Acta Acustica united with Acustica
Vol. 104 (2018) 1019 – 1029

DOI 10.3813/AAA.919268

Direct-To-Reverberant Ratio Threshold for Localization in Concert Halls

Aki Haapaniemi, Tapio Lokki
Department of Computer Science, Aalto University School of Science, 00076 Aalto, Finland.
aki.haapaniemi@aalto.fi

Summary

Direct-to-reverberant ratio threshold for localization was studied with anechoic orchestra instrument recordings in auralized concert halls. Excerpts of anechoic music were convolved with spatial impulse responses and reproduced with a multichannel loudspeaker system in anechoic chamber. Participants adjusted direct sound level in order to explore the transition range of localizability, with separate tasks for precise and imprecise localization. Halls with contrasting acoustics, and excerpts with pairs of instruments from the main orchestra instrument families, were used to estimate the range of influence of hall and excerpt. Statistical analysis was done using linear mixed models. Estimate for the localization threshold of halls was between −2.7 dB to 1.7 dB direct-to-reverberant ratio in the 700–4000 Hz frequency band. Hall had a significant effect on the threshold (around 3 dB), and excerpt had a significant effect on the imprecise threshold (around 4 dB). The interaction of hall and excerpt was also found significant for the imprecise threshold. Transition between imprecise and precise localization was associated with a 5–8 dB direct sound level difference.

PACS no. 43.55.Hy, 43.55.Gx, 43.66.Qp

1. Introduction

In concert halls, acoustical clarity is generally estimated from the room impulse response by calculating the amount of early energy relative to late energy, using a 80 ms crossover time. In addition to the direct sound, the early part of the impulse response typically includes first-order reflections from the side walls, ceiling, floor, and back wall, as well as some higher order reflections. Combined with the direct sound, the early reflections affect the loudness and tonal character, as well as the spatial extension and localization of the auditory event [1]. The spatial extension is linked with an increase in the vagueness of perceived source location, due to modulating interaural time- and level-differences [2]. In moderation, this effect is considered beneficial for acoustical quality. However, in some situations the early reflections may impede the directional cues of the direct sound, resulting in an inability to perceive the correct source location. Such situations may arise, for instance, near side walls that provide strong reflections. Similarly, the late reflections also tend to increase localization errors, especially when the direct-to-reverberant ratio is low [3].

A definition for acoustical clarity has been given as “the degree to which a listener can distinguish sounds in a musical performance” [4]. It may be divided into separability of simultaneously occurring sounds (vertical clarity), and separability of consecutive sounds (horizontal clarity). However, neither the definition of clarity, nor the related clarity index C_{80}, defined in the ISO3382-1 standard [5], consider localization explicitly. Yet, it is possible in the absence of direct sound to “distinguish sounds in a musical performance” from the early reflections, although localization will be more or less incorrect. Nevertheless, it could be argued that correct localization of sources is a prerequisite for adequate clarity. As localization appears to be based mainly on perception of instrument onsets and the directional cues (ITD, ILD, spectrum) of the direct sound, and reflections degrade both onsets [6] and the directional cues [7], it may be assumed that in reverberant rooms a threshold level could be found for the direct sound below which it is no longer possible to localize clearly/correctly. Recently, a measure called LOC has been proposed for predicting source localizability based on the relative amount of direct and early reflected sound [8]. Furthermore, it has been suggested that a distinct threshold may be found where perception changes from imprecise to precise localization. However, these claims have not yet been supported with perceptual data from formal listening experiments.

The present study explores a direct-to-reverberant ratio (DRR) threshold for localization in concert halls, and seeks to answer the following questions: 1. What is the necessary DRR for the correct localization of instrumental sources in concert halls? 2. Is the threshold dependent on the hall? 3. Is the threshold dependent on the source material (e.g. instrument characteristics, articulation, and com-

Received 6 June 2018,
accepted 31 August 2018.

positional style)? These questions are primarily relevant to the experience of listening to orchestral music in a concert hall. To this end, a listening experiment was conducted with anechoic orchestral instrument parts auralized with the acoustics of three measured halls, which were chosen for their different early/late response characteristics, and reproduced through a multichannel loudspeaker system in an anechoic chamber. A method of adjustment procedure was used where the participants adjusted the direct sound level, while the rest of the sound field was kept at a constant level. Precise and imprecise thresholds for localization (see definitions in Section 4.3) were measured in separate experimental runs, in order to estimate the localization threshold more reliably using the average of these thresholds, and to explore the transition range of localizability. As a pursuit of supplementary interest, a threshold for clear articulation was measured.

2. Background

Localization of instruments in a concert hall is facilitated by the precedence effect [9], a phenomenon whereby auditory localization is dominated by the direction of the first wavefront. The subsequent reflections that arrive within the echo threshold time limit are perceptually fused with the direct sound, and usually have a negligible effect on source localization. The echo threshold depends on the type of stimulus and especially its attack characteristics. For impulsive sounds such as clicks, the echo threshold is 5-10 ms, whereas for speech, it can be as high as 50 ms [9]. For orchestral music the echo threshold appears to be even longer, and has been thought to be around 80–100 ms [1].

Moreover, it has been shown that the precedence effect is most effective with short attacks (transients), and the range of attack times that aid localization in rooms extends to about 100 ms [10]. On the other hand, attack times must be at least 200 ms long to avoid influencing the lateralization of long duration tones [11]. Hence, the shapes of instrument onsets may be expected to affect auditory localization in concert halls. Typical attack time ranges for traditional orchestral instruments' (excluding percussion) are between 14 ms to 85 ms, when taken as averages over various tone frequencies and dynamics [12]. These are within the range of attack times useful for localization [10].

Furthermore, there appears to be an obligatory window of integration for localization cues, which is estimated to be in the order of 100 ms [13, 14, 15, 16]. To the extent that localization is based on perception of instrument onsets, instruments with longer attacks are more susceptible to interfering localization cues. In conditions of low DRR, the localizability of different instruments likely varies depending on attack time and spectral structure. For example, reverberation increases perceived attack times by smoothing instrument envelopes [6], and this influence is dependent on the shape of the amplitude envelope [17]. Thus, it may be expected that reverberation results in different amounts of temporal shift of perceived attack for different instruments. Thus, the hall's response can also be expected to affect localization, not only through degraded directional cues, but also through its effect on onset detection.

In summary, different source signals likely lead to different localization thresholds, particularly depending on their attack characteristics. Sounds with sharp attacks probably have lower thresholds than sounds with more gradual attacks, which may be more susceptible to the interference of early reflections and the envelope smoothing effect of reverberation. Thus, also the degree of liveness in a hall is likely to affect the threshold of localization.

3. Methods

3.1. Overview

The listening experiment was conducted using auralizations of anechoic music featuring two concurrent instrument parts (separate tracks). A lateral separation of 40° (±20° relative to directly ahead) was used for the sources to ensure separate localizability. The choice of two sources was a compromise between one source and a full ensemble. The challenge with a full ensemble is its complexity; a perceivable and unambiguous criterion for a localization threshold is difficult to define, and mutual masking is likely to introduce more confusion. In contrast, using only one source appears contrived and too simplistic, although the criterion will not present a problem. On the other hand, two sources already have a relationship, and the mutual distance of the sources can serve as an additional point of reference in the experiment. Furthermore, the maximum number of simultaneous sound sources that the human auditory system can identify and localize seems to be limited to around three or four, depending on the type of stimulus [18].

3.2. Room impulse responses

The spatial room impulse responses (RIRs) were taken from loudspeaker orchestra [19] measurements of three unoccupied concert halls; Cologne Philharmonie (CP, amphitheater style), Lahti Sibelius Hall (LS, contemporary shoebox), and Wuppertal Stadthalle (WS, classical shoebox). A single RIR from each hall, measured at the same distance in the stalls, was used for the auralizations (a RIR for the other instrument was created by copying the first RIR and changing left-right directions; see Section 3.3). The receiver position was in the stalls, at a 15 m distance from the stage and 2 m left of the hall's midline, and the measurement source was 2 m from the stage edge and 4 m to the right of the stage center (with respect to the receiver). This geometry yields a 20 degree off-centre angle for the sound source in auralization. Figure 1 depicts the measurement arrangement in hall WS.

The RIRs were measured by playing back logarithmic sine sweeps from the LSO loudspeakers on stage, and recording with a 6-channel intensity probe at the listener position. The 6-channel RIRs were subjected to directional analysis using the spatial decomposition method [20], which decomposes the RIR to image sources in short analysis time-windows. The image sources were assigned to the nearest reproduction loudspeakers as a multichannel convolution reverb with sparse impulse responses. Since the measurement and spatial analysis/synthesis methods are not in the focus of this article, and have been described in detail elsewhere [21], further discussion is left out for brevity.

The RIRs were selected from a larger set of measured halls to encompass a range of extremes with regard to the amount of early reflections and late reverberation, and to explore how the localization threshold is affected by these variables. Figure 2 shows objective parameters that depict the amount of early (G_{early}) and late (G_{late}) energy, excluding the direct sound. The octave bands above 500 Hz are expected to be most relevant for localization of instrument sounds. Table I lists also the corresponding ISO3382-1 [5] acoustical parameters for reference.

Each of the selected halls represents an extreme in some acoustical property. CP has a rather dry response, which is evident in low relative levels of both early and late reflected energy (G_{early} and G_{late}). LS and WS are similar with respect to late reverberation in the mid frequency bands (G_{late}), but WS has more low frequencies, while LS has more high frequencies. However, with respect to the amount of early energy (G_{early}), WS is in the middle among the halls for most frequency bands, whereas LS has a prominent early response that is also considerably lateral (G_{early} with figure-of-eight weighting).

3.3. Direct sound replacement and RIR mirroring

The RIRs were modified prior to the experiment. The direct sound portion (0–5 ms) was replaced with a free-field response of the source loudspeaker used in the RIR measurement. The purpose of the replacement was to have a clearly defined direct sound that is exactly the same for all halls, and thus rule out differences in the direct sound as a potential influence in the experiment. The free-field response was measured in anechoic chamber, and air absorption was added based on the measurement distance of the RIRs. The same free-field direct sound was used at equal level in all RIRs, normalized to a nominal level of equal energy with the original direct sound measured in LS. The direct sound replacement removes some of the initial seat-dip effect, but the perceptual impact of this change is bound to be negligible [22].

Since the experiments were done using two concurrent sources, a second RIR was created for each hall by taking the original RIR and changing signs of azimuth locations (left–right flip). The purpose of this was to create two adequately spatially separated (azimuth ±20°) sources that have the same time-structure in their impulse responses, and thus to ensure identical direct-to-reverberant ratios. The 40° spatial separation was chosen since 1) it is approximately the maximum realistic separation for the listening distance when players are at opposite edges of the stage, and 2) sufficiently large to allow easy separation in the listening experiment. On the other hand, it might be argued that the RIR mirroring leads to an unrealistic situation, in that the two instruments play, in principle, in two displaced mirror-image halls. However, the mirroring procedure was deemed necessary, as it eliminates the time-energy differences between individual RIRs as a biasing factor in the results. In preliminary testing this arrangement was not found to present problems with regard to perceptual plausibility.

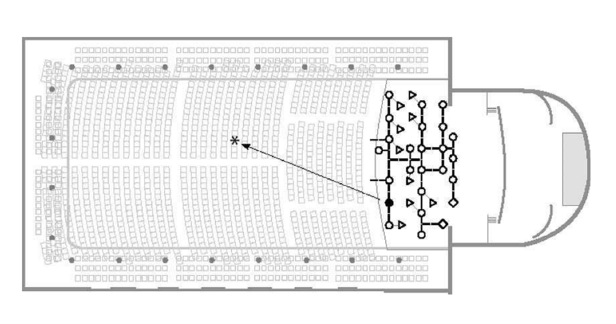

Figure 1. The measurement arrangement with the loudspeaker orchestra in Wuppertal Stadthalle (WS). The presently relevant source loudspeaker is marked with ● and receiver position with ∗.

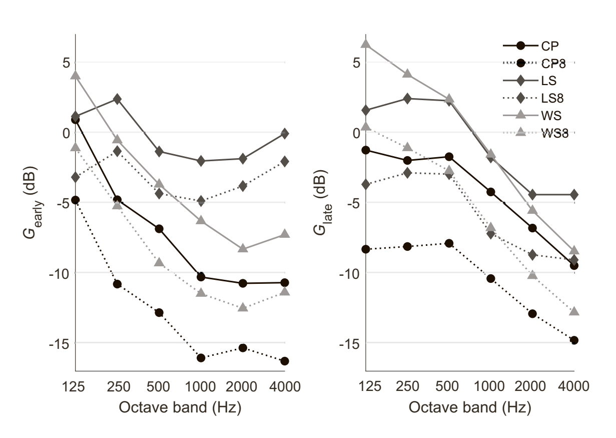

Figure 2. Early (5–80 ms) and late (80 ms – ∞) strength parameters calculated from the RIRs. Suffix '8' denotes energy weighting corresponding to a figure-of-eight directional pattern, with null pointed at the source.

Table I. Acoustical parameters calculated according to ISO3382-1:2009: sound strength (G), early decay time (EDT), reverberation time (T_{30}), clarity index (C_{80}), early lateral energy fraction (J_{LF}), and late lateral sound level (L_J). The parameters are averages of 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.

	G (dB)	EDT/T_{30} (s)	C_{80} (dB)	J_{LF}	L_J (dB)
CP	0.8	1.8/1.8	1.3	0.09	-8.6
LS	3.7	2.0/2.2	0.6	0.28	-3.9
WS	3.0	2.7/2.7	-1.1	0.18	-1.9

3.4. Stimuli

Three musical stimuli were chosen from anechoic multichannel orchestra instrument recordings (monophonic single instruments):

- Strings (violin and cello; Puccini),
- Brass (trumpet and trombone; Bruckner),
- Woodwinds (two clarinets; Beethoven).

The music excerpts were chosen as representatives of different families of traditional orchestra instruments, in a realistic setting with continuous music and running reverberation. They were also selected to cover some range in variables that might affect the localization threshold; playing dynamics, technique, range of attack times, and tendency to fuse. The excerpts were expected to reveal roughly the extent of differences in the localization thresholds due to source type. However, it should be noted that they may not be considered strictly representative of how localization works with a full symphony orchestra, which is a far more complex issue that involves auditory scene analysis and its grouping laws, as well as mutual masking between instruments.

The strings were extracted from the first eight bars (22.6 s) of the beginning of the aria *'O mio babbino caro'* from Giacomo Puccini's opera Gianni Schicchi, multitrack-recorded in quasi-anechoic conditions by D'Orazio et al. [23]. The excerpt features a *legato* melody for the violin, and an accompanying counterpoint for the cello and it is played in *pianissimo* with *andantino* tempo. The strings excerpt represent instruments with slow attacks (average attack times reported as 80 and 85 ms for violin and cello, respectively [12]).

The brass were extracted from a 11.4 s excerpt (bars 53–61) of an anechoic multitrack recording of the second movement of Anton Bruckner's Symphony No. 8 [24]. The excerpt features a repetitive rhythmic pattern with predominantly *staccato* playing, where the parts both overlap and alternate, and it is played in *fortississimo* with *allegro moderato* tempo. The brass excerpt represents instruments with relatively fast attacks (average attack times reported as 30 and 33 ms for trumpet and trombone, respectively [12]).

The woodwinds were extracted from a 7.2 s excerpt (bars 23–24) of an anechoic multitrack recording of the first movement of Ludwig van Beethoven's Symphony No. 7 [24]. The instruments play a harmonic accompaniment for the melody (played by oboe and not in excerpt), both following the same rhythmic pattern with different notes at close intervals. The playing is in *piano*, following a sustained (*poco sostenuto*) style. The excerpt was looped twice before convolution to create a longer (14.3 s) excerpt without gaps, for listening convenience. The woodwinds excerpt also represent faster attacks (average attack time reported as 30 ms for clarinet [12]), but differs from the other two excerpts in that the two similar instruments playing similar parts have a tendency to fuse perceptually.

The listening level for each excerpt was set to a comfortable level with the aim of making detailed listening as easy as possible while avoiding fatigue. Thus, the quieter levels for the strings and woodwinds excerpts, due to the *pianissimo* and *piano* dynamics and differences in instrument spectra, were partially compensated with additional gains to optimize listening levels. Care was taken not to overdo the compensation so as to prioritize ecological validity of the stimuli. In order to monitor the listening levels, L_{Aeq} values were measured with a Sinus Tango (class 1) sound level meter over the duration of the excerpts, with the direct sound set to the natural level. The measured levels were between 54–56 dB for the strings excerpt, 59–61 dB for brass, and 49–51 dB for the woodwinds.

Table II. A list of the reproduction loudspeaker directions; elevation 0° / azimuth 0° is directly in front of the listener and positive azimuth is counterclockwise.

elevation	azimuth
90°	0°
45°	0°, ±90°, 180°
22°	0°, ±30°, ±55°
0°	0°, ±10°, ±20°, ±30°, ±40°, ±60°, ±75°, ±90°, ±105°, ±120°, ±135°, ±150°, 180°
−22°	0°, ±30°
−45°	0°, ±90°, 180°

In the experiment, the direct sound levels of both sources were linked and adjusted with a single dial in the interface. While in theory there could be a difference between the localization thresholds of the individual instruments, the pairs were always from the same instrument family, played in similar style, and had similar attack characteristics. The feasibility of determining a single threshold with this arrangement was confirmed in preliminary testing.

3.5. Reproduction system

The experiment was conducted in an anechoic chamber ($V \approx 200$ m^3) with a multichannel system consisting of 41 loudspeakers (Genelec 8030B) located at a 2.2 m nominal distance from the listening position. The loudspeaker directions are listed in Table II. The channel gains were calibrated to ± 0.5 dB tolerance of target level (L_{AS}), measured at the listening position using static pink noise. An acoustically transparent dark curtain was drawn around the listening position, and the lights were dimmed, in order to exclude the loudspeakers and the room from participants' field of view. The background noise level (L_{AS}) in the anechoic chamber was found to be below the minimum measurement level (22 dB) of the sound level meter.

4. Listening experiment

4.1. Participants

The experiment was done by 14 participants (mean age 30.6 years; one female, 13 males). The participants reported normal hearing, except for one that had a narrow dip at 4 kHz. This participant's results were nevertheless retained in the analysis, since examination of this participant's results and consideration of the nature of the experimental task did not give any reasons to believe performance would have been compromised by the minor impairment.

4.2. Procedure

A method of adjustment procedure was used since it requires less listening than the more conventional forced choice methods, and thus reduces potential of fatigue. Also, by having control over the direct sound level themselves, the participants could explore and re-check the range of perceptual effects at will. To contain the risk of ambiguity in measuring a single threshold, two thresholds were measured: a threshold for precise localization, and a threshold for imprecise localization (see below for definitions). The estimate for the localization threshold was then taken as the average over the estimates of these two thresholds. The combination of the thresholds also yields an estimate for the transition range of localizability.

The listening experiment consisted of seven blocks; one block for each threshold type and excerpt combination, and an additional block for *threshold of clear articulation* (see below for definition), which was measured with one excerpt only as the last block of the experiment. Each block consisted of 12 trials: one musical excerpt auralized in all three halls, with four repetitions each presented in random order. Between each block was a short break. The precise/imprecise blocks for each excerpt were adjacent, and each participant always did the blocks in the same order for all excerpts, that is, with either precise or imprecise first. The order of excerpts and precise/imprecise blocks was balanced between participants, but the articulation block was not considered in the balancing so as not to distract from the main part of the experiment. A training session was completed in the beginning of the experiment, and before introducing each new excerpt in the experiment. Following is an example of one experiment sequence: 1) training/precise/imprecise (woodwinds); 2) training/precise/imprecise (brass); 3) training/precise/imprecise (strings); 4) articulation (brass).

The participants set the level of the direct sound component with an endlessly rotating virtual dial in the graphical user interface (Max/MSP on iPad). The virtual dial had a 0.25 dB granularity, with 10 dB per full revolution, and was made large enough to allow for adjustment precision within the granularity of the dial. The starting level of the direct sound was set with a randomization of ±10 dB around the nominal level, using rounding to 0.5 dB steps. The excerpts were continuously looped while the participants did the adjustments. To avoid creating an inconveniently long gap between each loop repetition, the terminal reverberation tails of the stimuli were truncated after convolution.

4.3. Instructions and training

The participants were given written instructions that specified the context and objective of the experiment and told them what they were adjusting. The following criteria and explanations were included in the written instructions. The last sentences (in bold) were also shown in the listening test graphical interface as a reminder for the participants in each corresponding block of the experiment.

– *Threshold of precise localization*

  This is a point where, when adjusting the direct sound level gradually from a level below the threshold, you become aware of a shift in the localization percept where the localization of the instruments becomes precise.

  Do this adjustment so that both instruments are precisely located.

– *Threshold of imprecise localization*

  This is a point where, when adjusting the direct sound level gradually from a level above the threshold, you become aware of a shift in the localization percept where the localization of the instruments becomes imprecise.

  Do this adjustment so that either or both instruments are imprecisely located.

– *Threshold of clear articulation*

  This is a point where, when adjusting the direct sound level gradually from a level below the threshold, you notice that the articulation of the instruments becomes clear / details of the instrument sounds are perceivable.

  Do this adjustment so that both instruments are clearly articulated.

The criteria *for both sources* for the precise threshold and *for either or both source(s)* for the imprecise threshold were chosen to correspond to 1) the ideal situation in a hall, where the listener can localize the sound sources correctly and perceive their mutual distance, and 2) the compromised situation, where either of the source locations has become vague. It is noted that these threshold definitions are primarily relevant for the context of concert hall acoustics. The terms precise and imprecise correspond to definitions for the scale of localizability given in [25].

Training sessions were completed prior to introducing every new excerpt in the experiment. First, the participants were given a demonstration of the differences between imprecise localization and precise localization. They were instructed to pay attention to the change in the localization percept while the direct sound level was adjusted. The participants were then instructed how to use the interface, and they were encouraged to explore extremes, and practice adjusting the direct sound for as long as they needed, in order to familiarize themselves with the excerpts and the interface. The training was done with the same samples as in the actual experiment. No additional training was done for the articulation block.

5. Results and discussion

The listening experiment yielded a total of 1176 trials, consisting of two thresholds × three halls × three excerpts + articulation threshold × three halls × one excerpt, by 14 participants for four repetitions. The threshold values are in the form of gain values relative to the nominal level of the direct sound.

One participant consistently confused the *imprecise* and *precise* tasks, which resulted in inverted thresholds for all the cases, and therefore the results of this participant were rectified. Three trials were discarded due to a bug in the program/interface that resulted in loss of sync between audio tracks of direct and reverberant sound, and one trial was excluded because a participant pressed the *next* button prematurely. Also, when a trial result reached ±20 dB (22 trials out of 1176; 5 for imprecise, 7 for precise, and 10 for articulation), they were discarded due to overshooting. These extreme results were thought to be the outcome of either a momentary lapse of focus or the participant not following the instructions properly. The following analysis is based on the remaining 1150 trials.

5.1. Overview of results

Figure 3 shows an overview with data aggregated across participants and repetitions. It is clear that there is considerable dispersion. This is partly due to the challenging task, but also due to the participants having mixed views on the thresholds, which is reflected in baseline differences among them (see analysis of random effects in Section 5.2). Nevertheless, the data shows some trends, and a few general observations can already be made based on the *median* threshold values:

1. The difference between the *imprecise* and *precise* thresholds are about 5-8 dB, with a typical value of slightly less than 7 dB, and the *articulation* thresholds are about 2-4 dB higher than the corresponding *precise* thresholds, suggesting that precise localization may not guarantee clear articulation.
2. CP has lowest thresholds for every case, while values for LS and WS are quite similar. This suggests that the hall response has an effect on localization (CP is drier than LS and WS; see Figure 2).
3. Woodwinds have always the highest, and brass the lowest thresholds, except for one case where the strings have just slightly lower median threshold (precise threshold in hall WS). This suggests a role for the excerpt characteristics as well.

5.2. Statistical analysis

The statistical analysis was performed with linear mixed model regression, using the R software version 3.4.2 [26] and the *lme4* package [27]. Advantages of using mixed models for analysis of repeated measures data are described by Quené and van den Bergh [28], and Baayen et al. [29]. Examples of application of mixed model analysis in research related to acoustics may be found in papers by Ferguson and Quené [30], and Kuusinen and Lokki [31]. The analysis was performed using data from 993 out of the total of 1008 trials for the *imprecise* and *precise* thresholds. The articulation threshold results were not considered in the analysis since they were only available for one of the excerpts.

Table III. Fixed effects coefficient estimates and random effects variance components for the models specified for imprecise and precise threshold data. The reference levels for dummy-coding are CP (hall) and brass (excerpt).

fixed effects coefficient	**imprecise thr.** est. (std. error)	**precise thr.** est. (std. error)
(intercept)	-5.97 (1.20)	1.24 (1.16)
hallLS	1.00 (0.69)	1.61 (0.58)
hallWS	2.59 (0.63)	3.32 (0.55)
excerptwoodwinds	4.42 (1.63)	2.25 (1.04)
excerptstrings	0.06 (0.95)	1.10 (1.26)
hallLS:excerptwoodwinds	0.03 (0.88)	0.81 (0.76)
hallWS:excerptwoodwinds	0.07 (0.88)	-0.66 (0.76)
hallLS:excerptstrings	2.61 (0.87)	1.02 (0.75)
hallWS:excerptstrings	1.85 (0.88)	-0.73 (0.75)
random effects (participant) coefficient	var. (std. dev)	var. (std. dev)
(intercept)	17.63 (4.20)	16.75 (4.09)
hallLS	1.31 (1.15)	0.73 (0.85)
hallWS	0.15 (0.39)	0.27 (0.52)
excerptwoodwinds	31.68 (5.63)	11.21 (3.35)
excerptstrings	7.32 (2.71)	18.26 (4.27)
residual	10.54 (3.25)	7.83 (2.80)

Two mixed models were fitted separately for the imprecise and precise threshold data, using maximum likelihood and the model specification

$$\text{thresholdvalue} \sim 1 + \text{hall} + \text{excerpt} + \text{hall} : \text{excerpt} \\ + \big(1 + \text{hall} + \text{excerpt} \,\big|\, \text{participant}\big),$$

which estimates an intercept, fixed effects coefficients for the hall and excerpt factors, as well as their interaction, and variance components for the participant random effect, including a random intercept and random slopes for hall and excerpt. The estimation procedure does not make any prior assumptions about the structure of the variance-covariance matrix, but instead estimates the parameters directly from the data and allows for correlation among the random effects. The basic idea of the model coefficients is that the fixed effect estimates show the differences in the means between cases, while the random effects variance components represent the variability among participants.

Model assumptions were checked for both models. Linearity and homoscedasticity were examined by plotting residuals against the fitted values; the residuals for both models indicated that the assumptions were adequately met. Histograms of the residuals were also found to be approximately normal.

The coefficient estimates are given in Table III. The dummy coding uses CP (hall) and brass (excerpt) as reference levels, so that the fixed effects intercept is the population mean estimate for case CP/brass. Following is an example of how to read the estimates for the other cases: the imprecise threshold estimate for LS with strings is -5.97 (intercept) + 1.00 (hallLS) + 0.06 (excerptstrings) + 2.61 (hallLS:excerptstrings) = -2.3. Calculating in this manner

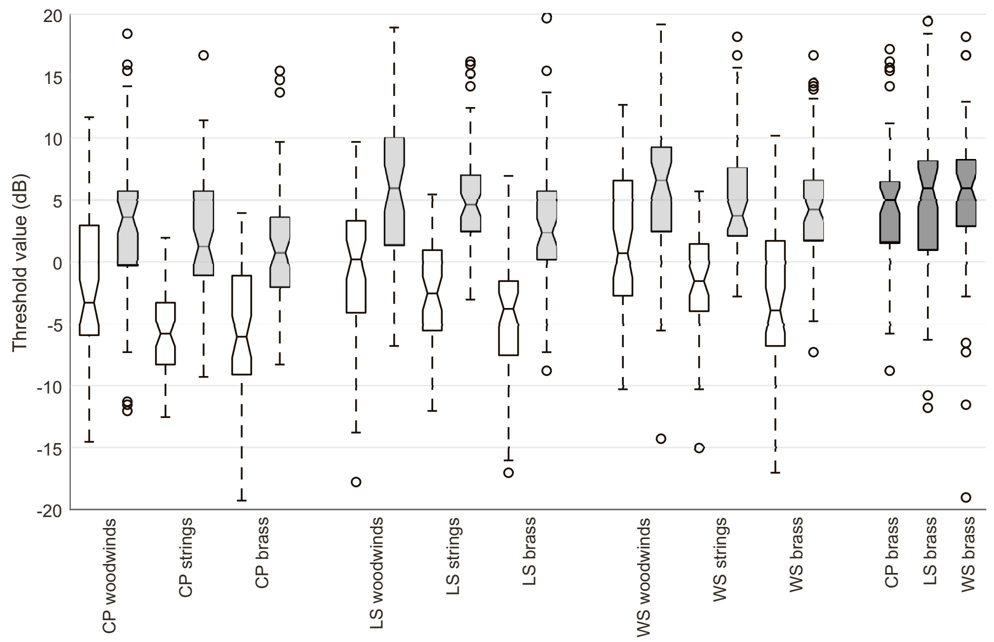

Figure 3. The data aggregated across participants and repetitions (note that the threshold values are in the form of gain values relative to the nominal level of the direct sound); box colors: white (imprecise threshold), light gray (precise threshold), dark gray (articulation threshold). The box outlines the interquartile range (IQR), horizontal line shows the median value, notches show the comparison intervals, and the dashed lines extend to a maximum of 1.5 × IQR below/above the IQR limits. Past the line ends the data points are considered potential outliers (denoted with circles).

for all the cases for both models give differences of about 5–8 dB between the corresponding imprecise and precise thresholds. The estimated range of influence of hall is in the vicinity of 3 dB for both thresholds, with CP giving the lowest, and WS the highest values for both thresholds. The effect of excerpt is found somewhat different for the two thresholds, but there seems to be a particularly clear effect for the imprecise threshold with woodwinds. There also appear to be interaction effects between hall and excerpt that are worth a further look.

Significance (p-values) of fixed effects was established with nested model comparisons. A reduced model was created by removing a fixed effect (and its associated interaction when applicable), and then comparing to the full model using a bootstrapped likelihood ratio test with $n = 1000$ simulations (function *PBmodcomp* from the *pbkrtest* package [32]). Random effects structure was the same between models. Table IV lists the χ^2 statistic and p- values associated with the fixed effects for both models. Hall was found to affect the threshold significantly for both thresholds. The excerpt, as well as an interaction between excerpt and hall, was found significant for the imprecise threshold, but not for the precise threshold.

Table IV. χ^2 statistic and p-values for the fixed effects for both models. For hall and excerpt, the values relate to there being either a main effect or an interaction.

	imprecise thr.	**precise thr.**
fixed effect	χ^2 (df) / p-value	χ^2 (df) / p-value
hall	43.415 (6) / 0.001	35.5500 (6) / 0.001
excerpt	19.114 (6) / 0.004	11.6580 (6) / 0.087
hall:excerpt	12.297 (4) / 0.019	6.4068 (4) / 0.209

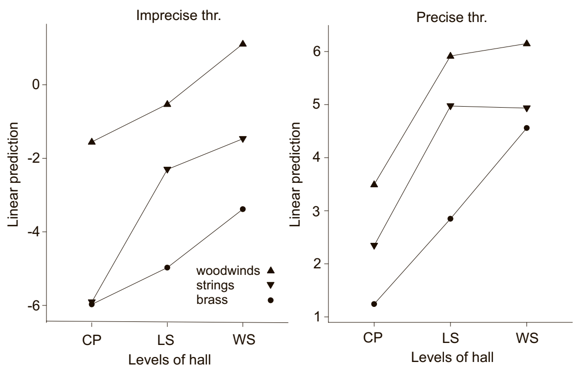

Figure 4. Plot of the *hall:excerpt* interactions for the two models as least-squares means based on the fitted models.

Figure 4 visualizes the interaction between hall and excerpt for both models. Least-squares means were computed based on the fitted models for combinations of levels in the hall and excerpt factors (function *lsmip* in the *lsmeans* package [33]). For the precise threshold, the predictions are elevated with woodwinds and strings in hall LS, while brass is not similarly affected. This might be due to an effect of the strong early reflections in LS; possibly the brass excerpt is not similarly affected due to its prominent short and bright attack transients. Although woodwinds also have a similarly short attack time, they are also

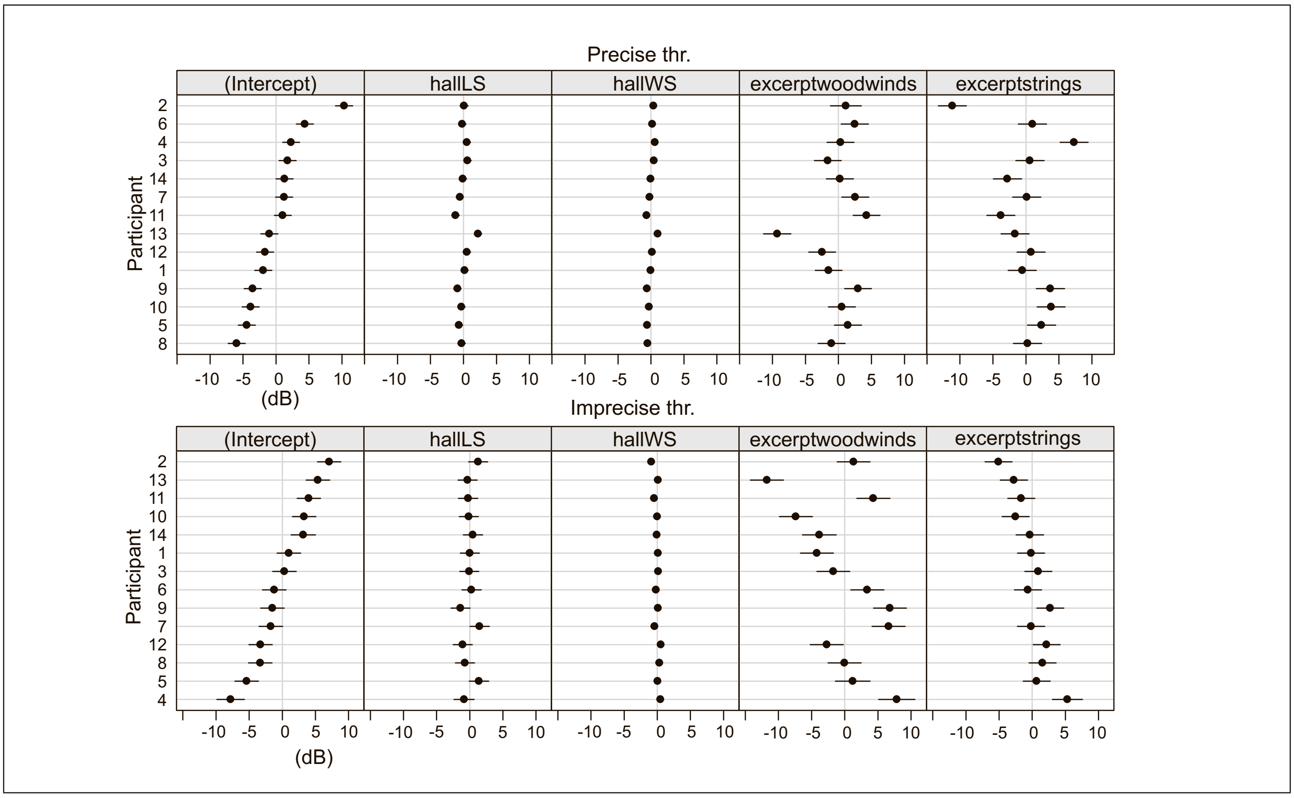

Figure 5. Conditional means and 95% prediction intervals of the random effects for precise threshold (top) and imprecise threshold (bottom) models. Note that the ordering of participants is from lowest to highest intercept, and that the ordering is different for the two models.

less spectrally bright, and in this case there is also a tendency of fusion due to two similar instruments and orchestration, which might be expected to elevate the thresholds. For the imprecise threshold, the values in CP are similar with strings and brass, although for the other halls there is a clear difference between the two excerpts. For this effect there appears to be no simple explanation.

As for the random effects, the variance components (except for hall) have notably larger values than the fixed effects coefficients. It seems likely that the effect of direct sound level on the localization percept may be less salient in the context of real concert halls than previously thought [8], given that the individual differences are so prominent. Especially the excerpt random slope has a large influence on the results, most notably for the imprecise threshold with woodwinds. For further analysis of individual differences, Figure 5 shows plots of the random effect intercepts and slopes per participant for both models (plotted using *ranef* and *dotplot* functions). These are the conditional means for the random effects given the model estimates. It is clear that the random variance associated with different levels of hall is a minor influence for all participants. However, the effect of excerpt is much greater. Furthermore, for the imprecise threshold there is an evident negative correlation between the intercept and both woodwinds and strings excerpts, i.e. there is a tendency for participants with higher intercepts to also have somewhat lower thresholds with these two excerpts. For this there appears to be no clear explanation.

The statistical analysis concludes that the hall affects both the imprecise and precise thresholds significantly, and the excerpt affects the imprecise threshold significantly, but not the precise threshold, although somewhat similar type of effects were observed. There is also an interaction effect between hall and excerpt that is found significant only for the imprecise threshold. The effect of hall is in the vicinity of 3 dB for both models, but may be also somewhat larger because of the interaction with the excerpt. The excerpt has an effect of around 4 dB for the imprecise threshold, and similarly to the hall effect, the assessment of the extent of the effect is made harder by the interaction. Examination of the hall/excerpt interaction for the two models suggests a possible effect of early reflections on less prominent instrument attacks. The difference between the imprecise and precise thresholds were found to be around 5–8 dB for the various hall/excerpt combinations.

5.3. Thresholds as room acoustic parameters

In order to look at the results from the point of view of room acoustic parameters, the thresholds for each hall and excerpt combination were transformed to direct-to-reverberant ratio (DRR) and LOC values, by taking the direct sound at threshold level and substituting it into the impulse response prior to parameter calculation. The thresholds were taken as averages of imprecise/precise thresholds, calculated based on the mixed model coefficient estimates presented in Table III. Since LOC is calculated from a RIR filtered with a 2nd order Butterworth with cutoff frequencies of 700 Hz and 4 kHz, the DRR values were also calculated based on this filtered RIR. Both parameters take the direct sound as the first 5 ms of the RIR. DRR was calculated as

$$\mathrm{DRR} = 10\log_{10}\frac{\int_0^{0.005} p(t)^2\,\mathrm{d}t}{\int_{0.005}^{\infty} p(t)^2\,\mathrm{d}t}, \qquad (1)$$

where $p(t)$ is the filtered impulse response. LOC was calculated according to [34], except that the -1.5 fudge factor was omitted as per later development of the parameter

$$\mathrm{LOC} = S + 10\log_{10}\int_0^{0.005} p(t)^2\,\mathrm{d}t \qquad (2)$$
$$-\frac{1}{D}\int_0^{D-0.005} POS\left(S + 10\log_{10}\int_{0.005}^{\tau} p(t)^2\,\mathrm{d}t\right)\mathrm{d}\tau,$$

where

$$S = 20 - 10\log_{10}\int_{0.005}^{\infty} p(t)^2\,\mathrm{d}t, \qquad (3)$$

D is the 100 ms evaluation window length, and POS denotes taking positive values only. According to Griesinger [34]:

> The measure simply counts the nerve firings that result from the onset of direct sound above 700 Hz in a 100 ms window, and compares that count with the number of nerve firings that arise from the reflections in the same 100 ms window.
>
> ...S is a constant that establishes a sound pressure at which nerve firings cease, assumed to be 20dB below the peak level of the sum of the direct and reverberant energy.
>
> The first integral in LOC is the log of the sum of nerve firings from the direct sound, and second integral is the log of the sum of nerve firings from the reflections.

Figure 6 shows the parameter values (“omni”). All the LOC values are above 0 dB at threshold, but most of them are in the vicinity of 10 dB. These elevated values would appear to be in conflict with Griesinger’s proposed localization thresholds. However, it should be noted that the proposed values are based on assumptions that differ from the present conditions in several ways. Specifically, LOC is designed for localization of speech, and has only been – as far as the present authors are aware – perceptually tested with ideal diffuse reverberation, with no overlapping reverberation from previous sounds, and is furthermore calculated from only one channel of a binaural impulse response. The present situation differs with respect to these assumptions by RIRs that include discrete early reflections, continuous music including overlapping reverberation from previous notes, and LOC values calculated from an omnidirectional impulse response.

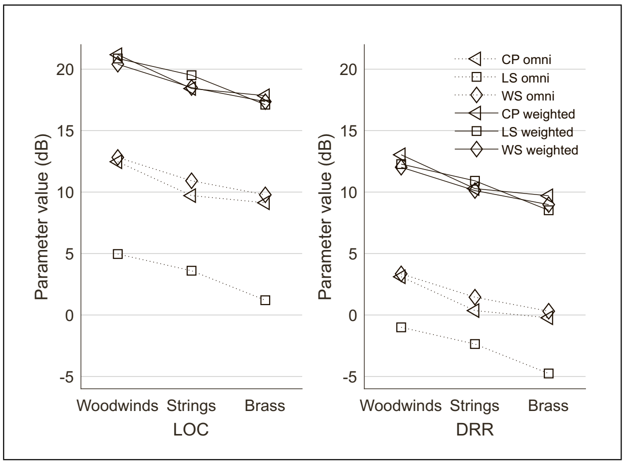

Figure 6. Illustration of the relative fit of omnidirectional and optimally weighted parameter values at threshold between the halls. LOC (left plot) and DRR (right plot) values correspond to the average threshold results for the woodwinds, strings, and brass excerpts. The weightings are *fig8* to the eighth power for LOC, and *fig8H* to the fourth power for DRR (see text for explanation).

With both DRR and LOC (“omni”) the results look overall similar, albeit with a different distance between the highest and lowest (hall LS) values. The notable discrepancy between the threshold results for LS compared to the two other halls might be due to the hall’s prominent early lateral reflections having a lower perceptual weight in localization. It has been noted that reflections from directions close to the source may be especially influential in reducing clarity [35]. It is also possible that this applies to localization. As an informal pursuit of this idea, spatial weighting was applied to the RIRs prior to parameter calculation. The candidate spatial weightings were figure-of-eight (*fig8*), a half figure-of-eight (*fig8H*; zero weight for angles beyond absolute value of 90°), and a modified figure-of-eight pattern that is dependent on the cosine of azimuth only (ignoring the elevation angle), and thus does not marginalize elevation angles. All of the candidate weightings had their maxima pointed to the direction of the direct sound. Additionally, the spatial weightings were raised to powers from 1 to 10, in order to include higher degrees of directional selectivity. The various weightings were pitted against each other by scoring the degree to which the parameter values of the halls clustered together, with a small score indicating a better fit. The score was calculated by taking excerpt-wise squared differences of pairs of parameter values between each of the three halls, and then taking a sum over the differences over all the pairs and excerpts. LOC had the minimal score (3.9) with *fig8* to the eighth power, while DRR was close (4.9) with *fig8H* to the fourth power. Both parameters gave values that cluster the results reasonably well together (see Figure 6 “weighted”). Therefore, it would seem that the results make better sense if it is assumed that the interference of reflections on the localization of sources is dependent on the direction of arrival relative to the source, with localization being less sensitive to lateral reflections. However, this is only an indication and should not be mistaken for direct evidence; the issue may be better approached by future studies.

5.4. General threshold estimates

As a general estimate for the hall-dependent localization threshold, the average of imprecise/precise threshold estimates averaged over the excerpts in the 700-4000 Hz band (full band result in parentheses) yield DRR values of CP: 1.1 (1.6) dB, LS: −2.7 (−2.5) dB, WS: 1.7 (1.7) dB. The differences between the imprecise and precise threshold estimates were about 5–6 dB for CP, 6–8 dB for LS, and 7–8 dB for WS.

It is notable that these estimates are rather high in comparison to typical DRR values found in halls. The corresponding DRR values (700–4000 Hz band) with the nominal direct sound level applied gives values of 2.1 for CP, −3.7 for LS, and −0.3 for WS. Only the value for CP is above the estimated threshold, while the others are 1–2 dB below. This would imply that at 15 m distance from stage in halls LS and WS, the precise auditory localizability of sources is already at a critical limit. It might be expected that as one moves out to further distances, localization is gradually compromised. This might be seen as contrary to the common experience that source localization is more robust in a real concert hall. However, the robustness may be attributed at least partly to the influence of visual cues on auditory localization, the so-called "ventriloquist effect" [36, 37]. With the aid of visual cues, the overall localization is likely satisfactory for most listeners in normal circumstances. Also, even in the absence of a direct sound, the early reflections provide the listener cues about the source arrangement, and often aid in sustaining a frontal localization percept. However, it is possible that precise auditory localizability is limited to the front part in most halls. The question how critical these precisely localized auditory sources are for the enjoyment of music in concert halls is left for future studies, and may well be a matter of subjective preference.

6. Conclusions

The direct-to-reverberant source localization threshold was studied in auralized concert halls, with musical excerpts featuring two spatially separated concurrent orchestra instruments. Thresholds were estimated for precise and imprecise localization using a method of adjustment procedure, where the participants adjusted the direct sound level while the rest of the sound field was kept fixed. Statistical analysis was performed using linear mixed effect models. The analysis concluded that the hall significantly affects the threshold (by about 3 dB). The choice of excerpt was found to affect the imprecise threshold significantly (by about 4 dB), but the effect on the precise threshold was not significant. Furthermore, an interaction of hall and excerpt was found significant for the imprecise threshold, but not for the precise threshold. The difference between precise and imprecise thresholds was found to be about 5–8 dB, depending on the hall and excerpt. This range, and the high variance between individuals evident in the random effects, suggest that the effect of the direct sound level over the localization percept within real concert halls is a more subtle phenomenon than previously suggested [8].

Expressed as a direct-to-reverberant (DRR) ratio in the 700–4000 Hz frequency band, the general localization threshold estimate for the halls was found to be between −2.7 dB to 1.7 dB, averaged over the excerpts. DRR and LOC values corresponding to the average threshold results showed a discrepancy between a hall with strong lateral reflections and the two other halls. Applying a weighting function that marginalized lateral directions of arrival improved the match of the parameters between the halls, suggesting that the ability of reflections to interfere with localization of sources may be dependent on the direction of arrival relative to the source.

Acknowledgments

This research was supported by the Academy of Finland, project no. 296393. Thanks to Jukka Pätynen for helpful comments on the manuscript.

References

[1] M. Barron: The subjective effects of first reflections in concert halls — the need for lateral reflections. J. Sound Vib. **15**, (1971) 475–494.

[2] J. Blauert, W. Lindemann: Auditory spaciousness: Some further psychoacoustic analyses. J. Acoust. Soc. Am. **80** (1986) 533–542.

[3] W. Hartmann: Localization of sound in rooms. J. Acoust. Soc. Am. **74** (1983) 1380–1391.

[4] L. Beranek: Concert halls and opera houses: Music, acoustics, and architecture. Springer, New York, 2004, pp. 24–26.

[5] ISO 3382-1:2009: Acoustics — Measurement of Room Acoustic Parameters. I: Performance Spaces. International Standards Organization, 2009.

[6] G. Naylor: A laboratory study of interactions between reverberation, tempo and musical synchronization. Acust. **75** (1979) 256–267.

[7] B. Shinn-Cunningham, N. Kopco, T. Martin: Localizing nearby sound sources in a classroom: binaural room impulse responses. J. Acoust. Soc. Am. **117** (2005) 3100–3115.

[8] D. Griesinger: The physics of auditory proximity and its effects on intelligibility and recall. 141st Convention of Audio Engineering Society, Los Angeles, 2016, paper no. 9659.

[9] R. Litovsky, H. Colburn, W. Yost, S. Guzman: The precedence effect. J. Acoust. Soc. Am. **106**, 1633–1654 (1999).

[10] B. Rakerd, W. Hartmann: Localization of sound in rooms, III: Onset and duration effects. J. Acoust. Soc. Am. **80** (1986) 1695–1706.

[11] H. Kunov, S. Abel: Effects of rise/decay time on the lateralization of interaurally delayed 1-kHz tones. J. Acoust. Soc. Am. **69** (1981) 769–773.

[12] D. Luce, M. Clark: Durations of attack transients of nonpercussive orchestral instruments. J. Audio Eng. Soc. **13** (1965) 194–199.

[13] D. Grantham, F. Wightman: Detectability of varying interaural temporal differences. J. Acoust. Soc. Am. **63** (1978) 511–523.

[14] M. Akeroyd, A. Summerfield: A binaural analog of gap detection. J. Acoust. Soc. Am. **105** (1999) 2807–2820.

[15] S. Boehnke, S. Hall, T. Marquardt: Detection of static and dynamic changes in interaural correlation. J. Acoust. Soc. Am. **112** (2002) 1617–1626.

[16] A. Kolarik, J. Culling: Measurement of the binaural temporal window using a lateralisation task. Hear. Res. **248** (2009) 60–68.

[17] G. Naylor: Some effects of signal and noise modulation on rhythm detection, and relations to musical performance in rooms. Acust. **73** (1979) 208–214.

[18] X. Zhong, W. Yost: How many images are in an auditory scene? J. Acoust. Soc. Am. **141** (2017) 2882–2892.

[19] J. Pätynen: A virtual symphony orchestra for studies on concert hall acoustics. Ph.D. dissertation, Aalto University School of Science, Espoo, Finland, 2011.

[20] S. Tervo, J. Pätynen, A. Kuusinen, T. Lokki: Spatial decomposition method for room impulse responses. J. Audio Eng. Soc. **61** (2013) 17–28.

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

[22] H. Tahvanainen, A. Haapaniemi, T. Lokki: Perceptual significance of seat-dip effect related direct sound coloration in concert halls. J. Acoust. Soc. Am. **141** (2017) 1560–1570.

[23] D. D'Orazio, S. De Cesaris, M. Garai: Recordings of Italian Opera orchestra and soloists in a silent room. Proc. of 22nd International Congress on Acoustics, Buenos Aires, Argentina, 2016, paper no. 732.

[24] J. Pätynen, V. Pulkki, T. Lokki: Anechoic recording system for symphony orchestra. Acta Acust. united Ac. **94** (2008) 856–865.

[25] ITU-R BS.2399-0: Methods for selecting and describing attributes and terms, in the preparation of subjective tests. International Telecommunications Union, 2017.

[26] R Core Team: R: A language and environment for statistical computing (version 3.4.2) [computer program]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/, 2017.

[27] D. Bates, M. Mächler, B. Bolker, S. Walker: Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67** (2015) 1–48.

[28] H. Quené, H. van den Bergh: On multi-level modeling of data from repeated measures designs: a tutorial. Speech Commun. **43** (2004) 103–121.

[29] R. Baayen, D. Davidson, D. Bates: Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang **59** (2008) 390–412.

[30] S. Ferguson, H. Quené: Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. J. Acoust. Soc. Am. **135** (2014) 3570–3584.

[31] A. Kuusinen, T. Lokki: Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics. J. Acoust. Soc. Am. **138** (2015) 3148–3159.

[32] U. Halekoh, S. Højsgaard: A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - The R package pbkrtest. J. Stat. Softw. **59** (2014) 1–30.

[33] R. Lenth: Least-squares means: The R package lsmeans. J. Stat. Softw. **69** (2016) 1–33.

[34] D. Griesinger: The relationship between audience engagement and the ability to perceive pitch, timbre, azimuth and envelopment of multiple sources. Proc. of the International Symposium on Room Acoustics, Melbourne, Australia, 2010.

[35] E. Kahle, C. Mulder, T. Wulfrank: Perceptual relevance of location of reverberation in a concert Hall. Psychomusicology: Music, Mind & Brain **25** (2015) 326–330.

[36] H. Witkin, S. Wapner, T. Leventhal: Sound localization with conflicting visual and auditory cues. J. Exp. Psychol. **43** (1952) 58–67.

[37] C. Choe, R. Welch, R. Gilford, J. Juola: The "ventriloquist effect": Visual dominance or response bias? Percept. Psychophys. **18** (1975) 55–60.