Spaced A-B placements of higher-order Ambisonics microphone arrays: Techniques for recording and balancing direct and ambient sound

Florian Grond*, Jack Kelly and Wieslaw Woszczyk

McGill University Schulich School of Music,
555 Sherbrooke St. W. Montreal, Quebec, Canada H3A 1E3
(Received 18 September 2021, Accepted for publication 7 December 2021)

Abstract: In acoustic music recording, spaced microphone positions typically give an impression of spaciousness and depth in the rendered sound scene. Inter-channel decorrelation, particularly at low frequencies, is thought to contribute to this effect, and cannot be achieved through coincident techniques. Higher order Ambisonics microphones have seen a rapid increase in popularity, despite their limitation as a coincident recording technique. In this article, recording and mixing strategies for the A-B spaced placement of HOA arrays are described combining the advantages of both: sound stage depth and stability, and the compatibility of b-format based mixing with respect to rendering for binaural and immersive 3D reproduction loudspeaker systems. Two tests were conducted to evaluate a) the suitability of various A-B spacings with regards to direct and/or ambient sound, and b) a novel two-dimensional mixing control tool, to study direct and ambient sound optimization. Results show that increased A/B spacing of HOA receivers correlates to more optimal presentation of ambient sound. Furthermore, users interacted with the mixing control tool differently when asked to optimize the sound scene for either direct or ambient sound, or both, providing support for our microphone placement and mixing techniques. This work points to new directions for the integration of HOA microphones with traditional recording techniques towards improved spatial representation of mediated music.

Keywords: Musical acoustics, Higher order Ambisonics, Spaced microphone arrays

1. INTRODUCTION

Many stereo and surround sound recording techniques take advantage of a variety of spaced microphone arrays, generally resulting in inter-channel decorrelation between microphone signals to support a well differentiated spatial image and a wide soundstage [1]. Classic stereo examples for this approach distinguish typically between near coincident configurations with varying microphone spacings and angles such as NOS or ORTF [2], and wider A-B spaced techniques, such as Decca Tree [3]. These established techniques for stereo productions were then expanded to microphone setups for surround recordings such as the Fukada Tree [4], Double ORTF [5], the Hamasaki square [6] and the IRT cross [7] to name only a few. While near coincident techniques orient their spacing on the scale of between-ears distances, wider A-B spacings leverage the decorrelation of the signal across the whole spectrum to generate a sense of envelopment [8,9]. A common feature of these and other surround recording systems is the distinction between channels devoted to direct sound, and those devoted to ambience capture, the latter being typically used to enhance listener envelopment. For a general discussion of 3D arrays see [10].

Ambisonics microphones, in contrast, are a coincident technique the principles of which were established in the 1970ies [11]. Ambisonics has only recently gained momentum due to two factors. Firstly, the increasing importance of AR and VR technologies with their need for creating and manipulating isotropic sound scene representations that integrate with immersive and interactive 360-degree video content. Secondly, HOA microphone arrays have recently become commercially available at an affordable price point. For sound recording, mixing, and rendering, Ambisonics is both a way of recording the acoustic field at a specific point in a sound scene, and a way of manipulating a multichannel capture through the channel independent intermediary b-format. For a recent compilation of Ambisonics signal processing techniques see [12].

By modeling of the recorded sound scene as an acoustic field, Ambisonics typically approximates one selected perspective. In contrast, traditional channel-based tech-
niques capture sounds from various positions and blend the signals together at the mixing stage. Today, multipoint HOA and 6DOF recordings [13,14] bridge these two approaches. This convergence generates interest in investigating how several HOA capture points distributed throughout a sound scene can be mixed such that traditional aesthetic expectations from channel-based recording practice are not only met but augmented through the integration of emerging spatial recording and mixing technologies. In this article the results of an experiment exploring the perception of direct sound and ambience as a function of A-B spacings of HOA microphones are presented. A dedicated mixing technique and tool for A/B spaced HOA microphones is proposed, in which multiple HOA signals are processed prior to summation in B-format and decoding. Finally, results of a use case study using the proposed mixing technique are presented and discussed.

2. COMPARING BOTH APPROACHES

To situate the work, a brief comparison of channel-based spaced techniques with coincident microphone arrays is given. From an aesthetic as well as technical perspective, the following observations generally hold true.

2.1. Channel-based Spaced Techniques and Coincident Array Recordings

Spaced microphone setups are often optimized for specific channel-based reproduction systems. As an example, A-B or A-B-C techniques are typically conceived for traditional two- and three-channel stereo. A-B-C is typically applied to L-R-C channels in 5.1 surround sound. Channel-based spaced techniques can be modified to accommodate limited changes in the reproduction format. However, this process is not automatic and requires manual remixing to be done correctly. Despite this rigid relationship between capture and reproduction formats, one key advantage to using spaced microphone arrays is that they can create wide sound stages through full spectrum inter-channel decorrelation, as each microphone captures a different acoustic perspective.

In contrast, recordings from coincident Ambisonics microphone arrays can flexibly be decoded to any immersive reproduction setup, where they may deliver a unified (for some listeners, constrained) sense of envelopment. While the spatial resolution and precision of the sound image improves with the order of the Ambisonics microphone array, the disadvantage is that the b-format of a single coincident array does not contain time-based decorrelated signals from the sound scene. Despite the potentially compelling sense of envelopment, this leads to a vastly different sound image in the reproduction when compared with spaced techniques and established production aesthetics. Also, the spatial resolution of low frequencies, primarily a function of the diameter of the microphone array, is typically limited to 1st order directivity. It vanishes completely at very low frequencies creating a mono spatial impression. Similarly, above the spatial aliasing frequency, determined through the distance between the capsules, the spatial differentiation disintegrates and becomes diffuse. These are likely some of the reasons why single point HOA recordings do not compare favourably with traditional channel-based approaches for music recordings, as for instance reported in [15]. In order to extend spatial resolution across the whole frequency range, dual layer microphone array designs are being investigated [16], yet currently exist only as prototypes.

2.2. HOA Microphones in Spaced Arrays

To leverage the advantages of coincident and spaced techniques, A-B stereo set ups can integrate coincident systems at each side, such as M-S or X-Y, offering added adjustability and sonic characteristics. Dobson and Woszczuk [17] experimented with multiple FOA (1st order Ambisonics) tetrahedral receivers in spot and A-B spaced- apart configurations to integrate the attributes of coincident and spaced techniques. In practice, substituting traditional microphones in spaced arrays (ORTF-3D and ESMA-3D) with 2nd order Ambisonics microphones have been explored by Len Moskowitz from Core Sound [18] and recently with 3rd order arrays by Przemek Danowski [19]. The added value being that flexible directional polar patterns can be emulated for each coincident capture point. While this gives more flexibility in post-production, it does not fully leverage the advantage of HOA recordings as it collapses the rich spatial information of the diffuse field continuum, which according to [20] is an important part of spatial impression. Another recent development termed Bilateral Ambisonics [21–23] tries to leverage the spatially rich information of spaced HOA capture points. Here the two Ambisonics microphones are placed at ear-to-ear distance and adapted sets of HRTFs are applied to account for the interaural time delay, which yields a substantial improvement in perceived binaural reproduction quality even for lower Ambisonics orders. Bilateral Ambisonics can be considered as an extension of near-coincident techniques, as it orients the microphone spacing on ear-to-ear distance. Contributing to these developments, this article explores the influence of the A-B spacing in HOA capture of a sound scene, whether spacings with larger-than-ear distances are advantageous, and whether they can lead to desirable results. Further we demonstrate how A-B spaced HOA microphones can be mixed such that an intermediary b-format is preserved in order to be able to decode the resulting master file to various immersive reproduction formats.
2.3. Multiple HOA Perspectives in One Mix

Recording and mixing with more than one HOA microphone necessarily breaks the isotropic symmetry of a single Ambisonics capture. For the context of this article, spaced Ambisonics capture in a music concert setting typically has a defined and preferred orientation, we hence address perceptual and aesthetic concerns and are less interested in the correct capture and reconstruction of an isotropic acoustic field. The aim of this article is to bring multipoint HOA recording solutions to meet listening expectations, as those created through established recording and reproduction systems, and to leverage the advantages of a single master b-format for flexible yet effective surround and immersive sound production workflows.

Based on the brief review of channel based and Ambisonics recording techniques, a recordings and mixing strategy that combines the advantages of both was developed by using two HOA microphone arrays in spaced A-B positioning, capturing the scene from multiple auditory perspectives. The goals were:

1. Preserving the audible source separation and the depth of the sound stage of spaced techniques through decorrelated signals of time-based stereophony.
2. Preserving the spatially resolved diffuse sound field information and the aurally enveloping characteristics of HOA sound recordings.
3. Preserving the compatibility of the HOA b-format with respect to multiple surround reproduction systems by combining the different perspectives of HOA capture points in a single master b-format.

This single master b-format could then be decoded to either channel based immersive formats, as e.g. beds in Dolby Atmos, or be used as is in formats that accommodate scene-based approaches such as for instance MPEG-H. In both cases it combines the aesthetics of spaced microphone techniques inspired from channel-based traditions and the flexibility of Ambisonics for capturing and mixing.

2.4. HOA Arrays as A-B Microphone Systems

Initial explorations in the development of our mixing approach started with the capture of a concert performance by the Pronto Musica Orchestra in the Church of St. John The Evangelist, Montreal, December 3, 2019. A photo of the setup from a recording is shown in Fig. 1. To achieve the above goals, two 3rd order HOA microphone arrays (ZM1-D from Zylia) were placed in the vicinity of two conventional A-B omnidirectional microphone positions.

These established A-B positions capture a good balance between direct sound and reverberation. Monitoring the signal of each of the two HOA microphone arrays during rehearsal also suggested this placement provided adequate source separation. The 19 channel recordings of each microphone were then converted into two 16 channel 3rd order b-format files and synced in the DAW Reaper for initial explorations.

3. MIXING A-B HOA STREAMS

The problem of mixing two HOA streams can be considered analogous to the challenge of collapsing non-coincident stereo pairs into mono. This may lead to undesirable results due to the two stereo signals being similar and offset in time. Expanding this to two HOA streams adds a position shift in space to the offset in time, as any given source is represented in the two HOA streams with different incident angles (azimuth and elevation).

HOA A-B positions therefore cannot simply be mixed into one single master b-format because the time-based separation of sources across the A-B positions and sources’ differing angles of incidence, may cause a doubling of the source image in both time and space, and an extra cognitive task for perception. A direct layering approach creates conflicting cues instead of sharpening the sound image and increasing its transparency.

The conventional mixing approach for multiple HOA signals in one music scene is to emulate virtual microphones with varying on-axis directions and polar patterns. Blending two or several HOA capture positions as done in 6DoF HOA interpolations [13] returns a b-format but is concerned with the correct interpolation of the acoustic field and does not provide the mixing engineer with the necessary perceptual controls to adjust the recording in a typical production workflow.

It was hence necessary to shape the b-format through spatial filtering to separate the signals from A and B, much as bilateral Ambisonics [18] caters to the left and right ear separately. The difference being that our approach aims for larger microphone spacings in the A-B pairs. Preserving of a single b-format representing both positions for convenient decoding into various surround setups was essential. Inspiration for the spatial filtering of the b-format came from Zotter and Frank elaborating in [12] “what is most important when mixing, is the option to treat the gains of
different directions differently: it might be necessary to attenuate directions of ... disturbing content while boosting [other] directions ...” The MCFX plugin from Kronlachner [24] achieves this with a neutral directional re-mapping, which in the simplest version, is implemented in the ambix directional loudness plugin, where separate gain functions can be applied to two mutually exclusive spatial regions. Within a directional region the gain can be boosted or lowered compared to the region outside. The ambix directional loudness plugin outputs b-format to the audio FX chain. When rendered, the spatially filtered signal of microphone A defines the left side of the sound field, and microphone B the right side. Through this spatial filtering approach, conflicting cues from temporal and spatial offsets were minimized by attenuating the right hemisphere on the left and the left hemisphere on the right. More precisely, the half-sided attenuation was complemented with an equivalent boost in gain on the opposite side to maintain equal loudness in the overall signal. These two half-sided spatially offset representations of the sound field were then summed to form a single b-format. This b-format contained time-based stereo differentiations along the A-B axis with non-conflicting rich spatial cues and a structured diffuse field for either side of the A-B setup. A screen shot of the ambix directional loudness plugin with the discussed setting for position A is shown in Fig. 2.

3.1. Informal Evaluation

After implementing spatial filtering through directional loudness adjustments, A and B were auditioned in isolation, decoded over a speaker array as well as over headphones. This processing gave a noticeably clearer left and right differentiation of the sound image across the whole frequency range, not only in separation but also when mixed into one b-format. This result confirmed that:
1. The sound stage was stable and wide as usually achieved through spaced channel-based recordings.
2. The envelopment and the sense of immersion due to a rich diffuse field was strong, as expected from an Ambisonics reproduction.
3. Perceptually similar decoding results of the b-format over headphones as well as over loudspeaker arrays proved to be a basis for a generally applicable master b-format regardless of the reproduction system.

This preliminary and informal evaluation indicated that the proposed mixing technique of A-B spaced HOA microphones corrects imaging deficiencies of single coincident HOA arrays. It also offers another interesting parameter to optimize the mix as shown in Fig. 3.

The angular orientation of the spatially filtered half-spheres can be systematically changed along the horizontal axis (azimuth). Variations of this parameter led to salient changes in multiple perceptual dimensions. Firstly, one could emphasise direct sound in front, or emphasise more lateral reflections from the recording environment when oriented towards the side (±90°) in a continuous manner. Secondly, varying this parameter influenced the perceived width of the sound stage by focusing the listeners attention on laterally arriving reflections.

While slightly speculative as interpretation, changing the orientation of the spatially filtered half-spheres is reminiscent of the capacities of some species to intentionally perk or even orient their ears towards directions of interest in the acoustic field. As has been shown in [25] a still existing human rudimentary auriculomotor activity indicates the direction of auditory attention in humans and hints at why such a manipulation of the sound field can be meaningfully interpreted even if it is not a naturalistic representation of the acoustic field.

In Fig. 4, A-B spatial cues start to overlap as the half-spheres toe-in from left to right. A-B microphone spacings would likely have an influence on whether this overlap would result in conflicting cues. This angular adjustment of directional loudness, as well as the contrast in gain between the half-spheres, became the two most promising parameters to be tested more formally in an experiment.
The final parameters to be evaluated were:
- the degree (depth) of spatial filtering (in dB),
- the directional orientation of the halve spheres (in degrees), and,
- the A-B distance (in meters).

4. DATA ACQUISITION

SOUND RECORDING

In August 2020, the musical trio of Papadimitriou (double bass), AbouAfach (viola), Altobelli (accordion) was recorded in Pollack Hall at McGill University. Two systematic variations of A-B microphone spacings were made with two different distances (2 m and 4 m) between the microphone line up and the musicians. The 2 m distance between the ensemble and microphones was chosen for the testing stimuli, as the 4 m distance did not provide adequate representation of the instrument’s direct sound compared to the room’s reverberation. Figure 4 shows a photo of the setup with the 2 m distance between the musicians and the microphone line up. The stage size was reduced through placing two absorbers on the side and a diffuser in the back as shown in the aerial diagram in Fig. 5. After the recording, all files were synchronized and converted to b-format.

4.1. Technical Implementation

The listening test software used to evaluate the three parameters (degree of (depth of) spatial filtering, directional orientation, and A-B distance) was implemented in MaxMSP. The 3rd order b-format files of all the 9 microphone positions (144 channels) were sample-accurately streamed from the hard drive directly, and the signal chain sent selected A-B pairs to two VST objects each hosting an ambix directional loudness plugin. A simple graphical user interface allowed participants to switch from one A-B distance to another. The two directional loudness parameters (boost and attenuation of the half-spheres), as well as the principal directions of half-spheres, were controlled through a 2-D slider as depicted in Fig. 8. The vertical axis of this slider represented the boost/attenuation, and the horizontal axis the angular orientation. Both dimensions could also be independently controlled with two 1-D sliders arranged along the respective axes. The exposed VST parameters were mapped using ambix directional loudness to have a symmetric boost and attenuation function. The half-spheres could reach a maximum difference of ±20 dB. All directions from ±90° to 0° orientation of the half-spheres for both A-B positions could be adjusted continuously. Once the b-format files were loaded, a visual representation of the W channel from the center microphone was presented on a timeline, allowing users to set the music playback regions and loop them. All the parameters of interest (A-B distance, half-sphere attenuation, directional orientation) could be changed without interrupting the playback. Binauralization was implemented using the BinauralDecoder VST from IEM [26].

5. EXPERIMENT

The experiment consisted of 3 parts and was conducted with 10 subjects, 7 male and 3 female (within the age range of 25 to 40), all expert listeners and graduate students in McGill University’s Sound Recording Program. The stimulus was a musical loop that was kept consistent for all conditions throughout the whole experiment. It was chosen because all 3 instruments were equally present, the dynamic range and length were appropriate, and the loop did not become irritating after repeated listening. Restrictions on in-person testing due to COVID 19 limited the sample size, as the testing platform could not be adapted for at-home testing due to the high number of 3rd order HOA b-format files. Subjects used professional quality...
open-backed headphones and digital/analog conversion set up in a quiet dedicated listening room.

5.1. Part 1. Evaluating the Impact of A-B Spacings

In the first part, subjects were asked to rate the sound of binaural renderings of the A-B mixes for 4 different A-B spacings named A, B, C, and D, where A was the anchor (rendering only a single central microphone) with only the orientation of the two half spheres showing an effect. This was simply applying spatial loudness filtering on one HOA microphone. B was the smallest spacing of 50 cm, C with a spacing of 100 cm, and D the largest spacing of 150 cm between the microphones. For all A-B microphone spacings, the sound came from the same music loop of 26.3 seconds duration and the parameters of the A-B spatial filtering were set to an attenuation of $-20$ dB, and both half-spheres were panned to left and right at $90^\circ$.

Subjects rated aspects of the sound scene they perceived in the continuing music loop. The ratings were expressed on a Likert scale [27] based on 3 considerations: direct sound of instruments only, ambient sound of room acoustics only, and both of these considerations combined. Through iterative pilot testing the statements that were presented to the subjects were refined, aiming for enough specificity in order to direct each subject’s focus to the relevant sonic qualities of the recording (i.e., perception of direct and ambient sound). Thus, our experimental questions reflect general appreciation of these considerations, as experienced during mixing. Kelly et al. highlight the interdependency between a listener’s impression of the sound source, room acoustics, and themselves in relation to the sensation of presence in the immersive reproduction of acoustic music [28]. While evaluations of source and room impressions may be treated independently, the listener’s impression of how they integrate with one another is also of interest. With this in mind, the following test questions were formulated:

1. Regarding the direct sound from the sources: The presentation of the instruments is optimal.
2. Regarding the ambient sound (disregard the sources): The presentation of the room acoustics is optimal.
3. Regarding Direct Sound and Ambient sound: The balance of instruments and room acoustics is optimal.

The available responses on the Likert scale were: (0) strongly disagree, (1) disagree, (2) somewhat disagree, (3) neither agree nor disagree, (4) somewhat agree, (5) agree, (6) strongly agree.

Subjects gave their Likert scale response via a horizontal slider as shown in Fig. 6. By touching the slider for microphone spacing, A, they would hear the audio related to this specific microphone pair, by switching to slider B, they would seamlessly continue with B from the current playback position within the loop. This allowed for a direct comparison between all four A-B microphone spacings without interruptions in the flow of the music. For each rating, subjects were also asked to provide a rationale, which they could type into a textbox next to the rating slider. This part of the experiment took on average 15 to 20 minutes.

5.2. Part 2. Exploring a New Mixing Tool

Part 2 of the study focused on evaluating our A-B HOA mixing tool in a balancing process following a research-creation approach. Participants were given the opportunity to interactively explore the effect of the A-B spatial signal filtering. Two balancing parameters could be controlled either integrally with a 2-dimensional X/Y controller, or separately using vertical and horizontal sliders, respectively. The vertical axis, which controlled the dB boost and attenuation, was described as CONTRAST. The horizontal axis, which controlled the orientation of the attenuated half spheres was described as WIDTH, with wide on the left and narrow on the right. These descriptions of the dimensions of the control interface were inspired by their perceptual effects. The control interface is shown in Fig. 7 as displayed on a computer screen.

Our initial plan was to provide no information on the link between the control interface and the parameter space...
because the subjects were not informed about the underlying transformations of the signal. During pilot testing, however, some indications were required to help the subjects form a mental model of how they interacted with the sound. Color coding was explained to the subjects (Fig. 8, on the left), which was to indicate that a narrow-wide differentiation (red - blue) would only emerge when a sufficient difference in contrast was reached (moving out of the green area). Subjects were encouraged to explore the parameter space freely and to ask questions until they thought they had a solid grasp of the sonic balancing effect and how to control it through the interface. This exploratory part of the experiment lasted typically 5 minutes and was based on the microphone spacing D (largest spacing).

5.3. Part 3. Mixing/Balancing with the New Tool

In the third part of the study, subjects were asked to use the control interface to find the optimal setting with respect to the three considerations they had already rated earlier: direct sound, ambient sound, and the balance of both. They were told to expect to hear stimuli A, B, C, and D again. Randomization was considered during pilot testing, but it was determined that the presence of the anchor having no effect on the contrast dimension, was too confusing if it happened to be right after the exploration session. Also, while pilot subjects reported that the mixing task was very stimulating in terms of the variations they heard, they found it more difficult to conceptualize it with regards to what they knew from mixing/balancing in stereo, particularly since the anchor A only showed variations in width by narrowing the perspective to the front of a single HOA capture point. To support subjects’ sense of aural control in using the tool, the stimuli were presented systematically with decreasing microphone spacings D, C, B, and A, which is the ordering subjects were already familiar with from the rating exercise. With this presentation order, the difficulty of the task increased as spacings narrowed. For each microphone spacing, subjects had to find the optimal position for the resulting mix in the control interface with regards to direct sound, ambient sound, and the balance of both. The mixing/balancing objective was highlighted under the control interface. Subjects were encouraged to comment verbally on what they heard while they were optimizing the sound using the control interface, and why they made certain mixing choices. Live comments were recorded, though subjects reporting was not structured, therefore inconsistent across individuals.

6. RESULTS AND ANALYSIS

The following data were collected during the experiment. In part 1: The Likert scale responses for all four A-B spacings A, B, C, and D regarding direct sound, ambient sound, and the balance of both. For each Likert response, a justification was collected in the form of qualitative written statements. In part 3: trajectory data was recorded while subjects navigated the parameter space to optimize for specific mixing targets (direct, ambient, balance of both), including the endpoint of the trajectory representing the subject’s final choice. Live statements were again recorded during the mixing phase.

6.1. Results, Part 1 — Likert Response Data

The Likert scale responses are depicted in the boxplot in Fig. 8. The microphone pairs (A, B, C, and D) are arranged on the x axis from left to right, and the Likert response with regards to the optimal quality of the presented mix (0, strongly disagree, to 6 strongly agree) on the y axis. The attributes are color coded: blue for direct sound, brown for ambient and red for both.

Kruskal-Wallis tests were conducted to examine the differences on Likert rating according to microphone spacing per attribute. A significant difference (α = 0.05) was found for Ambience (chi-squared = 9.5571, df = 3, p = 0.0249) and Direct (chi-squared = 9.3413, df = 3, p = 0.0251) optimization. No significant difference was noted for Both optimization (chi-squared = 6.8061, df = 3, p = 0.108).

6.2. Results, Part 1 — Qualitative Data

Listed below are a selection of quotes given by
participants during the Likert rating phase. Considering the multi-dimensional nature of the tasks, the questions posed to the subjects were left open-ended by design, in an effort to uncover their motivations through qualitative assessment. As such, a wide array of responses was expected as participants used their own internal isomorphic mapping strategies to explain their assessments of the attributes in question. Therefore, while the excerpts below should be considered conjectural and not statements of fact, they give valuable insight into participant’s awareness and motivation.

### 6.3. Discussion, Part 1

While the assessment of the quality of the direct sound remains mostly the same for the microphone pairs A, B, and with a shift upwards in C, it drops noticeably for the widest spacing D. The rating for the perceived quality of the ambient sound raises constantly with the increase in spacing between the microphone pairs and is rated best for the widest distance D. The balance of both, direct and ambient sound is rated best for the microphone spacing C and drops strongly for the widest spacing D. Except for the attribute ambient sound, microphone spacing C received the best Likert responses, suggesting that it is close to an optimal A-B spacing in this context. The poor rating for the attributes direct sound and direct-ambient balance in D strongly suggests that in this spacing the coherent spatial sound image disintegrated.

Quantitative analysis confirmed that microphone spacing did in fact influence the participants rating for direct sound and ambient sound. When looking at the qualitative responses, remarkably diverse justifications were reported. Some subjects referred to the overall quality of the sound scene, and some seemed to base their response on the perceived quality of the presence of specific instruments in the trio. With the double bass being in the middle of the trio, its variation in timbre as a function of the microphone spacing was brought up several times in comments, as larger A-B spacings between the microphones also increased the perceived distance to the double bass. Similarly, as the A-B microphone spacing widened, participants also reported a change in the presence of the accordion and the viola. Some less favorable Likert rating justifications (rating 0 meaning strongly disagree to rating 6 strongly agree) for the ambient sound component for the anchor directly responded to our initial assumption about the need to differentiate the low frequency correlation: (rating 1) “room is audible but very little.” (rating 3) “dull reverb not much realism compared to the other samples”; but were contradicting more favorable ratings (rating 4) also given for the assessment of the ambient sound for the anchor: “reverb has more character … it sounds like there is an actual space”. Most explicit reference to our perceptual assumptions was found in the statement (rating 3) for the spacing C “reverb sounds central in the stereo image whereas the instruments are wide panned. reverb itself sounds balanced in frequency range but not realistic.”

While the exploratory nature with the deliberately open questions and the resulting diverse answers did not lend

---

**Table 1** Spacing A.

| Likert Rating | Examples of participant comments for optimization of Direct Sound |
|---------------|---------------------------------------------------------------|
| 1             | “Too much high end in the accordion”                         |
| 2             | “High frequencies and very low frequencies are a bit exaggerated” |
| 4             | “Balanced sound, fair amount of room in mix”                 |
| 5             | “Most linear”                                                |
| 6             | “Clearest bass picking representation string transients best represented.” |

**Table 2** Spacing B.

| Likert Rating | Examples of participant comments for optimization of Ambient Sound |
|---------------|---------------------------------------------------------------|
| 1             | “Reverb tail present however not a sense of a clear external room” |
| 2             | “Room acoustics audible especially at the tail may still be too little” |
| 4             | “Still nice like C but I like C better. I’m not sure why.” |

**Table 3** Spacing C.

| Likert Rating | Examples of participant comments for optimization of both Direct and Ambient Sound |
|---------------|----------------------------------------------------------------------------------|
| 3             | “Reverb sound sounds central in the stereo image whereas the instruments are wide panned. reverb itself sounds balanced in frequency range but not realistic” |
| 4             | “I would like things to sound a bit closer”                                      |
| 5             | “Also an appropriate balance between direct and room more of the room is present in this one though — concert hall experience…?” |

**Table 4** Spacing D.

| Likert Rating | Examples of participant comments for optimization of Ambient Sound |
|---------------|---------------------------------------------------------------|
| 0             | “Too ambient/far back reverb is too high in level and frequency. Lows are missing?” |
| 4             | “Long tail affects the bass too much”                           |
| 6             | “Most sense of space here the accordion doesn’t sound like it’s sitting in my ear canal” |
itself to an in-depth analysis of the collected qualitative statements, the justifications for the Likert ratings show that what expert listeners report depends on what you ask them to listen for. This is not to say that their responses are without grounding but are simply a function of how they interpret the task they are being asked to perform.

6.4. Results, Part 3 — Quantitative Optimization Data

Figure 9 shows the response data from the optimization task for all subjects, and correlates to the testing interface presented in Fig. 7. Density plots are given for each parameter (azimuth of orientation for each hemisphere, and level boost/attenuation in dB) along the X and Y axis to show how subjects responses cluster within the parameter space depending on which aspect of the sound scene (direct, ambient, or both) they were instructed to optimize. Spacing type is plotted using shapes.

To further illustrate this point, the data set was transformed such that all responses for the ‘both’ optimization were centered at the origin of a 2D cartesian graph. Each corresponding ‘direct’ and ‘ambient’ data point underwent a proportional transformation. The result shows two separate point clouds, indicating that they are pulled from distinct distributions.

![Figure 9](image)

6.5. Results, Part 3 — Qualitative Data

With regards to the qualitative statements, some participants commented a lot on what they were doing while others said very little. Some participants, who were very verbal about how they mixed, did not comment on certain stimuli. Below is a selection of comments that stand for general tendencies observed and reported. They indicate the level of aural awareness with which subjects are able to control the tool to achieve specified results.

When optimizing direct sound “Width one third across is optimal. Too high in the contrast and the instruments take a back seat to the room sound”

When optimizing for larger AB microphone spacings: “What I notice in this trial is that they are not completely in sync.”

With respect to the sound image, one participant found that “The width seems to control the position of the outside instruments.”

The top left corner representing maximum contrast and maximum width was described as “If the contrast is maxed out and the width maxed out... the reverb sound almost becomes its own instrument, I can really notice it the more I tend to gravitate to the top corner, that’s when I hear it clear as day, I can hear the decay of the reverb sound it does not sound the most realistic but here the tone of the reverb tail is the best.” and similarly, “over in this corner I have this wonderful spacious separation, the bass in the middle and the accordion here and the viola right there.”

For the narrow setting in contrast one participant noticed: “It is somewhat crowded here”

When trying to apply the effect to the anchor condition one subject noticed: “I like where it sits but I’m not sure moving around affects the ambient sound.”

6.6. Discussion Part 3

When evaluating the trajectories of how subjects explored the parameter space, different individual styles of optimization were found. Some explored the parameter space in circular motion to narrow down towards the optimum. Others varied the setting systematically in horizontal and vertical lines sometimes jumping from one extreme of the parameter space to the other to simulate a wet/dry type of listening comparison.

The endpoints for all three conditions (direct, ambient, and the balance of both) are plotted for all mic spacings in Fig. 9. Some clustering emerges when looking at the azimuth distribution. Most of the endpoints for the ambient sound optimization condition (in brown) tends to cluster with a broad peak around the 75-degrees of angular orientation termed WIDTH. The endpoints for the direct sound condition (blue) are broadly distributed across the whole orientation scale. While the optimization for both (red) seems to represent a compromise between the direct
and the ambient, peaking around 62 degrees. The positioning of the endpoints on the vertical axis (termed CONTRAST in the GUI interface) shows a unimodal distribution for the direct sound (blue) and two bimodal distributions for the ambient (brown) and balance of both conditions (red). The partially multimodal distribution along the two-axes suggests complex interaction between the mixing target, and the two dimensions that were termed Width and Contrast. It’s likely that the A-B distance plays a role in this, however when analyzing subsets of the data, split by A-B spacings, the sample size of 10 participants turned out to be too small to show a significant interaction.

This clustering described above was examined with further exploratory data analysis. We conclude that all subjects had their own preferred area of mixing and while these areas were distributed across the whole control interface of mixing parameters, subjects still made their choices in a consistent manner relative to these areas of their own preference. The optimization for both mixing targets (red) seems to tend toward the middle between both. Because of this, the individual offset was removed to create a within-subject paradigm by centering all endpoints for the optimization of both direct and ambient sound. In effect, normalizing the data of all subjects treating this as the reference condition. The result is shown in Fig. 10. Because of the multimodal nature of the distributions in the horizontal and vertical axis, a two-dimensional Kolmogorov-Smirnov (KS) test was used, which makes no assumption about the underlying distribution. With ‘both’ optimization data centered (see Fig. 11), results of the KS test show $D = 0.5699$ and $p < 0.000$, indicating that both distributions are significantly different.

This confirmed that from a within-subject perspective, the ‘ambience’ optimization and ‘direct sound’ optimization endpoints were drawn from significantly different distributions. This implies that subjects went to specific areas of the parameter space consistently to optimize the sound-scene for either direct or ambient sound.

The positions of the end points of the mixing trajectory show primarily that the effect created by this mixing approach is a practical and systematic way of integrating two HOA signals. Almost all endpoints had a positive offset along the boost/attenuation axis which backed our initial observation that this approach diminishes conflicting cues that would be created by simply overlaying two HOA signals. There also seems to be a systematic positioning with respect to subjects’ individually chosen areas of preferences in the mixing interface. Ambience is optimized in the mix by orienting the spatial filters towards the extreme left and right angles, whereas direct sound is optimized by turning the spatially boosted areas towards the stage, in frontal angles. Unlike with the results from part 1 of the experiment a systematic connection with the microphone spacings could not be seen due to a) the small sample size of subjects and b) the open nature of the task.

These comments all confirm observations the authors had made themselves when exploring the mixing interface and are a testimony to the multifaceted potential that this interface affords the mixing engineer. Some of the observations related to the coherence of the image as a
function of the A-B distance are related to traditional channel-based spaced microphone techniques; others such as the perceived spaciousness and clarity of the room impression are related to the novel mixing technique employing HOA source signals.

7. APPLYING THE RESULTS IN A DEMO MIX

The mixing technique was applied in recording and postprocessing the performance of the trio where the A-B HOA mixing approach with two different A-B settings was used (one narrow and closer to the trio and one wider-apart and more distant). The narrow A-B setting had an orientation of the spatial filtering that generated a wider image and conversely the wide A-B setting at greater distance profited from an orientation of the spatial filtering that narrowed the spatial image. This image control applied to the two different acoustic perspectives offered a fluid shaping of complementary aural perspectives onto one integrated scene.

Additionally, close HOA microphones were placed next to the performers. These closer microphones were also spatially filtered using directional loudness and added to the mix in order to further improve the balance between images of direct and reverberant sound, adjusted dynamically in time during mixing. The binaural rendering of this mix can be auditioned here: https://soundcloud.com/nordestjazz/esperanza-live-pollack-hall

8. CONCLUSION

The authors presented a recording and mixing/balancing technique for two HOA microphones in A-B spaced-apart positioning for musical recordings in a concert setting, which they tested with a trio of musicians in a performance hall. This mixing technique and a corresponding tool were developed to leverage the aesthetics of traditional channel-based techniques for HOA multipoint recordings. With two broadly spatially filtered HOA streams, the overall richness of diffuse information of the acoustic field and spatial resolution of source positioning could be preserved and offers a new way of mixing immersive sound using spaced HOA microphone techniques. This new approach combines the sense of continuous spatial envelopment from HOA with the enhanced impression of depth and width of space due to low frequency decorrelation between the A-B positioned receivers.

Our results showed that increasing the A-B spacing between microphones correlated with the increased rating of the perception of room ambience on a Likert scale. Complementary to the current work on bilateral Ambisonics, the evaluation of our proposed recording and mixing technique equally underlines the importance of the A-B spacing as an essential ingredient for the listener to judge and agree that the presentation of ambient sound is optimal. However, differently to bilateral Ambisonics, this evaluation also demonstrates that A-B spacings larger than average ear-to-ear distances (producing larger differences of perspective) are desirable for the perception of ambience which is generally related to the sense of envelopment and immersion as is traditionally the case for spaced techniques in channel-based music recordings. The reported results were derived from listening tests conducted with headphones, and it is expected that in tests over loudspeakers A-B spacings wider than ear-to-distance will also be preferred.

A practice-led part of the evaluation has shown that with only two parameters a perceptually interesting space can be systematically and usefully controlled through this new mixing approach. The analysis of the trajectory mixing end-points showed that despite individually different mixing styles, participants used the proposed control interface in a consistent and systematic way to identify and adjust different attributes in the sound scene. The results were applied in a demo mix of a musical trio, in which direct sound was balanced against ambience as a function of HOA arrays’ distance to the source and spacing of the A-B pair.

Most importantly the proposed mixing technique for multiple HOA streams can be integrated into a single HOA b-format in order to create a master file that combines several acoustic perspectives into one. While this mixing approach was evaluated more formally as a binaural rendering, the master b-format is a convenient basis that can be rendered to various 3D immersive sound setups or distribution formats like Dolby Atmos or MPEG-H. Moreover, several A-B pairs of HOA arrays could be integrated into the single master b-format, each pair presenting a scene, or a section from a different perspective. This work translates the established know-how of traditional legacy techniques into new recording technologies. While it may open more questions than it answers we hope to point the way towards further research on how best to translate multiple HOA capture points into compelling musical, and experiential results.

REFERENCES

[1] R. King, Recording Orchestra and Other Classical Music Ensembles, 1 online resource (272 pages) vols. (Routledge, New York, 2016). [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1435471
[2] C. Ceoen, “Comparative stereophonic listening tests,” AES Convention 2ce (1972). [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=1818
[3] D. M. Huber and R. E. Runstein, Modern Recording Techniques, 9th ed. (Routledge, Taylor & Francis Group, New York, London, 2018).

141
[4] A. Fukada, “A challenge in multichannel music recording,” AES 19th Int. Conf., Schloss Elmau, Germany (2001).

[5] H. Wittek and G. Theile, “Development and application of a stereophonic multichannel recording technique for 3D Audio and VR,” AES 143rd Convention, New York, USA (2017).

[6] K. Hamasaki and K. Hiyama, “Reproducing spatial impression with multichannel audio,” AES 24th Int. Conf. Multichannel Audio, Banff, Canada (2003).

[7] Schoeps Mikrofone, ‘IRT Cross,’ Available at: http://www.schoepts.de/en/products/irt-cross-set, 2017, online; Accessed on 15/07/21.

[8] K. Hamasaki and W. Van Baelen, “Natural sound recording of an orchestra with three-dimensional sound,” AES 138th Convention, Warsaw, Poland (2015).

[9] W. Howie, R. King and D. Martin, “A three-dimensional orchestral music recording technique, optimized for 22.2 multichannel sound,” AES 141st Convention, Los Angeles, USA (2016).

[10] H. Lee, “Multichannel 3D microphone arrays: A review,” J. Audio Eng. Soc., 69, 22 (2021).

[11] A. M. Gerzon, “The design of precisely coincident microphone arrays for stereo and surround sound,” AES 50th Convention, London, UK (1975).

[12] F. Zotter and M. Frank, Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality (Springer International Publishing, 2019).

[13] J. G. Tylka and E. Y. Choueiri, “Comparison of techniques for binaural navigation of higher-order Ambisonic sound fields,” AES 139th Convention, New York, USA (2015).

[14] E. Patricio, A. Ruminski, A. Kuklasinski, L. Januszskiewicz and T. Zernicki, “Toward six degrees of freedom audio recording and playback using multiple ambisonics sound fields,” AES 146th Convention, Dublin, Ireland (2019).

[15] W. Howie, R. King, D. Martin and F. Grond, “Subjective evaluation of orchestral music recording techniques for three-dimensional audio,” AES 142nd Convention, Berlin, Germany (2017).

[16] P. Grandjean, C. Girin, P.-A. Gauthier and A. Berry, Spherical microphone array design and validation: double-layer versus simple-layer array (2019).

[17] A. Dobson and W. Woszczyk, “Tetrahedral microphones: An effective A-B main system,” AES 147th Convention, New York, USA (2019).

[18] “Core Sound OctoMic™ 2nd-order Ambisonic Microphone,” Core Sound. https://www.core-sound.com/products/octomic (accessed 19 Aug. 2021).

[19] P. Danowski, INEXISTSENTS (Personal correspondence).

[20] D. Kombloc, C. Guastavino and P. Depalle, “Perceptual thresholds for non-ideal diffuse field reverberation,” J. Acoust. Soc. Am., 140, 3908–3916 (2016), doi: 10.1121/1.4967523.

[21] Z. Ben-Hur, D. L. Alon, R. Mehra and B. Raufay, “Binaural reproduction using bilateral Ambisonics,” AES Int. Conf. Audio for Virtual and Augmented Reality, Online (2020).

[22] Z. Ben-Hur, D. L. Alon, R. Mehra and B. Raufay, “Binaural reproduction based on bilateral Ambisonics and ear-aligned HRTFs,” IEEE ACM Trans. Audio Speech Lang. Process., 29, 901–913 (2021), doi: 10.1109/TASLP.2021.3055038.

[23] I. Engel, D. Goodman and L. Picinali, “Improving binaural rendering with bilateral Ambisonics and MagLS,” DAGA (2021).

[24] M. Kronlachner, “Plug-in suite for mastering the production and playback in surround sound and Ambisonics,” AES Students Design Competition 136th AES Convention, Berlin, Germany (2014).

[25] D. J. Strauss, F. I. Corona-Strauss, A. Schroer, P. Flotho, R. Hannemann and S. A. Hackley, “Vestigial auriculomotor activity indicates the direction of auditory attention in humans,” eLife, 9, e54536 (2020).

[26] C. Schörkhuber, M. Zauschhirm and R. Höldrich, “Binaural rendering of Ambisonic signals via magnitude least squares,” Fortschr. Akust., DAGA (2018).

[27] R. Likert, “A technique for the measurement of attitudes,” Arch. Psychol., 22(140), 55 (1932).

[28] J. Kelly, W. Woszczyk and R. King, “Are you there?: A literature review of presence for immersive music reproduction,” AES 149th Convention, Online (2020).

Florian Grond Ph.D. is an interaction designer working as a research associate in the Sound Recording Department of the Schulich School of Music at McGill University. As a media artist with a focus on sound art and sonic ethnographies, he has been recording and mixing in Ambisonics for many years. From 2015–2017 he conducted postdoctoral research funded through the B5 research-creation fellowship from FRQSC, with the focus on the creative explorations of immersive sound recordings with microphone arrays. He has exhibited as a media artist globally and has received funding for his artistic creations from the Canada Council for the Arts, the Conseil des Arts et des Lettres du Québec and the Conseil des Arts de Montréal. He is a CIRMMT collaborator (Centre for Interdisciplinary Research in Music, Media, and Technology). His research and design interests are multimodal participatory design in the context of disability, the arts, and assistive technology.

Jack Kelly is Ph.D. candidate at the Schulich School of Music, McGill University. His thesis research centers on the influence of spatial room impulse response convolution technologies (channel-based and HOA arrays) on the sensation of physical presence in immersive music production. He received a BFA in Electro-acoustic Studies at Concordia University in 2012, and an M.Mus. in Sound Recording (Tonmeister degree) at McGill University in 2016. Jack has been an active recording engineer in Montréal, Canada since 2013, working with numerous artists under labels such as Analekta, Atma Classique, Centrediscs, Deutsche Grammophon, Leaf Music, and Pentatone. He is a CIRMMT collaborator (Centre for Interdisciplinary Research in Music, Media, and Technology), and AES student member.

Wieslaw Woszczyk holds the James McGill Professor Research Chair position and a full professorship at McGill University in Montreal, Canada, and is the founding director of the Graduate Program in Sound Recording (1978) and founding director of the CIRMMT Centre for Interdisciplinary Research in Music Media and Technology, established at McGill University in 2001. An AES member since 1976, Woszczyk is a Fellow of the Audio Engineering Society (1996) and the former Chair of its Technical Council (1996–2005). He served as AES Governor (1991–1993 and 2008–2010) and President (2006–2007). Woszczyk received the Board of Governors Awards in 1991 and 2015, and a Group Citation Award in 2001 for “pioneering the technology enabling collaborative multichannel performance over the broadband Internet.”