Cognition 244 (2024) 105696

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Generalization of auditory expertise in audio engineers and instrumental musicians

Francesco Caprini [a,*], Sijia Zhao [b], Maria Chait [c], Trevor Agus [d], Ulrich Pomper [e], Adam Tierney [a,1], Fred Dick [f,1]

[a] Department of Psychological Sciences, Birkbeck, University of London, UK
[b] Department of Experimental Psychology, University of Oxford, UK
[c] University College London (UCL) Ear Institute, UK
[d] School of Arts, English and Languages, Queen's University Belfast, UK
[e] Department of Cognition, Emotion, and Methods in Psychology, Universität Wien, Austria
[f] Department of Experimental Psychology, University College London (UCL), UK

ARTICLE INFO

Keywords:
Expertise
Learning
Attention
Auditory scene analysis
Speech in noise
Musicians

ABSTRACT

From auditory perception to general cognition, the ability to play a musical instrument has been associated with skills both related and unrelated to music. However, it is unclear if these effects are bound to the specific characteristics of musical instrument training, as little attention has been paid to other populations such as audio engineers and designers whose auditory expertise may match or surpass that of musicians in specific auditory tasks or more naturalistic acoustic scenarios. We explored this possibility by comparing students of audio engineering ($n = 20$) to matched conservatory-trained instrumentalists ($n = 24$) and to naive controls ($n = 20$) on measures of auditory discrimination, auditory scene analysis, and speech in noise perception. We found that audio engineers and performing musicians had generally lower psychophysical thresholds than controls, with pitch perception showing the largest effect size. Compared to controls, audio engineers could better memorise and recall auditory scenes composed of non-musical sounds, whereas instrumental musicians performed best in a sustained selective attention task with two competing streams of tones. Finally, in a diotic speech-in-babble task, musicians showed lower signal-to-noise-ratio thresholds than both controls and engineers; however, a follow-up online study did not replicate this musician advantage. We also observed differences in personality that might account for group-based self-selection biases. Overall, we showed that investigating a wider range of forms of auditory expertise can help us corroborate (or challenge) the specificity of the advantages previously associated with musical instrument training.

1. Introduction

1.1. Musical expertise

1.1.1. Current literature

The training of a professional musician normally begins very early in life and is estimated to entail over 10,000 h of training by early adulthood (Ericsson & Charness, 1994; Macnamara & Maitra, 2019). Playing a musical instrument involves a multifaceted ensemble of skills, including acoustic processing (e.g., pitch, duration, timbre), cognitive processing of melodic and harmonic content (e.g., selective attention, grouping, auditory scene analysis), fine motor control, sensory-motor coordination, sequence memorisation, musical interpretation and expression, and development of learning strategies (Hallam, 2001, 2010; Norton et al., 2005). Over the past few decades, the perceptual and cognitive advantages associated with musical training have been studied extensively. Reported examples of such advantages include pitch perception (Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Spiegel & Watson, 1984), temporal information processing (Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012; Güçlü, Sevinc, & Canbeyli, 2011; Rammsayer & Altenmüller, 2006), phonological processing (Chobert, François, Velay, & Besson, 2014; Tierney, Krizman, Kraus, & Tallal,

---

* Corresponding author.
E-mail addresses: fcapri01@mail.bbk.ac.uk, f.dick@ucl.ac.uk (F. Caprini).

[1] Shared senior authorship.

https://doi.org/10.1016/j.cognition.2023.105696
Received 2 July 2021; Received in revised form 4 December 2023; Accepted 13 December 2023
Available online 30 December 2023
0010-0277/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2015; Wong, Skoe, Russo, Dees, & Kraus, 2007), attention (Kaganovich et al., 2013; Román-Caballero, Martín-Arévalo, & Lupiáñez, 2020; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Strait, Slater, O'Connell, & Kraus, 2015; Zendel & Alain, 2009), speech in noise perception (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Slater & Kraus, 2016; Tierney, Rosen, & Dick, 2020; Yoo & Bidelman, 2019; for a review, see e.g. Coffey, Mogilever, & Zatorre, 2017), statistical learning (Mandikal Vasuki, Sharma, Demuth, & Arciuli, 2016; Schön & François, 2011), working memory (Bugos, Perlstein, McCrae, Brophy, & Bedenbaugh, 2007; Talamini, Altoè, Carretti, & Grassi, 2017, p. 201), auditory-motor synchronisation (Chen, Penhune, & Zatorre, 2008; Zatorre, Chen, & Penhune, 2007), visuospatial cognition (Douglas & Bilkey, 2007; Hassler, Birbaumer, & Feil, 1985; Lidji, Kolinsky, Lochy, & Morais, 2007), reading (Flaugnacco et al., 2015; Tierney & Kraus, 2013), and metacognition (Hallam, 2001). The educational and clinical implications of these findings (François, Grau-Sánchez, Duarte, & Rodriguez-Fornells, 2015) and their relevance in the study of brain plasticity and learning (Hyde et al., 2009; Zatorre, 2005) are some of the reasons that underlie the widespread adoption of musicianship[2] as a model of how expertise in one domain might develop and affect supposedly unrelated[3] and/or more general domains of perception and cognition. This phenomenon is known as transfer of expertise (either near or far based on the relatedness of the skills' contexts and cognitive demands; Barnett & Ceci, 2002; Mestre, 2005), and commonly serves as the epistemological construct underlying the literature on the effects of music training on both musical and nonmusical abilities.

1.1.2. Limitations

However, the conclusions that can be drawn from the current literature on the topic are somewhat limited by conflicting evidence and theoretical issues. An example is speech-in-noise perception. As noted above, a number of studies have reported musician advantages in perceiving speech in noisy or distracting environments, but equally, several studies have failed to detect an association with musical training across multiple experimental conditions (e.g. Boebinger et al., 2015; MacCutcheon et al., 2020; Madsen, Marschall, Dau, & Oxenham, 2019; Madsen, Whiteford, & Oxenham, 2017; Ruggles, Freyman, & Oxenham, 2014). It has been suggested that the advantage of musicians for speech-in-noise perception might depend on the relevance of pitch discrimination for the given task (Fuller, Galvin, Maat, Free, & Başkent, 2014), along with rhythmic skills (Slater et al., 2018) and the presence of spatial cues (Bidelman & Yoo, 2020; Clayton et al., 2016; but see Madsen et al., 2019), and may be partially negated by musicians' high levels of chronic noise exposure (Skoe, Camera, & Tufts, 2019). Importantly, the musician advantage for speech-in-noise perception could also be mediated by other and possibly preexisting cognitive abilities (e.g., working memory, attention) rather than being a direct effect of musical experience (Escobar, Mussoi, & Silberer, 2020; Schellenberg, 2015, 2019; Yoo & Bidelman, 2019). Thus, despite the interest in the topic and promising clinical applications (e.g. the rehabilitation of sensorineural and age-related hearing loss; Alain, Zendel, Hutka, & Bidelman, 2014; Lo, Looi, Thompson, & McMahon, 2020; Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011), current evidence does not unequivocally support the hypothesis that musical training enhances speech-in-noise perception.

Another example is the musicians' advantage for auditory sequence memorisation and reproduction (Krishnan, Carey, Dick, & Pearce, 2021; Tierney, Bergeson-Dana, & Pisoni, 2008), which Carey et al. (2015) did not replicate using the same general paradigm, despite testing a relatively large number of highly trained violinists and pianists. More generally, many studies have only observed expertise-related skill transfer to contexts closely related to the original training context (for a review, see Green & Bavelier, 2008), although a lack of granularity in the definition of population characteristics and behavioural measurements might make it difficult to reach conclusive and replicable results (Green, Strobach, & Schubert, 2014).

For example, simple comparisons of musically trained and untrained individuals cannot explain whether any of the observed advantages are specifically associated with unique features of musical training or could instead be observed (or even enhanced) with other types of training. Evidence from single-task randomised controlled training studies on non-musicians shows that several auditory perceptual thresholds (i.e., pitch, duration, intensity, interaural time and level difference) can indeed be individually improved with training (Wright & Fitzgerald, 2005; Wright & Sabin, 2007), subsequently matching those of musicians (Micheyl, Delhommeau, Perrot, & Oxenham, 2006).

Non-musical forms of auditory training have been investigated in the enhancement of speech intelligibility in adults with hearing loss (e.g. Fu, Nogaki, & Galvin, 2005; Henshaw & Ferguson, 2013; Whitton, Hancock, Shannon, & Polley, 2017; but see Stacey & Summerfield, 2007), language processing in children with learning difficulties (for a review, see Loo, Bamiou, Campbell, & Luxon, 2010), as well as neurocognitive improvements of psychiatric patients (Adcock et al., 2009; Bettison, 1996; Fisher, Holland, Merzenich, & Vinogradov, 2009). Training profile variations within the musician population also appeared to be associated with specific perceptual advantages. A number of studies have reported perceptual differences between musicians who play different instruments and genres, such as lower (i.e. better) frequency discrimination thresholds for classical musicians compared to jazz musicians (Kishon-Rabin et al., 2001; cf. Tervaniemi, Janhunen, Kruck, Putkinen, & Huotilainen, 2016; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012), a frequency discrimination advantage for players of a variable pitch instrument (i.e. string and woodwind) compared to a fixed-pitch instrument (Micheyl et al., 2006) or percussion instruments (Zaltz, Globerson, & Amir, 2017), an instrument-specific preference for a musical temperament (i.e. tuning system; Carey et al., 2015), and better perception of speech harmonics for vocalists as opposed to speech timing for percussionists (Slater, Azem, Nicol, Swedenborg, & Kraus, 2017).

Additionally, other types of musical performers such as professional club disk jockeys have been shown to match trained percussionists in rhythmic ability (Butler & Trainor, 2015). Neuroplastic and behavioural correlates of other forms of auditory expertise unrelated to musical training have also been studied. For instance, 60 min of birdsong identification training was shown to lead to a decrease in early (200–300 ms) neural activity in left superior temporal gyrus and middle frontal gyrus in response to trained stimuli, but also a later (500–550 ms) increase in activity in the cingulate cortex bilaterally for untrained songbird stimuli (De Meo, Bourquin, Knebel, Murray, & Clarke, 2015). Additionally, scalp topography of P2 auditory-evoked potentials of songbird experts revealed a more frontal positivity than naive participants in response to not only birdsongs, but also voice and environmental stimuli, which might reflect a generalised difference in processing strategy (Chartrand, Filion-Bilodeau, & Belin, 2007). Another example is learning to decode Morse code, which has been associated with an increase in neural activity in the inferior and medial parietal cortex bilaterally and in grey matter density in the fusiform gyrus (Schmidt-Wilcke, Rosengarth, Luerding, Bogdahn, & Greenlee, 2010), while musicians have been shown to reproduce Morse code at variable speeds more accurately than non-musicians after training at a static speed (Slayton, Romero-Sosa, Shore, Buonomano, & Viskontas, 2020). Nonetheless, very little attention has been paid to other populations whose profession depends on high levels of auditory sophistication, such as audio engineers.

---

[2] At least in the compartmentalised or quasi-Platonic western notions of "music" and "being a musician." (Cross, 2012; Wiggins, Müllensiefen, & Pearce, 2010)

[3] These domains might in fact share perceptual and cognitive processing in the brain, despite appearing superficially unrelated (e.g. the OPERA hypothesis for music and speech processing; Patel, 2011; Patel, 2014).

1.2. Audio engineering

1.2.1. Population characteristics

Audio engineers attempt to create, capture, and modify sound in order to resolve technical issues and meet multiple artists' objectives (e.g., a musician, a producer, or their own), ultimately curating the listener's experience (Zwicker & Zwicker, 1991). This process can involve the discrimination and manipulation of psychoacoustic attributes such as pitch and timbre via equalisation and filtering, loudness and dynamic range via compression and expansion, but also synchronicity, phase, filtering, masking, and spatial features via custom configurations of hardware and software tools (Corey & Benson, 2016). Other than professional practice, this level of perceptual expertise is usually achieved via technical ear training, which involves exercises designed to improve the ability to focus on and identify discrete elements of auditory sensations, and associate them with objective acoustical measurements (Corey, 2013; Iwamiya, Nakajima, Ueda, Kawahara, & Takada, 2003; Letowski, 1985), although this practice is not yet fully standardised (Kaniwa et al., 2011; Kim, Kaniwa, Terasawa, Yamada, & Makino, 2013; Marui & Kamekawa, 2013, 2019; McKinnon-Bassett & Martens, 2013). Additionally, audio engineers must learn to deliberately direct their attention to individual elements of sounds or auditory scenes, and to maintain them in memory. For example, the practice of mixing in music production can involve listening to a complex auditory scene (e.g. an instrument group), scanning the scene to identify a source of potential acoustic issues in the global sound (e.g. phase interference, tonal imbalance, lack of definition or "muddy" sound, timing issues, etc.), applying a fix at the level of individual instruments or elements, reintegrating them into the scene, and reevaluating the updated auditory scene (for a detailed description of what mixing entails, see e.g. Case, 2012; Izhaki, 2008). This process intuitively should require considerable sustained selective attention (auditory scene segregation and integration) and auditory working memory (mental sound manipulation and pre-post comparison); the relevant tasks are supported by visual cues provided by screening devices like spectrum analysers.

1.2.2. A different model of auditory expertise

Musicians who play in ensembles must also be able to track the auditory scene and, in large ensembles, interpret the conductor's cues to synchronise with the group and adapt their sound to the collective performance. By comparison, audio engineers are responsible for several sound sources at the same time, have a much larger toolbox for acoustic manipulation that is not constrained by the physical construction of a musical instrument and can work either synchronously (e.g., live performance) or asynchronously (e.g., studio work). Furthermore, there can be multiple ways of achieving similar acoustic outcomes depending on the available gear, personal workflow, and creative process (De Man et al., 2015). For instance, the adjustment of a sound's intensity could correspond to the turn of a knob or a push of slider on a mixing board, the click of a mouse in a digital audio workstation, or the repositioning of a microphone. Moreover, these gestures can affect sound in real time or with any amount of delay. Conversely, the correspondence between an instrumentalist's gestures and acoustic outcomes is narrower in terms of range of motion and temporal co-occurrence of action and sound, which may promote auditory-motor coupling (Alluri et al., 2017; de Manzano, Kuckelkorn, Ström, & Ullén, 2020; Li et al., 2018; Palomar-García, Zatorre, Ventura-Campos, Bueichekú, & Ávila, 2017; Zatorre et al., 2007).

Audio engineers are also equipped with domain-specific knowledge such as signal processing, electronics, audio theory, and psychoacoustics (Howard & Angus, 2009), as well as technical language and professional jargon (Porcello, 2004), which can provide context and assist the interpretation of sensory perception. Taken together, the skills of these professionals correspond to a model of auditory expertise that is very different in nature from that of musical instrument training. In contrast to performing musicians, audio engineers do not need high proficiency in playing a musical instrument to excel in their profession. These unique characteristics of audio engineers can be exploited to test the specificity of some of the auditory advantages associated with musical training described in the literature, in particular fine auditory perception and auditory scene analysis.

1.3. Current study

The current study contrasts two different ecologically valid, auditory-based forms of expertise: audio engineering and playing a musical instrument. First, we tested the hypothesis that both audio engineers and musicians would show superior auditory skills compared to matched controls across a broad set of auditory-based measures that are both associated with musical training and essential for the practice of audio engineering. We included 6 psychophysical measurements (i.e., frequency, duration, intensity, sinusoidal amplitude modulation, interaural level difference, and interaural time difference) and 4 measurements of auditory scene analysis. The latter were: 1) a sustained auditory selective attention task (Laffere, Dick, & Tierney, 2020) where participants discriminate between two concurrent streams of tonal sequences; 2) a working memory and sound segregation task that involves the memorisation and matching of three concurrent sounds varying in frequency and amplitude modulation with a target sound; 3) a task that involves the detection of changes in the statistical properties of an auditory scene; and 4) a diotic speech-in-babble-noise task.

Second, we ran a set of exploratory analyses to identify and describe the unique attributes of our auditory expert cohorts. To complement the observational nature of this study and detect cohort qualities that may contribute to self-selection and performance, we also included self-report measures of personality and musical sophistication. The latter is particularly important as musicians and audio engineers can present partially overlapping forms of auditory expertise, thus posing a challenge to the interpretation of observational data. It is possible, for instance, for audio engineers to be excellent instrumentalists and, vice-versa, for musicians to be knowledgeable in the field of audio engineering, although we aimed to partially reduce the overlap between these two populations by explicitly recruiting musicians with no expertise in audio engineering, including recording, mixing, and mastering. We then evaluated the associations between different levels of audio engineering experience, musical experience, and auditory skill.

Third, we explored whether, and to what extent, low level perceptual ability, auditory scene analysis, and speech-in-babble perception correlate with each other and compared the manner in which these associations manifest between groups.

2. Experiment 1: Methods

2.1. Participants

Participants ($n = 64$) were undergraduate students of either audio engineering, a musical instrument degree, or any other non-musical degree. All participants were native English speakers between 19 and 26 years old and reported no history of hearing impairments. Audiometric thresholds were verified manually (see 2.2.1). Because audio engineers have not previously been studied as an expert auditory group, and therefore no well-motivated effect size could be estimated, N per group was determined by reviewing the literature reviewed above with musicians, which has shown quite consistent, musician-specific effects on aspects of auditory perception (Bidelman and Yoo (2020), Boebinger et al. (2015), Clayton et al. (2016), Escobar et al. (2020), Fuller et al. (2014), Kaganovich et al. (2013), Kishon-Rabin et al. (2001), MacCutcheon et al. (2020), Madsen et al. (2019), Ruggles et al. (2014), Slater et al. (2018)). The average group size (N(musicians) + N (non-musicians)/2) across all these studies was $N = 20$ (range 14–30); we recruited on this basis.

2.1.1. *Audio engineering students*

Students of audio engineering (n = 20, 17 M; age range = 19–26, mean (SD) = 21.3 (1.9)) were recruited first through email and flyer advertising. At the time of testing, they were enrolled full time (year 1, n = 2; year 2, n = 8; year 3, n = 10) in the Music Technology and Sonic Arts (BSc) program at Queen's University Belfast, where they were tested in a sound-insulated recording studio. They reported having on average 3.9 years of experience with audio recording, mixing or mastering (SD = 1.7, range = 1–7; see Table 1).

2.1.2. *Musical instrument students*

Musicians (n = 24, 16 M; age range = 20–26, mean (SD) = 23.9 (1.69)) were students of a musical instrument degree (see Table 1 for instruments) recruited in London through flyer advertising and UCL/Birkbeck SONA systems. Recruitment criteria included the practice of any musical instrument other than percussion for 4 or more years, with

Table 1
Demographics, years of formal training (i.e. musical instrument lessons), years of regular practice of a musical instrument, and years of audio engineering experience.

GroupID	Age	Sex	Course	Formal Training	Regular Practice	Audio Engineering
E1	20	Male	Music Technology	3–5	2	6
E2	20	Male	Music Technology	3–5	4–5	4
E3	19	Female	Music Technology	10+	10+	1
E4	21	Male	Music Technology	0	4–5	3
E5	26	Male	Music Technology	0	3	7
E6	21	Male	Music Technology	1	4–5	4
E7	22	Male	Music Technology	0	4–5	1
E8	21	Male	Music Technology	6–9	6–9	4
E9	19	Male	Music Technology	10+	10+	6
E10	22	Male	Music Technology	2	6–9	3
E11	21	Female	Music Technology	10+	10+	3
E12	22	Male	Music Technology	0	6–9	5
E13	20	Male	Music Technology	6–9	10+	3
E14	20	Male	Music Technology	0	4–5	3
E15	21	Male	Music Technology	6–9	6–9	6
E16	20	Male	Music Technology	0	6–9	3
E17	23	Male	Music Technology	1	4–5	6
E18	21	Female	Music Technology	10+	10+	3
E19	21	Male	Music Technology	0	1	3
E20	26	Male	Music Technology	3–5	6–9	3
C1	20	Female	Psychology	0	0	0
C2	24	Male	Anthropology	0	0	0
C3	23	Female	Pharmacy	0	1	0
C4	22	Female	Medicine	0.5	0	0
C5	20	Male	History	0	0	0
C6	22	Male	Mathematics	0	0	0
C7	25	Male	Management	0	0	0
C8	21	Male	Computing	0.5	0	0
C9	24	Male	Medicine	1	0	0
C10	22	Male	Social Policy	0.5	0	0
C11	23	Male	Medicine	0	0	0
C12	21	Male	Computer Science	0	0	0
C13	24	Male	Banking and Finance	0	0	0
C14	20	Male	Mathematics	1	1	0
C15	19	Male	Economics	0	0	0
C16	20	Male	Engineering	0	0	0
C17	23	Male	Jewellery Design	1	0	0
C18	21	Male	Natural Sciences	0	0	0
C19	19	Male	Neuroscience	0	0	0
C20	19	Male	History and Politics	2	1	0
M1	25	Female	Piano	6–9	10+	0
M2	25	Female	Piano	6–9	10+	0
M3	22	Male	Trumpet	6–9	6–9	0
M4	26	Female	Voice	6–9	10+	0
M5	25	Male	Piano	10+	10+	0
M6	24	Male	Oboe	6–9	6–9	0
M7	24	Male	Guitar	3–5	4–5	0
M8	22	Male	Piano	3–5	4–5	0
M9	26	Male	Guitar	6–9	10	0
M10	24	Female	Piano	6–9	6–9	0
M11	26	Female	Piano	6–9	6–9	0
M12	22	Male	Violin	3–5	6–9	0
M13	23	Female	Violin	3–5	6–9	0
M14	24	Male	Oboe	3–5	6–9	0
M15	26	Male	Guitar	6–9	10+	0
M16	25	Male	Piano	3–5	4–5	0
M17	26	Male	Violin	3–5	4–5	0
M18	23	Male	Violin	3–5	6–9	0
M19	25	Male	Piano	3–5	6–9	0
M20	23	Male	Guitar	6–9	10+	0
M21	20	Male	Piano	6–9	10+	0
M22	21	Female	Piano	6–9	10+	0
M23	24	Male	Guitar	3–5	4–5	0
M24	23	Female	Oboe	6–9	6–9	0

an average daily practice of at least 2 h a day and no experience with audio engineering, mixing, mastering, or recording. Despite efforts being made to match all cohorts' demographics, participants in the musician group included 5 more female participants and were on average 2.5 years older than engineers and controls. The effects of these potential confounds on the auditory measurements were evaluated post-hoc via nonparametric univariate testing (see 2.4.2) for gender and Spearman correlations for age. No association was found for either demographic variable.

2.1.3. Control group

Controls ($n = 20$, 17 M; age range = 19–25, mean (SD) = 21.6 (1.9)) were also recruited in London through the UCL and Birkbeck SONA systems. They were undergraduate students of non-musical degrees (i.e., psychology, anthropology, pharmacy, history, management, mathematics, social sciences, finance, jewellery design, computer science, medicine), with no formal training or history of regular practice playing a musical instrument or audio engineering, mixing, mastering, or recording. Both music instrumentalists and controls were tested in a quiet testing booth at Birkbeck, University of London.

2.2. Procedure

The test battery was composed of one audiometric screening, 10 behavioural tasks, and 2 questionnaires. Each testing session lasted up to 2 h, with average duration being about 1 h and 45 min. To minimise differences across individuals due to task order, tasks and questionnaires were run in the same order for all participants, which is the order in which they are presented below. All auditory tasks were piloted by three expert raters who determined the ideal headphone volume in terms of task difficulty, clarity (i.e., task not loud enough), and comfort (i.e., task too loud). Once set, volume was kept constant for all tasks and participants. The study was approved by the Birkbeck Department of Psychological Sciences ethics committee (approval number 111228) and all participants gave their informed consent before the start of the experiment.

2.2.1. Audiology

Two different tools were used to measure audiometric thresholds. Students of audio engineering were tested with a Kamplex KC35 Audiometer, while musicians and non-musicians were tested with an Otopod system paired with Symphony software on a Windows XP laptop. In both cases, a 10 dB-down, 5 dB-up adaptive staircase procedure (British Society of Audiology, 2018) was used, and thresholds were measured using pure tones from a range of frequencies presented in this order: 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, 8 kHz, 125 Hz, 250 Hz, 500 Hz, 750 Hz. After manually checking that they could hear a sample sound from both ears, participants were asked to listen carefully and to press the provided response button whenever they could hear a tone, starting at 10 dB HL. All frequencies were presented monaurally starting with the left ear. For each frequency, a threshold was determined when the participant performed 2 reversals at the same intensity.

2.2.2. Speech in babble noise (SIN)

Participants were instructed to listen carefully to a target sentence in the presence of four-talker babble and repeat that sentence out loud to the experimenter. Target sentences, spoken by a British male, were sampled from the Bamford-Kowal-Bench Speech-in-Noise (BKB-SIN) sentences (Bench, Kowal, & Bamford, 1979; Etymotic Research, 2005) and included 3 keywords. All stimuli were presented diotically. Participants were encouraged to repeat any word they heard, regardless of whether that was a single word or an entire sentence. The experimenter marked the number of correct words that were repeated. Unlike the original BKB-SIN test, we estimated speech/babble SNR thresholds using an in-house adaptive staircase procedure implemented in MATLAB (2013b; The MathWorks Inc, 2013). The initial SNR value was set to +10 dB and changed adaptively up or down according to participants' response (1-up 1-down). A response was considered correct if at least 2 keywords were identified. After recording the participant's response, a new sentence was presented. The first step size was set to 8 dB and reduced to 6 dB, 4 dB, and 2 dB after each reversal. SNR changes were obtained by increasing or decreasing the amplitude of the target sentence, while the amplitude of the babble mask was kept constant. The experiment terminated after 6 2-dB-step reversals or when the limit of 20 sentences was reached. A final SNR threshold was calculated as the average SNR ratio of the stimuli presented after the first 3 reversals (i.e., the final set of stimuli presented with a 2 dB step size).

2.2.3. Sustained auditory selective attention (SASA)

This task was designed to quantify participants' sustained selective attention (Dick et al., 2017; Holt, Tierney, Guerra, Laffere, & Dick, 2018; Laffere et al., 2020). Each block consisted of a stream of 30 short sequences, each made of six 125 ms cosine-ramped sine tones sampled with replacement from two frequency bands in an alternating pattern (Fig. 1). Each band was composed of three tones set two semitones apart: 185, 207.7, and 233.1 Hz (F#3, G#3, and A#3) for the lower band and 370, 415.3, and 466.2 Hz (F#4, G#4, and A#4) for the higher band (i.e., one octave above). Tones were presented at regular intervals at a rate of 8 Hz followed by a 250 ms pause and the first tone was always sampled from the lower band. As higher-frequency stimuli tend to be perceived as louder, a difference of 8 dB was set between the amplitudes of the tones in the high and low bands. A total of 30 sequences was presented in each block. For the first 10 blocks, participants were asked to respond by pressing the space bar when they heard 2 consecutive identical sequences in the high band. After a short break, participants completed another 10 trials, this time detecting repetitions in the low band while ignoring tones in the high band. Each trial included between 3 and 6 repetitions. The experiment was preceded by 4 training blocks for each condition, during which the amplitude of the confounding stream was initially set to zero and linearly increased until it reached its final amplitude. Answers were evaluated within a 1 s window starting at the onset of the third tone of a sequence (i.e., between 0.5 s and 1.5 s after a sequence's onset). Participants received feedback on screen immediately after responding. Sensitivity to repetitions in the attended band was calculated as d'.

2.2.4. Goldsmiths musical sophistication index (gold-MSI)

A digital version of the Goldsmiths Musical Sophistication Index (Gold-MSI) (Müllensiefen, Gingras, Musil, & Stewart, 2014) was administered. This extensively normed questionnaire quantifies individual differences in musical sophistication according to five dimensions, Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, Emotions, and one common factor, General Sophistication. Participants rated on a 7-point Likert scale how much they agreed with a statement that described their experience with music. Scores for each dimension were calculated as the sum of the ratings given to each item belonging to that dimension after inverting negative-score items.

2.2.5. Ten item personality inventory (TIPI)

A computerised version of the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) was administered. In this brief questionnaire, each of the Big Five personality dimensions (i.e., extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience) is represented by two pairs of adjectives, one positive (e.g., "sympathetic, warm" for agreeableness) and one negative (e.g., "reserved, quiet" for extraversion). Participants were asked to indicate how much they identified with each pair of adjectives on a scale from 1 (Strongly disagree) to 7 (Strongly agree). The final scores were calculated by taking the average of the 2 items representing each dimension after inverting the ratings of the negative items.

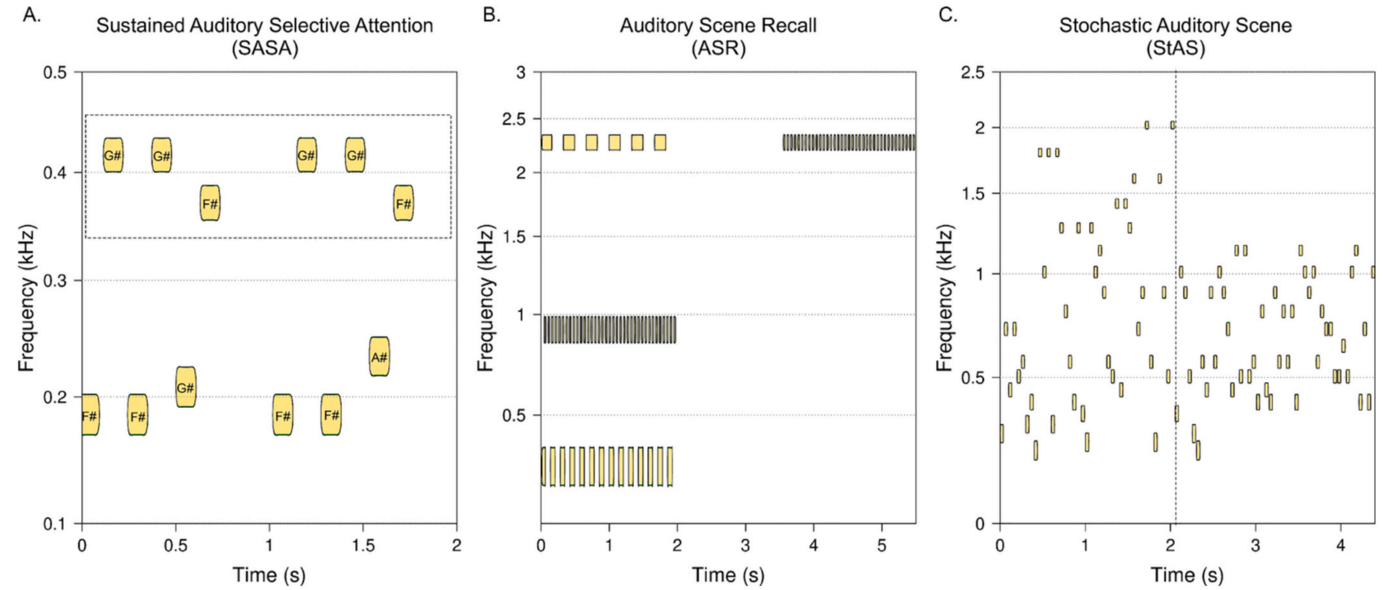


Fig. 1. Schematic representation of auditory scene analysis stimuli. **A**. Sustained auditory selective attention. Three-tone repetition in the high band marked by black rectangle. **B**. Auditory scene recall. Three tones with different frequencies and square-wave amplitude modulation rates followed by a target tone with a new combination of frequency and modulation rate. **C**. Stochastic auditory scene. Example of a “full to middle” (F-M) transition. Vertical dotted line represents the change in frequency sampling pool for the random tones.

2.2.6. *Auditory scene recall (ASR)*

This task was designed to measure participants’ ability to segregate different sounds in an auditory scene analysis and maintain them in memory for a short period of time (Pomper, Curetti, & Chait, 2023, in press). Each trial was made of three phases. During the first phase (“encoding phase”), participants listened to an auditory scene (2 s) composed of three amplitude-modulated pure tones (“streams”) drawn from a fixed pool of 20 log-spaced frequencies between 200 Hz and 3000 Hz, with square-wave amplitude modulation rates set at either 3 Hz, 7 Hz, or 19 Hz, applied with 5 m cosine ramps at the onset and offset of each pulse. Tone frequencies and modulation frequencies were set so that they would not be multiples of each other. The second phase consisted of 1.5 s of silence. During the third phase (“test phase”), a single stream (2 s) was presented: in half the trials, the stream was identical in both frequency and amplitude modulation rate to one of the streams presented in the encoding phase, whereas in the other half it had a new unique combination of frequency (sampled from the three frequencies presented in the encoding phase) and modulation rate (Fig. 1). For each trial, participants were asked to memorise the three streams presented simultaneously in the encoding phase and determine whether the single stream heard in the retrieval phase was one of the three tones they memorised. This is analogous to a reversed delay match-to-sample task, in that the options are presented before the sample. Participants responded by pressing the “F” key if they believed the target tone was present in the encoding phase, and “D” key if it was not. Participants were allowed to respond as soon as they heard the target stream and up to 4 s after the stream offset. Before the task was administered, the experimenter played several sample sounds to make sure participants understood the task. 100 trials were generated for each participant using MATLAB (2015b; The MathWorks Inc, 2015). Stimuli were generated at a 44.1 kHz sampling rate, saved as WAV files, and subsequently presented to participants in the form of two blocks of 50 trials each, with a break in between the two blocks. Visual feedback was provided for each trial and a summary score of false alarms, correct, and invalid responses was displayed at the end of each block. Target detection sensitivity was calculated as d’ following a “1/2 N” correction for extreme proportions of hit or false alarm rates (Macmillan & Kaplan, 1985; Stanislaw & Todorov, 1999).

2.2.7. *Psychophysics*

Six psychophysical tasks were administered using the Maximum Likelihood Procedure (MLP) for auditory threshold estimation implemented in the Psychoacoustics toolbox (Soranzo & Grassi, 2014) in MATLAB (2013b; The MathWorks Inc, 2013) running on a MacBook computer. During the pitch discrimination (PD), duration discrimination (DD), intensity discrimination (ID), and sinusoidal amplitude modulation detection (SAMD) tasks, participants were asked to listen carefully to three randomly ordered sounds in a sequence (3AFC): 2 standard sounds set to a fixed parameter value, and 1 target sound whose parameter changed adaptively across trials. They then identified the sound that differed (‘odd one out’) from the standard sounds by pressing 1, 2 or 3 on the keyboard. For the interaural level difference (ILD) and interaural time difference (ITD) tasks, only two sounds were presented (2AFC), and participants were asked to identify whether the first of the two sounds was perceived as coming from the left or from the right (with the second sound having the same parameter magnitude but coming from the opposite side). All six psychophysical tasks were administered in two blocks of 20 trials each and no feedback was provided. Details of all six psychophysical tasks are reported in Table 2.

The MLP aims to achieve a fast estimate of psychophysical thresholds through a nonparametric adaptive procedure. After each trial, the procedure identifies the logistic function that best fits the expected psychometric function of each participant based on their current responses and other fixed variables such as the function slope, expected error rates (e.g. due to attentional lapses), and chance level (e.g. 0.33 for a 3AFC). The toolbox’s default function slopes and expected error rates were used for all tasks. The procedure then calculates the next target stimulus as the parameter corresponding to a certain probability of a correct answer (i.e., 0.73 for 3AFC, 0.81 for 2AFC) in the previously estimated psychometric function. Details of the psychometric function estimation and stimulus selection are described in Grassi & Soranzo, 2009). This procedure is very sensitive to attentional lapses, particularly at the beginning of each block, as early estimations of the participants’ psychometric functions affect all remaining estimations within a given block. For this reason, if participants failed to identify the target stimulus in the first trial, which was always the easiest and therefore expected to elicit a correct answer, the corresponding block was marked as invalid. Final thresholds were calculated as the average of the two blocks after

Table 2
Details of the six psychophysical tasks.

Task name	Stimulus	Parameter	Parameter range: Min (standard), Max	Target stimulus
Pitch discrimination (PD)	250 ms complex tones with four harmonics	Frequency	330 Hz, 390 Hz	Highest-pitched tone
Duration discrimination (DD)	Complex tones with four harmonics ($f0 = 330$ Hz)	Duration	250 ms, 450 ms	Longest tone
Intensity discrimination (ID)	Complex tones with four harmonics ($f0 = 330$ Hz)	Intensity	−30 dB FS, −20 dB FS	Loudest tone
Sinusoidal amplitude modulation detection (SAMD)	500 ms Gaussian noise with 60 Hz sinusoidal amplitude modulation	Depth of modulation expressed as 20log(m)	$m \approx 0.14$ (−40 dB), $m \approx 0.78$ (−5 dB)	Amplitude-modulated tone
Interaural level difference (ILD)	5000 Hz, 250 ms pure tones	Intensity (opposite signs for left and right ear)	±0.1 dB, ±5 dB	First tone (either left or right)
Interaural time difference (ITD)	330 Hz, 250 ms pure tones	Phase (opposite signs for left and right ear)	±0.0001 ms, ±0.3 ms	First tone (either left or right)

excluding blocks that were determined to be invalid.

Finally, we calculated the difference between participants' thresholds and the standard values of the Pitch Discrimination, Duration Discrimination, and Intensity Discrimination tests. For instance, a threshold of 335 Hz for the Pitch Discrimination task, which has a standard value of 330 Hz, would correspond to a difference of −5 Hz. This was done to facilitate data visualization across auditory tasks by having greater values always correspond to a greater sensitivity.

2.2.8. *Stochastic auditory scene (StAS)*

This task aimed at measuring participants' sensitivity to statistical changes in auditory sequences. Participants were presented with random sequences of concatenated 50 ms tone pips (gated on and off with 5 ms raised cosine ramps), selected with replacement from a pool of 20 distinct log-spaced frequencies between 222 and 1912 Hz (12% steps or 1/6 of an octave). All trials began with a series of randomly selected tones drawn from the pool. In half the trials, after 40–50 tones (with the number drawn randomly per trial), the sequence would then switch to a halved pool of only 10 frequencies for 40 tones (i.e., 2 s). There were two conditions: in the "full-to-middle" (F-M) condition, the halved pool consisted of the 10 middle frequencies (391-1085 Hz) of the original pool, whereas in the "full-to-edge" (F-E) condition it consisted of the five highest (1215-1912 Hz) and five lowest (222-349 Hz) frequencies (Fig. 1). Listeners were instructed to press the spacebar as soon as they heard a change in the auditory scene. Although they were not given information on what exactly would change, participants were provided with several examples and one practice trial per condition, as well as receiving visual accuracy feedback on the screen at the end of each sequence. Overall detection sensitivity was obtained by calculating d' for the two conditions, correcting for extreme proportions of hit and false alarm rates (Macmillan & Kaplan, 1985; Stanislaw & Todorov, 1999), and averaging them.

2.3. *Data preprocessing*

Scores for all behavioural tasks were screened for univariate outliers and patterns of missing data using JMP 15.2.1.

2.3.1. *Outliers*

Extreme data points were evaluated manually based on overall data distributions, previous benchmarks, and a combination of robust measures of centre and spread. More specifically, values over one interdecile range from the first or ninth decile or over 4 robust spreads from the centre (M-estimates; Huber, 1973; Huber, 2011) were initially flagged as extreme. A total of 5 individual data-points were flagged as outliers: pitch discrimination (29.52 Hz, or about 8.9% of the 330 Hz reference stimulus) for participant M6; intensity discrimination (7.08 dB SPL) for participant C20; sinusoidal amplitude modulation detection (−4.25 dB, 20log(m)) for participant M9; interaural level difference (4.54 dB) for participant C15; and speech-in-babble (−24.5 dB SNR) for participant M22. The first 4 observations correspond to exceptionally high (i.e., poor) psychoacoustic threshold estimates, by far higher than any other participant or benchmark (e.g. Kidd, Watson, & Gygi, 2007). Further inspection revealed that these were due to mistakes (e.g., attentional lapse, wrong button pressed, random guessing) made by participants within the first few trials of both blocks, to which the MLP staircase procedure is particularly sensitive (Soranzo & Grassi, 2014). For this reason, these measurements were judged as invalid and excluded from further analyses. The speech-in-babble outlier, on the other hand, corresponds to an extremely low SNR threshold (i.e., exceptionally good) which cannot be ruled out as a measurement error and so it was retained. None of the other potential outliers identified by manual inspection of data distributions could be attributed to technical error and so they were retained as valid measurements.

2.3.2. *Missing data*

A total of 15 out of 640 (2.3%) missing data points were identified across the behavioural dataset, due to either outlier exclusion, technical issues during testing, or time constraints. Missing data points were distributed across nine participants who failed to complete one tasks each, and one participant (E1) from the audio engineers group who failed to complete six tasks (auditory scene recall and all psychophysical tasks except interaural time difference). Gold-MSI data for 3 participants from the musician group was also missing due to data corruption during the online questionnaire saving process. When using statistical methods that require complete data vectors for every participant (i.e. variable importance, see 2.4.3), Participant E1 (audio engineers) was entirely excluded; for the remaining participants, multivariate normal imputation based on a least squares prediction from the non-missing variables with shrinkage (Schäfer & Strimmer, 2005) was calculated for the rest of the dataset using JMP (15.2.1) and employed as an alternative to listwise deletion in order to retain as much information as possible (Schafer, 1999). While imputation can generate redundancy in data and increase the risk of Type 1 error, listwise deletion can increase Type 2 error and reduce statistical power (Cheema, 2014; Mishra & Khare, 2014). Where applicable, both methods were utilised and results compared to verify whether missing data would cause critical differences in statistical analyses. Since results obtained with both procedures were nearly identical, for simplicity, only the results obtained with imputation of missing data are reported. Pairwise deletion was instead employed when calculating correlations (see 2.4.5).

2.4. *Statistical analyses*

Statistical analyses are divided in two sections. First, we tested the a priori global hypothesis that groups do not come from the same auditory population, followed by the more specific hypothesis that auditory experts (i.e., musicians and audio engineers) will outperform controls across each of the auditory tasks, with emphasis on inferential statistics and control of Type 1 error. Second, we ran a series of exploratory analyses to uncover any meaningful patterns in the dataset, as well as to test finer-grained hypotheses regarding the specific differences between the musician and audio engineer populations, relationships between auditory tasks, and the role of musical and audio engineering experience. Methods for data exploration included graphical methods, descriptive statistics, point estimates of relevant sample statistics, and data-driven models (Behrens, 1997; Szucs & Ioannidis, 2017). Any a posteriori hypothesis formulated during data exploration was made explicit in order for the associated confidence intervals and *p*-values to be interpreted as per their descriptive content (Amrhein, Korner-Nievergelt, & Roth, 2017; Lavine, 2014), rather than as confirmatory evidence for inference at the population level (Cohen, 1994; Gaus, 2015). Robust metrics and/or nonparametric methods were preferred across all statistical analyses to accommodate for differences in distribution characteristics across tasks and groups, unbalanced classes, heteroscedasticity, and presence of outliers without recurring to arbitrary data transformations or post-hoc analytic choices. Finally, the signs of all psychophysical and speech-in-babble thresholds were reversed before analyses so that a greater number always represented better performance across all tasks to improve readability.

2.4.1. *Multivariate differences (nonparametric MANOVA)*

The global null hypothesis of no group differences in auditory skills was tested with the *nonpartest* function in the *npmv* R package (version 2.4.0; Ellis, Burchett, Harrar, & Bathke, 2017), which employs a multivariate ANOVA-type test statistic based on ranks (Brunner, Dette, & Munk, 1997; Brunner & Munzel, 2000) and *p*-values calculated via an asymptotic F-distribution approximation (Bathke & Harrar, 2008) or resampling. This is a nonparametric equivalent of a MANOVA.

2.4.2. *Univariate multiple comparisons and relative effects*

In the case of a rejection of the multivariate null hypothesis, a set of univariate tests was planned to test whether experts outperform controls in each auditory task. This was done with a rank-based nonparametric multiple contrast test procedure (MCTP) implemented in the *mctp* function in the *nparcomp* R package (version 3.0; Konietschke, Placzek, Schaarschmidt, & Hothorn, 2015; Noguchi, Abel, Marmolejo-Ramos, & Konietschke, 2020). This procedure was selected for all univariate comparisons as it does not make assumptions about distribution shape, heteroscedasticity, or class imbalance. The MCTP tests hypotheses of stochastic inequality, that is the probability of a random observation from one sample to be larger (or smaller) than a random observation from another sample. This operationalises the notion that one group will tend to outperform another without reference to measures of central tendency and spread (Cliff, 1993; Delaney & Vargha, 2002). This probability is referred to as relative effect and was calculated for each group against a reference unweighted mean distribution of all group distributions, so that a random measurement from one group is always evaluated in the context of the entire dataset. Relative effects were used to formulate hypotheses about group inequalities. Specifically, for each auditory task, we tested the one-tailed null hypothesis that control participants will show equal or better performance compared to musicians or audio engineers, that is an equal or higher relative effect. The rejection of a null hypothesis for a given task would then support the alternative hypothesis that one or both auditory expert cohorts scored significantly higher than controls for that task. This was done by setting *type* = *"Dunnett"* (i.e., many-to-one comparisons) and *alternative* = *"greater"* in the *mctp* function. In addition to the simple difference between relative effects, a point estimate of a transformed log odds-type effect size comparable in magnitude to Cohen's d was also calculated and reported to facilitate interpretation (Noguchi et al., 2020). The MCTP is a single step procedure, in that overall and specific contrasts are evaluated at the same time with no contradiction (i.e., a statistically significant omnibus test always corresponds to a significant "post-hoc" test and vice-versa) and under strong control of the family-wise error rate (FWER). Asymptotic estimates of adjusted *p*-values and simultaneous confidence intervals were calculated following a multivariate t-based approximation with adjusted degrees of freedom (Noguchi et al., 2020). The *p*-values of the overall effects, which always correspond to the lowest p-value of any pairwise comparison, were further corrected following the Benjamini-Hochberg false discovery rate (FDR) (Benjamini & Hochberg, 1995) adjustment implemented in the *p.adjust* function from the stats R package. An equivalent testing procedure for simple pairwise comparisons (i.e., a studentised permutation test (Neubert & Brunner, 2007) with the *npar.t.test* function from the same package) was used to complement plots and descriptive statistics during exploratory analyses between audio engineers and musicians. In these cases, *p*-values were left uncorrected and explicitly reported as such to suggest an appropriate interpretation.

2.4.3. *Classification of musicians and audio engineers: Variable importance*

To further explore the different characteristics of our expert cohorts on a multivariate basis, we extracted variable importance from a random forest classifier (Breiman, 2001) trained with personality scores, Gold-MSI sub-dimensions, and auditory measures as predictors. Random forests are non-parametric algorithms that aggregate predictions from binary decision trees constructed on bootstrap samples or sub-samples of the original dataset and random subsets of predictors (for an overview, see e.g. Strobl, Malley, & Tutz, 2009). We selected a class of random forests that utilises conditional inference trees as base classifiers (Torsten, Kurt, & Achim, 2006). These perform permutation tests (Strasser & Weber, 1999) at each node to identify the predictor most strongly associated with the response variable along with the optimal split point that maximises the discrepancy between the subnodes (Torsten et al., 2006). This method, when applied with subsampling without replacement, has been shown to be unbiased to the nature of a predictor (e.g., categorical, scale, ordinal). This differs from other types of binary decision trees that rely on measures of impurity reduction such as classification and regression trees (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). This feature is particularly important as our predictors include both continuous variables and low-cardinality questionnaire data. We grew our forest with *cforest* from the *partykit* R package (Hothorn & Zeileis, 2015), with hyper parameters set to *ntree* = 10,000 (number of trees in the forest), *mtry* = 4 (number of random predictors tested at each node of a tree; default is $\sqrt{p}$ where p is the number of predictors), and perturb set to a subsampling fraction of 0.632 with no replacement in order to achieve unbiasedness to predictor type (see above). Trees in the forest were allowed to fully grow by setting *minsplit* = *minbucket* = 1 (minimum size of a node), only limited by a minimum significance of a permutation test set with *mincriterion* = 0.95 (1-*p*-value). These were set with the goal of achieving a compromise between variance (i.e., node size of 1) and bias (i.e., high criterion of 0.95) (see guidelines in Probst, Boulesteix, & Bischl, 2019). The importance of each predictor in the model was calculated as conditional permutation importance (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). Permutation importance corresponds to the mean decrease in prediction accuracy when the values of a predictor are randomly permuted. Conditional permutation importance also accounts for collinearity between variables by measuring associations between predictors and permuting collinear ones together. This was calculated using the *varimp* function in *partykit* with parameters *nperm* = 5 (number of permutations), conditional = TRUE, and threshold = 0.95. As per default, prediction accuracy and importance were calculated on the "out-of-box" data (i.e., OOB = TRUE), that is on the data excluded during subsampling. Random forests were employed here as a fully nonparametric tool for data exploration (Jones & Linder, 2015) which, given a high number of predictor variables and low number of observations, specifically serves the purpose of identifying and ranking a subset of variables (i.e., feature selection) that can best describe the differences between musicians and audio engineer. As multiple imputations and listwise deletion lead to interchangeable results, only results following listwise deletion are reported. This corresponds to a total of 40 participants, 19 audio engineers and 21 musicians. Variables with importance above 2.5%, corresponding to mean decrease in accuracy equivalent to at least one participant (i.e., 100%/40), were included in an alternative reduced model. For the purpose of replicability, results were obtained using a random seed of 1112.

2.4.4. Musical and audio engineering experience

To draw a more direct comparison between musicians and audio engineers with a similar musical background, we clustered participants in two groups based their score in the Musical Training sub-dimension of the Gold-MSI questionnaire. Specifically, apart from one musician who scored 31, musicians scored between 37 and 49 (Fig. 2). Therefore, using a cut-off of 37, we were able to match all but this one musician with 8 audio engineers with a similar musical background. The underlying meaning of this cutoff was further examined using two items of the Gold-MSI questionnaires that contribute the musical training score, namely "I engaged in regular, daily practice of a musical instrument (including voice) for__years" and "I have had__years of formal training on a musical instrument (including voice) during my lifetime," in order to qualify possible differences in formal or informal training between cohorts (Fig. 2). We then re-examined differences in behavioural measures between musicians and engineers with a similar level of musical training, as well as audio engineers with different levels of musical training, using the same methods described in paragraph 2.4.2. Additionally, we explored associations between mixing and mastering experience and behavioural measures with Spearman correlations.

2.4.5. Correlations between auditory tasks

Monotonic relationships between behavioural variables were estimated using Spearman's rank correlations coefficients (ρ) separately for each group. Empirical confidence intervals for individual bivariate ρ were calculated via bootstrapping (Haukoos & Lewis, 2005; Wright, London, & Field, 2011). Relevant correlations, as well as their differences across groups, were assessed graphically with a series of correlograms as well as bivariate scatterplots on both raw data and ranked data. To facilitate comparisons between groups, data were ranked within group and centred at the median rank before plotting.

3. Experiment 1: Results

3.1. Auditory expertise: Multivariate and univariate tests

The multivariate null hypothesis that participants come from the same population was rejected (ANOVA-type test statistic = 4.254, df1 = 11.616, df2 = 301.082, p-value <0.0001), confirming that groups do indeed exhibit overall different degrees of auditory ability. After FDR correction, the null hypothesis of stochastic equality between experts and controls was rejected at the 0.05 level on all tasks except duration discrimination, intensity discrimination, and stochastic auditory scene (full details of test statistics can be found in Table 3). On perceptual tasks, both students of audio engineering and musical instrumentalists had significantly lower thresholds for pitch discrimination and interaural time difference tasks than controls. Musicians also showed significantly lower thresholds than controls on sinusoidal amplitude modulation discrimination and interaural level difference tasks (Fig. 3). On auditory scene tasks, musicians were more accurate than controls on the sustained auditory selective attention task, while audio engineers were more accurate than controls on the auditory scene recall task. Finally, musicians, but not engineers, showed significantly lower SNR thresholds for the speech-in-babble-noise task (Fig. 4). Pitch discrimination had the largest expertise-related effect size across all auditory tasks for both expert cohorts compared to the control group, with median thresholds for audio engineers (median = 3 Hz, or 0.9% difference reference tone, MAD = 1.659 Hz (0.5%) and musicians (median = 3 Hz (0.9%), MAD = 1.248 Hz (0.37%) being approximately half of those of control participants (median = 6.667 Hz (2%), MAD = 3.983 Hz (1.2%).

3.2. Differentiating auditory expert cohorts: Exploratory analyses

3.2.1. Random forests: Variable importance

To summarise the variables in our dataset that can best discriminate between musicians and audio engineers and rank their relevance, we calculated conditional permutation importance - i.e., mean decrease in classification accuracy following a permutation of a given predictor - of a random forest classifier built on all variables in our dataset. The overall accuracy of the full model including all 20 predictors was 80%. A reduced model (Fig. 5) which only included variables with importance above 2.5% had an accuracy of 82.5%. The predictor with the largest influence on prediction accuracy was singing abilities (25.6%), followed by speech-in-babble-noise thresholds and musical training (~15%), and emotional stability (5.4%). Minor contributions between 2.5% and 5% were obtained for active engagement and 2 psychophysical tasks, interaural time difference and duration discrimination. Bivariate Spearman correlations among the top three predictors revealed that while singing abilities and musical training were strongly correlated for both musicians (ρ = 0.53, 90% CI [0.14, 0.83]) and audio engineers (ρ = 0.62, 90% CI [0.20, 0.86]), speech-in-babble-noise thresholds had no correlation with either predictor.

3.2.2. Auditory tasks

Data plots (i.e., Fig. 3, Fig. 4) and descriptive statistics were used to integrate the results from the random forest importance classification and interpret the directionality of its prediction. In terms of behavioural variables, speech-in-babble-noise thresholds of musicians (median = −9.87 dB SNR, MAD = 1.82 dB SNR) were significantly lower than both controls (median = −8.61 dB SNR, MAD = 1.42 dB SNR) and engineers (median = −8.15 dB SNR, MAD = 1.62 dB SNR; post-hoc Brunner-Munzel, effect size = 0.674, test statistic = 3.347, p = 0.003), although musicians were also the most inconsistent within group and displayed the largest range (20.5 dB SNR) of responses on this task - a point we return to below, and in Experiment 2. As for the other auditory scene performance tasks which did not add a unique contribution to classification accuracy according to the random forest model, median sustained auditory selective attention d' was marginally higher for the musician group (0.777, MAD = 0.173) than audio engineers (0.709, MAD = 0.148), while the opposite was true for the auditory scene recall task (audio engineers: median = 1.411, MAD = 0.307; musicians: median = 1.187, MAD = 0.519), although these differences were not statistically significant.[4] As for psychophysical tasks, with the exception of sinusoidal amplitude modulation discrimination, audio engineers' median thresholds were the lowest across all tasks, albeit by also a very small margin. The most apparent difference between expert cohorts (Fig. 3) was duration discrimination (audio engineers: median = 29.03 ms, MAD = 5.98 ms; musicians: median = 32.55 ms, MAD = 9.47 ms), although a post-hoc test showed this difference was also not statistically significant at the 0.05 level (post-hoc Brunner-Munzel, effect size = 0.408 test statistic = 1.927, p = 0.063).

3.2.3. Musical expertise and personality

Musical sophistication (Fig. S1) and personality traits (Fig. S2) were among the most important variables in the discrimination of musicians and audio engineers. Unsurprisingly, musicians scored substantially higher than audio engineers in the musical training (post-hoc Brunner-Munzel, effect size = 0.693, test statistic = 3.311, p = 0.004) and singing abilities dimensions (post-hoc Brunner-Munzel, effect size = 1.1486983, test statistic = 6.505, p < 0.001) of the Gold-MSI questionnaire, but also marginally higher in the perceptual abilities (post-

---

[4] An experiment with much larger sample sizes would be needed to appropriately test the statistical significance of such small effect sizes.

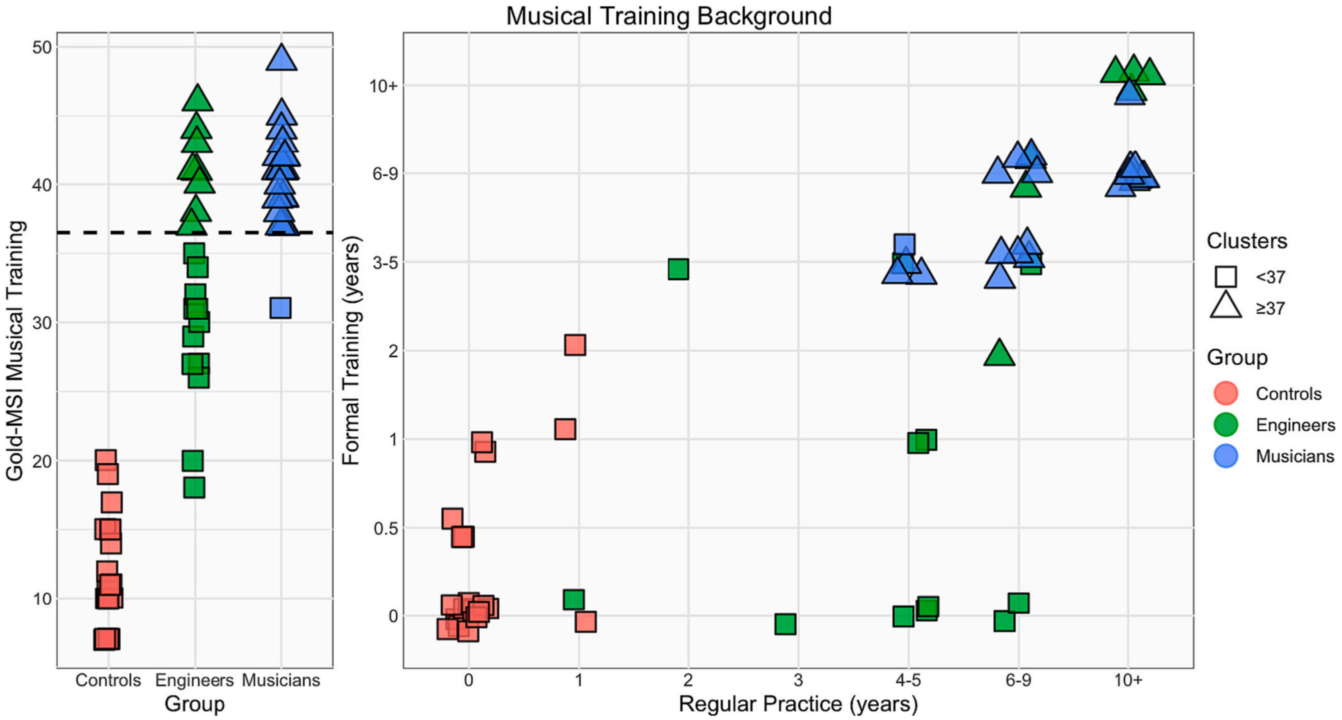


Fig. 2. Musical training background. Left plot represents Musical Training dimension scores from the Gold-MSI questionnaire. Data points above the dashed line correspond to musicians and audio engineers with a matching degree of musical training, defined by a Gold-MSI score higher or equal to 37, which captures all but one musician. Right plot shows the musical training background of the three cohorts, as well as musical training clusters, in terms of years of formal training and regular practice of a musical instrument. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Results of many-to-one testing procedure between audio engineers (E) and musicians (M) compared to controls (C). Tasks: pitch discrimination (PD), duration discrimination (DD), intensity discrimination (ID), sinusoidal amplitude modulation discrimination (SAMD), interaural level difference (ILD), interaural time difference (ITD), sustained auditory selective attention (SASA), auditory scene recall (ASR), stochastic auditory scene (StAS), speech in babble noise (SIN). H_a: Alternative hypotheses expressed as the probability that a random participant from the audio engineer group (E > C) or musician group (M > C) would have a higher score than a random participant from the control group. Rel. Effect [95% CI]: relative effects with one-tailed 95% confidence interval. Effect size: log-odds type effect size comparable in magnitude to Cohen's d. Statistic: test statistic. p: test significance with strong control of the family-wise error rate within each task. p_{omni}: significance of the omnibus test. p_{FDR}: significance of the omnibus test corrected for false discovery rate across all tasks (bolded if $p < 0.05$).

Task	H_a	Rel. Effect [95% CI]	Effect Size	Statistic	p	p_{omni}	p_{FDR}
Pitch	E > C	0.376 [0.236; 1.000]	0.938	5.316	<0.001	<0.001	**<0.001**
Discrimination	M > C	0.313 [0.170; 1.000]	0.782	4.336	<0.001		
Duration	E > C	0.081 [−0.098; 1.000]	0.192	0.888	0.288	0.288	0.320
Discrimination	M > C	−0.073 [−0.259; 1.000]	−0.173	−0.770	0.885		
Intensity	E > C	0.150 [−0.031; 1.000]	0.355	1.631	0.094	0.094	0.118
Discrimination	M > C	0.130 [−0.048; 1.000]	0.309	1.438	0.131		
Sin. Amplitude	E > C	0.136 [−0.046; 1.000]	0.324	1.466	0.125	0.016	**0.030**
Modulation Detection	M > C	0.220 [0.046; 1.000]	0.525	2.479	0.016		
Interaural Level	E > C	0.177 [−0.019; 1.000]	0.423	1.783	0.072	0.025	**0.036**
Difference Discrimination	M > C	0.208 [0.029; 1.000]	0.497	2.296	0.025		
Interaural Time	E > C	0.235 [0.079; 1.000]	0.564	2.984	0.005	0.005	**0.016**
Difference Discrimination	M > C	0.172 [0.004; 1.000]	0.413	2.026	0.045		
Sustained Auditory	E > C	0.132 [−0.034; 1.000]	0.318	1.575	0.109	<0.001	**0.004**
Selective Attention	M > C	0.275 [0.123; 1.000]	0.663	3.606	<0.001		
Auditory Scene	E > C	0.231 [0.049; 1.000]	0.554	2.491	0.016	0.016	**0.030**
Recall	M > C	0.146 [−0.032; 1.000]	0.349	1.606	0.097		
Stochastic	E > C	−0.084 [−0.278; 1.000]	−0.199	−0.866	0.918	0.689	0.689
Auditory Scene	M > C	−0.004 [−0.169; 1.000]	−0.009	−0.044	0.689		
Speech in Noise	E > C	−0.063 [−0.220; 1.000]	−0.151	−0.785	0.900	0.018	**0.030**
	M > C	0.207 [0.040; 1.000]	0.5	2.447	0.018		

hoc Brunner-Munzel, effect size = 0.498, test statistic = 2.393, $p = 0.028$) and emotions (post-hoc Brunner-Munzel, effect size = 0.448, test statistic = 2.134, $p = 0.041$) dimensions. However, a comparable level of active engagement with music was present in musicians compared to audio engineers. Results from the TIPI questionnaire revealed significant differences in emotional stability (post-hoc Brunner-Munzel, effect size = −0.523, test statistic = −2.569, $p = 0.015$), with musicians on average seeing themselves as less emotionally stable than audio engineers. Musicians and audio engineers also appeared to cluster around equally higher scores compared to controls in the openness to experience dimension, which included an item about creativity.

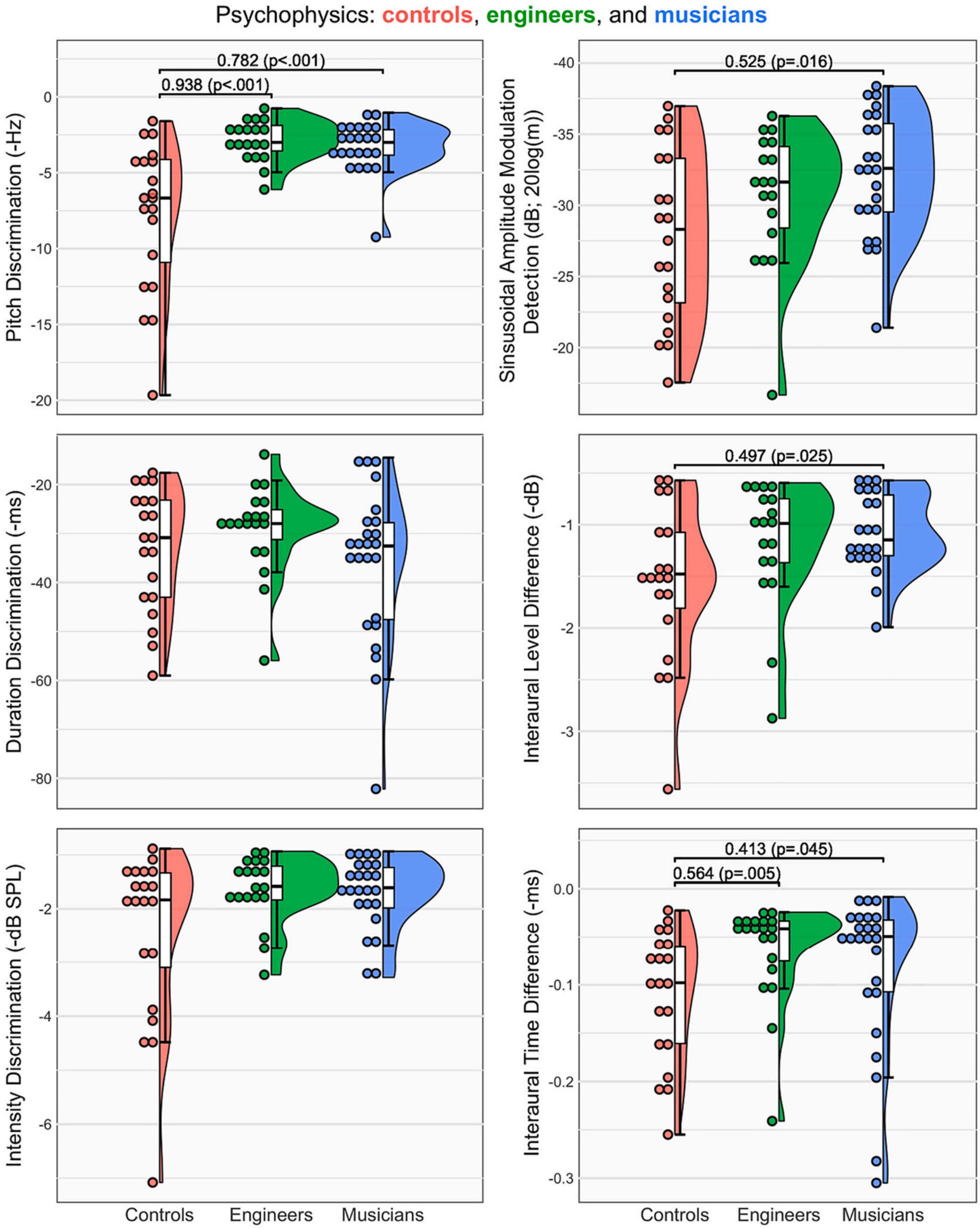


Fig. 3. Dot plots, same area violin plots, and box plots for all psychophysical measures by group. Just noticeable differences are reported on the y axes with opposite signs in order for a positive effect size to consistently correspond to a better performance across tasks. Brackets above graphs display log-odds-type effect size and one-tailed p values when $p < 0.05$.

3.3. *Musical training and audio engineering experience*

Clustering participants based on their musical training background (see 2.4.5) did not affect previous results: musicians displayed lower speech-in-babble thresholds than audio engineers with a matched degree of musical training (post-hoc Brunner-Munzel, effect size = 0.733, test statistic = 2.93, $p = 0.019$) and there were no significant differences in auditory ability between audio engineers with different musical backgrounds (nonparametric MANOVA, permutation test of ANOVA-type statistic with 10,000 replications, $p = 0.687$). On the other hand, audio engineering experience was moderately correlated with both stochastic auditory scene ($\rho = 0.43$, 90% CI [0.08, 0.70]) and speech in babble noise ($\rho = 0.49$, 90% CI [0.13, 0.78]) performance, although even the most trained participants' scores fell within the range of control participants.

3.4. *Associations between fine perception, auditory scene analysis, and speech in noise*

Among the auditory scene tasks, sustained auditory selective attention d' scores appeared to be the most consistently (i.e., across groups) associated with psychophysical thresholds, in particular with pitch discrimination, intensity discrimination, and interaural time difference (Fig. 6). Correlations between speech-in-babble-noise thresholds and psychophysical tasks were mixed across groups and overall negligible. Correlations between the auditory scene and speech-in-babble-noise

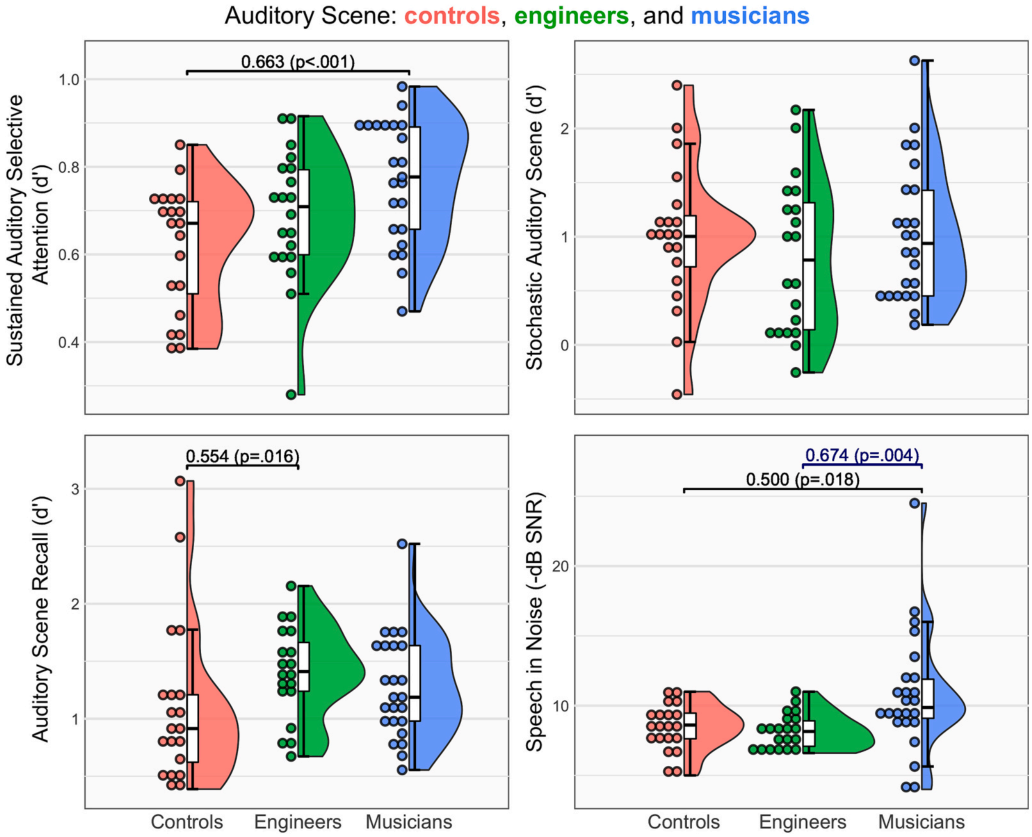


Fig. 4. Dot plots, same area violin plots, and box plots for all auditory scene tasks by group. Speech-in-babble thresholds are reported with opposite signs in order for a positive effect size to consistently correspond to a better performance across tasks. Brackets above graphs display log-odds-type effect size and one-tailed p values when $p < 0.05$. Values in blue brackets correspond to post-hoc two-tailed tests and are not corrected for multiple comparisons. Note that for the Auditory Scene Recall task, audio engineers' d' is significantly higher than that of controls overall, despite the two outlier control participants showing high d' values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

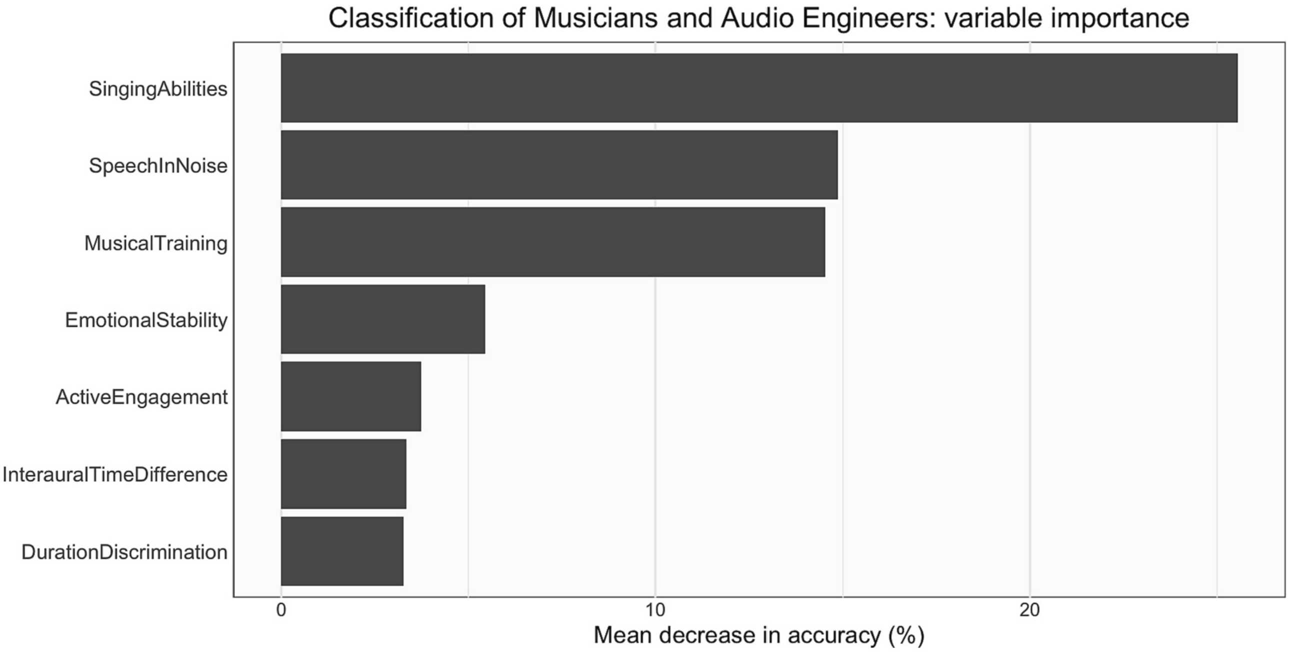
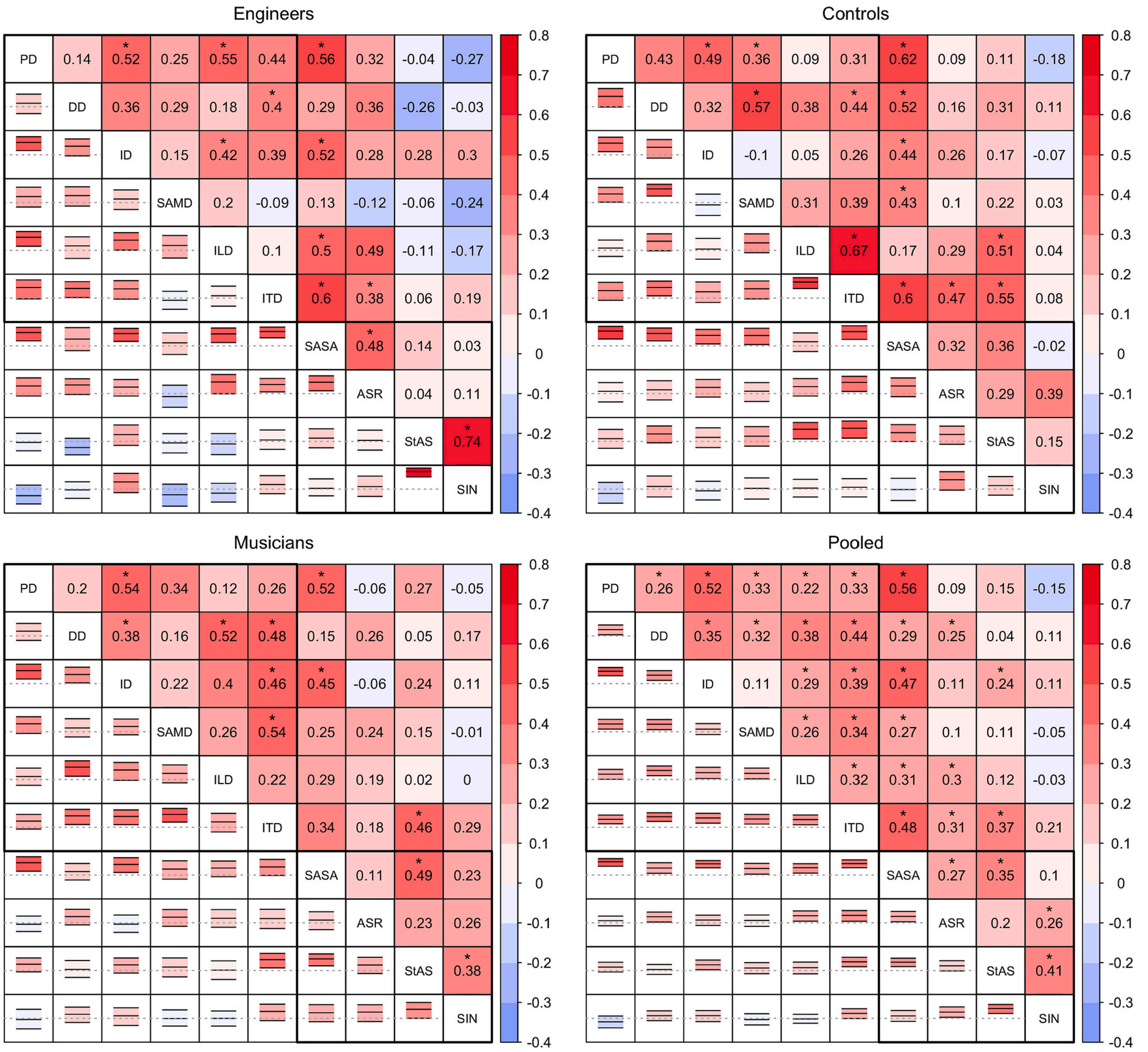


Fig. 5. Mean decrease in prediction accuracy (i.e., conditional permutation importance) of a random forest classifier with predictors including auditory tasks, personality traits, and musical sophistication sub-dimensions. Only the predictors with importance over 2.5% are shown.



Engineers

PD	0.14	*0.52	0.25	*0.55	0.44	*0.56	0.32	-0.04	-0.27
	DD	0.36	0.29	0.18	*0.4	0.29	0.36	-0.26	-0.03
		ID	0.15	*0.42	0.39	*0.52	0.28	0.28	0.3
			SAMD	0.2	-0.09	0.13	-0.12	-0.06	-0.24
				ILD	0.1	*0.5	0.49	-0.11	-0.17
					ITD	*0.6	*0.38	0.06	0.19
						SASA	*0.48	0.14	0.03
							ASR	0.04	0.11
								StAS	*0.74
									SIN

Controls

PD	0.43	*0.49	*0.36	0.09	0.31	*0.62	0.09	0.11	-0.18
	DD	0.32	*0.57	0.38	*0.44	*0.52	0.16	0.31	0.11
		ID	-0.1	0.05	0.26	*0.44	0.26	0.17	-0.07
			SAMD	0.31	0.39	*0.43	0.1	0.22	0.03
				ILD	*0.67	0.17	0.29	*0.51	0.04
					ITD	*0.6	*0.47	*0.55	0.08
						SASA	0.32	0.36	-0.02
							ASR	0.29	0.39
								StAS	0.15
									SIN

Musicians

PD	0.2	*0.54	0.34	0.12	0.26	*0.52	-0.06	0.27	-0.05
	DD	*0.38	0.16	*0.52	*0.48	0.15	0.26	0.05	0.17
		ID	0.22	0.4	*0.46	*0.45	-0.06	0.24	0.11
			SAMD	0.26	*0.54	0.25	0.24	0.15	-0.01
				ILD	0.22	0.29	0.19	0.02	0
					ITD	0.34	0.18	*0.46	0.29
						SASA	0.11	*0.49	0.23
							ASR	0.23	0.26
								StAS	*0.38
									SIN

Pooled

PD	*0.26	*0.52	*0.33	*0.22	*0.33	*0.56	0.09	0.15	-0.15
	DD	*0.35	*0.32	*0.38	*0.44	*0.29	*0.25	0.04	0.11
		ID	0.11	*0.29	*0.39	*0.47	0.11	*0.24	0.11
			SAMD	*0.26	*0.34	*0.27	0.1	0.11	-0.05
				ILD	*0.32	*0.31	*0.3	0.12	-0.03
					ITD	*0.48	*0.31	*0.37	0.21
						SASA	*0.27	*0.35	0.1
							ASR	0.2	*0.26
								StAS	*0.41
									SIN

Fig. 6. Correlograms of behavioural tasks for all groups and pooled correlations obtained by median-centring ranks by group. Top triangles: Spearmans' ρ. Positive correlations correspond to red-colored cells, negative correlations to blue-colored cells, while colour saturation reflects correlation magnitude. Correlations whose 90% empirical confidence interval does not include the null are marked with *. Bottom triangles: 90% empirical confidence intervals. Dashed horizontal lines represent ρ = 0. Thicker black margins identify psychophysical tasks and auditory scene tasks. Acronyms: PD = pitch discrimination; DD = duration discrimination: ID = intensity discrimination; SAMD = sinusoidal amplitude modulation discrimination; ILD = interaural level difference; ITD = interaural time difference; SASA = sustained auditory selective attention; ASR = auditory scene recall; StAS = stochastic auditory scene; SIN = speech in babble noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tasks also appeared fairly inconsistent across groups, with stochastic auditory scene showing the largest correlation with speech-in-babble thresholds in both engineers and musicians.

4. Experiment 2: Online coordinate response measure (CRM) task

The apparent advantage for musicians perceiving speech in babble versus both audio engineers and controls could potentially be related to the difference in dialect between audio engineers (Irish) and musicians and controls (English). Also, as noted above, there was considerable variability in musician performance on this task, with most musicians performing like the other groups, but a minority performing well at very challenging SNR levels. Therefore, we conducted an online coordinate response measure experiment as a follow-up test of the performing musician advantage we observed in recognizing words in the presence of competing babble. It also served as a test of the potential effect of native (Irish versus English) accent on perceiving a southern English accent in challenging conditions.

4.1. Methods

4.1.1. Participants

We recruited online participants via email advertisement from

Queen's University Belfast ($n = 32$) and via the online recruitment portal Prolific ($n = 84$) using the following criteria: no hearing impairments, English-speaking monolingual, age range 18–35, nationality (English; Irish or Northern Irish), experience with musical instrument (musicians ≥ 5 years; non-musicians ≤1 year). To further validate the screening criterion relating to musical training, we collected data on the number of years of regular practice (mean ± SD: musicians, 9 years ±4.8; non-musicians, 0.8 years ±1.8) and formal training (musicians, 5.9 years ±4.6; non-musicians, 0.7 years ±1.8) in any musical instrument. Additionally, we asked participants to describe their English background in their own words (e.g. *"I grew up in London, my family spoke Punjabi and English at home to me."*). Five participants whose accent was neither English nor Irish were excluded from further analyses. The resulting four groups were English musicians ($n = 24$, 10F, age = 29.6 years old ±4.2), English non-musicians ($n = 30$, 16F, age = 28.2 years old ±5.1), Irish musicians ($n = 27$, 15F, age = 23.9 years old ±4.3), and Irish non-musicians ($n = 32$, 23F, age = 26.3 years old ±5.2).

4.1.2. Procedure

The speech-in-noise task was an online implementation of the adaptive Coordinate Response Measure (CRM) (Bianco, Mills, de Kerangal, Rosen, & Chait, 2021; Bolia, Nelson, Ericson, & Simpson, 2000; Messaoud-Galusi, Hazan, & Rosen, 2011), in which the target stimuli were sentences including a colour and a number following the format: "show the dog where the [colour] [number] is". We chose this speech-in-noise task as a self-administered alternative to the one used in the first experiment, which instead requires interaction with an experimenter. Each stimulus was masked by the same speech babble used in the previous speech in multi-talker babble task (see paragraph 2.2.2), applied using an adaptive 1-up 1-down staircase procedure. The initial talker-masker ratio and step size were set to +20 dB and + 9 dB respectively. Step size was reduced by 2 dB at each reversal, with each run stopping after 4 reversals. Each participant completed 4 runs in total. To perform the task, participants clicked on boxes with the corresponding combination of number and colour. All combination of numbers (1 to 9, except the bisyllabic 7) and colours (black, white, green, red, blue, pink) were displayed at all times. Thresholds were calculated as the average SNR of the last 3 reversals. Before the start of the experiment, we used an online sound-level-setting paradigm (Zhao, Brown, Holt, & Dick, 2022) to help participants set the amplitude of the stimulus at an average of 74 dB SPL (range 67–80 dB SPL).

4.2. Results

A 2 × 2 ANOVA was conducted to examine the effect of native accent and musical training on speech-in-noise perception thresholds (Fig. 7). There was no statistically significant main effect for either variable (Accent, F(1,112) = 0.447, $p = 0.505$; Musical training, F(1,112) = 0.031, $p = 0.860$), nor their interaction (F(1,112) = 0.473, $p = 0.493$). Furthermore, there was no correlation between SNR thresholds and years regular instrument practice (rho = −0.08) or formal musical training (rho = −0.05). As can be seen in Fig. 7, there is essentially complete overlap in the range and distribution of thresholds across groups.

5. Discussion

5.1. Results summary

This study introduced a novel population of auditory experts: audio

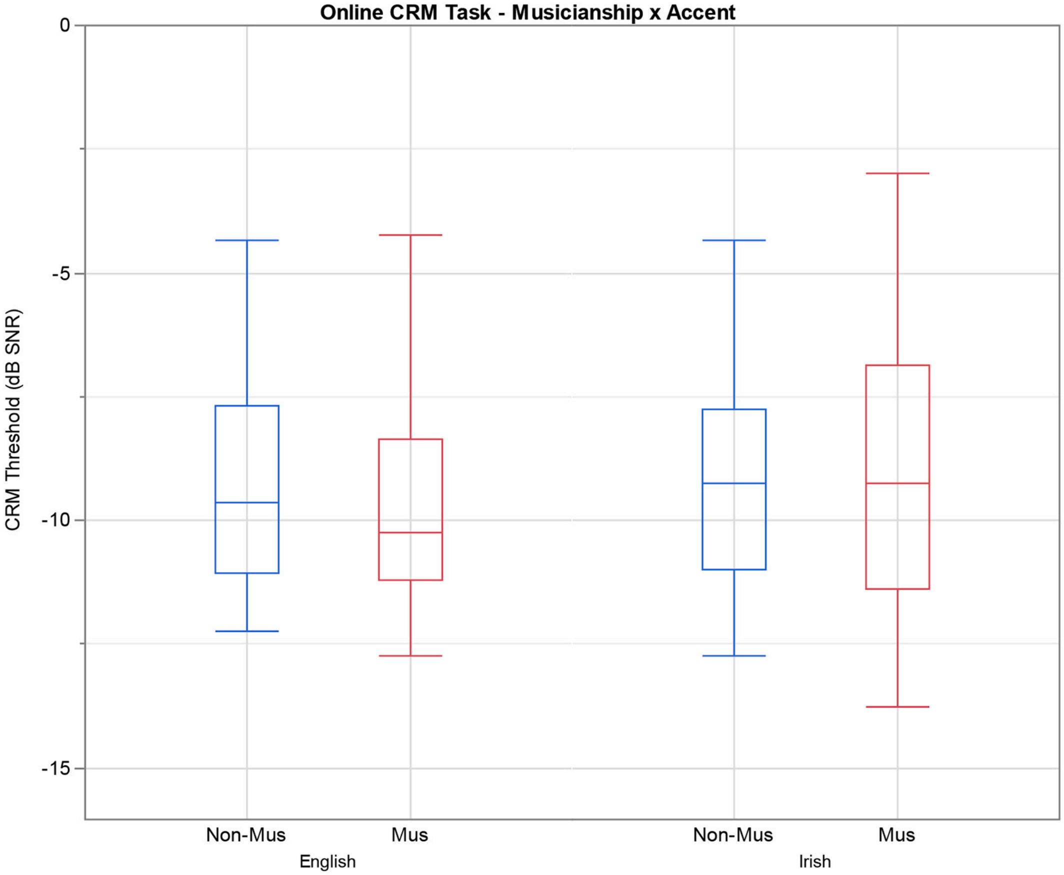


Fig. 7. Results of the online CRM task. Boxplots and jittered individual data points represent participants' speech-in-babble-noise SNR thresholds across 2 × 2 levels of musicianship (musicians, non-musicians) and native accent (English, Irish).

engineers. We first tested the hypothesized superiority of their fine perceptual and auditory scene analysis skills in relation to naive participants and contrasted their performance to that of music instrumentalists. We found that, when compared to naive participants, both auditory expert cohorts had lower thresholds for pitch discrimination and interaural time difference discrimination, while musicians also had lower thresholds for sinusoidal amplitude modulation discrimination and interaural level difference discrimination. Audio engineers performed better than controls in auditory scene recall, which requires participants to determine whether a target sound matches one of three sounds presented earlier in terms of pitch and amplitude modulation frequency. On the other hand, musicians outperformed controls in sustained auditory selective attention, during which participants identified repetitions of three-tone sequences in an auditory stream while ignoring a competing stream one octave apart. Musicians also had significantly lower thresholds for speech-in-babble-noise perception than both naive participants and audio engineers (a result that did not extend to the new sample of musicians and non-musicians in Expt 2). Both auditory expert cohorts showed higher levels of openness to experience and audio engineers had higher levels of emotional stability compared to musicians. Audio engineers had a wider variety of musical backgrounds, although controlling for this did not affect previous conclusions about group differences. The number of years of audio engineering experience was moderately associated with better sensitivity in the stochastic auditory scene task and lower speech-in-babble-noise thresholds, but overall scores for both tasks fell within the normal range. Finally, psychophysical scores were the most associated with sustained auditory selective attention and speech in babble noise was associated stochastic auditory scene, particularly for audio engineers.

In sum, we gathered evidence supporting the hypothesis that audio engineers' auditory expertise, similarly to musical training, corresponds to generalised advantages in fine auditory discrimination and auditory scene analysis. The apparent musician-selective advantage for speech in babble - a topic that has been of considerable debate in the literature - was more ephemeral, with the initial result from Experiment 1 not extending to the online study of Experiment 2.

5.2. Fine perception

Musicians and audio engineers showed superior fine auditory perception, with pitch discrimination having the largest effect size and clear-cut separation between experts and controls. Thresholds for the control group followed generally a wider distribution, as reflected by a higher median absolute deviation across all psychophysical tasks, and the top performers always matched the performance of the expert groups. These results reflect one challenging aspect of designing a control group for expert populations in a cross-sectional study, as pseudo-randomly sampling from the general population will unavoidably correspond to a wider spectrum of responses and include highly skilled individuals, despite controlling for musical training (Law & Zentner, 2012). Overall, we could not detect a clear advantage of musicians over audio engineers or vice versa in fine auditory perception.

5.3. Auditory scene analysis and speech in babble noise

While musicians significantly outperformed controls in sustained auditory selective attention (SASA), audio engineers performed better than controls in auditory scene recall (ASR).[5] In addition to the differences in cognitive loads for each task (i.e., SASA relies more heavily on sustained selective attention and ASR on working memory), SASA stimuli are comparatively more "musical," in that the two competing auditory streams are constructed from the first three tones of a major scale separated by an octave, which might resemble competing melodies. ASR stimuli, on the other hand, are simple pure tones defined by a pitch and an amplitude modulation frequency but have no tonal relation with each other. Audio engineers' selective attention ability in this case might benefit from a more technical understanding of sound components and a more generalised experience working with any type of sound, musical and non-musical. Furthermore, this task required participants to analyse and maintain the whole auditory scene (i.e., three sounds) in memory, as no distinction between target and foil can be made until the target sound is heard. This type of mental sound manipulation and asynchronous pre-post comparison is common during mixing practices (see paragraph 1.2.1) and could in part account for the audio engineers' advantage.

In Experiment 1, musicians showed significantly lower SNR thresholds for speech-in-babble-noise perception compared to both controls and audio engineers, even accounting for differences in musical training. However, musical training and general musical sophistication (as measured by the Gold-MSI questionnaire) showed no association with speech-in-babble-noise thresholds within each group, implying that the musician effect might be due to characteristics intrinsic to the group not detected by our test battery - a possibility we explored in Experiment 2. Indeed, the only behavioural measure in our data that showed a fairly consistent positive correlation across groups with speech-in-babble-noise perception was sensitivity to statistical changes in a stochastic auditory scene. This could be explained by a better ability to detect changes in higher-order statistics of a sound sequence (Barascud, Pearce, Griffiths, Friston, & Chait, 2016; Skerritt-Davis & Elhilali, 2018), spectral entropy (Overath et al., 2007; Stilp & Kluender, 2010) or, more generally, informational content in a noisy signal, which is a strategy implemented for instance in speech-in-noise recognition algorithms (e.g. Misra, Ikbal, Bourlard, & Hermansky, 2004; Toh, Togneri, & Nordholm, 2005). However this is speculative and a dedicated experiment is needed to test this specific hypothesis. In this vein, Oberfeld and Klöckner-Nowotny (2016) found that individual differences in selective attention — measured in both auditory and visual modes — could explain variations in speech-in-noise perception abilities in their sample. However their stimuli consisted of two individual competing speakers presented binaurally and one central target speaker, which might more explicitly depend on the ability to pay selective attention to one of multiple intelligible elements. Similarly, Tierney et al. (2020) found a correlation between non-verbal sustained selective attention and the perception of speech masked by one distractor talker. On the other hand, De Kerangal, Vickers, and Chait (2021) found an association between musical training and sustained attention, but not between sustained attention and speech in noise perception.

Furthermore, musical training and general musical sophistication (as measured by the Gold-MSI questionnaire) showed no association with speech-in-babble-noise thresholds within each group, implying that the musician effect might be due to characteristics intrinsic to the group not detected or controlled by our test battery. As noted above, one of these characteristics is the different regional accent of audio engineers (Belfast, UK) and musicians (London, UK). It has been shown that listening to speech in an unfamiliar accent can negatively affect the perception of speech in noisy enviroments even after long-term familiarization, possibly due to an increase in processing cost (e.g. Smith, Holmes-Elliott, Pettinato, & Knight, 2014). For this reason, we hypothesized that audio engineers might have been at a disadvantage, given that target sentences were spoken in a Southern British English accent, and conducted a second experiment to test this hypothesis.

5.4. Speech in babble noise, musical expertise, and native accent

Data from our second online speech-in-noise experiment suggest that neither native accent nor musical training affect the perception of speech in babble noise, which appears to contradict our previous findings. On the one hand, it is possible that results from the first study were simply due to random sampling. On the other hand, the studies differ in several qualities which might have affected participants' SNR thresholds. Specifically, in the first study, responses were directly communicated to the experimenter in an acoustic environment that was controlled and kept constant across participants (i.e. quiet room, headphones, sound pressure levels), which might have led to more precise measurements. Additionally, there was a much greater variability in musical training in the online experiment due to the limitations and self-report nature of the screening procedure, indicating that the "musician" groups in the two experiments might not come from the same expert population. Nonetheless, when examining musical expertise directly, we did not observe any significant correlations between SNR thresholds and years regular instrument practice nor formal musical training.

---

[5] The effect sizes for the direct comparison between expert cohorts were much smaller and did not reach statistical significance.

In summary, our observations appear to be in line with the current literature on this topic, namely the difficulty in capturing a consistent advantage for speech-in-noise perception in musicians. If the hypothesis on the musician advantage is actually true, several variables that are assumed to be constant across studies might affect measurements to a greater extent than musical training itself, making studies not directly comparable.

5.5. Personality and musical sophistication

Not surprisingly, students of audio engineering had higher levels of musical sophistication compared to the general population, with about half of the participants reporting a similar degree of formal musical training as musicians. According to random forest variable importance, Gold-MSI Singing Abilities is the measure that can best discriminate between musicians and audio engineers in our dataset. Items that contribute to this sub-dimension include questions on singing itself (e.g. "I am able to hit the right notes when I sing along with a recording"), but also melodic memory (e.g. "I only need to hear a new tune once and I can sing it back hours later" or "I can sing or play music from memory") and performance anxiety (e.g. "I don't like singing in public because I'm afraid that I would sing wrong notes"). In terms of personality, both musicians and audio engineers scored higher in openness to experience, which is associated to creative abilities (McCrae, 1987) and musical sophistication (Greenberg, Müllensiefen, Lamb, & Rentfrow, 2015) and has been shown to predict auditory and musical abilities by predicting engagement with music and musical training Corrigall, Schellenberg, and Misura (2013). Neuroticism, which is the reverse of emotional stability, was significantly higher in musicians than audio engineers. The association between musicianship and neuroticism has been observed before Gillespie and Myors (2000), although the connection between the two is not yet fully understood (for a review, see Miranda, 2020). These findings imply that there can be several covariates specific to the musician population that are not normally controlled during the recruitment process or considered in the interpretation of musicians' data. For instance, differences in musical sophistication profiles and personality could be interpreted as an effect of self-selection of creative individuals (i.e., high openness) who chose a stage-oriented career as music instrumentalists as opposed to a more studio-oriented or "behind-the-scenes" profession such as audio engineering (i.e., emotional stability and singing abilities).

5.6. Limitations and future directions

One limitation of this study was the inclusion of students of audio engineering who might still be relatively inexperienced, as they reported having between 1 and 6 years with recording, mixing, and mastering, while musicians had from 4 to over 10 years of regular practice of a musical instrument. For instance, speech-in-babble-noise thresholds showed an association with the years of audio engineering experience, although the performance of even the more experienced audio engineers in our sample was entirely within the range achieved by controls. Data from more experienced professionals could clarify whether audio engineering training can be associated with speech-in-noise perception abilities beyond the levels of the general population. Additionally, the inclusion of only one diotic speech-in-babble-noise test somewhat limits the generality of the conclusions that can be reached with our data. The inclusion of a wider range of tests in future experiments will allow us to determine whether cohort differences should be interpreted at a construct level rather than at a single task level (Green et al., 2014) and to pinpoint which auditory abilities might benefit specific aspects speech-in-noise perception. Finally, cross-sectional experiments like the one presented in this paper cannot enable conclusions to be drawn about the causality of an observed group effect. Despite this, the legitimacy of causal inference in this category of music training studies has been often erroneously assumed (Schellenberg, 2019), underestimating the complexity of the interaction between individual differences and environment (Schellenberg, 2015). For instance, the association between musical training and IQ could be explained by genetic pleiotropy (Mosing, Madison, Pedersen, & Ullén, 2016) and the undertaking and duration of music practice itself can be predicted by general cognitive ability, personality, socioeconomic status (e.g. Corrigall et al., 2013; Schellenberg, 2011; Swaminathan & Schellenberg, 2018), and genetics (Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014). Genetic variability accounts for individual difference across several musical skills (Gingras, Honing, Peretz, Trainor, & Fisher, 2015) and even in the absence of actual musical training, auditory and musical abilities are associated respectively with enhanced neural encoding of speech (Mankel & Bidelman, 2018) and emotion recognition (Correia et al., 2020). In the current study, by contrasting musicians with another population of auditory experts, we were able to draw a more nuanced and specific picture of musicians' profile in terms of auditory ability, personality, and musical sophistication. More generally, despite not being able to directly test causality, we showed that the inclusion of additional control groups and covariates in cross-sectional studies on musical expertise can help clarify the implicit assumptions about the musically trained population, challenge the specificity of the observed perceptual or cognitive advantages, and form new hypotheses about the potential source of such advantages beyond musical training itself.

CRediT authorship contribution statement

Francesco Caprini: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing. **Sijia Zhao:** Methodology, Software, Writing – review & editing. **Maria Chait:** Conceptualization, Methodology, Software, Validation, Writing – review & editing. **Trevor Agus:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Ulrich Pomper:** Methodology, Software. **Adam Tierney:** Conceptualization, Methodology, Project administration, Software, Supervision, Validation, Writing – review & editing. **Fred Dick:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

None.

Data availability

Data available upon request

Acknowledgements

This work was funded by the Department of Psychological Sciences at Birkbeck College, University of London, the British Psychological Society (Postgraduate Study Visit Scheme), and the Society for

Education, Music and Psychology Research (Arnold Bentley New Initiatives Fund). We would also like to thank Christina Makri who helped with data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105696.

References

Adcock, R. A., Dale, C., Fisher, M., Aldebot, S., Genevsky, A., Simpson, G. V., … Vinogradov, S. (2009). When top-down meets bottom-up: Auditory training enhances verbal memory in schizophrenia. *Schizophrenia Bulletin, 35*(6), 1132–1141. https://doi.org/10.1093/schbul/sbp068

Alain, C., Zendel, B. R., Hutka, S., & Bidelman, G. M. (2014). Turning down the noise: The benefit of musical training on the aging auditory brain. *Hearing Research, 308*, 162–173. https://doi.org/10.1016/j.heares.2013.06.008

Alluri, V., Toiviainen, P., Burunat, I., Kliuchko, M., Vuust, P., & Brattico, E. (2017). Connectivity patterns during music listening: Evidence for action-based processing in musicians. *Human Brain Mapping, 38*(6), 2955–2970. https://doi.org/10.1002/hbm.23565

Amrhein, V., Korner-Nievergelt, F., & Roth, T. (2017). The earth is flat ($p > 0:05$): Significance thresholds and the crisis of unreplicable research. *PeerJ, 2017*(7), 1–40. https://doi.org/10.7717/peerj.3544

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences, 113*(5), E616–E625. https://doi.org/10.1073/pnas.1508523113

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Bathke, A. C., & Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference, 138*(3), 588–610. https://doi.org/10.1016/j.jspi.2006.11.004

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*(2), 131–160. https://doi.org/10.1037/1082-989X.2.2.131

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13*(3), 108–112.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological, 57*(1), 289–300. https://doi.org/10.2307/2346101

Bettison, S. (1996). The long-term effects of auditory training on children with autism. pdf. *Journal of Autism and Developmental Disorders, 26*, 361–374.

Bianco, R., Mills, G., de Kerangal, M., Rosen, S., & Chait, M. (2021). Reward enhances online participants' engagement with a demanding auditory task. *Trends in Hearing, 25*. https://doi.org/10.1177/23312165211025941, 23312165211025940.

Bidelman, G. M., & Yoo, J. (2020). Musicians show improved speech segregation in competitive, multi-talker cocktail party scenarios. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01927

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America, 137*(1), 378–387. https://doi.org/10.1121/1.4904537

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America, 107*(2), 1065–1066. https://doi.org/10.1121/1.428288

Breiman, L. (2001). Random forests. *Random Forests, 45*(1), 5–32. https://doi.org/10.1201/9780429469275-8

British Society of Audiology. (2018). *Pure-tone air-conduction and bone- conduction threshold audiometry with and without masking*.

Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association, 92*(440), 1494–1502. https://doi.org/10.1080/01621459.1997.10473671

Brunner, E., & Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation - Brunner - 2000 - biometrical journal - Wiley online library. *Biometrical Journal, 42*, 17–25.

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health, 11*(4), 464–471. https://doi.org/10.1080/13607860601086504

Butler, B. E., & Trainor, L. J. (2015). The musician redefined: A behavioral assessment of rhythm perception in professional club DJs. *Timing & Time Perception, 3*(1–2), 116–132. https://doi.org/10.1163/22134468-03002041

Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition, 137*, 81–105. https://doi.org/10.1016/j.cognition.2014.12.005

Case, A. (2012). *Mix smart: Professional techniques for the home studio*. Focal Press.

Chartrand, J.-P., Filion-Bilodeau, S., & Belin, P. (2007). Brain response to birdsongs in bird experts. *Neuroreport, 18*(4), 335–340.

Cheema, J. R. (2014). Some general guidelines for choosing missing data handling methods in educational research. *Journal of Modern Applied Statistical Methods, 13*(2), 53–75. https://doi.org/10.22237/jmasm/1414814520

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience, 20*(2), 226–239. https://doi.org/10.1162/jocn.2008.20018

Chobert, J., François, C., Velay, J. L., & Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex, 24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience, 32*(3), 1056–1060. https://doi.org/10.1523/JNEUROSCI.3411-11.2012

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS One, 11*(7), 1–17. https://doi.org/10.1371/journal.pone.0157638

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114*(3), 494–509.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research, 352*, 49–69. https://doi.org/10.1016/j.heares.2017.02.006

Cohen, J. (1994). *The Earth Is Round ($p < .05$)*.

Corey, J. (2013). Technical ear training: Tools and practical methods learning acoustic phonetics by listening, seeing, and touching proceedings. *The Journal of the Acoustical Society of America, 19*, 25020. https://doi.org/10.1121/1.4795853

Corey, J., & Benson, D. H. (2016). Audio production and critical listening: Technical ear training. In *In Audio Production and Critical Listening* (2nd ed.). Technical Ear Training. https://doi.org/10.4324/9781315727813.

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion.*. https://doi.org/10.1037/emo0000770

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology, 4*(April), 222. https://doi.org/10.3389/fpsyg.2013.00222

Cross, I. (2012). Cognitive science and the cultural nature of music. *Topics in Cognitive Science, 4*(4), 668–677. https://doi.org/10.1111/j.1756-8765.2012.01216.x

De Kerangal, M., Vickers, D., & Chait, M. (2021). The effect of healthy aging on change detection and sensitivity to predictable structure in crowded acoustic scenes. *Hearing Research, 399*, Article 108074.

De Man, B., Boerum, M., Leonard, B., King, R., Massenburg, G., & Reiss, J. D. (2015). Perceptual evaluation of music mixing practices. In *, 1. 138th Audio Engineering Society Convention 2015* (pp. 129–136).

De Meo, R., Bourquin, N. M.-P., Knebel, J.-F., Murray, M. M., & Clarke, S. (2015). From bird to sparrow: Learning-induced modulations in fine-grained semantic discrimination. *Neuroimage, 118*, 163–173.

Delaney, H. D., & Vargha, A. (2002). Comparing several robust tests of stochastic equality with ordinally scaled variables and small to moderate sized samples. *Psychological Methods, 7*(4), 485–503. https://doi.org/10.1037/1082-989X.7.4.485

Dick, F. K., Lehet, M. I., Callaghan, M. F., Keller, T. A., Sereno, M. I., & Holt, L. L. (2017). Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. *Journal of Neuroscience, 37*(50), 12187–12201. https://doi.org/10.1523/JNEUROSCI.1436-17.2017

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience, 10*(7), 915–921. https://doi.org/10.1038/nn1925

Ellis, A. R., Burchett, W. W., Harrar, S. W., & Bathke, A. C. (2017). Nonparametric inference for multivariate data: The R package npmv. *Journal of Statistical Software, 76*(1). https://doi.org/10.18637/jss.v076.i04

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725–747. https://doi.org/10.1037/0003-066x.49.8.725

Escobar, J., Mussoi, B. S., & Silberer, A. B. (2020). The effect of musical training and working memory in adverse listening situations. *Ear and Hearing, 41*(2), 278–288. https://doi.org/10.1097/AUD.0000000000000754

Etymotic Research. (2005). *Bamford–Kowal–Bench speech-in-noise test (version 1.03) [Audio CD]*.

Fisher, M., Holland, C., Merzenich, M. M., & Vinogradov, S. (2009). Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. *American Journal of Psychiatry, 166*(7), 805–811. https://doi.org/10.1176/appi.ajp.2009.08050757

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One, 10*(9). https://doi.org/10.1371/journal.pone.0138715

François, C., Grau-Sánchez, J., Duarte, E., & Rodriguez-Fornells, A. (2015). Musical training as an alternative and effective method for neuro-education and neuro-rehabilitation. *Frontiers in Psychology, 6*(APR), 1–15. https://doi.org/10.3389/fpsyg.2015.00475

Fu, Q. J., Nogaki, G., & Galvin, J. J. (2005). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *JARO - Journal of the Association for Research in Otolaryngology, 6*(2), 180–189. https://doi.org/10.1007/s10162-005-5061-6

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience, 8*(8 JUN), 1–16. https://doi.org/10.3389/fnins.2014.00179

Gaus, W. (2015). Interpretation of statistical significance - exploratory versus confirmative testing in clinical trials, epidemiological studies, meta-analyses and toxicological screening (using Ginkgo biloba as an example). *Clinical and Experimental Pharmacology, 05*(04). https://doi.org/10.4172/2161-1459.1000182

Gillespie, W., & Myors, B. (2000). Personality of rock musicians. *Psychology of Music, 28* (2), 154–165. https://doi.org/10.1177/0305735600282004

Gingras, B., Honing, H., Peretz, I., Trainor, L. J., & Fisher, S. E. (2015). Defining the biological bases of individual differences in musicality. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 370*(1664). https://doi.org/10.1098/rstb.2014.0092

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Grassi, M., & Soranzo, A. (2009). MLP: A MATLAB toolbox for rapid and reliable auditory threshold estimation. *Behavior Research Methods, 41*(1), 20–28. https://doi.org/10.3758/BRM.41.1.20

Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging, 23*(4), 692–701. https://doi.org/10.1037/a0014345

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research, 78*(6), 756–772. https://doi.org/10.1007/s00426-013-0535-3

Greenberg, D. M., Müllensiefen, D., Lamb, M. E., & Rentfrow, P. J. (2015). Personality predicts musical sophistication. *Journal of Research in Personality, 58*, 154–158. https://doi.org/10.1016/j.jrp.2015.06.002

Güçlü, B., Sevinc, E., & Canbeyli, R. (2011). Duration discrimination by musicians and nonmusicians. *Psychological Reports, 108*(3), 675–687. https://doi.org/10.2466/11.22.27.PR0.108.3.675-687

Hallam, S. (2001). The development of metacognition in musicians: Implication for education. *British Journal of Music Education, 18*(01), 27–39. https://doi.org/10.1017/S0265051701000122

Hallam, S. (2010). Transitions and the development of expertise. *Psychology Teaching Review, 16*(2), 3–32.

Hassler, M., Birbaumer, N., & Feil, A. (1985). Musical talent and visual-spatial abilities: A longitudinal study. *Psychology of Music, 13*(2), 99–113.

Haukoos, J. S., & Lewis, R. J. (2005). Advanced statistics: Bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic Emergency Medicine, 12* (4), 360–365. https://doi.org/10.1197/j.aem.2004.11.018

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence. *PLoS One, 8*(5). https://doi.org/10.1371/journal.pone.0062836

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research, 366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Hothorn, T., & Zeileis, A. (2015). Partykit : A toolkit for recursive Partytioning. *Journal of Machine Learning Research, 16*, 3905–3909.

Howard, D. M., & Angus, J. A. S. (2009). Acoustics and psychoacoustics. In *In Acoustics and Psychoacoustics: Fifth Edition* (4th ed.).

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics, 1*(5), 799–821. https://doi.org/10.1214/aos/1176342503

Huber, P. J. (2011). Robust statistics. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1248–1251). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-04898-2$_5$94.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience, 29*(10), 3019–3025. https://doi.org/10.1523/JNEUROSCI.5118-08.2009

Iwamiya, S. I., Nakajima, Y., Ueda, K., Kawahara, K., & Takada, M. (2003). Technical listening training: Improvement of sound sensitivity for acoustic engineers and sound designers. *Acoustical Science and Technology, 24*(1), 27–31. https://doi.org/10.1250/ast.24.27

Izhaki, R. (2008). *Mixing audio: Concepts, practices, and tools*. Focal Press. https://doi.org/10.5860/choice.49-6905

Jones, Z., & Linder, F. (2015). Exploratory data analysis using random forests. In *73rd annual MPSA conference* (pp. 1–31).

Kaganovich, N., Kim, J., Herring, C., Schumaker, J., MacPherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: An ERP study. *European Journal of Neuroscience, 37*(8), 1295–1307. https://doi.org/10.1111/ejn.12110

Kaniwa, T., Kim, S., Terasawa, H., Ikeda, M., Yamada, T., & Makino, S. (2011). Towards a personalized technical ear training program: An investigation of the effect of adaptive feedback. In *Proceedings of the 8th sound and music computing conference, SMC 2011*.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America, 122*(1), 418–435. https://doi.org/10.1121/1.2743154

Kim, S., Kaniwa, T., Terasawa, H., Yamada, T., & Makino, S. (2013). Inter-subject differences in personalized technical ear training and the influence of an individually optimized training sequence. *Acoustical Science and Technology, 34*(6), 424–431. https://doi.org/10.1250/ast.34.424

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology, 12*(2), 125–144. https://doi.org/10.1515/JBCPP.2001.12.2.125

Konietschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. A. (2015). Nparcomp: An R software package for nonparametric multiple comparisons and simultaneous confidence intervals. *Journal of Statistical Software, 64*(9), 1–17. https://doi.org/10.18637/jss.v064.i09

Krishnan, S., Carey, D., Dick, F., & Pearce, M. T. (2021). Effects of statistical learning in passive and active contexts on reproduction and recognition of auditory sequences. *Journal of Experimental Psychology. General, 151*(3), 555–577. https://doi.org/10.1037/xge0001091

Laffere, A., Dick, F., & Tierney, A. T. (2020). Effects of auditory selective attention on neural phase: Individual differences and short-term training. *NeuroImage, 213*, Article 116717. https://doi.org/10.1016/j.neuroimage.2020.116717

Lavine, M. (2014). Comment on Murtaugh. *Ecology, 95*(3), 642–645. https://doi.org/10.1890/13-1112.1

Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One, 7*(12), Article e52508.

Letowski, T. (1985). Development of technical listening skills: Timbre solfeggio. *AES: Journal of the Audio Engineering Society, 33*(4), 240–244.

Li, Q., Wang, X., Wang, S., Xie, Y., Li, X., Xie, Y., & Li, S. (2018). Musical training induces functional and structural auditory-motor network plasticity in young adults. *Human Brain Mapping, 39*(5), 2098–2110. https://doi.org/10.1002/hbm.23989

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189

Lo, C. Y., Looi, V., Thompson, W. F., & McMahon, C. M. (2020). Music training for children with sensorineural hearing loss improves speech-in-noise perception. *Journal of Speech, Language, and Hearing Research : JSLHR, 63*(6), 1990–2015. https://doi.org/10.1044/2020_JSLHR-19-00391

Loo, J. H. Y., Bamiou, D. E., Campbell, N., & Luxon, L. M. (2010). Computer-based auditory training (CBAT): Benefits for children with language- and reading-related learning difficulties. *Developmental Medicine and Child Neurology, 52*(8), 708–717. https://doi.org/10.1111/j.1469-8749.2010.03654.x

MacCutcheon, D., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., & Ljung, R. (2020). Investigating the effect of one year of learning to play a musical instrument on speech-in-noise perception and phonological short-term memory in 5-to-7-year-old children. *Frontiers in Psychology, 10*(January). https://doi.org/10.3389/fpsyg.2019.02865

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Macnamara, B. N., & Maitra, M. (2019). The role of deliberate practice in expert performance: Revisiting Ericsson, Krampe & Tesch-Römer (1993). *Royal Society Open Science, 6*(8), Article 190327.

Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports, 9*(1), 1–10. https://doi.org/10.1038/s41598-019-46728-1

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports, 7*(1), 1–9. https://doi.org/10.1038/s41598-017-12937-9

Mandikal Vasuki, P. R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research, 342*, 112–123. https://doi.org/10.1016/j.heares.2016.10.008

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences of the United States of America, 115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

de Manzano, Ö., Kuckelkorn, K. L., Ström, K., & Ullén, F. (2020). Action-perception coupling and near transfer: Listening to melodies after piano practice triggers sequence-specific representations in the auditory-motor network. *Cerebral Cortex, 30* (10), 5193–5203. https://doi.org/10.1093/cercor/bhaa018

Marui, A., & Kamekawa, T. (2013). Towards the development of objective difficulty measure in technical ear training tasks. *Proceedings of Meetings on Acoustics, 19*, 1805. https://doi.org/10.1121/1.4800094

Marui, A., & Kamekawa, T. (2019, May). Does spectral flatness affect the difficulty of the peak frequency identification task in technical ear training?. In *AES 146th international convention*.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52*(6), 1258. https://doi.org/10.1037/0022-3514.52.6.1258

McKinnon-Bassett, M., & Martens, W. L. (2013). Experimental comparison of two versions of a technical ear training program: Transfer of training on tone colour identification to a dissimilarity-rating task. *Proceedings of the AES International Conference*, 54–63.

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals? *Journal of Speech, Language, and Hearing Research: JSLHR, 54*(6), 1682–1701. https://doi.org/10.1044/1092-4388(2011/09-0261)

Mestre, J. (2005). Is transfer ubiquitous or rare? New paradigms for studying transfer. *AIP Conference Proceedings, 790*, 3–6. https://doi.org/10.1063/1.2084687

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Miranda, D. (2020). The emotional bond between neuroticism and music. *Psychomusicology: Music, Mind, and Brain, 30*(2), 53–63. https://doi.org/10.1037/pmu0000250

Mishra, S., & Khare, D. (2014). On comparative performance of multiple imputation methods for moderate to large proportions of missing data in clinical trials: A simulation study. *Journal of Medical Statistics and Informatics, 2*(1), 9. https://doi.org/10.7243/2053-7662-2-9

Misra, H., Ikbal, S., Bourlard, H., & Hermansky, H. (2004). Spectral entropy based feature for robust ASR. In *, 1. 2004 IEEE international conference on acoustics, speech, and signal processing*. https://doi.org/10.1109/ICASSP.2004.1325955. I–193.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science, 25*(9), 1795–1803. https://doi.org/10.1177/0956797614541990

Mosing, M. A., Madison, G., Pedersen, N. L., & Ullén, F. (2016). Investigating cognitive transfer within the framework of music practice: Genetic pleiotropy rather than causality. *Developmental Science, 19*(3), 504–512. https://doi.org/10.1111/desc.12306

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One, 9*(2). https://doi.org/10.1371/journal.pone.0089642

Neubert, K., & Brunner, E. (2007). A studentized permutation test for the non-parametric Behrens-Fisher problem. *Computational Statistics and Data Analysis, 51*(10), 5192–5204. https://doi.org/10.1016/j.csda.2006.05.024

Noguchi, K., Abel, R. S., Marmolejo-Ramos, F., & Konietschke, F. (2020). Nonparametric multiple comparisons. *Behavior Research Methods, 52*(2), 489–502. https://doi.org/10.3758/s13428-019-01247-9

Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., & Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain and Cognition, 59*(2), 124–134. https://doi.org/10.1016/j.bandc.2005.05.009

Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife, 5*, Article e16747. https://doi.org/10.7554/eLife.16747

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., … Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology, 5*(11), Article e288. https://doi.org/10.1371/journal.pbio.0050288

Palomar-García, M.Á., Zatorre, R. J., Ventura-Campos, N., Bueichekú, E., & Ávila, C. (2017). Modulation of functional connectivity in auditory-motor networks in musicians compared with nonmusicians. *Cerebral Cortex, 27*(5), 2768–2778. https://doi.org/10.1093/cercor/bhw120

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*, 653–661.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS One, 6*(5). https://doi.org/10.1371/journal.pone.0018082

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308*, 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Pomper, U., Curetti, L. Z., & Chait, M. (2023). Neural dynamics underlying successful auditory short-term memory performance. *European Journal of Neuroscience, 58*(8), 3859–3878. https://doi.org/10.1111/ejn.16140

Porcello, T. (2004). Speaking of sound: Language and the professionalization of sound-recording engineers. *Social Studies of Science*. https://doi.org/10.1177/0306312704047328

Probst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research, 20*, 1–32.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception, 24*(1), 37–48. https://doi.org/10.1525/mp.2006.24.1.37

Román-Caballero, R., Martín-Arévalo, E., & Lupiáñez, J. (2020). Attentional networks functioning and vigilance in expert musicians and non-musicians. *Psychological Research*, 1–15.

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One, 9*(1). https://doi.org/10.1371/journal.pone.0086980

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology, 4*(1), 1–30. https://doi.org/10.2202/1544-6115.1175

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3–15. https://doi.org/10.1191/096228099671525676

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology, 102*(3), 283–302. https://doi.org/10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G. (2015). Music training and speech perception: A gene-environment interaction. *Annals of the New York Academy of Sciences, 1337*(1), 170–177. https://doi.org/10.1111/nyas.12627

Schellenberg, E. G. (2019). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000263

Schmidt-Wilcke, T., Rosengarth, K., Luerding, R., Bogdahn, U., & Greenlee, M. W. (2010). Distinct patterns of functional and structural neuroplasticity associated with learning Morse code. *NeuroImage, 51*(3), 1234–1241. https://doi.org/10.1016/j.neuroimage.2010.03.042

Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology, 2*(July), 1–9. https://doi.org/10.3389/fpsyg.2011.00167

Skerritt-Davis, B., & Elhilali, M. (2018). Detecting change in stochastic sound sequences. *PLoS Computational Biology, 14*(5), Article e1006162. https://doi.org/10.1371/journal.pcbi.1006162

Skoe, E., Camera, S., & Tufts, J. (2019). Noise exposure may diminish the musician advantage for perceiving speech in noise. *Ear and Hearing, 40*(4), 782–793. https://doi.org/10.1097/AUD.0000000000000665

Slater, J., Azem, A., Nicol, T., Swedenborg, B., & Kraus, N. (2017). Variations on the theme of musical expertise: Cognitive and sensory processing in percussionists, vocalists and non-musicians. *European Journal of Neuroscience, 45*(7), 952–963. https://doi.org/10.1111/ejn.13535

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79–87. https://doi.org/10.1007/s10339-015-0740-7

Slater, J., Kraus, N., Woodruff Carr, K., Tierney, A. T., Azem, A., & Ashley, R. (2018). Speech-in-noise perception is linked to rhythm production skills in adult percussionists and non-musicians. *Language, Cognition and Neuroscience, 33*(6), 710–717. https://doi.org/10.1080/23273798.2017.1411960

Slayton, M. A., Romero-Sosa, J. L., Shore, K., Buonomano, D. V., & Viskontas, I. V. (2020). Musical expertise generalizes to superior temporal scaling in a Morse code tapping task. *PLoS One, 15*(1), Article e0221000. https://doi.org/10.1371/journal.pone.0221000

Smith, R., Holmes-Elliott, S., Pettinato, M., & Knight, R.-A. (2014). Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization. *Quarterly Journal of Experimental Psychology, 67*(3), 590–608. https://doi.org/10.1080/17470218.2013.822009

Soranzo, A., & Grassi, M. (2014). Psychoacoustics: A comprehensive MATLAB toolbox for auditory testing. *Frontiers in Psychology, 5*(JUL). https://doi.org/10.3389/fpsyg.2014.00712

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America, 76*(6), 1690–1695. https://doi.org/10.1121/1.391605

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *The Journal of the Acoustical Society of America, 121*(5 Pt1), 2923–2935. https://doi.org/10.1121/1.2713668

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stilp, C. E., & Kluender, K. R. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences, 107*(27), 12387–12392. https://doi.org/10.1073/pnas.0913625107

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research, 261*(1–2), 22–29. https://doi.org/10.1016/j.heares.2009.12.021

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience, 12*, 94–104. https://doi.org/10.1016/j.dcn.2015.01.001

Strasser, H., & Weber, C. (1999). On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics, 8*, 220–250. https://doi.org/10.1016/S0304-0208(08)70741-2

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*, 1–11. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*. https://doi.org/10.1186/1471-2105-8-25

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*(4), 323–348. https://doi.org/10.1037/a0016973

Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports, 8*(1), 1–7. https://doi.org/10.1038/s41598-018-27571-2

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience, 11*(August). https://doi.org/10.3389/fnhum.2017.00390

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS One, 12*(10), 1–21. https://doi.org/10.1371/journal.pone.0186773

Tervaniemi, M., Janhunen, L., Kruck, S., Putkinen, V., & Huotilainen, M. (2016). Auditory profiles of classical, jazz, and rock musicians: Genre-specific sensitivity to musical sound features. *Frontiers in Psychology, 6*(JAN), 1900. https://doi.org/10.3389/fpsyg.2015.01900

The MathWorks Inc. (2013). *MATLAB version: 8.2 (R2013b)*. Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.

The MathWorks Inc. (2015). *MATLAB version: 8.6 (R2015b)*. Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.

Tierney, A., Rosen, S., & Dick, F. (2020). Speech-in-speech perception, nonverbal selective attention, and musical training. *Journal of Experimental Psychology*.

Learning, Memory, and Cognition, 46(5), 968–979. https://doi.org/10.1037/xlm0000767

Tierney, A. T., Bergeson-Dana, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review: EMR, 3*(4), 178–186.

Tierney, A. T., & Kraus, N. (2013). Music training for the development of reading skills. *In Progress in Brain Research.*. https://doi.org/10.1016/B978-0-444-63327-9.00008-4

Tierney, A. T., Krizman, J., Kraus, N., & Tallal, P. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences of the United States of America, 112*(32), 10062–10067. https://doi.org/10.1073/pnas.1505114112

Toh, A., Togneri, R., & Nordholm, S. (2005). Spectral entropy as speech features for speech recognition. In *Proceedings of PEECS*.

Torsten, H., Kurt, H., & Achim, Z. (2006). Unbiased recursive partitioning: A conditional interference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651–674.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia, 50*(7), 1432–1443. https://doi.org/10.1016/j.neuropsychologia.2012.02.028

Whitton, J. P., Hancock, K. E., Shannon, J. M., & Polley, D. B. (2017). Audiomotor perceptual training enhances speech intelligibility in background noise. *Current Biology, 27*(21), 3237–3247.

Wiggins, G. A., Müllensiefen, D., & Pearce, M. T. (2010). On the non-existence of music: Why music theory is a figment of the imagination. *Musicae Scientiae, 14*(1_suppl), 231–255. https://doi.org/10.1177/10298649100140S110

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422. https://doi.org/10.1038/nn1872

Wright, B. A., & Fitzgerald, M. B. (2005). Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes. In *Auditory signal processing* (pp. 509–515). Springer.

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: How much daily training is enough? *Experimental Brain Research, 180*(4), 727–736. https://doi.org/10.1007/s00221-007-0898-z

Wright, D. B., London, K., & Field, A. P. (2011). Using bootstrap estimation and the plug-in principle for clinical psychology data. *Journal of Experimental Psychopathology, 2* (2), 252–270. https://doi.org/10.5127/jep.013611

Yoo, J., & Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hearing Research, 377*, 189–195. https://doi.org/10.1016/j.heares.2019.03.021

Zaltz, Y., Globerson, E., & Amir, N. (2017). Auditory perceptual abilities are associated with specific auditory experience. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02080

Zatorre, R. (2005). Music, the food of neuroscience? *Nature, 434*(7031), 312–315. https://doi.org/10.1038/434312a

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews. Neuroscience, 8*(7), 547–558. https://doi.org/10.1038/nrn2152

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience, 21*(8), 1488–1498. https://doi.org/10.1162/jocn.2009.21140

Zhao, S., Brown, C. A., Holt, L. L., & Dick, F. (2022). Robust and efficient online auditory psychophysics. *Trends in Hearing, 26*. https://doi.org/10.1177/23312165221118792, 23312165221118790.

Zwicker, E., & Zwicker, U. T. (1991). Audio engineering and psychoacoustics. Matching signals to the final receiver, the human auditory system. *AES: Journal of the Audio Engineering Society, 39*, 115–126.