Meyer-Kahlen, Nils; Piiroinen, Petra; Vishwanath, Gautam; Juntunen, Petri; Tiainen, Eero; Schlecht, Sebastian

Inside The Quartet-A first-person virtual reality string quartet production

Published in:
Proceedings of the Audio Engineering Society Convention 154

Published: 13/05/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Inside The Quartet - A first-person virtual reality string quartet production

Nils Meyer-Kahlen[*], Petra Piiroinen[†], Gautam Vishwanath, Petri Juntunen, Eero Tiainen, Sebastian J. Schlecht[‡]

March 6, 2024

"Inside the Quartet" is a virtual reality experience that provides users with a first-person perspective of playing in a string quartet. The user can see and hear the music performance from the different players' perspectives, experiencing how musicians communicate with gestures and eye contact from within the quartet. The VR experience shows recordings of the Kamus string quartet playing Sibelius's string quartet in D-Minor op. 56 "Voces Intimae" in Ainola, Jean Sibelius's home in Järvenpää, Finland. The audio production employed spot mics encoded to 5th-order Ambisonics, which were dynamically decoded to headphones during playback. The installation was shown at various music and VR festivals. Here, we describe the audio and video production, rendering workflow, and the challenges and future perspectives for similar VR experiences.

1 Introduction

"Inside the quartet" is a virtual reality experience, allowing the user to observe a string quartet from the perspective of one of the four performers. The vision of this VR experience is to highlight the communication aspect of playing in a small chamber music formation. For example, the user experiences eye contact and gestures helping to synchronize the expression of the music. The spatial audio reproduction captures the intimate acoustics of a historic domestic environment rather than that of a concert hall. This provides the chance to listen to musical details in a highly professional string quartet performance with unusual clarity.

To achieve this, the Kamus string quartet was recorded at Ainola, the personal home of Jean Sibelius in Järvenpää, Finland, playing Sibelius's string quartet in D-Minor op. 56 "Voces Intimae". Four recordings were made, each removing one player from the 360-degree video using a production trick described below. To connect the spectator to the first-person experience, it is important to convey the idea that one will be, in fact, a member of the string quartet. The performance is preceded by a short documentary introducing the place and music to the spectator.

Various audio recording techniques were explored, including spherical microphone arrays, spot microphones, and clip-on microphones. After testing different approaches, the audio production employed spot mics encoded to 5th-order Ambisonics. During playback, the Ambisonics audio is dynamically decoded to headphones. The final installation shows the VR experience on four Oculus Quest 2 arranged in a square, similar to the performance setup. Users can choose the instrument by choosing the seat in the performance. Fig. 1 shows the spectator's view when selecting the second violin's perspective. The installation has been shown at multiple music and VR festivals and is due to tour abroad.

We first describe the artistic concept in Section 2. Section 3 discusses the technical realization, including audio and video production, as well as the the mixing and rendering workflow. Section summarizes user experiences and mentions challenges and perspectives for similar VR experiences. Section 5 concludes the report.

A short trailer of the experience is available online at `https://www.sebastianjiroschlecht.com/project/insidethequartet/`

2 Artistic Concept

This section presents the artistic idea and the VR experience's structure.

2.1 Background and Artistic Idea

In 2018, the Kamus string quartet (Terhi Paldanius and Jukka Untamala, violin; Jussi Tuhkanen, viola and Petja Kainulainen, cello) started their tenure as the artistic directors of Our Festival (in Finnish: Meidän Festivaali), an annual chamber music event arranged in Järvenpää and Tuusula by Lake Tuusula, Finland.

As one of their first intiatives, Kamus suggested using VR as a medium to shift the perspective of

Figure 1: 360 degree screenshot from inside the VR experience. Here, the user took the place of the second violin.

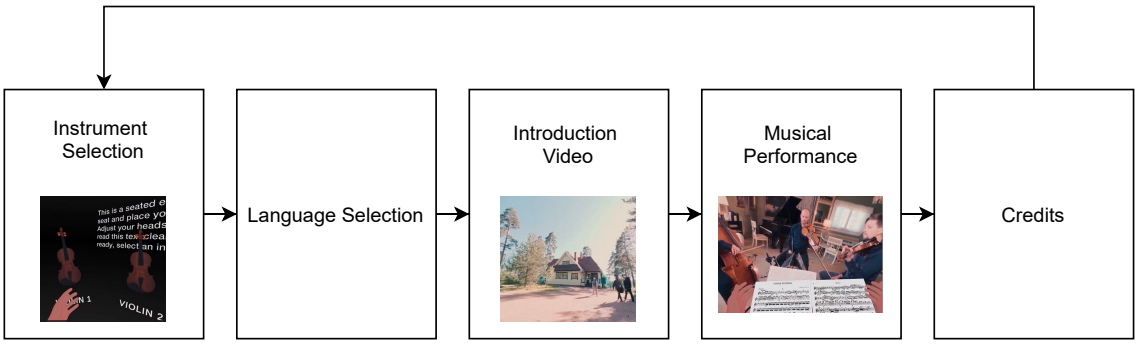

Figure 2: Structure of the experience.

the festival audience in chamber music performances. Usually, the audience is only allowed to assume the role of the spectator. The concept of this production was to allow the audience to experience a string quartet from within.

A similar approach was taken for the "Virtual Orchestra" by the London Philharmonia, a project in which the user views the performance from the middle of the symphony orchestra [1]. "Inside the Quartet" takes this experience into a more intimate chamber music realm. The artistic goal of the project was to break the barrier between the audience and the artist intentionally and to offer Our Festival's audiences an intimate view of Kamus' artistic work.

While the main goal of the experience was to transport the user into the string quartet, utilizing the opportunity to introduce the user to one of Our Festival's venues was considered important additional value. Our Festival organizes its annual concerts in historic artist villas in the lake Tuusula area. Finland's most notable artists took residency by the lake in the 1900s.

The experience was realized at Ainola[1], the home of the composer Jean Sibelius' family from 1904 to 1969. On the one hand, recording Sibelius' music in his own living room added to the experience. On the other hand, capturing in a protected museum environment, with limited time available posed a challenge to the production.

The music choice was largely affected by the venue choice. The Voces Intimae (eng. "Inner Voices") quartet is Sibelius's only full-length string quartet work and a staple of the Finnish string quartet repertoire. Kamus recorded the Voces Intimae in 2015 and performs it regularly, which made it reasonably easy to rehearse for capturing in one take. Furthermore, the work has clear and continuously shifting roles for all players, making it an interesting one to explore from the musicians' point of view.

In the early stages of the artistic planning, it was also proposed that the user could have an opportunity to choose a piece from a classical or contemporary option, offering very differing affects to be experienced: a more linear, melodic music or a more complex composition, both requiring somewhat different skills from the musicians. The final solution of playing the first and second movement of Sibelius' Voces Intimae offers a similar proposition, as the first movement is played in tempo marking andante – allegro molto

[1] https://www.ainola.fi/

moderato (moderately slow – moderately fast) and the second movement in vivace (fast, lively).

2.2 Structure

The VR experience has four main parts: 1) instrument slection menu, 2) language selection menu, 3) introduction video, and 4) music performance, ilustrated in Fig. 2. The instrument menu lets the spectator choose one of the string quartet instruments by touching a corresponding virtual object. The language menu gives the option to choose between English and Finnish. Note that when the experience is shown as an installation with four headsets (shown in Figure 6), the instrument selection is ommited. There, the choice of one of the four headsets determines the instrument.

The introduction shows the quartet members infront of the house, in the kitchen and rehearsing in the living room. It contains a voice-over of the string quartet member who plays the selected instrument. They introduce the quartet, Ainola as a place, and also explain the role of the selected instrument in the music.

Once the introduction video is over (users also have the oppertunity to skip to the end), during the warm-up and tuning, the spectator is handed a visual representation of a virtual string instrument which can be manipulated by hand tracking to enhance the role-play further. Holding the instrument should signal to the user that they will indeed follow the performance as one of the players. As soon as the performance starts, the instrument is replaced by sheet music. The music performance of both movements lasts about 11 minutes and ends with a credit screen. In total, the whole VR experience lasts about 15 minutes.

3 Technical Realization

The technical production encompassed both the recording and postprocessing of 360 videos and spatial audio.

3.1 360 Video Recordings

The stereoscopic 360 video recording was recorded with an Insta 360 Pro [2]. Upon stitching, the video was saved in 360 stereoscopic format with 3840 × 3840 pixels. The introduction scene contains recordings outside and inside of Ainola, e.g., the kitchen and library. The music performance was recorded in the living room.

For the performance, the main challenge was to remove the selected player from the image to evoke a sense of presence in the spectator. At the same time, the string quartet needed to play live for a high sense of interaction and spontaneity. We used the following production trick, which is illustrated in Fig. 3. The video camera was placed in the natural position of the replaced player, while the actual player was sitting behind the video camera (as seen by the other string quartet members). Fig. 4a shows the recording of the second violin perspective. The distance between the replaced player and the video camera was kept short, to keep the interaction natural and capture eye contact between the players faithfully. The rear part of the 360 images, including the player that is to be deleted, was then replaced by a still image of the empty living room. This also allowed for hiding the production staff and other technical equipment such as lighting, microphones, and the production desk, see left of Fig. 4b.

3.2 Audio Recording

The scene was recorded using several different techniques. A few centimeters behind and a few centimeters above the 360 camera, in the angular range in which the video was later replaced with the static image, both an Eigenmike em32 and a Soundfield ST350 microphone array were placed. For the Eigenmike, all 32 microphone signals were recorded, so that encoding to up to fourth order could be done in post-production. The soundfield microphone's four signals were directly encoded to first-order Ambisonics using the matching preamplifier unit.

Additionally, every instrument was equipped with a DPA IMK 4060 clip microphone and an AKG C414 large diaphragm condenser microphone as a spot microphone. Recording were made using a Fireface UFX+ interface and Cockos Reaper.

3.3 Mixing

Several options were tested, like using the Eigenmike, the soundfield microphone, and encoding the spot mics to fourth order Ambisonics. Another tested option involved convolving recordings made with clip-on mics with Ambisonics room impulse responses, obtained by processing measured responses with the Spatial Decomposition Method (SDM) [2].

After playing several variants to the musicians and collecting their comments, the final choice was to use the following setup. It is a good example of realizing an Ambisonics production, with all its benefits like portability and efficient rotation, without using dedicated microphone arrays. Only freely available plugins were used. Fig. 5 shows the signal flow used in this production. It can be seen as one example of the general Ambisonics production signal flow, as it is shown in [3].

The three AKG C414 spot mics for violin 1, violin 2, and viola were encoded to fifth-order Ambisonics using one channel of the IEM Plugins' [3] StereoEn-

[2] `https://www.insta360.com/product/insta360-pro/`

[3] https://plugins.iem.at/

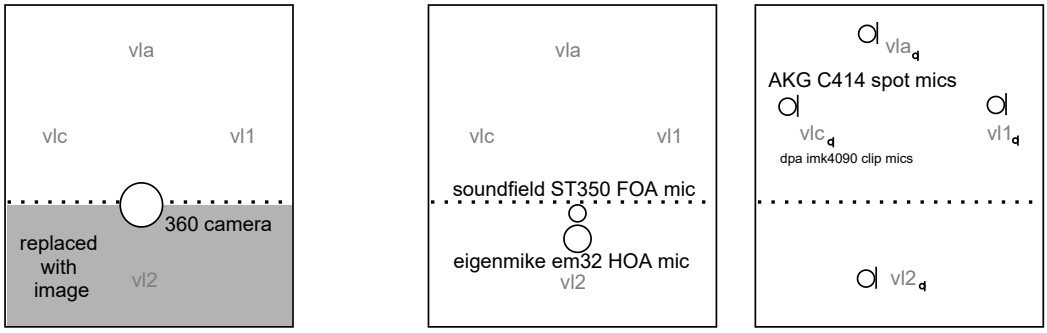

Figure 3: Placement of the equipment. The spot microphones shown on the right were used for the final production.

(a) Recording one of the scenes seen from the dining room. The second violinist on the left is behind the camera and will later be removed.

(b) Part of the setup as seen from the piano room. Setup with the camera on the left and microphone arrays behind it. Spot microphones for the three musicians that are in front of the camera.

Figure 4: Recording setup as seen from different angles at Ainola.

coder VST plugin each. The plugin performs linear Ambisonics encoding (and therefore has the same functionality as Sparta AmbiEnc [4], the Ambix encoder [4] or comparable tools).

For each of the four versions, the encoding positions were selected such that they correspond to the position of the players in the video. For the signal of the own, selected instrument, the DPA spot mic was used and encoded to fifth order as well. A high pass with a cut-off frequency of approximately 50 Hz was applied, as well as a high shelf reducing frequencies above 1 kHz by 3 dB. For the violins and the viola, the own intrument was placed on the side of the listener, slightly below the horizontal plane. The cello was placed in infront, also with a negative elevation.

All the encoded signals were sent to a 36-channel Ambisonics bus. Additionally, all encoded instrument channels except the own one were sent to a custom version of the IEM FDNReverb. It was set to use 64 channels and had a reverberation time of 0.8 s. What sets the custom version apart from the version available in the toolbox is that it has improved fade-in time control, as described in [5].

During mixing, the SPARTA AmbiBIN decoder was used for monitoring. Upon bouncing, the decoder was bypassed, and the final file was saved to a 36-channel wav file, using Ambisonics Channel Numbering (ACN) and

3.4 Audio Rendering

Upon loading the Ambisonics file in Audiokinetic wwise, it was encoded to the opus format (as it is also used in [6]), because the file size of an uncompressed Ambisonics file would otherwise have been to large to run the experience. Lossy coding of Ambisonics scenes is expected to cause very minor spatial and some timbral degredations in complex scenes [7], depending on the complexity of the scene and the used bitrate.

Scene rotation of the Ambisonics material is set to follow the orientation of the camera, i.e., the user.Together with the soundtrack of the introduction video and required button sounds, a soundbank was created that is then incorporated into the unity project.

3.5 Final VR Production

The audio-visual material was combined with interactive elements using the Unity game engine. The

[4] http://www.matthiaskronlachner.com/?p=2015

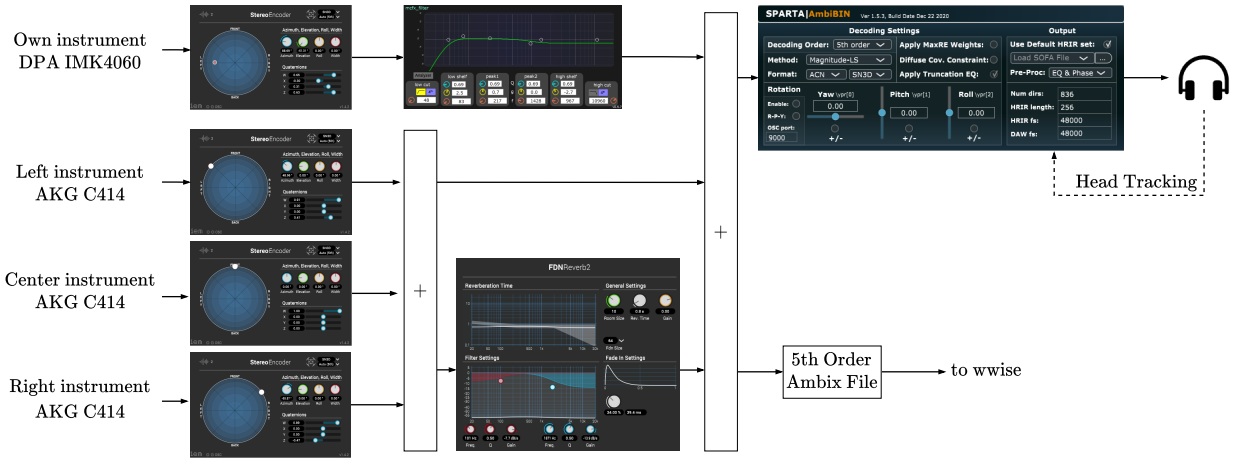

Figure 5: Signal flow

Figure 6: Installation

user interface can be navigated by hand tracking and touching virtual objects. The virtual instrument that is displayed at the beginning of the performance can be moved by placing the hands into a typical playing position. During the performance, hand tracking allows for holding the sheet music. For the installation version, the finalized VR app was uploaded to four Oculus Quest 2 headsets. Therein, the music is listened to via noise-canceling headphones (Sony WH-1000XM).

4 User Experience

The experience was produced first and foremost as a tool for audience development for Our Festival, both for audience outreach and as an educational tool to deepen the experience of the festival's existing audience. The installation of four VR headsets with four headphones, set up with music stands to reminisce the user of the string quartet, is very mobile and easy to tour by design to reach audiences anywhere. In Our Festival 2022, the experience was displayed alongside several Festival concerts in the concert venues and as pop-ups in public libraries in the area, also functioning as a visible marketing tool. It has also visited elderly homes in the Lake Tuusula area with great success. The production has attracted interest in the art music field and has been showcased at several industry conferences in Finland and internationally. Currently, the installation is also gaining international interest in concert venues and festivals in Europe. See Fig. 6 for a picture from the installation in Budapest.

The artistic goal was to break the barrier between the audience and the performers, allowing a string quartet experience from within. Following conversations with users of the experience so far, the production has met this goal. Users did understand that they were part of the quartet. Some highlighted the audio quality and the opportunity to listen to fine details of the playing. Spectators who were able to read music were often more engaged, as they were able to follow the score. For non-musicians and people with low interest in music, the 11 min performance might seem a bit long, although leaving the experience before the ending was very rare.

Besides a lot of positive feedback, some users have expressed confusion about the instrument model that is shown in the beginning of the performance. They felt pressured to do something with it and thought that the model would be able to produce sound. Also, while hand tracking in the instrument selection menu often worked well, some users experienced difficulties reaching the models and understanding that they needed to touch them for a few moments before the instrument is selected. Especially first time VR users needed some encouragement to reach out for a virtual object. For a more in-depth analysis, formal evaluation regarding the quality of experience would need to be conducted in a dedicated study.

Furthermore, a future version of the experience is imaginable, in which the own instrument sound is not played back, so that interested musicians, for example students, could play along. Similar experi-

ences could be built for students at different levels of musical skill. For some instruments and less advanced musicians, such experiences may become even more relevant as pass-through functionality is becoming more accesible with new mixed reality devices, allowing players to see their own, real instrument.

5 Conclusion

We have decribed the concept and realization of the first-person, virtual reality string quartet experience "Inside the quartet".

It allows the user to experience the Kamus string quartet from within. The video is realized using a 360 degree camera recording, in which part of the scene is replaced by a static image. The audio production uses spot microphones that are then encoded to 5th order Ambisonics, which is dynamically decoded during playback. The experience has found an audience in Finland and abroad, receiving positive feedback from the majority of users.

Acknowledgements

We would like to thank the Kamus string quartet for the opportunity to realize the production and Julia Donner from Ainola, home of Aino and Jean Sibelius for welcoming us. Furthermore we are thankful for project funding from Jenny and Antti Wihuri Foundation supporting Our Festival in this project. NMK's contribution was funded by VRACE under the EU's Marie Skłodowska Curie actions (grant nr. 812719).

References

[1] "Philharmonia - Immersive," `https://philharmonia.co.uk/what-we-do/digital-immersive/immersive/`, 2014, accessed: 2023-03-23.

[2] Tervo, S., Tynen, J. P., Kuusinen, A., and Lokki, T., "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, 61(1), pp. 17–28, 2013.

[3] Frank, M., Zotter, F., and Sontacchi, A., "Producing 3D Audio in Ambisonics," in *AES 57th International Conference: The Future of Audio Entertainment Technology – Cinema, Television and the Internet*, 2015.

[4] McCormack, L. and Politis, A., "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *AES International Conference on Immersive and Interactive Audio*, York, UK, 2019.

[5] Meyer-Kahlen, N., Schlecht, S. J., and Lokki, T., "Fade-In Control for Feedback Delay Networks," in *DAFx*, 2020.

[6] Deppisch, T., Meyer-Kahlen, N., Hofer, B., Łatka, T., and Żernicki, T., "HOAST: A Higher-Order Ambisonics Streaming Platform," in *148th AES Convention*, 2020.

[7] Rudzki, T., Hening, P., Gomez-Lanzaco, I., Stubbs, J., McKenzie, T., Skoglund, J., Murphy, D., and Kearney, G., "Perceptual Evaluation of Bitrate Compressed Ambisonic Scenes in Loudspeaker Based Reproduction," 2019.