Article:

Cairns, Patrick, Hunt, Anthony, Johnston, Daniel et al. (4 more authors) (2023) Evaluation of Metaverse Music Performance With BBC Maida Vale Recording Studios. Journal of the Audio Engineering Society. pp. 313-325. ISSN: 0004-7554

https://doi.org/10.17743/jaes.2022.0086

Evaluation of Metaverse Music Performance with BBC Maida Vale Recording Studios*

PATRICK CAIRNS,[1] *AES Student Member,* (patrick.cairns@york.ac.uk)
ANTHONY HUNT,[1] *AES Student Member,* (anthony.hunt@york.ac.uk)
DANIEL JOHNSTON,[1] *AES Member,* (daniel.johnston@york.ac.uk)
JACOB COOPER,[1] *AES Student Member,* (jacob.cooper@york.ac.uk)
BEN LEE,[1] *AES Associate Member,* (benjaminlee1297@gmail.com)
HELENA DAFFERN,[1] (helena.daffern@york.ac.uk)
GAVIN KEARNEY,[1] *AES Member* (gavin.kearney@york.ac.uk)

[1]*AudioLab, School of Physics Engineering and Technology, University of York, York, UK*

This paper details a case study evaluation of a recording experience in a networked XR simulation of the renowned BBC Maida Vale Recording Studios. The system allows multiple remote musicians to connect over a network, providing a shared virtual acoustic space, with interactive immersive audio, XR display, and low-latency throughput. A four-piece rock band used this system in a live recording session, performing under different latency and audio conditions. Technical setup and case study protocol is detailed. Evaluation is provided in the form of Quality of Experience rating, tempo analysis, and a semi-structured exit interview.

0 INTRODUCTION

The metaverse [1] presents a new frontier for the music industry, where online virtual spaces may be accessed remotely using Virtual and Extended Reality (VR/XR) technology and used to engage in individual or group musical experiences. Opportunities presented by the metaverse extend beyond the delivery of pre-produced content to include streaming of live and interactive performances taking place in real-time, and delivered in shared virtual spaces.

Music industry interest in the opportunities presented by the metaverse is demonstrated by the engagement of mainstream artists in large-scale live music events, and the engagement of online communities on mainstream social VR platforms. Examples of live metaverse concert include 360° video events, such as the Foo Fighters Super Bowl concert[1]. Other examples merge live video with virtual elements, such as Jean-Michel Jarre's 'Welcome to the Other Side' experience in a virtual re-imagining of Notre Dame[2]. Entirely virtual live events also take place in popular multiplayer game spaces, such as Marshmello's Fortnite concert[3]. Live performances are also delivered regularly in metaverse such as Sansar[4] and social VR platforms like AltSpaceVR[5].

Beyond the delivery of live performances to online audiences, the metaverse also offers the opportunity for shared virtual spaces to be used for the live creation of musical content by remote musicians [2, 3]. Such work falls into the category of Network Music Performance (NMP) [4]. Consideration of audio is paramount to the integration of VR/XR within an NMP context. Audio latency is important to the NMP experience, where high latency impairs the ability of remote performers to play in time with one another. The introduction of new modality presented by VR/XR systems has the potential for interplay with latency perception. This is true with respect to visual performance timing cues (which are expected to influence performance timing in live [5], video [6] and XR [7] cases), and with respect to the potential for virtual acoustics to affect collaborative music-making experiences [8]. As such, further exploration of the NMP experience and effect of latency is prudent in this context.

This paper expands on the presentation of an online XR recording studio experience, focusing on the implementation of a simulation of Maida Vale Studio 4 (MV4) [9]. The system allows performers to immerse themselves inside a virtual model of MV4 using a Head Mounted Display (HMD) and low-latency immersive audio to see and

*To whom correspondence should be addressed, e-mail: patrick.cairns@york.ac.uk. Last updated: Oct 11, 2022

[1]https://www.oculus.com/blog/catch-foo-fighters-in-vr-horizon-venues-concert-to-air-february-13-after-the-big-game/

[2]https://jeanmicheljarre.com/live/welcome-to-the-other-side

[3]https://fortnite.fandom.com/wiki/Showtime

[4]https://events.sansar.com/

[5]https://account.altvr.com/events/main

hear each other as if they were in the same space, while being physically located at different endpoints over a network. XR display allows performers to see both their instruments in the real world, and the virtual environment and networked performers' avatars in the virtual space. A four piece rock band, *Wolph*, participated at a case study to evaluate the XR performance experience.

1 RELATED WORK

1.1 NMP and Latency

NMP systems allow remote musicians to play together by transporting audio across networks. The transport of audio, as well as other audio throughput contributions, such as hardware capture and playback buffers or plugin process buffers, incur latency between the remote performers. NMP research has demonstrated that excessive delay results in an impairment of the ability of an online ensemble to perform together synchronously [4]. The level of delay at which performance becomes impaired is typically estimated in the range of 20-30ms one-way throughput between performers [4]. This is demonstrated by previous tapping studies [10, 11]. Above this threshold the effect of latency is generally characterised by tempo deceleration (though tempo deceleration can also occur at lower delays [12]) and poor synchrony [4].

It is possible to maintain stable synchrony above the 20-30 ms threshold. This requires the performance to be adapted to counteract the effect of latency [4, 10, 13], or the application of a tempo descriptor such as a metronome [14]. Adaption to the effects of latency is generally considered to improve with practice [4, 15, 13]. The experience of latency has some dependencies. Timbral qualities of instrumentation can influence the experience, where the spectral flatness appears to result in more acute influence of latency on performance [16]. Dynamics are also relevant, with slow attack times being associated with more stable tempi ratio between performers than fast attack times [17]. Effects of tempo and event density in performance are also suggested as influencing the NMP experience, where higher values are associated with more acute influence of latency on performance [16].

1.2 NMP and Visual Display

NMP systems which provide video display and interface (or *telepresence* systems) acknowledge the importance of visual timing cues in musical performance. The transport of video data will typically involve high delays [18]. High video delay can result in noticeable asynchrony between video and audio display [6], which may result in synchrony issues [6], and can lead to networked musicians ignoring video display during performance [19]. Where telepresence NMP systems attempt to avoid these issues using low-latency video [18, 20] it is acknowledged that bandwidth requirements are beyond what consumer internet can typically support.

VR/XR HMD-rendered visual display presents the potential for similar issues with respect to visual latency and audiovisual asynchrony. Experiments with point-cloud capture and rendering of remote performers in an XR NMP context [7] conclude that the approximately 400 ms video latency experienced on a Local Area Network (LAN) caused audiovisual asynchrony issues for musicians, making this approach unusable for NMP. Other XR NMP work used a motion capture and avatar animation approach to visual display and interface [9, 21]. This is useful in practical Wide Area Network (WAN) contexts, as avatar control data may be transported with low bandwidth and latency, and is not sensitive to packet loss. An XR performance system [21] which uses OptiTrack[6] motion capture for animation of networked musician avatars measured motion-to-photon latency of approximately 74 ms on LAN. It was considered that the visual display and interface may enhance the sensation of presence in the experience despite visual latency and audio-visual asynchronies.

VR NMP has been evaluated for delivery of therapeutic group singing [22] and was found to be therapeutically useful for spinal cord injury patients, and deliver a sense of presence in the experience. The limitations of avatar rendering with respect to social cues such as facial expression, and the delivery of visual timing cues for group singing, was also highlighted. Other VR NMP evaluation has provided introductory Quality of Experience (QoE) ratings comparing audio-only, telepresence and VR visual display using an immersive audio system [23], in which the importance of being able to see real instruments is discussed, highlighting the usefulness of XR NMP systems. Preliminary QoE results suggest that VR conditions may be preferable to telepresence systems as an NMP visual modality, and that the immersive audio rendering gave the best spatial impression of the virtual performance space when combined with VR rather than telepresence.

1.3 NMP and Immersive Audio Display

Using immersive audio rendering on top of NMP audio streaming is another method of providing a shared virtual space for online performance, here through interactive auditory display and interface. This may be implemented with no visual display and interface [8, 24, 25, 26], or alongside telepresence [23, 27, 28, 29] or VR/XR [9, 21, 23] visual display and interface. This is particularly relevant to VR/XR NMP with respect to the importance of coherence between audio and visual display and interface. Immersive audio methods vary in NMP, however current work appear to favour Ambisonic approaches (though some systems appear to use purely binaural approaches [11, 26]).

Ambisonic encoding of direct sound is used to spatialise sources in several cases [24, 27, 28]. Other Ambisonic systems provide virtual room simulation through convolution with measured Spatial Impulse Response (SIR) [8, 23, 25], or by modelling of acoustics via analysis of IR capture [29]. In other cases a hybrid approach is used, such as rendering direct sound using Ambisonic encoding methods,

[6] https://optitrack.com/

and rendering reverberation through SIR convolution [9]. Image source modelling of early reflections with convolution for late reverb is used in one case to auralise sources in acoustically coupled spaces [21]. Though audio may be rendered over loudspeakers [24], the Ambisonic approach is useful in the VR/XR NMP context as it provides efficient 3 Degrees of Freedom (3DoF) through sound field rotation and decoding to binaural. Binaural playback is specified for VR/XR NMP as headphones are the most common audio delivery method, and as headphone playback incurs no additional air propagation delay [9].

The immersive audio context introduces new variables to the NMP experience, namely spatial perception of source and room attributes [30]. Simulation of virtual acoustic spaces is expected to influence the ensemble performance experience, as is known to be the case in live performance [31]. Indeed, in a recent study using a system for VR singing [32] it was shown that virtual reverberant conditions were almost always preferred to virtual anechoic conditions, and that there was a majority preference for dynamic virtual acoustics over static virtual acoustics.

Investigation of reverberation in NMP contexts has, however, shown varying results. In one study drummers compare non-reverberant and 'artificial reverb' conditions at different network delays [33]. Here it is reported that in the case where latency is high, and considered to impair performance quality, a preference is stated for non-reverberant conditions, claiming better clarity in note onsets. In contrast, tapping experiments comparing virtual anechoic conditions with virtual reverberant conditions rendered with static Binaural Room Impulse Responses (BRIR) demonstrate improved performance synchrony in reverberant conditions [11].

Evaluation of immersive audio NMP over WANs provided comparison of performance synchrony between duet singers in different virtual rooms, and observed more stable tempi ratio in a medium hall simulation compared to studio booth and cathedral alternatives [8]. With respect to spatialisation factors, one study conducted with percussive instrumentation suggests that source panning to provide seperation may influence better performance synchrony in NMP contexts [34]. Though the effect of immersive audio and virtual acoustics on the NMP experience remains ambiguous, it is certainly the case that further research is required, especially for VR/XR NMP systems.

2 EXPERIMENT

2.1 Study Overview

To investigate the NMP experience with immersive audio XR systems, a case study was undertaken of a live band recording with a XR NMP system which simulates MV4 of BBC Maida Vale Studios. This was done using a previously developed system [9] based upon measurement of the real space [35]. A four piece rock band, *Wolph*, were invited to participate. The band consists of Lead Vocals/Keys (keys), Vocals/Guitar (guitar), Vocals/Bass synthesiser (bass), and Vocals/Drums (drums). These performers had participated in a previous case study using the XR NMP system at Maida Vale Studios [9].

The study was designed in order to evaluate the XR NMP experience under different latency conditions, and to evaluate the immersive audio rendering system relative to mono baseline audio conditions. Evaluation is provided in QoE rating from the band, tempo analysis of the kick drum from recordings, and a semi-structured exit interview. Each of the four band members were isolated in different rooms (Figure 1), and used the XR NMP system (Figure 3) to undertake a musical performance under different latency and audio system conditions. All analogue inputs were recorded to a studio console for controlled recording. Though the system was in this case deployed over LAN, configuration was the same as over WAN (minus port-forwarding), and latency conditions reflect those achievable with WAN deployment.

2.2 XR Networked System

A prototype implementation of the system was made using open source technology, original environment and avatar models, acoustic measurements of the real space of MV4, and an XR application. The technical details of the system [9] and acoustic survey of Maida Vale Studios [35] are available in prior publication.

The system specifies a low-latency User Datagram Protocol (UDP) streaming method for audio transport, which is common in NMP technology. Avatar control data is transported using the Open Sound Control (OSC) protocol [36]. Avatar control data at each endpoint is streamed from HMD to a bridge application hosted on a local desktop. The bridge application passes head tracking data to local audio rendering processes, and to instances of the bridge application at remote endpoints to be forwarded to remote HMDs for avatar animation.

The XR application provides a virtual environment consisting of a model of MV4, avatars which are customised resemble band members, and pass-through 'windows'. The 'windows' are rectangular game objects which can be spawned, resized and placed as an overlay on the virtual environment. Through these 'windows' performers see pass-through display, allowing them to see their instruments in a real space while inhabiting the virtual environment.

Immersive audio rendering provides a performance scene from the perspective of each performer, auralised to simulate the acoustics of MV4. This is achieved using a combination of Ambisonic methods. Direct sound is auralised using an Ambisonic encoding approach, with the exception of the performers own vocals, where analogue direct monitoring is used. It is noted that direct sound auralisation does not model source directivity. Reverberation is rendered by convolution with measured 3rd Order Ambisonic (3OA) SIR. Ambisonic rotation is applied to provide 3DoF. Binaural decoding for delivery over headphones is achieved using the virtual loudspeaker approach. Contrary to system description in [9] the system implementation in this study rotates singers own vocals reverberation, and uses no separate late reverb convolution.

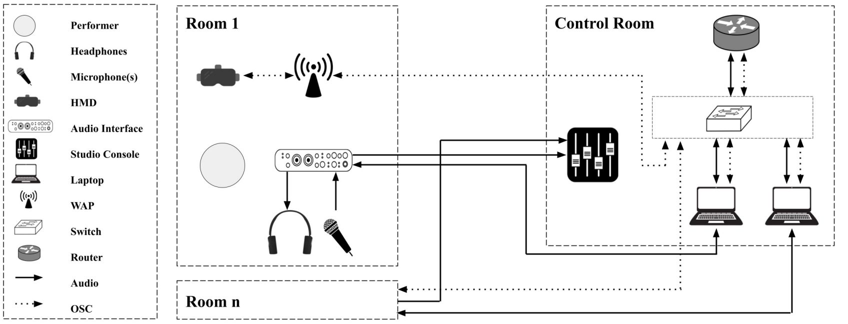

Fig. 1. Equipment setup for the XR MV4 component of the Maida Vale case study.

2.3 Hardware and Software Resource

Meta Quest 2 HMDs with the XR application were provided to the each performer. *ASUS RT-AX 55* routers were used as Wireless Access Points (WAPs) for streaming of avatar control data from HMDs, and as the router for the LAN. Audio rendering processes at each endpoint were hosted on a laptop (*i7-11800H, 2.3 GHz, 32GB RAM*). Audio interfaces (*Focusrite Scarlet 18i20 and 2i2*) were used for audio capture and playback. Microphones for audio capture were: *Neumann u87* (Lead guitar vocals) *Rode NT2a* (Bass vocals), *Shure SM58* (Drum vocals), *Shure Beta 52a* (kick), *Shure SM57* (snare), *Rode M5* (hi hat), *Sennheiser e604* (toms) and *AKG C414* (overheads). Keyboard, bass synthesiser and guitar were captured via Direct Injection (DI) boxes. Audio was played back over *Beyerdynamic DT770* headphones.

JackTrip [37] provided audio streaming functionality between endpoints, transporting individual instrument channels. Jack Audio Connection Kit (JACK)[7] was used to route audio between JackTrip, Reaper[8], and hardware buffers. Audio rendering processes were hosted in Reaper. Ambisonic encoding was accomplished using AmbiX Encoder [38]. Convolution was computed using X-MCFX Convolver[9]. Ambisonic rotation was performed using IEM Scene Rotator[10]. Binaural decoding was achieved using a dual band weighting filter[11], and AmbiX Binaural Decoder [38]. HRTFs were sourced from the SADIE II database [39] (KU100 measurements), as was the virtual loudspeaker configuration (26-point lebedev grid). 3OA SIR used in auralisation were sourced from an acoustic survey of BBC Maida Vale Recording Studios [35].

2.4 Evaluation Conditions

Latency conditions describe one-way full system throughput at 19 ms, 24 ms and 29 ms. Latency is calculated based upon the summing of hardware, processing, and transport latency. Hardware latency incurred in the capture and playback buffers was a static value of 5.3ms (48 kHz, 128 samples). Immersive audio rendering incurs a static latency of 1ms (a result of a pre-ring on HRTFs, with zero processing latency). JackTrip was configured to connect between endpoints in a peer-peer mesh, with a jitter buffer sized at 4 blocks of 128 samples, and Forward Error Correction redundancy of 2 packets.

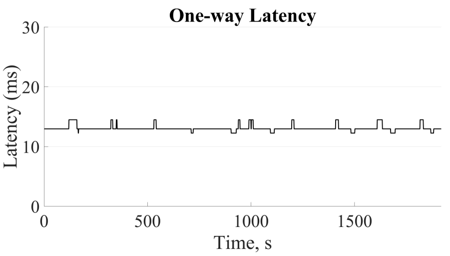

Fig. 2. Exemplary one-way latency estimate from measured RTT.

A half-hour exemplary latency measurement (Figure 2) was performed using JackTrip in this configuration, and suggests latency of 13.0 ms. This was estimated by measuring the Round Trip Time (RTT) with a 1 Hz pulse train and halving the measured delay to estimate the one-way latency. An error of +1.5/-0.8 ms was measured in this latency trace. This error is below any Just Noticeable Difference (JND) and is considered negligible in this study. System throughput is therefore measured at a baseline of 19 ms (rounded from 19.3 ms). Incrementation of latency was performed by adding delays on remote performer inputs at each endpoint.

No clock synchronisation was applied between endpoints. This can cause infrequent periodic audio glitches [40] and result in minor latency variance [41]. GPS-derived clock synchronisation is viable [40], however requires additional equipment which may not be available in common NMP practice, considering accessibility to a typical home user. Head-tracking and motion to photon latencies are unknown in this study, however are assumed relatively constant, with the only point of jitter being wifi data transport between HMD and laptop.

Audio conditions were varied between immersive audio conditions as rendered by the XR NMP system, and baseline mono audio conditions. Immersive audio conditions

[7] https://github.com/jackaudio
[8] https://www.reaper.fm/
[9] https://github.com/JB-Luke/X-MCFX
[10] https://git.iem.at/audioplugins/IEMPluginSuite
[11] https://github.com/trsonic/dual-band-filter

Fig. 3. The band *Wolph*, captured live while using the XR MV4 system, and represented as avatars in the virtual environment.

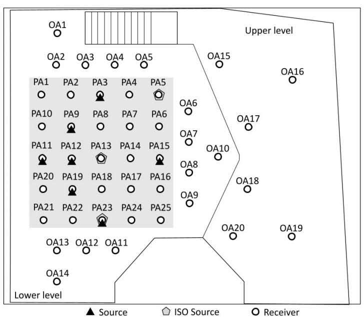

Fig. 4. Floor plan of MV4 outlining performer positions PA03, PA11, PA15 and PA23. For further detail see [35].

use directivity processed SIR measured in the real space of MV4 [35], which match the positions of performers in the virtual environment as detailed in Figure 4. Positions PA03, PA11, PA15 and PA23 represent the positions of the performers for bass, drums, keys and guitar respectively. Positions PA09, PA12 and PA19 represent a 2D plane which is used to position drum sources for auralisation of reverberation as detailed in [9]. ISO-3382 standard [42] acoustic measurements for MV4 can be found in [35]. Although a metronome was previously used in the XR NMP system [9], in order to observe performance without the stabilising influence on tempo this study did not use a metronome.

2.5 Testing Protocol

The band *Wolph* performed the single *"Shoulders of Giants"* for studio recording using the XR NMP system. This task was repeated 3 times in each discrete combination of latency and audio conditions in a randomised sequence. The band are professional musicians with previous experience using the XR NMP system [9], and practised during a 'sound check' before the protocol began. After each performance, each performer provided response to a QoE questionnaire. Audio recording of each performance was captured. The QoE items being rated by performers were:

1. We can play together as though in the same room.
2. I feel like I am in the same space as the other performers.
3. I have a sense of where other musicians are located in the space.
4. The other players sound as though they are playing where I see their avatars.
5. I enjoyed the virtual acoustics during performance.
6. The audio quality was of a professional standard.
7. I am conscious of audio delay between performers.
8. I engaged with the virtual environment and avatars.
9. I engaged with the pass-through display during performance.

These were rated on a 7 point Likert scale, with 1, 4 and 7 corresponding to 'strongly disagree', 'neutral' and 'strongly agree' respectively. The items evaluate aspects of the experience, including: NMP QoE items [13, 15, 43] such as perception of latency and synchrony: spatial audio qualities [30, 44], such as localisation and room perception; and immersive experience items [45] such as Presence. The band participated in a semi-structured exit interview[12] addressing similar topics and encouraging wider discussion. Parallel to this study data was collected for a discrete tapping study. The influence on the testing protocol is that prior to conducting each performance to which QoE rating relates, participants took part in a brief clapping exercise. Each performer therefore had brief experience of the audio and latency conditions for each take prior to each performance to which QoE ratings relate.

3 RESULTS

QoE ratings were evaluated by comparison of means across groups, with audio system and latency as within-subject factors, and musician as a between-subjects factor. The sample size in this case study (3 repetitions for each discrete combination of factors) is too small for statistical analysis. It should be noted that variance of means in this case study represents casual observation of patterns in QoE response. Another limitation of this study is that the variable 'musician' may equally relate to the discrete individual or the discrete instrument associated with each group within this factor.

[12] https://github.com/SpaceCadetAlba/MVXRJAES

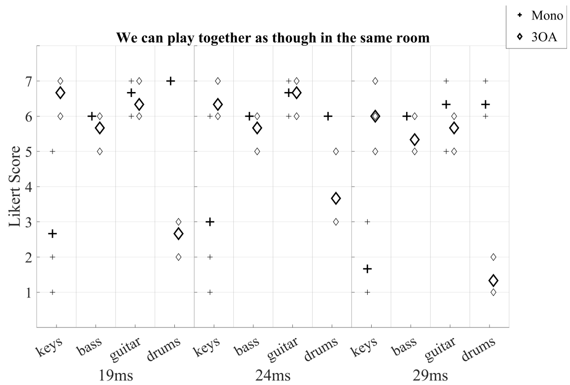

Fig. 5. Likert response to questionnaire item 1.

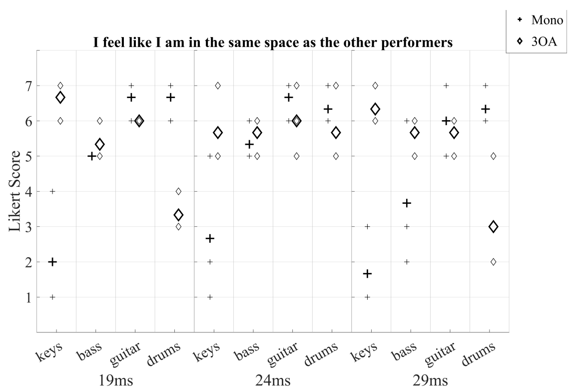

Fig. 6. Likert response to questionnaire item 2.

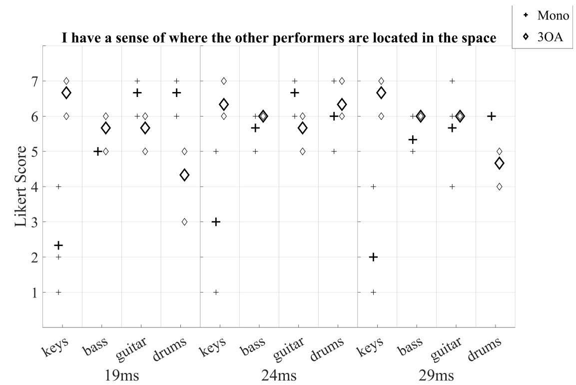

Fig. 7. Likert response to questionnaire item 3.

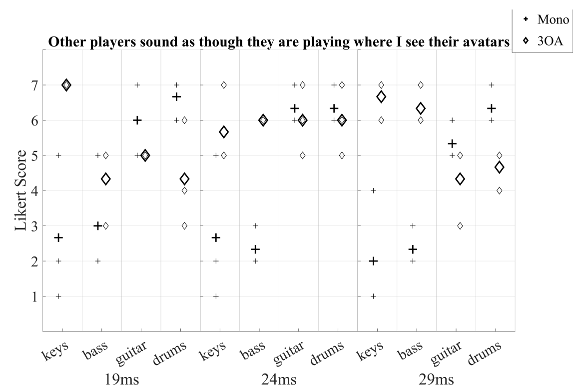

Fig. 8. Likert response to questionnaire item 4.

3.1 QoE Item 1

For the item ‘We can play together as though in the same room’ (Figure 5) a difference can be observed in audio system rating from the keyboard player and the drummer. Across all latency levels the keyboard player provides mean ratings in disagreement for mono audio conditions, while rating with means of slight-strong agreement for immersive audio conditions. The disagreeable rating of mono conditions here contrasts with means of agreement-strong agreement from all other musicians. The drummer rates across all latency levels with means between neutrality and strong disagreement for immersive audio conditions, and means of agreement-strong agreement for mono audio conditions. Low rating of immersive audio conditions here deviates from the response of all other musicians, where means range between slight and strong agreement.

3.2 QoE Item 2

For the item ‘I feel like I am in the same space as the other performers’ (Figure 6) differences in audio system rating are again observed in response from the keyboard player and drummer. Across latency levels the keyboard player rates mono conditions with means of slight-strong disagreement, while rating immersive audio conditions with means of slight-strong agreement. Low rating of mono conditions is the deviation from the trend here, with other musicians rating mono conditions with means of slight-strong agreement across latency conditions. At 19 ms and 29 ms latency, the drummer rates immersive audio conditions poorly with means of approximately slight disagreement, whereas means for mono conditions are between agreement and strong agreement. In this case the low rating of immersive audio conditions is different from the response of other musicians, who rate immersive audio conditions with means of slight-strong agreement.

3.3 QoE Item 3

Rating for the item, ‘I have a sense of where the other performers are located in the space’ (Figure 7) also demonstrates differences in response from the keyboard and drum musicians. Response from the keyboard player across latency levels provides means of agreement-strong agreement for immersive audio conditions, while mean ratings for mono conditions range between slight disagreement and disagreement. The low rating of mono conditions contrasts with means of slight-strong agreement in rating from other musicians. At 19 ms the drummer rates immersive audio conditions with a near-neutral mean, while a mean of near-strong agreement is observed for mono conditions. Here the neutral rating of immersive audio conditions is lower than mean rating of slight-strong agreement provided by other musicians.

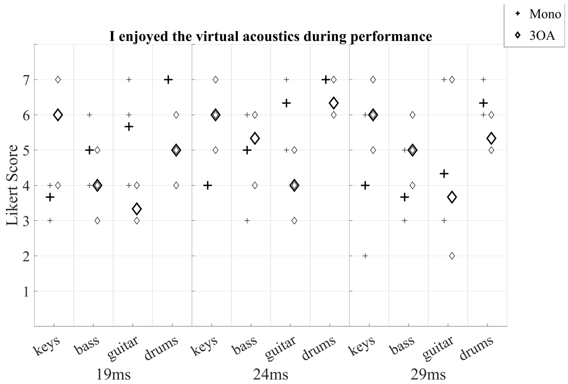

Fig. 9. Likert response to questionnaire item 5.

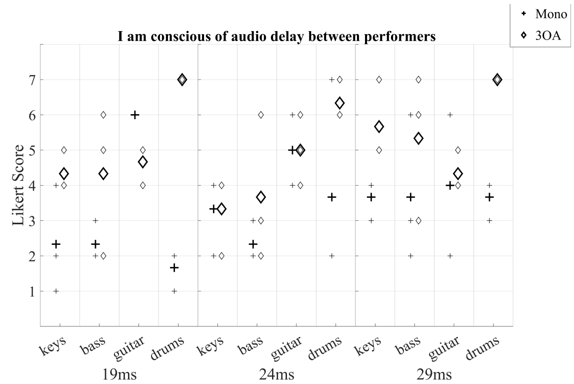

Fig. 11. Likert response to questionnaire item 7.

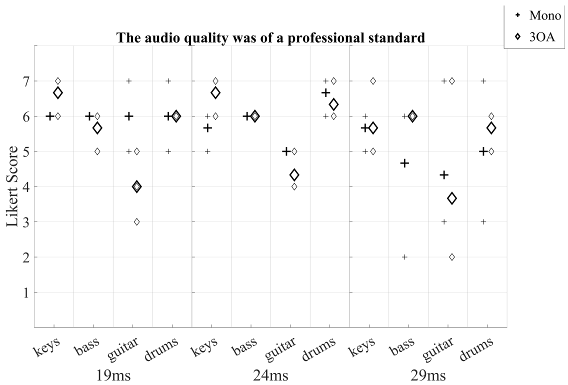

Fig. 10. Likert response to questionnaire item 6.

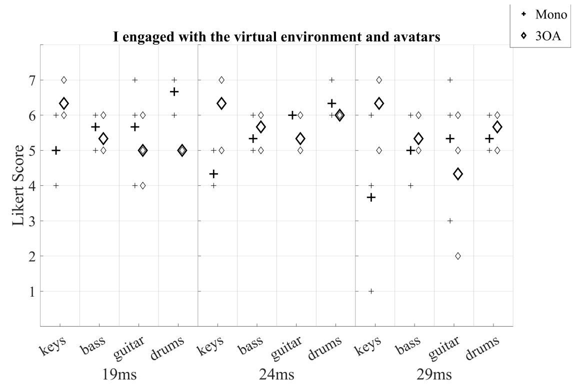

Fig. 12. Likert response to questionnaire item 8.

3.4 QoE Item 4

In response to the item ‘The other players sound as though they are playing where I see their avatars’ (Figure 8) difference can again be observed in rating of audio system. At all latency levels the keyboard player responds with means of slight-strong agreement for immersive audio, while mono conditions are rated with means between slight disagreement and disagreement. At 24 ms and 29 ms latency the bass synthesiser musician demonstrates similar discernment between audio systems, rating immersive audio with means of slight-strong agreement, and mono conditions with means in disagreement. In these cases the low rating of mono conditions contrasts with response from the guitarist and drummer, who rate with means of slight-strong agreement. At 19 ms and 29 ms latency the drummer rates mono conditions with means between agreement and strong agreement, while rating immersive audio with means of neutral-slightly agree. At 29 ms the low rating of immersive audio conditions contrasts with means of high agreement from the keyboard and bass synthesiser musicians.

3.5 QoE Item 5

Response to the item ‘I enjoyed the virtual acoustics during performance (Figure 9) again shows differences in audio system rating. Across all latency levels the keyboard player provides mean ratings of agreement for immersive audio conditions, while rating mono conditions with near-neutral means. The high rating of immersive audio here contrasts with means of slightly disagree-neutral from the bass synthesiser musician at 19 ms, and from the guitarist across all latency levels. At 19 ms the drummer rates mono conditions with strong agreement, and immersive audio conditions with a lower mean of slight agreement. The high rating of mono conditions here contrasts with the less favourable ratings from the keyboard and bass synthesiser musicians.

3.6 QoE Item 6

For the item ‘The audio quality was of a professional standard’ (Figure 10), difference in audio system rating between the guitarist and other musicians is observed. At all levels of latency the guitarist provides approximately neutral mean rating of immersive audio conditions, which contrasts with mean ratings ranging of slight-strong agreement provided by other musicians. At 24 ms latency the guitarist provides mean rating of slight agreement for mono audio conditions, which contrasts with mean rating between agreement and strong agreement provided by the drummer.

3.7 QoE Item 7

The item ‘I am conscious of audio delay between performers’ (Figure 11) shows difference in mean rating of audio system within musician response and between musi-

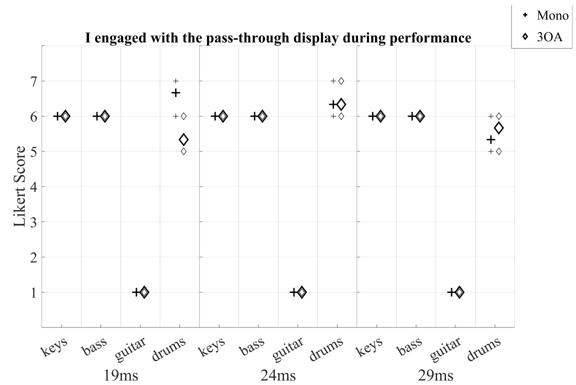

Fig. 13. Likert response to questionnaire item 9.

cians. At 19 ms the keyboard, bass synthesiser, and drum performers all rate mono conditions with means close to slight disagreement. Immersive audio conditions are rated with means of neutral-slight agreement by the keyboard and bass synthesiser musicians, and with a mean of strong agreement from the drummer. Here the low rating of mono conditions by these 3 instruments contrasts with the guitarist response, who rates a mono conditions with agreement. At 24 ms and 29 ms latency, the drummer rates mono conditions with near-neutral means, while rating immersive audio conditions with means of slight-strong agreement. At 29ms the keyboard player also rates mono conditions with a near-neutral mean, while rating immersive conditions with a mean between slight agreement and agreement. A trend is observed for immersive audio conditions to be rated higher than mono conditions in the response of all musicians except the guitarist.

3.8 QoE Item 8

For item 8, 'I engaged with the virtual environment and avatars' (Figure 12s), slight variation of means is observed in audio system rating. At 24 ms and 29 ms the keyboard player rates mono conditions with approximately neutral means, and immersive audio conditions with means between agreement and strong agreement. This low rating of mono conditions contrasts with response from the guitarist and drummer at 24 ms, where means between agreement and strong agreement are given.

3.9 QoE Item 9

Variance in means between musicians is apparent for the item 'I engaged with pass-through display during performance' (Figure 13). Here the guitarist provides mean rating of strong disagreement across all audio system and latency conditions. This contrasts with means ranging between slight and strong agreement for all audio and latency conditions from all other musicians.

3.10 Tempo Slope Analysis

To verify reports of noticeable delay and tempo deceleration (an identifier of effect of latency [4, 12]) tempo analysis of the introduction section of each performance was

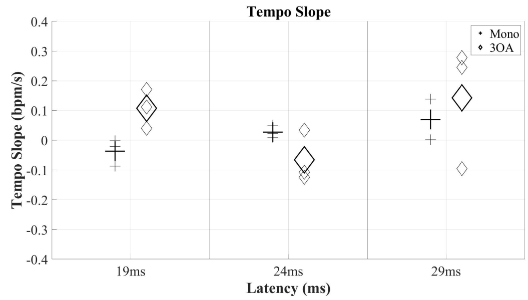

Fig. 14. Kick drum sample tempo slope for mono and 3OA audio conditions across all latency levels.

conducted using kick drum recordings. This section contains no intended tempo change and was performed with consistency across recordings. Tempo slope was calculated using MATLAB[13] with MIR Toolbox [46] for onset detection. Onset time differences were used for tempo calculation, and linear fitting is used to calculate tempo slope. Results (Figure 14) indicate that instances of deceleration do occur in mono conditions at 19 ms, and in immersive audio conditions at 24 ms and 29 ms. Mono conditions seem consistently closer to stable tempo than immersive audio conditions. Tempo also seems to become less stable as latency increases. In some cases tempo acceleration is present, which may be a result of the drummer intentionally speeding up to counteract deceleration in other parts by playing ahead to stabilise performance.

4 DISCUSSION

4.1 Effect of Latency

Although latency was varied across this study no emergent pattern was observed in results. One possibility is that the drummer simply adopted latency-coping strategies in each performance, stabalising the effect of latency for other performers. Indeed previous research [10] has considered that in the NMP context, where no metronome device is present, drums may typically elect to lead performance, effectively acting as a click track. Notably responses to QoE item 7 (Figure 11) contain no means of strong disagreement. This could imply that musicians are aware of delay to at least some degree under all conditions.

In the exit interview musicians commented that latency was a potential area for improvement. Indeed, all performers responded affirmatively to the questions *"did you find (the amount that latency impaired performance varied)?"* and *"was latency more noticeable some times than others?"* which would indicate that some perceptual differences were present, despite lack discernment in QoE results. It may be that the 10 ms latency range and 5 ms increments in this study are too small for a perceptual difference to be acknowledged in QoE rating. In other work latency variance of 40 ms appears to elicit perceptual difference in NMP QoE items such as rating of audio delay or perceived

[13] https://uk.mathworks.com/products/matlab.html

synchrony [43]. Alternatively, it may be the case that the sample size in this study was too small to identify any discernible difference, especially considering a previous study noted synchrony variance across the 19-29 ms range under mono audio conditions [10]. A similar controlled tapping study may yield similar results under immersive audio conditions.

4.2 Effect of Musician

Musicians' ratings consistently varied across QoE items. As previously mentioned, these differences may relate either to the musical instrument or the individual performer. Previous research has shown that NMP synchrony varies with timbral [16, 43] or dynamic [17] parameters of musical sources. It is therefore likely that some differences in rating relate to instrumentation in this study. Another contributing factor may be the system being experienced differently between musicians. For example the drummer has much greater loudness in the room than other performers which will be heard in addition to auralisation. Also each musician is presented with a unique perspective both visually and aurally, which could also influence QoE rating.

Rating of strong disagreement of QoE item 9 by the guitarist is certainly related to instrumentation. In the exit interview they noted that they found playing in VR *"very difficult"* stating that one of the difficulties encountered was *"the visuals"* and *"tiny lag on the (pass-through)"*. They state *"there was a couple of occasions where I moved forward and hit the microphone...and my guitar was out the portal...I had to bring it back in and readjust"* and *"the cameras on the VR...they are not HD...it makes it a little bit harder"*.

Here the perspective provided by HMD cameras used in pass-through display and 2-dimensional nature of pass-through portals made it difficult for the guitarist to view and interact with their instrument and effects pedals. The bass synthesiser player also notes poor video quality, stating that *"you loose granular detail, playing the moog theres obviously lots of controls, it's a bit of a challenge"*. In contrast to guitar, the keyboard, bass synthesiser and drums performers all seemed able to view and interact with using pass-through, as is reflected in high rating from these musicians for QoE item 9. It may be that the pass-through method used in this paper are problematic for instruments such as guitar. Pass-through quality and perspective may be improved through hardware upgrade. Alternatively systems which merge telepresence and XR methods to couple real spaces with virtual environments [21] may be a good solution here.

4.3 Effect of Audio System

QoE results show multiple differences in rating between immersive audio and mono conditions within musicians. This would indicate that immersive audio rendering methods influence the NMP experience relative to a mono baseline. The most overt discernment is noted in response from keys and drums. High rating of immersive audio conditions and generally poor rating of mono conditions by the keyboard player is consistent across QoE items 1-5. Indeed, in the exit interview the keyboard player expressed discernment between these audio conditions, noting *"you could tell there were certain takes which were 2D and some which were 3D"*. Additionally the keyboard player seems to indicate appreciation of immersive audio conditions, stating *"it's like I could hear the drums are set up...like a real drum kit...I think it's really cool"*.

For QoE items 1-3 guitar and bass synthesiser musician results are consistent with high rating of immersive audio in keys results. It is noted, however, that for these QoE items, guitar and bass synthesiser musician results show little discernment between mono and immersive audio conditions, rating both highly. Despite this, both bass synthesiser and guitar performers express awareness of the variance in audio system across the study in exit interview response.

The guitarist states *"when the 3D audio was on, and it all clicked in, and I just instantly turn around to turn it down (a guitar amp), it's quite weird. It's like: 'it's not there is it?'...it sounded like it was on the floor"*. The bass synthesiser musician comments *"even (vox/keys), who was off at an angle, went from being, it felt like hard left, kind of behind where their avatar was, to certain takes where it was bang on...it did feel like they were there, the sound was coming from where their actual avatar (was)...sometimes it just felt like a stereo mix"*. It can be noted that QoE items 1-3 do not specifically relate to audio system, and ratings here may reflect other factors of the experience such as visual display.

QoE item 4 specifically relates to audio, namely localisation. Here, ratings from the bass synthesiser musician conform with the keyboard player response, rating immersive audio conditions generally superior to mono conditions. Coherent rating of QoE item 4 between keyboard and bass synthesiser musicians may indicate that this discernment between audio systems is not simply an individual preference from the keyboard player. It is possible that the discernment between mono and immersive audio conditions expressed here may reflect a beneficial aspect of immersive audio display in comparison to the mono baseline. In the exit interview the keyboard performer commented on the experience *"when you have the separation I know (moog) is over there...if you are a bit out of time you can pull yourself back in, because you can hear if 3 people are in time you can pull back in with them"*. This statement may reflect the findings of a previous study which suggests that stereo source panning may benefit performance synchrony in the NMP context [34].

In the exit interview the drummer indicates an ability to discern between mono and immersive audio conditions, stating *"it feels like you are in the room listening to something...it has that ambience"*. Despite this, the drummer provides rating of QoE items 1-4 where mono conditions are consistently preferred to immersive audio conditions, which tend to be rated neutrally or disagreeable. One possible reason for this is the fact that the drummer has high loudness in the room from their instrument, which has potential to interfere with monitoring of auralised audio over headphones. In previous research it is suggested that clarity may be useful in achieving synchronous performance in the NMP context [33]. It may be the case that perception of audio system may be dependant on instrumental context, and that in the case of drums the clarity of a mono signal makes this preferable to immersive audio rendering.

Notably QoE item 6 reveals a trend for musicians to be more conscious of delay in immersive audio conditions than the mono baseline. This indicates that the audio rendering system used in immersive NMP experiences may influence perception of latency in the delay-sensitive context.

4.4 Wider XR Experience

In the exit interview musicians highlight several aspects of the XR performance experience which are not addressed in the QoE evaluation. When asked about their reaction to playing in XR, the keyboard player comments *"The first time...we felt a bit queasy and weird about it, but coming back to it, it did feel a bit more natural"* highlighting the importance of prior experience as a study control in the XR NMP context. Performers also described several additional desirable features for the system. This included 6DoF, virtual instruments and effects, individual level control, lightweight HMDs, visual metronomes, pass-through portals which could track real instruments such as the guitar.

Concerning visual display, it was noted that though the avatar representation was limited (rendering only head and torso), there was still benefit to the modality of visual contact between performers. Indeed, the bass synthesiser player stated in the interview that *"Even though the avatars were very limited range of motion and lack of expression...little bits of visual interaction did positively influence performance"*. Other performers noted that the shared virtual environment provided a method of cohabiting a space while maintaining isolation between instruments, in contrast to separate booths in typical recording sessions. Generally positive ratings of QoE items 1-3 could be indicative of presence in the XR NMP experience. QoE item 8 also expresses general agreement, which could suggest that virtual environments and avatar representation are an engaging method of visual contact and shared space in the NMP context. This would conform with expectations from other work in the field which suggests a degree of presence and usefulness within the VR/XR modality [21, 22].

5 FURTHER WORK

Improved pass-through display, avatar representation, and haptic control are the topic of current developments, as is investigation of the influence of visual elements on the XR NMP experience. This includes robust motion-photon measurement. Other current work evaluates networked musical interactions under the influence of different virtual acoustic environments and explores control of virtual acoustics in immersive NMP contexts. Improvements in immersive NMP QoE rating methods are also the topic of future development. Alternative immersive audio systems should also be evaluated in the NMP context, including different Ambisonic decoding methods and different auralisation methods.

6 CONCLUSION

The presented case study investigating the QoE of an XR NMP performance identifies several potential areas which merit further investigation in this emerging field. The Likert scale rating of QoE varied considerable between subjects, although it is ambiguous in many cases as to whether this relates to individual preferences or instrumentation. Certainly instrumentation is likely to impact QoE and needs to be considered in future studies.

Results revealed some discernment between immersive audio and the mono baseline in the XR NMP context, and that this can include perception of latency. It is therefore recommended that the introduction of immersive audio systems to the NMP context requires further critical evaluation in order to identify how immersive audio systems may affect the experience.

General positive ratings of QoE items across this study may indicate that XR NMP is viable as a novel modality for online music-making experiences. Feedback from professional artists in an ecological use-case context also guides technical improvements for XR NMP systems. This study provides an introductory investigation into the experience of a professional band using an XR-NMP system. The results from QoE evaluation provide valuable insight for the future development of XR-NMP systems and considerations of future research to understand the perceived quality of XR interactions for NMP.

7 Acknowledgements

This work was supported by the EPSRC IAA (EP/R51181X/1) MINERVA project. This work was supported in part by the UK AHRC XR Stories project, grant no. AH/S002839/1, and in part by a University of York funded PhD studentship. Authors wish to express their thanks to the band *Wolph*, Emma Young and the BBC R&D team, and the staff at Maida Vale Studios, for collaboration on this project.

8 REFERENCES

[1] L. Turchet, "Musical Metaverse:vision, opportunities, and challenges," *Personal and Ubiquitouts Computing* (2023 Jan), doi:https://doi.org/10.1007/s00779-023-01708-1.

[2] L. Turchet, R. Hamilton, and A. Çamci, "Music in Extended Realities," *IEEE Access*, vol. 9, pp. 15810–15832 (2021 Jan), doi:10.1109/ACCESS.2021.3052931.

[3] B. Loveridge, "Network Music Performance in Virtual Reality: Current Perspectives," *Journal of Network Music and Arts*, vol. 2, no. 1 (2020).

[4] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An Overview on Networked Music Performance Tech-

nologies," *IEEE Access*, vol. 4, pp. 8823–8843 (2016 Dec), doi:10.1109/ACCESS.2016.2628440.

[5] S. D'Amario, H. Daffern, and F. Bailes, "Synchronization in Singing Duo Performances: The Roles of Visual Contact and Leadership Instruction," *Frontiers in Psychology*, vol. 9, p. 1208 (2018 July), doi:10.3389/fpsyg.2018.01208.

[6] S. D. Monache, L. Comanducci, M. Boccoli, M. Zanoni, A. Sarti, E. Pietrocola, *et al.*, "A Presence and Performance Driven Framework to Investigate Interactive Networked Music Learning Scenarios," *Wireless Communications and Mobile Computing* (2019 Aug), doi:https://doi.org/10.1155/2019/4593853.

[7] L. Turchet, N. Garau, and N. Conci, "Networked Musical XR: where's the limit? A preliminary investigation on the joint use of point clouds and low-latency audio communication," presented at the *Audio Mostly* (2022 Sep).

[8] P. Cairns, H. Daffern, and G. Kearney, "Parametric evaluation of ensemble vocal performance using an immersive network music performance audio system," *Journal of the Audio Engineering Society*, vol. 69, no. 12, pp. 924–933 (2021 Dec), doi:https://doi.org/10.17743/jaes.2021.0040.

[9] P. Cairns, A. Hunt, J. Cooper, D. Johnston, B. Lee, H. Daffern, *et al.*, "Recording music in the metaverse: a case study of XR BBC Maida Vale recording studios," presented at the *AES 2022 International Conference on Audio for Virtual and Augmented Reality* (2022 Aug).

[10] C. Chafe, J.-P. Cáceres, and M. Gurevich, "Effect of Temporal Separation on Synchronization in Rhythmic Performance," *Perception*, vol. 39, no. 7, pp. 982–992 (2010 Jan), doi:10.1068/p6465.

[11] S. Farner, A. Solvang, A. Sæbo, and U. P. Svensson, "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments," *Journal of the Audio Engineering Society*, vol. 57, no. 12, pp. 1028–1041 (2009 Dec), doi:http://www.aes.org/e-lib/browse.cfm?elib=15235.

[12] C. Chafe and M. Gurevich, "Network Time Delay and Ensemble Accuracy: Effect of Latency, Asymmetry," presented at the *AES 117th Convention* (2004 Oct).

[13] M. Bosi, A. Servetti, C. Chafe, and C. Rottondi, "Experiencing remote classical music performance over long distance: a Jacktrip concert between two continents during the pandemic," *Journal of the Audio Engineering Society*, vol. 69, no. 12, pp. 934–945 (2021 Dec), doi:https://doi.org/10.17743/jaes.2021.0056.

[14] R. Hupke, L. Beyer, M. Nophut, S. Preihs, and J. Peissig, "Effect of a global metronome on ensemble accuracy in networked music performance," presented at the *AES 147th Convention* (2019 Oct).

[15] R. Hupke, D. Jan, N. Werner, and J. Peissig, "Latency and quality-of-experience analysis of a networked music performance framework for realistic interaction," presented at the *AES 152nd Convention* (2022 May).

[16] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, "Feature-Based Analysis of the Effects of Packet Delay on Networked Musical Interactions," *Journal of the Audio Engineering Society*, vol. 63, no. 11 (2016 Dec), doi:http://dx.doi.org/10.17743/jaes.2015.0074.

[17] A. Barbosa and J. Cordeiro, "The Influence of Perceptual Attack Times in Network Music Performance," presented at the *AES 44th International Conference* (2011 Nov).

[18] A. Carot, C. Hoene, H. Busse, and C. Kuhr, "Results of the The Fast Music Project - Five Contributions to the Domain of Distributed Music," *IEEE Access*, vol. 8, pp. 47925–47951 (2020 March), doi:10.1109/ACCESS.2020.2979362.

[19] M. Iorwerth and D. Knox, "The application of Networked Music Performance to access ensemble activity for socially isolated musicians," presented at the *Web Audio Conference 2019 – Diversity in Web Audio* (2019 Dec).

[20] C. Drioli, C. Allocchio, and N. Buso, "Network Music Performance and Natural Interaction via LOLA: Low Latency High Quality AV Streaming," presented at the *International Conference on Information Technologies and Performing Arts* (2013 April).

[21] R. Hupke, S. Preihs, and J. Peissig, "Immersive room extension environment for networked music performance," presented at the *AES 153rd Convention* (2022 Oct).

[22] J. Tamplin, B. Loveridge, K. Clarke, Y. Li, and D. J. Berlowitz, "Development and feasibility testing of an online virtual reality platform for delivering therapeutic group singing interventions for people living with spinal cord injury," *Journal of Telemedicine and Telecare*, vol. 26, no. 6, pp. 365–375 (2020 July), doi:10.1177/1357633X19828463.

[23] A. Hunt, *Creating a VR recording live music system*, Masters Thesis, University of York, York (2021).

[24] M. Gurevich, D. Donohoe, and S. Bertet, "Ambisonic spatialization for network music performance," presented at the *Interactive Audio Systems Symposium* (2011 June).

[25] P. Cairns, H. Daffern, and G. Kearney, "Immersive Network Music Performance: Design and Practical Deployment of a system for Immersive Vocal Performance," presented at the *149th AES Convention* (2020 Oct.).

[26] H. von Coler, N. Tonnätt, V. Kather, and C. Chafe, "Sprawl: A network system for enhanced interaction in musical ensembles," presented at the *Proceedings of the 18th Linux Audio Conference* (2020 Nov).

[27] W. Ritsch, "ICE - towards distributed networked computermusic ensemble," presented at the *2014 International Computermusic Conference* (2014 Sep).

[28] D. Rossetti and C. C. Bomfim, "Audiovisual compositions and telematic performance: Collaborations During the Pandemic," *Journal of Network Music and Arts*, vol. 3, no. 1 (2021), doi:https://commons.library.stonybrook.edu/jonma/vol3/iss1/.

[29] R. Hoy and D. Van Nort, "A Technological and Methodological Ecosystem for Dynamic Virtual Acoustics in Telematic Performance Contexts," presented at the *Audio Mostly 2021* (2021 Sep), doi:10.1145/3478384.3478425.

[30] J. Francombe, T. Brookes, and R. Mason, "Evaluation of spatial audio reproduction methods (part 1): elicitation of perceptual differences," *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 198–211 (2017 March), doi:https://doi.org/10.17743/jaes.2016.0070.

[31] M. Kob, S. Amengual Garí, and Z. Kalkandjiev, *Room Effect on Musicians' Performance*, pp. 223–249 (Springer Nature Switzerland AG) (2020 Aug).

[32] A. Niimura, S. Serafin, and N. Kaplanis, "Real-time dynamic acoustics in 6dof vr for amateur singers," presented at the *AES 2022 International Conference on Audio for Virtual and Augmented Reality* (2022 Aug).

[33] A. Carot, C. Werner, and T. Fischinger, "Towards a Comprehensive Cognitive Analysis of Delay-Influenced Rhythmical Interaction," presented at the *International Computer Music Conference* (2009 Aug).

[34] R. Hupke, J. Peissig, A. Genovese, S. Sridhar, and A. Roginska, "Impact of Source Panning on a Global Metronome in Rhythmic Networked Music Performance," presented at the *2020 27th Conference of Open Innovations Association (FRUCT)* (2020 Sep).

[35] G. Kearney, H. Daffern, P. Cairns, A. Hunt, B. Lee, J. Cooper, *et al.*, "Measuring the Acoustical Properties of the BBC Maida Vale Recording Studios for Virtual Reality," *Acoustics*, vol. 4, no. 3, pp. 783–799 (2022 Sep), doi:10.3390/acoustics4030047, URL `https://www.mdpi.com/2624-599X/4/3/47`.

[36] M. Wright, A. Freed, and A. Momeni, "2003: OpenSound Control: State of the Art 2003," *A NIME Reader*, pp. 125–145 (2017).

[37] J.-P. Cáceres and C. Chafe, "JackTrip: Under the Hood of an Engine for Network Audio," *Journal of New Music Research*, vol. 39 (2010 Nov), doi: 10.1080/09298215.2010.481361.

[38] M. Kronlachner, "Ambisonics plug-in suite for production and performance usage," presented at the *Linux Audio Conference* (2013 May).

[39] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, vol. 8, no. 11 (2018 Oct), doi:10.3390/app8112029.

[40] P. Ferguson, C. Chafe, and S. Gapp, "Trans-Europe Express Audio: testing 1000 mile low-latency uncompressed audio between Edinburgh and Berlin using GPS-derived word clock, first with Jacktrip then with Dante," presented at the *AES 148th Convention* (2020 May).

[41] R. Hupke, S. Sridhar, A. Genovese, M. Nophut, S. Preihs, T. Beyer, *et al.*, "A Latency Measurement Method for Networked Music Performances," presented at the *AES 147th Convention* (2019 Oct).

[42] International Organization for Standardization, "ISO 3382-1:2009; Acoustics-Measurement of Room Acoustic Parameters—Part 1: Performance Spaces," www.iso.org (accessed Dec. 15, 2022).

[43] K. Tsioutas and G. Xylomenos, "On the impact of audio characteristics to the quality of musicians' experience in network music performance," *Journal of the Audio Engineering Society*, vol. 69, no. 12, pp. 914–923 (2021 Dec), doi:https://doi.org/10.17743/jaes.2021.0041.

[44] F. Rumsey, "Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm," *Journal of the Audio Engineering Society*, vol. 50, no. 9, pp. 651–666 (2002 Sep).

[45] H. Lee, "A Conceptual Model of Immersive Experience in Extended Reality," (2020 Sep), doi:10.31234/osf.io/sefkh, URL `psyarxiv.com/sefkh`.

[46] O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," presented at the *10th International Conference on Digital Audio Effects* (2007 Sep).

THE AUTHORS


Patrick Cairns


Anthony Hunt


Daniel Johnston


Jacob Cooper


Ben Lee


Helena Daffern


Gavin Kearney

Patrick Cairns received a BSc in Audio Systems Engineering from Glasgow Caledonian University (2019), and MSc by Research in Audio Technology at the AudioLab, University of York (2021), where they continue their studies as a PhD student. Patrick's research explores the development, application and evaluation of network music performance systems, with a focus on immersive audio.

Anthony Hunt graduated from University of York in 2021 with a Masters of Science degree in Audio and Music Technology. He is currently a postgraduate researcher at the AudioLab, University of York, investigating the design and impact of virtual avatar representation on social presence, in the context of group VR-NMP.

Daniel Johnston is a lecturer of audio systems engineering within the AudioLab, University of York. His research explores sound interactions within virtual environments, with particular focus on the application of immersive audio into digital healthcare interventions for mental health.

Jacob Cooper is a recent graduate of the AudioLab who, following the completion of his MSc in Audio & Music Technology, has gone on to work as a Research Technician within the lab. His expertise is in immersive audio for Extended Reality (XR) experiences using Ambisonics.

Ben Lee graduated from University College London in 2018 with a BSc in Physics, followed by a MSc in Audio and Music Technology at the University of York. After obtaining his MSc, he worked for a year in acoustic consultancy, then returned to the University of York's AudioLab as a research technician evaluating the Opus Codec in immersive audio and VR contexts.

Helena Daffern is a Professor in the AudioLab at the University of York where she graduated with a BA (Hons.) and MA degree in Music, and a PhD in Music Technology in 2009. She then completed postgraduate singing training at Trinity College of Music, London. Her interdisciplinary research uses virtual reality technology to investigate voice science and acoustics, particularly, performance, choral singing and singing for health and wellbeing.

Gavin Kearney is a Professor of Audio Engineering at the University of York. He obtained his PhD degree in Audio Signal Processing from Trinity College Dublin in 2010. He has written over 100 research articles and patents on different aspects of immersive and interactive audio. He is currently Vice-Chair of the AES Audio for Games Technical Committee and an active sound engineer and producer of immersive audio experiences.