Psychomusicology: Music, Mind, and Brain

The Preferred Level Balance Between Direct, Early, and Late Sound in Concert Halls

Aki Haapaniemi and Tapio Lokki

The Preferred Level Balance Between Direct, Early, and Late Sound in Concert Halls

Aki Haapaniemi and Tapio Lokki
Aalto University School of Science

The preferred level balance between the direct, early, and late components of the sound field in a concert hall was studied by letting listeners manipulate sound field auralizations in an adjustment procedure featuring acoustics of two halls of contrasting designs—Berlin Konzerthaus (shoebox) and Berlin Philharmonie (vineyard). The room impulse responses measured at two positions in each hall were encoded for multichannel reproduction, split into direct, early, and late components, convolved with anechoic recordings of a symphony orchestra, and reproduced in a multichannel listening room. The listeners' task consisted of optimizing the level balance between the direct, early, and late components by adjusting the level of two of the components relative to one that was kept fixed. The results reveal a strong tendency to emphasize the early component for both positions in Berlin Philharmonie, where the early reflections contain less energy due to the fan-shaped frontal auditorium and comparatively small effective side wall areas. In contrast, the median results for one position in Berlin Konzerthaus were within 1.5 dB of the actual acoustics of the hall, with statistically insignificant deviations. The relationship between the direct sound and early reflections is seen as a potential subject of further study.

Keywords: concert hall acoustics, early reflections, late reverberation, spatial impulse responses, perceptual studies

The auditory perception of a symphony orchestra playing in a concert hall can be understood with respect to two main percepts: the source presence and the room presence (Kahle, 2013). The source presence is the continuous perception of the sound sources in the hall, whether it be through a unified source perception (i.e., the full orchestra as one auditory entity), groups of instruments, or each instrument as a separate entity in the perceptual domain. The formation of the auditory streams comes about through stream segregation (Griesinger, 1997) and is subject to the perceptual grouping laws therein (Moore, 2012). The early reflections are perceptually grouped with the source streams through the precedence effect (Litovsky, Colburn, Yost, & Guzman, 1999), and affect the width, loudness, and timbre of the auditory events

(Blauert, 1997). In this way, the direct sound of the orchestra and the early reflections of the hall combine to make up the source presence. The late reflections form the room presence, that is, the context and space for the music which lends the music support, embellishment, and a sense of depth, and provides the listener with a sense of envelopment.

The balanced relationship of the early sound (i.e., direct sound and early reflections) and the late reflections is a necessary requirement for a concert hall (Meyer, 2009). It is usually expressed with the concept of "clarity" and the corresponding standard parameter C_{80} (ISO 3382–1:2009, 2009), which measures the amount of early sound energy against that of late energy. Adequate clarity is needed for appreciation of musical detail (Barron, 2009), but it is also known that the late reflections enhance the blending of the instruments of the orchestra into a "closed overall sound" (Meyer, 2009).

The crossover time for early and late reflections is often taken to be 80 ms, as is also the case for the ISO 3382–1 parameters. However, it is a simplification. In terms of the source presence, an early reflection is one which contributes to it. From the studies concerning the precedence effect, it is known that the limit of fusion is dependent on the nature of the sound stimuli (Litovsky et al., 1999). The limit is shorter for transients such as clicks, and longer for less abrupt and more continuous sounds, such as speech. The significance for concert hall acoustics is that the limit is more or less dependent on program material. It has also been suggested that the fusion limit for early reflections depends on frequency (Soulodre, Lavoie, & Norcross, 2003), with low frequency limits longer than 80 ms and high frequency limits somewhat shorter. However, the 80 ms limit may be seen as a plausible perceptual average for most symphonic music.

The relationship of the direct sound to the early reflections is also of interest. The direct sound is afflicted by the seat dip effect (Schultz & Watters, 1964; Sessler & West, 1964), which is caused by interference from diffracted and reflected sound from the audience seats. It is often seen as a broad region of attenuation in the low frequencies with a center between about 100 Hz to 300 Hz and a maximum attenuation of about 10 to 20 dB. The early reflections contribute to fill the dip in the frequency response (Pätynen, Tervo, & Lokki, 2013). Lateral early reflections also broaden the width of the auditory sources, which is a crucial effect for good concert halls (Barron & Marshall, 1981). The extent of this effect is measured by the early lateral energy fraction (standard parameter J_{LF}), which is the fraction of sound energy arriving from lateral directions within 5 to 80 ms compared with the whole sound energy arriving within 0 to 80 ms.

While the ISO 3382–1 standard parameters are useful for classifying concert halls and establishing optimum criteria for guidance, several studies point toward shortcomings of the parameters (Bradley, 2011; Kirkegaard & Gulsrud, 2011; Lokki, Pätynen, Tervo, Siltanen, & Savioja, 2011). In an earlier study, it was shown that in the most general situation with continuous music, the early sound of a hall is more salient in characterizing a hall compared with the late reflections, although the standard parameters showed less changes dependent on the early sound (Haapaniemi & Lokki, 2014). Some studies also show emergence of at least two distinct preference groups that emphasize different aspects of the perceptual space. For example, Barron (1988) found that his listeners were divided into two groups that preferred either intimacy or reverberance, while giving secondary importance to envelopment, whereas Kuusinen, Pätynen, Tervo, and Lokki (2014) found a division between listeners that preferred either definition or reverberance, while proximity was found to be the main driver of preference.

In this study, the question of optimal acoustics is explored with a listening test using an adjustment procedure with auralized acoustics of two existing concert halls. The listeners adjusted levels of the direct, early, and late sound components in order to reach a subjective optimum balance between the components based on their preferences. Two positions from two unoccupied concert halls of different types were used—Berlin Konzerthaus (shoebox; Figure 1a) and Berlin Philharmonie (vineyard; Figure 1b). Both of the halls are found within the highest class in Beranek's "Rank-orderings of acoustical quality of 58 concert halls developed from interviews and questionnaires" (Beranek, 2004), populating positions 4 and 16, respectively.

Methods

Measurements

The room impulse responses for the listening test were derived from measurements of two listening positions in two halls with a 24-channel loudspeaker orchestra (LSO) (Pätynen, 2011)—see stage areas in Figure 1. The LSO simulates the positioning of the various instrument groups on stage and their directivity in an approximate manner with 34 calibrated and carefully positioned loudspeakers. The impulse responses were measured for the same geometrical configuration of sources in each hall, for position P1 at 11 m and P2 at 19 m (see Figure 1), allowing for direct comparability between the halls. A logarithmic sweep was recorded for each channel of the LSO with a 3-D microphone array of six omnidirectional microphones (G.R.A.S Vector Intensity Probe VI50).

Processing

The impulse responses were analyzed with the spatial decomposition method (SDM; Tervo, Pätynen, Kuusinen, & Lokki, 2013) to estimate the direction of arrival of sound as a function of time. SDM is based on the assumption that an impulse response can be represented by a limited number of image sources. The average arrival direction of each discrete sample is estimated in short time windows to produce the set of image sources. Depending on the reproduction method, the image sources are assigned to one or more reproduction channels to produce a multichannel impulse response for convolution. In this study, the image sources were assigned to the reproduction loudspeakers that were closest to the estimated location. This approach was chosen over more elaborate panning methods because it is free of the coloration caused by panning the image sources between two or more loudspeakers (Pätynen, Tervo, & Lokki, 2014).
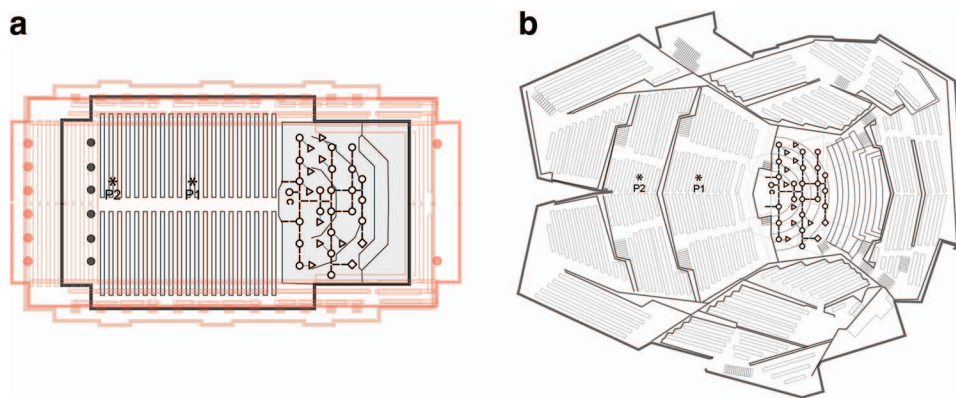

Figure 1. Positioning of the loudspeaker orchestra and the positions P1 and P2 in (a) Berlin Konzerthaus and (b) Berlin Philharmonie. See the online article for the color version of this figure.

In SDM, only one of the measured omnidirectional microphone channels is used for rendering. The sum of the reproduction channel impulse responses equals the original impulse response, and the frequency response at the sweet spot of the loudspeaker setup is identical to the original response at the measurement microphone. SDM has recently been used successfully, for example, in a study of preference of critical listening environments among mixing and mastering engineers (Tervo, Laukkanen, Pätynen, & Lokki, 2014). In the study, all but one of the sound engineers that had previous experience working in the studied control rooms were able to recognize their own control room from the total of nine reproduced control rooms.

The reproduction impulse responses representing each channel of the LSO were individually split into direct (0–15 ms), early (15–80 ms), and late (80 ms onward) components, with a 1-ms crossfade between the components to avoid discontinuities. The length of the initial silence preceding the impulse responses varied between LSO channels due to distance differences between the measurement positions and the individual loudspeakers. The 0-ms point was therefore taken to be the start of the impulse in the individual channel. After splitting the channel impulse responses into components, the relative timing was restored by adding the correct initial silence for each LSO channel. The 15-ms limit (corresponding to roughly 5 m traveled by sound) of the direct component was chosen in order to include the seat dip effect, but to exclude any wall reflections (Bradley, 1991). The 80-ms crossover time between the early and late reflections was chosen following the ISO 3382–1:2009 standard.

Acoustics of the Halls

Figures 2 and 3 present the frequency (magnitude) responses and the lateral energy distributions for the reproduction impulse responses, respectively. The analysis is based on the techniques presented by Pätynen et al. (2013), but is done separately for direct, early, and late components. The lateral energy distributions are shown as the relative amount of sound energy reproduced through the loudspeakers of the reproduction system. Note that the frequency responses of Figure 2 would be identical whether calculated directly from the measured room impulse responses or the reproduction impulse responses.

The shapes of the direct component frequency responses are quite similar for positions P1 and P2 within each hall (see Figure 2), and the overall level difference between them is an expected effect of distance. However, the frequency responses are remarkably different between the halls. In Berlin Konzerthaus (BK), the seat dip attenuation is seen to affect the frequencies between about 200–1000 Hz, with a maximum attenuation exceeding 10 dB, but below that the level of bass is still considerable. In Berlin Philharmonie (BP), the seat dip effect is very different, potentially due to the sloped seating, and it is seen as loss of sound energy below about 100 Hz. However, above that the midbass and low mid frequencies are intact and the stage floor reflection contributes more energy to the direct component due to the sloped seating area, resulting in a strong and bright direct component. The slopedness of the seating area in BP also causes more of the nearby seatbacks behind the listening positions to be exposed to the sound from the LSO. This may be seen in Figure 3 as more reflected/
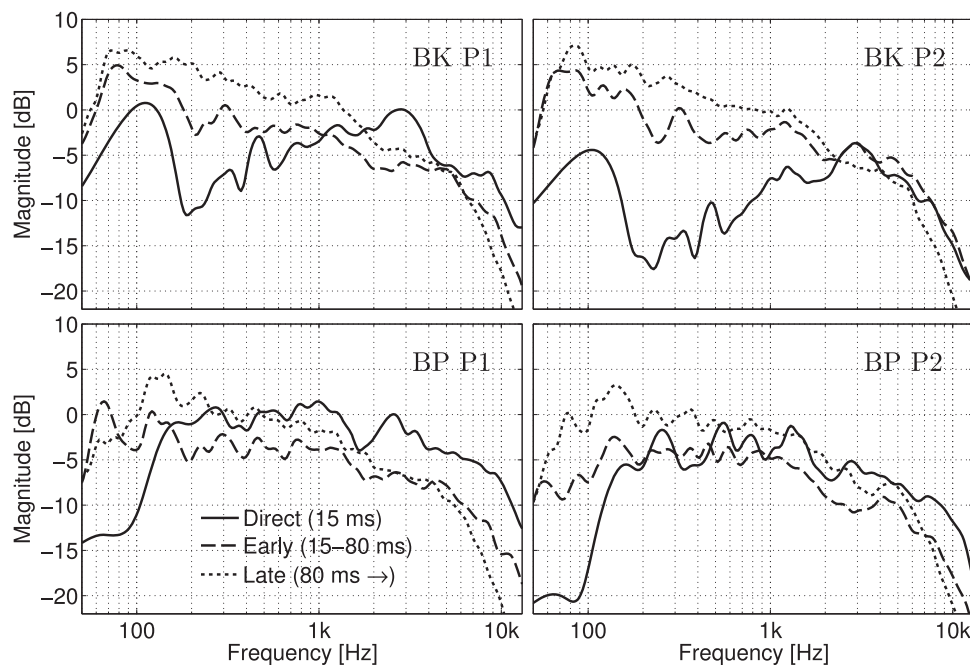

Figure 2. The frequency (magnitude) responses at positions P1 and P2 in Berlin Konzerthaus (BK) and Berlin Philharmonie (BP) for the direct (0–15 ms), early (15–80 ms), and late (80 ms →) components. The responses were averaged over 24 channels of the LSO and 1/3 octave smoothed. Scaling was done so that the full response corresponds to standard G values.
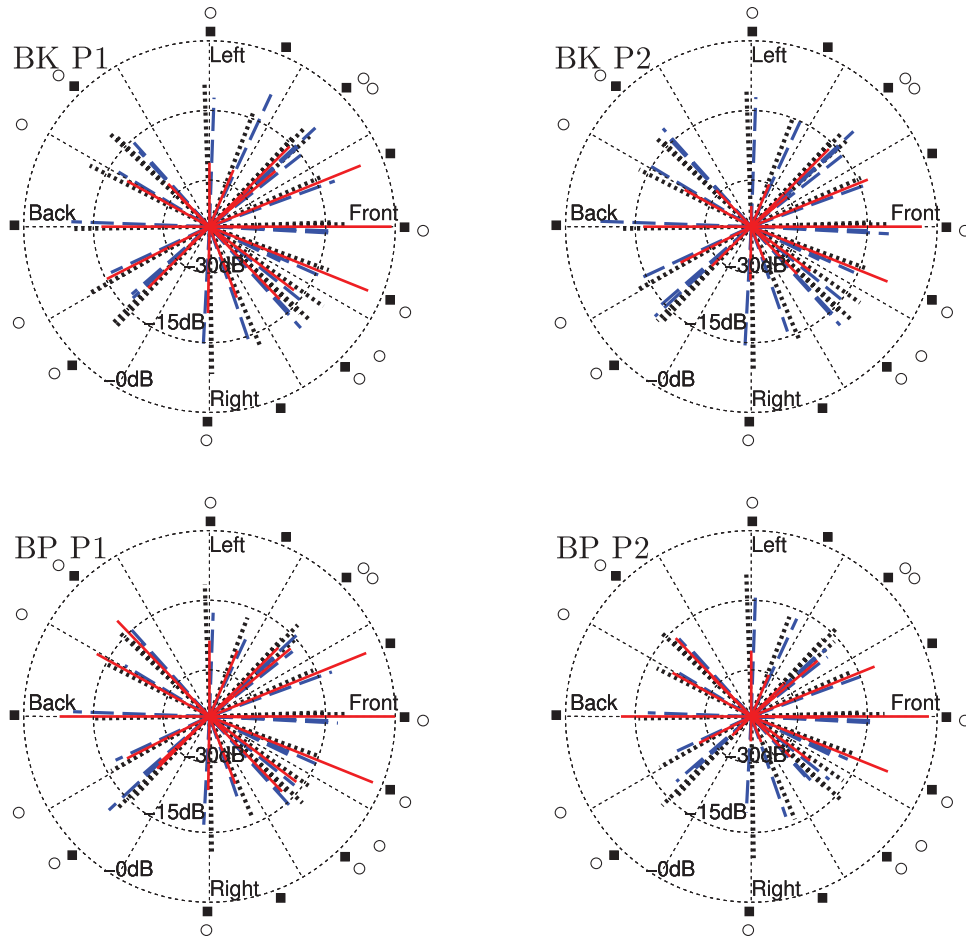
Figure 3. The lateral energy distribution as the relative sound energy reproduced through the loudspeakers of the reproduction system for the direct (0–15 ms) [solid line (red)], early (15–80 ms) [dashed line (blue)], and late (80 ms →) [dotted line (black)] components for positions P1 and P2 in Berlin Konzerthaus (BK) and Berlin Philharmonie (BP) as calculated from the multichannel reproduction impulse responses. The measured azimuth positions of the loudspeakers are shown for reference, with filled squares (■) denoting loudspeakers on the lateral plane, and empty circles (○) denoting loudspeakers at positive/negative elevation. Cosine weighting was used to compute the contribution of elevation loudspeakers to the sound energy in the lateral plane. Note that the radial distances are disregarded in depicting the loudspeaker angles, and that the lines corresponding to the early and late components are rotated 2° in opposite directions to avoid exact overlap. See the online article for the color version of this figure.

diffracted sound energy arriving from between the *back* and *left* directions for the direct component in BP compared with BK. The asymmetrical amount of energy arriving from between the *back* and *left/right* directions in BP is due to the orientation of the seating area (see Figure 1b).

The early component frequency responses are similar within BK, having a gradual downward slope from low to high frequencies, but P2 has slightly more energy above 1 kHz (see Figure 2). BP P1 has a similar early component frequency response level and shape in the mid to high frequencies compared with BK P1, but below 1 kHz the magnitude is 1 to 3 dB lower. BP P2 has the quietest early component of the four positions, with loss of energy at 100 Hz and below, and between 2 to 4 kHz. Figure 3 also shows differences in the early reflections between the halls; BK has stronger early reflections and especially those that come from between the *front* and the *left/right*

directions. This is a typical difference between shoebox and vineyard/fan-shaped halls (Pätynen, Tervo, Robinson, & Lokki, 2014). At BK P2, also the reflections that come from between the *back* and *left/right* directions are strong.

The late component frequency response is of a similar shape between the positions within each hall, but the level is somewhat higher at P1 compared with P2 (see Figure 2). Overall, the late component is stronger in BK compared with BP, and it has more of the low frequencies that are lacking in BP. Looking at the lateral distribution of sound energy in Figure 3, it is evident that the late component energy in BK is more evenly distributed among different directions compared with BP.

The ISO parameters for the five subjective aspects: reverberance (early decay time, EDT), clarity (C_{80}), apparent source width (early lateral energy fraction, J_{LF}), level of sound (sound strength,

Table 1

ISO 3382–1 Parameter Values for the Four Positions

Position	EDT (s)	C_{80} (dB)	J_{LF}	G (dB)	L_J (dB)
BK P1	2.1	−1.5	0.27	2.9	−2.1
BK P2	2.1	−2.3	0.29	1.8	−3.0
BP P1	2.1	2.4	0.11	2.5	−5.3
BP P2	1.9	0.2	0.15	1.2	−5.7

Note. The parameters are averaged over the loudspeaker orchestra channels and further averaged over the 500 Hz and 1 kHz octave bands except for J_{LF} and L_J (energy averaged), which are averaged over the 125 Hz, 250 Hz, 500 Hz, and 1 KHz octave bands. EDT = early decay time; BK = Berlin Konzerthaus; BP = Berlin Philharmonie.

G), and envelopment (late lateral sound level, L_J) were also calculated for the four positions (see Table 1). EDT is similar across both positions and halls, predicting sufficient reverberance, but L_J is 2 to 3 dB less in BP, which indicates that the envelopment aspect of reverberation in BP may be compromised. C_{80} in BK is quite low and therefore possibly insufficient, whereas in BP it is 2 to 4 dB higher. J_{LF} in BK is in the higher end of the typical range (0.05; 0.35) and BP in the lower end/middle, and both are characteristic of shoebox and vineyard type halls, respectively. The G values are about 1 dB higher for positions P1, and slightly higher in BK compared with BP.

Stimuli

The direct, early, and late components were separately convolved with 24-channel anechoic recordings of a symphony orchestra (Pätynen, Pulkki, & Lokki, 2008). The recordings were of bars 33–61 (40 s) of the second movement of Bruckner's Symphony No. 8, and bars 23–30 (26 s) of the first movement of Beethoven's Symphony No. 7. The Bruckner excerpt represents a forceful type of playing by a large orchestra with a strong brass section, including a dramatic crescendo, whereas the Beethoven piece features strings and woodwinds playing a relatively quiet and delicate, spacious passage. The Bruckner excerpt was attenuated 6 dB relative to the Beethoven excerpt prior to convolution in order to optimize the listening levels. The excerpt lengths were chosen to present coherent musical phrases in order to facilitate a holistic

listening attitude, and to prevent the psychological fatigue often encountered in listening tests of extremely repetitive nature.

Reproduction

The convolved samples were reproduced in an acoustically treated listening room with a 24-channel setup consisting of 20 Genelec 8020B and four Genelec 1029A loudspeakers, which were calibrated with 0.1 dB accuracy. The loudspeakers were positioned on five elevation levels: at 0° (ear level) [azimuth 0°, ±22.5°, ±45°, ±67.5°, ±90°, ±135°, 180°], 30° [azimuth 0°, ±45°, ±135°], 45° [azimuth ±90°], 90° (directly over the listening position), and at −35° [azimuth ±40°, ±150°]. The nominal loudspeaker distance was 1.5 m from the listening position, but the loudspeaker directly on top and those elevated on either side of the listening position were positioned at 1.2 m for practical reasons. The differences in distance were compensated with appropriate delays and attenuation in the signals fed to the loudspeakers.

The reverberation time of the listening room is 0.1 s at midfrequencies and the background noise level (A-weighted, slow) is 30 dB. The peak-to-peak level difference between the direct sound and the strongest early reflection in the 1–8 kHz band was found to be 12.8 dB on average across the loudspeakers. Therefore, the setup complies with the ITU-R BS.1116–1 (1997) recommendation for subjective audio evaluation systems, which states that the early reflection peaks should be at least 10 dB below the direct sound peak.

The nominal levels of the excerpts were measured with a Sinus Tango (Class 1) sound level meter by measuring the A-weighted equivalent SPL for the duration of the samples with each of the direct, early, and late components set at a neutral gain of 0 dB. The measured values were between 73 to 75 dB for the Bruckner excerpt, and between 61 to 64 dB for the Beethoven excerpt.

Listening Test

The listening test consisted of three adjustment tasks, one with each component (direct, early, and late) fixed, for both positions in both halls with both music pieces, that is, a total of $3 \times 2 \times 2 \times 2 = 24$ tasks, which were given to the listeners in random order. In each task, the level of one of the components was kept fixed

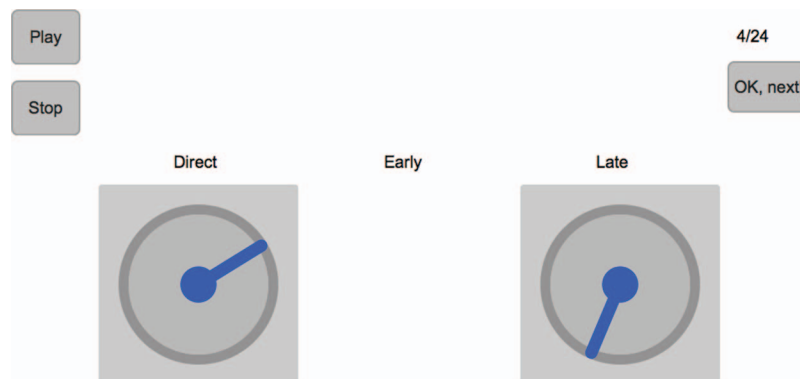

Figure 4. The listening test graphical user interface (GUI) with the endless encoders for the direct and late components shown and the encoder for the early (fixed) component hidden. See the online article for the color version of this figure.

while the listeners adjusted the levels of the other two components according to their preference, from a starting level of relative silence.

The test was implemented in Max/MSP and the listeners adjusted the levels of the components using a touchpad to control the two available endless encoders within the GUI (see Figure 4). The encoder for the fixed component was always hidden. The nominal starting levels of the adjustable components were set at -15 dB (relative to the fixed component) with a further random fluctuation of ± 3 dB to ensure that memorizing the encoder positions would be of no assistance in setting the levels similarly across tasks. The maximum possible gain of each component was set at $+15$ dB and the lower limit was set at -25 dB. However, the listeners had no means of knowing exactly when they had reached the limits, as the encoders were endless.

Before the test, the listeners were given brief written instructions that detailed the task and the interface. They were told that the task was to adjust levels of the direct, early, and late components to optimize the level balance between the components, following their own preference. A supervised practice round of two tasks was completed before the start of the test to ensure that the listeners understood the task and the user interface. They could spend as much time as they needed on the test, and most listeners spent around 30 to 60 min. After the test, the listeners filled out a questionnaire about their criteria and procedures for setting the levels where they also had the opportunity to comment on the test.

Seventeen listeners of ages between 21 and 43 with normal hearing took part in the test, including the authors. All had some form of background in music and/or acoustics, previous experience as participants in listening tests, and were at least occasional concert goers.

Results

Figure 5 shows the results of the test in the form of median values and 95% confidence intervals, grouped alternatively according to the music excerpt, listening position, and hall. Jarque–Bera and Lilliefors tests (Jarque & Bera, 1987; Lilliefors, 1967) were run for the results in order to assess whether normal distributions could suitably describe the level data. Many of the level distributions failed the tests at the 5% significance level, and therefore the confidence intervals for the median values were derived using bootstrapping (Efron, 1979) with 10,000 samples.

From Figure 5a, it can be seen that the level results are quite similar for both music excerpts. The median values for the different conditions (i.e., with a different component fixed) are generally quite near the original acoustics of the halls, but the early component was set on average above $+2$ dB whenever it was available for adjustment. From Figure 5b, it becomes clear that the tendency for the relative amplification of the early component is more prominent for position P2 compared with P1. For P1, the direct component was also set on average below -1 dB whenever it was adjustable. However, Figure 5c shows that the greatest difference in the results is seen between the halls: the listeners amplified the early component in BP more than in BK. Also, the direct component in BP tends slightly toward attenuation, and the late component toward amplification, but for both of the components the effect is significant only for one of the conditions.
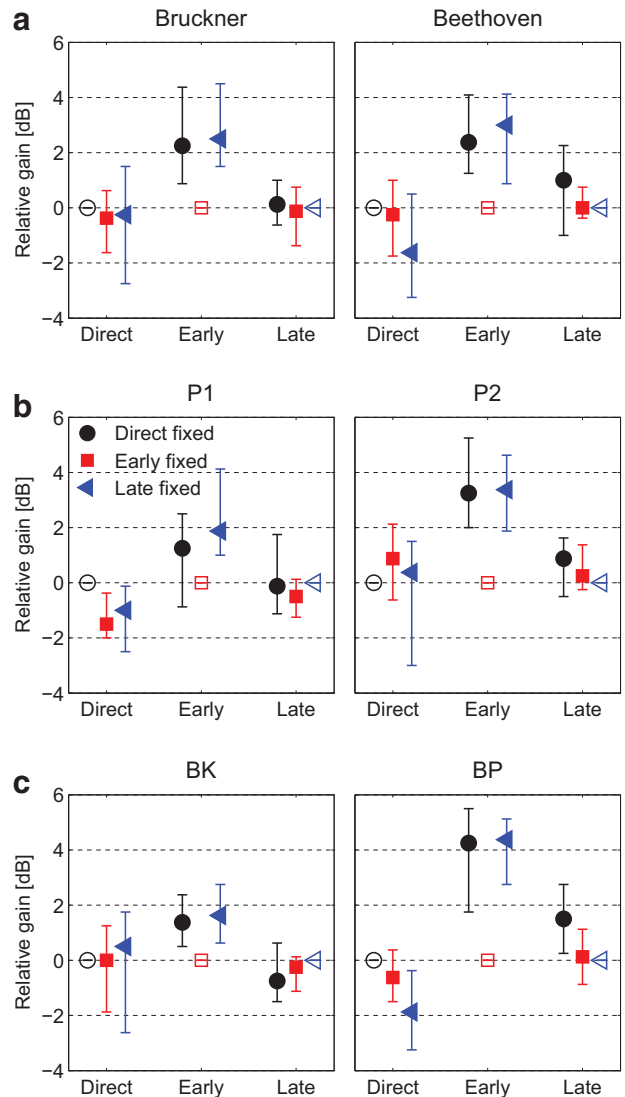

Figure 5. The median levels and corresponding 95% bootstrap confidence intervals for the results grouped according to (a) music excerpt, (b) listening position, and (c) hall. Different markers denote median levels for different conditions (with either direct component, early component, or late component fixed at 0 dB). Fixed levels are denoted by empty markers for visual clarity. BK = Berlin Konzerthaus; BP = Berlin Philharmonie. See the online article for the color version of this figure.

Seeing that the music excerpt had a minor influence on the results (Figure 5a), the results with both excerpts were grouped together and analyzed separately for all four positions (see Figure 6). For BK P1, the results are overall close to the original acoustics, and none of the deviations from 0 dB are statistically significant. In BK P2, the results are also quite close to the original acoustics, but the early component tends toward amplification whenever it was adjustable. In BP P1, the direct sound tends to be slightly attenuated, and the early reflections are clearly amplified. In BP P2, the early reflections are amplified even more, while the only other significant result is the amplification of the late component when the direct component was fixed.
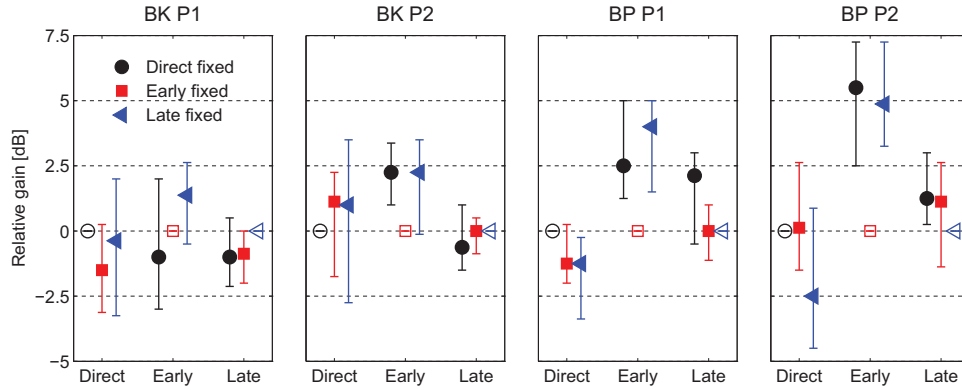
Figure 6. Median levels and 95% bootstrap confidence intervals for the results with both music excerpts grouped together. Different markers denote median levels for different conditions (with either direct, early, or late component fixed at 0 dB). Fixed levels are denoted by empty markers for visual clarity. BK = Berlin Konzerthaus; BP = Berlin Philharmonie. See the online article for the color version of this figure.

Figure 7 shows an alternative view of the results with the music excerpts grouped together to facilitate comparison between the different conditions. The results with the early fixed condition are the closest to the original acoustics, showing only relatively minor and statistically insignificant deviations from 0 dB. In contrast, the early component for both positions in BP have been significantly amplified for both the direct fixed and late fixed conditions. Furthermore, the rank order of the median early component levels for the four positions, and the lower and upper limits of their confidence intervals, is the same between the direct fixed and late fixed conditions. On the other hand, the smallest average deviations from 0 dB are seen in the late component levels.

In order to assess the possible existence of preference groups in the results, the data were partitioned with k-means clustering using squared euclidean distances and k-means++ algorithm (Arthur & Vassilvitskii, 2007), for two and three groups, respectively. The analysis showed some spreading of the results that could be interpreted as emerging separation into groups, but the limited amount of data and the complexity of the results allow no further conclusions to be drawn.

Table 2 shows the amount of change seen in the ISO 3382–1 parameters of the four positions based on the results with both music excerpts grouped together (as shown in Figures 6 and 7). The changed parameter values were calculated using the median component gains. The shown values are the three values corresponding to each fixed condition, the average of these, and the original value for reference.

The changes in EDT are predominantly negative, owing to the amplification of early reflections which makes the early part of the Schroeder integration curve steeper. In terms of the 5% just noticeable difference (JND; ISO 3382–1:2009, 2009), the average change in EDT is perceptually relevant for BK P2 and both positions in BP. The average changes in C_{80} are mostly positive and exceed the 1 dB JND for positions P2 in both halls. For J_{LF} (JND 0.05) and G (JND 1 dB), the only perceptually relevant average changes are for BP P2, and they are positive. While the

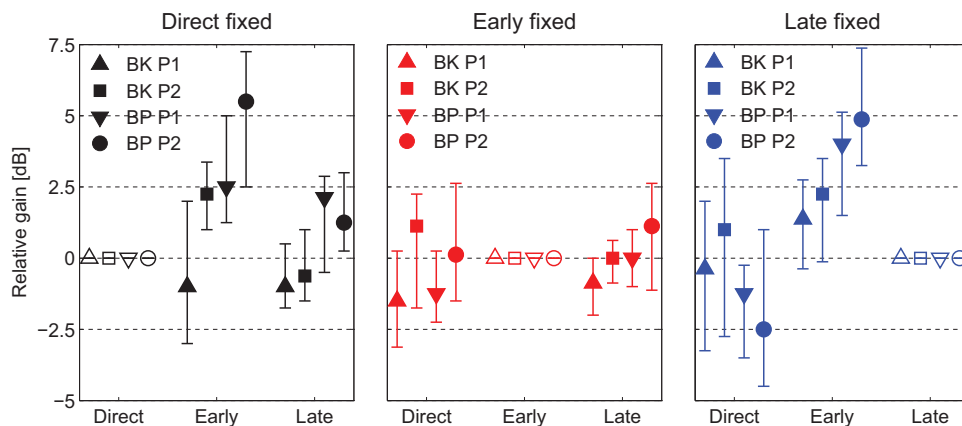*Figure 7.* Median levels and 95% bootstrap confidence intervals for the results with both music excerpts grouped together (alternative view). Different markers denote the median levels for different listening positions and fixed levels are denoted by empty markers for visual clarity. BK = Berlin Konzerthaus; BP = Berlin Philharmonie. See the online article for the color version of this figure.

Table 2

ISO 3382–1 Parameter Value Changes Based on the Median Values Shown in Figures 6 and 7

Position	EDT (s)					C₈₀ (dB)					J_LF					G (dB)					L_J (dB)				
	dir.	ear.	late	*avg.*	orig.	dir.	ear.	late	*avg.*	orig.	dir.	ear.	late	*avg.*	orig.	dir.	ear.	late	*avg.*	orig.	dir.	ear.	late	*avg.*	orig.
BK P1	0	0	−0.1	*0*	(2.1)	+0.3	+0.4	+0.8	*+0.5*	(−1.5)	−0.01	+0.01	+0.02	*+0.01*	(0.27)	−0.8	−0.7	+0.4	*−0.4*	(2.9)	−0.9	0	−1.0	*−0.7*	(−2.1)
BK P2	−0.2	0	−0.2	*−0.1*	(2.1)	+0.2	**+2.6**	**+2.1**	***+1.7***	(−2.3)	0	−0.01	0	*0*	(0.29)	+0.5	+0.1	+0.9	*+0.5*	(1.8)	0	0	−0.6	*−0.2*	(−3.0)
BP P1	0	0	−0.4	*−0.1*	(2.1)	−0.8	−1.1	**+1.2**	*−0.2*	(2.4)	+0.02	+0.01	**+0.05**	*+0.03*	(0.11)	**+1.4**	−0.5	+0.8	*+0.6*	(2.5)	0	0	+2.2	*+0.7*	(−5.3)
BP P2	−0.3	+0.1	−0.4	*−0.2*	(1.9)	−1.0	**+2.0**	**+2.2**	***+1.1***	(0.2)	**+0.06**	0	**+0.08**	***+0.05***	(0.15)	**+2.3**	+0.6	**+1.2**	***+1.4***	(1.2)	+1.1	0	+1.3	*+0.8*	(−5.7)
	dir.	ear.	late	avg.	orig.	dir.	ear.	late	avg.	orig.	dir.	ear.	late	avg.	orig.	dir.	ear.	late	avg.	orig.	dir.	ear.	late	avg.	orig.

Note. The three values for each listening position were calculated based on the median values for fixed direct, early, and late component (marked as dir, ear., and late), respectively. The average of the three calculated values is shown in italic (avg.), and the original parameter value (orig.) in parentheses for reference. Changes that are equal to or greater than the known JND are shown in bold. Note that the JND for late lateral sound level L_J is not known. BK = Berlin Konzerthaus; BP = Berlin Philharmonie; JND = just noticeable difference.

JND of L_J is not known, it can be seen that the average changes are negative for BK and positive for BP.

Discussion

The tendency to amplify the early component is clear for both positions in BP, especially P2. This maybe due to the hall providing insufficient early side wall reflections. The direct component magnitude responses of BP shown in Figure 2 present a further clue: The direct sound lacks low frequencies and is very bright, possibly even masking running reverberation due to the relatively low level of the late component, whereas the early component may be able to provide the missing low frequencies and some softening for the spectrum. The early reflections are originally low in magnitude compared with the direct sound, so a considerable positive gain difference between the early and direct components does not appear unreasonable. The changes in J_{LF} (see Table 2) reflect this: The values for changed BP still fall short of the original values in BK for all conditions.

In other regards, the results show quite a lot of dispersion. However, it is good to keep in mind that the listeners had no point of reference as to what was the true acoustics of the halls within the course of the test. Many of the listeners also remarked afterward that the test was challenging. In this light, it is interesting to see that the median results for BK P1 were within 1.5 dB of the actual acoustics of the hall, and none of the deviations were statistically significant. This indicates that out of the four positions, BK P1 is closest to optimal acoustics, at least in the context of the listening test and its participants. It is also interesting to see that the results were essentially independent of the music excerpt even though they represented contrasting styles and playing dynamics.

The room impulse responses were captured in unoccupied halls, where the reverberation is generally more prominent than in occupied halls, as the audience acts as a further sound-absorbing element. A tendency to set the late component at slightly attenuated levels on average might therefore be expected. However, the late levels had the smallest deviations from 0 dB on average, and were thus most near the original acoustics, and the median results were both above and below the original levels. Thus, it does not seem likely that the unoccupied condition of the halls had any significant effect on the results.

Seeing that the late component levels were the most stable of all, the late fixed condition was taken as an example to look into how the relationships of the component frequency responses change based on the median component gains. In Figure 8, it is seen that for positions BK P2 and both positions in BP, the relative amount of sound energy provided by the early and late components has changed markedly. In the original balance for these positions (see Figure 2), the relative amount of late energy is seen to be clearly above that of the early component for most of the frequency range up to about 2 kHz (and higher for BP P2). In the changed balance (see Figure 8), the early component energy is approximately matched with the late component energy below about 1 or 2 kHz, and is above the late component level at higher frequencies. For both positions in BP the direct component level has also changed so that the amount of energy is below that of the early component for most frequencies,
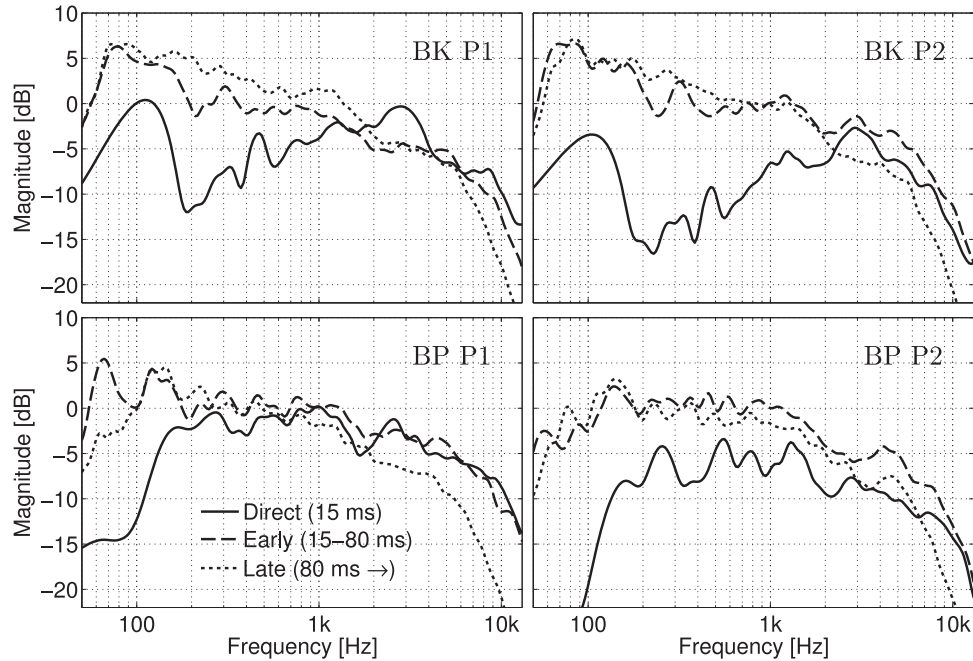
Figure 8. The relative levels of the frequency (magnitude) responses at positions P1 and P2 in Berlin Konzerthaus (BK) and Berlin Philharmonie (BP) for the direct (0–15 ms), early (15–80 ms), and late (80 ms →) components with the median component gains for the fixed late condition applied.

albeit to a lesser degree for P1. Also, for BP P1, the early component supplies some additional low frequency energy under 100 Hz.

When asked about their criteria and procedure for adjusting the component levels, the listeners fell into roughly two groups in terms of their preferred order of adjustment. Most that reported their method of adjustment balanced the levels between the direct and early components first. Others preferred to start by setting the amount of late component (or alternatively the relationship between direct and late). The listeners associated the direct component with localization, sharpness, distance, presence, clarity, and timbre. The early component was associated with width, clarity, punch, balance between voices, timbre, and envelopment. The late component was associated with blending of sound, envelopment, size of space, spaciousness, and warmth.

While the amplification of the early component in BP is statistically significant for both positions and for both direct fixed and late fixed conditions, the difference in the relative levels with the early fixed condition raises questions. Why did the listeners tend not to set the direct and late components at attenuated levels relative to the early component in order to create a similar kind of level balance as with the direct fixed or late fixed condition? One explanation could be the listening level: when the early component is fixed, the other components do not tend to be set at attenuated levels because it would reduce the total loudness of the source and room too much. It could be that a certain amount of direct sound and late reflections are needed to create a plausible listening experience. A further concern is that starting the adjustment from the relatively silent levels instead of fully randomized levels might have influenced the results. However, many of the listeners

reported making their adjustments by bringing the adjustable components initially clearly too high in level in order to hear the contribution of the component properly, and then proceeded to make smaller adjustments. Such a strategy reduces the potential effect of the starting levels. In any case, it is hard to say for certain what caused the difference, and it may well be a further research question in itself. However, the overall similarity of the results between the direct fixed and late fixed conditions may be seen as evidence of significance, at least in the case of Berlin Philharmonie.

A relevant notion for the present analysis has emerged from the work of Gade, who explains that the listener expectation for concert hall acoustics has been molded by the way recordings of classical music are made (Beranek et al., 2011). Microphones are often placed within the reverberation radius of the instruments, which results in a higher direct-to-reverberant ratio and therefore a higher C_{80} than would occur naturally at most of the seats in a hall. Furthermore, artificial reverberation is added in the signal chain which lengthens the reverberation time. The increase in C_{80} is also seen in the current study: for P2 in both halls the average change in C_{80} exceeded the JND (see Table 2). However, the increase was primarily due to an increase in the level of the early reflections (see Figure 6). At BK P2, the direct sound was also slightly amplified on average, albeit not as much as the early reflections, but at BP P2, the direct sound was even reduced on average. The early component was also amplified more for BP P2 compared with BP P1, although J_{LF} was already higher for P2 (0.15) compared with P1 (0.11). Therefore, it is unlikely that the desire for more source width was the sole reason for the increased level of the early component. The original C_{80} values might appear to present a plausible expla-

nation (0.2 dB at P2 compared with 2.4 dB at P1) in the form of listeners wanting more clarity at P2. But then the question is—why did they not amplify the direct component when the early component was fixed? Instead, on average, they amplified the late component, and the change in C_{80} was actually -1 dB.

It seems therefore that there may be other perceptual aspects that determined how the relative level between the direct component and early component was set. Timbre is a potential candidate. While many listeners reported using the direct component for sharpness/localization, several also reported using the early reflections to improve the timbre. Neither of the parameters C_{80} and J_{LF} address timbre. As the current state of concert hall acoustics stands, there is no analysis criteria that directly measures the relationship between the direct sound and the early reflections. The notion of the importance of having such a measure and understanding of its relative importance has been brought up in the context of small listening rooms (Toole, 2006). Considering the results of the present study, such an understanding might also be useful for concert hall acoustics.

Conclusions

The preferred levels for the direct, early, and late components of the sound field were studied with an adjustment procedure using auralized acoustics based on measurements of two listening positions in two existing concert halls of different designs, Berlin Konzerthaus (shoebox) and Berlin Philharmonie (vineyard). ISO 3382–1 standard parameters and the frequency responses and spatial distribution of the energy for the direct, early, and late components were used to aid understanding. The results were found to be essentially independent of the musical stimuli. The largest differences in the results were seen between the halls, and a lesser degree of differences was found between the positions. The results show a significant tendency to emphasize the early reflections in the vineyard hall. Furthermore, the effect was greater for the position that was further from the stage (P2 at 19 m compared with P1 at 11 m). In contrast, the median results for position P1 in Berlin Konzerthaus were within 1.5 dB of the original acoustics, and none of the deviations were statistically significant, suggesting that the position was already close to optimal in terms of the balance between the direct, early, and late components. Development of new measures and understanding about the relationship between the direct sound (including the seat dip effect) and the early reflections could be useful in furthering the understanding of concert hall acoustics.

References

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). New Orleans, LA.

Barron, M. (1988). Subjective study of British symphony concert halls. *Acustica, 66,* 1–14.

Barron, M. (2009). *Auditorium acoustics and architectural design* (2nd ed.). New York, NY: Spon Press, p. 40.

Barron, M., & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration, 77,* 211–232.

Beranek, L. L. (2004). *Concert halls and opera houses: Music, acoustics, and architecture* (2nd ed.). New York, NY: Springer, pp. 494–496.

Beranek, L. L., Gade, A. C., Bassuet, A., Kirkegaard, L., Marshall, H., & Toyota, Y. (2011). Concert hall design — present practices. *Building Acoustics, 18,* 159–180.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization* (revised ed.) Cambridge, MA: MIT Press, pp. 223–224.

Bradley, J. S. (1991). Some further investigations of the seat dip effect. *Journal of the Acoustical Society of America, 90,* 324–333.

Bradley, J. S. (2011). Review of objective room acoustics and future needs. *Applied Acoustics, 72,* 713–720.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7,* 1–26.

Griesinger, D. (1997). The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acta Acustica United with Acustica, 83,* 721–731.

Haapaniemi, A., & Lokki, T. (2014). Identifying concert halls from source presence vs room presence. *Journal of the Acoustical Society of America, 135,* EL311–EL317.

ISO 3382–1:2009. (2009). *Acoustics–measurement of room acoustic parameters–Pt. 1: performance spaces.* Geneva, Switzerland: International Standards Organization.

ITU-R BS. 1116–1. (1997). *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.* Geneva, Switzerland: ITU Radiocommunication Assembly.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review, 55,* 163–172.

Kahle, E. (2013, June 9–11). Room acoustical quality of concert halls: Perceptual factors and acoustic criteria – return from experience. In *International Symposium on Room Acoustics (ISRA 2013).* Toronto, Canada. (Paper No. P074)

Kirkegaard, L., & Gulsrud, T. (2011). In search of a new paradigm: How do our parameters and measurement techniques constrain approaches to concert hall design? *Acoustics Today, 7,* 7–14.

Kuusinen, A., Pätynen, J., Tervo, S., & Lokki, T. (2014). Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls. *Journal of the Acoustical Society of America, 135,* 239–250.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62,* 399–402.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *Journal of the Acoustical Society of America, 106,* 1633–1654.

Lokki, T., Pätynen, J., Tervo, S., Siltanen, S., & Savioja, L. (2011). Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *Journal of the Acoustical Society of America, 129,* EL223–EL228.

Meyer, J. (2009). *Acoustics and the performance of music* (5th ed.). New York, NY: Springer, pp. 203–204.

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Bingley, UK: Emerald, pp. 283–312.

Pätynen, J. (2011). *A virtual symphony orchestra for studies on concert hall acoustics* (PhD thesis). Aalto University School of Science, Espoo, Finland.

Pätynen, J., Pulkki, V., & Lokki, T. (2008). Anechoic recording system for symphony orchestra. *Acta Acustica United with Acustica, 94,* 856–865.

Pätynen, J., Tervo, S., & Lokki, T. (2013). Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *Journal of the Acoustical Society of America, 133,* 842–857.

Pätynen, J., Tervo, S., & Lokki, T. (2014). Amplitude panning decreases spectral brightness with concert hall auralizations. In *Aes 55th international Conference on Spatial Sound.* Helsinki, Finland.

Pätynen, J., Tervo, S., Robinson, P. W., & Lokki, T. (2014). Concert halls with strong lateral reflections enhance musical dynamics. *Proceedings of the National Academy of Sciences USA, 111,* 4409–4414.

Schultz, T. J., & Watters, B. G. (1964). Propagation of sound across audience seating. *Journal of the Acoustical Society of America, 36,* 885–896.

Sessler, G. M., & West, J. E. (1964). Sound transmission over theatre seats. *Journal of the Acoustical Society of America, 36,* 1725–1732.

Soulodre, G. A., Lavoie, M. C., & Norcross, S. G. (2003). Objective measures of listener envelopment in multichannel surround systems. *Journal of the Audio Engineering Society, 51,* 826–840.

Tervo, S., Laukkanen, P., Pätynen, J., & Lokki, T. (2014). Preferences of critical listening environments among sound engineers. *Journal of the Audio Engineering Society, 62,* 300–314.

Tervo, S., Pätynen, J., Kuusinen, A., & Lokki, T. (2013). Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society, 61,* 17–28.

Toole, F. E. (2006). Loudspeakers and rooms for sound reproduction – a scientific review. *Journal of the Audio Engineering Society, 54,* 451–476.