



HAL
open science

Binaural Headphone Monitoring to Enhance Musicians' Immersion in Performance

Valentin Bauer, Dimitri Soudoplatoff, Leonard Menon, Amandine Pras

► **To cite this version:**

Valentin Bauer, Dimitri Soudoplatoff, Leonard Menon, Amandine Pras. Binaural Headphone Monitoring to Enhance Musicians' Immersion in Performance. *Advances in Fundamental and Applied Research on Spatial Audio*, IntechOpen, 2022, 978-1-83969-006-8. 10.5772/intechopen.104845 . hal-04447847

HAL Id: hal-04447847

<https://universite-paris-saclay.hal.science/hal-04447847v1>

Submitted on 8 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,800

Open access books available

183,000

International authors and editors

195M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Binaural Headphone Monitoring to Enhance Musicians' Immersion in Performance

Valentin Bauer, Dimitri Soudoplatoff, Leonard Menon and Amandine Pras

Abstract

Musicians face challenges when using stereo headphones to perform with one another, due to a lack of audio intelligibility and the loss of their usual benchmarks. Also, high levels of click tracks in headphone mixes hinder performance subtleties and harm performers' aural health. This chapter discusses the approaches and outcomes of eight case studies in professional situations that aimed at comparing the experiences of orchestra conductors and instrumentalists while monitoring their performances through binaural versus stereo headphones. These studies assessed three solutions combining augmented and mixed reality technologies that include binaural with head tracking to conduct a large film-scoring orchestra and jazz symphonic with a click track; binaural without head tracking to improvise in trio or on previously recorded takes in the studio; and active binaural headphones to record diverse genres on a click track or soundtrack. Findings concur to show that better audio intelligibility and recreated natural-sounding acoustics through binaural rendering enhance performers' listening comfort, perception of a realistic auditory image, and musical expression and creativity by increasing their feeling of immersion. Findings also demonstrate that the reduction of source masking effects in binaural versus stereo headphone mixes enables performers to monitor less click track, and therefore protect their creative experience and aural health.

Keywords: headphone monitoring, binaural audio, music performance, creativity, studio recording, immersion, acoustic realism

1. Introduction

While musicians are performing on stage or in the studio, monitoring on headphones interferes with their instrument embodiment, the auditory feedback of their sound within room acoustics, and their interactions with other musicians. Indeed, wearable monitoring devices disturb the physical and technical ease that performers have acquired over a long, multi-sensory process to play their instruments or conduct ensembles at their best level. By covering their ears, headphones also jeopardize musicians' ability to control the parameters of their sound production. For instance, singers "suffer the most from the dislocation of sound that headphones engender [...] because the sound is produced in their bodies, resonating in the chest cavity and sinuses" [1]. As another example, the absence of direct auditory feedback

compromises “the production of high-quality trumpet tone [that] is achieved by a combination of the correct vocal tract position, the lip-reed mechanism, and the player’s breath control” [2]. Moreover, headphone monitoring obstructs collective soundscapes and established ways of listening and playing music together. To mitigate these challenges, performers sometimes remove one earcup [1] to attenuate their feeling of exclusion from the acoustic environment or to compensate for the lack of externalized sources that wearable monitoring devices as opposed to onstage speaker monitors induce [3]. In this chapter, we examine orchestra conductors’ and music improvisers’ experiences with wearable monitoring devices, and we discuss three binaural technology solutions that overcome stereo headphone monitoring challenges for a range of professional performance contexts.

Headphone monitoring was introduced in recording studios where it was necessary to isolate sound sources and synchronize performances on cue tracks while enabling musicians to hear themselves and others. Whereas this technology offers flexibility and creative possibilities such as overdubbing on previously recorded takes, it calls for the use of visual cues through windows and red lights, and for the setup of talk-forward and talkback microphones that may expose musicians to the others’ comments on their performances. In such a technological environment for music creation, sound engineers control both the quality of headphone mixes and the communication system in the studio. Williams highlighted how the setup of the communication system increases stress and may result in tensions between musicians and engineers during recording sessions [1]. Also, adding headphone monitoring as yet another layer of engineers’ sound control may worsen experiences of gendering and microaggressions in the commercial recording studio [4]. Therefore, although “the number of available headphone mixes becomes a status marker reflecting the professional standing of the studio among competing facilities” [1], using a high number of headphone mixes may negatively impact the production workflow and the social climate of the workplace. Our approach consists of adapting technologies to specific performance contexts to enhance musicians’ immersion in their artistic tasks, and thus reduce stress and other adverse sociopsychological effects of headphone monitoring.

The audio content of monitoring systems influences all aspects of musicians’ performances, in positive and negative ways. For instance, balancing harmonic versus rhythmic sections in a singer’s or a melodic instrumentalist’s monitoring mix impacts their comfort in finding their best tuning, rhythmic placement, and dynamics. Furthermore, signal processing like equalization, dynamic range compression, delays, and reverberation is commonly used to facilitate ensemble cohesion. As an example, boosting the attack of the kick drum in a bassist’s monitoring mix can enhance the groove of a band. Also, a study showed that monitoring different reverberation lengths of room acoustics affects orchestra conductors’ tempo, timbre, and appreciation of the performance quality when listening to recorded takes [5]. Findings from a PhD thesis about live engineering on Broadway underline how engineers are responsible for “sonic colors” that represent “the unique resonant characteristics of sound sources associated with music-making, but also to invoke “color” as a broader metaphor for social difference and identity” [6]. From this perspective, both the sound capture system and mixing approach of monitoring systems must meet the cultural expectations and genre conventions of specific performance contexts. For each of our three binaural solutions, we detail how we designed the monitoring technology, the sound capture system, and the mixing approach to satisfy the requirements of specific performance contexts.

Our interdisciplinary team of four researchers who are also experienced sound engineers and music performers aim at examining the following research questions:

1. What are the main challenges of using a wearable device for monitoring while performing music? And to what extent do these challenges differ between conducting large ensembles versus improvising?
2. Could binaural headphone monitoring technologies that are adapted to specific performance contexts enhance musicians' listening comfort, perception of a realistic auditory scene, musical expression and creativity?
3. Could binaural headphone monitoring systems decrease the click-to-music ratio compared to stereo headphones?

Before we present a fresh perspective on the methods and results of a series of three studies that were published in the proceedings of *Audio Engineering Society Conventions* [7–9], we highlight previous research on delivering synchronization auditory cues to performers; augmented and mixed reality audio applications; and binaural music production that informed our solution designs. Then, we discuss the methods and outcomes of two online surveys about orchestra conductors' and improvisers' experiences when monitoring through headphones. The survey findings serve as a basis to support the design of eight case studies that aimed to compare binaural versus stereo headphones in recording or rehearsal situations.

Because musicians rely on the auditory cues that their monitoring systems convey to elaborate their performance process, comparing the influence of binaural versus stereo monitoring on musicians' performances requires researchers to design "ecologically valid" experimental protocols and technologies that address creative cognition [10, 11]. Hence, we carried out our eight case studies in real-life performance situations.

With experienced musicians, to test three binaural monitoring solutions that we designed to meet the esthetic and cultural context of three distinct performance situations. Finally, we provide ideas for future research with audio augmented and mixed reality applications to facilitate musicians' immersion in the performance.

2. Literature review

2.1 Delivering synchronization auditory cues to music performers

The use of a click track in music performance was first documented for the soundtrack recording of *Fantasia* (Disney, 1940). Maestro Leopold Stokowski, who was an audio engineer at Bell Labs, experimented with new recording workflows to synchronize different sections of orchestra and principals on a multitrack device [12]. While the need for a click track was justified by such a creative innovation, its extensive use in music performance comes with downsides. Like sirens or fire alarms, click sounds are designed to grab attention with a lot of high-frequency energy. Therefore, long exposures to high levels of click tracks contribute to the risks of musicians' hearing loss [13]. Although click samples can be changed in digital audio workstations to accommodate musicians' preferences, the mechanical nature of the click implies that "overall, playing with a click track means playing with the metronome" [14]. According to Cardassi [15], "a click track is likely the most dreaded synchronization tool in music," and it generates performers' "angst and unpleasantness." Drawing upon Blauert [16]'s theory, spatial audio applications offer greater source discrimination possibilities than a stereo image. Therefore, binaural technologies provide sound engineers with more mixing space than stereo, which implies more source-positioning options and the need for less equalization and dynamic range compression [17] to avoid masking effects among sound

sources. Consequently, we suggest that binaural headphone monitoring solutions allow for lower click track levels and less processed instrumental and vocal sources for performers to synchronize with each other, on a soundtrack or a movie, meanwhile protecting their aural health and improving their creative experience.

Previous research suggests that a generalized use of click tracks has homogenized creativity and globalized music cultures. For instance, an analysis of tempo across the past 60 years of U.S. Billboard Hot 100 #1 Songs revealed that a 5-beat average standard deviation from 1955 to 1959 decreased to 1-beat between 2010 and 2014 [18]. Moreover, Éliézer Oubda, a music producer and sound engineer who owns *Hope Muziks Studio* in Ouagadougou, Burkina Faso, trains his assistants in explaining to Western African musicians how to perceive the downbeat in the click track in the same way Europeans and North Americans do.^{1,2} To minimize “the straight-jacket feeling” [14] induced by click tracks, composers, performers, and studio professionals can collaborate on developing alternative cue tracks and monitoring systems. For instance, customized tracks may combine pre-recorded fragments from the parts to be performed with vocal instructions or relevant pitches. These may also feature excerpts of embedded click tracks within the pre-existing layers of audio to provide additional guidance at specific times only. These cue improvements can be accompanied by a context-dependent choice of the monitoring system. While high-fidelity technologies may not always be the best solution³, the selection of a wearable device requires some attention. Typical studio headphones consist of closed-back headphones that are “designed to block out environmental noise using a passive acoustic seal” [19]. Mostly found in live scenarios, in-ear monitors provide a more drastic acoustic isolation, with visual discretion and stability benefits in situations where the performer frequently moves their head. With non-isolated ear cups, open-back headphones offer a more natural or “speaker-like sound” [19] with a more pleasant spatial image, and less risk of performers feeling isolated and disconnected from the environment. Whereas we did not consider using open-back headphones for our monitoring applications because their audio content would leak into the microphones, their benefits have inspired our binaural solutions to overcome the auditory feedback challenges of stereo closed-back headphones and in-ear monitors.

Two types of technologies exist to deliver synchronization auditory cues to musicians while providing them with direct access to their own sound production and acoustic environment, namely acoustic-hear-through and microphone-hear-through monitoring systems [20]. Primarily developed to improve the safety of outdoor runners when they are listening to music, acoustic-hear-through monitoring systems, also known as bone conduction headphones leave the users’ ear canal free by conveying the auditory cues “from the vibration of the bones of the skull [or jaw] that is transmitted to the inner ear” [21]. Whereas this technology eliminates

¹ See 15:00-20:10 of the roundtable discussion about “De-colonizing the Digital Audio Workstation” with Éliézer Oudba and Eliot Bates facilitated by Menon and organized by Pras and Kirk McNally: <https://www.canal-u.tv/chaines/afrinum/roundtable-discussion-about-de-colonizing-the-digital-audio-workstation>

² During a jembe workshop taught by Issa Traoré alias Ken Lagaré, an arranger and sound engineer who owned *Authentik Studio* in Bamako, Mali, graduate student Leo Brooks and percussion instructor Adam Mason explained the fact that European and North American musicians struggle to hear the downbeat in Western African music (see 45:50–47:10): <https://www.canal-u.tv/chaines/afrinum/percussion-workshop-with-ken-lagare>

³ For example, Oubda gave the example of rural musicians from Burkina Faso who got intimidated when they heard themselves through high-fidelity headphones for the first time (see 17:00–17:40): <https://www.canal-u.tv/chaines/afrinum/roundtable-discussion-about-de-colonizing-the-digital-audio-workstation>

disconnection feelings from the acoustic surroundings, like open-back headphones, the monitoring mix may leak into the microphones. Indeed, May and Walker [22] reported “approximately 12 dB A (total) of ‘leakage’” in the context of listening tests. Also, Cardassi, who tested a bone conduction headphone to record an electro-acoustic album on piano and vocals, could only use it for pieces that did not require the use of a close vocal microphone, and whose cue tracks did not include any click.⁴ Primarily used as hearing aids devices, microphone-hear-through monitoring systems consist of mounted microphones on the users’ headset that capture what they would hear without headphones [23]. Cooper and Martin [2] designed a microphone-hear-through monitoring system named Acoustically Transparent System (ATH) that combines the binaural rendering of the signal captured from two headset-mounted microphones with the synchronization cues. In performance situations, they observed that the ATH has “a notable impact on both quality of tone production and the confidence of the [trumpetist]” [2]. Their findings confirm the relevance of designing binaural technologies to improve musicians’ experience while performing with headphone monitoring.

2.2 Mixed and augmented reality applications with binaural technology

Audio Mixed Reality (AMR) applications aim at recreating new auditory spaces for listeners by balancing the proportion of real and virtual elements. Also, Audio Augmented Reality (AAR) applications aim at achieving listeners’ experiences of acoustic transparency, as if there was no headset, to interleave virtual sounds with an unaltered reality [24]. Drawing upon Milgram and Kishino [25]’s “virtuality continuum” of visual displays, McGill et al. [26] define AAR as “auditory headset experiences intended to [...] exploit spatial congruence with real-world elements.” From this perspective, AAR sits at the edge of AMR that encapsulates “any auditory VR and AR experiences.” These definitions mirror the recording esthetics continuum from “attempting realism” to “creating virtual worlds” produced through different sound capture systems and mixing approaches [27]. While mixing for stereo recordings differs from mixing for AMR and AAR applications, we applied our knowledge of sound capture systems to best meet the cultural expectations and genre conventions of the performance contexts. Specifically, we primarily used microphone arrays that captured the acoustic environment for our five AAR case studies, versus close mono microphones that focused on the instruments’ direct sound for our three AMR case studies.

To enhance listeners’ perception of auditory spaciousness through headphone monitoring, König [28] conceptualized one of the first four-channel headphones that positioned an additional speaker driver near the tragus to diffuse reverberation, and thus allow for a more accurate spatial image with less sound pressure level on the ear axis. Further developments intending to simulate surround and multi-channel loudspeaker systems have led to the design of multi-driver headphones that position multiple speaker drivers within the ear cup, employing the shape of the listener’s ear and pinna to influence the filtering of high frequencies as they enter in the ear canal [19]. Meanwhile, most of today’s AAR and AMR headphone applications use binaural filtering with head-related transfer functions (HRTF) that enable listeners to externalize sound sources while wearing regular headphones. Theoretically, delivering accurate intelligibility, localization, and externalization

⁴ Cardassi first tested a bone-conduction headphone in Fall 2017 for the recording of *Ramos* (Redshit, 2019) with Pras as music producer and sound engineer. While she could only use it for the recording of a few pieces in Rolston Hall at the Banff Centre, she enjoyed preparing for the sessions with it at home. This was confirmed through personal email communication on March 9, 2021.

of sound sources through headphones requires the binaural rendering of sound sources via individualized HRTFs transmitted through high fidelity open-back headphones [29]. Nevertheless, according to a review of sound externalization studies, adding reverberation-related cues, and/or dynamic binaural rendering that matches listeners' self-initiated head movements, facilitates the localization and externalization of binaural cues [30], which may compensate for the use of non-individualized HRTFs and closed-back headphones. Whereas dynamic binaural ensures the success of AAR applications for users who move a lot in the real-world environment, such as orchestra conductors, we suggest that static binaural may be more relevant for AMR applications where most of the binaural cues are out of sight, such as recording sessions with musicians performing in separate rooms. In this view, static binaural might still provide users with a better source intelligibility and a more spatial experience compared to stereo systems since there is less masking effect among sources, even though the localization accuracy and externalization of binaural cues remain compromised, for example, generating front-back confusions. In fact, a study showed that "short training periods involving active learning and feedback" facilitate listeners' ability to externalize sources while using binaural systems with non-individualized HRTFs [31]. In this chapter, we present the concept of two distinct dynamic binaural AAR setups and one static binaural AMR setup that involved a short training tutorial for listeners.

Besides the popularity of noise cancellation headphones that filter the real acoustic environment out for listeners to focus on music or other virtual elements [26], AAR and AMR microphone-hear-through devices are primarily developed for single users' experiences in non-musical applications, for example, for audio gaming [32]; street navigation [33]; and soundwalks that immerse listeners in sonic art compositions [34]. Only a few collaborative AAR experiences have been tested [35], for example, a four-player interactive audio experience [36]; a two-player audio game called *eidola multiplayer* [37]; and creative artworks dedicated to multi-users, such as *Listen* for museum visits [38] or *SoundDelta* devoted to large public outdoor events [39]. Also, to our knowledge, very few AAR musical applications besides Copper and Martin's ATH [2] have been designed. For instance, a Master thesis showed that members of a rock band preferred performing with AAR dynamic setups compared to mono and stereo headphones [40]; a study with methodological shortcomings tested AAR dynamic in-ear monitors for members of an acoustic ensemble [41]; and the *Architexture Series* brought new music composers, sound engineers, and architects to collaborate on site reconstruction [42, 43]. Our eight music performance case studies, therefore, contribute to AAR and AMR research by assessing two AAR setups that aim at overcoming performers' social interaction challenges when wearing headphones, and one AMR setup that aims at enhancing social interactions among performers when being remotely located.

2.3 Binaural music production

Sound engineers increasingly use binaural technology in the recording studio in parallel with the development of new plugins and devices that enable listeners' sound externalization on headphones with and without the tracking of their head movements, for example, binaural simulation of surround sound mixes in control rooms that do not have a 5.1 speaker system [44]. Although binaural audio is optimized for headphone listening which is the primary music listening mode of our time, so far only few binaural music productions have been released on the market. For instance, Williams and Reiser walked us through the binaural capture and rendering processes of sources for the production of "*GoGo Penguin [untitled]*"

(Blue Note Records 2020), which was released in stereo and not yet in binaural.⁵ They used three *Neuman KM 100* dummy heads to *overdrive* space in the main live room and the drum room, and to immerse listeners within the piano sound. At the mixing stage, they also used *dear VR* plugins to externalize specific sources. They underlined that binaural production techniques are the best fit to convey virtuosic performances of high-level musicians in contemporary jazz and classical music because the recording of their performances requires little signal processing in terms of equalization and dynamic range compression. Indeed, extensive signal processing does not work well with binaural rendering, and equalization and compression should only be used for creative purposes since there are less source masking effects than in stereo [17]. We thus assessed our three binaural solutions in professional-level performance contexts whose esthetics did not require much signal processing, with five out of the eight case studies primarily involving classical and jazz musicians.

Whereas binaural has not yet succeeded commercially as a release format, more and more public European radios offer binaural programs, for example, *Hyperradio* on Radio France, which primarily broadcasts audio plays and electronic music live shows. To broadcast classical orchestral recordings for *BBC Proms* on BBC Radio 3, Parnell and Pike [45] reported on using IRCAM's *Panoramix* to enhance the positioning and ambiance of the auditory scene captured with a Schoeps ORTF-3D microphone array that features two coincident layers of four microphones. Results from their audience study showed that binaural mixes were rated as "more enjoyable" by 79% of respondents, whilst 75% said that the experience was "somewhat" or "absolutely" like being there in person. These findings contrasted with previous research that found that overall, the stereo listening experience was preferred to binaural for a range of musical genres [46]. Also, the outcomes of a study about binaural mixing for hip-hop production suggest that listeners can be disoriented by this unfamiliar immersive format [47]. In particular, the main sources of the beat seem more effective when not externalized. We used this knowledge to capture and mix sound sources in the performers' binaural headphones for our eight case studies.

3. Online survey on music performers' experiences with headphone monitoring

3.1 Online survey methods

A combination of two online surveys further examined the challenges that music performers face when wearing monitoring devices in the studio or on stage [7, 8].

3.1.1 Respondent demographics

We recruited 12 orchestra conductors and 12 music improvisers from our respective networks by email to fill out a survey on an unpaid, volunteer basis. These 24 professional respondents included 20 males and four females living in seven countries (Australia, Canada, France, Germany, Netherlands, Switzerland, the UK, and the USA). They had at least 5 years on the job, except for one who reported having between one and 5 years on the job. More than half (15 out of 24)

⁵ <https://mupact.com/seminar-program-may-jul-2020/aesthetic-manifestos-and-binaural-integration-an-investigation-of-pre-in-session-and-post-production-techniques-employed-in-gogo-penguins-self-titled-2020-album-release/>

were touring internationally; the other nine were primarily working in France. All 24 respondents had experienced headphone monitoring while performing. Half of the conductors primarily performed for studio recording sessions with acoustically isolated instruments and/or the need to overdub on previous recordings; five for live concerts of film-scoring or new music compositions with electronic components; and one for both kinds of performance situations. Nine of the improviser respondents reported wearing headphones for more than half of their studio recording sessions; and three of them for 30% or less. Improvisers played a variety of instruments and included a singer; an acoustic bassist; a trombonist; a hornist; two saxophonists; one flutist and electronic artist; one multi-instrumentalist who played sousaphone, saxophone, clarinet, and flute; two drummers (one also conducted ensembles); and two pianists (one also played electronic keyboards and produced recordings, the other one also sang and played prepared piano). About musical genres, improvisers primarily performed jazz and/or world music (53%); pop-rock subgenres including French *variété* (27%); experimental, improvised, or contemporary music (9%).

3.1.2 Questionnaire

Both surveys used similar semi-directed questionnaires because Bauer et al. [8] adapted Soudoplatoff and Pras [7]’s methods from the context of orchestra conducting to the context of music improvisation. In this chapter, we focus on the analysis of the respondents’ answers to four questions that were featured in both questionnaires. These questions are slightly reworded here to encompass both performance contexts (i.e., orchestra conducting and recording improvisations):

1. According to your previous studio recording session experiences, how would you describe an ideal headphone monitoring system?
2. Think about one of your best studio recording sessions with headphone monitoring. Start by describing the context of this session (ensemble, production, location, etc.). Why do you think this session was a success?
3. Think about one of your worst studio recording sessions with headphone monitoring. Start by describing the context of this session (ensemble, production, location, etc.). Why do you think this session went this way?
4. When recording in the studio, do you have a particular way of wearing headphones? If so, why?

3.1.3 Qualitative data analysis

Respondents’ verbal descriptions were analyzed using a *Grounded Theory* approach [48] drawing from previous research on studio practices (e.g., [49]). This approach consists of extracting meaningful phrasings from the free-format verbal descriptions to be classified into concepts and categories without preconceived themes. Specifically, the constant comparison technique of Grounded Theory called for a minimum of two researchers to review each other’s classification and to draw parallels between findings from the different questions, to gradually refine the emerging concepts and categories as well as to identify consensus and contradictions among outcomes of different questions. Results will be presented with the count of phrasing occurrences for each concept and category.

3.2 Online survey analysis

3.2.1 Ideal headphone monitoring system

We identified 50 phrasings from the respondents' free-format verbal descriptions of their ideal headphone monitoring system. These phrasings were classified into three major categories, namely Sound Quality (n = 24), System Technical Quality (10), and Physical Properties (10); and into three minor categories, namely Click (3), Ambiance in the Studio (2), and Forgetting the Headphones (1). The most-reported concepts for each major category were Realism (8), Instrument balance (8), and Control over monitoring (6). **Figure 1** displays the classification into emerging concepts and categories of the 31 phrasings coming from improvisers and the 19 phrasings coming from conductors separately, since a Yates' chi-squared test revealed a significant difference between the answers' distribution into the six categories for conductors and improvisers ($\chi^2(5,50) = 2,97, p < 0.05$).

3.2.2 Positive and negative experiences when performing with headphones

In total, we collected 129 phrasings, 70 from improvisers and 59 from conductors about their positive and negative experiences when performing with headphones. A Yates's chi-squared test revealed no significant difference between the answers' distribution into the nine categories for conductors and improvisers ($\chi^2(8,129) = 15,91, p > 0.05$). Nevertheless, we chose to keep both populations of performers distinct in **Figure 2**, to stay consistent with the

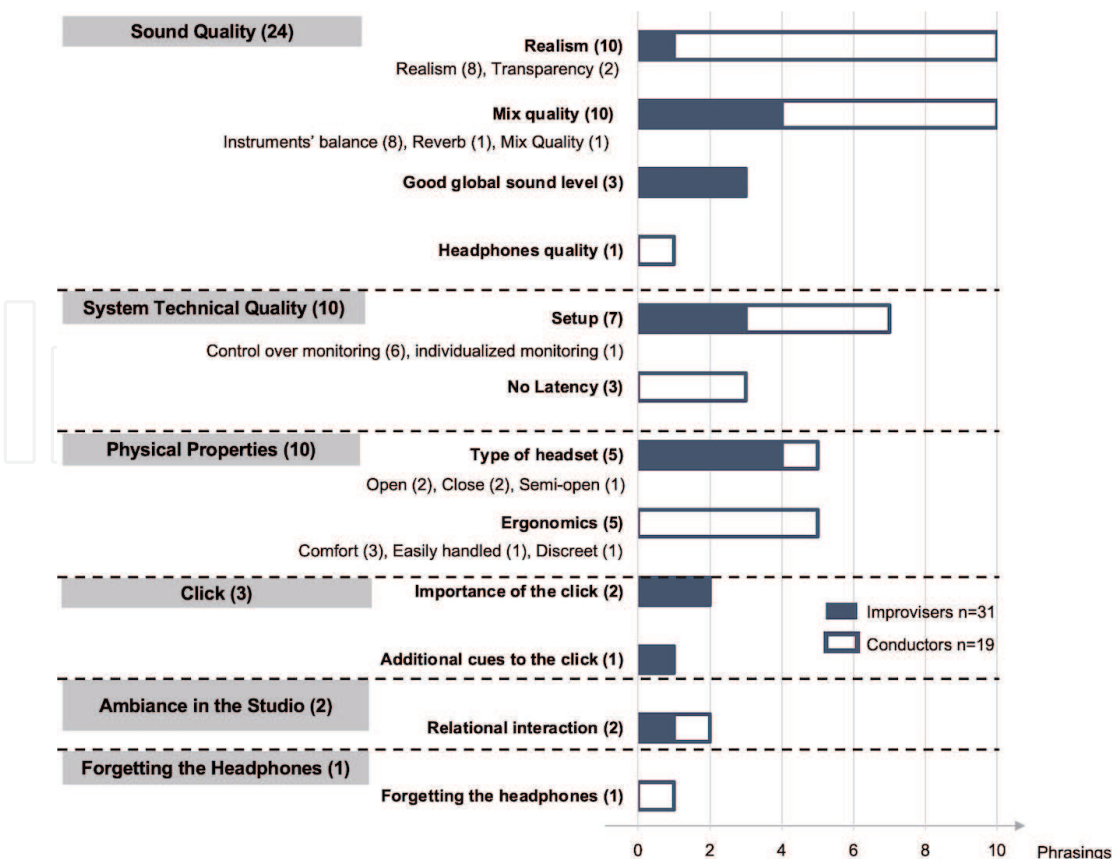


Figure 1. Classification of phrasings extracted from the 12 improvisers' and 12 conductors' free-format verbal descriptions accounting for their ideal headphone monitoring system.

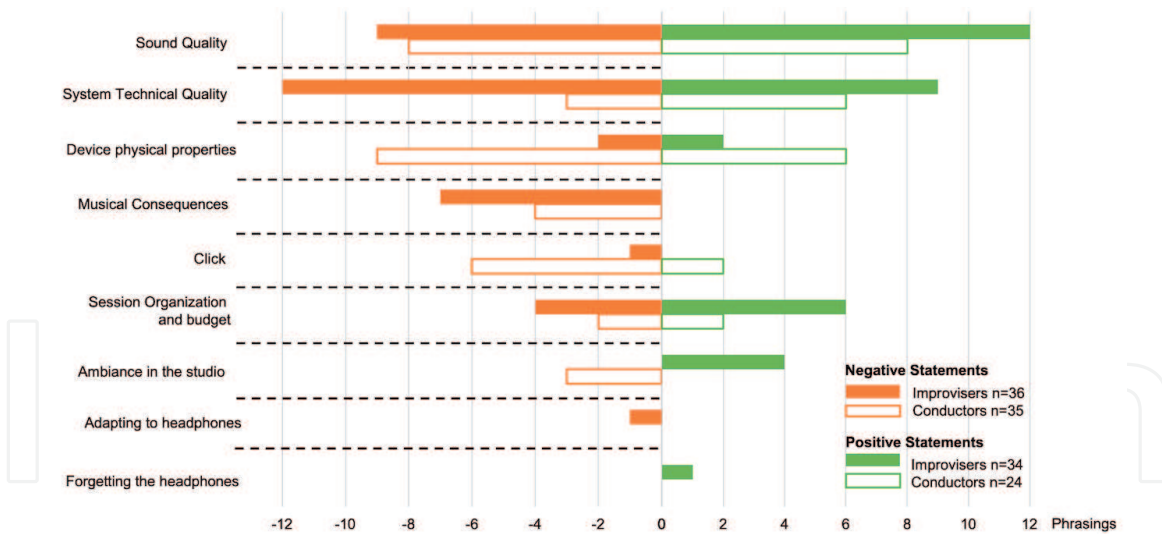


Figure 2. Classification of phrasings extracted from the 12 improvisers' and 12 conductors' free-format verbal descriptions accounting for their positive (green) versus negative (orange) monitoring experiences.

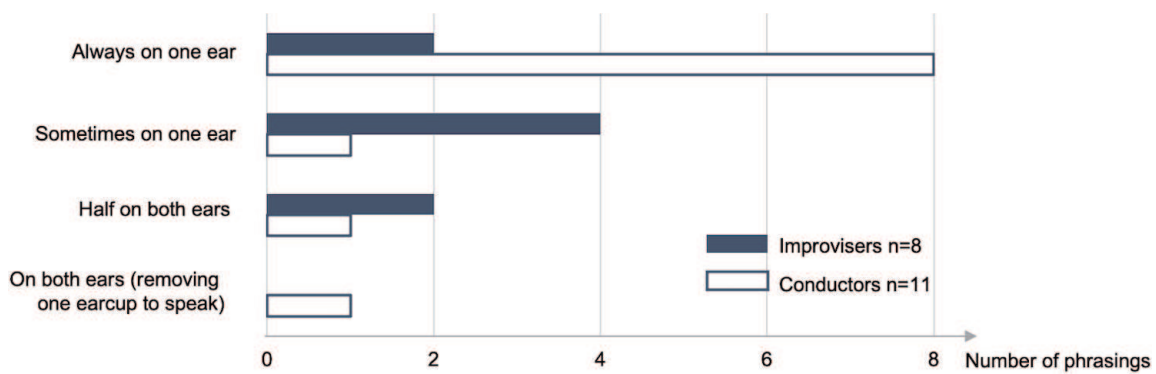


Figure 3. Classification of phrasings extracted from the 12 improvisers' and 12 conductors' free-format verbal descriptions accounting for their usual ways of wearing headphone monitoring devices.

other figures in this section. Regarding positive experiences, 58 phrasings were identified, 34 from improvisers and 24 from conductors, and classified into the major category Sound Quality (20), followed by System Technical Quality (15). Regarding negative experiences, 71 phrasings were identified, 36 from improvisers and 35 from conductors, and classified into three major categories, namely Sound Quality (17), System technical Quality (15), and (negative) Musical consequences (11).

3.2.3 Ways of wearing headphones

We collected 19 phrasings, eight from improvisers and 11 from conductors about their ways of wearing headphones. A Yates's chi-squared test revealed a significant difference between the answers' distribution into the four different habits of wearing headphones for improvisers and conductors ($\chi^2(3,19) = 6,42, p < 0.05$). Hence, **Figure 3** presents the classification of phrasings for the improvisers and conductors separately. The main habit that we identified consisted in always (Improvisers: 2; Conductors: 8) or sometimes (I: 4; C: 1) wearing the device on one ear only.

4. Assessment of three binaural headphone monitoring technologies in a performance situation

4.1 Technology design and experimental protocols

We designed three binaural headphone monitoring solutions to enhance musicians' cognitive engagement in performance (**Table 1**). For each solution, we adapted the augmentation type, sound capture system, and mixing approach to the esthetic and cultural context of distinct performance situations (**Table 2**). Then, we conducted eight case studies that involved two renowned conductors of symphonic ensembles with a click track in large acoustics [7]; seven emerging music improvisers in solo or trios in separated dry rooms or overdubbing alone in small acoustics [8]; and three music students and one touring musician who recorded for a range of musical genres alone or in a duo with a click track or a soundtrack in medium-size acoustics [9]. These eight case studies were all carried out in real-life performance situations at the Paris Conservatoire (CNSMDP), Radio France, and the University of Lethbridge (ULeth).

Our mixed methods of assessing these solutions draw upon Agrawal et al. [50]'s definition of *immersion* as a psychological state that enables an individual's mental absorption in the world and in the tasks that are presented to them. Therefore, for each performance case study, we determined which auditory information would be the most important for the users to monitor in order to perform at their best, in other words, which auditory information would be "immersive enough" [51] to achieve a sense of "being there together" [52].

4.1.1 Description of three binaural headphone monitoring solutions

Table 1 highlights the binaural rendering pipelines and augmentation technologies that we chose to best adapt to performers' needs for each context. To enable conductors to monitor large ensembles on headphones, Soudoplatoff and Pras [7] designed a Binaural with Head Tracking (BHT) system that rendered a *JML* tree [53], that is, a main five-microphone array with specific dimensions, and integrated spot microphones. This system used *Bipan*⁶ software [54] coupled with *Hedrot*,⁷ that is, a head tracker located on the conductors' headphones. In *Bipan*, the LISTEN database [55] was used with the HRTF pair n°1040, as advised in a previous study [56], since this HRTF pair satisfied most users during public demonstrations of the software [57]. *Bipan* had a latency of 5.3 ms when used with a buffer size of 256 samples. According to previous research, a monitoring system latency below 42 ms should be acceptable [58]. Furthermore, *Hedrot* had a latency of 48.1 ± 4.3 ms [54], which should provide conductors with accurate localization cues since the head tracking latency does not hinder the stability of virtual sounds within complex auditory scenes under 71 ms [59], even if it could be noticeable when superior to 30 ms [60]. The assessment tests required the use of the *TotalMix* application, which has a meaningless latency of three samples (equal to about 68 μ s at 44.1 kHz), to digitally convert the microphone signal and send it to the BHT via a *RME MADiface*.⁸

⁶ 3D audio technology developed in-house at the Paris Conservatoire in collaboration with IRCAM as part of the Bili project: <http://www.bili-project.org/>. More details can be found here: <https://alexisbas-kind.net/fr/bipan-binaural/>

⁷ <https://abaskind.github.io/hedrot/>

⁸ <https://www.manualslib.com/manual/1310692/Rme-Audio-Madiface-Usb.html?page=70>

	Performance context		Binaural rendering pipeline		Augmentation related to the context		
	Ensembles [# users]	Sync. cues	Microphone system	Binaural Rendering	Acoustic venues	Type of immersion	Static vs. dynamic
BHT	2 Conductors of large Ensembles [2]	Click & Rhythmic section	5-microphone array + spot mics	Bipan with LISTEN HRTF pair 1040 + Hedrot head tracker with latency of 48.1 ± 5.3ms	Large acoustics	AAR	Dynamic
BMR	1 Solo & 2 Trios [7]	Overdubbing & Rhythmic section	Close mics	KF with proprietary anechoic HRTFs	Isolated dry rooms	AMR	Static
ABH	2 Soli & 1 Duo [4]	Click or Soundtrack	2x2-mounted microphones	KV with proprietary anechoic HRTFs	Medium acoustics	AAR	Dynamic

Table 1.

Performance context, binaural rendering pipeline, and augmentation principles of the three binaural headphone monitoring technologies.

To enhance the intelligibility of improvisers' subtle expressive gestures, Bauer et al. [8] developed a Binaural Mixed Reality (BMR) system that rendered close mono microphones through KLANG: fabrik (KF) hardware. KF was chosen for its convincing externalization of sources and sound quality⁹ as well as its latency of less than 3 ms.¹⁰ Indeed, the set of KLANG-proprietary HRTFs was preferred to HRTFs from the LISTEN database that features a low sampling resolution, introduces noise artifacts, and present amplitude errors, for example, for the HRTF pair of subject IRC_1034 [61]. The BMR had a total latency (KF latency plus ProTools latency) of 4 ms for the two trios. Regarding the world music performer, the technical setup between the microphone signal and the monitoring system included several digital devices, and the measured total latency of the chain was 14.1 ms. The musician specified that he did not notice it, and confirmed that the system latency did not hinder his performance.

To attempt acoustic transparency of the recording auditory space, Menon [9] built an Active Binaural Headphones (ABH) system with two 150°-angled small condenser microphones mounted on each earcup. Based on Bauer et al. [8]'s satisfying findings, the signal coming from the four mounted microphones was binaurally rendered through KLANG: vier (KV) hardware, which features the same sonic and latency properties than KF¹¹. The ABH total latency was inferior to 16.8 ms. The assessment tests required the use of the *CueMix* application that has no latency to digitally convert the microphone signal and send it to the ABH via a *MOTU 896*

⁹ The researchers were able to evaluate the quality of this equipment as they are experienced sound engineers, in both stereo and 3D audio production techniques.

¹⁰ https://www.klang.com/en/products/klang_fabrik

¹¹ https://www.klang.com/en/products/klang_vier

*mk3*¹² that has a latency of under 13 ms, and the *Aviom* personal monitor mixer that has a latency of 0.88 ms to amplify the headphone signal.¹³

In summary, the BHT and ABH are two AAR systems with dynamic binaural because for both of their applied contexts, performers primarily needed to monitor sound sources while being in the same room as their peers, and thus required a technology that accurately conveyed source localization. On the other hand, the BMR is an AMR system with static binaural because improvisers primarily needed to monitor their previous recordings or their band members who were playing in separate rooms; thus, the re-creation of a virtual space that facilitated their immersion was more desirable than accurate source localization.

For all three technologies, closed-back headphones were used to minimize sound leakage into the microphones. Both Bipan and KLANG used anechoic HRTFs, and so enabled us as sound engineers to generate spatial images with re-created acoustics that fit the acoustics of the performance space.¹⁴ These HRTFs were also non-individualized and thus required performers' listening training [31] and/or dynamic binaural rendering [29] to optimize source externalization and mitigate timbre artifacts. Therefore, one week before conducting the case studies that assessed the BMR setup, which is static, the improviser participants were instructed to listen to three-to-five binaural audio productions over headphones that were selected from *Hyperradio* podcasts by Bauer (total duration of around 25 mn), to get used to the binaural rendering. All of them confirmed to Bauer at the beginning of their recording session that they had listened to at least three of these productions. This consists of a total listening experience of 15 mn at minimum for each participant.

4.1.2 Case study procedures for binaural solution assessment

Table 2 details the locations, genres, and instrument line-ups of the eight case studies in chronological order for testing our BHT, BMR, and ABH technologies in rehearsal or studio recording situations. Thirteen performers agreed to participate in these comparative tests without financial compensation. The first two tests that involved symphonic ensembles were organized at the institutional level as part of a pedagogical project. For the other five tests, Bauer and Menon volunteered to mix the recordings, which the performers could use to promote their music.

To assess the three headphone technologies that are described in the previous section, two conductors, seven improvisers, and four musicians who perform a range of musical genres compared binaural against *traditional* stereo headphones, that is, the monitoring systems commonly used in each of the performance venues. The experimental procedures for each case study are summarized in **Table 2**. Because “an experimental protocol is ecologically valid if the participants react [...] as if they were in a natural situation” [10], Soudoplatoff organized the first two case studies during rehearsals of programmed productions with large ensembles. Specifically, for the last two days of a week of film-scoring rehearsals, Maestro Laurent Petitgirard agreed to swap headphone conditions five times during breaks that occurred every 90 min, which led him to test each condition three times. Unfortunately, the comparison could not be carried out with the jazz symphonic ensemble due to a conjunction of acoustic and organization issues (see Section 4.2

¹² <https://motu.com/techsupport/technotes/what-is-the-latency-of-my-motu-audio-interface>

¹³ https://www.aviom.com/library/User-Guides/36_A-16D-User-Guide.pdf

¹⁴ Using non-anechoic HRTFs implies generating a binaural image that emulates the externalization of sources in specific room acoustics. This may be enjoyable for the listener and can be creative in the context of music production. However, it is likely to be confusing for the musician in the context of headphone monitoring when performing.

	Institution	Venue	Audio latency	Musical genres	Instruments	Performance purpose	Comparison procedure [# cases; duration]
BHT	CNSMDP	Art lyrique	5.3 ms	Film-scoring	Symphonic orchestra	Rehearsal	B S B S B S [6; 90 mn]
		GPO	5.3 ms	Symphonic jazz	Symphonic orchestra with a jazz big band and non-acoustically amplified instruments	Studio recording	<i>Comparison not possible</i>
BMR	Radio France	Studio 115	14.1 ms	World music	Voice, bass, various percussions, small guitar	Studio overdubbing	B S [2; 2 h]
	CNSMDP	240/244/245	4 ms	Jazz trio	Double bass, drums, electric guitar	Studio recording	B S [2; 45 mn]
			4 ms	Free improvisation trio	Drums, clarinet/ bass clarinet, accordion	Studio recording	S B [2; 45 mn]
ABH	ULeth	Studio one	3.8 ms	Singer-songwriter	Voice & banjo	Studio recording	S B [2; When musicians were pleased with the stereo takes]
			3.8 ms	Pop-rock	Drums & electric guitar	Studio recording	S B [2; When musicians were pleased with the stereo takes]
			3.8 ms	Electroacoustic	Piano and acoustically amplified soundtrack	Studio recording	S B [2; When musicians were pleased with the stereo takes]

Table 2. Location, genre, instrumentation, and comparison procedure of the eight case studies—B refers to the binaural condition and S to the stereo condition.

for explanations). Bauer and Menon ensured the ecological validity of their experiments by inviting performers to record in the studio with the incentive of getting a demo that they could use to promote their music. In this context, the world music performer and two improvisation/jazz trios accepted to test the BMR system in a counterbalanced order, and each switched conditions once, after 2 h and 45 min, respectively. Also, a singer-songwriter, a rock duo, and a pianist who performed with electronics accepted to test the ABH system once they were satisfied with their takes using the traditional stereo system of the studio.

For the seven case studies during which performers compared binaural and stereo headphones, the researchers took notes on users' behaviors and comments during the tests. Whereas Soudoplatoff asked Maestro Petitgirard to react spontaneously after each trial, Bauer conducted post-test focus group interviews, and Menon carried out

individual post-test written surveys with the performers at the end of the recording session. For all case studies, performers were asked to compare both types of headphones in terms of comfort, playfulness, benchmarks, and perception of the spatial image. For the recording sessions only, performers were asked to compare the perception of their own instrument in relation to others'. Moreover, a few weeks after the recording sessions of the world music performer and the two improvisation/jazz trios, Bauer sent stereo mixes of all the takes to the performers, and he asked them to select their favorite take for each piece (or their favorite improvisation). Based on a previous performance study in the recording studio in jazz [62], collecting musicians' choice of takes that were recorded in different conditions has the potential to inform the impact of the BMR on creativity and musical results. The context of Soudoplatoff's and Menon's tests did not allow for this additional collection of data.

4.1.3 Click-to-music loudness ratio measurements

To investigate the extent to which the reduction of the sound masking effect in binaural enabled musicians to monitor less synchronization cues, for each of his three case studies¹⁵, Menon [9] compared the click-to-music (CMR) loudness ratio between the headphone mix recordings of the takes using his ABH and those using the traditional stereo monitoring system of the studio. For each of the takes recorded with the ABH, he copied the musicians' KV interface settings into a "second user," so that he could print the monitoring mix that featured the binauralization of the four headphone-mounted microphones and the synchronization cues. For each of the takes recorded with the stereo headphones, he captured the signal from the headphone output of the Aviom personal monitor mixer by using a stereo jack into two unbalanced jack adapters and two Direct Input (DI) boxes. Then, because each monitoring mix replica would include a few seconds of synchronization cues before the beginning of the music performance, he could normalize the loudness of each replica with the synchronization cues as a reference. This data acquisition procedure enabled the visualization of the CMR throughout and across takes.

4.2 Experimental findings

For the seven case studies during which performers compared binaural and stereo headphones, all performers favored the binaural over the stereo condition. In the following sections, we detail comparison findings for the main criteria that emerged from our analysis of performers' comments and take choices, namely Listening comfort; Perceived realism; and Musical expression and creativity.

For the symphonic jazz ensemble recording session, the comparison could not be conducted as planned due to several challenges that highlighted the limitation of the BMH [7]. This large ensemble combined orchestral and big band instruments with electric guitars and keyboards that were not amplified in the room, as well as drums that were semi-isolated in the room. Consequently, the electric guitars, keyboards, and the double-bass' quiet acoustic sound were not captured by the main 5-microphone array so they could not be homogeneously integrated into the auditory scene. Also, the main array captured a lot of drum leakage, which damages the intelligibility of the auditory scene. Moreover, the

¹⁵ Menon [9] conducted a fourth case study with a classical pianist who tested the ABH and compared it to stereo headphones to monitor a metronome while performing Beethoven's Piano Sonata No. 2, Op. 57. Because this piece would not be performed with a metronome in professional situations, we excluded this fourth study from this chapter.

complexity of the situation generated communication challenges between the electric instrument players, the sound engineer, and the conductor, therefore the conductor did not feel comfortable enough to use the BHT for the session. In the discussion, we provide ideas to overcome the BHT limitations for conducting large ensembles that blend different types of instruments in large acoustics.

4.2.1 Listening comfort

All eleven performers who participated in comparative studies in the recording studio preferred the auditory feedback quality of their own sound production in the binaural conditions. In particular, two improvisers who tested the BMR and all performers who tested the ABH reported having more control of their own instruments. For instance, the world music performer kept both earcups in the binaural condition but removed one earcup to control his voice in relation to the room acoustics in the stereo condition. Also, the double bass player of the jazz trio perceived a more realistic “physical-auditory contact” with his instrument in the binaural condition.

The conductor and seven instrumentalists expressed being more comfortable while performing in the binaural condition. In particular, whereas Maestro Petitgirard was a bit reluctant to try the BHT in the beginning, he mentioned feeling comfortable with it as soon as he started using it. Also, two out of the four performers who tested the ABH stated that they were able to forget about the device while monitoring in binaural. Furthermore, three out of the seven performers who tested the BMR reported that the binaural condition was less tiring in comparison with the stereo condition. Only the world music performer was disturbed during the first hour by this new kind of monitoring.

All performers perceived better sound quality in the binaural condition that they described as more natural than stereo in terms of spatial realism and audio clarity. With the BMR, all performers perceived the binaural mix as more intelligible, since they could better differentiate the details of the different instruments. In this view, free improvisers and jazz musicians reported “not having to force” to hear what they needed to react to their bandmates’ musical gestures. They could appreciate more subtleties in their playing, for example, the sounds of the fingers on the double bass and soft percussions, and the drummer said that the sound was more “accurate to what they would hear in their daily practice.” Also, the free improvisers who used the BMR and Maestro Petitgirard who used the BHT perceived more depth in the binaural mix compared to the stereo mix.

4.2.2 Perceived realism

Across the seven comparison studies, performers expressed that binaural monitoring was more realistic. However, the meaning of *realism* varied according to the type of augmentation that was used in the different studies. Regarding the two AAR systems, realism implied that the binaural rendering of the music signal was close to the real auditory environment in terms of source spatialization, room acoustics, and timbre quality. In contrast, regarding the BMR solution that is AMR, by realism performers meant that they could recreate familiar auditory situations in their mind, for example, to “be in the performance” and to connect with other players and their own instrument like in rehearsal. In the next paragraph, we illustrate these two meanings of realism with test observations.

When first trying the BHT, Maestro Petitgirard thought that he was only hearing the click track, and Soudoplatoff had to convince him that the orchestra

was also rendered in the headphone mix by muting the microphones for a few seconds. Similarly, all performers who experimented with the ABH mentioned that they perceived a more realistic spatial image in comparison with stereo monitors. Beyond the basic acknowledgment of the spatial authenticity that the ABH facilitated, performers commented explicitly on the efficacy of this enhanced acoustic realism. For instance, the pianist who performed the electroacoustic piece stated, "I felt myself making decisions in real-time, reacting to my own emotions and improvising some aspects of interpretation, whereas with the traditional headphones, I found my performance becoming stagnant." As for the AMR system, since 3D audio cues did not match the real auditory scene of the studio, realism was about the sound quality of recreated acoustics and the convincing spatialization of 3D audio cues. This led the world music performer to report that he "had the impression that the music was real around him." Moreover, two of the free improvisers had the impression that their bandmembers were next to them although they were in separated rooms. In particular, the clarinetist said: "It recreated a second room where we were all present in my head."

4.2.3 Musical expressivity and creativity

All performers who tested the BMR or the ABH stated that binaural monitoring positively impacted their musical playfulness and creative process. Whereas performers did not expand verbally on this impact, six out of the seven who used the BMR only selected takes that were recorded in the binaural conditions. Also, we observed that the takes that were recorded by the free improvisation trio with binaural monitoring lasted longer, and the clarinetist reported, "musical ideas came faster." Moreover, the guitarist of the jazz trio expressed being able to take more risks, and the world music performer reported being inspired by the binaural auditory space to build his composition in the studio. In contrast, whereas Maestro Petitgirard perceived the BHT as very pleasing, he said that the monitoring condition should not have impacted his way of performing as he had drawn well-established habits over years of conducting experience.

The free improvisers who used the BMR and all performers who used the ABH expressed that they performed more intimately in binaural conditions. For instance, the free improvisers noticed that they performed the only soft improvisation with many subtleties while monitoring in binaural. Similarly, the pianist who played an electroacoustic piece said that binaural monitoring facilitated a more sensitive performance. Moreover, synchronization cues were more easily perceived in the binaural condition. Indeed, the singer-songwriter who tested the ABH noted that keeping tempo was easier, and the drummer who used the BMR reported that there was better bass/drums cohesion in the binaural condition, which led to more swing.

4.2.4 Click-to-music loudness ratios

The Click-to-music-ratio (CMR)¹⁶ analyses were measured in relative Loudness Units (LU). These analyses across tests showed that the CMR was 4.2 LU to 17.4 LU lower when using the ABH compared to the stereo systems [9]. **Figure 4** displays the CMR in LU at key performance moments of the pop-rock duo for the drummer's monitoring mix with a click track (A), and at key performance timings of the electroacoustic piece for the pianist's monitoring mix with a soundtrack that included

¹⁶ Examples of the binaural versus stereo monitoring mixes that the musicians heard are available under this link: <https://www.youtube.com/watch?v=8c8IBCzJR-M>

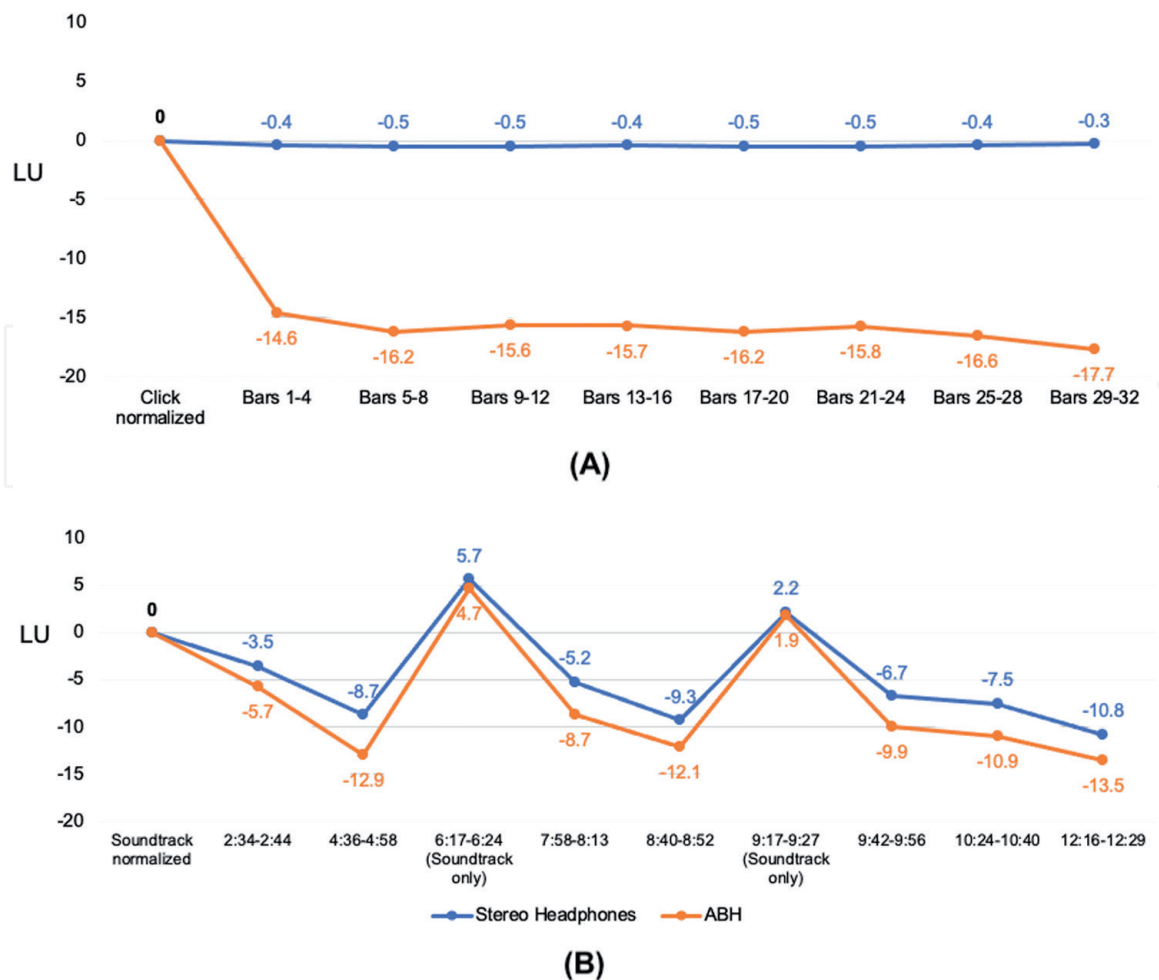


Figure 4. Click-to-music loudness ratios (A) in the drummer's monitoring mix at key sections of the guitar and drums pop-rock duo, and (B) in the pianist's monitoring mix at key timings of Nicole Lizée's Hitchcock Études.

a click track on the left channel (B). We observe that for the chorus of the pop-rock duo, the drummer monitored the click track at 17.7 LU lower than the music with the ABH, versus at nearly the same loudness as the music at 0.3 LU with the stereo headphones. While the CMR decrease was less noticeable for the pianist's mix, we could observe that the ABH enabled a more dynamic headphone mix, and so a more expressive balance between the piano and soundtrack than the stereo headphones.

5. Discussion

5.1 What are the main challenges of using a monitoring wearable device while performing music?

Results from the questionnaires expand previous findings regarding the challenges that musicians face when performing with wearable monitoring devices [1]. In addition to being very sensitive to the sound quality of the headphone mix, performers also strongly value the technical quality and physical properties of the monitoring system. Moreover, results confirm that they develop strategies to cope with their discomfort. Indeed, only one out of the 21 respondents who answered the fourth question reported wearing headphones on both ears while performing. It should be noted that wearing only one earcup or half of both earcups is tiring for performers due to the asymmetry or layer of the auditory feedback. These findings thus reinforce the need to find monitoring solutions that overcome the challenges of traditional stereo headphones.

A large number of phrasings about negative musical consequences show that musicians are aware of the impact of poor monitoring setups on their performance. In this view, instrumentalists' comments during the case performance studies confirm that many do not expect to get a comfortable headphone mix in the studio [1] and that some of them come to the studio mentally prepared to face monitoring challenges. For instance, the drummer of the jazz trio explained that he usually expects to experience latency issues. However, while we know that monitoring mixes lead to different ways of performing, be the impact positive or negative [3], survey respondents surprisingly did not mention any positive musical consequences. Similarly, we noticed a reluctance from the participants who tested the BMR to detail the positive effects of their preferred monitoring condition on their musicality. These findings indicate that musicians and sound engineers should communicate more about monitoring systems to transcend the status quo. Also, results show that improvisers conceptualize their ideal monitoring system differently than orchestra conductors do, which corroborates with the need for engineers to adapt the design of monitoring systems as well as recording and live engineering sound choices [27] to the esthetic and culture of the performance context.

5.2 Could binaural technologies that are adapted to specific performance contexts enhance musicians' listening comfort, perception of a realistic auditory scene, and musical expression and creativity?

Across the seven comparison case studies, performers appreciated the listening comfort in binaural compared to stereo, and they expressed that the binaural rendering was more realistic than the stereo rendering. Nevertheless, in keeping with AMR and AAR definitions from the literature review [24, 26], the meaning of the realism concept varied depending on the augmentation type, from a convincing recreated spatial auditory scene in AMR to an auditory scene close to what performers heard in the real acoustics in AAR. This AAR realism definition was further researched by Soudoplatoff and Pras [7] who asked 15 sound engineers to describe how real they perceived the superposition of the binaural rendering of two soundscapes that were captured in the same room with the same microphone setup. The two soundscapes featured a jazz duo performance that was happening live in real time on the other side of the studio window, and a crowded ambiance that was recorded a few days before to give the illusion of a bar soundscape. Results showed that participants perceived "scene realism and a well-established illusion of being in a crowd." These outcomes call for future research that would assess the relevance of superposing a binauralized pre-recording of the synchronization cues in the venue to the music in performers' monitoring mixes.

For all AAR and AMR case performance studies, findings highlighted that the binaural conditions enabled all participants to be more collectively and cognitively involved in their creative tasks compared with the stereo conditions. This implies that our AAR setups could overcome social interaction challenges when wearing headphones and that an AMR setup could enhance social interaction among participants remotely located. Specifically, we suggest that participants were more *immersed* in performance [50] and that the free improvisers experienced a state of *flow* [63] since they performed longer takes with binaural monitoring. These research outcomes are important from an artistic perspective and should be made broadly available to musicians. Indeed, Menon noted from his studio experience as a rock guitarist that when controlling a personal monitoring mixer consumes more time than desired, musicians would rather cope with whatever they are hearing than to fix these issues. Therefore, we believe that greater learning around the impact of monitoring systems on the musicians' ability to be immersed in performance would motivate them to always ensure an optimized headphone mix.

In keeping with the findings of the BBC study [45], we found that the binaural rendering of the main array worked well to recreate the auditory scene of the film-scoring orchestra. In contrast, the binaural rendering of the same array was problematic in the jazz symphonic ensemble situation that featured complex interactions of room and instrument acoustics. Here we propose three solutions that could have helped reduce the drum leakage and better integrate the electric instruments and the double bass within the auditory scene. First, the percentage of natural reverberation in the mix could be manipulated by changing the balance between the main array and the spot microphones. Second, reverberation could be artificially re-created by using a real-time binaural room simulator with an object-based mixing device to wet the electric instruments and double bass, and thus enable their acoustic homogeneity with the rest of the auditory scene. Finally, using transparent glass panels could minimize the leakage of the drums while maintaining visual contact and giving all band members the illusion of playing in the same room. Moreover, the jazz symphonic ensemble situation reminded us that flaws in the quality and intercommunication setup of a monitoring system can be detrimental to the performance situation, for instance by increasing performers' stress [1], and/or by disempowering musicians while reinforcing the engineers' sound control [4]. This negative experience demonstrates that technological adaptation to the music performance context requires sound engineers to consider the overall studio context and researchers to ensure the success of the first trial.

5.3 Could binaural headphone monitoring systems increase the click-to-music ratio compared to stereo headphones?

With the ABH and BMR, we observed that performers required fewer synchronization cues in their headphone mix compared to the stereo conditions, due to enhanced binaural intelligibility and less sound masking effects [17]. The less dynamic nature of the stereo headphones brings musician to monitor a louder headphone mix overall, which is likely to damage their aural health over time [13]. Indeed, in addition to perceiving a more realistic spatial image, the extreme dynamic range differences between the click track and the music signal are enhanced by the binaural rendering, leading musicians to set their monitor level at a comfortable volume to enjoy the dips and valleys of the musical scenario. In this view, the drummer who used the ABH system got concerned about losing his hearing when he realized how loud his click track was in stereo. He mentioned to Menon that he would consider purchasing a wearable metronome to avoid using audible click tracks at such high volumes in the future. Similar findings appeared with the BMR system, as the jazz musicians and free improvisers could hear the music cues more distinctly without forcing in binaural compared to stereo. In particular, the jazz drummer and guitarist asked Bauer to increase the bass level in their monitoring mix when switching from binaural to stereo, as they explained that the bass was masked by other music elements in stereo. Also, whereas the world music performer removed one earcup in stereo while overdubbing, as the headphone mix content gradually got denser, he kept both earcups on throughout the recording process with the binaural condition. These findings also illustrate the challenge of controlling the balance of headphone mixes in stereo.

One of Soudoplatoff's motivations in developing a binaural monitoring solution was that a conductor from his professional network suffered from hyperacusis and tinnitus due to working with loud in-ear monitors when she toured with a symphonic orchestra mixed with electronic music in a dozen representations over a two-week period. The results of our seven comparison tests with three binaural monitoring setups encourage us to pursue this research to improve music performers' working conditions. One next step would be to focus on better integration of the click track in

the monitoring mix by spatializing it to allow for its externalization with an appropriate localization distance. We believe that this advancement would enable musicians to monitor even less click track than with our BHT, BMR, ABH, and Copper and Martin [2]'s ATH, and thus to feel even more immersed in the performance. This should allow to reduce the cultural implications of the click track [14, 15]. Such an approach would thus treat the click track as AAR instead of AMR and would take full advantage of binaural unmasking capabilities.

6. Conclusion

Our research contributions that concur to show the potential of dynamic and static binaural monitoring solutions in enhancing performers' immersion in creative cognition are threefold. First, we constructed new theoretical knowledge on musicians' experiences when performing with headphones based on a multidisciplinary literature review and on survey responses from professional orchestra conductors and music improvisers. This knowledge provides acousticians with important insights to develop ecologically valid experimental protocols to assess innovative technologies in professional music performance contexts. Second, we designed one AMR and two AAR binaural monitoring solutions for which we detailed how we adapted their augmentation type, sound capture, and mixing approaches to distinct music performance contexts. Sound engineers can extend and modify these solutions to other contexts, within and beyond music performance. Third, we explored mixed-method approaches to assess our technologies in three different professional performance contexts through eight case studies. These approaches combined performers' feedback on their experience by comparing binaural versus stereo conditions, their choice of takes, and the measurement of click-to-music loudness ratios in both conditions. Discussed in terms of performers' listening comfort, their perception of a realistic auditory scene, and their musical expression and creativity, our case study outcomes could be integrated into close-ended feedback questionnaire in future studies that aim at assessing monitoring solutions based on performers' experience.

Because this series of studies drew out more insights into the positive influence of using binaural headphone monitoring for instrumentalists than for conductors, future comparisons between the two AAR solutions, namely BHT and ABH, will be pursued with professional conductors to determine which dynamic binaural solutions could best support their performance experience. Because our case study outcomes underline the positive influence of binaural monitoring over the music performance for instrumentalists but do not address the case of singers, further research will identify which binaural monitoring solution would best support professional singers' performance needs. Future research will also include tests with more musicians of different popular music genres to find solutions in terms of beat spatialization, which we know can be tricky in binaural, especially for hip hop [47].

To circumvent the hesitation to acknowledge the impact of monitoring technology on professional performers' musicality among practitioners, we encourage researchers to adopt a post-performance procedure, for instance through the analysis of performers' take choices a few weeks after the recording session to analyze the potential interconnection between the experimental condition and the best musical result [62]. This approach calls for conducting case studies that are fully integrated into real-life recording sessions that last several days. Also, future studies should further examine music performers' perception of *acoustic realism* when creating music in real-life situations with binaural monitoring, and how this perception depends on the cultural context (musical genre, ensemble's habits) and the acoustic situation (acoustic separation or not, size of the venue, amplified instruments or not). To that end, it would be

interesting to compare three recording setups, such as instrumentalists being in the same room without headphone monitoring, instrumentalists being in the same room with binaural monitoring to augment their natural hearing, and instrumentalists being in separate rooms and hearing each other through binaural monitoring. This investigation could also help to refine the potential of AAR and AMR approaches from a practical point of view, and thus inform the design of future headphone monitoring systems. Furthermore, and with respect to the emerging concept of acoustic realism, a large longitudinal study should be conducted to identify the duration requirements of a training procedure with non-individualized HRTFs to reach an optimal level of performers' ability to externalize binaural audio cues, as well as to appreciate the intelligibility and comfort of a binaural mix. In that respect, using complex musical stimuli during the training procedure is advised, instead of non-ecological stimuli such as pink noise. At last, only few binaural music productions have been released on the market so far, and we hope that our research will inspire more sound engineers to explore binaural mixing techniques, and the music industry to give a chance to this 3D audio format.

Acknowledgements

We would like to thank all the questionnaire respondents and case study participants for their time, expertise, and useful insight into this research, as well as Dr. Terri Hron for her English review of our first submission. Also, we acknowledge the assistance and contribution of Dr. Georg Boenn and Chris Morris from the University of Lethbridge; Hervé Déjardin from Radio France; and Dr. Pascal Dietrich, Phil Kamp, Benedikt Krechel, Markus Pesch, and Dr. Roman Scharrer from KLANG: technologies to our technology designs and performance case studies. Menon's research assistantship and the publication processing charges for this chapter were funded by Pras' Partnership Development Grant of the Social Sciences and Humanities Research Council of Canada (SSHRC).

Author details

Valentin Bauer^{1*}, Dimitri Soudoplatoff², Leonard Menon³ and Amandine Pras⁴

1 The Interdisciplinary Laboratory for Computer Sciences (LISN), Université Paris-Saclay, CNRS, France

2 Specialised Orchestral Conducting Master of Arts, Geneva University of Music (HEM), Switzerland

3 Schulich School of Music, McGill University, Canada

4 Department of Music, University of York, United Kingdom

*Address all correspondence to: valentin.bauer@lisn.fr

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Williams A. I'm Not Hearing What You're Hearing: The Conflict and Connection of Headphone Mixes and Multiple Audioscapes. In: Frith S, Zagorski-Thomas S, editors. *The Art of Record Production: An Introductory Reader for a New Academic Field*. 1st ed. Farnham: Ashgate; 2012. p. 113-128
- [2] Cooper A, Martin N. The impact of a prototype acoustically transparent headphone system on the recording studio performances of professional trumpet players. In: Hepworth-Sawyer R, Hodgson J, Paterson J, Toulson R, editors. *Innovation in Music: Performance, Production, Technology, and Business*. 1st ed. New York: Routledge; 2019. pp. 368-384. DOI: 10.4324/9781351016711-23
- [3] Berg J, Johannesson T, Löfdahl M, Nykänen A. In-ear vs. loudspeaker monitoring for live sound and the effect on audio quality attributes and musical performance. In: *Proceedings of the Audio Engineering Society Convention 142*; 20-23 May 2017; Berlin. New York: Audio Engineering Society; 2017
- [4] Brooks G, Pras A, Elafros A, Lockett M. Do we really want to keep the gate threshold that high? *Journal of the Audio Engineering Society*. 2021;69(4):238-260. DOI: 10.17743/jaes.2020.0074
- [5] Berg J, Jullander S, Sundkvist P, Kjekshus H. The influence of room acoustics on musical performance and interpretation—A pilot study. In: *Proceedings of the Audio Engineering Society Convention 140*; 4-7 June 2016; Paris. New York: Audio Engineering Society; 2016
- [6] Slaten WJ. *Doing Sound: An Ethnography of Fidelity, Temporality, and Labor Among Live Sound Engineers*. New York City: Columbia University; 2018
- [7] Soudoplatoff D, Pras A. Augmented reality to improve orchestra conductors' headphone monitoring. In: *Proceedings of the Audio Engineering Society Convention 142*; 20-23 May 2017; Berlin. New York: Audio Engineering Society; 2017
- [8] Bauer V, Déjardin H, Pras A. Musicians' binaural headphone monitoring for studio recording. In: *Proceedings of the Audio Engineering Society Convention 144*; 23-26 May 2018; Milan. New York: Audio Engineering Society; 2018
- [9] Menon L. Click-to-music ratio: Using active headphones to increase the gap. In: *Proceedings of the Audio Engineering Society Convention 149*; 27-30 October 2020; Online. New York: Audio Engineering Society; 2020
- [10] Guastavino C, Katz BF, Polack J, Levitin DJ, Dubois D. Ecological validity of soundscape reproduction. *Acta Acustica United With Acustica*. 2004;91:333-341
- [11] Donin N, Traube C. Tracking the creative process in music: New issues, new methods. *Musicae Scientiae*. 2016;20(3):283-286. DOI: 10.1177/1029864916656995
- [12] Klapholz J. Fantasia: Innovations in sound. *Journal of the Audio Engineering Society*. 1991;39(1/2):66-70
- [13] Chasin M. Musicians and the prevention of hearing loss. In: *Proceedings of the 2018 Audio Engineering Society International Conference on Music Induced Hearing Disorders*; 20-22 June 2018. Chicago Illinois. New York: Audio Engineering Society; 2018
- [14] Cardassi L. In search of expressive time in mixed media works: Composer-performer collaboration, synchronization cues and customized click-tracks. In: *Proceedings of the Encontros de Cognição Musical—Processos Criativos*; 18-20 November

2020; Salvador: Universidade Federal da Bahia; 2020

[15] Cardassi L. Balancing musical and mechanical cues for synchronization in mixed media works and a model for customized click-tracks. ART Music Review.

[16] Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. Revised ed. Cambridge: MIT Press; 1996. 502 p

[17] Oltheten W. Mixing with Impact: Learning to Make Musical Choices. 1st ed. New York: Routledge; 2018. 364 p. DOI: 10.4324/9781315113173

[18] Roessner S. The Beat Goes Static: A Tempo Analysis of US Billboard Hot 100# 1 Songs from 1955-2015. In: Proceedings of the Audio Engineering Society Convention 143; 18-20 October 2017; New York. New York: Audio Engineering Society; 2017

[19] Roginska A. Binaural audio through headphones. In: Roginska A, Geluso P, editors. Immersive Sound. 1st ed. New York: Routledge; 2017. pp. 88-123. DOI: 10.4324/9781315707525-5

[20] Lindeman RW, Noma H, De Barros PG. Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio. In: Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality; 13-16 November 2007; Nara. Washington, DC: IEEE; 2007

[21] Fujise A. Investigation of practical compensation method for bone conduction headphones with a focus on spatialization. In: Proceedings of the 2018 Audio Engineering Society International Conference on Spatial Reproduction-Aesthetics and Science; 7-9 August 2018; Tokyo. New York: Audio Engineering Society; 2018

[22] May KR, Walker BN. The effects of distractor sounds presented through bone

conduction headphones on the localization of critical environmental sounds. Applied Ergonomics. 2017;**61**: 144-158. DOI: 10.1016/j.apergo.2017.01.009

[23] Albrecht R, Lokki T, Savioja L. A mobile augmented reality audio system with binaural microphones. In: Proceedings of Interacting with Sound Workshop: Exploring Context Aware, Local and Social Audio Applications; 30 August 2011; Stockholm. New York: ACM; 2011. pp. 7-11

[24] Härmä A, Jakka J, Tikander M, Karjalainen M, Lokki T, Hiipakka J, et al. Augmented reality audio for mobile and wearable appliances. Journal of the Audio Engineering Society. 2004;**52**(6):618-639

[25] Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems. 1994;**77**:1321-1329

[26] McGill M, Brewster S, McGookin D, Wilson G. Acoustic transparency and the changing soundscape of auditory mixed reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; 25-30 April 2020; Honolulu. New-York: ACM; 2020. pp. 1-16

[27] Pras A, Guastavino C, Lavoie M. The impact of technological advances on recording studio practices. Journal of the American Society for Information Science and Technology. 2013;**64**(3):612-626. DOI: 10.1002/asi.22840

[28] König FM. New measurements and psychoacoustic investigations on a headphone for TAX/HDTV/Dolby-surround reproduction of sound. In: Proceedings of the Audio Engineering Society Convention 98; 25-28 February 1995; Paris. New York: Audio Engineering Society; 1995

[29] Møller H. Fundamentals of binaural technology. Applied Acoustics. 1992;**36**(3-4):171-218

- [30] Best V, Baumgartner R, Lavandier M, Majdak P, Kopčo N. Sound externalization: A review of recent research. *Trends in Hearing*. 2020;**24**: 1-14. DOI: 10.1177/2331216520948390
- [31] Mendonça C, Campos G, Dias P, Vieira J, Ferreira JP, Santos JA. On the improvement of localization accuracy with non-individualized HRTF-based sounds. *Journal of the Audio Engineering Society*. 2012;**60**:821-830
- [32] Chatzidimitris T, Gavalas D, Michael D. SoundPacman: Audio augmented reality in location-based games. In: *Proceedings of the 18th Mediterranean Electrotechnical Conference*; 18-20 April 2016; Limassol. New York: IEEE; 2016. pp. 1-6
- [33] Katz BF, Kammoun S, Parseihian G, Gutierrez O, Brilhault A, Auvray M, et al. NAVIG: Augmented reality guidance system for the visually impaired. *Virtual Reality*. 2012;**16**:253-269. DOI: 10.1007/s10055-012-0213-6
- [34] Naphtali D, Rodkin R. Audio augmented reality for interactive soundwalks, sound art and music delivery. In: Filimowicz M, editor. *Foundations in Sound Design for Interactive Media*. 1st ed. New York: Routledge; 2019. pp. 300-332. DOI: 10.4324/9781315106342
- [35] Mariette N. Human factors research in audio augmented reality. In: Huang W, Alem L, Livingston MA, editors. *Human Factors in Augmented Reality Environments*. 1st ed. New York: Springer; 2013. pp. 11-32. DOI: 10.1007/978-1-4614-4205-9
- [36] Nagele AN, Bauer V, Healey PG, Reiss JD, Cooke H, Cowlshaw T, et al. Interactive audio augmented reality in participatory performance. *Frontiers in Virtual Reality*. 2020;**1**:46. DOI: 10.3389/frvir.2020.610320
- [37] Moustakas N, Floros A, Grigoriou N. Interactive audio realities: An augmented/mixed reality audio game prototype. In: *Proceedings of the Audio Engineering Society Convention 130*; 13-16 May 2011; London. New York: Audio Engineering Society; 2011
- [38] Zimmermann A, Lorenz A. LISTEN: A user-adaptive audio-augmented museum guide. *User Modeling and User-Adapted Interaction*. 2008;**18**: 389-416
- [39] Mariette N, Katz BF, Boussetta K, Guillerminet O. SoundDelta: A study of audio augmented reality using WiFi-distributed Ambisonic cell rendering. In: *Proceedings of the Audio Engineering Society Convention 128*; 22-25 May 2010; London. New York: Audio Engineering Society; 2010
- [40] Goony A. Les HRTF appliquées au retour de scène par "in-ear monitors" [master thesis]. Saint Denis: École nationale supérieure Louis-Lumière; 2010
- [41] Zea E. Binaural In-Ear Monitoring of acoustic instruments in live music performance. In: *Proceedings of the 15th International Conference on Digital Audio Effects*; 30 November – 3 December 2012; Trondheim. Trondheim: Norwegian University of Science and Technology; 2012. pp. 1-8.
- [42] Field A. Hearing the past in the present: An augmented reality approach to site reconstruction through architecturally informed new music. In: Schofield J, Maloney L, editors. *Music and Heritage*. 1st ed. London: Routledge; 2021. pp. 212-221
- [43] Murphy D, Shelley S, Foteinou A, Brereton J, Daffern H. Acoustic Heritage and Audio Creativity: the Creative Application of Sound in the Representation, Understanding and Experience of Past Environments. *Internet Archaeology*. 2017;**44**. DOI: 10.11141/ia.44.12
- [44] Gebhardt M, Kuhn C, Pellegrini R. Headphones Technology for Surround

Sound Monitoring—A Virtual 5.1 Listening Room. In: Proceedings of the Audio Engineering Society Convention 122; 5-8 May 2007; Vienna. New York: Audio Engineering Society; 2007

[45] Parnell T, Pike C. An efficient method for producing binaural mixes of classical music from a primary stereo mix. In: Proceedings of the Audio Engineering Society Convention 144; 23-26 May 2018; Milan. New York: Audio Engineering Society; 2018

[46] Walton T. The overall listening experience of binaural audio. In: Proceedings of the 4th International Conference on Spatial Audio; 7-10 September 2017; Graz. VDT&IEM; 2017. pp. 170-177

[47] Turner K, Pras A. Is Binaural Spatialization the Future of Hip-Hop?. In: Proceedings of the Audio Engineering Society Convention 147; 21-24 October 2019; New York. New York: Audio Engineering Society; 2019

[48] Corbin J, Strauss A. Strategies for qualitative data analysis. In: Corbin J, Strauss A, editors. *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*. 3rd ed. New York: SAGE Publications Inc; 2008. 434 p. DOI: 10.4135/9781452230153.n4

[49] Pras A, Guastavino C. The role of music producers and sound engineers in the current recording context, as perceived by young professionals. *Musicae Scientiae*. 2011;**15**(1):73-95. DOI: 10.1177/1029864910393407

[50] Agrawal S, Simon A, Bech S, Bærentsen K, Forchhammer S. Defining immersion: Literature review and implications for research on immersive audiovisual experiences. *Journal of Audio Engineering Society*. 2019;**68**(6):404-417. DOI: 10.17743/jaes.2020.0039

[51] Cummings JJ, Bailenson JN, Fidler MJ. How immersive is enough? A

foundation for a meta-analysis of the effect of immersive technology on measured presence. *Media Psychology*. 2016;**19**:272-309. DOI: 10.1080/15213269.2015.1015740

[52] Heeter C. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*. 1992;**1**(2):262-271

[53] Messonnier JC, Lyzwa JM, Devallez D, de Boishéraud C. Object-based audio recording methods. In: Proceedings of the Audio Engineering Society Convention 140; 4-7 June 2016; Paris. New York: Audio Engineering Society; 2016

[54] Baskind A, Messonnier JC, Lyzwa JM. Bipan: An experimental mixing tool for 3D-audio on headphones with open-source head tracker. In: Proceedings of the 29th Tonmeistertagung-VDT International Convention; 17-20 November 2016; Cologne. Köln: Verband Deutscher Tonmeister e.V.; 2016

[55] Warusfel O. LISTEN HRTF database. Available from: <http://recherche.ircam.fr/equipes/salles/listen/> [Accessed: July 11, 2021]

[56] Hendrickx E, Stitt P, Messonnier JC, Lyzwa JM, Katz B, De Boishéraud C. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America*. 2017;**141**:2011-2023. DOI: 10.1121/1.4978612

[57] Nicol R, Gros L, Colomes C, Roncière E, Messonnier JC. Etude comparative du rendu de différentes techniques de prise de son spatialisée après binauralisation. In : Proceedings of the 13ème Congrès Français d'Acoustique; 11-15 April 2016; Le Mans. Paris: Société Française d'Acoustique; 2016. pp. 211-217

[58] Lester M, Boley J. The effects of latency on live sound monitoring.

In: Proceedings of the Audio Engineering Society Convention 123; 5-8 October 2007; New York. New York: Audio Engineering Society; 2007

[59] Stitt P, Hendrickx E, Messonnier JC, Katz B. The influence of head tracking latency on binaural rendering in simple and complex sound scenes. In: Proceedings of the Audio Engineering Society Convention 140; 4-7 June 2016; Paris. New York: Audio Engineering Society; 2016

[60] Brungart D, Kordik AJ, Simpson BD. Effects of headtracker latency in virtual audio displays. *The Journal of the Acoustical Society*. 2006;**54**(1/2):32-44

[61] Eley N. Classification of HRTFs using perceptually meaningful frequency arrays. In: Proceedings of the Audio Engineering Society Convention 147; 21-24 October 2019; New York. New York: Audio Engineering Society; 2019

[62] Pras A, Guastavino C. The impact of producers' comments and musicians' self-evaluation on perceived recording quality. *Journal of Music, Technology & Education*. 2013;**6**(1):81-101. DOI: 10.1386/jmte.6.1.81_1

[63] Csikszentmihalyi M. *Flow: The Psychology of Optimal Experience*. New York: Harper & Row; 1990. p. 303