Slattery, C. F., Paterson, R. W., Foulkes, A. J. M., Schott, J. M., Mummery, C. J., Crutch, S. J., & Warren, J. D. (2017). Music Perception in Dementia. *Journal of Alzheimer's Disease : JAD*, 55(3), 933–949.
<https://doi.org/10.3233/JAD-160359>
- Golden, H. L., Agustus, J. L., Goll, J. C., Downey, L. E., Mummery, C. J., Schott, J. M., Crutch, S. J., & Warren, J. D. (2015). Functional neuroanatomy of auditory scene analysis in

- Alzheimer's disease. *NeuroImage: Clinical*, 7, 699–708.
<https://doi.org/10.1016/j.nicl.2015.02.019>
- Gow, A. J. (2015). Intelligence and Aging. In N. A. Pachana (Ed.), *Encyclopedia of Geropsychology* (pp. 1–13). Springer Singapore. https://doi.org/10.1007/978-981-287-080-3_261-1
- Grimault, N., Michey, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *British Journal of Audiology*, 35(3), 173–182.
<https://doi.org/10.1080/00305364.2001.11745235>
- Guest, H., Dewey, R. S., Plack, C. J., Couth, S., Prendergast, G., Bakay, W., & Hall, D. A. (2018). The Noise Exposure Structured Interview (NESI): An Instrument for the Comprehensive Estimation of Lifetime Noise Exposure. *Trends in Hearing*, 22.
<https://doi.org/10.1177/2331216518803213>
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466–472.
<https://doi.org/10.1016/j.tics.2007.08.008>
- Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7(1), 3618.
<https://doi.org/10.1038/s41598-017-03586-z>
- Haywood, N. R., McAlpine, D., Vickers, D., & Roberts, B. (2024). Factors Influencing Stream Segregation Based on Interaural Phase Difference Cues. *Trends in Hearing*, 28, 23312165241293787. <https://doi.org/10.1177/23312165241293787>
- Heinrich, A., Henshaw, H., & Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers in Psychology*, 6, 782.
<https://doi.org/10.3389/fpsyg.2015.00782>
- Heinrich, A., Henshaw, H., & Ferguson, M. A. (2016). Only Behavioral But Not Self-Report Measures of Speech Perception Correlate with Cognitive Abilities. *Frontiers in Psychology*, 7, 576. <https://doi.org/10.3389/fpsyg.2016.00576>
- Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117(2), 842–849. <https://doi.org/10.1121/1.1836832>
- Henry, M. J., Herrmann, B., Kunke, D., & Obleser, J. (2017). Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. *Nature Communications*, 8, 15801. <https://doi.org/10.1038/ncomms15801>
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, 76(3), 486–502.
<https://doi.org/10.1016/j.neuron.2012.10.011>
- Hoffman, H. J., Dobie, R. A., Losonczy, K. G., Themann, C. L., & Flamme, G. A. (2017). Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years. *JAMA Otolaryngology--Head & Neck Surgery*, 143(3), 274–285. <https://doi.org/10.1001/jamaoto.2016.3527>
- Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *Journal of the Acoustical Society of America*, 123(2), 1140–1153. <https://doi.org/10.1121/1.2824018>
- Hugo, J., & Ganguli, M. (2014). Dementia and Cognitive Impairment: Epidemiology, Diagnosis, and Treatment. *Clinics in Geriatric Medicine*, 30(3), 421–442.
<https://doi.org/10.1016/j.cger.2014.04.001>
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, 18(7), 590–603. <https://doi.org/10.3766/jaaa.18.7.6>

- Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., Gates, G. A., Wilson, R. H., & Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal of the American Academy of Audiology*, 23(8), 635–666. <https://doi.org/10.3766/jaaa.23.8.5>
- Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, 123(6), 4369–4379. <https://doi.org/10.1121/1.2904826>
- Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravignani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-Cultural Work in Music Cognition: Challenges, Insights, and Recommendations. *Music perception*, 37(3), 185–195. <https://doi.org/10.1525/mp.2020.37.3.185>
- Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective & Behavioral Neuroscience*, 2(2), 121–140. <https://doi.org/10.3758/CABN.2.2.121>
- Janse, E., & Andringa, S. J. (2021). The roles of cognitive abilities and hearing acuity in older adults' recognition of words taken from fast and spectrally reduced speech. *Applied Psycholinguistics*, 42(3), 763–790. <https://doi.org/10.1017/S0142716421000047>
- Jerger, J., Brown, D., & Smith, S. (1984). Effect of peripheral hearing loss on the masking level difference. *Archives of Otolaryngology*, 110(5), 290–296. <https://doi.org/10.1001/archotol.1984.00800300018004>
- Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., & Lopez-Poveda, E. A. (2016). The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216516641055>
- Kannyo, I., & DeLong, C. M. (2011). The effect of musical training on auditory perception. In *Proceedings of Meetings on Acoustics* (p. 25002). Acoustical Society of America. <https://doi.org/10.1121/1.4733850>
- Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, 118(6), 3804–3815. <https://doi.org/10.1121/1.2109187>
- King, A. J., & Walker, K. M. M. (2020). Listening in complex acoustic scenes. *Current Opinion in Physiology*, 18, 63–72. <https://doi.org/10.1016/j.cophys.2020.09.001>
- Kimura, D. (1992). Sex differences in the brain. *Scientific American*, 267(3), 118–125. <https://doi.org/10.1038/scientificamerican0992-118>
- Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, 36(2), 217–228. <https://doi.org/10.1097/AUD.0000000000000112>
- Koehnke, J., Culotta, C. P., Hawley, M. L., & Colburn, H. S. (1995). Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear and Hearing*, 16(4), 331–353. <https://doi.org/10.1097/00003446-199508000-00003>
- Kothinti, S. R., & Elhilali, M. (2023). Are acoustics enough? Semantic effects on auditory salience in natural scenes. *Frontiers in Psychology*, 14, 1276237. <https://doi.org/10.3389/fpsyg.2023.1276237>
- Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, 1252(1), 100–107. <https://doi.org/10.1111/j.1749-6632.2012.06463.x>
- Krause, A. E., North, A. C., & Hewitt, L. Y. (2015). Music-listening in everyday life: Devices and choice. *Psychology of Music*, 43(2), 155–170. <https://doi.org/10.1177/0305735613496860>

- Kusy, A., & Châtillon, J. (2012). Real-world attenuation of custom-moulded earplugs: Results from industrial in situ F-MIRE measurements. *Applied Acoustics*, 73(6–7), 639–647. <https://doi.org/10.1016/j.apacoust.2012.02.001>
- Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(45), 14077–14085. <https://doi.org/10.1523/JNEUROSCI.2845-09.2009>
- Laffere, A., Dick, F., & Tierney, A. (2020). Effects of auditory selective attention on neural phase: Individual differences and short-term training. *NeuroImage*, 213, 116717. <https://doi.org/10.1016/j.neuroimage.2020.116717>
- Laitinen, H. M., Toppila, E. M., Olkinuora, P. S., & Kuisma, K. (2003). Sound exposure among the Finnish National Opera personnel. *Applied Occupational and Environmental Hygiene*, 18(3), 177–182. <https://doi.org/10.1080/10473220301356>
- Larsby, B., Hallgren, M., & Lyxell, B. (2008). The interference of different background noises on speech processing in elderly hearing-impaired subjects. *International Journal of Audiology*, 47(Suppl. 2), S83–S90. <https://doi.org/10.1080/14992020802301193>
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 30(22), 7604–7612. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010>
- Lentz, J. J., Humes, L. E., & Kidd, G. R. (2022). Differences in Auditory Perception Between Young and Older Adults When Controlling for Differences in Hearing Loss and Cognition. *Trends in Hearing*, 26, 23312165211066180. <https://doi.org/10.1177/23312165211066180>
- Leibold, L. J. (2012). Development of Auditory Scene Analysis and Auditory Attention. In L. Werner, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research. Human Auditory Development* (Vol. 42, pp. 137–161). Springer New York. https://doi.org/10.1007/978-1-4614-1421-6_5
- Lewald, J., & Hausmann, M. (2013). Effects of sex and age on auditory spatial scene analysis. *Hearing Research*, 299, 46–52. <https://doi.org/10.1016/j.heares.2013.02.005>
- Liberman, M. C. (2017). Noise-induced and age-related hearing loss: New perspectives and potential therapies. *F1000Research*, 6, 927. <https://doi.org/10.12688/f1000research.11310.1>
- Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, 308, 13–26. <https://doi.org/10.1016/j.heares.2013.04.009>
- Liu, M., Arseneau-Bruneau, I., Franch, M. F., Latorre, M.-E., Samuels, J., Issa, E., Payumo, A., Rahman, N., Loureiro, N., Leung, T. C. M., Nave, K. M., Handorf, K. M. von, Hoddinott, J. D., Coffey, E. B., Grahn, J., & Zatorre, R. J. (2024). *Auditory Working Memory Mediates the Relationship Between Musicianship and Auditory Stream Segregation*. <https://doi.org/10.1101/2024.10.24.619127>
- Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., & Simonsick, E. M. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, 173(4), 293–299. <https://doi.org/10.1001/jamainternmed.2013.1868>
- Lo, C. Y., Dubinsky, E., Wright-Whyte, K., Zara, M., Singh, G., & Russo, F. A. (2025). On-beat rhythm and working memory are associated with better speech-in-noise perception for older adults with hearing loss. *Quarterly Journal of Experimental Psychology (2006)*, 17470218241311204. <https://doi.org/10.1177/17470218241311204>
- Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., & Lawlor, B. A. (2018). Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and

- Dementia: A Systematic Review and Meta-analysis. *JAMA Otolaryngology-- Head & Neck Surgery*, 144(2), 115–126. <https://doi.org/10.1001/jamaoto.2017.2513>
- Lopez-Poveda, E. A. (2014). Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech. *Frontiers in Neuroscience*, 8, 348. <https://doi.org/10.3389/fnins.2014.00348>
- Lunner, T., Hellgren, J., Arlinger, S., & Elberling, C. (1997). A digital filterbank hearing aid: Three digital signal processing algorithms - user preference and performance. *Ear and Hearing*, 18(5), 373–387. <https://doi.org/10.1097/00003446-199710000-00003>
- Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, 9(1), 10404. <https://doi.org/10.1038/s41598-019-46728-1>
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18. <https://doi.org/10.1177/2331216514558271>
- Maillard, E., Joyal, M., Murray, M. M., & Tremblay, P. (2023). Are musical activities associated with enhanced speech perception in noise in adults? A systematic review and meta-analysis. *Current Research in Neurobiology*, 4, 100083. <https://doi.org/10.1016/j.crneur.2023.100083>
- Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). The Effect of Timbre and Loudness on Melody Segregation. *Music Perception*, 30(3), 259–274. <https://doi.org/10.1525/mp.2012.30.3.259>
- Marozeau, J., Innes-Brown, H., Grayden, D. B., Burkitt, A. N., & Blamey, P. J. (2010). The effect of visual cues on auditory stream segregation in musicians and non-musicians. *PloS One*, 5(6), e11297. <https://doi.org/10.1371/journal.pone.0011297>
- Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19(9), 1453–1463. <https://doi.org/10.1162/jocn.2007.19.9.1453>
- McAdams, S., & Bregman, A. S. (1979). Hearing Musical Streams. *Computer Music Journal*, 4(3), 26–43. MIT Press.
- McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E. A., Jacoby, N., Godoy, R. A., & McDermott, J. H. (2020). Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nature Communications*, 11(1), 2786. <https://doi.org/10.1038/s41467-020-16448-6>
- McKay, C. M. (2021). No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review. *Trends in Hearing*, 25, 2331216520985678. <https://doi.org/10.1177/2331216520985678>
- Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science (New York, N.Y.)*, 366(6468). <https://doi.org/10.1126/science.aax0868>
- Mehrkian, S., Moossavi, A., Gohari, N., Nazari, M. A., Bakhshi, E., & Alain, C. (2022). Long latency auditory evoked potentials and object-related negativity based on harmonicity in hearing-impaired children. *Neuroscience Research*, 178, 52–59. <https://doi.org/10.1016/j.neures.2022.01.001>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Middlebrooks, J. C. (2017). Spatial Stream Segregation. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *Springer Handbook of Auditory Research. The Auditory*

- System at the Cocktail Party* (Vol. 60, pp. 137–168). Springer International Publishing.
https://doi.org/10.1007/978-3-319-51662-2_6
- Mishra, S. K., Panda, M. R., & Raj, S. (2015). Influence of musical training on sensitivity to temporal fine structure. *International Journal of Audiology*, 54(4), 220–226.
<https://doi.org/10.3109/14992027.2014.969411>
- Moore, B. C. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *British Journal of Audiology*, 19(3), 189–201.
<https://doi.org/10.3109/03005368509078973>
- Moore, B. C. J. (2021). Effects of hearing loss and age on the binaural processing of temporal envelope and temporal fine structure information. *Hearing Research*, 402, 107991.
<https://doi.org/10.1016/j.heares.2020.107991>
- Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing AIDS and paired comparisons. *Ear and Hearing*, 32(5), 556–568.
<https://doi.org/10.1097/AUD.0b013e31820b5f4c>
- Moore, B. C. J., Glasberg, B. R., & Hopkins, K. (2006). Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure. *Hearing Research*, 222(1–2), <https://doi.org/10.1016/j.heares.2006.08.007>
- Moore, B. C. J., & Gockel, H. (2002). Factors Influencing Sequential Stream Segregation. *Acta Acustica United with Acustica*(88), Article 3.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2). <https://doi.org/10.1371/journal.pone.0089642>
- Møller, A. R. (2006). Neural plasticity in tinnitus. *Progress in Brain Research*, 157, 365–372.
[https://doi.org/10.1016/S0079-6123\(06\)57022-0](https://doi.org/10.1016/S0079-6123(06)57022-0)
- Nesselroade, J. R., & Ford, D. H. (1987). Methodological considerations in modeling living systems. In M. E. Ford & D. H. Ford (Eds.), *Humans as self-constructing living systems: Putting the framework to work* (pp. 47–79). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nitsan, G., Banai, K., & Ben-David, B. M. (2022). One Size Does Not Fit All: Examining the Effects of Working Memory Capacity on Spoken Word Recognition in Older Adults Using Eye Tracking. *Frontiers in Psychology*, 13, 841466.
<https://doi.org/10.3389/fpsyg.2022.841466>
- Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife*, 5.
<https://doi.org/10.7554/eLife.16747>
- Olsen, W. O., & Noffsinger, D. (1976). Masking level differences for cochlear and brain stem lesions. *Annals of Otology, Rhinology & Laryngology*, 85(6 Pt. 1), 820–825.
<https://doi.org/10.1177/000348947608500605>
- Ouda, L., Profant, O., & Syka, J. (2015). Age-related changes in the central auditory system. *Cell and Tissue Research*, 361(1), 337–358. <https://doi.org/10.1007/s00441-014-2107-2>
- Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, 12(4), 316–331.
<https://doi.org/10.1177/1084713808325881>
- O'Connor, K. (1990). Towards a process paradigm in psychophysiology. *International Journal of Psychophysiology*, 9(2), 209–223.
- O'Brien, I., Driscoll, T., & Ackermann, B. (2013). Sound exposure of professional orchestral musicians during solitary practice. *The Journal of the Acoustical Society of America*, 134(4), 2748–2754. <https://doi.org/10.1121/1.4820900>
- Paredes-Gallardo, A., Innes-Brown, H., Madsen, S. M. K., Dau, T., & Marozeau, J. (2018). Auditory Stream Segregation and Selective Attention for Cochlear Implant Listeners:

- Evidence From Behavioral Measures and Event-Related Potentials. *Frontiers in Neuroscience*, 12, 581. <https://doi.org/10.3389/fnins.2018.00581>
- Park, D. C., & Bischof, G. N. (2013). The aging mind: Neuroplasticity in response to cognitive training. *Dialogues in Clinical Neuroscience*, 15(1), 109–119. <https://doi.org/10.31887/DCNS.2013.15.1/dpark>
- Pausch, V., Müllensiefen, D., & Kopiez, R. (2022). Musikalischer g-Faktor oder multiple Faktoren? Struktur und Leistungskennwerte der musikalischen Hörfähigkeit von Jugendlichen. *Jahrbuch Musikpsychologie*, 30, Article e89. <https://doi.org/10.5964/jbdgm.89>
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, 50(1), 302–313. <https://doi.org/10.1016/j.neuroimage.2009.12.019>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
- Phillips, S. L., & Mace, S. (2008). Sound level measurements in music practice rooms. *Music Performance Research*, 2(1993), 36–47.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37 Suppl 1, 5S-27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, Cognition, and Healthy Aging: Social and Public Health Implications of the Links between Age-Related Declines in Hearing and Cognition. *Seminars in Hearing*, 36(3), 122–139. <https://doi.org/10.1055/s-0035-1555116>
- Popelka, G. R., Moore, B. C. J., Fay, R. R., & Popper, A. N. (2016). *Hearing Aids* (Vol. 56). Springer International Publishing. <https://doi.org/10.1007/978-3-319-33036-5>
- Pressnitzer, D., Suied, C., & Shamma, S. A. (2011). Auditory scene analysis: The sweet music of ambiguity. *Frontiers in Human Neuroscience*, 5, 158. <https://doi.org/10.3389/fnhum.2011.00158>
- Profant, O., Jilek, M., Bures, Z., Vencovsky, V., Kucharova, D., Svobodova, V., Korynta, J., & Syka, J. (2019). Functional Age-Related Changes Within the Human Auditory System Studied by Audiometric Examination. *Frontiers in Aging Neuroscience*, 11, 26. <https://doi.org/10.3389/fnagi.2019.00026>
- Ragert, M., Fairhurst, M. T., & Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PloS One*, 9(1), e84085. <https://doi.org/10.1371/journal.pone.0084085>
- Rajasingam, S. L., Summers, R. J., & Roberts, B. (2021). The dynamics of auditory stream segregation: Effects of sudden changes in frequency, level, or modulation. *The Journal of the Acoustical Society of America*, 149(6), 3769. <https://doi.org/10.1121/10.0005049>
- Regev, J., Zaar, J., Relaño-Iborra, H., & Dau, T. (2023). Age-related reduction of amplitude modulation frequency selectivity. *The Journal of the Acoustical Society of America*, 153(4), 2298. <https://doi.org/10.1121/10.0017835>
- Relaño-Iborra, H., & Dau, T. (2022). Speech intelligibility prediction based on modulation frequency-selective processing. *Hearing Research*, 426, 108610. <https://doi.org/10.1016/j.heares.2022.108610>

- Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review*, 24(3), 355–370.
<https://doi.org/10.1007/s11065-014-9270-9>
- Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, 51(1), 160–172.
[https://doi.org/10.1044/1092-4388\(2008/012\)](https://doi.org/10.1044/1092-4388(2008/012))
- Rodrigues, M. A., Freitas, M. A., Neves, M. P., & Silva, M. V. (2014). Evaluation of the noise exposure of symphonic orchestra musicians. *Noise & health*, 16(68), 40–46.
<https://doi.org/10.4103/1463-1741.127854>
- Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4(4), 525–553. <https://doi.org/10.1111/j.1756-8765.2012.01223.x>
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278), 367–373. <https://doi.org/10.1098/rstb.1992.0070>
- Rossi-Katz, J., & Arehart, K. H. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research*, 52(2), 435–453.
[https://doi.org/10.1044/1092-4388\(2008/07-0229\)](https://doi.org/10.1044/1092-4388(2008/07-0229))
- Rudner, M. (2016). Cognitive Spare Capacity as an Index of Listening Effort. *Ear and Hearing*, 37 Suppl 1, 69S-76S. <https://doi.org/10.1097/AUD.0000000000000302>
- Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences of the United States of America*, 108(37), 15516–15521. <https://doi.org/10.1073/pnas.1108912108>
- Russo, F. (2019). Pitch combinations and grouping. In D. J. Levitin & P. J. Rentfrow (Eds.), *Foundations in music psychology: Theory and research* (pp. 121–148). MIT Press. ISBN: 9780262039277.
- Russo, F. A., Behar, A., Chasin, M., & Mosher, S. (2013). Noise exposure and hearing loss in classical orchestra musicians. *International Journal of Industrial Ergonomics*, 43(6), 474–478. <https://doi.org/10.1016/j.ergon.2012.11.001>
- Rönnerberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, 47 Suppl 2, S99-105. <https://doi.org/10.1080/14992020802301167>
- Safety Research Corporation of America (2020). *Noise and hearing protection fact sheet*. Retrieved July 5, 2023, from <https://www.srca.net/Blog/NoiseAndHearingProtectionFactSheet.aspx#:~:text=Why%20Can%E2%80%99t%20I,as%20adequate%20protection>
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–428. <https://doi.org/10.1037/0033-295X.103.3.403>
- Sallat, S., & Jentschke, S. (2015). Music Perception Influences Language Acquisition: Melodic and Rhythmic-Melodic Perception in Children with Specific Language Impairment. *Behavioural Neurology*, 2015, 606470. <https://doi.org/10.1155/2015/606470>
- Sandell, G. J., & Darwin, C. J. (1996). Recognition of concurrently-sounding musical instruments with different fundamental frequencies. *The Journal of the Acoustical Society of America*, 100, 2683. <https://doi.org/10.1121/1.416079>
- Sauvé, S. A., Marozeau, J., & Rich Zendel, B. (2022). The effects of aging and musicianship on the use of auditory streaming cues. *PloS One*, 17(9), e0274631.
<https://doi.org/10.1371/journal.pone.0274631>

- Schwartz, G. E. (1986). Emotion and psychophysiological organization: A systems approach. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 354–377). New York, NY: Guilford Press.
- Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, 75, 87–128. <https://doi.org/10.1146/annurev-psych-032323-051354>
- Schmidt, J. H., Pedersen, E. R., Juhl, P. M., Christensen-Dalsgaard, J., Andersen, T. D., Poulsen, T., & Bælum, J. (2011). *Sound exposure of symphony orchestra musicians. Annals of Occupational Hygiene*, 55(8), 893–905. <https://doi.org/10.1093/annhyg/mer055>
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18(7), 559–572. <https://doi.org/10.3766/jaaa.18.7.6>
- Schurig, E., Hake, R., Birke, M., Derks, D., Siedenburg, K., & Kreutz, G. (2024). Hearing health literacy among professional and amateur musicians. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-79875-1>
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, 51(1), 450–461. <https://doi.org/10.1016/j.neuroimage.2010.02.023>
- Searchfield, G. D., Morrison-Low, J., & Wise, K. (2007). Object identification and attention training for treating tinnitus. *Progress in Brain Research*, 166, 441–460. [https://doi.org/10.1016/S0079-6123\(07\)66043-9](https://doi.org/10.1016/S0079-6123(07)66043-9)
- Searchfield, G. D., Kobayashi, K., & Sanders, M. (2012). An adaptation level theory of tinnitus audibility. *Frontiers in Systems Neuroscience*, 6, 46. <https://doi.org/10.3389/fnsys.2012.00046>
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4), 283–299. <https://doi.org/10.1177/1084713808325306>
- Shinn-Cunningham, B. (2017). Cortical and Sensory Causes of Individual Differences in Selective Attention Ability Among Listeners With Normal Hearing Thresholds. *Journal of Speech, Language, and Hearing Research : JSLHR*, 60(10), 2976–2988. https://doi.org/10.1044/2017_JSLHR-H-17-0080
- Siedenburg, K., Goldmann, K., & van de Par, S. (2021). Tracking Musical Voices in Bach's The Art of the Fugue: Timbral Heterogeneity Differentially Affects Younger Normal-Hearing Listeners and Older Hearing-Aid Users. *Frontiers in Psychology*, 12, 608684. <https://doi.org/10.3389/fpsyg.2021.608684>
- Siedenburg, K., Röttges, S., Wagener, K. C., & Hohmann, V. (2020). Can You Hear Out the Melody? Testing Musical Scene Perception in Young Normal-Hearing and Older Hearing-Impaired Listeners. *Trends in Hearing*, 24, 2331216520945826. <https://doi.org/10.1177/2331216520945826>
- Silas, S., Müllensiefen, D., Gelding, R., Frieler, K., & Harrison, P. M. C. (2022). The Associations Between Music Training, Musical Working Memory, and Visuospatial Working Memory. *Music Perception*, 39(4), 401–420. <https://doi.org/10.1525/mp.2022.39.4.401>
- Slater, K. D., & Marozeau, J. (2016). The effect of tactile cues on auditory stream segregation ability of musicians and nonmusicians. *Psychomusicology: Music, Mind, and Brain*, 26(2), 162–166. <https://doi.org/10.1037/pmu0000143>
- Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18(1), 1–13. <https://doi.org/10.1162/089892906775250021>
- Smith, N. A., Folland, N. A., Martinez, D. M., & Trainor, L. J. (2017). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate

- auditory and visual object. *Cognition*, 164, 1–7.
<https://doi.org/10.1016/j.cognition.2017.01.016>
- Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear and Hearing*, 32(1), 75–83.
<https://doi.org/10.1097/AUD.0b013e3181eccfe9>
- Souza, P., Hoover, E., Blackburn, M., & Gallun, F. (2018). The characteristics of adults with severe hearing loss. *Journal of the American Academy of Audiology*, 29(8), 764–779.
<https://doi.org/10.3766/jaaa.17050>
- Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, 4(4), 653–667. <https://doi.org/10.1111/j.1756-8765.2012.01215.x>
- Susini, P., Jiaouan, S. J., Brunet, E., Houix, O., & Ponsot, E. (2020). Auditory local-global temporal processing: Evidence for perceptual reorganization with musical expertise. *Scientific Reports*, 10(1), 16390. <https://doi.org/10.1038/s41598-020-72423-7>
- Susini, P., Wenzel, N., Houix, O., & Ponsot, E. (2023). Psychophysical characterization of auditory temporal and frequency streaming capacities for listeners with different levels of musical expertise. *JASA Express Letters*, 3(8), Article 084402.
<https://doi.org/10.1121/10.0020546>
- Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *Journal of Speech, Language, and Hearing Research : JSLHR*, 60(10), 2989–3000.
https://doi.org/10.1044/2017_JSLHR-H-17-0041
- Sussman, E., Steinschneider, M., Lee, W., & Lawson, K. (2015). Auditory scene analysis in school-aged children with developmental language disorders. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 95(2), 113–124. <https://doi.org/10.1016/j.ijpsycho.2014.02.002>
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, 5, 11628. <https://doi.org/10.1038/srep11628>
- Syka, J. (2020). Age-Related Changes in the Auditory Brainstem and Inferior Colliculus. In K. S. Helfer, E. L. Bartlett, A. N. Popper, & R. R. Fay (Eds.), *Springer Handbook of Auditory Research. Aging and Hearing* (Vol. 72, pp. 67–96). Springer International Publishing.
https://doi.org/10.1007/978-3-030-49367-7_4
- Szalárdy, O., Bendixen, A., Böhm, T. M., Davies, L. A., Denham, S. L., & Winkler, I. (2014). The effects of rhythm and melody on auditory stream segregation. *The Journal of the Acoustical Society of America*, 135(3), 1392–1405. <https://doi.org/10.1121/1.4865196>
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, 12(10), e0186773.
<https://doi.org/10.1371/journal.pone.0186773>
- Thomson, R. S., Auduong, P., Miller, A. T., & Gurgel, R. K. (2017). Hearing loss as a risk factor for dementia: A systematic review. *Laryngoscope Investigative Otolaryngology*, 2(2), 69–79. <https://doi.org/10.1002/lio2.65>
- Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology. Human Perception and Performance*, 37(4), 1253–1262.
<https://doi.org/10.1037/a0021925>
- Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1664), 20140089. <https://doi.org/10.1098/rstb.2014.0089>

- Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological Bulletin*, 145(3), 273–301.
<https://doi.org/10.1037/bul0000179>
- Uys, M., & Latzel, M. (2015). Long-Term Effects of Non-Linear Frequency Compression on Performance of Music and Speech Perception. *Journal of Communication Disorders, Deaf Studies & Hearing Aids*, 03(03). <https://doi.org/10.4172/2375-4427.1000139>
- Uys, M., Latzel, M., & van Dyk, M. (2016). The Performance of Music and Speech with Different Frequency Lowering Devices. *Journal of Communication Disorders, Deaf Studies & Hearing Aids*, 4(3). <https://doi.org/10.4172/2375-4427.1000163>
- Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *The South African Journal of Communication Disorders = Die Suid-Afrikaanse Tydskrif Vir Kommunikasieafwykings*, 58, 19–47.
<https://doi.org/10.4102/sajcd.v58i1.38>
- van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Doctoral dissertation). Technische Hogeschool Eindhoven. Published Eindhoven: Druk van Voorschoten.
- Vinay, Sandhya, & Moore, B. C. J. (2020). Effect of age, test frequency and level on thresholds for the TEN(HL) test for people with normal hearing. *International Journal of Audiology*, 59(12), 915–920. <https://doi.org/10.1080/14992027.2020.1783584>
- Walton, J. P. (2010). Timing is everything: Temporal processing deficits in the aged auditory brainstem. *Hearing Research*, 264(1-2), 63–69.
<https://doi.org/10.1016/j.heares.2010.03.002>
- Wenhart, T., Hwang, Y.-Y., & Altenmüller, E. (2019). Enhanced auditory disembedding in an interleaved melody recognition test is associated with absolute pitch ability. *Scientific Reports*, 9(1), 7838. <https://doi.org/10.1038/s41598-019-44297-x>
- Whittom, A., Chauvette, L., Bégin, A., Blanchette, I., Tremblay, P., & Sharp, A. (2024). Music Perception in Older Adults With Hearing Loss: Protective Effect of Musical Experience. *Ear and Hearing*. Advance online publication.
<https://doi.org/10.1097/AUD.0000000000001615>
- Williamson, J. M. (2018). Individual Differences. In *Teaching to Individual Differences in Science and Engineering Librarianship* (pp. 1–10). Elsevier.
<https://doi.org/10.1016/B978-0-08-101881-1.00001-7>
- Windle, R., Dillon, H., & Heinrich, A. (2023). A review of auditory processing and cognitive change during normal ageing, and the implications for setting hearing aids for older adults. *Frontiers in Neurology*, 14, 1122420.
<https://doi.org/10.3389/fneur.2023.1122420>
- Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, 7(3), 175–182.
- Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., Huottilainen, M., Näätänen, R., & Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20), 11812–11815. <https://doi.org/10.1073/pnas.2031891100>
- Woodcock, R. W. (1990). Theoretical Foundations of the Wj-R Measures of Cognitive Ability. *Journal of Psychoeducational Assessment*, 8(3), 231–258.
<https://doi.org/10.1177/073428299000800303>
- Woods, K. J. P., & McDermott, J. H. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of Sciences of the United States of America*, 115(14), E3313–E3322. <https://doi.org/10.1073/pnas.1801614115>

- Yurgil, K. A., Velasquez, M. A., Winston, J. L., Reichman, N. B., & Colombo, P. J. (2020). Music Training, Working Memory, and Neural Oscillations: A Review. *Frontiers in Psychology*, 11, 266. <https://doi.org/10.3389/fpsyg.2020.00266>
- Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl 2(Suppl 2), 10430–10437. <https://doi.org/10.1073/pnas.1301228110>
- Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S. M. G., & Houtgast, T. (2009). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognition system. *Ear and Hearing*, 30(2), 262–272. <https://doi.org/10.1097/AUD.0b013e3181987063>
- Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8), 1488–1498. <https://doi.org/10.1162/jocn.2009.21140>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, 27(2), 410–417. <https://doi.org/10.1037/a0024816>
- Zhao, J. (2024). Memory, attention and creativity as cognitive processes in musical performance: A case study of students and professionals among non-musicians and musicians. *Attention, Perception & Psychophysics*, 86(6), 2042–2052. <https://doi.org/10.3758/s13414-024-02944-0>
- Zhang, J. D., & Schubert, E. (2019). A Single Item Measure for Identifying Musician and Nonmusician Categories Based on Measures of Musical Sophistication. *Music Perception*, 36(5), 457–467. <https://doi.org/10.1525/mp.2019.36.5.457>
- Zündorf, I. C., Karnath, H.-O., & Lewald, J. (2011). Male advantage in sound localization at cocktail parties. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 47(6), 741–749. <https://doi.org/10.1016/j.cortex.2010.08.002>

LIST OF PUBLICATIONS

The following publications are related to the work in this thesis^e.

Published peer-reviewed journal paper(s)

Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburg, K. (2024). A Survey on Hearing Health of Musicians in Professional and Amateur Orchestras. *Trends in Hearing*, 28, 1–20. <https://doi.org/10.1177/23312165241293762>

Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., & Siedenburg, K. (2024). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods*(56), 5456–5481. <https://doi.org/10.3758/s13428-023-02279-y>

Accepted for publication in peer-reviewed journal paper(s)

Hake, R., Bürgel, M., Lesimple, C., Vormann, M., Wagener, K. C., Kühnel, V., & Siedenburg, K. (2025). Music perception with hearing aids: Compression differentially affects musical scene analysis and musical sound quality. *Trends in Hearing* (accepted on 20. February 2025).

Submitted to peer-reviewed journal paper(s)

Hake, R., Müllensiefen, D., & Siedenburg, K. (2025). Auditory scene analysis in music and speech: Individual differences in listening abilities. Manuscript submitted for publication in *Scientific Reports*.

Other publications in peer-reviewed journals, that are indirectly related to this thesis.

Schurig, E., **Hake**, R., Birke, M., Derks, D., Siedenburg, K., & Kreutz, G. (2024). Hearing health literacy among professional and amateur musicians. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-79875-1>

Bürgel, M., Mencke, I., Benjamin, A., Dechert, M., Derks, D., Gerdes, K., **Hake**, R., Jacobsen, S., & Siedenburg, K. (2023). Unifying concert research and science outreach. *Musicae Scientiae*, Article 10298649231182078. <https://doi.org/10.1177/10298649231182078>

^e A declaration of the author's individual contribution to thesis related publications is provided on the disclaimer page at the beginning of each chapter.

AUTHOR'S DECLARATION

I, the undersigned Robin Hake (Robin Hake) hereby declare that I am the sole author of this thesis. To the best of my knowledge, this thesis contains no material previously published by any other person, except where due acknowledgement has been made. This thesis contains no material which has been accepted as part of the requirements of any other academic degree or non-degree program, in English or in any other language, unless explicitly acknowledged. I further declare that I have written this dissertation independently and that all sources and aids used have been fully and accurately disclosed. Large language models, including AI-based tools, were used solely for editorial purposes such as language editing, rewording, and formatting assistance. The principles of good scientific practice of the Carl von Ossietzky University of Oldenburg have been fully considered. Parts of this dissertation have been previously published. A complete list of these publications is provided in the appendix. I confirm that no commercial intermediary or consultancy services (i.e., doctoral advisory services) were used in connection with this dissertation. This is a true copy of the thesis.

Date: 24.04.2025
Name (printed letters): Robin Hake
Signature: 