Determining the area of the sweet spot in a surround loudspeaker setup for various microphone techniques

Nils Peters

April 4, 2007

Abstract

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 the arrays are targeted to produce an accurate spatial impression at the sweet spot, the listening area of a speaker arrangement in which the spatial cues are reproduced with sufficient accuracy. The aim of this investigation is to determine how different microphone techniques affect the size of the sweet spot. In particular, the common belief that spaced techniques lead to larger sweet spot areas than coincidence and near-coincident techniques is investigated. For this purpose recordings which was made with 3 different microphone techniques simultaneously are presented through a surround loudspeaker set-up in two different large rooms. Binaural recordings were produced to capture sound fields at different positions inside the listening area. In a psychoacoustic experiment, listeners were asked to rate these binaural recordings according to the spatial impression provided by different recording techniques. The results of the listening test are also compared to the data of an binaural auditory model simulation.

1 Introduction

blah blah blah

1.1 Sweet spot

The sweet spot can be defined as the spatial area where the listener maintains the desired perception of virtual sound sources. Beside the primary demand of a stable sound localization this definition also includes a plausible sound reproduction in terms of timbre or source width. The quality of creating spaciousness and listener envelopment is also important for multichannel sound setups.
IEC 60268 defines 3 factors under the heading ”overall spatial quality” which correlating strong with the meaning of the sweet spot:

- *Image localization*: perceived spatial location of a reproduced sound source. The image may be well defined or blurred.
- *Image stability*: 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.

- *Width homogeneity*: stereophonic image should be distributed uniformly between loudspeakers.
  "The sweet spot [...] is the area in which listeners get a good and consistent impression of the directions and distances to individual sources of a mix. The sweet spot is also the reference point of the loudspeaker arrangement." [Lund, 2000]

1.2 Recent studies

Several studies were undertaken to evaluate the perceptual differences in between different surround microphone techniques. In [Pulkki and Hirvonen, 2005] the directional quality of different reproduction techniques for multi-loudspeaker setups were estimated using a binaural auditory model in the best listening position. The tested concepts were 1st and 2nd order Ambisonic, spaced omni microphones and pair-wise panning in a ITU 5.0 and a 8 channel uniformly distributed speaker setup. The simulated results were verified with psychoacoustical listening tests, in which the subjects adjusted an auditory pointer emanating broad-band noise to the same direction as their perception of the virtual source containing-band noise at five different frequencies. The main focus of all these studies lies on the the listening position at the sweet spot of the speaker system amd not much effort were done to evaluate reproduction techniques for listening position away from that point like this study does. The following studies were taken off-center listening position into account:
- [Landone and Sandler, 1999]
- [Griesinger, 2001]

2 Hypothesis

A spaced microphone recording technique produces a larger sweet spot compared to a sound reproduction based on a coincident microphone or near coincident recording technique.

3 Experimental Conditions

3.1 Music Material

Two sets of 5.0 multichannel recordings of different kinds of sounding bodies, a piano and a sinfonie performance, were used. Each musical performances were recorded with different multichannel microphone techniques simultaneous to make sure that the musical quality of the performance remains consistent. The recording techniques are more described in section 3.2 of this report.

Piano performance Recordings of a Piano solo performances which were performed by Thomas Plaunt and recorded in Pollack Hall, McGill University Montreal in 2006 were used. The recording procedure is described in detail in [Kim et al., 2006].

- J.S. Bach: Variation 13, Goldberg Variationen (BWV 988),
- F. Schubert, 1. Allegro ma non troppo, Sonata in A minor (D537)
- J. Brahms , "Ballade in d minor" (op. 10 no. 1)

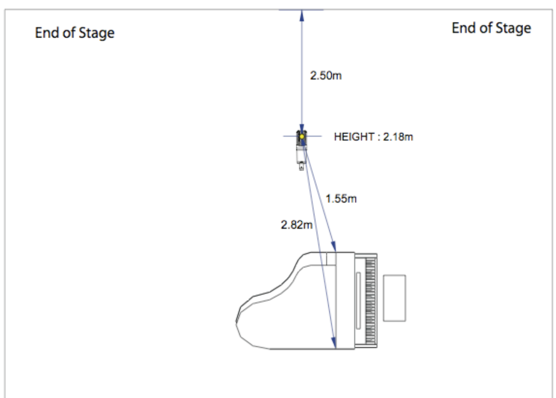
Orchestra performance Extracts of multichannel recordings of a Mozart sinfonie ("Maurische Trauermusik" c-minor (KV 477)) were taken [Camerer and Sodl, 2001]. The sinfonie was performed by the RSO Vienna in Austrian Radios Grosser Sendesaal (large broadcast hall).
All together 6 differrent excerpts were carefully chosen to have a separated variety of different musical expressions. Each excerpt have a length of around 7 seconds to ease the signal analyzes of a binaural model and to shorten the listening test duration.

3.2 Multichannel Microphone Techniques

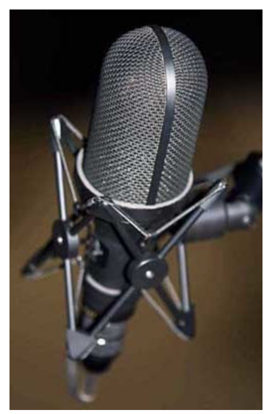
For this study, different multichannel microphone techniques were taken into account. The chosen techniques differ in their strategy to avoid interchannel crosstalk which is either done due to microphone directivity pattern and/or due to displacement of the microphones.

3.2.1 Coincident Microphone Technique - Ambisonics

Ambisonic is a recording technique which aims to capture a soundfield in a single point. Originally it goes back to [Blumlein, 1931] and extends Blumlein's ideas about coincident recording and reproduction techniques: By adding an omnidirectional microphone to the pair of figure eight units (for left-right respectively front-back), it can be shown (e.g. [Poletti, 2000]) that this setup captures all information that it 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 from the other two, also the up-down component of the soundfield is taken into account. The soundfield is encoded in terms of velocity and pressure components in the so called B-Format, 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 has to be suitable decoded according to the corresponding speaker setup. The most applied decoder to transform B-format into 5.1 is the so called *Vienna decoder* introduced in [Gerzon and Barton, 1992].



(a) Ambisonic recording setup, [Kim et al., 2006]



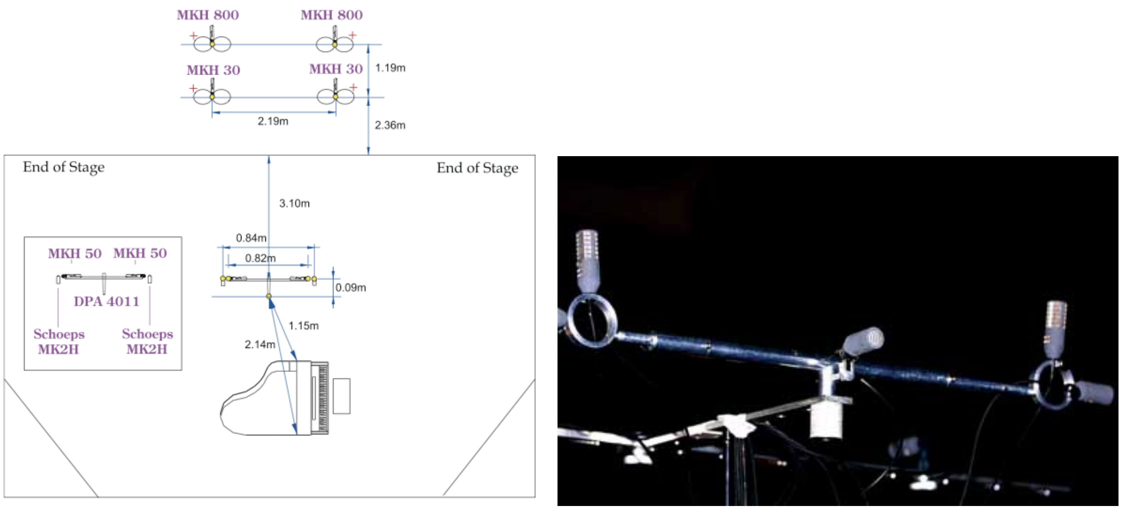
(b) Soundfield MKV

Figure 1: Ambisonic microphone Soundfield MKV

3.2.2 Near-Coincident Microphone Technique - OCT

The Optimized Cardioid Triangle (OCT) was first proposed in [Theile, 2001]. This technique is known to reduce crosstalk between channels by incorporating two outer hyper-cardioid microphones facing $\pm 90°$ side wards (see 2). The center microphone is a cardio capsular which is supposed to be placed 8 centimeter forward.

Optional low-passed 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 this frontal array is usually combines with several rear techniques such as: OCT surround, IRT cross, or Hamasaki Square [Hamasaki, 2003]. In all musical examples of this study a OCT and Hamasaki square combination was used.



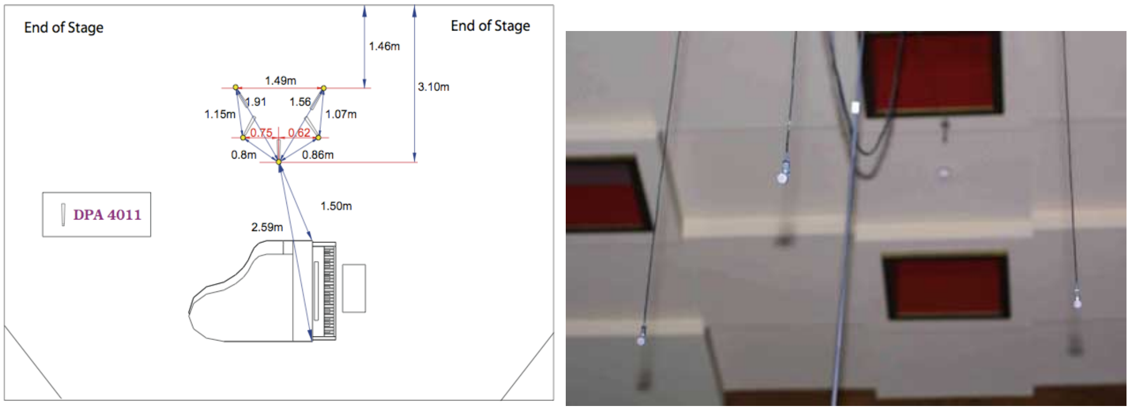
(a) OCT with Hamasaki Square, [Kim et al., 2006] (b) OCT with Hamasaki Square, [Camerer and Sodl, 2001]

Figure 2: Optimized Cardioid Triangle

3.2.3 Spaced Omni Microphone Technique

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 source like orchestra the system ca be extended with additional omni microphones to the side, the so-called ”outriggers”. 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 a IRT cross, or the Hamasaki Square, like it is done in the studied recordings of [Camerer and Sodl, 2001].

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. According to the IT BS.775-1 recommendation,the microphones are arranged in a large circle, whereby the position corresponds to the azimuthal angles of the speaker in the ITU BS.775-1 recommendation [BS., 1992]. The Polyhymnia Pentagon was applied in the recordings by [Kim et al., 2006].



(a) Polyhymnia Pentagon, [Kim et al., 2006]

(b) Decca Tree, used in [Camerer and Sodl, 2001]

Figure 3: Spaced Omni Microphones

4 Experiment A - Tanna Schulich Hall

4.1 Preparation of the stimuli

The Stimuli for the listening experiment were recorded in Tanna Schulich Hall, McGill University Montreal. The Tanna Hall has 121 seats and a reverberation time (t_{60}) of ca. 2.9 sec. It was built in 2005. To reproduce the multichannel recordings introduced in section 3.1 the installed 5 channel full range loudspeaker system of the hall was used (Kling & Freitag CA 1515 for L,C & R, two Kling & Freitag CA 1001 for surrounds) Two days before recording the stimuli the loudspeaker system was re-calibrated by the technicians of the hall. The positions of the speakers differs from the ITU-BS 1116-1 standard [ITU, 1994] in terms of the azimuth angle and the recommended reverberation time. Moreover the seats ascends in the hall. Also the Center Speaker was noticeable elevated (see fig.4).

A B&K dummyhead with shoulder damping fabric was placed on 13 different Positions in the hall (see fig.5) The green marked seat in figure 5 is supposed to be in the sweet spot of the 5 channel loudseaker system. Thus fore every reference stimuli there are 12 recordings beside the reference point. The measurements were focused to one half of the hall - due to the symmetrical shape of the hall it can be assumed that there will be no significant differences in sound reproduction degradation for the ”mirrored” seats. There were 243 stimuli recorded at 48 kHz. (13 positions · 6 different musical performances · 3 recording techniques) An omnidirectional microphone (Earthworks QTC1) and an a microphone with a figure-of-8 directivity pointed to the lateral sides (Sennheise MKH 30) were placed above the dummyhead to measure the lateral energy in relation the the overall energy (see figure 4(b)). The SPL during that recordings varied between 74 dBA and 77 dBA depending on the position in the hall. The measured SNR was ca. 50 dBA. Additionally we measured the impulse responses of each loudspeaker at every position with the sine swept technique [Farina, 2000].

4.2 Listening experiment

Subjects 20 subjects with normal hearing participated in the experiment. They were expert listeners of both gender, either studying or working in the field of acoustics. They have experience in listening to sound in a critical way. The participants were students from sound recording programs at McGill University Montreal as well as students and stuff from the sound recording work-study program of the Banff Centre, Alberta. Their ages ranged from 23 to 44 (M=28) and they work from 1 to 23 years (M=7.4) in audio recording.

(a) Loudspeaker setup with elevated center speaker

(b) Dummy head

Figure 4: Recordings in Tanna Schulich Hall

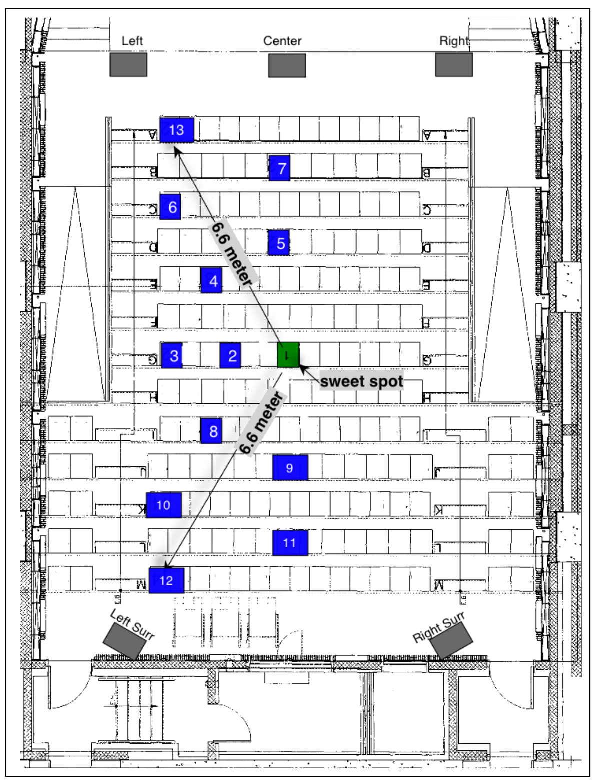


Figure 5: tested positions with the dummy head

In average the subjects are used to use headphones 1.7 hours per day. The subjects were tested at two sites: a) McGill University, Music Technology Area, Music Perception and Cognition Lab or b) facilities of the Banff Centre. All subjects were tested under similar conditions. Each subject completed a questionnaire concerning the amount and kind of musical practice, headphone listening and sound recording experience.

Stimuli The independent variables for that experiment can be seen in table 1 and yield to 84 experimental conditions. As can be seen from that table, the number of musical excerpt has to be reduced to two excerpts due to decrease the duration for this experiment. Each excerpts has a duration of ca. 7 sec. and were taken from:

Independent Variable	Rank
Musical Excerpt	2
Recording Techniques	3
Positions in Tanna Hall	14

Table 1: Independent variables for experiment A

- J.S. Bach: Variation 13, Goldberg Variationen (BWV 988),
- W.A. Mozart: "maurische Trauermusik" c-minor (KV 477)

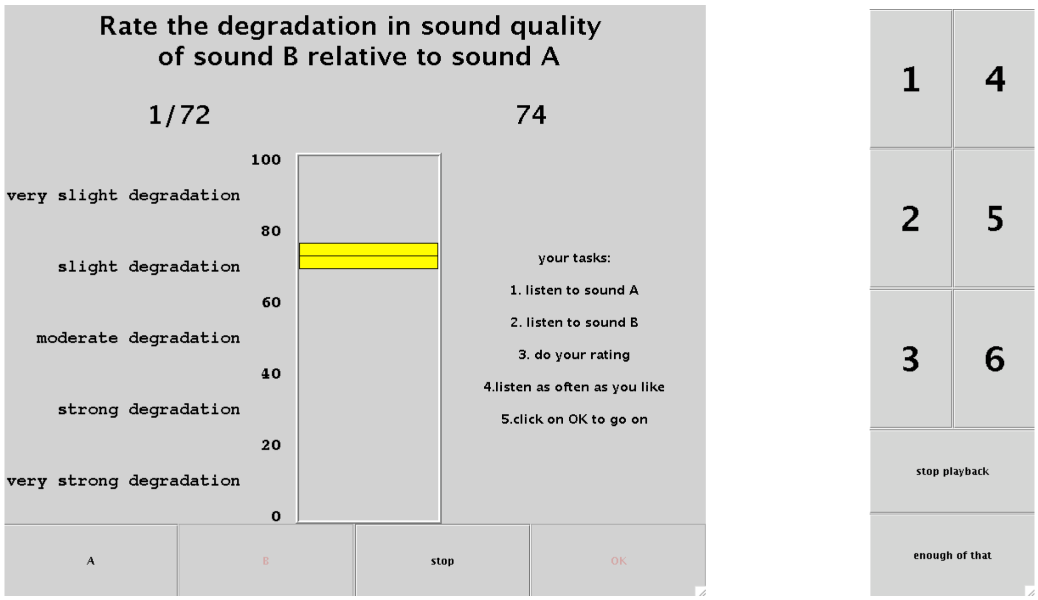
Procedure For the listening experiment a computer aided listening test was designed using PsyExp (see figure 6). The experiment consisted of a training phase, a familiarization phase and the experimental phase. The subject read the experimental instructions (see appendix) and asked any questions necessary for clarification. For the training phase the subjects could learn on 5 training trails how to use the user interface which will be presented in the experimental phase. The subjects were told to fulfill the following task:

"Rate the degradation in sound quality of the sound B relative to sound A."

Sound A represents the reference sound recorded in the sweetspot, while sound B is one of the following stimuli:

- One of the 12 recorded stimuli outside the sweetspot
- The same sound than sound A - a hidden reference
- A monaural processed version of the seat A12 - a hidden anchor of the worst quality (see fig. 5)

The purpose of the hidden reference and anchor is the ability to evaluate the subject's ability to detect small impairments. 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 trials were presented in random order (double blind single stimulus test). Within the presented pair the listener could switch freely from reference to the stimuli back and forth. In the training phase the played pairs of sounds where musical excerpts independent from these which are presented in the exerimental phase. Furthermore the subjects were informed that that these ratings would not be included in the analysis.
After the training phase a representative collection of 5 samples of each group (3 recording techniques $\cdot$ 2 musical excerpts) were randomized presented to familiarize the subject with the stimuli (see user interface in figure 6(b)). The subjects were asked to use the familiarization phase to be able to use the full scale in making their judgments in the following experimental phase.
The experimental phase has an effective length of approximately 45 minutes. To avoid bad data due to fatigue the participants were free to take a break whenever they wanted to. Every participants has executed that experiment twice to increase the reliability. To keep a realistic impression of the binaural recordings the subjects were told to look in frontal direction and not to move their heads either while they are listening to the sound samples. The stimuli were presented with Sennheiser HD 600 Headphones at normal listening level in a sound proof room (McGill) respectively an acoustical treated audio edit suite (Banff Centre).



(a) Experimental Phase Interface

(b) Listening Phase Interface

Figure 6: PsyExp user intefaces

5 Experiment B - Telus Studio

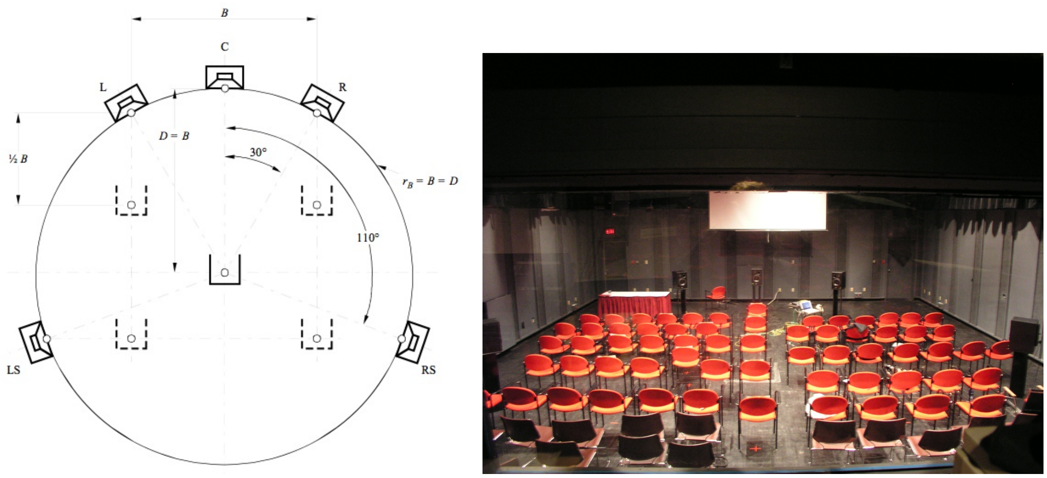
5.1 Preparation of the stimuli

A similar experiment was prepared under the same conditions in a second listening experiment. The purpose of that experiment is to see if the results of the first experiment are also valid in another listening environment. Compared to the recordings of experiment A, the following conditions varied:

Room Size The new set of stimuli were recorded in *Telus Studio* at the Banff Centre in Banff, Alberta. This Studio, which is usually used as a recording room for medium large ensembles or as a film set, has a floorspace of ca. $140m^2$ and a Volume of ca. $800m^3$. The estimated reverberation time t_{60} is about 600 ms (A-weighted). The measured SNR was ca. 45 dBA. These values fits to the ITU.BS 1116-1 recommendation [ITU, 1994] for multichannel loudspeaker setups in (large) listening rooms[??? check again with Berg-book].

Loudspeaker layout This ITU.BS 1116-1 recommendation is the common guideline to set up a five speaker system in the meaning of having the speaker at the positions: Left, Center, Right, Left Surround, and Right Surround. The five speakers were placed like it is shown in figure 7. The speaker lying on a circle which has a radius of $B = 4.2$ meter. The ITU recommendation indicates also the limits of the listening area in that setup (see dashed seats in figure 7). According to that suggestion the position for the binaural recordings were chosen as it is shown in figure 9.

Loudspeaker Five self-powered two-way, Dynaudio BM15A loudspeakers were used.



(a) Loudspeaker setup recomended in ITU BS-1116-1 (b) Telus Studio, Banff Centre

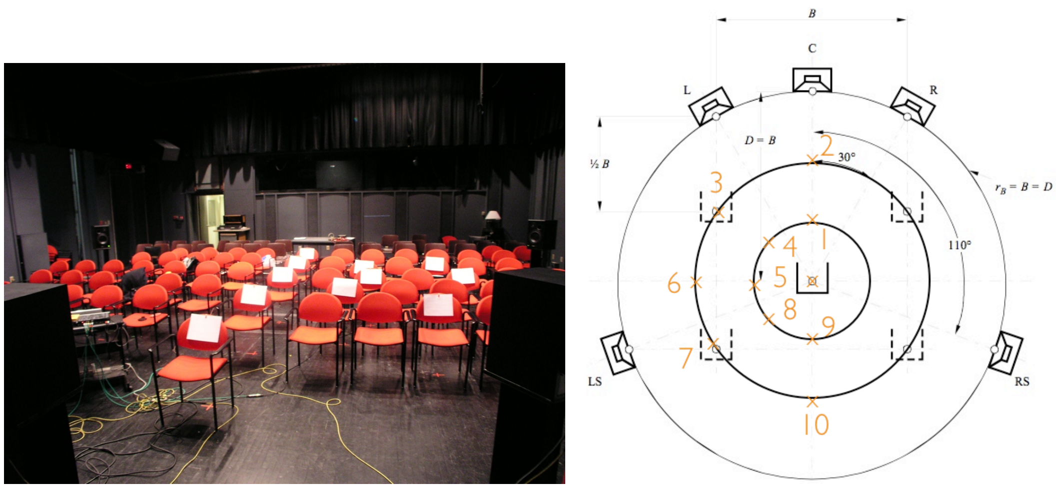
Figure 7: Loudspeaker setup

(Dummy) head For that recording the B&K dummy head was replaced by the head of the author: DPA 4060 omnidirectional miniature microphones were placed at the beginning of the author's ear channels to record the stimuli (see figure 8). It has to be mentioned that the dummy head microphones and the microphones in that experiment are from the same high quality and even produced by the same company. To avoid uncontrolled head movements which might cause artifacts during the listening experiment, a neck-wrist was used to tight the neck.

(a) r (b)

Figure 8: Original Head Recording

The tested positions are marked in figure 9(b). All together there were 198 stimuli recorded at 48 kHz. (11 positions · 6 different musical performances · 3 recording techniques). In addition for every listening position the impulse responses from each of the 5 loudspeaker were measured. The SPL of the reference listening position was calibrated calibrated to have the same level than the SPL of the reference position in experiment A. Again, due to the symmetrical conditions of the speaker layout in the room only one half of the listening area needed to be tested.



(a) tested positions marked with a white paper

(b) tested positions in the ITU sketch

Figure 9: Recordings in Tanna Schulich Hall

5.2 Listening experiment

10 subjects which also took part in the first experiment participated in this study. The age of this population varies in between 24 and 44 (M=30) and the work experience is in between 1 and 23 years (M=9). The experimental design and the experimental conditions were exactly the same than in the first listening experiment, just the stimuli were changed. The loudness of the reference signals were adjusted to the loudness of the reference signals of the first experiment by the author.

Independent Variable	Variantion
Musical Excerpt	2
Recording Technique	3
Position in Tanna Hall	12

Table 2: Independent Variables for Experiment B

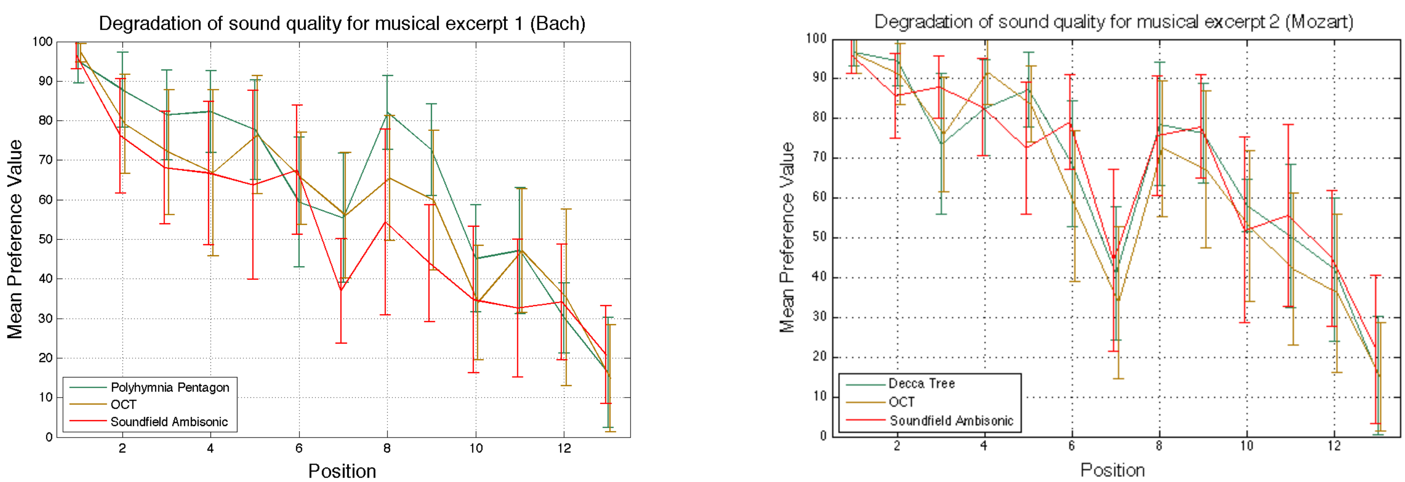
6 Results

Outliers A given Rating was considered as an outlier if its value was further than twice the standard deviation away from the mean value for that particular stimuli. In this case the outlier was replaced by that mean value. 57 ratings from Experiment A were detected as outliers which in 3.8% of all given ratings. In experiment B 23 ratings (1.6%) were considered.

6.1 Experiment A

Variable	Within-Subjects Effects	Sig.	Measure of effect size η_P^2 all positions	Measure of effect size η_P^2 for pos. 2-5,8,9
musical excerpts (exc)	$F(1, 18) = 18.253$	$p < .001$	.503	.537
recording technique (rt)	$F(2, 36) = 11.441$	$p < .001$	.389	.628
listening position (pos)*	$F(3.937, 7.873) = 156.676$	$p < .001$	.897	.538
exc · rt	$F(2, 36) = 34.657$	$p < .001$	.658	.567
exc · pos	$F(12, 216) = 8.834$	$p < .001$	.329	.116
rt · pos	$F(24, 432) = 6.034$	$p < .001$	.251	.190
exc · rt · pos	$F(24, 432) = 6.342$	$p < .001$	.261	.273

Table 3: Results of robust ANOVA; * indicates Greenhouse Geisser Correction



(a) (b)

Figure 10: Mean Results of Experiment A, corresponding to the seat coding shown in figure 5, the hidden reference is Nr. 1, whereby the hidden anchor is Nr. 13; errorbars: 95% confidence intervals for robust correction.

6.2 Experiment B

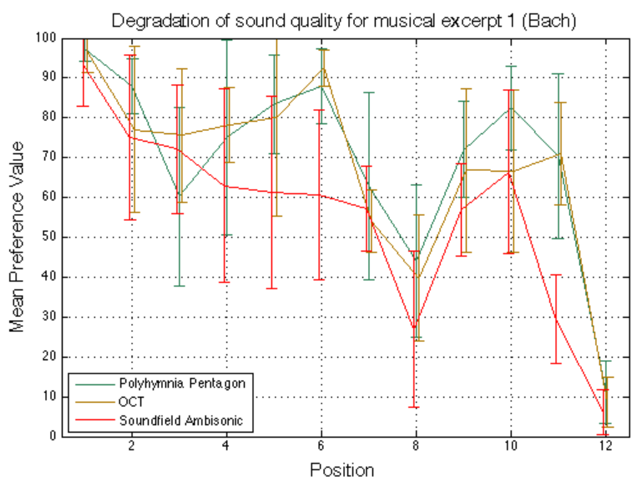
7 Verbal explanation

The participants were asked to explain their strategy how they have done the rating task. The following explanations were given:

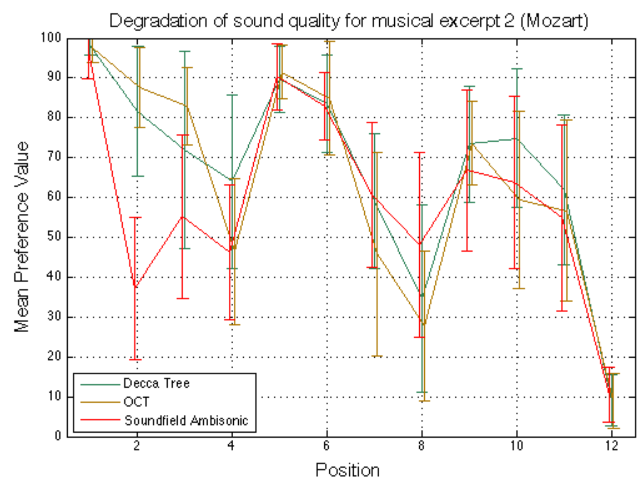
- dummy
- dummy

Variable	Within-Subjects Effects	Sig.	Measure of effect size η_P^2
musical excerpts (exc)	$F(1,9) = .112$	.745	.012
recording technique (rt)*	$F(1.286, 11.576) = 37.834$	$p < .001$	.808
listening position (pos)	$F(11,99) = 78.383$	$p < .001$	.897
exc · rt	$F(2,18) = 5.274$	.0016	.370
exc · pos	$F(11,99) = 6.545$	$p < .001$	.421
rt · pos	$F(22,198) = 6.471$	$p < .001$	.418
exc · rt · pos	$F(22,198) = 5.862$	$p < .001$	.394

Table 4: Results of robust ANOVA; * indicates Greenhouse Geisser Correction



(a)



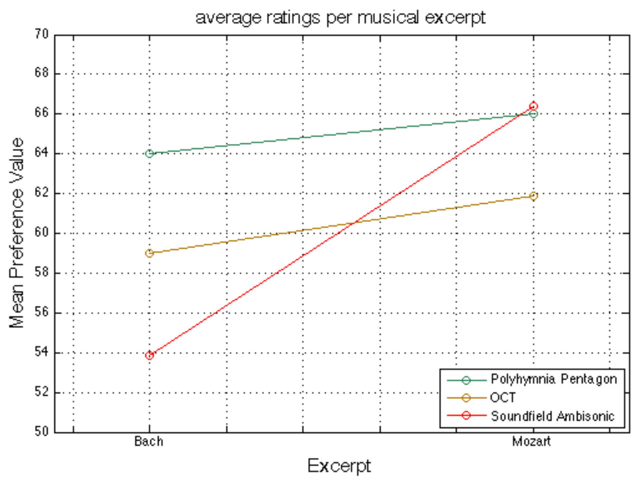
(b)

Figure 11: Mean Results of Experiment B, corresponding to the seat coding shown in figure 9; errorbars: 95% confidence intervals for robust correction.

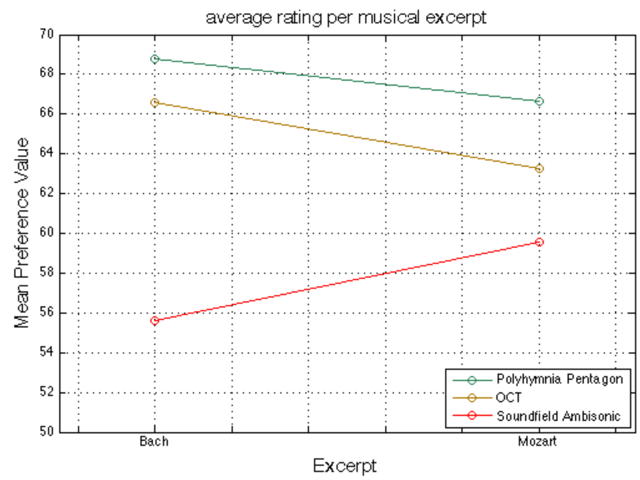
- dummy

From the results of the listening experiment we extract following conclusion:

- The overall ratings for spaced omni microphone techniques are higher than for near-coincident and coincident microphone techniques
- According to figure 12 , Spaced Omni and OCT following the same tendency in terms of preference ratings for both tested halls
- Ratings of the Ambisonic recording differs in between the musical excerpts and causing a moderate effect in experiment A



(a) Tana Schulich Hall



(b) Telus Studio

Figure 12: Interaction betweeen prefered recording technique and musical excerpts

8 Analyzation of the binaural stimuli

In order to find reasonable explanation for the ratings of the listening experiment the presented stimuli are analyzed with a binaural model, [Braasch, 2005]. This part will be done in April.

9 Conclusion

Recording Technique	Compared Recording Technique	Tanna Schulich Hall all Pos. Sig.	Tanna Schulich Hall all Pos. Std. Error	Tanna Schulich Hall for pos. 2-5,8,9 Sig.	Tanna Schulich Hall for pos. 2-5,8,9 Std. Error	Telus Studio Sig.	Telus Studio Std. Error
1	2	<.01*	.876	<.01*	1.105	.015*	.749
	3	.002*	1.162	<.01*	1.209	<.01*	1.621
2	1	<.01*	.876	<.01*	1.105	.015*	.749
	3	1.0	1.242	.075	1.561	<.01*	1.271
3	1	.002*	1.162	<.01*	1.209	<.01*	1.621
	2	1.0	1.242	.075	1.561	<.01*	1.271

Table 5: Pairwise comparison of the microphone techniques (Bonferroni), * indicates significant dissimilarities between the techniques based on estimated marginal means, 95% confidence interval

References

A. D. Blumlein. U.k. patent 394,325, reprinted in stereophonic techniques, 1986, aes, new york, 1931.

Jonas Braasch. A binaural model to predict position and extension of spatial images created with standard sound recording techniques. In *119th AES Convention, preprint 6610*, New York, NY, USA, October 7 - 10 2005.

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

F. Camerer and C. Sodl. Classical Music in Radio and TV–a multichannel challenge. *InterBEE Tokyo, Nov*, 2001. URL http://www.hauptmikrofon.de/orf.htm.

Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *108th AES Convention, preprint 5093*, Paris, France, 2000.

Michael A. Gerzon and Geoffrey J. Barton. Ambisonic decoders for hdtv. In *92nd AES Convention, preprint 3345*, Vienna, Austria, 1992.

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.

Kimio Hamasaki. Multichannel recording techniques for reproducing adequate spatial impression sion. In *Proceedings of the Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality*, Banff, Canada, 2003.

ITU. BS 1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. *International Telecommunications Union, Geneva Swizerland*, 1997, 1994.

Sungyoung Kim, Martha de Francisco, Kent Walker, Atsushi Marui, and William 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.

Christian Landone and Mark Sandler. Issues in performance prediction of surround systems in sound reinforcement applications. In *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects, DAFx99*, Trondheim, Norway, 1999.

Thomas Lund. Enhanced localization in 5.1 production. In *109th AES Convention, preprint 5243*, Los Angeles, CA. USA, September 2000.

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

Ville Pulkki and Toni Hirvonen. Localization of virtual sources in multichannel audio reproduction. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, 13(1):105 – 119, 2005.

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