applied sciences

MDPI

*Case Report*

# Creating a Remote Choir Performance Recording Based on an Ambisonic Approach

**Bartłomiej Mróz [1,2], Piotr Odya [1,*] and Bożena Kostek [2]**

[1] Department of Multimedia Systems, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 11/12 Narutowicza St., 80-233 Gdansk, Poland; bartlomiej.mroz@pg.edu.pl
[2] Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 11/12 Narutowicza St., 80-233 Gdansk, Poland; bokostek@audioakustyka.org
* Correspondence: pioodya@pg.edu.pl

**Featured Application: It is shown that quality of experience (QoE) techniques, such as eye-tracking, might be applicable to measure the ease of performing in a remote-like simulated environment setting. This is backed up by the results of the eye-tracker-based examination with choristers, which reveal that regardless of the virtual location of the recording, almost all persons generally focus on the conductor, similarly as in regular performance situations. The fixation point remained on the conductor's face, but less on his hands, which were used to convey expressiveness inscribed within the music notation. In such a setting, participants' familiarity with the sung piece and their ability to read music notes fluently may be accounted for.**

**Abstract:** The aim of this paper is three-fold. First, the basics of binaural and ambisonic techniques are briefly presented. Then, details related to audio-visual recordings of a remote performance of the Academic Choir of the Gdansk University of Technology are shown. Due to the COVID-19 pandemic, artists had a choice, namely, to stay at home and not perform or stay at home and perform. In fact, staying at home brought in the possibility of creating and developing art at home while working online. During the first months of lock-down, the audience was satisfied with music performances that were fairly far from the typical experience of a real concert hall. Then, more advanced technology was brought to facilitate joint rehearsal and performance of better quality, including multichannel sound and spatialization. At the same time, spatial music productions benefited from the disadvantage of remote rehearsal by creating immersive experiences for the audience based on ambisonic and binaural techniques. Finally, subjective tests were prepared and performed to observe performers' attention behavior divided between the conductor and music notation in the network-like environment. To this end, eye-tracking technology was employed. This aspect is related to the quality of experience (QoE), which in the performance area–and especially in remote mode–is essential.

**Keywords:** ambisonics; virtual concert; remote music performance; eye-tracking; quality of experience; networked music performance



**Citation:** Mróz, B.; Odya, P.; Kostek, B. Creating a Remote Choir Performance Recording Based on an Ambisonic Approach. *Appl. Sci.* **2022**, *12*, 3316. https://doi.org/10.3390/app12073316

Academic Editor: Alexandros A. Lavdas

Received: 11 February 2022
Accepted: 23 March 2022
Published: 24 March 2022



## 1. Introduction

In recent years, advances in technology have resulted in the increasing use of real-time remote working devices. Nevertheless, remote creative and artistic work, particularly musical performance, remains a challenge. One-way connections, through which artistic events are transmitted in real-time, have become available using networks capable of supporting high bandwidth, low-latency packet routing, and guaranteed quality of service (QoS) [1–3]. These can be described as unidirectional, allowing artists to remotely connect only between the performance venue and the audience, not each other. It is much more difficult when performers in two or more remote locations attempt to perform an established composition

or improvisation together in real-time. Weinberg [4] calls this way of performing music the "bridge approach". In such cases, the inevitable delay caused by the physical transit time of network packets is known to affect performance [5–10]. Therefore, attempts have been made to account for these delays by design or composition [11–13].

This paper briefly provides an overview of distance concert performance and the difficulties involved. It also refers to the theoretical basis of ambisonics, which is one of the sound reproduction formats that focuses on listener immersion in the sound scene. The process of making the recording of the Academic Choir of Gdansk University of Technology concert is shown in detail, starting from the preparation stage, through recording the musicians, to postproduction. It includes a description of the sound and video engineer's workshop and discusses problems related to sound synchronization. Finally, the eye-tracker is employed to observe performers' attention behavior divided between the conductor and music notation in the network-like environment. This aspect is related to the quality of experience (QoE), which in the performing domain is essential.

The summary presents overall conclusions concerning the design of subjective tests to evaluate the realized recordings and further plans related to the Academic Choir of Gdansk University of Technology concerts.

## 2. Performing Remote Concerts

It is worth mentioning that the first implementations of remote concerts were applied to academic networks, and an experimental concert with artists (AMFC Vocal Consort, Schola Cantorum Gedanensis, and Capella Cracoviensis) from different cities in Poland took place in 2001 [14]. On the occasion of the 10th anniversary of the Internet in Poland, the Research and Academic Computer Network (NASK) and the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw (ICM), in collaboration with the Fryderyk Chopin Academy of Music, organized an experimental Internet concert performed in three cities (Warsaw, Gdansk, and Cracow) using Internet connectivity on 14 September 2001. The concert included the world premiere of Stanisław Moryto's piece Ad Laudes, written especially for the 10th anniversary of the Internet, but at the same time to commemorate the victims of the terrorist attacks of 11 September 2001. The sound and image were transmitted via the Internet via the POL-34 network. The musicians saw each other on monitors while the audience gathered in the three cities watched the concert on plasma screens. Simultaneously, the broadcast was carried out by all public Polish Television regional channels. The artistic direction was assumed by the conductor R. Zimak [14]. The technical layer of the network connection is also worth mentioning. The channel bitrate was set to 15 Mbps. The video signal was encoded in the MJPEG standard, which ensured minimal delay in transmission—the most critical factor for the success of the experiment. Realistically, delays of about 180 ms were obtained between the cities in one direction, of which about 70 ms were consumed by encoding and decoding the signal.

In streaming, however, the spatial layer of sound sources has only recently become technically available. Binaural and ambisonic techniques as tools for localizing remote audio sources have settled fastest in teleconferencing applications [15–20]. We should mention the AltSpaceVR platform [21], in which teleconference meetings, demonstrations, presentations, classes, or social gatherings can be conducted; the audio of all objects in the platform is subject to auralization and binauralization.

In the case of the so-called Networked/Remote Music Performance (NMP) [22], successful attempts have already been made to transmit multichannel audio [23–26]. Transmissions using ambisonics have also been performed [27]. Ambisonics can be thought of as a transmission format for creating 3D spatial audio. It is based on the representation of the sound field by decomposing it into orthonormal basis functions—called "spherical harmonics". This representation enables a flexible production process that is independent of the target playback system (speaker system or headphones). 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 [28–30]. 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 the ICSA 2015 conference 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 [28].

However, all of these presentations were experimental. To become a standard for recording or broadcasting, they must be implemented on commonly used platforms. For example, on the YouTube platform, only live streaming with 360° video alone is possible (sound must be in stereo). First-order ambisonic audio can only be used when uploading a finished recording [31]. Another platform, Facebook, supports live streaming with 360° images and first-order ambisonics [32], while the uploaded finished recording can contain second-order ambisonic sound [32,33]. There is also a completely new platform focused on the presentation of content with spherical video and audio mixed in higher-order ambisonics called "HOAST" [34]. It is worth noting, however, that all of these technologies are not strictly about connecting musicians located in separate locations but rather about virtually mapping the stage with musicians for audiences located in distant concert halls. The stumbling block to full NMP continues to be latency on the order of 25 ms, particularly when using popular consumer connections such as ADSL. Only 5G technology offers hope to overcome these difficulties [35].

Therefore, remote musical performances are present in the form of recordings created in postproduction based on recordings of individual parts of songs sent by their performers. The best known are the achievements of composer E. Whitacre, who has been organizing concerts of the so-called virtual choir since 2009 [36]. Six editions of these sessions have been created, and thousands of people from all over the world have participated in each of them. Especially the last one, the 6th edition, received significantly more interest, which was due to the fact that it took place in the spring of 2020, when COVID-19 was considered a global pandemic. At that time, musicians, in particular, began to investigate alternative, remote performance methods; the demand for such solutions can be confirmed by the fact that Eric Whitacre's Virtual Chorus, sixth edition, attracted over 40,000 choristers from 145 countries. Nevertheless, no remotely performed recording has been spatialized using ambisonics. It was not until the COVID-19 pandemic that the first such productions were made [37–41]. One of them is Giovanni Pierluigi da Palestrina's piece, Sicut Cervus, performed by the Academic Choir of the Gdansk University of Technology in a pioneering recording, called "Virtual Cathedral" [42]. A discussion of the production of this recording will be presented in the following sections.

It should be pointed out that investigations concerning performing remotely in the context of quality of experience (QoE) [43] are rare or non-existent in the literature. They appear mainly in identifying problems in the learning environment [44,45], where a student's attention or collaboration is essential. In contrast, such remote strategies are well-adapted to medical applications, where they are especially used for surgical planning with simulation instruments, including VR (Virtual Reality) and AR (Augmented Reality), are used on a daily basis [46]. Moreover, checking audio-visual quality often employs QoE methodology [47–51]. When discussing QoE, one should consider at least several factors that quality encompasses. These are overall perception, acceptability, presence, immersion, to name a few [52]. Meghanathan et al. measured some of these aspects by employing a VR headset with an eye tracker [50].

It is known that one of the parameters evaluating artists' performances, whether a choir or an orchestra, is called 'ensemble'. Moreover, the conductor guides all the performers and strengthens the bond between the artists and the audience. When artists perform remotely and separately, such an ambiance may not be present. So, a question arises about how to measure QoE in such a case. In many studies, head-/gaze-/eye-tracking technology,

a video-based eye-tracking relying on locating pupil center to measure gaze, is a highly exploited technology in the context of QoE and objectivization of audio-visual experiments and subjective tests [44,49,51,53–57], so we decided to apply an eye-tracker to observe the choristers' performance.

## 3. Basics of Ambisonics

Ambisonics was proposed by M. Gerzon, P. Felgett, and G. Barton in the early 1970s at the Universities of Oxford and Surrey [58]. Ambisonics refers to "periphony", a technique for reproducing sound both vertically and horizontally around the listener. As defined as follows: Ambisonics is a full-sphere surround sound format referring to a soundstage represented by a set of so-called spherical harmonics. In contrast to conventional stereo and surround formats (which rely on the principle of sending sound signals to specific speakers), ambisonics captures full directionality information for each sound wave that is captured by the microphone—including in the vertical plane [59]. In ambisonics, it is important to distinguish between the so-called A-Format and the B-Format. The A-Format is a microphone-specific recording of the signal—it is a direct recording of the signal coming from each microphone capsule. In this form, it is not yet usable, but each manufacturer provides methods for converting their microphone to the ambisonic domain, i.e., to B-Format. In B-Format, the following two methods of channel numbering are distinguished: FuMa (Furse-Malham) and ACN (Ambisonic Channel Number). FuMa numbering uses letter designations, and ACN numbering uses numerical designations; in addition, these channels are given in a different order. For example, the first row of ambisonics has four channels written in FuMa notation WXYZ (W—omni-directional, the so-called omni, X—forward-backward, Y—left-right, Z—up-down). In ACN notation, these channels will be 0, 3, 1, 2, respectively. The number of channels in 2D ambisonics is equal to 2N + 1, while in 3D ambisonics—respectively, as follows: $(N + 1)^2$.

Figure 1 shows an illustration of sixteen (i.e., up to 3rd order) spherical harmonics together with the FuMa and ACN ambisonic channel numbers.

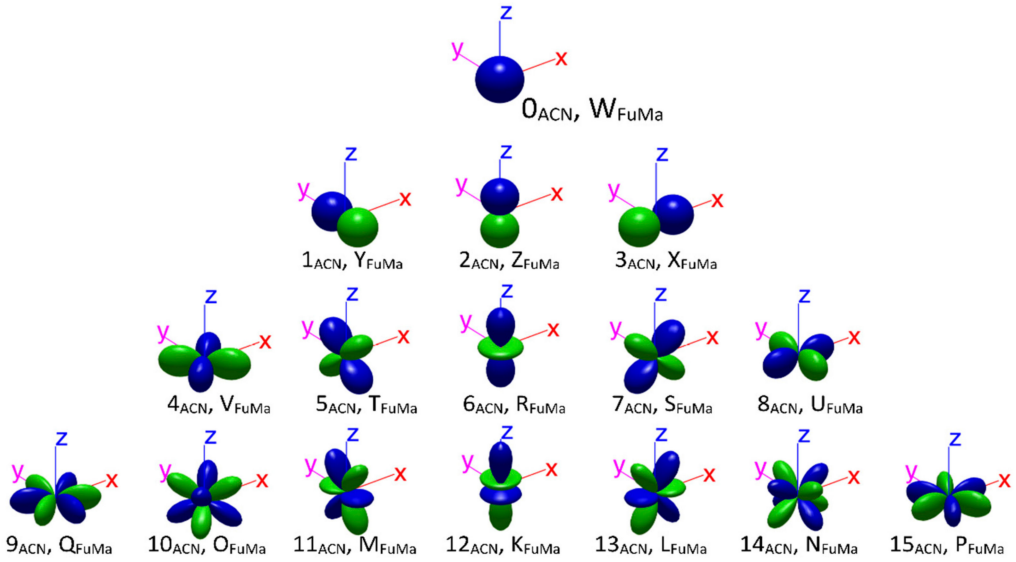


**Figure 1.** Illustration of spherical harmonics with the FuMa and ACN numbering corresponding to channels in ambisonics [60].

Moreover, worth mentioning are the recording formats, ranging from historical UHK to the following current ones: AMB—Microsoft Wave Format Extensible (WAVE-EX), a B-Format recording with the FuMa weighting factors. The main drawback of this file format based on the WAV container developed by Microsoft is the file size limitation of 4 GB. Ambisonic Exchange (AmbiX), on the other hand, uses Apple's (Core Audio Format) as a container.

The channels are normalized according to the format in which the following signals are recorded:

1. MaxN—normalizes each individual component not to exceed a gain of 1.0 for a panned mono source—used in FuMa;
2. N3D—similar to SN3D—orthonormal basis for 3D decomposition. Provides equal power to the encoded components for a perfectly scattered 3D field;
3. SN3D—(in ACN channel order) is widely used. Unlike N3D, no component will ever exceed the peak value of the 0th order component for single point sources. This scheme has been adopted in the AmbiX coding format and is widely used.

However, decoding includes algorithms and tools such as AllRAD [61]/AllRAD2 [62], Harpex [63], and DirAC [64,65]. AllRAP (All-Round Ambisonic Panning) is an algorithm for arbitrary speaker arrangements aimed at creating apparent sources with stable loudness and adjustable width. The AllRAD (All-Round Ambisonic Decoding) method fits the ambisonic format concept. Conventional ambisonic decoding is straightforward only with optimal speaker settings, for which direction-independent energy and energy dissipation, estimated loudness, and width of the apparent source are obtained. The AllRAP/AllRAD algorithm is still simple but more versatile and uses a combination of virtual optimal loudspeaker settings with VBAP (Vector-Base Amplitude Panning). The Harpex (High Angular Resolution Planewave Expansion) decoder, on the other hand, is a tool that combines the spatial sharpness of parametric methods with the physical correctness of linear decoding without introducing audible artifacts. The Directional Audio Coding (DirAC) algorithm uses the Spatial Impulse Response Rendering (SIRR) method. SIRR analyzes room impulse responses in frequency bands, their dispersion, and time-dependent direction of arrival. Based on the analytical data, a multichannel response suitable for reproduction with any chosen surround speaker configuration is synthesized. It should also be added that a particular example of ambisonic decoding is the use of the HRTF function to decode an ambisonic signal into a binaural format [66].

## 4. Realization of Remote Music Recording

There are several important references to audio realization in the context of both concert streaming and the possibility of immersive audio reproduction, i.e., inducing the viewer/listener to be "immersed in sound" [67–69]. Among them is a chapter by S. Meltzer, A. Murtaz, and G. Pietrzyk, entitled "The MPEG-H 3D Audio standard and its applications in digital television" [70]. The authors point out that the immersive audio transmission uses the MPEG-H Audio standard based on Higher-Order Ambisonics (HOA) as the optimal input to the 3D audio codec. The ambisonic technique is designed to record and later reproduce the whole sound field using sound objects. This requires transmitting information about the object's position in space and time in the form of a meta-description. The meta-description enables the correct rendering of objects on the playback side [70].

In this Section, the implementation stage of a concert recording is to be presented, from the preparation stage to the postproduction of the recording.

### *4.1. Preparing a Remote Audio-Visual Recording*

The first step in an audio-visual recording is to discuss the following musical material with the performers: to present the conductor's or ensemble leader's musical interpretation of the material, to indicate the places in the score (possibly marking them) that require special attention, etc. In addition, it is important to discuss the recording technique, the aesthetics of the frame, the location of the recording device, and all the details related to it. On the side of the sound engineer, it is crucial to provide a repository for sending files with recordings by project participants. This is crucial for the sake of work organization—in this way, the performers will not have to find a way to send recordings on their own, which will significantly improve communication in the recording project.

The next step is to create a "model recording"—this is what the performers of the piece will use as a basis. The "model recording" is actually a way to conduct the choir or

ensemble remotely or to provide the musicians with a remote rehearsal. This way, it is easier for singers to reproduce what they hear. Moreover, this is an artistic interpretation of the music by the conductor, specific to this very time and place and not from some other past recording. Very good guidance can be a visual recording, in which the conductor conducts the piece and performs a given piece on the piano from the piano transcription (or another instrument, although the classical piano usually accompanies choral rehearsals—the sound layer can also be performed by the accompanist). Then, with the material constructed in this way—*sound + vision = conductor + accompaniment*—the musicians presenting their parts create their own audio-visual recordings. For example, for a choir, these are soprano, alto, tenor, and bass; for a string quintet, the following: first violin, second violin, viola, cello, and double bass. It is vital that these recordings should be a bit more expressive than usual; to do so, the players of such a "template" will obtain an extra layer of information from which they can reconstruct a common interpretation of the piece. Moreover, it can be beneficial to arrange with the performers to tap out 1–2 bars of tempo (e.g., 1–2 bars before the piece begins) before performing their part in an agreed-upon manner. This will enable the players joining the "pattern" to enter into the rhythm and prepare for the recording. Moreover, in postproduction, it is a good reference for synchronizing the tracks—what is more, it will give a proper space before the start of the performance to edit in postproduction, so the sound engineer will avoid the situation where the recording is too short because someone was not ready to perform their part yet, for example. It is also possible to use a metronome; however, in the authors' experience, this is an unnecessary distraction for the performer due to the handling of an additional device. Additionally, in the recordings presented here, the metronome would probably be implemented using a telephone, which is typically used in the recording in other ways (e.g., as a device to play the pattern or to conduct the recording). Figure 2 illustrates the idea of this approach.

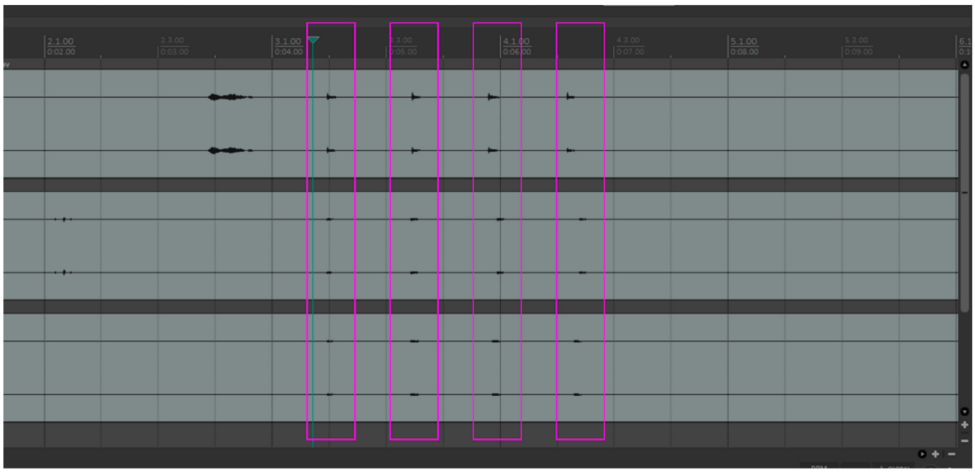

**Figure 2.** Pre-sync tracks to four claps from the beginning of the recording [60].

### *4.2. Technical Considerations*

It is of importance to consider in such a recording hardware specifications. Ideally, each musician should be recorded in a controlled recording environment with hi-fidelity studio equipment. Even more welcome for an Ambisonic realization would be creating such a recording with a spherical microphone grid, thus creating a holographic sound image [71–73]. This would allow reproducing the directivity patterns of the recorded instruments with the highest realism.

The other solution would require mono recordings, preferably anechoic; this would allow for simulating the virtual acoustics as well as directivity of audio objects in the Ambisonic domain. The latter approach was employed in the aforementioned "Virtual Cathedral" recording [42]. However, due to at-home COVID-19 pandemic conditions, the individual tracks are far from perfect. Only a few choir members on this recording had a

semi-professional condenser microphone at home (i.e., Behringer C-1U, Genesis Radium 200, Zoom H2n, Samson C01U Pro, Maono AU-A03); the vast majority could only use smartphones (iPhone 11, iPhone 7, Xiaomi, and other mobile devices with their internal microphones (i.e., iPad 5, Dell Inspiron 15, Samsung Galaxy A5/Samsung Galaxy S7, Samsung Galaxy Note 8, Samsung A40, Nikon D3300, Canon PowerShot G7X mk1, Huawei P20 Pro). These devices have different microphone characteristics, as well as each recording was realized in different acoustic conditions, including background noises such as household appliances working in the distance and passing vehicles. However, because of the more directional characteristics of the microphone, fewer background noises were picked up in the signal. Overall, due to the large variability of sound material quality, each recording required de-noising and pitch, timing, and level corrections (as described in further chapters). A remark may, however, be included here, the literature sources indicate that smartphones are used in music learning and recording, and that there exist records produced entirely with a smartphone, iPad, carrying recording software, etc. [74,75].

### *4.3. Recording of Musicians*

The prepared model recording should be the base to which the other members of the ensemble will add their parts. It is very convenient to put the examples on the YouTube platform as non-public recordings so that during the performance of their part, the musician will be able to plug in small headphones and keep the phone in sight (e.g., close to the music notation) and continuously control the hand movements of the conductor and the person's 'exemplary performance' for their musical part. The illustration in Figure 3 presents a recording situation under the described conditions. The recordings submitted by the musicians should be verified by the conductor and/or the producer of the project in terms of both artistic and technical aspects especially with regard to the aesthetics of the frame and the fidelity of the reproduction of the pattern, especially on the rhythmic side. The more faithful the performance, the fewer corrections need to be applied later in the recording release process.

**Figure 3.** An example of setting up a recording in a home environment; the screen shows the score, the model recording, and a separate device (e.g., laptop) recording the sound [60].

### *4.4. Postproduction of Soundtrack*

It is necessary to correct each of the submitted soundtracks manually in the postproduction phase to achieve the best effect. Unfortunately, as already mentioned, most of the musicians in the academic choir did not have the possibility at home to make a professional recording. Most participants produced their recordings with a smartphone, therefore, most often, they required noise removal. This was performed with noise gates and FIR filters.

Another issue is corrections to the melodic-rhythmic layer itself. While intonation irregularities in large performance groups can be obscured by other more correct performances, rhythmic irregularities are much more noticeable and detract from a good impression of

the production reception. To rectify these problems, available VST plug-ins were used, including the ReaTune plug-in built into the Reaper digital audio workstation (DAW) [76] for correcting melodic lines. Moreover, the Melodyne Studio plug-in [77] appears to be a potent tool that allows improving not only the pitch or duration of a sound but also has algorithms for editing formants, attack time of individual sounds, or transitions between sounds. An illustration of working in this environment is shown in Figure 4.

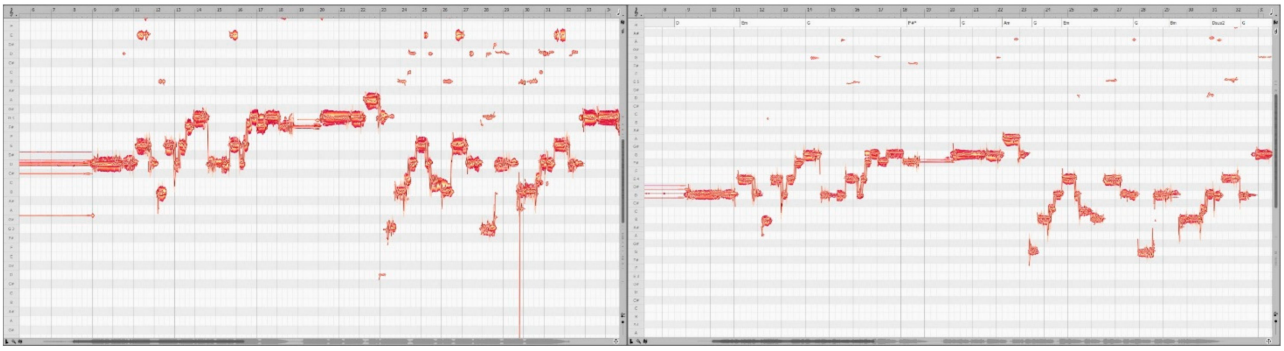

**Figure 4.** Before and after the recording correction in Melodyne Studio (the alto part) [60].

Large performance groups are most often recorded in large concert halls rather than individually, each with their own individual voices. A recording realized this way enables the tracks to be treated objectively and spatialized in ambisonics. It is important to ensure that the acoustics of both the virtual room and the rendered "objects," i.e., the choir members, are as faithful as possible. To this end, a directivity pattern can be given to each object source. In addition, the virtual acoustics can be modeled using apparent reflections and also by adding diffuse reverberation. In the considered recording, it was decided to use the VST plug-ins from the IEM Plug-in Suite [78], in particular, DirectivityShaper [79], RoomEncoder [80], and FDNReverb [81]. Figure 5 shows an example setup of these three VST plug-ins.


**Figure 5.** VST from IEM Plug-in Suite: DirectivityShaper, RoomEncoder, FDNReverb [60].

### *4.5. Postproduction of the Visual Layer*

When producing a video of a classical virtual choir, much attention needs to be paid to planning all the shots and transitions between them. For this recording, it was decided to use a static 360° image, on which the recordings of the choristers would be "suspended". The same people are visible throughout the recording, as transitions between them would be unsightly. Moreover, in a 360° recording, the creator has no control over where the viewer is looking at any given moment. Therefore, the image in the entire sphere was left unchanged. Instead, a refined model was used to generate the background. For that purpose, the 3D model of the cathedral was created, and then Blender was used to render the interior of the cathedral from a given point—a specific camera setting was employed. The camera was set to spherical mode, or more precisely, to cylindrical equirectangular projection. In this way, the resulting visual background could be used as a background for the video timeline in the nonlinear editing software—DaVinci Resolve [82]. Figure 6 presents an arrangement of choristers arranged on the background of the developed graphics.

**Figure 6.** Arrangement of the choir on the background of the rendered view of the cathedral interior model [60].

The model created in this way can then be combined with the audio in B-format. In the case of first-order ambisonics, this is a 4-channel audio track. The FB360Encoder program [83] provided by the Facebook platform can be used for this purpose, as it also supports audio-video encoding for the YouTube platform. After uploading to this platform, the resulting file is automatically recognized as a 360° video with ambisonic sound based on the metadata contained in the file. Once processed by the platform, it is ready for playback.

## 5. Eye-Tracker-Based Examination

The aim of the designed examination was to check whether eye-tracker technology may be employed in observing performers' divided behavioral attention in the context of looking at the conductor and music notation, both visible on the screen in the network-based environment. This is related directly to the guidance role of a conductor. Secondly, it was to evaluate the aspect of quality of experience (QoE).

The examination setup was designed to replicate the at-home recording situation. As shown in Figure 3, a typical home-recording setup consists of a screen displaying the conductor's model recording and, at the same time, showing a sheet of music notes. Moreover, participants used headphones to provide auditory input from the model recording. Therefore, the eye-tracked environment attempted to replicate these conditions as accurately as possible.

The testing environment comprised a Tobii EyeX eye-tracker, a PC running Windows 10 along with the eye-tracking software, and a pair of consumer-grade closed-back headphones (namely, Sennheiser HD380 Pro). Moreover, an Apogee ONE USB audio interface was mounted on a tripod, used for convenience, as it has an easily accessible volume knob, so the participants could comfortably adjust the loudness of the audio played back. The environment is presented in Figure 7. In Figure 8, the visual input is shown.

**Figure 7.** The setup of the eye-tracking environment.

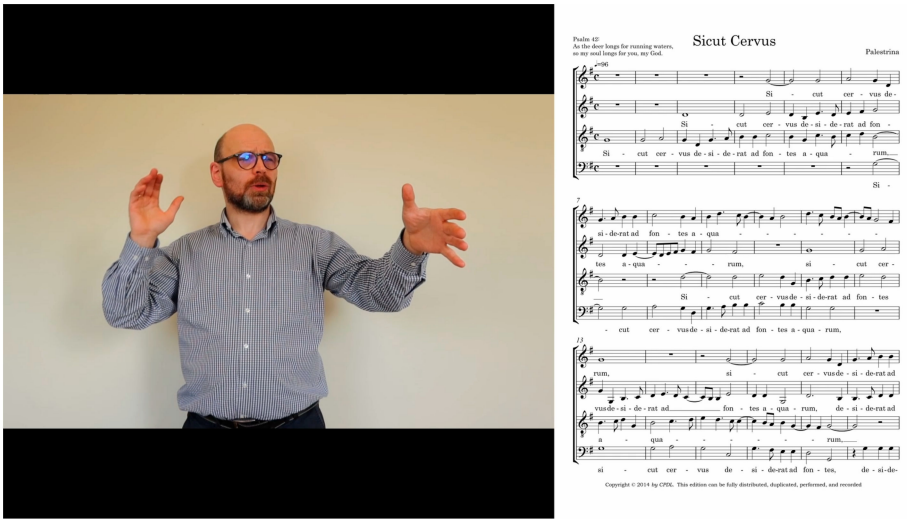


**Figure 8.** Visual input.

The examination started with a calibration of the eye-tracker. Then, the audio-visual recording was displayed on the screen, and the participant was performing their vocal part. The recording for the examination consisted of the first page of G. P. da Palestrina's Sicut Cervus and took 57 s in total. The calibration, the video playback, and collecting the eye-tracker data were automated via scripts written in Python 3.

Figure 9 shows the heat map obtained based on averaging all individual performers' heat maps. The bar on the right indicated the intensity of fixation. It is evident that in the case of the averaged results, the primary focus was on the conductor. The ROI (Region of Interest) is on the conductor's face but less on his hands, which are used to convey expressiveness inscribed within the music notation. This may be related to the properties

of human vision. By analyzing the placement of the test stand in terms of the chorister's field of view (Figure 10), it can be estimated that with a typical distance of the chorister from the screen ($d$) of about 60–70 cm and a space between the conductor's hands ($2w$) of 20 cm, the chorister's vision focus lies within +/− 8–10 degrees. For the calculation of the viewing angle ($\alpha$), we used the following Equation (1):

$$\propto = arc \tan \frac{w}{d}, \quad (1)$$

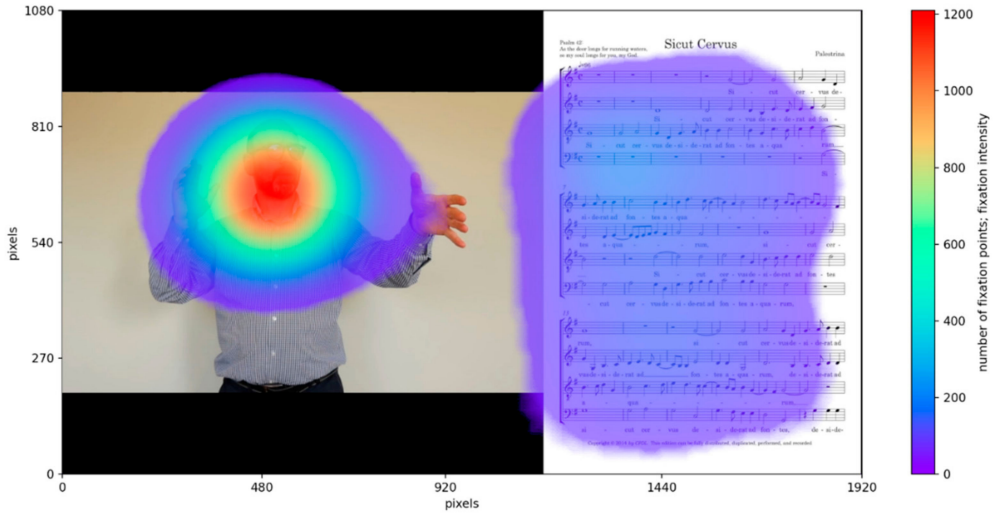


**Figure 9.** Average ROI for 20 choristers.

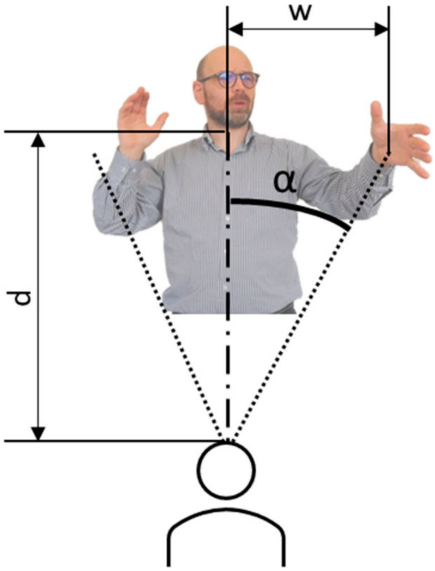


**Figure 10.** Spatial relationships between the chorister and the conductor.

Moreover, these angles are similar to the real situation, where the conductor is about 5 m away from the chorister and the distance between the conductor's hands is about 1.5 m.

Such values of visual field angles suggest that the conductor's hands are on the edge between central and peripheral vision [84,85]. Thus, by focusing their gaze on the conductor's face, choristers can easily see the conductor's hand movements clearly. Focusing the gaze on one of the hands would definitely make it challenging to observe the other hand (especially while both hands are in motion). Furthermore, following the conductor's facial expressions makes singing easier for choristers [86].

In contrast, Figure 11 shows the individual results of a chorister unfamiliar with the sung piece; hence, this person focuses much more on the right side, where the sheet of music notes was presented. It should be, however, noted that this person has a relatively fluent ability to read music notes *a vista*.

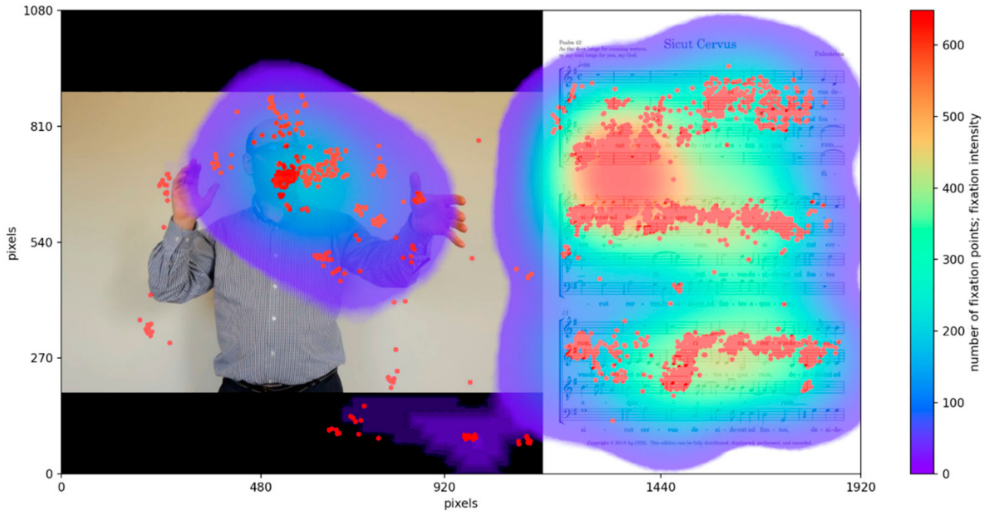


**Figure 11.** Region of interest and all fixation points for a participant with lesser familiarity with the sung piece but a relatively fluent ability to read music notes *a vista*.

## 6. Summary

This paper presents and briefly discusses what has been achieved in the field of remote audio-visual recording. The difficulties related to remote music performance are presented. The examples of spatial transmission and remote production are also discussed. The production process of a remote choral recording performed by the Academic Choir of Gdansk University of Technology is presented in detail. Despite the fact that the existing platforms for remote communication do not provide the full possibility of music recording, which is offered by immersive technologies, this recording presents a high artistic level—it received a number of positive opinions and comments not only from the conductor, choir members, and listeners of the Academic Choir of the Gdansk University of Technology but also from the jury of the following virtual choir festival: Bandung Choral Society World Virtual Choir Festival 2021 in Indonesia, where the recording won a gold award for the following: "[ . . . ] outstanding classical performance". Additionally, this means that providing the listeners with new artistic experiences resulting from immersion in virtual reality may emphasize the virtues of virtual choirs, especially their innovative character and new creative opportunities that may become an added value in 21st-century culture. Undoubtedly, the outlined aspect requires ensuring a careful process of postproduction of the prepared recording.

Initial examination of the eye-tracker data leads to interesting conclusions. Regardless of the virtual setting of the recording, the participants were still generally focused on the conductor, similarly as in regular singing situations. The fixation point remained on the conductor's face, which allowed observation of the conductor's hand movements simultaneously. However, one of the factors unaccounted for was the participants' familiarity with the sung piece and their ability to read music notes fluently. This is especially important as most members of the academic choir are not trained musicians, but students of engineering degrees, and their ability in this regard varies significantly.

The next stage of this artistic project will be performing subjective tests, especially in evaluating various aspects of the ambisonic recording by the audience. The context of familiarity with the piece and fluency in reading notes will also be examined.

**Author Contributions:** Conceptualization, B.M. and B.K.; methodology, B.M., B.K. and P.O.; software, B.M. and P.O.; validation, B.M. and P.O.; formal analysis, B.M.; investigation, B.M., B.K. and P.O.; resources, B.M.; data curation, B.M.; writing—original draft preparation, B.M. and B.K.; writing—review and editing, B.M., B.K. and P.O.; visualization, B.M.; supervision, B.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study, due to the following type of tests: listening to music at the normal level of loudness and using the eye-tracker as recommended by the manufacturer–no possible harm entailed by tests.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank the Academic Choir of Gdansk University of Technology and the conductor, Mariusz Mróz, for their participation in the recordings and the experiment. Thanks are also due to the reviewers for their thoughtful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Caceress, J.-P.; Chafe, C. JackTrip/SoundWIRE meets server farm. *Comput. Music J.* **2010**, *34*, 29–34.
2. Chafe, C.; Wilson, S.; Leistikow, R.; Chisholm, D.; Scavone, G. A simplified approach to high quality music and sound over IP. In Proceedings of the Conference on Digital Audio Effects, Verona, Italy, 7–9 December 2000; pp. 159–164.
3. Chafe, C. Tapping into the Internet as an acoustical/musical medium. *Contemp. Music Rev.* **2009**, *28*, 413–420.
4. Weinberg, G. Interconnected musical networks: Toward a theoretical framework. *Comput. Music J.* **2005**, *29*, 23–39.
5. Bartlette, C.; Headlam, D.; Bocko, M.; Velikic, G. Effect of network latency on interactive musical performance. *Music Percept.* **2006**, *24*, 49–62.
6. Bouillot, N.; Cooperstock, J.R. Challenges and performance of High-Fidelity audio streaming for interactive performances. In Proceedings of the 9th International Conference on New Interfaces for Musical Expression, Pittsburgh, PA, USA, 4–6 June 2009; pp. 135–140.
7. Chafe, C.; Caceres, J.P.; Gurevich, M. Effect of temporal separation on synchronization in rhythmic performance. *Perception* **2010**, *39*, 982–992.
8. Gu, X.; Dick, M.; Kurtisi, Z.; Noyer, U.; Wolf, L. Network-centric music performance: Practice and experiments. *IEEE Commun. Mag.* **2005**, *43*, 86–93.
9. Kapur, A.; Wang, G.; Davidson, P.; Cook, P.R. Interactive network performance: A dream worth dreaming? *Organ. Sound* **2005**, *10*, 209–219.
10. Lazzaro, J.; Wawrzynek, J. A case for network musical performance. In Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video, New York, NY, USA, 25–26 June 2001; pp. 157–166.
11. Bouillot, N. nJam user experiments: Enabling remote musical interaction from milliseconds to seconds. In Proceedings of the 7th International Conference on New Interfaces for Musical Expression, New York, NY, USA, 6–10 June 2007; pp. 142–147.
12. Caceres, J.-P.; Hamilton, R.; Iyer, D.; Chafe, C.; Wang, G. To the edge with China: Explorations in network performance. In Proceedings of the 4th International Conference on Digital Arts, Porto, Portugal, 10–12 September 2008; pp. 61–66.
13. Gurevich, M. JamSpace: A networked real-time collaborative music environment. In Proceedings of the CHI'06 Extended Abstracts on Human Factors in Computing Systems, Montréal, Canada, 22–27 April 2006; pp. 821–826. [CrossRef]
14. 10th Anniversary of the Internet in Poland, Internet Concert (In Polish). Available online: http://www.internet10.pl/koncert.html (accessed on 1 February 2022).
15. Aoki, S.; Cohen, M.; Koizumi, N. Design and control of shared conferencing environments for audio telecommunication using individually measured HRTFs. *Presence* **1994**, *3*, 60–72.
16. Buxton, W. Telepresence: Integrating shared task and person spaces. In Proceedings of the Graphics Interface '92, Vancouver, Canada, 11–15 May 1992; pp. 123–129. [CrossRef]
17. Durlach, N.I.; Shinn-Cunningham, B.G.; Held, R.M. Supernormal auditory localization. *Presence* **1993**, *2*, 89–103.
18. Durlach, N. Auditory localization in teleoperator and virtual environment systems: Ideas, issues, and problems. *Perception* **1991**, *20*, 543–554.
19. Jouppi, N.P.; Pan, M.J. Mutually-immersive audio telepresence. In Proceedings of the 113th Audio Engineering Society Convention, Los Angeles, CA, USA, 5–8 October 2002.
20. Wenzel, E.M.; Wightman, F.L.; Kistler, D.J. Localization with non-individualized virtual acoustic display cues. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 27 April–2 May 1991; pp. 351–359.
21. AltSpaceVR. Available online: https://altvr.com/ (accessed on 1 February 2022).

22. Rottondi, C.; Chafe, C.; Allocchio, C.; Sarti, A. An Overview on Networked Music Performance Technologies. *IEEE Access* **2016**, *4*, 8823–8843. [CrossRef]
23. The University of Texas at Austin and Internet2 to Host First Virtual Concert Experiment. 2004. Available online: https://news.utexas.edu/2004/09/27/the-university-of-texas-at-austin-and-internet2-to-host-first-virtual-concert-experiment-tuesday-sept-28/ (accessed on 1 February 2022).
24. Sawchuk, A.; Chew, E.; Zimmermann, R.; Papadopoulos, C.; Kyriakakis, C. From Remote Media Immersion to Distributed Immersive Performance. In Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence, Berkeley, CA, USA, 7 November 2003. [CrossRef]
25. Xu, A.; Woszczyk, W.; Settel, Z.; Pennycook, B.; Rowe, R.; Galanter, P.; Bary, J.; Martin, G.; Corey, J.; Cooperstock, J.R. Real-time streaming of multichannel audio data over Internet. *J. Audio Eng. Soc.* **2000**, *48*, 627–641.
26. Zimmermann, R.; Chew, E.; Ay, S.A.; Pawar, M. Distributed musical performances: Architecture and stream management. *ACM Trans. Multimedia Comput. Commun. Appl.* **2008**, *4*, 1–23. [CrossRef]
27. Gurevich, M.; Donohoe, D.; Bertet, S. Ambisonic spatialization for networked music performance. In Proceedings of the 17th International Conference on Auditory Display, Budapest, Hungary, 20–23 June 2011.
28. Frank, M.; Sontacchi, A. Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broadcasts. *J. Audio Eng. Soc.* **2017**, *65*, 749–756. [CrossRef]
29. PURE Ambisonics Concert & the Night of Ambisonics. Available online: https://ambisonics.iem.at/icsa2015/pure-ambisonics-concert (accessed on 1 February 2022).
30. Rudrich, D.; Zotter, F.; Frank, M. Efficient Spatial Ambisonic Effects for Live Audio. In Proceedings of the 29th Tonmeistertagung—VDT International Convention, Cologne, Germany, 17–20 November 2016.
31. YouTube Help—Use Spatial Audio in 360-Degree and VR Videos. Available online: https://support.google.com/youtube/answer/6395969 (accessed on 1 February 2022).
32. Facebook 360 Spatial Workstation—Creating Videos with Spatial Audio for Facebook 360. Available online: https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/CreatingVideosSpatialAudioFacebook360.html (accessed on 1 February 2022).
33. Facebook 360 Spatial Workstation—Using an Ambisonic Microphone with Your Live 360 Video on Facebook. Available online: https://facebookincubator.github.io/facebook-360-spatial-workstation/KB/UsingAnAmbisonicMicrophone.html (accessed on 1 February 2022).
34. Deppisch, T.; Meyer-Kahlen, N.; Hofer, B.; Latka, T.; Zernicki, T. HOAST: A Higher-Order Ambisonics, Streaming Platform. In Proceedings of the 148th Audio Engineering Society Convention, Online, 25–28 May 2020.
35. Carôt, A.; Sardis, F.; Dohler, M.; Saunders, S.; Uniyal, N.; Cornock, R. Creation of a Hyper-Realistic Remote Music Session with Professional Musicians and Public Audiences Using 5G Commodity Hardware. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, UK, 6–10 July 2020. [CrossRef]
36. Eric Whitacre's Virtual Choir. Available online: https://ericwhitacre.com/the-virtual-choir/about (accessed on 1 February 2022).
37. A Socially-Distanced, 360 Performance of Puccini's Turandot (Royal Opera House Chorus and Orchestra). Available online: https://youtu.be/VwOpNf8eHeY (accessed on 1 February 2022).
38. Georgia Symphony Chorus, Georgia On My Mind—360° Virtual Choir with Adaptive Audio in 8K. Available online: https://youtu.be/BrXZ63nOUhU (accessed on 1 February 2022).
39. I(solace)ion (Juliana Kay & Exaudi)|360°—Exaudi. Available online: https://youtu.be/HkiIUeuugk8 (accessed on 1 February 2022).
40. J. S. Bach-Koncert na Dwoje Skrzypiec BWV 1043 [360°] (J. S. Bach—Concerto for Two Violins BWV 1043 [360°]). Available online: https://youtu.be/mQXNneuRG3s (accessed on 1 February 2022).
41. Socially Distant Orchestra Plays "Jupiter" in 360°. Available online: https://youtu.be/eiouj6HkjfA (accessed on 1 February 2022).
42. Sicut Cervus-Virtual Cathedral #StayAtHome #SingAtHome [4k 360°]. Available online: https://youtu.be/4dwSRNxUrlU (accessed on 1 February 2022).
43. Hewage, C.; Ekmekcioglu, E. Multimedia Quality of Experience (QoE): Current Status and Future Direction. *Future Internet* **2020**, *12*, 121. [CrossRef]
44. Kunka, B.; Czyżewski, A.; Kostek, B. Concentration tests. An application of gaze tracker to concentration exercises. In Proceedings of the 1st International Conference on Computer Supported Education, Lisboa, Portugal, 23–26 March 2009.
45. Ramírez-Correa, P.; Alfaro-Pérez, J.; Gallardo, M. Identifying Engineering Undergraduates' Learning Style Profiles Using Machine Learning Techniques. *Appl. Sci.* **2021**, *11*, 10505. [CrossRef]
46. Jo, Y.-J.; Choi, J.-S.; Kim, J.; Kim, H.-J.; Moon, S.-Y. Virtual Reality (VR) Simulation and Augmented Reality (AR) Navigation in Orthognathic Surgery: A Case Report. *Appl. Sci.* **2021**, *11*, 5673. [CrossRef]
47. Becerra Martinez, H.; Hines, A.; Farias, M.C.Q. Perceptual Quality of Audio-Visual Content with Common Video and Audio Degradations. *Appl. Sci.* **2021**, *11*, 5813. [CrossRef]
48. Kunka, B.; Kostek, B.; Kulesza, M.; Szczuko, P.; Czyżewski, A. Gaze-Tracking Based Audio-Visual Correlation Analysis Employing Quality of Experience Methodology. *Intell. Decis. Technol.* **2010**, *4*, 217–227. [CrossRef]
49. Kunka, B.; Kostek, B. Exploiting Audio-Visual Correlation by Means of Gaze Tracking. *Int. J. Comput. Sci.* **2010**, *3*, 104–123.

50. Meghanathan, R.N.; Ruediger-Flore, P.; Hekele, F.; Spilski, J.; Ebert, A.; Lachmann, T. Spatial Sound in a 3D Virtual Environment: All Bark and No Bite? *Big Data Cogn. Comput.* **2021**, *5*, 79. [CrossRef]
51. Zhu, H.; Luo, M.D.; Wang, R.; Zheng, A.-H.; He, R. Deep Audio-visual Learning: A Survey. *Int. J. Autom. Comput.* **2021**, *18*, 351–376. [CrossRef]
52. Tran, H.T.T.; Ngoc, N.P.; Pham, C.T.; Jung, Y.J.; Thang, T.C. A Subjective Study on User Perception Aspects in Virtual Reality. *Appl. Sci.* **2019**, *9*, 3384. [CrossRef]
53. Brungart, D.S.; Kruger, S.E.; Kwiatkowski, T.; Heil, T.; Cohen, J. The effect of walking on auditory localization, visual discrimination, and aurally aided visual search. *Hum. Factors* **2019**, *61*, 976–991. [CrossRef]
54. Hekele, F.; Spilski, J.; Bender, S.; Lachmann, T. Remote vocational learning opportunities—A comparative eye-tracking investigation of educational 2D videos versus 360° videos for car mechanics. *Br. J. Educ. Technol.* **2022**, *53*, 248–268. [CrossRef]
55. Kostek, B.; Kunka, B. Application of Gaze Tracking Technology to Quality of Experience Domain. In Proceedings of the MCSS 2010: IEEE International Conference on Multimedia Communications, Services and Security, Kraków, Poland, 6–7 May 2010; pp. 134–139.
56. Kostek, B. Observing uncertainty in music tagging by automatic gaze tracking. In Proceedings of the 42nd International Audio Engineering Society Conference Semantic Audio, Ilmenau, Germany, 22–24 July 2011; pp. 79–85.
57. Poggi, I.; Ranieri, L.; Leone, Y.; Ansani, A. The Power of Gaze in Music. Leonard Bernstein's Conducting Eyes. *Multimodal Technol. Interact.* **2020**, *4*, 20. [CrossRef]
58. Gerzon, M.A. What's wrong with Quadraphonics. *Studio Sound* **1974**, *16*, 50–56.
59. RØDE Blog—The Beginner' 's Guide To Ambisonics. Available online: https://www.rode.com/blog/all/what-is-ambisonics (accessed on 1 February 2022).
60. Mróz, B.; Odya, P.; Kostek, B. Multichannel Techniques in the Application of Remote Concerts and Music Recordings at a Distance (in Polish). In *Research Advances in Audio and Video Engineering. New Trends and Applications of Multichannel Sound Technology and Sound Quality Research*; Opieliński, K., Ed.; Wroclaw University of Technology Publishing House: Wroclaw, Poland, 2021; pp. 67–82. (In Polish). [CrossRef]
61. Zotter, F.; Frank, M. All-round Ambisonic panning and decoding. *J. Audio Eng. Soc.* **2012**, *60*, 807–820.
62. Zotter, F.; Frank, M. Ambisonic decoding with panning-invariant loudness on small layouts (allrad2). In Proceedings of the 144th Audio Engineering Society Convention, Milan, Italy, 24–26 May 2018.
63. Berge, S.; Barrett, N. High angular resolution planewave expansion. In Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris, France, 6–7 May 2010.
64. Murillo, D.; Fazi, F.; Shin, M. Evaluation of Ambisonics decoding methods with experimental measurements. In Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3–4 April 2014. [CrossRef]
65. Pulkki, V.; Merimaa, J. Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests. *J. Audio Eng. Soc.* **2006**, *54*, 3–20.
66. Wiggins, B.; Paterson-Stephens, I.; Schillebeeckx, P. The analysis of multichannel sound reproduction algorithms using HRTF data. In Proceedings of the 19th International AES Surround Sound Convention, Schloss Elmau, Germany, 21–24 June 2001; pp. 111–123.
67. Beack, S.; Sung, J.; Seo, J.; Lee, T. MPEG Surround Extension Technique for MPEG-H 3D Audio. *ETRI J.* **2016**, *38*, 829–837. [CrossRef]
68. Herre, J.; Hilpert, J.; Kuntz, A.; Plogsties, J. MPEG-H 3D Audio—The New Standard for Coding of Im-mersive Spatial Audio. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 770–779. [CrossRef]
69. Meltzer, S.; Neuendorf, M.; Sen, D.; Jax, P. MPEG-H 3D Audio—The Next Generation Audio System. *IET Commun.* **2014**, *8*, 2900–2908. [CrossRef]
70. Meltzer, S.; Murtaza, A.; Pietrzyk, G. MPEG-H 3D Standard. Audio and its applications in digital television (in Polish). In *Research Advances in Audio and Video Engineering. New Trends and Applications of Multimedia Technologies*; Kostek, B., Ed.; Academic Publishing House EXIT: Warsaw, Poland, 2019; pp. 16–44. (In Polish).
71. Zotter, F.; Frank, M. Does it Sound Better Behind Miles Davis' Back?—What Would It Sound Like Face-to-Face? Rushing through a Holographic Sound Image of the Trumpet. Available online: https://acoustics.org/2paaa4-does-it-sound-better-behind-miles-davis-back-what-would-it-sound-like-face-to-face-rushing-through-a-holographic-sound-image-of-the-trumpet-franz-zotter-matthias-frank/ (accessed on 18 March 2022).
72. Hohl, F.; Zotter, F. Similarity of musical instrument radiation-patterns in pitch and partial. In Proceedings of the DAGA 2010, Berlin, Germany, 15–18 March 2010; pp. 701–702.
73. Pätynen, J.; Lokki, T. Directivities of Symphony Orchestra Instruments. *Acta Acust. United Acust.* **2010**, *96*, 138–167. [CrossRef]
74. Waddell, G.; Williamon, A. Technology Use and Attitudes in Music Learning. *Front. ICT* **2019**, *6*, 11. [CrossRef]
75. Ruby, R. How to Record High-Quality Music with a Smartphone. Available online: https://rangeofsounds.com/blog/how-to-record-music-with-a-smartphone/ (accessed on 18 March 2022).
76. Reaper. Available online: https://www.reaper.fm/ (accessed on 1 February 2022).
77. Melodyne Studio. Available online: https://www.celemony.com/en/melodyne/what-is-melodyne (accessed on 1 February 2022).
78. IEM Plug-in Suite. Available online: https://plugins.iem.at/ (accessed on 1 February 2022).

79. IEM Plug-in Suite—DirectivityShaper. Available online: https://plugins.iem.at/docs/directivityshaper/ (accessed on 1 February 2022).
80. IEM Plug-in Suite—RoomEncoder. Available online: https://plugins.iem.at/docs/plugindescriptions/#roomencoder (accessed on 1 February 2022).
81. IEM Plug-in Suite—FDNReverb. Available online: https://plugins.iem.at/docs/plugindescriptions/#fdnreverb (accessed on 1 February 2022).
82. DaVinci Resolve. Available online: https://www.blackmagicdesign.com/products/davinciresolve/ (accessed on 1 February 2022).
83. Facebook 360 Spatial Workstation. Available online: https://facebook360.fb.com/spatial-workstation/ (accessed on 1 February 2022).
84. Strasburger, H.; Rentschler, I.; Jüttner, M. Peripheral vision and pattern recognition: A review. *J. Vis.* **2011**, *11*, 13. [CrossRef]
85. Simpson, M.J. Mini-review: Far peripheral vision. *Vis. Res.* **2017**, *140*, 96–105. [CrossRef] [PubMed]
86. Wöllner, C. Which Part of the Conductor's Body Conveys Most Expressive Information? A Spatial Occlusion Approach. *Music. Sci.* **2008**, *12*, 249–272. [CrossRef]