Audio Engineering Society
Convention Paper
Presented at the 123rd Convention
2007 October 5–8 New York, NY

Evaluating Off-Center Sound Degradation in Surround Loudspeaker Setups for Various Multichannel Microphone Techniques

Nils Peters[1], Stephen McAdams[1], and Jonas Braasch[2]

[1] *McGill University, Schulich School of Music and CIRMMT, 555 Sherbrooke Street West, Montréal, PQ, H3A 1E3, Canada*

[2] *Rensselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, NY, 12180-3590, USA*

Correspondence should be addressed to Nils Peters (`nils.peters@mcgill.ca`)

ABSTRACT
Many listening tests have been undertaken to estimate listeners' preferences for different multichannel recording techniques. Usually these tests focus on the sweet spot, the spatial area where the listener maintains optimal perception of virtual sound sources, thereby neglecting to consider off-center listening positions. The purpose of the present study is to determine how different microphone configurations affect the size of the sweet spot. A perceptual method is chosen in which listening impressions achieved by three different multichannel recording techniques for several off-center positions are compared with the listening impression at the sweet spot. Results of this listening experiment are presented and interpreted.

1. INTRODUCTION
Several types of microphone techniques exist to record music performances for surround sound reproduction. Variations between different techniques are found in the distance and angle between the microphones, and the choice of directivity patterns. All arrays are targeted to produce an accurate spatial impression at the reference listening position, the sweet spot. In general, multichannel surround is known as a non-democratic reproduction technique: Only a listener in the sweet spot perceives sound quality, whereas listeners outside that spot perceive a degraded image [9]. When a surround recording is presented to the audience in a large room where almost all listeners are outside the sweet spot, this raises the question how much of degradation in

sound quality for off-center position is actually occurs. This study explores the influence of three different recording techniques on the amount of sound degradation for off-center listening positions in large rooms.

1.1. Sweet spot

Although every audio recording and reproduction technique refers to the idea of having a sweet spot, the definition is ambiguous. Over the years several authors have attempted to define the term ”acoustical sweet spot” based on their specific understanding, and this has led to a general consensus that the sweet spot is the point in space where an individual is fully capable of hearing the audio mix in the way it was intended to be heard by the executive sound engineer or Tonmeister. In [19] the sweet spot is more technically defined as *the point where the combined wavefront generated by the loudspeaker in the reproduction layout is coherent, or, alternatively, where the listener is roughly equidistant from the radiators.* Similar to the latter definition, the International Telecommunication Union defined in their well-known ITU-R BS.1116-1 recommendation [13] for L/C/R and LS/RS multichannel sound systems the reference listening position–or, sweet spot–at the center of the loudspeaker setup as Figure 1 shows. However, there does not seem to be any true standardization of the term ”sweet spot” in audio reproduction systems. For our purposes, ”sweet spot” will be considered as the point where all loudspeakers are equidistant at equal volume from a listener.

1.1.1. Off-center listening position

At off-center listening positions, sound degradation may affect the following properties [6]:

- Image localization: perceived spatial location of a reproduced sound source;
- Image stability: perceived location of the reproduced sound source, may change with pitch, loudness, or timbre;
- Width homogeneity: uniform distribution of the image.

There are three main effects that cause sound degradation:

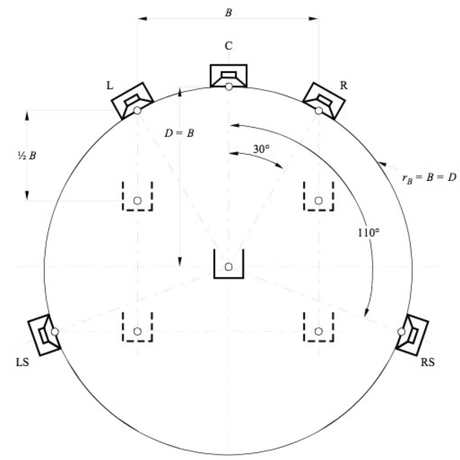


Fig. 1: Recommendation for 5.0 speaker systems (ITU-R BS.1116-1 [13]). Markings for reference listening position (solid lines) and worst case listening positions (dashed lines).

Precedence Effect If a listener in a multichannel sound system is located on an off-center position, the loudspeakers are not equidistant from the listener anymore (e.g. in Figure 1 at a position marked with the dashed lines). The different speaker distances yield different time-of-flights and, therefore, a signal which is transmitted by all loudspeakers at the same time will not conjointly arrive at the listener. In this case, the perceived audio image can be distorted, depending on the relative delays, direction, loudness, and the spectral content of these multiple wavefronts. This distortion can be perceived as an image shift of the phantom source toward the direction of the first arriving wavefront. The perceived location might also become unfocused, or frequency dependent [19]. In addition, precedence can also be observed for spatial features other than direction [3]. Also, the auditory interpretation of room reflections depends on the precedence effect which is demonstrated, for example, in [29] or in [23]. Several studies on the precedence effect have been reviewed in [20].

Comb-filter effect Timbre change through the comb-filter effect occurs by summing two or more similar, but slightly independently delayed audio signals. With respect to surround reproduction in large rooms, this effect can appear through room reflections [10], but also at off-center listening positions between the loudspeaker signals as it was explained above. The absolute threshold for an audible timbre change rises when these delays increase and thus, when the spacing between the comb-filter peaks becomes narrower [17]. On the other hand it can also be seen that very high but very dense irregular room reflections suppress timbre changes, because the different delays produce a high temporal diffusivity [1].

Proximity Effect All loudspeakers used in practice have an angular directional dependence of radiation which usually increases for high frequencies. For the central listening position, this pattern might not be relevant in the listening experience because all loudspeakers are supposed to face that center point. At off-center listening positions the speakers are off-axis to the off-center listener and thus this off-axis radiation pattern of the speakers becomes apparent. Furthermore, the proximity of a listening position to a certain loudspeaker engenders an increase in the experience of loudness. Also, the perception of the curvature of the emitting wavefront is heightened. Thus, the loudspeaker and its location becomes more and more detectable—an undesirable result when producing phantom sound sources.

2. METHODOLOGY

This investigation addresses off-center listening positions to determine how different microphone techniques affect the size of the sweet spot in a 5.0 multichannel sound system. In a listening experiment, the reproduced soundfield at different off-center positions is compared with the soundfield a listener perceives at the sweet spot of the system. In this study, to capture sound fields at different listening positions, two sets of 5.0 multichannel recordings of different kinds of musical performances (see section 3.2) were used. These recordings were made simultaneously with three different multichannel microphone techniques, which are described in section 3.1. These recordings were presented in two different large rooms through a 5.0 multichannel sound system and binaural stimuli were recorded at different listening positions (see sections 4 and 5). For the listening experiment, a computer-aided listening test was designed (see section 3.4). In a sound-proof booth and using headphones, trained listeners with normal hearing were asked to compare these binaural stimuli under controlled conditions. For easy comparison, the interface allowed listeners to select between two stimuli in real time.
This method has the advantage that listening positions can be compared quickly and randomly in a double-blind test because the subject remains seated while the listening position changes virtually. A memory problem can also be avoided which might appear if participants had to change listening position physically to make these comparisons. Removing the need to change listening position by isolating and presenting the auditory stimuli through headphones also circumvents the potential for sound quality prejudice on behalf of the participants.

2.1. Former studies on evaluating surround microphone techniques

Several perceptual listening tests have been performed to assess differences among various surround microphone techniques at the reference listening position (e.g. [15], [14] or [2]). The work by Camerer and Sodl [5] as well Kim et al. [16] is highlighted here because the present study uses a selection of their surround recordings to evaluate off-center sound degradation. In both studies, experienced Tonmeisters Martha de Francisco in [16] and Florian Camerer in [5] recorded musical performances with several multichannel microphone techniques simultaneously. In both, an attempt was made to optimize each microphone technique in terms of sound quality to ensure a fair comparison. For the listening tests, expert and trained listeners were asked to judge these recordings according to several spatial and timbral aspects. Although these listening tests used different questionnaires and different musical materials, the results followed the same tendency: For the reference listening position, spaced microphone techniques were preferred above coincident microphone techniques, the latter represented by the Ambisonics approach. However, all of these studies mainly focused on listening position at the sweet spot and excluded off-center listening.
An approach that is somehow related to our method was proposed in [18]. Here, the quality of loudspeaker spatialization techniques at different listening positions could be predicted by headphones, whereby the binaural cues were synthesized in order

to simulate the pressures at the listener's ears corresponding to a sound field produced by loudspeakers. Griesinger [9] as well as Martin [21] suggested that inter-channel decorrelation increases the sweet spot, which can be achieved by spacing the microphones. To our knowledge, no formal listening tests were performed to confirm this hypothesis.

3. EXPERIMENTAL DESIGN

3.1. Multichannel recording techniques

For this study, three well-known multichannel microphone techniques were tested [25]. The chosen techniques differ in their strategy for avoiding inter-channel crosstalk, which is either realized due to the microphone directivity pattern to increase amplitude separation and/or due to displacement of the microphones to create inter-channel time delays. The following section gives a short overview of these recording techniques and their setup within the recordings used in our study. Sketches to describe the microphone setup can be found in the appendix of this paper (Figures 17 - 19). To ease the reading of this study, we shorten the term multichannel recording technique to RT.

3.1.1. Coincident Microphone Technique - Ambisonics

Ambisonics is a recording technique which aims to capture a soundfield at a single point. Originally it goes back to [4] and extends Blumlein's ideas about coincident recording and reproduction techniques: By adding an omnidirectional microphone to the pair of Figure-of-Eight units (for left-right and front-back), it can be shown (e.g. [24]) that this setup captures all information that can be extract from the horizontal soundfield at that point. It is assumed that the microphone capsules are spatially coincident, meaning that all three capsules are acoustically at exactly the same spot in the soundfield. By adding a third Figure-of-eight unit perpendicular to the other two, the up-down component of the soundfield is also taken into account. The soundfield is encoded in terms of velocity and pressure components in the so called B-Format and contains the 4 channels (W,X,Y,Z). This format does not include any information about the reproduction setup. Therefore, to reproduce the soundfield, the B-format channels have to be suitably decoded according to the corresponding speaker setup. The most frequently applied decoder to transform B-format into 5.1 is the so-called *Vienna decoder* introduced in [8]. The Ambisonics versions of the used recordings were made with a SoundField MKV microphone and encoded via the SoundField model SP451 surround processor.

3.1.2. Spaced Cardioid Microphone Technique - OCT

The Optimized Cardioid Triangle (OCT) was first proposed in [28]. This technique is known to reduce crosstalk between channels by incorporating two outer hyper-cardioid microphones facing ±90° side-ways (see Figure 18). The center microphone is a cardio capsular which is supposed to be placed 8 cm forward. Optional omnidirectional pressure microphones may also be used for enhanced low-frequency response. Signals from these optional microphones are low-pass filtered and summed with the high-passed filtered portions of the left and right microphones. For a surround recording, the OCT array is usually combined with several rear techniques such as: OCT surround, IRT cross, or Hamasaki Square [11]. In both tested musical excerpts, an OCT and Hamasaki square combination was used.

3.1.3. Spaced Omni Microphone Technique

The omnidirectional microphones applied for this recording technique are widely spaced, thus creating larger inter-channel time delays.

Decca Tree + Hamasaki-Square The Decca Tree is originally designed for 3 Neumann M50 omnidirectional microphones arranged in a triangle. The center microphone is placed 70 cm to 100 cm forward, whereby the right and left capsulars are spaced in a distance ranging from 1.4 m to 2 m depending on the intended recording angle. For large sound sources, such as an orchestra, the system can be extended with additional omnidirectional microphones to the side, the so-called "out-riggers". The Decca Tree has been widely used for large-scale recordings and is a favorite among film scoring mixers because of its ability to maintain imaging and separation through the various matrix systems employed in the distribution of film soundtracks. To feed the rear channels in a surround speaker setup, the Decca Tree is usually expanded with an IRT cross, or the

Hamasaki Square, which was what Camerer used to record the Mozart symphony (EXC 2).

Polyhymnia Pentagon This technique, invented by Polyhymnia International (formerly the Philips Classics Recording Department), uses five widely spaced omnidirectional microphones. It is often described as a multichannel version of the stereophonic Decca-Tree. The microphones are arranged in a large circle, in which the position corresponds to the azimuthal angles of the speaker in the ITU-R BS.775-1 recommendation [12]. The Polyhymnia Pentagon was used for the Bach piano recording (EXC 1).

3.2. Surround recordings

Two sets of 5.0 multichannel recordings of different kinds of sounding bodies were used: a piano and a symphony performance. Each musical excerpt was recorded with the previously described RTs simultaneously to make sure that the musical quality of the performance remains consistent. For this study, musical excerpts of the following pieces were used:

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

2. W.A. Mozart "Maurische Trauermusik" (KV 477) c-minor

The Bach piece (EXC 1) was performed by Thomas Plaunt and recorded in Pollack Hall, McGill University Montréal in 2006. The recording procedure is described in detail in [16]. The Mozart symphony (EXC 2) was performed by the RSO Vienna in Austrian Radio "Grosser Sendesaal" (large broadcast hall) and recorded under the supervision by Florian Camerer ([5]).

3.3. Recording of the binaural stimuli

These two musical excerpts, each recorded with three RTs described in section 3.1, were presented through surround loudspeaker setups. At specific positions binaural recordings were made to capture the soundfield a potential listener would perceive while sitting there. At each of these positions, six binaural recordings (2 EXC · 3 RT) were captured. Each excerpt has a duration of about 7 s to shorten the duration of the listening test. Additionally, we

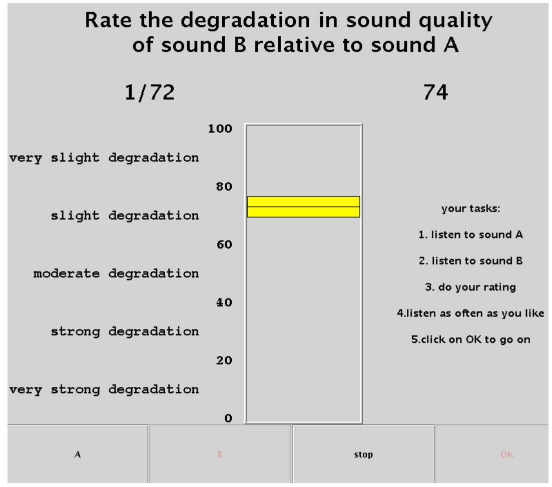


Fig. 2: Graphical user interface

measured room impulse responses for each loudspeaker at each listening position according to [7]. The signals were recorded at 48 kHz with an RME Fireface 800. An Apple Macbook running Pure Data was used for controlling playback and recording processes.

3.4. Listening experiment

To evaluate the binaural recordings, a graphical user interface was designed using *PsiExp* [26] (see Figure 2). The subjects were asked to perform the following task: *"Rate the degradation in sound quality of the sound B relative to sound A."* Sound A represents one of the six sweet spot recordings (EXC $\times$ RT), while sound B could be: a) one of the off-center recordings with the same EXC $\times$ RT; b) the same sweet spot recording as sound A (a hidden reference); or c) a monaural recording taken at a off-center listening position (a hidden anchor). The purpose of the hidden reference and anchor is to set best- and worse-case references for the rating scale. The trials were presented in random order (double blind single stimulus test). Within the presented pair, the listener could switch between sound A and sound B at will. The ratings were made on a slider with a continuous scale from 0 to 100, where 0 corresponds to the bottom (total degradation) and 100 to the top of the scale (no degradation) using a computer mouse. The slider of the scale was auxiliary marked by descriptors in the following order: very strong degradation - strong degradation - moderate degradation - slight degradation - very slight degradation. The listening experiments took place at McGill University, Music Technology Area, Music Perception and Cognition Lab, and the Banff Centre facilities. All subjects were tested under similar conditions. Each subject completed a questionnaire concerning their musical practice, and their headphone listening sound recording experiences. Furthermore, after the listening experiment they were asked to verbally describe their strategy for performing the rating task.

3.4.1. Procedure

The experiment consisted of a training phase, a familiarization phase and the experimental phase. The subject read the experimental instructions and asked any questions necessary for clarification. For the training phase five trails were presented to train the subjects on the operation of the user interface. The musical excerpts used for training were different from those presented in the experimental phase. Furthermore, the subjects were informed that these ratings would not be recorded. After the training phase a representative collection of five samples of each group (2 EXC $\cdot$ 3 RT) were presented to familiarize them with the stimuli. They were told that the familiarization phase would give them a sense of the range of variation in sound degradation, so they could subsequently use the full scale in making their judgments in the experimental phase. This latter phase lasted approximately 60 minutes. The participants were free to take breaks whenever they wanted. Each stimuli was presented twice. To keep a realistic impression of the binaural recordings, the subjects were told to face the frontal direction and not to move their heads while listening to the sounds samples. The stimuli were presented over Sennheiser HD 600 Headphones at normal listening level (65 dBA LAeq for the sweet spot recording) in a sound-proof booth (McGill University) or alternatively in an acoustically treated audio editing suite (Banff Centre for the Arts).

4. EXPERIMENT A - TANNA SCHULICH HALL

The Stimuli for Experiment A were recorded in Tanna Schulich Hall, McGill University Montréal. The Tanna Schulich Hall, built in 2005, has a floor space of ca. $240m^2$ with 188 seats. The reverberation time for this recording may be seen in Table 1. The 5-channel loudspeaker system installed in the hall was used (Kling & Freitag CA 1515 for Left ,Center & Right, two Kling & Freitag CA 1001 for the surround speaker) and was calibrated in terms of optimal sound quality for the reference listening position (see Figure 4). The positions of the speakers differ from the ITU-R BS.775-1 recommendation [12] according to azimuthal angle; instead the surround speakers are placed at ± 150 with an arc of ca. 8.2 meter, measured from the reference listening position. Furthermore the center speaker was noticeably elevated (see Figure 3). Due to the graded seating in the hall, the listening perspective relative to the elevated speakers varies. Therefore this layout can be seen as a non-optimal, but somewhat practical speaker setup.

A B&K Head and Torso Simulator (HATS) with shoulder damping fabric was placed at 12 listening positions in the hall (see Figure 4). An omnidirectional microphone (Earthworks QTC1) and a microphone with a Figure-of-Eight directivity pointing to the sides (Sennheiser MKH 30) were placed above the dummy-head for additional measurements. The recordings were limited to one side of the hall because it can be assumed that symmetrical degradation occurs due to the symmetrical shape of the room and the symmetrical speaker setup. The SPL during the recordings varied between 74 dBA and 77 dBA depending on the position in the hall. The measured SNR was ca. 50 dBA.

Fig. 3: Loudspeaker setup with elevated center speaker, in the background the graded seats with the B&K HATS at the reference listening position.

Frequency [Hz]	125	250	500	1k	2k	4k	8k	16k
RT_{60} [sec.]	0.87	0.82	0.86	0.77	0.79	0.76	0.66	0.58

Table 1: Reverb time RT_{60} in octave bands measured in Tanna Schulich Hall

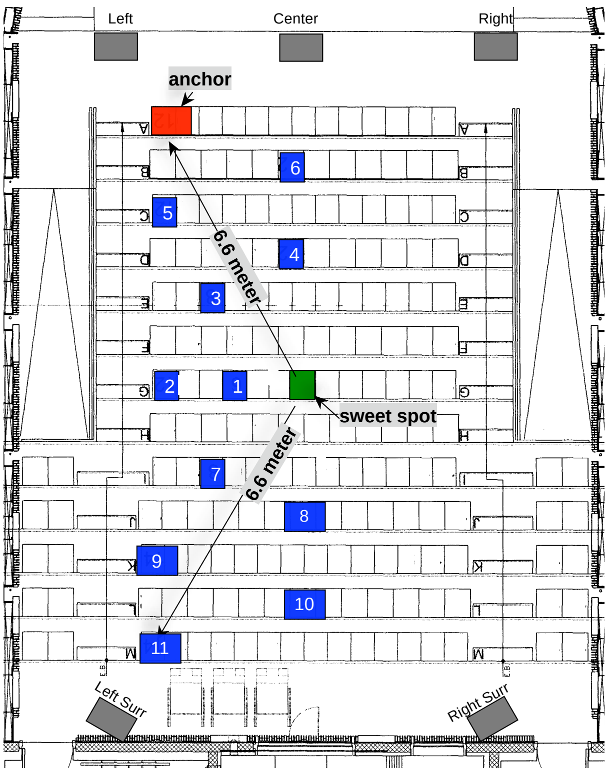


Fig. 4: Recorded listening positions in Tanna Schulich Hall.

4.1. Participants & Variables

Nineteen subjects of both gender with normal hearing participated in the experiment. They were trained listeners, either studying or working in the field of acoustics or sound recording. They were trained in listening to sound in a critical way. The participants were students from the sound recording programs at McGill University Montreal, as well as students and stuff from the sound recording work-study program of the Banff Centre for the Arts, AB. Their ages ranged from 23 to 44 (M=28) and their work experience from 1 to 23 years (M=7.4). On average the subjects were used to using headphones 1.7 hours/day. The independent variables for that experiment seen in Table 2 and yields 72 experimental conditions. The hidden anchor was a monaural recording taken at the position marked in Figure 4.

Independent Variables		Variations
Musical excerpt	EXC	2
Recording technique	RT	3
Positions in Tanna S. Hall	POS	11+2*

Table 2: Independent variables for Experiment A. * +2 represents hidden reference and hidden anchor

4.2. Results of Experiment A

Outline data were removed from the analysis. A given rating was considered as an outlier if its value was more than two standard deviations away from the mean value for that particular stimulus. In this case, the outlier was replaced by that mean value. Fifty-seven ratings were detected as outliers, or 3.8% of all ratings.

A repeated-measures analysis of variance (ANOVA) shows that all independent variables and their interactions are significant ($p < 0.001$) (see Table 3). The ANOVA was calculated for two conditions: 1) Taking all positions into account, and 2) Taking only the closest six positions to the reference position into account (positions 1-4,7,8 see Figure 4).

For the "all positions" analysis, the main effect POS, the interaction EXC × RT, as well as EXC have a bigger effect size than the RT effect. From all tested effects, the interaction RT × POS seems to be least important.

For the second ANOVA (the center area) the effect sizes changed, but all variables are still significant. Here, POS has the biggest effect size, followed by EXC, RT and the EXC × RT interaction. As in the first ANOVA, RT × POS seems to be least important from all tested effects.

The big effect size of POS in both analyses can be interpreted as POS having the most influence on the perceived sound degradation. This is expected

and shows the areal limits of the sound reproduction within a five loudspeaker system. Contrariwise and surprisingly, the effect size of EXC is bigger than the size of RT. The increase of RT's effect size for the center area shows that the recording technique influences the ratings for inner positions more than it does across all tested listening positions.
The ratings of the listening experiment can be seen in the Figures 6 and 8 which show interpolated contour-plots based on the mean ratings at different positions. Especially, the Figure 6 demonstrates how the sound quality degrades radially from the reference listening position but with varying degrees of speed across the tested RTs. The plots for the Bach piano excerpt (Figure 6) presents obviously better ratings for Spaced Omnis at off-center positions than for the Ambisonics recording. The largest differences for EXC 1 between RT can be found at position 8. The results for the Mozart symphony recording (Figure 8) are less uniform and therefore unclear with the largest differences between RT at position 5. Comparing the mean ratings for EXC 2 across all RTs, it surprises that especially at position 2 and 5 the Ambisonics recording was rated best. In particular, the Ambisonics recording of (EXC 2) was rated better by the participants than the Ambisonics recording of piano (EXC 1). Figure 5 demonstrates this interaction by showing the mean ratings as a function of the musical excerpt for each recording technique across all listening positions. This interaction is unexpected. The results of a Tukey-Kramer HSD post-hoc test, which reveals listening positions that are not rated significantly differently from the reference positions, are shown for each musical excerpt in Figures 7 and 9. The results for EXC 1 in Figure 7 displays the biggest reference listening area produced by the Spaced Omnis microphone technique. The results for EXC 2 presented in Figure 9 shows equivalent sound degradation for Spaced Omnis and the OCT recordings.

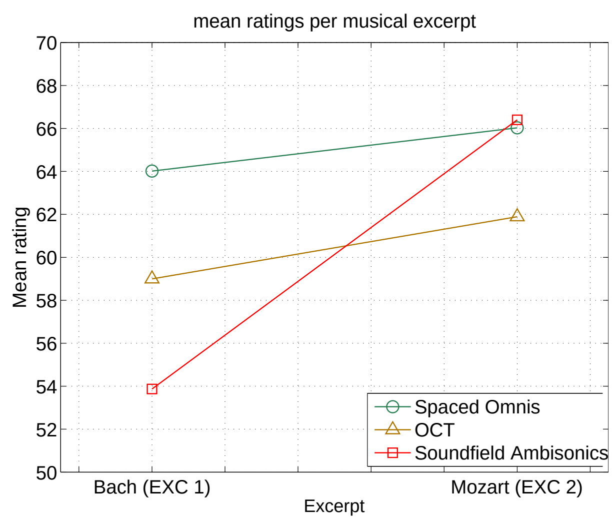


Fig. 5: Experiment A: Mean ratings across all positions as a function of RT and EXC. Interaction between RT and EXC.

Effect	df	F	GG_ϵ	p	η_P^2	
Musical excerpt (EXC)	$1, 18$	19.3	-	$< .001$	0.517	[3]
Recording technique (RT)	$2, 36$	13.7	-	$< .001$	0.433	[4]
Listening position (POS)	10, 180	107.1	.344	$< .001$	0.856	[-]
POS*	$3.4, 61.9$	107.1	-	$< .001$	0.856	[1]
EXC × RT	$2, 36$	40.1	-	$< .001$	0.690	[2]
EXC × POS	$10, 180$	8.2	.592	$< .001$	0.313	[-]
EXC × POS*	5.9, 106.6	8.2	-	$< .001$	0.313	[5]
RT × POS	$20, 360$	5.4	.406	$< .001$	0.229	[-]
RT × POS*	8.1, 146.1	5.4	-	$< .001$	0.229	[7]
EXC × RT × POS	$20, 360$	6.1	.461	$< .001$	0.254	[6]
EXC × RT × POS*	9.2, 165.5	6.1	-	$< .001$	0.254	[6]

Table 3: Experiment A: Robust ANOVA of data for "all positions". * indicates application of the Greenhouse Geisser correction for violations of sphericity. [ ] shows the ranking of the measured effect size η_P^2.

Effect	df	F	GG_ϵ	p	η_P^2	
Musical excerpt (EXC)	$1, 18$	29.7	-	$< .001$	0.662	[2]
Recording technique (RT)	$2, 36$	20.4	-	$< .001$	0.531	[3]
Listening position (POS)	5, 90	85.8	.529	$< .001$	0.827	[-]
POS*	$2.6, 47.6$	85.7	-	$< .001$	0.827	[1]
EXC × RT	$2, 36$	19.1	-	$< .001$	0.515	[4]
EXC × POS	$5, 90$	3.8	.581	.003	0.175	[-]
EXC × POS*	$2.9, 52.3$	3.8	-	.016	0.175	[6]
RT × POS	$10, 180$	3.7	-	$< .001$	0.169	[7]
EXC × RT × POS	$10, 180$	6.4	-	$< .001$	0.262	[5]

Table 4: Experiment A: Robust ANOVA of data for the Center Area (Pos. 1-4,7,8 in Figure 4). * indicates application of the Greenhouse Geisser correction for violations of sphericity. [ ] shows the ranking of the measured effect size η_P^2.

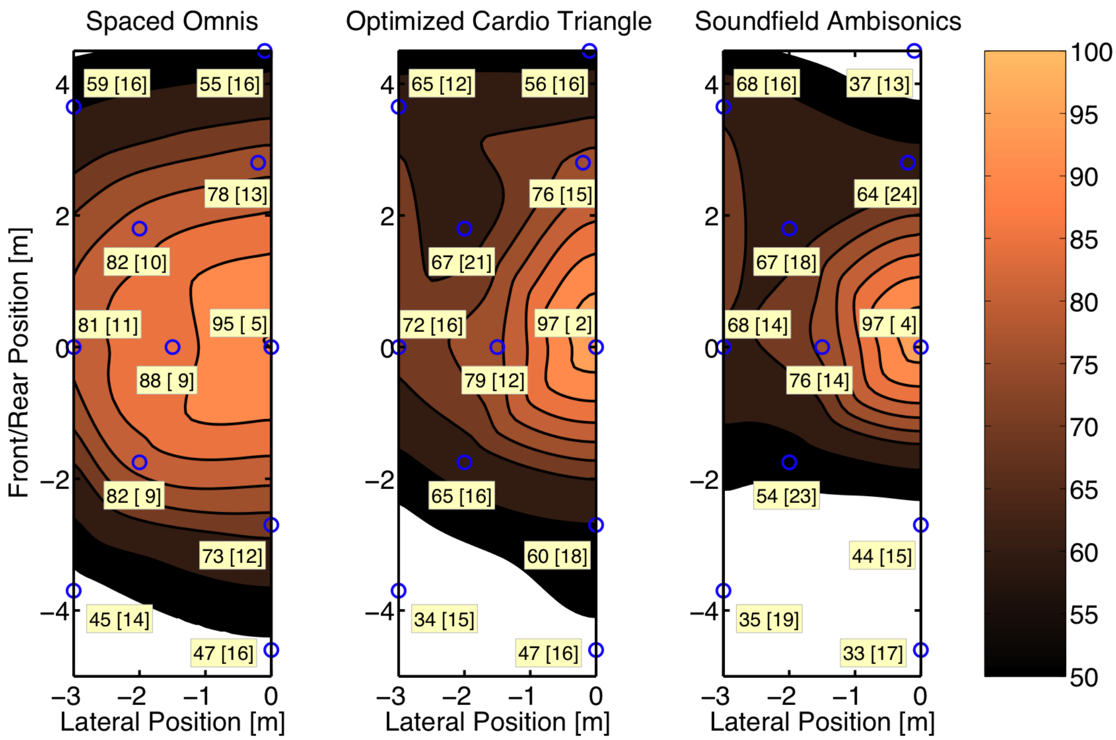


Fig. 6: Experiment A: Mean ratings for the Bach piano recording (EXC 1). Referring to Figure 4 the tested positions are marked with blue circles, standard deviation in [ ]. Contour-plots were created with cubic interpolation.

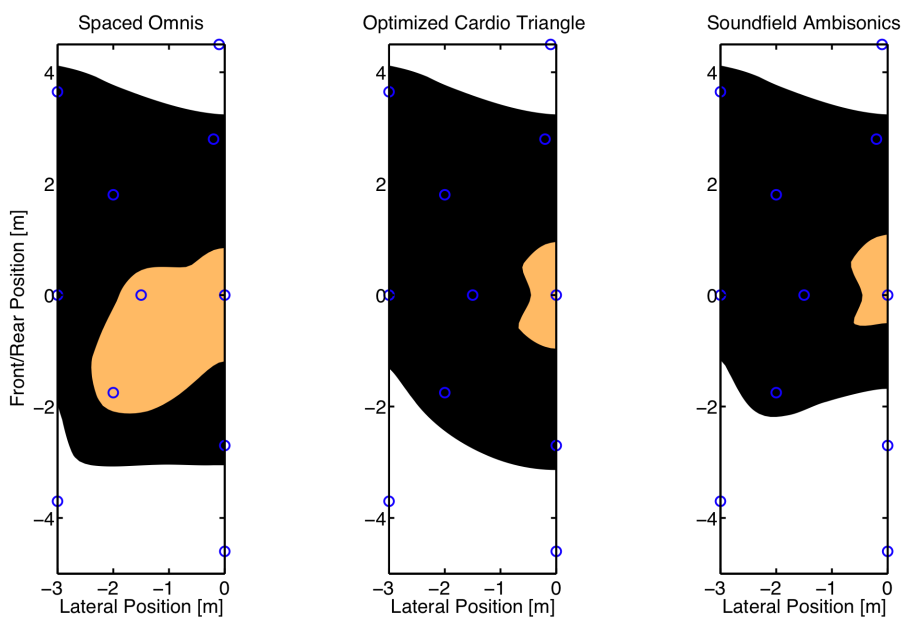


Fig. 7: Experiment A: Contour plot of the Tukey-Kramer HSD post-hoc analysis, for EXC 1.

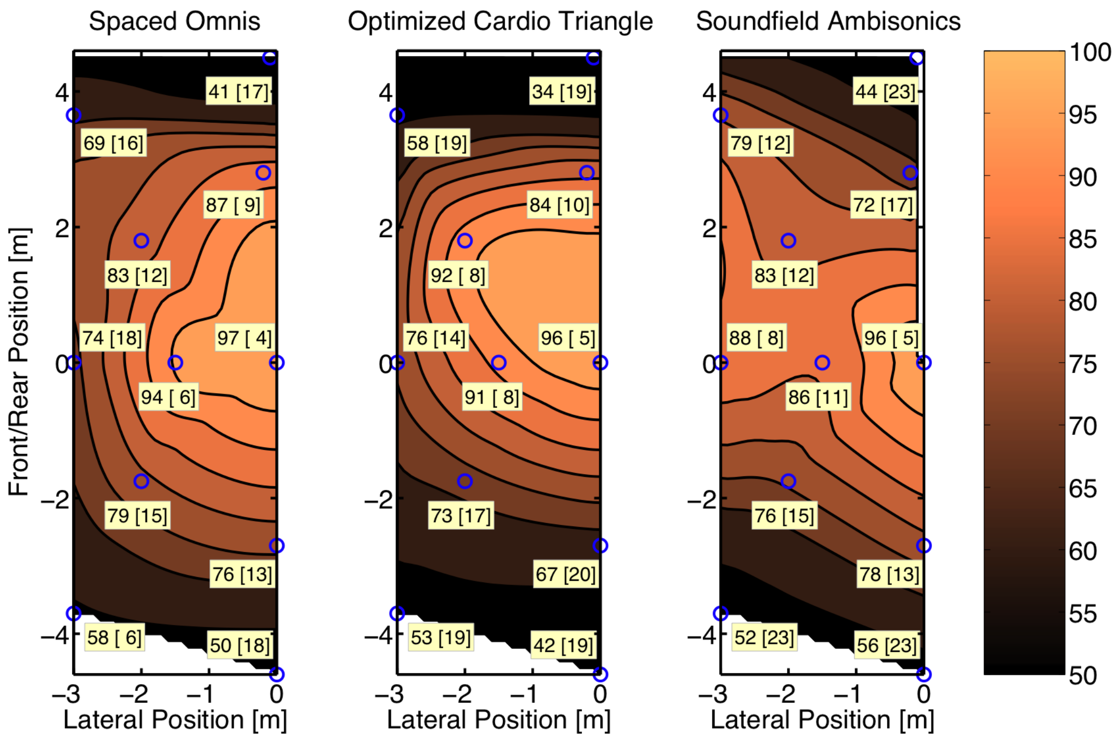


Fig. 8: Experiment A: Mean ratings for the Mozart symphony recording (EXC 2). Referring to Figure 4 the tested positions are marked with blue circles, standard deviation in [ ]. Contour-plots were created with cubic interpolation.

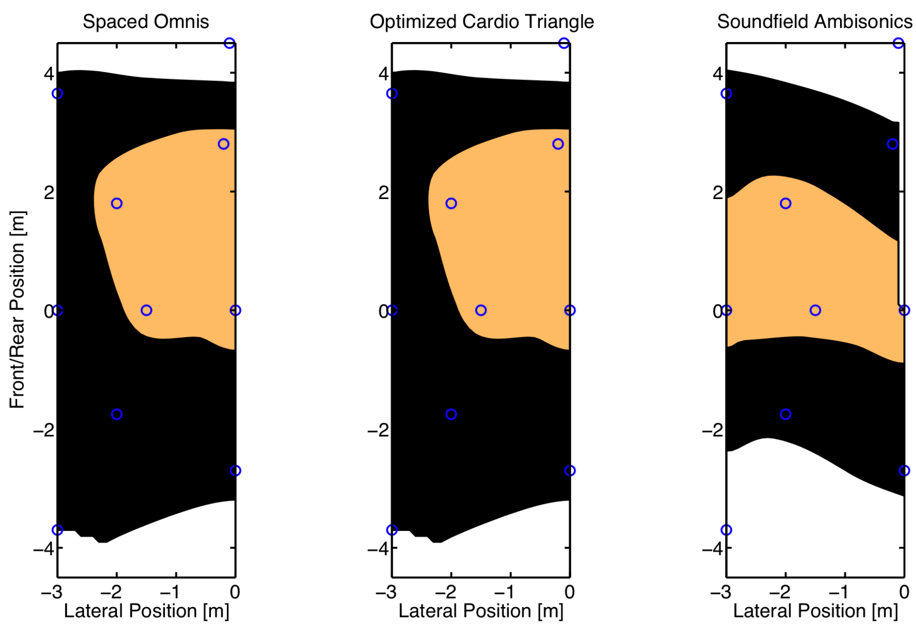


Fig. 9: Experiment A: Contour plot of the Tukey-Kramer HSD post-hoc analysis, for EXC 2.

5. EXPERIMENT B - TELUS STUDIO

The stimuli for Experiment B were recorded in Telus Studio at the Banff Centre for the Arts in Banff, AB. This studio, which is usually used as a recording room for medium-large ensembles or as a film set, has a floor-space of ca. $140m^2$ and a volume of ca. $800m^3$. The estimated reverberation times Rt_{60} may be seen in Table 5. The measured signal-to-noise-ratio (SNR) was ca. 45 dBA. In terms of Rt_{60} and SNR, the Telus Studio fits to the ITU-R BS.1116-1 recommendation [13] for multichannel loudspeaker setups for larger listening rooms. According to Figure 1, five self-powered, two-way, Dynaudio BM15A loudspeakers at a height of 1.2 meter were placed on an arc with a radius of $B = 4.2$ m. The setup can be seen in Figure 10. To record the signals, the

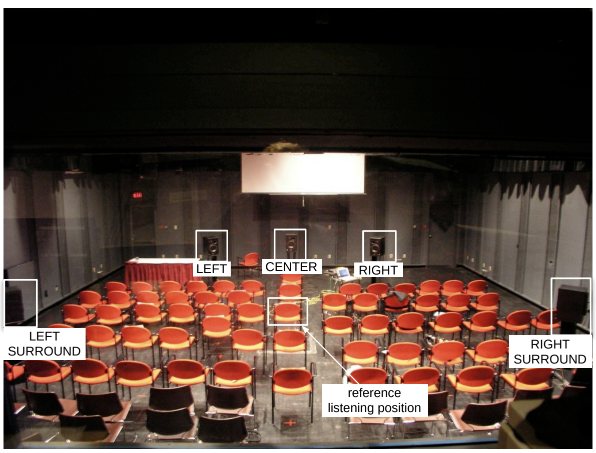


Fig. 10: Telus Studio with ITU speaker setup

the B&K HATS was replaced by the head of the first author: DPA 4060 omnidirectional miniature microphones were placed at the beginning of the ear canals to record the stimuli. To avoid uncontrolled head movements that might cause artifacts for the listening experiment, a neck-brace was used. According to the recommended maximal listening area of a ITU 5.0 setup (see best case and worst case positions in Figure 1), the positions to record the binaural stimuli were chosen as it can be seen in Figure 11. The SPL during the recordings varied between 73.5 dBA and 79 dBA depending on the position in the studio. The sound pressure level (SPL) of the reference listening position was calibrated to have the same level as the SPL of the reference position in Experiment A (75 dBA).

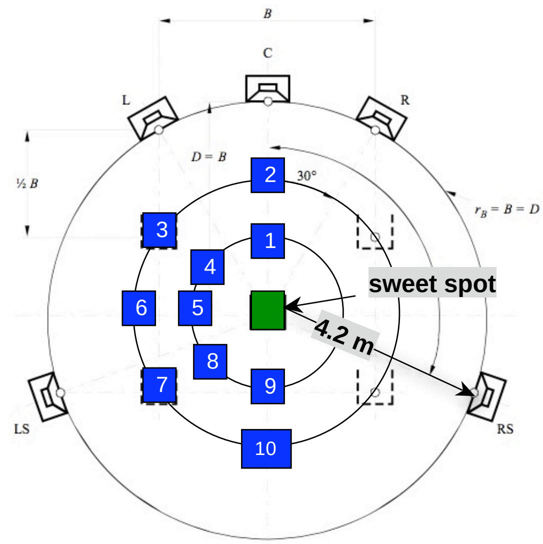


Fig. 11: Recorded positions in Telus Studio (marked in the ITU plot [13]).

Frequency [Hz]	RT_{60} [sec.]
125	0.74
250	0.67
500	0.59
1k	0.54
2k	0.53
4k	0.50
8k	0.43
16k	0.34

Table 5: Reverb time RT_{60} in octave bands measured in Telus Studio

5.1. Participants & Variables

The loudness of the recorded binaural stimuli were adjusted to the loudness of the stimuli presented in Experiment A by estimating the gain level needed to equalize the loudness of the sweet spot recordings of Experiments A and B. The independent variables can be seen in Table 6. A monaural recording of listening position 10 was chosen as the hidden anchor. Ten trained listeners were tested. All of them also participated in Experiment A. The age of this population varies between 24 and 44 (M=30) and the

Effect	df	F	GG_ϵ	p	η_P^2	
Musical excerpt (EXC)	1, 9	0.2	-	.684	.019	[7]
Recording technique (RT)	2, 18	43.3	-	$< .001$	.828	[1]
Listening position (POS)	9, 81	31.0	.401	$< .001$	.775	[-]
POS*	3.6, 32.5	31.0	-	$< .001$	.775	[2]
EXC × RT	2, 18	4.2	-	.032	.317	[6]
EXC × POS	9, 81	7.3	.498	$< .001$	.450	[-]
EXC × POS*	4.5, 40.4	7.3	-	$< .001$	.450	[3]
RT × POS	18, 162	6.0	.313	$< .001$	.402	[-]
RT × POS*	5.6, 50.7	6.0	-	$< .001$	.402	[5]
EXC × RT × POS	18, 162	6.1	.30	$< .001$	.404	[-]
EXC × RT × POS*	5.4, 48.6	6.1	-	$< .001$	.404	[4]

Table 7: Robust ANOVA, experiment B; * indicates Greenhouse Geisser Correction, [ ] shows ranking of the measured effect size

work experience is between 1 and 23 years (M=9).

Independent Variables		Variations
Musical excerpt	EXC	2
Recording technique	RT	3
Positions in Telus Studio	POS	10+2*

Table 6: Independent variables for Experiment B. * +2 represents hidden reference and hidden anchor

5.2. Results of experiment B

As described in section 4.2, twenty-three outliers (1.6% of all ratings) were rejected. A repeated-measures ANOVA was performed and shows that the EXC effect is not significant. All other tested effects are significant ($p < 0.001$), whereas the EXC × RT interaction is relatively weaker (see Table 7). A measure of effect size (η_P^2) demonstrates that RT (0.828) and the POS (0.775) are the largest effects. Figure 13 for EXC 1 and Figure 15 for EXC 2 show interpolated contour-plots based on the average ratings at the tested positions. The largest difference for EXC 1 between RT can be found at position 5 (see Figure 11) and for EXC 2 at position 1. As it could be observed for Experiment A in Figure 15, a similar sound degradation also occurs for EXC 1 in Figure 13. Figures 14 and 16 demonstrate the results of a Tukey-Kramer HSD post-hoc test for each RT and EXC. The ratings for the Bach piano excerpt (EXC 1) reveal that the Spaced Omni microphone technique produces a larger area where the reference

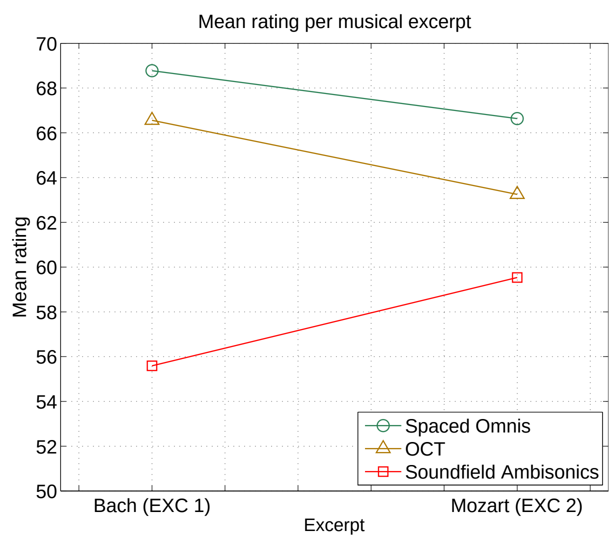


Fig. 12: Mean ratings across all positions as a function of RT and EXC for Experiment B

sound quality can be perceived. For the Mozart symphony recording (EXC 2), the Ambisonics recording produces the worst results.

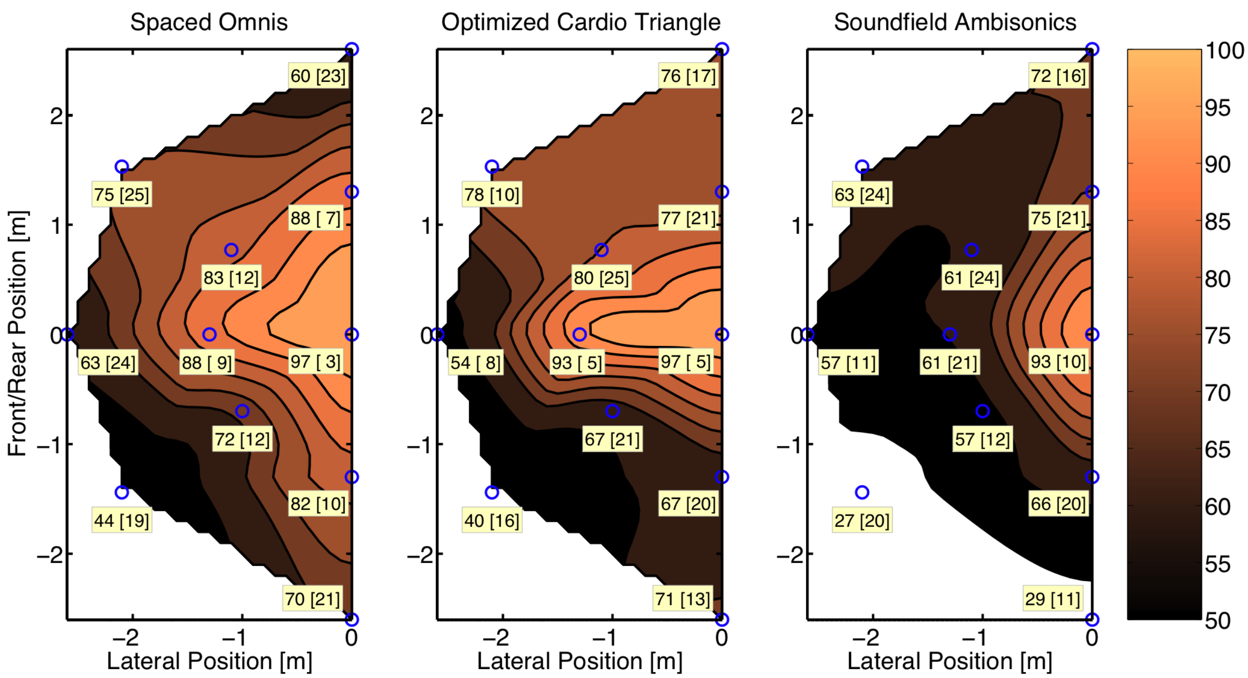


Fig. 13: Experiment B: Mean ratings for the Bach piano recording (EXC 1) in the Telus Studio. Referring to Figure 11 positions the tested are marked with blue circles, standard deviation in [ ]. Contour-plots were created with cubic interpolation.

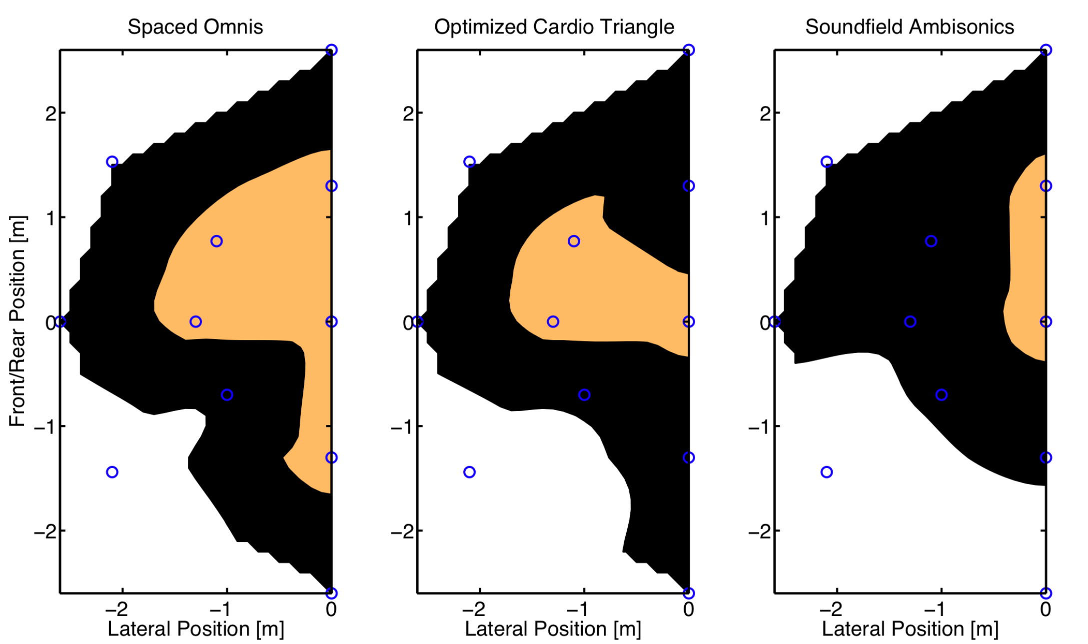


Fig. 14: Experiment B: Contour plot of the Tukey-Kramer HSD post-hoc analysis, for EXC 1.

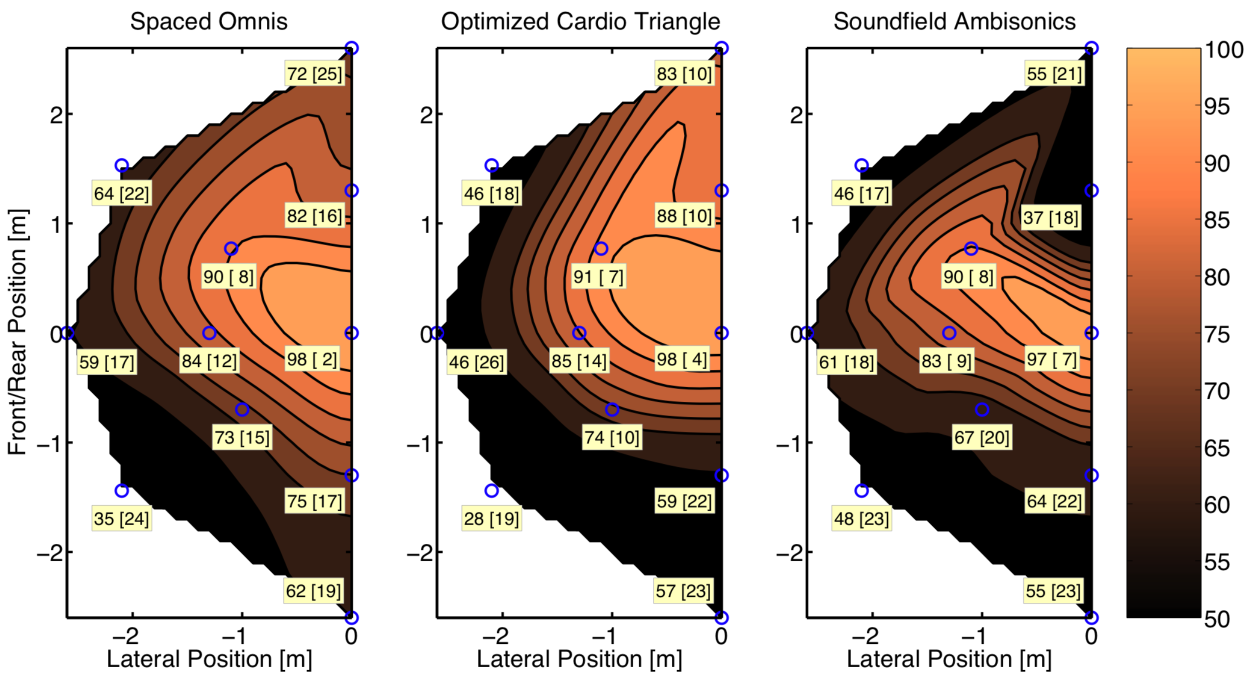


Fig. 15: Experiment B: Mean ratings for the Mozart symphony recording (EXC 2) in the Telus Studio. Referring to Figure 11 the tested positions are marked with blue circles, standard deviation in [ ]. Contour-plots were created with cubic interpolation.

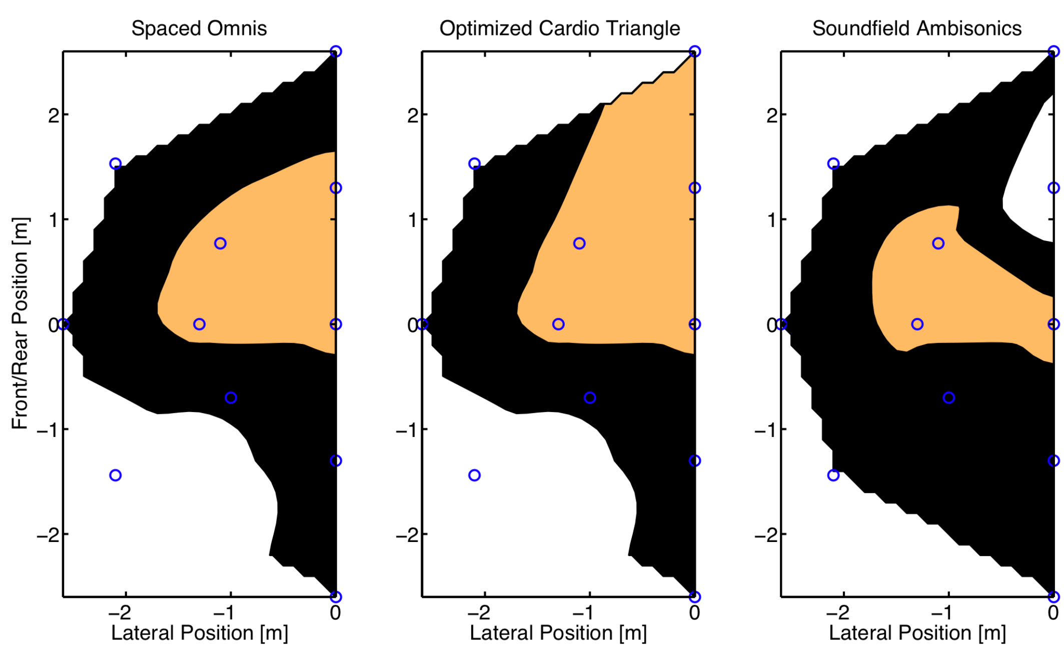


Fig. 16: Experiment B: Contour plot of the Tukey-Kramer HSD post-hoc analysis, for EXC 2.

6. COMPARISON & DISCUSSION

All pairs of recording technique were compared using a Bonferroni procedure as well as a Tukey-Kramer HSD post-hoc test for different conditions. Both tests gave equivalent results (see Table 8). The Bach piano recordings (EXC 1) show perceivable differences between all techniques in both experiments, except between OCT and Spaced Omnis in Experiment B. For the Mozart symphony recording (EXC 2) in Experiment A, we found significant differences for the condition "all position", but no differences for for the "center area". In Experiment B, the comparison methods found significant differences between Spaced Omnis and Ambisonic.

Comparing Figure 5 (Experiment A) with Figure 12 (Experiment B) reveals that, in both experiments, mean ratings across all positions for RT generally differ between EXC. Especially, the ratings for the Ambisonics recording increase from EXC 1 to EXC 2 and cause an interaction with RT in experiment A. On the other hand, the ratings for Spaced Omnis and OCT follow the same tendency in both experiments, whereas the ratings for Spaced Omnis are higher than for OCT.

The different results between both experiments can be attributed to unequal conditions, such as different loudspeaker setup (model and position), room acoustic properties, and also the binaural recording technique. Although both musical excerpts were recorded by similar multichannel recording techniques, the variation in results for EXC might be caused by differences related to instrumentation, the size of the sounding bodies, different microphone models, and hall acoustics. In particular, the EXC 2 recordings included more reverberation and ambience than EXC 1, while the strings in EXC 2 produced a natural "chorus" effect. Perhaps these effects mask the perception of sound degradation in EXC 2. We are aware that the used binaural recording techniques are limited and cannot reproduce a fully spatial impression due to the lack of freedoms in terms of head movement, and to the mismatch between the participants HRTFs and those which where used for the recording process. An evaluation of artificial heads in [22] shows that the localization errors for binaural recordings primarily occur in the median plane. Moreover trained listeners have the ability to adapt to a "foreign" sets of HRTF after a short learning phase. Nevertheless, an improvement concerning head movements could be done by applying an auralization technique called binaural room scanning (BRS) at each listening position [27].

Compared RT	OCT / Spaced Omnis	Ambisonics / Spaced Omnis	OCT / Ambisonics
Experiment A			
all POS, EXC 1	$\neq$	$\neq$	$\neq$
all POS, EXC 2	$\neq$	$=$	$\neq$
center area, EXC 1	$\neq$	$\neq$	$\neq$
center area, EXC 2	$=$	$=$	$=$
Experiment B			
EXC 1	$=$	$\neq$	$\neq$
EXC 2	$=$	$\neq$	$=$

Table 8: Results of the comparison of RT. $\neq$ indicates a significant dissimilarity; $=$ indicates similarity; 95% confidence interval.

7. CONCLUSION

An evaluation of off-center sound degradation in surround loudspeaker setups for various multichannel microphone techniques was presented, whereby the spatial impression at different off-center listening positions is captured through binaural recording processes. The results obtained by the listening tests are generally consistent with the theory of spatial sound reproduction using a limited number of loudspeakers in that we did demonstrate an increases in sound degradation at off-center positions. It can be stated that the recording technique is an integrated part of the surround sound reproduction process and the tested microphone techniques affect the strength of the sound degradation at off-center positions as well as the size of the spatial area where the reference sound is equally perceivable. It is hard to generalize the results due to the strong effect of the musical excerpts on the results. This is somehow surprising and future work including the interpretation of the data with an auditory model will hopefully clarify this influence.

8. ACKNOWLEDGMENT

This work has been funded by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Banff Center for the Arts. Special thanks to Günther Theile and Sungyoung Kim for providing the multichannel recordings which were used for creating the binaural stimuli. The authors would also like to thank Theresa Leonard, William L. Martens, Bruno L. Giordano, and Georgios Marentakis for fruitful advice and assistance in this project.

Bibliography

[1] W. Ahnert and F. Steffen. *Sound Reinforcement Engineering: Fundamentals and Practice*. Spon Press (UK), 1999.

[2] J. Berg and F. Rumsey. Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques. In *112th AES Convention, Preprint 5593*, Munich, Germany, 2002.

[3] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, Mass., rev. edition, 1997.

[4] A. D. Blumlein. U.K. patent 394,325, Reprinted in Stereophonic Techniques, 1986, AES, New York, 1931.

[5] F. Camerer and C. Sodl. Classical Music in Radio and TV–a multichannel challenge; The IRT/ORF Surround Listening Test. *http://www.hauptmikrofon.de/orf.htm*, 2001.

[6] I. E. Commission. *Draft IEC 60268: Sound System Equipment —- Part 13: Listening Tests on Loudspeakers*. International Electrotechnical Commission, Geneva, Switzerland, 1997.

[7] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *108th AES Convention, Preprint 5093*, Paris, France, 2000.

[8] M. A. Gerzon and G. J. Barton. Ambisonic Decoders for HDTV. In *92nd AES Convention, Preprint 3345*, Vienna, Austria, 1992.

[9] D. Griesinger. The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique. *AES 19th Intern. Conference*, pages 182–200, 2001.

[10] T. Halmrast. Orchestral Timbre: Comb-Filter Coloration from Reflections. *Journal of Sound Vibration*, 232:53–69, Apr. 2000.

[11] K. Hamasaki. Multichannel recording techniques for reproducing adequate spatial impressionsion. In *Proceedings of the Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality*, Banff, Canada, 2003.

[12] International Telecommunication Union. *ITU-R BS.775-1, Multichannel stereophonic sound system with and without accompanying picture*, volume 1994. International Telecommunication Union, Geneva, Switzerland, 1992.

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

[14] R. Jacques, M. Fleischer, S. Fuhrmann, B. Steglich, U. Reiter, and H. Kutschbach. Empirischer Vergleich von Mikrofonierungsverfahren für 5.0 Surround [Empirical Comparison of Microphone Setups for 5.0 Surround]. In *22. Tonmeistertagung VDT International Audio Convention*, Hannover, Germany, 2002.

[15] R. Kassier, H.-K. Lee, T. Brookes, and F. Rumsey. An informal comparison between surround-sound microphone techniques. In *Convention of the AudioEng. Soc. 118, Preprint 6429*, Barcelona, Spain, 2005.

[16] S. Kim, M. de Francisco, K. Walker, A. Marui, and W. L. Martens. An examination of the influence of musical selection on listener preferences for multichannel microphone technique. In *AES 28th international conference*, Piteå, Sweden, 2006.

[17] H. Kuttruff. *Room Acoustics*. Spon Press (UK), 4th edition, 2000.

[18] C. Landone and M. Sandler. Applications of binaural processing to surround sound reproduction in large spaces. *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The*

2000 IEEE International Symposium on, 3:217–220 vol.3, 2000.

[19] C. Landone and M. Sandler. Surround sound impact over large areas. In *Convention of the AudioEng. Soc. 110, Preprint 5373*, Amsterdam, The Netherlands, 2001.

[20] R. Y. Litovskya, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106(4):1633 – 1654, 1999.

[21] G. Martin. *Introduction to Sound Recording*. available at http://www.tonmeister.ca/main/textbook, online version Oct. 2006.

[22] H. Møller, D. Hammershøi, C. Jensen, and M. Sørensen. Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc.*, 47(3):83–100, 1999.

[23] M. Morimoto. The Relation Between Spatial Impression and the Precedence Effect. In *Proceedings of International Conference on Auditory Display (ICAD)*, pages 297–306, Kyoto, Japan, 2002.

[24] M. Poletti. A unified theory of horizontal holographic sound systems. *J. Acoust. Soc. Am.*, 48(12):1155 – 1182, 2000.

[25] F. Rumsey. *Spatial Audio*. Focal Press, 2001.

[26] B. Smith. PsiExp: An environment for psychoacoustic experimentation using the IRCAM Musical Workstation. In *Proceedings of the Society for Music Perception and Cognition Conference*, pages 83–84, 1995.

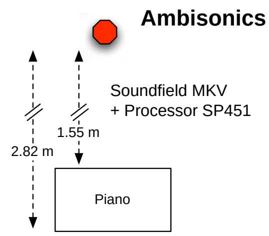
[27] G. Spikofski and M. Fruhmann. Optimization of Binaural Room Scanning (BRS): Considering interindividual HRTF-characteristics. In *AES 19th Intern. Conference*, pages 124–135, 2001.

[28] G. Theile. Multichannel Natural Music Recording Based on Psychoacoustic Principles. In *extended version of the paper presented at AES 19th International Conference*, 2001.

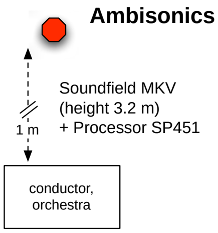
[29] B. Wagener. Räumliche Verteilung der Hörrichtungen in synthetischen Schallfeldern. *Acustica*, 25:203 – 219, 1971.

1. MULTICHANNEL MICROPHONE ARRAYS SET-UPS USED TO RECORD THE MUSICAL EXCERPTS

1.1. Coincident Microphone Technique: Ambisonic



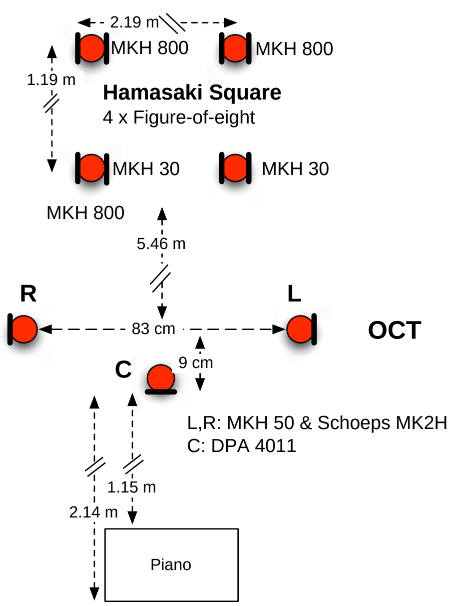
(a) Ambisonics recording setup, used for EXC 1



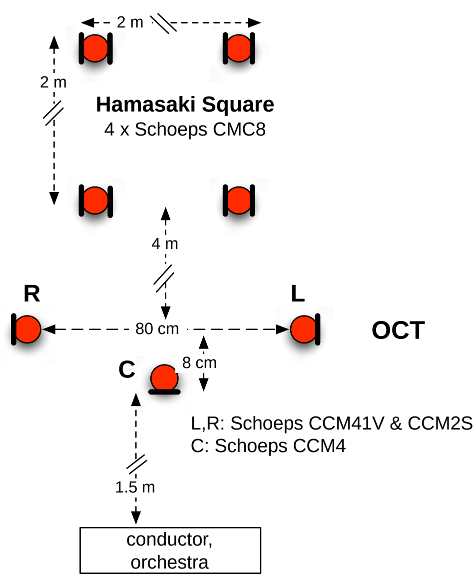
(b) Ambisonics recording setup, used for EXC 2

Fig. 17: Ambisonics Microphone Soundfield MKV

1.2. Spaced Cardioid Microphone Technique: OCT+Hamasaki Square



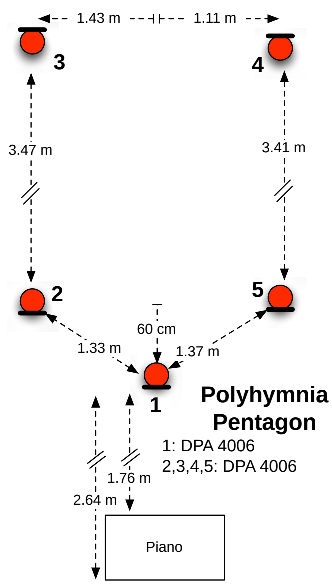
(a) OCT+Hamasaki Square, used for EXC 1



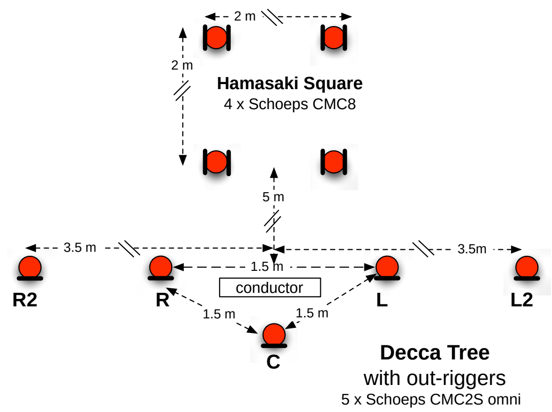
(b) OCT+Hamasaki Square, used for EXC 2

Fig. 18: Optimized Cardioid Triangle

1.3. Spaced Omni Microphone Technique



(a) Polyhymnia Pentagon, used for EXC 1



(b) Decca Tree, used for EXC 2, at 4m height above the conductor with the center mic down 40cm

Fig. 19: Spaced Omni Microphones