www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

ARCHIVES OF ACOUSTICS Online First June 11, 2024 doi: 10.24425/aoa.2024.148798

Research Paper

Assessing Spatial Audio: A Listener-Centric Case Study on Object-Based and Ambisonic Audio Processing

Paweł MAŁECKI*, Joanna STEFAŃSKA, Maja SZYDŁOWSKA

Department of Mechanics and Vibroacoustics, AGH University of Krakow
Kraków, Poland

*Corresponding Author e-mail: pawel.malecki@agh.edu.pl

(*received July 18, 2023; accepted March 20, 2024; published online June 11, 2024*)

The research explores the production and critical evaluation of two distinct mixes of "Dancing Ends", a musical composition by Łukasz Pieprzyk. These mixes were engineered using two cutting-edge spatial sound technologies: Dolby Atmos and Ambisonics. The recording process incorporated overdub and multitrack recording techniques. Once created, the mixes were evaluated using a method of direct rating, based on an average rank system from 1 to 5, adhering strictly to the (ITU-R, 2015) BS.1116-3 and (ITU-R, 2019) BS.1284-2 standards. Evaluation criteria included factors such as mix selectivity, depth, width, and height of the sound stage, sound envelopment, tonal brightness, and quality of source localization. Additionally, some criteria were specifically tailored to evaluate characteristics unique to the composition. The evaluations were performed on three different listening systems and environments: surround systems of 5.1 and 7.1.4, and binaural listening. Although Ambisonics' mix received higher ratings in several categories, Dolby Atmos' mix was preferred across all listening environments. The results underscore the potential benefits of employing spatial sound technologies in music production and evaluation, offering insight into the capabilities of Dolby Atmos and Ambisonics.

Keywords: spatial sound technologies; Dolby Atmos; Ambisonics; music production; sound evaluation.

1. Introduction

The origins of spatial sound experiments date back to the 1930s (SPORS *et al.*, 2013). Presently, with the advancement in sound engineering, extraordinary possibilities are being achieved in the field of three-dimensional sound technology. Modern technologies not only ensure precise reproduction in the frequency domain but also allow for faithful representation of the sound space. Nowadays, immersive audio format is available in a binaural standard for the mass user (Apple, 2023). The binaural format enables users to experience three-dimensional audio over standard headphones, enhancing the listening experience in everyday use.

In 2012, a new spatial sound format, Dolby Atmos, was introduced, which uses audio objects (KELLY *et al.*, 2020). This innovative format added a new dimension of height to surround sound systems, offering a more immersive and realistic sound experience. The technology comprises two basic elements: bed tracks and objects. Bed tracks are channel-based buses that can be decoded to various standard configurations such as 2.0, 5.1 or 7.1 but all with fixed locations, strictly defined by the speaker layout. On the other hand, objects refer to sound elements individually mapped on a hemisphere, independent of the reproduction system. These consist of an audio stream that is sent to the Dolby Atmos Renderer and a metadata stream that carries panning information to determine the location in the space (Dolby Laboratories, n.d.). The Dolby Atmos Renderer application is a pivotal component of any Dolby Atmos mixing system. In configuration with a digital audio workstation (DAW), it generates positional metadata that enables spatial representation of an audio mix in a playback environment. The number of input channels that can be configured depends on the renderer itself and the sampling frequency of the session. When operating at 48 kHz, the format supports 128 monophonic input channels, while at a sam-

pling frequency of 96 kHz, it handles 64 input channels. By default, channels 1–10 are configured as a 7.1.2 bed track, and channels 11–128 are objects; 7.1.2 notation refers to a speaker layout with seven main channels around the listener, one subwoofer channel for low-frequency effects, and two overhead or height channels.

Apart from Dolby Atmos, AURO-3D has emerged as another significant immersive audio technology. AURO-3D enhances the sound field by adding an additional "height" layer, creating a three-dimensional experience. The original AURO-3D approach was channel-based, with an emphasis on vertical sound layering to produce a more enveloping audio experience. This format has been recognized for its ability to produce a natural and realistic listener experience. Subsequent iteration AURO-Cx, have introduced a versatile engine that supports not only the original channel-based approach but also object-based and scene-based (Ambisonics) audio, along with scalable channel-based configurations, thereby broadening the potential applications of AURO-3D technology in various listening environments (AURO-3D, 2023).

Wave field synthesis (WFS), on the other hand, is a spatial audio rendering technique that uses many speakers to recreate an acoustic environment. It enables the synthesis of sound waves to form a continuous wave front, creating a sound field that can simulate sounds both inside and outside the listener's space, offering a heightened sense of realism. While this approach to spatial sound provides listeners with an exceptional level of immersion by accurately reproducing the way sound interacts with the environment, it is accompanied by technical and financial challenges. Implementation is complex and costly, and it can be prone to truncation error and spatial aliasing, which are limitations that can affect sound quality and spatial accuracy (Wittek, 2013).

Ambisonics is a sophisticated spatial audio technology that enables the encoding and decoding of sound fields in a full-sphere around the listener. It is grounded in the principles of spherical harmonics, which are used to mathematically represent complex sound fields (Zotter, Frank, 2019). Contrary to channel-based audio systems that transmit signals to designated speakers, Ambisonics represents the sound field in a speaker-independent format, utilizing the physical properties of the sound to create a scene-based audio experience. It encodes sound waves in a way that captures their complete directional information, which can be decoded by a corresponding array of loudspeakers.

The evaluation of spatial sound quality comprises several different components, the attributes of which include source location, perceived source width, and listener envelopment (Power, 2015). According to (Rumsey *et al.*, 2005), spatial attributes account for over one-third of all quality ratings in listening tests and are therefore crucial in determining the quality of a system. In the work of (Francombe *et al.*, 2017), an evaluation of spatial sound reproduction methods was conducted, and it was found that the listening test results are influenced by the test subject's experience with multichannel formats. The most common attributes used by experienced and inexperienced listeners to describe auditory impressions were identified. Experienced listeners used the depth of the sound field, surroundings, and spectral clarity, while inexperienced listeners determined the position of the sound source, its transparency, and the space. Oramus and Neubauer (2020) conducted studies comparing an object-based and channel-based panning models. Tests were conducted with 127 subjects to compare the perceived positions of six audio samples, each of which was reproduced in 5.1, 7.1, and Dolby Atmos. The results did not show an increase in spatial location precision when using object sound, however, listeners demonstrated greater confidence in determining the position of the sound object compared to conventional channel-based playback. Cengarle (2013) compared the Ambisonic technique and 5.1 surround – he stated that first-order Ambisonics is suitable for diffuse sounds. Furthermore, he concluded that the higher-order Ambisonic technique used in a spatial system enhances perceived realism compared to the 5.1 system. Kleczkowski *et al.* (2015) examines the perceptual effect of the separation of the components of direct and reflected sound impulse responses in multichannel systems, using phantom sound sources. The findings reveal a more consistent perceptual advantage of separation, particularly among experienced listeners. Many research papers point to the significant superiority of spatial sound over stereo productions. In a prior study by the authors of the current work (Malecki *et al.*, 2020), the focus was on electronic music. This earlier study involved creating a spatial remix of a stereophonic composition using Ambisonics. This was followed by a subjective comparative analysis between the original stereophonic version and the spatially remixed Ambisonic version. The primary objective was to explore the potential of spatial dimensions and an extended music scene. The subjective evaluation involved a group of experts and predicted playback in stereo and Ambisonic configurations, as well as binaural listening. The subjects evaluated aspects such as spatiality, selectivity, timbre, dynamics, and overall impression. On the basis of the listening tests conducted, a preference for spatiality and selectivity of the Ambisonic production was established. On the basis of a comparison of the stereophonic and binaural render of the Ambisonic mix, a clear preference for spatiality in the binaural version was demonstrated.

The purpose of this study is to produce and conduct auditory evaluations of music mixes created using two distinct spatial sound technologies: Ambisonics and Dolby Atmos. This comparative analysis aims

to understand the nuances in listener perception and audio quality between these advanced sound reproduction methods. By doing so, the study explores the effectiveness and immersive qualities of each technology in the context of classical music production.

In the first stage, music mixes were created using the two technologies. The first mix was prepared using the Dolby Atmos Production Suite. The Pro Tools Ultimate digital workstation was used to create the mix. The Dolby Atmos Renderer application, which communicates with the DAW software, was used to generate positional metadata that allows for precise spatial reproduction of the mix. The second mix was performed using the Ambisonic technique in REAPER software. To compare the final materials, a key task was to reproduce the first production. The IEM Plug-in Suite (2023), which includes a set of open-source Ambisonic plug-ins, was used to create the mix. The next section of the study describes the survey conducted of the participants in terms of auditory impressions. It was investigated how listeners perceive the selected audio formats in terms of spatiality and sound quality in various loudspeaker system configurations and in binaural listening. The following chapter presents a statistical analysis of the results of listening tests in terms of technology preference and listening system.

The article represents a significantly expanded continuation of the work (Malecki *et al.*, 2023) that was presented and discussed at a conference. This current paper encompasses a considerably more detailed description of the conducted experiments, additional results, and an in-depth statistical analysis. The manuscript emphasizes its novel contributions through new results and expanded analysis, clearly delineating its incremental advancements.

2. Production of spatial sound mixes

The composition selected for spatial mixing and subsequent evaluation is "Dancing Ends for Symphony Orchestra and Piano" by Pieprzyk (2023). This score belongs to the genre of film music, originally produced in a stereophonic format. Łukasz Pieprzyk is an alumnus of the Krakow Academy of Music, where he studied composition under the tutelage of Professors Zbigniew Bujarski and Krzysztof Penderecki. The production phase aimed at a natural representation of the musical stage, striving for a realistic placement of a symphony orchestra ensemble. It was assumed that the listener's position as shown in Fig. 1 within the sound space would mirror that of a conductor in a concert hall, characterized by a typical reverberation time of approximately 2 s. No dynamic compression was used during the mixing process. Only slight timbre equalizations and the manual adjustments of volume levels over time within DAW software were done. This technique was employed to maintain the appropriate balance of the orchestra and to accurately reflect the dynamic changes in the music. These adjustments were made track-by-track and were crucial for preserving the natural dynamics and expression of the orchestral performance.

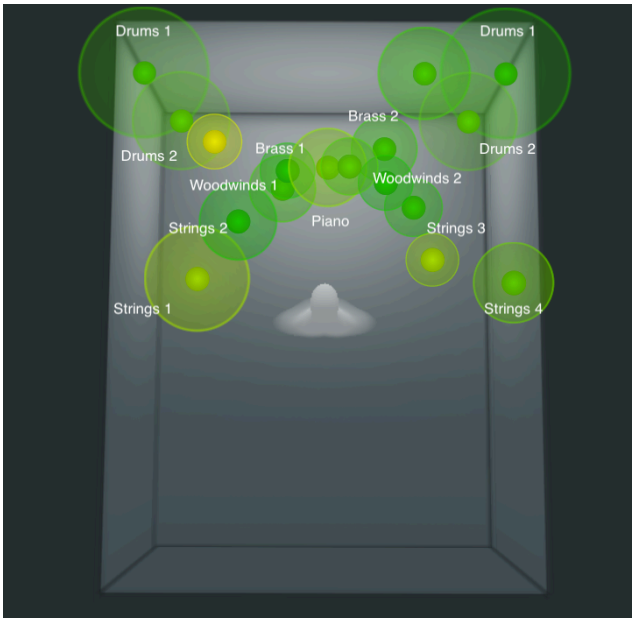

Fig. 1. Main source panning as represented in Dolby Atmos Panner.

The composition was produced in the early period of the SARS-COV-2 pandemic, and the recording was made during its ongoing course; thus the recording was made using the overdub method at the Kotłownia Recording Studio (2023) of the AGH University of Krakow. Each individual section or instrument of the orchestra was recorded separately to a guide track (pilot) provided by the composer. The ensemble included a diverse array of percussion instruments – a casa, daiko, snare drum, toms, tam-tam, kettledrums, along with various types of bells, cymbals, and smaller percussive elements, cumulatively forming two distinct percussion sets, each recorded on separate stem together with individual mono tracks with main percussion elements. Also, the orchestration featured an array of woodwinds and brass, including flutes, oboes, clarinets, bassoons, French horns, trumpets, tubas, and trombones. The string section comprised first violins, solo violin, second violins, violas, cellos, and double basses, complemented by a grand piano. During these sessions, the musicians wore one-ear headphones, through which the pilot track was played. This setup enabled the musicians to hear the guide track in one ear while still maintaining a natural perception of their own instrument. Additionally, a conductor was present in front of the musicians during the recording. The conductor also wore headphones to follow the pilot track and led the musicians through their performance, en-

suring coherence and musicality akin to a traditional orchestral recording.

In the recording of each instrumental section performing in unison (tutti), the Blumlein pair microphone technique was utilized, involving a pair of Neumann U87 microphones. This method was selected to ensure the capture of the rich acoustic detail and spatial characteristics of the ensemble's performance. The placement of the microphones relative to the instruments was determined based on the Recording Angle of the stereo pair, which provided a balance between direct sound and ambient reflections and avoided the proximity effect typically associated with "close miking". For soloists or smaller sections, such as tubas and trombones, individual microphones were used as necessary, with options including the AKG C414 or Schoeps MK4. Drum instruments were similarly captured individually and were further enhanced by an overhead stereo microphone configuration. The entirety of the recording process took place in the live room of Kotlownia Studio, which boasts roughly 80 m^2 of the floor space and an average ceiling height of 8 m. The studio's reverberation time is around 1 s for the acoustic bandwidth, thereby providing a controlled yet resonant acoustic setting ideal for high-fidelity recordings.

The preparation of stems is a very important element in the process of creating a spatial mix. Stems are mono- or stereo-audio files that create subgroups of similar sound sources and represent specific elements of the mix. Typically, a stem represents a group of instruments, such as strings, percussion, or woodwinds. However, this is not a strict rule, and instruments are grouped depending on the genre of material being produced. By default, stems take into account the signal processing chain applied to their components. When played together, they create a full musical mix. Stems have found widespread use in the film industry and are typically divided into dialogues, music, and sound effects. For the execution of the surround mixes, encompassing both Dolby Atmos and Ambisonic formats, a total of 27 stereo and mono stems were rendered from the original Pro Tools stereo mix session. In alignment with our production concept, the rendering of these stems was deliberately executed without incorporating any previously applied equalization, dynamic processing, or reverberation effects. This approach ensured the preservation of the raw, unaltered essence of each instrument group, allowing for greater precision and creativity in the subsequent spatial mixing phase.

2.1. Dolby Atmos mix

The initial plan was to create a first mix using Dolby Atmos technology. Once this was established, a parallel approach was utilized to process the material using Ambisonics, based on the preliminary decisions derived from the Dolby Atmos execution. The first part of the composition contained only percussion instruments. Due to the relatively small number of sources, sound objects were widely panned as shown in Fig. 2. In the second part, the entire symphony orchestra participated. The sound space was divided into plans where individual sections of the instruments were placed. Sections of string instruments, woodwind and brass were distinguished. According to the standard orchestra layout, string instruments were placed at the front, followed by woodwinds and then brass. Figure 1 shows the panning position of the main instrument sections. The figure does not show reverb panning or the drums during the introduction, nor the sound effects of the composition's outro. To achieve greater selectivity and separation between objects, the height of the sources was added; the wind instruments were positioned slightly above the string section. Assigning varied sizes to the instruments aided in unifying their sound while also imparting a distinctive character to each.

Fig. 2. Percussion instruments panning as represented in Dolby Atmos Panner.

Once the mixing process was completed, files were generated to facilitate playback on the intended systems. The renders were made based on the main Master File. It was created in real time during the recording process from Pro Tools to Dolby Renderer. The Master File contained three files with extensions: .atmos, .atmos.metadata, and .audio. The first of them, the top-level file, provided basic information about the project. The second contained all the 3D position coordinates for the object sound in the .audio file, while the last contained audio data for all bed track signals and objects. The rendering was performed using Dolby Atmos presets for the 5.1, 7.1.4, and binaural formats. The binaural audio was rendered statically, without any additional diffuse-field or free-field equalization applied. Furthermore, no head tracking

was employed in the rendering process, implying that the binaural audio output remained consistent regardless of the listener's head movements.

The preliminary version of the spatial mix in Dolby Atmos technology was made in the AGH Music Studio Kotłownia (Fig. 3). The listening room system is based on a 5.1 configuration, standardized by the (ITU-R, 2022) BS.775 standard. The system includes the following speaker models: Genelec 1034 BM (*L*, *R* channels), Genelec 1034 BC (*C* channel), Genelec 1038B (*Ls* and *Rs* channels) and Genelec 7360 (LFE channel).

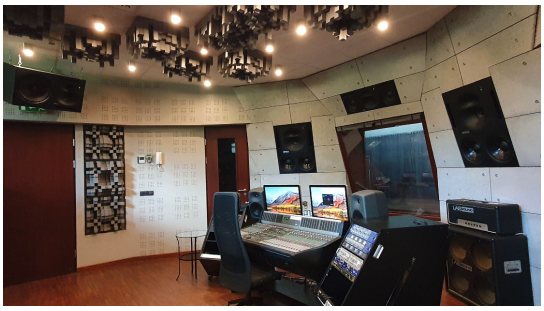

Fig. 3. Control room of the Kotłownia recording studio of the AGH University of Krakow (Kotłownia Recording Studio, 2023).

The validation and additional spatial enhancement of the mix were carried out in the ATMOS Sound Truck (Fig. 4), equipped with an SSL T80 unit and a Dolby Atmos 7.1.4 monitoring system, based on Genelec speakers. The setup included Genelec 8351 (*L*, *R* channels), Genelec 8331 (*C* channel), Genelec 8320 (*Ls*, *Rs*, *Lrs*, *Rrs*, *Ltf*, *Rtf*, *Ltr*, *Rtr* channels) and Genelec 7360 (LFE channel) (Fig. 5). The process involved a thorough check of several key aspects to ensure the mix's quality and spatial accuracy. Selectivity and localization precision, ensuring that each sound element was clearly distinguishable and accurately positioned within the sound field. The balance between direct sound and reverberation. This involved fine-tuning the mix to achieve the right blend of clarity and spatial depth, ensuring that the reverberation did not overpower the direct sound but rather complemented it to enhance the overall spatial impression. The balance of levels across the mix was meticulously adjusted that all elements were at appropriate levels relative to each other, maintaining a harmonious and cohesive soundstage.

Fig. 4. 120 dB ATMOS Sound Truck (120db Sound Engineering, n.d.).

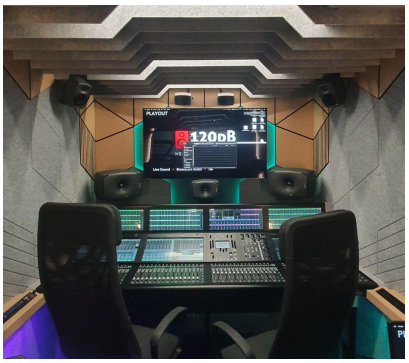

Fig. 5. Inside of 120 dB ATMOS Sound Truck (120db Sound Engineering, n.d.).

After finishing the first mix and preparing it for different listening setups, work on an Ambisonics version of the audio has been started. This phase involved carefully directing the audio signals and using special tools for encoding and decoding.

2.2. Ambisonic mix

The Ambisonic mix was carried out in the Auralization Laboratory of the Department of Mechanics and Vibroacoustics at the AGH University of Science and Technology. The room is equipped with a 16-channel system arranged in a spherical layout (Fig. 6). The system consists of sixteen Genelec 6010 speakers set in a radius of 1.5 m from the center of the sphere.

Fig. 6. Auralization Laboratory at the AGH University of Krakow.

The loudspeakers are arranged in three layers relative to the listener's ear level. In the horizontal plane, eight loudspeakers (channels 1 to 8) are positioned at ear level, at azimuthal angles of 45° increments. Above the listener, four loudspeakers are placed at an elevation angle of 45°, also spaced at 90° intervals azimuthally. Similarly, below the listener, four loudspeakers are situated at an elevation angle of −45°.

The configuration of the IEM AllRAD (IEM Plug-in Suite, 2023) decoding plugin mirrored the physical speaker setup with basic decoding of third-order Ambisonics.

The Ambisonic mix was fundamentally intended to mirror the mix in Dolby Atmos technology. This en-

tails that specialized Ambisonic tools were employed to craft the optimal mix, ensuring adherence to the assumptions and the results obtained from the Dolby Atmos mix. Therefore, from the Dolby Atmos session, 27 stems were generated, incorporating signal processing elements such as automatic volume, correction, and compression. This procedure was carried out to ensure the coherence of sound and dynamics of the signals. Based on the created stems, a spatial Ambisonic mix was created. In order to achieve better source separation and spatiality, some instruments were positioned slightly below the listener's head level, which could not be accomplished using Dolby Atmos technology.

The positioning of signals in the space and their simultaneous encoding in the Ambisonic domain was achieved using the StereoEncoder plugin (IEM Plug-in Suite, 2023). Instruments were intended to be placed as similar as possible to the positioning in the Dolby Atmos mix. To adapt the created Ambisonic mix to the 5.1 and 7.1.4 playback systems, the AllRADecoder plugin was used to decode the third-order Ambisonic signal to selected arrangements. To appropriately design the decoder, JSON configuration files were created for the 5.1 and 7.1.4 setups, containing information on the coordinates of all speakers and their corresponding channel numbers. For the binaural version of the Ambisonics mix, an HRTF set from the Neumann KU 100 dummy head was used by Binaural Decoder (IEM Plug-in Suite, 2023). No additional headphone correction was applied.

3. Listening evaluation

3.1. Tests in a 5.1 surround sound system

The subsequent phase of the research involved the preparation of a protocol for subjective tests. Two listening rooms were chosen for the study: the AGH Kotłownia Music Studio and the AGH Auralization Laboratory. The former room was used to perform tests on the surround 5.1 system. The placement of individual channels was designed according to the (ITU-R, 2022) BS.775 standard. The speaker setup consisted of the following models: Genelec 1034 BM (*L*, *R* channels), Genelec 1034 BC (*C* channel), Genelec 1038B (*Ls* and *Rs* channels) and Genelec 7360 (LFE channel) as shown in Fig. 7. The distance from each speaker to the sweet spot equals 2.8 m. The system is based on the AVID HDX PCIe Card sound card. The room is characterized by complete acoustic adaptation, ensuring appropriate conditions in the listening space by RFZ (reflection free zone) solution (Fig. 3). The room is symmetrical with respect to the vertical plane and the floor surface has a trapezoidal shape. The room's area meets the requirements specified for a multichannel system, measuring over 35 m^2 according to (ITU-R, 2015) BS.1116 standard.

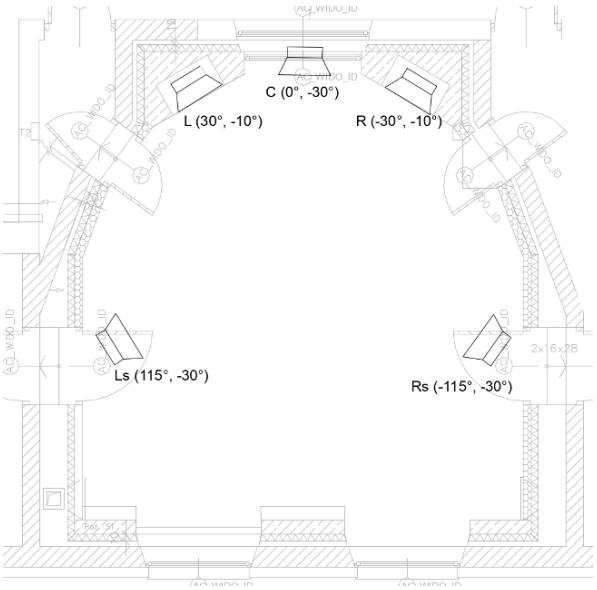

Fig. 7. Speaker placement in the control room of the Kotłownia Recording Studio (2023).

The mean measured reverberation time in third-octave bands from 200 Hz to 4 kHz falls within the range of 0.2 s to 0.4 s according to (EBU, 2004) Tech 3276 S1.

3.2. Tests in 7.1.4 surround sound system

The 7.1.4 configuration for listening tests was implemented in the AGH Auralization Laboratory. The system was equipped with 11 Genelec 6010 speakers, arranged in a radius of 1.5 m from the sweet spot as shown in Fig. 8, and a PSI Sub A225-M subwoofer, all according to Dolby Atmos recommendations. The room has basic acoustic adaptation. The average RT20 (reverberation time) is 0.15 s, calculated for 500 Hz and 1000 Hz. The dimensions of the laboratory are 3.9 m × 6.7 m × 2.8 m. The room meets most of the (ITU-R, 2022) BS.775 standard or has parameters very close to recommended. It meets the criteria for floor area and the ratio of dimensions. The RT is much

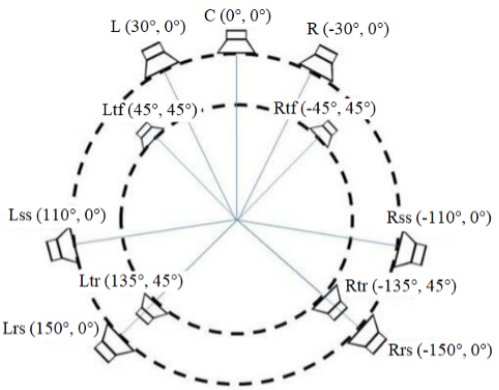

Fig. 8. Speaker placement in the control room of the Auralization Laboratory.

shorter than the recommended 0.56 s and the noise level is 19 dB(A) and meets the most stringent rating, NC15.

3.3. Binaural tests

For binaural listening, a set-up that incorporates a computer equipped with DAW software, Focusrite Scarlett 8i6 audio interface, and the Beyerdynamic DT770 Pro 250 Ohm headphones were used.

3.4. Subjective tests protocol

To evaluate the audio samples, both absolute (direct) and relative (comparative) evaluation methods were implemented, guided by the ITU-R (2019) BS.1284 and ITU-R (2015) BS.1116 standards. This included a global assessment of the overall quality or differences in the given objects and a parametric examination of individual sound attributes such as clarity, spatiality, and timbre. The test plan incorporated three distinct evaluation methods: detection, ordinal, and assignment procedures, each chosen according to the nature of the question. The detection method involved questions related to compatibility evaluation (determining if samples are identical or different) or situations that required a choice (identifying the differing sample). Ordinal evaluation was used for questions regarding ranking (intensity of a certain feature), preferences (better/worse), and similarity (most similar/different). The assignment method allowed for numerical estimation across different types of scales. Listening tests were organized as surveys, in which participants analyzed the material and made choices between the music excerpts presented. The scales for rating were discrete, graphical, and accompanied by labels. Using Google Forms, a survey consisted of eight questions:

1) Rate the following sound properties:
   a) selectivity
   b) depth of soundstage,
   c) width of soundstage,
   d) height of soundstage,
   e) sound immersion,
   e) clarity of sound,
   f) localization quality.
2) Evaluate in which mix you can better locate the flute?
3) Assess whether the piano's position aligns precisely in both mixes?
4) Evaluate in which mix you rate the balance between the string section and the brass section better?
5) Assess if any of the mixes more realistically represented the placement of musicians in the space?
6) Which mix is more balanced in terms of frequency?
7) Which mix do you prefer?
8) What aspects differentiate these mixes the most? (choose two):
   a) selectivity,
   b) timbre,
   c) source location,
   d) listener's perspective,
   e) width of the soundstage,
   f) sound envelopment.

The sound samples for each question were carefully selected to ensure signal diversity and reduce listener fatigue. Each music excerpt was designed to be directly related to a specific question. To optimize the accuracy of the results, each listener was tested individually with an interactive signal presentation that allowed for unlimited repetitions of each excerpt. All the sound samples were logically arranged to avoid abrupt endings and presented in random order. The first question followed a single signal presentation principle (parameter evaluation), whereas the subsequent questions used a paired comparison (preference or difference evaluation). In the first and the last questions, signals were played one after another, whereas in the other questions, switching between signals was enabled.

Following the guidelines of ITU-R (2019) BS.1284 for conducting subjective tests, the participant group comprised a so-called "expert group" of at least ten individuals. For this experiment, twelve people were involved for the 5.1 system and binaural listening, and ten people for the 7.1.4 configuration. Each participant had a higher education degree in acoustics, had basic skills in sound production, and previous experience in listening tests. Some participants also did first- or second-degree music education. All were otologically normal, which means that they were free of diagnosed diseases or pathologies of the auditory system.

To minimize the potential influence of the participants' emotions and attitudes on their judgment, questions were precisely articulated, signals were equalized to the same level (SPL A-weighted equal to 80 dB), and all listeners received training prior to the listening tests. Initial preparation included familiarization with the survey structure, rules for presenting music excerpts, and methods of answering individual questions. Further clarification of the evaluation parameters of the first question was provided in a document at the beginning of the study, minimizing any misunderstanding of the applied concepts.

4. Results

Listening tests were conducted with the intent to compare the Ambisonic technique and Dolby Atmos in a selected speaker configurations. On the basis of the characteristics of the constructed questions, a distinction was made between qualitative and quantitative variables. Depending on the features under examination, the data were classified on an interval or a nominal scale.

4.1. Question 1

The construction of the first question indicated quantitative variables assigned to the interval scale. Following this assumption, it was necessary to investigate whether the results demonstrated characteristics of a normal distribution (the Shapiro–Wilk test), determine if variables were correlated, and verify if the variances of variables across populations were equal (Levene's test). Depending on the final assignment of data, an independent t-Student test or a Mann–Whitney U test was performed (Table 1). When verifying the statistical hypotheses, a significance level α of 0.05 was adopted in all tests. Assessments were made based on the responses collected for eight distinct questions.

The averaged parameter ratings (question 1) did not show significant differences between technologies. The Shapiro–Wilk test, carried out to assess the *selectivity* parameter, reached statistical significance of a normal distribution only for Dolby Atmos technology in the 7.1.4 system. In other tests, the null hypothesis was rejected. To evaluate statistical significance, Mann–Whitney tests were performed. It was found that the evaluation of *selectivity* of the mixes presented in the 5.1, 7.1.4 systems and in binaural listening, did not show significant differences depending on technology.

For both technologies, Levene's tests proved that the variances of the *depth of soundstage* parameter are homogeneous so the null hypothesis was accepted. Also, based on the t-Student test, no significant differences depending on the technology used was found.

The distribution of the *width of soundstage* parameter was found to be normal only in the 5.1 system. The t-Student test also showed no significant differences in terms of the *width of soundstage* but within binaural and the 7.1.4 configuration, the performed tests showed significant differences in terms of the evaluated property within the two technologies. In both cases, Ambisonic signals was rated higher than Dolby Atmos as shown in Fig. 9. The evaluation of the *Height of the soundstage* and *sound immersion* also did not show significant differences between the compared technologies. The distribution of the results for the *clarity of sound* was not normal for any listening technology. In the 7.1.4 configuration, the results showed significant differences depending on whether the Dolby Atmos or the Ambisonics was presented to the listeners, in favor of the Ambisonic system (Fig. 10). No statistically significant differences were obtained in the remaining systems.

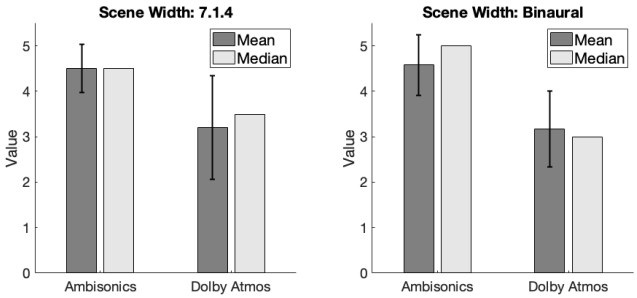

Fig. 9. *Width of soundstage* parameter rating for 7.1.4 system and binaural listening. Error bars represent one standard deviation from the mean.

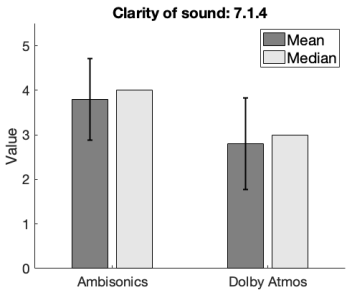

Fig. 10. *Clarity of sound* parameter rating for 7.1.4 system. Error bars represent one standard deviation from the mean.

Table 1. Significance test results for all listening configurations for question 1.

Parameter	5.1		7.1.4		Binaural	
	Test	p-value	Test	p-value	Test	p-value
Selectivity	M–W	0.667	M–W	0.155	M–W	0.116
Depth of soundstage	t-test	1.000	t-Test	0.492	t-test	0.292
Width of soundstage	t-test	0.239	M–W	**0.010***	M–W	**0.001***
Height of soundstage	M–W	0.976	M–W	0.345	M–W	0.707
Sound immersion	M–W	0.951	M–W	0.097	M–W	0.066
Clarity of sound	M–W	0.206	M–W	**0.030***	M–W	0.763
Localization quality	M–W	0.140	M–W	**0.018***	M–W	0.233

M–W – Mann–Whitney test, t-test – t-Student test.

*Significant values ($p < 0.05$).

The Shapiro–Wilk test, conducted to assess the *localization quality* parameter, reached statistical significance of the normal distribution assumption for Dolby Atmos technology in the 5.1 and 7.1.4 systems. However, for none of the systems in Ambisonics technology, a normal distribution was obtained for the examined feature. The Mann–Whitney significance test, performed for the 5.1 system and binaural listening, showed that the characteristic analyzed does not show significant differences depending on the technology. The analysis for the 7.1.4 configuration showed significant differences in the assessment of the *localization quality* and Ambisonic coding was rated better than Dolby Atmos (Fig. 11).

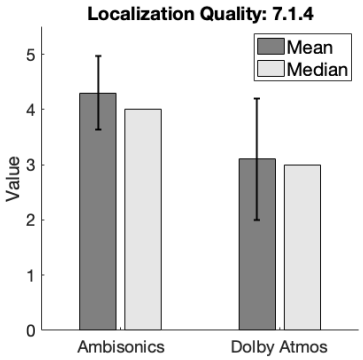

Fig. 11. *Localization quality* parameter rating for 7.1.4 system. Error bars represent one standard deviation from the mean.

In addition, a statistical comparison of the two technologies was conducted in terms of the significance of the system used on the results for all the evaluations received for the 7 perceptual parameters in the question 1. The evaluation results of a given parameter were significantly dependent on the listening system when the p-value was less than the significance level α set to 0.05. Ambisonic mix evaluations for the 5.1 system compared to the 7.1.4 system showed significant differences only for the *width of soundstage* parameter (p = 0.011, the Mann–Whitney test). For evaluations of Dolby Atmos mix, statistical significance tests did not show differences between listening in 5.1 and 7.1.4.

When compared against binaural listening with respect to the 7.4.1 system, the evaluations of the Ambisonic system differed statistically for the *height of the soundstage* parameter (p = 0.034, the Mann-Whitney test). When comparing binaural listening to the 5.1 system, statistically significant differences were only obtained for the *width of soundstage* parameter (p = 0.005, the Mann–Whitney test).

For the mix in Dolby Atmos technology, the assessment of the *sound clarity* parameter depended on whether the hearing tests were conducted in binaural listening or in the 7.1.4 system (p = 0.042, the Mann–Whitney test). When comparing headphone listening with the 5.1 system for Dolby Atmos mix, no statistically significant differences were demonstrated for any parameter.

4.2. Questions 2-7

The results for questions 2–7 are shown in Fig. 12. In the question 2, study participants were asked to select the technology in which they could more accurately locate the flute. In each listening system, a larger percentage of respondents indicated that Dolby Atmos technology allows for more precise localization. During tests conducted in the 5.1 system, 83 % of respondents chose the mix made in Dolby Atmos, in the 7.1.4 configuration the same answer was indicated by 70 % of people, whereas during binaural listening – 58 %.

The next question concerned the placement of the piano. The subjects determined whether the location of the instrument matched in both presented pieces of music. According to the majority of listeners, it did not – respectively, 75 % and 80 % of respondents for the 5.1 and 7.1.4 systems. In the binaural listening, those who responded that the location of the instrument had changed were in the minority – 33 %. The listeners' answers indicated that there is a discrepancy between the mixes in the location of the piano, even though the instrument was positioned directly in front of the listener (Fig. 12).

In the question number 4, listeners chose the technology in which a better balance between the string section and the brass section was obtained. In the 5.1 configuration, 58 % of respondents chose Dolby Atmos and 42 % chose Ambisonics as the technology providing a better balance between the selected sections. The same percentage results were obtained for binaural listening, where the majority of respondents chose Dolby Atmos technology. The opposite situation occurred for the 7.1.4 system, where the balance of the Ambisonic mix was better assessed – 80 % of listeners pointed out this technology. It was noticed that the preference for a given technology was associated with the listening room in which this technology was implemented (Fig. 12).

In the following question, the respondents were asked whether any mix reflected the arrangement of musicians in the space, in a more realistic way. The answers obtained in the question 5 were very diverse. According to the respondents, in the 5.1 system, the Dolby Atmos mix reflected the arrangement of musicians in a more realistic way, while in the 7.1.4 configuration, the Ambisonic mix did. Ambisonics was better assessed in the room where the mix in this technology was implemented, similarly for Dolby Atmos. In binaural listening, no technology was distinguished that would enable arranging musicians in a more realistic way (Fig. 12).

In evaluating the presented musical materials for frequency balance, the majority of respondents in both

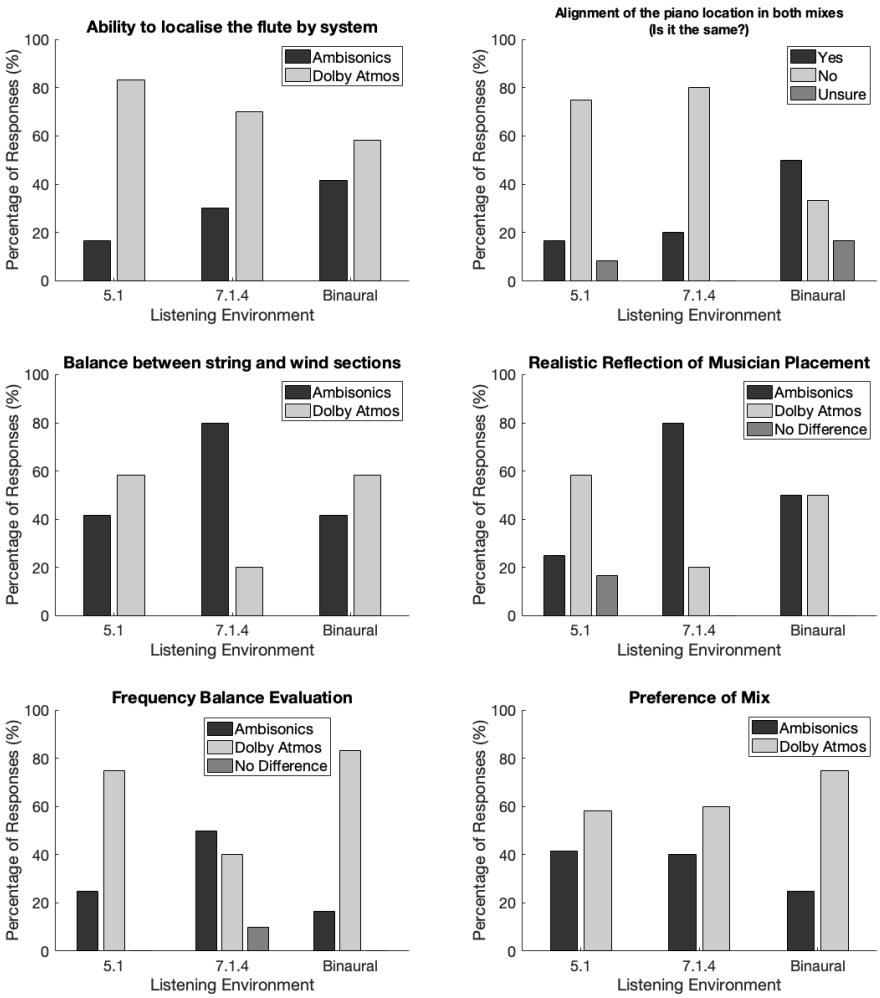

Fig. 12. Results from questions 2–7: Q2. Which mix provides better flute localization? Q3. Does the piano's position align in both mixes? Q4. Which mix offers better balance between the string and brass sections? Q5. Which mix more realistically represents the musicians' spatial placement? Q6. Which mix is more balanced in terms of frequency? Q7. Which mix do you prefer?

5.1 surround and binaural listening selected Dolby Atmos technology (Fig. 12). The significant advantage of one technology over another in headphone listening might have been influenced by the use of different binaural rendering algorithms. In the 7.1.4 setup, 10 % of participants indicated that they heard no difference between the mixes, 40 % chose the Dolby Atmos mix, and 50 % chose the Ambisonic mix.

The last question of this set of the results (the question 7) aimed to collect information about listener preferences. In all listening systems, most respondents decided that they prefer the Dolby Atmos mix. The largest advantage of the Dolby Atmos mix over the Ambisonic mix was obtained during binaural listening 75 % (Fig. 12).

4.3. Question 8

The question 8 required the identification of two aspects that most differentiated the mixes made in two different technologies. When identifying the most differentiating aspects between the samples, listeners most often chose *sound envelopment* for the 5.1 system and binaural listening, and *width of the soundstage* for 7.1.4 arrangement (Fig. 13). According to the respondents, significant differences between the mixes in the

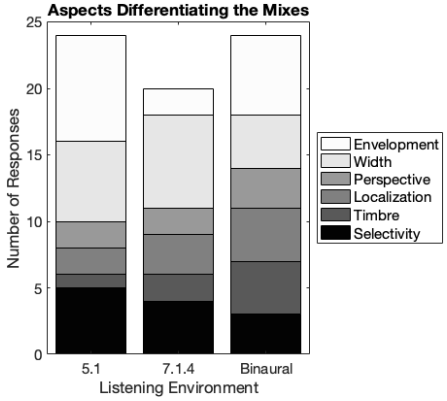

Fig. 13. Results from question 8: Which two aspects differentiate these mixes the most?

5.1 setup were also shown in the *selectivity* and *Width of the soundstage* parameters. In contrast to the 5.1 channel listening and binaural listening, in the 7.1.4 configuration, a small number of respondents voted for the *sound envelopment* parameter. In headphone listening, the respondents' answers were diverse, and there was no second dominating feature differentiating the presented musical materials.

Based on the conducted χ^2 independence test, it was found that the variables are independent, and the listening system does not affect the preference for a given technology ($p > \alpha$). In statistical terms, the evaluations were not dependent on either the technology or the speaker configuration. For questions 2 through 8, independent χ^2 tests (variables unlinked in the nominal scale) were performed and shown in Table 2. The adopted null hypothesis for $p > \alpha$ stated that the listening system does not affect the preference for a given technology and there is no significant relationship between the variables. Based on the calculations made, a decision was taken to reject the null hypothesis.

Table 2. Results of the χ^2 independence test for questions 2 to 8.

Question	χ^2 value	p-value	D.F.
2	1.809	0.405	2
3	6.865	0.143	4
4	4.163	0.125	2
5	8.929	0.063	4
6	6.246	0.182	4
7	0.867	0.648	2
8	7.311	0.696	10

5. Discussion

The implementation of this study involved making a sequence of critical decisions that unquestionably influenced the outcomes obtained. The process of creating a musical mix in each instance was deeply subjective and rooted in individual auditory perceptions. Numerous stages of material representation work were contingent on the personal judgment of the sound engineer, thus introducing an element of subjectivity.

This is noteworthy that the ratings gathered in the subjective tests were influenced by a variety of uncontrolled variables. These variables could be anything from the time of day when the tests were conducted, the listener's mood, or their prior experience with spatial audio. Such variables, though not directly controlled or manipulated in the study, could still exert significant effects on the results.

Additionally, another crucial decision, that was in essence arbitrary but had a potential bearing on the results, was the initial choice of starting the mix with Dolby Atmos technology instead of Ambisonics. Following this, there was an attempt to replicate the effect achieved with Dolby Atmos using Ambisonics. This approach, although logical in its structure, may have inadvertently introduced a bias towards the Dolby Atmos technology.

Furthermore, the basic mix was initially crafted in the 5.1 system and was subsequently examined and upgraded in a system specifically dedicated to Atmos. The selection of the 5.1 system as the starting point, followed by enhancement in the Atmos-specific system, was yet another decision that could have a substantial impact on the final outcome of the study. This sequence of decisions reinforces the fact that the results of the study, although comprehensive, are influenced by subjective choices and uncontrolled variables.

In the study, unconventional loudspeaker configurations, such as 5.1 or 7.1.4, were employed for the reproduction of Ambisonic recordings since it was intended to compare with Dolby Atmos system that is limited to standarized layouts. These configurations can impact the accurate rendering of the Ambisonic field due to their irregular spacing and positioning, which may not align with the standard Ambisonic decoding formats that are designed for uniform speaker layouts. Such irregular setups could potentially introduce spatial anomalies, especially when reproducing higher-order Ambisonics that rely on precise speaker placement to convey detailed spatial information. Converting third-order Ambisonic recordings to a 5-loudspeaker array, might result in spatial aliasing or spatial distortions. This is because the downmixing process does not preserve the higher resolution of spatial cues encoded in the third-order Ambisonic format, leading to a less accurate sound field reproduction.

6. 6. Summary

The focal point of this research was a meticulous comparative evaluation of spatial audio mixing executed in two contemporary technologies – Ambisonic and Dolby Atmos. The study involved conducting an auditory examination of musical materials processed by these technologies. The breadth of the work was extensive and encompassed a detailed narrative of mix realizations, implementation of listening tests, in-depth statistical analysis, and a comprehensive interpretation of the data collected from the research.

The study was fundamentally rooted in the subjective analysis of the participants. The methodology involved executing surveys about the psychoacoustic impressions of the respondents across two different speaker configurations and binaural listening. These tests and the resultant feedback painted an interesting picture about the relative efficacy of these technologies.

Despite the Ambisonic mix scoring higher on many critical criteria (width of soundstage, clarity of sound,

localization quality, the Mann–Whitney test ($p < 0.05$)) for evaluating the quality of the musical material, the subjective tests pointed towards a general preference for the mix produced by Dolby Atmos technology across all the listening systems (preferences: 59 % for 5.1 and 7.1.4; 75 % for binaural). This preference resonated irrespective of the speaker setup and was observed even in binaural listening.

Delving into the statistical aspect of the research, interesting results were observed. In the case of the 5.1 system, no many significant differences were noted between examined variables. However, when it came to the 7.1.4 setup, the data showed substantial disparities in the evaluation of the scene width, sound clarity, and localization quality (the Mann–Whitney test, ($p = 0.01, p = 0.03, p = 0.018$)). Moreover, in binaural listening, the Scene width was marked with significant differences between loudspeaker systems ($p = 0.001$).

Another very important conclusion is that the ambiguity of the results obtained suggests that the difference between the systems is not significant. This opens up possibilities for the production of high-quality spatial signals using open technology and free tools. Also, it can be stated that the purpose of the study was achieved. Although not all conclusions were statistically confirmed, the study successfully identified general trends in the auditory evaluation of the two technologies. Furthermore, this investigation underscores the high quality of both Ambisonics and Dolby Atmos technologies, highlighting their respective strengths and capabilities in spatial audio reproduction. The insights gained from this comparative analysis provide valuable contributions to the field of spatial sound and its application in music production.

Currently, spatial sound is on the rise, experiencing significant technological evolution and witnessing increased utilization in various fields. The subject matter addressed in this study warrants a broader examination considering different music genres and a range of listening systems. The findings of this research can serve as a robust foundation for further, more diversified analyses in the field of spatial audio technology.

Acknowledgments

The research presented in this paper was conducted as part of the statutory activities of the Faculty of Mechanical Engineering and Robotics at AGH University of Science and Technology, under the project number 16.16.130.942. We would like to extend our sincere gratitude to the anonymous reviewers for their meticulous and insightful feedback, which greatly contributed to the enhancement of this manuscript. Their expertise and constructive criticism have been invaluable in refining our work and guiding its development to meet the high standards of academic rigor.

References

1. 120db Sound Engineering (n.d.), 120dB ATMOS Sound Truck, https://www.120db.pl/atmos (access: 12.06.2023).
2. Apple (2023), About Spacial Audio with Dolby Atmos in Apple Music, https://support.apple.com/en-us/HT212182 (access: 12.06.2023).
3. AURO-3D. (2023), AURO®-CX™. Advanced next generation audio codec, NEWAURO BV, https://www.auro-3d.com/wp-content/uploads/2023/08/Auro-Cx-White-Paper-rev1-20230714.pdf (access: 30.11.2023).
4. Cengarle G. (2013), *3D audio technologies: Applications to sound capture, post-production and listener perception*, Unpublished Ph.D. Thesis, Universitat Pompeu Fabra.
5. Dolby Laboratories (n.d.), The Dolby Atmos essentials course, https://learning.dolby.com/course/info.php?id=191 (access: 12.06.2023).
6. European Broadcasting Union (2004), *Listening conditions for the assessment of sound programme material*, EBU Tech 3276-E, Supplement 1.
7. Francombe J., Brookes T., Mason R., Woodcock J. (2017), Evaluation of Spatial Audio reproduction methods (Part 2): Analysis of listener preference, *Journal of the Audio Engineering Society*, **65**(3): 212–225, doi: 10.17743/jaes.2016.0071.
8. IEM Plug-in Suite (2023), https://plugins.iem.at/ (access: 12.06.2023).
9. International Telecommunication Union (2015), *Methods for the subjective assessment of small impairments in audio systems*, Recommendation ITU-R BS.1116-3.
10. International Telecommunication Union (2019), *General methods for the subjective assessment of sound quality*, Recommendation ITU-R BS.1284-2.
11. International Telecommunication Union (2022), *Multichannel stereophonic sound system with and without accompanying picture*, Recommendation ITU-R BS.775-4.
12. Kelly J., Woszczyk W., King R. (2020), Are you there?: A literature review of presence for immersive music reproduction, [in:] *Audio Engineering Society Convention 149*.
13. Kleczkowski P., Król A., Małecki P. (2015), Reproduction of phantom sources improves with separation of direct and reflected sounds, Archives of Acoustics, **40**(4): 575–584, doi: 10.1515/aoa-2015-0057.
14. Kotłownia Recording Studio (2023), https://kotlownia.agh.edu.pl/ (access: 12.06.2023).
15. Malecki P., Piotrowska M., Sochaczewska K., Piotrowski S. (2020), Electronic music production in ambisonics-case study, *Journal of the Audio Engineering Society*, **68**(1/2): 87–94, doi: 10.17743/jaes.2019.0048.

16. Małecki P., Stefańska J., Szydłowska M., Tęczyńska Kęska M. (2023), A listening test evaluation of spatial sound technologies in music production: Dolby Atmos and ambisonics, [in:] *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*.

17. Oramus T., Neubauer P. (2020), Comparison of perception of spatial localization between channel and object based audio, [in:] *Audio Engineering Society Convention 148*.

18. Pieprzyk Ł. (2021), Dancing Ends for Symphony Orchestra and Piano (feat. Gajusz Keska), Pedagogical University in Cracow, https://open.spotify.com/track/3fCCmIQvxLeAb0WdIuSgtj (access: 12.06.2023).

19. Power P.J. (2015), *Future spatial audio: Subjective evaluation of 3D surround systems*, University of Salford.

20. Rumsey F., Zieliński S., Kassier R., Bech S. (2005), On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality, *The Journal of the Acoustical Society of America*, **118**(2): 968–976, doi: 10.1121/1.1945368.

21. Spors S., Wierstorf H., Raake A., Melchior F., Frank M., Zotter F. (2013), Spatial sound with loudspeakers and its perception: A review of the current state, [in:] *Proceedings of the IEEE*, **101**(9): 1920–1938, doi: 10.1109/JPROC.2013.2264784.

22. Wittek H. (2013), *Perceptual differences between wavefield synthesis and stereophony*, Ph.D. Thesis (unpublished), University of Surrey.

23. Zotter F., Frank M. (2019), *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer Nature, doi: 10.1007/978-3-030-17207-7.