Perceptual significance of seat-dip effect related direct sound coloration in concert halls

Henna Tahvanainen,[a] Aki Haapaniemi, and Tapio Lokki
Department of Computer Science, Aalto University School of Science, P.O. Box 15500, FI-00076 Aalto, Finland

(Received 28 September 2016; revised 24 January 2017; accepted 10 February 2017; published online 7 March 2017)

In concert halls, the spectrum of direct sound (here 0 to 15 ms) is influenced by the seat-dip effect that causes selective low frequency attenuation. The seat-dip effect has been considered to be detrimental to the acoustic quality of halls, yet there is little evidence about the perceptual significance of the effect. This paper studies the discrimination and preference of seat-dip effect related changes in the direct sound, with realistic auralization of multichannel anechoic orchestra recordings in halls measured with the loudspeaker orchestra. Comparisons are made with a free-field direct sound and direct sound magnitude changes typically associated with the seat-dip effect. Overall, the differences were not significantly audible, except with a subgroup of participants in one out of four halls, and two out of three comparisons. Furthermore, participants' preference for the uncolored direct sound was significant in the halls with less reflected energy, but non-significant in the halls with more reflected energy. The results imply that for most seats in adequately reverberant halls, typical seat-dip effect related coloration in the direct sound can be perceptually negligible.

© *2017 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4977188]

[BFGK] Pages: 1560–1570

I. INTRODUCTION

When sound arrives to the seating area at small grazing angles it undergoes complicated patterns of diffraction and reflections across the seat rows and the floor, resulting in an attenuation centered at around 100 to 300 Hz, and in some cases extending up to 1 kHz.[1–3] This phenomenon, called the seat-dip effect (SDE), colors the spectra of both the direct sound and some of the early reflections. The SDE varies with the direction of arrival,[1–4] and therefore the direct sound and the early reflections can have quite different low-frequency spectra.

The SDE is most prominent in the direct sound until about 15–20 ms, when the first reflections arrive from the room boundaries. In general, subsequent reflections tend to level the initial SDE in the room impulse response (RIR) spectrum,[5,6] but it is not clear to what extent the leveling affects its audibility. Therefore, it is essential to focus on the perceptibility of the *initial* SDE since it has architectural implications: if it presents a perceptual problem, it would be useful to know if it can be rendered inaudible with additional reflected energy that levels the RIR spectrum, or whether it is necessary to design the seating area in a way that reduces the initial SDE itself.

Some studies consider the SDE to have an adverse effect on sound quality, and therefore have focused on removing or at least reducing the effect.[4,7,8] Yet, no perceptual studies have directly addressed sound quality or preference related to the SDE. This research focuses on the perception of the initial SDE in existing halls representing different degrees of initial SDE leveling. To this end, discrimination and preference tests were performed with realistic auralizations of symphonic music at a single representative listening position in measured unoccupied concert halls. The free-field and SDE-filtered direct sound components are modeled according to existing measured data. Differences other than the direct sound were eliminated as a potential source of discrimination by preserving the subsequent reflections present in the original hall RIRs.

II. BACKGROUND

Ever since the first reports of the seat-dip effect (SDE), the discussion on its perception has been divided. Sessler and West[2] recommended measures to remove the effect as they believed it might degrade the early sound and sense of envelopment, while Schultz and Watters[1] suggested that the SDE may not be a serious problem provided that the late reverberation is sufficiently strong. Following this discussion, Barron,[10] based on his analysis of British concert halls, proposed that increasing the low frequency reverberation time would level the SDE. Later, Davies *et al.*[9] suggested that unattenuated early reflections can mask the perception of the SDE. Based on time-frequency analysis of the impulse responses in measured concert halls, Bradley[4] suggested that strong ceiling reflections can level the SDE, while Pätynen *et al.*[5] proposed that lateral reflections may serve that purpose better with less coloration. In general, uniform accumulation of spatial energy appears to level the SDE.[6]

Perceptual studies specifically addressing the SDE are limited to the threshold of perception study by Davies *et al.*[9] They used a 3D reproduction system with eight loudspeakers in anechoic chamber, delay units to create individual reflections, and a digital reverberator for the late sound. The seat-

---

[a] Electronic mail: henna.tahvanainen@aalto.fi

dip filter was composed of two cascaded 1/3 octave band equalizers to create an approximately even attenuation across the 200 Hz octave band. The same filter was applied to the direct sound and five of the early reflections, of which four were lateral. Davies *et al.* obtained a threshold value of -3.8 ± 0.2 dB as the smallest change in early energy at the 200 Hz octave band that 50% of the subjects noticed. They also concluded that the difference in the threshold result between the different reverberant conditions was not significant.

However, the general usability of the threshold obtained by Davies *et al.* has some limitations: (1) In their study, the SDE is confined to one octave band (200 Hz) while the variety of SDE spectra encountered in real halls span more than one octave band. In addition, identical SDE are applied for the direct sound and SDE-filtered reflections, while in reality the reflections are affected differently. In particular, the kind of wide spectral attenuation seen, e.g., in some shoebox halls, affects the low-mid and mid-frequency ranges that may affect the tonal balance, and not necessarily the perceived strength of bass. Therefore, a different threshold could apply in cases where more general changes in timbre take place. (2) The underlying assumption of equal importance for the direct sound and all reflections arriving within the first 80 ms is not supported by perceptual data. For instance, the direct sound spectrum might possibly be weighted more than the reflected energy in the perceptual integration. For example, Soulodre *et al.* have suggested that the perceptual integration limit is actually frequency dependent, and for low frequencies may be closer to 160 ms.[11] (3) The range of reverberant conditions studied was rather limited. While Davies *et al.* do not state whether they simulated a listening position at a specific distance, the C_{80} value of the simulator in the *normal* reverberant condition was high (3.0 dB), and even the *high* reverberant condition had a C_{80} of 0.2 dB, which is more than one would expect for a highly reverberant hall. For example, the best-rated concert halls in Beranek's studies had C_{80} values between -1 and -5 dB.[12] Alternatively, one can consider the results to represent a relatively close listening position, about 10 m from the stage. However, it is doubtful whether the threshold can be considered representative of the situation in a highly reverberant concert hall, especially further than 10–15 m from the stage, where at least half of the seats in most halls are.

Other studies that may be relevant for the perception of the SDE include that of Bradley *et al.*,[13] which indicates that the effect of the late energy is significant for the perceived strength of bass. In their experiments with simulated sound fields, they found that the ratings of bass strength were more sensitive to the late low frequency sound level than the early level. However, when the late low frequency level was low, the early level had more perceptual weight. Thus, designing for sufficient reflections seems a feasible alternative for ameliorating the potential lack of bass caused by the SDE, provided that the reflection spectra are diverse enough. Fortunately, this seems to be the case with many halls.[5,6] Possible more general effects on timbre or tonal balance may require different measures.

Another study related to direct sound coloration by Takahashi *et al.*[14] investigated the tonal balance between music excerpts convolved with RIRs composed of either direct sound only, or direct sound with simulated reflections arriving within less than 20 ms of direct sound from two flat surfaces. They found that the excerpt with no reflections was rated as having the best tonal balance. It seems plausible therefore that the SDE could also impair the tonal balance of music, unless the effect is masked or leveled by later reflections.

III. SETUP

The following describes the methods used to generate and reproduce the stimuli for the listening tests. The listening tests were conducted first in an acoustically treated listening room, and afterwards in an anechoic chamber to alleviate concerns over the potential effect of the environment on the results, and to further augment the data. Concerning the room impulse response (RIR) measurements, their analysis, and the listening room reproduction system and its validation, a detailed description and a block diagram of the stimulus generation has recently been provided in a paper by Pätynen and Lokki,[15] and therefore only briefly discussed here.

A. Reproduction systems

1. Listening room

The reproduction system in the listening room consists of 24 loudspeakers, of which twelve (Genelec 8020B) are at ear level, eight (8020B) above the ear level, and four (1029A) below the ear level (see Table I). The loudspeakers are located at a 1.5 m nominal distance from the listening position. The peak-to-peak level difference between direct sound and the first strong reflection over 1–8 kHz is 12.8 dB, and the mid-frequency mean reverberation time is 0.11 s, both in compliance with the ITU-R BS.1116-3 (Ref. 20) recommendations. Room acoustical parameters calculated from the reproduced hall RIRs in the listening room correlate well with the parameters calculated directly from the concert hall RIRs.[15] The background noise level measured at the listening position L_{AS} was 22 dB with the ventilating system at

TABLE I. Loudspeaker directions for the reproduction systems in the listening room and the anechoic chamber, denoted as azimuth degrees at various elevations. Elevation/azimuth 0°/0° is directly in front of the listener and positive azimuth is to the left.

elevation	listening room	anechoic chamber
90°	0°	0°
45°	±90°	0°, ±90°, 180°
30°	0°, ±45°, ±135°	
22°		0°, ±30°, ±55°
0°	0°, ±22.5°, ±45°, ±67.5°, ±90°, ±135°, 180°	0°, ±10°, ±20°, ±30°, ±40°, ±50°, ±60°, ±75°, ±90°, ±105°, ±135°, 180°
−22°		0, ±30°
−35°	±40°, ±150°	
−45°		0, ±90°, 180°

minimum. The ventilation system occasionally produces noise that increases the measured value by a few decibels. The listener was surrounded by an acoustically transparent curtain for a neutral visual appearance.

2. Anechoic chamber

The reproduction system in the anechoic chamber ($V \approx 200\,\mathrm{m}^3$) consists of 39 loudspeakers (Genelec 8030B) located at a 2.2 m nominal distance from the listening position. The loudspeaker positions are detailed in Table I. The channel levels were calibrated to within 0.5 dB of target level measured at the listening position. An acoustically transparent dark curtain was drawn around the listening position and the lights were dimmed in order to exclude loudspeakers and most of the room from field of view.

B. Room impulse responses

The room impulse responses (RIRs) were taken from loudspeaker orchestra[16] (LSO) measurements of four unoccupied concert halls (listed in Table II).

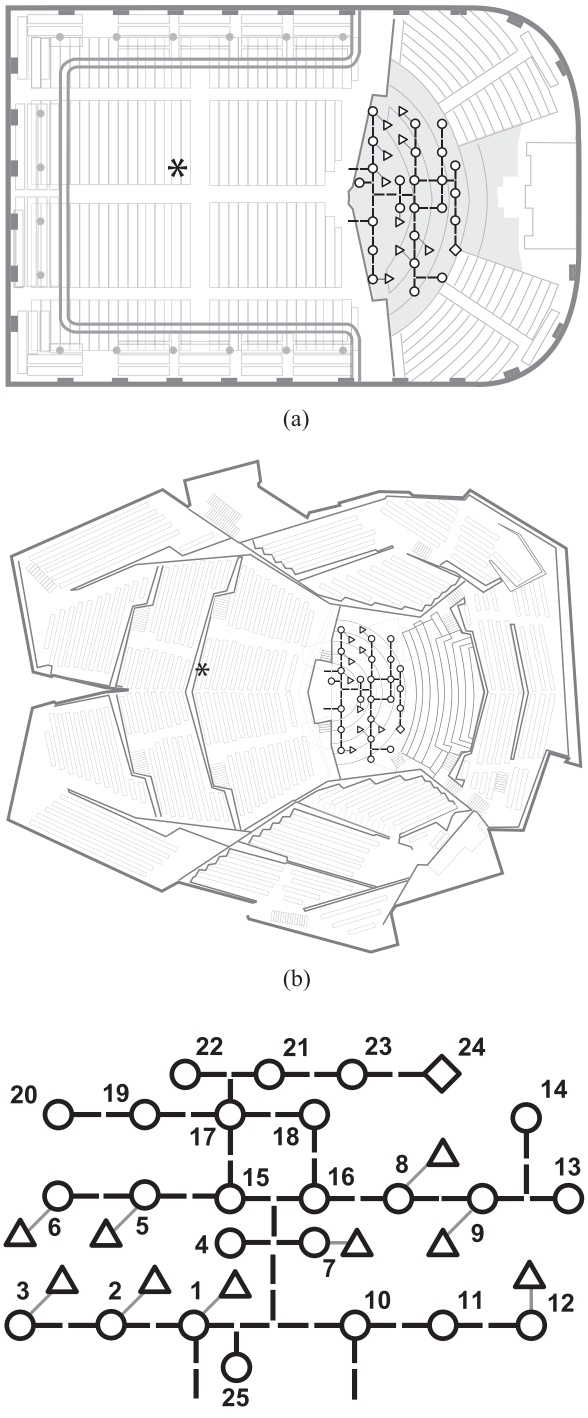
The LSO is a measurement tool to capture concert halls' acoustics. It consists of 34 loudspeakers set on stage to approximate the directivity of a symphony orchestra. The loudspeakers represent 25 channels, of which nine feature auxiliary loudspeakers directed towards the ceiling to approximate directivity of string sections. Measurement of a listening position consists of recording an impulse response with a 3D microphone array (G.R.A.S 50VI-1 Vector Intensity Probe) separately for each of the 25 channels of the LSO. The LSO is set up similarly in each hall, and thereby enables close comparisons between halls. Figure 1 shows the LSO channel numbering, and a schematic of the positioning of LSO and listener/measurement position in two halls.

The measured impulse responses were decomposed into image sources with the spatial decomposition method.[18] The image sources were allocated to the spatial sound reproduction systems as a multichannel convolution reverb by spatially discretizing the image source directions to the nearest reproduction loudspeakers.

From each hall, RIRs at a listening position at 15 m from the edge of the LSO, and 2 m away from the midline of the hall were used. This position was chosen because it is a representative listening position at approximately midway in the stalls, and it is within the range where the SDE attenuation is at maximum, which is typically between 15–19 m in concert halls.[1,2] Table II shows room acoustic parameters for the halls, taken as averages over the LSO channels.

TABLE II. Room acoustic parameters (LSO channel averages; channels 19, 20, and 25 excluded) for the original hall RIRs. Following ISO 3382-1:2009 (Ref. 17), the values are octave band averages calculated over the 500 and 1000 Hz bands except J_{LF} for which the averages were calculated over the 125 to 1000 Hz bands. $\mathrm{D/R}_{15}$ is a modified direct-to-reverberant ratio with 15 ms direct sound integration time to correspond with the present study.

Hall	G (dB)	C_{80} (dB)	EDT (s)	J_{LF}	$\mathrm{D/R}_{15}$ (dB)
Amsterdam Concertgebouw (AC)	3.1	−2.5	2.4	0.24	−7.7
Berlin Philharmonie (BP)	2.5	2.2	1.9	0.10	−2.1
Helsinki Music Centre (HMC)	3.0	1.7	2.0	0.22	−2.2
Munich Herkulessaal (MH)	3.0	−0.7	2.1	0.24	−7.6



(a)

(b)

(c)

FIG. 1. Diagrams of the loudspeaker orchestra (LSO) and the listening positions (∗) in (a) AC and (b) BP, and (c) of the LSO channel numbering (grid spacing 1 m). Different symbols denote different loudspeaker models.

C. Direct sound

In order to study the perceptual effect of direct sound spectrum coloration, it was necessary first to establish a neutral direct sound that corresponds to the source loudspeakers in free-field conditions (hereafter referred to as *freefield*).

The *freefield* direct sound was created by removing the first 15 ms response from each LSO channel RIR, and replacing it with the anechoic response of the corresponding loudspeaker(s). The original between-channel timing of the LSO was preserved in the process. A window of 15 ms was used because in the presently investigated concert halls the seat-dip effect (SDE) forms between 7 and 10 ms after the start of the impulse, and remains prominent until about 15 ms. This window may include additional stage and floor reflections, but excludes wall reflections for most source-receiver configurations. It can be assumed that any stray reflections within this time window are perceptually fused into a single source event, based on the law of the first wavefront.[19]

Finally, the loudness of the *freefield* direct sound was equalized to the same level as the original direct sound in order to preserve the orchestral balance. This was done separately for each LSO channel in each hall. Both direct sound versions were convolved with the corresponding anechoic orchestra track, and the loudness of the *freefield* version was matched with the original. The loudness was computed with the algorithm described in ITU-R BS.1770-4.[21] The calculation is essentially an equivalent sound level calculation with prefiltering to take approximately into account the effect of the human head and the frequency dependence of hearing sensitivity.

D. Seat-dip filters

The analysis of the initial seat-dip effect (SDE) in previous measurements with the LSO showed roughly two different types of average spectra.[6] First, open seats on a flat floor enable multiple paths with varying lengths between the seats, which typically results in a wide dip with maximum attention around 200 Hz, and attenuation range that may extend up to about 1 kHz. Second, closed seats on a raked floor allow only a few fixed paths, thus generating a narrow dip at around 100 Hz that is considerably deeper than in the case of the open seats. Based on these observations, two types of initial SDE were modelled approximately after magnitude responses measured at 15 m, averaged across the LSO sources and several concert halls, with (1) open seats and flat floor (*wide* filter) and (2) closed seats and raked floor (*narrow* filter).

A finite impulse response (FIR) filter was designed for both types of SDE by specifying the target magnitude responses and using a frequency sampling method. Approximating the initial SDE with a fixed filter response is a reasonable estimate due to the perceptual fusion of sound events within the 15 ms time window. Furthermore, the initial SDE averaged over a large number of sources is better approximated with a fixed filter response than the initial SDE of a single source, as the former varies less from seat to seat, and the dip minimum and frequency vary more smoothly over time compared to the latter.[6]

Since the aim was to selectively target the low frequency range, long FIRs were necessary. To keep the filter length within reasonable limits, minimum phase spectral factors were computed and the responses were truncated by windowing. Figure 2 shows the magnitude and phase responses of the two filters. The wide filter has approximately 15 dB maximum attenuation at around 200 Hz, and a wide attenuation range with upper cutoff frequency (-3 dB) of 900 Hz. The narrow filter has about -21 dB maximum attenuation around 110 Hz, and an upper cutoff of 200 Hz. Both filters have lower cutoffs at around 50 Hz.

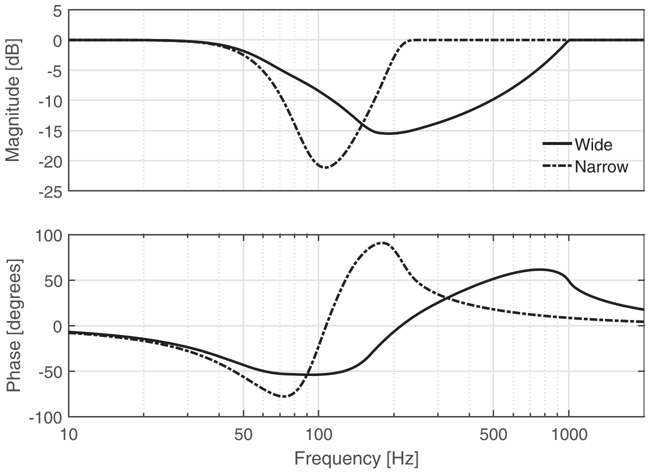


FIG. 2. Magnitude and phase responses of the *wide* and *narrow* seat-dip filters.

Each LSO channel *freefield* direct sound component was processed with the same filter, which amounts to each source on the stage undergoing the same initial SDE. No further loudness equalization was applied after the filtering since potential loudness differences are an intrinsic part of the SDE. It should be noted that this approach covers the initial SDE in an average sense and therefore enables a well-defined SDE spectrum based on measured data.

E. Anechoic orchestra excerpts and generation of stimuli

Two excerpts were selected from multichannel anechoic recordings of orchestral music,[22] based on their suitable low frequency content. The excerpts were taken from A. Bruckner's Symphony No. 8, 2nd movement, bars 36–40 (BR1, duration 5.5 s), and bars 58–61 (BR2, duration 4.6 s). BR1 features strings, flutes and oboes, and tuba. BR2 features the same instruments and in addition trumpets, trombones, bassoons, and timpani. BR2 also includes French horns, but they were left out because the corresponding LSO speakers radiate backwards to better approximate the directivity of the instruments. It was considered that the direct sound replacement method would alter them rather unnaturally. Because the excerpts did not contain audio on all channels, and because of the exclusion of the French horns, LSO channels 19, 20, and 25 were left out for both excerpts, and additionally channels 21–23 were left out for BR1 (empty tracks). The lowest notes in both excerpts is played by tuba; B1 (62 Hz) in BR1, and Eb2 (78 Hz) in BR2.

The multichannel excerpts were convolved with the corresponding LSO channel direct sound (0 to 15 ms) and reflection responses (15 ms onward) for each hall. Direct sound and reflection components were summed channelwise to form the complete LSO channel stimuli, and finally they were all combined into a single stimulus for the experiments.

In preliminary listening, it was noted that BR2 was considerably louder than BR1, and was attenuated by 4 dB relative to BR1 to optimize listening levels in tests D1 and P1. In order to have a similar experience of loudness for excerpt BR2 in the anechoic chamber, the listening levels were set slightly higher compared to the reproduction in the listening room. This difference is reflected in the measured L_{Aeq} levels that are somewhat higher in the anechoic chamber. Table III shows the L_{Aeq} ranges for the various stimuli measured in both listening spaces.

IV. LISTENING TESTS

A series of listening tests (summarized in Table III) was conducted in order to study the perceptual effect of the initial SDE. The discrimination tests (D) and preference tests (P) were run in both a listening room with acoustical treatment (1) and in an anechoic chamber (2). The latter set of tests was performed to alleviate concerns about potential effects of reproduction room acoustics on the results, and to augment the data.

The discrimination of direct sound differences was studied between the *freefield* version and two seat-dip filtered versions: *wide* and *narrow*. The purpose of the preference tests was to find out which of the direct sound versions would be preferred, if any. Since testing for preference is meaningful only when stimuli can be discriminated, only participants with the best discrimination scores, and the halls that showed a tendency for discrimination results above threshold were selected for the preference tests.

A. Methods

The discrimination tests were conducted as triangle tests, and a threshold above the chance level was chosen as the criterion for discrimination. This was done because using significant difference from chance level is a questionable criterion since the significance of the proportion correct depends on the panel size. The setup and analysis of such tests are described by Lawless and Heymann.[23] Above the threshold, the initial SDE was considered to be discriminable enough to present a potential perceptual problem. A 50% threshold was chosen as it is considered a working definition for a threshold in psychophysical tests,[23] and because it facilitates comparison with the results of Davies *et al.*,[9] who also used a 50% threshold to define a threshold of perceptibility. After adjustment for the 1/3 chance probability in the triangle test, the 50% discrimination threshold corresponds to 66.7% proportion correct in the actual test.

In order to obtain the panel size required by the set effect size, replications were used. However, to allow pooling of the results without adjustments, the replications must be independent. The independence of the results was assessed with Tarone's Z-test.[24] When the Z-test result is not significant, the replications can be pooled without further considerations. In the listening room (D1), the Z-test was found significant in AC for the *freefield-narrow* comparison and in BP for the *freefield-wide* comparison when both themes were combined. In the anechoic chamber (D2), the Z-test was significant in MH for the *narrow-wide* comparison for the tests. For these cases, the overdispersion γ was calculated according to Liggett and Delwiche[25] and used to adjust the panel size.

The preference tests were conducted as paired comparison tests, and the results were analysed with BTL, a probabilistic choice model developed by Bradley and Terry[26] and Luce.[27] When the data of the participants and repetitions are aggregated, the resulting model indicates the probability of preferring one stimulus over the others. The goodness of fit of the BTL-model can be tested by comparing its likelihood to that of a saturated model that fits the data perfectly.[28] In all the cases, the BTL-model was found to account well for the preference data as its likelihood was not significantly different from a saturated model at $\alpha = 0.05$.

Prior to all tests, the participants were given written instructions on paper. They were told that they would hear orchestral music in the acoustic setting of concert halls, and that their task was to pick the odd one out of a set of three stimuli (D1, D2), or to choose their preferred stimulus among two options (P1, P2). The nature of the differences between the stimuli was not specified, in order to prevent biasing the participants' focus on any specific attribute. After the discrimination test, the participants were asked to write down the types of differences they heard between the stimuli. For the preference test, the participants wrote down the reason for their preferences after each comparison.

B. Tests in the listening room

1. Participants

Thirteen participants took the discrimination test D1 (mean age 31, SD 4) and eight of them took part in the preference test P1. The participants were all male and work as researchers in the field of signal processing and acoustics,

TABLE III. Summary of the listening tests. *Time* is the approximate average duration of the tests. The obtained panel size is *number of participants* × *number of repetitions* per excerpt. The required panel size shown in parenthesis has been calculated with $\alpha = 0.05$ and $\beta = 0.05$ according to Eq. (5.1) in Ref. 23. Discr. = discrimination, pref. = preference.

Test	Space	Type	Halls	Excerpts	L_{Aeq}	Time	Panel size (req.)
D1	Listening room	Discr. (triangle)	AC BP	BR1 BR2	70–73 dB	20 min	$13 \times 2 = 26$ (22)
P1	Listening room	Pref. (paired comp.)	BP	BR1 BR2	70–73 dB	15 min	$8 \times 3 = 24$
D2	Anechoic chamber	Discr. (triangle)	AC BP HMC MH	BR2	74–75 dB	30 min	$11 \times 4 = 44$ (22)
P2	Anechoic chamber	Pref. (paired comp.)	AC BP HMC MH	BR2	74–75 dB	30 min	$7 \times 3 = 21$

and had previous experience participating in listening experiments. One participant reported a mild hearing impairment within the 1–2 kHz octave bands, but was kept in the panel because of his good discrimination performance. Furthermore, a review by Bech and Zacharov[29] of the pertinent literature concludes that there is no clear relationship between listening test performance and hearing acuity as measured by a standard audiometric test.

2. Discrimination of seat-dip filtered direct sound (D1)

The results in Fig. 3 show that for AC, none of the cases significantly exceeded the chance level, and were therefore not at all discriminable. In BP, *freefield-narrow* was not discriminable, but the discrimination threshold was exceeded in case of *freefield-wide* for both musical excerpts, and for *narrow-wide* with BR2. However, none were significantly above the threshold level, and therefore not clearly discriminable. Many participants commented that the test was hard. The most commonly perceived differences mentioned by the participants were timbre, bass, attack, articulation, and reverb.

3. Preference of seat-dip filtered direct sound (P1)

In P1, only the stimuli for BP were used since D1 results were below the discrimination threshold for AC in all cases. This was true also when only the results of the participants selected based on the best discrimination scores were analyzed. Although the threshold was not exceeded in every case for BP, all the cases were used in the preference test in order to obtain a complete BTL-model.

The results in Fig. 4 show a significant preference for *freefield* over *narrow* and *wide* for both excerpts. The adjusted χ^2-test at $\alpha = 0.05$ indicates that the preference probabilities differ significantly for the excerpts both separately and combined [BR1: $\chi^2(2) = 32.9$, $p < 0.0001$; BR2: $\chi^2(2) = 48.9$, $p < 0.0001$; BR1 and BR2 combined: $\chi^2(2) = 38.4$, $p < 0.0001$].

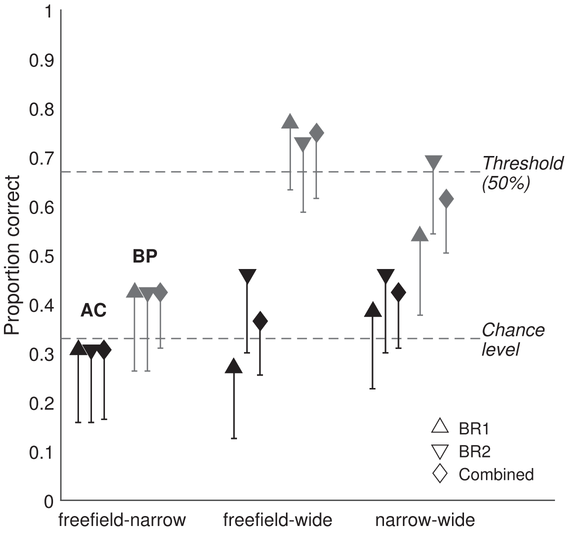


FIG. 3. The results for the discrimination test (D1) between different direct sound versions in AC and BP with the corresponding 95% confidence intervals. The interval is one-tailed since the question is whether results are significantly above the 50% discrimination threshold (perceptibly different).

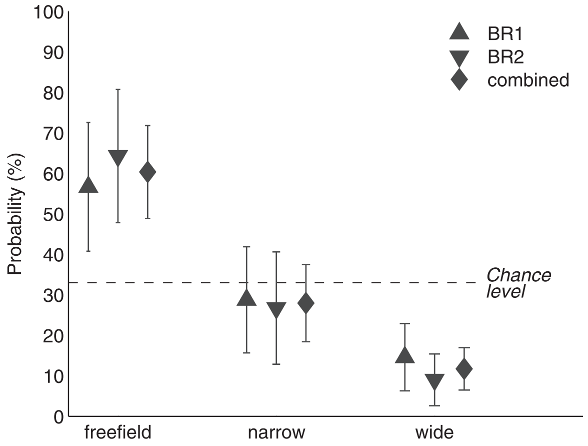


FIG. 4. The preference test (P1) results for BP as the probabilities obtained from the BTL-model with the corresponding 95% confidence intervals.

The highest agreement between the participants was that of the cases where they chose *freefield* over either *wide* or *narrow*, 30% were linked with a higher perceived amount of bass. The second highest agreement was with timbre-related attributes (18%), followed by articulation and spaciousness/envelopment. Between *narrow* and *wide*, overall the participants preferred the *narrow* version.

Interestingly, a clear order of preference was observed in P1, although none of the cases for BP clearly exceeded the discrimination threshold in D1. Particularly, *freefield* was significantly preferred over *narrow*, although in D1 the discrimination results for this pair were not significantly different from the chance level. A number of reasons may explain this difference. One is that only the participants with the highest discrimination rate in D1 were chosen for P1, and could therefore probably hear the differences more reliably. Also, the paired comparison test is less taxing since it involves less listening, and also due to the different cognitive strategies involved in making the decisions.[25]

C. Tests in the anechoic chamber

On closer inspection it was noticed that some of the loudspeakers reproducing the direct sound had somewhat attenuated low frequency magnitude responses in the listening position due to the response of the listening room. Since this was seen as a potential cause for the low discrimination performance, a second set of tests was conducted in an anechoic chamber to confirm and augment the results.

Since the results in D1/P1 for both excerpts BR1 and BR2 were qualitatively similar, the second series of tests were performed only for BR2, and the number of halls was increased to four in order to obtain a more general view on the matter.

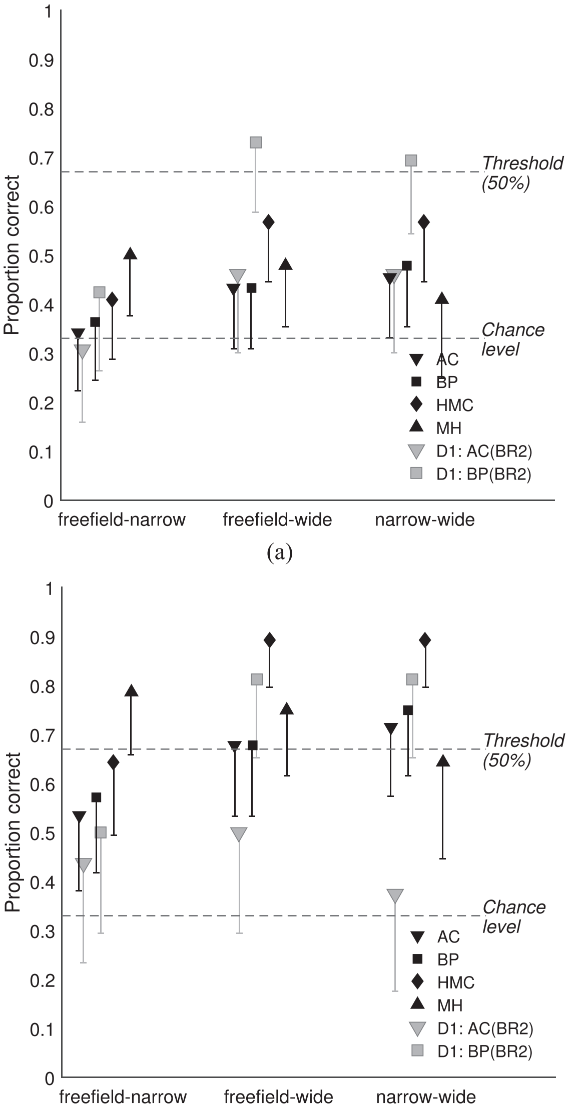
1. Participants

Eleven participants (mean age 30, SD 3) took the listening test D2, and seven of them took the test P2. The participants were all male and work as researchers in the field of signal processing and acoustics, and had previous experience participating in listening experiments. Five of these participants had also taken part in the previous tests in the listening room.

2. Discrimination of seat-dip filtered direct sound (D2)

The results in Fig. 5(a) show that none of the cases in any of the halls was clearly discriminable. Compared to D1 results, shown with light gray markers, the results of D2 are closely similar in AC for all cases, but considerably different in BP between freefield-wide and narrow-wide cases. Again, many participants commented that the test was hard. They found differences in timbre, bass, attack, articulation, and reverb, among others. There was less consensus on the attributes compared to D1.

However, when only the results of the seven best performing participants are considered, the discrimination performance is markedly improved [Fig. 5(b)]. Note that the order of the D2 results between halls and cases stays the same between Figs. 5(a) and 5(b), which indicates that the selection of the results of the best discriminating participants may be considered as a removal of noise from the results. However, only the results in HMC for the *freefield-wide* and *narrow-wide* cases are significantly above the threshold, and therefore clearly discriminable. The threshold is also exceeded, but not significantly, in MH for the *freefield-narrow* case, in all halls for *freefield-wide* case, and in AC, BP, and HMC for the *narrow-wide* case. Apart from BP in the *freefield-wide* and *narrow-wide* case, the discrimination performance is better in D2 compared to D1, when only the selected participants' results are considered.



(a)

(b)

FIG. 5. The results for the discrimination test D2 with the corresponding 95% confidence intervals for (a) all eleven participants, and (b) seven best performing participants. The light grey markers show the corresponding results from D1 in the listening room for comparison with (a) all thirteen participants, and (b) eight best performing participants. The confidence interval is one-tailed since the question is whether results are significantly above the 50% discrimination threshold (perceptibly different).

3. Preference of seat-dip filtered direct sound (P2)

All four halls were included in the preference test, as two out of three cases were discriminated above the threshold in each hall by the seven best performing participants.

The results in Fig. 6 show a preference for the *freefield* case and the least preference for the *wide* case, similarly to the results in the listening room. The significance of this preference is visually apparent only for HMC, but according to the adjusted χ^2-test at $\alpha = 0.05$, the preference probabilities differ significantly also in the other vineyard hall BP [BP: $\chi^2(2) = 10.02$, $p = 0.007$; HMC: $\chi^2(2) = 49.67$, $p < 0.0001$]. A significant preference was obtained for BP in both listening spaces, although the discrimination results were not significant.

Similarly to P1, the most frequently quoted attribute for preferring the *freefield* version (in any hall) was higher perceived amount of bass (30%). One participant used the term fullness, but in the analysis it was grouped together with bass. Otherwise the comments were varied. In both shoebox halls AC and MH, there was no significant preference towards any direct sound version.

V. DISCUSSION

The results of the discrimination tests show that although some seat-dip effect (SDE) related direct sound

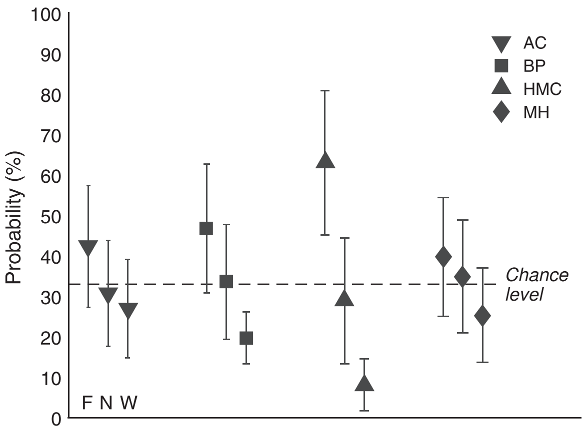


FIG. 6. The preference test (P2) results as the probabilities obtained from the BTL-model with the corresponding 95% confidence intervals. For each hall the order of the direct sound versions is *freefield* (F), *narrow* (N), and *wide* (W).

colorations may be discriminable, they are probably not very salient. Looking at the results of all participants, the 50% discrimination threshold was not significantly exceeded between any of the direct sound versions in any of the concert halls [Fig. 5(a)]. However, when taking into account only the best discriminators, the result was significant in HMC between *freefield-wide*, and *narrow-wide* [Fig. 5(b)].

Furthermore, it appears that the *wide* version was more readily discriminated from *freefield* than the *narrow* version in all halls except MH. This observation is in line with a result obtained by Bücklein, who found in his study of colorations that wide dips were easier to detect than narrow dips when white noise, speech, and music were presented diotically.[30]

In order to compare the present results with the detection thresholds obtained by Davies *et al.*,[8,9] the energy changes (ΔE) caused by the SDE filtering were calculated for the *freefield-wide* and *freefield-narrow* cases. The energy changes were calculated in various low-frequency bands since the modelled SDEs span a wider range of frequencies. Also, several time windows were used, including the full length of the impulse response, in order to show the cumulative leveling effect of the early and late reflected energy on the initial SDE. The original threshold value reported by Davies *et al.* (hereafter referred to as $\mathrm{THR}_{80,200}$) was calculated over the 200 Hz octave band in a test that studied the perceptual effect of changes in early energy (80 ms) in that octave band exclusively. Therefore, the threshold cannot be considered directly applicable to all kinds of seat-dip spectra, nor to octave bands other than 200 Hz, but here it is used for reference since no other result exists in the literature. Davies *et al.*[8] also extrapolated a threshold for the direct sound only from their results using an integration limit of 0–18 ms (here $\mathrm{THR}_{18,200}$). The time limit corresponds to the present 15 ms direct sound window, since it encompasses the energy change related to the initial SDE filtering.

Figure 7 shows the ΔE values calculated as the total change in the energy sum of the LSO channels in halls where the direct sound versions were least (AC) and most discriminated (HMC). The values are shown only for two halls since the other two are quite similar; the changes in MH are similar to those in AC with the exception that the energy increase is faster at all frequency bands, and the changes in BP resemble those in HMC, except at the 125 Hz octave band, where the energy increase is faster in BP. Note that the ΔE_{15} values are independent of the hall.

Both ΔE_{15} and ΔE_{80} values at 200 Hz are similarly positioned with their respective thresholds in all halls, which indicates that as such the thresholds proposed by Davies *et al.* cannot account for the differences in the discrimination results between the halls. This is understandable since the changes caused by the SDE filters are not limited to the

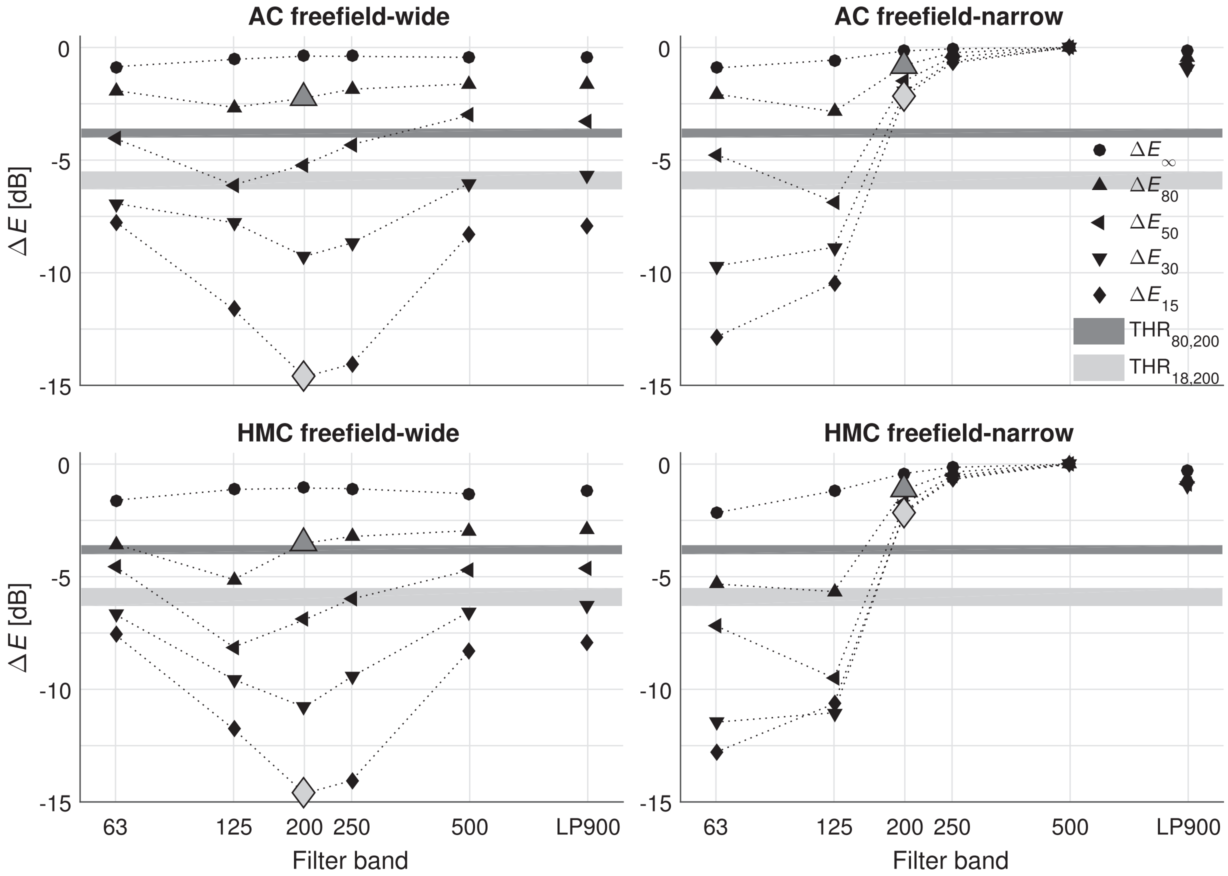


FIG. 7. The changes in the energy sum over the LSO channels between the SDE-filtered versions (*wide*/*narrow*) and the *freefield* versions calculated over various low frequency octave bands, and the overall low frequency band (lowpass-filtered at $f_c = 900$ Hz), within various time windows (ΔE_{15} to ΔE_∞) for AC and HMC. Larger grey markers denote the values corresponding to the octave band and time limits used by Davies *et al.*, and shaded horizontal areas show the corresponding perceptual threshold limits ($\mathrm{THR}_{80,200} = -3.8 \pm 0.2$ dB and $\mathrm{THR}_{18,200} = -5.9 \pm 0.4$ dB), reported in their two papers (Refs. 8 and 9).

200 Hz octave band. However, comparison of the thresholds with the ΔE_{80} values at other octave bands shows that the values at 125 Hz seem to be in line with the discrimination results. Namely, in the *freefield-wide* case, the ΔE_{80} value is below $\mathrm{THR}_{80,200}$ at 125 Hz in HMC, but not in any other hall. However, the same applies for the *freefield-narrow* case at 125 Hz, although the discrimination result was below the threshold for HMC. As to the ΔE_{80} values at 63 Hz, they are below $\mathrm{THR}_{80,200}$ for both HMC and BP, although discrimination test results for BP were not significantly above the threshold in any case. Thus, estimating the perceptibility of the initial SDE with a single threshold seems problematic.

Regarding the preference of different direct sound versions, in the vineyard halls (BP and HMC) *freefield* was favored over the filtered versions, and based on participants' comments, the preference was most commonly linked with stronger bass (P1, P2) and quality of timbre (P1). The preference for the uncolored *freefield* direct sound is in line with the results of Takahashi *et al.*,[14] which show that the direct sound with no reflections has the best tonal balance. Of the filters, *narrow* was significantly preferred over *wide*, which further supports the notion that participants preferred minimal coloration. In the shoebox halls, there was no significant preference.

The difference in the preference results between the vineyard and shoebox halls could be related to the difference in the ratio of the direct and reflected energy. Namely, AC and MH have clearly lower C_{80} and $\mathrm{D/R}_{15}$ values than BP and HMC (see Table II). In addition, the final ΔE_{∞} values in Fig. 7 are closer to 0 dB for AC (and MH) than HMC (and BP) at the frequency bands affected by the SDE filtering. Thus, the SDE-filtered direct sound represents a smaller proportion of the overall energy in the shoebox halls than in the vineyard halls. Overall, the results show that the perceptual severity of the direct sound changes appear to be reduced by the presence of adequate reflected energy. However, no conclusions can be made about the relative importance of early and late reflected energy, since shoebox halls seem to have more of both compared to the vineyard halls. In some cases (like MH), the direct sound changes may be borderline audible, but insignificant from the point of view of sound quality since there is no clear preference associated with the changes.

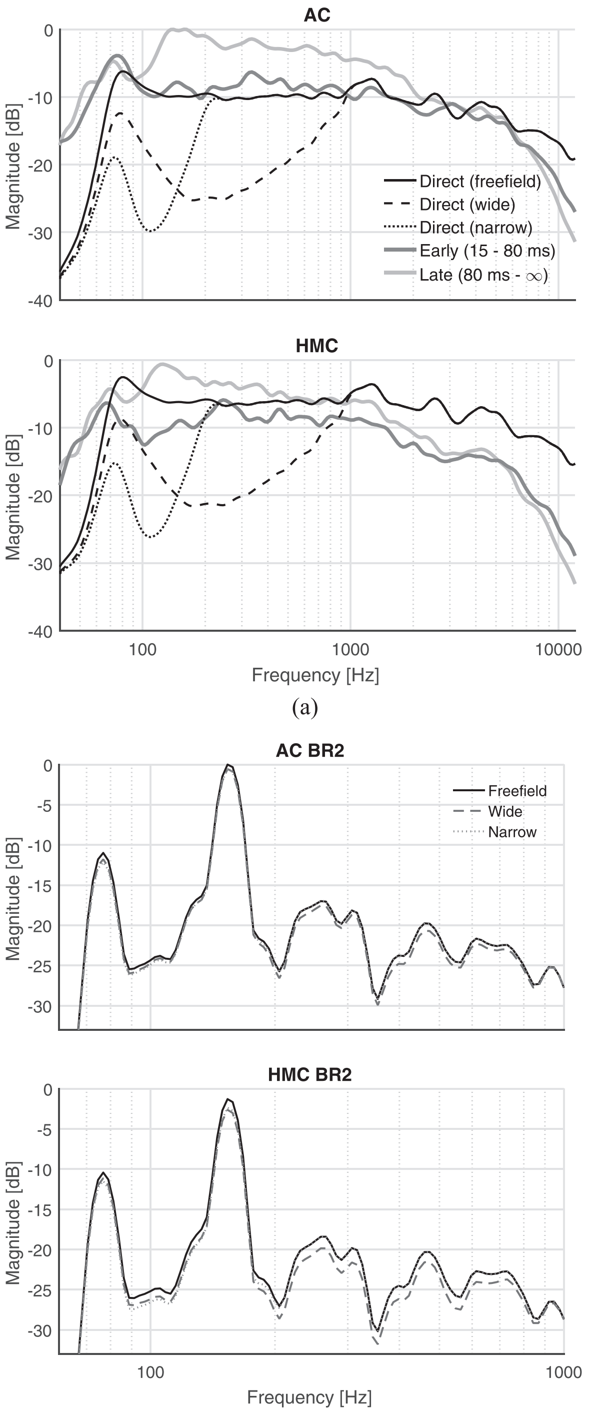
The ratio between direct and reflected energy changes also within the hall as a function of receiver distance, and therefore the discrimination results might look different at a different receiver position. In fact, informal listening in the listening room showed that the difference between SDE-filtered and freefield direct sound was notably easier to discriminate at 11 m than at 15 m, likely because closer to the source the direct sound represents a larger proportion of the overall sound. Therefore, if the differences are audible at 15 m, they would also likely be audible at 11 m. Although coloration of direct sound is probably easier to detect closer to the stage, the initial SDE is also less prominent.[2,6,31] Based on the present results, the audibility of the initial SDE may present a problem in concert halls at distances closer than about 15 m. However, at distances further than this, it is not likely to be a problem since the initial SDE remains approximately constant while the direct-to-reverberant ratio decreases. It is therefore possible that SDE-related direct sound colorations are perceptually relevant only in a fairly small proportion of the seats, as at least half of the seats at the stalls of a typical concert hall are further away than 15 m, and balconies tend to have a reduced or non-existent SDE.[2]

While the preference results in both listening spaces were similar, there were some differences between the discrimination tests. First, AC was better discriminated in D2 than in D1 when considering the best performing listeners. Second, BP was better discriminated in D1 than in D2, but among the best performing listeners the results are more similar between D1 and D2. This disparity may be caused by differences in the listening space/reproduction system, in the number of direct sound reproduction speakers (D1: three, D2: five or six), or differences in the participants.

In order to understand the rather low discriminability of the initial SDE, it is worth considering how it is portrayed by different analysis methods. Figure 8(a) shows the magnitude responses of various direct sound versions separately alongside the early and late reflections for two of the halls: it is clear that the SDE-filtered direct sound magnitude responses appear severely compromised. However, the direct sound plays a minor role in the magnitude responses taken over the full excerpts as can be seen in Fig. 8(b); the largest differences between the spectra amount to about 2 dB in HMC between *freefield* and *wide*, and the corresponding differences in AC are minor. Based on the present series of listening tests, it appears that the latter method of analysis gives a more realistic estimate of the perceptual effect.

There are a few other factors that may affect the perceptual salience of the SDE. On one hand, compositions with particularly strong low frequency content are rather rare in the classical repertoire, and similarly the low end of the excerpts used in the present study are confined to a few notes. This is likely to further reduce the salience of the initial SDE, as far as the perceived strength of bass is concerned. On the other hand, the perceived strength of bass may also be affected by a low frequency boost below 100 Hz associated with seats with underpasses,[6,32] but this phenomenon and its severity in the presence of an audience is not yet fully understood. In general, the main concern in the discussion of perceptual consequences of the SDE has been insufficient bass. However, comments from the participants of the current tests imply that a more general change in timbre may well be, in some cases, the more salient perceptual effect.

To answer the question presented in the Introduction—whether it is enough to design for sufficient reflections, or whether the initial SDE needs to be reduced by seating design—the listening test results suggest that the former may be the case, as also originally hypothesized by Schultz and Watters.[1] Furthermore, in halls where the reflections are insufficient to level the initial SDE, the design of the seating area might be more critical. Given that the *narrow* filter represents a concert hall with closed seats and raked floor, and the *wide* filter a concert hall with open seats and a flat floor, both preference test results favor the former design. However, any conclusions on the perceptibility of the SDE



(a)

(b)

FIG. 8. The low frequency magnitude spectra (50–1000 Hz) corresponding to *freefield*, *wide*, and *narrow* versions of the direct sound of (a) RIRs split into direct sound and reflections (1/12-octave smoothed), and (b) the excerpts convolved with the full RIRs (1/3-octave smoothed for easier comparison).

as a whole based on this study should be treated with caution since only the *initial* SDE was controlled and the early reflections were kept unique to the hall. This means, for example, that for the case with *freefield* direct sound, the grazing-angle reflections retain their original SDE-affected spectra. Especially with closed seats and raked floor, the coloring due to SDE is strong and focused in a narrow frequency range.[6] Such coloring may be more audible when applied to both direct sound and reflections rather than just to the direct sound, especially under low reverberation conditions. Thus, further studies are required to confirm the perception of SDE-related changes in both the direct sound and the first important reflections, as well as the relative contributions of early and late energy to the perception of the initial SDE.

VI. CONCLUSIONS

The discrimination and preference between uncolored (*freefield*) and two filtered (*wide, narrow*) direct sound spectra representing the initial SDE of two different seating areas were studied with realistic concert hall auralizations. For a listening position 15 m from the orchestra edge, and 2 m from the hall midline, the discrimination tests showed that the difference between direct sound spectra was overall hard to distinguish; the discrimination results were significant only for one of four halls (HMC), two out of three cases (*freefield-wide* and *narrow-wide*, but not *freefield-narrow*), and for a subgroup of participants.

The preference was significant towards the uncoloured direct sound in the vineyard halls (BP, HMC), while no clear preference was obtained in the shoebox halls (AC, MH). Comments gathered from the participants included many attributes; most commonly cited ones were higher perceived strength of bass and timbre quality. In the literature, the SDE has been discussed primarily as having an effect on perceived bass, but the potential more general effects on timbre seem to have been overlooked. The present results indicate that it is a question worth investigating further.

The results suggest that the initial SDE can be rendered perceptually negligible by architectural design that ensures sufficient reflected energy. Moreover, the initial SDE is likely to be a concern for a rather small proportion of the seats at the stalls, as the direct-to-reverberant ratio grows with distance from the stage while the initial SDE stays approximately constant after 15–19 m.

ACKNOWLEDGMENTS

This research has received funding from the Academy of Finland, project No. [257099]. The work of H.T. has also been supported by the Emil Aaltonen Foundation. H.T. and A.H. contributed equally to this work.

[1] T. Schultz and B. Watters, "Propagation of sound across audience seating," J. Acoust. Soc. Am. **36**, 885–896 (1964).

[2] G. Sessler and J. West, "Sound transmission over theatre seats," J. Acoust. Soc. Am. **36**, 1725–1732 (1964).

[3] K. Ishida, "Investigation of the fundamental mechanism of the seat-dip effect—Using measurements on a parallel barrier scale model," J. Acoust. Soc. Jpn. (E) **16**, 105–114 (1995).

[4] J. S. Bradley, "Some further investigations of the seat dip effect," J. Acoust. Soc. Am. **90**, 324–333 (1991).

[5] J. Pätynen, S. Tervo, and T. Lokki, "Analysis of concert hall acoustics via visualisations of time-frequency and spatiotemporal responses," J. Acoust. Soc. Am. **133**, 842–857 (2013).

[6] H. Tahvanainen, J. Pätynen, and T. Lokki, "Analysis of the seat-dip effect in twelve European concert halls," Acta Acust. Acust. **101**, 731–742 (2015).

[7] T. Cox and W. J. Davies, "Amelioration of seat dip attenuation," WIT Trans. Built Env. **11**, 73–80 (1995).

[8] W. J. Davies and T. J. Cox, "Reducing seat dip attenuation," J. Acoust. Soc. Am. **108**, 2211–2218 (2000).

[9] W. J. Davies, T. J. Cox, and Y. W. Lam, "Subjective perception of seat dip attenuation," Acta Acust. Acust. **82**, 784–792 (1996).

[10] M. Barrow, "Bass sound in concert auditoria," J. Acoust. Soc. Am. **97**, 1088–1098 (1995).

[11] G. Soulodre, M. Lavoie, and S. Norcross, "Objective measures of listener envelopment in multichannel surround systems," J. Audio Eng. Soc. **51**, 826–840 (2003).

[12] L. L. Beranek, *Concert and Opera Halls: Music, Acoustics, and Architecture*, 2nd ed. (Springer, New York, 2004), p. 527.

[13] J. S. Bradley and G. A. Soulodre, "Factors influencing the perception of bass," J. Acoust. Soc. Am. **101**, 3135 (1997).

[14] D. Takahashi, K. Togawa, and T. Hotta, "Objective measures for evaluating tonal balance of sound fields," Acoust. Sci. Technol. **29**, 2–8 (2008).

[15] J. Pätynen and T. Lokki, "Concert halls with strong and lateral sound increase the emotional impact of orchestra music," J. Acoust. Soc. Am. **139**, 1214–1224 (2016).

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

[17] ISO 3382-1:2009; *Acoustics—Measurement of Room Acoustic Parameters. I: Performance Spaces* (International Organization for Standardization, Geneva, Switzerland, 2009).

[18] S. Tervo, J. Pätynen, and T. Lokki, "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**, 1–13 (2013).

[19] J. Blauert, *Spatial Hearing*, 2nd ed. (MIT Press, New York, 1996), p. 527.

[20] ITU-R BS. 1116-3, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems* (International Telecommunications Union, Geneva, Switzerland, 2015).

[21] ITU-R BS. 1770-4; *Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level* (International Telecommunications Union, Geneva, Switzerland, 2015).

[22] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," Acta Acust. Acust. **94**, 856–865 (2008).

[23] H. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices*, 2nd ed. (Springer, New York, 2010), pp. 101–126.

[24] R. E. Tarone, "Testing the goodness of fit of the binomial distribution," Biometrika **66**, 585–590 (1979).

[25] R. Liggett and J. Delwiche, "The beta-binomial model: Variability in overdispersion across methods and over time," J. Sens. Stud. **20**, 48–61 (2005).

[26] R. Bradley and M. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," Biometrika **39**, 324–345 (1952).

[27] R. Luce, *Individual Choice Behaviour. A Theoretical Analysis* (Wiley, New York, 1959), pp. 1–176.

[28] F. Wickelmaier and C. Schmid, "A Matlab function to estimate choice model parameters from paired-comparison data," Behav. Res. Methods Instrum. Comput. **36**, 29–40 (2004).

[29] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application*, 1st ed. (Wiley, Chichester, 2006), pp. 123–125.

[30] R. Bücklein, "The audibility of frequency response irregularities," J. Audio Eng. Soc. **29**, 126–131 (1981).

[31] W. J. Davies and Y. W. Lam, "New attributes of seat dip attenuation," Appl. Acoust. **41**, 1–23 (1994).

[32] W. J. Davies, "The effects of seating on the acoustics of auditoria," Ph.D. thesis, University of Salford, England, 1992.