ON STUDYING AUDITORY DISTANCE PERCEPTION IN CONCERT HALLS WITH MULTICHANNEL AURALIZATIONS

Antti Kuusinen,
Aalto University School of Science
Dept. of Computer Science
Espoo, Finland
`antti.kuusinen@aalto.fi`

Tapio Lokki,
Aalto University School of Science
Dept. of Computer Science
Espoo, Finland
`tapio.lokki@aalto.fi`

ABSTRACT

Virtual acoustics and auralizations have been previously used to study the perceptual properties of concert hall acoustics in a descriptive profiling framework. The results have indicated that the apparent auditory distance to the orchestra might play a crucial role in enhancing the listening experience and the appraisal of hall acoustics. However, it is unknown how the acoustics of the hall influence auditory distance perception in such large spaces. Here, we present one step towards studying auditory distance perception in concert halls with virtual acoustics. The aims of this investigation were to evaluate the feasibility of the auralizations and the system to study perceived distances as well as to obtain first evidence on the effects of hall acoustics and the source materials to distance perception. Auralizations were made from measured spatial impulse responses in two concert halls at 14 and 22 meter distances from the center of a calibrated loudspeaker orchestra on stage. Anechoic source materials included symphonic music and pink noise as well as signals produced by concatenating random segments of anechoic instrument recordings. Forty naive test subjects were blindfolded before entering the listening room, where they verbally reported distances to sound sources in the auralizations. Despite the large variance in distance judgments between the individuals, the reported distances were on average in the same range as the actual distances. The results show significant main effects of halls, distances and signals, but also some unexpected effects associated with the presentation order of the stimuli.

1. INTRODUCTION

Virtual acoustics and auralizations offer many possibilities to study the perceptual aspects of room acoustics. For example, it is possible to perform instantaneous side-by-side comparisons of different acoustic conditions; comparisons that are impossible to perform in real environments. Currently, there are a number of spatial sound technologies including ambisonics [1], wave-field synthesis [2], directional audio coding [3] and spatial decomposition method (SDM) [4] which can be used to recreate a measured or simulated sound field. Reproduction is realised either binaurally by headphones or by a multichannel loudspeaker setup in a (semi-)anechoic listening room. This article discusses a listening experiment on auditory distance perception (referred to with 'distance perception' or similar terms in the rest of the text) in concert halls with auralizations produced by SDM and reproduced in a multichannel setup.

Multichannel auralizations have been previously used in a number of formal listening experiments (e.g. [5]). Because the possible perceptual biases due to signal processing are the same in all stimuli, it is and it has been reasonable to assume that the perceptual differences between the auralizations of different halls and seating positions are representative of the differences between real conditions. Although the previous investigations together with the myriad of discussions with the test subjects and various experts who have listened to these auralizations do not give any reason to doubt the validity of this assumption, it is still formally unclear whether the auralized sources are actually perceived as being at the distance where they would be perceived in the real concert hall. Formal evidence on realistic perception of the distance to sound sources would greatly enhance the credibility of the results of any study - past and future - where these auralizations are being used as stimuli.

The motivation for the present study is twofold. On one hand, the work is motivated by the need to further test our system and auralizations with different anechoic source materials and especially with listeners who have little or none prior experience with spatial sound systems. Specific focus is on the very first perceptions of the auralizations, before any perceptual adaptation or calibration and/or learning has taken place. The first perception data - although susceptible to be more variable and less accurate - may be the most unbiased indication that the auditory perceptions of the sound sources in the auralizations are comparable to the perceptions of their real counterparts.

On the other hand, the previous studies [5, 6] on concert hall acoustics have indicated that “proximity” (or “intimacy” [7]), that is, the feeling of being close to the performers is one of the most important aspects of the listening experience in concert halls. This aspect is possibly linked to perceived distance, but due to the lack of substantial evidence on distance perception concerning large spaces, little is certain. Thus, we also seek to obtain preliminary evidence on the differences in perceived distances between concert halls before continuing with more detailed investigations in this respect.

2. BACKGROUND ON AUDITORY DISTANCE PERCEPTION

Considering distance perception to sound sources outside a few meters range from the listener, the main acoustic distance cues are intensity (or loudness), direct-to-reverberant energy ratio (DRR) and frequency spectrum [8, 9]. Intensity has been found to act as a relative cue, whereas DRR seems to act as an absolute distance cue [10, 11] - at least when a sound is perceived the first time. People might also use, or weight, cues differently with different signals. For instance it has been found that intensity cue is weighted more with speech whereas the distance to a noise source was determined more by DRR [12]. In the context of concert hall acoustics, the sound strength G is commonly used instead of sound intensity to measure the perceived loudness of the sound field. G is normalised by the source sound intensity at 10 meters in free field, and thus, reflects the contribution of the room. G is also used in this article instead of sound intensity.

Also the inter-aural level and -time differences (ILDs, ITDs) have been found to act as distance cues for sources near the listener. In larger spaces, such as concert halls, the effectiveness of these inter-aural cues is unclear, but they are related to inter-aural cross-correlation (IACC), which measures the similarity of incoming sounds between the two ears. IACC have been linked to various perceptual qualities of concert hall acoustics, such as, width and envelopment.

Familiarity to source characteristics has also been found to influence distance judgments [8]. In most cases listeners have some a priori (long term) knowledge about how the sound is perceived at different distances and/or at different output levels. Musical instruments are typical sources that produce sounds which vary systematically in spectral content with playing dynamics. Moreover, the changes in spectral information, for instance, the attenuation of higher frequencies due air absorption, has been found to serve as a relative distance cue, independent of the variation in overall sound level [13].

Finally, the relationship between the physical distances and the perceived distances is known to follow a power function in the form $p = kr^a$ where k is a linear scaling factor and a is the exponent indicating the amount of "compression" ($a < 1$) or "expansion" ($a > 1$) of the perceived distances (p) compared with real physical distances (r) [8]. Average values for the k have been reported being around 1.32 and for the exponent a around 0.54, but because most previous studies included only distances up to around 10 meters, it is unclear whether these values are also representative of larger spaces.

3. METHODS AND MATERIALS

3.1. Spatial room impulse response measurements

Acoustic measurements were made with a calibrated array of loudspeakers, i.e., a loudspeaker orchestra [14] as a sound source and an array of six omnidirectional microphones as a receiver. Measurements and the processing with SDM were performed separately per each loudspeaker source on stage. The room impulse responses were performed with the swept-sine technique [15]. Details of the loudspeaker orchestra, the measurement technique and the technical equipment have been described in previous publications [16, 17].

Spatial room impulse responses (SRIRs) were obtained with SDM [4] which extracts the spatio-temporal evolution of the sound field from the impulse responses captured by the microphone array. The estimation of the directions of the arriving sounds is based on a combination of time-of-arrival and time-difference-of-arrival estimates calculated from the six pressure values of the omnidirectional microphone sensors. In the current implementation, the analysis is carried out in a sliding temporal window of 2 ms, with a hop-size of one sample (99 % overlap). This window length has been chosen to be in-line with the current knowledge about temporal resolution of human hearing [18]. By considering the echo density [19] of these concert halls and the length of the time window, it is possible to approximate a time instant when it is probable that more than one reflection occurs in the analysis window. The time instants for present halls with volumes of 15000 m^3 (BK) and 16000 m^3 (SB) are 120 ms and 124 ms, respectively. If this approximation is valid, SDM will yield accurate estimates for the direct sound and the early reflections. When the echo density increases, the sound field becomes more diffuse and stochastic, and more and more reflections from different directions will be included in the analysis window. When this happens, SDM will still yield only one direction estimate per analysis window, but these estimates in overall behave more or less randomly depending on the properties of the sound field. Therefore, in a diffuse sound field, SDM yields stochastic (diffuse) estimates.

By employing a hop size of one sample, this spatio-temporal analysis gives each sample in the impulse response a direction, as well as a pressure value which is obtained from one of the omnidirectional microphone responses. It is noteworthy that, for the microphone array used here, the reported RMSE error of the estimated locations of loudspeaker sources in real concert hall was 2 degrees in direction and 0.4 meters in distance [20].

SRIRs containing the information on directions and pressures for each sample are translated to the reproduction loudspeakers in the listening room. In this study, the direction of each sample in the spatial impulse response is used to position that sample in the direction of the nearest (the smallest angle difference) loudspeaker in the listening room setup, thus, distributing the original spatial impulse response to the spatial configuration of the reproduction loudspeakers. One alternative to this direct panning technique is amplitude panning between the loudspeakers [21], but it has been found to decrease the spectral brightness in the auralizations of concert hall acoustics [17].

After translating the impulse responses to the reproduction loudspeakers, they are convolved with anechoic instrument recordings [22]. Naturally, the convolutions with anechoic recordings were made in respect to the instrument positions in a real orchestra although practically any arrangement is possible. Finally, the convolved signals of each instrument are summed per reproduction channel to produce final samples with 24 channels for playback. Max/MSP 6 software was used to playback the audio and to set the output at a comfortable listening level.

A noteworthy difference between these auralizations and in-situ listening of a real orchestra is that all the perceptual cues in auralizations are less dynamic than in reality as they are produced by stationary sound sources with directivities that differ from those of real instruments [14]. It is currently unknown how such dynamic aspects affect the perception of sound sources. It is also unclear how the distance cues from different sources are used together and weighted when the task is to evaluate the distance to multiple distributed sound sources, instead of just one single source.

3.2. Concert halls

Strong reflections from the side has been found to enhance musical dynamics [23], what seems to be one governing aspect differentiating traditional shoebox shaped halls from other designs, such as the fan shape. To obtain some evidence whether the shape has an effect also on perceived distances, this study included a shoebox shaped Konzerthaus in Berlin (seats = 1575, volume = 15000 m^3) and an irregular fan shaped Beethovenhalle in Stuttgart (seats = 2000, volume = 16000 m^3), both measured without an audience. Concert hall layout and cross section are illustrated in Fig. 1. On-site measured sound pressure A-levels with pink noise played back simultaneously from all the sources on stage were 84.4 dB (14 m to the center of the orchestra) and 82.6 dB (22 m) in BK, and 83 dB and 81.4 dB in SB. Taking these values at 14 meters, we calculated the theoretical value for 22 meters in free field conditions, and subtracted those from the measured SPLs. For both halls, the boost in SPLs at 22 meters was about 3 dB compared with the theoretical free field conditions.

The values for room acoustical parameters are tabulated in Table 1 and the values of G and DRR, which are the main distance cues in this context are also illustrated in Fig. 2.

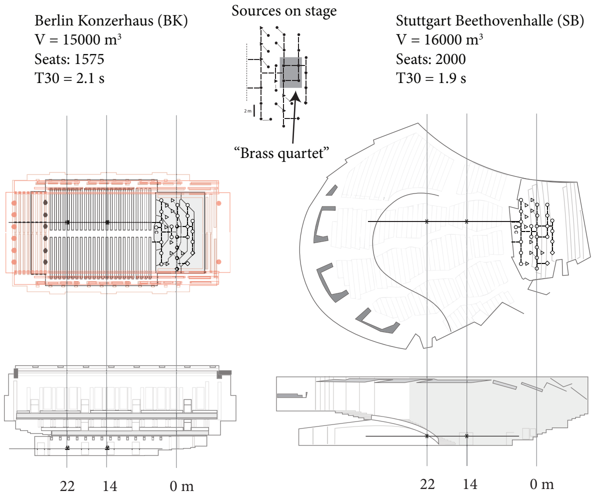

Figure 1: *Layouts and cross sections of the studied halls, and the positions of the sources in the loudspeaker orchestra on stage.*

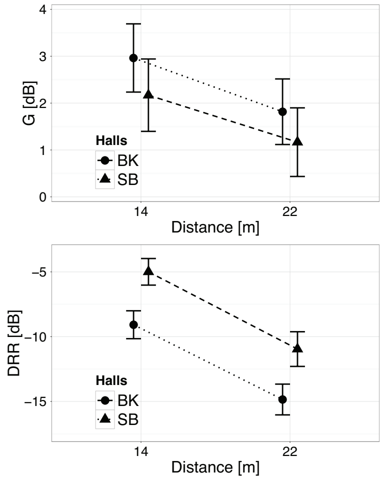

Figure 2: *G and DRR values in each hall over the distances used in the study. Values are averaged over the 500 Hz and 1 kHz octave bands. Bars repsresent the 95 % confidence intervals calculated over the 24 sources and the two frequency bands.*

3.3. Listening room

Listening room meets the ITU BS.1116 recommendation [24], with the exception that only the noise rating (NR) 30 is met, whereas NR 15 is recommended. Also the listening distance from the reproduction loudspeakers is less than the recommended two meters, about 1.5 meters on average. 24 reproduction loudspeakers are positioned around the listening position in 3-D layout. There are a few more loudspeakers in the frontal direction to account for the fact that human sound localisation accuracy is the greatest in the front. Details of the loudspeaker setup have been provided by Haapaniemi and Lokki [16]. The listening room and the loudspeaker layout are also illustrated in Fig. 3, where ITD contours are depicted in the figure for illustration purposes.

Because the listening room and reproduction setup might affect the relative loudness differences in the auralizations compared with real halls, we made additional auralizations with the same pink noise used in on-site measurements and measured the difference in SPLs between these auralizations and the on-site measurements. For on-site measurements, the differences between 14 and 22 meters were 1.8 dB and 1.6 dB for BK and SB, respectively. The same differences in the listening room measurements were 1.4 dB and 1.2 dB indicating the compression of 0.4 dB due to the listening room. Thus, the relative differences in real conditions might actually be a little greater than can be perceived in our listening room. In addition, the listening room itself is not anechoic but damped with absorbers and has a reverberation time of 0.2 seconds at mid-frequencies (averaged over 0.5-1 kHz frequency bands). Hence, this reverberation is confounded in sounds evaluated by the listeners.

For the listening, the test samples were set at a comfortable listening level with the average of approximately 78 to 80 dB across all samples. The actual sound levels were highly variable due to the nature and duration of the anechoic signals. The output level was determined by informal listening to not to overwhelm any listeners during the experiment. It is worth to mention that the focus here is in the relative differences between the conditions, without attempting to calibrate the listening levels to correspond to any 'true' references. We also considered to systematically randomise the output levels, but it was decided to be left out from this study.

3.4. Anechoic signals

ID	Descr.	Dur.	Sources	Configuration
M	Bruckner symphony	60 s	1 - 24	classic orc. config.
F	musical cacophony	48 s	1 - 24	classic orc. config.
B	stream of brass quartet	26 s	15, 16 17, 18	2 trumpets & 2 trombones
N	pink noise	5 s	1- 24	classic orc. config.

Table 2: Summary of the anechoic signals used in the listening experiment.

Table 2 summarises the types of anechoic signals used in the listening experiment. The source positions on stage are illustrated in Fig. 1. The signal abbreviated as 'M' is one minute excerpt

Hall	EDT (s)	DRR (dB)	G (dB)	C_{80} (dB)	J_{LF}	L_J (dB)	$IACC$
BK	2.1, 2.1	-4.0, -7.6	2.9, 1.8	-1.7, -2.3	0.25, 0.26	-2.6, -3.4	0.39, 0.31
SB	2.3, 2.1	-1.5, -5.4	2.1, 1.1	0.6, -1.2	0.11, 0.11	-5.5, -5.8	0.56, 0.48

Table 1: Values for acoustical parameters for the two distances D1 (14 m) and D2 (22 m) in each hall. Values are averaged over the 500 Hz and 1 kHz octave bands except for J_{LF}, which is averaged over the 125 Hz, 250 Hz, 500 Hz, and 1 kHz octave bands and L_J which is energy averaged over the 125 Hz, 250 Hz, 500 Hz, and 1 kHz octave bands.

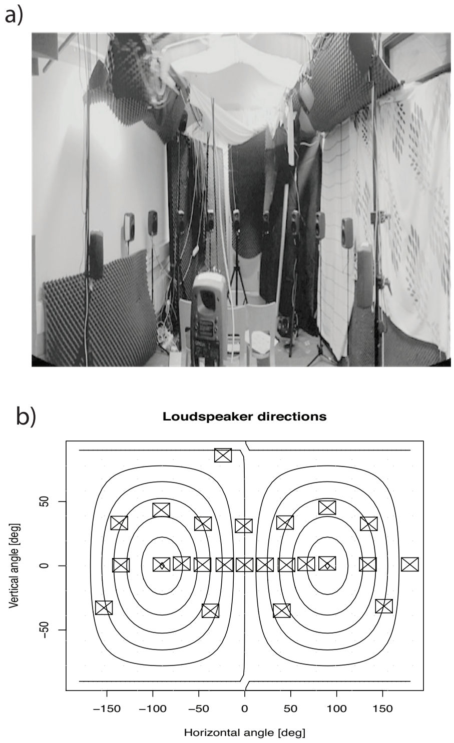

Figure 3: *a) Photo of the listening room. The listening position, i.e., the sweet spot is in the middle of the room. b) Illustration of the loudspeaker positions in the listening room. The ITD contours are drawn to better illustrate the directions of the loudspeakers in relation to the ears of the listeners.*

of Bruckner's 8th symphony, second movement. Shorter excerpts of this composition have been used in our previous investigations [26, 5], but here we wanted to provide test subjects with enough time for making their distance judgment.

In contrast to this excerpt of Bruckner's symphony, two signals ('F' and 'B') were constructed according to a procedure proposed by Kuusinen [27]. In brief, the anechoic recordings of excerpts of symphonies by Mozart, Beethoven, Mahler and Bruckner, where each individual instrument has been recorded separately, were first cut into variable length (short) segments. Then randomly selected segments (per each instrument) were concatenated to produce each instrument track. A certain amount (30 % of the total duration) of silence was required to be left in the individual tracks in order to reduce the cacophony observed in informal listening to various outputs of this procedure. Here, two different configurations of instruments was used. For the signal denoted as 'F', the configuration of the virtual orchestra approximated the configuration of a full classical orchestra. 'B' signal represents a brass quartet consisting of two trumpets and two trombones located in the middle of the stage, see Fig. 1. The reason to select brass instruments was that the directivity patterns of these instruments [28] are similar to those of the loudspeakers used as sources in the acoustic measurements.

Because a single randomisation procedure might yield an output which is far from an 'average' output, this procedure was run 100 times per each configuration. The acoustic features in these 100 samples were then analysed with MIRToolbox [29] in Matlab, using the *mirfeatures*-function, which extracts the averages of a number of acoustic features from the signals (see [29] for details). Next, this feature data was subjected to principal component analysis with varimax rotation on the first 5 components. The sample scores on these components were used to find the individual sample nearest to the center of the sample space in terms of Euclidean distances. The values of the acoustic features themselves are not investigated here, as they were only used as a means for sample selection.

The fourth sample type was a 5-second long pink noise, which was generated separately for each source to avoid unwanted additive effects when convolved with the spatial impulse responses.

3.5. Subjects

Forty subjects (12 women and 28 men in ages between 21 and 58 years) were recruited from people working in the department of Computer Science and from the building lobby. The recruitment process was facilitated by restricting the total duration of the experiment under 20 minutes, meaning that the listening itself was to be performed in 15 minutes, and that some compromises in the experimental design were to be made. Seven subjects reported having some prior experience in listening experiments, and four subjects had been in the listening room beforehand. The participants did not report any known hearing impairments when asked. This was the only screening criterium although it was not formally tested due to the time restrictions.

3.6. Listening experiment

Inspired by the experimental design used by Mershon and King [10], the experiment was based on an idea that each subject can give only one unbiased absolute distance estimate, that is, for the very first sound they hear in the experiment. All the other sounds are judged in relation to the first judgment and thus biased by the 'priming' or 'anchoring' effect of the first sound. Accordingly any type of training (e.g., sequences or otherwise) was also excluded from the experiment. Also, due to restricting the duration of the listening to 15 minutes, the repetitions of any of the stimuli were not included in the experiment.

Before entering the listening room, written and verbal instructions of the test procedure was presented to the test subjects. The main instructions were given as:

"Shortly, you will be blindfolded and guided to the listening room to listen to a set of sound samples. After each sample, your task is to verbally report the distance to the source/s of the sound in meters. If you hear multiple sources, report the distance to the center of the sources. You will hear each sound sample only once. Try to be as accurate as you can, but give the answer at least in one meter accuracy."

Then the subjects were blindfolded and guided to a chair at the listening position in the listening room. The experimenter stayed at the corner of the room to playback the sound samples from a laptop and to write down listener's verbal responses. The sound samples were presented one by one and after each sample the subject reported the distance to the sound sources in meters. Experimenter verbally repeated the reported distance to avoid any mistakes in the responses. An alternative to having the experimenter present in the room was to provide the listeners with a writing pad and a pen, as was done by Mershon and King [10], but this was considered unfeasible in practice without expanding the duration of the experiment.

The first part (i.e., first 4 sounds) of the experiment was a balanced between subjects block design with the one-minute long excerpt of Bruckner's symphony ('M'). The subjects were first divided into two groups of 20 people, according to the hall designated to the first two samples, that is, either BK or SB. In each group, half of the subjects (n = 10) were first presented with the closer sound sample and the other half with the further one. After the first four samples, the listening continued with the other three source materials ('F', 'B' and 'N') in random order. The presentation orders of the acoustic conditions within these signals were randomised.

4. RESULTS

The data consists of total of 640 observations, each being a combination of subject (N = 40), signal ('M','F', 'B' and 'N'), hall (SB and BK), and seating position (i.e., distances D1 and D2). Thus, there are 320 data points for each hall, 320 for each positions (abbr. pos), 160 for each signal and 40 for each combination of hall, distance and signal, referred to as 'sample' in the following. The 5 %, 25 %, 50 %, 75 %, 95 % percentiles for the whole data are 3.0, 7.4, 15.0, 28.0, and 50.0 meters, respectively. Fig. 4 a) and b) show histograms and normal quartile-quartile (QQ) residual plots for the data. Histogram illustrates that listeners often used the resolution of 5 meters instead of one meter which was given in the instructions. This behaviour is also seen as "steps" in the QQ-plot, which in addition shows that the data is approximately normal when the results are transformed to logarithmic coordinates. The geometric means for two distances across subjects and samples are 12.0 m for D1 and 14.8 m for D2 in BK and 13.2 m and 15.9 m in SB. Fig. 4 c) and d) shows the data per each subject and per each sample, repectively. Subjects have used very different scales in their responses, and few subjects reported some "wild" distance estimates (max = 240 m) in the case of pink noise.

Analysis of variance (anova) was carried out in log-transformed responses. Note that any formal test of normality will fail on this data set, because the data is effectively discrete (with resolution of 1 meter in untransformed coordinates). However, the analysis of variance and the corresponding F-tests, are known to be robust against the violations of the normality [30]. The main results were checked with non-parametric tests (Kruskal- Wallis) with the same results, but those are omitted here for brevity.

One main target was to investigate the distance judgements to the first presented stimuli. The first four stimuli were presented in a balanced between subjects block design, where the blocking factor was the presentation order of two halls. Anova results of these four samples are tabulated in Table 3. There is a significant main effect of position (F(1,38) = 11.8, $p < 0.01$) as well as an interaction effect between hall and the between subjects blocking factor (hall BK or hall SB first; F(1,38) = 11.5, $p < 0.01$)). The block effect is illustrated in Fig. 5 a) where each hall is judged closer/further depending on which one has been presented first. This indicates that the order of presentation had a major influence on the results.

Source	df	SS	MS	F	p(>F)
Between:					
block	1	0.34	0.34	0.11	0.70
residuals	38	123.3	3.24		
Within:					
hall	1	0.19	0.19	1.2	0.27
pos	1	0.71	0.71	11.8	**< 0.01**
hall*pos	1	0.02	0.02	0.24	0.62
block*hall	1	1.7	1.7	11.5	**< 0.01**
block*pos	1	0.01	0.01	0.25	0.62
block*hall*pos	1	0.02	0.02	0.32	0.57
residuals	38	2.34	0.06		

Table 3: Anova results regarding the first 4 samples. Significant effects ($p < 0.05$) are bold-faced.

Next, we included the rest of the data set in the analysis. The results of this repeated measures anova are tabulated in Table 4. Significant differences were found between the two positions, the two halls as well as between the four signals. There is also a significant interaction between the position and signals, meaning that the differences between the positions were influenced by the properties of the source material. The results are illustrated in Fig. 5 b).

Due to the significant interaction between the between subjects blocking factor and hall in the first anova, we investigated the order effect more closely. Within each signal, the samples were identified by their presentation order, and the factor 'order' was included in anova-model. The results revealed significant order effects for all source material except for the 'F', which is the random cacophony produced by "full orchestra". The order effects are illustrated in Fig. 5 c) and tabulated in Table 4.

5. DISCUSSION

It is noteworthy that the experiment was designed so that a single listening session could be performed in 15 minutes. On the one hand, this restriction facilitated the recruitment of 40 test subjects, but on the other hand, only a very limited number of stimuli could be included. Also, the stimuli were chosen to be long enough in duration in order to give listeners enough time to adapt to the sound field and to decide on their answer. As many as four different source materials were included in the experiment, what excluded

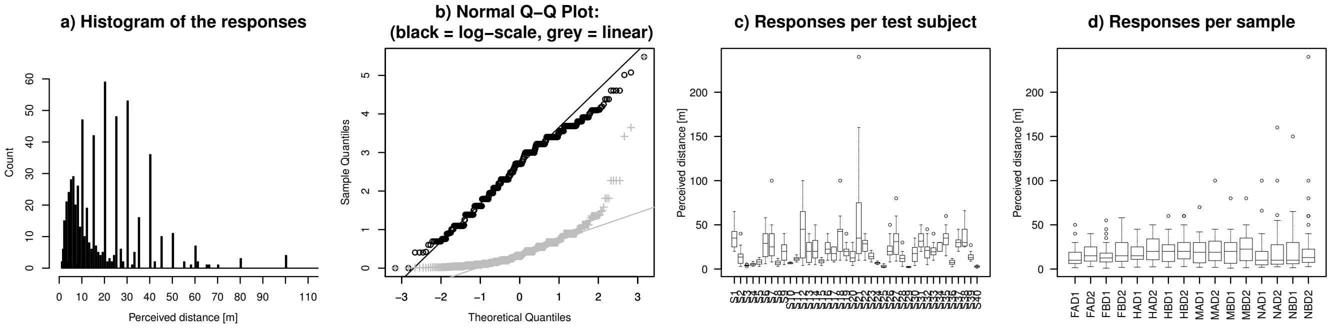

Figure 4: a) Histogram of the data. b) Normal quartile-quartile (QQ) plot of the residuals (grey: linear scale, black: logarithmic scale). c) Data per each test subject. d) Data per each combination of source material (F,H,M,N), hall (A or B) and distance (D1 (14 m) and D2 (22 m).

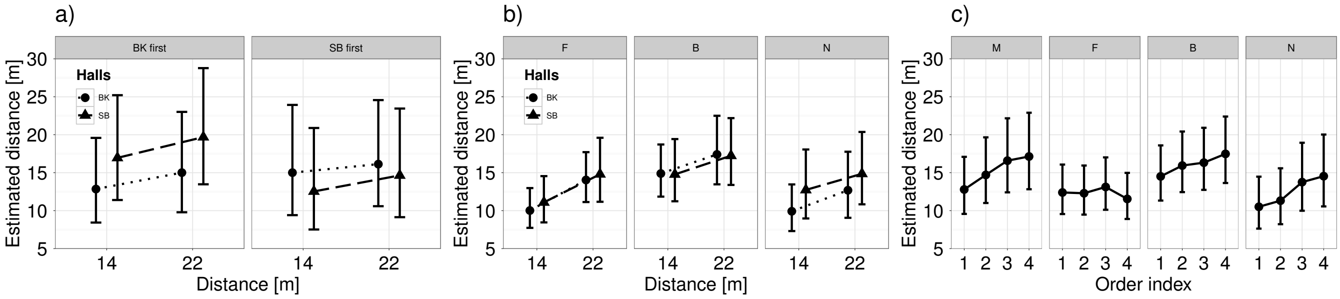

Figure 5: Results: a) The excerpt of Bruckner's symphony ('M') with a balanced between subjects design where half of the listeners listened to BK first (left) and half listened to SB first (right). Also within these groups, half of the subjects were first presented with the closer distance and the other half with the further distance. b) Other source materials. c) Order effect per each source material.

Source	df	SS	MS	F	p(>F)
Between:					
residuals	39	361.8	9.23	-	-
Within:					
pos	1	6.42	6.42	51.2	**< 0.01**
hall	1	1.16	1.16	7.40	**0.01**
signal	3	8.88	2.96	3.6	**0.02**
pos*signal	3	0.77	0.26	3.55	**0.02**
hall*signal	3	0.93	0.31	2.67	0.051
hall*pos	1	0.03	0.03	0.42	0.52
hall*pos*signal	3	0.10	0.03	0.44	0.73
residuals	585	139.5	0.24		
Order effects:					
sig. M	3	2.12	0.71	8.16	**< 0.01**
sig. F	3	0.37	0.12	1.3	0.28
sig. B	3	0.69	0.23	5.7	**< 0.01**
sig. N	3	2.5	0.83	7.2	**< 0.01**

Table 4: Results of overall analysis of variance. Significant effects ($p < 0.05$) are bold-faced.

collecting repeated judgments and thus, the assessment of the reliability of the distance judgments. Since we targeted the unbiased judgments when the sounds are listened to the very first time, it is possible that the results reflect the perceptions made during the period of perceptual calibration, adjustment and learning. This interpretation is supported by the observed order effect. However, when most of the perceptual studies include training sequences to the experimental design, it is sometimes worthwhile to experiment with alternative ways, because outside the laboratory such explicit training sequences hardly exist.

It is also noteworthy that the experimenter was present in the corner of the listening room - about 2.5 meters behind the listeners - verbally repeating the reported distances back to the listeners. Thus, the listeners in fact had an implicit reference from the experimenter's voice although this distance was never asked or told to the subjects. Nevertheless, for some people this reference may have anchored them to the dimensions of the listening room. There were few people who reported all distances to the range of the reproduction loudspeakers, what also speaks against the desired illusion of being in a larger space.

With these considerations in mind, the results showed that the majority of the listeners reported distances which were well beyond the dimensions of the listening room. Previous studies on distance perception have reported large inter-individual variances [8] and this study was not an exception. Especially the convolved pink noise was proven to be a difficult sample to judge as illustrated by the very large variance between the listeners. Nevertheless, in overall the judgement are well in the range of the actual physical distances. The perceptual distances were also underestimated which is in-line with the previous knowledge.

Considering the first four samples, there is a peculiar relationship between the measured DRR and strength parameter G, as illustrated in Fig. 2 and the distance estimates, as illustrated in Fig. 5a. When BK was presented first, the judgments follow G, and the higher DRR in SB seems to be counteracted by the lesser overall strength of the sound field. When SB was presented first, the judgments seem to follow DRR, so that the greater relative reverberant sound energy in BK (lower DRR) seems to make the sources sound further away, even though G is greater in BK. In other words, when the first two samples had overall loudness greater than the two subsequent samples, according to the values of G, the distance judgments were based on the loudness differences. When the later presented samples actually had greater G, they were still judged as being further away as if the relative judgments were based on differences in DRR and not on overall loudness.

However, because the results showed that the order of presentation influenced the judgments, as illustrated in Fig. 5c and that the subjects did not have any possibility to make repeated comparisons between the auralizations, we suspect that this peculiarity is related to the perceptual calibration to the task as well as to the stimuli. The order effect was observed for all signal types except for the signal which was described as "musical cacophony" ('F'-signal). Remembering that the experimental design consisted of the combination of balanced between subjects design (first four samples), and a randomised design within the other signals, the order effect seems to be robust but dependent on the properties of the source material. So, this order effect seems to explain the peculiar relationship between the distance judgments and DRR and G in the first four samples. Nevertheless, the reasons why the sources in later presented auralizations were judged as being further is an open question and deserves some speculation.

For the three signals which exhibit the order effect, it is possible that repeated exposure to the same musical source material allows for a gradual attentional shift from the musical features to the room acoustic cues initially overshadowed by the attentional (musical) targets in the signals, that is, by the musicality of the signals. By informal listening it was observed that 'F'-signal was actually sparser than the other signals providing separate unconnected sound elements which possibly allowed the listeners to directly focus on the room acoustics and the relevant auditory distance cues without being distracted by the musical features. Considering that this is the first study where this systematic stimulus production procedure [27] has been employed in a listening experiment, the 'F'-signal gave results which warrant for further investigations with this approach.

The main objective of this study was to evaluate the feasibility of these auralizations to study distance perception in concert halls and the results give a good premise to continue with more detailed experiments. Although the results did not show significant interaction effect between the hall and distance, it is possible that including more seating positions, that is, more variation in distance and collecting the distance judgments with a more rigorous and systematic testing procedure could reveal more on the effect of hall acoustics on auditory distance perception.

6. CONCLUSIONS

Spatial decomposition method (SDM) based multichannel auralizations were used to study auditory distance perception in concert halls with forty naive test subjects. There were large inter-individual differences in the reported distances, and many listeners used the resolution of five meters what reflects the inaccuracy of our auditory distance perception for longer distances. Still, the results indicated that these auralizations produced significantly different distance estimates between the two halls and two positions under evaluation. However, the results failed to show an interaction effect between distance and hall acoustics. Possible reason is that only two distances and two halls were included while a more systematic variation in distances and halls might reveal such an interaction. Overall, the present experiment yielded distance estimates which are reasonably close to the real distances indicating that these auralizations may well be used to study perceived distances in concert halls, but that the following studies should use a more rigorous experimental approach.

Finally, the results showed an interesting effect of the order of presentation which was observed for all but one source material. The subjects gave increasing distance estimates with repeated exposure to the same source material independent of the auralized acoustic conditions. Such an effect is possibly related to the musicality or other properties of the source material, but the definitive reasons for this effect remain unclear and open for future work.

7. ACKNOWLEDGMENTS

The research leading to these results have received funding from the Academy of Finland, project no. 257099. Also many thanks to the great number of anonymous reviewers!

8. REFERENCES

[1] M. A. Gerzon, "The design of precisely co- incident microphone arrays for stereo and surround sound," in *Proc. of The 50th AES Convention (Abstracts)*, vol. 23, pp. 402–404, 1975.

[2] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[3] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[4] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[5] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3148–3161, 2012.

[6] A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki, "Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 239–250, 2014.

[7] L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*, Springer Science & Business Media, 2004.

[8] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.

[9] N. Kopčo and B. G. Shinn-Cunningham, "Effect of stimulus spectrum on distance perception for nearby sourcesa)," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1530–1541, 2011.

[10] D. H. Mershon and L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception & Psychophysics*, vol. 18, no. 6, pp. 409–415, 1975.

[11] S. H. Nielsen, "Auditory distance perception in different rooms," *Journal of the Audio Engineering Society*, vol. 41, no. 10, pp. 755–770, 1993.

[12] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1832–1846, 2002.

[13] A. D. Little, D. H. Mershon, and P. H. Cox, "Spectral content as a cue to perceived auditory distance," *Perception*, vol. 21, no. 3, pp. 405–416, 1992.

[14] J. Pätynen, *A virtual symphony orchestra for studies on concert hall acoustics*, Ph.D. thesis, Aalto University School of Science, November 2011.

[15] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. of the 108th AES Convention*, 2000.

[16] A. Haapaniemi and T. Lokki, "Identifying concert halls from source presence vs room presence," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. EL311–EL317, 2014.

[17] J. Pätynen, S. Tervo, and T. Lokki, "Amplitude panning decreases spectral brightness with concert hall auralizations," in *the 55th International AES Conference: Spatial Audio*, 2014.

[18] C. J. Plack, *The sense of hearing*, Psychology Press, 2013.

[19] H. Kuttruff, *Room acoustics*, Spon Press, London and New York, 2009.

[20] S. Tervo, T. Lokki, and L. Savioja, "Maximum likelihood estimation of loudspeaker locations from room impulse responses," *Journal of the Audio Engineering Society*, vol. 59, no. 11, pp. 845–857, 2011.

[21] Ville Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[22] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 856–865, Dec. 2008.

[23] J. Pätynen, S. Tervo, P. W. Robinson, and T. Lokki, "Concert halls with strong lateral reflections enhance musical dynamics," *Proceedings of the National Academy of Sciences*, vol. 111, no. 12, pp. 4409–4414, 2014.

[24] ITU-R BS.1116-1, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, 1997.

[25] R. Y Litovsky, H. S. Colburn, W. A. Yost, and S. J . Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[26] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," *J. Acoust. Soc. Am.*, vol. 130, no. 2, pp. 835–849, 2011.

[27] A. Kuusinen, "An anechoic audio corpus for room acoustics and related studies," in *Proc. of International Symposium of Auralization and Ambisonics, Berlin, Germany*, 2014.

[28] J. Pätynen and T. Lokki, "Directivities of symphony orchestra instruments," *Acta Acustica United with Acustica*, vol. 96, no. 1, pp. 138–167, 2010.

[29] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data analysis, machine learning and applications*, pp. 261–268. Springer, 2008.

[30] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, pp. 114–133, 1931.