Audio Engineering Society

Conference Paper 14

Presented at the International Conference on
Spatial and Immersive Audio
2023 August 23–25, Huddersfield, UK

A commonly-accessible toolchain for live streaming music events with higher-order ambisonic audio and 4k 360 vision

Bartłomiej Mróz[1,2], Piotr Odya[1], Przemysław Danowski[3], and Marek Kabaciński[2]

[1]*Gdańsk University of Technology, Department of Multimedia Systems, Narutowicza 11/12, 80-233 Gdańsk, Poland*
[2]*Zylia Sp. z o. o., Mostowa 27, 61-854 Poznań, Poland*
[3]*Chopin University of Music, Department of Sound Engineering, Okólnik 2, 00-368 Warsaw, Poland*

Correspondence should be addressed to Bartłomiej Mróz (`bartlomiej.mroz@pg.edu.pl`)

ABSTRACT

An immersive live stream is especially interesting in the ongoing development of telepresence tools, especially in the virtual reality (VR) or mixed reality (MR) domain. This paper explores the remote and immersive way of enabling telepresence for the audience to high-fidelity music performance using freely-available and easily-accessible tools. A functional VR live-streaming toolchain, comprising 360 vision and higher-order ambisonic audio, is presented and described in detail. Two separate recordings with two different approaches are outlined. The results gathered from the audiences' impressions are discussed. Finally, possible improvements to the immersive live streaming and recording processes were discussed.

1 Introduction

The immersive sound produced via loudspeakers or headphones simulates the experience of sound in 3-dimensional space. This can improve the quality of immersive experiences, such as communication and entertainment. However, most importantly it also allows to match audio with full-sphere visual stimuli.

It is worth mentioning that the first implementations of remote concerts were applied to academic networks. One such experimental concert with artists from different cities in Poland took place in 2001 [1]. Transmissions using ambisonics have also been performed [2]. The concert, *"PURE Ambisonics Concert & the Night of Ambisonics"*, organized in the framework of an international conference (3rd International Conference on Spatial Audio) in Graz in September 2015, was undoubtedly one of the pioneering events of its kind [3],[4],[5]. During the concert evening, an ambisonic format was used to distribute the concert to different venues and broadcast it in real-time (including the following radio transmissions: nationwide terrestrial and satellite radio broadcasts). The concert hall was prepared for recording and transmission using a 23-channel speaker system, a 5.1 mix, and a binaural mix for headphone listening. The concert from ICSA 2015 encouraged the authors of this idea to conduct a live 3D concert with Al Di Meola in July 2016, including real-time spatial effects and transmission to another interior [3]. However, all of these presentations were experimental. To become a standard for recording or broadcasting, they must be implemented on commonly used platforms.

Support for spatial audio formats in commonly used live-streaming tools has only recently become available. Binaural and ambisonic techniques as tools for localizing remote audio sources have settled fastest in teleconferencing applications [6],[7],[8],[9],[10],[11].

It is also important to mention networked music performances in the context of remote performances. Especially in the recent intensification of lockdown restrictions due to the COVID-19 pandemic has accentuated the demand for innovative methods of remote teamwork in the music sector. This includes music academies, conservatories, musical groups, and solo artists, all of whom stand to gain from having the appropriate tools to create high-quality, real-time collaborative music in a dispersed manner.

Networked Music Performance (NMP) and live streaming represent two distinct forms of digital musical interaction, differentiated largely by their communication mode and latency requirements. NMP is a duplex system, meaning it supports bidirectional communication between all participants simultaneously. This real-time interaction requires very low latency, often in the range of tens of milliseconds, to maintain the synchronicity of performance and allow musicians to react to each other as if they were in the same physical location. Conversely, live streaming is a unidirectional transmission where performers broadcast their music to an audience without receiving any real-time feedback. As such, latency in live streaming, although it can often be quite significant, is generally of lesser concern since the delay doesn't impact the performers' ability to play together. The audience's perception of the performance remains consistent regardless of slight broadcast delays. In the case of the NMPs [12], successful attempts have already been made to transmit multichannel audio [13],[14],[15],[16],[17],[18],[19]. Especially it is worth to mention the *"JackTrip concert between two continents during the pandemic"*, during which Bosi et al. [20] have connected musicians from Italy, Germany and United States. To the best author's knowledge, this was the first documented attempt to perform a classical music repertoire in a distributed concert setting where physical separation between locations is in the order of 10,000 km while ensuring high-fidelity audio quality. One additional concert which also was a worth-mentioning achievement in the area of networked music performances employed Dante audio networking technology by Audinate. In June 2021, the Blue Note Entertainment Group and Peltrix collaborated on an industry-first demonstration for long-distance, networked audio-video links [21],[22]. They demonstrated their solution linking three locations up to 750 miles apart with multiple streams of audio and Dante video links. In this demonstration, system latency was so low, musicians were able to comfortably play together as if they were in the same room. No click tracks or guides were used, and timing was natural for performers. Employment of Dante technology with precise GPS clocking among locations allowed musicians seeing each other, taking visual cues, read body language and engaging with each other seamlessly.

2 Commonly available live-VR-capable devices and platforms

Creating a 360 live video stream is widely supported by both popular video streaming platforms and manufacturers of 360 cameras. Live streaming of 4-channel immersive audio is slightly less easy to come by, primarily due to two main reasons. Firstly, ambisonic microphones require A- to B-Format conversion, usually done by a dedicated VST plugin hosted by a digital audio workstation (DAW) running on a PC.

There are microphones capable of onboard ambisonic conversion, such as the Zoom H2n (4ch spatial audio, no height channel) [23], the Zoom H3-VR (1OA - 1st-order ambisonics) [24], or the Voyage Audio Spatial Mic (2OA - 2nd-order ambisonics) [25]; however, even with B-format ambisonic output from an ambisonic microphone, most 360 cameras do not support connecting multi-channel external audio sources. The only two identified during this research 360 cameras which support attaching external multi-channel devices were the Insta360 Pro2 [26] and the Insta360 Titan [27] – both are professional-grade 360 cameras. Both devices comprise a USB port and support connecting UAA-compliant devices, but only up to four channels.

Secondly, among the popular video streaming platforms, only Facebook allows 360 videos with 4-channel 1OA audio [28]. Its competitor, YouTube, accepts only stereo with 360 live videos [29]. The majority of 360 content on popular video platforms is not created as live broadcasts; hence, a real-time joining of 360 video with ambisonic audio is a very niche problem. However, with growing interest in telepresence, especially for VR environments, real-time solutions are becoming of interest in the audio research community.

3 Experiments with VR live streaming of music events

A recent expansion of open-source tools capable of processing higher-order ambisonic (HOA) audio also included development in streaming, transcoding, and viewing platforms. Hence, it has become feasible to combine these tools and achieve a functional toolchain for live streaming with 360 video and HOA audio.

One such music event - *"Koncert Muzyki Dawnej"* [30] - was live-streamed in 2020 during the internal tests conducted at the Chopin University of Music (UMFC). This event was also transmitted to the Facebook platform. Unfortunately, rewatching the streamed concert is no longer possible on PCs, as Facebook's spatial audio binauralization has been broken since 2021. Nonetheless, it still works fine on smartphone devices.

The extension of this toolchain has been developed by the authors for two music events. The first one was a symphonic concert featuring John Rutter's *"Magnificat"*, [31] which was experimentally broadcasted in March 2022 on the occasion of the 70th anniversary of the Faculty of Electronics, Telecommunications, and Informatics of the Gdańsk University of Technology. Furthermore, it was also used during a percussion and drum instruments concert in December 2022 with Iannis Xenakis' *"Persephassa"* and *"Pleiades"* at the UMFC [32]. The UMFC is equipped with a VR projection booth that is designed as an endpoint for VR content distribution; this booth is located in the foyer of the concert hall and thus it is frequently visited by concertgoers. The idea was to stream a live music concert directly to the booth and ask the audience to compare it with the in situ experience.

The extended 360+HOA live streaming toolchain used in March and December 2022 is described in detail in the following sections.

4 Open-source, 360+HOA live streaming toolchain

The first tool of the toolchain is an enthusiast-developed version of a popular, open-source platform for video streaming – Open Broadcaster Software Music Edition [33]. This version provides three critical features not present in the regular OBS version: (1) support for ASIO devices, (2) capability of hosting VST plugins, and (3) multichannel mappings, supporting up to 16 channels as of March 2022 (the regular OBS supports up to 8 channels; the most recent version of the OBS Music Edition as of March 2023 supports 22.2 channel mapping, i.e., 24-channel audio). Furthermore, two more open-source tools were recently presented – both in the year 2020. These tools are needed for transcoding and presenting the live stream in a way that is easily accessible - such as a video player in a web browser.

The first tool was developed by the Envelop company and is called Earshot – Envelop Ambisonic RTMP Streaming Higher-Order Transcoder [34]. Earshot is a containerized multichannel RTMP-to-DASH transcoder based on Nginx [35]. The last tool required for 360+HOA live streaming is a viable web player. HOAST360 is an open-source, HOA-capable, 360° video player which dynamically outputs a binaural audio stream from up to fourth-order ambisonic (4OA) audio content [36],[37]. HOAST360 player works with the Opus audio codec [38], which is natively supported on almost all browsers – including smartphone ones. The notable browser which does not support the Opus audio codec is the Safari browser, developed by Apple company. It is worth mentioning that the Opus audio codec is capable of a total of 255 audio channels; however, for ensuring proper sound source localizability, it is recommended to provide at least 32 kbps of bandwidth per channel [39].

4.1 HOA audio and 4k 360 2D video capture of the first concert

The first concert was captured using the Insta360 Pro2 camera and the Zylia ZM-1 microphone. The Insta360 Pro2 camera consists of 6 equidistant, horizontally spaced lenses and is capable of 8k 2D 360 recording or 6k 3D 360 recording with simultaneous on-board stitching at 4k resolution (2D or 3D, depending on the selected recording format) [40]. This camera provides a multitude of wired and wireless connectivity options, one of which is Ethernet connectivity, allowing a very distant yet stable connection. The Ethernet connectivity enables real-time capture of the aforementioned onboard-stitched 4k material; the recorded higher-resolution material requires additional stitching in post-production. All the connectivity and recording capabilities require a lot of onboard processing power which in turn generates a considerable amount of heat. In the case of the Insta360 Pro2 camera, this heat is dissipated by the cooling fans. In the recording mode, the

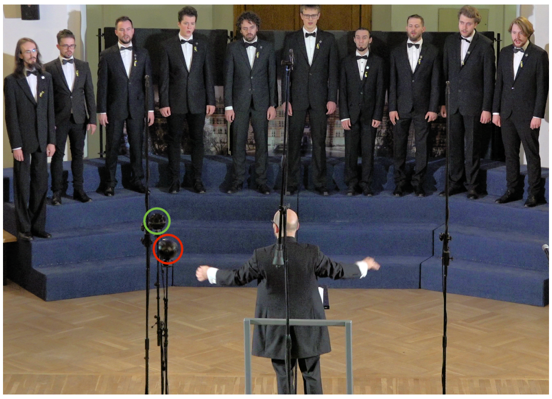

Fig. 1: 360 camera (red) and ambisonic microphone (green) placement during the first concert at the Gdańsk University of Technology

Insta360 Pro2 is capable of up to 15 minutes of fanless operation. However, during a live streaming operation, the cooling fans work all the time at maximum speed.

As for the Zylia ZM-1 microphone – it is a 19-capsule ambisonic microphone array capable of third-order ambisonic (3OA) recording [41]. This microphone features a USB connection, limiting issues with the number of audio cables required for typical multi-capsule microphone arrays. Additionally, a USB-over-Ethernet adapter was used in order to provide more distant connectivity, similar to the one provided by the camera. The connectivity over a high distance was required for such a large concert. The camera was positioned between the first row of seats and the stage at the height of approximately 200 cm, providing the so-called *"super-viewer"* position. In order to suppress the noise of the camera's cooling fans, the microphone was placed 50 cm in the back and 30 cm above the camera – so the viewer's rotations were only slightly off between video and audio rotations. The positioning of camera and microphone is presented in Figure 1.

The camera and the microphone were connected to a workstation with sufficient computational and graphic processing capabilities (i.e., i7-7700k and GTX1080), allowing for running necessary drivers, audio and video processing tools, and audio-video synchronization with subsequent transmission of the live content to a web server. As previously mentioned, OBS Music Edition allowed for 16-channel audio at the time of the realization of this streaming concert. Therefore, a separate tool was required to provide a 19-channel A-Format to 16-channel B-Format conversion. A digital audio workstation (DAW) widely adopted by audio researchers is Reaper, developed by the Cockos company. It is not only a very viable audio-editing software, but its Windows version includes a ReaRoute ASIO driver, allowing for a virtual intra-system connection between ASIO-capable applications [42]. This feature permitted using Reaper for hosting VST plugins: not only A- to B-format conversion but also ambisonic equalization, compressor, and limiter, ensuring a decent audio quality without distortions or clippings. The ambisonic signal was processed in real-time and then sent to OBS Music Edition via ReaRoute ASIO connection. Then the video was transmitted with 17 Mbps of bandwidth; the 16-channel audio was transmitted with 1024 Kbps, ensuring 64 Kbps per channel. Regarding audio bitrate compression for HOA content, it is important to note that Rudzki et al. [43],[44],[45] concluded that using strong bitrate compression will not affect the auditory localization in scenes encoded using Opus compression when compared to uncompressed Ambisonic presentations; however, the timbral fidelity aspect of the compressed audio should be considered as well. In authors' experience, the timbral fidelity aspect was sufficiently provided with the selected bandwidth.

4.2 HOA audio and 4k 360 2D video capture of the second concert

The second event was also captured using the Insta360 Pro2 camera; however, the sound recording setup combined two layers. The first layer comprised capturing sound from all directions in a channel-based manner using a 1OA microphone (namely, SoundField ST450 mkII). The second layer was the object-based layer of spot microphones for individual musical instruments. The object-based layer was panned and mixed with the channel-based layer using third-order ambisonic VST plugin (namely, IEM's *MultiEncoder* [46], which evaluates the spherical harmonics for given direction; then, it uses these weights for distributing the input signal to the ambisonic channels) in Reaper DAW. The channel-based 1OA layer was not upmixed or upscaled in any way. Routing of the sound from the concert hall to the mixing room was done using the Dante network. The receiving workstation was using Dante Virtual Soundcard which allows to send and receive up to 64 channels [47]. The positioning of ambisonic microphone, spot microphones and the camera is presented in Figure 2.

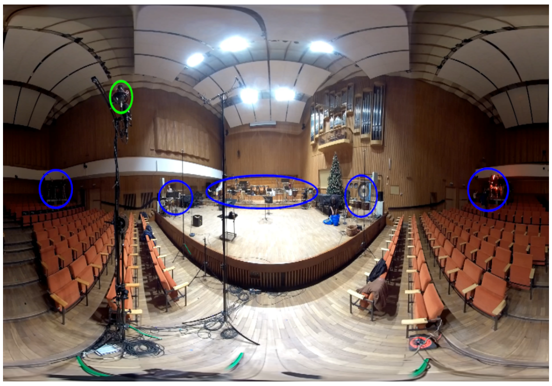

Fig. 2: Ambisonic microphone (green) and spot microphones / individual instrument zones (blue) during the second concert at the Chopin University of Music. The 360 camera is the source of the photo and thus is not visible

Spatial resolution disparity between the immersive channel-based and the object-based layers prevents the spatial disintegration effect [48] - an offset resulting in the incoherent auditory image, i.e., two different sound source locations of the same sound source that is audible during the head rotations. Using two sources with high spatial resolution, such as third-order ambisonic recording and third-order synthesized objects, can produce problems with the spatial matching of the same sources and consequently disjoint one music object into two separate, individually reacting to the head rotations. This effect is especially significant when the sound source is the human voice. Thus using a lower resolution for immersive channel-based capture and higher for objects creates a more coherent spatial image of the scene.

4.3 Audio-video synchronisation

Furthermore, OBS Music Edition allows for delaying audio signals. Thus, synchronization of audio and video signals was possible; however, it could only be done in a manual fashion. Moreover, this synchronization varied with each turning off/on the microphone or the camera. Fortunately, once the signals were synchronized, they remained stable for the entire time of the capture. The delay of the microphone signals with regard to the video signal from the Pro2 camera varied around 10 seconds.

4.4 RTMP encoding, DASH transcoding

The synchronized and encoded 4k 360+HOA material was subsequently transmitted to the web server hosting necessary services for a multi-client reception of this live stream. The necessary services were hosted on the Multimedia Systems Department server, namely, the Earshot RTMP-DASH transcoder and HOAST360 player. The HOAST360 player was the service publicly available to the viewers; it was the last part of the toolchain. The block diagram of the toolchain is presented in Figure 3.

4.5 Configuring grid geometry for improved 360 projection

Grid geometry is an essential component in rendering 360° videos for virtual reality experiences. The choice of grid geometry affects the overall video quality, projection accuracy, and the viewer's immersive experience. Initial testing of the toolchain revealed the necessity to optimize the grid geometry of the 360 video projection of the HOAST360 player, which uses the videojs-xr [49] plugin based on videojs-vr [50]. Both are using three.js [51] library for generating sphere geometries. Sphere geometry in three.js is based on 3 main parameters: radius, number of width segments, and number of height segments. In HOAST360, the initial configuration uses a spherical geometry with a radius of 256 units, and 32 width and height segments. During testing, however, the projection quality appeared to be unsatisfactory – especially at the bottom of the video - as presented in Figure 4. A test video configuration with the use of Orion360 test content [52] was created to test the projection more thoroughly.

After a series of tests and adjustments, an optimized configuration was achieved with a sphere buffer geometry of 64 width segments and 64 height segments. The increased number of segments allowed for a more detailed projection of the video, particularly noticeable in the logo projection at the nadir of the video. The optimized configuration enhanced the overall viewing experience - as presented in Figure 5.

4.6 Audience and the questionnaire regarding the first concert

Since the whole toolchain setup was highly experimental, the announcement about the first concert was sent

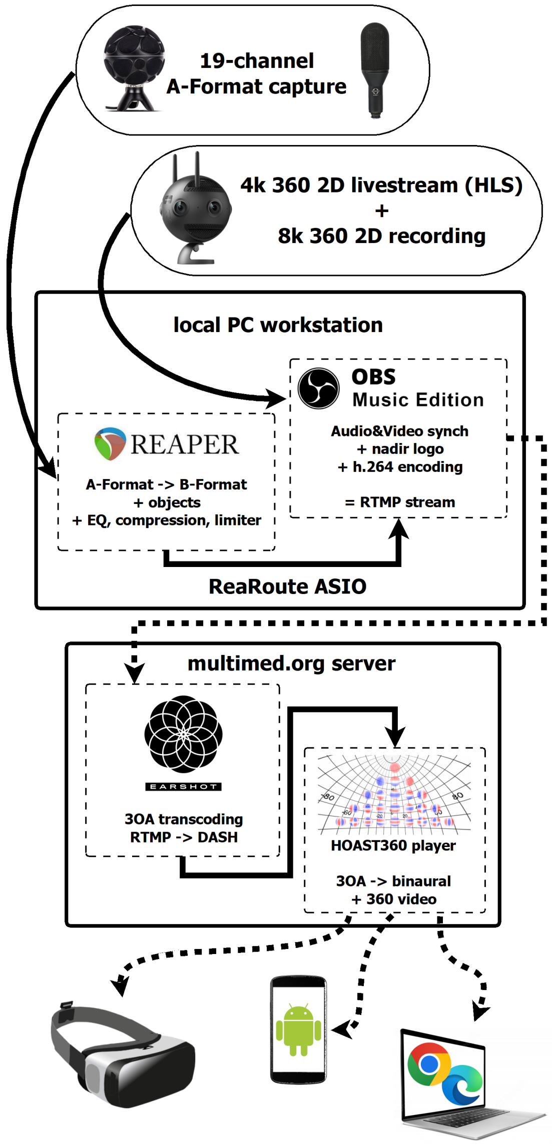

Fig. 3: 4k 360 + 3OA live stream toolchain

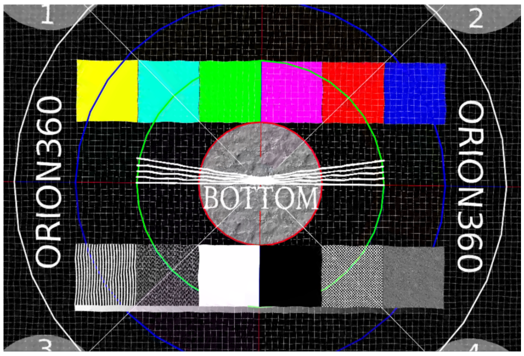

Fig. 4: Initial HOAST360 spherical geometry render testing. Visible artifacts especially at straight parallel lines and in joining between different sections

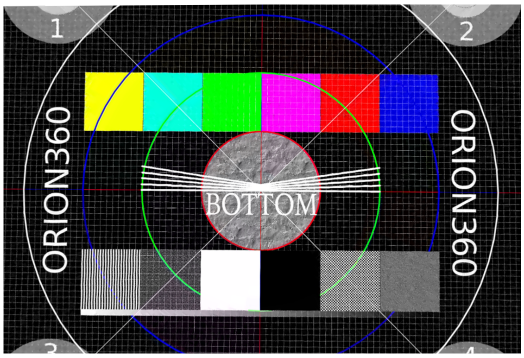

Fig. 5: Adjusted HOAST360 spherical geometry render testing. Much smoother lines and overall image quality are visible.

only to immersive media community groups. No specific mechanism was implemented in order to track the exact number of connected viewers; however, a server log provided a specific approximation of this number. It is estimated that 50-70 viewers from all around the world watched at least some part of the concert. Additionally, the audience was asked to fill in a short questionnaire after the live stream; however, only eight participants provided answers. The questionnaire comprised questions about participants' background, the devices used, as well as the rating of different aspects of the live stream on an 11-point Likert scale. Furthermore, the participants were allowed to write open comments about the whole experience. Most of the participants used a PC and a web browser; three of them used VR goggles. All of the participants were using hi-fidelity headphones. All of the participants work either professionally or scientifically in the audio industry. The written feedback stated a very enthusiastic reception of the whole event and the technologies employed; the most stated issues were occasional – but very few – freezes and relatively poor video quality (especially disturbing were the stitching and compression artifacts). No participant reported a noticeable disparity

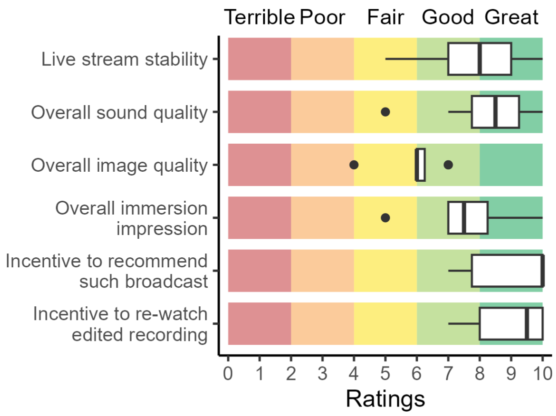

Fig. 6: Viewers' ratings of different 360+HOA live stream aspects. The ratings were given on an 11-point (0-10) Likert scale

between audio and video rotations, despite a slight displacement between the ambisonic microphone and 360 camera. The detailed Likert ratings are presented in Figure 6. The participants who filled in the questionnaire were located in Gdańsk, Poznań, Warsaw (Poland), Montreal (Canada), Derby (United Kingdom), Parma (Italy), New York City (United States), and Sweden.

4.7 Audience and feedback regarding the second concert

During the second concert, the feedback from the participants was collected in verbal form, right after they visited the VR booth. Reactions were generally positive with high regard to the sound but with comments regarding the quality of the video. The artifacts from the first concert were not as strongly perceived since the camera was placed a little farther away. However, the second concert was held in a much darker space than the first one, therefore the image quality was limited due to the sensor size (1/2.74-inch) in low light conditions and high compression at the camera.

The recorded fragment of the second concert was binauralized (DearVR Ambi Micro, binaural format using KU100 HRTFs [53]). As for the video, only the 360 photo was captured (live-streaming was not recorded during the concert this time). In order to visualise the sound scene, the output of Blue Ripple Sound's O3A Flare VST plugin [54] was captured. Such a combination of binauralized audio and 2D video was subsequently uploaded to `https://youtu.be/niRuKz6EUmc` and is available for viewing.

4.8 Post-production

The original, live-streamed material from the first concert included stitching artifacts, such as bending or flickering the conductor's hands or the soprano soloist's arm. Since they are two leading performers on which the viewer's attention is highly focused, such artifacts could not remain in the final version of the recorded material. Other strongly pronounced artifacts included split or blurred tripods of other event installments, unnatural shapes of floor and ceiling lines, etc. The stitching software provided with the Insta360 Pro2 camera does not allow to correct the stitching artifacts, even in post-production.

Therefore, a specialized software dedicated to VR stitching – Mistika VR – was employed for this purpose. This software allows for precise control over stitching parameters, including manipulating the stitching lines between recordings from the camera's lens [55]. Figure 7 shows the difference between uncorrected stitching and the corrected one, with visible modified stitching lines.

The recorded material from the first concert was post-produced with corrected stitching and in a higher resolution of 7680x3840 pixels (8k 360 2D). The post-produced version was subsequently uploaded to `https://multimed.org/weti70/` and is available for viewing.

5 Summary

The presented toolchain showcases a scalable, open-source-based solution for high-fidelity immersive live streaming. An enthusiastic reception from the remote audience, underlined by ratings given in the questionnaire, exhibits high anticipation and demand for such realizations . The described issues with 360 stitching artifacts could be addressed by using a 2-lens 360 camera, or a 360 camera from a different manufacturer who provides better on-board stitching capabilities. Additionally, there is a general issue with 360 cameras and their support for multi-channel higher-order ambisonic audio. However, with further development of such streaming platforms, it might be happening in the near future that such support would be included in professional and consumer 360 cameras, similarly as a stereo

Fig. 7: (a) original, uncorrected stitching with artifacts; (b) the stitching modifications. The red lines indicate the stitching between the recordings from different camera lenses. The green lines indicate the non-stitched areas within which the image originates from a single lens

input is almost always present even in action cameras. On the other hand, using a dedicated workstation with DAW software might be helpful for combining higher-order ambisonic signals from multiple microphones, as well as from traditional microphones with further object-like encoding. Such an approach might lead to a multi-point solution, allowing for six degrees of freedom. However, such systems working in real-time are yet to be seen.

References

[1] Faculty of Physics of University of Warsaw, NASK, "10th Anniversary of the Internet in Poland, Internet Concert (in Polish)," `www.internet10.pl/koncert.html`, 2001, accessed: 2023-04-05.

[2] Gurevich, M., Donohoe, D., and Bertet, S., "Ambisonic spatialization for networked music performance," in *Proceedings of the 17th International Conference on Auditory Display (ICAD-2011)*, International Community for Auditory Display, 2011.

[3] Frank, M. and Sontacchi, A., "Case study on ambisonics for multi-venue and multi-target concerts and broadcasts," *Journal of the Audio Engineering Society*, 65(9), pp. 749–756, 2017.

[4] International Conference on Spatial Audio, "PURE Ambisonics Concert & the Night of Ambisonics," `ambisonics.iem.at/icsa2015/pure-ambisonics-concert`, 2015, accessed: 2023-04-05.

[5] Rudrich, D., Zotter, F., and Frank, M., "Efficient spatial ambisonic effects for live audio," in *Proceedings of the 29th Tonmeistertagung—VDT International Convention, Cologne, Germany*, pp. 17–20, 2016.

[6] Aoki, S., Cohen, M., and Koizumi, N., "Design and Control of Shared Conferencing Environments for Audio Telecommunication Using Individually Measured HRTFs," *Presence: Teleoperators and Virtual Environments*, 3(1), pp. 60–72, 1994, doi:10.1162/pres.1994.3.1.60.

[7] Buxton, W., "Telepresence: Integrating shared task and person spaces," in *Proceedings of graphics interface*, volume 92, pp. 123–129, Citeseer, 1992.

[8] Durlach, N., Shinn-Cunningham, B., and Held, R., "Supernormal Auditory LocalizationI. General Background," *Presence: Teleoperators and Virtual Environments*, 2(2), pp. 89–103, 1993, doi: 10.1162/pres.1993.2.2.89.

[9] Durlach, N., "Auditory Localization in Teleoperator and Virtual Environment Systems: Ideas, Issues, and Problems," *Perception*, 20(4), pp. 543–554, 1991, doi:10.1068/p200543, pMID: 1771137.

[10] Jouppi, N. P. and Pan, M. J., "Mutually-immersive audio telepresence," in *Audio Engineering Society Convention 113*, Audio Engineering Society, 2002.

[11] Wenzel, E. M., Wightman, F. L., and Kistler, D. J., "Localization with non-individualized virtual acoustic display cues," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 351–359, 1991.

[12] Rottondi, C., Chafe, C., Allocchio, C., and Sarti, A., "An Overview on Networked Music Performance Technologies," *IEEE Access*, 4, pp. 8823–8843, 2016, ISSN 2169-3536, doi:10.1109/ACCESS.2016.2628440.

[13] Jacuzzi, G., Brazzola, S., and Kares, J., "Approaching Immersive 3D Audio Broadcast Streams Of Live Performances," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[14] von Coler, H., Tonnätt, N., Kather, V., and Chafe, C., "SPRAWL: A Network System for Enhanced Interaction in Musical Ensembles," in *Proceedings of the 18th Linux Audio Conference (LAC-20)*, 2020.

[15] Rossetti, D. and Bomfim, C. C., "Live Electronics, Audiovisual Compositions, and Telematic Performance: Collaborations During the Pandemic," *Journal of Network Music and Arts*, 3(1), p. 3, 2021.

[16] Hoy, R. and Nort, D. V., "A Technological and Methodological Ecosystem for Dynamic Virtual Acoustics in Telematic Performance Contexts," in *Audio Mostly 2021*, ACM, 2021, doi:10.1145/3478384.3478425.

[17] Cairns, P., Hunt, A., Cooper, J., Johnston, D., Lee, B., Daffern, H., and Kearney, G., "Recording Music in the Metaverse: A case study of XR BBC Maida Vale Recording Studios," in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Audio Engineering Society, 2022.

[18] Hupke, R., Preihs, S., and Peissig, J., "Immersive Room Extension Environment for Networked Music Performance," in *Audio Engineering Society Convention 153*, Audio Engineering Society, 2022.

[19] Cychnerski, J. and Mróz, B., "Architecture Design of a Networked Music Performance Platform for a Chamber Choir," in *European Conference on Advances in Databases and Information Systems*, pp. 437–449, Springer, 2022, doi:10.1007/978-3-031-15743-1_40.

[20] Bosi, M., Servetti, A., Chafe, C., and Rottondi, C., "Experiencing Remote Classical Music Performance Over Long Distance: A JackTrip Concert Between Two Continents During the Pandemic," *Journal of the Audio Engineering Society*, 69(12), pp. 934–945, 2021, doi:10.17743/jaes.2021.0056.

[21] Young, C., "Blue Note, Audinate, Peltrix Team to Beat Latency for Long-Distance Performances," `mixonline.com/live-sound/blue-note-audinate-peltrix-latency-danteav`, n.d., accessed: 2023-04-05.

[22] Audinate, "Webinar: Peltrix takes Dante over distance with Blue Note Entertainment Group," `audinate.com/webinar-peltrix-takes-dante-over-distance-with-blue-note-entertainment-group`, n.d., accessed: 2023-04-05.

[23] Zoom Corporation, "H2n handy recorder - spatial audio for VR," `zoomcorp.com/en/jp/handy-recorders/handheld-recorders/h2n-handy-recorder/`, n.d., accessed: 2023-04-05.

[24] Zoom Corporation, "H3-VR audio recorder," `zoomcorp.com/en/us/handheld-recorders/handheld-recorders/h3-vr-360-audio-recorder/`, n.d., accessed: 2023-04-05.

[25] Voyage Audio, "Spatial Mic," `voyage.audio/spatialmic/`, n.d., accessed: 2023-04-05.

[26] Insta360, "Insta360 Pro2 manual - recording equipment," `onlinemanual.insta360.com/pro2/en-us/video/prepare/4`, n.d., accessed: 2023-04-05.

[27] Insta360, "Insta360 Titan manual - recording panoramic sound," `onlinemanual.insta360.com/titan/en-us/video/production/2`, n.d., accessed: 2023-04-05.

[28] Facebook Audio 360, "Using an Ambisonic Microphone With Your Live 360 Video on Facebook," `facebookincubator.github.io/facebook-360-spatial-workstation/KB/UsingAnAmbisonicMicrophone.html`, n.d., accessed: 2023-04-05.

[29] YouTube Help, "Live encoder settings, bitrates, and resolutions," support.google.com/youtube/answer/2853702, n.d., accessed: 2023-04-05.

[30] Chopin University of Music, "UMFC VR - Koncert Muzyki Dawnej," facebook.com/umfcvr/videos/715647085996863/, 2020, accessed: 2023-04-05.

[31] Gdańsk University of Technology, "Anniversary concert for the 70 years of the Faculty of Electronics, Telecommunications and Computer Science," eti.pg.edu.pl/70-lat/wydarzenia/koncert-akademickiego-choru-politechniki-gdanskiej-10032022, 2022, accessed: 2023-04-05.

[32] Chopin University of Music, "Pleiades / Persephassa: Iannis Xenakis in memoriam," chopin.edu.pl/szczegoly-wydarzenia/524_pleiades-persephassa, 2022, accessed: 2023-04-05.

[33] pkviet, "OBS Studio Music Edition," github.com/pkviet/obs-studio, n.d., accessed: 2023-04-05.

[34] Lindsay, R., "Earshot - Envelop Ambisonic RTMP Streaming Higher-Order Transcoder," github.com/EnvelopSound/Earshot, n.d., accessed: 2023-04-05.

[35] Nginx, "NGINX-based Media Streaming Server," github.com/arut/nginx-rtmp-module, n.d., accessed: 2023-04-05.

[36] Deppisch, T. and Meyer-Kahlen, N., "HOAST360 - higher-order Ambisonics, 360° video player," github.com/thomasdeppisch/hoast360, n.d., accessed: 2023-04-05.

[37] Deppisch, T., Meyer-Kahlen, N., Hofer, B., Łatka, T., and Żernicki, T., "HOAST: A higher-order ambisonics streaming platform," in *Audio Engineering Society Convention 148*, Audio Engineering Society, 2020.

[38] Valin, J.-M., Maxwell, G., Terriberry, T. B., and Vos, K., "High-Quality, Low-Delay Music Coding in the Opus Codec," 2016, doi:10.48550/ARXIV.1602.04845.

[39] Narbutt, M., O'Leary, S., Allen, A., Skoglund, J., and Hines, A., "Streaming VR for immersion: Quality aspects of compressed spatial audio," in *2017 23rd International Conference on Virtual System & Multimedia (VSMM)*, IEEE, 2017, doi:10.1109/vsmm.2017.8346301.

[40] Insta360, "Insta360 Pro2 - 8K 3D professional 360 camera," insta360.com/product/insta360-pro2/#pro2_specs, n.d., accessed: 2023-04-05.

[41] Zylia, "ZM-1 3rd order ambisonic microphone array," zylia.co/zylia-zm-1-microphone.html, n.d., accessed: 2023-04-05.

[42] Reaper, "ReaRoute ASIO Driver and ReaStream Network Audio Plugin," reaper.blog/2022/01/rearoute-reastream-tutorial/, n.d., accessed: 2023-04-05.

[43] Rudzki, T., Gomez-Lanzaco, I., Stubbs, J., Skoglund, J., Murphy, D. T., and Kearney, G., "Auditory Localization in Low-Bitrate Compressed Ambisonic Scenes," *Applied Sciences*, 9(13), p. 2618, 2019, doi:10.3390/app9132618.

[44] Rudzki, T., Gomez-Lanzaco, I., Hening, P., Skoglund, J., McKenzie, T., Stubbs, J., Murphy, D., and Kearney, G., "Perceptual Evaluation of Bitrate Compressed Ambisonic Scenes in Loudspeaker Based Reproduction," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[45] Lee, B., Rudzki, T., Skoglund, J., and Kearney, G., "Context-Based Evaluation of the Opus Audio Codec for Spatial Audio Content in Virtual Reality," *Journal of the Audio Engineering Society*, 71(4), pp. 145–154, 2023, doi:10.17743/jaes.2022.0068.

[46] IEM Plug-in Suite, "MultiEncoder plug-in descrpition," plugins.iem.at/docs/plugindescriptions/#multiencoder, n.d., accessed: 2023-04-05.

[47] Audinate, "Dante Virtual Soundcard, pro AV networking software," audinate.com/

products/software/dante-virtual-soundcard, n.d., accessed: 2023-04-05.

[48] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkman, F., and Weinzierl, S., "A Spatial Audio Quality Inventory (SAQI)," *Acta Acustica united with Acustica*, 100(5), pp. 984–994, 2014, doi:10.3813/aaa.918778.

[49] Deppisch, T., "Plugin for using WebXR with videojs, based on videojs-vr," github.com/thomasdeppisch/videojs-xr, n.d., accessed: 2023-04-05.

[50] Barstow, A., "A plugin to add 360 and VR video support to video.js." github.com/videojs/videojs-vr, n.d., accessed: 2023-04-05.

[51] ThreeJS, "ThreeJS - JavaScript 3D Library," threejs.org/, n.d., accessed: 2023-04-05.

[52] Rantakokko, T., "A repository for Orion360 SDK's test content," github.com/FinweLtd/orion360-sdk-test-content, n.d., accessed: 2023-04-05.

[53] Dear Reality, "DearVR AMBI MICRO - ambisonic toolbox," dear-reality.com/products/dearvr-ambi-micro, 2022, accessed: 2023-04-05.

[54] Blue Ripple Sound, "O3A Flare," blueripplesound.com/plugin/o3a_flare, n.d., accessed: 2023-04-05.

[55] Soluciones Gráficas por Ordenador, "MistikaVR - Industry-adopted Optical Flow Image Stitching," sgo.es/mistika-vr/, n.d., accessed: 2023-04-05.