Playing With Others Using Headphones: Musicians Prefer Binaural Audio With Head Tracking Over Stereo

Matteo Tomasetti ⬤ and Luca Turchet ⬤, *Senior Member, IEEE*

Abstract—**Immersive listening systems have grown significantly over the past decade and are now an established area of scientific, artistic, and industrial research. However, scarce research has been conducted on musicians' preferences for playing through headphones over binaural spatialization systems with the addition of head tracking, as opposed to classical stereophonic systems. This comparison is essential to optimally support the playing experience with others for cases of remote collaborative playing, individual instrumental practice, individual recreational music-making using backing tracks, and studio recording sessions. In this article, we study the preferences of playing musicians for a stereophonic system versus a binaural head-tracking system composed of Ambisonics technology and binaural synthesis with generalized head-related transfer functions. We conducted two experiments, each with 30 expert musicians, where participants were asked to rate and compare the 2 listening conditions while playing their instrument either seated or standing. Overall, the quantitative and qualitative results indicated a generalized preference for the binaural system with head tracking over the stereophonic system, with higher ratings for localization, immersion, social presence, realism, and connection with other musicians. Moreover, participants moved their heads significantly more in the binaural conditions. This phenomenon may be explained by the higher engagement and arousal due to the improved auditory experience, or alternatively by the presence of embodied music cognition mechanisms that cause a higher degree of exploration to better understand the action–perception loop. These findings highlight the need for progressing current commercial hardware and software systems used by musicians while playing over headphones.**

Index Terms—**Binaural audio, head tracking, music playing, music spatialization, networked music performance (NMP) systems.**

I. INTRODUCTION

IMMERSIVE audio has witnessed in the last decade a widespread interest in both academia and industry, as testified by a large amount of commercial products, patents, and scientific publications [1]. Applications of this technology are manifold, ranging from virtual and augmented reality [2], [3], [4], [5], to music listening [6], [7], to art installations [8], and mobile web contents [9].

Nowadays, the most accessible form of immersive audio is the binaural one because most people possess all the technologies necessary for its reproduction [10]. Binaural audio highly depends on sound source localization, which is related to acoustic cues, such as interaural time differences, interaural level differences, and acoustic filtering. The latter is the spectral information that depends on the physical conformity of each human being (e.g., the shape of ears, head, shoulders, and torso) [11], [12]. These acoustic cues in immersive audio software systems are rendered using head-related transfer functions (HRTFs). HRTF is the acoustic transfer function that depends on the directionality of a sound source to the listener's eardrum. The transfer function encodes the acoustic information used to localize sound and is crucial in the perception of sound localization in headphones [13]. HRTFs illustrate the changes in the sound spectrum as it enters the ear canal and depends on the diffraction and reflection of the physical conformation of each human body. That is why they are unique to each individual [14], [15].

HRTFs can be extrapolated from acoustic measurements [16] and organized into databases [17]. It is possible to measure each person's HRTF using costly recording systems, specialized facilities, and hardware [18], [19]. Because of the high cost, shrinking portability of the 3-D audio system, and computation challenges, generic HRTFs are used with lower accuracy and a higher margin of error in sound localization [19]. Measuring personal (or individualized) HRTFs is still impractical for the general public. For this reason, many binaural systems adopt generic HRTFs, which are measured on a simulated torso along with a binaural head or are calculated by averaging a set of people [20].

The fact that many binaural systems use generic HRTFs means that users listen to inadequate spatial cues [21]. In addition to HRTF-dependent localization errors, binaural systems have problems with front–back confusion, and externalization, as analyzed by Faller and Breebaart [22]. Nonetheless, the attention of researchers has been devoted to investigating use cases for binaural audio technology. In particular, different scholars investigated its benefits compared to the conventional stereo system, showing that binaural technology is preferred under certain conditions such as studio monitoring [23]. Musicians typically

face challenges when using stereo headphones to play with one another. This is mostly due to a lack of audio intelligibility and the loss of their usual localization benchmarks [24]. Nevertheless, in terms of listening experience, in a recent study by Morell and Lee [25], several participants were asked to evaluate stereophonic and binaural mixes. Specifically, they were asked to evaluate the all-inclusive immersive experience and perceived timbral and spatial aspects. Results showed that binaural mixes were rated lower than stereo mixes, especially for pop and rock genres.

However, to the best of authors' knowledge, today it is still unknown whether musicians prefer playing with others (offline or in real-time) using a binaural system with head tracking compared to a conventional stereo system. Unraveling a preference for a binaural with head-tracking setup would entail the need for progressing current systems used by musicians while playing over headphones. This would be relevant in various use cases such as: 1) recording sessions in a studio; 2) remote collaborative playing using a networked music performance (NMP) system; 3) individual instrumental practice, e.g., learning improvisation over backing tracks; 4) individual recreational music making using backing tracks. On the other hand, if stereo systems are preferred, this would indicate that binaural systems and head tracking are unnecessary for playing musicians.

In this article, we aim to understand musicians' preferences for binaural versus stereo setups and the underlying motivations. While we are aware of the inevitable problems with using generic HRTFs, we decided to use them in our study to be as generic as possible, since nowadays it is challenging for every musician to have individualized HRTFs. To minimize localization errors and ambiguities in binaural systems, we included in our study the use of head tracking, as suggested by Katz and Picinali [26]. Furthermore, we aim to investigate whether the use of the head tracking is considered as a useful tool by musicians playing with binaural audio. Specifically, our investigations were driven by the following four main research questions.

1) Is binaural audio with head tracking preferred over stereo while musicians play with other recorded musicians? If so, to which extent?
2) What is the experience of musicians playing with a binaural system and head tracking compared to a conventional stereo setup?
3) Are generic HRTFs sufficient to provide a better spatialization experience compared to the stereo while playing with others?
4) Do musicians explore more the sonic space (i.e., move their heads more) with a binaural system with head tracking compared with a stereo setup?

To address these questions, we conducted two experiments involving expert musicians, who played their main instrument using different sound spatialization conditions provided via headphones. Experiment 1 assessed the preference for a binaural plus head-tracking setup versus a stereo one, whereas Experiment 2 investigated more deeply the related experiences and perceptions of users, as described in Section IV-B. We carried out two different experiments because, in the first, we wanted to probe the direct and immediate preference of the musicians without investigating their motivations after each trial. In the

second, however, we wanted to understand with a deeper level of detail the actual motivations regarding preference by investigating different perceptual dimensions. Notably, our study focused exclusively on the auditory dimension, simulating other connected playing musicians. By the term "connected," we refer to the fact that in both experiments, we simulated the presence of other musicians playing in real-time through audio recordings. We were not interested in the investigation of audio–visual aspects resulting from the addition of a synchronous 2-D or 3-D visual representation of the connected musicians. Nevertheless, our study is important as it can provide a ground truth for future studies focusing on the more complex multisensory perception of audio–visual spatializations in musical interactions (e.g., using video streaming or virtual/augmented reality).

II. RELATED WORK

A. Noncommercial Binaural Systems

The IEM Plug-in Suite [27] is an open-source suite of plug-ins developed by students and researchers at the Institute of Electronic Music and Acoustics located in Graz. This suite makes it possible to find many plug-ins to encode and decode Ambisonic signals up to the seventh order [1]. The suite includes the "BinauralDecoder" plug-in for the binaural synthesis. Such a plug-in exploits the magLS approach proposed by Schörkhuber et al. [28], which renders the Ambisonics-encoded input signal into a binaural headphone signal. The 3-D Tune-in Toolkit [29] is a standard and open-source C++ library developed by teams at the University of Malaga and Imperial College London for sound spatialization (binaural or in loudspeakers), hearing loss simulation, and hearing aids. The Sparta & Compass [30] is a collection of flexible plug-ins developed by members of the Acoustics Lab at Aalto University, Finland, for the production, playback, and visualization of spatial audio.

B. Binaural Spatialization in Musical Contexts

In a previous study by Ueno et al. [31], comparisons of the same sound sources were made through listening tests between stereophonic playback with stereo width control and binaural audio with the HRTFs provided by Tohoku University [32]. Results showed that binaural audio has an advantage in reproducing the entire spatial image. The stereo width control suggests a better spatial impression in relationship with single-sound sources.

In Walton's report [10], a web-based study explored the impact of binaural audio on the overall listening experience, which is a measure used in assessing the quality of audio experience. The results revealed that binaural audio affects negatively the overall listening experience in a slight but notable way, compared with stereophonic reproduction in headphones.

C. NMP Systems

NMP systems aim to interconnect musicians over a wired or wireless network link [33] achieving the same conditions as instrumental on-site performances. A number of either hardware- or software-based solutions for NMP are currently available either at experimental or commercial stage. Noticeable examples

are LoLa [34], jacktrip [35], and Elk LIVE [36]. Most of these systems provide only the audio data streaming while others integrate video streaming.

Current NMP systems are not equipped with a set of independent channels, one for each sound source representing a connected musician. Existing systems only provide a stereo mix of the remotely connected musicians. For a binaural spatialization accounting for the rendering of the designated position of the connected musicians, it is necessary to provide at the receiver side the unmixed signals of each sound source. To enable such a scenario, it is necessary to advance the hardware and software components of NMP systems.

III. EXPERIMENT 1: PREFERENCE SELECTION

Experiment 1 assessed the musicians' preference between a binaural plus head-tracking system compared to a conventional stereo system.

A. Participants

Thirty participants took part in the evaluation (21 males and 9 females, aged between 19 and 47, mean age = 30.8, standard deviation = 7.5). All participants were highly expert musicians and reported normal hearing. No participant was knowledgeable of binaural audio technologies nor had previous experience with it. Moreover, none of them had any experience with NMPs. Furthermore, all participants had experience with the proposed genres in the experiment. During the experiment, seven played the electric guitar, five the electric bass, three the trumpet, two the double bass, three the violin, three the tenor saxophone, three the clarinet, two the keyboard, and two the flute. The reason for involving different musical instrument was motivated by our interest in assessing whether the type of instrument could have had an influence on the evaluations. Participants took on average one hour and a half to complete the experiment.

B. Apparatus and Stimuli

The setup consisted of a Dell Alienware x15 R2 laptop running the Windows 10 operating system, a pair of Beyerdynamic DT-770 Pro 80 Ohm headphones, one MetaMotionRL (MMRL) head-tracker by Mbient Lab, one RME Fireface UFX sound card, one AKG C414 XLII microphone (used for the acoustic instruments involved in the experiments), and a TRS-TRS cable used to connect each participant's instrument to the sound card. Half of the experiments were conducted at the participants' homes, each in optimal auditory conditions. The other half was conducted at a laboratory of the University of Trento, each equally in optimal auditory conditions. The MMRL was placed on top of the headphones, at the center of the head. For the software part, we used the Digital Audio Workstation (DAW) Cockos Reaper, Cycling '74 Max 8, and the MMRL2OSC application.

Concerning the stimuli, the multitracks of four different songs were collected. Each song belonged to different genres: Roots Rock, Funk, Blues, and Rock. Each song was edited to last 1 minute and 30 seconds, and all the instruments in the songs were recorded in a professional recording studio with acoustic treatment. These diverse musical genres were selected to assess whether genre could have an effect on participants' evaluations. The tracks composing the songs were each associated with a different instrument, namely drums, bass, keyboard, and guitar. Two mixes for each multitrack were created: one stereophonic and one binaural using Ambisonics. Specifically, we utilized the IEM Plug-in Suite [27] as it is a widely used, state-of-the-art system. For the encoding part, we created a multichannel routing with fifth-order Ambisonic and used the "BinauralDecoder" plug-in for the binaural synthesis. The results were eight mixes performed with the Reaper DAW, four using the binaural system with the head tracker, and four using stereo. The "Stereo Encoder" plug-in was used for spatialization and source encoding of each group for the binaural system with the head tracker. For spatialization, the sources were placed in the azimuth at this position: the drum at $0°$, the keyboard at $-45°$, the bass at $90°$, and the guitar at $-90°$. The exact positions were used to pan the instrument groups in the stereo mix, obviously using only the left and right pan provided. For example, if the bass in the binaural system was positioned at $90°$, in the stereo mix, we placed it at 100% panning in the left channel and so on for the other groups of musical instruments.

Furthermore, before the signal reached the "BinauralDecoder," the "SceneRotator" plug-in made all individual tracks communicate via open sound control (OSC) [37] with the MMRL. The "SceneRotator" plug-in was developed for the Ambisonics rotation of signals. For our experiment, we used rotation via quaternions and set the order of the rotations within the plug-in to yaw, pitch, and roll. This plug-in communicated with the head tracker and rotated the entire binaural scene based on the musician's head movements. Both binaural and stereophonic mixing achieved the same loudness (-12 LUFS). There was no perceivable latency in the system, including the Reaper DAW, that could prevent participants from playing normally. The head tracker transmitted the quaternion values into the DAW every 20 ms due to the Bluetooth connection's latency, the plug-ins' audio buffer, and the refresh rate of 200 Hz introduced by the MMR2OSC software used for OSC streaming.

Notably, the spatial parameter of elevation was not used in the binaural system because there is no equivalent parameter in the stereo system and we aimed at a fair comparison between both systems. Particularly, reverberation was not used in both systems because we wanted to investigate musicians' preferences without this variable, as the latter is difficult to replicate equally in both systems. Especially, the head tracker used did not allow any calibration, and the quaternion was sent via the head tracker's default system. For calibration, we mean that the head tracker default system communicated via OSC protocol with the spatialization software previously described. For this reason, it was not possible to experiment with different settings. Notably, the used headphones were not flat in equalization. They had equalization, especially in the high frequencies (above 5000 Hz). We decided to avoid applying any filtering because we were interested in assessing the two spatialization setups in the most general case, where the average musician does not have access to flat headphones or a software system to make headphones flat.

A pilot test was conducted with two professional musicians (with more than 15 years of musical experience), and they played electric guitar and bass. First, during the pilot test, we asked the participants whether they could perceive the differences between the two systems. Both answered that they immediately understood the differences between the two systems. Second, the two musicians consistently reported that the experience of playing while standing with the binaural system and the head tracker was much more immersive, engaging, and enjoyable than playing while seated. Therefore, we decided to investigate the sitting and standing positions. Furthermore, thanks to the feedback from the participants in the pilot test, it was decided to avoid spatializing the musical instrument being played and to have it always placed in the center of both systems. This decision was motivated by the fact that in this way, musicians could better recognize their contribution at all times. In addition, the pilot test participants were asked whether they heard the ambient sound (which means the reverberation) in both systems. Both participants responded that they heard no reverberation and that they heard the instruments clearly. The participants in the pilot test were not involved in the experimental sessions.

C. Procedure

The evaluation procedure consisted of the following steps. First, participants were briefed about the experiment and signed a consent form. Then, they were asked to interact with both systems by playing with their instruments over the "Roots Rock" genre condition (such a stimulus was not involved in the experimental sessions). This familiarization phase consisted of four trials, where the participant played standing or seated with both systems for each trial. After each trial, they were asked to express their preference as detailed below. During this phase, we also asked participants whether they were able to distinguish the two systems. All participants reported that they clearly understood the differences between the two systems. Neither during the familiarization phase nor the experiment participants were informed about the study purpose or the use of stereo or binaural conditions.

Subsequently, participants underwent the experimental trials. Each trial was composed of two parts. In the first part, one of the two systems (stereo or binaural with head tracking) was utilized, in the second, the other system. Both parts involved same conditions for position (standing, seated) and genre (Funk, Blues, and Rock). Specifically, in half of the trials, participants were exposed first to the stereo setup and then to the binaural with head-tracking system, in the other half vice versa. The resulting 12 trials were repeated twice in randomized order across participants, for a total of 24 trials. After playing over both systems, participants were asked to express their preference for either the first or second system (with a forced choice, i.e., *Do you prefer system 1 or system 2?*). After the first 12 trials, participants took a 10-min break to rest. After having completed all trials, participants underwent a semistructured interview, which involved the following open-ended questions.
1) *What was it like to play with the two systems? Why do you think the system you prefer is the best?*

2) *Is your ability to locate other musicians using the headphones an added value when playing with others, e.g., at a distance? Why?*

IV. EXPERIMENT 2: PREFERENCE DIMENSIONS

Experiment 2 aimed at investigating more deeply the reasons for the preferences for the binaural system and head tracker expressed by musicians in Experiment 1, by investigating different perceptual dimensions.

A. Participants

Thirty participants took part in the evaluation (24 males and 6 females, aged between 19 and 57, mean age $= 30.1$, standard deviation $= 9.8$). All participants were highly qualified expert musicians. None of these participants was involved in Experiment 1. During the experiment, seven played the electric guitar, five the electric bass, three the trumpet, two the double bass, three the violin, three the tenor saxophone, three the clarinet, two the keyboard, and two the flute (i.e., the same instruments as in Experiment 1, with the same number of participants per instrument). Similarly to Experiment 1, we aimed to assess whether the type of instrument could have had an influence on participants' ratings. Participants took on average one hour and a half to complete the experiment.

B. Apparatus, Stimuli, and Procedure

The involved apparatus and stimuli were the same as in Experiment 1. The familiarization phase followed the same structure as in Experiment 1. During the experiment, in each trial participants were asked to play over one of the three pieces (Funk, Blues, and Rock) with one of the two systems (stereo, binaural with head tracking), in one of the two positions (seated, standing). The resulting twelve trials were repeated twice in randomized order across participants, for a total of 24 trials. After the first 12 trials, participants took a 10-min break to rest. After each trial, participants were asked to fill in a questionnaire composed of the following items to be assessed on 11-point semantic differential scales.
1) *Social Presence: While playing I felt like if I was in the same room together with the other musicians.* [not present at all, very present]
2) *Localization: While playing I was able to localize the other musicians.* [not accurately at all, very accurately]
3) *Immersion: While playing I felt immersed in the acoustic scene.* [not immersed at all, very immersed]
4) *Ease: I felt at ease when playing.* [not at ease, very much at ease]
5) *Naturalness: I played in a natural way.* [not natural, very much natural]
6) *Connection: I felt connected with the other musicians.* [not connected, very much connected]
7) *Own Contribution: While playing I could distinguish well my contribution from that of the other musicians.* [little distinction, much distinction]
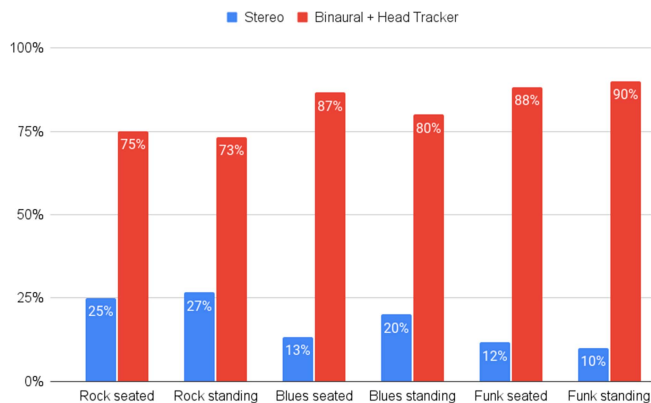
Fig. 1. Percentages of the expressed preferences of Experiment 1.

Fig. 2. Mean and standard error of the quantity of motion measured by the head tracker in experiments 1 (left) and 2 (right). Legend: $* = p < 0.05$; $** = p < 0.01$.

8) *Realism: The acoustic scene was similar to reality.* [not realistic, very much realistic]

9) *Sound Quality: Assess the sound quality.* [low quality, high quality]

After having completed all trials, participants underwent a semistructured interview, which involved the following open-ended questions:

1) *What was the difference you noticed between the two systems? What was the system where you found it best?*

2) *Is being able to locate other musicians in headphones an added value when playing at a distance? Why?*

Finally, participants were given the opportunity to leave an open comment.

V. RESULTS

A. Experiment 1: Quantitative Results

1) Questionnaire: Fig. 1 illustrates the percentages of the expressed preferences for each position condition in each music genre investigated.

2) Head-Tracker Data: The data from the head tracker were analyzed to search for differences between the various spatialization conditions. Specifically, we computed the quantity of motion in each trial to investigate whether participants moved more (or less) with the binaural setup compared to the stereo one. For this purpose, we computed the cumulative sum of the absolute values of the derivatives of the norm of the quaternions (saved at intervals of 100 ms). Data were analyzed with a linear mixed effect model, which had the quantity of motion variable and condition (modality, i.e., seated versus standing, and system, i.e., stereo versus binaural) as fixed factors, and subject as a random factor. A significant main effect was found for factor modality ($F(1,687) = 40.50$, $p < 0.001$, Cohen's $d = 0.42$), which indicates that participants moved their heads more when standing compared to when sit. A significant main effect was found also for the factor technique ($F(1,687) = 2.42$, $p < 0.05$, $d = 0.39$). As illustrated in Fig. 2, the post-hoc tests revealed that for both sit and standing modalities, the binaural technique had a higher quantity of motion compared to the stereo technique (respectively, $p < 0.001$, $d = 0.44$ and $p < 0.05$, $d = 0.27$).
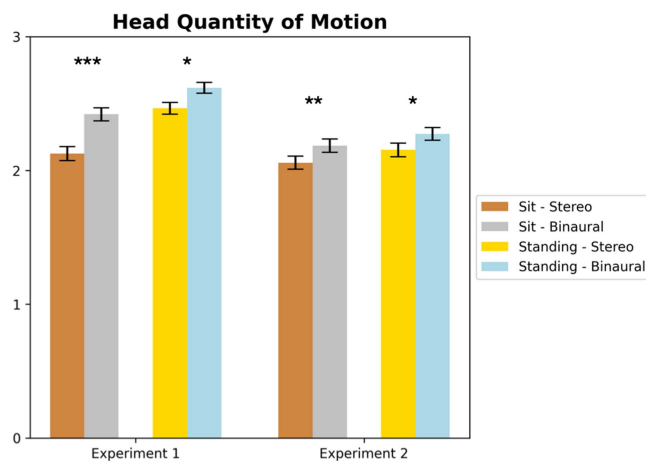
B. Experiment 2: Quantitative Results

1) Questionnaire: The results of the quantitative questions of the administered questionnaire are reported in Fig. 3. The ratings were analyzed with different linear mixed effect models, one for each response variable (Social Presence, Localization, Immersion, Ease, Naturalness, Connection, Own Contribution, Realism, Sound Quality). Specifically, each model had the response variable and condition (modality and technique) as fixed factors, and subject as a random factor. Post-hoc tests were performed on the fitted model using pairwise comparisons adjusted with Tukey's correction.

Regarding Social Presence, a significant main effect was found for the factor technique ($F(1,687) = 80.35$, $p < 0.001$, $d = 0.5$). The post-hoc tests revealed that for both sit and standing modalities, the binaural technique had higher scores of Social Presence compared to the stereo technique (respectively, $p < 0.001$, $d = 0.63$ and $p < 0.001$, $d = 0.38$).

Concerning Localization, a significant main effect was found for the factor technique ($F(1,687) = 118.67$, $p < 0.001$, $d = 0.62$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Localization compared to the stereo technique (respectively, $p < 0.001$, $d = 0.67$ and $p < 0.001$, $d = 0.57$).

As for Immersion, a significant main effect was found for the factor technique ($F(1,687) = 69.67$, $p < 0.001$, $d = 0.48$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Immersion compared to the stereo technique (respectively, $p < 0.001$, $d = 0.64$ and $p < 0.001$, $d = 0.34$).

Regarding Connection, a significant main effect was found for the factor technique ($F(1,687) = 65.89$, $p < 0.001$, $d = 0.38$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Connection compared to the stereo technique (respectively, $p < 0.001$, $d = 0.41$ and $p < 0.001$, $d = 0.35$).
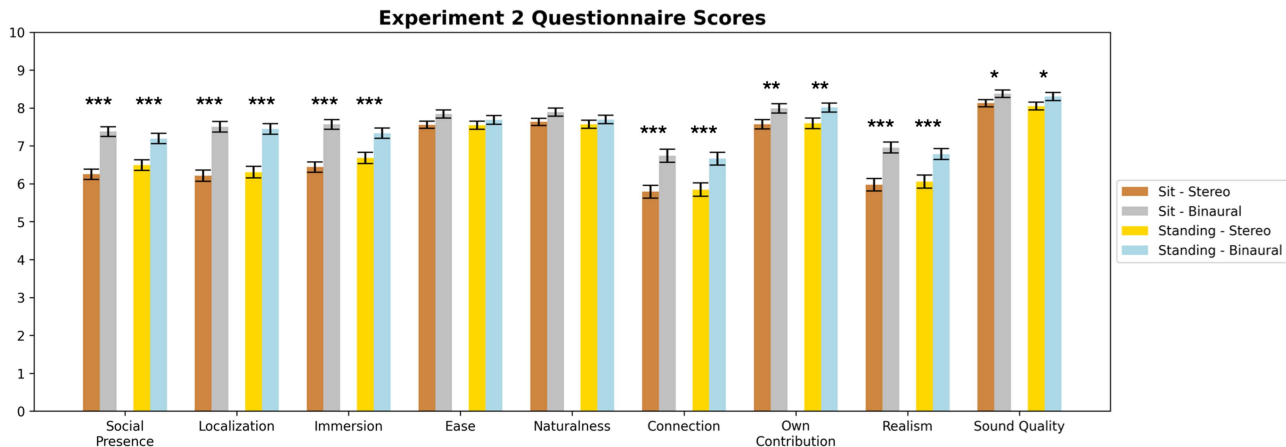
Fig. 3. Mean and standard error of the questionnaire scores of Experiment 2. Legend: $* = p < 0.05$; $** = p < 0.01$; $*** = p < 0.001$.

Concerning Own Contribution, a significant main effect was found for the factor technique ($F(1,687) = 21.75$, $p < 0.001$, $d = 24$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Own Contribution compared to the stereo technique (respectively, $p < 0.01$, $d = 0.25$ and $p < 0.01$, $d = 0.23$).

As for Realism, a significant main effect was found for the factor technique ($F(1,687) = 59.22$, $p < 0.001$, $d = 0.4$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Realism compared to the stereo technique (respectively, $p < 0.001$, $d = 0.46$ and $p < 0.001$, $d = 0.33$).

Regarding Sound Quality, a significant main effect was found for the factor technique ($F(1,687) = 13.04$, $p < 0.001$, $d = 0.18$). The post-hoc tests revealed that for both standing and seated modalities, the binaural technique had higher scores of Sound Quality compared to the stereo technique (respectively, $p < 0.05$, $d = 0.18$ and $p < 0.05$, $d = 0.17$).

No significant main effects were found for Ease and Naturalness.

2) Head-Tracker Data: The quantity of motion analysis was performed in the same way as for Experiment 1. A significant main effect was found for factor modality ($F(1,687) = 10.34$, $p < 0.001$, $d = 0.13$), which indicates that participants moved their heads more when standing compared to when sitting. A significant main effect was found also for the factor technique ($F(1,687) = 18.59$, $p < 0.001$, $d = 0.18$). The post-hoc tests revealed that for both sit and standing modalities, the binaural technique had a higher quantity of motion compared to the stereo technique (respectively, $p < 0.01$, $d = 0.19$ and $p < 0.05$, $d = 0.18$).

C. Qualitative Results for Both Experiment 1 and 2

Participants' answers to the open-ended questions of the questionnaires in both experiments were jointly analyzed using an inductive thematic analysis [38]. Such analysis was conducted by generating codes, which were further organized into the following four subgroups (perception, action, emotion, and musical task) and their themes that reflected patterns.

1) Perception:
Usefulness of immersion and realism. Thirty-seven participants noticed that the binaural plus head-tracking system made them more immersed in the music scene, which created a greater sense of realism. They commented that such immersion allowed them to express their musical contributions in a better way (e.g., *"The immersion that is achieved with this system gives me that possibility to express my contribution as if I was really there along with the other musicians."; "Immersiveness is the key to understanding this system. With this possibility, I can feel inside the music scene, and I can hear all the sounds in a surrounding way as if the musicians were really around me."; "It is a more straightforward system that mimics real concert situations. I imagined myself inside a venue, feeling immersed with the other musicians."; "I felt immersed with other musicians like a real-life situation, such as a rehearsal room."*).

Benefit of wide spatialization. Thirty-six participants noted that the binaural with head-tracker system caused the perception of a broader space, with more significant spatial expansion in headphones, which is impossible to perceive with the conventional stereo system (e.g., *"I noticed that the space is extensive and more defined. This expansion, I think, is better because playing on it, I can really hear all the musical contributions of the other musicians with the broader space."; "The wide spatialization completely enveloped me, and I felt like I was on stage with my band. This spatial dilation made me sound much better." "The wide spatialized system imitates what happens in reality. Each musician's contribution gave the groove because each musician had their own highly recognizable position in space."*).

Instruments are "alive". Twenty-five participants commented that the binaural plus head-tracking system gives the instruments and the music scene a "life" (e.g., *"This system gives much more vitality to the music scene. The precise location and the movement of my head conditioning the position of the musicians give that added value to the music scene that I think is very important. Everything in the music scene seems more alive, more real."*).

Improved sound quality. Eighteen participants explicitly indicated that the overall sound quality was better in the binaural

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TOMASETTI AND TURCHET: PLAYING WITH OTHERS USING HEADPHONES: MUSICIANS PREFER BINAURAL AUDIO 7

plus head-tracking system compared to the stereo one (e.g., *"In this system, I could hear the harmonics better. Even though everything was more open in space, I felt that the overall sound quality was better."; "This system sounds divinely mixed. I can hear all the instruments very well, and the quality seems really good to me."*).

Interaction as an enhancement of entertainment. Sixteen participants described the binaural plus head-tracking system as interactive. By this term, they referred to the fact that the element of interactivity enhances the involvement and enjoyment in the act of playing together (e.g., *"The interactive system adds that poetic layer to playing at a distance that made me more involved and entertained."; "If I have to think about jamming, I would choose the system with interactivity because it is much more fun and engaging and tries to replicate what happens in the rehearsal room."*).

Improved dynamics perception. Fourteen participants reported that in the binaural plus head-tracking system, they could perceive more significant dynamics of all instruments (i.e., the volume variations), which led to a more accurate and intense listening experience (e.g., *"With this system, I could perceive the dynamic range of the instruments, especially the guitar. This dynamic is something that I have never been able to perceive with the classic system of headphone listening"; "With this system, I can acutely distinguish the differences between pianissimo and fortissimo. It is like there is a slow, rounded compression that seasons everything. It is much more pleasant to listen to."*).

Training period. Fifteen participants noted that before thoroughly enjoying the binaural plus head-tracking system, it took some time to familiarize and get used to it (e.g., *"It was amazing how my perception changed in the second repetition. The space seemed more realistic, and I was immersed differently. It was the habit I had gotten into with the system."; "Before a musician can completely enjoy the system and make his musical contribution, he has to get used to the new way of listening. Since we are used to mostly static music through headphones, getting used to the new system takes some time. Once a musician gets used to it, he can really notice the extraordinary beauty of playing on this immersive and interactive system."*).

Issues with sound movements. Thirteen participants commented that on some occasions, the movements of the instruments (which followed the head movements) were too fast, and their positions were sometimes perceived strangely. These two issues caused a certain discomfort (e.g., *"To play, I found better with the spatial one, even if it moved too fast, so in my opinion, it must be calibrated better."; "Small head movements caused sudden changes in the instruments' spatialization, and thus this caused the system to be artificial."; "A simple shift of my head would send the hi-hat of the drum from one ear to the other. Some sound movements are unnatural and do not reflect what happens in reality."*).

Influence of musical genre. Twelve participants stated that declaring their preference for one system over the other was highly influenced by the genre of music they were playing. In particular, they stated that with Rock, for example, the spatialized system was not as successful as in other genres because it is a characteristic of Rock to play very loudly. Moreover, this genre typically brings all instruments to the center, which are not easily distinguishable (e.g., *"The Rock genre is better with the stereo system. I want to have the Rock genre directly in the face because the most important thing about this genre is the loudness."; "The preference toward a system depends on the musical genre in which the musician plays. Funk and Blues are perfect for a spatialized system. However, I prefer Rock with more loudness and the sound coming to me as compact as stereo listening."*).

Narrow stereo image. Eleven participants commented that the instruments were poorly separated in the stereophonic system, and all the sound came from the center. Thus, the sound image was very narrow compared to binaural audio. These perceptions caused more difficulty in playing because the individual contributions of the other musicians could not be perceived correctly (e.g., *"I had to build the groove myself in the static system. The other musicians did not help me because they mixed up in the center."; "In this system, the keyboard was completely covered, and I could not really distinguish what chords it was doing."; "The drums hid the guitar...everything was so central that it created a very narrow, enclosed image for me."*).

Self-location difficulties. Six participants commented that they could locate other musicians precisely but manifested some difficulties in locating themselves (e.g., *"I can locate other musicians exactly, but I cannot locate myself. I can always hear myself everywhere I go, with the same volume."; "I found it strange that the other instruments were moving, but my bass always felt present, central. This thing in the rehearsal room does not happen. If I move, my perception of my bass should also change."*).

2) Action:

Body movements. Twenty-five participants commented that the binaural plus head-tracking system invites the musicians to move their bodies more during the musical performance, allowing even musicians who usually stand still to start moving to explore the music scene better (e.g., *"By moving more in the spatialized system, I was able to recreate the interplay between musicians that I usually perceive when I play in a concert. A musician has to move to benefit from this system to the fullest."; "If a musician uses this system as a consequence, he moves more because the system itself invites him to move, to discover things that normally, in headphones, it is impossible to discover."; "Usually, when I play, I do not move much. On the other hand, this system invited me to move, to browse."*).

Personalized spatialization and the possibility of choosing desired instruments. Twenty-one participants commented that they could find their custom spatialization, meaning that by turning in a particular position, they heard specific instruments better than others. In particular, they made home to the guitar and keyboard harmonies, the hi-hat, and the kick drum. They were very comfortable playing with the found personal spatialization because they could hear everything precisely the way they wanted. This more defined perception gave them greater confidence and stability in playing (e.g., *"I found a particular panning by turning to the left, like 45°. Once I found my panning, the one I thought was fascinating, I did not want to move from that position anymore because I could hear everything perfectly,*

and I was in comfort with the music scene."; "At one point, when I turned around, I started to hear the musicians differently, and as a result, it stimulated me to play differently. It was a beautiful experience."; "If I lose the rhythm, I turn to the drums and, hearing it better, I get it back. It is incredible to choose the instruments you want to listen to the most based on the necessity of the musical moment."; "According to how I positioned myself, I decided the ideal panning; this thing, in my opinion, is the novelty."; "It was nice because, at one point, I turned to the left, and it was like I had the keyboard player there, and consequently, I responded to him in musical terms in a different way. This action is exciting because I have more stimulation in playing by hearing everything more defined and localized."; "Turning towards the window, I noticed that I had a perfect spatialization and could hear the rhythmic pulse of the hi-hat perfectly, something that hardly happens to me. This precise perception gave me security and stability, allowing me to play more fluidly, spontaneously, and giving me a great sense of ease.").

Performative behavior. Nine participants emphasized that the binaural plus head-tracking system offers the possibility to behave during an actual performance and brings the concept of physical expressiveness into NMP. Even when playing at a distance, this interactivity and defined localization of musical sources make it possible to play more naturally, as it happens in live musical performances on large stages and concerts (e.g., *"Playing in NMP with this system, in my opinion, opens up the possibility of authentic performance. I am finally not sitting in my room, but I could move around and give physical expressiveness to my body like I do when I am on a stage."; "With the broader spacing system, I am much more elastic and feel much looser in playing. It is like I am playing live on a stage, and therefore I am also giving my best in performativity with the body.").*

3) Emotion:

Engagement and enjoyment. Thirty-four participants commented that their experience using binaural audio and head tracking was enjoyable, engaging, positive, and attractive. They reported to have played very actively compared to stereo listening through headphones. They stated that with the binaural system and the head tracker, the musician, being more immersed, feels more involved in the musical scene and consequently tries to play differently, creating phrasing, accompaniments, solo, and improvisations that they would not normally do (e.g., *"The involvement in this system is great. I felt like I was playing with the other musicians, like I was jamming in the rehearsal room."; "This system is really immersive, and personally, being a guitarist who improvises a lot, I think it is exciting because it makes me feel involved in the musical scene and more engaged in the performance with the other musicians."; "I had much fun with this system, I was wrapped up in the sounds, more involved, and time passed me by faster."; "This system is engaging and fun because it is possible to recognize the musicians' position, and also it has high quality. Consequently, it allows musicians to express themselves more naturally.").*

Feeling of social connection. Twenty-four participants reported that the binaural plus head-tracking system creates the illusion that they have social connections as if the other musicians are only really there around them at that precise moment (e.g., *"I felt like I had musicians around me. As I turned around, it was as if they were there, which is why I felt wholly enveloped as if I was really playing with them."; It was fantastic because I felt like I had the bass player beside me. With the phrasing I was doing on the clarinet, I responded to what the bass player was doing the whole time.").*

Stability of the stereo system. Six participants commented that the stereophonic system gave a sense of comfort and stability since it is the system they have the most experience with (e.g., *"The static system is stable anyway. I cannot say that I do not like it as I play every day over this system, giving me that sense of stability."; "The stereophonic sound gave me that sense of stability and security...I think it is related to the fact that I have always listened to music this way.").* Three participants commented that the stereo system sounded more direct and powerful, allowing them to play more naturally, without technical difficulties. This is because they did not have to get used to a new type of listening since they were already used to stereo listening *"Although the overall sound is flatter having the drums so powerful and direct has helped me feel comfortable and play more spontaneous."; "Because the separation between the left and right channels is clear and the sound is more direct, I can play naturally from the start because I have been used to this system since childhood."*).

Lack of visual feedback. Seven participants emphasized that without visual feedback, the binaural plus head-tracking system does not bring a significant added value to the NMP (e.g., *"When playing together we experience visual and musical experiences since music played together is made up of visible glances and anticipations. Without the visual dimension, the usefulness of the spatialized system is limited."; "To create a natural interplay in the music, I missed the visual feedback from the other musicians when playing live or in the rehearsal room.").*

4) Musical Tasks:

Localization as an added value in NMP. Thirty-eight participants reported that the precise localization of instruments in NMP is an added value because it creates a series of perceptions such as pleasantness, immersiveness, involvement, and more focus. These perceptions help improve attention as well as better perceive both the own musical contribution and that of other musicians. These attributes were also judged to help avoid the feeling of boredom when playing at a distance (e.g., *"I consider localization fundamental because by perceiving better the dynamics and the origin of the various instruments I can project myself more from a performance and emotional point of view."; "Localization is fundamental because it allows a musician to have his own space where he can spread his sound and at the same time be able to hear the other musicians perfectly."; "I think it is crucial to locate the musicians precisely because it is a matter of interplay, in a sense that if I want to go and find the drummer, as happens in reality, I turn to him and hear what he is playing."; "Localization is essential because, for example, for following the harmony, I follow the bass. Knowing that the bass player is in that space's position is a significant added value because it reassures me.").*

Improvisation and recreational music-making. Twenty-six participants stated that they would use the binaural plus head-tracking system to make recreational music together, to improvise together or over backing tracks (e.g., *"I would use that system for improvising over backing tracks because it powerfully simulates what it feels like when a musician plays with others on big stages or in rehearsal rooms."; "Suppose I have to jam, improvise, or have fun for a couple of hours. In that case, I will use the spatialized system because I can simulate interplay with other musicians and imitate our perception in real-life rehearsal room situations."; "I would use the spatialized system for all jamming, improvisation, or recreational music-making contexts. This preference is because, with that system, a musician has an extreme position's definition of all the instruments. It also allows a musician to level the technique with which a musician plays because it allows hearing all the instruments clearly and well separated."*).

Recording studio dimension. Eighteen participants commented that the stereophonic system is unbeatable if a musician needs to record his/her contribution for a record or production (e.g., *"With the static system, which I think is the classic stereo system, everything is so straight. That system cannot be beaten in recording studios because it represents the cultural coding we have had for almost a hundred years."; "This system clearly distinguishes between left and right channels and catapults me into the studio's recording dimension. In that dimension, it remains unbeatable."; "If I have to record, it would be more convenient to have the musicians fixed in the headphones because I have fewer distractions than the spatialized system."*).

Spatialization not fundamental for NMPs. Twelve participants commented that spatialization and exact localization of the various instruments was not a fundamental added value in NMPs. According to them, to play well in NMP, it is sufficient to have an equal distribution of sounds between the left and right channels, good overall sound quality, and no latency problems. Wide spatialization and localization would only be a slight improvement that they do not consider fundamental (e.g., *"There is no need for precise localization in the NMPs to play well because, at the moment, more significant problems have to be solved, such as latency, sound quality, and not seeing each other."; "I consider it a small improvement because many musicians are used to playing with not even two loudspeakers or with good home setups."*).

VI. DISCUSSION

Concerning Experiment 1, from the quantitative and qualitative results, it is evident that the vast majority of participants prefer the binaural audio with the head tracker compared to a classical stereo setup. This system was deemed to help engage, immerse, and entertain the musician to a greater extent, projecting him/her as if inside a space shared with other musicians. Furthermore, the binaural system with the head tracker was preferred for playing in NMPs, as well as practicing at home (i.e., improvising over backing tracks). However, the stereophonic system was deemed to be better suited for the recording studio dimension. Most of the participants stated that the stereophonic setup is a more stable system because every musician has grown up playing over stereo tracks and listening to stereo music.

Regarding Experiment 2, the obtained results were in line with those of Experiment 1. The binaural plus head-tracking system received for almost all dimensions higher evaluations than stereo (see Fig. 3). In particular, the binaural system led to a greater sensation of being together with other musicians and a better ability to localize them, as well as a higher degree of perceived realism and immersiveness in the acoustic scene. Moreover, a higher sound quality was perceived for the binaural conditions compared to the stereo ones.

Notably, The results concerning Connection indicate that the spatialization technique had an effect on the sense of being connected to others. Relatedly, the results regarding the preference for the binaural system over stereo are essential for the design of new NMP systems. Current NMP systems only provide a stereo mix of the sounds from each remotely connected musician [33]. To achieve a binaural spatialization, it is necessary to provide at each of the receivers' sides the independent channels related to each musician. Therefore, this technical issue represents a requirement for advancing the hardware and software components of NMP systems.

There was no substantial difference in the dimensions of Ease and Naturalness. Participants reported that they did not notice any significant difference in these dimensions as these are related to their own musical contributions and to their way of playing. They reported to have felt comfortable when using both systems. They could play naturally and at ease without any difficulty because they could clearly hear and recognize their own musical contribution and distinguish it from those of the other instruments.

Interestingly, in both experiments, on average participants moved their heads to a greater extent in the binaural condition than in the stereo one. The fact that participants moved more in the binaural conditions could be explained in different ways. First, this phenomenon could be caused by the benefits added by binaural spatialization as indicated by the evaluations along the investigated dimensions (including realism, localization, and social presence) as well as by the qualitative results emerged from the thematic analysis. As such, binaural spatialization could have improved the overall experience, increasing engagement and arousal, and as a consequence corporeal articulation.

It is also possible to explain such results in terms of the theory of embodied music cognition, which claims that bodily involvement is crucial in human interaction with music and, therefore, also in our understanding of this interaction [39]. Thus, in light of this theory, more motor engagement with the binaural conditions possibly increased participants' overall understanding of the acoustic scene. What distinguishes stereo from binaural audio playback is essentially related to the relationship that users have with these kinds of displays. For stereo, users might take a third-person perspective and need to adapt themselves to the display. For binaural, users take a first-person perspective where the display adapts automatically to the user's behavior. In terms of embodied experience, this makes a lot of difference, in particular, related to the coupled action-perception mechanisms at work [40], [41]. Therefore, an explanation for our finding

might be that the first-person perspective triggered musicians to explore the possibilities of the binaural display more extensively, as if they were "sampling" the acoustic environment to figure out action–perception relationships.

Notably, musicians' preferences for the binaural spatialization emerged for some of the possible use cases investigated, but not for all. The preferences were confirmed for remote collaborative playing using an NMP system, individual instrumental practice (such as learning improvisation over backing tracks), and individual recreational music-making using backing tracks. On the contrary, musicians reported to prefer the conventional stereo setup for studio recording sessions. This is not in agreement with the results reported in Bauer et al.'s study [23], which employed a binaural spatialization without head tracking. This discrepancy might be due to the fact that the head tracking adds a new perceptual value that many musicians have not experienced and, therefore, might cause a sort of distraction in a recording session compared to the stereo system's stability, as some participants reported in their comments.

The results of our study are in line with those reported previously [24], which also show that the binaural system improves the comfort of performers' listening, perception of a realistic sound image, creativity, musical expression, and sense of immersion. However, our results are not consistent with those reported in a study by Morell and Lee [25], where binaural mixes were rated lower than stereo mixes. This disagreement could be related to the perception that changes significantly in a listening situation compared to a situation in which a musician is playing.

It is worth noticing that this study has some limitations. First, most of the participants were Italians, most of them were males, and a relatively limited number of musical instruments and musical genres were involved. A wider set of musicians from different countries and cultures, a more balanced gender distribution, as well as a more diversified pool of musical instruments and genres involved would confer the results with a higher level of generalizability. Second, generalized HRTFs were involved. Third, our study did not use the spatial parameter of elevation in the binaural system and no reverberation was added. Moreover, we did not apply filtering to make headphones flat. It is plausible to expect that even better results for the binaural audio system could have been achieved using personalized HRTFs, the elevation parameter, and the equalization filtering. Furthermore, the use of reverberation can improve distance perception, externalization, and localization accuracy [25], [42]. Finally, musicians were not playing together over the actual network using an NMP system, but such a condition was de facto simulated. Moreover, no visual representation of the connected musicians was provided. Investigating such a condition would allow studying musicians' preferences and experiences in multisensory contexts.

VII. CONCLUSION

This article compared musicians' preferences and experiences for collaborative playing with a binaural plus head-tracking system versus a conventional stereo setup. The results reported in the two conducted experiments consistently indicate that musicians prefer playing their instrument in conjunction with binaural spatialization (combined with head tracking) of other musicians' sounds. Nevertheless, such a preference appears to be relevant for some scenarios, but not for all. The preference for a binaural spatialization was confirmed for remote collaborative playing using an NMP system, individual instrumental practice (such as learning improvisation over backing tracks), and individual recreational music-making using backing tracks. Conversely, in the case of studio recording sessions, musicians reported to generally prefer the conventional stereo setup.

Our results also showed that musicians moved more their heads during the binaural conditions compared to the stereo ones. This phenomenon could have different explanations. First, the higher engagement and arousal due to the improved auditory experience. Second, based on the tenets of embodied music cognition theory, the different relationship users have with the two kinds of systems, where the first-person perspective caused by the binaural display would inherently lead to a higher degree of exploration to better understand the action–perception loop.

Moreover, the findings reported in this study entail the need for progressing current systems used by musicians while playing over headphones. First, musicians could be empowered to use a binaural system, which could also account for the personalization of the position of the sound source representing the connected musicians. Second, our results have strong implications for the design of NMP systems, which need to be extended at the hardware and software level to provide independent channels in order to support the binaural spatialization of remotely connected musicians.

Several avenues are possible for future work to extend the results of this study. First, we plan to implement a novel NMP system supporting at least four independent channels, which will enable us to perform a comparison between binaural and stereo spatializations when remotely connected musicians actually play together. Second, we plan to integrate such a system with an augmented reality setup to provide a fully immersive experience, including a visual representation of the other musicians. This will enable us to investigate the role of binaural audio in a multisensory setup. Finally, we plan to conduct a study where the different reverberations and the simulation of acoustics of different rooms (also with the integration of binaural room impulse response) will be investigated. We aim to investigate the perceptual differences of musicians concerning stereophonic and binaural setups with and without the addition of reverb, which should help further improve localization, sense of immersion, and involvement in binaural systems.

REFERENCES

[1] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality.* Berlin, Germany: Springer Nature, 2019.

[2] M. Naef, O. Staadt, and M. Gross, "Spatialized audio rendering for immersive virtual environments," in *Proc. ACM Symp. Virtual Reality Softw. Technol.*, 2002, pp. 65–72.

[3] J. Susal, K. Krauss, N. Tsingos, and M. Altman, "Immersive audio for VR," in *Proc. Audio Eng. Soc. Conf.*, 2016, pp. 1–8.

[4] S.-N. Yao, "Headphone-based immersive audio for virtual reality headsets," *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 300–308, Aug. 2017.

[5] T. Chatzidimitris, D. Gavalas, and D. Michael, "SoundPacman: Audio augmented reality in location-based games," in *Proc. 18th Mediterranean Electrotechnical Conf.*, 2016, pp. 1–6.

[6] X. Amatriain, T. Höllerer, J. Kuchera-Morin, and S. T. Pope, "Immersive audio and music in the AlloSphere," in *Proc. Int. Comput. Music Conf.*, 2007, pp. 276–283.

[7] Apple music announces spatial audio with Dolby Atmos; will bring lossless audio to entire catalog, May 2021. [Online]. Available: www.apple.com/newsroom/2021/05/apple-music-announces-spatial-audio-and-lossless-audio/

[8] G. Eckel, "Immersive audio-augmented environments: The listen project," in *Proc. 5th Int. Conf. Inf. Visualisation*, 2001, pp. 571–573.

[9] V. K. Wu and R. H. Campbell, "3D audio interface for rich mobile web experiences," in *Proc. Int. Conf. Mobile Comput., Appl., Serv.*, 2010, pp. 1–16.

[10] T. Walton, "The overall listening experience of binaural audio," in *Proc. 4th Int. Conf. Spatial Audio*, 2017, pp. 170–177.

[11] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300–321, 1995.

[12] C. C. Berger, M. Gonzalez-Franco, A. Tajadura-Jiménez, D. Florencio, and Z. Zhang, "Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity," *Front. Neurosci.*, vol. 12, 2018, Art. no. 21.

[13] A. Kulkarni, S. Isabelle, and H. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoustical Soc. Amer.*, vol. 105, no. 5, pp. 2821–2840, 1999.

[14] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.

[15] P. Strumillo, *Advances in Sound Localization*. Paris, France: BoD–Books on Demand, 2011.

[16] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a head-related transfer function database with high spatial resolution," in *Proc. 7th Forum Acusticum*, 2014, pp. 1–6.

[17] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic HRTF database," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2001, pp. 99–102.

[18] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoustical Soc. Amer.*, vol. 110, no. 5, pp. 2440–2448, 2001.

[19] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, "HRTF database at FIU DSP lab," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 169–172.

[20] O. Rummukainen et al., "Listening tests with individual versus generic head-related transfer functions in six-degrees-of-freedom virtual reality," in *Proc. Int. Conf. Spatial Audio*, 2019, pp. 55–62.

[21] P. Stitt, L. Picinali, and B. F. Katz, "Auditory accommodation to poorly matched non-individual spectral localization cues through active learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, 2019.

[22] C. Faller and J. Breebaart, "Binaural reproduction of stereo signals using upmixing and diffuse rendering," in *Proc. Audio Eng. Soc. Conv.*, 2011, pp. 1–8.

[23] V. Bauer, H. Déjardin, and A. Pras, "Musicians' binaural headphone monitoring for studio recording," in *Proc. Audio Eng. Soc. Conv.*, 2018, pp. 1–10.

[24] V. Bauer, D. Soudoplatoff, L. Menon, and A. Pras, "Binaural headphone monitoring to enhance musicians' immersion in performance," in *Adv. in Fundam. and Applied Res. on Spatial Audio*, B. F. Katz and P. Majdak, Eds. Rijeka, Croatia: IntechOpen, 2022.

[25] P. A. Morell and H. Lee, "Binaural mixing of popular music," in *Proc. Audio Eng. Soc. Conv.*, 2021, pp. 414–418.

[26] B. F. Katz and L. Picinali, "Spatial audio applied to research with the blind," in *Advances in Sound Localization*. Rijeka, Croatia: InTech, 2011, pp. 225–250.

[27] M. Frank, F. Zotter, and A. Sontacchi, "Producing 3D audio in ambisonics," in *Proc. Audio Eng. Soc. Conf. Audio Eng. Soc.*, 2015, pp. 1–8.

[28] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proc. DAGA*, 2018, pp. 339–342.

[29] M. Cuevas-Rodríguez et al., "3D tune-in toolkit: An open-source library for real-time binaural spatialisation," *PLoS One*, vol. 14, no. 3, 2019, Art. no. e0211899.

[30] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Proc. Audio Eng. Soc. Conf.*, 2019, pp. 1–12.

[31] Y. Ueno, M. Mizumachi, and T. Horiuchi, "Perceptual evaluation of binaural rendering and stereo width control in headphone reproduction," in *Proc. Audio Eng. Soc. Conv.*, 2020.

[32] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical Sci. Technol.*, vol. 35, no. 3, pp. 159–165, 2014.

[33] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[34] C. Drioli, C. Allocchio, and N. Buso, "Networked performances and natural interaction via LOLA: Low latency high quality a/v streaming system," in *Proc. Int. Conf. Technol. Performing Arts, Media Access, Entertainment*, 2013, pp. 240–250.

[35] J. Cáceres and C. Chafe, "JackTrip: Under the hood of an engine for network audio," *J. New Music Res.*, vol. 39, no. 3, pp. 183–187, 2010.

[36] L. Turchet and C. Fischione, "Elk audio OS: An open source operating system for the internet of musical things," *ACM Trans. Internet Things*, vol. 2, no. 2, pp. 1–18, 2021.

[37] M. Wright, A. Freed, and A. Momeni, "OpenSound control: State of the art 2003," in *Proc. Conf. New Interfaces Musical Expression*, 2017, pp. 125–145.

[38] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.

[39] M. Leman, P.-J. Maes, L. Nijs, and E. Van Dyck, "What is embodied music cognition?," in *Springer Handbook of Systematic Musicology*. Berlin, Germany: Springer, 2018, pp. 747–760.

[40] M. Leman and P.-J. Maes, *Music Perception and Embodied Music Cognition*. Evanston, IL, USA: Taylor & Francis, 2014, pp. 81–89.

[41] I. Findlay-Walsh, "Hearing how it feels to listen: Perception, embodiment and first-person field recording," *Organised Sound*, vol. 24, no. 1, pp. 30–40, 2019.

[42] L. Picinali, A. Wallin, Y. Levtov, and D. Poirier-Quinot, "Comparative perceptual evaluation between different methods for implementing reverberation in a binaural context," in *Proc. Audio Eng. Soc. Conv.*, 2017, pp. 1–7.
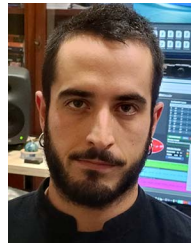
Matteo Tomasetti received the B.A. degree in electronic music and sound design from the Conservatory of Pesaro, Pesaro, Italy, in 2017, and the M.A. degree in audiovisual composition, 3-D audio, and virtual reality from the Conservatory of Frosinone, Frosinone, Italy, in 2021, and the IEM (Institut für Elektronische Musik und Akustik), Graz, Austria, with a thesis focused on interactive music composition experiences in VR environments. He is currently working toward the Ph.D. degree in musical XR and spatial audio with the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy.

His research interests include extended reality, music composition, human–computer interaction, and immersive audio.

Luca Turchet (Senior Member, IEEE) received the master's degrees (*summa cum laude*) in computer science from the University of Verona, Verona, Italy, in 2006, in classical guitar and composition from the Music Conservatory of Verona, Verona, and in electronic music from the Royal College of Music of Stockholm, Stockholm, Sweden, and the Ph.D. degree in media technology from Aalborg University Copenhagen, Copenhagen, Denmark, in 2013.

He is currently an Associate Professor with the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. His scientific, artistic, and entrepreneurial research has been supported by numerous grants from different funding agencies including the European Commission, the European Institute of Innovation and Technology, the European Space Agency, the Italian Minister of Foreign Affairs, and the Danish Research Council. He is the co-founder of the music-tech company, Elk. His main research interests include music technology, Internet of Things, human–computer interaction, and perception.

Dr. Turchet is an Associate Editor for IEEE ACCESS and the *Journal of the Audio Engineering Society*.