This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Kuusinen, Antti; Lokki, Tapio

Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics

Published in:
Journal of the Acoustical Society of America

DOI:
[10.1121/1.4935388](https://doi.org/10.1121/1.4935388)

Published: 01/01/2015

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Kuusinen, A., & Lokki, T. (2015). Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics. *Journal of the Acoustical Society of America*, 138(5), 3148-3159.
<https://doi.org/10.1121/1.4935388>

Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics

Antti Kuusinen, and Tapio Lokki

Citation: [The Journal of the Acoustical Society of America](#) **138**, 3148 (2015); doi: 10.1121/1.4935388

View online: <https://doi.org/10.1121/1.4935388>

View Table of Contents: <http://asa.scitation.org/toc/jas/138/5>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences](#)

[The Journal of the Acoustical Society of America](#) **140**, 551 (2016); 10.1121/1.4958686

[Concert halls with strong and lateral sound increase the emotional impact of orchestra music](#)

[The Journal of the Acoustical Society of America](#) **139**, 1214 (2016); 10.1121/1.4944038

[Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls](#)

[The Journal of the Acoustical Society of America](#) **135**, 239 (2014); 10.1121/1.4836335

[Concert hall acoustics assessment with individually elicited attributes](#)

[The Journal of the Acoustical Society of America](#) **130**, 835 (2011); 10.1121/1.3607422

[Assessing auditory distance perception using virtual acoustics](#)

[The Journal of the Acoustical Society of America](#) **111**, 1832 (2002); 10.1121/1.1458027

[Perceptive and objective evaluation of calibrated room acoustic simulation auralizations](#)

[The Journal of the Acoustical Society of America](#) **140**, 4326 (2016); 10.1121/1.4971422

Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics

Antti Kuusinen^{a)} and Tapio Lokki

Department of Computer Science, Aalto University School of Science, P.O. Box 15500, Otaniementie 17,
FI-00076 Aalto, Finland

(Received 26 February 2015; revised 15 September 2015; accepted 26 October 2015; published online 19 November 2015)

Virtual acoustics with multichannel sound reproduction was used to study auditory distance perception in four concert halls with multiple sound sources on stage. Eight subjects reported apparent auditory distances in five seating positions from 10 to 26 m to the middle of the sources on stage. The distance estimates were collected by absolute distance estimation procedure as well as a free modulus estimation procedure including both within and between halls evaluations. In addition, pairwise preferences were collected for two positions within each hall and for one position between halls. Results reveal that the perception of distance is dependent on the hall acoustics and show how the strength factor G and direct-to-reverberant energy ratio covary in relation to perceptual distances in these halls. The results also indicate that in such large spaces the overestimation of short distances may continue up to and further than 10 m from the sound sources. Preference results show that closer seats were liked more than further ones and that the strength of this preference is associated with the difference in perceptual distances. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4935388>]

[MV]

Pages: 3148–3159

I. INTRODUCTION

A fundamental aspect of our hearing is the perception of the location of sounds, but among the mechanisms of sound localization, the perception of auditory distances in rooms has perhaps been studied the least. Yet, many aspects are known: loudness is an effective relative distance cue, and reverberation facilitates the perceptual estimation of absolute distances¹ with accuracy improving upon repeated exposure to the room acoustical conditions.² In general, it is thought that the perceived distances correspond to the actual distances via a psychophysical power function in the form $p = kr^a$, where p is the perceived distance, r is the actual physical distance, k is a linear scaling factor, and a expresses the amount of non-linear scaling, i.e., compression ($a < 1$) or expansion ($a > 1$).³

The previous studies of auditory distance perception (ADP) in rooms have shown that short distances up to around 2 m are commonly overestimated while further distances are progressively underestimated. Typical values for the scaling factor k have been reported being around 1.32 [standard deviation (SD)=0.75] and for the exponent a around 0.54 (SD=0.21). There have been large differences between individuals in terms of the perceived auditory distances with inter-subjects SDs increasing up to 60% of the physical distance (in logarithmic coordinates) with increasing distance. Thus, the further we are from the sound sources, the less accurate and the less homogeneous responses we tend to give. Nevertheless, perceived distances have been generally well approximated (average $R^2 = 0.91$, SD = 0.13) by the power

function indicating that this psychophysical relationship can be assumed as being valid.³

Most of the previous studies have been conducted in small and middle-sized spaces and with distances up to 10 m.³ There are only a few studies^{4–6} which have focused on large spaces such as auditoriums or concert halls where typical distances between the performers and the audience are in the range of 10 to 30 m. Evidence from these studies suggests, for example, that in certain conditions sound sources can be perceived as being further than the actual distances instead of being perceived as closer.⁴ From technological viewpoint, these studies show that auralizations made from measured impulse responses from real halls and reproduced with a restricted number of loudspeakers can produce perceptions and distance estimates which are comparable to the actual distances.⁵ Regarding the power function, the evidence indicates that the scaling factor k is somewhat greater in larger spaces, while the exponent a is similar to those obtained in smaller rooms.⁶ However, whether there are differences between different spaces in terms of ADP and in terms of the power function parameters k and a is still an open question.

The present study is motivated by the incompleteness of our knowledge on the differences in ADP in different concert halls and by a previous finding that acoustic “proximity” (or “intimacy”⁷) is a shared preferred quality among listeners⁸ while different “tastes” manifest in other factors, e.g., reverberation and clarity.^{8,9} These terms, “intimacy” and “proximity” (which are treated here as synonyms) have clear inherent connotations to the perception of distance and there has been some discussion about the linkage between “intimacy” and the perception of distance,¹⁰ but without information about ADP in concert halls, little is certain.

^{a)}Electronic mail: antti.kuusinen@aalto.fi

Although “intimacy” is not directly in focus in the present investigation, it is worthwhile to note that the experience of hearing the music as though being “near the performers” or “in a small room,” has been described as one of the most important factors of acoustic quality.⁷

The investigation consists of a multiphase listening experiment including (1) a preference test with paired comparisons, (2) an absolute distance estimation task where listeners reported the distances to the center of multiple sound sources without an explicit reference, and (3) a free modulus distance estimation including both within- and between-halls designs. The studied conditions include four concert halls, five distances from 10 to 26 m to the center of stage, and two different anechoic sound materials: a 1-min excerpt of symphonic music as well as a stream of sounds from a brass ensemble. The different phases are intended as complementary to each other, with a common aim to provide evidence on how these concert halls might differ in terms of ADP and whether closer perceived distances can be associated with higher preference.

II. MATERIALS AND SETUP

A. Acoustic measurements and auralization methodology

Acoustic measurements were performed with an array of six omnidirectional microphones (G.R.A.S. 50VI-1 Vector intensity probe) as the receiver, and a calibrated array of 33 loudspeakers (Genelec models 1029A, 8020A and 1032A), i.e., “a loudspeaker orchestra”¹¹ as the source. The receiver was placed at the same distances in different halls, by measuring the distances to three sources on stage with a laser meter. The height of the receiver was set at the ear height of a sitting person, approximately at 1.2 m from the floor. After calibrating the loudspeakers on stage, room impulse responses (RIR) were captured by the microphone array from each source on stage with the swept-sine technique.¹²

For each source on stage, RIRs were analyzed with Spatial Decomposition Method¹³ to derive spatial (S)RIR containing directional information (azimuth and elevation angles) of the sound field. This information was linked with each pressure value (i.e., sample) in the RIR captured by the microphone pointing up in the array.

These SRIRs were used to derive secondary IRs corresponding to the locations of the 24 reproduction loudspeakers in the listening room. The present study used a direct panning technique, where SRIRs were distributed to the reproduction setup by positioning the samples to the loudspeaker with the smallest angle difference from the actual angle. Although this technique might slightly modify the directional information in the sound field, it has been found to provide a better overall spectral balance in the context of concert hall acoustics,¹⁴ than, e.g., an amplitude panning technique.¹⁵

Finally, the spatially distributed impulse responses were used as convolution reverbs, so that, each anechoic instrument recording was separately convolved with the SRIR corresponding to a particular source on stage. The anechoic materials used in these auralizations are discussed in Sec. IID

B. Concert halls and acoustic measures

Figure 1 illustrates the studied concert halls with plans and cross sections as well as the volumes and seating capacities indicated for each hall. The selected halls represent different architectural designs. Konzerthaus in Berlin (BK) is a traditional shoebox shaped hall, Beethovenhalle in Stuttgart (SB) exhibits a slightly asymmetric fan design, Music Centre in Helsinki (MT) is a modern vineyard type of design where audience areas surround the orchestra and Palais des Beaux Arts in Brussels (BB) is a smaller oval shaped hall with a quite low ceiling height. The receiver/listening positions are also marked in the figure. It is worth mentioning that one major difference between Music Centre and the three other halls is the steeply rising audience areas in Music

FIG. 1. (Color online) Plans and cross sections of the studied concert halls showing also the receiver positions R1–R5 at the distances 10 to 26 m to the center of stage. The figures are in scale relative to each other. The source layout is illustrated in the middle, with the indication on the source positions of the brass ensemble.

Centre. The elevation in the further positions would have resulted in a downward shift in the overall sound image, but this shift was compensated for in the auralizations.

In this article, the discussion about acoustic measures is restricted to the two previously well evidenced ADP cues in rooms: direct-to-reverberant energy ratio (DRR) and sound intensity at the listener's ears or loudness, which is here represented by the strength factor G . G is commonly used in the context of performance spaces as a measure of subjective sound level.¹⁶ Loudness in this context is therefore defined as perceived intensity or perceived strength at the listener's ears, as opposed to perception of source power or sound source loudness.¹⁷ We also acknowledge that other measures may be associated with ADP, but because such associations are currently unknown, the treatment here is restricted to DRR and G .

Figure 2 represent the values of DRR and G , calculated from the impulse responses in the receiver positions R1 to R5. These values are the averages of octave frequency bands centered at 63 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz. Bars represent 95% confidence intervals in respect to the values of all source positions and thus, represent the variability that can be associated to the stages and the stage houses.

DRR values were calculated by dividing the impulse responses into the direct and reverberant sound with a 6 ms time window and 1 ms non-linear fades ("Hamming window"). The direct sound energy was integrated over the first 6 ms and the reverberant energy was integrated over the rest of the response.

G was calculated according to ISO 3382-1:2009¹⁶ with the exceptions that the sources in the measurements were two-way directional loudspeakers (Genelec models 1029A, 8030, 1032A)¹¹ instead of omnidirectional ones and the calculations were made on the frequency range described above instead of only 500 Hz and 1 kHz octave bands.

DRR values between these halls, as illustrated in Fig. 2, are inside the reported JND of 5–6 dB¹⁸ indicating little effect on the perceived distances between these halls in this respect. There might be a small overall effect in this respect, so that, the sound sources in MT would be perceived the closest and the ones in BK the furthest. The variation across positions is comparative between the halls with the

exception of BB, where the decrease in DRR seems to stop after R3. It is probably due to the low ceiling, which produces a reflection with time delay shorter than the direct sound integration window. The fact that MT exhibits the largest values, especially in R1 and R2, may be due to the sloping seating, where the directivity of the loudspeakers increases the direct sound energy in these positions which are on the same height with the loudspeakers on stage. This interpretation was verified by analyzing the direct sound energy only (not shown).

Considering the variation in G values in Fig. 2 across the receiver positions, it can be observed as being dependent on the hall. Particularly, the differences in G between R1 and R5 are the greatest in MT, a little over 5 dB, whereas for other halls, the change is approximately half of that, around 2.5 dB. There is a noteworthy dip in MT between positions R4 and R5 in comparison to other halls, what can be expected to be influential for ADP considering that the known JND for G is in the range of 1 dB.¹⁶

C. Listening room and reproduction setup

The listening room was damped with absorbers and had a reverberation time of 0.2 s. The room met the recommendations of ITU BS.1116-1 standard with the exceptions that the noise rating exceeds the recommended NR15 being under NR30 and the loudspeakers are approximately 1.5 m from the listening position, when 2 m is recommended. The reproduction setup consisted of 24 loudspeakers arranged around the listening position in a three-dimensional layout. The spacing of loudspeakers was more dense in the frontal direction than behind the listener reflecting the localization accuracy of human hearing. More details about the loudspeaker grid have been provided by Haapaniemi and Lokki.¹⁹ The loudspeakers were connected to three DA-converters (Presonus Digimax FS) which were synced and connected to an audio interface (Rme Digiface) which in turn was connected to a laptop (Macbook Pro) running a custom designed listening test program developed with Max/MSP 6 software. The laptop screen was shared with an iPad to display a graphical user interface, which was used to perform the evaluations.

D. Anechoic materials

Two anechoic source materials, referred to with the term "signals" in the following text, were used in the experiment. The first signal was a 1-min excerpt of Bruckner's 8th symphony, 2nd movement. Shorter excerpts of this piece have been employed in other studies of concert hall acoustics.^{14,20} For this material, the source layout and the instrumentation approximated a classical orchestra configuration.

The second signal was specially designed for the current study. The idea was to produce a continuous and variable stream of sounds played by a brass ensemble composed of two trumpets and two trombones. The rationale was to direct listener's attention to the location of the sources, i.e., the distance to the instruments, by providing them with a stream of sounds which would be perceptually unpredictable and would require the listener to concentrate on where the

FIG. 2. Direct to reverberant-ratio and strength G per receiver position and hall. The values are calculated as the average of octave frequency bands centered at 63 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz. Bars represent 95% confidence intervals in respect to the sources on stage and the frequency bands.

sounds are originating from, instead of listening to what is being played. Because producing auralizations in real time is not yet feasible, and it was not sensible to auralize very long individual excerpts due to restrictions in playback buffer sizes, the following approach was used:

- Four anechoic tracks were produced by concatenating randomly selected segments of anechoic recordings of real instruments. In this case these were recordings of two trumpets and two trombones playing symphonic music, namely excerpts of compositions by Beethoven, Bruckner, Mahler, and Mozart (more details about these recordings can be found in Ref. 11). The produced anechoic tracks differed in duration (15, 16, 19, and 23 s).
- Each anechoic track was auralized separately with the SRIR corresponding to the selected source and receiver position. See Sec. II A for details.
- Auralized samples were played back simultaneously in continuous loops. Because each loop differ in duration, the overall stream changes with time.
- To prevent any repetition at the beginning of the playback, the starting position of each loop was randomized for each block in the experimental design.

Because of the special nature of the brass ensemble, and the fact that the test subjects could not choose or loop repeatedly any specific part from the samples, subjects were instructed to concentrate on the apparent distances in one sample at a time and try to form a clear percept of the distance by listening to the sample for a longer period, instead of switching back and forth between the samples in a fast pace. Most of the test subjects reported that this listening strategy greatly facilitated the task and helped them to focus their attention to the location of the sources.

The overall listening level was set at a comfortable listening level while making sure that the samples were neither too quiet nor too loud for long periods of listening. The overall level was the same for all listeners and in all parts of the study, so all the changes in loudness were due to the properties of anechoic materials and the auralized conditions. A-weighted sound pressure levels (SPL) measured at the listening position varied between about 53 to 84 dB (125 ms integration time) for the excerpt of Bruckner's symphony, and between 51 to 75 dB for the brass material. Measured SPLs produced by a real orchestra in various halls and with different compositions and configurations have been reported to be in a similar range, between 45 to 95 dB (B-weighted, fast)²¹ and mostly in the range of 60 to 90 dB.

E. Subjects

Eight subjects (6 men, 2 women, ages between 21 and 37 years old) participated to the experiment. All participant had been part of a preliminary ADP study and hence, they had some prior knowledge about the purpose of this research and knew that the experiment was part of the research on auditory perception in concert halls. All subjects were screened with audiometry and had normal hearing. Three subjects considered themselves as critical listeners and two subjects

actively practiced an instrument and had practiced more than 10 years. All subjects were right handed.

III. DESIGN OF EXPERIMENT

Excluding the audiometry, the experiment was separated into the following parts, which were also run in this order: (1) preference test 1 (PT1); (2) absolute distance estimation 1 (ADE1); (3) free modulus distance estimation (FMDE); (4) absolute distance estimation 2 (ADE2); and (5) preference test 2 (PT2). PT2 and ADE2 were replications of PT1 and ADE1, respectively. The term "absolute" is used here to distinguish and highlight the difference between ADE and FMDE.

Subjects were given written and verbal instructions describing their tasks in each part of the experiment (see details in Secs. III A–III C). Subjects were required to take 5 to 15 min long breaks between each part and in between the sequences during FMDE. The total duration of the experiment ranged from 2 to 4 h depending on the subject. The subjects were given an option to divide the experiment into two separate sessions, but only one subject actually did. During the breaks, subjects were offered coffee, tea, and other refreshments, and they were asked if they felt much fatigue from the listening. Only a little fatigue was reported by the subjects, except for the one who decided to finish the rest of the listening tasks in the following day. The details of each part of the test procedure are presented next.

A. Preference tests

Preference tests were performed by pairwise comparisons. The subjects were instructed as follows: "*Choose the one of the two acoustics / samples you like better.*". Subjects were free to switch between the samples as they liked and to perform the comparisons at their own pace. The preference test included two parts which were interleaved in the randomized sequence of the sample pairs.

In part (a), halls were compared against each other in the receiver position R3 (18 m). The preferences for each pair were collected with both presentation orders and with both source materials making up a total of 24 ($6 \times 2 \times 2$) pairs to compare.

In part (b), receiver positions R2 (14 m) and R4 (22 m) were compared against each other within each hall in one presentation order. The comparisons between R2 and R4 were performed only with the excerpt of Bruckner's symphony due to a small mistake in the coding of the design, but now the comparisons were made two times. Thus, in this part there were a total 8 pairs (4×2) to evaluate.

The presentation order of all the pairs, i.e., both part (a) and part (b) (32 pairs in total), was fully randomized between subjects. As mentioned above, the preference test was conducted at the beginning of the experiment (PT1) and replicated at the end of the experiment (PT2) in order to assess whether extensive listening had an effect on the preferences.

B. Absolute distance estimation

The absolute distance estimation was considered as a designed familiarization phase, where most of the distances and acoustic conditions included in the study were presented to the listeners. The listener's task and the instructions were to "report the distance to the centre of the sound sources in meters." This part was designed so that distances were first given to R2 and R4 in BK and in SB with Bruckner in a random order which was balanced between the subjects. Distances were then given for the brass material with the same two halls. Then the sequence continued with all four halls and receiver positions R1 and R5 and with both signals. The presentation order of these samples was fully randomized between subjects. The absolute distance estimation was repeated at the end of the experiment before PT2.

C. Free modulus distance estimation

Free modulus distance estimation is a direct magnitude estimation²² procedure where the magnitude of the stimulus is the perceived distance. In magnitude estimation paradigms it is common to employ an arbitrary ratio scale set by numbers such as 100 or 10, but as distance is very tangible with a natural scale in meters (or feet), such arbitrary scaling procedures were not needed. The term "free modulus" refers to allowing each subject to set their own reference distance, i.e., a "modulus" and to compare and judge other distances relative to this modulus. Aim is to direct attention to the relative differences between the stimuli. This aim was also reflected in the instructions given to the subjects:

"When you see the text 'Set reference,' your task is to indicate the perceived distance to the center of the sound sources in the reference sample in meters. Set your answer by pressing the arrows on the left of the user interface. The reference sample is always on the left side of the screen. The reference sample will remain the same until you are prompted to set the distance to a new reference in regular intervals (of 4 or 6 comparisons).

When you see the text 'Compare,' your task is to compare the two samples and to indicate the perceived distance to the center of the sound sources in the comparison sample (always on the right). Focus on the relative difference between the reference and the comparison sample. For example, you can think as follows: If these sound sources (the reference) are at a distance of X meters (i.e., the distance which you have set), then these ones (the comparison) are at a distance of ?-meters."

Because of the nature and duration of the samples, the procedure was implemented as pairwise comparisons and the subjects were required to use the meter scale. The stimulus associated with the modulus could be played back at any time in a sequence. The perceived distance to the reference, that is, a new modulus was set at specific intervals in the sequence according to the experimental design.

FMDE actually included two designs where distances were estimated either *within* each hall, or *between* the halls in one receiver position at a time. These two parts were separately designed, but interleaved in the actual presentation sequence. The "within halls"-design included all the receiver positions from R1 to R5, whereas "between halls"-design only positions from R2 to R5. This decision was made in order to shorten the total duration of the experiment.

In the "within halls"-design, R2, and R4 were both used as the references for the moduli, because the choice of these stimuli might influence the results and the use of extreme stimuli for this purpose is not recommended.²³ In order to assess the reliability of the judgments and whether any systematic bias would influence the judgments, each "within hall" sequence included a hidden reference. Thus, a single FMDE sequence within a hall, included setting the apparent distance for the modulus, which was either R2 or R4, and then judging the distances to the five other samples (R1–R5) in relative to the modulus. The order of presentation was randomized within this sequence.

One complete sequence per each hall consisted of 24 pairwise evaluations (5 comparisons $\times$ 2 moduli $\times$ 2 signals + setting the moduli $\times$ 4). Giving one to 2 min per evaluation, and considering other parts of the experiment and the required breaks, it was apparent that one individual could assess a maximum of three halls in a practically sensible timeframe.

In order to include a total of four halls in this study and to simultaneously restrict the total duration of the experiment per each listener, a balanced incomplete block design (BIBD)²⁴ was employed. In this BIBD, each of eight subjects (blocking factor) performed the magnitude estimation for three halls. This way the distances in each hall were estimated by six subjects and each pair of halls were estimated by three subjects. As described above, the distances were collected twice (each R2 and R4 as modulus) and in addition to both source materials (Brass and Bruckner) resulting in a total of 24 observations per each distance in each hall. This within halls-design required each subject to make a total of 36 (3×12) judgments per signal and 72 judgments overall. It was acknowledged that reliability of a single individual could not be rigorously addressed by this design which yielded only two observations per each signal-hall-distance combination per individual. However, the focus here was not on the individuals but on the differences between the halls. As mentioned, the hidden reference was included to provide some information on the reliability of the subjects.

In the "between halls"-design, halls were compared between each other in receiver positions R2, R3, R4 and R5. Now, four halls and four distances yielded a total of 16 different moduli, which were distributed so that each subject was assigned with six different moduli. In contrast to the "within halls"-design, this was not a BIBD and it resulted in unbalanced data set. However, the linear mixed effects approach used in the data analysis, see Sec. IV B, can take into account this type of unbalanced data. To further decrease the duration of the experiment, a hidden references were not included in the "between halls" sequences. This

way there were 24 separate judgments (i.e., 6 comparisons per each position) per each of the two signals yielding a sequence of 48 judgments for each subject.

The “within halls” and “between halls”-designs were interleaved in the overall presentation sequence. Thus, in total each sequence included 120 (72 + 48) evaluations. This sequence of 120 was divided into four parts of 30 evaluations each.

IV. RESULTS

Data analysis was performed in R statistical programming language.²⁵ The most important packages used in this study were lme4²⁶ for linear mixed effects analysis, and ggplot2²⁷ for generating the figures.

A. Preferences

The results of the pairwise preference judgments were analyzed with χ^2 tests of independence and exact binomial tests. Also McNemar’s²⁸ test was used to check whether the preferences changed between the PT1 and PT2.

First, we analyzed the data from the preference judgments between halls in the receiver position R3. To analyze whether the preferences changed between the PT1 and PT2, the data were first inspected at the level of individual subjects for each combination of hall and source material. At this level there were only six observations per combination in each PT1 and PT2, so statistical methods were not used, but the data were manually checked for dramatic discrepancies in the preference counts. In 11 cases (out of 64) the change in preference counts were three or more (i.e., over 50%) and the changes were mostly observed for the Brass signal. Aggregating the data over all subjects enabled performing McNemar’s χ^2 test for each combination of hall and signal. McNemar’s test takes into account the fact that the preferences were collected from the same individuals.

The only significant result was observed for MT with Brass material [$\chi^2(1, N: 96) = 4.17, p = 0.008$] indicating that MT was less liked at the end of experiment than at the beginning. Aggregating over signals and performing McNemar’s test per each hall, did not however indicate any substantial changes. Thus, it was concluded that the changes in preferences between PT1 and PT2 were insignificant, and the data from the two sessions were merged.

Next, the differences between subjects were analyzed with χ^2 test of independence to see whether the group agreed or disagreed on their preference judgments. The test was performed separately with each signal and then also for aggregated data. The test results were approached with caution as the χ^2 approximation may not hold for these sample sizes. Although some indications of differences between individuals were found in the case of Bruckner ($\chi^2(21, N: 192) = 39.7, p = 0.008$ and for overall results [$\chi^2(21, N: 384) = 37.1, p = 0.017$], this was clearly caused by one subject who differed from the rest of the group in his/her preferences between MT and SB.

χ^2 and McNemar’s tests were also used to investigate the overall differences between source materials. Although

χ^2 test gave only little support ($\chi^2(3, N: 384) = 7.02, p = 0.07$), McNemar’s test per each hall indicated that BB was preferred more ($\chi^2(1, N: 96) = 16.01, p < 0.001$) while MT ($\chi^2(1, N: 96) = 9.1, p = 0.002$) and SB ($\chi^2(1, N: 96) = 8.1, p = 0.004$) were preferred less with Bruckner than with Brass. Exact binomial tests with the expected probability of 1/4 were also performed in terms of the preference counts of each hall, and the results are illustrated in Fig. 3(a). SB was the least preferred, MT is less preferred than BB and BK with Bruckner, and these three halls were equally preferred with the brass signal.

A similar analysis was performed for the preferences between R2 and R4 within each hall. The analysis of differences between PT1 and PT2 showed indications of significant changes only in the case of BB, where R2 was preferred more in PT1 than R4 while it was the contrary in PT2 [McNemar’s $\chi^2(1, N: 32) = 8.1, p = 0.004$]. This being the only significant result between the PT1 and PT2, the data sets from PT1 and PT2 were merged in order to simplify the subsequent analysis.

The data were first inspected for apparent differences between subjects, but only little differences were noted, and the data were aggregated over subjects. Binomial tests were performed separately for each hall. Results are illustrated in Fig. 3(b) showing that R2 is preferred over R4, and that this division is the strongest in MT, while it is observed in lesser degree in BK and SB. For BB the difference between R2 and R4 is insignificant and at the level of chance (i.e., probability of 1/2). The grey lines represent different individuals showing that the judgments were almost unanimous in BK and MT, while more heterogeneity is observed for BB and SB. Overall the results show a clear general tendency to prefer R2 over R4, and that hall acoustics influences how strong this tendency is.

FIG. 3. (a) Preferences between halls in R3. (b) Preferences between R2 and R4 within each hall. Grey lines and points represent the preferences of each individual.

B. Absolute distance estimation

All absolute estimation as well as free modulus estimation data were first subjected to a modulus equalization procedure²³ to account for inter-individual differences in the scaling behavior. It is well evidenced that ADP data may exhibit large inter-subject variances,³ which are needed to be accounted for when focusing on the relative differences between the stimuli. Thus, the data of each individual were transformed to logarithmic coordinates and scaled toward the overall average of the whole data by the difference between each individual's own average and the overall mean. This equalization procedure was performed separately to absolute estimation and free modulus estimation data.

Absolute distance estimation data were analyzed with the repeated measures analysis of variance (ANOVA) with factors "session," "hall," "position," and "signal." Analysis did not show significant session effects, so all results below are aggregated, respectively.

Results of the analysis corresponding to Fig. 4(a) showed a significant difference between R2 and R4 [$F(1, 7) = 7.82, p = 0.027$] and the main effect of hall [$F(1, 7) = 12.28, p = 0.001$]. The main effect of signal [$F(1, 7) = 2.14, p = 0.187$] and the interaction between hall and position [$F(1, 7) = 1.023, p = 0.346$] were not significant nor was the interaction between signal and hall [$F(1, 7) = 4.5, p = 0.072$].

The analysis for positions R1 and R5, illustrated in Fig. 4(b), revealed significant main effects for hall [$F(3, 21) = 5.195, p = 0.008$] and position [$F(1, 7) = 52.88, p < 0.005$] but not for signal [$F(1, 7) = 0.355, p = 0.57$]. The interaction between hall and position was now also significant [$F(3, 21) = 13.6, p < 0.0001$]. *Post hoc* analysis by Tukey's Honestly Significant Differences (HSD),²⁹ verified that MT was estimated closer than BK and SB in R1, and further than BB and BK in R5. The differences between BB and BK were not significant. Also, SB was perceived further than BB and MT in R1 and further than BB and BK in R5. The greatest compression of perceived distances compared with physical distances is shown for BK. It is also notable that these results indicate that all halls except MT were perceptually little overestimated at the distance of 10 m, but largely underestimated at the distance of 26 m.

FIG. 4. Results of absolute distance estimation. (a) Positions R2–R4. Note that these conditions were only performed with BK and SB. (b) Positions R1–R5.

C. Free modulus distance estimation

1. Within halls

The comparison sequence in the within halls–design included hidden references, which were used to check for bias and the reliability of the results. The mean differences between the moduli and the hidden references were 0.3, -1.3 , 0.0 , 1.3 , 0.4 , 0.0 , -0.2 , and 0.0 m for each of the eight subjects, respectively. Considering that the resolution of the scale was 1 m, these differences were regarded as minor and were not compensated in the data, because the data were also treated with the above mentioned modulus equalization procedure. Before the subsequent analysis, the judgments given to the references, i.e., the moduli (but not to the hidden references) were removed in order to balance the data across all positions.

After these preprocessing stages, a linear mixed effects³⁰ analysis was performed to investigate the relationships in the data. The visual inspection of residuals did not show clear deviations from homoscedasticity or normality. Signals (two levels), halls (four levels), and either position (five levels), or a distance covariate in logarithmic coordinates were entered as fixed effects. Distance covariate was used to obtain the estimates for the power function parameters k and a . The random effects terms reflected the structure of the experimental design. When position was used as categorical variable, the random effects structure included intercepts for each level of nesting: positions nested in halls, halls nested in signals and signals nested in subjects. When distance was entered in the model as a covariate, the random effect terms were investigated more closely to assess whether the best fit was obtained by entering the slope and intercept terms as independent or dependent from each other. The best fit was obtained by entering the intercept and the slope as dependent and conditional to subject as well as conditional to the hall.

This linear mixed effects analysis was first used to perform repeated measures ANOVA (type III, Satterthwaite approximation of degrees-of-freedom) which was of primary interest when position was entered in the model as a categorical fixed effect. The results showed significant main effects for hall [$F(3, 31.85) = 2.6, p = 0.007$] and position [$F(4, 400) = 54.1, p < 0.0001$] as well as significant hall-position interaction effect [$F(12, 400) = 2.3, p = 0.007$]. *Post hoc* analysis with Tukey's HSD revealed that perceived distances for SB were in overall further than in the other halls, which were not different from each other. Considering different positions, R1 was rated closer than all other positions, R2 was closer than R4 and R5, but not significantly different from R3. R3 was perceived closer than R5 but not closer than R4. R4 and R5 did not differ significantly. These results are illustrated in Fig. 5(a).

Performing the same analysis with the position variable replaced by the distance covariate corroborated these results by indicating significant differences between halls in terms of both intercepts (k) and slopes (a). Respective differences between signals were not significant and the division is thus omitted in the Fig. 5(b). It is observed, that auditory distance perception in MT was significantly more linear than in the

FIG. 5. Results of the free modulus distance estimation within halls–design. The division into signals is omitted because the analysis did not reveal significant differences between the signals. Straight grey lines represent linear correspondence. (a) The geometric means. (b) Estimated power functions. The parameter values and the confidence intervals are as follows: BB, $k=4.1$ (3.0 – 5.7), $a=0.41$ (0.30 – 0.51); BK, $k=5.0$ (3.6 – 6.9), $a=0.33$ (0.22 – 0.44); MT, $k=1.8$ (1.3 – 2.4), $a=0.70$ (0.60 – 0.82); SB, $k=4.2$ (3.0 – 5.8), $a=0.45$ (0.33 – 0.55).

other halls, with average k and a estimated to be 1.8 and 0.70, respectively. BK ($k=5.0$, $a=0.33$) and BB ($k=4.1$, $a=0.41$) had similar estimates, while in SB distances were perceived further in overall and in a little more linear fashion ($k=4.2$, $a=0.45$). It is notable that the estimates for k , except for MT, were greater than reported in most previous studies, and combined with the estimates for a indicated that the overestimation of auditory distances may continue up to 14 m. However, in all cases the auditory distances at the furthest seats (R4, R5) are greatly underestimated as expected. The overall means across halls were 3.5 [standard error (s.e.) 1.16] and 0.47 (s.e. 0.05), for parameters k and a , respectively.

In order to investigate the results of each individual more closely, the data of each subject were also analyzed by linear mixed effects model where signals were entered as a random factor and hall as a fixed factor with distance as a fixed covariate. The parameter k and a values ranged from 1.23 to 11.43 and 0.12 to 0.85, respectively, so there were large differences between individuals.

2. Between halls

In the between halls–design, each subject gave a total of 48 distance estimates. Each of four halls were evaluated a total of 96 times and each signal a total of 192 times. However, because of erroneous design, the data (overall N of 384) were unbalanced across some of the subject, hall, signal, and position combinations. There were differences between receiver positions: R2 position had a total of 100 responses (25 per hall), R3 96 responses (24 per hall), R4 84 (21 per hall), and R5 104 responses (26 per hall). In addition the subjects differed in terms of how many times they had evaluated each position (4, 8, 12, 16, or 20 times), and there were some differences between signal and position combinations (N was 40, 44, 48, or 52).

After modulus equalization procedure, the data were visually inspected for departures from normality and

homoscedasticity, without noticing any apparent violations. A full mixed effects model specified with signals (2 levels), halls (4 levels), and positions (4 levels) entered as fixed effects. The same random intercept terms were specified as in within halls analysis but, halls were now entered as nested in positions, which is the very manner how the data were collected. The corresponding analysis of variance indicated significant main effects for signals [$F(1, 13) = 16.0$, $p = 0.002$], and halls [$F(3, 302) = 32.0$, $p < 0.001$] as well as the interaction between halls and positions [$F(9, 302) = 9.0$, $p < 0.001$]. Interestingly a significant main effect for position was not found [$F(3, 35) = 1.45$, $p = 0.24$]. The results were also investigated by the least significant differences *post hoc* analysis which showed that in overall BB was estimated closer than the other halls, MT was estimated closer than SB in R2 and R3, and further than BB in R4 and R5 and further than BK in R5. These differences were noted at a conservative significance level of 0.001. The results are illustrated in Fig. 6.

D. Distance judgment variability

In order to evaluate the variability of distance judgments throughout the study, SDs were calculated for all original (i.e., “unequalized”) distance judgments given to the same position and hall combination per each individual and by aggregating the data over individuals. Considering that the focus in judgment variability is in terms of distances and halls, the possible effect of signals is ignored. The SDs per subject were 3.9, 11.5, 3.2, 2.2, 4.9, 1.7, 1.9, and 3 m. Inspecting the data, subject 2 had apparently changed his/her scaling behavior drastically between different parts of the study resulting in the observed large SD. Otherwise, these SDs with the average of 4 m can be regarded as moderate considering that distance judgment based on audition is a difficult task.

Aggregating the data over subjects, the SDs were calculated for each distance in each hall. SDs in MT were little greater in overall (average 10.5 m) than in the other halls (BB, 7.6 m; BK, 8.5 m; and SB, 7.9 m). The SDs for each distance were between 7.9 m (R1) and 9.1 m (R5), but contrary to previous studies,³ there was not a systematic increase in SD from R1 to R5 (R2, 9 m; R3, 8.3 m; R4, 9 m; R5, 9.1 m).

FIG. 6. Distance estimation results from the free modulus between halls–design. Bars represent 95% confidence intervals.

FIG. 7. DRR and G values plotted against perceptual distances [(a) and (b)] and against each other (c). Note that the axes in (c) are inverted. G and DRR values decrease when moving right and up on the x and y axes, respectively.

E. Investigating the effects of G and DRR

The relationships between the perceptual distances and G and DRR were investigated by plotting the mean parameter values against perceptual distances and against each other in Fig. 7. In Fig. 7(a), it is interesting to note that the variation in DRR values in MT is the smallest among these halls, while perceptual distances vary the most. Regarding the values of G represented in Fig. 7(b), it seems that the range of G is associated with the range of perceived distances. In Fig. 7(c) these parameters are plotted against each other. Notable is that the small change in G is associated with a great change in DRR in BB, BK, and SB while it is the opposite in MT.

The effects of G and DRR were quantified by dividing the data from FMDE “within halls”-design into three subsets, so that the subsets included positions R1 to R3, R2 to R4 and R3 to R5. In each case, the data were analyzed with a simple additive linear (mixed) model as: $Z_p = \beta_0 + \beta_G Z_G + \beta_{DRR} Z_{DRR} + e$, where Z_p stands for z-transformed perceptual distances (in logarithmic coordinates), Z_G and Z_{DRR} refer to z-transformed parameters. A mixed effects approach was used again, and the residuals term e , in this case, represents random intercepts for each subject besides the residual error. The interaction between G and DRR was found insignificant in all cases, so that term was removed from the model. The coefficients β_G and β_{DRR} can be interpreted as indicating how the increase of 1 SD of the parameter influences the distance perception while the other one is kept constant. Thus, when the data are analyzed

separately in the different distance ranges, the coefficient shows the relative weightings of G and DRR when moving from close positions to further ones. Standardizing ensures that the estimates are comparable, regardless of the differences in the actual ranges of the parameter values and perceptual distances. Recognizing the likely issue of collinearity between G and DRR, the partial correlations between the perceptual data and the parameters were also calculated. The regression coefficient estimates are represented in Fig. 8, and partial correlations are tabulated in Table I.

The collinearity of G and DRR can be noted in Table I as in many cases neither of the parameters are significant when the collinearity with the other is taken into account. Similarly, when the regression analysis was performed by including only G or DRR, the results were the same as those obtained by the sum $G + DRR$, shown in the lowest row of Fig. 8. Still, the partial correlations in Table I indicate that G had more influence than DRR.

It is interesting that the positive coefficient estimates for DRR in the closest distance range (R1-R3) in BB, BK and MT indicate that the increase in DRR would in fact result in an increase in perceptual distance and not in a decrease. The coefficient estimates for BK in R2-R4 are easier to explain by the relative increase in reverberant energy which affects strongly both G and DRR measures. Also, in accordance with the perceptual distances, the coefficients in the lowest row of the Fig. 8 show how the net effect of G and DRR continue up to the furthest range in the case of MT, while it becomes insignificant in other halls.

FIG. 8. (Color online) Regression coefficients for Z_G (denoted as G), Z_{DRR} (DRR) as well as their sum $Z_G + Z_{DRR}$ in each range R1-R3, R2-R4, R3-R5 and full range R1-R5 (ALL Rs). Although the coefficients were significant in only a few cases, the differences between the halls can be observed. Particularly, the net effect of these cues can be noted to continue for MT, while it vanishes for other halls when moving from the R1-R3 to R3-R5. Significance ($p < 0.001$) is denoted by *.

V. DISCUSSION

The authenticity and realism of the auralizations used in the present investigation have been evidenced by extensive formal^{14,20} and informal listenings and still one can not be

TABLE I. Partial correlations between perceptual distances and either G or DRR, while controlling for one or the other.

	All Rs		R1-R3		R2-R4		R3-R5	
Hall	G	DRR	G	DRR	G	DRR	G	DRR
BB	-0.07	-0.05	-0.13	0.04	-0.07	-0.03	-0.07	-0.06
BK	-0.09	0.01	-0.16	0.08	0.03	-0.04	0.06	-0.1
MT	-0.19	-0.05	-0.32	0.12	-0.21	-0.01	-0.06	-0.1
SB	-0.25	0.07	-0.01	-0.06	-0.16	0.08	-0.03	0.02
ALL	-0.48 ^a	-0.25 ^a	-0.42 ^a	-0.30 ^a	-0.34 ^a	-0.18	-0.33 ^a	-0.08

^aSignificant correlations ($p < 0.05$, with Bonferroni correction).

certain that the results are fully representative of real conditions. It is reasonable, however, to assume that the prominent differences between the acoustic conditions under study were preserved in these auralizations, because any systematic bias due to signal processing would have been the same for all stimuli. The stimuli were produced by combining anechoic musical sounds with the SRIRs obtained with directional loudspeaker sources on stage, what means that generalizing the results to other sounds (e.g., speech) and other types of sources and source configurations has to be made with caution. It is also noteworthy that the subjects were told that the study is about concert hall acoustics, and hence, they may have adjusted their judgments to meet with the expectations induced by this knowledge. While keeping in mind these limitations of this study, and a critical view toward the generalizability of these results, we strongly believe that these findings correspond well to the auditory perceptions in real concert halls.

The observed strength of preference for R2 over R4 for different halls implies that there are more seats in BB, BK, and SB with similar acoustical quality with R2 than there are in MT. The tendency to prefer closer seats seems to increase with the difference of perceptual distances, which is the greatest in MT. Considering that there are a number of acoustic factors that change with distance the evidence from studies on opera theatres^{31,32} suggests that the determining factor is in fact loudness. Observing the G values presented in Fig. 2 and the greatest difference for MT, it is possible that loudness might be the determinant factor of preference also in the present study.

If loudness was the main determinant, setting the output to a comfortable listening level may have excluded conditions where the sounds would have been judged too weak or too loud in a real setting. For example, if the overall level was set on a higher level, the closer seats could have been perceived as too loud, and the preference results would have been very different. If the level was set lower, it is likely, however, that the results would not have been dramatically different, because now the further positions would have been perceived too weak and the direction of the preferences would have remained the same. Therefore, these preference results can be argued to be representative of cases where the loudness at the closer seats is not overwhelming, what in the author's opinion, is the general case in real concerts. The influence of the overall listening level to ADP is more difficult to assess, but for instance Nielsen³³ has concluded that the overall listening level is not a factor for ADP in reverberant conditions. It is likely that an analytical evaluation of the distance by using the auditory cues is not much affected by the overall listening level given that the level is adequate.

The free modulus estimation was performed in two interleaved designs where distances were judged "within" and "between" halls. While the results showed that relative differences between the halls in each position are roughly the same for both designs, it is interesting that in the "between halls"-design the positions were not found significantly different from each other. This observation highlights the nature of the free modulus magnitude estimation protocol, which targets the relative differences between a set of

stimuli. The subjects were specifically told to direct their attention to the relative differences, what might have resulted in giving more arbitrary distances for the moduli. It is also possible that the inconsistencies in the number of observations per condition in the "between halls" data may have influenced the results.

All the free modulus and the absolute distance judgments resulted in similar trends and patterns where MT exhibited the most linear change in auditory distances when moving from R1, where MT was rated as the closest among these halls, to R5 where MT was perceived as the furthest one. The results for Palais de Beaux Arts in Brussels (BB), and Konzerthaus in Berlin (BK) were similar with the highest compression of the perceptual distances in relation to the actual distances, whereas the auditory distances in SB were judged generally the furthest across the studied distances. Overall, the underestimation of the furthest distances was observed for all subjects, all four halls and with both source materials. The absolute distance judgments illustrated in Fig. 4 give some reason to speculate that the differences in perceived distances between the halls are pronounced with the full orchestra, although statistically significant differences between the signals were not observed. For the fitted power functions, the grand average for the exponent a was 0.49 (s.e., 0.05), which is well in-line with the values reported by others.^{3,6} The average of constant k was 3.5 (s.e., 1.16) which is somewhat higher than what has been previously reported.

The mixed effects analysis with different random term configurations also indicated that the parameters k and a are in fact dependent on each other. Dependency of k and a was observed as conditional to both subject and hall. The conditionality to subject can be explained by the fact that each individual might use a different evaluation strategy that is reflected in the results. After all, inter-subject variability of k and a parameters was substantial as the estimated values ranged from 1.23 to 11.43 and 0.12 to 0.85, respectively.

The conditionality to hall is perhaps more interesting as this suggest that each hall would have a characteristic power function with specific parameter values indicating how auditory distances may be expected to change inside the hall when moving further away from the stage. If the dependency between k and a holds, a high value of k would be associated with a low value of a , and thus, with greater compression of perceived distances. It is interesting that Anderson and Zahorik⁶ have previously discussed that it is the constant k that may be associated with the amount of reverberation in a room, because the exponent a has been observed to have similar values across different experiments and acoustical conditions. The current estimates of k seem to give further evidence to this suggestion. Moreover, the results imply that greater amount of reverberation is linked to more compression via a smaller value for exponent a . For instance, the low level of reverberant sound energy in relation to the direct sound in MT is associated with a low estimate for k and a high estimate for a , while the contrary is observed with other halls. It is tempting to speculate a connection between k and a and reverberation and overall sound strength. The validity of making such connection is however left open for future

research, because in this study, the dependency between k and a might actually be a by product of the free modulus evaluation procedure and the experimental design.

The fitted power functions can be also used to approximate and compare the crossover points where the perceived distances match the actual physical distances. Here, the crossover points were 11 m (both BB and BK), 6.5 m (MT), and 13 m (SB). Again, if k and a are in fact dependent on each other and conditional to the acoustics of the room, the closer the crossover distance is, the more linear fashion the distance will change. Overall, these crossover distances together with the perceptual distance judgments suggest that in concert halls distances up to and over 10 m may be overestimated, what is in clear contrast to the results from smaller rooms.³

Considering the analysis of G and DRR, the results highlight how these measures covary and their relationship with the perceived distances. The decrease in G was associated with the relative increase in perceptual distances, but it was observed that DRR (i.e., reverberation) might moderate this increase. In BB, BK, and SB at positions R3 to R5, a relatively large change in DRR values was observed together with only little decrease in G values while perceptual distances also converged. In MT where the perceptual distances continued to increase up to R5, the change in DRR values is less than in the others halls, while the change in G is the greatest. Investigating the relative weightings of G and DRR in more detail implied that in some cases G and DRR cues might actually counteract each other. The overall balance of these cues together seems to determine how much and in which direction the perceptual distances tend to change.

Overall these results illustrate an important perceptual consequence of the sound energy relations³⁴ within and between concert halls with different architectural designs. A more detailed discussion and treatment of these energy relations and other acoustic parameters is however outside the scope of this study and is left for future consideration, as is the computational modeling of ADP^{35,36} on this data. At present these results corroborate the evidence from other studies^{4,5} indicating that ADP is dependent on the hall acoustics and that the parameter estimates of the power function vary due to the acoustical properties of the surrounding environment.

One motivation for this study was that “intimacy” (or “proximity”) has been previously found as one main driver of preference⁸ and a determinant of sound quality⁷ in concert halls. The results seem to corroborate these findings inside a single hall, where closer seats were preferred to further ones. The preferences between the halls, however, were collected only in the position R3, and because they were collected only in this one central position, it remains unclear whether closer/further auditory distances are always more/less preferred and in which conditions other factors may come into play.

VI. CONCLUSIONS

This study with virtual acoustics showed that auditory distance perception in concert halls is affected by the

acoustics of the hall. Among the studied four halls, the greatest amount of perceptual compression was observed in Palais des Beaux Arts in Brussels (BB), and in Konzerthaus in Berlin (BK), where the sources at the furthest distance at 26 m on stage were estimated as being around 15 m. The same distance in Beethovenhalle in Stuttgart (SB) and Music Centre in Helsinki (MT) was estimated to be around 18 m. The sources in SB were perceived as the furthest throughout the four other seating positions at 10, 14, 18, and 22 m. MT, in turn, was perceived the closest at the position nearest to the stage, and exhibited the most linear change in terms of perceptual distances when moving further away from the sources. The data were fitted with the power function $p = kr^a$ and the estimated values for the linear constant k were 4.1 (BB), 5.0 (BK), 1.8 (MT), and 4.2 (SB) and for the exponent a 0.41 (BB), 0.33 (BK), 0.70 (MT), 0.45 (SB). The average values were 3.5 and 0.49 for k and a , respectively. The parameter estimates for MT were significantly different from the other halls.

Strength factor G and DRR were used to quantify the observed differences between the halls. The greatest range in the values of G in MT was associated with the greatest range of perceived distances among these halls. G and DRR values and their relationship with the perceptual distances imply that the reverberant energy at the furthest positions may be associated with the convergence of perceptual distances. It seems that reverberant sound energy may limit the relative increase in perceptual distances by compensating for the decrease in overall strength due to the attenuation of the direct sound when moving further away from the sources.

Results from the preference test showed that closer positions were preferred within each hall, and that the greater the perceptual distance difference, the stronger was the preference for the closer position. This result is, however, restricted to such situations where the loudness in the closer positions is not too loud, because the overall listening level was set at a comfortable level. When preferences were collected between the halls at the distance of 18 m from the center of the stage, SB was the least liked possibly due to the lack of sound strength indicated by a low value of G . Overall this study highlights important differences between concert halls in terms of auditory distance perception which is a fundamental aspect of human hearing and an integral part of auditory experience.

ACKNOWLEDGMENTS

We thank Dr. Sakari Tervo and Dr. Jukka Patynen for discussions. The research leading to these results has received funding from the Academy of Finland (project No. 257099), Emil Aaltonen's Foundation and Finnish Foundation for Technology Promotion.

¹D. H. Mershon and L. E. King, “Intensity and reverberation as factors in the auditory perception of egocentric distance,” *Percept. Psychophys.* **18**, 409–415 (1975).

²D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, “Effects of room reflectance and background noise on perceived auditory distance,” *Perception* **18**, 403–416 (1989).

³P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, “Auditory distance perception in humans: A summary of past and present research,” *Acta Acust.* **91**, 409–420 (2005).

- ⁴D. Cabrera, A. Nguyen, and Y.-J. Choi, "Auditory versus visual spatial impression: A study of two auditoria," in *The 10th Meeting of the International Conference on Auditory Display* (2004).
- ⁵D. Cabrera, A. Azzali, A. Capra, A. Farina, and P. Martignon, "Perceived room size and source distance in five simulated concert auditoria," in *Proceedings of the 12th International Congress on Sound and Vibration* (2005).
- ⁶P. W. Anderson and P. Zahorik, "Auditory/visual distance estimation: accuracy and variability," *Front. Psych.* **5**(1097), 1–11 (2014).
- ⁷L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York, 2002), pp. 503–535.
- ⁸A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki, "Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls," *J. Acoust. Soc. Am.* **135**, 239–250 (2014).
- ⁹M. Schroeder, D. Gottlob, and K. Siebrasse, "Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters," *J. Acoust. Soc. Am.* **56**, 1195–1201 (1974).
- ¹⁰J. R. Hyde, "Acoustical intimacy in concert halls: Does visual input affect the aural experience?," in *Proceedings of Institute of Acoustics* (2002), Vol. 24, Pt. 4.
- ¹¹J. Pätynen, "A virtual symphony orchestra for studies on concert hall acoustics," Ph.D. thesis, Aalto University School of Science, Espoo, Finland (2011).
- ¹²A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention* (Audio Engineering Society, New York, 2000), Vol. 108.
- ¹³S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.* **61**, 17–28 (2013).
- ¹⁴J. Pätynen, S. Tervo, and T. Lokki, "Amplitude panning decreases spectral brightness with concert hall auralizations," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio* (Audio Engineering Society, New York, 2014).
- ¹⁵V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.* **45**, 456–466 (1997).
- ¹⁶ISO 3382-1: *Acoustics—Measurement of room acoustic parameters – Part 1: Performance spaces* (International Organization for Standardization, Geneva, Switzerland, 2009).
- ¹⁷P. Zahorik and F. L. Wightman, "Loudness constancy with varying sound source distance," *Nature Neurosci.* **4**, 78–83 (2001).
- ¹⁸P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *J. Acoust. Soc. Am.* **112**, 2110–2117 (2002).
- ¹⁹A. Haapaniemi and T. Lokki, "Identifying concert halls from source presence vs room presence," *J. Acoust. Soc. Am.* **135**, EL311–EL317 (2014).
- ²⁰T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *J. Acoust. Soc. Am.* **132**, 3148–3161 (2012).
- ²¹F. W. Winckel, "Optimum acoustic criteria of concert halls for the performance of classical music," *J. Acoust. Soc. Am.* **34**, 81–86 (1962).
- ²²S. S. Stevens, "The direct estimation of sensory magnitudes: Loudness," *Am. J. Psychol.* **69**, 1–25 (1956).
- ²³S. S. Stevens, "Issues in psychophysical measurement," *Psychol. Rev.* **78**, 426–450 (1971).
- ²⁴D. C. Montgomery, *Design and Analysis of Experiments* (John Wiley and Sons, New York, 2008), pp. 168–176.
- ²⁵R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2014).
- ²⁶D. Bates, M. Maechler, B. Bolker, and S. Walker, *lme4: Linear mixed-effects models using Eigen and S4*, R package version 1.1-7 (R Foundation for Statistical Computing, Vienna, Austria, 2014).
- ²⁷H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
- ²⁸Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika* **12**, 153–157 (1947).
- ²⁹H. Abdi and L. J. Williams, "Tukeys honestly significant difference (hsd) test," in *Encyclopedia of Research Design* (Sage, Thousand Oaks, CA, 2010), pp. 1–5.
- ³⁰R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *J. Memory Lang.* **59**, 390–412 (2008).
- ³¹J. Y. Jeon, Y. H. Kim, D. Cabrera, and J. Bassett, "The effect of visual and auditory cues on seat preference in an opera theater," *J. Acoust. Soc. Am.* **123**, 4272–4282 (2008).
- ³²S. Sato, S. Wang, Y. Zhao, S. Wu, H. Sun, N. Prodi, C. Visentin, and R. Pompoli, "Effects of acoustic and visual stimuli on subjective preferences for different seating positions in an Italian style theater," *Acta Acust. Acust.* **98**, 749–759 (2012).
- ³³S. H. Nielsen, "Auditory distance perception in different rooms," *J. Audio Eng. Soc.* **41**, 755–770 (1993).
- ³⁴M. Barron and L.-J. Lee, "Energy relations in concert auditoriums. I," *J. Acoust. Soc. Am.* **84**, 618–628 (1988).
- ³⁵A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature* **397**, 517–520 (1999).
- ³⁶A. W. Bronkhorst, "Modeling auditory distance perception in rooms," in *Proceedings of the EAA Forum Acusticum Sevilla* (2002).