RECORDING TECHNIQUES AND THEIR EFFECT ON SOUND QUALITY AT OFF-CENTER LISTENING POSITIONS IN 5.0 SURROUND ENVIRONMENTS

Nils Peters[1], **Jonas Braasch**[2], **Stephen McAdams**[3]
[1] Centre for Interdisciplinary Research in Music Media and Technology, McGill University, nils@music.mcgill.ca
[2] School of Architecture, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180, braasj@rpi.edu
[3] Schulich School of Music, McGill University, 555 Sherbrooke St. W., Montreal, QC H3A 1E3, smc@music.mcgill.ca

ABSTRACT

Assessments of listener preferences for different multichannel recording techniques typically focus on the sweet spot, the spatial area where the listener maintains optimal perception of the reproduced sound field. The purpose of this study is to explore how multichannel recording techniques affect the sound quality at off-center (non-sweet spot) listening positions in medium-sized rooms. Listening impressions of two musical excerpts created by three different multichannel recording techniques for multiple off-center positions are compared with the impression at the sweet spot in two different listening room environments. The choice of a recording technique significantly affects the sound quality at off-center positions relative to the sweet spot, and this finding depends on the type of listening environment. In the studio grade listening room environment featuring a standard loudspeaker configuration, the two tested spaced microphone techniques were rated better at off-center positions compared to the coincident Ambisonics technique. For the less controlled room environment, the interaction between recording technique and musical excerpt played a significant role in listener preference.

SOMMAIRE

L’ évaluation par des auditeurs de préférences entre différentes techniques d’enregistrement multi-canal se focalisent typiquement sur la zone idéale *(sweet spot)*, la région de l’espace où l’auditeur maintient une perception idéale du champ sonore reproduit. L’objectif de cette étude est de comprendre comment les techniques d’enregistrement multi-canal affectent la qualité sonore à des endroits hors de la zone idéale dans des salles de taille moyenne. Dans deux salles différentes, les impressions à l’écoute de deux extraits de musique créés par trois techniques d’enregistrement multi-canal à plusieurs endroits hors de la zone idéale sont comparées avec l’impression obtenue dans la zone idéale. Le choix d’une technique d’enregistrement affecte significativement la qualité sonore dans des zones non-idéales par rapport la zone idéale. Ce résultat dépend du type d’environnement d’écoute. Dans un studio d’écoute avec une configuration d’enceintes standard, les deux techniques utilisant des microphones espacés créent une moindre perception de dégradation sonore dans les zones non-idéales comparées à la technique Ambisonics. Dans un environnement moins contrôlé, l’interaction entre la technique d’enregistrement et l’extrait musical joue un rôle significatif dans la préférence des auditeurs.

1 INTRODUCTION

A concert hall is designed to enhance natural sound sources and produce a plurality of listening positions with perceptually good sound images of those sources [1]. In spatial audio reproduction, however, a best listening point is usually implied and limits quality surround-sound reproduction to small audiences. Although several types of microphone techniques exist for surround-sound recordings, and all techniques aim to give listeners the impression of *being there*, they favor the centralized listener and yield a degraded sound image for the others. Understanding the delivery of an improved sound image across the audience is critical. Off-center locations may be more representative of typical listening situations, and research on non-ideal listening positions “may provide significant information regarding the general performance of the [audio] system" [2].

In the past, listening tests have assessed the differences among surround microphone techniques primarily at the central listening position (e.g., [3, 4, 5, 6]) and excluded off-center positions. Also in a closely related field (the evaluation of sound reproduction environments), the effect of the listening position was primarily studied for localization errors (e.g., [7, 8]), neglecting all other perceptual dimensions. This paper investigates off-center listening, specifically, the degradation in sound quality as a function of the recording technique used for capturing a recording. Recording techniques generally differ in their strategy for creating phantom sources and for reducing undesired inter-channel correlation. Strategies may involve spacing of microphones and/or increasing the microphones’ directivities. Griesinger [9] suggests that decorrelation of the loudspeaker feeds increases the listening area, which can be achieved, for instance, by spacing the microphones. To our knowledge, no formal listening tests have investigated Griesinger’s hypothesis.

In the following section, we define the terms Center and Off-center Listening Position and identify acoustical properties in the spatial relationship of an off-center listener to the loudspeaker setup that cause a variety of perceptual artifacts. Our methodology and the experimental conditions are explained in Section 2. Listening experiments in two different listening rooms are analyzed in Sections 3 and 4. We conclude in Section 5 with a final discussion.

1.1 Center and Off-Center Listening Positions

Audio recording and reproduction techniques usually refer to a reference listening point, called the *sweet spot*, which draws from perceptual or geometric concepts. The perceptual concepts suggest a vague consensus that the sweet spot is the point in space where a listener is fully capable of hearing the intended audio recording, the spatial bubble of head positions where the listener maintains the desired perception. For scientific use, such a definition is imprecise, because the intended sound design is unknown to most listeners. The sweet spot has also been described as the point in space where the listener is equidistant from all speakers (or at least maximally distant from them if they do not form a circle).

To avoid the ambiguous meaning of the sweet spot, we will use the term *Central Listening Position* (CLP) to describe the reference listening point where all loudspeakers are equidistant and equally calibrated in Sound Pressure Level (SPL). An *Off-Center Listening Position* (OCP) refers to all other positions within the loudspeaker array. Our definition is compliant with ITU recommendation BS.1116-1 [10], which places the reference listening point in the center of the surround loudspeaker setup (Fig. 1). This recommendation also points to the least recommended listening positions.

1.2 Loudspeakers - Listener Relation

In spatial sound reproduction, speaker feeds from multiple directions create signals at the listener's ears, uniquely for each listening position. We will briefly introduce the underlying physical relationships.

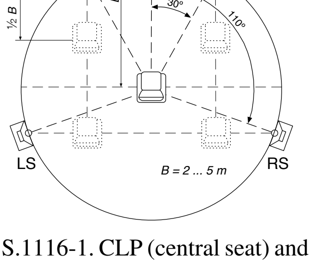


FIGURE 1 – ITU BS.1116-1. CLP (central seat) and worst case OCPs (dotted). The recommended listening area is within 0.7 m of the CLP.

Unbalanced Sound Pressure Level (SPL). A closer loudspeaker will produce a higher SPL than a loudspeaker that is farther away. For a conventional loudspeaker, the attenuation of the direct sound is ca. 6 dB SPL per doubled distance for the direct sound component. Thus, the SPL changes very quickly near a loudspeaker, which makes this effect most prominent at off-center positions in small speaker setups. Loudspeaker level differences at off-center positions also depend on room characteristics and on loudspeaker directivity due to the contribution of reflected sound energy. For uncorrelated sounds that contribute to envelopment the attenuation is closer to 3 dB SPL per doubled distance, which causes variations in off-center sound degradation across audio content [11].

Time-of-Arrival Differences (ToA). Loudspeaker feeds will arrive at an off-center position with different temporal delays due to distance differences. The maximal temporal delay is calculated from the distance of the closest and farthest loudspeakers and the speed of sound. The further away the off-center position is from the center, the greater the ToA differences.

Direction of Arriving Wavefronts. At the central listening position, a wavefront emitted by the right speaker (*R* in Fig. 1) arrives from a direction of 30°, whereas for a listener at the upper right dashed seat, the same wavefront impinges from the front.

1.3 Perceptual Artifacts

Localization. Depending on all three physical circumstances, the sound image might shift or even collapse toward the direction of the most prominent speaker feed. The Precedence Effect may explain this perception (see, e.g., [12] for a review). Although the Precedence Effect is primarily investigated for indoor localization (since it is related to localization processes in the presence of early reflections), it is also important in multichannel sound reproduction. An important distinction between these two scenarios is that a real sound source has one direct wavefront, from which directional information is decoded via summing localization and multiple (to-be-inhibited) early reflections. In multichannel audio, the location of a virtual sound source is perceived by the superposition of wavefronts emitted from several loudspeakers. At off-center positions, the auditory system may fuse and inhibit the wrong set of wavefronts. Each loudspeaker can also cause individual reflections in the listening room that will be superimposed upon the early reflections of the room in which the recording was made. Localization of reproduced sound over loudspeakers in listening rooms was specifically investigated by Olive and Toole and later by Bech. Olive and Toole [13] measured the energy of room reflections that is necessary to shift the image of the reproduced sound under three different room acoustic conditions. For early reflections (< 30 ms) this image-shift threshold was similar across all three conditions, but for reflections later than 30 ms, the reverberation time of the room had a strong influence, with the thresholds for the delayed reflection rising sharply with each move to a more reflective listening space. Bech [14] found that the amount of reflected spectral energy above 2 kHz contributes to audibility, and a strong first-order floor reflection can significantly affect spatial aspects of the reproduced sound field.

Image Stability. The perceived location of the reproduced sound source may change with pitch, loudness, or timbre. It may also change as a function of listener position, head rotation, or other normal movements. If these effects are small, the image will be stable [15]. Image Stability is one of three factors in the definition of Overall Spatial Quality by the IEC [16]. Other related spatial descriptors are Spatial Clarity, Readability, Locatedness, and Image Focus. For virtual sound sources, Lund [17] derived a localization-consistency score from the related descriptors Robustness, Diffusion, and Certainty of Angle.

Spatial Impression comprises Apparent Source Width (ASW) and Listening Envelopment (LEV). ASW describes the spatial extent of a sound source influenced by early lateral room reflections (up to 80 ms). ASW was found to be primarily generated by frequencies above 1 kHz and is correlated with the Inter-Aural Cross-correlation Coefficient (IACC) calculated from the early energy [18]. The authors of [19] found that for many, but not all sounds, the ASW is closely related to Image Stability. LEV describes the fullness of sound images around the listener due to late lateral reflections. LEV depends on the front/back energy ratio, the direction of the speakers' wavefronts, and the spectral content primarily below 1 kHz [20]. At off-center positions, the LEV can become unstable and compromises the envelopment illusion.

Timbral Effects. The relative importance of timbre and spatial aspects in audio reproduction was examined by Rumsey et al., [21]. Timbral fidelity has a weight of ca. 70% on the overall sound quality, whereas spatial factors accounted for ca. 30% of the variance. It was found that naive listeners valued surround spatial fidelity over frontal spatial fidelity, which was found to be the inverse for expert listeners [22]. Especially relevant for surround reproduction, Olive et al. [23] showed that listeners are less sensitive to the timbral effects of loudspeakers in multichannel setups compared to one-channel sound reproduction.

At off-center positions, the misalignment of the loudspeaker wavefronts (see ToA differences) can also lead to audible comb filtering [24]. The absolute threshold for an audible timbre change rises with increasing delays, whereas complex reflection patterns (responsible for ASW and LEV) and a binaural decoloration mechanism [25] can mask timbre changes. Rakerd [26] hypothesized that the auditory system may combine binaural and spectral cues for localization, so that a timbre change causes a localization change of an auditory event.

2 GENERAL METHODS

In two listening experiments, the reproduced sound field at different off-center listening positions is compared with the sound field a listener perceives at the center. We chose two sets of previously produced 5.0 multichannel content (EXC). Each 5.0 multichannel content was simultaneously recorded with three different multichannel microphone techniques (RT). All content was recorded, mixed, and produced by experts who used them in their own experiments on recording technique evaluation (see [4, 3]). To study off-center sound degradation as an effect of listening position we reproduced their content in two different rooms through 5.0 multichannel loudspeaker systems, and captured binaural stimuli at multiple listening positions (POS). Each binaural stimulus was captured at 48 kHz and had a duration of about 7 s. In total, for each tested listening position, six binaural stimuli were captured (2 excerpts × 3 recording techniques). In a soundproof booth, these binaural stimuli were compared by trained listeners wearing diffuse-field equalized headphones.

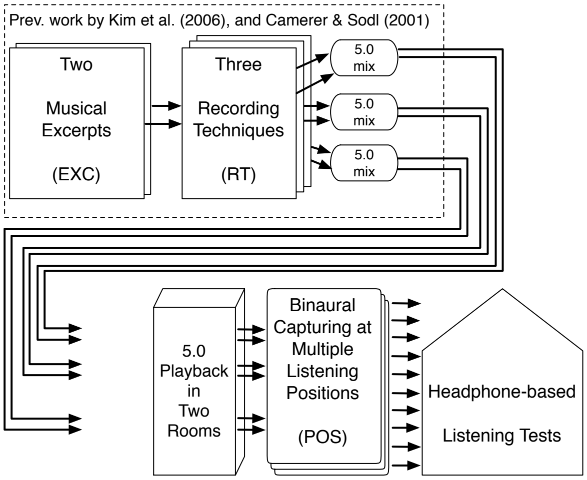


FIGURE 2 – General experimental method.

To study off-center sound degradation as an effect of the reproduction environment, our sets of binaural stimuli were captured in two very different reproduction environments. Both environments are actively used for multichannel sound reproduction for larger audiences. The first reproduction environment (Telus Studio) is a medium size room with a standard 5.0 full-range loudspeaker setup to meet the ITU requirements for multichannel loudspeaker setups for listening rooms. The second reproduction environment (Tanna Schulich Hall) is a small multi-purpose concert venue, a non-ideal, ecologically valid sound reproduction environment. The reproduction environments differ in terms of the room acoustic condition, loudspeaker type, and loudspeaker arrangement (see Fig. 3 for comparison of the reverberation time). Practical reasons led us to create two *most-different* scenarios for our study of perceived off-center sound degradation in 5.0

surround sound environments as a function of the recording technique. A detailed explanation of each reproduction environment is provided in Sections 3 and 4. This general method is depicted in Fig. 2. We discuss the challenges faced when using real-world sound reproduction environments for this type of auditory research in Section 5.

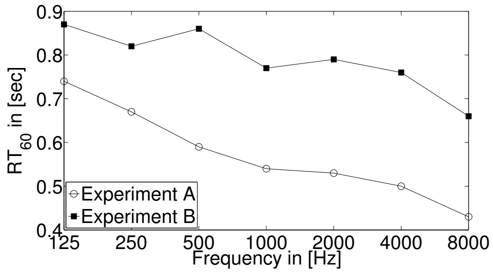


FIGURE 3 – Reverberation time RT_{60} in Telus Studio (Exp. A) and Tanna Schulich Hall (Exp. B).

2.1 Musical Excerpts — EXC

Each musical excerpt was a 5.0 multichannel recording created from the perspective of a concert audience facing a stage with the instrument sounds arriving from the front and ambient sounds and room response from the sides and behind. The excerpts were:

EXC 1: J.S. Bach "Variation 13", Goldberg Variationen for solo piano (BWV 988).

EXC 2: W.A. Mozart "Maurische Trauermusik" in c-minor for symphony orchestra (KV 477).

Detailed information regarding the recording and mixing procedures for these two excerpts are given by Kim [4] for excerpt 1 and by Camerer [3] for excerpt 2. An overview of these recording techniques follows.

2.2 Recording Techniques — RT

Each musical excerpt was recorded with three prominent multichannel recording techniques. These techniques differ in their strategy for reducing correlation across the channels. We provide a short overview of these techniques including drawings of the recording setups in Fig 4. Detailed descriptions on all three recording techniques can be found in [27].

Coincident Microphone Technique — Ambisonics. Ambisonics extends Blumlein's coincident recording technique. An omnidirectional microphone is added to the pair of perpendicularly oriented figure-eight units. The vertical component of the sound field is captured by adding a third figure-eight unit perpendicular to the others. All microphone capsules are meant to be at exactly the same spatial location. Thus, amplitude differences between the microphones are created. For both excerpts a Soundfield MKV microphone was used. The microphone signals are encoded into the so-called B-format. To reproduce the sound field, the B-format signals are decoded with respect to a specific loudspeaker setup. Although Ambisonics is theoretically best reproduced on regular loudspeaker layouts, algorithms exist to create an optimized decoder for an irregular loudspeaker setup. For instance, the (irregular) 5.0 loudspeaker setup is supported since Gerzon's *Vienna decoder* [28]. In both excerpts the Soundfield SP451 processor [29] was used for 5.0 decoding.

Spaced Omnis Microphone Technique. The omnidirectional microphones are widely spaced, primarily creating inter-channel time differences. To account for the different source widths in EXC 1 and EXC 2, slightly different variations of this technique were used.

Polyhymnia Pentagon (used for EXC 1): This technique uses five widely spaced omnidirectional microphones and is often described as a multichannel version of the Decca Tree. The microphones are arranged in a large circle and their positions correspond to the azimuthal angles of the 5.0 loudspeakers.

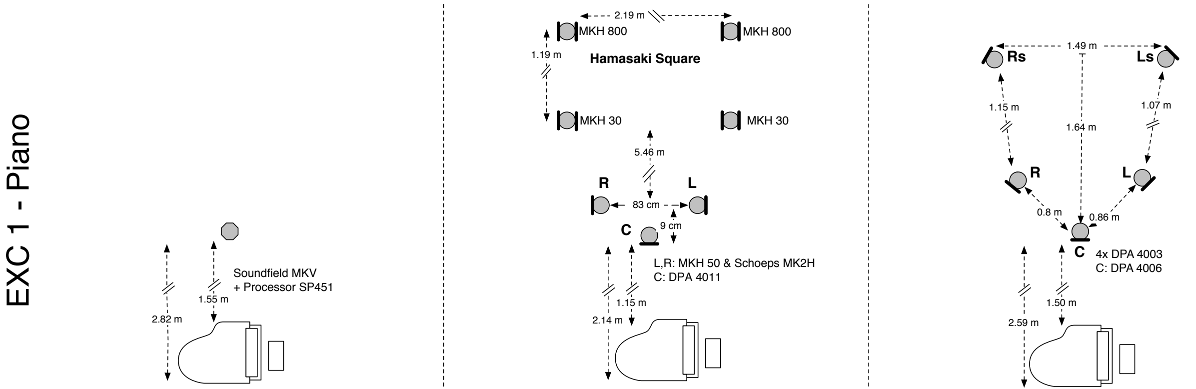
Decca Tree + Hamasaki-Square (used for EXC 2): The Decca Tree consists of three omnidirectional microphones arranged in a triangle. The center microphone is placed 0.7 to 1 m forward, whereas the right and left capsules are spaced at a distance ranging from 1.4 to 2 m. In the recording of EXC 2, two additional lateral microphones were used to capture the entire width of the orchestra. Furthermore, the sound field for the two 5.0 surround channels was recorded with a Hamasaki Square.

Spaced Cardioid Microphone Technique. The Optimized Cardioid Triangle (OCT) reduces channel crosstalk by creating both inter-channel amplitude and inter-channel time differences. Two outer hyper-cardioid microphones face $\pm 90^\circ$ sideways from the center cardioid microphone, which is usually placed 8 cm forward. For both excerpts, the OCT array was extended with a Hamasaki Square to feed the two 5.0 surround channels.

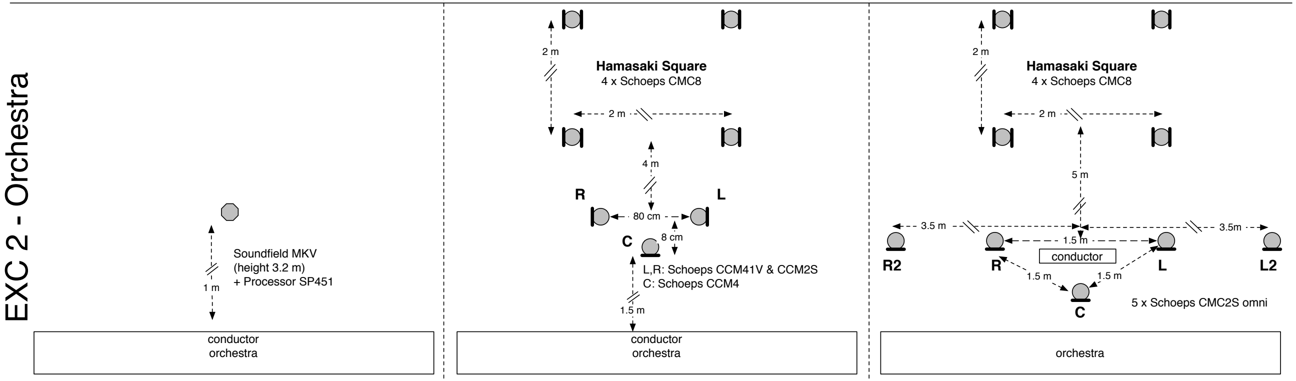
2.3 Procedure and Apparatus

The listeners were asked to *Rate the degradation in sound quality of sound B relative to sound A.* Sound A represented one of the six central listening position (reference) stimuli, whereas sound B could be: a) one of the off-center stimuli of the same musical excerpt and recording technique as sound A ; b) the hidden reference (the same central listening position stimulus as sound A) ; or c) the hidden anchor, which is a monaural stimulus captured at a very off-center position, where the left audio channel was presented to both ears. The purpose of the hidden reference and anchor was to set best- and worse-case references for the rating scale and to validate listeners' reliability.

Listeners are typically asked to rate the absolute difference (or similarity) between stimuli. Absolute diffe-



(a) Ambisonics (b) Optimized Cardioid Triangle (OCT) (c) Polyhymnia Pentagon



(d) Ambisonics (e) Optimized Cardioid Triangle (OCT) (f) Decca Tree + Hamasaki Square

FIGURE 4 – Multichannel microphone array setups used in the recordings of the musical excerpts. (a)-(c) for EXC 1 adapted from [4] and (d)-(f) for EXC 2 adapted from [3].

rence/similarity does not necessarily indicate preference or quality. We chose to ask listeners to rate *sound quality degradation*. Rating perceived sound degradation explicitly asks the listener about quality (better, worse), and is therefore more meaningful for describing preference in one listening position over another.

The pairwise comparison trials were presented in random order. A graphical user interface was employed and the ratings were made with a computer mouse on a slider with a continuous scale from 0 (total degradation) to 100 (no degradation). The scale was also marked by the following descriptors: *very strong degradation - strong degradation - moderate degradation - slight degradation - very slight degradation*. This scale corresponds roughly to an analogical-categorical scale, found in psychophysical research to increase response reliability [30]. Within the presented pair, listeners could switch between sounds A and B at will and could listen as many times as necessary.

The experiments consisted of a training phase (phase 1), a familiarization phase (phase 2), and the experimental phase (phase 3). In phase 1, five trials with musical excerpts that were different from those presented in phase 3 were presented for interface training. Listeners were informed that these ratings would not be recorded. In phase 2, a representative collection of 30 binaural stimuli were used to familiarize them with the musical material. They were told that phase 2 would give them the range of variation in sound degradation so they could subsequently use the full scale for their judgments in the experimental phase, which lasted about 60 min. To increase the reliability of the data, each stimulus pair appeared twice. We used Sennheiser HD 600 headphones at a normal listening level (70 dB(A) for the recording at the central listening position). Besides diffuse-field headphone equalization, no additional filtering was applied. The listeners were told to face the frontal direction and to keep their heads steady. Breaks were allowed.

2.4 Discussion of Experimental Method

The ideal test design for this experiment would make participants listen and relocate from seat to seat in the actual listening room. Unfortunately, such an *in situ* design has various drawbacks: it would be almost impossible to allow for double-blind, comparative, and repeatable evaluations in a reasonable time-frame ; for the participants it would also be extremely challenging to memorize the perceived sound quality while physically changing listening positions. Our method allowed listeners to switch between two binaural stimuli in real time, and thus had the advantage that listening positions could be compared quickly and repeatedly in a double-blind test while minimizing cognitive challenges. Furthermore, by isolating and presenting the binaural stimuli via headphones, the potential for sound quality biases based on visual cues on the part of the listeners was also circumvented.

Our method relies on the assumption that the presentation of the binaural stimuli can evoke all perceptually important elements of the captured sound field as they would have been perceived by a subject directly. Toole [31] discussed the potential and drawbacks of using a binaural reproduction system in listening experiments. In particular, the absence of head movements in static binaural recordings and non-individual HRTF cues may cause localization errors mainly in the median plane and in the region of the cone-of-confusion. Therefore we acknowledge that not all perceptual dimensions may be perfectly reproduced by the binaural system. However, because the binaural reproduction conditions were equal for all stimuli in the listening experiment, we think that the effect generates a constant bias for all stimuli, and thus, the relative differences are preserved. Despite these constraints, several related studies have successfully used similar methods. In [32], for one test listeners rated loudspeakers *in situ* in different rooms. In a second test, listeners were asked to rate via headphones binaural recordings of these loudspeakers captured in each room. Although some differences in the ratings between the two experiments occurred, the pattern of results was essentially the same.

As an alternative to static binaural recordings, a binaural room-scanning system (BRS) could have been used [33]. BRS allows head movements through head tracking in the binaural reproduction system, reduces localization errors and increases out-of-head localization. However, those two advantages diminish when room reflections are included in the capturing process [34], as is the case in this presented study.

3 EXPERIMENT A — TELUS STUDIO

The Telus Studio at the Banff Centre for the Arts has a floor-space of ca. 140 m^2 and a volume of ca. 800 m^3 and is used for lectures, film presentations, and as a recording room for medium-large ensembles. For the reverberation times (Fig. 3) and SNR, the Telus Studio marginally meets the recommendation by the ITU [10] as well by the IEC [16] for multichannel loudspeaker setups for listening rooms. The Schroeder frequency, below which the modal density distribution dominates, is about 53 Hz. For the 5.0 loudspeaker setup, five Dynaudio BM15A loudspeakers were placed at a height of 1.2 m on an arc with a radius of 4.2 m. To capture the binaural stimuli, omnidirectional probe microphones (DPA 4060) were placed at the entrance of the first author's ear canals. To avoid uncontrolled head movements during recordings, a neck-brace was used. The ten tested positions were chosen as depicted in Fig. 5 and included the best- and the two left-sided worst-case listening positions as shown previously in Fig. 1. The listening positions cover only the left side of the listening area because one expects that a quasi-symmetrical sound field occurs due to the symmetrical shape of the room and the loudspeaker setup. In total, 72 pairwise comparisons were prepared for the listening experiment (2 excerpts × 3 recording techniques × 12 positions). A monaural recording at position 10 was used as the hidden anchor. The SPL varied between 73.5 and 79 dB(A) depending on position and was 75 dB(A) at the central listening position.

Ten trained listeners (8 male, 2 female) with normal hearing were tested. They were sound recording students with technical ear training and work experience between 1 and 23 years (Median=9). Their age varied between 24 and 44 (Median=30).

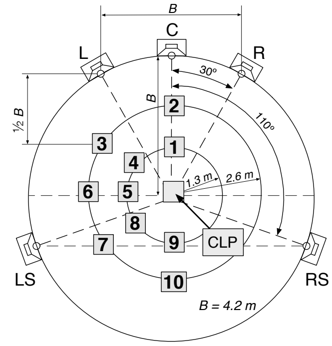


FIGURE 5 – Listening positions in Experiment A.

3.1 Results

The hidden reference and the hidden anchor were used to post-screen the behavioral data for potential outliers. There was a strong agreement across listeners for the rating of the hidden reference (M=95.6, SD=5.1) and the hidden anchor (M=11.9, SD=11.7). After excluding the ratings for the hidden reference and the hidden anchor, an EXC(2)×RT(3)×POS(10) repeated-measures analysis of variance (ANOVA) was performed. Besides the EXC main effect and the EXC×RT interaction (Table 1), all effects are significant ($p < .001$), The effect size measure η_p^2 indicates that the recording technique (RT) and the listening positions (POS) have by far the largest effects.

TABLE 1 – ANOVA results for Experiment A.

Effect	df	F	p	η_p^2	η_p^2-Rank
EXC	1, 9	0.7	.794	.01	7
RT	2, 18	34.6	$< \mathbf{.001}$	.80	1
POS	9, 81	26.9	$< \mathbf{.001}$	.75	2
EXC×RT	2, 18	3.5	.054	.28	6
EXC×POS	9, 81	7.1	$< \mathbf{.001}$	.44	3
RT×POS*	18, 162	5.4	$< \mathbf{.001}$	.37	5
EXC×RT×POS*	18, 162	5.2	$< \mathbf{.001}$	.37	4

* Greenhouse-Geisser correction for violation of sphericity

To determine statistical differences across recording techniques pairwise comparisons (Bonferroni-Holm adjus-

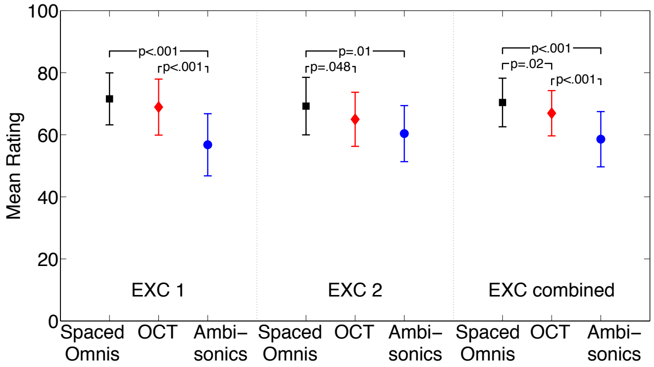


FIGURE 6 – Experiment A: Mean ratings and 95% confidence interval as a function of recording technique and excerpt: Brackets show significant differences between two recording techniques evaluated with Bonferroni-Holm-adjusted pairwise comparisons.

ted) were performed. The results are depicted in Fig. 6. As indicated by the ANOVA results (EXC main effect not significant), the group means and the 95% confidence intervals of all recording techniques have a similar trend across musical excerpts with Spaced Omnis rated best and Ambisonics rated worst. When combining the behavioral data for both excerpts, the pairwise comparisons indicate significant differences ($p < .05$) between all three recording techniques (right section in Fig. 6).

Figure 7 visualizes the sound-quality mean ratings across the listening area. A spatial cubic interpolation was used to estimate the sound degradation between the tested listening positions. Starting at the central listening position, a radially diminishing sound quality can be observed for all three recording techniques. The slope of this radial degradation however varies across recording techniques and is steepest for Ambisonics. An opposite trend can be observed for the standard deviation of the rating, which tends to increase the more off-center a listening position is. Therefore, one can say that the agreement among listeners is higher the better the sound quality is and the closer the listening position is to the center.

The so-called sweet area, the listening area around the central listening position that was rated equally well, was estimated by a Tukey-Kramer HSD post-hoc test (see white lines in Fig. 7). The largest sweet area for EXC 1 was created by the Spaced Omnis recording technique and for EXC 2 by the Optimized Cardioid Triangle. For both excerpts, Ambisonics produced the smallest sweet area. For the Optimized Cardioid Triangle and Ambisonics, the listening area of EXC 2 (orchestra) seems to be slightly wider than for EXC 1 (solo piano). Interestingly, the sweet area shows different shapes across recording techniques and musical excerpts and is never front/back symmetric.

The largest difference between the different recording techniques can be found at listening positions 5 and 10 for EXC 1 and at positions 1 and 2 for EXC 2.

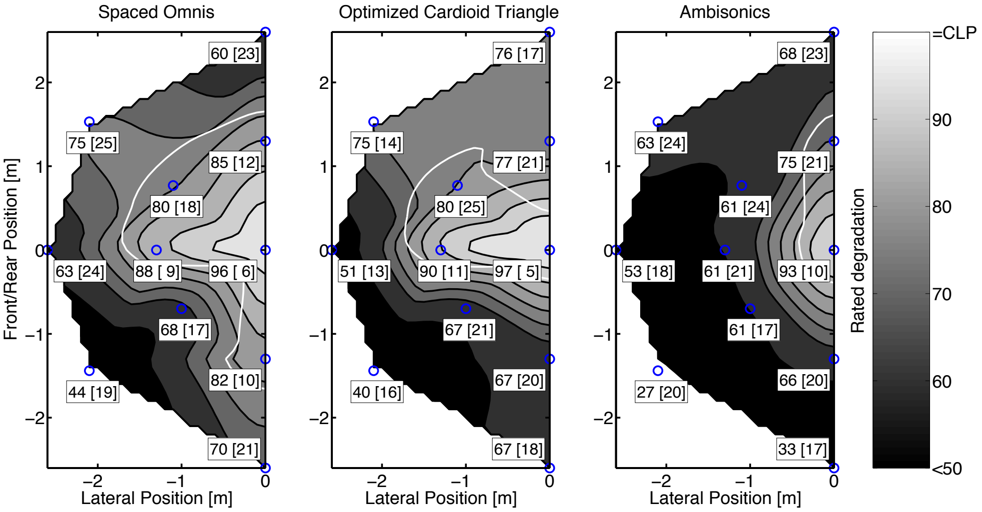
3.2 Discussion

The results of the ANOVA suggest that recording technique (RT) followed by the listening position (POS) are the two largest effects in the behavioral data. The effect of POS is expected and confirms the consensus among listeners and audio engineers concerning the limited ideal listening area of surround-sound reproduction systems. It is surprising that the largest ANOVA effect size was found for the RT main effect. This finding suggests that choosing the right multichannel recording technique during the sound recording process is an essential parameter to reduce off-center sound quality degradation. The pairwise comparisons across recording techniques (Fig. 6) show that in both excerpts the Spaced Omnis microphone technique significantly outperformed its contenders OCT and Ambisonics most of the time considering the ratings of all 10 listening positions. Nevertheless, with respect to the sweet area, the OCT recording technique created a larger sweet area than the Spaced Omnis technique for EXC 2.

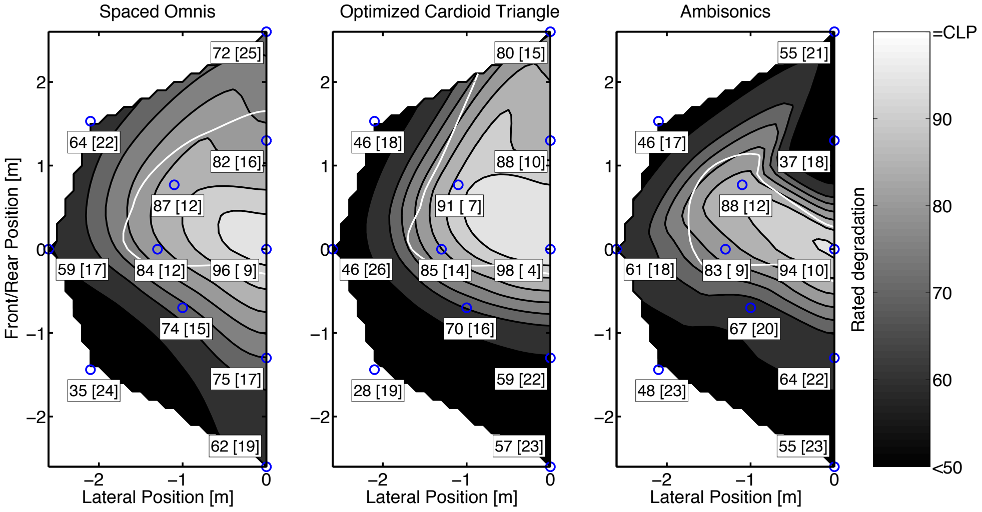
The third-largest ANOVA effect was found for the EXC×POS interaction effect, which can be observed by studying the sound degradation maps in Fig. 7, e.g., comparing the ratings at listening position 1 between both excerpts. The ratings for listening positions 3 and 7 are particularly interesting, because both positions are classified in ITU-R BS.1116 [10] as worst-case positions. In all six EXC×RT conditions, position 7 always received the lowest ratings of all tested positions (M=37), making position 7 the least desired seat. In comparison, in all but the Ambisonics recording of the orchestra, position 3 was rated 65% better than position 7 (M=61).

4 EXPERIMENT B — TANNA SCHULICH HALL

Tanna Schulich Hall (McGill University) has a floor space of ca. 240 m^2 with 188 seats and a volume of ca. 1400 m^3. It is used for jazz and chamber music performances, as a lecture hall, and for electroacoustic and mixed music concerts with multi-loudspeaker arrays. It is known for its intimacy and short reverberation time (Fig. 3). The Schroeder frequency is about 47 Hz. The hall's 5-channel surround loudspeaker system was used and calibrated for optimal sound quality at the central listening position (Kling & Freitag CA 1515 for the front and CA 1001 for the surround). Due to the rectangular shape of the room, the positions of the loudspeakers differ from ITU-R BS.1116-1: instead of ±110°, the rear speakers are placed at ±150° with an arc of ca. 8.2 m, measured from the central listening position. Because of this displacement, the expected effect of the surround loudspeaker (to enhance listener envelopment) may be reduced. Further, the center speaker is noticeably elevated to



(a) EXC 1 - piano



(b) EXC 2 - orchestra

FIGURE 7 – Sound degradation maps for Experiment A. Referring to Fig. 5, the listening positions are marked with circles. At position (0,0) the rating of the hidden reference. Each position shows the mean rating and [standard deviation]. The size of the sweet area (estimated with Tukey-Kramer HSD) is shown by white contours.

account for an optional projection screen. Due to the raked seats in the hall, the listening perspective relative to the elevated speakers varies. This entire layout we consider as a non-ideal, yet ecologically valid real-world setup. A B&K dummy head was placed at 13 positions (see Fig. 8). In concordance with experiment A, the sound pressure at the central listening position was calibrated to 75 dB(A) and varied between 74-77 dB(A) depending on the listening position. The independent variables for the experiment yield 78 conditions (2 excerpts × 3 recording techniques × 13 positions). The hidden anchor was a monaural recording of the position marked as “anchor” in Fig. 8.

Nineteen trained listeners (16 male, 3 female) with normal hearing participated in the experiment, including all of the listeners from experiment A. Ages ranged from 23 to 44 (Median=27) and work experience within the sound recording field varied from 1 to 23 years (Median=7).

4.1 Results

Similar to Experiment A, there was a strong agreement across listeners how to rate the hidden reference (M=95.5, SD=3.9) but a less strong agreement for the hidden anchor (M=19.6, SD=16.1). After removing the ratings for the hidden reference and anchor, a EXC(2)×RT(3)×POS(11) repeated-measures ANOVA was performed on the sound degradation ratings. Results are shown in Table 2. All main effects (EXC, RT, POS) and all interactions were found to be significant ($p < .05$). The POS main effect has the largest η_p^2 effect size followed by the EXC×RT interaction and the RT main effect.

TABLE 2 – ANOVA results for Experiment B.

Effect	df	F	p	η_p^2	η_p^2-Rank
EXC	1, 18	10.5	**.004**	.37	4
RT	2, 36	15.2	**< .001**	.46	3
POS*	10, 180	69.0	**< .001**	.79	1
EXC × RT	2, 36	28.2	**< .001**	.61	2
EXC × POS*	10, 180	5.6	**< .001**	.23	5
RT × POS*	20, 360	4.2	**< .001**	.19	7
EXC×RT×POS*	20, 360	5.1	**< .001**	.22	6

* Greenhouse-Geisser correction for violation of sphericity

The mean ratings and 95% confidence interval as a function of the recording technique and the musical excerpt are shown in Fig. 9. This figure displays also the results of a Bonferroni-Holm-adjusted pairwise comparison to evaluate the recording techniques against one another. For both excerpts, the Spaced Omnis technique was rated significantly higher than the OCT technique. The EXC×RT interaction revealed in the ANOVA can be attributed to the Ambisonics technique: While in both excerpts there is a similar relation of Spaced Omnis to OCT, for excerpt 1 (solo piano), Spaced Omnis and OCT were both rated better than Ambisonics, but in excerpt 2 (orchestra), the Ambisonics technique received the higher scores. When combining the ratings from both excerpts, the Spaced Omnis technique is significantly better rated than OCT and Ambisonics ($p < .001$) while OCT and Ambisonics are statistically similar. The recording technique with the lowest mean rating (Ambisonics for EXC 1 and OCT for EXC 2) also has the largest confidence intervals. In contrast, the data for the Spaced Omnis recording technique have the smallest confidence interval of all three recording techniques for both excerpts, which means that listeners were more in agreement than for the other two techniques.

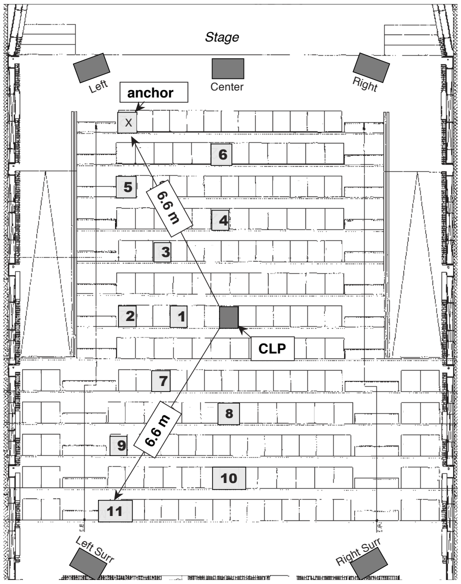


FIGURE 8 – Listening positions in Tanna Schulich Hall.

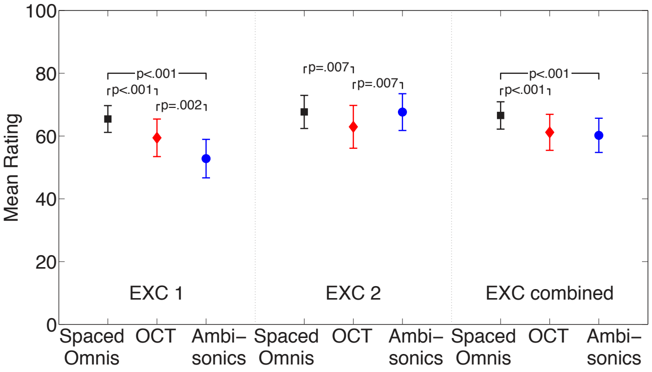
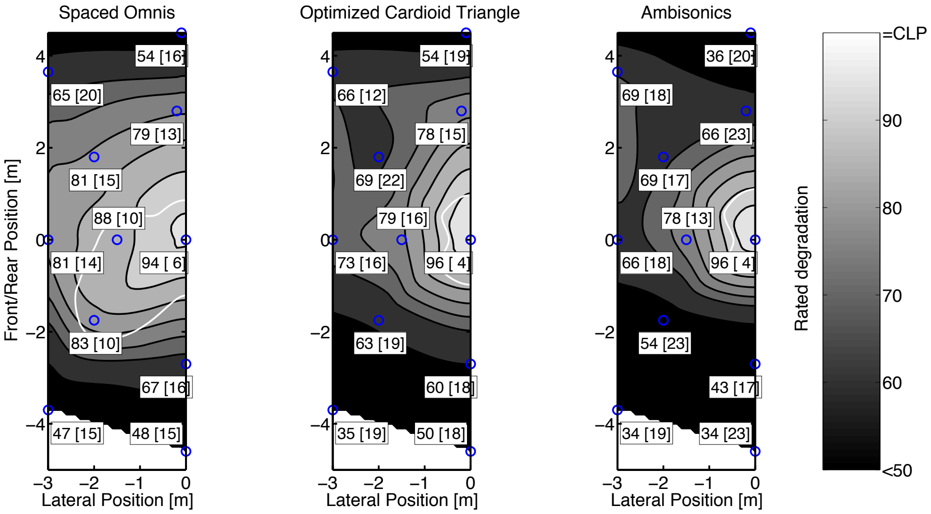


FIGURE 9 – Experiment B: Mean ratings and 95% confidence interval as a function of the recording technique and excerpt. Brackets show significant differences between two recording techniques indicated by pairwise comparisons (Bonferroni-Holm adjusted).

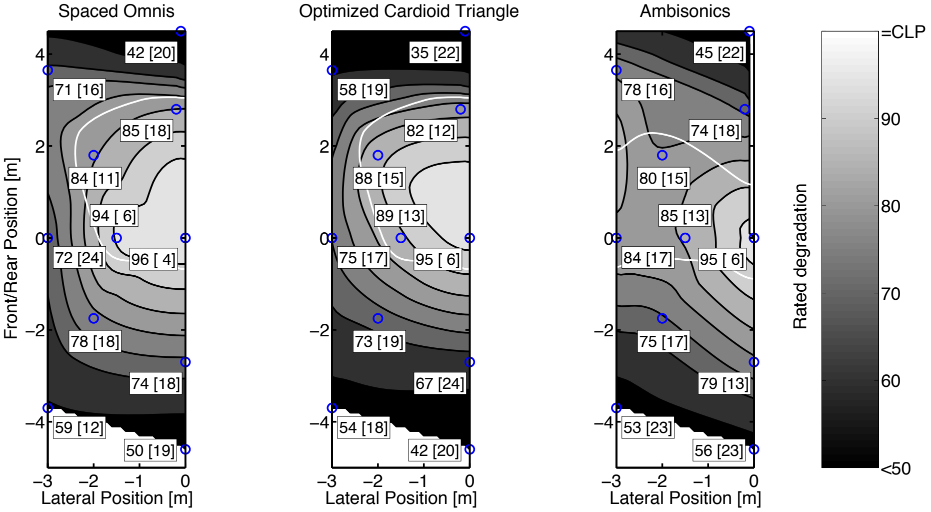
The sound quality maps based on the average ratings of the listening area are visualized in Fig. 10. The white line indicates the sweet area, the listening area around the central listening position that was rated equally well, identified with a Tukey-Kramer HSD post-hoc test. For EXC 1 (piano, Figure 10(a)), the biggest reference listening area was created by the Spaced Omnis technique. Our post-hoc analysis suggested a similarly sized reference listening area for the other two recording techniques. The largest differences for EXC 1 between recording techniques can be found at listening positions 7 and 8. For EXC 2 (orchestra, Figure 10(b)) the contours are less uniform and show less pronounced differences across recording techniques. Generally for all three recording techniques, the reference listening area around the central listening position is bigger in EXC 2 than in EXC 1. Further, the plots show equivalent sound quality degradation for Spaced Omnis and the OCT. Ambisonics was rated in EXC 2 much better than in EXC 1, in particular for position 8. Interestingly, at positions 2 and 5, the Ambisonics recording produced the best off-center sound quality across all three techniques.

4.2 Comparison with Experiment A

Because the experimental design did not involve a direct comparison of listening positions between Experiment A and Experiment B, we cannot compare the behavioral data of these two experiments directly, but we can compare the relative performance of each recording technique with each musical excerpt. This relative comparison is visualized in Fig. 11. The mean ratings already shown in Fig. 6 and 9 were ranked and show that the Spaced Omnis microphone technique performed best overall in three out of the four visualized condi-



(a) EXC 1 - piano



(b) EXC 2 - orchestra

FIGURE 10 – Sound degradation maps for Experiment B.: Referring to Figure 8 the listening positions are marked with circles. Each position shows the the mean rating and standard deviation. The size of the sweet area (estimated via Tukey-Kramer HSD) is shown by white contours.

tions. In two of these three conditions, the difference between the first and the second best recording technique (OCT) is also significant. However, in Experiment B for EXC 2, the Ambisonics recording received the highest mean rating, but when compared to Spaced Omnis (second highest rated technique), Ambisonics is not significantly better. In all other three conditions the Ambisonics technique was always ranked third. The OCT recording technique was rated second best in three conditions and ranked third in one condition.

By comparing the sound quality maps from Experiments A and B (Figs. 7 and 10), one sees that in both reproduction environments EXC 2 is perceived to have a wider area with good sound quality than EXC 1 regardless of recording technique. Further, one sees a similar radial shape of the sound degradation in the two reproduction environments for all the recording techniques, except for the Ambisonics recording for EXC 2. Here, in both reproduction environments the shape of the sweet area is more lateral than radial. Further, the large ANOVA effect size of listening position (POS) in both reproduction environments shows that the listening position has the most influence on the perceived sound degradation. Contrary to Experiment A, the EXC×RT interaction in Experiment B, clearly visible in Fig. 9, is significant and has the second largest effect size, which suggests that in this non-ideal listening environment, the off-center sound quality depends on the combination of recording technique and actual content.

Also in contrast to Experiment A, the mean ratings of sound quality are higher for EXC 2 than for EXC 1 (compare Fig. 6 with Fig. 9), meaning that listeners were less critical in their judgements for EXC 2. This might indicate that in

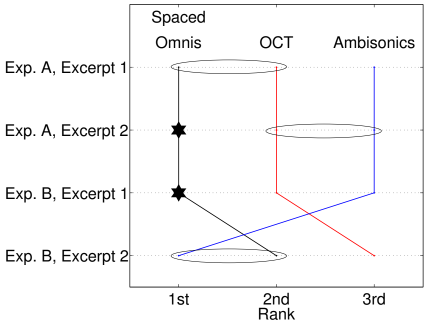


FIGURE 11 – Ranked overview of the recording technique mean ratings. A star indicates significant differences between first ranked and second ranked recording technique, an ellipse indicates similarly rated recording technique (Bonferroni-Holm pairwise comparison).

non-ideal (more reverberant) reproduction environments for complex musical material, such as in the case of an orchestra reproduced in Tanna Schulich Hall (Experiment B, EXC 2), a variety of perceptual artifacts are at play in the listeners' evaluations. Another explanation could be that the task (analyze a complex acoustic scene such as a symphonic excerpt in a reverberant room) is more demanding than it is for a less complex acoustic scene, e.g., a solo piano excerpt. Listeners may be attending more to listening envelopment than localization. A diffused sound image is by definition less localizable and unstable, perhaps democratizing sweet spot as a function of listening position. In view of Rumsey's scene-based approach to spatial quality evaluation [15], listeners may pay more attention to ensemble-related sound quality aspects, e.g., the apparent source width or brilliance of the orchestra, rather than to aspects related to the individual sound sources, such as the location of an instrument within the orchestra. Other studies have also found that sound quality preference judgments depend on the audio material, (e.g., [1, 33]) or on the acoustical conditions (e.g., room reverberation [35]).

To make a direct comparison about the sound quality between recording techniques at off-center listening positions, another experiment is necessary in which all three recording techniques at every off-center position are compared with each other. Such experimental design would result in three times as many pairwise comparisons as in our experiments, might be exhausting for the listeners, and may not even be necessary. Consider that our study measured the sound degradation at all tested listening positions relative to the central listening position for all three recording techniques, and previous studies [3] and [4] (from which we borrowed our musical excerpts) evaluated the sound quality of the recording techniques at the central listening position. Putting the results of these previous studies and our study into dialogue, we can make an informed prediction about the absolute sound quality for each recording technique at off-center positions. From Kim et al. [4] we know that for the piano excerpt (EXC 1) the preferred recording technique (at the central listening position) was Spaced Omnis, followed by OCT and Ambisonics. For the orchestra excerpt (EXC 2) Camerer [3] tested nine perceptual aspects of the recordings at the central listening position. The rating of the Spaced Omnis and OCT were comparable and both techniques were rated better than Ambisonics regarding "image stability", "sound colour", or "room impression". The Ambisonics recording of the orchestra was rated as having too little "presence of room information".

5 SUMMARY AND DISCUSSION

The off-center sound degradation in two different listening room environments was investigated with respect to three recording techniques (RT), two classical musical excerpts (EXC), and multiple off-center listening positions (POS). We found that the tested recording techniques significantly affect the sound degradation strength at off-center listening positions and the size of the sweet area. In most conditions, a somewhat radial sound degradation from the central listening position occurs, but with varying slope across the recording techniques. With increasing distance to the central listening position, the agreement across listeners (indicated by the standard deviation per listening position) also tends to decrease. In all but one condition, spaced microphone techniques create less sound degradation at off-center positions than the coincident Ambisonics techniques (see Fig. 11), supporting Griesinger's hypothesis that time-delay-based decorrelation among the loudspeaker feeds (Interchannel Time Differences) increases the listening area [9].

In a listening environment featuring a standard loudspeaker configuration (experiment A), the worst listening position was typically near the rear surround speaker (Pos. 7 in Fig. 5). For this position, the rear surround speaker dominates the sound image (unbalanced SPL) to the extend that the Listening Envelopment (LEV) is compromised. Future work needs to identify recording and reproduction methods that create a more balanced SPL across the listening area.

In a non-ideal listening environment (Experiment B), the interaction between recording technique and musical excerpt played a significant role in listener preference. Our data suggest that in a more reverberant listening room, a more diffuse sound material (e.g., an orchestra recording) is likely to be better reproduced at off-center listening positions than a recording with more precise source images (e.g., a piano recording). Regarding the reproduction environment, our study shows that when reverberant, classical, multichannel recordings are reproduced in a medium-sized, moderately reverberant space, the usable listening area is larger than it is in a smaller, less reverberant space. Better understanding of listener preference for the Ambisonics recording technique in the EXC 2 (orchestra) condition of Experiment B is required, especially since this recording technique was least preferred for all other conditions. It seems possible that the space itself is adding credible reflected sounds to the mix of sounds arriving at the listeners' ears and that the space favors sound sources that are reproduced by relatively uncorrelated loudspeaker feeds. It may also be possible that the non-ideal loudspeaker configuration in Experiment B constrained reproduction quality of both excerpts.

Uncoupling all of the variables that differentiated experiment A and experiment B (room acoustics, loudspeaker type, and loudspeaker arrangement) would yield better understanding of the interactions. One approach could be to use auditory virtual environments that can simulate a multichannel recording scenario (e.g., [36]). The trade-off involves more controlled variables but less ecological validity [37]. In future work, we hope to examine the instrument–microphone–room interaction at the recording site embedded in musical excerpts. Generating impulse responses from the instrument's position (similar to the loudspeaker orchestra approach in [38]) and capturing them with the tested microphone arrays would provide insights into sound propagation characteristics and performance of microphone arrays. Such impulse responses do not exist for the musical excerpts used in our study.

The selection of the musical excerpts was constrained by the limited availability of content that was simultaneously captured with different recording techniques. A significant amount of equipment, time, and effort is necessary to create such material. While we limit our findings and discussion to two of the most popular genres of surround recordings (solo piano and orchestra), the question remains whether our findings can be generalized to other content types, e.g., ambience recordings for broadcast and film. Although ambience is captured with a variety of microphone arrays, including those used in our study (see e.g., [39]), further work is needed to generalize our findings.

Between the two musical excerpts, the recording techniques slightly differ in positioning, type, and brand of microphone (see Fig. 4, especially the different arrangements of the Spaced Omnis in (c) and (f)). These differences exist to optimize the recording technique for a specific recording environment and musical material, but they also make it difficult to compare directly the perceptual experience of the recorded material. Using exactly the same arrangement to capture both musical excerpts would have made the experimental conditions more controlled but less meaningful, because the recordings would not represent what Tonmeisters actually record and mix in these situations. Our aim was to extend previous work and explore how perceptual data from off-center listening positions. Comparing these musical excerpts and recording techniques within this paradigm is reasonable, considering the small amount of prior work in this area. Our study is exploratory, and we consider our work as a starting point for further discussion and future research.

6 ACKNOWLEDGMENTS

This work was funded by a grant from the Canadian Natural Sciences and Engineering Research Council and the Canada Council for the Arts (NSERC, CCA, STPGP 322947) to Stephen McAdams and Jonas Braasch, an NSERC grant (RGPIN 312774) to Stephen McAdams, and a student award from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Nils Peters. Many thanks to Bruno L. Giordano, Georgios Marentakis, Sungyoung Kim and William L. Martens for fruitful advice. The authors would like to to thank the reviewers for their helpful comments.

REFERENCES

[1] Y. Ando. *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners* (Springer Verlag, New York, 1998).

[2] S. Bech, N. Zacharov. *Perceptual Audio Evaluation—Theory, Method and Application* (Wiley & Sons, New York, 2006).

[3] F. Camerer, C. Sodl. "Classical Music in Radio and TV - a multichannel challenge ; The IRT/ORF Surround Listening Test." http://www.hauptmikrofon.de (2001).

[4] S. Kim, M. de Francisco, K. Walker, A. Marui, W. L. Martens. "An examination of the influence of musical selection on listener preferences for multichannel microphone technique." In "Proc. of the AES 28th International Conference," (Piteå, Sweden, 2006).

[5] H. Irimajiri, T. Kamekawa, A. Marui. "Correspondence relationship between physical factors and psychological impressions of microphone arrays for orchestra recording." In "123rd AES Convention, Preprint 7233," (New York, US, 2007).

[6] Y. Yun, T. Cho. "The study on surround sound production beyond the stereo techniques." In "Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition," pp. 133–140 (Springer, 2012).

[7] G. Marentakis, N. Peters, S. McAdams. "Auditory resolution in virtual environments: Effects of spatialization algorithm, off-center listener positioning and speaker configuration (A)." J. Acoust. Soc. Am., **123**, 3798 (2008).

[8] V. Pulkki. "Microphone techniques and directional quality of sound reproduction." In "112th AES Convention, Preprint 5500," (Munich, Germany, 2002).

[9] D. Griesinger. "The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique." In "AES 19th Intern. Conference," pp. 182–200 (Schloss Elmau, Germany, 2001).

[10] ITU. *Recommendation BS.1116-1, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems* (International Telecommunication Union, Geneva, Switzerland, 1997).

[11] F. E. Toole. *Sound reproduction: loudspeakers and rooms* (Focal Press, Oxford, UK, 2008).

[12] R. Y. Litovsky, H. S. Colburn, W. A. Yost, S. J. Guzman. "The precedence effect." J. Acoust. Soc. Am., **106**, 1633–1654 (1999).

[13] S. E. Olive, F. E. Toole. "The detection of reflections in typical rooms." J. Audio Eng. Soc., **37**, 539–553 (1989).

[14] S. Bech. “Spatial aspects of reproduced sound in small rooms.” J. Acoust. Soc. Am., **103**, 434–445 (1998).

[15] F. Rumsey. “Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm.” J. Audio Eng. Soc., **50**, 651 – 666 (2002).

[16] IEC. *Draft 60268: Sound System Equipment - Part 13: Listening Tests on Loudspeakers* (International Electrotechnical Commission, Geneva, Switzerland, 1997).

[17] T. Lund. “Enhanced localization in 5.1 production.” In “Proc. of the 109th AES Convention, Preprint 5243,” (Los Angeles, US, 2000).

[18] T. Okano, L. Beranek, T. Hidaka. “Relations among interaural cross-correlation coefficient $(IACC_E)$, lateral fraction (LF_E), and apparent source width (ASW) in concert halls.” J. Acoust. Soc. Am., **104**, 255–265 (1998).

[19] H.-K. Lee, F. Rumsey. “Investigation into the effect of interchannel crosstalk in multichannel microphone technique.” In “Proc. of the 118th AES Convention, Preprint 6374,” (Barcelona, Spain, 2005).

[20] J. Bradley, G. Soulodre. “Objective measures of listener envelopment.” The Journal of the Acoustical Society of America, **98**, 2590–2597 (1995).

[21] F. Rumsey, S. Zieliński, R. Kassier, S. Bech. “On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality.” J. Audio Eng. Soc., **118**, 968–976 (2005).

[22] F. Rumsey, S. Zieliński, R. Kassier, S. Bech. “Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences.” J. Acoust. Soc. Am., **117**, 3832–3840 (2005).

[23] A. Devantier, S. Hess, S. Olive. “Comparison of loudspeaker-room equalization preferences for multichannel, stereo, and mono reproductions: Are listeners more discriminating in mono ?” In “124th AES Convention,” (Amsterdam, The Netherlands, 2008).

[24] D. Clark. “Measuring audible effects of time delays in listening rooms.” In “Proc 74th AES Convention, Preprint 2012,” (Audio Eng. Soc., New York, US, 1983).

[25] M. Brüggen. “Coloration and binaural decoloration in natural environments.” Acta Acustica united with Acustica, **87**, 400–406 (2001).

[26] B. Rakerd, W. Hartmann, J. Hsu. “Echo suppression in the horizontal and median sagittal planes.” J. Acoust. Soc. Am., **107**, 1061–1064 (2000).

[27] F. Rumsey. *Spatial Audio* (Focal Press, Oxford, UK, 2001).

[28] M. A. Gerzon, G. J. Barton. “Ambisonic Decoders for HDTV.” In “92nd AES Convention, Preprint 3345,” (Vienna, Austria, 1992).

[29] Soundfield SP451. www.soundfield.com/products/sp451.php (accessed Sept. 2013).

[30] R. Weber. “The continuous loudness judgement of temporally variable sounds with an “analog" category procedure.” In “Results of the third Oldenburg Symposium on Psychological Acoustics. A. Schick.”, pp. 267–294 (Oldenburg, Germany, 1990).

[31] F. E. Toole. “Binaural record/reproduction systems and their use in psychoacoustic investigations.” In “Proc. of the 91st AES Convention, Preprint 3179,” (New York, US, 1991).

[32] S. E. Olive, P. L. Schuck, S. L. Sally, M. Bonneville. “The variability of loudspeaker sound quality among four domestic-sized rooms.” In “Proc. of the 99th AES Convention, Preprint 4092,” (New York, US, 1995).

[33] S. E. Olive. *Interaction Between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction*. Ph.D. thesis, McGill University, Montreal, Canada (2008).

[34] D. Begault, E. Wenzel, M. R. Anderson. “Direct comparison of the impact of head tracking, reverberation, and individualized hrtfs on the spatial perception of a virtual speech source.” J. Audio Eng. Soc., **49**, 904 – 916 (2001).

[35] C. Gilford. “The acoustic design of talks studios and listening rooms.” Proc. Inst. of Electrical Eng., **106**, 245–258 (1959).

[36] J. Braasch, N. Peters, D. L. Valente. “A loudspeaker-based projection technique for spatial music applications using virtual microphone control.” Computer Music Journal, **32**, 55–71 (2008).

[37] C. Guastavino, B. Katz, J. Polack, D. Levitin, D. Dubois. “Ecological validity of soundscape reproduction.” Acta Acustica united with Acustica, **91**, 333–341 (2005).

[38] J. Pätynen, S. Tervo, T. Lokki. “A loudspeaker orchestra for concert hall studies.” In “Proc. of the Institute of Acoustics,” volume 30, pp. 45–52 (2008).

[39] H. Wittek. “Microphone techniques for 2.0 and 5.1 ambience recording.” In “Proc. of the VDT International Convention,” (2012).


Blachford
ACOUSTICS LABORATORY