Latency of spatial audio plugins: a comparative study

Matteo Tomasetti
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
matteo.tomasetti@unitn.it

Angelo Farina
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
angelo.farina@unipr.it

Luca Turchet
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
luca.turchet@unitn.it

***Abstract*—The use of spatial audio plugins (SAPs) with Ambisonics processing and binaural rendering has become widespread in the last decade, thanks to their increased accessibility and usability. SAPs are particularly relevant in scenarios involving real-time music playing with headphones, such as networked music performance and individual recreational music-making using backing tracks. However, a crucial issue that has been largely overlooked thus far is the measurement of the processing latency introduced by currently available SAPs. Identifying which SAPs are the fastest is essential to enable designers, musicians, and researchers to create time-sensitive applications involving 3D audio. To bridge this gap, we compared nine systems formed by different SAPs that enable 3D audio management. We measured the latency of each system throughout the third-order Ambisonics plugins pipeline: encoding, room simulation, sound scene rotation, and binaural decoding. In particular, the measurements were performed utilizing different buffer sizes. Results showed that to achieve a minimization of the latency, it is necessary to use a combination of different SAPs from different systems. Based on our measurements, we propose two spatial audio systems that mix different SAPs. Considering a sampling rate of 48 kHz, a Dell Alienware x15 R2 laptop running the Windows 10 operating system, and an RME Fireface UFX sound card, the two systems achieved an overall latency of 0.33 ms and 0.94 ms respectively.**

***Index Terms*—Spatial audio plugins, latency measurements, binaural audio, Ambisonics, 3D audio**

I. Introduction

Immersive audio has evolved significantly in the past two decades, with applications in concert halls, theaters, home cinema, and beyond [1]. Nowadays, this technology finds wide-ranging utility across other various domains, including music listening [2], extended reality [3], video-on-demand services [4], and web-browser content [5]. Spatial awareness and human interaction with the environment are greatly influenced by hearing, which plays a vital role in making sense of one's surroundings and experiences in everyday life [6].

Humans demonstrate exceptional accuracy in localizing the position of sound sources by employing a diverse range of acoustic cues [7]. The latter are formed by Interaural Time Differences (ITDs), Interaural Level Differences (ILDs), and acoustic filtering, which are fundamental aspects related to sound source localization. This is especially relevant in binaural audio, which, to date, stands as the form of immersive audio most widely used by people since everyone possesses the technology (the headphones) necessary for its reproduction [8]. Binaural audio rendering systems employ head-related transfer functions (HRTFs) to generate acoustic cues, which are acoustic transfer functions that rely on the spatial position of a sound source relative to the listener's head. HRTFs encode the necessary acoustic information for sound localization in headphones and play an essential component in auditory perception [9]. HRTFs capture the alterations in the sound spectrum as it enters the ear canal and are influenced by the diffraction and reflection resulting from each individual's unique physical characteristics [10]. These individualized or personalized HRTFs can be obtained through acoustic measurement and organized into databases [11], [12]. Due to the expensive measurement process, reduced portability of 3D loudspeaker systems for the recording, and computational difficulties, binaural decoders rely on generic HRTFs, resulting in lower sound localization accuracy and a more considerable margin of error [12]. Apart from the localization errors caused by HRTFs, binaural systems face issues with front-back confusion and externalization [13]. One solution to mitigate localization errors is using head-tracking devices that adapt the binaural rendering following the user's head movements [14].

Nowadays, the most common approach musicians, composers, and sound designers utilize to work with spatial audio (both for loudspeakers array and binaural) is the use of Higher Order Ambisonics (HOA) [15], which can be simply integrated as Virtual Studio Technology (VST) audio plugins [16] into classic digital audio workstations (DAWs), such as Reaper[1]. Introduced by Gerzon [17], Ambisonics is a sound reproduction technique that enables the creation of a complete 3D virtual acoustic environment with multiple moving sound sources using a determinate set of playback channels. A complete review of Ambisonics can be found in Zotter and Frank [18]. The first Ambisonics technique introduced by Gerzon employed only the 0th and 1st-order directional patterns (spherical harmonics), specifically the omnidirectional (W) and three dipole components (X, Y, Z), known as B-Format.

[1] https://www.reaper.fm/

979-8-3503-1104-4/23/©2023 IEEE

However, the limited spatial resolution of the 1st order restricts accurate sound field reconstruction to a small listening area. To overcome this limitation, Higher Order Ambisonics (HOA) [19] extends the B-Format by utilizing spherical harmonic decomposition of the sound field at higher orders, resulting in an expanded reproduction area, at the price of a much larger channel count [20]. Currently, it is possible to work with a max of 7th-order Ambisonics, as the best DAWs have a practical limit of 64-channels, but many others are limited to just 16 or 32.

We have recently shown that musicians prefer to play with others through headphones with binaural audio (created with HOA) and head-tracking, compared to classic stereophonic audio streams [21]. Such a result has highlighted the need of equipping current headphones-based systems with spatial audio capabilities in order to optimally support the playing experience with others for networked music performances [22], personal practice with musical instruments, music-making using backing tracks, and studio recording sessions. Especially for the case of networked music performances, knowing the latency introduced by such SAPs is paramount in order to create systems that do not exceed the maximum latency tolerable by musicians to play synchronously.

However, to the best of the author's knowledge, nowadays, it is unknown which suites of spatial audio plugins (SAPs) to produce HOA audio with binaural rendering are most efficient in terms of processing latency. The latter, in our case, means how long a particular SAP takes to perform that particular type of processing and with how much latency (expressed in audio samples and subsequently converted in milliseconds). Furthermore, with the term suite, we refer to SAPs that are developed by a single company or research center (see Section II-B). To bridge this gap, in this paper, we analyze the processing latency of the currently available SAP suites that treat HOA audio and binaural rendering. Notably, our interest is only in assessing SAP suites from the sole latency perspective, not from the perceptual quality standpoint. After the description of the measurements methods detailed in Sections III and IV, we propose in Section V-A two efficient workflows with two SAPs systems capable of performing HOA audio processing and binaural rendering (or binaural decoding) with the lowest possible processing latency. In that case, we use the term SAPs system because we selected different SAPs from different suites.

II. Related Work

This Section is divided into two parts: in Section II-A, we describe all the processes accomplished when working with the suites of SAPs available for Ambisonics and binaural rendering; in Section II-B, we describe the suites of SAPs most used today by musicians and composers to work with Ambisonics audio and binaural rendering within their DAWs. The audio signal of a given sound source must first be converted to Ambisonics, and this process is called encoding (Section II-A1). After that, the simulation of the reverberation of a specific environment is usually added, and we refer to it as room simulation (Section II-A2). After this step, specific SAPs are used to perform head-tracking by responding to the head's attitude information provided by head-tracking devices, and we refer to this part as rotation with head-tracking (Section II-A3). Finally, the Ambisonics signal is decoded into binaural through the process of binaural rendering (Section II-A4).

A. Ambisonics to binaural workflow

There are established methods for recording and reproducing 3D sound scenes. These involve using an array of microphones to capture the 3D audio scene and then directly mapping their signals to the corresponding channels of the planned playback system (which can be stereo, headphones, or surround formats). However, these methods are often inflexible when reproducing recordings on different playback systems or accounting for variations in the user's head orientation (for the circumstance of binaural microphone array recordings) [23]. It is precisely to solve these limitations and have more flexibility that Ambisonics is widely used, with its encoding and decoding processes.

1) ***Encoding:*** the first type of processing that is applied to the sound signal to convert it to the Ambisonics domain is usually called encoding. Here the input microphone signals are transformed into intermediate spherical harmonic (SH) signals using linear and signal-independent mapping. For a review about SH signals, please refer to the study of Rafaely reported in [24]. The encoding process allows for the selection of Ambisonics order, which determines the number of channels and, consequently, the spatial detail of the reproduced sound field and the computational load and memory needed [25]. To date, the encoding process is implemented in all the suites of SAPs analyzed in this paper (see Section II-B). Typically, encoding SAPs that convert the signal to SH allow the sound signal to be positioned in the 3D space within their user interface, usually by specifying the azimuth and elevation parameters in a reference system compliant with the ISO 2631 standard. It must be understood, however, that in some workflows, Ambisonics encoding does not represent the position of the source in the sound scene from the perspective of the listener; it represents the directivity and orientation of the sound source. In these cases, the resulting encoded multichannel stream is usually named O-format instead of B-format [26].

2) ***Room simulation:*** in binaural systems, the addition of room simulation is crucial to enhance sound source localization capabilities, improve distance perception (since the effective distance of the sound sources is not a directly controllable parameter in the Ambisonics domain), and mitigate externalization issues [27], [28]. A long tradition of research is present in the context of acoustics, modeling, room simulation, and their algorithms. For a comprehensive review of this scientific field, please refer to the study of Vorländer reported in [29]. Room simulation in the Ambisonics domain is generally implemented with one of two different techniques: algorithmic and convolution. The main differences between the two relate to the fact that in algorithmic ones, the early reflections and the reverberant sound field are created through specific algorithms. Conversely, in convolution techniques, the sound signal is convolved with a multichannel impulse response (IR) of a given environment (which is recorded through specific methods and microphone arrays). Convolution-based SAPs allow for loading different IRs. In some workflows, the source signal is in O-format, representing its directivity and orientation [30].

*3) **Rotation with head-tracking:*** in order to enhance the localization's accuracy and minimize errors and ambiguities within Ambisonics-binaural systems, it is fundamental to integrate wearable head-tracking devices [31]. These external head-tracking devices, commonly referred to as head-trackers, allow the relative three-dimensional rotation coordinates (yaw, pitch, and roll), or, better, the quaternion [32] to be sent via USB, Bluetooth, or WiFi, usually using the Open Sound Control (OSC) [33] protocol. Specific SAPs receive these data and rotate the entire Ambisonics sound stream in real-time. The data that are sent vary according to the attitude of the user's head. It is essential to note that another critical aspect is added during this step: the tracking latency. This paper does not focus on this type of latency since our primary purpose is to calculate the different processing latencies of the SAPs currently available.

*4) **Binaural rendering:*** the headphone signal is created through specific SAPs that allow for binaural rendering, converting Ambisonics signals into binaural signals. The binaural rendering SAPs aim to recreate the spatial attributes of sound sources recorded or generated in Ambisonics format through headphone listening. The field of binaural rendering for HOA signals is an active area of research, where the primary challenge lies in identifying an optimal filter matrix that can accurately renders the Ambisonics signals into the signals corresponding to each ear [34]. Binaural audio rendering through headphones presents a series of reasonably well-documented challenges. We have, in part, already presented them in Section I, and these should be necessarily taken into account during the binaural decoding process. Please, see the works of Faller and Breebaart [13], and Møller [35] for a more in-depth overview of the challenges.

B. Spatial audio plugins

In this Section, we describe all the suites of SAPs that we subsequently measured and analyzed in terms of processing latency, as reported in Sections III and IV. We highlight that some of the SAPs suites we have measured do not contain some essential components, such as SAPs to produce room simulation or sound scene rotation with head-trackers, but only allow to perform Ambisonics encoding and binaural decoding. Some other SAPs integrate two functions, for example, encoding and room simulation or rotation and binaural decoding. Moreover, some SAPs cannot be used with buffer sizes smaller than a given size; we detail these aspects in Sections III and IV. Furthermore, in our measurements, we also included the suite of SAPs developed by Facebook, named *Facebook Spatial Workstation*, which is no longer maintained by developers but is still been widely employed by many users, especially in the case of post-production for 360° video [36]. For each SAPs suite, we also describe the individual SAPs we used in our measurements that are utilized to produce encoding, room simulation, sound scene rotation with external head-trackers, and binaural decoding processes.

IEM Plug-in Suite (v.1.14.0) [15]: it is a collection of open-source SAPs created by students and researchers at the Graz Institute of Electronic Music and Acoustics (IEM, Austria). These SAPs offer a wide range of encoding and decoding capabilities for Ambisonics signals up to the 7th order. In this suite, among the SAPs, there is the *BinauralDecoder*, which utilizes the Magnitude Least-Squares (MagLS) approach introduced by Schörkhuber et al. [34] that enables the conversion of Ambisonics-encoded input signals into binaural headphone signals. Particularly noticeable is also the *Room Encoder* SAP, which performs a rectangular room simulation where the source and listener can be arbitrarily positioned and moved, processing a O-format input signal and delivering a B-format rendered signal, which takes into account source position and orientation, and listener position and orientation, with the capability of adjusting the room size and the amount of absorption on its surfaces. We measured the processing latency of the following SAPs of this suite: *StereoEncoder*, *RoomEncoder*, *SceneRotator*, and *BinauralDecoder*.

Spatial Audio Real-Time Applications (SPARTA, v.1.6.2) [37]: it is a collection of flexible and signal-dependent SAPs developed by the Acoustics Lab at Aalto University (Finland). These SAPs aim to enhance immersive audio production, reproduction, and visualization beyond traditional linear Ambisonics algorithms. They extract and utilize parameters over time to map the input to the output, resulting in adaptive and informed spatial audio processing. This is the main difference between parametric and linear algorithms. We measured the processing latency of the following SAPs of this suite: *ambiENC*, *ambiRoomSim*, *sparta rotator*, and *ambiBIN*. We underline that several decoding methods are available within the *ambiBIN* SAP. We chose the MagLS method for our measurements because it is also used by the IEM *BinauralDecoder*, and we left the other *ambiBIN* default parameters unchanged. Furthermore, *ambiBIN* can also perform head-tracking, making the usage of *sparta rotator* unnecessary.

03A CORE (v.2.2.1): it is a collection of SAPs developed by the *Blue Ripple Sound*[2] company in London. These SAPs offer all the necessary tools to create an HOA audio mix. It is available as a free download, providing users with access to the essential components developed by the *Blue Ripple Sound* company. However, SAPs for room simulation and binaural rendering are unavailable in the freeware part of this suite. For this reason, we used the corresponding older TOA plugins (an older version of O3A, which differs just for channel ordering and normalization but employs the very same processing algorithms). So the plugins that were tested in the

[2] http://www.blueripplesound.com/index

O3A setup were the following: *O3A panner - hemisphere*, *TOA reverb*, *O3A Rotation*, and *TOA Decoder - Headphones*.

ambiX (v.0.3.0) [16]: these are open-source SAPs, created by Mathias Kronlachner[3], enabling the production of Ambisonics content and post-production work on recordings, such as those captured by Soundfield microphones. Different SAPs are provided for Ambisonics orders equal to 1, 3, 5, and 7, offering flexibility to the users. Working with binaural audio within this SAPs suite requires specific preconfigurations (presets), including filter matrices and binaural IRs provided by the developer, and must be inserted into the *ambiX binaural* SAP. Room simulation is performed using the *mcfx convolver* and a preset containing measured impulse responses in Ambisonics B-format. We measured the latency processing of the following SAPs of this suite: *ambix encoder*, *mcfx convolver*, *ambix rotator*, and *ambix binaural*.

ICST (v.2.3.0)[4]: they are SAPs developed at the University of Fine Arts in Zurich (Switzerland) and are versatile tools for creating Ambisonics content. They support the simultaneous spatialization (encoding) of up to 64 audio sources (up to 7th-order Ambisonics) and decoding with up to 64 loudspeakers. Noteworthy features include an interactive graphical radar view and the ability to record positioning information in the encoding SAP through the user interface or via OSC protocol. However, there is no binaural rendering SAP within this suite, so we have used Dear Reality[5] company's *DearVR Ambi Micro* for the binaural rendering since the developers of this suite recommended it[6]. *DearVR Ambi Micro* currently only supports audio signals reaching the 3rd Ambisonics order. For the measurements, we used the following SAPs of this suite, which allow only for the encoding and binaural rendering processes: *ICST AmbiEncoder o3*, and *DearVR Ambi Micro*.

X-MCFX (v1.0.4): it is a SAP performing any task that can be expressed as a matrix of finite impulse response (FIR) filters, capable of processing a large number of inputs and outputs (up to 128x128). It is an enhanced version of the *mcfx convolver* SAP created by Mathias Kronlachner within the MCFX SAPs suite[7]. *X-MCFX* offers improved performance, an easier way of loading the FIR filter matrix (using a single multichannel WAV file), and provides an expanded number of channels. With this SAP, users can specify the folder path containing the FIR filter matrices in WAV format for producing the convolution within the SAP. Furthermore, users can select a desired filter matrix from a drop-down list or navigate to a specific file in another folder if necessary. Regarding the present work, *X-MCFX* can be used, with a suitable FIR filter matrix, for performing 3 tasks: encoding, room simulation by convolution with a MIMO filter matrix, and binaural decoding. Rotation is not easily implemented with *X-MCFX*. However, whatever the task performed, *X-MCFX* has an optional setting allowing it to operate in "zero latency" mode at the cost of some additional CPU load. Consequently, the results of latency measurements done on *X-MCFX* were independent of the host buffer size: the algorithm is slightly different than the original *mcfx convolver*, and internal buffering has been removed. But, as this is a convolver, the user gets a latency if this is embedded in the FIR filters employed. This is particularly relevant for room simulation, as using measured impulse responses, typically one gets a latency given by the *time-of-flight*, which can only be removed by manually editing the impulse response WAV file, removing all the silence before the direct sound. This kind of latency is independent of the host's buffer size.

FB360 (v.3.3.3)[8]: this suite encompasses a collection of SAPs tailored for DAWs, a virtual reality video player, and a versatile native engine compatible with multiple platforms. This SAPs suite simplifies creating and delivering content for cinematic virtual reality and 360° video projects, offering a seamless and efficient solution. Within this suite, it is possible only to work with the Ambisonic signal up to the 3rd order. The SAPs we measure are the *FB360 Spatialiser*, where it is possible to do the encoding, and there is a control (on/off) for turning on the room simulation. There is the *FB360 Converter*, which is the binaural renderer SAP, and the *FB360 Control*, where it is possible to control the room simulation's parameters and perform conversion to binaural with head-tracking commanded by a head-mounted display (HMD).

The following two SAPs that we measured are the only ones that do not use Ambisonics; the rendering is "direct" from the virtual source position to binaural, including room effect and reverb. Not passing through Ambisonics provides generally sharper localization and proper Interaural Time Difference. Nevertheless, these aspects are not the focus of the present paper. Not going through spherical harmonics, it is much more difficult to perform head-tracking: in fact, the following two SAPs do not support natively head-trackers. Head-tracking could be added if the host program manages the OSC data received from the head-tracker and adjusts the azimuth-elevation of each sound source accordingly, modifying them through the automation parameters which are exposed by the SAP. One host program capable of such processing is Max[9]. This indeed does not alter the processing latency, so we ignored this possibility in the present work.

3D Tune-In Toolkit (v.1.1.4) [38]: it is an open-source C++ library provided also as a VST plugin. Developed collaboratively by teams at the University of Malaga and Imperial College London, this toolkit serves as a comprehensive solution for sound spatialization (both binaural and loudspeakers), hearing loss simulation, and hearing aids. It offers a standard platform for various applications related to immersive audio and auditory research. In this paper, we used the single SAP of this toolkit provided by the developers, which does all the

[3] https://www.matthiaskronlachner.com/

[4] https://ambisonics.ch/

[5] https://www.dear-reality.com/

[6] https://ambisonics.ch/post/icst-ambiplugins-in-logic-pro

[7] https://www.matthiaskronlachner.com/?p=1910

[8] https://github.com/facebookarchive/facebook-360-spatial-workstation

[9] https://cycling74.com/products/max

spatial audio processes in one single SAP.

Anaglyph (v.0.9.4) [39]: it is an audio engine that stands as a VST audio plugin tailored for binaural spatialization. It was crafted not only to aid in ongoing research endeavors, but also to extend the benefits of this research to audio engineers within conventional DAW environments. Noteworthy features within Anaglyph comprise a customizable morphological ITD model, adjustments for near-field ILD and HRTF parallax, a Localisation Enhancer, an Externalisation Booster, and support for head-related impulse response (HRIR) via *Spatially Oriented Format for Acoustics* files.

III. Material and Method

A. Setup

In this Section, we present the setup employed for performing latency measurements of all the SAPs described above. We used the Plogue Bidule[10] host software, which allows users to create, connect, and manipulate various audio and MIDI modules in a visual and flexible environment, a Dell Alienware x15 R2 laptop running the Windows 10 operating system, and an RME Fireface UFX sound card. We decided to set the sample rate at 48 kHz because it is a standard audio format. Furthermore, we added a condition in the measurement process: the buffer size (BS), set to 64, 128, 256, and 512 samples. We decided to add this condition to investigate whether the BS would impact the different processing latencies. In order to perform the measurements, we decided to simulate the binaural-Ambisonics workflow described in Section II-A for each suite of SAPs analyzed in Section II-B. In detail, we measured latencies at each step, from encoding to room simulation to sound scene rotation with head-tracking (not considering the positional latency given by the external head-tracker) to binaural rendering. Moreover, we made these measurements with signals in 3rd order Ambisonics (16 channels). Furthermore, in addition to measuring latency for each step and thus for each specific SAP, we summed up the different processing latencies to show the overall processing latency of that particular suite of SAPs. The results are summarized in Fig. 1 and Fig. 2 and described in Section IV. Fig. 1 shows the results in samples from analyzing individual SAPs of each suite with the different BSs described. 3D Tune-In Toolkit is not included in Fig. 1, as it consists of a single SAP, not employing Ambisonics, that performs the three main tasks of encoding, room simulation, and binaural decoding. 3D Tune-In Toolkit, on the other hand, is included in Fig. 2, which, instead, shows the overall results both in samples and in milliseconds (ms) derived from the sum of each SAP that constitutes a suite. As a consequence, Fig. 2 aims to draw up which suites of SAPs have the lowest processing latency considering the different BSs.

B. Measurement process

To measure the processing latencies, we used a Dirac's Delta (an impulse) as the audio signal, and we processed it according to all the steps described in Section II-A. By steps, we mean that in each SAPs suite described, there is a specific SAP that deals with that particular step. We recorded the audio signal at the output of each SAP, and we saved the file and performed the analysis within the Adobe Audition[11] software. The file we saved is an audio file that contains 5 audio channels (commonly called stems): the first is the direct signal, the second is the signal after the encoding SAP, the third is the signal after the room simulation SAP, the fourth is the signal after the sound scene rotation SAP, and the fifth is the signal after the binaural rendering SAP.

C. Algorithmic and convolution SAPs

In this Section, we specify what types and parameters we have configured in the room simulations SAP available in the suites. In IEM (*RoomEncoder* SAP), SPARTA (*AmbiRoomSim* SAP), O3A (*TOA Reverb* SAP), 3D Tune-In Toolkit, Anaglyph, and FB360 (*FB360 Control* SAP), we have the algorithmic one, while in ambiX (*mcfx convolver* SAP) and *X-MCFX* SAP we have the convolution one. In the ICST suite, there is no possibility to simulate room reverberation.

In algorithmic room simulations, we set parameters related to the room dimension, the listener's position, and the sound source's position. For setting the room dimension parameters, we relied on the results of Rindel's article [40], which shows that the ratio between length and width must be between 1.15 and 1.45 and indicate that the height can be chosen more freely without compromising the acoustic quality. For our parameters, we set the length-to-width ratio of 1.30, the width-to-height ratio of 1.25, and the room's length to 5 meters. By doing so, we have subsequently set the room's width to 3.85 meters and the value of the height to 3.08 meters for all the described algorithmic room simulations SAPs. Regarding the source and listener positions, instead, we configured the source position to be 1.80 meters away in the X-axis from the listener (which, in turn, we then set to be at 0m, that is, in the center of the room). Then we put the value 0m in the Y-axis (of both the listener and source position) and 1.5m in the Z-axis (of both the listener and source position). We chose these values to simulate an ideal condition when two musicians are together in the same room. In addition, we enabled the *Direct Path Zero Delay* parameter in all the algorithmic room simulations SAPs to avoid the latency introduced by the *time-of-flight* of the reverberation. The *time-of-flight* refers to the duration it takes for the initial sound impulse, in the case of IR measurements, to travel from the sound source to the listener. It describes the elapsed period between the emission of the impulse and the first detection of it at the receiver. In the 3D Tune-In Toolkit and Anaglyph, we left the values of the reverberation parameters with the default setting since these SAPs do not allow controlling the values of the above parameters. In algorithmic reverbs, one can remove it, while in convolution reverbs, one has to manually remove it from the recorded file (i.e., the first milliseconds of silence at the

[10] https://www.plogue.com/products/bidule.html

[11] https://www.adobe.com/it/products/audition.html

beginning of the file before the first impulse representing the direct sound).

In ambiX *mcfx convolver* and *X-MCFX* SAPs, on the other hand, there is no ability to control any type of parameter, as they are specific SAPs where IRs, mainly recorded with microphone arrays in real environments, are inserted and thus vary according to the recordings made. For the measurement of both these SAPs, we downloaded one Ambisonics 1st order IR from this website[12]. In detail, we use the *hm2_000_bformat_48k.wav* file for the latency measurements. The measured IR was edited, removing as much as possible the silence before the arrival of the direct sound (the initial 530 samples of *time-of-flight* were removed).

It is worth noticing that we chose this set of IRs in the convolution reverbs and these settings in the algorithmic reverb parameters on the basis of a pilot study conducted with three professional musicians. Consistently, all of them stated that they could hear the room simulation contribution well with the selected settings and IRs compared to others tested during the pilot study.

D. Binaural decoder SAPs

In this Section, we point out that some parameters should also be considered regarding SAPs that deal with binaural decoding, particularly in those where it is possible to produce binaural decoding by convolution, that is, by loading a file representing a filtering matrix related to HRIR, which if the Fourier transform is applied becomes the so-called HRTF. The binaural decoding by convolution is done within the ambiX *binaural* SAP, Angelo Farina's *X-MCFX* SAP, and in the binaural rendering user interfaces of the Anaglyph and 3D Tune-In Toolkit SAPs. In detail, we would like to show that Angelo Farina's *X-MCFX* SAP and the AmbiX *binaural* SAP also have zero sample processing latency. This means that these SAPs use as a minimum partition the same as the BS received from the host application (in our case, the one we set on Plogue Bidule). So the SAP does not add additional processing latency. However, if a FIR filter [41] not having a minimum phase is used in these SAPs for binaural convolution, there is a latency imposed by the filter's coefficients. This scenario is for ambiX and *X-MCFX* SAPs cases, where the user can load in the filtering matrix of his/her choice. For the measurements of the ambiX's *binaural* SAP, we utilized and loaded in the plugin the HRIR preset *icosahedron 3h3v* that we found available for download in the developer's website[13]. We used the following FIR filter matrices *ViveCinema-Ambix2Bin-256.wav* for the measurements of the *X-MCFX* SAP, which is available on the developer's website[14]. On the other hand, in the 3D Tune-In Toolkit and Anaglyph, the processing latency depends on the different filtering matrices (or HRIRs) that are available as presets in the user interface of the SAP that deals with the binaural renderer.

IV. Results

The results, illustrated in Fig. 1 and Fig. 2, indicate that the SAPs suite with the lowest latency processing for the Ambisonics-binaural workflow (described in Section II-A) is the *X-MCFX*, and secondly, FB360, which currently is no longer maintained as the developers finished supporting its development as of May 2022. In order to accomplish the full Ambisonics to binaural processing, the *X-MCFX* employs 53 samples in total, which, converted to ms at a sampling rate of 48KhZ, is equal to 1.10ms. Secondly, the FB360 SAPs suite employs 80 samples in total (1.66 ms). However, the Anaglyph SAP, which does not perform the Ambisonics-binaural workflow since it does not pass through the SH domain, appears to be one of the best performing in terms of processing latency. It achieves an overall processing latency of 22 samples, which, converted to ms at a sampling rate of 48 kHz, is equal to 0.46ms with all the buffer sizes we tested. Instead, the SAPs suite with the highest latency processing is the SPARTA, which performed the Ambisonics-binaural workflow with 1909 samples, which is equal to 39.77ms.

Below we present the results for each SAP suite we measured, analyzing them separately for more clarity, and we present the low processing latency SAPs systems that we selected in Section V-A and in Fig. 3 and Fig. 4, which combines the different SAPs of the analyzed suites.

1) ***IEM Plug-in Suite****:* in this suite, the overall processing latency time is 321 samples at all different BSs, which equals 6.69ms. The *StereoEncoder* and *SceneRotator* SAPs have 0 samples processing latency, while the *RoomEncoder* with our settings described in Section III-C has a processing latency of 195 samples (4.06 ms). The *BinauralDecoder*, instead, is of 126 samples (2.63 ms). The different BSs had no impact whatsoever in this suite.

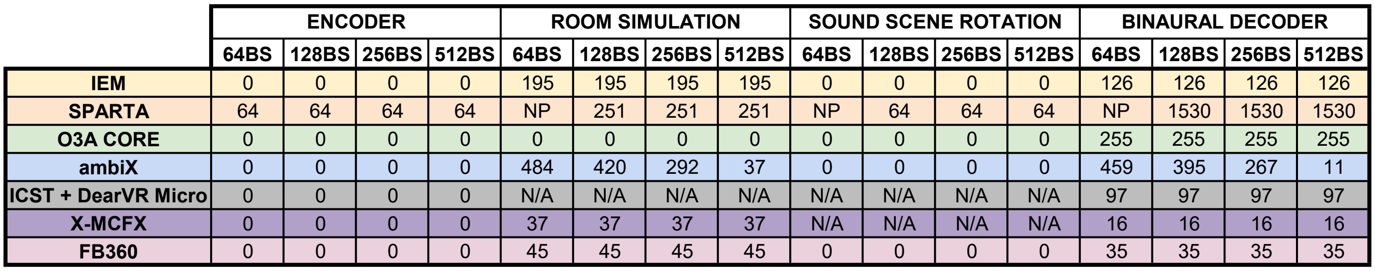
2) ***SPARTA****:* in this suite, we noticed that all SAPs except *ambiENC* cannot be used with the BS set to 64. Except for this case, the BSs had no impact. The encoding (*ambiENC*) and sound scene rotation (*sparta rotator*) SAPs are with 64 samples of processing latency (1.33 ms). The *ambiRoomSim* has 251 samples (5.23 ms), and the *ambiBIN* has 1530 samples (31.88 ms).

3) ***O3A CORE****:* we measured the processing latencies of the *O3A panner - hemisphere* and *O3A Rotation* SAPs, and they have 0 samples of processing latency. However, the *O3A Rotation* SAP does not have support for communicating via OSC with external head-trackers, but it does allow for manual rotation via automation on yaw, pitch, and roll parameters (so OSC data must be managed by the host program). *TOA reverb* has a latency of zero samples because, being a parametric reverb in which the direct sound simply makes a “passthrough” within the SAP, the direct sound does not experience any processing latency. After the direct sound follows the reverberant tail, but the direct sound suffers no processing latency. The *TOA Decoder - Headphones* has a latency of 255 samples (5.31ms). Within this suite, the various BSs had no impact.

[12] https://www.openair.hosted.york.ac.uk/?page_id=502

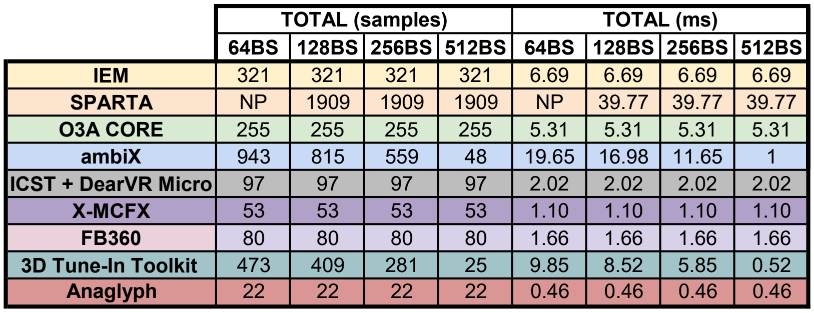
[13] https://www.matthiaskronlachner.com/?p=2015

[14] http://www.angelofarina.it/Public/ViveCinema/



	ENCODER				ROOM SIMULATION				SOUND SCENE ROTATION				BINAURAL DECODER			
	64BS	128BS	256BS	512BS	64BS	128BS	256BS	512BS	64BS	128BS	256BS	512BS	64BS	128BS	256BS	512BS
IEM	0	0	0	0	195	195	195	195	0	0	0	0	126	126	126	126
SPARTA	64	64	64	64	NP	251	251	251	NP	64	64	64	NP	1530	1530	1530
O3A CORE	0	0	0	0	0	0	0	0	0	0	0	0	255	255	255	255
ambiX	0	0	0	0	484	420	292	37	0	0	0	0	459	395	267	11
ICST + DearVR Micro	0	0	0	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	97	97	97	97
X-MCFX	0	0	0	0	37	37	37	37	N/A	N/A	N/A	N/A	16	16	16	16
FB360	0	0	0	0	45	45	45	45	0	0	0	0	35	35	35	35

Fig. 1. Representation of analysis results on processing latencies expressed in samples. **N/A** means *unavailable - the SAP is not present in the suite*, and **NP** means *not possible with this SAP at this buffer size.*



	TOTAL (samples)				TOTAL (ms)			
	64BS	128BS	256BS	512BS	64BS	128BS	256BS	512BS
IEM	321	321	321	321	6.69	6.69	6.69	6.69
SPARTA	NP	1909	1909	1909	NP	39.77	39.77	39.77
O3A CORE	255	255	255	255	5.31	5.31	5.31	5.31
ambiX	943	815	559	48	19.65	16.98	11.65	1
ICST + DearVR Micro	97	97	97	97	2.02	2.02	2.02	2.02
X-MCFX	53	53	53	53	1.10	1.10	1.10	1.10
FB360	80	80	80	80	1.66	1.66	1.66	1.66
3D Tune-In Toolkit	473	409	281	25	9.85	8.52	5.85	0.52
Anaglyph	22	22	22	22	0.46	0.46	0.46	0.46

Fig. 2. Representation of the total sum in terms of processing latency for all steps (encoding, room simulation, sound scene rotation, and binaural decoding) of each suite of SAPs described except the 3D Tune-In Toolkit and Anaglyph, which perform the binaural rendering in one single SAP since they do not pass through Ambisonics. Sums are expressed in samples and milliseconds for all buffer sizes analyzed.

*4) **ambiX**:* from the results of the measurements, it emerged that the BS has a significant impact on this suite. This is because *mcfx convolver* and *ambix binaural* have their internal buffer, which is hardcoded and which we noticed to be at 512. In the *mcfx convolver* or *ambix binaural*, for example, the SAP explicitly states that one is introducing processing latency if one changes the buffer size other than 512 samples. Of course, as we described in Section III-C, the *mcfx convolver* SAP is at 0 samples processing latency, and the latency we calculated depends on the IR we used (where, in turn, we removed its *time-of-flight*). The same scenario happens with the *ambix binaural*, where the processing latency depends on the filtering matrices (HRIRs) and their coefficients that are loaded into the SAP. Nevertheless, we noticed also that the *ambix encoder* and *ambix rotator* have 0 samples processing latency at all the BSs, but it is not possible to communicate via OSC via external head-trackers. It is only possible to control the yaw, pitch, and roll parameters manually via automation.

The results from the *mcfx convolver* measurements are: 64BS = 484 samples (10.08ms); 128BS = 420 samples (8.75ms); 256BS = 292 samples (6.08ms); 512BS = 37 samples (0.77ms). Indeed, the results from the *ambix binaural* measurements are: 64BS = 459 samples (9.56ms); 128BS = 395 samples (8.23ms); 256BS = 267 samples (5.56ms); 512BS = 11 samples (0.23ms);

*5) **ICST plus DearVR Ambi Micro**:* in this suite, we measured the processing latency of the *ICST AmbiEncoder o3* and *DearVR Ambi Micro*. The *ICST AmbiEncoder o3* SAP is at 0 samples processing latency, while the binaural rendering SAP is 97 samples (2.02ms). In this suite, the different BSs have no impact on processing latency.

*6) **X-MCFX**:* this matrix convolution SAP operates in "zero latency" mode, but the measured latency comes from the FIR filters employed. Encoding can be done at zero latency if using minimum-phase filters simply expressing the gain for each channel, represented by the theoretical encoding formulas[15]. Some samples of latency come from the room impulse response employed for reverb, where, despite cutting away almost all of the *time-of-fight*, the direct sound has some pre-ringing, causing the main peak to be delayed by 37 samples (0.77ms). Similarly, the binaural decoding filter matrix employed has the peak at the 16th sample, causing some further latency of 0.33ms. Overall, however, *X-MCFX* SAP can perform all the three main tasks (encoding, room reverb, and binaural decoding) with very small latency, which can be further reduced by operating on the filters employed.

*7) **FB360**:* the results that emerged from the analysis of the processing latencies of this suite show that the different BSs have no impact. The results are: *FB360 Spatialiser* is at 0 samples processing latency, *FB360 Converter* is 35 samples (0.73ms), and of *FB360 Control* is 45 samples (0.94ms).

V. Discussion

Our findings demonstrate how the spatial audio plug-ins (SAPs) involved in producing Higher Order Ambisonics (HOA) encoding are all at 0 samples processing latency in all

[15] http://www.angelofarina.it/Aurora/HOA_explicit_formulas.htm

the buffer sizes (BSs) we tested, with the exception of the SAP for encoding within the SPARTA suite, which has 64 samples of processing latency.

Regarding room simulation, the difference between the two types described in Sections II-A2 and III-C is significant because the algorithmic one allows for relatively low processing latency, and this one highly depends on the setting of parameters within the room simulation SAPs. In contrast, the IR-dependent convolution reverb one has the *time-of-flight* that must always be considered. It would be necessary for future research to understand which of the two types of binaural reverberation is preferred by musicians or listeners in perceptual and quality perspectives.

In terms of sound scene rotation with external head-trackers, it is significant to report that the *sparta rotator* SAP has the same scenario reported in the encoding process, which is the 64 samples of processing latency, but it supports the communication via OSC with external head-trackers devices. Furthermore, in the ICST suite, Anaglyph, and 3D Tune-In Toolkit, the SAP for sound scene rotation is not implemented, while ambiX, FB360, and O3A have the SAP for sound scene rotation. However, the OSC protocol is not implemented, so it is impossible to communicate with external head-tracker devices. For this reason, using the sound scene rotation SAP to communicate via OSC to external head-trackers is possible only in the IEM and SPARTA suites. It is essential to point out that, in the processing latency measurements of the sound scene rotation SAPs, we have not measured the latency of data transmissions (tracking latency), which depends on the different head-trackers used, as this does not affect the audio processing latency.

Regarding the SAPs that deal with binaural rendering, the SAPs with the most negligible processing latency are in ascending order: *X-MCFX*, *Anaglyph*, *FB360 Converter* (FB360), then that of *DearVR Ambi Micro*, and finally, IEM *BinauralDecoder*. We do not include the ambiX *binaural* SAP and the 3D-Tune In Toolkit SAP among the best in terms of processing latency because the processing latency values change significantly as different BSs change within this SAP. We highlight the fact that the processing latencies greatly depend on the type of approach that is used in the algorithm itself and on the filtering matrices or HRIRs and their relative filtering coefficients that are used in the part of convolution, which also have a considerable impact on the calculation of the processing latency. More research should emerge in the future on the perceptual and quality study of the different binaural rendering SAPs available.

In terms of BSs, we noticed that it impacts the processing latency, especially in the following SAPs suites: SPARTA, ambiX, and the 3D Tune-in Toolkit. This factor about the different BSs is essential, as it demonstrates that some SAP suites can be used with low processing latencies, as in the case of musicians who have to play with others in networked music performances or practice alone with the musical instrument over backing tracks. In these cases, it is fundamental to use the Ambisonics-to-binaural workflow, which allows for the lowest processing latency with the lowest possible BS. This is to avoid perceiving the unpleasant effect of playing out of real-time and, therefore, to perceive the latency between the produced sound and the processed spatial sound. In the case of composers and sound designers, on the other hand, who have to compose in DAWs in a deferred way, it is not necessarily advisable to use SAPs with low BSs.

A limitation of this study lies in the fact that we have calculated the processing latencies of the convolution room simulations (e.g., ambiX *mcfx convolver*) and the convolution binaural decoders SAPs (e.g., *ambix binaural*) only with the presets that we have described in Sections III-C and III-D. Moreover, we only made these measurements with a Windows 10 laptop. It would be interesting to investigate the variance of the results by making the same measurements also in other machines and with other platforms. Furthermore, it would also be interesting to investigate other configurations in algorithmic reverbs and other IRs in convolution reverbs than the ones employed in this study.

A. The two selected SAPs systems

In this Section, we present two spatial audio systems that we selected among the different SAPs suites for performing the Ambisonics-to-binaural workflow with the lowest processing latency and that resulting from our measurements. These two systems are summarized in Fig. 3 and Fig. 4. We use the word system in this case as the following SAPs are selected from different suites. We present two of them as the first system is based on the highest-performing SAPs regarding processing latency that we selected from different suites. The second system, instead, is essentially based on the SAPs of FB360, which has proved to be the second most performing in terms of processing latency but which, at the same time, is no longer available and maintained by the developers. In particular, we chose to avoid using the Anaglyph SAP, even though it is the plugin that has been shown to perform as one of the best in terms of processing latency. The reason is that for the following spatial audio systems, we selected and mixed together only different SAPs that enable to perform the Ambisonics-binaural workflow described in Section II-A.

1) ***First selected SAPs system:*** as for the SAPs for encoding, we selected the *ICST AmbiEncoder o3* SAP. Regarding the room simulation, the SAP for the reverberation with the first lowest processing latency is the *TOA reverb* SAP of the O3A suite. When using *TOA reverb*, it must be understood that the positioning of the source in space is made by *TOA reverb*, and the task of the *ICST AmbiEncoder* becomes that of defining the directivity and the aiming of the source, not its position in space. As *ICST AmbiEncoder* has no control of how much the sound is "spread", hence how wide is the beam radiated from the source, it could be advisable to employ another encoding SAP if the user wants to control the beamwidth of the source directivity. Alternatively, one can reduce the Ambisonics order of *ICST AmbiEncoder* for widening the beam. As with the first system, for the rotation of the sound scenes with an external head-tracker that communicates via OSC, we have selected

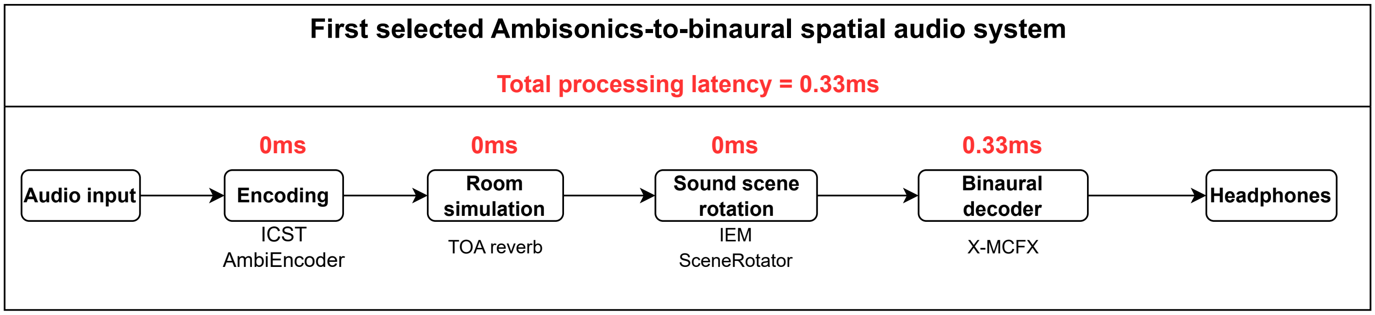
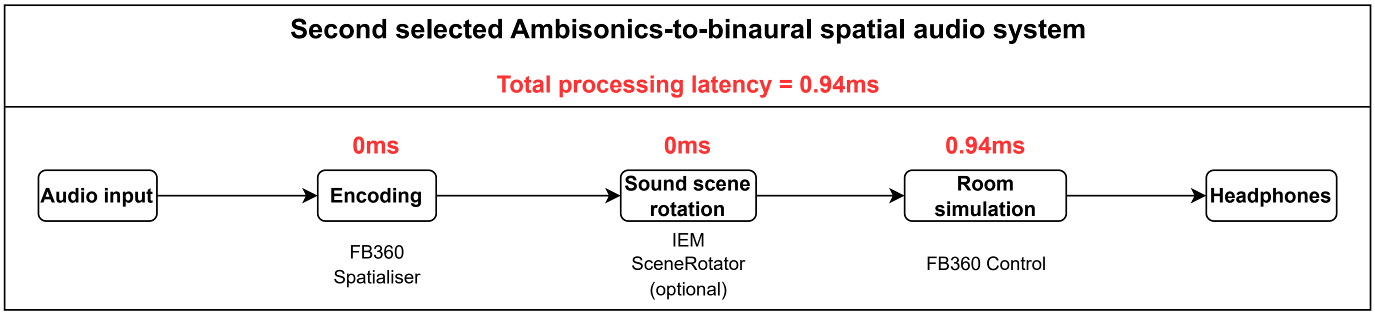


Fig. 3. Representation of the first spatial audio system we selected among the different SAPs, which allows having a SAPs system with the first lowest processing latency to make the Ambisonics-to-binaural workflow. *Audio Input* means the digital audio's real-time input signal (e.g., guitar, bass, vocals).



Second selected Ambisonics-to-binaural spatial audio system
Total processing latency = 0.94ms
0ms
0ms
0.94ms
Audio input
Encoding
Sound scene rotation
Room simulation
Headphones
FB360 Spatialiser
IEM SceneRotator (optional)
FB360 Control

Fig. 4. Representation of the second spatial audio system we selected with the SAPs taken mainly from the FB360 suite, which allows having a SAPs system with the second lowest processing latency to make the Ambisonics-to-binaural workflow. *Audio Input* means the digital audio's real-time input signal (e.g., guitar, bass, vocals).

the IEM *SceneRotator* SAP. As a binaural rendering SAP, however, we have chosen *X-MCFX* since, from what emerged from our measurements, it is the first of the SAPs for binaural decoding that has the lowest processing latency.

*2) **Second selected SAPs system:*** as for the SAPs for encoding, we selected the *FB360 Spatialiser* because it is heavily linked to the (optional) video player and to room simulation and binaural rendering with an HMD described below. Regarding the room simulation, the SAP for parametric reverberation with the second lowest processing latency is the *FB360 Control* SAP. For the rotation of the sound scenes with an external head-tracker that communicates via OSC, we inserted the *SceneRotator* SAP of the IEM, which provides a processing latency of 0 samples. As a binaural decoder, we employed the one already included in the *FB360 Control* SAP because it is also the fastest in terms of processing latency. The result in this system's processing latency is even smaller than what we have shown for the whole FB360 suite, as we removed the *FB360 Converter* SAP, as the binaural rendering can be already performed inside the *FB360 Control* SAP, so there is no need to use a separate SAP for this task. It must be noticed that if an HMD is employed instead of a standard OSC head-tracker, it is possible to get rid also of the IEM *SceneRotator* SAP, as the *FB360 Control* SAP already performs head-tracking when an HMD is employed for watching the associated panoramic video.

VI. Conclusion

In this paper, we measured the processing latency introduced by nine different suites of nowadays available SAPs. We measured them with different buffer sizes (BSs), and the sample rate was set at 48 kHz. In addition to presenting the measurement results, we proposed two spatial audio plugins (SAPs) systems selected on the basis of the lowest processing latency possible. The first of the two SAPs systems achieves an overall processing latency of just 0.33 ms for accomplishing the Ambisonics to binaural workflow we described in Section II-A. The second system, instead, achieves an overall processing latency of 0.94 ms.

Our findings are fundamental to determining which SAPs are the most immediate regarding processing latency, enabling musicians, designers, and researchers to develop time-sensitive applications affecting 3D audio. Knowing the processing latencies of the currently available SAPs is particularly relevant in scenarios involving real-time music playing with headphones. These include networked music performances and individual recreational music-making using backing tracks, where typically, the different SAPs are utilized. Moreover, these results are relevant to support researchers and companies working in the convolution reverbs SAPs as well as in immersive networked music performances since many suites of SAPs can also be implemented in embedded systems (e.g., Linux-based platforms typically used in networked performance systems).

There are various avenues to expand the findings of this study. One of these consists of repeating the same study but on different machines and platforms to investigate how much they impact the processing latencies of the SAPs. Another avenue for future work concerns the inclusion of musicians to assess the qualities of the SAPs from the point of view of the musicians' perception. Indeed, in this study, we have measured the SAPs only from the point of view of processing latencies, not of head-tracking latency. Finally, we plan to measure the latencies with other settings and IRs in the room simulations and with different filtering matrices regarding the SAPs dealing with binaural decoding.

References

[1] M. F. Davis, "History of spatial coding," *Journal of the Audio Engineering Society*, vol. 51, no. 6, pp. 554–569, 2003.

[2] J. Kelly, W. Woszczyk, and R. King, "Are you there?: A literature review of presence for immersive music reproduction," in *Audio Engineering Society Convention 149*. Audio Engineering Society, 2020.

[3] L. Turchet, R. Hamilton, and A. Çamci, "Music in Extended Realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[4] S. George, S. Zielinski, and F. Rumsey, "Feature Extraction for the Prediction of Multichannel Spatial Audio Fidelity," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1994–2005, 2006.

[5] A. McArthur, C. Van Tonder, L. Gaston-Bird, and A. Knight-Hill, "A survey of 3D Audio through the browser: practitioner perspectives," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–10.

[6] F. Rumsey, *Spatial Audio*. Taylor & Francis, 2012.

[7] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[8] T. Walton *et al.*, "The overall listening experience of binaural audio," in *Proceeding of the 4th International Conference on spatial audio (ICSA 2017), Graz*, 2017.

[9] A. Kulkarni, S. Isabelle, and H. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.

[10] P. Strumillo, *Advances in Sound Localization*. BoD–Books on Demand, 2011.

[11] D. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF personalization using anthropometric measurements," in *2003 IEEE workshop on applications of signal processing to audio and acoustics (IEEE Cat. No. 03TH8684)*. Ieee, 2003, pp. 157–160.

[12] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, "HRTF database at FIU DSP lab," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 169–172.

[13] C. Faller and J. Breebaart, "Binaural reproduction of stereo signals using upmixing and diffuse rendering," in *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.

[14] H. Wierstorf, A. Raake, and S. Spors, "Binaural assessment of multichannel reproduction," *The technology of binaural listening*, pp. 255–278, 2013.

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

[16] M. Kronlachner, "Plug-in suite for mastering the production and playback in surround sound and Ambisonics," *Gold-Awarded Contribution to AES Student Design Competition*, 2014.

[17] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[18] F. Zotter and M. Frank, *Ambisonics: A practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Nature, 2019.

[19] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," in *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[20] S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with Higher Order Ambisonics - Objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, 2006, pp. 20–23.

[21] M. Tomasetti and L. Turchet, "Playing with others using headphones: Musicians prefer binaural audio with head tracking over stereo," *IEEE Transactions on Human-Machine Systems*, pp. 1–11, 2023.

[22] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on Networked Music Performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[23] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric Ambisonic encoding of arbitrary microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2062–2075, 2022.

[24] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[25] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2711–2721, 2013.

[26] D. Menzies, "W-Panning and O-Format, tools for object spatialization," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

[27] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3D Ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Audio Engineering Society, 2003.

[28] L. Picinali, A. Wallin, Y. Levtov, and D. Poirier-Quinot, "Comparative perceptual evaluation between different methods for implementing rever-beration in a binaural context," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[29] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Nature, 2020.

[30] D. Menzies and M. Al-Akaidi, "Ambisonic synthesis of complex sources," *Journal of the Audio Engineering Society*, vol. 55, no. 10, pp. 864–876, 2007.

[31] B. F. Katz and L. Picinali, "Spatial Audio applied to research with the blind," *Advances in sound localization*, pp. 225–250, 2011.

[32] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "Ambix - A suggested Ambisonics format," in *Ambisonics Symposium*, vol. 2011, 2011.

[33] M. Wright, A. Freed, and A. Momeni, "2003: OpenSound Control: State of the Art 2003," *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*, pp. 125–145, 2017.

[34] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of Ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, vol. 44, 2018, pp. 339–342.

[35] H. Møller, "Fundamentals of binaural technology," *Applied acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[36] J. Paterson and O. Kadel, "Immersive audio post-production for 360° video: workflow case studies," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, 2019.

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

[38] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PloS one*, vol. 14, no. 3, p. e0211899, 2019.

[39] D. Poirier-Quinot and B. F. G. Katz, "The Anaglyph binaural audio engine," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[40] J. H. Rindel, "Preferred dimension ratios of small rectangular rooms," *JASA Express Letters*, vol. 1, no. 2, p. 021601, 2021.

[41] T. Saramaeki, "Finite Impulse Response Filter Design," *Handbook for digital signal processing*, p. 155, 1993.