CARL VON OSSIETZKY universität OLDENBURG

MUSIC MIXING PREFERENCES AND SCENE ANALYSIS ABILITIES OF HEARING-IMPAIRED LISTENERS

Von der Fakultät für Medizin und Gesundheitswissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels eines

Doktor der Naturwissenschaften Dr. rer. nat.

eingereichte Dissertation

von Aravindan Joseph Benjamin

gebören am 04.10.1985 in Colombo (Sri Lanka)

Erstbetreuer : Prof. Dr. Kai Siedenburg

November 13, 2025

Acknowledgments

I would like to express my heartfelt gratitude to all those who supported me throughout my PhD journey and contributed to our shared understanding of human perception. Above all, I am profoundly grateful to my supervisor, Prof. Kai Siedenburg, whose exceptional scientific insight, generous mentorship, and unwavering support have been the cornerstone of my research. His guidance sharpened my scientific thinking and opened doors to valuable opportunities and collaborations that deeply enriched my PhD experience.

I wish to express my deepest gratitude in advance to Prof. Steven van de Par and Prof. Lorenzo Picinali for kindly agreeing to serve as members of my thesis committee. I am truly thankful for their generosity in sharing their time, insight, and expertise, and for their willingness to provide thoughtful guidance and constructive feedback on my work. Their support is invaluable, and I deeply appreciate their contribution to this important stage of my doctoral journey. Moreover, I would like to thank Prof. Brian Moore for his invaluable feedback in one of my publications.

I am also grateful for the opportunity to have known and worked with my colleagues at the Music Perception and Processing lab. Their collaboration, encouragement, and camaraderie have been a constant source of learning and joy.

My deepest thanks go to my mother for her extraordinary love, resilience, and unwavering support, which have shaped the person I am today. I also wish to honor my late father, whose guidance and example continue to inspire me, and I dedicate this work to him with all my gratitude.

Special thanks are due to the Freigeist Fellowship of the Volkswagen Foundation and the SFB project of the Deutsche Forschungsgemeinschaft for generously funding my research and making this work possible.

Oldenburg, November 13, 2025

Aravindan Joseph Benjamin

Abstract

The practice of multitrack mixing has become central to contemporary music production with its origins dating back to developments in the 1960s. Despite the increasing prevalence of hearing loss, particularly among older adults, mixing conventions have largely been optimized for the perceptual needs of normal-hearing (NH) listeners. Commercially available music mixes may be altered to improve music enjoyment for hearing-impaired (HI) listeners, nevertheless. In order to assess the impact of manipulating music mixes on accessibility for HI listeners with mostly moderate hearing loss or higher, three studies were conducted in this dissertation.

Across the studies, a progressive relationship was observed between pure-tone audiometry thresholds and perceptual outcomes. According to results from remixing tasks in Study 1, elevated thresholds were associated with preferences for enhanced lead-vocal levels relative to the accompaniment and more exaggerated spectral contrast adjustments through higher % EQ-transform preferences. The EQ-transform exaggerates or minimizes the track-specific power spectrum in mixes by linearly extrapolating it with respect to a smooth, reference spectrum. This reference is an ensemble average power spectrum taken over a number of different tracks from the open-source Medley database. With 100 % referring to the spectral contrast of a single track in the original mix, a 200 % EQ-transform essentially doubles the power level difference between this original and reference, exaggerating its contrast as a result. HI listeners also tended to favor high-frequency amplification (>1 kHz) when unaided. However, with bilateral hearing-aid use, preferences for lead-vocal

levels and contrast adjustments were reduced, while that for greater amplification shifted to frequencies below 1 kHz. Moreover, variability in EQ-transform and high-frequency amplification preferences increased as hearing thresholds worsened in unaided HI listeners. Among aided listeners on the other hand, these observations were mitigated.

As observed in the top-down selective attention task in Study 2, musical scene analysis (MSA) abilities declined steadily with increasing hearing loss, reaching mere chance-level performance at hearing thresholds associated with moderately severe hearing loss. Target instrument category emerged as a dominant predictor of MSA performance with lead vocals being most salient but disproportionately affected by hearing loss. On the other hand, detection of bass guitar targets was poorest overall yet was least affected by hearing loss. Despite no observable effect of EQ-transform, performance of HI listeners saw an improvement for musical scenes in which mixes with sparser power spectra and lower roll-off points compared to the targets were presented. Variance in MSA performance increased as hearing thresholds worsened, suggesting that individuals with greater hearing loss not only performed uniformly worse but also demonstrated more varied selective listening abilities. In the audio quality appraisal task in Study 3, NH listeners were more critical of spectral contrast adjustments than HI listeners. Notably, the latter became less critical in their quality judgments for similar changes in spectral shape as hearing thresholds worsened. Furthermore, unlike in NH, observations for HI suggested an associative relationship between MSA and perception of quality.

The overall findings draw attention to the challenges involved in creating music mixes for individuals with moderate or greater hearing loss. As such, this work emphasizes the need for specialized, multifaceted mixing strategies tailored to the perceptual and cognitive profiles of such listeners, unlike the more conventional *"Best Practices"* typically followed by mixing engineers for mainstream audiences.

Zusammenfassung

Die Praxis des Multitrack-Mixings hat sich seit ihren Ursprüngen in den 1960er-Jahren zu einem zentralen Bestandteil der zeitgenössischen Musikproduktion entwickelt. Trotz der zunehmenden Verbreitung von Hörverlusten, insbesondere bei älteren Erwachsenen, sind gängige Mischkonventionen weitgehend auf die Wahrnehmungsbedürfnisse normalhörender (NH) Hörerinnen und Hörer zugeschnitten. Dennoch können kommerziell verfügbare Musikmischungen modifiziert werden, um das Musikerleben für Menschen mit Hörbeeinträchtigungen (HI) zu verbessern. Um den Einfluss solcher Anpassungen auf die Barrierefreiheit für HI-Hörerinnen und -Hörer mit überwiegend moderatem oder stärkerem Hörverlust zu untersuchen, wurden im Rahmen dieser Dissertation drei Studien durchgeführt.

Über die Studien hinweg zeigte sich ein progressiver Zusammenhang zwischen den Reinton-Audiometrieschwellen und den wahrnehmungsbezogenen Ergebnissen. Die Ergebnisse der Remix-Aufgaben in Studie 1 verdeutlichten, dass erhöhte Hörschwellen mit Präferenzen für verstärkte Lead-Gesangspegel im Verhältnis zur Begleitung sowie mit stärker ausgeprägten spektralen Kontrastanpassungen in Form erhöhter % EQ-Transform-Präferenzen verbunden waren. Der EQ-Transform verstärkt oder reduziert das track-spezifische Leistungsspektrum in Mischungen, indem er dieses linear in Bezug auf ein geglättetes Referenzspektrum extrapoliert. Dieses Referenzspektrum stellt ein Ensemble-Mittelwert des Leistungsspektrums über eine Vielzahl verschiedener Titel aus der Open-Source-Medley-Datenbank dar. Während 100 % dem spektralen Kontrast eines einzelnen Tracks in der Originalmischung

entsprechen, verdoppelt ein 200 %-EQ-Transform im Wesentlichen die Leistungspegeldifferenz zwischen Original und Referenz und betont damit den Kontrast entsprechend. HI-Personen neigten zudem im unausgeglichenen Zustand zu einer Präferenz für Hochfrequenzverstärkung (>1 kHz). Mit beidseitiger Hörgeräteversorgung reduzierten sich jedoch die Präferenzen für Lead-Gesangspegel und Kontrastanpassungen, während sich die Präferenz für stärkere Verstärkung auf Frequenzen unterhalb von 1 kHz verlagerte. Darüber hinaus nahm die Variabilität in den EQ-Transform- und Hochfrequenzverstärkungspräferenzen bei unbehandelten HI-Personen mit zunehmendem Hörverlust zu; bei versorgten Zuhörern war dieser Effekt hingegen abgeschwächt.

Wie in der Top-Down-Selektionsaufgabe zur Musikalischen Szenenanalyse (MSA) in Studie 2 gezeigt, nahm die MSA-Leistungsfähigkeit mit steigendem Hörverlust kontinuierlich ab und erreichte bei Schwellenwerten, die einem moderat ausgeprägten Hörverlust entsprechen, lediglich Zufallsniveau. Die Zielinstrumentenkategorie erwies sich als dominanter Prädiktor der MSA-Leistung: Lead-Gesang war am salientesten, jedoch überproportional stark vom Hörverlust betroffen. Die Detektion von Bassgitarren-Zielen war hingegen insgesamt am schwächsten, wurde jedoch vergleichsweise geringfügig durch den Hörverlust beeinflusst. Trotz des Fehlens eines nachweisbaren Effekts des EQ-Transforms zeigte sich bei HI-Personen eine Leistungsverbesserung in musikalischen Szenen, in denen Mischungen mit dünneren Leistungsspektren und niedrigeren Abfallpunkten im Vergleich zu den Zielsignalen präsentiert wurden. Mit zunehmendem Hörverlust stieg zudem die Varianz der MSA-Leistung, was darauf hindeutet, dass Personen mit stärkerem Hörverlust nicht nur insgesamt schlechter abschnitten, sondern auch größere interindividuelle Unterschiede in selektiven Hörfähigkeiten aufwiesen.

In der Aufgabe zur Beurteilung der Audioqualität in Studie 3 reagierten NH-Personen kritischer auf spektrale Kontrastanpassungen als HI-Personen. Bemerkenswerterweise wurden letztere mit zunehmendem Hörverlust in ihren Qualitätsurteilen

gegenüber ähnlichen spektralen Veränderungen weniger kritisch. Anders als bei NH-Personen deuteten die Beobachtungen bei HI-Personen zudem auf eine assoziative Beziehung zwischen MSA-Leistung und Qualitätswahrnehmung hin.

Die Gesamtergebnisse machen auf die Herausforderungen bei der Erstellung von Musikmischungen für Personen mit mittelgradigem oder stärkerem Hörverlust aufmerksam. Dementsprechend betont diese Arbeit die Notwendigkeit spezialisierter, multifaktorieller Mischstrategien, die auf die perzeptuellen und kognitiven Profile dieser Hörer zugeschnitten sind, im Gegensatz zu den eher konventionellen *"Best Practices"*, denen Mischtoningenieure typischerweise für das Massenpublikum folgen.

Contents

List of figures

1. Introduction

Music, a cultural universal, has evolved from ancient traditions to the contemporary global stage. Among its many forms, popular music plays a prominent role in reflecting the trends and values of the present era, spanning genres enjoyed by listeners worldwide (Shuker, 2012). *'Pop'* music, a genre of popular music, has been especially central in garnering universal attention, largely by virtue of the mainstream media such as television and radio (Boyle et al., 1981). Multi-track mixing which is an essential part of audio production, is instrumental in optimizing the sonic attributes and enriching the listening experience of such musical genres (Owsinski, 2014). This practice entails the layering of individual sound elements on separate channels or *tracks*, which then undergo spectro-temporal manipulations by a professional mixing engineer to create a composite and cohesive music *'mix'*. The advent of multi-track mixing was a turning point for music production in the midst of the 20th century. Musicologist, Shara Rambarran states:

> *"As multi-track recording advanced and expanded, it revolutionized the creation and production of music, particularly from the 1960s onward, with the works of the Beatles and Rolling Stones being prime examples, but, of course, there are many, many others as well."* (Rambarran, 2021, p. 15)

Nevertheless, multi-track mixes are created primarily for individuals without a diagnosed hearing impairment, commonly referred to as normal-hearing (NH) lis-

teners. Despite the steady rise in hearing loss, especially in older individuals (Golovanova et al., 2019), music mixes are rarely adapted for such listeners. Therefore, this dissertation will examine how modified music mixes affect the subjective preference, perceived quality, and auditory scene analysis abilities among listeners with largely moderate or greater hearing loss.

1.1 Music perception with hearing aids

Hearing impairments present themselves in varying degrees of severity, ranging from mild to profound hearing loss. On a global scale, World Health Organization (2021) reports that a staggering 1.5 billion individuals experience some degree of hearing impairment, with the majority of them (approximately 1.2 billion) having mild hearing loss. Among the 430 million individuals with higher degrees of hearing loss (moderate or higher), those with moderate hearing loss make up more than 60%.

As a non-invasive intervention, individuals with mild-moderate hearing loss are often prescribed hearing aids (HAs) (Ferguson et al., 2017) which facilitate improved audibility by amplifying sounds (Hoppe and Hesse, 2017). On the other hand, invasive cochlear-implants (CI) are generally reserved for those with profound levels of hearing loss who do not benefit from the frequency dependent amplification of HAs (Zheng et al., 2022). A number of studies substantiate the benefits of HAs on speech perception (Suatbayeva et al., 2024; Abdi, 2020; Cox et al., 2014), thus improving the quality of life among hearing-impaired (HI) individuals (Garcia et al., 2016).

Nevertheless, the implications of the use of HAs on music perception remain unclear. A study by Leek et al. (2008) showed that, concerns pertaining to music enjoyment via HAs saw a drop greater than 79% in elderly listeners over a period of 20 years, which can be attributed to improvements in wide-dynamic-range compres-

sion technologies. This beneficial trend notwithstanding, there were still existing problems, where around 30% of the listeners reported that hearing-aid use had a detrimental effect on their enjoyment of music. With the aid of over 500 hearing-aid users, Madsen and Moore (2014) showed that a vast majority of them found that HAs were beneficial when listening to both reproduced and live music. Despite this observation, there were reports of acoustical feedback, distortions brought on by clipping, a compromised frequency response, and improper gain that depreciated enjoyment. Looi et al. (2019) showed that as listeners suffered from higher levels of hearing loss, their enjoyment of music through HAs diminished, so much so that the HAs themselves were reported to compromise musical melody. Although no significant differences in pleasantness ratings for listening to music through HAs were shown between listeners with mild and moderate hearing loss, those with severe hearing loss elicited much lower ratings than the former group. Greasley et al. (2020) showed that the lack of audiological focus on music perception with HAs, may contribute to the challenges faced by hearing-aid users in listening to music. Furthermore, such shortcomings were implicated in a poorer quality of life among hearing-aid users which sometimes resulted in their detachment from music altogether.

On the other hand, Chasin and Russo (2004) argue that despite dedicated signal processing settings for music signals or *'music programs'*, HAs do not confer desirable musical fidelity unless their electroacoustical parameters are also specifically calibrated for music. To that end, Sandgren and Alexander (2023) evaluated how NH rated the quality of music excerpts under simulated hearing-aid conditions. They found that improvements to quality by virtue of activating the music program was observed in only some hearing-aid brands simulated. Similar findings were reported by Vaisberg et al. (2017) with HI listeners, where music programs improved sound quality in only two out of the 5 HAs tested. More recently, Lesimple et al. (2024) demonstrated that the manner in which HAs process music signals depends

heavily on the frequency-domain properties and dynamic range of the signals, irrespective of the genre.

1.2 Music pre-processing and intelligent mixing

Owing to the ambiguous nature of hearing-aid use on music perception, research on pre-processing methods for improving music listening and appreciation with HAs is paramount. However, most of the studies on music signal pre-processing have been conducted within the purview of *'intelligent'* mixing. This novel field of automating multi-track mixing practices, is aimed at facilitating computerized means which circumvent the involvement of a trained mixing engineer. Intelligent mixing pioneer Joshua Reiss explains:

> *"By 'intelligent', we mean that these tools are expert systems that perceive, reason, learn and act intelligently. This implies that they must analyze the signals upon which they act, dynamically adapt to audio inputs and sound scene, automatically configure parameter settings, and exploit best practices in sound engineering to modify the signals appropriately. They derive the parameters in the editing of recordings or live audio based on analysis of the audio content and on objective and perceptual criteria."* (Reiss, 2016, p. 226)

The manner in which intelligent mixes are derived is predominantly by manipulating spectro-temporal characteristics of music signals. The signal processing methods that entail such manipulations are referred to as digital audio effects, in the context of digital multi-track music (Verfaille et al., 2006). The commonly used audio effects are: Stereo panning (Tzanetakis et al., 2007), dynamic range compres-

sion (DRC) (Maddams et al., 2012), level balancing (Bromham, 2016), equalization (EQing) (Hodgson, 2010), and reverb (Valimaki et al., 2012). Using subjective preference ratings for stereo panning in multi-track mixes, Perez Gonzalez and Reiss (2010) showed that semi-autonomous panning was preferred over that from inexperienced mixing engineers. The term autonomous or *'fully autonomous'* in the context of intelligent mixing refers to the manner in which optimal mixing parameters are estimated and the desired mix is created without input from the user (De Man et al., 2019). Interestingly, Perez Gonzalez and Reiss (2010) also showed that preferences for the semi-autonomously panned mixes and those panned by professional mixing engineers were not significantly different. This was similarly the case with an intelligent DRC paradigm proposed by Ma et al. (2015), where DRC was applied to individual tracks based on features extracted from all the constituent tracks in the mix. DRC applied using the paradigm elicited similar preference ratings and, in some cases, was even preferred over that applied by semi-professional mixing engineers. Wichern et al. (2015) showed that the level balance applied autonomously to multi-track music, using the standard energy based loudness model outlined in BS.1770-3 (ITU., 2011), elicited better preference ratings than that achieved using psychoacoustical loudness models proposed by Moore et al. (1997) and Glasberg and Moore (2002). Furthermore, these preference ratings were comparable with that accrued for the mixes engineered by a professional.

Adjustments to EQing have been shown to be effective at reducing simultaneous masking in multi-track mixes (Hafezi and Reiss, 2015). Ronan et al. (2018) proposed a masking reduction paradigm where intelligent mixes were created using automonous EQing and DRC to optimally reduce simultaneous masking in individual tracks. However, when compared to professional mixes, these intelligent mixes were not noticeably better in terms of subjective preference. Chourdakis and Reiss (2017) created intelligent mixes in which artificial reverb was applied to voice, saxophone, and bass tracks with the aid of supervised learning of previous listener preferences. By doing so, they demonstrated that these mixes were preferred over

those without any reverb. However, they received significantly poorer ratings than professionally mixed tracks. Although these studies adequately demonstrate the merit of music pre-processing methods in enhancing the perception of multi-track mixes, they have primarily considered listeners without a diagnosed hearing impairment. Furthermore, a large number of these studies were conducted on NH who are professional mixing engineers themselves. In one such study by Steinmetz et al. (2021), a deep neural network based mixing paradigm was used to automate EQing, compression, and reverb of mixes taken from the École Nationale Supérieure des Télécommunications (ENST) database for Drums (Gillet and Richard, 2006). Based on a subjective evaluation conducted on audio professionals, it was found that the preference ratings for the automated mixes were only marginally lower and in some cases even higher than for the original mixes.

1.3 Music mixing for cochlear-implant users

To date, the subjective preferences of pre-processed music for listeners with a sensorineural hearing impairment have been investigated mostly among CI users. Buyens et al. (2014) were among the first to assess preferences of music mixing among such listeners. They showed that compared to NH, CI users preferred pop music mixes with higher levels of the vocals. Specifically, CI users preferred mixes with vocal levels of around 6 dB higher than the other instruments in the mix or the *'accompaniment'*. Secondly, for mixes with louder and clearer vocals with respect to the accompaniment, CI users preferred higher levels of Bass or Drums over other instruments. Based on their observations, Buyens et al. (2014) argue that music mixes that are available to the general public, may not be suitable for listeners with CI. With the same music mixes, Pons et al. (2016) validated the findings by Buyens et al. (2014), where CI users preferred vocals to be 6 dB louder than the accompaniment. This was assessed using a remixing paradigm that relies on source separation algorithms based on a fast non-negative matrix factorization method (Lee and Se-

ung, 2000) and a Deep Recurrent Neural Network (DRNN) trained by Huang et al. (2014), to separate vocals from the accompaniment in monophonic mixes. Tahmasebi et al. (2020) made a similar investigation using a multi-layer perceptron to separate vocals from a mix. They showed that, contrary to the 10 NH controls who preferred no level differences between the vocals and the accompanying instruments in the mix, vocal level preferences of around 8 dB higher were observed among the 13 CI users tested.

In addition to the implications of distinct level preferences associated with the music enjoyment of CI users, previous research has also showed that they may have difficulties processing more complex mixes. In a mixing task assigned to CI users, Buyens et al. (2014) observed that the subjective difficulty increased with the number of mixing channels made available, making it more challenging for the users to create their preferred mix. In order to evaluate the listening experience among a sample of 16 NH and 9 CI users, Kohlberg et al. (2015) presented several variants of the mix *'Milk Cow Blues'* by Angela Thomas Wade to both participant groups. The listening experience was assessed by way of subjective scores elicited for pleasantness, musicality, and naturalness of the mix. As for the variants presented, the original consisted of more than 10 tracks while the modified variants had fewer tracks. Among NH, the overall listening experience was best for the original mix while CI users preferred the modified mixes with 1-3 tracks. Interestingly, a similar, albeit less pronounced an effect was observed for NH under simulated CI conditions.

CI users also tend to prefer music mixes which are of reduced spectral complexity. Nagathil et al. (2017) investigated the subjective preference of classical chamber music mixes with reduced spectral complexity in a sample of 14 CI users. The reduced complexity was achieved using a low-rank approximation of the Constant-Q-Transform (CQT). They showed that a blind approximation performed using Principal Component Analysis (PCA) accrued better preference ratings compared to a source separation and remixing method used, despite the former even introducing

timbral distortions to the final mix.

The distinct music mixing preferences of CI users compared to NH notwithstanding, it is imperative to acknowledge the unique hearing modalities afforded to these listeners. As CIs prioritize speech perception, they adequately convey slowly varying envelope cues, usually through 12-22 frequency channels, which are sufficient for the purpose (McDermott, 2004). However, CI technologies so far, fail to reliably convey rapidly varying temporal fine structure (TFS) information to the user (Imennov et al., 2013). TFS cues have been implicated in pitch (Smith et al., 2002) and melody perception (Moon and Hong, 2014). Heng et al. (2011) showed that TFS cues may also play an important role in musical timbre perception and that CI users have a poorer ability to judge timbre using mainly TFS cues. Looi et al. (2004) showed that between two notes, CI users required a minimum deviation in their fundamental frequencies by more than 20% or greater than a minor third, in order to correctly discern changes in the pitch direction in Western musical pieces. However, in Western music, these deviations can be as little as 6% or roughly a semitone (Huron, 2001). Marozeau (2021) suggests that such challenges posed by CIs may contribute to the diminished ability of its users to track musical melody. Similarly, Kang et al. (2009) showed that CI users required on average a pitch difference of three semitones to identify the direction of pitch in complex tones, compared to just one semitone sufficient for NH. Importantly, CI users were significantly worse than NH at discriminating melody and timbre.

Studies suggest that HI listeners without CI, may possess better music perception abilities than CI users, even when subjected to hearing-aid use. In one such study, Looi et al. (2008) showed using a sample of 15 CI and hearing-aid users that, the latter had superior musical pitch and melody discrimination abilities in spite of performing similarly in discerning rhythm. Furthermore, hearing-aid users demonstrated superior overall music perception abilities over CI users, despite having similar hearing thresholds. These observed differences notwithstanding, the perception

of music in both groups was noticeably poor. However, given the limited research on the distinct preferences for music mixes among hearing-aid users, this dissertation will initially explore the mixing preferences of popular music among a sample of mild to moderately hearing-impaired listeners, with and without HAs.

In spite of there being only a handful of studies exploring the spectro-temporal mixing preferences of HI, even fewer studies have investigated the ability of individuals with a sensorineural hearing impairment to selectively *hear-out* musical targets amid competing musical maskers in a multi-track arrangement. In other words, Musical Scene Analysis (MSA) as opposed to Auditory Scene Analysis, as a function of hearing loss has received very little attention so far.

1.4 Auditory scene analysis

Auditory Scene Analysis (ASA) is a conceptual framework that describes the process by which the auditory system organizes complex sounds into perceptual components or streams which convey more meaning to the listener about the auditory scene. Cognitive psychologist Albert Bregmann having coined the term, describes it as follows:

> *"Let me clarify what I mean by auditory scene analysis. The best way to begin is to ask ourselves what perception is for. Since Aristotle, many philosophers and psychologists have believed that perception is the process of using the information provided by our senses to form mental representations of the world around us. In using the word representations, we are implying the existence of a two-part system: one part forms the representations and another uses them to do such things as calculate appropriate plans and actions. The job of perception, then, is to take the sensory input and to derive a useful representation of reality from it."* (Bregman, 1994, p. 3)

Importantly, ASA aims to explain the neural and cognitive mechanisms by which the auditory system organizes and separates complex real-world sounds that overlap in both time and frequency into distinct, segregable streams. The process of integration and segregation, collectively referred to as auditory streaming, is central to ASA. By virtue of auditory streaming, an individual is able to attend to a stream of interest in the complex sound event (Calcus, 2024). This remarkable ability of humans to segregate complex sound events into comprehensible auditory streams is fostered from infancy (Sussman and Steinschneider, 2009).

1.5 Auditory streaming

Auditory streaming has been extensively investigated among humans and non-human species. Experiments to understand the process are usually conducted using pure tone sequences of mainly two distinct frequencies. Miller and Heise (1950) pioneered such investigations, where a pure tone is subjected to square-wave frequency modulation to create tone pairs with varying frequency separation. They demonstrated that when the separation between the frequencies was small, listeners perceived the tones as a continuous fluctuation between high and low pitches, through a process referred to as integration. However, as the frequency separation increased, the two tones were eventually perceived as distinct and separate streams - a phenomenon referred to as segregation. Van Noorden (1975) systematically quantified the boundaries of integration and segregation for tone sequences. Importantly, his findings laid the basis for future studies investigating the role of consistent temporal patterns in auditory stream segregation, which later contributed to the understanding of *'temporal coherence'*. Shamma et al. (2011) hypothesized that when an external stimulus elicits temporally coherent neural responses in the auditory cortex, they are perceived as a unified singular stream (integration). On the other hand, concurrently existing stimuli that trigger incoherent neural responses form perceptually separate streams (segregation). Studies by Oh et al. (2022) and

Rajasingam et al. (2021) demonstrated the importance of pitch and timbre cues in the segregation and integration of pure tones. Szalárdy et al. (2014) showed that there was an increased likelihood of integration for a pair of tone sequences with greater temporal overlap and a tonal musical structure, with the former playing a major role in stream segregation. On the other hand, segregation of the sequences was more likely in the presence of melodic familiarity, indicating the influence of long-term memory on auditory streaming.

Although many previous studies have investigated auditory streaming with the aid of simple tones, very few of them do so for complex real-world sounds such as speech or even music. Contrary to that observed for simple tones (Micheyl et al., 2013; Rezaeizadeh and Shamma, 2021), Lee and Oxenham (2024) showed that segregating speech in noise and speech maskers did not depend as strongly on temporal coherence. The segregability of speech in the presence of echoes was investigated by Gao et al. (2024). They demonstrated that, although echoes degrade the slow modulations in speech that are critical for segregation, the auditory cortex of NH remained highly effective in tracking the target speech, even pre-attentively. However, speech segregation was significantly compromised when TFS cues were eliminated.

In that light, the tedium of listening to and tracking speech in the presence of the so-called *'babble'* noise has garnered substantial interest within the purview of ASA. The difficulties faced by listeners in such an auditory scene, were initially formulated in what is the popular *'Cocktail party problem'* by Colin Cherry (Cherry, 1953). The problem originates from what is a typical cocktail party setting, where an individual attempts to focus and understand a speaker of interest amidst a cacophony of competing sounds. The challenges posed by such a setting manifest mainly because of the multiple speakers who simultaneously produce overlapping acoustic signals, thus giving rise to the irksome babble. The brain is therefore required to sift through this complex auditory landscape, isolating the voice of interest while suppressing those which are irrelevant.

The available literature on auditory streaming notwithstanding, relatively few studies investigate the neural implications of streaming in the cocktail party problem. Even fewer studies consider the effect of hearing loss on auditory streaming. In one such study, Bayat et al. (2013) assessed the so-called *'fission-boundaries thresholds'* of NH and HI with mild to moderate hearing loss. In the context of the two-tone streaming experiment, this threshold represents the maximum frequency separation between tones at which they are still perceived as a single stream, often accompanied by a percept of pitch fluctuations or a *'gallop'* (Rose and Moore, 2000). The study showed that the HI listeners had significantly elevated fission-boundary thresholds, particularly at higher frequencies. This finding suggests poorer frequency discriminability in HI, ergo, reduced stream segregation ability which declines with increasing frequency. A similar finding was made by David et al. (2018) for speech, where segregation of interleaved speech sequences was observably poorer for older HI compared to both younger and older NH participants. However, very few studies have explored the effect of sensorineural hearing loss on segregating musical targets in the presence of musical maskers within the context of musical scene analysis (MSA).

1.6 Musical scene analysis

A subset of ASA, MSA deals with the processes by which the auditory system segregates music instruments or *'target'* amid the complex musical ensemble. Much like ASA, integration and segregation play a similarly pivotal role in the perception of the musical scene, allowing the listener to identify individual musical elements whilst perceiving their collective interplay. In MSA, integration relates to the percept of either a combination of individual notes as a unified chord or the process by which a combination of harmonics of a note form a cohesive musical timbre, among many other examples. Similarly, segregation in the context of music is where dissimilar pitch, timbral, rhythmic, and timing cues form separate perceptual streams

(McAdams and Bregman, 1979). By doing so, specific instruments or vocals are perceived as separate, irrespective of evoking the same note.

Particularly in this dissertation, the role of sensorneural hearing loss is investigated in the ability of an individual in effectively segregating and tracking a specific target in the music mix, in a multi-track arrangement; a metric we refer to as MSA performance.

1.7 Aims of this dissertation

Given the limited research on mixing practices for individuals with hearing loss, this dissertation will make early attempts to assess how multi-track mixes can be effectively modified for such listeners. Since these listeners benefit from frequency-dependent amplification, our main goal is to particularly explore how spectrally manipulated multi-track music is received by them.

In light of the common clinical practice of prescribing hearing aids for mild to moderate hearing loss (Ferguson et al., 2017), it is essential to evaluate how these devices influence subjective preferences for music mixing. Understanding these preferences could facilitate the improvement of hearing-aid technologies to enhance music perception in aided listening; a challenge that remains poorly addressed to date.

In spite of subjective preferences, the efficacy of remixed music through spectral alterations should also be objectively assessed. The MSA performance is therefore used as an objective measure of scene analysis abilities among the listeners. As many previous studies demonstrate the diminished ability of hearing-impaired individuals in distinguishing frequencies in tone sequences, this dissertation aims to underpin this assertion with musical scenes. Importantly, this dissertation investigates whether spectral contrast modifications to music can improve the reduced scene analysis abilities that result from diminished cochlear frequency discriminabil-

ity in these listeners.

Lastly, this dissertation investigates how hearing loss affects listeners' ability to appraise changes in spectral shape through subjective audio quality ratings. Furthermore, it examines the possible relationship between musical scene analysis and the listener's perception of audio quality. By exploring these metrics, the work aims elucidate the multifaceted nature of music perception among hearing-impaired individuals.

1.8 Dissertation structure

This dissertation explores the effects of multi-track music mixes subjected to level and spectral adjustments on listeners with mild to moderate hearing loss. The focus is particularly on spectral alterations to music mixes on the subjective preferences and scene analysis abilities. The peer-reviewed studies conducted by the author that address the research questions are provided in Chapters 2-4. Each chapter begins with an introduction to the corresponding study and ends with a contextual summary.

Chapter 2 presents a study in which level-and spectrum-based modifications to pop music mixes and their effect on the subjective preferences are assessed by virtue of a remixing task. The task entails altering the broadband level of the lead vocals relative to the accompaniment, low-high frequency balance, and spectral contrast of the mixes. Changes to spectral contrast are achieved using the EQ-transform which alters the spectral shape of a track with respect to a smooth reference spectrum. This transform is shown to bring about significant changes to the objective frequency-domain sparsity, as measured using the Gini index. In the study, normal-hearing and hearing-impaired listeners with bilateral hearing aids were tested. The results show that hearing-aid users prefer higher levels of lead vocals than normal-hearing listeners, a finding consistent with that observed among cochlear-implant users in

previous studies. Furthermore, among the hearing-impaired listeners, tracks with sparser power spectra and elevated weightings for higher frequencies are favored during hearing-aid disuse. Overall, a higher degree of hearing loss is linked to stronger effects for lead vocal level and contrast preferences. The results underpin the necessity of bespoke mixes for individuals with hearing loss.

Building on the contrast preferences observed in Chapter 2, **Chapter 3** presents a study evaluating scene analysis abilities of listeners for music scenes subjected to the EQ-Transform. The study assesses the listeners' ability to accurately detect a cued target instrument amid a multi-track musical arrangement of mainly pop music excerpts. This so-called MSA performance is evaluated in a sample of young normal-hearing and older hearing-impaired listeners with predominantly moderate hearing loss. The findings reveal an inferior overall performance of hearing-impaired listeners. As observed for normal-hearing listeners, hearing-impaired listeners tend to detect lead vocals with the highest accuracy while doing the opposite for bass guitar, notwithstanding changes to the spectral contrast or shape. Despite non-significant effects of the EQ-transform, the MSA performance, particularly among the hearing-impaired listeners, is sensitive to changes in the spectral descriptors of both the target and the mix. Therefore, the study supports the validity of spectral adjustments to multi-track music as a potential means of improving MSA abilities in listeners with hearing loss.

In **Chapter 4**, a study relating MSA performance and perceived audio quality ratings of multi-track music among a sample of normal-hearing and moderately hearing impaired listeners is presented. In spite of the observation in Chapter 3 where scene analysis abilities remained robust to changes to spectral contrast, audio quality ratings are sensitive to the contrast changes, especially in normal-hearing listeners. Interestingly, the strong positive correlation of MSA performance and quality ratings in hearing-impaired listeners highlights the mutually reinforcing benefits of these two metrics to music perception in listeners with hearing loss.

Chapter 5 provides a general discussion of the key findings from the studies described in chapters 2-4 and presents an overarching conclusion. Importantly, direct comparisons of the results are made with existing literature. At the end of the chapter, potential future research directions are outlined with respect to the overall scope of this dissertation.

References

Abdi, S. (2020). Timely application of hearing aids helps preserve speech discrimination ability. *Journal of Hearing Science*, 10(1):27–32.

Bayat, A., Farhadi, M., Pourbakht, A., Sadjedi, H., Emamdjomeh, H., Kamali, M., and 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).

Boyle, J. D., Hosterman, G. L., and Ramsey, D. S. (1981). Factors influencing pop music preferences of young people. *Journal of Research in Music Education*, 29(1):47–55.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bromham, G. (2016). How can academic practice inform mix-craft? In *Mixing music*, pages 265–276. Routledge.

Buyens, W., Van Dijk, B., Moonen, M., and Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International journal of audiology*, 53(5):294–301.

Calcus, A. (2024). Development of auditory scene analysis: a mini-review. *Frontiers in Human Neuroscience*, 18:1352247.

Chasin, M. and Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, 8(2):35–47.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979.

Chourdakis, E. T. and Reiss, J. D. (2017). A machine-learning approach to application of intelligent artificial reverberation. *Journal of the Audio Engineering Society*.

Cox, R. M., Johnson, J. A., and Xu, J. (2014). Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. *Gerontology*, 60(6):557–568.

David, M., Tausend, A. N., Strelcyk, O., and Oxenham, A. J. (2018). Effect of age and hearing loss on auditory stream segregation of speech sounds. *Hearing research*, 364:118–128.

De Man, B., Stables, R., and Reiss, J. D. (2019). *Intelligent music production*. Focal Press.

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., and Hoare, D. J. (2017). Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database of Systematic Reviews*, (9).

Gao, J., Chen, H., Fang, M., and Ding, N. (2024). Original speech and its echo are segregated and separately processed in the human brain. *Plos Biology*, 22(2):e3002498.

Garcia, T. M., Jacob, R. T. d. S., and Mondelli, M. F. C. G. (2016). Speech perception and quality of life of open-fit hearing aid users. *Journal of Applied Oral Science*, 24:264–270.

Gillet, O. and Richard, G. (2006). Enst-drums: an extensive audio-visual database for drum signals processing. In *International Society for Music Information Retrieval Conference (ISMIR)*.

Glasberg, B. R. and Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5):331–342.

Golovanova, L., Boboshko, M. Y., Kvasov, E., and Lapteva, E. (2019). Hearing loss in adults in older age groups. *Advances in Gerontology*, 9(4):459–465.

Greasley, A., Crook, H., and Fulford, R. (2020). Music listening and hearing aids: perspectives from audiologists and their patients. *International Journal of Audiology*, 59(9):694–706.

Hafezi, S. and Reiss, J. D. (2015). Autonomous multitrack equalization based on masking reduction. *Journal of the Audio Engineering Society*, 63(5):312–323.

Heng, J., Cantarero, G., Elhilali, M., and Limb, C. J. (2011). Impaired perception of temporal fine structure and musical timbre in cochlear implant users. *Hearing research*, 280(1-2):192–200.

Hodgson, J. (2010). A field guide to equalisation and dynamics processing on rock and electronica records. *Popular Music*, 29(2):283–297.

Hoppe, U. and Hesse, G. (2017). Hearing aids: indications, technology, adaptation, and quality control. *GMS current topics in otorhinolaryngology, Head and Neck Surgery*, 16.

Huang, P.-S., Kim, M., Hasegawa-Johnson, M., and Smaragdis, P. (2014). Singing-voice separation from monaural recordings using deep recurrent neural networks. In *ISMIR*, pages 477–482.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19(1):1–64.

Imennov, N. S., Won, J. H., Drennan, W. R., Jameyson, E., and Rubinstein, J. T. (2013). Detection of acoustic temporal fine structure by cochlear implant listeners: Behavioral results and computational modeling. *Hearing research*, 298:60–72.

ITU., R. (2011). Algorithms to measure audio programme loudness and true-peak audio level. In *International Telecommunication Union Radiocommunication Assembly*.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B., and Rubinstein, J. (2009). Development and validation of the university of washington clinical assessment of music perception test. *Ear and hearing*, 30(4):411–418.

Kohlberg, G. D., Mancuso, D. M., Chari, D. A., and Lalwani, A. K. (2015). Music engineering as a novel strategy for enhancing music enjoyment in the cochlear implant recipient. *Behavioural neurology*, 2015(1):829680.

Lee, D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13.

Lee, J. and Oxenham, A. J. (2024). Testing the role of temporal coherence on speech intelligibility with noise and single-talker maskers. *The Journal of the Acoustical Society of America*, 156(5):3285–3297.

Leek, M. R., Molis, M. R., Kubli, L. R., and Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, 19(06):519–526.

Lesimple, C., Kuehnel, V., and Siedenburg, K. (2024). Hearing aid evaluation for music: Accounting for acoustical variability of music stimuli. *JASA Express Letters*, 4(9).

Looi, V., McDermott, H., McKay, C., and Hickson, L. (2004). Pitch discrimination and melody recognition by cochlear implant users. In *International Congress Series*, volume 1273, pages 197–200. Elsevier.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and hearing*, 29(3):421–434.

Looi, V., Rutledge, K., and 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.

Ma, Z., De Man, B., Pestana, P. D., Black, D. A., and Reiss, J. D. (2015). Intelligent multitrack dynamic range compression. *Journal of the Audio Engineering Society*, 63(6):412–426.

Maddams, J. A., Finn, S., and Reiss, J. D. (2012). An autonomous method for multi-track dynamic range compression. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, pages 1–8.

Madsen, S. M. and Moore, B. C. (2014). Music and hearing aids. *Trends in Hearing*, 18:2331216514558271.

Marozeau, J. (2021). Why people with a cochlear implant listen to music. In *Perception, Representations, Image, Sound, Music: 14th International Symposium, CMMR 2019, Marseille, France, October 14–18, 2019, Revised Selected Papers 14*, pages 409–421. Springer.

McAdams, S. and Bregman, A. (1979). Hearing musical streams. *Computer Music Journal*, pages 26–60.

McDermott, H. J. (2004). Music perception with cochlear implants: a review. *Trends in amplification*, 8(2):49–82.

Micheyl, C., Kreft, H., Shamma, S., and Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. *The Journal of the Acoustical Society of America*, 133(3):EL188–EL194.

Miller, G. A. and Heise, G. A. (1950). The trill threshold. *The Journal of the Acoustical Society of America*, 22(5):637–638.

Moon, I. J. and Hong, S. H. (2014). What is temporal fine structure and why is it important? *Korean journal of audiology*, 18(1):1.

Moore, B. C., Glasberg, B. R., and Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240.

Nagathil, A., Weihs, C., Neumann, K., and Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners. *The Journal of the Acoustical Society of America*, 142(3):1219–1228.

Oh, Y., Zuwala, J. C., Salvagno, C. M., and Tilbrook, G. A. (2022). The impact of pitch and timbre cues on auditory grouping and stream segregation. *Frontiers in Neuroscience*, 15:725093.

Owsinski, B. (2014). *The mixing engineer's handbook*. Course Technology, Cengage Learning.

Perez Gonzalez, E. and Reiss, J. (2010). A real-time semiautonomous audio panning system for music mixing. *EURASIP Journal on Advances in Signal Processing*, 2010:1–10.

Pons, J., Janer, J., Rode, T., and Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, 140(6):4338–4349.

Rajasingam, S. L., Summers, R. J., and 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):3769–3784.

Rambarran, S. (2021). *Virtual Music: Sound, Music, and Image in the Digital Era*. Bloomsbury Publishing USA.

Reiss, J. D. (2016). An intelligent systems approach to mixing multitrack audio. In *Mixing music*, pages 246–264. Routledge.

Rezaeizadeh, M. and Shamma, S. (2021). Binding the acoustic features of an auditory source through temporal coherence. *Cerebral cortex communications*, 2(4):tgab060.

Ronan, D., Ma, Z., Namara, P. M., Gunes, H., and Reiss, J. D. (2018). Automatic minimisation of masking in multitrack audio using subgroups. *arXiv preprint arXiv:1803.09960*.

Rose, M. M. and Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, 108(3):1209–1214.

Sandgren, E. and Alexander, J. M. (2023). Evaluating the efficacy of music programs in hearing aids. *The Journal of the Acoustical Society of America*, 153(3_supplement):A40–A40.

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in neurosciences*, 34(3):114–123.

Shuker, R. (2012). *Popular music culture: The key concepts*. Routledge.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):87–90.

Steinmetz, C. J., Pons, J., Pascual, S., and Serra, J. (2021). Automatic multitrack mixing with a differentiable mixing console of neural audio effects. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 71–75. IEEE.

Suatbayeva, R., Toguzbayeva, D., Taukeleva, S., Mukanova, Z., and Sadykov, M. (2024). Speech perception and parameters of speech audiometry after hearing aid: Systematic review and meta-analysis. *Electronic Journal of General Medicine*, 21(1).

Sussman, E. and Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia*, 47(3):771–785.

Szalárdy, O., Bendixen, A., Böhm, T. M., Davies, L. A., Denham, S. L., and 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.

Tahmasebi, S., Gajcki, T., and Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Frontiers in Neuroscience*, 14:434.

Tzanetakis, G., Jones, R., and McNally, K. (2007). Stereo panning features for classifying recording production style. In *ISMIR*, pages 441–444.

Vaisberg, J. M., Folkeard, P., Parsa, V., Macpherson, E., Froehlich, M., Littmann, V., and Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *Audiology Online*.

Valimaki, V., Parker, J. D., Savioja, L., Smith, J. O., and Abel, J. S. (2012). Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5):1421–1448.

Van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences.

Verfaille, V., Zolzer, U., and Arfib, D. (2006). Adaptive digital audio effects (a-dafx): A new class of sound transformations. *IEEE Transactions on audio, speech, and language processing*, 14(5):1817–1831.

Wichern, G., Wishnick, A., Lukin, A., and Robertson, H. (2015). Comparison of loudness features for automatic level adjustment in mixing. In *Audio Engineering Society Convention 139*. Audio Engineering Society.

World Health Organization (2021). World report on hearing. *World Health Organization*.

Zheng, Y., Swanson, J., Koehnke, J., and Guan, J. (2022). Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: a review. *American journal of audiology*, 31(3):819–834.

2. Exploring level- and spectrum-based music mixing transforms for hearing-impaired listeners

In this work, the effects of hearing loss and bilateral hearing-aid use on the subjective mixing preferences, particularly with respect to spectrum and broadband level modifications to multi-track music mixes, are investigated. In order to do so, normal-hearing controls and individuals with mild to moderate hearing loss were tested. The preferences were elicited through remixing tasks of mostly pop music mixes of 8-seconds duration. Fundamentally, the study aims at identifying the distinct preferences in individuals with hearing loss and the influence of hearing aids on these preferences. Importantly, this investigation serves as a foundation for understanding the implications of spectral contrast or shape changes in music mixes on listeners with hearing loss. Furthermore, the study validates the benefits of customized music mixes for moderately hearing-impaired listeners, as shown in previous studies for cochlear-implant users. The findings from this study lay the necessary groundwork for further investigation into the efficacy of modified music mixes for hearing-impaired listeners.

2.1 Study 1

The study included in this chapter was published as: Aravindan Joseph Benjamin, Kai Siedenburg; Exploring level- and spectrum-based music mixing transforms for hearing-impaired listeners. *J. Acoust. Soc. Am.* 1 August 2023; 154 (2): 1048–1061. *https://doi.org/10.1121/10.0020269*. The content of this chapter is identical to the published work.

Author Contributions: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

_____ _____

(name) Date

Supervisor

2.2 Abstract

Multi-track mixing is an essential practice in modern music production. Research on automatic-mixing paradigms however has mostly tested samples of trained, normal hearing (NH) participants. The goal of the present study was to explore mixing paradigms for hearing-impaired (HI) listeners. In two experiments we investigated the mixing preference of NH and HI listeners with respect to the parameters of lead-to-accompaniment level ratio (LAR) and the low-to-high frequency spectral energy balance. Furthermore, preferences of transformed equalization (EQ-transform) were assessed, achieved by linearly extrapolating between the power spectrum of individual tracks and a reference spectrum. Multi-track excerpts of popular music were used as stimuli. Results from Experiment 1 indicate that HI participants preferred an elevated LAR compared to NH participants but did not suggest distinct preferences regarding spectral balancing or EQ-transform. Results from Experiment 2 showed that bilateral hearing aid (HA) disuse among the HI participants saw higher weighting for LARs, stronger weighting of higher frequencies as well as sparser EQ-transform settings compared to that with HA use. Overall, these results suggest that adjusting multi-track mixes may be a valuable way for making music more accessible for hearing-impaired listeners.

2.3 Introduction

Multi-track mixing is a practice whereby separate audio recordings intended for an envisioned piece of music or the mix are combined upon being spectro-temporally modified by a mixing engineer. The process involved in creating a mix often follows the recording phase where raw recordings are made in a studio. For ease of processing, the recordings are conducted in a manner where different sources or instruments are recorded as separate tracks. With the separate tracks from the recording phase, the mixing engineer is then tasked with creating a coherent mixdown version whilst emphasizing the artistic visions of the parties involved (Case, 2011). Furthermore, it is also incumbent upon them to consider the audibility or transparency of all sources in the mix (Moylan, 2014). For a number of reasons including time and costs, it may be be beneficial to automate certain steps of this practice (Reiss, 2016). In fact, research on automatic mixing has made significant progress in the last 15 years (De Man et al., 2019). However, to the best of our knowledge these approaches have mostly been studied on expert listeners and have not considered hearing-impaired listeners. Given that hearing aid users continue to be dissatisfied with how hearing aids transmit music (Madsen and Moore, 2014; Greasley et al., 2020), developing strategies to pre-process music signals for hearing aid users is an important task for music processing research. Here, we evaluate the preferences of normal-hearing (NH) and hearing-impaired (HI) listeners with regards to level- and spectrum-based mixing effects.

To understand the processing chain of mixing from a mathematical perspective, several authors have put forward models to automate the mixing process. Izhaki (2017) describes the mixing process as being a correlative one where the processing needed on one track depends entirely on the presence or upon the introduction of other tracks. Jillings and Stables (2017) monitored user interactions on a browser-based Digital Audio Workstation (DAW). Upon investigation, it was apparent that the degree of changes made by the users decreased as they moved towards the latter

stages of their mixing project with the DAW. Accordingly, an iterative model of the mixing process was suggested, where one goes from an early *coarse* stage to a later *fine* stage requiring more fine-grained attention through continuous refinement of earlier mixing decisions. Alternatively, Ma (2016) modeled mixing as an optimization problem where, given a finite set of variable parameters, the final mix is created by virtue of arriving at an optimal solution to a system of equations that describes the process best. Reiss (2011) proposed topologies with which automatic mixing can be achieved. In these topologies, necessary spectral and temporal features of the raw or unmixed tracks are first extracted and processed to create modified or processed tracks making up the final mix.

The field of automatic mixing has been emerging in significance with the advent of novel techniques in engineering and signal processing (Moffat and Sandler, 2019). A number of studies have been conducted on intelligent mixing systems, employing techniques to create mixes that mimic that of a trained engineer. An important question in automatic mixing concerns the reduction of auditory masking, that is, the mutual overshadowing of sound sources in a mix. Hafezi and Reiss (2015) explored an automatic EQ-based masking reduction paradigm, which was tested on a sample of 11 NH participants between the ages of 20 and 42 years. The paradigm was implemented both using an off-line and a real-time approach. To objectively evaluate the efficacy of masking reduction, the masker to unmasked ratio (Aichinger et al., 2011) was evaluated. Based on this measure, the paradigms showed improvements in the masking reduction when compared to the un-mixed or raw versions of eight songs. A subjective evaluation conducted on four songs showed that some of the implementations created mixes that were preferred over manual mixes created by novice mixing engineers. Notably, listeners with sensorineural hearing impairment have been shown to have higher masking thresholds even at frequencies where they show less than 30 dB hearing loss (Smits and Duijhuis, 1982). This could suggest that such listeners may benefit from greater masking reduction through higher masker to unmasked ratios.

In a later study, Ronan et al. (2018) presented a paradigm to optimize masking reduction via EQ and Dynamic Range Compression (DRC) controls. The optimization of the controls was achieved with the aid of a particle swarm optimizer with and without sub-grouping tracks in the mix. The particle swarm optimization is performed by moving a set of candidate solutions to a problem called particles via a search space, usually a high dimensional Cartesian space. Each particle's velocity and step size are iteratively updated with time and also depend on the location of the particle with the best or most optimal position at a given time (Kennedy and Eberhart, 1995). An optimal solution is reached upon minimizing a cost function which in this case is the L2 norm of multi-track masking at frequencies between 500 Hz and 2 kHz. In sub-grouping, tracks belonging to similar instruments were grouped and the controls were applied to each of these groups to create secondary mixes which were then summed to create the final mix. When there was no sub-grouping, the controls were applied to all the un-mixed tracks at once. Objective evaluation of the cross-adaptive multi-track masking reduction showed that the use of sub-grouping achieved greater reduction in masking with fewer iterations of the optimizer. This was supported by a listening test with 24 NH participants (between the ages of 23 to 52 years) with 5 songs. The mixes created using this paradigm with sub-grouping elicited better preference and clarity ratings than when no sub-grouping was considered. As an alternative, Tom et al. (2019) demonstrated a masking reduction paradigm via frequency-dependent panning. Here, two approaches were explored with one being a real-time implementation and the other an offline implementation. In the real-time implementation, a palindromic Siegel-Tukey type ordering (Siegel, 1956) with respect to masking was performed, such that the last stem in the ordering would require the least or no panning intervention for masking reduction. In the off-line implementation, panning position of a track to optimize masking reduction was estimated using a particle swarm optimizer. Objectively the resulting panning positions were comparable to those in professional mixes. A subjective evaluation conducted on 25 trained NH participants showed that both

modes out-performed available panning based masking reduction paradigms. Steinmetz et al. (2021) demonstrated an early application of deep learning for creating automatic mixes: With a subjective evaluation conducted on 16 audio engineers, it was demonstrated that the implementation had promise by virtue of being rated closely to factory or target mixes especially when tested on audio samples from the ENST drums database (Gillet and Richard, 2006).

Although successful paradigms have been formulated for automatic mixing, these have only been tested on listeners with no reported hearing impairment and on a relatively small number of audio samples. This beckons the question of whether HI listeners would benefit from mixes specifically tailored towards their needs. Kohlberg et al. (2015) showed that CI users prefer reduced music complexity with fewer instruments in the mix than NH listeners. Similarly, Pons et al. (2016) indicated that cochlear implant (CI) users may benefit from individualized mixes and higher vocal levels specifically. Buyens et al. (2014) also showed that CI users preferred the lead vocal levels to be enhanced with respect to the rest of the instruments in the mix and also intimated that the mixes generally available to the public might not be suitable for such listeners. More recently, Tahmasebi et al. (2020) proposed a deep-neural-network-based and real-time capable source separation for music remixing to enhance music perception of CI listeners. The implementation was aimed at separating lead vocals from the rest of the mix in popular western music. The stimuli were presented to 13 bilateral CI users and 10 NH participants under realistic conditions with and without visual cues. It was evident that the CI users preferred elevated lead vocal levels with respect to that of the other instruments irrespective of the reverberance or the presence of visual cues. As such, the lead vocal level preferences were 8 dB higher among the CI users than the NH participants tested.

Nagathil et al. (2017) demonstrated a paradigm where by the spectral complexity of classical chamber music pieces were reduced through low rank approximation of their constant-Q transforms. This paradigm yielded significantly higher prefer-

ence ratings among a sample of 14 CI users over a source separation and remixing paradigm. In the latter, the lead vocals were separated from the mix and remixed with elevated levels as suggested by Buyens et al. (2014). Based on the fact that the low rank paradigm brings about timbre distortions and yet outperforms the source separation and remixing paradigm that faithfully reconstructs the lead vocals, it can be maintained that the study underpins the priority of reduced spectral complexity over timbral fidelity in chamber music among CI users.

Whether such findings would generalize to HI participants wearing hearing aids (HA) instead of CIs is yet to be ascertained. It is well-known that HI listeners are fraught with psychoacoustical limitations. These limitations include impaired frequency selectivity of HI listeners (Glasberg and Moore, 1986), impaired temporal fine structure sensitivity (Hopkins and Moore, 2011) and impaired sound localization abilities (Warnecke et al., 2020). Studies of music perception have shown that listeners with moderate hearing-impairment have drastically reduced abilities to hear out melodies or instruments from a mixture (Siedenburg et al., 2020). In a later study by Siedenburg et al. (2021b), the ability to track if a reference voice in a mixture had tremolo artificially introduced by amplitude modulation was investigated among young normal hearing listeners and older hearing impaired listeners. The latter were tested with and without their hearing aids. It was discovered that the ability of the older hearing impaired listeners to track the existence of tremolo in the reference voice did not improve with the use of their hearing aids.

A number of aforementioned studies allude to the fact that cochlear implant users prefer the enhancement of lead vocal level relative to the other instruments in the mix. Accordingly, Bürgel et al. (2021) showed that lead vocals serve as powerful attractors for garnering auditory attention in popular music. In a more recent study, Knoll and Siedenburg (2022) showed that a mixed group of NH and HI participants preferred elevated lead vocal levels than those available in original mixes. To investigate if the distinct preferences reported in the literature on CI

listeners hold true when NH and moderately HI listeners are compared directly, individual preferences regarding the level based mixing effect lead-to-accompaniment ratio (LAR) were recorded in this study.

Concerning spectral mixing effects, a study by Hornsby and Ricketts (2006) showed that the speech information use at high frequencies did not improve with frequency dependent amplification among participants with sensorineural hearing impairment. However, in order to assess the contribution of frequency-dependent loudness weighting in music preference, we use a spectral based effect changing the balance of spectral energy of the audio signal in this paper. Furthermore the consequence of reduced frequency selectivity brought on by hearing impairment (Florentine et al., 1980) on music preference need be assessed from the perspective of mixing. EQing may well serve as the effect which can be manipulated to make such an assessment. However, Izhaki (2017) highlights the challenges associated with the appropriate application of EQing in popular music. Here, the author emphasizes the effect of EQing on *tonality* achieved by accentuating and attenuating levels at different frequency bands. He also highlights that it serves as the cardinal tool with which the engineer may alter the timbre of the instruments in the mix. This can be used to convey different emotions to the listener, so much so that it must be performed with the greatest care by a trained professional.

In this study, we aim at emphasizing or downplaying the spectral distinctiveness of individual tracks via the EQ-transform. Through assessing the preferences of these mixing effects, our goal is to characterize mixing preferences of HI listeners in comparison to NH listeners, as well as the effect of HA use on these preferences from a general perspective, as opposed to doing so as a function individualized settings of the HA. Overall, these concerns indicate several issues in the production of multi-track mixes for HI listeners, let alone their automation. Therefore, it behooves research in this direction to explore their mixing preferences prior to creating dedicated automatic mixing paradigms. Moreover, whether the use of bilateral hear-

ing aids has an influence on mixing preferences among HI listeners is still open for discussion. To answer these questions, this study evaluated listeners' preferences with regards to characteristic mixing effects. In Experiment 1, we tested a sample of NH and HI participants with and without bilateral Hearing Aids (HAs). In Experiment 2, a sample of HA users were tested to compare their preferences with and without HAs. Based on previous studies, we hypothesize that HI participants may show elevated lead vocal level preferences and high frequency weighting in the mixes than NH participants. Furthermore, reduced frequency selectivity among HI participants may manifest as greater affinity towards spectrally sparser mixes. HA use is also hypothesized to bring about significant differences in these preferences among HI listeners.

2.4 Mixing effects

Here we outline the mixing effects used in the present study[1]. The rationale is to provide effects that have one free parameter only and thus may be easily apprehended and used by the participants of our study. As motivated above, we seek to test the LAR and also explore spectral-based effects and the way which participants adjust these effects according to their preference.

2.4.1 Lead-to-Accompaniment Ratio

The Lead-to-Accompaniment ratio (LAR) in dB is varied by accentuating or attenuating the broadband level of the lead vocal track with respect to that of the accompanying instruments considered *enmasse*. By this, we merely consider all the tracks other than the lead vocal tracks in the mix, which we refer to as the *accompaniment*. We also disregard backing vocal tracks entirely. The manipulation herein affects only the level of the lead vocals in the multi-track excerpts, leaving the relative levels of the accompanying tracks unperturbed as they were in the original

[1]For sound examples, see:
uol.de/en/music-perception/sound-examples/mixing-transforms-for-hearing-impaired-listeners

mix. This was done to avoid bringing out unnatural level relationships between the accompanying tracks (to prevent accentuating the level of low energy or transient tracks which may inadvertently bring about the audibility of background noise). A LAR of 0 dB would imply that the level of the lead vocals and that of the accompaniment are identical. To avoid alterations brought on to panning in the mix, the weighting applied to the left and right channels of the lead vocal tracks were identical in that the lead vocal levels of both channels were altered in unison. Furthermore, as the broadband levels were an average evaluated over the entire duration of the excerpts, the silent or low-energy portions were also part of the calculation.

2.4.2 Spectral balance

In this spectral filtering effect described by Siedenburg et al. (2021a), the weightings of bands of a filter bank is altered. Effectively, this shifts the spectral slope between 125 Hz and 8 kHz of the final mix. That is, positive values for spectral balance increase the auditory brightness of the signal. A spectral balance of 0 dB/Oct means the filter applied is an all pass filter with no change in the spectral centroid. The slope is only applied between 125 Hz and 8 kHz with a balancing point at 1 kHz (implying a gain at this frequency of 0 dB). The filter is applied on the final stereo mix which encompasses all tracks.

2.4.3 EQ-transform

In transformed mixing effects, the effect of interest is acquired from a factory mix which is made originally available by the mixing engineer (*factory* effect). The effect in question is then extrapolated linearly with respect to a reference for that effect. Therefore, the extrapolation is performed with the reference effect corresponding to 0% and the factory effect to 100%. A participant wishing double the effect available in the factory mix would do so by choosing a 200% transform. This is to say that if the power level of a given track, at a given frequency bin was 5 dB above the reference level (level at the 0% transform), a 200 % EQ transform would then transform this

difference to 10 dB, and a 300 % transform to 15 dB and so on. By doing so, power levels above the reference in each track are accentuated and notches falling below the reference are commensurately attenuated thus affecting the frequency domain or spectral sparsity of the track. To gauge changes in such sparsity brought on by the transform, a constant Q-transform based method to evaluate the Gini-index or coefficient as a measure of sparsity was used. The Gini-index-based measure was used owing to its robust nature (Hurley and Rickard, 2009). Rickard and Fallon (2004) showed using the Gini coefficient that speech samples taken from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Garofolo, 1993) were sparser in the tempo-spectral domain than in the temporal domain. In a later study, Rickard (2006) showed that the Gini-index as a measure of time-frequency sparsity serves as a reliable indicator of mathematical separability of sources in a mixture by serving as a reasonable surrogate for *disjoint orthogonality* of the sources. In this context, disjoint orthogonality indicates how strongly the energy of one source dominates the other sources in the mix at a particular point in time and frequency. Zonoobi et al. (2011) demonstrated the superiority of the Gini coefficient as a measure of sparsity over conventional norm based measures when reconstructing randomly generated one dimensional signals and real images with and without additive white gaussian noise. The signals and the images were reconstructed from compressed samples using sparse estimates with aid of the norm based measures and the Gini coefficient. Those reconstructed with the aid of the latter had the lowest mean square errors with respect to their original versions. Furthermore, noise corrupted images reconstructed with the aid of the Gini coefficient had the highest peak signal to noise ratios. More recently, Orović et al. (2022) demonstrated the reliability of the Gini coefficient as a measure of energy distribution obtained from time - frequency representations of non-stationary signals.

Figure 2.2(A) illustrates the manner in which the sparsity was objectively measured in this study. Glasberg and Moore (1986) showed that hearing impairment lead to broader auditory filters indicating poor frequency selectivity. This was also

shown by Zurek and Formby (1981) where sinusoidal frequency modulation discriminability depreciated with higher levels of hearing loss. The EQ-transform was therefore conceived as a potentially valuable tool to assess the effects of spectral sparsity in multi-track mix preferences among HI listeners owing to such shortcomings. In

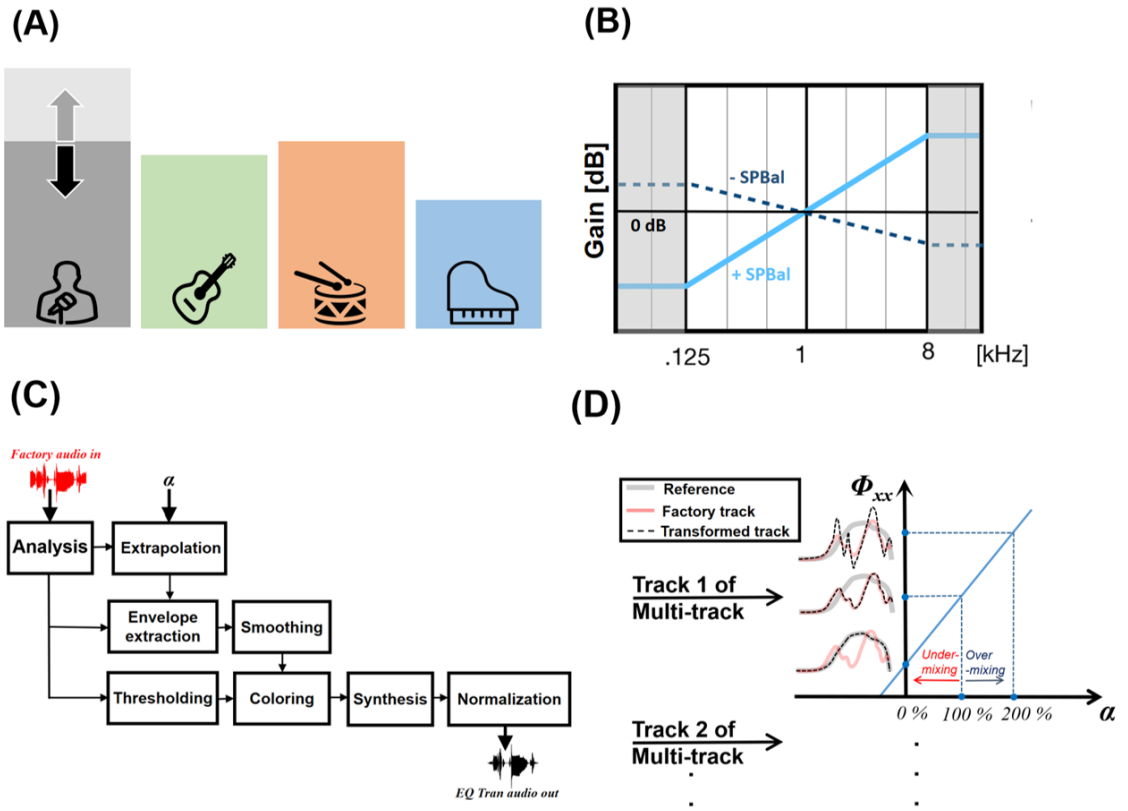


Figure 2.1: (A) Illustration of Lead-to-Accompaniment implemented here. Only the broadband level of the lead vocals are varied whilst those of the accompaniment are left unaltered. (B) Spectral balance as adapted from Siedenburg et al. (2021a); licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0; https://creativecommons.org/licenses/by/4.0/) license. (C) Illustration of the EQ-transform processing chain showing the transform being performed in eight steps for a given audio input. The input to the process is always the original from the mix referred to here as the *factory* audio. (D) Illustration of the extrapolation phase where the transformed power spectrum is derived.

the EQ-transform, each stereo-channel of each track was processed independently. Specifically, the one-sided power spectra (using a 352800-point FFT corresponding to the 8 second duration and sampling frequency of 44100 Hz) of the factory tracks were linearly extrapolated between themselves serving as the 100% transform and the reference power spectrum corresponding to a 0% transform as shown in Fig-

ure 2.1. The so called reference spectrum was the ensemble average of that of the lead vocals and of the most commonly occurring instruments (piano, guitar, bass guitar, drums, percussion, and synth instruments), extracted from the open source databases used in this study (discussed in the next section). As the EQ-transform process depends heavily upon the factory power levels with respect to the reference levels, the latter is normalized such that it has the same overall power as the track undergoing the transform.

The EQ transform was completed in eight steps illustrated in Figure 2.1(C). Each track was independently subjected to these steps in the following manner : 1) The power spectrum of a track was obtained using the FFT (Analysis). 2) The transformed power spectrum of the track was derived as shown in Figure 2.1(D). Here, the so called transformed power spectrum for the track is derived by linearly extrapolating between the reference spectrum and the factory power spectrum of that track (Extrapolation) . 3) The residual power spectrum (difference between the transformed spectrum from step 2 and the factory spectrum from step 1 was evaluated (Envelope extraction). 4) This residual spectrum was then smoothed using a Savitsky-Golay filter (Schafer, 2011) (Smoothing). The smoothing was performed to avoid temporal smearing in the transformed mix. 5) Bands of the factory spectrum with power not less than 90 dB below the global maximum for the track were evaluated (Thresholding). 6) The smoothed spectrum from step 4 was used to color the factory spectrum in the bands evaluated in step 5 (Coloring). 7) The time domain representation of the colored spectrum is then obtained using the inverse fourier transform (Synthesis). 8) The the resulting signal is normalized so that it has the same broadband level of the factory track from which it was derived (Normalization). Finally, the EQ-transformed mixdown were created by merely adding the normalized signals for each track.

As shown in Figure 2.2(A), Constant-Q-Transform (CQT) was applied to an audio input with a frequency resolution of 3rd of an octave and the hamming window

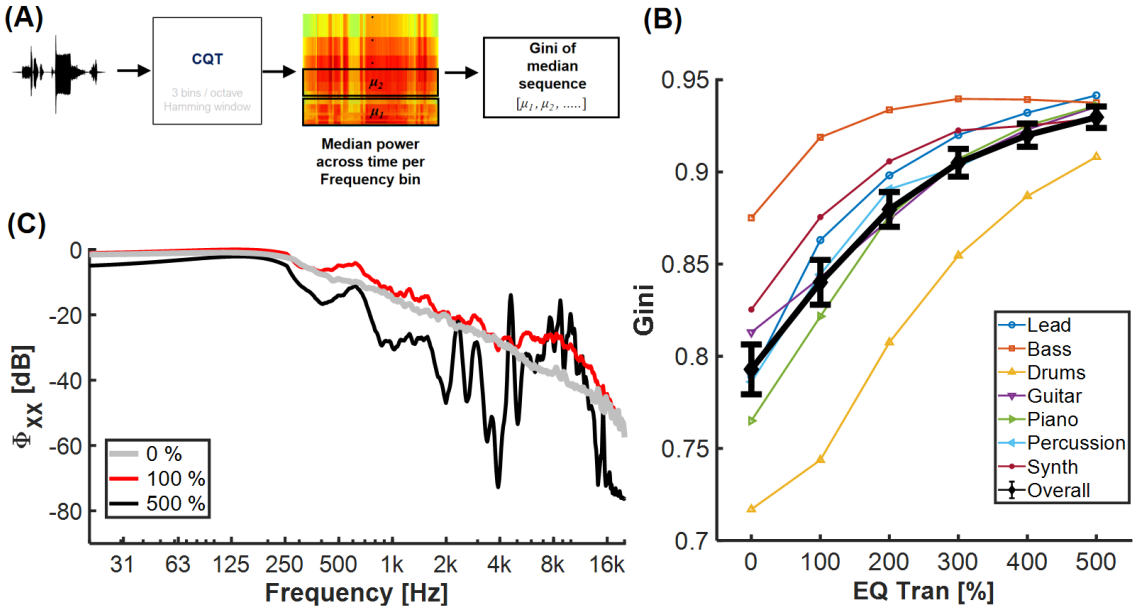


Figure 2.2: (A) The Constant-Q-Transform (CQT) based method to objectively evaluate frequency-domain sparsity illustrated here for an EQ-transformed lead vocals track. (B) Plot of sparsity measure using the mean Gini coefficient as a function of % EQ-transform for the commonly occurring tracks in the Medley and Cambridge database. The higher the Gini index, the greater the sparsity in the frequency domain. (C) Envelope of the power spectra of an example multi-track mix (Hold on you by James May, see supplementary material[2]) after 0%, 100% (Factory mix), and 500% (over-mixed) EQ-transform.

used as a windowing function. Upon performing the CQT, the median power across time was obtained for each frequency bin. The Gini coefficient was then evaluated for the resulting sequence of the median power values. Figure 2.2(B) illustrates the Gini coefficient evaluated in this manner as a function of the applied EQ-transform for 75 lead vocal, 78 bass, 84 drum, 35 guitar, 18 percussion, 40 piano, and 24 synth tracks extracted from both of the databases used in this study. From Figure 2.2(B), a monotonically increasing mean Gini coefficient with respect to over-mixing can be observed. A higher Gini indicates higher frequency domain sparsity alluding to the fact that the EQ-transformation implementation used in this study gives rise to greater spectral sparsity with over-mixing in the individual tracks. A noticeable increase in spectral density with under-mixing, particularly between 0% and 100% is also apparent. Going from 0% reference, there appears significant increase in the Gini index going towards 100% and onward showing higher spectral sparsity

with increasing % EQ-transform. Figure 2.2(C) shows the envelopes of the power spectral densities of an example mixdown after 0%, 100% (Factory mix), and 500% EQ-transform. The excerpt contained 4 tracks from lead vocals, bass guitar, drums, and guitar. An apparent increase in the spectral sparsity of the mixdown upon over-mixing as shown for individual tracks is visible here.

Summary of definitions

Abbreviation	Definition
CI	Cochlear implant(s)
EQ Tran	EQ Transform
HA	Bilateral hearing aids
HI	Hearing impaired listeners / participants
HL	Hearing loss level
LAR	Lead to Accompaniment Ratio
NH	Normal hearing participants
SPBal	Spectral balance about 1 kHz
wHA	Hearing impaired participants with bilateral hearing aids
woHA	Hearing impaired participants without bilateral hearing aids
BTE	Behind-the-ear type hearing aids
ITE	In -the-ear type hearing aids

2.5 Experiment 1

In this first experiment, we compared effect preferences in a between-subjects design with distinct groups of normal-hearing participants, hearing-impaired participants who did not wear hearing aids, and hearing-impaired participants who wore hearing aids.

2.5.1 Methods

Participants

A sample of 25 normal hearing (NH) and 20 hearing impaired (HI) participants took part. Among the HI participants, ten participants were bilateral hearing aid users (wHA) and other ten participants did not use any hear aids (woHA). HI participants were recruited via Hörzentrum gGmbH. However, the distinction between wHA and woHA was made post-recruitment. The NH participants were recruited using an online advertisement with no reference to hearing impairment whatsoever. However, there were a few woHA participants who were recruited through the advertisement. They were classified as hearing impaired with the aid of pure-tone audiometry performed on all participants. All of the participants were compensated at the rate of 12 Euros per hour for their involvement in the study.

The NH participants were on average 28 years old ($SD = 9.6$), wHA participants were 70 years old ($SD = 9.5$), and woHA participants were 67 years old ($SD = 14.5$).The age difference between wHA and woHA participants was not significant. Among the NH participants, there were 15 female and 10 male participants. Among wHA participants, there were seven male and three female participants. In the woHA group, there were four female and six male participants. Figure 2.3(A) shows individual and median hearing level (HL) across the respective participant groups. The HL was assessed via pure-tone audiometry using puretones at 125 Hz, 250 Hz,

500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz frequencies. From the audiogram, it can be observed that the wHA participants indeed had more than 10 dB greater HL on average (M = 42 dB, SD = 11 dB) compared to the woHA participants (M = 29 dB, SD = 3 dB). This can also be further observed in Figure 2.3(B), displaying age and average HLs (arithmetic mean across all tested frequencies). Here, it is apparent that all of the woHA participants had mild hearing impairment (25 dB $<$ HL $\leq$ 40 dB) as compared to 50 % of the wHA participants with moderate to severe hearing impairment (HL $>$ 40 dB) (Clark, 1981). The NH participants did not have elevated HL, as expected (M = 1.5 dB, SD = 5 dB).

Figure 2.3(B) shows the relationship between average hearing loss and age among the participants. A significant positive correlation was observed, $r(43) = 0.9$, $p <$ .001. According to the Gold MSI musical training subscale, participants had mean scores of M = 29 (SD=10) for NH, M = 26 (SD = 12) for wHA, and M = 31 (SD = 13) for woHA participants and there were no significant differences between any of the three groups of participants ($p >$ .3).

Concerning the HA of the wHA participants, all nine of the ten participants from whom the data was made available after the study wore behind-the-ear (BTE) hearing aids. Among them, five of them wore closed-fit type HAs (with a tube connecting the BTE case to a ear mold customized for the participant). The participants using the closed-fit HAs reported having used them between 2 and 15 years (M = 9.4 yrs, SD = 4.9 yrs). The other four participants were using open-fit type hearing aids where (with a tube connecting the BTE case to a dome, leaving the ear canal open). This allows for unamplified low frequency sound to enter. The participants reported having used these open-fit type hearing aids between 4 and 17 years (M = 7.8 yrs, SD = 6.2 yrs). The HA use of the remaining wHA participant was unavailable at the time collection after the study was completed. Refer to the supplementary material for information pertaining to individual HA use.[3]

[3]See section 6.1 of the supplementary material at Appendix A for information pertaining to HA use among wHA in Experiment 1.

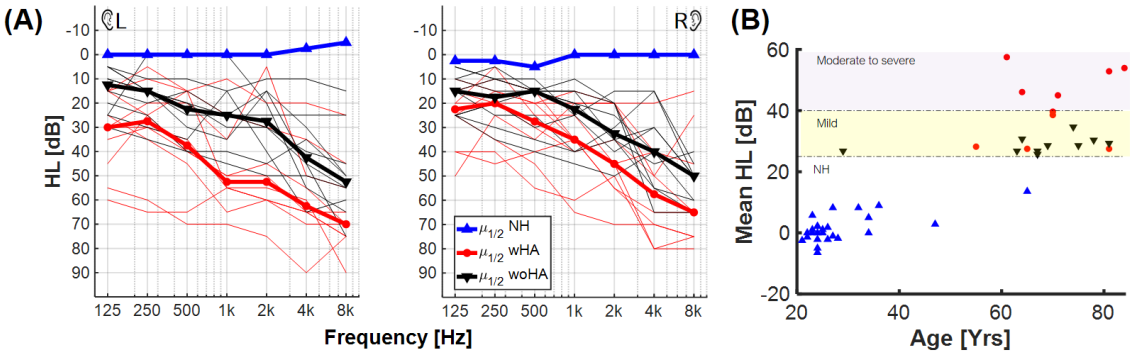


Figure 2.3: (A) The audiograms for the NH and the two HI participant groups taken for the left and right ears. (A) The relationship between hearing loss level (HL) averages and participants' age. Hearing loss categories indicated according to (Clark, 1981).

Stimuli and apparatus

The audio excerpts used as stimuli were 8 seconds long and were taken from the Medley database (Bittner et al., 2014). All of the audiometry in this study was conducted using a portable AD528 audiometer by Interacoustics. The audio playback was realized over a pair of ESI activ 8" near-field studio monitors in a low reflection chamber at the University of Oldenburg, Germany. The monitors were separated by a 90° angle and 2 m distance from the listener's seat. The overall playback level was adjusted to 80 dBA at the participant position. The monitor levels at the participant positions were calibrated to these levels with a stationary noise shaped with an average spectrum acquired from commonly available instruments from both of the aforementioned databases as with the reference spectrum used for the EQ-transform discussed earlier. Please see section 5 of the supplementary material for sound pressure levels of the individual excerpts presented in the study at the participant position[4]. Due to the fluctuating levels of music signals, we have also provided the minimum (LAFmin) and maximum (LAFmax) levels of our stimuli. The measurements were made using a Nor140 precision sound level meter from Norsonic AS. The sound pressures were measured over a full minute during which the excerpts were looped. It was evident herefrom that none of the excerpts used in this study

[4]See section 5 of the supplementary material at Appendix A for minimum and maximum sound pressure levels of the excerpts measured at the participant position.

exceeded 90 dBA at maximum level. The audio transforms were realized using a standalone desktop computer running MATLAB 2021b. The desktop computer was connected to the monitors via an RME Fireface UFX audio-interface.

Procedure

Prior to commencing with the training phase of the experiment, the participant was asked to fill the Gold MSI musical training questionnaire (musical training subscale) (Müllensiefen et al., 2014) in order to estimate their level of musical training. Upon completing the questionnaire, the participant was guided to the training phase of the experiment which was used merely to acquaint the participant with the graphical user interface and the concept of the experiment.

The training phase consisted of a single block of 10 trials where an audio excerpt that was not included in the main experiment was repeated. In each trial, one of the three afore-discussed mixing effects would be manipulated by the participant via the rotation of an ungraduated virtual dial. Between each trial the initial position of the virtual dial was randomized. As the mixing effect would take effect immediately upon the dial change, the participant would be instructed to set the dial where they preferred the audio playback over the loudspeakers best.

After completion of the training phase, the main phase of the experiment ensued after a break during which the participant was requested to retain or wear their hearing aids throughout the rest of the experiment if they were indeed hearing aid users. The main phase of the experiment was grouped into 3 blocks where each block was dedicated to one of the mixing effects. Each block comprised 10 trials with different multi-track excerpts being presented per trial. Within a given block, no excerpt was presented more than once. There were 20 distinct multi-track excerpts taken only from the Medley database. Among them, the first 10 were used exclusively for the level based effect and the other 10 excerpts were used for the two spectral based effects. To avoid order biases, both the order of the blocks and the

order in which the excerpts were presented within a block were randomized. The starting position of the virtual dial in each trial was also assigned randomly. The dial position set by the participant was stored upon proceeding to the next trial.

Data analysis

A linear mixed effects model (LME) was used to estimate relationships between mixing effect preferences and participant groups, the 10 audio excerpts, and the interaction effects between the two factors. Such a model was used for its advantages over repeated measures ANOVA for unbalanced sample sizes between groups to avoid list-wise deletion (Lohse et al., 2020). To summarize the main effects and interactions, results are presented in the form of classic ANOVA statistics for the ease of interpretation, derived from the LME models via MATLAB's `anova` function.

To underpin the differences shown in the mixed effects model, a post-hoc, independent samples t-test was used. The test was applied on average preferences of each participant, calculated across the ten presented excerpts. The resultant p-values were subjected to a controlled-Holm procedure, for it is uniformly more powerful than the Bonferroni correction (VanderWeele and Mathur, 2019). As a measure of effect size, Cohen's d (Lakens, 2013) was used.

As a way to assess individual differences within groups, a test of variance compared the mean preferences taken across all excerpts for each participant to mean preferences taken across participants within the group. In other words, evaluating variances of each distribution in Figure 2.4(B) were evaluated in a given participant group and that in Figure 2.4(C) for the same group for comparison. A single-tailed test of variance was then used to determine if the variance of the former was significantly larger than the latter. A significantly higher variances of mean preferences of participants would indicate significant individual differences.

2.5.2 Results and discussion

As mentioned earlier, the preferences for a given mixing effect were elicited from each of the participants for ten distinct audio excerpts. Figure 2.4(B) illustrates 95% confidence interval plots and mean LAR preferences elicited from each participant for the ten excerpts. LAR preferences pertaining to each of the excerpts averaged across participants belonging to the respective groups are presented in Figure 2.4(C). Figure 2.4(A) shows the resulting averages across excerpt per participant. The preferences recorded for both of the spectral mixing effects are also presented in this manner in Figures 2.5 and 2.6. According to the one-sample t-test conducted on means shown in Figure 2.4(A), LAR preference among NH participants were slightly negative (M = -0.92 dB, SD = 2.33 dB), whereas wHA participants preferred positive levels of LAR (M = 1.68 dB, SD = 2.76 dB), similar to woHA participants (M = 1.32 dB, SD = 4.78 dB). Yet, none of these deviations from zero were statistically robust, neither for NH participants ($t(24) = 2$, $p = 0.06$), nor for HI participants ($t(9) > 0.9, p > 0.4$).

In a direct comparison between groups, the LME model showed significant inter-group effects $F(2,420) = 3.64$, $p = .02 < .05$. Furthermore, there were inter-excerpt effects $F(9,420) = 16.52$, $p < .001$, together with a significant interaction effect $F(18,420) = 1.67$, $p = .04 < .05$. The post-hoc independent samples t-test showed that mean LAR preferences among wHA participants were significantly higher than that among NH participants, $t(33) = 2.83$, $p = .008 < .02$, $d = 1.06$ (large effect size) as can be observed in Figure 2.4(A). When comparing NH and woHA participants, no significant differences were shown, $t(11) = 1.4$, $p = .18$. However, the two-tailed test of variance shows that the variances of the mean LAR preferences between NH and woHA are significantly different ($F = 0.24$, $p = .005 < .01$) and the degrees of freedom were therefore adjusted from 33 to 11. Finally, no significant differences between the wHA and woHA were shown, $t(18) = 0.2$, $p = .84$. None of the groups showed any significant within-group individual differences ($p > .08$).

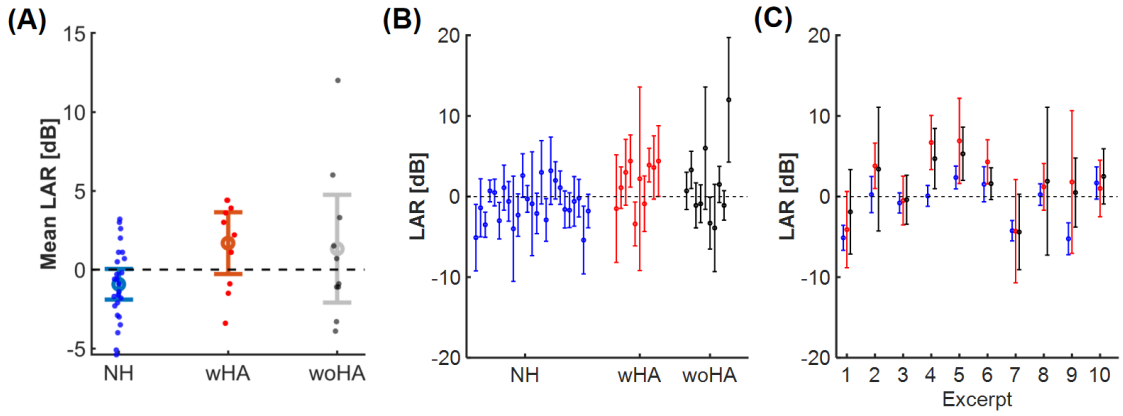


Figure 2.4: (A) Means and respective 95 % confidence interval plots of mean LAR preferences taken over all 10 excerpts. Dots correspond to individual data of each participant within the groups. (B) Means and respective 95 % confidence interval plots of LAR preferences taken for each participant over the 10 excerpts within the groups. (C) Means and the respective confidence interval plots of LAR preferences averaged across participants within each group for each of the 10 excerpts presented.

Similarly, the one-sample t-test performed on the mean spectral balance preferences illustrated in Figure 2.5(A) did not reveal significant deviations of preferences from the 0 dB/Oct reference (where the unweighted factory mix-down is presented) for NH participants ($M = 0.2$ dB/Oct, $SD = 0.8$ dB/Oct), $t(24) = 1.3$, $p = 0.2$. This was similarly the case for preferences among wHA participants ($M = 0.1$ dB/Oct, $SD = 0.6$ dB/Oct), $t(9) = 0.6$, $p = 0.6$. However, woHA participants ($M = 0.6$ dB/Oct, $SD = 0.7$ dB/Oct) preferred a significantly elevated preference favoring weighting higher frequencies about 1 kHz more, $t(9) = 2.4$, $p = .04 < .05$, $d = 0.77$ (medium effect).

Here, the LME model did not show significant inter-group effects of preferences, $F(2{,}420) = 1.13$, $p = .32$ between the participant groups. Furthermore significant inter-excerpt effects $F(9{,}420) = 11.07$, $p < .001$ but no interaction effects $F(18{,}420) = 1.1$, $p = .34$, were shown. However, unlike with LAR preferences earlier, the variance of mean SPBal preferences taken across all excerpts for each participant was significantly larger than that taken across all participants for each excerpt for NH participants ($F = 6$, $p = .004 < .01$). This indicates stark differences in SPBal preferences owing to the NH participants. This can be visualized in Figures 2.5(B &

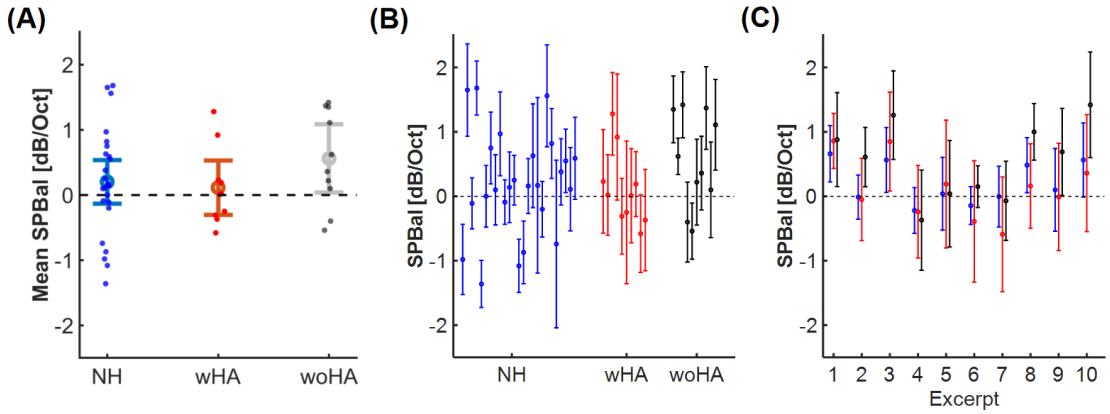


Figure 2.5: (A) Means and respective 95 % confidence interval plots of mean spectral balance about 1 kHz (SPBal) preferences taken over all 10 excerpts. Dots correspond to individual data of each participant within the groups. (B) Means and respective 95 % confidence interval plots of SPBal preferences taken for each participant over the 10 excerpts within the groups. (C) Means and the respective confidence interval plots of SPBal preferences averaged across participants within each group for each of the 10 excerpts presented.

C), where one can observe drastic differences between preferences of NH participants, but relatively small differences across excerpts.

As for the EQ-transform, the mean preferences illustrated in Figure 2.6(A) for wHA (M = 83.5 %, SD = 33.8 %), and woHA participants (M = 89.1 %, SD = 43.5 %) were not significantly different from factory settings ($t(9) = 1.5$, $p = 0.15$ and $t(9) = 0.8$, $p = 0.4$ respectively). However, that for NH participants (M = 80.5 %, SD = 28.3 %) showed significantly reduced EQ transform preferences compared that presented to them in the original mix, $t(24) = 3.5$, $p = .002 < .01$, $d = 0.7$ (medium effect).

The LME model used did not show an effect of participant group, $F(2,420) = 0.26$, $p = .8$. However, inter-excerpt effects $F(9,420) = 6$, $p < .001$ and interaction effects $F(18,420) = 1.8$, $p = .02 < .05$ were observed. Barring non-significant group effects, interesting trends can be observed here as illustrated in Figure 2.6(A & B). It is evident that the participants from all of the groups mostly preferred under-mixing or more specifically a transform between 0 and 100%. All participant groups showed similar mean EQ-transform preferences with no significant differences between their

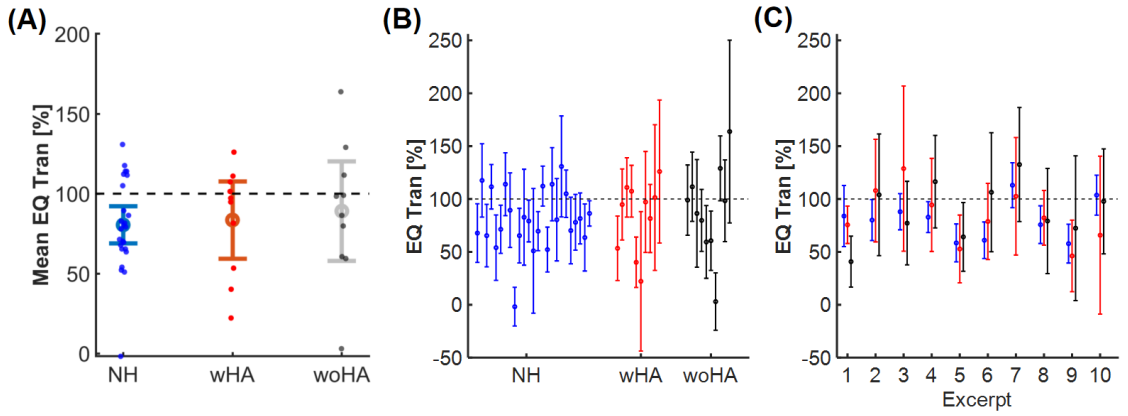


Figure 2.6: (A) Means and respective 95 % confidence interval plots of mean EQ-transform preferences taken over all 10 excerpts. Dots correspond to individual data of each participant within the groups. (B) Means and respective 95 % confidence interval plots of EQ-transform preferences taken for each participant over the 10 excerpts within the groups. (C) Means and the respective confidence interval plots of EQ-transform preferences averaged across participants within each group for each of the 10 excerpts presented.

variances. Furthermore, no significant individual differences were observed within the groups ($p > .09$). Refer supplementary material for the illustration of p-values for inter-excerpt comparisons of the respective preferences for all the 3 groups pooled in Experiment 1[5]. An illustration of raw error residuals from the LME model used are shown in the supplementary material[6].

In summary, we find that the wHA participants preferred a significantly elevated level of the lead vocals in the mixes presented to them compared to the NH participants. These preferences were significantly more diverse among the woHA participants than among the NH participants. When spectral balance preferences of the mixes were assessed, there were significant individual differences among NH participants. On average, wHA participants preferred the factory settings in the mixes with an almost 0 dB/Oct preference. However woHA participants favored weighting higher frequencies in the mixes more by way of significantly elevated SPBal preferences than factory settings. All three participant groups preferred spectrally denser

[5]See section 3 of the supplementary material at Appendix A for an illustration of p-values from inter-excerpt comparisons in Experiment 1.

[6]See section 4 of the supplementary material at Appendix A for the error residual plots of the LME (Linear Mixed Effects) model used.

mixes than those presented to them by way of an EQ-transform preference below 100%. This observation was significant for NH participants.

A clear limitation of Experiment 1 was that the degree of hearing-loss confounded the between-subjects distinctions of hearing aid use, see Fig. 3(B). For that reason, we sought to follow up on these findings using a more controlled within-subjects design, where the mixing effect preferences were assessed with and without HA use among a sample of HI participants different from that in Experiment 1. Here from, we aimed at assessing the role of hearing aid use alone on preferences of music processing strategies.

2.6 Experiment 2

This experiment was identical to the previous in setup and was implemented to evaluate the effect of hearing aid use on the mixing effects preferences. To that end, this experiment only targeted participants who were bilateral hearing aid users who completed the experiment once with (wHA) and once without (woHA) their hearing aids on.

2.6.1 Methods

Participants

A sample of 18 participants with a mean age of 73 years participated in this experiment. 14 of them were moderate to severely HI (M = 50 dB HL, SD = 7 dB HL) and had a mean age of 75 years. Only four participant were mild HI (M = 33 dB HL, SD = 5 dB HL) with a mean age of 67 years. Participants with bilateral hearing aids were specifically recruited via a subjects database from Hörzentrum gGmbH. There were 12 male and 6 female participants. Musical training estimated as in the first experiment was on average 18 points on the Gold-MSI musical training subscale (Müllensiefen et al., 2014) (SD = 9). Figure 2.7(A) shows the median hearing loss of

the participants evaluated through puretone audiometry. Similar to that observed with the participant pool in Experiment 1, there was a significant linear correlation between age and mean hearing loss, $r(16) = 0.7$, $p = .003 < .01$, as visible in Figure 2.7(B).

Sixteen of the participants wore BTE type HA. Among these participants, twelve of them wore open-fit type HAs with a reported duration of use between 1 and 23 years ($M = 9$ yrs, $SD = 5.4$ yrs). Four of them wore closed-fit type HAs with a reported duration of use between 5 and 18 years ($M = 13.5$ yrs, $SD = 5.8$ yrs). Only one participant wore a full shell in-the-ear (ITE) type HA with a reported length of use spanning 35 years. Data pertaining to the hearing aid use from only one remaining participant was unavailable upon collection post-study. Refer supplementary material for individual information pertaining to the HA use[7].

Stimuli

Here, a total of 60 distinct tracks (10 per block) from the Medley dB (Bittner et al., 2014) and Cambridge-MT (Senior, 2010) databases were used to provide distinct tracks for each participant to average out excerpt specific biases. See the supplementary material for further information [8].

Procedure

Unlike in the previous experiment where HA users were advised to wear their hearing aids for the entire duration of the experiment, here, the participants were asked to leave their HA on in a given phase consisting of the ten trials for each of the three mixing effects and then remove them in a second phase of the experiment. Whether the participant was to wear their hearing aids in a former or latter phase of the experiment was counterbalanced across participants. Furthermore, the blocks and

[7]See section 6.2 of the supplementary material at Appendix A for information pertaining to HA use among wHA in Experiment 2.

[8]See section 2 of the supplementary material at Appendix A for the list of audio excerpts used in experiment 2.

the choice of the respective audio excerpts were completely randomized.

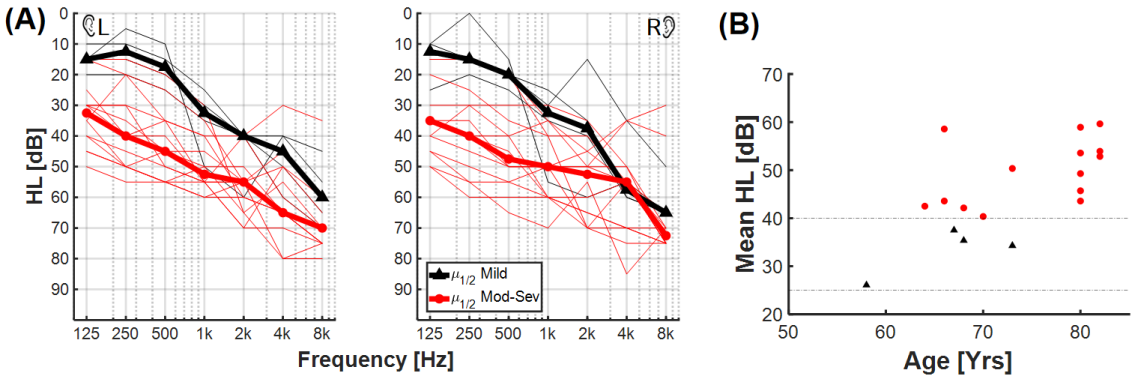


Figure 2.7: (A) Audiograms measured for the participants in Experiment 2 using pure tone audiometry for left and right ears. Thick lines indicate median hearing loss. (B) Mean HL and participant age. Dashed lines indicate thresholds for mild and moderate to severe hearing impairments respectively.

2.6.2 Results and Discussion

Figures 2.8(A-C) illustrate the results from Experiment 2. A paired t-test was performed to assess significant differences of preferences in the wHA and woHA conditions. For LAR, the test indicated significantly elevated average preferences without HA ($M = 13$ dB, $SD = 8$ dB) than with HA ($M = 8$ dB, $SD = 8$ dB), $t(17)$= 2.4, $p = .028 < .05$, $d = 0.6$ (medium effect). This trend is also seen in spectral balance preferences where an elevated SPBal without HA use (M = 0.64 dB/Oct, SD = 1 dB/Oct) than with HA use (M = -0.74 dB/Oct, SD = 1.2 dB/Oct), $t(17)$ = 3.62, $p = .002 < .01$, $d = 0.9$ (large effect) is apparent. Finally EQ-transform preferences were similarly elevated without HA use (M = 156 %, SD = 101.3 %) when compared to that with HA use (M = 74 %, SD = 70.4 %), $t(17)$= 2.3, p = .03 $< .05$, d = 0.5 (medium effect). That is, an increase from undermixing to overmixing in EQ-transform preferences can be observed with the removal of their HA. In tandem with the previously made assertion, this observation suggests that the removal of their HA resulted in the participants preferring spectrally sparser mixes.

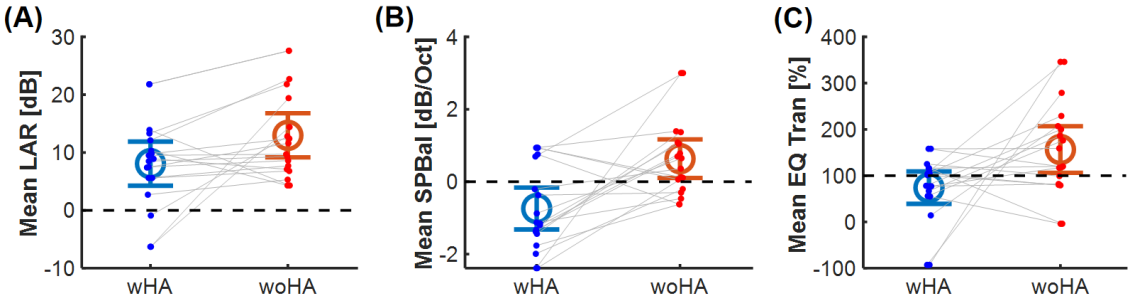


Figure 2.8: Preferences elicited from the Experiment 2 for (A) LAR, (B) SPBal, and (C) EQ-transform effects. Participants were tested either with (wHA) or without (woHA) their bilateral HAs.

To analyze the relationship between mixing effect preferences and the level of hearing loss, the data from both experiments were pooled by using the data of NH participants and woHA participants from Experiment 1 and the data of the woHA conditions from Experiment 2. Here, it can be observed further that the participants from the second experiment had around ($M = 46$ dB HL, $SD = 9$ dB HL) 17 dB HL higher hearing levels compared to the woHA participants the first experiment. A significant positive correlation between mean HL and LAR preferences was observed, $r(51) = 0.7$, $p < .001$. Mean HL and EQ-transform preferences were similarly positively correlated, $r(51) = 0.5$, $p < .001$ and a marginal correlation was observed for SPBal preferences $r(51) = 0.3$, $p = 0.06$. Figures 2.9(A-C) provide an illustration of the correlation in the pooled data between the two experiments. Taken together, the data from both experiments thus suggest a monotonic relationship between the degree of hearing loss and mixing effects preferences for LAR and the EQ-transform. With participants wearing HAs, somewhat expectedly, this relation does not hold any more since HAs arguably compensate the hearing-impairment. That is, it seems highly beneficial to seek different music processing strategies for HI participants with and without HAs.

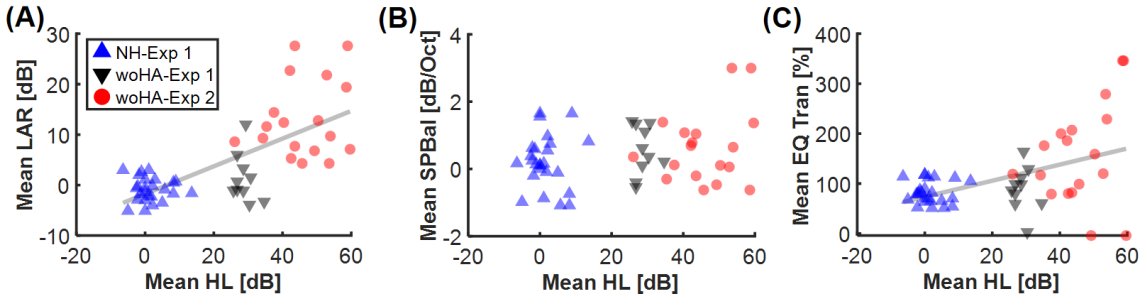


Figure 2.9: Linear correlations between mean HL and (A) LAR, (B) SPBal, and (C) EQ-transform preferences derived from NH data from Experiment 1 and woHA data pooled from both Experiment 1 and Experiment 2. Correlation lined indicate significant linear correlation.

2.7 General Discussion

In this study, we sought to establish the preferences with regards to basic mixing effects in hearing impaired individuals. Therefore, individual preferences of the lead-to-accompaniment ratio (LAR), spectral balance, and EQ-transform effects were assessed in a sample of normal hearing and hearing impaired listeners. In Experiments 1, the HI participants were grouped into those who did and those who did not wear bilateral hearing aids. We observed that HI participants with and without bilateral hearing aids preferred an increased LAR compared to NH participants. We did not observe pronounced effects of spectral balance between groups, but rather substantial individual differences for NH participants. The EQ-transform implemented in this study (linearly extrapolating between available EQ in the mix and a reference spectrum) was shown to significantly affect frequency domain or spectral sparsity measured here using the Gini-index. Results showed that all three participant groups preferred under-mixing or an EQ-transform setting of less than 100% on average (i.e., less sparse than the original mix). Yet, this observation was only significant among the NH participants who preferred mean EQ-transforms of 20% below the factory settings. In Experiment 2 targeting only participants with bilateral hearing aids, preferences were recorded with and without their HA. The use of HA resulted in a 5 dB reduction in LAR. Moreover, HA use also yielded a significant reduction in the spectral balance preference. A -0.7 dB/Oct balance

favoring low frequency with HA use and a similar positive balance of around + 0.7 dB/Oct favoring high frequencies with no HA use was observed. When the NH and woHA data from Experiment 1 and the woHA data from Experiment 2 were pooled, a significant positive correlation between both LAR and EQ-transform preferences with respect to the mean hearing loss was observed. These results suggest that with increasing hearing loss, participants had a greater affinity towards louder lead vocals in the mix. Moreover, with increasing hearing loss, spectrally sparser mixes were also favored.

From Experiment 1, it was evident that on average, HI participants from both groups preferred a LAR of 2 dB with a statistically significant difference between NH and wHA participants. According to Pons et al. (2016), a small sample of cochlear implant users (CI) on average preferred a instrument to vocals ratio of -1.92 dB (translating into a 1.92 dB LAR), similar to that found here. This also underpins similar findings by Buyens et al. (2014). In the present experiment, the NH participant however preferred the lead vocals to be merely a decibel lower than that of the accompaniment, consistent with other recent findings (Tahmasebi et al., 2020). In Experiment 2, wHA participants also preferred the lead vocals louder than the accompaniment, but even more so when the HA were not used (yielding an increase of around 5 dB). Here, the fact that wHA participants preferred the LAR to be 8 dB louder than each of the accompaniment is similar to that shown by Tahmasebi et al. (2020) for CI users. Together with the within-subjects comparison of Experiment 2, our results suggest that the LAR in its effect on vocal level is an important feature to consider for adjusting music mixes for mild-to-severely unaided HI listeners, with a tendency of higher preferred LARs by listeners with higher degrees of hearing loss. This appears to be very plausible, given the exceptional status of vocals in popular music and their role in conveying the lyrics of songs (Condit-Schultz and Huron, 2015).

Spectral balance preferences dictate the frequency weighting of the audio excerpt, the perceptual effects of which have been described in terms of brightness perception (Saitis and Siedenburg, 2020). From the first experiment, it was observed that the woHA preferred an elevated spectral balance favoring higher frequencies in the mix. Both NH and wHA preferred spectral balance of 0 dB/Oct on average which was the unperturbed factory mix. Although NH participants also preferred similar settings of spectral balance, their choices were varied and the mean participant-specific choices over excerpts bore a significantly greater variance than the mean excerpt or condition specific choices. The strength of these individual differences is rather surprising, given that previous work showed rather steep psychometric functions of spectral balance in the range of -2 to 2 dB (Siedenburg et al., 2021a). Experiment 2 followed up on this observation by revealing a significant reduction of spectral balance values due to hearing aid use: The wHA preferred negative spectral balances (-0.7 dB/Oct) on average. When they removed their HA, a commensurately positive value of +0.7 dB/Oct was preferred on average. According to Thrailkill et al. (2019), loudness perception among hearing aid users with sensorineural hearing impairment is dominated by higher frequencies and did not change with their experience with these devices. A reduced spectral balance preference among such listeners as shown here, may highlight the fact that they may counter the compensation of high-frequency hearing loss brought on by the hearing aids.

The EQ-transform implementation discussed here was shown to bring about significant changes in spectral sparsity in the multi-track mixes. Although we did not observe statistically robust effects in Experiment 1, Experiment 2 showed a significant elevation in EQ-transform preferences when HA were removed, implying that HA users showed preferences towards spectrally denser mixes. Furthermore, correlation of the pooled data showed that there was a significant positive correlation between mean HL and preferred EQ-transform settings, indicating a preferences of greater spectral sparsity with increasing mean hearing loss levels. Recent studies (O'Grady et al., 2005; Abdulla and Jayakumari, 2022) have demonstrated that

spectro-temporal sparsity appears to be a pivotal factor in source separability in that sparse representations in both time or frequency domain improve the performance of blind source separation algorithms. Here, we first observed that HI listeners may prefer higher levels of spectral sparsity and with the so-called EQ-transform we presented an algorithm that yields according changes of multi-track mixes.

We acknowledge that a major limitation of this study was not to consider a more comprehensive full input/output characterization of hearing aids including insertion gains, which may have allowed us to derive a deeper understanding of some of the various auditory mechanisms involved in the perception of multi-track music. However, our primary goal was to suggest and explore novel strategies for re-mixing music that may lead to higher degrees of music appreciation among hearing-impaired subjects. Furthermore, this study does not address the issues pertaining to individual differences in loudness growth functions (e.g., Marozeau and Florentine, 2007). Specifically, Florence et al. (2017) showed that HI participants had steeper loudness growth than NH participants for sounds of low to moderate intensity, even when the former were fitted with compression HAs aimed at compensating for the reduced compressive nonlinearity of the cochlea as a result of sensorineural hearing impairment. Evaluating the loudness growth curves of participants in this study may have provided additional insight into individual preferences of the mixing effects. Finally, we wish to acknowledge that this study did not specifically control for loudness saturation in the HAs, brought on by possibly high crest factors of music signals. This can be a critical issue for music listening with HAs and is especially problematic in older HAs with narrower input headroom. It should be noted, however, that the maximal presentation levels of our stimuli did not appear to be problematic (see the supplementary materials), so that we do not expect hearing aid saturation as a critical issue in the present study, even though it is certainly an important factor to consider for real-life music listening with HAs.

Overall, with this study we made a first attempt to explore mixing preferences of NH and HI listeners with and without HAs. Despite substantial individual differences among NH and HI listeners, we observed consistent choices of mixing parameters that extend previous work on CI listeners (Buyens et al., 2014; Tahmasebi et al., 2020) towards mild to moderately HI listeners. Furthermore, with the so-called EQ-transform, a straight-forward spectral effect was introduced that appears to be a promising tool for the individualization of multi-track mixes. In follow-up studies, we seek to test objective performance of HI and NH listeners in music scene analysis tasks in order to assess the effects of our implementation on source transparency. Particularly, the participant's ability to determine if a cued instrument or vocal was present within a given excerpt of a mix will be assessed with different EQ transform settings yielding significant differences in spectral sparsity, as shown in this study. Furthermore, we seek to explore whether the combination of mixing effects, which in this study have only been tested in isolation, may provide synergistic results, which could provide participants with a richer palette of potential audio manipulations to adjust according to their preferences.

2.8 Conclusion

The main contribution of this study was to evaluate music mixing preferences in a sample of HI listeners with and without HAs. Besides suggesting that previous findings on LAR preferences of CI listeners extend towards HA users, it was also shown that there are distinct preferences regarding the setting of spectral mixing effects. Importantly, with the EQ-transform we proposed a new spectral transformation of multi-track music signals. Our results suggest that HI listeners prefer spectrally sparser mixes as evinced by preferences towards increased EQ-transform settings. Generally, our findings indicate that the individualization of both level- and spectral-based mixing effects may yield enhanced music appreciation for listeners with hearing loss.

2.9 Acknowledgments

The authors would like to thank all of the participants for having participated in this study and Hörzentrum Oldenburg gGmbH for their support. This study was funded by a Freigeist Fellowship to KS from the Volkswagen Stiftung.

References

Abdulla, S. M. and Jayakumari, J. (2022). Improving time–frequency sparsity for enhanced audio source separation in degenerate unmixing estimation technique algorithm. *Journal of Control and Decision*, pages 1–14.

Aichinger, P., Sontacchi, A., and Schneider-Stickler, B. (2011). Describing the transparency of mixdowns: The masked-to-unmasked-ratio. In *Audio Engineering Society Convention 130*. Audio Engineering Society.

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. P. (2014). Medleydb: A multitrack dataset for annotation-intensive mir research. In *ISMIR*, volume 14, pages 155–160.

Bürgel, M., Picinali, L., and Siedenburg, K. (2021). Listening in the mix: Lead vocals robustly attract auditory attention in popular music. *Frontiers in psychology*, page 6117.

Buyens, W., van Dijk, B., Moonen, M., and Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International journal of audiology*, 53(5):294–301.

Case, A. (2011). *Mix smart: Pro audio tips for your multitrack mix*. Focal Press.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *Asha*, 23(7):493–500.

Condit-Schultz, N. and Huron, D. (2015). Catching the lyrics: intelligibility in twelve song genres. *Music Perception: An Interdisciplinary Journal*, 32(5):470–483.

De Man, B., Stables, R., and Reiss, J. D. (2019). *Intelligent Music Production*. Routledge.

Florence, J., Prakash, P. H., Bhargavi, P., Krishna, Y., and Bellur, R. (2017). Comparison of loudness growth function in normal hearing individuals and impaired aided hearing. *Advanced Science Letters*, 23(3):1946–1948.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (1980). Frequency selectivity in normally-hearing and hearing-impaired observers. *Journal of Speech, Language, and Hearing Research*, 23(3):646–669.

Garofolo, J. S. (1993). Timit acoustic phonetic continuous speech corpus. *Linguistic Data Consortium, 1993*.

Gillet, O. and Richard, G. (2006). Enst-drums: an extensive audio-visual database for drum signals processing. In *International Society for Music Information Retrieval Conference (ISMIR)*.

Glasberg, B. R. and Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79(4):1020–1033.

Greasley, A., Crook, H., and Fulford, R. (2020). Music listening and hearing aids: perspectives from audiologists and their patients. *International Journal of Audiology*, 59(9):694–706.

Hafezi, S. and Reiss, J. D. (2015). Autonomous multitrack equalization based on masking reduction. *Journal of the Audio Engineering Society*, 63(5):312–323.

Hopkins, K. and Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, 130(1):334–349.

Hornsby, B. W. and Ricketts, T. A. (2006). The effects of hearing loss on the contribution of high-and low-frequency speech information to speech understand-

ing. ii. sloping hearing loss. *The Journal of the Acoustical Society of America*, 119(3):1752–1763.

Hurley, N. and Rickard, S. (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741.

Izhaki, R. (2017). *Mixing audio: concepts, practices, and tools*. Routledge.

Jillings, N. and Stables, R. (2017). Investigating music production using a semantically powered digital audio workstation in the browser. In *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE.

Knoll, A. L. and Siedenburg, K. (2022). The optimal mix? presentation order affects preference ratings of vocal amplitude levels in popular music. *Music & Science*, 5:20592043221142712.

Kohlberg, G. D., Mancuso, D. M., Chari, D. A., and Lalwani, A. K. (2015). Music engineering as a novel strategy for enhancing music enjoyment in the cochlear implant recipient. *Behavioural neurology*, 2015.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, 4:863.

Lohse, K. R., Shen, J., and Kozlowski, A. J. (2020). Modeling longitudinal outcomes: A contrast of two methods. *Journal of motor learning and development*, 8(1):145–165.

Ma, Z. (2016). *Intelligent Tools for Multitrack Frequency and Dynamics Processing*. PhD thesis, Queen Mary University of London.

Madsen, S. M. and Moore, B. C. (2014). Music and hearing aids. *Trends in Hearing*, 18:2331216514558271.

Marozeau, J. and Florentine, M. (2007). Loudness growth in individual listeners with hearing losses: A review. *The Journal of the Acoustical Society of America*, 122(3):EL81–EL87.

Moffat, D. and Sandler, M. B. (2019). Approaches in intelligent music production. In *Arts*, volume 8, page 125. MDPI.

Moylan, W. (2014). *Understanding and crafting the mix: The art of recording*. Routledge.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642.

Nagathil, A., Weihs, C., Neumann, K., and Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners. *The Journal of the Acoustical Society of America*, 142(3):1219–1228.

O'Grady, P. D., Pearlmutter, B. A., and Rickard, S. T. (2005). Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*, 15(1):18–33.

Orović, I., Stanković, S., Beko, M., et al. (2022). On the use of gini coefficient for measuring time-frequency distribution concentration and parameters selection. *Mathematical Problems in Engineering*, 2022.

Pons, J., Janer, J., Rode, T., and Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, 140(6):4338–4349.

Reiss, J. D. (2011). Intelligent systems for mixing multichannel audio. In *2011 17th International Conference on Digital Signal Processing (DSP)*, pages 1–6. IEEE.

Reiss, J. D. (2016). An intelligent systems approach to mixing multitrack audio. In *Mixing Music*, pages 246–264. Routledge.

Rickard, S. (2006). Sparse sources are separated sources. In *2006 14th European signal processing conference*, pages 1–5. IEEE.

Rickard, S. and Fallon, M. (2004). The gini index of speech. In *Proceedings of the 38th Conference on Information Science and Systems (CISS'04)*.

Ronan, D., Ma, Z., Namara, P. M., Gunes, H., and Reiss, J. D. (2018). Automatic minimisation of masking in multitrack audio using subgroups. *arXiv preprint arXiv:1803.09960*.

Saitis, C. and Siedenburg, K. (2020). Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories. *The Journal of the Acoustical Society of America*, 148(4):2256–2266.

Schafer, R. W. (2011). What is a savitzky-golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117.

Senior, M. (2010). The 'Mixing Secrets' Free Multitrack Download Library. https://cambridge-mt.com/ms/mtk/. [Online; accessed 2022-08-10].

Siedenburg, K., Barg, F. M., and Schepker, H. (2021a). Adaptive auditory brightness perception. *Scientific reports*, 11(1):1–11.

Siedenburg, K., Goldmann, K., and Van de Par, S. (2021b). 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*, page 1156.

Siedenburg, K., Röttges, S., Wagener, K. C., and 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.

Siegel, S. (1956). *Nonparametric statistics for the behavorial sciences*. McGraw-Hill.

Smits, J. and Duijhuis, H. (1982). Masking and partial masking in listeners with a high-frequency hearing loss. *Audiology*, 21(4):310–324.

Steinmetz, C. J., Pons, J., Pascual, S., and Serrà, J. (2021). Automatic multitrack mixing with a differentiable mixing console of neural audio effects. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 71–75. IEEE.

Tahmasebi, S., Gajcki, T., and Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Frontiers in Neuroscience*, 14:434.

Thrailkill, K. M., Brennan, M. A., and Jesteadt, W. (2019). Effects of amplification and hearing-aid experience on the contribution of specific frequency bands to loudness. *Ear and hearing*, 40(1):143.

Tom, A., Reiss, J. D., and Depalle, P. (2019). An automatic mixing system for multitrack spatialization for stereo based on unmasking and best panning practices. In *Audio Engineering Society Convention 146*. Audio Engineering Society.

VanderWeele, T. J. and Mathur, M. B. (2019). Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? *American journal of epidemiology*, 188(3):617–618.

Warnecke, M., Peng, Z. E., and Litovsky, R. Y. (2020). The impact of temporal fine structure and signal envelope on auditory motion perception. *Plos one*, 15(8):e0238125.

Zonoobi, D., Kassim, A. A., and Venkatesh, Y. V. (2011). Gini index as sparsity measure for signal reconstruction from compressive samples. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):927–932.

Zurek, P. and Formby, C. (1981). Frequency-discrimination ability of hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 24(1):108–112.

2.10 Summary

- In order to ascertain multi-track mixes created using standardized *best practices* are suitable for individuals with mild to moderate hearing loss, two experiments were conducted.
- In both experiments, the participants elicited their subjective preference for level and spectrally modified mixes.
- In experiment 1, normal-hearing (NH), hearing-impaired bilateral hearing-aid users (wHA), and non-users (woHA) were tested.
- wHA participants preferred elevated levels of the lead vocals in the mixes than NH.
- Preference of lead vocal levels were observably more varied among woHA compared to NH.
- woHA favored mixes with greater energy weightings at high frequencies (> 1 kHz).
- All three participant groups preferred mixes with constituent tracks that were of reduced spectral contrast compared to the original. This observation was most pronounced for NH.
- In the more controlled experiment 2, similar preferences were assessed among a matched sample of hearing-aid users with (wHA) and without (woHA) bilateral hearing aids.
- Hearing-aid disuse was associated with the preference for higher levels of lead vocals, greater energy weightings at high frequencies, and tracks with a sparser frequency domain representation or a more exaggerated spectral contrast.
- Based on the results from NH and woHA from both experiments, increasing hearing thresholds were associated with the preference for higher lead vocal

levels and spectral contrast.

- The study shows that spectral and level adjustments to music mixes may prove to be beneficial for hearing-impaired listeners.
- Additionally, bilateral hearing aids tend to have a noticeable effect on the music mixing preferences.
- The increasing preference for a more pronounced spectral contrast with higher hearing thresholds raises the question of whether these modifications could potentially improve musical scene analysis in moderately hearing-impaired listeners.

3. Effects of spectral manipulations of music mixes on musical scene analysis abilities of hearing-impaired listeners

Given the positive association between hearing thresholds and spectral contrast by way of higher EQ-transform preferences observed in the first study, the study in this chapter aims to assess the effects of the EQ-transform on musical scene analysis. The focus here is to evaluate the benefits of spectral modifications on the selective listening abilities of moderately hearing-impaired listeners for multi-track musical scenes. Furthermore, influence of the type of instrument in music mixes on the scene analysis abilities are also investigated.

3.1 Study 2

The study included in this chapter was published as: Benjamin AJ, 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*.

The content of this chapter is identical to the published work.

Author Contributions: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the necessary experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

_____ _____

(name) Date

Supervisor

3.2 Abstract

Music pre-processing methods are currently becoming a recognized area of research with the goal of making music more accessible to listeners with a hearing impairment. Our previous study showed that hearing-impaired listeners preferred spectrally manipulated multi-track mixes. Nevertheless, the acoustical basis of mixing for hearing-impaired listeners remains poorly understood. Here, we assess listeners' ability to detect a musical target within mixes with varying degrees of spectral manipulations using the so-called EQ-transform. This transform exaggerates or downplays the spectral distinctiveness of a track with respect to an ensemble average spectrum taken over a number of instruments. In an experiment, 30 young normal-hearing (yNH) and 24 older hearing-impaired (oHI) participants with predominantly moderate to severe hearing loss were tested. The target that was to be detected in the mixes was from the instrument categories Lead vocals, Bass guitar, Drums, Guitar, and Piano. Our results show that both hearing loss and target category affected performance, but there were no main effects of EQ-transform. yNH performed consistently better than oHI in all target categories, irrespective of the spectral manipulations. Both groups demonstrated the best performance in detecting Lead vocals, with yNH performing flawlessly at 100% median accuracy and oHI at 92.5% (IQR = 86.3 - 96.3%). Contrarily, performance in detecting Bass was arguably the worst among yNH (Mdn =67.5% IQR = 60 - 75%) and oHI (Mdn =60%, IQR = 50 - 66.3%), with the latter even performing close to chance-levels of 50% accuracy. Predictions from a generalized linear mixed-effects model indicated that for every decibel increase in hearing loss level, the odds of correctly detecting the target decreased by 3%. Therefore, baseline performance progressively declined to chance-level at moderately severe degrees of hearing loss thresholds, independent of target category. The frequency domain sparsity of mixes and larger differences in target and mix roll-off points were positively correlated with performance especially for oHI participants (r = .3, $p < .01$). Performance of yNH on the other

hand remained robust to changes in mix sparsity. Our findings underscore the multifaceted nature of selective listening in musical scenes and the instrument-specific consequences of spectral adjustments of the audio.

3.3 Introduction

What makes a good musical mix for listeners with a hearing loss? This seemingly straight-forward question hosts a plethora of questions in music production and psychoacoustics that research is only beginning to address. In fact, sounds from musical instruments tend to overlap substantially in time and frequency in music mixes and listeners with different hearing abilities (and of potentially different age ranges) vary in their ability to separate sound sources perceptually (Hake et al., 2023). Yet, little is known about how properties of the mix affect this process. Neither do we have substantiated knowledge on whether different groups of listeners such as listeners with and without hearing loss react in similar or dissimilar ways to manipulations of a musical mix. Here, we attempted to approach these questions by devising an experiment based on selective listening in musical mixtures, wherein listeners were tasked to detect a cued target sound in musical mixtures that were manipulated in the frequency domain.

Selective listening has been a thoroughly researched field within the context of auditory scene analysis (ASA) (Bregman, 1994). A very large part of research dealing with selective attention has been performed on the so called *'cocktail party problem'* (Bee and Micheyl, 2008). Here, the perceptual processes in a receiver involved in tracking and understanding one speaker amid competing speakers in a setting similar to a cocktail party are of focus (Bronkhorst, 2015). The terminology was coined by Collin Cherry who initially conducted experiments where participants were tasked separating different speech signals presented diotically (Cherry, 1953); he showed that separability of the signals in the presence of background noise depended upon the rate of the speech, its direction of arrival, the participants' gen-

der, and average pitch of the speech signals. However, given the focus on speech perception, the task of detecting and tracking musical targets in the presence of accompanying musical maskers remains underexplored (i.e., Musical Scene Analysis or MSA tasks), especially within the context of sensorineural hearing impairment.

3.3.1 Previous work

The study by Siedenburg et al. (2020) was among the first to investigate MSA ability as a function of hearing impairment. Here, melody and timbre discrimination abilities of a sample of young normal-hearing and older hearing-impaired listeners were investigated. In the melody discrimination task, first a reference signal which was a clarinet target along with Piano , Cello, or noise maskers was presented. This was followed by two, one second Clarinet excerpts with varying pitch sequences. The participants were tasked with detecting which one of the two Clarinet excerpts was in the reference. Similarly, in the timbre task, target and masker references with the same maskers were presented and followed by successive Trumpet or Flute excerpts with identical pitch sequences. Here, the participants had to determine which of the two instruments was in the reference. It was shown that the older hearing-impaired participants required signal-to-masker ratios 10 dB greater on average compared to the young normal-hearing participants. However, musical training among both groups brought about a reduction to these requirements. The older hearing-impaired listeners also did not utilize level drops in the maskers in both tasks unlike normal-hearing listeners. In a later study by Siedenburg et al. (2021), four target musical voices taken from Johann Sebastian Bach's *'The Art of the Fugue'* were presented through four spatially separated loudspeakers. The number of voices presented at a given time was varied between two and four. The cued voice was played first, and a tremolo by way of amplitude modulation was applied or not applied in a given trial and presented through one of the four loudspeakers along with the other voices. The participants were then tasked with detecting if a tremolo was present in the cued

voice in the subsequent presentation. It was observed that the timbral homogeneity of the instrumentation, directly related to spectral similarity of the target sounds with the accompanying voices. This in turn had detrimental effects on performance compared to a heterogeneous instrumentation.

In a study investigating MSA abilities among normal-hearing participants, Bürgel et al. (2021) showed that Lead vocals prevailed in attracting auditory attention when presented amid coherent multi-track mixes of popular music. Furthermore, detection performance of the participants depended on the order in which the target and the mix were presented in the trials and the target instrument category (e.g., Lead vocals, Guitar, Piano etc.). In a follow up study (Bürgel and Siedenburg, 2023), when frequency micro-modulations of Lead vocals were applied to other target instruments in a similar task, the effect of presentation order seen previously was mitigated, suggesting a role for frequency micro-modulations in auditory salience. More recently, Hake et al. (2023) developed a viable open source tool to assess MSA ability with the aid of a scoring system for listeners with varying degrees of hearing impairment. Here it was shown using a large sample of normal-hearing and hearing-impaired participants that detecting the presence of a cued target instrument in a mixture depended greatly upon the level ratios between the target and the mixture, the complexity of the mixture (i.e., the number of instruments), and the target instrument category. Stereo panning width of the individual instruments in the mix however had a smaller impact on the task performance. More importantly, they showed that hearing-impaired listeners with severe to profound hearing loss had significantly lower MSA abilities. However, the effect of spectral changes of the instruments in the mixture were not assessed.

3.3.2 Motivation

Despite the alarming rise in individuals living with sensorineural hearing impairment as per the WHO projection of over 900 million people by year 2050 (Davis and Hoffman, 2019), research into music processing methods among them remains surprisingly scarce. The music industry relies heavily upon mixing and mastering practices of trained professionals. Here, the so called mixing engineer is required to combine raw recordings of a myriad of instruments and vocals made available through separate tracks into a coherent mixdown or mix, which has been subjected to meticulous spectral and temporal manipulations (Case, 2011). The mix unlike its raw constituents should bear optimal transparency of the individual tracks while matching the aesthetic intentions of the artists.

With regards to mixing preferences, Nagathil et al. (2017) showed that cochlear-implant (CI) users preferred reduced spectral complexity in classical chamber music which was accomplished through a low-rank approximation of its constant-Q-transform. Using a music re-mixing task, Hwa et al. (2021) showed that CI users preferred an average increase of 7.1 dB in bass frequencies and 6.7 dB in treble frequencies compared to the original stimuli. The bass and treble frequencies were defined as those below and above the 50th percentile of the frequencies in the stimuli, respectively. Nevertheless, in a more recent study by Althoff et al. (2024), which utilized a similar remixing task, no significant differences in low/high pass filtering preferences for instrumental music were observed between CI users and normal-hearing controls. However, Benjamin and Siedenburg (2023) showed that spectral manipulations via the transformed equalization or EQ-Transform performed on constituent tracks of multi-track mixes, increased their objective frequency-domain sparsity by way of higher Gini indices. Importantly, hearing aid users, with mild to moderate hearing loss preferred mixes with spectrally sparser tracks. Sparser time-frequency representations have been postulated at overcoming the cocktail party problem (Asari et al., 2006). Furthermore, the effectiveness of source separation

algorithms, improves for sparse representations of music (Plumbley et al., 2009). From a psychoacoustical standpoint, cochlear hearing impairment has been shown to give rise to broader auditory filters and reduced frequency selectivity which may in turn depreciate the ability to separate sounds closer in frequency (Glasberg and Moore, 1986; Florentine et al., 1980) even when the impairment was mild (Rasidi and Seluakumaran, 2024). This assertion was underpinned in speech perception by Gaudrain et al. (2007) where normal-hearing listeners under simulated hearing loss performed better at identifying the order of a vowel sequence when the component vowels were smeared in the frequency domain. Importantly, studies by Lentz and Leek (2003) and Narne et al. (2020) showed that hearing-impaired listeners have a reduced ability to perceive changes in spectral shape.

Therefore, the specific question addressed by this study is whether spectral manipulations to music can be employed as effective means to enhance scene analysis performance among hearing-impaired listeners. To investigate this question within the context of music perception, we aim at assessing the ability of a participant at detecting a musical target in the presence of musical maskers in a mix, as a function of their level of hearing loss, musical training, and the degree of spectral manipulation applied to both target and the accompanying maskers in the mix. In other words, this work aims at assessing how alterations to the power spectral variation of popular music, affect the MSA abilities of listeners with cochlear hearing loss in multi-track musical scenes. Based on the implications of the EQ-transform on hearing-impaired listeners as shown in (Benjamin and Siedenburg, 2023), it will be used to bring about such alterations in this work. To control for other effects on MSA as demonstrated by Hake et al. (2023), complexity and level differences between target and mix were kept constant. For simplicity, we will refer to (Benjamin and Siedenburg, 2023) as our earlier work throughout this manuscript.

3.4 Methods

In this section we outline the participants recruited for the study and the equipment and stimuli used. We then describe the experiment design used to assess MSA performance among the participants. Furthermore, a brief overview of the EQ-transform will be given.

3.4.1 Participants

A sample of 30 young normal-hearing (yNH) participants and 24 older hearing-impaired (oHI) participants took part in this study. The yNH were recruited using an advertisement posted on-line and oHI were mostly recruited via Hörzentrum Oldenburg gGmbH (Oldenburg, Germany). The recruitment process started on the 9th of May, 2023 and ended on the 20th of September, 2023. All of the participants provided their informed consent in writing by signing a consent form, provided immediately upon participation. The participants could choose either a German or an English version of the consent form. After receiving their informed consent, we proceeded to assess their level of musical training using the Gold MSI musical training questionnaire (musical training subscale) proposed by Müllensiefen et al. (2014). Based on the assessment, the normal-hearing participants were significantly more musically trained than the hearing-impaired participants, $t(52) = 2.1$, $p = .04$, $d = 0.6$ (Medium effect). However, as apparent later in our analysis, musical training had no effect on MSA performance. Afterwards, the hearing loss levels (HL) using pure-tone audiometery at 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz frequencies were assessed for each of the participant using a portable audiometer for both ears. The mean hearing loss level (MHL) which was an arithmetic mean taken over all of the frequencies over both ears was used to categorise the participant groups as per the guidelines outlined in (Clark, 1981). Normal hearing were so classified with a MHL $\leq$ 25 dB (M = 6 dB HL, SD = 5.3 dB HL). Among the 24 oHI participants, three of them were classified as having mild hearing impair-

ment with 25 dB < MHL ≤ 40 dB (M = 35.2 dB HL, SD = 3.04 dB HL) and 21 of them had moderate to severe hearing impairment or MHL > 40 dB (M = 46.3 dB HL, SD = 4.3 dB HL). However, in this study, all of the oHI participants were grouped together (M = 45 dB HL, SD = 5.6 dB HL). Figure 3.1(A) shows an illustration of the hearing loss levels at the aforementioned frequencies for both participant groups. It can be observed from Figure 3.1(A) that the mean hearing loss among the hearing-impaired participants is around 40 dB greater than that of the normal-hearing participants.

The normal-hearing participants who took part in this study were relatively young with ages of 18 and 45 years (M = 26.6 yrs, SD = 6.1 yrs). The hearing-impaired participants on the other hand were significantly older and were between the ages of 26 and 82 years (M = 71 yrs, SD = 11.6 yrs). As all except one among the hearing-impaired participants were above the age of 50 years, in this study we use the abbreviation oHI (older hearing-impaired) to refer to this group. Among the oHI, 22 participants were bilateral behind-the-ear type hearing aid users with only one participant using bilateral in-the-ear type hearing aid. The last remaining participant had no history of hearing aid use. Figure 3.1(B) illustrates the linear relationship between the participant ages and their respective MHL. As illustrated, a significant linear correlation between age and hearing loss is apparent among yNH with a positive correlation between age and mean hearing loss $r(28) = .5$, $p = .003 < .01$. Hearing impaired participants showed a similar correlation only when considered altogether $r(22) = .5$, $p = .006 < .01$. However, when we disregard the youngest participant (i.e., < 50 yrs) with no history of hearing aid use, the correlation becomes non-significant ($r(21) = .26$, $p = .2$).

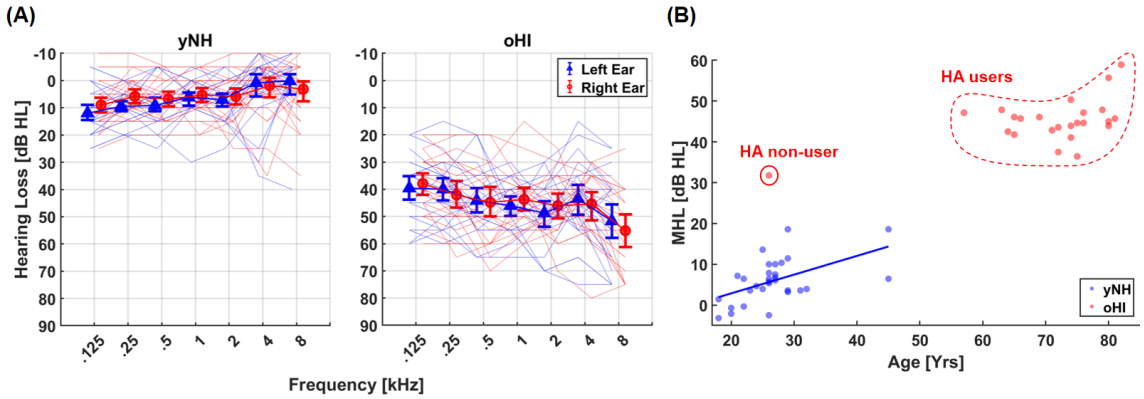


Figure 3.1: (A) Audiograms of the participant groups considered in this study mean hearing loss plotted in thick lines about bootstrapped 95% confidence intervals for the pure tone frequencies considered. The thin lines indicate individual audiograms. (B) The relationship between mean hearing loss level (MHL) and participants' ages. The significant linear correlation among the yNH is shown with a straight line. Among oHI participants, bilateral hearing aid (HA) users and the non-user are highlighted.

3.4.2 Stimuli and apparatus

All the audio excerpts used as stimuli were 2 seconds long and were taken from the Medley database (Bittner et al., 2014). The broadband target to mix level ratios were always kept constant at -10 dB. The number of vocals / instrument tracks in all of the mixes were kept at 5. The stimuli playback was presented over a pair of activ 8" near-field studio monitors by ESI Audiotechnik GmbH (Leonberg, Germany) in a low reflection chamber at the University of Oldenburg, Germany. The monitors were separated by a 90° angle and 2 m distance from the listener's seat. The overall playback level was calibrated to be 80 dBA (i.e., A-Weighted equivalent continuous sound pressure level over a measurement duration of 1 minute) at the participant position. This calibration was performed using white noise colored by the ensemble average spectrum of commonly occurring instrument classes and Lead vocals available in the Medley database. The processing on the stimuli were realized using a standalone desktop computer running MATLAB R2023a. This standalone machine was connected to the monitors with the aid of an RME Fireface UFX audio-interface. The puretone audiometry performed on all of the participants in this study was fulfilled with a portable AD528 audiometer from Interacoustics

(www.interacoustics.com/ad528).

3.4.3 Procedure

The experiment consisted of a detection task similar to that used by Hake et al. (2023). Here, the two-second target sound was presented first and after a second of silence, the two-second mix was presented. All stimuli were presented via a pair of loudspeakers in such a manner that the signals generated from both speakers were identical to avoid the influence of spatial cues on the detection task (Middlebrooks and Waters, 2020). All oHI participants using hearing aids were explicitly requested to take them off before the experiment.

The participant upon listening to the two excerpts, was tasked with identifying if the target sound was present in the mix through being prompted to click 'Yes' or 'No' to whether they heard the target in the mix presented to them as illustrated in Fig. 3.2(A). In any given trial where a target sound followed by a mix is presented, the target sound was taken from one of 5 different instrument or target classes: Lead vocals (Lead), Bass guitar (Bass), Drums, Guitar, and Piano. Altogether, 200 trials with distinct target and mix combinations were presented for the five target classes with 40 per target category. Within the 40 trials per target category, 20 trials contained the target and the remaining 20 trials did not contain the target. The 20 trials were then ramified into 4 sets of 5 trials with each set being subjected to a specific % EQ-transform (% EQ Tran) among the ones considered (i.e., 0 %, 100 % / Factory, 200 %, and 300 %). Figure 3.2(B), provides an illustration of EQ-transform and their implications on the power spectrum of a track to which the transform is applied. The transform estimates the transformed power spectrum by way of a linear extrapolation between the original or factory power spectrum (100 %) and a reference (0 %). The reference is an ensemble average spectrum taken over a number of tracks from a variety of instrument classes. The reference power spectrum is always energy normalized with respect to the factory spectrum under-

going the transform. By applying the 200 % EQ-transform on a track as shown in Figure 3.2(B), we essentially double the power level differences between the factory power spectrum and the reference in the transformed spectrum. A 300 % transform would commensurately triple this difference and so on. In our earlier work, we showed that for tracks taken from different instrument categories, frequency-domain sparsity sees a monotonic rise with increasing % EQ-Transform.

The % EQ-transform parameters used in the experiment were specifically chosen for they cover the range of those preferred by both yNH and oHI participants studied in our earlier work. The same % EQ-transform was applied to both the target and each track of the corresponding mix in a given trial. The order of the trials were randomized irrespective of target category, presence of the target in the mix, or % EQ-transform applied to avoid order biases in the trials presented between the participants.

3.4.4 Data analysis

A non-parametric approach was adapted here by way of evaluating bootstrapping sample means using 10^3 realizations with replacement (LaFontaine, 2021). A generalized linear mixed-effects model (GLME) was fitted on the data to estimate the main effects of the EQ-transform, mean hearing loss (MHL), the target instrument category, and the interaction between these effects on the dichotomous correct or incorrect responses to the trials. The model also accounted for the interaction between musical training scores and MHL. Furthermore, the model considered random intercepts to account for variability among the participants and the individual trials. The GLME was used as the analysis was performed on the dichotomous correct or wrong answer to those acquire from the participants (Parzen et al., 2011). Post-hoc analyses were conducted using the Mann-Whitney-U test (Nachar et al., 2008).

.

3.5 Results

Considering the proportion of correct responses over the total number of trials as a percentage (% Correct), overall median performance of yNH at 84.9% accuracy was around 5% better than oHI at 79.6%. Both groups performed best when detecting lead vocals with yNH performing at a perfect 100% median accuracy and oHI at 92.5%. The worst performance was observed for bass targets where, both yNH (67.5%) and oHI (60%) came closest to performing at a chance-level of 50%. Nevertheless, yNH consistently outperformed oHI across all target classes. The latter notwithstanding, both yNH and oHI were most accurate at factory settings (100 % EQ Tran), with median % correct responses of 88% and 81%, respectively. Figure 3.2(D) shows the distribution of % correct responses of yNH and oHI, under the different test conditions.

Here, we summarize the output of the GLME model fitted on our data, as classical ANOVA statistics for ease. These statistics are derived using MATLAB's `anova` function. Based on the model, musical training had neither an independent effect on performance (p=.98), nor did it interact significantly with MHL (p=.33). However, there was a main effect of MHL on performance $F(1,10434) = 24.4$, $p < .0001$, for which the model estimates an odds ratio of OR = 0.97 (95% CI :0.96-0.98). This means, for every unit increase in MHL (+1 dB HL), the model predicts a modest albeit progressive 3% drop in the odds of correctly detecting the target in the mix. The progressive effect of hearing loss on MSA performance can be observed in the conditional probability plots derived from the model, shown in Figure 3.2(C). Furthermore, while a significant main effect of target category was observed, $F(4,10434) = 30.6$, $p < .0001$, % EQ Tran did not independently affect performance (p=.07). Nevertheless, % EQ Tran and MHL interacted significantly, $F(3,10434) = 3.2$, $p = .02 < .05$. A significant two-way interaction effect between MHL and target cat-

egory was also observed $F(4,10434) = 11.7$, $p < .0001$ as well as % EQ Tran and target category, $F(12,10434) = 3.8$, $p < .0001$. Lastly a significant three-way interaction on performance was observed between target category, % EQ Tran, and MHL $F(12,10434) = 5.2$, $p < .0001$.

Post-hoc analysis was conducted on the % correct responses accrued over the respective test conditions for both participant groups. As suggested by the model, barring the effects of EQ-Transform and target category, % correct responses among yNH ($Mdn = 84.9\%$, $IQR = 82.9$ - 89.3%) were significantly higher than oHI ($Mdn = 79.6\%$, $IQR = 70.3$ - 84.3%), $U = 589$, $p < .0001$, $r = .54$ (Very large effect). The model nevertheless suggests a progressive decline in MSA performance with increasing hearing loss. From a baseline performance which was that taken for yNH at 84.9 % for a MHL of 6 dB HL and an estimated OR of 0.97 for the main effect of MHL, the model projects that the performance may fall to even chance-levels at an MHL of around 63 dB HL, characterized by moderately severe hearing loss. The negative effect of hearing loss on MSA performance was noticeable across all target classes. Interestingly, both yNH and oHI showed the highest accuracy in detecting Lead vocals. Most notably however, Lead vocals brought on the largest disparity in performance between the groups, with yNH having a perfect median score of a 100%, markedly outperforming the oHI ($Mdn = 92.5\%$, $IQR = 86.3$ - 96.3%), $U = 651$, $p < .0001$, $r = 0.73$ (Huge effect). In contrast, both groups performed worst at detecting Bass targets. For the latter, although yNH ($Mdn = 67.5\%$, $IQR = 60$ - 75%) performed significantly better than the oHI ($Mdn = 60\%$, $IQR = 50$ - 66.3%), the performance gap was the smallest observed across the target classes, $U = 507$, $p = .01 < .05$, $r = 0.3$ (Medium effect). After Lead vocals, Guitar targets elicited the best performance among yNH ($Mdn = 91.5\%$, $IQR = 85.8$ - 94.4%). The difference in performance compared to oHI ($Mdn = 81.6\%$, $IQR = 74.5$ - 87.2%) was also observably the second largest, $U = 573$, $p = .0002 < .001$, $r = 0.5$. Detection of Piano targets also saw a superior performance in yNH ($Mdn = 86.8\%$, $IQR = 81.7$ -

92%) which was significantly better than oHI (Mdn = 78.1%, IQR = 71.7 - 85.7%), U= 550, p = .001 < .01, r= 0.45 (Large effect). Both groups performed similarly for Drums: yNH (Mdn = 87.5%, IQR = 82.5 - 90%), oHI (Mdn = 78.8%, IQR = 75 - 85%), U= 531, p = .003 < .01, r= 0.4 (Large effect).

As suggested by the model, there were no significant main effects of EQ-Transform on the performance among both yNH and oHI (p>.2). Nevertheless, the performance disparity between the two groups was most pronounced at factory settings (100 % EQ Tran), where yNH (Mdn = 87.7%, IQR = 85.3 - 90%) significantly outperformed oHI (Mdn = 80.7%, IQR = 74.3 - 84.8%), U= 600, p < .0001, r= 0.6, (Very large effect). This difference was smallest for 0 % settings, with yNH (Mdn = 85.1%, IQR = 83.1 - 89.3%) still performing significantly better than oHI (Mdn = 76.8%, IQR = 70.7 - 86.3%), U= 538, p = .001 < .01, r= 0.4 (Large effect size). This observation was also made at 200 % settings with yNH (Mdn = 85.8%, IQR = 81.6 - 88%) performing significantly better than oHI (Mdn = 78.6%, IQR = 71.4 - 81.9%), U= 576, p < .0001, r= 0.51 (Large effect size). Performance at 300 % settings among both groups was comparable to that shown for 200 % with yNH (Mdn = 85.8%, IQR = 81.6 - 89.8%) performing similarly better than oHI (Mdn = 78.9%, IQR = 69.6 - 83.7%), U= 567, p = .0002 < .001, r= 0.49 (Large effect size).

Factoring in the interaction effect of target category and % EQ Tran, a flawless median performance of 100 % was observed for yNH across all % EQ Tran settings for Lead vocals. As such, yNH performed significantly better across all settings except at factory settings where oHI (Mdn = 100%, IQR = 90 - 100%), performed similarly well. At 300 % EQ Tran, performance of oHI (Mdn = 90%, IQR = 80 - 95%) saw the largest deviation from that of yNH, U = 602, p < .0001, r = 0.7 (Very large effect). Although the disparity in performance for Bass was relatively smaller across all settings, 300 % EQ Tran similarly brought about the largest deviation in the performance between yNH (Mdn = 70%, IQR = 60 - 80%) and oHI

(Mdn = 50%, IQR = 50 - 65%), U = 535, p = .001 < .01, r = 0.42 (Large effect) with the latter performing at almost chance-level. Performance of oHI was similarly close to chance-level at both factory settings (Mdn = 60%, IQR = 45 - 65%) and 200 % (Mdn = 55%, IQR = 45 - 65%). For Drums, performance of oHI (Mdn = 80%, IQR = 70 - 90%) remained consistent over all settings while yNH performed best at factory settings (Mdn = 90%, IQR = 90 - 100%) and 0 % (Mdn = 90%, IQR = 80 - 100%). The largest disparity in performance for Drums was observed between the groups for the latter setting, U = 560, p = .0002 < .01, r = 0.5 (Large effect). For Guitar, yNH (Mdn = 89%, IQR = 77.8 - 100%) performed similarly across all settings except at 0 % settings (Mdn = 100%, IQR = 87.5 - 100%), where they performed best. As such, the performance disparity compared to oHI (Mdn = 87.5%, IQR = 75 - 93.8%), was observably the largest U = 574, p < .0001, r = 0.6 (Very large effect). Across all settings except 300 %, where performance of oHI (Mdn = 89%, IQR = 78 - 89%) was comparable, yNH performed significantly better. On the other hand, for all settings for Piano except 0 %, yNH performed significantly better. This effect was most pronounced at factory settings (Mdn = 100%, IQR = 90 - 100%) where the largest deviation in performance from oHI (Mdn = 90%, IQR = 85 - 90%) was observed, U = 566, p < .0001, r = 0.5 (Very large effect).

In order to assess if a statistical trend in performance of oHI existed by virtue of a step-wise increase in % EQ-Tran, we conducted the Jonckheere-Terpstra test (Manning et al., 2023). The test was conducted on the % correct responses across the four degrees of spectral manipulation, going from 0% EQ Tran implying the lowest spectral contrast, to 300 % implying the highest. By doing so, a significant, monotonically decreasing trend in the performance of oHI in detecting Bass targets was observed, $J\text{-}T$ = 1416, Z = -2.08, p = .03 < .05, τ = -.172 (Medium effect). For the other target classes however, no such trend could be shown for oHI.

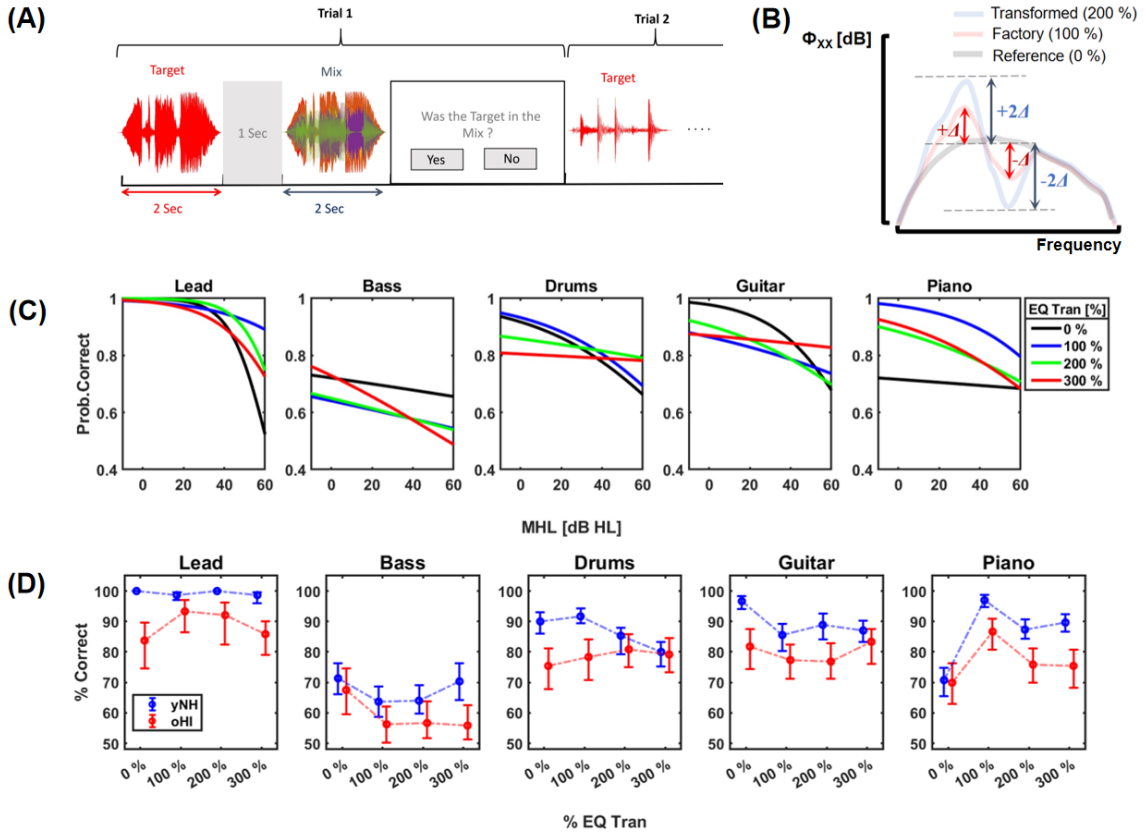


Figure 3.2: (A) Procedure of the experiment. (B) The effect of 200 % EQ-transform on the power spectrum. (C) The conditional probability plots illustrating the model output of probability correct for different target classes with respect to mean hearing loss and (D) means about bootstrapped 95% confidence intervals of correct answers as a percentage of the total trials per target category and % EQ-Transforms considered.

Although it can be observed from Figure 3.2(D), that manipulating the spectral constrast using the EQ-Transform does bring about changes to the performance, neither of the participant groups saw any improvement in their MSA performance with over-mixing (EQ Tran $>$ 100 %). In spite of these observations, it should be acknowledged that alterations to contrast may affect objective spectral descriptors , such as the frequency-domain sparsity of the stimuli nevertheless. However, as previously mentioned, the different % EQ Tran were not applied on the same tracks for an objective comparison in this work. Therefore, it cannot be ascertained that increasing the degree of spectral manipulations will give rise to higher objective frequency domain sparsity through higher Gini indices, because the global energy densities of the tracks may well vary. As shown in Figure 3.3, although there are

marginal changes in Gini indices for different % EQ Tran, neither the target nor the mix sparsity saw a significant increase ($p > .18$) by virtue of the higher degrees of spectral manipulations.

Apart from the Gini index, the spectral roll-off point (roll-off) was considered as another objective descriptor. This spectral descriptor provides the upper frequency limit below which 95% of the energy of the signal is contained, thereby indicating the rate at which the energies decay over frequency. Among other applications, this descriptor has been used widely within the context of genre classification (Li and Ogihara, 2005) in music and at discerning speech from music (Scheirer and Slaney, 1997). The overall effect of EQ-transform for this descriptor as shown in Figure 3.3 was significant only for the mix, $\chi^2(3) = 68$, $p < .0001$, η^2 =0.5 (Large effect). Moreover, a significant monotonical increase in the roll-off of the mixes with increasing % EQ-transform $p < .05$ was evident overall. The p-values reported herein were subjected to Bonferroni-Holm corrections to account for family-wise errors in multiple comparisons (Abdi, 2010).

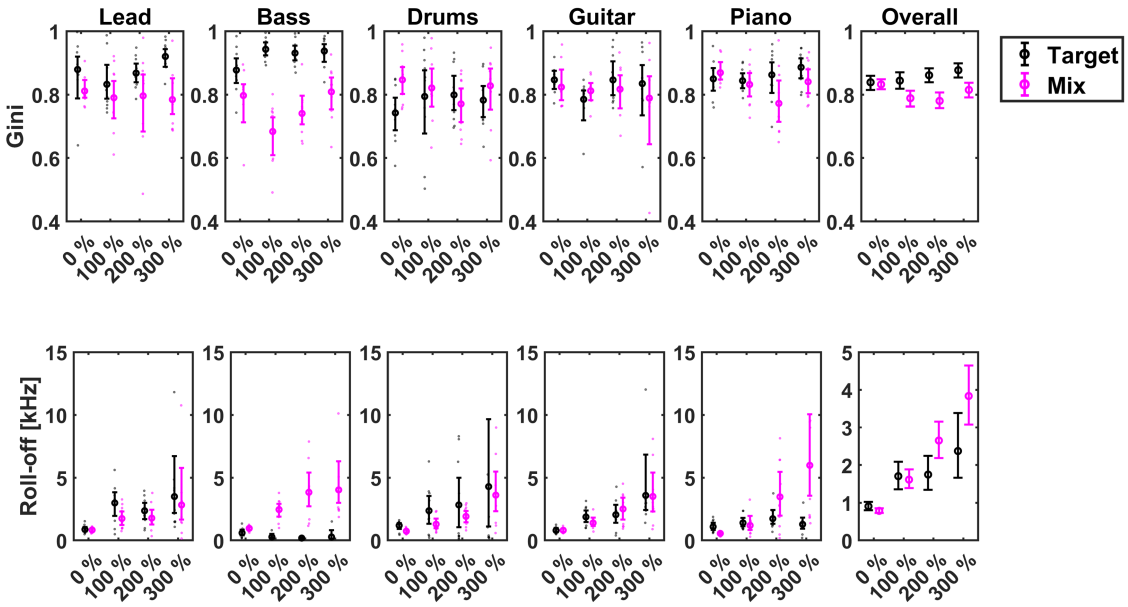


Figure 3.3: 95% Confidence interval plots about boot-strapped means illustrating the effects of % EQ-transform on the Gini indices and spectral roll-off points of the targets and mixes considered.

Unlike in the case of the Gini indices, the spectral manipulations through the EQ-transform do bring about changes to the roll-off points of the mixes. However, the question beckons if these spectral descriptors on their own, influence MSA performance among the participant groups. We therefore assessed a linear regression between the two descriptors and the MSA performance. Here, the % correct was that evaluated over all the participants in a given group (i.e. yNH and oHI) for the answers accrued over the individual trials bearing distinct values for the descriptors. Figure 3.4 gives an illustration of the linear correlation between the aforementioned descriptors and the % correct answers for the two groups. It is apparent that as the mixes became objectively sparser in the frequency domain, oHI participants saw an improvement in their performance $r = .3$, $p = .001 < .01$ as shown in Figure 3.4(A). As the roll-off does not indicate any transparency markers of the target amidst the mix, we took the vector difference between roll-off of the target and the masker (i.e. target roll-off - mix roll-off). In such a difference, a positive quantity indicates that the energy of the mix is distributed mostly over a smaller bandwidth than that of the target. A positive linear correlation between this roll-off difference and % correct answers is shown in Figure 3.4(B) among both the yNH $r = .3$, $p = .0001 < .001$. and the oHI $r = .3$, $p = .0003 < .001$. This may indicate that the narrower the range of frequencies over which the energy distribution of the mix becomes relative to that of the target, the less the target is energetically masked. Lastly, it was shown that musical training showed no significant correlation ($p > .3$) with MSA performance among neither participant groups. This assertion is supplemented by the model predictions outlined earlier where musical training had no effect on MSA performance.

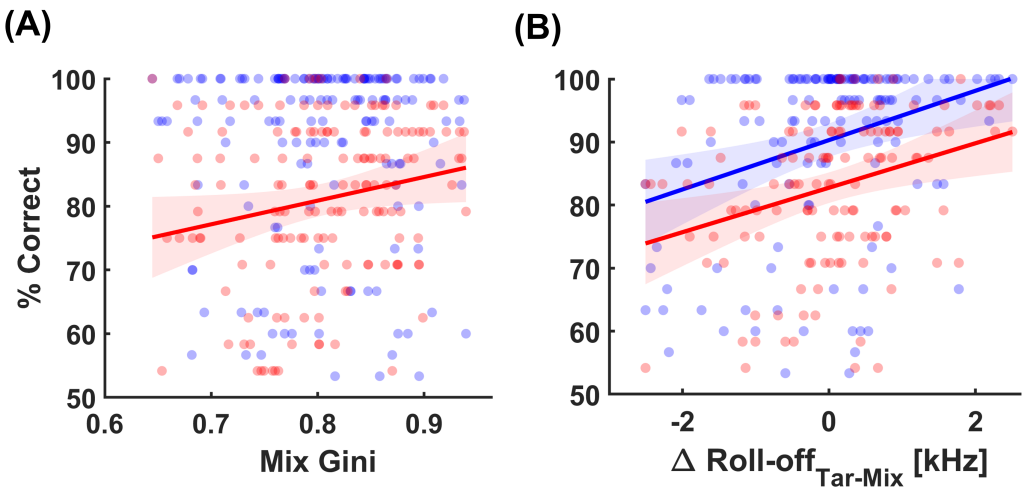


Figure 3.4: Linear correlations Gini index of the mixes (A), roll-off differences between target and mix (B) and % correct. Blue markers and trend lines indicate correlation among yNH and red oHI participants. Shaded regions shown with the trend lines correspond to the 95% confidence intervals.

3.6 Discussion

The findings of our detection task improve our understanding of a number of factors influencing the ability of listeners of varying degrees of hearing impairment to detect a target track within a musical mix. Notably, the type of instrument in the target track or target category emerged as a strong determinant of performance. Both participant groups performed best at detecting lead vocals while demonstrating the worst performance for bass, so much so that oHI participants performed at near chance-levels. This suggests that the perceptual salience of different instruments plays a pivotal role in the overall task performance. The level of hearing loss among participants demonstrated a modest negative influence on their detection ability. As such, MSA performance progressively declined with increasing hearing loss, reaching even chance-levels of performance at thresholds associated with moderately severe hearing impairment. The degree of spectral manipulation, while enhancing mixing preferences among oHI (Benjamin and Siedenburg, 2023), was found to have no noticeable effect on their MSA performance. Surprisingly, higher degrees of spectral manipulation even depreciated the performance of oHI in detecting Bass targets. In

Benjamin and Siedenburg (2024), we showed that although spectral manipulations to musical scenes did not confer benefits to MSA performance, the subjective sound quality ratings were heavily influenced by the degree of the manipulations, especially in yNH. Interestingly, MSA performance strongly correlated with the quality ratings in oHI participants. This suggests that improved scene analysis abilities among listeners with hearing loss, may have a favorable influence on their overall listening experience of multi-track music.

Furthermore, our results indicate that spectral manipulations on both the target and mix had varying degrees of influence on MSA performance. Unlike that of the target, the Gini indices of the mixes were associated with MSA performance of oHI participants, enhancing their ability to discern the target instrument within a mix. This finding suggests the potential utility of frequency domain sparsity as an adaptive strategy to improve auditory perception in individuals with hearing challenges. Furthermore, the higher roll-off points of the target with respect to that of the mix may well serve as an indicator of the energetic masking release of the former from the latter. Many previous studies elaborate upon the reduced frequency selectivity brought on by cochlear hearing loss (Plack, 2018). These discrepancies usually resulting from broader auditory filters may underpin our findings behind oHI listeners benefiting from sparser mixes with higher spectral contrast and higher roll-off point differences between target and mix. These higher positive roll-off differences allude to reduced energetic masking of the target by the mix, which may in turn provide an overall benefit in MSA performance irrespective of hearing loss.

3.7 Limitations

In the sample of participants considered in this study, the yNH participants were significantly younger than the oHI participants. Therefore the effect of hearing impairment on MSA performance observed, may also be due to a composite effect of hearing loss and age. Several studies allude to age being a critical determinant in

music perception despite hearing loss. In one such study, Bones and Plack (2015) showed that aging among normal-hearing listeners brought about a depreciation in neural representation responsible for differentiating between consonant and dissonant chords made up of two notes. Moreover, older participants rated dissonant chords as being more pleasant than consonant chords unlike their younger peers. In another study by Cohrdes et al. (2020), there were observable differences in valence and arousal perception brought on by age for musical stimuli but not for non-musical sounds. On an independent note, Goossens et al. (2017) showed that speech perception amid a masker has been shown to depreciate for participants above 50 years of age, even though they presented with normal-hearing loss levels up to 4 kHz. However, the same for music perception remains moot. Based on previous research, there is a possibility that our results could see differences if we considered yNH and oHI participants of similar ages. On that note, it is also important to highlight the fact that normal-hearing participants were more musically trained than their hearing-impaired peers. Previous literature compliments the advantages offered by musical training not only in music perception but also perceiving masked speech (Merten et al., 2021; Madsen et al., 2019). Chen et al. (2010) showed among children with cochlear implants that early introduction of musical perception training among such children may augment their ability to distinguish pitch differences between two successive Piano tones. Larrouy-Maestri et al. (2019) showed that the Gold-MSI musical training scores as that used here, correlated positively with the participants' ability to perceive mistuning or pitch shifts in lead vocal tracks with respect to that of the accompaniment in pop music. Musicians were also more likely to identify if a complex tone had a mistuned second harmonic (Zendel and Alain, 2009). Although musical training offers a plethora of advantages in music and speech perception, within the context of MSA ability, evidence supporting a similar assertion remains rather weak. Case in point, Hake et al. (2023) through investigating scene analysis abilities among participants with a wide rage of musical abilities showed that there was but a modest correlation between musical training

scores and MSA performance. This finding does not necessarily conflict with our analysis where musical training had no observable effects on the MSA performance. The absence of a significant effect brought on by musical training could therefore be due to the shortcomings brought on by comparatively small and unbalanced sample sizes between yNH and oHI. Nevertheless, controlling for musical training may potentially reduce the performance gap between the participant groups. On the note of classifying hearing impairment, we relied heavily upon the pure-tone audiometry. The area of auditory processing disorder addresses the phenomenon behind participants having difficulties in making out sounds amid a masker despite presenting with normal audiograms (Moore, 2006). As previous studies in the area suggest the implications of various factors in such shortcomings (Moore et al., 2013), more customized screening methods aimed at gauging these factors should be used in tandem with audiometry to more accurately assess the physiological and psychological causes of processing disorders impinging on MSA ability.

3.8 Conclusion

In an attempt to understand how music can be remixed for listeners with sensorineural hearing impairment to better facilitate their ability to successfully hear out a target instrument (MSA performance) within a coherent mix of several instruments and vocals, we applied the EQ-transform introduced in our previous study. This transform was used as a means of manipulating the spectral coloration on both the target and the individual tracks making up the mix. Despite having no effect on the objective frequency domain sparsity of the target and the mix measured using the Gini index, the transform did bring about significant and monotonical changes to the mix roll-off points. We assessed MSA performance as a function of hearing loss levels, musical training using the Gold-MSI questionnaire, and the level of spectral manipulation by way of the % EQ-transform. Although the participants reported varying degrees of musical training, its effect on MSA performance was negligible. As shown previously by Hake et al. (2023), the performance depended

strongly upon the category of the target instrument and to a much lesser extent on the hearing loss levels of the participants. Importantly, our results reveal a notable trend where as levels of hearing loss increased, MSA performance saw a steady decline, reaching even chance-level at hearing thresholds characteristic of moderately severe impairment or worse. This finding is similar to that shown by Hake et al. (2023) where listeners with severe to profound hearing loss performed at near chance-levels. Interestingly, the MSA performance among hearing-impaired participants saw a significant improvement for spectrally sparser mixes unlike that among normal-hearing participants which remained robust to changes in mix sparsity. Both participant groups benefited from target roll-offs being larger than that of the mix which may serve as a marker for reduced energetic masking of the former from the latter. Our findings thus far show that, within a musical scene analysis context, spectral manipulations to popular multi-track music may benefit hearing-impaired listeners. Given the complex interplay between hearing loss and various other factors in musical scene analysis, it would be beneficial to formulate auditory models to streamline our understanding of it. As a first step towards creating such models, in a future work, we aim to explore effective models of speech intelligibility and sound quality in their utility to predict performance in the music perception tasks investigated here.

References

Abdi, H. (2010). Holm's sequential bonferroni procedure. *Encyclopedia of research design*, 1(8):1–8.

Althoff, J., Gajecki, T., and Nogueira, W. (2024). Remixing preferences for western instrumental classical music of bilateral cochlear implant users. *Trends in Hearing*, 28:23312165241245219.

Asari, H., Pearlmutter, B. A., and Zador, A. M. (2006). Sparse representations for the cocktail party problem. *Journal of Neuroscience*, 26(28):7477–7490.

Bee, M. A. and Micheyl, C. (2008). The cocktail party problem: what is it? how can it be solved? and why should animal behaviorists study it? *Journal of comparative psychology*, 122(3):235.

Benjamin, A. J. and Siedenburg, K. (2023). Exploring level-and spectrum-based music mixing transforms for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 154(2):1048–1061.

Benjamin, A. J. and Siedenburg, K. (2024). Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music. *JASA Express Letters*, 4(11).

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. P. (2014). Medleydb: A multitrack dataset for annotation-intensive mir research. In *ISMIR*, volume 14, pages 155–160.

Bones, O. and Plack, C. J. (2015). Losing the music: aging affects the perception and subcortical neural representation of musical harmony. *Journal of Neuroscience*, 35(9):4071–4080.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, 77(5):1465–1487.

Bürgel, M., Picinali, L., and Siedenburg, K. (2021). Listening in the mix: Lead vocals robustly attract auditory attention in popular music. *Frontiers in Psychology*, 12:769663.

Bürgel, M. and Siedenburg, K. (2023). Salience of frequency micro-modulations in popular music. *Music Perception: An Interdisciplinary Journal*, 41(1):1–14.

Case, A. U. (2011). *Mix smart: Pro audio tips for your multitrack mix*. Focal Press Oxford.

Chen, J. K.-C., Chuang, A. Y. C., McMahon, C., Hsieh, J.-C., Tung, T.-H., and Li, L. P.-H. (2010). Music training improves pitch perception in prelingually deafened children with cochlear implants. *Pediatrics*, 125(4):e793–e800.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *Asha*, 23(7):493–500.

Cohrdes, C., Wrzus, C., Wald-Fuhrmann, M., and Riediger, M. (2020). "the sound of affect": Age differences in perceiving valence and arousal in music and their relation to music characteristics and momentary mood. *Musicae Scientiae*, 24(1):21–43.

Davis, A. C. and Hoffman, H. J. (2019). Hearing loss: rising prevalence and impact. *Bulletin of the World Health Organization*, 97(10):646.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (1980). Frequency selectivity in normally-hearing and hearing-impaired observers. *Journal of Speech, Language, and Hearing Research*, 23(3):646–669.

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing research*, 231(1-2):32–41.

Glasberg, B. R. and Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79(4):1020–1033.

Goossens, T., Vercammen, C., Wouters, J., and van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing research*, 344:109–124.

Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., and Siedenburg, K. (2023). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners.

Hwa, T. P., Tian, L. L., Caruana, F., Chun, M., Mancuso, D., Cellum, I. P., and Lalwani, A. K. (2021). Novel web-based music re-engineering software for enhancement of music enjoyment among cochlear implantees. *Otology & Neurotology*, 42(9):1347–1354.

LaFontaine, D. (2021). The history of bootstrapping: Tracing the development of resampling with replacement. *The Mathematics Enthusiast*, 18(1):78–99.

Larrouy-Maestri, P., Harrison, P. M., and Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, 51:663–675.

Lentz, J. J. and Leek, M. R. (2003). Spectral shape discrimination by hearing-impaired and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(3):1604–1616.

Li, T. and Ogihara, M. (2005). Music genre classification with taxonomy. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 5, pages v–197. IEEE.

Madsen, S. M., Marschall, M., Dau, T., and Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific reports*, 9(1):10404.

Manning, S. E., Ku, H.-C., Dluzen, D. F., Xing, C., and Zhou, Z. (2023). A non-parametric alternative to the cochran-armitage trend test in genetic case-control association studies: The jonckheere-terpstra trend test. *Plos one*, 18(2):e0280809.

Merten, N., Fischer, M. E., Dillard, L. K., Klein, B. E., Tweed, T. S., and Cruickshanks, K. J. (2021). Benefit of musical training for speech perception and cognition later in life. *Journal of Speech, Language, and Hearing Research*, 64(7):2885–2896.

Middlebrooks, J. C. and Waters, M. F. (2020). Spatial mechanisms for segregation of competing sounds, and a breakdown in spatial hearing. *Frontiers in neuroscience*, 14:571095.

Moore, D. R. (2006). Auditory processing disorder (apd): Definition, diagnosis, neural basis, and intervention. *Audiological Medicine*, 4(1):4–11.

Moore, D. R., Rosen, S., Bamiou, D.-E., Campbell, N. G., and Sirimanna, T. (2013). Evolving concepts of developmental auditory processing disorder (apd): a british society of audiology apd special interest group 'white paper'. *International journal of audiology*, 52(1):3–13.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality

of non-musicians: An index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642.

Nachar, N. et al. (2008). The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology*, 4(1):13–20.

Nagathil, A., Weihs, C., Neumann, K., and Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners. *The Journal of the Acoustical Society of America*, 142(3):1219–1228.

Narne, V. K., Jain, S., Sharma, C., Baer, T., and Moore, B. C. (2020). Narrow-band ripple glide direction discrimination and its relationship to frequency selectivity estimated using psychophysical tuning curves. *Hearing Research*, 389:107910.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G. M., Mallick, B. K., and Ibrahim, J. G. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *The annals of applied statistics*, 5(1):449.

Plack, C. J. (2018). *The sense of hearing*. Routledge.

Plumbley, M. D., Blumensath, T., Daudet, L., Gribonval, R., and Davies, M. E. (2009). Sparse representations in audio and music: from coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005.

Rasidi, W. N. A. and Seluakumaran, K. (2024). Simplified cochlear frequency selectivity assessment in normal-hearing and hearing-impaired listeners. *International Journal of Audiology*, 63(5):326–333.

Scheirer, E. and Slaney, M. (1997). Construction and evaluation of a robust multifeature speech/music discriminator. In *1997 IEEE international conference on acoustics, speech, and signal processing*, volume 2, pages 1331–1334. IEEE.

Siedenburg, K., Goldmann, K., and 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.

Siedenburg, K., Röttges, S., Wagener, K. C., and 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.

Zendel, B. R. and Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8):1488–1498.

3.9 Summary

- In an earlier study, we showed that individuals with elevated hearing thresholds favored greater spectral contrast in music mixes through higher % EQ-transform preferences.
- This follow-up study evaluated the effect of such spectral contrast changes on musical scene analysis (MSA) abilities.
- Listeners were required to identify or detect if a target excerpt played first was present in a coherent music mix played subsequently.
- % correct responses were used as a surrogate for MSA ability or performance.
- The target was of categories: lead vocals, bass guitar, drums, guitar, and piano.
- Young normal-hearing controls (yNH) and unaided older hearing-impaired listeners (oHI) with mostly moderate hearing loss were tested.
- Both participant groups performed best at detecting lead vocals and worst at detecting bass guitar.
- For bass guitar targets, oHI performed at near chance-levels of 50 %.
- Degree of spectral manipulations by way of the EQ-Transform had no influence on MSA performance.
- For bass guitar targets, a decreasing trend in MSA performance in oHI was observed with increasing % EQ-transform.
- Hearing loss brought on a progressive decline in MSA performance, so much so that baseline performance depreciated to mere chance levels at hearing thresholds characteristic of moderately severe hearing loss.

- Mixes with objectively sparser power spectra improved MSA performance in oHI.
- Musical scenes consisting of targets with broader roll-off points than the mixes benefited MSA performance in both groups.
- Unlike alterations to contrast, changes to frequency-domain sparsity and energetic masking in musical scenes may affect MSA performance in moderately hearing-impaired listeners.
- This opens the question of whether spectral contrast modifications may influence perceived audio quality.
- Particularly in hearing-impaired listeners, are there possible associations between scene analysis abilities and perceived audio quality for multi-track musical scenes ?

4. Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music

In this 3rd and final study, the emphasis is on evaluating potential associations between musical scene analysis and perceived audio quality in moderately hearing-impaired listeners. Particularly, the effect of spectral contrast modifications on audio quality appraisal of music mixes is investigated with the aid of the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) methodology. The study focuses on supplementing our understanding of the distinctive nature of music perception among hearing-impaired listeners.

4.1 Study 3

The study included in this chapter was published as: Benjamin AJ, Siedenburg K. Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music. JASA Express Lett. 2024 Nov 1;4(11):113202. *https://doi.org/10.1121/10.0032474*. The content of this chapter is identical to the published work.

Author Contributions: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the necessary experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

_____ _____

(name) Date

Supervisor

4.2 Abstract

This study assessed Musical Scene Analysis (MSA) performance and subjective quality ratings of multi-track mixes as a function of spectral manipulations using the EQ-transform (% EQT). This transform exaggerates or reduces the spectral shape changes in a given track with respect to a relatively flat, smooth reference spectrum. Data from 30 younger normal hearing (yNH) and 23 older hearing-impaired (oHI) participants showed that MSA performance was robust to changes in % EQT. However, audio quality ratings elicited from yNH participants were more sensitive to % EQT than those of oHI participants. A significant positive correlation between MSA performance and quality ratings among oHI showed that oHI participants with better MSA performances gave higher quality ratings, whereas there was no significant correlation for yNH listeners. Overall, these data indicate the complementary virtue of measures of MSA and audio quality ratings for assessing the suitability of music mixes for hearing-impaired listeners.

4.3 Introduction

Musical Scene Analysis (MSA) in multi-track music refers to perceptual and cognitive processes that allow listeners to discern and focus on specific musical elements within a complex multi-track musical arrangement. Multi-track mixing is a common component of contemporary music production. In a coherent multi-track mix, an assortment of vocal tracks and accompanying instruments are consolidated to create a richly layered mixture. The clarity of the mix, allowing listeners to identify and focus on individual instruments, seems important for both casual listeners and professional mixing engineers. Understanding how different listener groups, specifically those with a diagnosed hearing impairment, discern and appreciate mixes, may aid in creating specialized mixes for such listeners.

A handful of previous studies suggest that cochlear implant (CI) users may benefit from bespoke mixes other than the commercially distributed ones (Buyens et al., 2014; Gajecki and Nogueira, 2018; Tahmasebi et al., 2020). However, few studies exist that test mixing properties for non-CI users with cochlear hearing loss in terms of mix clarity or preference. One such study (Benjamin and Siedenburg, 2023) showed that hearing-impaired listeners with moderate to severe hearing loss had distinct level and balance preferences for mixes. Here, spectral manipulation using the so-called EQ-transform (EQT) was introduced. This transform applied to individual tracks of a mix enhanced or reduced their spectral coloration. The EQT had a significant effect on their objective frequency-domain sparsity measured using the robust Gini index (Hurley and Rickard, 2009). It was also shown that participants with higher levels of hearing loss preferred mixes with spectrally sparser tracks. A study conducted by Hake et al. (2023) tested MSA performance, that is, the ability to detect a target instrument in a multi-track mix, for participants with varying levels of hearing impairment. Performance depended strongly on the level differences between the target and the corresponding mix (that may or may not include the

target), the type of the target instrument (e.g. lead vocals, drums, guitar), and the number of tracks in the mix. However, the effects of spectral manipulations were not considered. In a more recent study (Benjamin and Siedenburg, 2025), the EQT was used to assess the effects of spectral manipulations of multi-track music on MSA performance. Although the MSA performance of normal hearing participants was unaffected by frequency-domain sparsity, hearing-impaired listeners performed better for mixes that were objectively sparser. However, the relation between subjective quality ratings of musical mixes and MSA performance remains poorly understood, especially for hearing-impaired listeners.

Auditory scene analysis (ASA) deals with individual abilities to discern and segregate sounds within a complex auditory scene. For musical stimuli, MSA is paramount in the context of appreciating and discerning individual instruments or lead vocals in polyphony (McAdams and Bregman, 1979). Notably, Hake et al. (2023) showed that there were relatively weak improvements in MSA abilities with increasing musical training, but stronger effects of cochlear hearing loss. The influence of cochlear hearing loss on MSA remains poorly explored, especially for multi-track music perception. Sensorineural hearing impairment, which mostly manifests itself with aging, affects auditory perception. The condition widely referred to as presbycusis, characterised by the gradual bilateral loss of hearing with increasing age, especially at high frequencies (Wu et al., 2020). According to Pichora-Fuller et al. (1995), presbycusis may compromise the ability to segregate sounds in noisy environments. Alain et al. (2001) suggested that among older individuals affected by presbycusis, neuroplasticity may aid in sound localization and pitch perception. Helfer and Freyman (2008) suggested that older individuals with hearing loss tend to rely considerably on spectral and contextual cues in a manner different from that for younger people with no hearing impairment. Importantly, Rasidi and Seluakumaran (2024) showed that even mild hearing impairment may reduce frequency selectivity. This may negatively impact the ability to detect spectral changes. Overall these

studies suggest that sensorineural hearing impairment may modify the perceptual weights allocated to different acoustic features, such as spectral content in scene-analysis tasks.

In multi-track mixing and production, equalization or EQing remains a fundamental method of spectral manipulation (Izhaki, 2017). Using EQing, the mixing engineer may manipulate the frequency content of a component track by enhancing or attenuating specific components in the frequency domain to achieve the desired audio quality (Senior, 2018). According to Izhaki (2017), EQing can be pivotal in determining the perceived quality of music. Studies by Gabrielsson et al. (1988) and Zielinski et al. (2008) demonstrate the sensitivity of listeners to changes in frequency balance and how it may dictate their judgments of audio quality. Yet, research on the influence of spectral manipulations and hearing loss on perceived audio quality in multi-track music remains scarce. More importantly, even fewer studies explore the connection between scene analysis abilities and perceived quality. Freyman et al. (2001) showed that the ability to separate target speech from a masker may be influenced by the quality of the speech and spatial cues. However, the manner in which scene analysis ability interacts with perceived audio quality of an auditory scene remains unclear, especially in the context of music perception. In this study, we aimed to understand the relationship between MSA and quality ratings as a function of hearing impairment and spectral manipulations using the EQT in multi-track mixes.

4.4 Methods

4.4.1 Participants

In the present study, 30 young normal hearing (yNH) (Ages, $M = 27$, $SD = 6$ yrs.) and 23 older hearing-impaired (oHI) participants (Ages, $M = 73$, $SD = 7$ yrs.) with predominantly moderate to severe cochlear hearing loss were tested. The

musical training of all of the participants was assessed using the Goldsmith Musical Sophistication Index (Gold-MSI) musical training subscale questionnaire proposed by Müllensiefen et al. (2014). The questionnaire consists of 7 questions. For each of the questions, a score between 1 and 7 was calculated based on the answers. The final score was calculated by summing the individual scores, a higher score indicating more musical training. **Supplementary material 1 in Appendix B** provides two specimen examples. The yNH were significantly more musically trained than the oHI, $t(51) = 2.2$, $p = 0.03$,$d = 0.6$ (Medium effect).

Pure-tone audiometry was conducted using a portable AD528 audiometer from Interacoustics GmbH (www.interacoustics.com/ad528). Based on the hearing loss for the ear with the lower arithmetic mean hearing thresholds (BEMHT) (taken over the pure tone frequencies 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz), the oHI ($M = 43$, $SD = 5$ dB HL) had approximately 40 dB higher hearing loss on average than yNH ($M = 4.3$, $SD = 6$ dB HL). Figure 4.1(A), shows audiograms for both groups for the better ear. According to Clark (1981), the following classifications were made: normal if BEMHT $\leq$ 25 dB, mild hearing loss if 25 dB $<$ BEMHT $\leq$ 40 dB, and moderate to severe hearing loss if BEMHT $>$ 40 dB. Thus, 16 oHI participants had moderate to severe hearing loss ($M = 45$, $SD = 4$ dB HL) with only 7 having mild hearing loss ($M = 38.3$, $SD = 2$ dB HL). Although a very strong trend of increasing hearing loss with age was apparent among yNH, $r = 0.6, p < 0.001$, $d = 1.5$ (Very large effect), the ages of oHI participants and their BEMHT were uncorrelated ($p = 0.5$). Figure 4.1(B) is a scatter plot of ages and BEMHT values.

4.4.2 Stimuli, apparatus, and procedure

The experiment was conducted in a low-reflection chamber with a pair of ESI active 8" near-field studio monitors by ESI Audiotechnik GmbH (Germany) (www.esi-audio.de/activ8) at the University of Oldenburg, Germany. These monitors were

separated by 90° and were 2 meters equidistant from the participant. The audio playback levels were calibrated using noise with the same long-term spectrum as the ensemble average of the power spectra taken from a myriad of instruments and lead vocals available in the open-source Medley database (Bittner et al., 2014). The calibration was such that the sum of sound pressures from both monitors at the position of the participant was 80 dB SPL(A). The stimuli were processed on a stand-alone desktop terminal using MATLAB R2023a. The terminal was linked to the monitors using an RME Fireface UFX audio-interface. The stimuli were taken from the aforementioned Medley database.

The experiment was conducted in two parts. In the first part, referred to as the MSA part, a target track or musical target consisting of a single instrument or vocal was presented followed by a mix after a one-second pause. The mix was an ensemble of instruments and lead vocal tracks. The participant was asked to indicate whether they heard the target in the mix that followed it. Both the target and the mix were of two seconds duration. The target-to -mix level difference was maintained at -10 dB and all mixes contained five distinct tracks. All oHI participants wearing hearing aids were requested to complete both parts of the experiment unaided. As shown in our previous study (Benjamin and Siedenburg, 2023), hearing-impaired listeners preferred boosting higher frequencies when unaided to compensate for the reduced audibility at these frequencies. Nevertheless, in this study, no such high-frequency amplification was provided as we aimed to investigate the effect of the EQT alone on MSA performance and quality ratings.

Both the target and component tracks of the mix were subjected to the EQT investigated by Benjamin and Siedenburg (2023). The EQT exaggerates or reduces the spectral variation of a given track in the frequency domain. This is achieved by linearly interpolating or extrapolating between the power spectrum of the stimulus undergoing the EQT (input signal) and a smooth reference spectrum, which is an

average spectrum of individual power spectra of more than 100 tracks taken from the aforementioned database. To perform the EQT operation, firstly the power spectrum of the input signal was calculated using the fast Fourier transform applied over its entire duration with a sampling frequency of 44.1 kHz. A rectangular window was used without any zero padding. Using the power spectrum of the input signal and the reference spectrum whose energy was normalized to that of the input signal, the power spectrum of the transformed signal was calculated by linearly interpolating between the two spectra. To mitigate audible artefacts in the transformed signal, firstly the difference between the transformed and the input power spectra was extracted. This noisy representation was then smoothed using a Savitsky-Golay filter (Schafer, 2011). The power spectrum of the input signal was colored with the smoothed power difference in the penultimate step to obtain the spectrum of the EQT signal. The EQT signal was obtained by applying the inverse Fourier transform to this spectrum. To illustrate, a 200% EQT would double the power differences between the reference spectrum and the power spectrum of the original stimulus, as illustrated in Figure 4.1(C). Transformed stimuli have an altered spectral contrast compared to the original. Conversely, a 0 % transform would reduce the coloration in the original spectrum. A detailed step-by-step explanation of the EQT process is provided in **Supplementary material 2, Appendix B**. In part one of the present experiment, both the target and the component tracks in the mix following it were subjected to either 0 %, 100 % (original), 200 %, or 300 % EQT. These degrees of spectral manipulation were chosen based on our previous work showing that unaided bilateral hearing aid users had % EQT preferences between 0 and 300 %.

The quality rating task in the second part of the experiment commenced after a voluntary break. Here, the participants were presented with a Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) interface (ITU-R, 2015) where they were tasked with providing their subjective audio quality ratings of 15 different music excerpts of two seconds duration, as in the first part. In every trial, the

participant would rate one of the 15 excerpts. The presentation order of excerpts was randomized. During a trial, the participant listened to EQT versions of that excerpt by clicking the play button and provided a quality rating for it on a scale from 0 to 100 with the rating slider. A rating of 0 corresponds to the worst quality and 100 to the best. The % EQT included in each trial were : -500 % serving as the anchor, -200 %, 0 %, 50 %, 100 %, 200 %, and 300 %. The participant was not allowed to provide their rating for a stimulus prior to listening to it at least once. Furthermore, the participant was only allowed to rate one item with a rating of 0 and one other item with a rating of 100 per trial. All of the 5 remaining items could only be rated between 0 and 100. The positions of the stimuli on the interface between trails were randomized so that the mixes transformed with a given % EQT did not always appear in the same locations. Only after providing the subjective quality ratings of all the EQT versions of a mix in one trial, could the participant proceed to the next trial where the process was repeated for another excerpt. The experiment as a whole came to an end once all of the 15 different excerpts were rated. Figure 4.1(D) illustrates the interface. The aim of the two parts was to ascertain the relationship between quality ratings and musical scene analysis ability as a function of the varying degrees of spectral manipulation provided by the EQT.

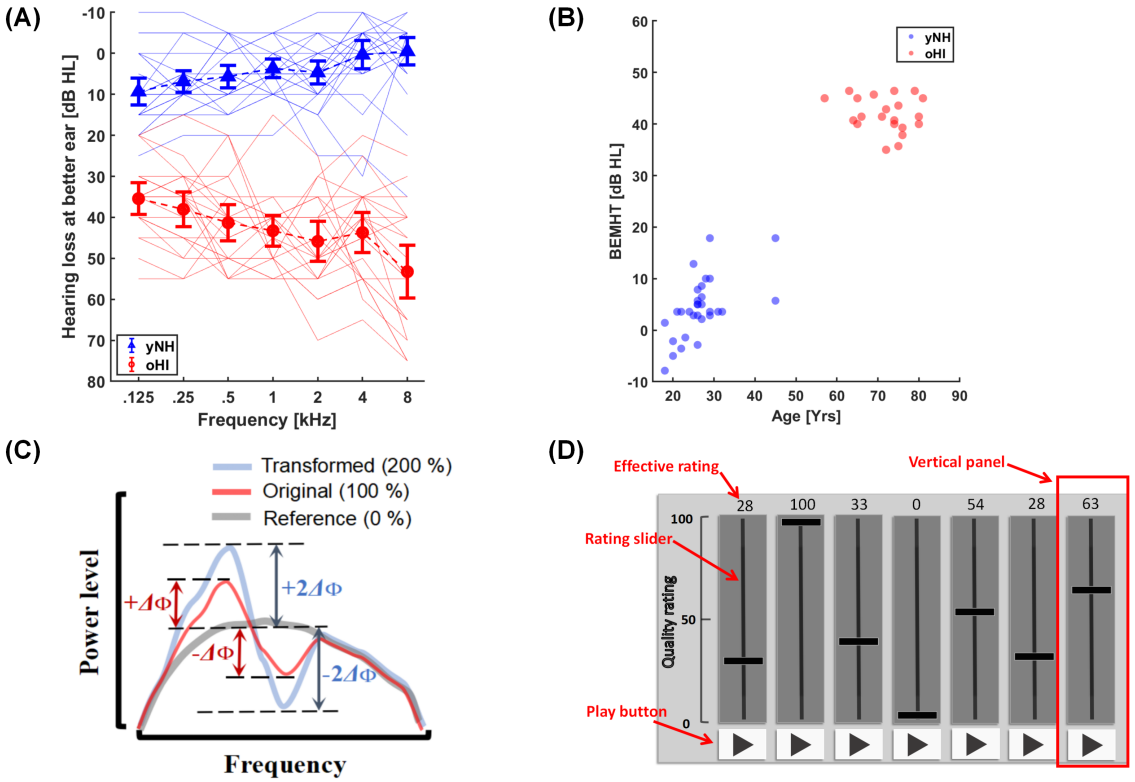


Figure 4.1: (A) Individual audiograms for the better ear of the participants (thin lines) and group mean hearing loss levels with 95% confidence intervals (thick lines). (B) Scatter plot of age and mean hearing loss for the better ear. Blue indicates yNH and Red oHI. (C) Effects of the EQT on the power spectrum of the original signal. Power level differences between original and flat reference are doubled with a 200 % EQT, as shown. (D) MUSHRA interface used in the quality rating task. The vertical panel where a particular stimulus appeared was randomized between trials. The participant rated the stimulus upon hearing it by clicking the play button and then rating its subjective quality using the rating slider provided in the corresponding vertical panel. Upon moving the slider, the effective rating indicated the actual position of the slider.

4.4.3 Statistical analysis

A non-parametric approach was used to analyse the MUSHRA data, which are prone to type 1 errors with parametric testing (Mendonça and Delikaris-Manias, 2018). We calculated bootstrapped means with 10^3 iterations with replacement (Davison and Hinkley, 1997). Lastly, to ascertain independent and interaction effects of musical training, % EQT, BEMHT, and MSA performance on the subjective quality ratings, a linear mixed effects model was used (Shek and Ma, 2011).

4.5 Results

4.5.1 Data

We first aimed at understanding the association between quality rating and MSA performance as a function of % EQT applied to the stimuli. Figure 4.2(A) shows the mean quality rating preferences and MSA performance and 95% confidence interval plots. For both groups, MSA performance was hardly affected by % EQT. However, the quality rating scores showed a quadratic dependence on % EQT. Furthermore, quality ratings was more affected by % EQT for yNH, $\chi^2(6) = 149$, $p < 0.0001$, $\eta^2 = 0.45$ (Large effect), than for oHI, $\chi^2(6) = 69.3$, $p < 0.0001$, $\eta^2 = 0.33$ (Medium effect). Among the oHI, quality rating scores were more closely related to MSA performance. This can be shown in Figure 4.2(B) where quality scores and MSA performance were averaged for each participant over % EQT values. Among the yNH, mean MSA performance was not correlated with the quality rating scores ($p = 0.75$) whereas for the oHI, there was a positive correlation, $r = 0.54$, $p = 0.009 < 0.01$, $d = 1.3$ (Very large effect). Lastly, mean quality ratings were significantly higher for the yNH than for the oHI participants, $t(51) = 4.7$, $p < 0.0001$, $d = 1.3$ (Very large effect).

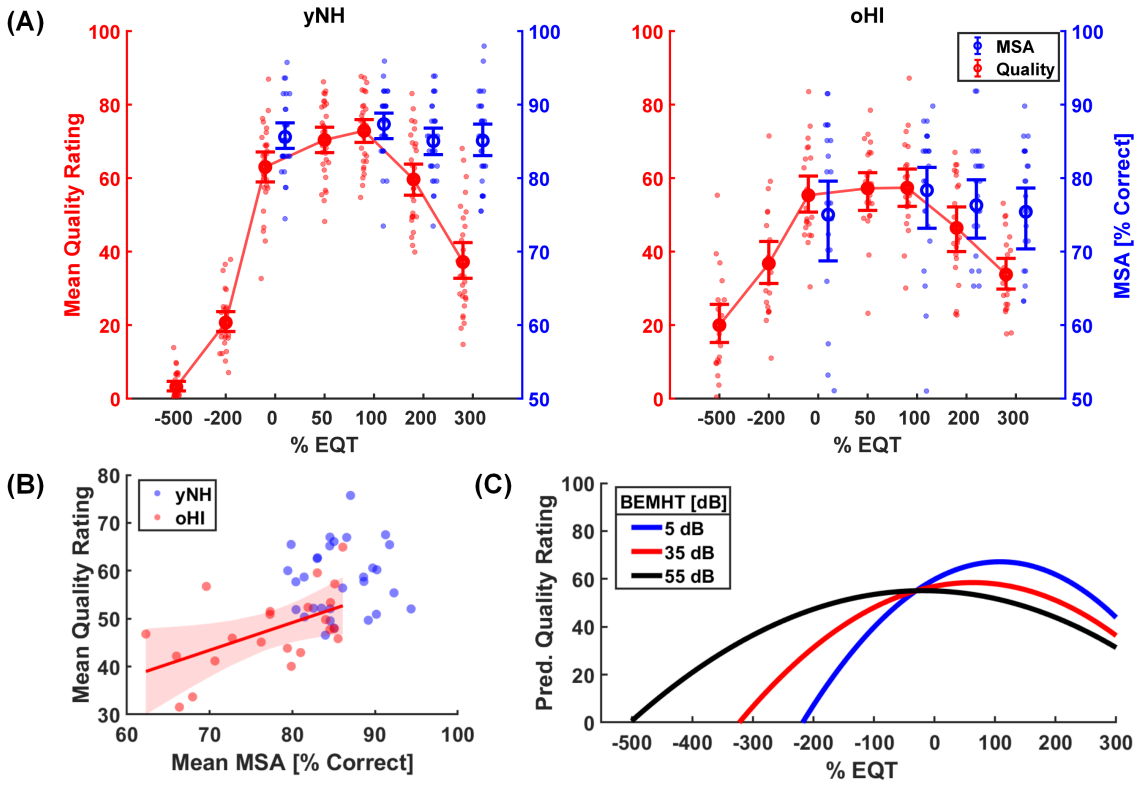


Figure 4.2: (A) Means and 95% confidence intervals for MSA performance and quality ratings scores calculated using bootstrapping for the data accrued over the % EQT values used. The plots are for yNH and oHI groups separately. (B) Linear correlation between mean MSA and quality ratings averaged over 0 %, 100 %, 200 %, and 300 % EQT for yNH and oHI groups. The shaded region accompanying the trend line corresponds to the 95% confidence interval. (C) Model predictions of quality ratings from the linear mixed effects model for the % EQT for different BEMHT.

4.5.2 Statistical model

A linear mixed-effect model was used to assess the influence of musical training, MSA ability, BEHMT, and % EQT on the subjective quality ratings obtained in the second part of the experiment. Owing to the parabolic pattern of quality rating scores with respect to % EQT observed in Figure 4.2(A), % EQT was included as a quadratic term in the model. Based on the model output, musical training did not have any effect on the quality ratings ($p = 0.4$). BEMHT had a significant independent effect on the quality ratings $F(1,199) = 21.1$, $p < 0.0001$, ${\eta_p}^2 = 0.1$ (Small effect). % EQT had a rather strong independent effect on the quality ratings $F(1,199) = 132.3$, $p < 0.0001$, ${\eta_p}^2 = 0.4$ (Large effect). The quadratic term of % EQT had a weaker albeit significant effect on the quality ratings, $F(1,199) = 55.4$, $p < 0.0001$, ${\eta_p}^2 = 0.22$ (Small effect). Although MSA performance had no significant impact on the quality ratings independently ($p = 0.23$), it had a significant interaction effect with BEMHT and % EQT, $F(1,199) = 4.4$, $p = 0.036 < 0.05$, ${\eta_p}^2 = 0.02$ (Very Small effect). Lastly, there was a modest yet significant two-way interaction effect between the quadratic % EQT term and BEMHT, $F(1,199) = 4.21$, $p = 0.042 < 0.05$, ${\eta_p}^2 = 0.02$ (Very small effect). Figure 4.2(C) provides the model prediction of quality ratings with respect to % EQT for different levels of BEMHT. Increasing BEMHT was associated with smaller changes in quality rating with changes in % EQT.

4.6 Discussion

In this study, we investigated the relationship between MSA performance and audio quality ratings for multi-track music stimuli subjected to spectral manipulations with the EQT. The EQT manipulates the spectral contrast and therefore the spectral shape of a signal as shown in an earlier study. Moreover, oHI participants preferred mixes with higher % EQT settings than yNH individuals. In this study, MSA perfor-

mance among both groups was robust across varying degrees of % EQT. Therefore, the ability to detect a target track in a complex musical mix was largely unaffected by spectral manipulations. This observed robustness could mean that spectral contrast alone may not influence MSA abilities in multi-track music, notwithstanding hearing impairment. However, the quality rating scores were more affected by % EQT for yNH than for oHI, suggesting that yNH listeners are more sensitive to alterations in the frequency domain and therefore more critical in their quality assessment of multi-track music. Lentz and Leek (2003) showed that hearing-impaired listeners had a reduced ability in processing alterations in spectral shape compared to normal hearing listeners. This was underpinned by Narne et al. (2020), where it was shown that hearing-impaired listeners had broader psychophysical tuning curves that correlated with a poorer ability to discriminate the ripple glide direction of narrow-band signals, ergo poorer discriminability of spectral shape. Huber et al. (2019) showed that age had a negative impact when detecting linear and non-linear distortions in speech while improved selective attention contributed positively. For music, only working memory had a significant positive effect at perceiving such distortions. In the present study, it was shown that MSA performance among yNH was independent of their quality rating scores, whereas a strong correlation between these two music perception metrics was observed for oHI. This suggests that oHI who were more adept at detecting the target track within the mix had a tendency to provide higher quality ratings. This could suggest that for individuals with cochlear hearing loss, improved scene analysis abilities may facilitate better listening experiences or vice versa. Future research should further investigate the validity of such a relationship.

4.7 Acknowledgments

The authors would like to thank Brian Moore for highly valuable comments on the manuscript and all of the participants for their interest in the study, and Hörzentrum Oldenburg gGmbH for their support. This study was funded by the

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330 A6 and by a Freigeist Fellowship to K.S. from the Volkswagen Stiftung.

4.8 Author Declarations

4.8.1 Conflict of Interest

The authors do not have any conflicts to disclose.

Ethics Approval

This study received approval from the Commission for Research Assessment and Ethics of the Carl von Ossietzky University of Oldenburg in Germany (Drs.EK/2019/092-01). An informed consent was obtained in written form from each and every participant.

4.9 Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the national academy of sciences*, 98(21):12301–12306.

Benjamin, A. J. and Siedenburg, K. (2023). Exploring level and spectrum-based music mixing transforms for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 154(2):1048–1061.

Benjamin, A. J. and Siedenburg, K. (2025). Effects of spectral manipulations of music mixes on musical scene analysis abilities of hearing-impaired listeners. *PloS one*, 20(1):e0316442.

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. P. (2014). Medleydb: A multitrack dataset for annotation-intensive MIR research. In *ISMIR*, volume 14, pages 155–160.

Buyens, W., Van Dijk, B., Moonen, M., and Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International journal of audiology*, 53(5):294–301.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *Asha*, 23(7):493–500.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Number 1. Cambridge university press.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109(5):2112–2122.

Gabrielsson, A., Schenkman, B. N., and Hagerman, B. (1988). The effects of different frequency responses on sound quality judgments and speech intelligibility. *Journal of Speech, Language, and Hearing Research*, 31(2):166–177.

Gajecki, T. and Nogueira, W. (2018). Deep learning models to remix music for cochlear implant users. *The Journal of the Acoustical Society of America*, 143(6):3602–3615.

Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., and Siedenburg, K. (2023). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods*, pages 1–26.

Helfer, K. S. and Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and hearing*, 29(1):87–98.

Huber, R., Rählmann, S., Bisitz, T., Meis, M., Steinhauser, S., and Meister, H. (2019). Influence of working memory and attention on sound-quality ratings. *The Journal of the Acoustical Society of America*, 145(3):1283–1292.

Hurley, N. and Rickard, S. (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741.

ITU-R (2015). ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems (ITU-R recommendation).

Izhaki, R. (2017). *Mixing audio: concepts, practices, and tools*. Routledge.

Lentz, J. J. and Leek, M. R. (2003). Spectral shape discrimination by hearing-impaired and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(3):1604–1616.

McAdams, S. and Bregman, A. (1979). Hearing musical streams. *Computer Music Journal*, 3(4):26–60.

Mendonça, C. and Delikaris-Manias, S. (2018). Statistical tests with MUSHRA data. In *Audio Engineering Society Convention 144*. Audio Engineering Society.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642.

Narne, V. K., Jain, S., Sharma, C., Baer, T., and Moore, B. C. J. (2020). Narrow-band ripple glide direction discrimination and its relationship to frequency selectivity estimated using psychophysical tuning curves. *Hearing Research*, 389:107910.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1):593–608.

Rasidi, W. N. A. and Seluakumaran, K. (2024). Simplified cochlear frequency selectivity assessment in normal-hearing and hearing-impaired listeners. *International Journal of Audiology*, 63(5):326–333.

Schafer, R. W. (2011). What is a savitzky-golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117.

Senior, M. (2018). *Mixing secrets for the small studio*. Routledge.

Shek, D. T. and Ma, C. M. (2011). Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *The scientific world journal*, 11(1):42–76.

Tahmasebi, S., Gajcki, T., and Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Frontiers in Neuroscience*, 14:514226.

Wu, P.-z., O'Malley, J. T., de Gruttola, V., and Liberman, M. C. (2020). Age-related hearing loss is dominated by damage to inner ear sensory cells, not the cellular battery that powers them. *Journal of Neuroscience*, 40(33):6357–6366.

Zielinski, S., Rumsey, F., and Bech, S. (2008). On some biases encountered in modern audio quality listening tests-a review. *Journal of the Audio Engineering Society*, 56(6):427–451.

4.10 Summary

- In the previous study, alterations to spectral contrast in multi-track music was shown to have no observable effects on the musical scene analysis (MSA) performance.
- In this study, the perceived audio quality for spectrally modified music mixes were evaluted.
- Young normal-hearing (yNH) and older hearing-impaired (oHI) were tested.
- oHI listeners were mostly moderately hearing-impaired.
- The perceived audio quality ratings were elicited for spectrally modified versions of music excerpts using the MUSHRA methodology.
- Contrary to that observed for MSA performance, perceived quality was sensitive to spectral contrast changes, especially in yNH.
- Predictions from a linear mixed-effects model fitted to the data suggested that as hearing thresholds increased, the distribution of quality ratings for spectrally modified mixes became more platykurtic.
- This reduced variability in audio quality ratings indicates that listeners with higher hearing thresholds may be less adept at discriminating changes to spectral contrast in music mixes.
- MSA performance correlated positively with quality ratings in oHI.
- Improved scene analysis abilities in hearing-impaired listeners may therefore enhance their listening experience of music or vice-versa.

5. Discussion

5.1 Summary

This dissertation investigates the effects of modifying commercially available music mixes on the subjective preferences and musical scene analysis in listeners with hearing impairment. While previous literature on cochlear implant (CI) users have highlighted the benefits of customized mixes for such listeners, this study considers hearing-impaired (HI) listeners who are non-CI users. Additionally, it evaluates whether spectral manipulations can improve musical scene analysis, specifically the capacity to selectively attend to individual musical elements within a multitrack arrangement. This is motivated by the broader auditory filters characteristic of cochlear hearing loss, which have been shown to affect the ability to resolve spectral detail in complex acoustic environments. Listening tests were conducted with normal-hearing (NH) controls and individuals with predominantly moderate or greater hearing loss to compare perceptual outcomes. The findings may aid in the development of novel audio pre-processing strategies for assistive technologies to enhance music appreciation among HI listeners. Importantly, any potential benefits of adjusting music mixes evident from this work may motivate the need to revise existing *best practices* followed by mixing engineers, to better accommodate the needs of such listeners.

5.1.1 Study 1: Exploring level- and spectrum- based music mixing transforms for hearing-impaired listeners

To investigate whether modifying commercially available music mixes created for NH listeners can benefit HI individuals, two experiments were conducted. In the first experiment, 25 NH listeners, 10 HI bilateral hearing-aid users, and 10 HI non-users were tested. In the second experiment, 18 bilateral hearing-aid (HA) users with predominantly moderate hearing loss were tested. Unlike the first experiment, each participant in experiment 2 was evaluated with and without their HAs. In both experiments, participants completed a remixing task in which they individually manipulated three audio effects in real-time. By doing so, they could provide their preferences for music mixes of 8 seconds duration, subjected to level and spectral adjustments. In the level-based adjustment referred to as the lead-to-accompaniment ratio (LAR), the broadband level of the lead vocals was altered, while that of the accompaniment was kept constant. In the second, spectrum-based adjustment involving spectral balance (SPBal), the spectral slope between 125 Hz and 8 kHz was adjusted in the final mix. This allowed the participants to manipulate the energy distribution of the mix around a 1 kHz center pivot point, effectively shifting its spectral centroid. In the third adjustment, also spectrum-based, the spectral contrast of individual tracks was altered using the EQ-transform. This transform calculates a power spectrum of a track with diminished or exaggerated contrast by linearly extrapolating the original spectrum relative to a smooth, musically average spectrum. The latter is an ensemble average of power spectra derived from a variety of vocal and instrumental excerpts taken from the open-source Medley database. This operation, which essentially downplays or emphasizes equalization applied during mixing, significantly affected the track-specific spectral sparsity, as measured by the Gini index of Constant-Q Transform (CQT) power spectra computed with a resolution of 3 bins per octave.

In Experiment 1, HA users preferred significantly higher levels of lead vocals compared to NH controls. Interestingly, these preferences among non-users exhibited substantial variability, indicating strong individual differences. SPBal preferences among non-users favored energy weightings of the mixes at frequencies above 1 kHz. While NH listeners showed a preference for reduced spectral contrast compared to that in the original mixes, no significant EQ-transform preferences were observed among HA users and non-users.

In Experiment 2, a clear distinctions between aided and unaided listening was shown. Hearing-aid disuse was associated with preferences for elevated lead-vocal levels, high-frequency amplification, and heightened spectral contrast in individual tracks. A pooled analysis of unaided listeners in both experiments revealed a robust positive correlation between hearing thresholds and preferences for level of the lead-vocals and spectral contrast. As hypothesized, the findings in this study suggests potential benefits in customizing music mixes for HI individuals.

5.1.2 Study 2: Effects of spectral manipulations of music mixes on musical scene analysis abilities of hearing-impaired listeners

Building upon the findings from the previous study where higher hearing thresholds were associated with exaggerated contrast preferences in the power spectra of individual tracks, this study aims to assess if greater spectral contrast may facilitate selective listening abilities for musical scenes or musical scene analysis abilities in HI listeners. To answer this question, a listening test with 30 young normal-hearing (yNH) controls and 24 older unaided hearing-impaired (oHI) individuals was conducted. The latter were listeners with largely moderate hearing loss. In the listening test, a target music excerpt of 2 seconds duration was presented first and was followed by a coherent music mix of similar duration after a one-second pause, in randomized trials. The target that preceded the mix was excluded in half of the

trials. Both the target and the constituent tracks in the mix were subjected to the EQ-transform used in the first study. The participants were required to report if they heard the preceding target in the mix. The target music excerpts presented were from one of five instrument categories : lead-vocals, bass guitar, drums, guitar, and piano.

Overall performance showed that NH listeners were more adept at selectively attending to target instruments in multi-track musical scenes. However, both groups performed best for lead-vocal targets and worst for bass guitar. For the latter, the oHI group performed almost at chance-levels of 50 % accuracy. The results indicated that poorer overall performance was observed for higher hearing thresholds, in that levels of hearing loss brought about a progressive decline in musical scene analysis abilities. Trials consisting of mixes sparser power spectra improved performance among the HI listeners. Moreover, mixes with lower spectral roll-off points compared to the target, improved performance in both groups. The outcomes of the study imply that spectral alterations to music mixes may serve as effective means of improving musical scene analysis abilities in moderately HI listeners.

5.1.3 Study 3: Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music

In the previous studies, it was shown that spectral contrast modifications to music mixes elicited higher subjective preferences from HI listeners. Furthermore, spectral descriptors indicating sparser frequency-domain representations and reduced energetic masking in multi-track musical scenes, improved scene analysis abilities in such listeners. However, perceptual quality ratings for music mixes with altered spectral contrast were not evaluated. To do so, 30 yNH and 23 unaided oHI listeners with predominantly moderate hearing-loss were tested. The quality ratings for music mixes subjected to varying degrees of contrast adjustments using the EQ-transform were assessed using the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) methodology. Although the degree of contrast alterations did not affect the upstream scene analysis abilities in both groups, the downstream quality ratings among NH listeners were observably more sensitive to such modifications. This suggests that HI listeners were less adept at identifying changes to spectral shape in music mixes, which became worse at higher hearing thresholds. Spectral contrast modifications notwithstanding, quality ratings and scene analysis were strongly associated with one another, in HI listeners. Such a finding may indicate a reciprocal relationship between musical scene analysis and perceived audio quality in individuals with moderate hearing loss.

5.2 Implications

The findings reported in this dissertation can suggest practical strategies in effectively customizing music mixes for individuals with moderate hearing loss or worse. These implications may be translated into potential improvements in the design of specialized music programs for hearing-aids and other assistive listening technologies aimed at enhancing music perception.

5.2.1 Vocal preference and salience in music mixes

A cardinal observation throughout Studies 1 and 2 was the noticeable role of lead vocals on both preference and MSA. Consistent with previous findings made for CI users (Buyens et al., 2014; Pons et al., 2016; Tahmasebi et al., 2020), aided HI listeners with moderate hearing loss or worse, favored louder lead vocals compared to NH through elevated LAR preferences. This trend became even more pronounced with HA disuse. Supplementary analysis in Appendix C4 indicated that LAR preferences did not vary significantly with worsening hearing thresholds, suggesting minimal individual differences in HI listeners, irrespective of HA use.

The observed preference for elevated lead vocals can be attributed to the subtle manner in which voice conveys unique semantically rich and emotionally salient speech information (Simon-Thomas et al., 2009; Scherer et al., 2017; Pinheiro, 2025). This includes high-frequency formant and consonant cues which are particularly vulnerable to the effects of sensorineural hearing loss (Horwitz et al., 2002; Carney et al., 2023). Furthermore, hearing loss has also been implicated in impaired fricative identification which relies heavily on information encoded in high frequencies (Scharenborg et al., 2015). Hearing loss is typically associated with reduced audibility in the 1–6 kHz range (Metidieri et al., 2013), broadened auditory filters (Bernstein and Oxenham, 2006), and reduced temporal resolution (Lorenzi et al., 2012; Baltzell et al., 2020). These auditory deficits degrade the ability of a listener

to separate the vocals from the accompaniment that occupy overlapping spectral regions, making them become more prone to energetic (Best et al., 2013) and forward masking (Brennan et al., 2015) than most other instruments. Therefore, the observed elevation in lead vocal levels in both aided and unaided HI may have been a means to counteract the effects of a compromised signal-to-masker ratio, which is only partially compensated for by HAs. Based on the findings in Study 1 for unaided listeners, this detrimental role of hearing loss on the perception of vocals becomes progressive as indicated by the preference for higher LAR with increasing hearing thresholds. Furthermore, the non-significant variability in LAR preferences brought on by hearing loss could suggest that such preferences may be driven largely by deficits brought on by hearing loss rather than any individual choices.

Similarly, a clear effect on scene analysis was observed in Study 2 in which, MSA performance in detecting lead vocals was comparatively the best, yet experienced the strongest decline as a result of hearing loss. Bürgel et al. (2021, 2023, 2024) conducted several studies examining top-down and bottom-up MSA abilities in NH listeners. They showed that although lead vocals prevailed in eliciting the best MSA performance, NH listeners were not particularly more efficient at attending to them, especially in the presence of spectrally similar competing sounds (Bürgel and Siedenburg, 2024). A rather interesting finding in (Bürgel et al., 2021) was that vocal salience for NH listeners remained unaffected by lowered energetic masking through reduced spectral overlap. For HI listeners on the other hand, investigations by Best et al. (2013) highlight the benefits of reduced spectral overlap in selectively attending to speech amid competing talkers. Although NH listeners tend to benefit more from energetic masking release than HI listeners in speech ASA tasks (Arbogast et al., 2005; Christiansen and Dau, 2012; Best et al., 2013), such benefits may have remained undetected in MSA tasks in Study 2 due to the near-perfect performance of NH listeners, especially for lead vocals. Considering these findings from speech ASA tasks in, the lack of a significant effect of spectral contrast modification

on overall MSA performance in HI listeners may be explained by their reduced sensitivity to the benefits of energetic masking release brought on by such modifications.

In summary, consistent with findings for CI users, our results indicate that elevating vocal levels in music mixes commensurate with hearing thresholds may enhance listening outcomes in unaided HI listeners. By doing so, beneficial effects to both scene analysis and listener satisfaction could be conferred in such a population. The uniform variance in LAR preferences across different hearing thresholds suggests that these level adjustments may require little customization beyond accounting for hearing loss. Furthermore, the conservative range of lead-vocal levels (De Man et al., 2014) and the documented decline in LAR over time (Gerdes and Siedenburg, 2023) in commercial mixes, indicate that current mixing guidelines should be reconsidered when calibrating vocal levels for HI listeners.

5.2.2 Diminished perceptibility of bass

In Study 2, a rather prominent effect of target instrument category on MSA performance was observed for both NH and HI listeners. Consistent with previous findings highlighting the high salience of lead vocals, this category was associated with the best overall performance. On the other hand, bass guitar consistently produced poor performance in both groups. This pattern aligns with the observations made by Bürgel et al. (2021), where even NH listeners approached chance levels in bottom-up selective attention tasks involving bass. In the top-down tasks used in Study 2, HI listeners similarly performed near chance, while NH listeners performed significantly better. However, the performance gap between the groups was arguably the narrowest among all categories. This suggests that hearing loss had a comparatively smaller impact on selective attention to bass guitar.

Bass guitars primarily occupy the lower frequency range and are susceptible to energetic masking from broadband, bass-heavy percussion instruments such as kick

drums and toms, which share overlapping frequency content (Savage, 2014) (see Figure 1 in Appendix C for average power spectra of targets and mixes in Study 2). Nevertheless, even when low-frequency audibility is relatively preserved despite age-related hearing loss as in this case (Otte et al., 2013), broader auditory filters and poorer temporal fine-structure encoding (Hopkins and Moore, 2011) can degrade the internal spectral contrast of bass elements, making them less distinct from competing low-frequency tracks. This degraded neural representation can hinder segregation of bass, which can explain the tedium of tracking bass guitar targets experienced by both groups, with NH listeners performing consistently better, albeit relatively close to chance levels.

Non-significant effects of EQ-transform on overall performance observed in Study 2 notwithstanding, for bass, higher contrast modifications were associated with a decreasing trend in MSA performance among HI listeners. Furthermore, higher % EQ-transform settings lowered the 95% roll-off points of bass targets, essentially narrowing their low-frequency bandwidth. One possible explanation for such observed deficits in performance could be a compromised ability of HI listeners to use phase-locking (Plyler and Ananthanarayan, 2001; Verschooten et al., 2019) and TFS processing (Strelcyk and Dau, 2009) at low frequencies effectively. As a consequence of reduced access to resolved low-order harmonic cues (Oxenham, 2008), they may rely on unresolved higher harmonics for the percept of pitch and timbre (Madsen et al., 2025). Narrowing the bandwidth can attenuate these higher harmonics, thereby reducing available cues, which may ultimately compromise the segregability of bass in complex musical scenes.

Another possible explanation involves the smearing of temporal envelopes caused by the broader auditory filters associated with cochlear hearing loss (Lorenzi et al., 2012). These envelopes convey important rhythmic cues (Peelle and Davis, 2012), which can be degraded as a result. Even in NH listeners, modulation detection

thresholds are poorer at low carrier frequencies (Viemeister, 1979). This could suggest that bass guitar may inherently carry temporal envelope cues that are more difficult to track. Such a limitation could help explain the relatively poor performance also observed in NH listeners. In HI listeners, additional envelope smearing could further degrade these already weak cues, potentially worsening deficits in performance.

5.2.3 Frequency balance: Effects on preference and MSA

In Study 1, the preferences among unaided HI listeners for greater energy weighting above 1 kHz in composite music mixes was observed. This effect reversed with HA use, suggesting that HA amplification may sufficiently compensate for the compromised high-frequency audibility in age-related hearing loss (Beamer et al., 2000), reducing the need for artificially boosted content at these higher frequencies. Moreover, we later show that in unaided listening, high-frequency weighting preferences became increasingly varied as hearing thresholds worsened, suggesting greater individual differences in such preference with more severe loss. This variability may reflect the wide range of high-frequency sensitivities found in age-related hearing loss (Wang et al., 2021), implying that different listeners may require distinct compensatory adjustments to restore access to high-frequency content. Importantly, the absence of such variability in aided listening may reflect the benefits of clinical HA fitting practices, in which amplification is individualized to the listener's audiometric profile (Kimlinger et al., 2015; Urbanski et al., 2021).

As for MSA performance measured in Study 2, both NH and HI listeners performed better at selectively attending to musical targets with higher 95% roll-off points than the mixes. Such scenes usually present targets whose spectra are more high-frequency-centric than that of the mix. As a result, the target is less likely to be energetically masked (Brungart et al., 2001) and easier to hear-out as a result. Villard et al. (2023) showed that energetic masking increased listening effort

in younger NH listeners. A reduction in cognitive load conferred by a reduced effort may be especially beneficial to upstream MSA processes in older HI individuals who have relatively poorer cognitive reserve (Uchida et al., 2019).

In sum, for musical scenes, our results suggest that bandwidth extension of target sounds by facilitating access to high-frequency information may confer benefits to MSA performance in unaided HI listeners. Taken together with the observed weighting preferences in such listeners, music mixes with greater emphasis on higher frequencies may enhance listener preference and scene analysis. This is supported by the fact that in mild to severely HI individuals, high-frequency HA amplification elicited better listening preferences (Plyler and Fleck, 2006) and extending such amplification to even higher frequencies improved SRT in speech-on-speech ASA tasks (Levy et al., 2015).

5.2.4 Effect of spectral shape changes on quality perception in music mixes

In Study 3, we showed that HI listeners had a reduced ability to differentiate between different EQ-transform settings applied to music mixes, compared to NH listeners in terms of perceived quality. This was evinced by the increasingly platykurtic distribution of their audio quality ratings, which became progressively flatter with higher hearing thresholds. We hypothesized that this finding reflects a diminished ability of individuals with hearing loss to discern alterations to spectral shape in music mixes.

Previous literature has demonstrated such limitations in HI individuals for speech (Shrivastav et al., 2006; Fogerty et al., 2023) and non-musical stimuli such as complex tones (Lentz and Leek, 2002, 2003; Lauer et al., 2009; Rahne et al., 2011) and narrow-band noise (Narne et al., 2020). However, very few studies have conducted similar investigations within musical contexts. In one such study, Emiroglu

and Kollmeier (2008) investigated timbre discrimination abilities for musical instruments in quiet and noise. By measuring just noticeable differences (JND) in timbre among NH and moderately HI listeners, they showed that HI individuals with steep patterns of hearing thresholds were less adept at distinguishing timbre than NH listeners, in both noisy and quiet conditions. However, when the signal-to-noise ratio was sufficiently high, the JNDs of HI individuals with flat or diagonal hearing thresholds were comparable to those of NH listeners. Kong et al. (2011) showed that CI users were less successful at using spectral cues than NH listeners to judge changes in timbre in synthesized instrument tones. In music, modifications to spectral shape are closely associated with changes in timbre perception (Grey and Gordon, 1978). Furthermore, alterations to spectral descriptors such as centroid or relative energy content between even an odd harmonics (McAdams et al., 1995; Kendall et al., 1999) and spectral flux (Krimphoff et al., 1994) can introduce shifts in perceived musical timbre. Fujinaga (1998) showed that the timbre classification performance of an exemplar based machine-learning paradigm depended strongly on the spectral centroid of steady-state portions of instrumental excerpts.

These studies adequately demonstrate that manipulations to spectral shape can induce changes to timbre perception in music. However, such investigations have mostly been conducted on NH listeners with varying degrees of musical competence or sophistication. In this light, Wei et al. (2022) argue that a greater emphasis should be placed on the role of hearing loss on the perception of timbre or spectral shape in music. Therefore, we aim to clarify whether sensorineural hearing loss impairs the ability to judge distortions to spectral shape in music mixes as alluded to in Study 3. In order to do so, we evaluated objective changes in spectral shape introduced by the EQ-transform using the log-spectral distance (LSD), and examined their effects on audio quality ratings in an extended analysis (see Appendix C1).

LSD has been widely used in speech processing (Gray and Markel, 1976), par-

ticularly in improving the estimation of speech in noise (Erell and Weintraub, 1990) and to train statistical models for speech synthesis (Wu and Tokuda, 2009). It has also been used repeatedly to quantify spectral distortion introduced by algorithms that improve speech perception not only in noisy, but also reverberant environments (Habets, 2007; Dong and Lee, 2018). Despite limitations such as sensitivity to low-pass filtering and anomalous behavior under bandwidth-limited conditions, LSD remains a valid and informative measure for capturing spectral shape changes in broadband speech signals (Prodeus and Kotvytskyi, 2017). In musical contexts, Arifianto and Pratiwi (2016) applied LSD to assess the effects of harmonic enhancement in melodic audio by comparing their log-spectra with the original reference signal. Higher spectral deviations indicated by larger LSD values corresponded with greater perceived degradation, as confirmed by lower subjective quality ratings from both musicians and non-musicians. Collectively, these studies suggest that LSD is a useful measure of spectral shape change in speech and music, particularly in full-bandwidth signals where perceptual structure is preserved (Moore, 2019).

Results from the supplementary analysis in Appendix C1 indicate that HI listeners require observably greater deviations in spectral shape before experiencing a decline in perceived audio quality compared to NH listeners for music mixes. This effect becomes increasingly pronounced with higher degrees of hearing loss, suggesting a progressive shift in sensitivity to spectral coloration in music mixes as auditory thresholds worsen. In essence, audio quality appraisal among listeners with hearing loss is less sensitive to spectral shape changes introduced through practices of equalization. This implies that individuals with higher thresholds of hearing manifest a greater tolerance for spectral distortions. Such an interpretation is consistent with the findings of Brons et al. (2014), who reported that listeners with mild to moderate hearing loss were less sensitive than NH listeners to distortions in target speech introduced by background noise reduction algorithms in HAs. In accordance with their observations, they suggest that despite introducing higher distortions, stronger

noise reduction can be applied with minimal impact on speech quality in HI listeners. In this regard, results from the supplementary analysis shows that compared to NH listeners, those with moderate hearing impairment begin to report noticeable degradations in quality only when the average deviation in the power spectrum exceeds 4 dB or an approximately threefold difference in power per frequency bin. This approaches 7 dB or an alarming fivefold difference in power for individuals with moderately severe hearing loss.

In Study 2, we showed that music mixes of sparser power spectra, consistently improved scene analysis performance, primarily in individuals with moderate hearing loss or worse, irrespective of the target instrument category. Furthermore, in Study 1, we showed that preferred EQ-transform settings correlated positively with average hearing thresholds over both ears, indicating a preference for more exaggerated EQing at higher hearing thresholds. Moreover, in Study 1, it was also demonstrated that more pronounced EQing was associated with objectively sparser power spectra of composite tracks in music mixes.

Considered together, these findings indicate that mixing practices incorporating more aggressive equalization strategies may improve selective listening in multi-track musical scenes for individuals with moderate or greater hearing loss, while incurring relatively minimal detriments to perceived audio quality. Although very little evidence exists in previous literature for music, several studies on speech perception point towards the benefits of spectral alterations in HI individuals. Plomp (1988) argues that reduced spectral contrast in speech has a negative effect on recognition in aided listening. Heightened spectral contrast has been shown to improve consonant identification in unaided individuals with moderate to severe hearing loss (Bunnell, 1990; Munoz et al., 1999). Similarly, differentiation of vowel formants improved in HI with higher spectral contrast emphasizing the second formant (Woodall and Liu, 2013). Enhanced spectral contrast has also been shown to improve speech

intelligibility in noise for HI listeners (Simpson et al., 1990; Baer et al., 1993). For non-speech stimuli such as harmonic tone complexes, HI individuals benefited from elevated spectral contrast in distinguishing between complexes with closely spaced peak frequencies (Dreisbach et al., 2005).

5.2.5 Audio quality perception and MSA

In Study 3, a clear positive association was observed, suggesting a synergistic relationship between MSA performance and perceived audio quality for HI listeners with predominantly moderate-to-severe hearing loss (see Appendix C2), whereas no such link was evident for NH controls. Moreover, supplementary analysis provided in Appendix C3 revealed a greater dispersion in MSA performances with higher hearing thresholds.

These findings suggest that HI listeners with greater hearing thresholds may not simply perform uniformly worse but also show a wider range of selective listening abilities for musical scenes. Although similar findings have not been reported for music, these findings are conceptually consistent with those of Humes (2021), who showed that pure-tone average (PTA) was a major predictor of individual differences in speech reception thresholds (SRT) among older adults. More specifically, Nuesse et al. (2018) showed that among older HI with mild to moderate hearing loss, PTA emerged as the strongest predictor of the explainable variance in SRT, notwithstanding the test conditions. Therefore, hearing loss could play a dual role in MSA, in that it may not only impair the ability to selectively attend to musical scenes but also emphasize individual variability within the HI population. More recently, Hake et al. (2025b) also showed that hearing loss emerged as a strong predictor of SRT, while highlighting large individual differences within older HI participants. The relatively weaker effect on MSA, may suggest that as with speech, unmeasured factors such as residual frequency selectivity (Moore, 2007), cognitive capacity (Akeroyd, 2008), or listening effort (Pichora-Fuller et al., 2016) that may differ substantially among

individuals with similar audiometric profiles, could contribute to the higher variance in MSA abilities in older HI listeners. Such individual differences in selective listening ability are likely to play a pivotal role in shaping the overall listening experience.

Listening effort, which is often elevated in HI individuals (Mobarakeh et al., 2025), becomes progressively demanding as hearing thresholds worsen (Kamal et al., 2025). Nevertheless, this increased effort is not captured with standardized diagnostic audiometry (Hussein et al., 2022). In speech perception, improvements in sentence intelligibility in noise were associated with reduced listening effort in both NH (Zekveld et al., 2010) and HI listeners (Sarampalis et al., 2009; Wendt et al., 2017; Ohlenforst et al., 2018). In the context music, higher MSA performance may be similarly indicative of reduced listening effort, which could in turn, enhance downstream audio quality appraisal owing to more available cognitive resources (Wingfield, 2016). According to Konecni and Sargent-Pollock (1976) and Madison and Schiölde (2017), musical excerpts which are cognitively more demanding may be less favored, whereas appreciation for them may increase when cognitive resources are more readily available.

Altogether, the observed interdependence of MSA and quality perception in music may underscore the difficulty in creating music mixes that are both accessible and enjoyable for listeners with moderate or greater hearing loss. As such, the "one-size-fits-all" approach commonly used for NH listeners may not be optimal. Therefore, bespoke mixing strategies that consider perceptual and cognitive requirements of HI individuals could potentially enhance their MSA performance and overall listener satisfaction.

5.2.6 Individual differences in mixing preferences: The influence of hearing loss and bilateral hearing-aid use

Based on the preferences for mixing effects in Study 1, it was shown that preferences for LAR and % EQ-transform settings among unaided listeners increased at higher degrees of hearing loss, while no such observations could be made for SPBal preferences. Supplementary analysis in Appendix C4 shows a significant variation in SPBal and EQ-Transform preferences in unaided HI listeners with increasing hearing thresholds, which may suggest greater individual differences in preferences at higher thresholds. Furthermore, a similar trend, although not as pronounced, was observed for EQ-transform preferences in aided HI. On the other hand, variations in LAR preferences remained uniform over different hearing thresholds. Despite the higher variability in mean LAR preferences among woHA listeners with mild HI compared to NH in Experiment 1, the uniformity across both experiments suggests that the variability in Experiment 1 may not be attributable to hearing loss. Taken together with the findings from Study 1, these results imply that bilateral HA use not only brings down the mean preferences for the mixing effects but may also mitigate the individual differences in spectral balance and contrast preferences among HI listeners.

For unaided HI with especially moderate hearing loss or greater, contrast-adjusted mixes may be beneficial. Such listeners tend to prefer more pronounced spectral contrast and attend to musical targets more effectively in mixes with objectively sparser power spectra and spectral content that decays more rapidly with frequency, both of which can be effectively modified via the EQ-transform as shown in Studies 1 and 2. While the transform builds on established best practices in mixing (Pestana et al., 2014), the observed variability in listener preferences and MSA abilities indicates that mixes created by simply over-emphasizing generic EQing templates alone may not be sufficient for optimizing mixes for such listeners. Instead, these

results support the assertion made by Pons et al. (2016) for CI-users, who benefited from more individualized mixes. The moderate association between speech-in-noise intelligibility and MSA abilities reported by Hake et al. (2025b), combined with the observed link between MSA and quality judgments in Study 3, further suggests that selective listening abilities in general, may serve as a valuable marker to help custom-tailor EQ strategies for HI listeners with higher hearing thresholds. Specifically, more aggressive EQing to yield spectrally sparser representations, may serve as an effective remixing approach for listeners with poorer scene analysis abilities, potentially alleviating listening effort (discussed in the next section). Doing so may improve accessibility to multi-track music beyond what is feasible through conventional mixing practices.

5.2.7 Effects of objective frequency-domain sparsity on MSA and listener preference

Throughout Studies 1 and 2, a notable observation was the influence of objective frequency-domain sparsity on mixing preference and MSA performance among HI listeners. In both studies, sparsity was quantified using the robust Gini index computed from CQT spectra at third-octave spacing. In Study 2, we showed that the instrument category notwithstanding, HI listeners had improved MSA performance for sparser mixes as indicated by higher Gini indices. However, NH listeners did not see any benefits in their near-ceiling MSA performance which remained robust to changes in mix sparsity. Broader auditory filters characteristic of cochlear hearing loss may compromise auditory signals (Souza et al., 2012), thereby elevating cognitive load and listening effort in older HI listeners (Martini et al., 2015; Pichora-Fuller et al., 2016; Uchida et al., 2019). Spectrally sparser representations may partially compensate for the reduced spectral resolution brought on by hearing loss, thereby reducing the associated listening effort in selective attention tasks (Shinn-Cunningham and Best, 2008; Winn et al., 2015). According to the cognitive

load hypothesis (Lavie, 1995), this reduced effort could free up cognitive resources, which in turn may improve upstream MSA owing to more available top-down cognitive bandwidth (Tun et al., 2009).

As for preferences elicited in Study 1, interesting trends in contrast modifications were shown by way of higher % EQ-transform preferences as hearing thresholds worsened. In addition to the significant effects on the sparsity of individual tracks reported in Study 1, the influence of the EQ-transform on the sparsity of the composite music mix had not been discussed previously. On that note, supplementary analysis in Appendix C5 shows that the EQ-transform is similarly effective in manipulating the sparsity of the overall mix. Moreover, the analysis reveals that as hearing thresholds increased, unaided HI listeners tended prefer sparser mixes, where as the preferences elicited by aided listeners was unaffected. Although EQ-transform preferences were elevated with HA disuse according to the findings from Experiment 2 in Study 1, the elevation in sparsity preferences observed in the supplementary analysis was non-significant. Therefore, while bilateral HAs may not directly alter preferences for objectively sparser music mixes, they may moderate the influence of hearing loss on such preferences.

Considered collectively, these findings indicate that HI listeners not only show improved MSA performance in spectrally sparser mixes but also exhibit a clear subjective preference for such mixes when unaided. This aligns with the idea that spectral sparsity may aid in offsetting the reduced spectral resolution and increased listening effort associated with hearing loss, allowing greater cognitive resources to be allocated to auditory scene analysis. The lack of similar effects in NH listeners suggests that the benefits of sparsity are specific to the perceptual and cognitive challenges posed by hearing impairment. Therefore, sparsity measures such as the Gini index can serve as objective functions for optimizing EQing strategies to create perceptually tuned music mixes for unaided HI listeners with moderate hearing loss

or worse. As HAs moderate the influence of hearing loss but not significantly lower sparsity preferences, aided listeners may still benefit from objectively sparser mixes. Therefore, such strategies could potentially inform back-end processing approaches for music programs. This may complement the benefits already conferred by digital noise reduction schemes to listening effort (Arehart et al., 2011; Croghan et al., 2012; Desjardins and Doherty, 2014) and by DRC technologies to MSA and downstream judgments in sound quality (Hake et al., 2025a) and preference (Croghan et al., 2014; Moore and Sek, 2016) in aided music perception.

5.3 Future work

This dissertation provides a foundation for optimizing music mixes for individuals with mostly moderate hearing loss or worse. As such, the scope of the research can indeed be expanded further. Specifically, instead of a remixing task where isolated audio effects are manipulated as in Study 1, a more compound task involving the adjustment of two or more audio effects could be tested. By doing so, a more associative influence of these audio effects on listener preference can be examined such as, lead vocal level preferences in music mixes with heightened spectral contrast. On the note of spectral contrast, we used the EQ-transform to modify contrast, ergo objective spectral sparsity in music mixes. Given the role of sparsity measured with the Gini index observed throughout this dissertation, the index can be used as an objective function to optimally set track-specific transform parameters. By doing so, an instrument or stem of choice can be rendered more salient in the mix.

Spatialization plays a crucial role in music production and has been explored extensively in automatic mixing approaches to create mixes that are both transparent (Tom et al., 2019) and enjoyable (Perez-Gonzalez and Reiss, 2010). As such, among the audio effects reported in Study 1, we also investigated the influence of stereo panning width and transformed panning, which leverages track-specific panning ap-

plied by the mixing engineer, on listener preference. However, we did not observe any noticeable differences in spatialization preferences across participant groups. Similarly, Hake et al. (2023) showed a relatively weak effect of stereo panning on MSA in NH and HI listeners. These lack of meaningful effects may stem from the limited scope offered by such conventional panning methods. In comparison, multichannel spatialization offers a greater sense of envelopment (George et al., 2008), although it may not necessarily elicit better listener preferences in NH listeners (Rees-Jones et al., 2015). On the other hand, implementing virtual reality (VR) based auditory environments through headphones may be more feasible and requires considerably less infrastructure than deploying multichannel speaker systems.

Sensorineural hearing loss limits the ability to localize sound sources in complex auditory scenes (Akeroyd, 2014; Lundbeck et al., 2018) and training offered through simulated virtual scenes may aid in overcoming such limitations brought by mild to moderate (Valzolgher et al., 2024) or even severe hearing loss (Parmar et al., 2024). Therefore, spartializing music mixes in computerized auditory environments as implemented in these studies can be used as an effective alternative to the relatively cumbersome multichannel techniques. Given its efficacy on NH listeners, as an initial step, non-inividualzed HRTFs as proposed by Steadman et al. (2019) can be used to render such environments for unaided HI listeners.

Based on the quality ratings elicited from the listeners, we hypothesized that HI listeners were less sensitive to changes in spectral shape. In order to validate this assertion, JND based studies as that conducted by Emiroglu and Kollmeier (2008) can be extended to measure spectral shape discriminability in popular music mixes. Furthermore, in an attempt to explain the positive association between MSA abilities and quality appraisal in music mixes, we suggested that higher MSA abilities correspond to a reduced listening effort based on similar evidence in previous literature on speech perception. In line with the cognitive load hypothesis (Pichora-

Fuller et al., 2016), we suggested that such benefits may confer enhanced listening experiences and as a result, elicit more favorable quality ratings. Therefore, future research could aim to specifically investigate this hypothesis. To do so, pupilometric and brain-imaging methods can be incorporated to measure listening effort (Peelle, 2018) in tandem with MSA and quality assessment tasks used in this work.

Lastly, speech intelligibility models such as those proposed by Jørgensen and Dau (2011) and Biberger et al. (2016, 2017) can be adapted to predict MSA performance. This adaptation may offer deeper insights into the psychoacoustical processes underlying MSA and help clarify the complex, multidimensional nature of selective listening abilities in individuals with hearing loss.

5.4 Conclusion

In order to assess how manipulated music mixes affect preference, scene analysis, and subjective quality appraisal in HI individuals with mostly moderate hearing loss or higher, three studies were conducted in this dissertation. By doing so, a number of interesting findings were made available.

A rather consistent pattern in the findings was the progressive effect of hearing loss assessed via standardized pure-tone audiometry on the afore-mentioned music perception metrics. Higher hearing thresholds were implicated in preferences for elevated lead-vocal levels and exaggerated spectral contrast adjustments. Furthermore, HI listeners tended to prefer high-frequency amplification when unaided. Bilateral hearing aids not only moderated these observations, but also reduced the individual variability in the preferences brought on by hearing loss. The only exception was in the case of lead-vocal levels, where variability in preferences remained uniform as hearing thresholds increased.

Although MSA abilities were not evaluated in aided listening, hearing loss yet again played a progressive role in eliciting poorer and more varied MSA performances among unaided listeners. As such, overall selective listening abilities depreciated to mere chance levels at hearing thresholds associated with moderately severe hearing loss levels. In addition to the relatively weaker effects of hearing loss on MSA abilities, the target instrument category emerged as the strongest predictor of performance. Lead-vocals appeared most salient, yet MSA performance involving lead-vocal targets was more strongly affected by hearing loss. This observation was reversed for bass guitars where performance was arguably the poorest but remained relatively unaffected by hearing loss, despite NH listeners performing significantly better. Non-significant effects of contrast modifications notwithstanding, HI listeners saw a significant improvement in their performance for the top-down selective attention tasks involving mixes with objectively sparser power spectra. Both NH and HI listeners performed better for scenes with mixes of lower spectral roll-off points relative to the target, indicating benefits of energetic masking release to MSA, in spite of hearing loss.

Despite having no observable impact on MSA, quality appraisal were affected by the degree of spectral contrast adjustments to music mixes. To that end, NH listeners were more critical of the adjustments than HI listeners, as evinced by their audio quality ratings. Importantly, HI listeners were less critical of similar changes to spectral shape; an observation which became more pronounced with increasing hearing thresholds. Interestingly, a significant linear association between MSA performance and quality ratings were observed in HI listeners, suggesting a synergetic relationship between the two music perception metrics among individuals with hearing loss, unlike in NH listeners where no such association could be shown.

The overarching findings support the hypothesis that commercially available music mixes may not be suitable HI individuals and that mixing for such listeners

requires a rather multi-faceted approach than that followed in the more generic best practices. Such an approach should account for physiological, perceptual, and cognitive consequences of sensorineural hearing loss, as well as other listener-specific factors.

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(sup2):S53–S71.

Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends in Hearing*, 18:2331216514560442.

Arbogast, T. L., Mason, C. R., and Kidd Jr, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 117(4):2169–2180.

Arehart, K. H., Kates, J. M., and Anderson, M. C. (2011). Effects of noise, nonlinear processing, and linear filtering on perceived music quality. *International Journal of Audiology*, 50(3):177–190.

Arifianto, D. and Pratiwi, E. W. (2016). Enhanced harmonics for music appreciation on cochlear implant. In *2016 IEEE Region 10 Conference (TENCON)*, pages 2167–2171. IEEE.

Baer, T., Moore, B. C., and Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of rehabilitation research and development*, 30:49–49.

Baltzell, L. S., Swaminathan, J., Cho, A. Y., Lavandier, M., and Best, V. (2020). Binaural sensitivity and release from speech-on-speech masking in listeners with and without hearing loss. *The Journal of the Acoustical Society of America*, 147(3):1546–1561.

Beamer, S. L., Grant, K. W., and Walden, B. E. (2000). Hearing aid benefit in patients with high-frequency hearing loss. *Journal of the American Academy of Audiology*, 11(08):429–437.

Bernstein, J. G. and Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *The Journal of the Acoustical Society of America*, 120(6):3929–3945.

Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (2013). Spatial release from masking as a function of the spectral overlap of competing talkers. *The Journal of the Acoustical Society of America*, 133(6):3677–3680.

Biberger, T. and Ewert, S. D. (2016). Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility. *The Journal of the Acoustical Society of America*, 140(2):1023–1038.

Biberger, T. and Ewert, S. D. (2017). The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking. *The Journal of the Acoustical Society of America*, 142(2):1098–1111.

Brennan, M. A., McCreery, R. W., and Jesteadt, W. (2015). The influence of hearing-aid compression on forward-masked thresholds for adults with hearing loss. *The Journal of the Acoustical Society of America*, 138(4):2589–2597.

Brons, I., Dreschler, W. A., and Houben, R. (2014). Detection threshold for sound distortion resulting from noise reduction in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 136(3):1375–1384.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5):2527–2538.

Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 88(6):2546–2556.

Bürgel, M., Picinali, L., and Siedenburg, K. (2021). Listening in the mix: Lead vocals robustly attract auditory attention in popular music. *Frontiers in psychology*, 12:769663.

Bürgel, M. and Siedenburg, K. (2023). Salience of frequency micro-modulations in popular music. *Music Perception: An Interdisciplinary Journal*, 41(1):1–14.

Bürgel, M. and Siedenburg, K. (2024). Impact of interference on vocal and instrument recognition. *The Journal of the Acoustical Society of America*, 156(2):922–938.

Buyens, W., Van Dijk, B., Moonen, M., and Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International journal of audiology*, 53(5):294–301.

Carney, L. H., Cameron, D. A., Kinast, K. B., Feld, C. E., Schwarz, D. M., Leong, U.-C., and McDonough, J. M. (2023). Effects of sensorineural hearing loss on formant-frequency discrimination: Measurements and models. *Hearing research*, 435:108788.

Christiansen, C. and Dau, T. (2012). Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. *The Journal of the Acoustical Society of America*, 132(3):1655–1666.

Croghan, N. B., Arehart, K. H., and Kates, J. M. (2012). Quality and loudness judgments for music subjected to compression limiting. *The Journal of the Acoustical Society of America*, 132(2):1177–1188.

Croghan, N. B., Arehart, K. H., and Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and hearing*, 35(5):e170–e184.

De Man, B., Leonard, B., King, R., and Reiss, J. D. (2014). An analysis and evaluation of audio features for multitrack music mixtures. In *Proceedings of the 137th AES Convention*, Los Angeles, CA. Audio Engineering Society.

Desjardins, J. L. and Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and hearing*, 35(6):600–610.

Dong, H.-Y. and Lee, C.-M. (2018). Speech intelligibility improvement in noisy reverberant environments based on speech enhancement and inverse filtering. *EURASIP Journal on Audio, Speech, and Music Processing*, 2018:1–13.

Dreisbach, L. E., Leek, M. R., and Lentz, J. J. (2005). Perception of spectral contrast by hearing-impaired listeners. *Perception*.

Emiroglu, S. and Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain research*, 1220:199–207.

Erell, A. and Weintraub, M. (1990). Recognition of noisy speech: Using minimum-mean log-spectral distance estimation. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*.

Fogerty, D., Ahlstrom, J. B., and Dubno, J. R. (2023). Recognition of spectrally shaped speech in speech-modulated noise: Effects of age, spectral shape, speech level, and vocoding. *JASA Express Letters*, 3(4).

Fujinaga, I. (1998). Machine recognition of timbre using steady-state tone of acoustic musical instruments. In *ICMC*.

George, S., Zielinski, S., Rumsey, F., and Bech, S. (2008). Evaluating the sensation of envelopment arising from 5-channel surround sound recordings. In *124th Convention of the Audio Engineering Society*.

Gerdes, K. and Siedenburg, K. (2023). Lead-vocal level in recordings of popular music 1946–2020. *JASA Express Letters*, 3(4).

Gray, A. and Markel, J. (1976). Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(5):380–391.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500.

Habets, E. A. P. (2007). Single-and multi-microphone speech dereverberation using spectral enhancement.

Hake, R., Bürgel, M., Lesimple, C., Vormann, M., Wagener, K. C., Kuehnel, V., and Siedenburg, K. (2025a). Perception of recorded music with hearing aids: Compression differentially affects musical scene analysis and musical sound quality. *Trends in Hearing*, 29:23312165251368669.

Hake, R., Müllensiefen, D., and Siedenburg, K. (2025b). Individual differences in auditory scene analysis abilities in music and speech. *Scientific Reports*, 15.

Hopkins, K. and Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, 130(1):334–349.

Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (2002). Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing. *The Journal of the Acoustical Society of America*, 111(1):409–416.

Humes, L. E. (2021). Factors underlying individual differences in speech-recognition

threshold (srt) in noise among older adults. *Frontiers in Aging Neuroscience*, 13:702739.

Hussein, A. B., Lasheen, R. M., Emara, A. A., and El Mahallawi, T. (2022). Listening effort in patients with sensorineural hearing loss with and without hearing aids. *The Egyptian Journal of Otolaryngology*, 38(1):99.

Jørgensen, S. and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, 130(3):1475–1487.

Kamal, N., El Kholy, W., ElKabarity, R., and Mohamed Salah, S. (2025). Assessment of listening effort experience in normal and hearing-impaired adult population. *The Egyptian Journal of Otolaryngology*, 41(1):91.

Kendall, R. A., Carterette, E. C., and Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, 16(3):327–363.

Kimlinger, C., McCreery, R., and Lewis, D. (2015). High-frequency audibility: The effects of audiometric configuration, stimulus type, and device. *Journal of the American Academy of Audiology*, 26(02):128–137.

Konecni, V. J. and Sargent-Pollock, D. (1976). Choice between melodies differing in complexity under divided-attention conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3):347.

Kong, Y.-Y., Mullangi, A., Marozeau, J., and Epstein, M. (2011). Temporal and spectral cues for musical timbre perception in electric hearing. *Journal of Speech, Language, and Hearing Research*, 54(3):981–994.

Krimphoff, J., McAdams, S., and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. ii. analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, 4(C5):C5–625.

Lauer, A. M., Molis, M., and Leek, M. R. (2009). Discrimination of time-reversed harmonic complexes by normal-hearing and hearing-impaired listeners. *Journal of the Association for Research in Otolaryngology*, 10:609–619.

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

Lentz, J. J. and Leek, M. R. (2002). Decision strategies of hearing-impaired listeners in spectral shape discrimination. *The Journal of the Acoustical Society of America*, 111(3):1389–1398.

Lentz, J. J. and Leek, M. R. (2003). Spectral shape discrimination by hearing-impaired and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(3):1604–1616.

Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C., and Puria, S. (2015). Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech. *Ear and hearing*, 36(5):e214–e224.

Lorenzi, C., Wallaert, N., Gnansia, D., Leger, A. C., Ives, D. T., Chays, A., Garnier, S., and Cazals, Y. (2012). Temporal-envelope reconstruction for hearing-impaired listeners. *Journal of the Association for Research in Otolaryngology*, 13(6):853–865.

Lundbeck, M., Grimm, G., Hohmann, V., Bramsløw, L., and Neher, T. (2018). Effects of directional hearing aid settings on different laboratory measures of spatial awareness perception. *Audiology Research*, 8(2):215.

Madison, G. and Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. *Frontiers in neuroscience*, 11:147.

Madsen, S., Kreft, H., Purmalietis, E., Dau, T., and Oxenham, A. (2025). Association between hearing loss but not age and mistuning perception in music. *Hearing Research*, page 109403.

Martini, A., Castiglione, A., Bovo, R., Vallesi, A., and Gabelli, C. (2015). Aging, cognitive load, dementia and hearing loss. *Audiology and Neurotology*, 19(Suppl. 1):2–5.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological research*, 58:177–192.

Metidieri, M. M., Rodrigues, H. F. S., de Oliveira, F. J. M. B., Ferraz, D. P., de Almeida Neto, A. F., Torres, S., et al. (2013). Noise-induced hearing loss (nihl): literature review with a focus on occupational medicine. *International archives of otorhinolaryngology*, 17(02):208–212.

Mobarakeh, Z. I. P., Amiri, M., Tavanai, E., and Rahimi, V. (2025). Mechanisms of listening effort in individuals with hearing loss. *Journal of Modern Rehabilitation*.

Moore, B. C. (2007). *Cochlear hearing loss: physiological, psychological and technical issues*. John Wiley & Sons.

Moore, B. C. (2019). The roles of temporal envelope and fine structure information in auditory perception. *Acoustical Science and Technology*, 40(2):61–83.

Moore, B. C. and Sek, A. (2016). Preferred compression speed for speech and music and its relationship to sensitivity to temporal fine structure. *Trends in Hearing*, 20:2331216516640486.

Munoz, C. A., Nelson, P. B., and Rutledge, J. C. (1999). Enhancement of spectral contrast in speech for hearing impaired listeners. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*.

Narne, V. K., Jain, S., Sharma, C., Baer, T., and Moore, B. C. (2020). Narrow-band ripple glide direction discrimination and its relationship to frequency selectivity estimated using psychophysical tuning curves. *Hearing Research*, 389:107910.

Nuesse, T., Steenken, R., Neher, T., and Holube, I. (2018). Exploring the link between cognitive abilities and speech recognition in the elderly under different listening conditions. *Frontiers in Psychology*, 9:678.

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of snr, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing research*, 365:90–99.

Otte, R. J., Agterberg, M. J., Van Wanrooij, M. M., Snik, A. F., and Van Opstal, A. J. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance. *Journal of the Association for Research in Otolaryngology*, 14(2):261–273.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends in amplification*, 12(4):316–331.

Parmar, B. J., Salorio-Corbetto, M., Picinali, L., Mahon, M., Nightingale, R., Somerset, S., Cullington, H., Driver, S., Rocca, C., Jiang, D., et al. (2024). Virtual reality games for spatial hearing training in children and young people with bilateral cochlear implants: the "both ears (bears)" approach. *Frontiers in Neuroscience*, 18:1491954.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and hearing*, 39(2):204–214.

Peelle, J. E. and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in psychology*, 3:320.

Perez-Gonzalez, E. and Reiss, J. (2010). A real-time semiautonomous audio panning system for music mixing. *EURASIP Journal on Advances in Signal Processing*, 2010(1):436895.

Pestana, P. D., Reiss, J. D., et al. (2014). Intelligent audio production strategies informed by best practices.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). *Ear and hearing*, 37:5S–27S.

Pinheiro, A. (2025). Behind a voice there is a speaker: Why vocal emotion research needs to become 'personal'. *Affective Science*, pages 1–13.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America*, 83(6):2322–2327.

Plyler, P. N. and Ananthanarayan, A. (2001). Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*, 12(10):523–533.

Plyler, P. N. and Fleck, E. L. (2006). The effects of high-frequency amplification on the objective and subjective performance of hearing instrument users with varying degrees of high-frequency hearing loss. *Journal of Speech, Language, and Hearing Research*, 49(3):616–627.

Pons, J., Janer, J., Rode, T., and Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, 140(6):4338–4349.

Prodeus, A. and Kotvytskyi, I. (2017). On reliability of log-spectral distortion measure in speech quality estimation. In *2017 IEEE 4th International conference*

actual problems of unmanned aerial vehicles developments (APUAVD), pages 121–124. IEEE.

Rahne, T., Böhme, L., and Götze, G. (2011). Timbre discrimination in cochlear implant users and normal hearing subjects using cross-faded synthetic tones. *Journal of neuroscience methods*, 199(2):290–295.

Rees-Jones, J., Brereton, J., and Murphy, D. (2015). Spatial audio quality and user preference of listening systems in video games. In *DAFx 2015-Proceedings of the 18th International Conference on Digital Audio Effects*, pages 1–8.

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction.

Savage, S. (2014). *Mixing and mastering in the box: the guide to making great mixes and final masters on your computer*. Oxford University Press.

Scharenborg, O., Weber, A., and Janse, E. (2015). Age and hearing loss and the use of acoustic cues in fricative categorization. *The Journal of the Acoustical Society of America*, 138(3):1408–1417.

Scherer, K. R., Sundberg, J., Fantini, B., Trznadel, S., and Eyben, F. (2017). The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *The Journal of the Acoustical Society of America*, 142(4):1805–1815.

Shinn-Cunningham, B. G. and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in amplification*, 12(4):283–299.

Shrivastav, M. N., Humes, L. E., and Kewley-Port, D. (2006). Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners. *The Journal of the Acoustical Society of America*, 119(2):1131–1142.

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., and Abramson,

A. (2009). The voice conveys specific emotions: evidence from vocal burst displays. *Emotion*, 9(6):838.

Simpson, A., Moore, B., and Glasberg, B. (1990). Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Oto-Laryngologica*, 109(sup469):101–107.

Souza, P., Wright, R., and Bor, S. (2012). Consequences of broad auditory filters for identification of multichannel-compressed vowels. *Journal of Speech, Language, and Hearing Research*, 55(2):474–486.

Steadman, M. A., Kim, C., Lestang, J.-H., Goodman, D. F., and Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Scientific Reports*, 9(1):18284.

Strelcyk, O. and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, 125(5):3328–3345.

Tahmasebi, S., Gajcki, T., and Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Frontiers in Neuroscience*, 14:434.

Tom, A., Reiss, J. D., and Depalle, P. (2019). An automatic mixing system for multitrack spatialization for stereo based on unmasking and best panning practices. In *Audio Engineering Society Convention 146*. Audio Engineering Society.

Tun, P. A., McCoy, S., and Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and aging*, 24(3):761.

Uchida, Y., Sugiura, S., Nishita, Y., Saji, N., Sone, M., and Ueda, H. (2019). Age-related hearing loss and cognitive decline—the potential mechanisms linking the two. *Auris Nasus Larynx*, 46(1):1–9.

Urbanski, D., Hernandez, H., Oleson, J., and Wu, Y.-H. (2021). Toward a new evidence-based fitting paradigm for over-the-counter hearing aids. *American Journal of Audiology*, 30(1):43–66.

Valzolgher, C., Capra, S., Sum, K., Finos, L., Pavani, F., and Picinali, L. (2024). Spatial hearing training in virtual reality with simulated asymmetric hearing loss. *Scientific Reports*, 14(1):2469.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C., Joris, P. X., Heinz, M. G., and Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing research*, 377:109–121.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, 66(5):1364–1380.

Villard, S., Perrachione, T. K., Lim, S.-J., Alam, A., and Kidd, G. (2023). Energetic and informational masking place dissociable demands on listening effort: Evidence from simultaneous electroencephalography and pupillometry. *The Journal of the Acoustical Society of America*, 154(2):1152–1167.

Wang, M., Ai, Y., Han, Y., Fan, Z., Shi, P., and Wang, H. (2021). Extended high-frequency audiometry in healthy adults with different age groups. *Journal of Otolaryngology-Head & Neck Surgery*, 50(1):52.

Wei, Y., Gan, L., and Huang, X. (2022). A review of research on the neurocognition for timbre perception. *Frontiers in Psychology*, 13:869475.

Wendt, D., Hietkamp, R. K., and Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and hearing*, 38(6):690–700.

Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and hearing*, 37:35S–43S.

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and hearing*, 36(4):e153–e165.

Woodall, A. and Liu, C. (2013). Effects of signal level and spectral contrast on vowel formant discrimination for normal-hearing and hearing-impaired listeners. *American journal of audiology*, 22(1):94–104.

Wu, Y.-J. and Tokuda, K. (2009). Minimum generation error training by using original spectrum as reference for log spectral distortion measure. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4013–4016. IEEE.

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and hearing*, 31(4):480–490.

Appendix A

Supplementary material

1 Multi-track excerpts used in Experiment 1

In this section we provide the list of songs used in Experiment 1. All quiet tracks during 8 second duration considered were discarded. Furthermore, all background vocal tracks were excluded for songs presented in the LAR block.

No.	Artist	Song	Tracks
1	Aimee Norwich	Child	1. Lead vocals 2. Bass 3. Drums 4. Piano 5. Wind instruments 6. Guitar
2	Alexander Ross	Bolero	1. Lead vocals 2. Bass 3. Drums 4. String instruments 5. 2 × Guitar 6. Etno 7. Wind instruments
3	Clara Berry and Wooldog	Stella	1. Lead vocals 2. Bass 3. Drums 4. 2 × Percussion 5. Piano 6. Synth instruments
4	Clara Berry and Wooldog	Air traffic	1. Lead vocals 2. Bass 3. Drums 4. 2 × Guitar 5. Percussion 6. Piano
5	A Classic Education	Night Owl	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Synth instruments
6	Little Tybee	The Alchemist	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. String instruments 6. Keys
7	Berlin	Roads	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Synth instruments 6. Keys
8	Dhaka Band	Soldier Man	1. Lead vocals 2. Bass 3. Drums 4. Guitar
9	Family Band	Again	1. Lead vocals 2. Bass 3. Drums 4. Etno 5. 4 × Guitar
10	Mutual Benefit	Not for nothing	1. Lead vocals 2. Bass 3. Drums 4. Piano 5. Guitar 6. String instruments

Table 1: Songs taken from the medeley database for the LAR block of Experiment 1

No.	Artist	Song	Tracks
1	Patrick Talbot	Fool	1. Lead vocals 2. Background vocals 3. Bass 4. Drums 5. 4 × Guitar 6. Acoustic Guitar 7. Keys 8. Percussion
2	Robert Hammon	The Elephant	1. Lead vocals 2. 3 × Background vocals 3. Bass 4. Drums 5. Guitar 6. Synth instruments
3	James May	Hold on you	1. Lead vocals 2. Bass 3. Drums 4. 2 × Guitar
4	Fruit Cathedral	Keep me running	1. Lead vocals 2. Background vocals 3. Bass 4. Drums 5. Guitar 6. Percussion
5	Liz Nelson	Cold war	1. Lead vocals 2. Guitar
6	Liz Nelson	Rainfall	1. Lead vocals 2. Background vocals 3. Guitar
7	Music Delta	Brit Pop	1. Lead vocals 2. Background vocals 3. Bass 4. Drums 5. Guitar
8	Music Delta	Beatles	1. Lead vocals 2. Drums 3. Guitar
9	Night Panther	Fire	1. Lead vocals 2. Background vocals 3. Bass 4. Drums 5. Synth instruments 6. Brass instruments 7. Keys 8. String instruments
10	Secret Mountain	High Horse	1. Lead vocals 2. Background vocals 3. Bass 4. Drums 5. Guitar 6. Piano 7. Pad

Table 2: Songs taken from the medeley database for the spectral blocks of Experiment 1

2 Multi-track excerpts used in Experiment 2

In this section we provide the list of songs used in Experiment 2. All quiet tracks during 8 second duration considered were discarded. Furthermore, all background vocal tracks were excluded.

No.	Artist	Song	Database	Tracks
1	Angela Thomas Wade	Milk Cow Blue	Cambridge MT	1. Lead vocals 2. Kick 3. Snare 4. Toms 5. Overheads 6. Bass 7. Guitar 8. Fiddle 9. Piano
2	Music Delta	Beatles	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
3	Avalon	All I know	Cambridge MT	1. Lead vocals 2, Kick 3. Snare 4. HiHat 5. Drums 6. Bass 7. 2 × Guitar 8. Piano 9. Synth instruments
4	Music Delta	Country 1	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
5	Enda Reilly	An Nasc Nua	Cambridge MT	1. Lead vocals 2. Bass 3. Guitar 4. Fiddles
6	Music Delta	Disco	Medley	1. Lead vocals 2. Bass 3. Guitar 4. Fiddle
7	Spektakulatius	Jeden Winter	Cambridge MT	1. Lead vocals 2. Kick 3. Snare 4. Overheads 5. Toms 6. 2 × Bass 7. Piano
8	Music Delta	Grunge	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
9	Finlay	Same kind of love	Cambridge MT	1. Lead vocals 2. 2 × Kick 3. Snare 4. 3 × Toms 5. Cymbal 6. Bass 7. 3 × Guitar 8. 2 × Piano 9. Organ
10	Music Delta	Rockabilly	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar

Table 3: Song list 1 used in Experiment 2.

No.	Artist	Song	Database	Tracks
1	Angels in Amplifiers	Im Alright	Cambridge MT	1. Lead vocals 2. Kick 3. Snare 4. Overheads 5. Toms 6. Percussion 7. Bass 8. Piano 9. 2 × Guitar
2	Clara berry and Wooldog	Air Traffic	Medley	1. Lead vocals 2. Bass 3. Drums 4. 2 × Guitar 5. Percussion 6. Piano
3	Berlin	Roads	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Synth instruments 6. Keys
4	Vieux	Farka Joure Ana	Medley	1. Lead vocals 2. Bass 3. Drums 4. 2 × Guitar
5	Night Panther	Fire	Medley	1. Lead vocals 2. Bass 3. Drums 4. Percussion 5. Brass 6. Keys 7. Strings 8. String Instruments
6	Egda Carloyn	Saudade Do Teu Beijo	Cambridge MT	1. Lead vocals 2. Loop 3. Kick 4. Snare 5. Cowbell 6. Shaker 7. Bass 8. Guitar 9. Synth instruments
7	Celestial Shore	Die for us	Medley	1. Lead vocals 2. Bass 3. Drums 4. Etno 5. 2 × Guitar 6. 2 × Synth instruments
8	Arise	Run Run Run	Cambridge MT	1. Lead vocals 2. Kick 3. Rim 4. HiHat 5. Overheads 6. 2 × Snare 7. Bass 8. 2 × Guitar 9. 2 × Organ 10. Piano
9	Liz Nelson	Rainfall	Medley	1. Lead vocals 2. Guitar
10	Patrick Talbot	Fool	Medley	1. Lead vocals 2. Bass 3. Drums 4. 5 × Guitar 5. Percussion 6. Keys

Table 4: Song list 2 used in Experiment 2.

No.	Artist	Song	Database	Tracks
1	Cat Martino	I promise	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Pad 6. Fx
2	Dead Milkmen	Prisoners cinema	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Percussion 6. Pad 7. Fx
3	Fruit Cathedral	Keep me running	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
4	Midnight Blue	Hunteing	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Piano
5	Midnight Blue	Stars are screaming	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
6	Justin Myles	Alone with you	Cambridge MT	1. Lead vocals 2. Kick 3. Snare 4. Overheads 5. Bass 6. 2 × Guitar
7	Peter Mathew Baeuer	You always look for someone else	Medley	1. Lead vocals 2. Bass 3. Guitar 4. Percussion 5. Pad
8	Steven Clark	Bounty	Medley	1. Lead vocals 2. Bass 3. Drums 4. Piano 5. Pad 6. Strings
9	Strand of Oaks	Space station	Medley	1. Lead vocals 2. Bass 3. Drums 4. Percussion 5. Piano 6. Synth instruments
10	The Districts	Vermont	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Keys

Table 5: Song list 3 used in Experiment 2.

No.	Artist	Song	Database	Tracks
1	Trevor and the Sound Waves	Alone and sad	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
2	Tourist	Kin	Medley	1. Lead vocals 2. Bass 3. Drums 4. Percussion 5. String instruments 6. Synth instruments 7. Piano 8. Fx
3	The Scarlet Band	Les Fleuers Du Mal	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
4	The Kitchenettes	Alive	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Piano 6. String instruments
5	Sweet Lights	You let me down	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Percussion 6. Piano 7. Synth instruments
6	Snowmine	Curfews	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Percussion 6. Pad 7. Keys
7	Purling Hiss	Lolita	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
8	Casandra Jenkins	Perfect Day	Medley	1. Lead vocals 2. Bass 3. Guitar 4. Percussion 5. Piano 6. String insruments 7. Wind instruments
9	Filthy Bird	Like to know	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Percussion
10	Lewis and Clarke	The Silver Sea	Medley	1. Lead vocals 2. Drums 3. Guitar 4. Keys 5. Pad 6. String instruments

Table 6: Song list 4 used in Experiment 2.

No.	Artist	Song	Database	Tracks
1	Music Delta	80s Rock	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
2	Anna Blanton	Rachel	Cambridge MT	1. Lead vocals 2. Congas 3. Bass 4. 2 × Ukelele 5. Viola 6. Cello
3	Music Delta	Brit Pop	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
4	Eddie Garrido	Una Semata Sin Ti	Cambridge MT	1. Lead vocals 2. Percussion 3. Wood block 4. Toms cymbals 5. Bass 6. Piano 7. Vibes 8. Strings 9. French horns
5	Music Delta	Country 2	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
6	Enda Reilly	Cur An Long Ag Seol	Cambridge MT	1. Lead vocals 2. Bass 3. Drums 4. Brushes 5. Fiddle 6. Mandolin 7. Guitar
7	Music Delta	Gospel	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Percussion
8	Spektakulatius	Wayfaring Stranger	Cambridge MT	1. Lead vocals 2. Kick 3. Snare 4. Overheads 5. Toms 6. 2 × Bass 7. Piano 8. Saxophone
9	Music Delta	Hendrix	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
10	Speak softly	Broken man	Cambridge MT	1. Lead vocals 2. 2 × Drums 3. Percussion 4. Piano 5. 3 × Rhodes 6. 3 × Synth instruments

Table 7: Song list 5 used in Experiment 2.

No.	Artist	Song	Database	Tracks
1	Aimee Norwich	Child	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Piano 6. Wind instruments
2	Alexander Ross	Bolero	Medley	1. Lead vocals 2. Bass 3. Drums 4. Strings 5. 2 × Guitar 6. Etno 7. Winds
3	Clara Berry and Wool dog	Stella	Medley	1. Lead vocals 2. Bass 3. Drums 4. 2 × Percussion 5. Piano 6. Synth instruments
4	Jay Menon	Through my eyes	Cambridge MT	1. Lead vocals 2. Bass 3. Kick 4. Snare 5. HiHat 6. Overheads 7. 2 × Cymbal Roll 8. 5 × Guitar 9. Piano 10. 2 × Pads
5	A Classic Education	Night Owl	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Synth instruments
6	Little Tybee	The Alchemist	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar 5. Strings 6. Keys
7	Mike Senior	Mystery	Cambridge MT	1. Lead vocals 2. 2 × Bass 3. 4 × Drums 4. Guitar 5. Synth instruments
8	Dhaka Band	Soldier Man	Medley	1. Lead vocals 2. Bass 3. Drums 4. Guitar
9	Family Band	Again	Medley	1. Lead vocals 2. Bass 3. Drums 4. Etno 5. 4 × Guitar
10	Mutual Benefit	Not for nothing	Medley	1. Lead vocals 2. Bass 3. Drums 4. Piano 5. Guitar 6. Strings

Table 8: Song list 6 used in Experiment 2.

3 p-value maps for pooled group for inter-excerpt comparison in Experiment 1

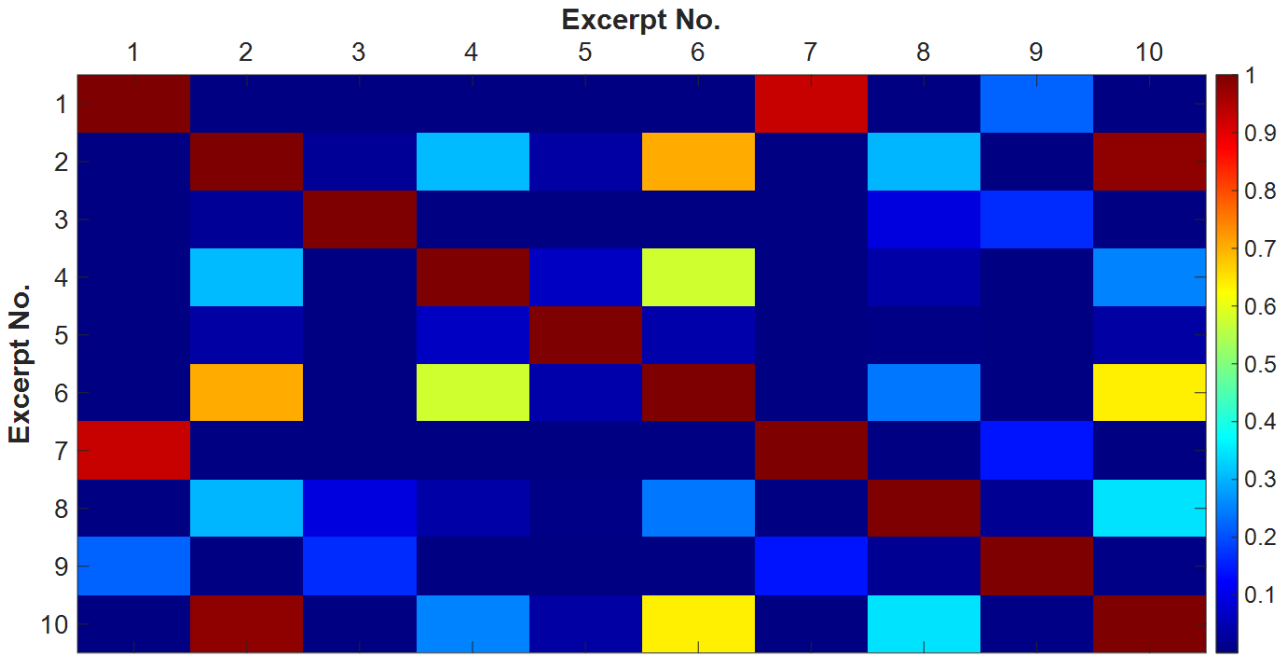


Figure S1: The p-values of a paired t-test applied on the pooled participant LAR preferences for each excerpt.

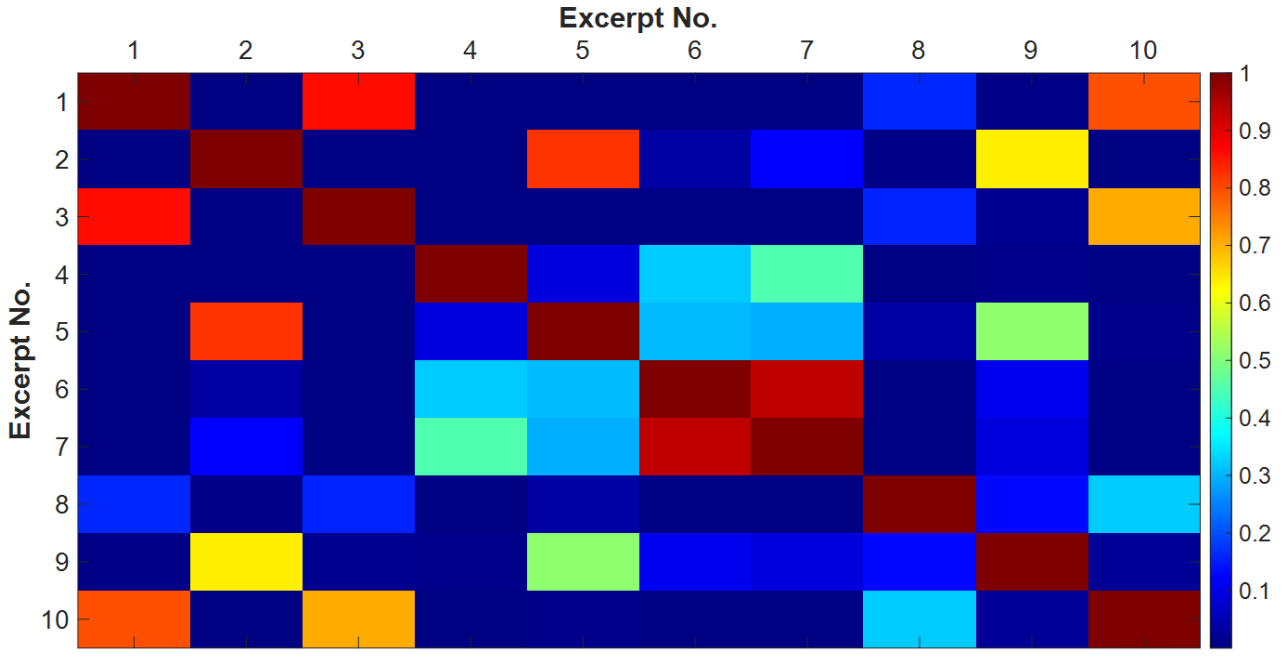


Figure S2: The p-values of a paired t-test applied on the pooled participant Spectral Balance preferences for each excerpt.

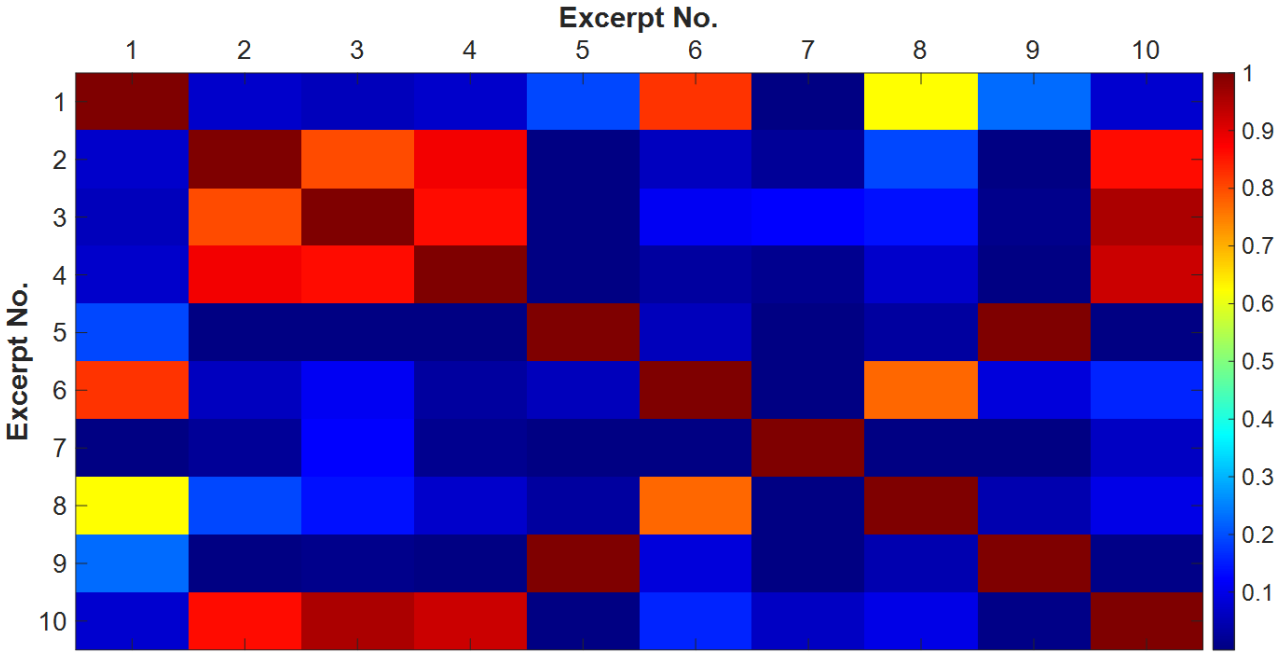


Figure S3: The p-values of a paired t-test applied on the pooled participant EQ-transform preferences for each excerpt.

4 Raw error residual plots for the linear mixed effects (LME) model used in Experiment 1

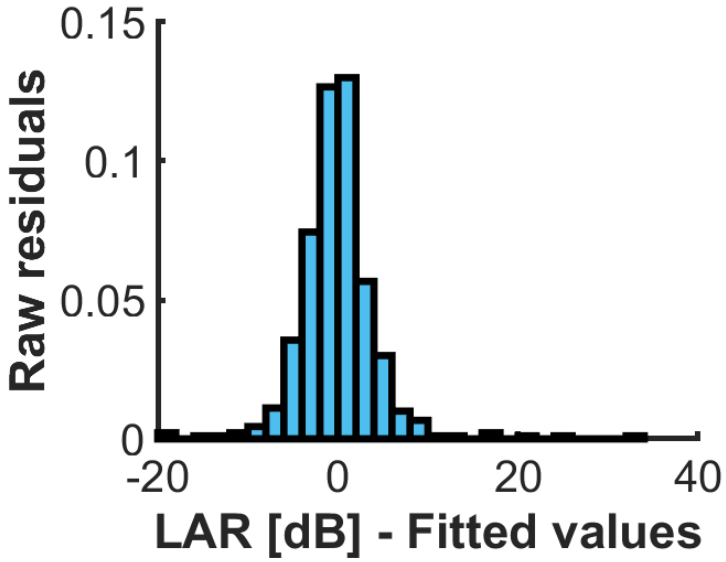


Figure S4: Histogram of the error residuals of the fitted data for LAR preferences using the linear mixed effects model.

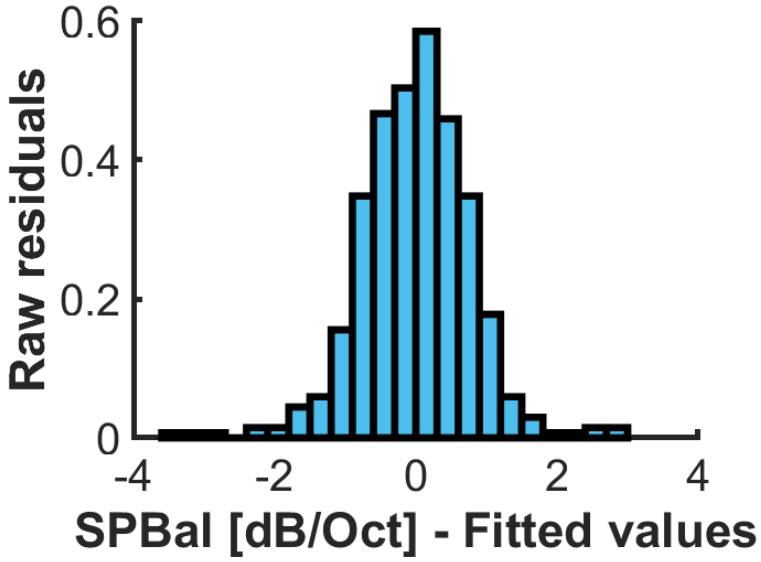


Figure S5: Histogram of the error residuals of the fitted data for SPBal preferences using the linear mixed effects model.

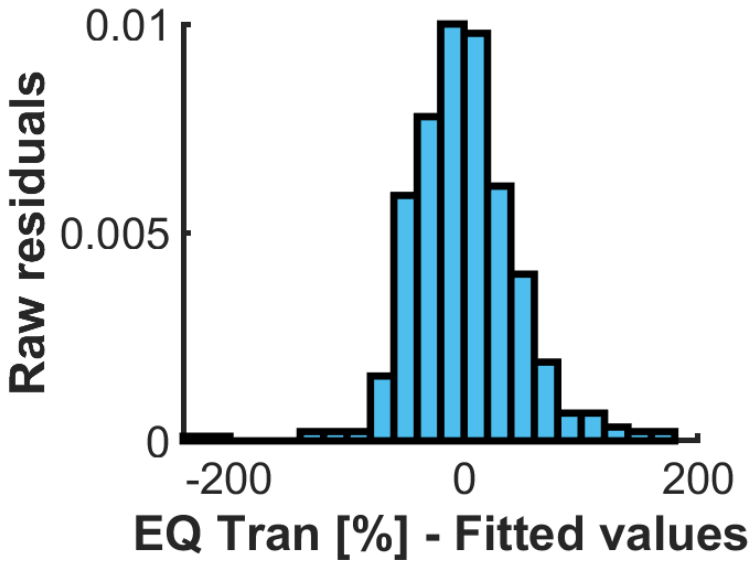


Figure S6: Histogram of the error residuals of the fitted data for EQ-transform preferences using the linear effects model.

5 Sound pressure levels of excerpts at listener/ participant position

In this section, the A-weighted minimum (LAFmin) and maximum (LAFmax) sound pressure levels using the fast time weighting (as per IEC 61672) measured for each of the songs used in the study are presented. The songs were looped and the measurements were made within a 1 minute window. The means and standard deviations included for LAFmax are those taken across excerpts. All of measurements were made at the participant position with the Nor140 Precision Sound Analyser from **Norsonic AS** (`https://web2.norsonic.com`).

5.1 Experiment 1

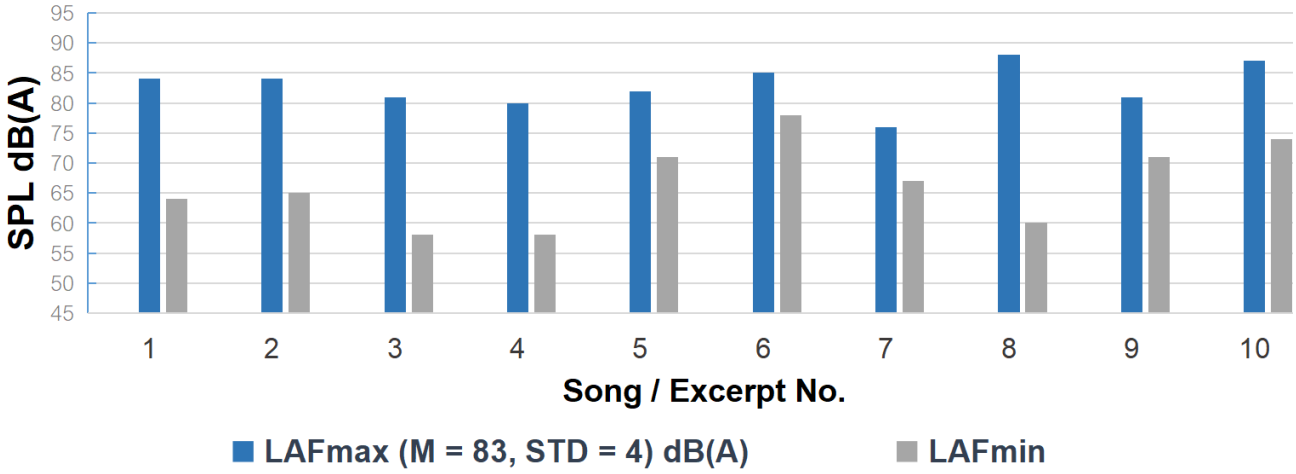


Figure S7: Sound pressure measurements for the stimuli presented in the level block.

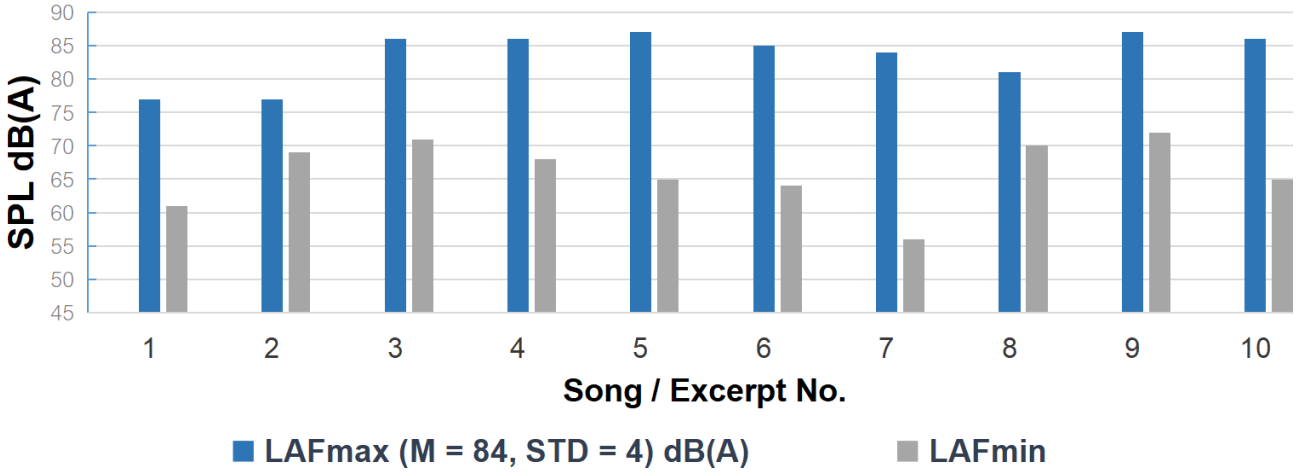


Figure S8: Sound pressure measurements for the stimuli presented in the spectral block.

5.2 Experiment 2

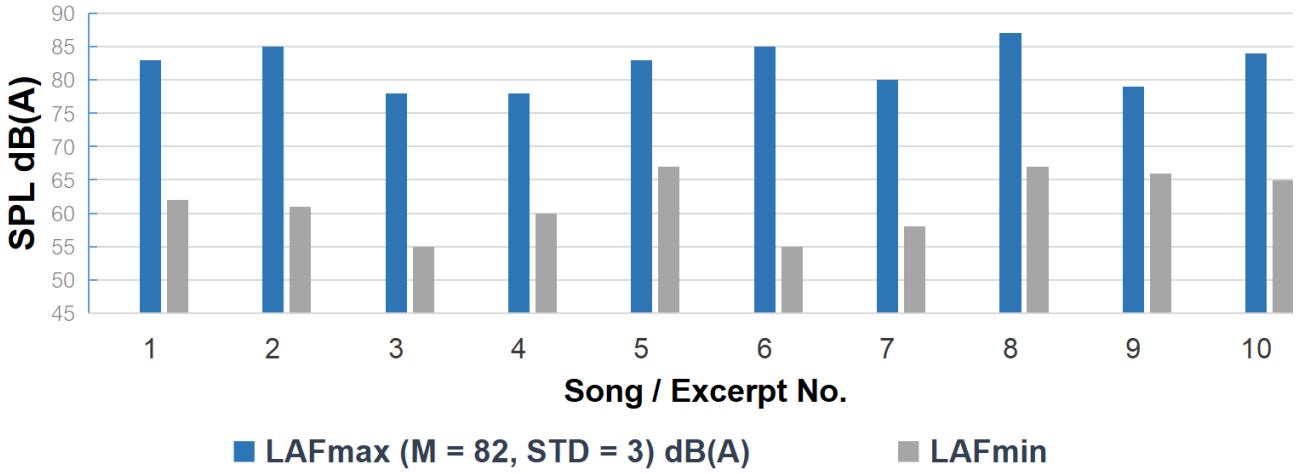


Figure S9: Sound pressure measurements for the stimuli presented in Song list 1.

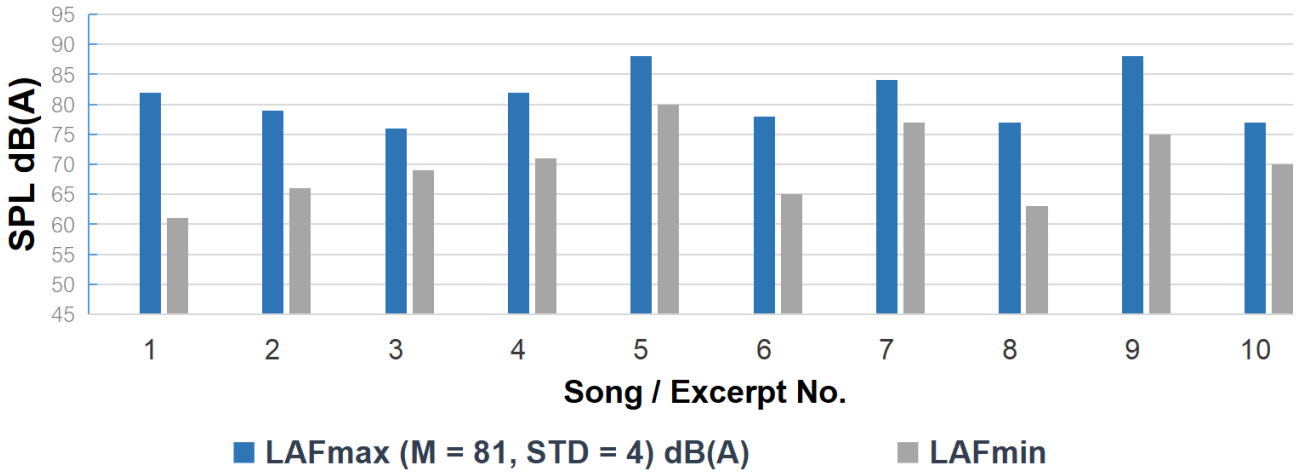


Figure S10: Sound pressure measurements for the stimuli presented in Song list 2.

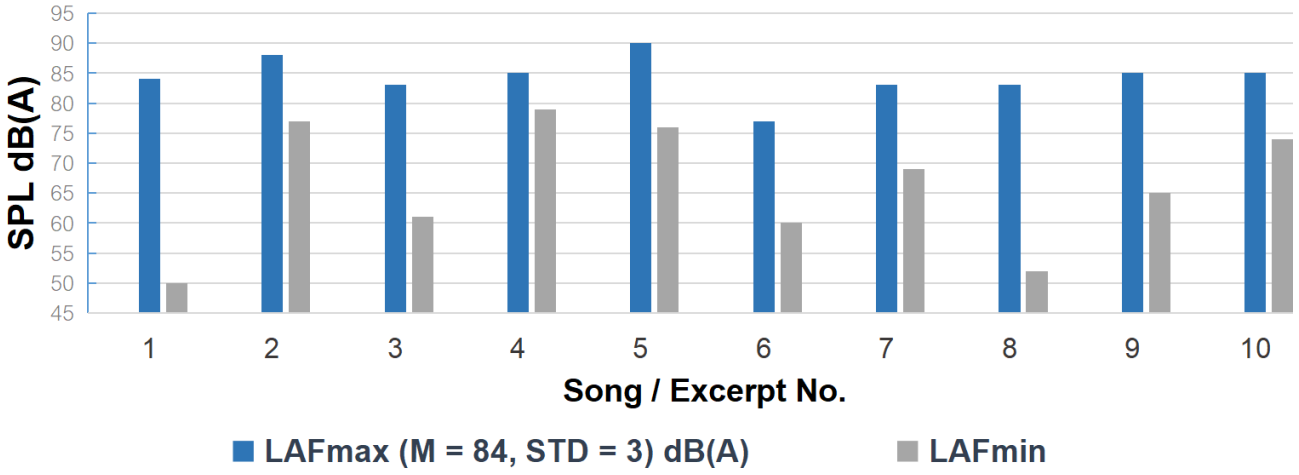


Figure S11: Sound pressure measurements for the stimuli presented in Song list 3.

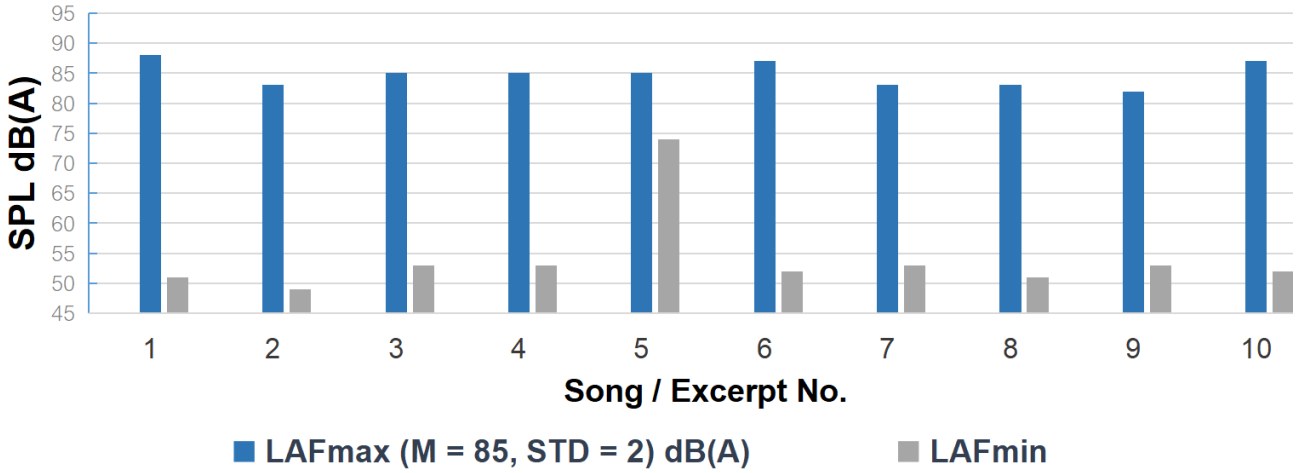


Figure S12: Sound pressure measurements for the stimuli presented in Song list 4.

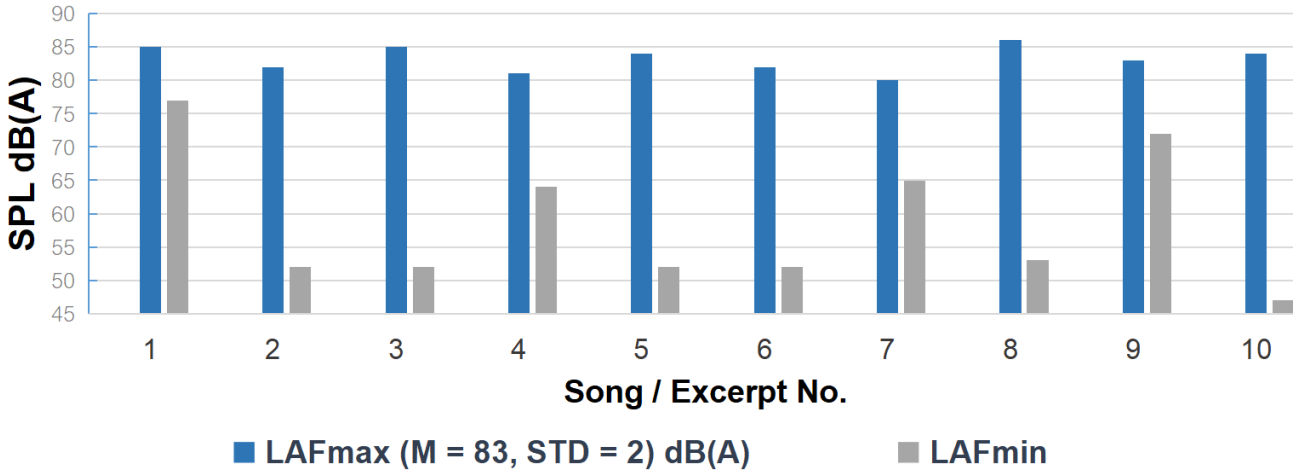


Figure S13: Sound pressure measurements for the stimuli presented in Song list 5.

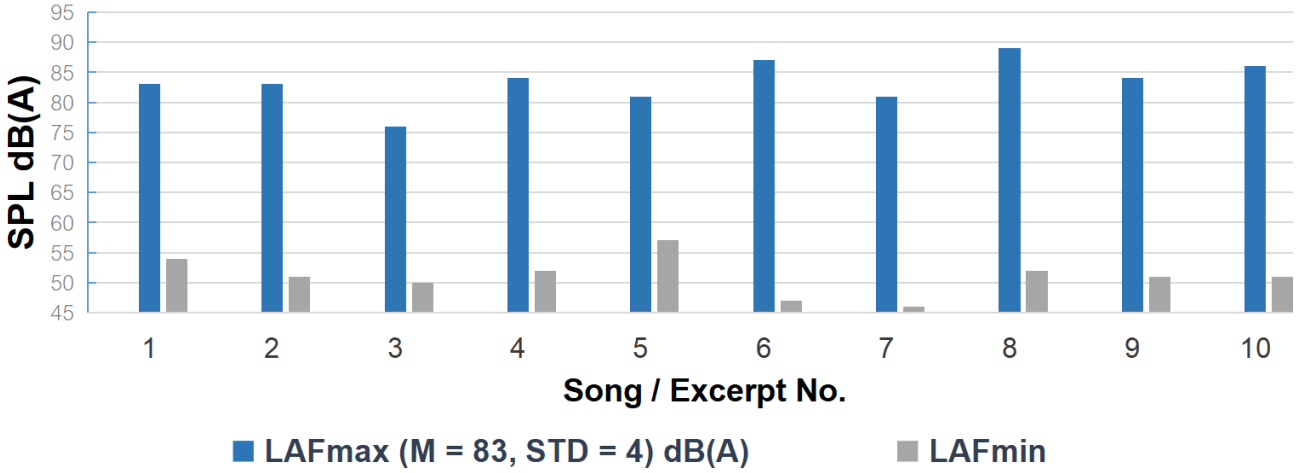


Figure S14: Sound pressure measurements for the stimuli presented in Song list 6.

6 Information about hearing aid use among wHA

6.1 Experiment 1

Participant No.	Gender	Hearing Aid type		Length of use [yrs.]	Mean HL [dB]		Mean preference		
					Left ear	Right ear	LAR [dB]	SPBal [dB/Oct]	EQ Tran [%]
1	M	BTE	*Open fit*	5	51.4	40.7	**-1.5**	**0.2**	**53**
2	M	BTE	*Closed fit*	2	23	32.1	**1.1**	**0**	**95**
3	M	BTE	*Open fit*	17	43	36	**3**	**1.3**	**111**
4	M	BTE	*Open fit*	4	25	30	**4.4**	**0.9**	**107**
5	M	BTE	*Closed fit*	8	40	37	**-3.4**	**-0.3**	**40**
6	M	BTE	*Closed fit*	10	53	55	**2.2**	**-0.2**	**22**
7	F	BTE	*Closed fit*	12	71	19.3	**-0.9**	**0**	**97**
8	M	BTE	*Open fit*	5	31	26	**3.9**	**0.2**	**82**
9	M	BTE	*Closed fit*	15	49	57	**3.6**	**-0.6**	**101**
10	F	*Data unavailable*			64	51	**4.4**	**-0.4**	**126**

Table 9: Individual hearing aid use, hearing loss, and mean preferences among wHA participants investigated in Experiment 1.

6.2 Experiment 2

Participant No.	Gender	Hearing Aid type		HA Length of use [yrs.]	Mean HL [dB]		Mean preference LAR [dB]		SPBal [dB/Oct]		EQ Tran [%]	
					Left ear	Right ear	wHA	woHA	wHA	woHA	wHA	woHA
1	M	**BTE**	*Open fit*	7	36.4	38.6	**-6**	**14.4**	**0.94**	**0.12**	**105**	**79**
2	M	**BTE**	*Open fit*	9	33.6	37.1	**7.4**	**11.6**	**-0.37**	**-0.3**	**56**	**176**
3	M	**BTE**	*Open fit*	11	35.7	32.9	**8.5**	**9.3**	**0.92**	**1.4**	**77**	**117**
4	M	**BTE**	*Closed fit*	18	43.6	37.1	**9.8**	**12.4**	**-1.38**	**1.08**	**78**	**200**
5	F	*Data unavailable*			45.7	38.6	**12.1**	**22.7**	**-1.33**	**0.7**	**67**	**186**
6	F	**BTE**	*Open fit*	11	26.4	25.7	**7.5**	**8.6**	**-0.2**	**0.4**	**158**	**119**
7	F	**BTE**	*Open fit*	7	43.6	43.6	**21**	**27.6**	**-2**	**-0.2**	**77**	**82**
8	F	**ITE**	*Full shell*	35	57.9	59.3	**-6.3**	**19.4**	**-1.76**	**-0.62**	**-93**	**346**
9	M	**BTE**	*Open fit*	5	51.4	56.4	**9.5**	**9.7**	**-0.88**	**0.6**	**104**	**229**
10	M	**BTE**	*Closed fit*	15	45	53.6	**5.6**	**6.8**	**-1.13**	**-0.5**	**116**	**-4**
11	M	**BTE**	*Closed fit*	5	53.6	52.1	**13.3**	**21.8**	**0.94**	**0.06**	**116**	**120**
12	M	**BTE**	*Closed fit*	16	55	64.3	**8.9**	**7.1**	**-1.44**	**1.4**	**55**	**-4**
13	M	**BTE**	*Open fit*	9	47.9	39.3	**5.6**	**7.7**	**-1.15**	**1.04**	**14**	**207**
14	F	**BTE**	*Open fit*	11	49.3	51.4	**-0.9**	**12.8**	**-2.4**	**0.1**	**158**	**159**
15	M	**BTE**	*Open fit*	1	44.3	40.7	**2.7**	**5.3**	**-1.2**	**0.8**	**101**	**80**
16	M	**BTE**	*Open fit*	4	55	62.9	**21.8**	**27**	**0.7**	**3**	**109**	**346**
17	M	**BTE**	*Open fit*	10	51.4	40	**13.9**	**4**	**0.76**	**-0.63**	**125**	**99**
18	F	**BTE**	*Open fit*	23	52.9	54.3	**10.3**	**4.3**	**-2.38**	**3**	**-93**	**279**

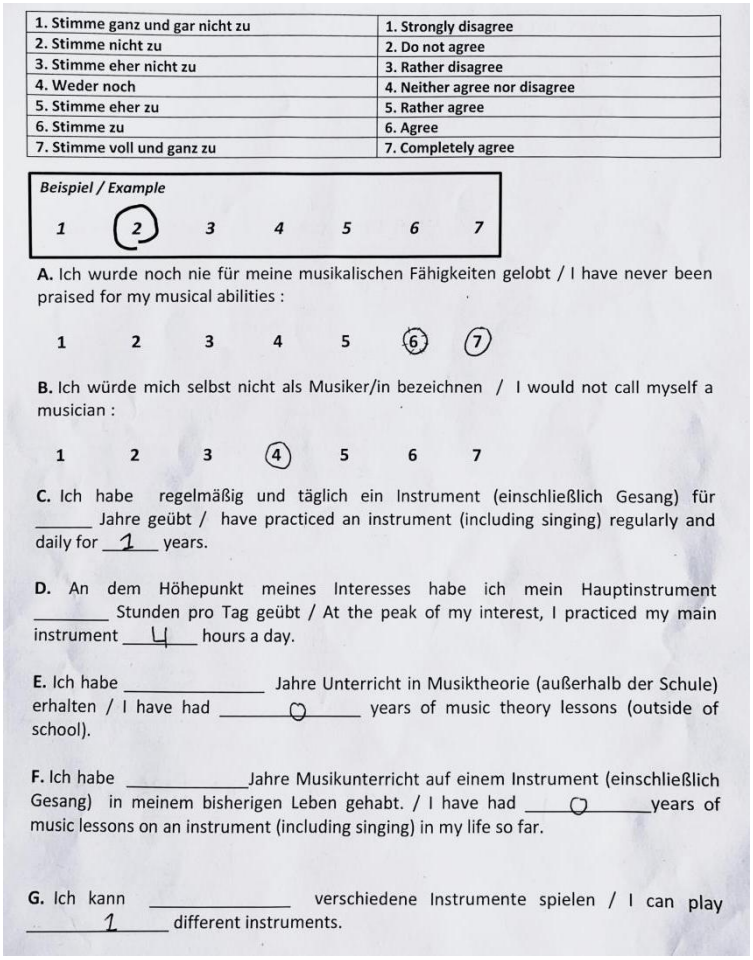
Table 10: Individual hearing aid use, hearing loss, and mean preferences among participants investigated in Experiment 2.

Appendix B

Supplementary Material 1

Evaluating Musical Training using the Goldsmith Musical Sophistication Index (Gold-MSI) Musical Training subscale Questionnaire.

Participant I



1. Stimme ganz und gar nicht zu	1. Strongly disagree
2. Stimme nicht zu	2. Do not agree
3. Stimme eher nicht zu	3. Rather disagree
4. Weder noch	4. Neither agree nor disagree
5. Stimme eher zu	5. Rather agree
6. Stimme zu	6. Agree
7. Stimme voll und ganz zu	7. Completely agree

Beispiel / Example

1 (2) 3 4 5 6 7

A. Ich wurde noch nie für meine musikalischen Fähigkeiten gelobt / I have never been praised for my musical abilities :

1 2 3 4 5 (6) (7)

B. Ich würde mich selbst nicht als Musiker/in bezeichnen / I would not call myself a musician :

1 2 3 (4) 5 6 7

C. Ich habe regelmäßig und täglich ein Instrument (einschließlich Gesang) für _____ Jahre geübt / have practiced an instrument (including singing) regularly and daily for __1__ years.

D. An dem Höhepunkt meines Interesses habe ich mein Hauptinstrument _____ Stunden pro Tag geübt / At the peak of my interest, I practiced my main instrument __4__ hours a day.

E. Ich habe _____ Jahre Unterricht in Musiktheorie (außerhalb der Schule) erhalten / I have had _____0_____ years of music theory lessons (outside of school).

F. Ich habe _____Jahre Musikunterricht auf einem Instrument (einschließlich Gesang) in meinem bisherigen Leben gehabt. / I have had _____0_____years of music lessons on an instrument (including singing) in my life so far.

G. Ich kann _____ verschiedene Instrumente spielen / I can play _____1_____ different instruments.

Figure 1 : Sample Gold-MSI questionnaire filled by Participant I who is a young normal hearing participant (yNH).

A. Participant chose ***(7) completely agree*** for : ***I have never been praised for my musical abilities:***
S(A) = **8 - 7 = 1**

B. Participant chose ***(4) neither agree nor disagree*** for : ***I would not call myself a musician:***
S(B) = **8 - 4 = 4**

C. Participant answered : ***x = 1 year*** for : ***I have practiced an instrument (including singing) regularly for ___x__ years***

Score of S(C) = x + 1 = 2 if x ≤ 3

Score of S(C) = 5 if 4 ≤ x ≤ 5

Score of S(C) = 6 if 6 ≤ x ≤ 9

Score of S(C) =7 if x ≥ 10

D. Participant answered : ***x = 4 hours a day*** for : ***At the peak of my interest, I practiced my main instrument ___x__ hours a day***

Score of S(D) = 1 if x = 0

Score of S(D) = 2 if 0 < x ≤ 0.5

Score of S(D) = 3 if x = 1

Score of S(D) = 4 if 1 < x ≤ 1.5

Score of S(D) = 5 if x =2

Score of S(D) = 6 if 3 ≤ x ≤ 4

Score of S(D) = 7 if x ≥ 5

E. Participant answered : ***x = 0 years*** for : ***I have had __x__ years of music theory lessons (outside of school).***

Score of S(E) = 1 if x = 0

Score of S(E) = 2 if x = 0.5

Score of S(E) = 3 if 0.5 < x ≤ 1

Score of S(E) = 4 if 1 < x ≤ 2

Score of S(E) = 5 if 2 < x ≤ 3

Score of S(E) = 6 if 4 ≤ x ≤ 6

Score of S(E) = 7 if x ≥ 7

F. Participant answered : ***x = 0 year*** for : ***I have had __x__ years of music lessons of an instrument (including singing) in my life so far.***

Score of S(F) = 1 if x < 6

Score of S(F) = 2 if 0.5 ≤ x < 1

Score of S(F) = 3 if 1 ≤ x < 2

Score of S(F) = 4 if 2 ≤ x < 3

Score of S(F) = 5 if 3 ≤ x < 6

Score of S(F) = 6 if 6 ≤ x < 10

Score of S(F) = 7 if x ≥ 10

G. Participant answered : ***x = 1 instrument*** for : ***I can play __x__ different instruments***

Score of S(F) = 1 + x = 2 if x < 6

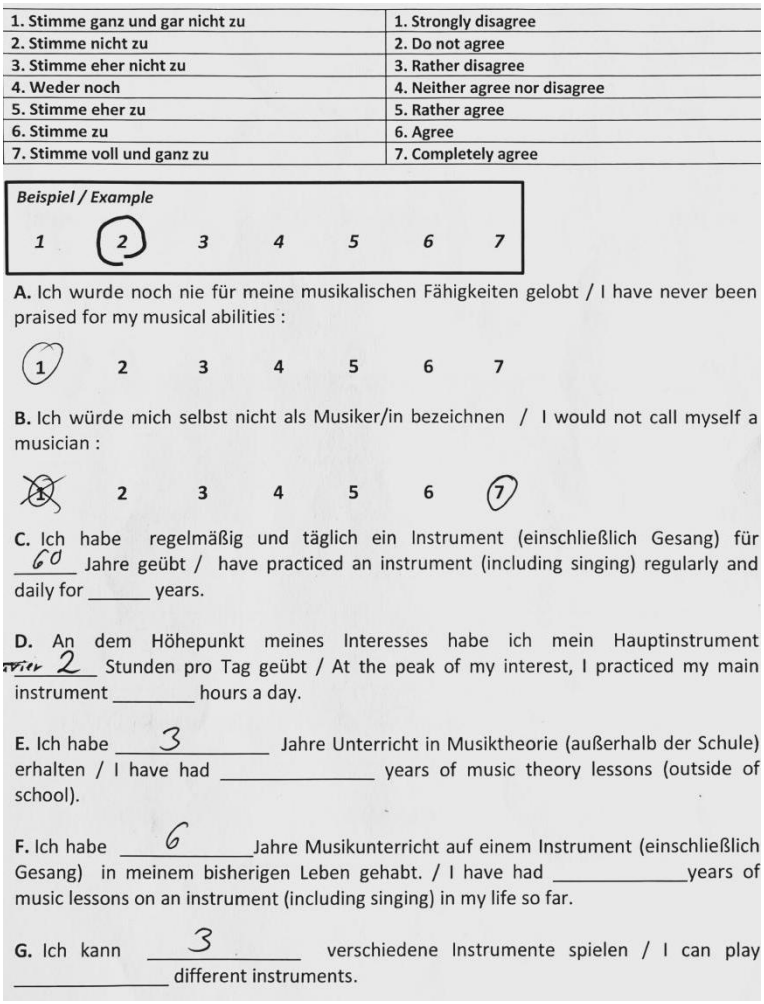
Score of S(F) = 7 if x ≥ 6

Musical Training score of participant I :

Participant I who is yNH has a musical training score of :

$$S(A) + S(B) + S(C) + S(D) + S(E) + S(F) + S(G) = 1 + 4 + 2 + 6 + 1 + 1 + 2 = 17$$

Participant II



1. Stimme ganz und gar nicht zu	1. Strongly disagree
2. Stimme nicht zu	2. Do not agree
3. Stimme eher nicht zu	3. Rather disagree
4. Weder noch	4. Neither agree nor disagree
5. Stimme eher zu	5. Rather agree
6. Stimme zu	6. Agree
7. Stimme voll und ganz zu	7. Completely agree

Beispiel / Example

1 (2) 3 4 5 6 7

A. Ich wurde noch nie für meine musikalischen Fähigkeiten gelobt / I have never been praised for my musical abilities :

(1) 2 3 4 5 6 7

B. Ich würde mich selbst nicht als Musiker/in bezeichnen / I would not call myself a musician :

(1) 2 3 4 5 6 (7)

C. Ich habe regelmäßig und täglich ein Instrument (einschließlich Gesang) für 60 Jahre geübt / have practiced an instrument (including singing) regularly and daily for _____ years.

D. An dem Höhepunkt meines Interesses habe ich mein Hauptinstrument vier 2 Stunden pro Tag geübt / At the peak of my interest, I practiced my main instrument _____ hours a day.

E. Ich habe 3 Jahre Unterricht in Musiktheorie (außerhalb der Schule) erhalten / I have had _____ years of music theory lessons (outside of school).

F. Ich habe 6 Jahre Musikunterricht auf einem Instrument (einschließlich Gesang) in meinem bisherigen Leben gehabt. / I have had _____ years of music lessons on an instrument (including singing) in my life so far.

G. Ich kann 3 verschiedene Instrumente spielen / I can play _____ different instruments.

Figure 2 : Sample Gold-MSI questionnaire filled by Participant II who is an older hearing impaired participant (oHI).

A. Participant chose ***(1) Strongly disagree*** for : ***I have never been praised for my musical abilities:***
S(A) = **8** - 1 = **7**

B. Participant chose ***(1) Strongly disagree*** for : ***I would not call myself a musician:***
S(B) = **8** - 1 = **7**

C. Participant answered : $x = 60$ ***years*** for : ***I have practiced an instrument (including singing) regularly for ___x__ years***

$$\textit{Score of } S(C) = x + 1 \quad \textit{if } x \leq 3$$

$$\textit{Score of } S(C) = 5 \quad \textit{if } 4 \leq x \leq 5$$

$$\textit{Score of } S(C) = 6 \quad \textit{if } 6 \leq x \leq 9$$

$$\textit{Score of } S(C) = 7 \quad \textit{if } x \geq 10$$

D. Participant answered : ***x = 2 hours a day*** for : ***At the peak of my interest, I practiced my main instrument ___x__ hours a day***

Score of S(D) = 1 if x = 0

Score of S(D) = 2 if 0 < x ≤ 0.5

Score of S(D) = 3 if x = 1

Score of S(D) = 4 if 1 < x ≤ 1.5

Score of S(D) = 5 if x =2

Score of S(D) = 6 if 3 ≤ x ≤ 4

Score of S(D) = 7 if x ≥ 5

E. Participant answered : ***x = 3 years*** for : ***I have had __x__ years of music theory lessons (outside of school).***

Score of S(E) = 1 if x = 0

Score of S(E) = 2 if x = 0.5

Score of S(E) = 3 if 0.5 < x ≤ 1

Score of S(E) = 4 if 1 < x ≤ 2

Score of S(E) = 5 if 2 < x ≤ 3

Score of S(E) = 6 if 4 ≤ x ≤ 6

Score of S(E) = 7 if x ≥ 7

F. Participant answered : ***x = 6 years*** for : ***I have had __x__ years of music lessons of an instrument (including singing) in my life so far.***

Score of S(F) = 1 if x < 6

Score of S(F) = 2 if 0.5 ≤ x < 1

Score of S(F) = 3 if 1 ≤ x < 2

Score of S(F) = 4 if 2 ≤ x < 3

Score of S(F) = 5 if 3 ≤ x < 6

Score of S(F) = 6 if 6 ≤ x < 10

Score of S(F) = 7 if x ≥ 10

G. Participant answered : ***x = 3 instruments*** for : ***I can play __x__ different instruments***

Score of S(F) = 1 + x = 4 if x < 6

Score of S(F) = 7 if x ≥ 6

Musical Training score of participant II :

Participant II who is oHI has a musical training score of :

$$S(A) + S(B) + S(C) + S(D) + S(E) + S(F) + S(G) = 7 + 7 + 7 + 5 + 5 + 6 + 4 = 41$$

Remarks : Based on the above assessment oHI Participant II is more musically trained than yNH Participant I.

Contrary to the specimen examples, the musical training scores of yNH ($M = 19, SD = 9.9$) was significantly higher than that among oHI ($M = 13, SD = 9$), $p = 0.03$, $d = 0.6$ *(medium effect)*. Figure 3 shows the mean and 95 % confidence intervals of Gold MSI scores for the participant groups.

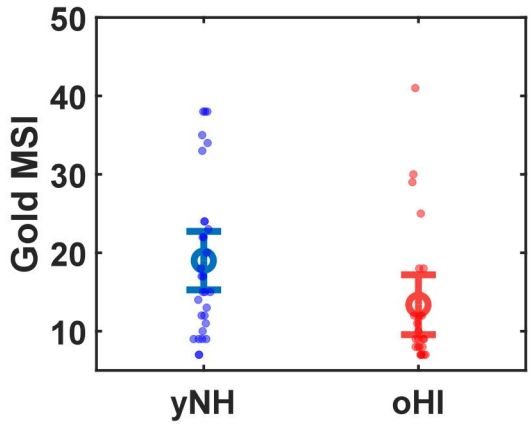


Figure 3 : Mean musical training scores and 95 % confidence intervals for yNH and oHI. Individual scores are provided alongside.

Supplementary Material 2

The transformed equalization or EQ-transform (EQT) process

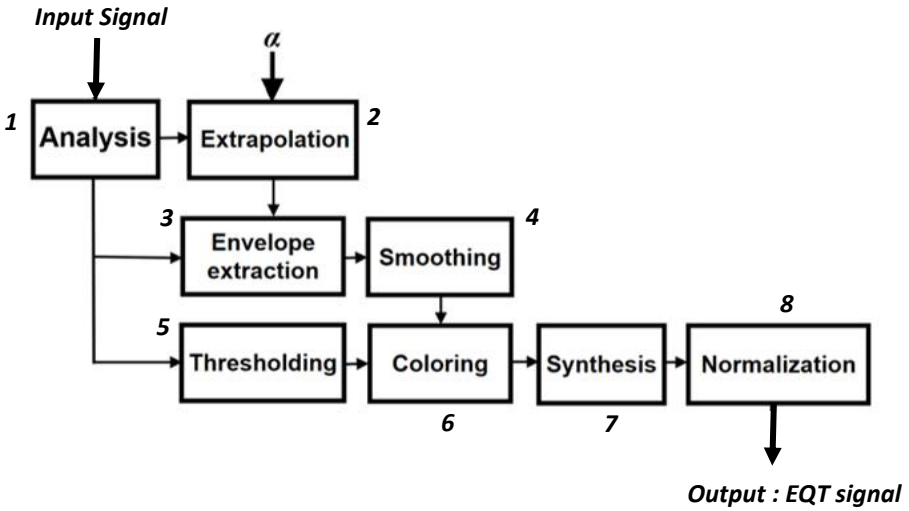


Figure 1 : Block diagram of the step-by-step process of the EQ-transform applied on a single track. The order in which each step is applied is numbered.

Step 1 (Analysis) :

The a fast-fourier transform is applied on the full 2 second duration of the input signal with rectangular windowing applied and no zero padding to calculate the input power spectrum $\mathbf{Or}(\boldsymbol{f})$. The sampling frequency used is 44.1 kHz

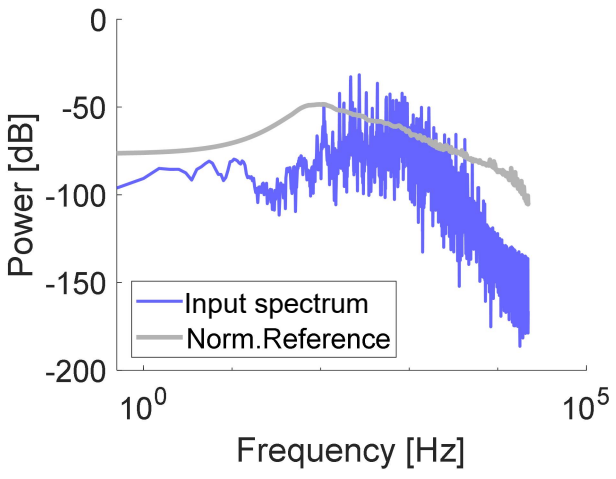


Figure 2 : The power spectrum of a 2 second piano track evaluated along with the energy-normalzed reference spectrum which has the same total energy as the input signal.

Step 2 (Extrapolation) :

The energy-normalized smooth reference spectrum **Ref(*f*)** which has the same total energy as the input power spectrum $\mathbf{Or}(f)$ calculated in step 1, and $\mathbf{Or}(f)$ are used to calculate the transformed spectrum. This is done by linearly interpolating between the two spectra as shown in Eq.1 :

$$\mathbf{Tr}(\boldsymbol{\alpha}, f) = \frac{\mathbf{Or}(f) - \mathbf{Ref}(f)}{\mathbf{100}} \times \boldsymbol{\alpha} + \mathbf{Ref}(f) \qquad \text{(Eq .1)}$$

In Eq.1, **T(.)** represents the transformed spectrum obtained by interpolating between the original power spectrum **Or(*f*)** and the energy-normalized reference spectrum **Ref(*f*)**. The factor $\boldsymbol{\alpha}$ is the desired degree of EQT as a percentage. $\mathbf{Tr}(\alpha, f)$ is the transformed power spectrum over frequencies f at an EQT of $\boldsymbol{\alpha}$ **%.**

Step 3 (Evelope extraction) :

The power difference between the transformed spectrum and the original that is :

$$\Delta\mathbf{E}_{\mathbf{Noisy}}(\boldsymbol{\alpha}, f) = \mathbf{Tr}(\boldsymbol{\alpha}, f) - \mathbf{Or}(f) \qquad \text{(Eq .2)}$$

is evaluated in this step. Figure 3 shows such a power difference which is a noisy sequence.

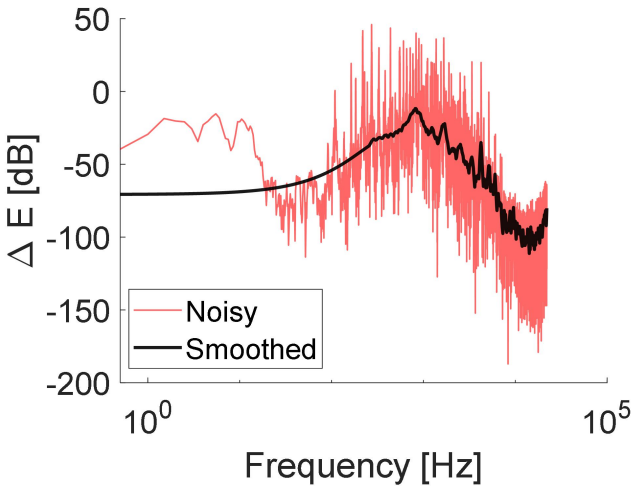


Figure 3 : The difference between the transformed and original power spectra. Both noisy and the Savitsky-Golay filtered (Smoothed) power differences are shown for an α of 300 %.

Step 4 (Smoothing) :

The noisy power difference derived in **Step 3** is smoothed using a ***Savitsky - Golay*** filter. Figure 3 shows the noisy $\Delta\mathbf{E}_{\mathbf{Noisy}}(\boldsymbol{\alpha}, \boldsymbol{f})$ and smoothed version $\Delta\mathbf{E}_{\mathbf{Smoothed}}(\boldsymbol{\alpha}, \boldsymbol{f})$ of the power difference.

Step 5 (Thresholding) :

Significant bands $\hat{f}$ where the power is at most 90 dB less than the global maximum of $\mathbf{Or}(\boldsymbol{f})$ is evaluated within the audible range [20 Hz, 20 kHz]. That is :

$$\hat{\boldsymbol{f}} \subseteq \boldsymbol{f}, \textit{ where } \quad \boldsymbol{Or}(\hat{\boldsymbol{f}}) \geq \boldsymbol{max}\{\boldsymbol{O}(\boldsymbol{f})\} - \mathbf{90} \qquad \text{(Eq .3)}$$

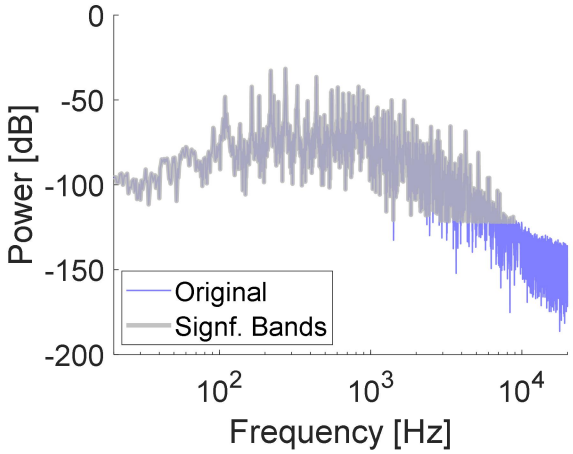


Figure 4 : The significant bands where the energy is at most 90 dB less than the global maximum within the audible range.

Step 6 (Coloring) :

The smoothed power difference is used to color $\mathbf{Or}(\boldsymbol{f})$ in the significant bands to calculate the frequency domain representation of the final EQT signal or the EQT spectrum $\tilde{\mathbf{Tr}}\,(\boldsymbol{\alpha}, \boldsymbol{f})$, as shown in Eq. 4 :

$$\tilde{\mathbf{Tr}}\,(\boldsymbol{\alpha}, \boldsymbol{f}) = \Delta\mathbf{E}_{\mathbf{Smoothed}}(\boldsymbol{\alpha}, \boldsymbol{f}) + \mathbf{Or}(\boldsymbol{f}), \textit{ if } \quad \boldsymbol{f} = \hat{\boldsymbol{f}} \qquad \text{(Eq . 4)}$$

$$\tilde{\mathbf{Tr}}\,(\boldsymbol{\alpha}, \boldsymbol{f}) = \mathbf{Or}(\boldsymbol{f}), \textit{ otherwise}$$

Figure 5 shows an illustration of the transformed spectrum derived for this example.

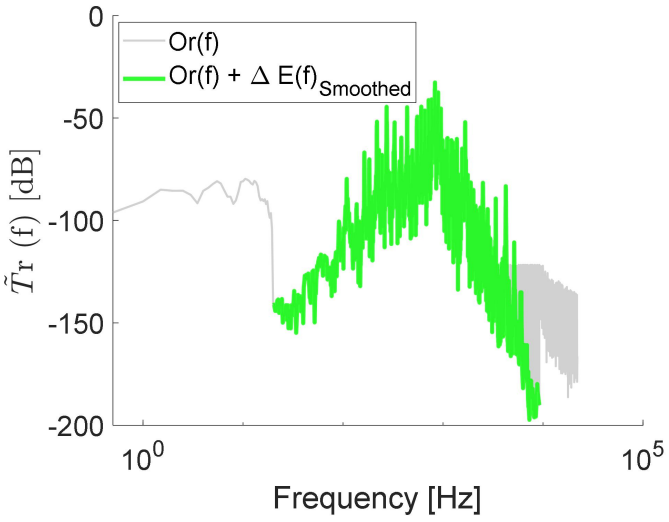


Figure 5 : The EQT spectrum which is the frequency-domain representaion of the final output .

Step 7 (Synthesis) :

An inverse Fourier transform is applied to the EQT spectrum to calculate the final EQT signal. The phase response of the input signal is used here.

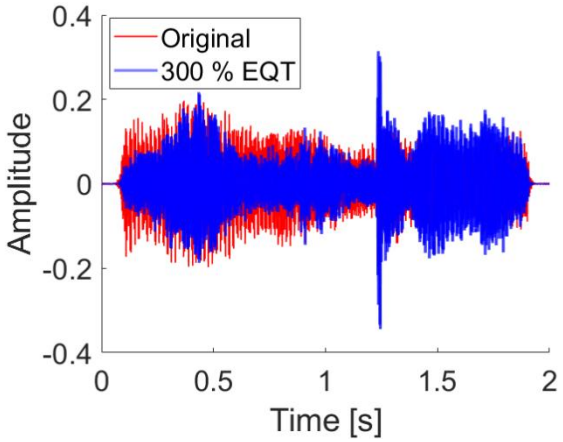


Figure 6 : The final EQT signal and the orignal signal prior to the transform. The EQT signal is energy normalized to the orignal in Step 8.

Step 8 (Normalization) :

Here, the EQT signal from step 7 is energy normalized to the input signal in the time-domain using a linear weight.

The EQT has a significant impact on the frequency domain sparsity of the signal as higher % EQT give rise to sparser representations in the frequency domain as shown in Figure 7.

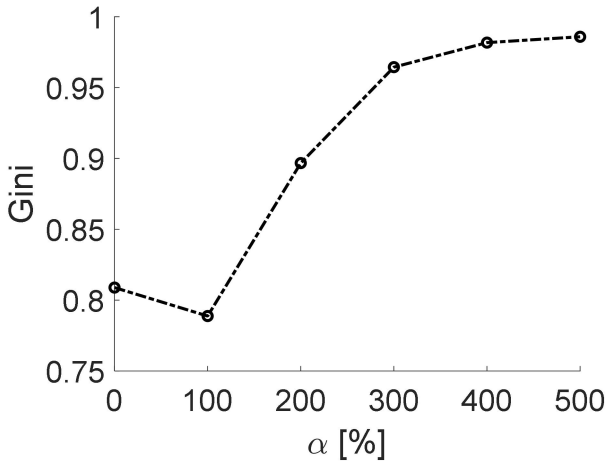


Figure 7 : The frequency domain sparsity of the piano track calculated using the Gini coefficient using a constant Q-transform based method at 3rd octave bands can be shown to significantly increase with higher % transforms.

To apply the EQT for a mix of coherent tracks, steps 1 - 8 are applied to each track independantly at the same % **α** . To derive the final EQT mix, the EQT tracks are added as shown in Figure 8.

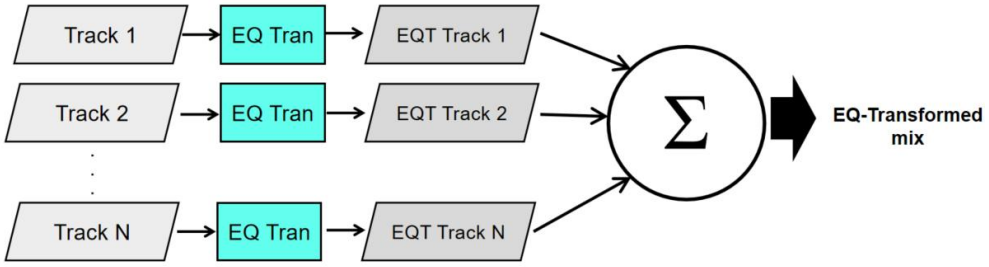


Figure 8 : Each track in a mix is subjected to the EQ Transform (EQ Tran) through steps 1-8 with the same % **α** independantly. The final EQT mix is derived by summing the individual EQT tracks as shown.

Appendix C

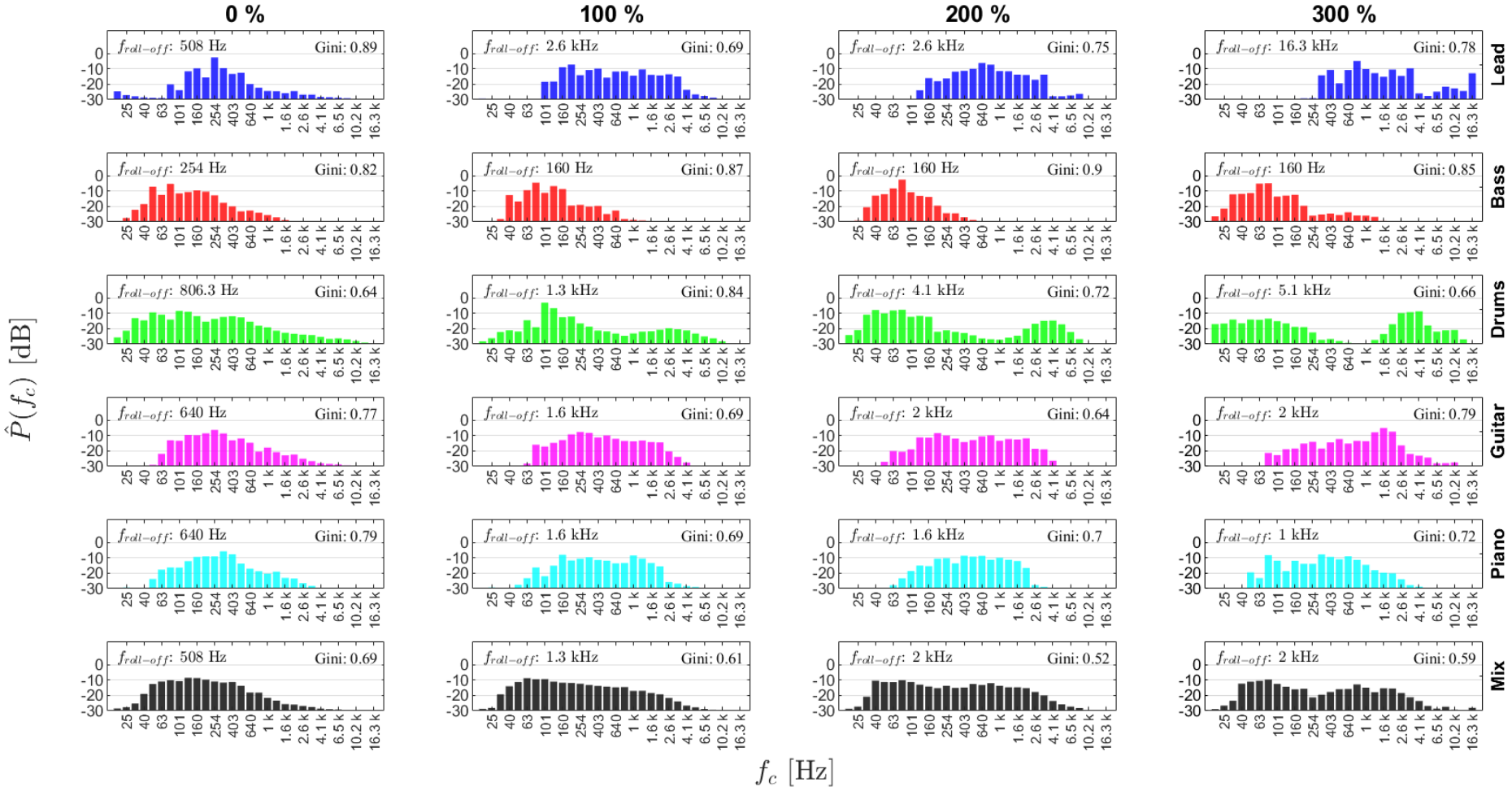


Figure 1: The normalized ensemble average power spectra at 3rd octave band center frequencies for the stimuli presented in Study 2. Corresponding sparsity (Gini) and 95% roll-off points provided are for the ensemble average spectra illustrated.

Supplementary Analysis

C1 EQ-transform and objective changes in spectral shape: Influence on audio quality ratings (Study 3)

In order to measure objective changes in spectral shape, the Log-Spectral Distance (LSD) between the Constant-Q transform (CQT) power spectra of the transformed mix $\widehat{\Phi}$ and the original mix (100 % EQT or factory settings) Φ_{orig} were evaluated in this analysis. We used the euclidean norm-based measure shown in equation 1 described by Batri (1998) to quantify distortions brought on by vector quantization and linear predictive coding of speech. In the below equation, the distance is measured over N frequency bins of the CQT spectra at third-octave spacing.

$$LSD = \sqrt{\frac{1}{N}\sum_{k=0}^{N-1}\left[10\,log_{10}\left(\frac{\Phi_{orig}(k)}{\widehat{\Phi}(k)}\right)\right]^2} \tag{1}$$

Figure 2(A) shows the LSD values calculated for the EQ-transformed mixes. The changes in LSD illustrated are statically robust in that the EQ-transform introduces significant changes to spectral shape. Specifically, higher % EQ-transform compared to factory settings introduced larger deviations in shape ($p < .0001$, $d > 2.8$, huge effect). Figure 2(B), shows the correlation of quality ratings and LSD for the NH,

mild, and moderate-severe HI listeners. For all of the participant groups, larger LSD values were associated with lower quality ratings. This indicates a consistent perceptual degradation with increasing spectral distortions. The Pearson's correlation between mean quality ratings and LSD indicated a strong linear association for all three participant groups: NH ($r(103)$ = -0.8, p <.0001, d = 2.8, huge effect), mild HI ($r(103)$ = -0.68, p <.0001, d = 1.8, very large effect), and moderate-severe HI ($r(103)$ = -0.64, p <.0001, d = 1.7, very large effect).

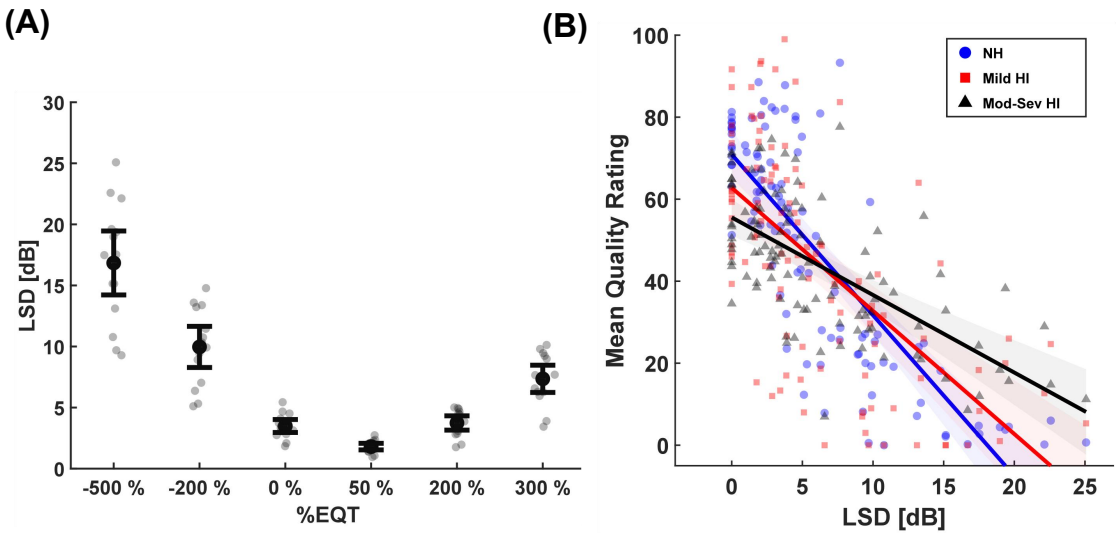


Figure 2: (A) The log-spectral distances (LSD) from original (100 % EQT) for the EQ-Transformed mixes. The LSD values were calculated from CQT power spectra with a resolution of 3 bins per octave. (B) Scatter plot of mean quality ratings and LSD taken for normal-hearing (NH), mild, and moderate-severely (Mod-Sev) HI listeners. Trend lines indicate a significant linear correlation with shaded 95% CI regions.

A Fisher's r-to-z transformation (Diedenhofen and Musch, 2015) showed that the depreciation in quality ratings with increasing LSD did not differ significantly between mild and moderate-severe HI (p = .6). However, this negative association observed for NH listeners was significantly stronger than that for both mild (z = -2.2, p = .02 < .05) and moderate-severe HI (z = -2.7, p = .007 < .01). These results suggest that while HI listeners can perceive spectral alterations in music mixes, they appear less affected by such distortions in their quality judgments compared to NH

listeners.

To evaluate the sensitivity to spectral changes grounded in the quality ratings as a function of hearing thresholds, a linear mixed-effects model was fitted to the data. The model included fixed effects for LSD, mean hearing threshold measured at the better ear (BEMHT), and their interaction to assess the influence of spectral shape changes on quality ratings for the different thresholds of hearing. The model also included random intercepts for the participants and mix, as well as a random slope for LSD for each participant. This model formulated in such a manner captures both population-level effects of hearing loss and spectral shape changes, along with individual differences in perceptual sensitivity. Furthermore, the inclusion of the random slopes assumes that the impact of spectral shape changes on perceived quality across the participants may not necessarily be uniform. The model estimates were such that all independent and interaction effects were statistically significant ($p < .0001$).

With the aid of the model estimates, we aim to predict the smallest change in LSD for which a significant drop in quality score is observed, as a function of BEMHT. Based on the data, in NH listeners, a minimum change of 3.5 dB in LSD brought about a significant 10-point drop in mean quality ratings from 73 to 63, $t(880) = 5.5$, $p < .0001$, $d = 0.4$ (Small effect). HI listeners showed a similarly significant drop of 11 points from 57 to 46, for a minimum change of 3.7 dB in LSD, $t(718) = 4.6$, $p < .0001$, $d = 0.3$ (Small effect). This approximate 10-point change observed for NH and HI listeners may bear perceptual relevance according to empirical evidence. Specifically, Mendonça and Delikaris-Manias (2018) recorded quality ratings using the 100-point MUSHRA scale from NH audio professionals for speech and music signals that were convolved with measured (reference) and synthesized room impulse responses. They showed that the smallest drop in the ratings from those elicited for reference signals was approximately 10-15 points. Similarly, Yao et al. (2024) illustrated that NH listeners showed a minimum deviation of approxi-

mately 10 points in their quality ratings for a reference when evaluating noise bursts colored with spectrally distorted head-related transfer functions (HRTFs). These finding underscore the perceptual relevance of a 10-point shift on the MUSHRA scale, albeit in listeners without a hearing impairment. Therefore, a 10-point drop in quality was adopted in our analysis as a perceptual threshold to estimate the smallest detectable change in LSD as a function of hearing loss, using the fixed effects coefficients from the model. As illustrated in Figure 3, the model predicts that as BEMHT increases, the minimum LSD required to produce a 10-point decline in quality ratings rises exponentially. The predictions for NH listeners suggest that they may begin to perceive a noticeable drop in quality only when the LSD is larger than 2 dB. This algins with the findings from Yao et al. (2024) who showed that the spectrally distorted HRTFs did not significantly affect quality ratings while resulting LSDs were 2 dB or less.

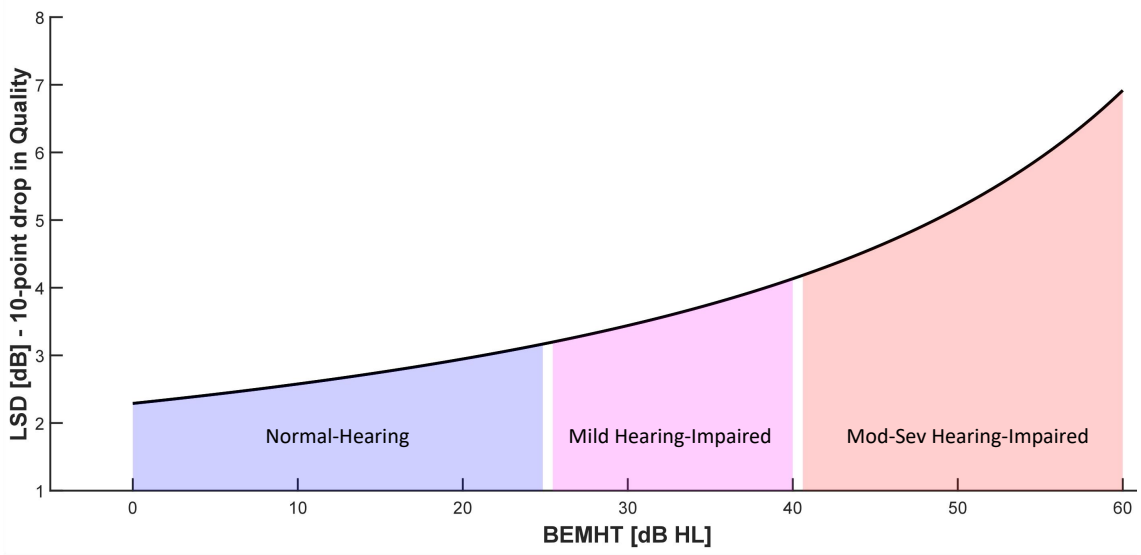


Figure 3: Model predictions of log-spectral distance from the original mix (LSD) that brings about a 10-point drop in quality ratings with increasing mean hearing thresholds measured at the better ear (BEMHT).

C2 Association between MSA and quality ratings in moderate-severe HI (Study 3)

The moderately strong association between MSA and quality ratings observed in Study 3 ($r = 0.54$; (Akoglu, 2018)) underscored the heightened inter-dependency of the music perception metrics among HI individuals, especially in the moderate-severe group ($r = .62$, $p = .01 < .05$, $d = 1.6$; very large effect), while the relationship appeared weak for listeners with mild HI ($p = .8$) as shown in Figure 4. Although the variability quality ratings did not differ between listener groups ($p = .3$), significantly higher variability in the mean MSA performance was observed in moderate-severe HI compared to NH ($F(29,15) = 0.2$, $p < .0001$). This is illustrated Figure 4 where relatively larger spread in mean MSA performance is observed among the moderate-severe HI.

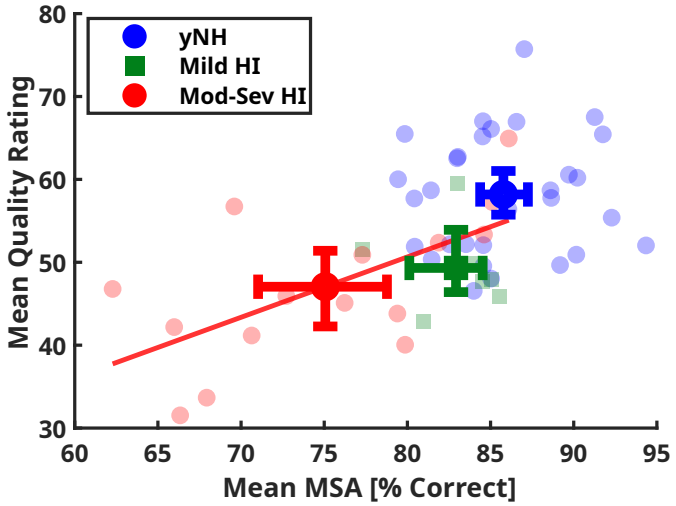


Figure 4: Scatter plot of mean MSA and quality ratings. Participant groups in Study 3 are separated into yNH, Mild, and Moderate-Severe (Mod-Sev) HI. Error bars indicate 95 % bootstrap confidence intervals for mean MSA and quality scores. The trend line shows a significant linear relationship for only the Mod-Sev group.

C3 Hearing loss and dispersion in MSA performance (Study 2)

To examine the observed dispersion in mean MSA performances associated with hearing loss in observed in Figure 4 in section C2, an Ordinary Least Squares (OLS) regression model was used as part of a Breusch–Pagan-style heteroskedasticity test. This approach was chosen because the Breusch–Pagan test is specifically defined for the residuals of OLS regression models (Breusch and Pagan, 1979).

Consistent with Study 2, the model showed that BEMHT had a significant negative effect on mean MSA performance $F(1,206) = 65.3$, $p < .0001$, $\eta^2{}_p = 0.24$, indicating that hearing thresholds at the better ear explained approximately 24% of the variance in mean MSA performance. On the other hand, musical training and % EQ-transform had no significant effect ($p > .1$), accounting for merely 1.3% and 1.8% of the variance, respectively. In order to assess the influence of the predictors on the residual variance of mean MSA performance, the heteroskedasticity test was performed by regressing the squared residuals in an auxiliary OLS regression. Based on the parameter estimates of the auxiliary model, every additional decibel increase in BEMHT was associated with an approximately 1.8-unit increase in the estimated residual variance of mean MSA performance. Furthermore, the test indicated that the assumption of homoskedasticity is violated, $LM(5) = 20.1$, $p < .01$. Neither % EQ-transform nor musical training significantly affected the residual variance, and no significant interaction effects were found. Thus, BEMHT was independently associated with the observed heteroskedasticity. This observation is further supported by the significant positive correlation between the squared residuals of mean MSA performance and BEMHT, as illustrated in Figure 5(A).

C4 Hearing loss and dispersion in mixing preferences (Study 1)

A similar heteroskedasticity analysis as that performed in the earlier section C3 was extended to mixing preferences observed in Study 1. Preferences from unaided HI listeners from both experiments revealed that the residual variance in SPBal and EQ-transform preferences increased as hearing thresholds worsened ($p < .02$). Based on the auxiliary regression, mean hearing thresholds taken over both ears (MHL) explained approximately 21 % ($R^2 = .207$) of the systemic variation in the residual variance in SPBal preferences, while explaining a substantial 41 % ($R^2 = .406$) of the variation in the residual variance of EQ-transform preferences in unaided listeners. However, when considering aided listeners from both experiments, the Breusch –Pagan test did not reject the assumption homoskedasticity for SPBal preferences ($LM(1) = 2.01,\ p = 0.2$). Nevertheless, EQ-transform preferences remained significantly varied as hearing thresholds worsened, despite MHL explaining a smaller portion (17.2 %). On the other hand, mean LAR preferences failed to reject the assumption of homoskedasticity in both aided ($LM(1) = 1.8$, $R^2 = .06$, $p = 0.2$) and unaided listening conditions ($LM(1) = 2.3,\ R^2 = .08,\ p = 0.13$). Figures 5(B-D), provides an illustration of the squared residuals of average mixing preferences and MHL as a scatter plot. The linear correlation shown between the two is consistent with the heteroskedasticity assessment.

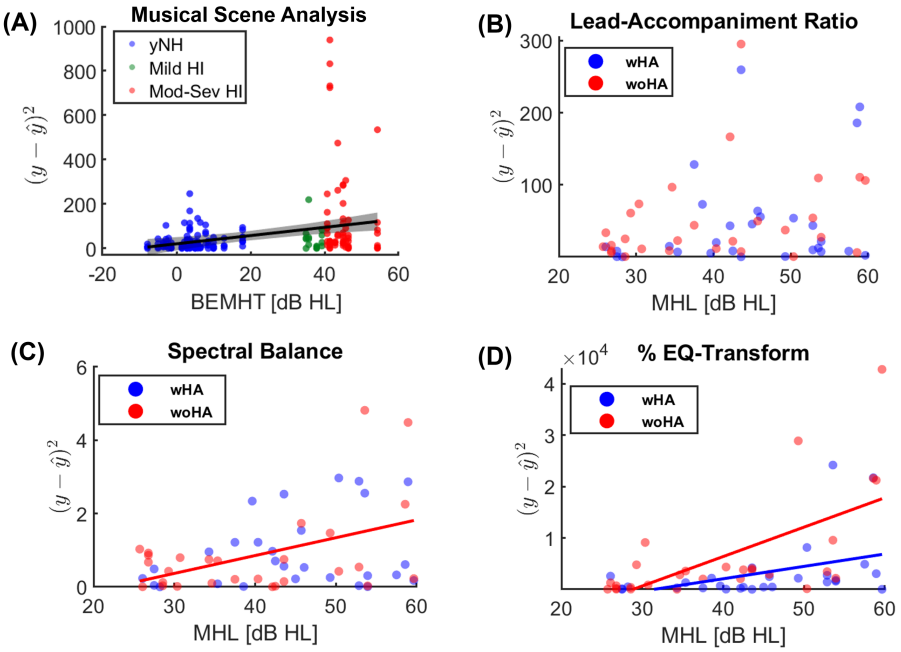


Figure 5: Scatter plot of hearing thresholds and squared residuals, used as a surrogate measure of residual variance of mean MSA (A) and mixing preferences (B-D). For the latter, aided (wHA) and unaided (woHA) conditions taken from both experiments in Study 1 are shown. Trend lines indicate a significant linear rise in the residual variance with hearing thresholds. Shaded region in (A) highlights the 95% CI. Significant linear correlations observed for: (A) MSA (r = .29, p < .0001), (C) Spectral balance for woHA (r = .45, p < .05), and (D) % EQ-transform for wHA (r = .41, p < .05) and woHA (r = .63, p < .001).

C5 EQ-transform effects on mix sparsity preferences (Study 1)

As performed on individual tracks in Study 1, Gini indices from CQT spectra (3 bins per octave) of 60 music mixes were computed in this analysis for different % EQ-transform settings. By doing so, we could observe a significant monotonic trend of increasing sparsity at higher % EQ-transforms, indicating greater spectral contrast, $F(5,354) = 61.7$, $p < .0001$, $\eta_p{}^2 = 0.47$ (Very large effect). This trend was particularly pronounced for over-mixing (i.e, > 100% EQ-transform; Figure 6A). These observations indicate that the EQ-transform modifies not only the frequency-domain sparsity of individual tracks but also that of the overall mix.

Interestingly, when we express % EQ-transform preferences from both experiments in Study 1 in terms of Gini indices, a significant positive correlation could be shown between MHL and sparsity preferences across NH and unaided HI, $r(51) = .686$, $p < .0001$, $d = 1.89$ (very large effect), and within unaided HI alone, $r(26) = .692$, $p < .0001$, $d = 1.92$ (very large effect), as illustrated in Figure 6B. For aided HI listeners on the other hand, this association was non-significant ($r = .26$, $p = .2$) and significantly weaker than that observed within unaided HI ($z = 2.06$, $p = .02 < .05$). Furthermore, hearing-aid use ($M = 0.76$, $SD = 0.03$) did not significantly alter mix sparsity preferences compared to non-use ($M = 0.78$, $SD = 0.04$), $p = .12$, based solely on the data from Experiment 2 (Figure 6C).

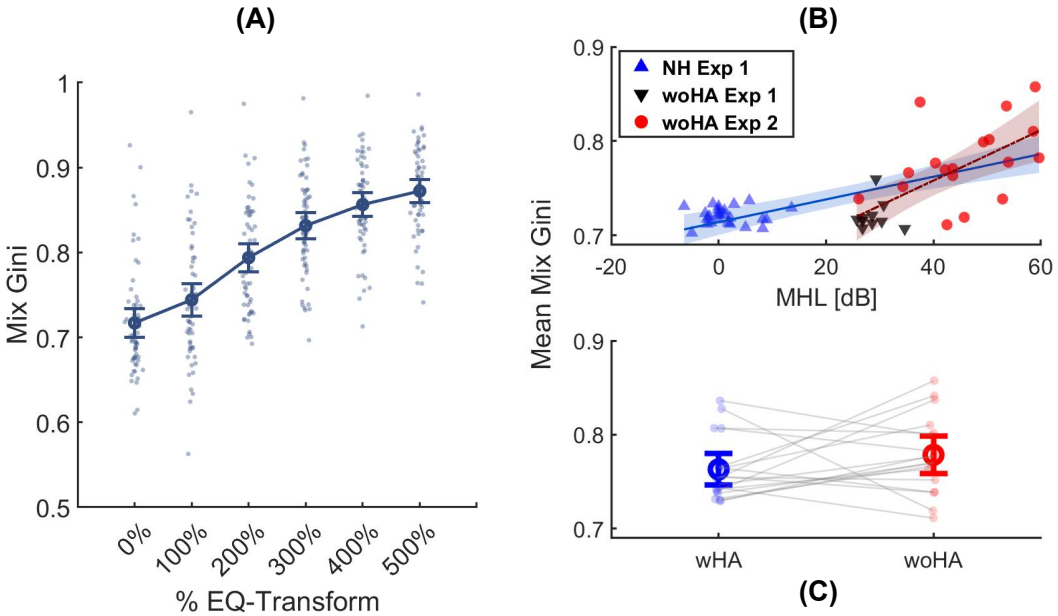


Figure 6: (A) Mean Gini coefficients and 95% CI over 60 music mixes for different % EQ-transforms. Individual points correspond to the Gini coefficient of each mix. (B) Scatter plot of mean Gini coefficients derived from % EQ-transform preferences of NH and unaided HI listeners in Study 1, Experiments 1 and 2 with trend lines indicating significant linear correlation (solid: all participants; dashed: HI listeners) and shaded 95% CI regions. (C) Mean Gini coefficients and 95% CI from % EQ-transform preferences for wHA and woHA conditions in Study 1, Experiment 2.

Supplementary References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3):91–93.

Batri, N. (1998). Robust spectral parameter coding in speech processing.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294.

Diedenhofen, B. and Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PloS one*, 10(4):e0121945.

Mendonça, C. and Delikaris-Manias, S. (2018). Statistical tests with mushra data. In *Audio Engineering Society Convention 144*. Audio Engineering Society.

Yao, D., Zhao, J., Liang, Y., Wang, Y., Gu, J., Jia, M., Lee, H., and Li, J. (2024). Perceptually enhanced spectral distance metric for head-related transfer function quality prediction. *The Journal of the Acoustical Society of America*, 156(6):4133–4152.

List of publications by author

Aravindan Joseph Benjamin, Kai Siedenburg; Exploring level- and spectrum-based music mixing transforms for hearing-impaired listeners. *J. Acoust. Soc. Am.* 1 August 2023; 154 (2): 1048–1061. *https://doi.org/10.1121/10.0020269*

Benjamin AJ, 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

Benjamin AJ, Siedenburg K. Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music. JASA Express Lett. 2024 Nov 1;4(11):113202. *https://doi.org/10.1121/10.0032474*

Declaration of own contribution

I hereby confirm that Aravindan Joseph Benjamin contributed to the aforementioned studies as stated below:

Article: Aravindan Joseph Benjamin, Kai Siedenburg; Exploring level- and spectrum-based music mixing transforms for hearing-impaired listeners. *J. Acoust. Soc. Am.* 1 August 2023; 154 (2): 1048–1061.
https://doi.org/10.1121/10.0020269

Author contribution: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the necessary experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

Article: Benjamin AJ, 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

Author contribution: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the necessary experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

Article: Benjamin AJ, Siedenburg K. Evaluating audio quality ratings and scene analysis performance of hearing-impaired listeners for multi-track music. JASA Express Lett. 2024 Nov 1;4(11):113202. *https://doi.org/10.1121/10.0032474*

Author contribution: Aravindan Joseph Benjamin formulated the research question, was involved in the design of the study, conducted the necessary experiments, performed the analysis on the data and drafted the final paper. Kai Siedenburg formulated the research question, guided the design of the study and the data analysis, and performed revisions to the manuscript.

_____ _____

(name) Date

Supervisor

Declaration of adherence to good scientific practice

I hereby declare that this dissertation is my own independent work, prepared using only the resources indicated, with all sources duly acknowledged through references. It has not been published or submitted, in whole or in part, for assessment in any other doctoral procedure at another university. I confirm that I have complied with the guidelines for good scientific practice of the Carl von Ossietzky University of Oldenburg and have not used any commercial placement or consulting services in connection with this work.

_____ _____

(name) Date