Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses[a)]

Jukka Pätynen,[b)] Sakari Tervo, and Tapio Lokki
Department of Media Technology, Aalto University School of Science, P.O. Box 15500, Otaniementie 17, FI00076 Aalto, Finland

(Received 12 July 2012; revised 8 November 2012; accepted 26 November 2012)

Acousticians and other practitioners alike often describe acoustic conditions in performance spaces with standard objective parameters. Apart from a few exceptions, the parameters are calculated by integrating the sound energy of the impulse responses over time; this makes them inadequate for researching the acoustics in detail, especially in the early part of the room impulse response. This paper proposes a method based on time-frequency and spatiotemporal presentations to overcome the lack of detail in the standard analysis. In brief, the proposed methods visualize the cumulative development of the sound field as a function of frequency or direction by forward-integrating the energy in the impulse response in short time frames. Analysis on the measurements from six concert halls concentrates particularly on interpreting the results in light of the seat dip effect. Earlier research has concluded that the seat dip effect is reduced by reflection from low overhead surfaces. In contrast, the current results indicate that the seat dip attenuation in the frequency response is corrected the best when the hall provides most lateral reflections. These findings suggest that the proposed analysis is suitable for explaining concert hall acoustics in detail.
 [http://dx.doi.org/10.1121/1.4770260]

PACS number(s): 43.58.Gn, 43.55.Mc, 43.55.Gx [NX] Pages: 842–857

I. INTRODUCTION

Objective room acoustics is an essential part of the evaluation of the acoustics of performance venues. Fundamentally, an objective acoustic parameter reflects the subjective perception of the acoustics of a space in some manner. Although the research on objective room acoustics has continued for about a century and even a measurement standard[1] has been established, in general the objective measures cannot accurately describe the acoustic details due to their inherent simplifications.[2] For example, the parameter C_{80}, the measure for clarity, has identical values for an infinite number of different spatial early reflection patterns.[3] This shows that C_{80} is not sensitive enough for more detailed use.

This paper presents an objective analysis technique for detailed inspection of room acoustics to overcome the drawbacks of the traditional measures. The analysis is based on forward integration of the room impulse response energy with respect to frequency or direction. In that matter, the analysis combines principles applied in room acoustic studies in the past.[4] Due to these similarities, it is possible to draw connections between, for example, the standard parameters and the proposed analysis, as is shown later on in this paper.

The method demonstrates the contribution of the arriving energy to the cumulative frequency and spatial responses in short time steps. Six measured concert halls are studied with the proposed techniques. Instead of numeric values for acoustic parameters, the paper presents the results in a visual form, enabling an efficient comparison between different acoustic conditions, that is, concert halls. As a case study, the results from the analyses are applied together for examining the seat dip effect. These investigations enable us to obtain more detailed insight on the behavior of the seat dip effect than with the standard measurements.

II. RELATED RESEARCH

A. Objective room acoustic parameters

The subjective quality of concert hall acoustics is traditionally formed around various descriptors, such as loudness, reverberance, clarity, definition, and envelopment in the sound field. Well known standardized objective parameters,[1] such as strength (G), reverberation time (T_{30}), reverberance (EDT), clarity (C_{80}), and lateral energy fraction (LEF) have been developed to quantify different subjective aspects in the acoustics. The parameters are mostly based either on estimating the slope of sound decay or calculating amounts or ratios of integrated sound energy in the impulse response.[5,6] Overviews covering a large portion of room acoustic parameters have been summarized by Lacatis *et al.*,[4] Bradley,[5] and Abdou and Guy.[7]

Despite the various approaches to develop objective measures, the use of such parameters contains prominent downsides. First, concert hall acoustics is usually described with around five objective parameters at octave-band resolution, compressing the acoustic features into some thirty values. Data are often reduced further by averaging values over middle frequency bands.[1] Second, some objective parameters contain ambiguity, for instance the linear fit of EDT.[2] Third, many of the common parameters based on energy integration

[a)]Part of this work was presented in Proceedings of IOA Auditorium Acoustics 2011, Dublin, Ireland.
[b)]Author to whom correspondence should be addressed. Electronic mail: Jukka.Patynen@aalto.fi

have high degrees of correlation. This has been presented in several studies, e.g., by Beranek[8] (p. 529), Bradley,[5] Kahle,[9] Hidaka,[10] and Hidaka *et al.*[11]

B. Time-frequency and spatiotemporal analysis

Time-pressure or time-energy diagrams are the basic method for displaying room impulse responses.[12,13] Spectrography is a generic approach to representing the instantaneous frequency content in sound. The spectrum of the reverberation is sometimes investigated using waterfall plots. These approaches give information on the instantaneous amount or spectrum of energy during sound decay, but the contribution to the overall response is obscured. Of the more elaborate methods, Davies and Lam[14] have analyzed the magnitude of the spectral minimum in the impulse response up to 50 ms after the initial direct sound. This approach gives, however, only the single value from the spectrum at a time. Spectra of integrated early energy at octave bands have been used as an analysis tool in a number of publications.[15,16]

Spatiotemporal analysis of room acoustics requires the measurement of a spatial room impulse response. The need for understanding the spatial development of sound in rooms in general has yielded various approaches for studying the spatiotemporal responses.[7,17–20] Yamasaki and Itow[21] presented one of the earliest approaches for studying the directional responses in detail. Similar visualizations of sound intensity vectors have also been employed by Abdou and Guy[7] and, more recently, by Bassuet.[18] Merimaa *et al.*[17] presented a combined spatiotemporal and time-frequency analysis by overlaying the directions of the sound intensity vectors on top of a spectrogram. The intensity vectors within time windows were visualized in planes and in three-dimensional (3-D) projections. Gover *et al.*[19] have utilized a spherical microphone array for obtaining directivity-type visualizations in planes. Yet another example of spatiotemporal analysis is the inverse tracing of acoustic reflections.[20] In this approach, the reflections are first localized and then traced back to the source by applying ray-tracing inversely.

An increasingly popular way to study the room acoustics and sound in general is the acoustic camera.[22] It overlays the sound energy over a time window on top of a photographic image. Several different microphone arrays, such as sound intensity microphones, spherical microphone arrays, and planar microphone arrays, have been applied for the acoustic camera.

Overall, the earlier directional studies have presented analyses of sound fields with instantaneous or integrated time scopes without the aspect of cumulative energy. Most of the spatial visualizations have utilized only one or few source positions analyzed in separated time windows. Moreover, the majority of the earlier research on the development of the frequency response in concert halls has concentrated in the few milliseconds after the arrival of the direct sound.

C. Seat dip effect

The seat dip effect is a phenomenon that affects particularly the direct sound in concert halls.[23,24] It occurs when the sound propagates in near grazing incidence over seat rows. The sound is diffracted from the seat backrests in complex patterns, also including reflections from the floor.[25] Superposition of the direct and delayed sounds causes destructive interference. This results in a substantial attenuation at low frequencies. The affected frequency band is associated with the seat dimensions as the seat height corresponds to a resonator. Typically, the seat dip is effective around 100–300 Hz, but the attenuation can span up to 1000 Hz. With measurements of scale models and concert halls, the increases in the height of the source or receiver have been found to reduce the severity of the attenuation.[14,15]

Different variants of parameter G have been introduced to describe the seat dip effect. The amount of early energy between 0 and 40 ms (G_{40}) at low octave bands was proposed for measuring the severity of the initial seat dip attenuation.[15] Soulodre and Bradley proposed the G_{50} value at low octave bands for measuring the perceived bass strength and the seat dip attenuation.[16]

Earlier studies have proposed various approaches to mitigating the seat dip effect. Bradley presented findings where overhead reflecting panels or ceilings at lower heights reduce the seat dip attenuation.[15] In contrast, Davies *et al.* have discussed the risk of coloration and decreased proportion of lateral energy related to overhead reflections.[26] In this regard, recent listening tests have shown that ceiling reflections provide a lower acoustic quality in comparison to lateral reflections.[3] Barron suggested that the attenuation of the low frequencies due to the seat dip effect should be compensated by increasing the amount of late energy at the affected frequencies.[12] In turn, this approach has been later criticized by Bradley[27] based on the insufficient compensation by the late energy. Further alternatives for reducing the seat dip effect include changing the absorption at the seating area.[14,25]

Generally in room acoustics, the attenuation of low frequencies is not given sufficient attention. Here the seat dip effect is considered an important factor regarding the temporal development of the frequency response in concert halls. Given the nature of the phenomenon, a study on the seat dip effect is an ideal application for the proposed time-frequency and spatiotemporal analyses.

III. CONCERT HALL MEASUREMENTS

To study the time-frequency development and the seat dip effect, acoustic measurements were conducted in six unoccupied Finnish concert halls of various sizes and shapes. The halls are denoted VE (Vanaja hall), SI (Sibelius hall), PR (Promenadi hall), KT (Culture house), FT (Finlandia hall), and MT (Music centre). Their basic measures and properties are given in Table I. Five measurement positions were chosen so that in each hall the corresponding positions were at the same distances (7, 11, 15, 19, and 23 m) from the array of sources.

Figures 1 and 2 show the plans and cross sections of the measured halls with the measurement positions denoted with R1–R5. A characteristic feature in SI is the solid concrete wall surrounding the front stalls. In addition, the stage differs from most typical concert halls by having an absorbing surface behind the stage. In PR, the lower part of the hall has a

TABLE I. Basic measures and properties of the measured concert halls. Reverberation times are averages of 250 Hz to 2 kHz octave bands, averaged over all source and receiver positions. (N, number of seats; A, surface area; V, Volume; T_{30}, reverberation time).

Abbreviation	VE	SI	PR	KT	FT	MT
Hall name	Vanaja hall	Sibelius hall	Promenadi hall	Culture house	Finlandia hall	Music centre
Location	Hämeenlinna	Lahti	Pori	Helsinki	Helsinki	Helsinki
Geometry	Rectangular	Rectangular	Rectangular	Fan	Fan	Vineyard
N	700	1250	860	1390	1700	1700
A (m^2)	3400	4800	3790	3200	4500	10500
V (m^3)	9500	15500	10000	7800	15000	24000
T_{30} (s)	1.7	2.2	2.5	1.6	2.0	2.4
Seat upholstery	Light	Moderate	Moderate	Light	Heavy	Moderate
Seat separation	Moderate	Contiguous	Small	Contiguous	Contiguous	Small

nearly flat floor, and the cross sections of the surrounding walls are shaped to resemble maximum length sequences. In MT, the inclined seating sections cause the furthest receiving position to be at the elevation of 5.5 m above the stage. Front stalls are in between massive side walls that splay outward in a manner of a fan shape. Audience areas at different distances are separated with walls across the stalls. A massive canopy is suspended above the stage in SI and MT. VE and PR have arrays of lightweight overhead reflectors. Apart from MT, these halls, measured near position R2, have been recently included in extensive subjective evaluations.[28]

The measurement source in the current investigations was a loudspeaker orchestra consisting of a total of 34 two-way loudspeakers positioned on the stage.[29] The layout shown in Fig. 1 simulates a traditional symphony orchestra seating order with the first violins on the left and celli on the right. The applied Genelec two-way loudspeaker models are not entirely omnidirectional as imposed in the ISO 3382-1 standard.[1] Measurements with type 1029A have shown that they are within 2 dB from the allowed directivity variation up to 1 kHz. The frequency response of the direct sound is well comparable to common omnidirectional sources with the loudspeakers facing to the audience area. The radiated sound level in other directions is attenuated with respect to the increasing angle from the front direction. As a consequence, reflected sound energy from behind the orchestra is expected to be lower. To improve the high frequency radiation with the channels associated with the string instrument positions, two loudspeakers were connected to a single channel. They are shown with thin lines between two loudspeakers in Fig. 1. The principal loudspeaker at 1.18 m height was directed to the audience, and the auxiliary loudspeaker was lying on the floor facing upward. Such combined loudspeakers represent the directivity of the actual string instruments better than a single loudspeaker.[30] The source configuration also corresponds loosely to the directivity and mean power response exhibited by the symphony orchestra instruments on average.[30]

According to the data measured with directional sound sources in Ref. 8 (pp. 585-613), the measured total sound from the 1 to 4 kHz octave bands energy (G) drops 1.5 dB on average in 25 selected concert halls of various sizes and designs. In comparison, the same difference in the six present halls is 2.0 dB with the current non-standard source setup. On this basis, we can support the assumption that the difference from omnidirectional measurements is tolerable. Furthermore, the present setup has been used in the auralization of concert halls.[28,31]

Swept sinusoid excitation was used as the measurement signal with a sampling frequency of 48 kHz. The response from each loudspeaker channel was measured with a six-channel intensity probe (G.R.A.S. Type 50 VI) at the ear height of average seated listener (1.15 m) above the floor. The open microphone probe has three co-centric pairs of omnidirectional capsules in orthogonal directions.[32] The spacing between capsules is 100 mm. In the monaural analysis, only the signal from the topmost omnidirectional microphone in the 3-D probe is utilized.

IV. METHODS

A. Time-frequency analysis

A majority of the objective parameters concentrate on measuring the decaying sound in enclosures. However, the structure of the early part in the impulse response is regarded crucial for the acoustic impression.[3,33] Therefore the proposed method analyzes the development of the frequency content in the monaural impulse response over time. The resulting forward integration emphasizes the significance of the direct sound and the early reflections.

The absolute frequency (i.e., magnitude) response for a single measurement source $l = 1,\ldots,L$ is defined as

$$H_l(\omega,\tau) = |DFT\{h_l(t)\cdot W(t,t_0^l,\tau)\}|, \tag{1}$$

where $h_l(t)$ denotes the pressure of the monaural impulse response for source l, ω indicates frequency, and $W(t,t_0^l,\tau)$ is a rectangular window function

$$W(t,t_0^l,\tau) = \begin{cases} 1 & t_0^l \le t \le t_0^l+\tau, \quad \tau > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The time t_0^l indicates the instant of arrival of the direct sound for each source l. DFT denotes discrete Fourier transform of fixed length independent of τ.

In the present analysis, the frequency response averaged over all sources is employed. All values are scaled so that the frequency response over the full impulse response corresponds to standard G (strength) values. This is accomplished

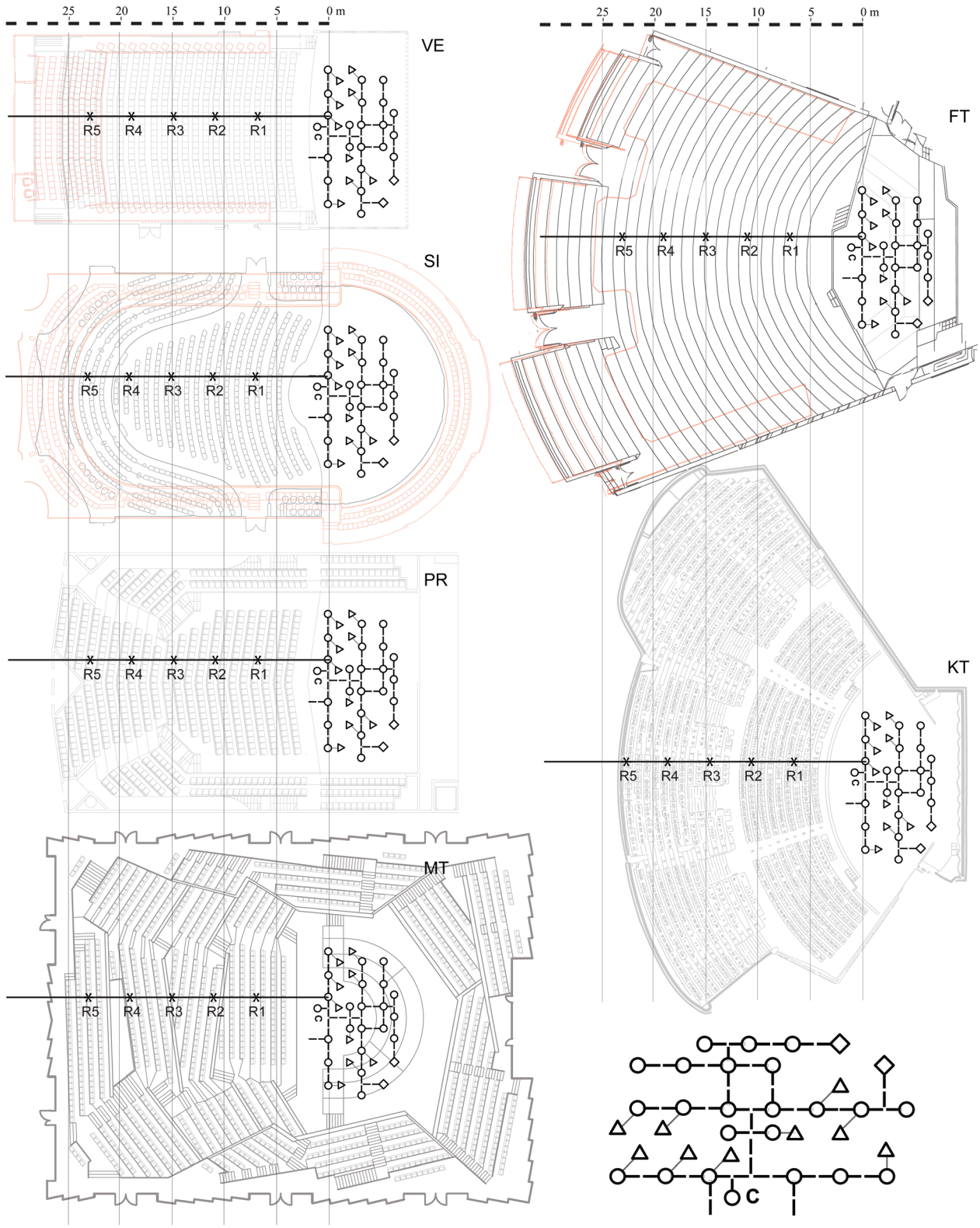

FIG. 1. (Color online) Floor plans of the measured concert halls with receiving positions and source layout in relative scale. Balcony elevations with overhang are overlaid in color. Bottom right: Layout of the source positions in the loudspeaker orchestra. Each thick bar in the diagram denotes a 1 m distance. Symbol "C" indicates the imaginary position of the conductor next to the soloist position. Different symbols denote the models of applied Genelec loudspeaker models 1029 A (◯), 8020 A (△), and 1032 A (◇).

by first calculating the reference frequency response of the direct sound at 10 m distance in free field $H_{\mathrm{ref},l}(\omega)$. The free-field reference varies for different source channels, hence the dependency for source l. The free-field references were measured in the direction of main audience area, regarding the source positioning on stage. Finally, the measured frequency responses are divided by the frequency-averaged reference $|H_{\mathrm{ref},l}(\omega)|_{\omega}$ for frequency-independent scaling:

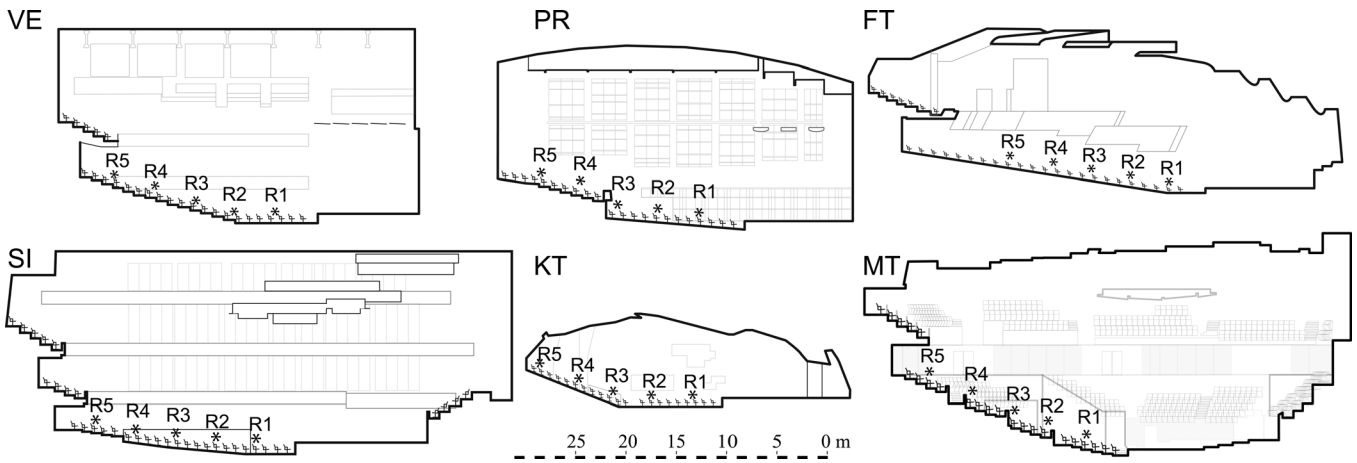

FIG. 2. Cross sections of the measured concert halls with receiving positions in relative scale.

$$G_{0,\tau}(\omega) = 10\log_{10}\left[\frac{1}{L}\sum_{l=1}^{L} H_l(\omega,\tau)^2/|H_{\mathrm{ref},l}(\omega)|_\omega^2\right]. \quad (3)$$

In the following the parameter ω for frequency dependency is omitted for brevity.

The proposed method windows the impulse responses before applying the discrete Fourier transform. DFT operation of a rectangular window function yields more side lobes in the frequency response in comparison to, e.g., Hann or Hamming window functions. However, using window functions with varying fade-in and fade-out times is not feasible. Application of different window functions with constant fade-in and fade-out times was studied. The differences were found to be ± 0.2 dB in the resulting frequency responses, and they were considered negligible. To treat various window lengths equally, rectangular window was chosen for the analysis.

The resolution of the time-frequency analysis depends on the selection of suitable values of τ. There is a natural trade-off between too many τ decreasing the readability and too few τ sacrificing the temporal resolution. Here, suitable values of τ are proposed for

$$\tau = 10n\,[\mathrm{ms}],\ \text{where}\ \ 2 \le n \le 20, \quad (4)$$

that is, from 20 to 200 ms with 10 ms intervals. Selecting 20 ms as the length of the first time window is supported by two factors. First, the seat dip effect is allowed to form at varying measurement distances.[14,15] Second, the precedence effect in human hearing fuses the very early reflections with the direct sound.[34] This effect is cited to span up to 10-30 ms after the direct sound.[3] The upper limit of 200 ms is associated with the upper limit of -5 dB in commonly used linear fit for reverberation time estimates. For ideal exponential decay with a reverberation time of 2.4 s, sound pressure level decreases 5 dB during the first 200 ms.

An example of the proposed analysis is given in Fig. 3 using measurements from MT position R4. On the right hand side, the impulse responses $h_l(t)$ for sources $l = 1,\ldots,25$ are windowed with each of the selected window lengths shown as the thick bars. The impulse responses windowed with each time window are transformed with DFT, and the absolute values are averaged over all sources. The increasing level of the curves demonstrates the contribution of each added time segment. Hence the method visualizes efficiently the development of the frequency response in the receiving position.

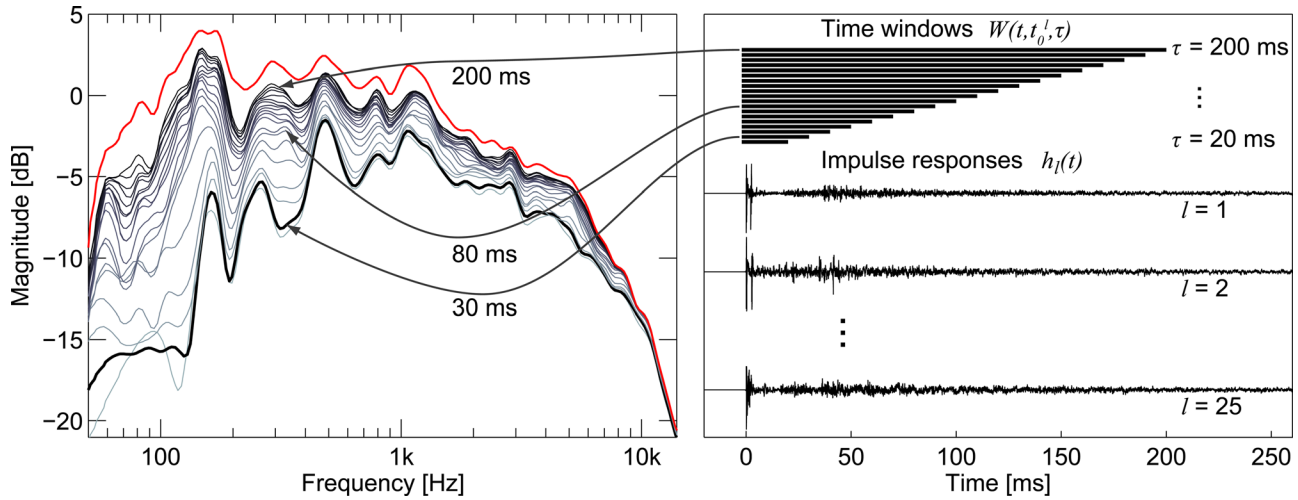

FIG. 3. (Color online) Illustration of the time-frequency representation. Each curve on the left shows the average of the magnitude responses of the impulse responses within the respective time window.

The lowest curves indicate the frequency response soon after the direct sound. In addition, the 30 ms curve is drawn with thick line to improve the readability. At 30 ms, the seat dip effect has stabilized,[14] and in most cases, only few early reflections are included in this window. The second highest curve show the cumulative frequency response after 200 ms after which the arriving sound energy can be assumed to consist of only diffuse late reverberation. The topmost curve shows the overall frequency response at the receiving position. For reference, having two loudspeakers connected in some source channels is found to have no mentionable comb-filter effect on the frequency responses.

B. Spatiotemporal analysis

So far the proposed analysis considers only the monaural frequency response. By employing analysis techniques for spatial audio, the method can be further extended to take into account the direction from which the sound is arriving at each time instant. Spatial impulse responses recorded with the 3-D microphone probe are utilized for this purpose (see Sec. III). Here the analysis and examples are shown only with wide-band impulse responses. However, the method can be implemented also for separate frequency bands.

Several methods exist for directional analysis from spatial impulse responses. In the following, the directional analysis is based on unconstrained least-squares solution of time-difference-of-arrival (TDOA) estimates. Because the actual signal analysis is outside the principal scope of this paper, a brief description of the method is given in the appendix. The employed method estimates the direction of incidence of the sound energy at each discrete time instant in very short time windows. Consequently, this gives a direction estimate for each audio sample in a room impulse response measured in the geometric mean of the 3-D microphone probe. Hence instead of having monaural impulse response $h_l(t)$ as in Eq. (1), the directional analysis provides spatial information for the room impulse response

$$h'_l(t|\hat{\theta}_l(t),\hat{\phi}_l(t)) \triangleq [h_l(t),\hat{\theta}_l(t),\hat{\phi}_l(t)] = \mathrm{DIR}\{\mathbf{h}_l(t)\}, \tag{5}$$

where $\hat{\theta}$ and $\hat{\phi}$ denote the azimuth and elevation angle estimates, respectively. In addition, $\mathbf{h}_l(t)$ is a set of impulse responses, i.e., a spatial impulse response, measured with a microphone array. $\mathrm{DIR}\{\cdot\}$ denotes the directional estimation algorithm, described in the Appendix.

By using directional estimates, energy in the impulse response $h'_l(t)$ can then be categorized with a histogram function over azimuth θ or elevation ϕ angles:

$$G^{DIR}_{0,\tau}(\theta) = 10\log_{10}\frac{1}{L}\sum_{l=1}^{L}\int_{t=t_0}^{\tau+t'_0}[h'_l(t|\hat{\theta}_l(t)=\theta)]^2dt - G_{\mathrm{ref}}, \tag{6}$$

for $\theta \in [-\pi,\pi)$. This equation yields the spatial development of the arriving energy over azimuth angles, marginalizing over elevation angles. G_{ref} is the normalization term for a free-field response at 10 m distance averaged over different source channels. t_0 denotes the time instant of the earliest initial direct sound from all sources. Here one should note that, unlike in the time-frequency analysis, the impulse responses from different sources are not time-aligned with respect to the direct sounds. This is analogous to a situation where the whole orchestra plays a note at exactly the same time. As a directional energy histogram in the time domain, this method does not take into account phase-related phenomena. For example, the energy of the slightly delayed interferences that cause the seat dip effect are not separable in the wide-band spatiotemporal analysis. Instead, the time-frequency analysis is used for that.

For improving the separation between the lateral and median planes, energy in the analyzed impulse responses is weighted with functions that reduce the effect of energy arriving from angles perpendicular to the analysis plane. Hence histogram functions for the lateral plane ($\theta \in [-\pi,\pi)$) and elevation plane ($\phi \in [-\pi/2,\pi/2)$) assume forms

$$G^{\mathrm{DIR}}_{0,\tau}(\theta) = 10\log_{10}\frac{1}{L}\sum_{l=1}^{L}\int_{t=t_0}^{\tau+t'_0}[w_{\mathrm{lat}}(\hat{\phi}_l(t))h'_l(t|\hat{\theta}_l(t) = \theta,\hat{\phi}_l(t))]^2dt - G_{\mathrm{ref}} \tag{7}$$

and

$$G^{\mathrm{DIR}}_{0,\tau}(\phi) = 10\log_{10}\frac{1}{L}\sum_{l=1}^{L}\int_{t=t_0}^{\tau+t'_0}[w_{\mathrm{med}}(\hat{\theta}_l(t),\hat{\phi}_l(t))h'_l(t|\hat{\theta}_l(t),\hat{\phi}_l(t)=\phi)]^2dt - G_{\mathrm{ref}}, \tag{8}$$

respectively. Here, the following toroidal weighting functions are used for the lateral and median planes, respectively:

$$w_{\mathrm{lat}}(\phi) = |\cos(\phi)|, \tag{9}$$

$$w_{\mathrm{med}}(\theta,\phi) = |1-\sin(\phi)|\cdot|\cos(\theta)| + |\sin(\phi)|. \tag{10}$$

Thus in the inspection of the lateral plane, energy from directly above is neglected and vice versa.

Figure 4 presents an example of the spatiotemporal analysis over azimuth [Eq. (7)] and elevation angles [Eq. (8)] with one source in PR. The visualized responses are overlaid with the corresponding floor plan and cross section and centered to the receiver position. The same time windows are used as in Fig. 3. The dashed circle demonstrates the physical 10 m radius from the current receiver position. Simultaneously, this dashed circle indicates also the sound energy at 10 m in the free field in the manner of ISO 3382-1 parameter G. This relates to the normalization of the polar plots with the term G_{ref} in the preceding equations. Dotted circles for ± 6 dB levels are shown for reference only, and they are not bound to physical distance.

V. RELATION BETWEEN TIME-FREQUENCY ANALYSIS AND MONAURAL OBJECTIVE PARAMETERS

The time-frequency analysis method is based on integrating the sound energy within time windows. This is also

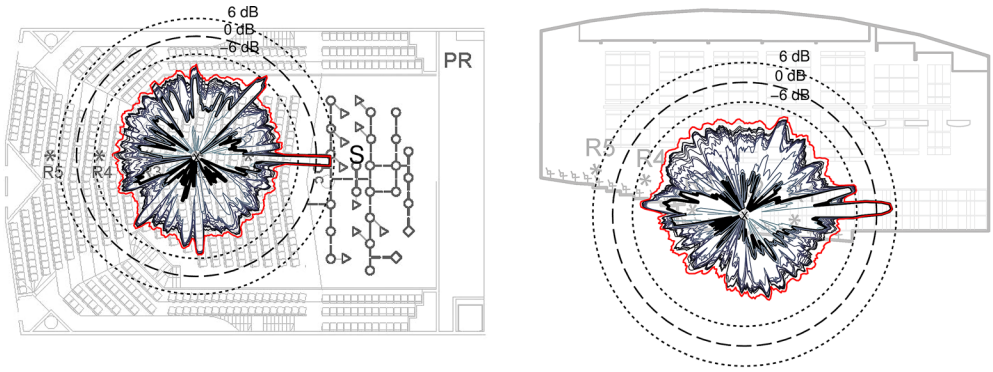

FIG. 4. (Color online) Example visualization of the spatiotemporal response at receiver position R2 in hall PR from a single source denoted with "S." Time increment in window length is 10 ms. The bold line shows the cumulative spatial energy at 30 ms after the direct sound. Dashed circle shows 10 m distance in the hall plan and the sound pressure reference for calibrating G. (a) Lateral plane with elevation weighting [Eq. (7)]. (b) Median plane with azimuth weighting [Eq. (8)]. Histogram curves are smoothed with 5 deg sliding average.

true for the ISO 3382-1 parameters. Therefore it is important to establish connections between the time-frequency representation and the standardized monaural objective measures. The following sections explain these relations for facilitating the interpretation of the results from the proposed analysis.

Strength (G) is a straightforward measure in room acoustics as it indicates the level of sound amplification in frequency bands, i.e., the frequency response. The connection of G to the proposed analysis method is obvious as the current method provides information on the frequency response only in a more detailed manner (see Fig. 3).

A. Reverberation and early decay time

Reverberation time T measures the time in which the level of decaying sound is reduced by 60 dB (Ref. 1) in ideal diffuse conditions. Despite belonging to the earliest acoustic parameters, reverberation time is possibly the most cited parameter when describing the acoustics of concert halls. Usually the decay slope is estimated using backward-integration[35] between -5 and -35 dB, which is denoted as T_{30}. Early decay time (EDT) indicates the rate in which the sound decays from 0 to -10 dB. In practice, the value is obtained by fitting a line to the corresponding region of the backward-integrated sound energy. EDT is considered more important than T_{30} for reflecting the subjective impression of reverberance during music.[1]

The following derivations study a hypothesis according to which the time-frequency visualizations can be used to estimate reverberation time and EDT values.

Given an ideal, exponentially decaying impulse response, the instantaneous sound energy at time instant t is[6] (p. 308)

$$h^2(t) = g\exp\left(\frac{-a}{T}t\right), \tag{11}$$

where g indicates the initial level of the decay at $t = 0$, i.e., the level of the direct sound. Hence, assuming unity gain $g = 1$, the total sound energy during ideal exponential reverberation depends only on reverberation time T. Constant a is defined $a = \log(10^6) \approx 13.8$ so that the exponential decay satisfies

$$10\log_{10}\left(\exp\left(\frac{-at}{T}\right)\right) = -60[\text{dB}]\,,\ t = T. \tag{12}$$

The energy of the impulse response within time instants t_1 and t_2 is defined as E_{t_1,t_2}. Given an ideal exponential decay, theoretical derivation for this energy is

$$E_{t_1,t_2} = \int_{t_1}^{t_2}\exp\left(\frac{-a}{T}t\right)\mathrm{d}t \tag{13}$$

$$= \frac{-T}{a}\left[\exp\left(\frac{-a}{T}t_2\right) - \exp\left(\frac{-a}{T}t_1\right)\right]. \tag{14}$$

Visualization in Fig. 3 shows the cumulative increase in the sound pressure level. The increase between two time-frequency curves at time instants t_1 and t_2 is denoted G_{t_1,t_2} to avoid confusion with the ISO 3382-1 parameter G,

$$G_{t_1,t_2} = 10\log_{10}[E_{0,t_2}/E_{0,t_1}][\text{dB}]. \tag{15}$$

An intermediate time instant t_{int} is introduced for dividing the energy of the impulse response into the early ($E_{0,t_{int}}$) and late ($E_{t_{int},\infty}$) part. The energy increase during the late part $G_{t_{int},\infty}$ is used to estimate the reverberation time T,

$$\frac{E_{t_1,t_2}}{E_{t_1,t_{int}}} = \frac{-\frac{gT}{a}\left(\exp\left(-\frac{at_2}{T}\right) - \exp\left(-\frac{at_1}{T}\right)\right)}{-\frac{gT}{a}\left(\exp\left(-\frac{at_{int}}{T}\right) - \exp\left(-\frac{at_1}{T}\right)\right)} \tag{16}$$

$$\lim_{\substack{t_1=0\\ t_2\to\infty}}\frac{E_{t_1,t_2}}{E_{t_1,t_{int}}} = -\frac{1}{\exp\left(\frac{-at_{int}}{T}\right) - 1}. \tag{17}$$

Solving this for T as a function of $G_{t_{int},\infty}$ leads to

$$T = \frac{-at_{int}}{\log\left(1 - \frac{E_{0,t_{int}}}{E_{0,\infty}}\right)} = \frac{-at_{int}}{\log\left(1 - \exp\left(-\frac{G_{t_{int},\infty}\log 10}{10}\right)\right)}. \tag{18}$$

The preceding equality appears non-trivial. However, with typical reverberation times of 1–3 s, the given relation between $G_{t_{int},\infty}$ and T is close to linear. Therefore the larger increase in the late cumulative sound energy indicates a longer reverberation time (see Fig. 3). Because T and EDT values are equal with an ideal exponential decay, the preceding formulation is applicable in theory to the estimation of EDT as well. These relations are investigated next with the values for T and EDT measured from the six halls described in Sec. III.

Reverberation times (T_{30}) are calculated from each source channel using the backward integration method[35] for octave-band filtered monaural impulse responses. Values for each receiving position are averaged over the reverberation times from all source channels. $G_{200,\infty}$ values are obtained from the time-frequency representation (see Fig. 3) by averaging the smoothed curve segments that correspond to the frequencies within the standard octave bands. Averaging from non-smoothed data could provide more accurate values, but the motivation here is to demonstrate the relation between the easily readable one-third octave-smoothed visualization and the standardized objective parameters.

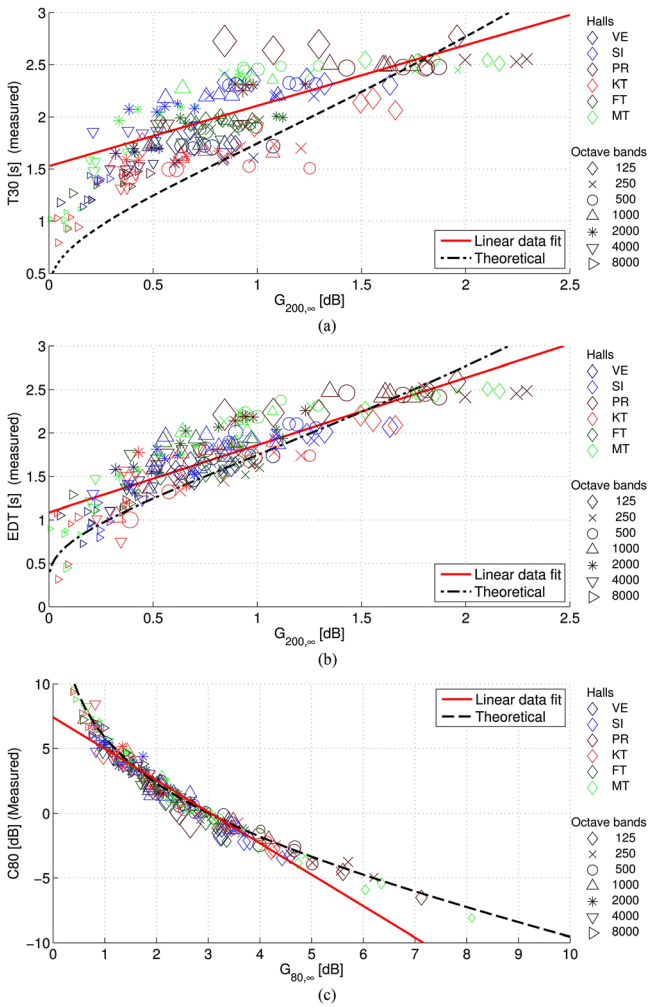

FIG. 5. (Color online) Comparison of measured objective parameters and values from the relation of respective curves. Five receiver positions are shown for each hall. Octave bands are plotted with different symbols. The size of the symbol depends on $G_{0,200}$ values that are calculated similarly to $G_{200,\infty}$. (a) Reverberation time (T_{30}). (b) Early decay time (EDT). (c) Clarity (C_{80}).

TABLE II. Correlation coefficients between values obtained from the time-frequency visualizations and measured objective parameters. Values outside and inside parentheses are for octave bands 125 Hz to 8 kHz and 500–2000 Hz, respectively. Relevant and particularly strong correlations are shown in bold typeface. For all given correlation coefficients, significance level $P \ll 0.001$.

	ISO 3382-1 parameter			
	G	T_{30}	EDT	C_{80}
$G_{0,\infty}$	**0.93 (0.91)**	0.69 (0.19)	0.7 (0.36)	−0.67 (−0.38)
$G_{200,\infty}$	0.62 (0.46)	**0.81 (0.65)**	**0.93 (0.92)**	**−0.88 (−0.85)**
$G_{80,\infty}$	0.58 (0.47)	0.73 (0.53)	0.89 (0.87)	**−0.95 (−0.95)**
$G_{30,\infty}$	0.54 (0.48)	0.62 (0.39)	0.76 (0.71)	**−0.86 (−0.81)**

Results in octave bands are plotted in Fig. 5(a). The horizontal axis indicates the vertical distance between the topmost time-frequency curve (i.e., the final frequency response at the receiving position) and the curve below that, thus the increase after 200 ms. Most data points are on the left side of the theoretical curve of Eq. (18). This indicates that in reality the level of reverberation is below the ideal exponential decay. The overall trend, however, is clearly visible. The correlation coefficient between $G_{200,\infty}$ and T_{30} is $r = 0.81$. Other relevant correlation coefficients are given in Table II. In addition, the best linear fit to the data, calculated with the MATLAB function robustfit, is plotted with a solid line in Fig. 5(a) for visualizing the degree of correlation.

Figure 5(b) shows the corresponding relation between $G_{200,\infty}$ and measured EDT values. The correlation coefficient between $G_{200,\infty}$ and EDT is $r = 0.93$. Again, best linear fit is visualized for assessing the correlation.

B. Clarity measures

Clarity of sound in rooms is measured with parameter C_{80}, which gives the relation of the sound energy before and after 80 ms.[1] Given the previous ideal exponential decay, C_{80} can also be expressed as a function of the reverberation time T:

$$C_{80} = 10\log_{10}\left(1 - \frac{1}{E_{80,\infty}}\right) \tag{19}$$

$$= 10\log_{10}\left(\exp\left(1 - \frac{a}{T}t_{\text{int}}\right)\right), \tag{20}$$

where $t_{\text{int}} = 0.08$ [s]. However, expressed in terms of G_{t_1,t_2} notation,

$$C_{80} = -10\log_{10}(10^{G_{80,\infty}/G_{0,80}} + 1). \tag{21}$$

As C_{80} depends on the sound energy ratios, it cannot be expressed solely based on energy *increase*. Hence C_{80} cannot be directly evaluated with the visualization. Figure 5(c) shows the relation between measured values of $G_{80,\infty}$ and C_{80}. The dashed line displays the same theoretical relation in terms of reverberation time using Eqs. (14) and (20). Measured values follow closely the theoretical relation. Analysis shows that $G_{80,\infty}$ alone and C_{80} have a high degree of inverse

correlation ($r = -0.95$). Correlation coefficients were calculated for other time boundaries as well (see Table II).

The definition parameter (D_{50}) is more associated to clarity of speech. It is defined as the ratio between early and total energy of the impulse response. Therefore it is inversely proportional to $G_{50,\infty}$, which is directly readable in the visualization as the area between 50 ms and final response curves,

$$D_{50} = \left[\int_0^{50\,\mathrm{ms}} h^2(t)\mathrm{d}t \Big/ \int_{0\,\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t\right] \tag{22}$$

$$= \frac{1}{10^{(G_{0,\infty}-G_{0,50})/10}} \tag{23}$$

$$= \frac{1}{10^{(G_{50,\infty})/10}} \quad . \tag{24}$$

Center time (t_s) is used as a measure of balance between early and late energy. It is defined as the center of gravity of the squared impulse response:

$$t_s = \int_0^{\infty} t h^2(t)\mathrm{d}t \Big/ \int_0^{\infty} h^2(t)\mathrm{d}t. \tag{25}$$

In decibel scale, this is equivalent to finding the curve nearest to 3 dB below the final level:

$$10\log_{10}\left[\int_0^{t_s} h^2(t)\mathrm{d}t\right] = 10\log_{10}\left[\frac{1}{2}\int_0^{\infty} h^2(t)\mathrm{d}t\right]. \tag{26}$$

This can be expressed using the earlier definition in Eq. (15) as

$$G_{0,t_s} = G_{0,\infty} - 3\,[\mathrm{dB}]. \tag{27}$$

The accuracy of such an estimate depends on the selected interval of τ in the frequency response plots. With a 10 ms interval, as used in the examples, the accuracy is comparable to the JND for t_s, which has been found out to be approximately 8.6 ms.[36]

C. Discussion

Correlation coefficients for measured omnidirectional ISO 3382-1 parameters and visualized energy increases in the time-frequency analysis are shown in Table II. Notably the correlation coefficient for parameter G and $G_{0,\infty}$ is not unity. The reason for this is the above-mentioned use of one-third octave-smoothed curves for calculating $G_{0,\infty}$. For correlation coefficients on individual octave bands $r > 0.99$.

The strong correlation $r = 0.93$ between EDT and $G_{200,\infty}$ is particularly interesting, as the actual EDT is calculated from the initial sound decay between 0 and approximately 300 ms instead of the late reverberant energy. While this could result from the complementarity of the energy build-up and decay,[37] the authors do not have direct explanation for this phenomenon.

In addition to $G_{80,\infty}$, correlations between measured C_{80} values and other energy increases were calculated. Results with 30, 80, and 200 ms calculations are given in Table II. All these late energy increases exhibit strong negative correlations with measured C_{80} values. This suggests that all energy increase after the direct sound has an adverse effect on the objective clarity in the measured halls.

In general, the high correlation coefficients in Table II concur with earlier research[8] (p. 529) showing that several of the ISO 3382-1 parameter measure the same attributes in the acoustic response. Results from the analytical derivations and correlation analysis confirm the hypothesis that the visualized time-frequency responses communicate the standardized parameters.

VI. ANALYSIS OF CONCERT HALL MEASUREMENTS

A. Time-frequency analysis

In the previous sections, the connection between the visualizations and the common objective parameters was established. Next, a study comparing six measured concert halls in five respective receiving positions is presented by utilizing the time-frequency analysis. The phenomena related to the seat dip effect are discussed in particular. Figure 6 shows the time-frequency responses in the six measured halls in receiving positions R1–R5 (see Figs. 1 and 2).

In the closest position R1, the direct sound dominates the sound field in all halls. Thus the curves are not spread much. However, the seat dip clearly affects the direct sound in most halls. The attenuation center frequency is the highest in VE, as the removable chairs with ample spacing below correspond to a half-wave resonator. They yield a seat dip center frequency one octave higher in comparison to similarly sized seats that are closed below,[24] such as in SI. There the initial frequency response suffers heavily from the seat dip. However, the attenuation is well recovered within the first 70 ms. In contrast, the steep attenuation in MT is not remedied nearly as much, and the seat dip has its signature still in the overall frequency response. The situation is similar also in VE and FT. With regard to the late reverberation, overall strength after 200 ms increases the most in PR, indicating the most reverberant acoustics.

In the position R2 and further, the significance of the reflected energy increases as the level of the direct sound is lower.[38] Thus hall geometries and surface properties begin to have more influence on the accumulating sound field. SI is very different from the other halls, as the flatness of the frequency response improves prominently during 30–60 ms. Again, this suggests intense early reflections. In PR, the prominent development over time affects gradually the whole frequency band below 1000 Hz. The relatively slow increase in the cumulative energy also refers to low values of C_{80}. In MT, the reverberance is emphasized at the low frequencies, while VE, KT, and FT are comparably dry.

SI is particularly interesting in the position R3, where the frequency response changes substantially between 30 and 50 ms. At 50 ms, the seat dip attenuation becomes nearly corrected, and the overall frequency response is flat in comparison to other halls. Energy at the seat dip frequencies increases far more rapidly than at other frequencies. Traces

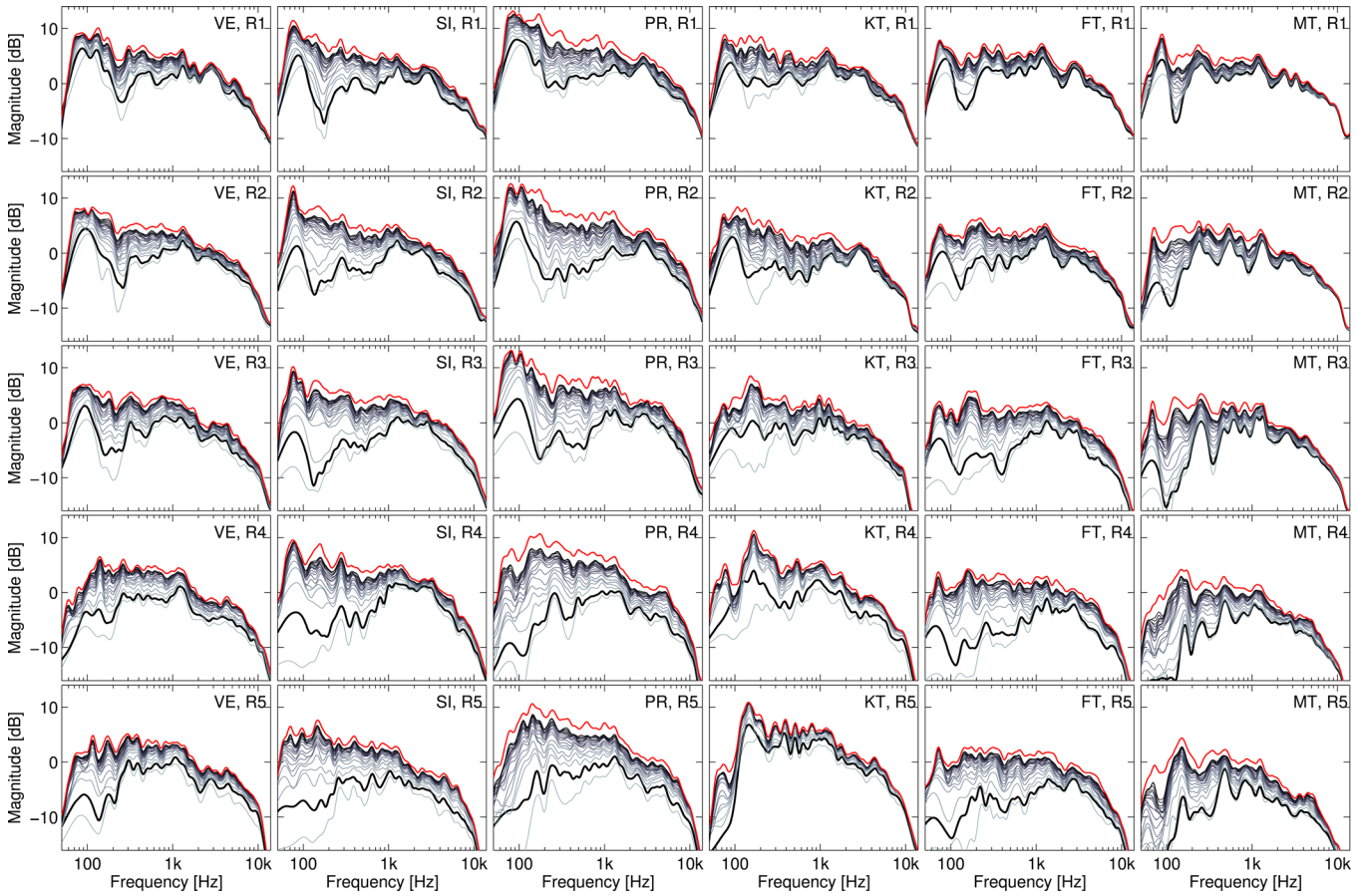

FIG. 6. (Color online) Visualizations of the frequency response development over time in six concert halls measured in five positions.

of similar effect are present in VE. This does not occur to the same extent in other halls. Notably, the pronounced attenuation in MT is not recovered. This phenomenon is discussed further in the next section.

In the furthest positions R4 and R5, the time-frequency curves are more dispersed, suggesting a high contribution by reflected energy. The time-frequency development in KT shows that very little reflections or reverberant energy arrive after 40 ms. Position R5 in KT is unusual, as the ceiling and rear wall create a focus-like geometry, as shown in Fig. 2. Almost all energy is received within the first 20 ms.

In general, the measured halls can be characterized as follows. In VE, KT, FT, and MT, the direct sound dominates the sound field at most receiving distances. The initial low frequency response in MT is the most impaired. In SI and PR, the contribution of the reflected energy is more prominent. The development of the response in PR is more gradual in comparison to SI, where the early reflected energy accumulates within a short time window. PR is also the most reverberant at a wide frequency band.

1. Discussion

The correction of the seat dip is the most substantial in hall SI. Earlier research suggests that the ceiling reflections are beneficial for such an effect.[15] Davies and Lam[14] have demonstrated that the frequency of the seat dip attenuation changes as a function of the horizontal and vertical angles of incidence. According to Schultz and Watters,[23] with large angles of incidence, e.g., early reflections from the ceiling, the attenuated frequency band is narrow in comparison to direct sound or lateral early reflections. Based on these findings, energy reflected from the ceiling or a canopy or from the sides has a different contribution to the development of the frequency response. A reflection from the median plane, unaffected by the seat dip, increases the cumulative energy at all frequencies. In contrast, lateral reflections have attenuations at other frequencies around the seat dip of the direct sound.[14] Hence, the cumulative frequency response becomes increasingly more flat with lateral reflections, and eventually the initial seat dip attenuation is diminished. Next, this hypothesis is reasoned in light of the hall measurements.

In position R1, a distinct seat dip occurs in VE, SI, FT, and MT. In all these halls, the distance to the ceiling or canopy is expected to yield a wide-band reflection at approximately 40–60 ms after the direct sound. Based on the time-frequency visualization, the best correction of the frequency response takes place in SI where the geometry is also the most favorable for lateral reflections: A shoebox shape, prominent reflective surfaces, and nearly flat stalls. However, FT and MT do not have these properties. In FT and MT, the opening fan-shape does not provide as much early lateral reflections. Instead ceiling structures still provide strong, wide-band reflections from above. In these halls, the signature from the seat dip remains visible in the overall frequency response. Hence it is proposed that the ceiling

reflection alone is not sufficient to counteract the initial seat dip effect, but lateral reflections are also required.

Again in receiving positions R2 and R3, ceiling or canopy reflections is estimated to arrive within 60–70 ms after the direct sound. In these positions, the shorter distance to the curved wall in SI allows lateral reflections to arrive within 50 ms after the direct sound. Despite the ceiling reflection in the compared conditions, hall SI is the only one where the frequency response is largely corrected. Furthermore, in hall SI the seat dip attenuation is initially very prominent; this emphasizes the significance of the correcting effect.

The strength at low frequencies after 50 ms has been found to correlate with the perceived amount of bass by Soulodre and Bradley.[16] Inversely, an increase of the low frequency level after 50 ms does not contribute to the perceived bass. While the study does not state the effect of the later increase in low-frequency level, the most plausible consequence is the decrease in clarity. Bradley[27] has proposed against correcting the overall frequency response by increasing the late energy at frequency bands originally suffering from the seat dip. In this light, the visualized time-frequency responses from SI and PR demonstrate a major difference in the bass clarity.

The presented visual analysis suggests that in many of the studied cases, using G_{40} at the low octave bands,[15] or calculating the narrowband minimum of the magnitude spectrum,[14] does not provide enough information on the seat dip effect for two reasons. First, the frequency bands affected by the seat dip vary strongly, and the octave-band resolution does not suffice to describe narrow-band attenuations. Second, as seen in Fig. 6, the minimum in the frequency spectrum can shift significantly as the frequency response develops over time. The time-frequency analysis benefits from averaging over a large number of sources. With only few sources, the extending window length can include reflections in the opposite phase that cause more fluctuation in the visualized curves over time. Averaging over more sources reduces such a possibility, improving readability of the figures.

Quantifying the recovery from the seat-dip attenuation with numeric values is possible. However, the varying attenuation bandwidth and magnitude combined with a nonlinear recovery rate present a problem with at least three dimensions. Hence condensing the phenomenon into a single parameter is a challenge. As an example, different approaches were devised for parameterizing the development of the seat-dip effect in the current data. First, calculating the center time (t_s) (Ref. 1) for the frequency bin of the maximum initial attenuation reveals the overall rate of energy increase. Second, it is possible to estimate the time in which the magnitude of the most attenuated frequency bin increases a certain amount. Third, a more complex alternative is to compare the accumulating energies at two frequencies—at the seat-dip attenuation and at 1 kHz, for instance. This provides information about the seat-dip recovery time, and also about the flatness of the frequency response. For reference, the time delay it takes for the accumulating energy at the seat-dip frequency to reach a level 3 dB below the accumulating energy at 1 kHz frequency varies among 27 ms (FT), 49 ms (SI), and 480 ms (MT) in position R3. Nevertheless, the visualized time-frequency analysis yields considerably more insight on the phenomenon.

As for the effect of the room geometry to the assumed reflections and their contribution, the preceding discussion is based only on the interpretation of the floor plans and cross sections. Next, the monaural time-frequency analysis is extended into the spatiotemporal analysis. It provides insight to the development of the room response, and substantial evidence for the preceding discussion.

B. Spatiotemporal analysis

Spatiotemporal responses of selected measurement conditions are presented in Figs. 7 and Figs. 8. The cumulating spatial energy in azimuth and elevation planes is the mean of all sources on the stage, following Eqs. (7) and (8).

The direction and influence of the direct sound are easily noticed in the direction of the sources in all conditions (see Fig. 7). At the corresponding receiver positions, the energy of the direct sound is nearly the same. Reflected or reverberant sound has only a negligible contribution to the cumulative energy in the direction of the sources. As proposed previously, the two topmost examples in Fig. 7 show the reflection from above in position R2 in SI and MT. This confirms that the canopy reflections to arrive at 50 and 60 ms after the direct sound, respectively. The figures show that the lateral energy within 30–60 ms is the most substantial difference between these two halls. The lateral reflections in SI begin to accumulate after 20 ms from a wide region on the left hand side and slightly delayed also on the right. In contrast, in MT, most of the reflected energy arrives very soon after the direct sound from behind. As shown in the floor plans in Fig. 1, different sections of the stalls are separated with transverse walls. Although not included in the figure, positions closer to these walls receive even more emphasized reflection from behind.

The energy development in these positions is also different in the lateral back left and back right directions (circa ± 135 deg). In both halls, the energy after 30 ms is approximately the same. After that, the response in SI accumulates strongly up to 100 ms, whereas in MT, the development is only minor. The analysis indicates that the difference between the total accumulated energy is nearly 6 dB in these directions.

In addition to the canopy reflection, the topmost median plane visualization reveals also distinct reflections from the rear wall and from upper elevations.

The third example shows the development of the spatiotemporal response in position R3 in SI, where the seat dip correction is particularly strong. As seen in Fig. 6, from 30 to 40 ms, all frequencies below 1 kHz are amplified, but the shape of the frequency response is not modified. The elevation plot indicates that the canopy reflection arrives during that time window. However, it is not seen to improve the relative level of the seat dip frequencies. Instead the correction in the frequency response takes place during later time windows, which coincides with a considerable amount of reflected lateral energy.

The bottom example in Fig. 7 presents spatiotemporal development in position R4 of FT. From the left hand side

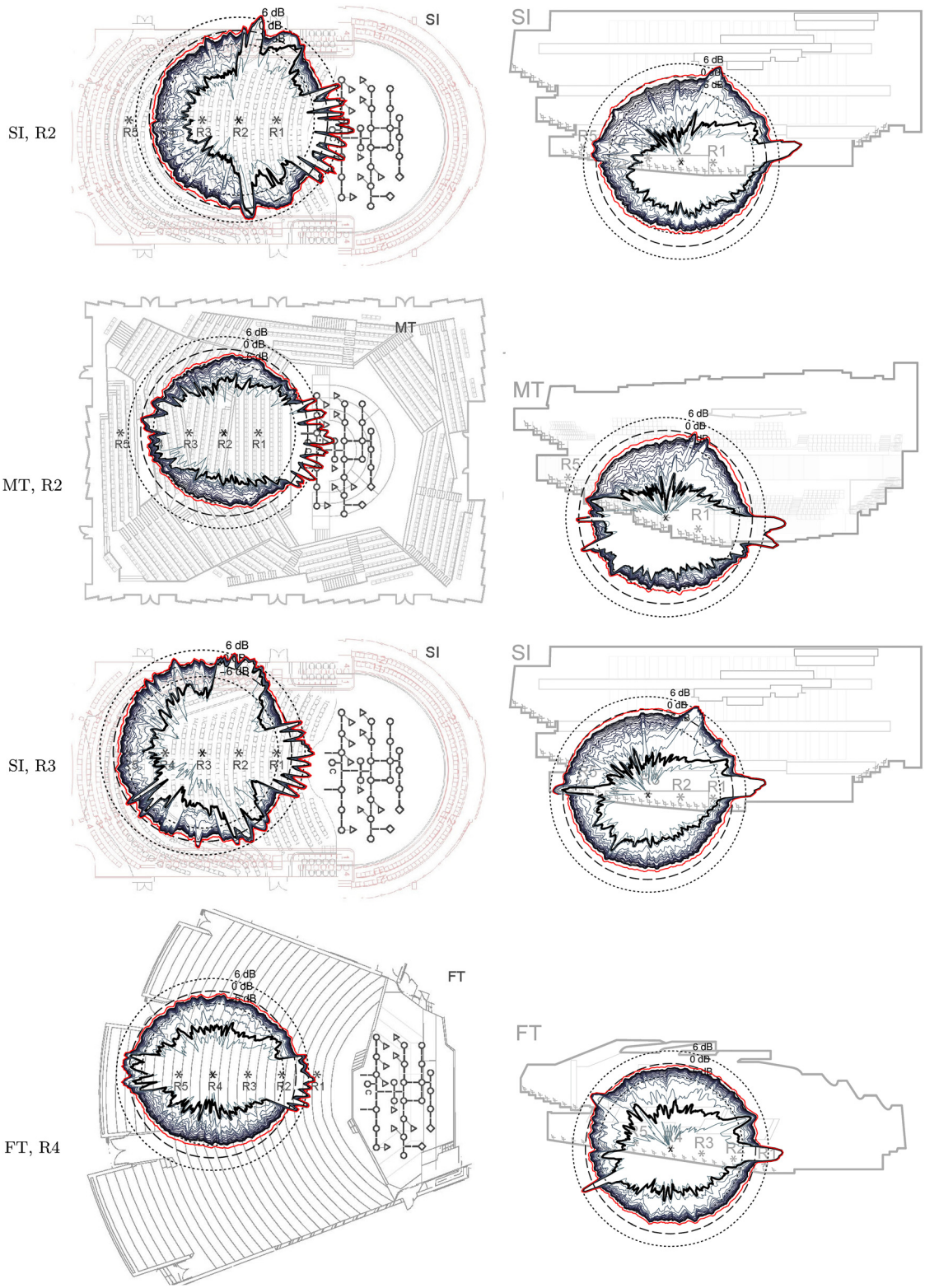

FIG. 7. (Color online) Visualized wide-band spatiotemporal responses of given halls and positions at 10 ms intervals. Spatial responses at 30 ms are plotted in bold. Histogram curves are smoothed with 3 deg sliding average.

plot it can be noticed that the early sound field consists only of the direct sound and reflections from behind. The oval-shape of the spatial response indicates that the amount of energy from the sides is low, and the distinct lateral side reflections are absent. The elevation plot shows that the ceiling reflections arrive soon after the direct sound from a comparably low elevation. The tilted balcony fronts provide also a distinct reflection.

Figure 8 visualize three more examples of spatiotemporal analysis. The topmost pair of plots shows the development in VE position R2. In comparison to the halls included in Fig. 7, this hall has an array of reflectors above the stage.

The influence of the overhead reflectors is visible between 20 and 30 ms in the median plane. In contrast to halls with a canopy, the reflection from the overhead array is not as focused. Slightly later in time, the energy increase in higher elevation suggests reflections from the technical balconies. In the lateral plane, the straight side walls provide early reflections. Compared to the same position in SI, the effect is considerably less pronounced. In addition, the amount of overall energy from behind is relatively low, and most of the energy increase after the first reflections comes from above.

The plots in the middle show the corresponding development in position R5. Here the lateral energy plot assumes again the oval shape as the reflections from the side walls arrive from a narrower angle. In the median plane, the balcony overhang causes strong reflections.

While not shown in the figures, the spatiotemporal responses in PR correspond to those in VE. In the front stalls in PR, the lateral reflections occur earlier in time due to the narrower rectangular geometry. The amount of late energy increase in all directions is very strong.

The bottom plots in Fig. 8 show a special case in KT, which has very wide fan shape. The amount of lateral energy from the sides is small as in FT (see Fig. 7). In contrast, the relatively low ceiling causes multiple early reflections very closely in time. The effect is emphasized even further in position R5, where the median plane response is nearly round. Other positions exhibit the oval shape in the lateral plane, and a strong ceiling reflection with limited late energy increase.

1. Discussion

In general, the spatiotemporal visualizations correspond to many features in the hall geometries. Shoebox-shape halls have visible areas of early reflections on the sides, and the horizontal angle of the reflections decrease with further receiver distances. Ceiling structures yield considerable reflected energy, while massive canopies result in distinct reflections.

The oval shape is found regularly in the lateral planes of fan-shape halls. However, also MT with the vineyard shape exhibits the narrow lateral energy distribution. A natural explanation for this is the outward-splayed walls that effectively function as a fan shape. Together these findings indicate that the overall geometry has a major effect on the lateral distribution of the early sound energy.

Importantly, lateral responses in MT resemble little those seen in SI. The plots in Fig. 7 show the lateral energy within 30-60 ms to be the most substantial difference between these two halls. As suspected in Sec. VI A 1, SI provides substantially more early energy from the sides than MT, while the energy distributions in the median plane are

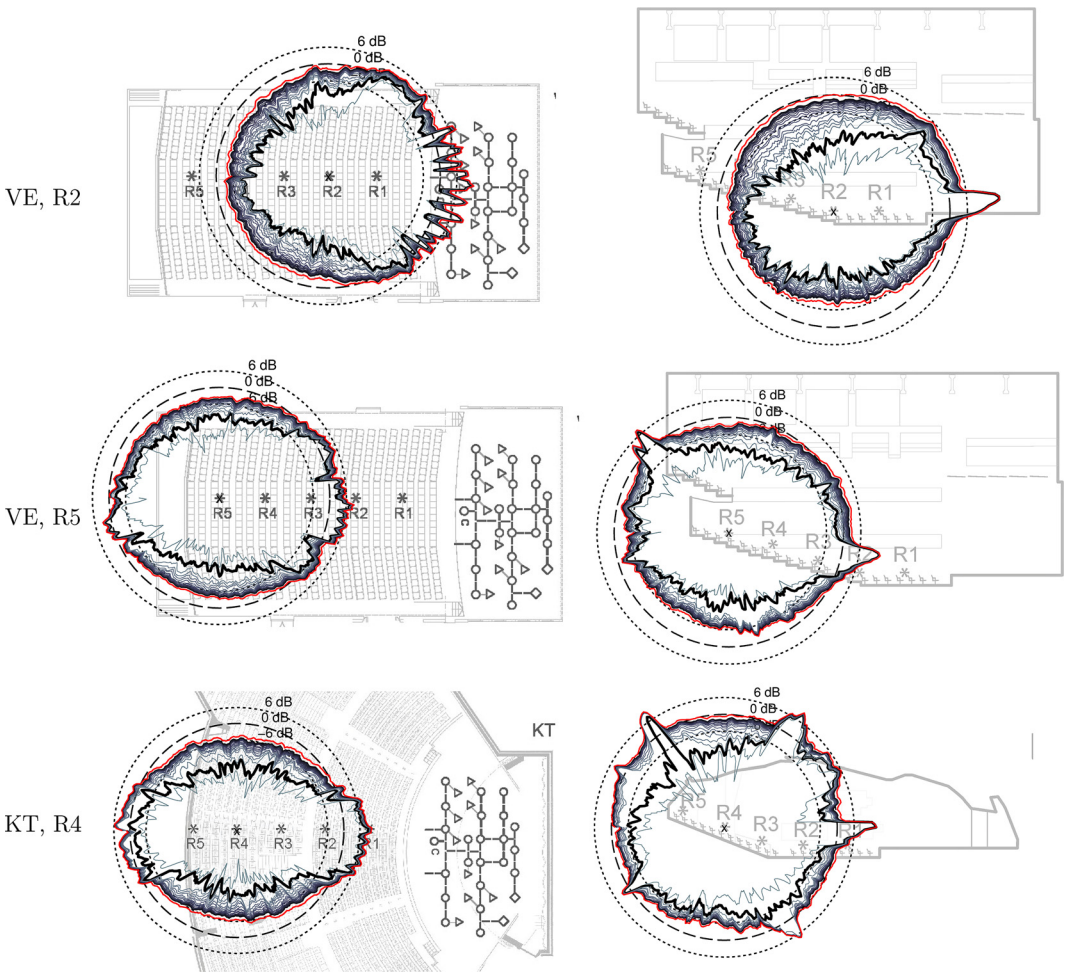

FIG. 8. (Color online) Visualized spatiotemporal responses of three positions in two halls. The visualization corresponds to Fig. 7. Floor plan of KT is truncated due to the width of the hall.

more similar. Again considering the seat dip effect, this result supports the hypothesis where the early reflections from the side would be beneficial for cumulatively correcting the attenuation from the seat dip in the direct sound.

From another aspect, the results suggest in halls with more lateral energy, the effective bass response is usually better. One can draw such a connection from time-frequency and spatiotemporal analysis in positions R2 in different halls. This theory is supported by results in VE, where the shape of the lateral energy with increasing distance resembles more those seen in fan-shape halls in the spatiotemporal analysis (see Fig. 8). Simultaneously, the overall bass level is decreased more compared to shoebox halls.

The spatiotemporal analysis reveals that in unoccupied measurements the seats behind reflect a considerable amount of energy. Further analysis with low-pass filtered impulse responses shows that the phenomenon occurs also at frequencies below 400 Hz. At such frequencies, the difference between unoccupied and occupied seats is considered small. Hence it is conceivable that this effect is present also in occupied situations. Disregarding positions that are close to vertical surfaces, the effect was prominent in all positions in FT and KT and in the front-most positions in other halls. Reasons for this variation remain inconclusive as the measurements in different halls were conducted in an unchanged manner.

Unlike with time-frequency responses, the presented spatial visualizations do not convey ISO 3382-1 parameters as such. On the one hand, the standardized spatial objective employs a combination of omnidirectional and figure-of-eight microphones or respective weighting for calculation of lateral energy fractions. On the other hand, inter-aural cross-correlation relies on binaural measurements and calculation on sample basis. However, the presented visualizations reveal considerably more information in an intuitive manner.

It should be noted that for the proposed analysis, a relatively large number of source positions is preferred due to their smoothing effect for improved readability. This can be noticed by comparing visualizations in Figs. 4 and 7. In contrast to the objective parameters that reduce data by linearizing the measured decay or by integrating energy over time, the spatiotemporal analysis aims to reduce data by averaging spatial features over several source positioning. Thus the spatial features affecting the whole source area become emphasized in the visualizations.

Based on the results from the time-frequency and spatiotemporal analysis from hall measurements, hard, smooth walls are found to contribute to the early correction of the initial seat dip attenuation in the cumulative frequency response. In contrast, diffusing side walls do not seem to provide this effect. These findings complement earlier research where a similar effect has been proposed for a reduced ceiling height.[15]

VII. CONCLUSIONS AND FUTURE WORK

Two methods for representing the acoustic response in concert halls were presented. The first method is based on representing the cumulating frequency response within extending time windows. The time-frequency analysis shows efficiently the development of the monaural sound field at different frequencies. This enables particularly studying the changes in the spectrum of the early energy, that is, the direct sound, early reflections, and the seat dip effect. The second method presents the development of the spatial sound field over time. This spatiotemporal analysis in lateral and median planes provides intuitive means to convey the most substantial spatial features concerning the receiving positions.

Impulse responses from six concert halls were measured using a large number of loudspeakers distributed in a constant pattern, representing a symphony orchestra. Five receiver positions at corresponding distances in each hall were used to demonstrate the time-frequency and spatiotemporal analysis methods. Comparison between the objective parameters and the time-frequency analysis showed that the time-frequency representation communicates the ISO 3382-1 objective parameters. Furthermore, different omnidirectional parameters were shown to correlate strongly with the late energy increase in the impulse responses. The time-frequency analysis presented large differences between measured halls in the seat dip attenuation and the subsequent recovery after the direct sound. Hypotheses explaining these differences were supported by a spatiotemporal study from which the effective directions and time instants for early reflections could be determined. The results show that the best correction of the frequency response after the initial seat dip attenuation takes place in conditions where the lateral reflections are the strongest. Moreover, subjective features of the concert halls found with recent listening tests are demonstrated by a combination of time-frequency and spatiotemporal visualizations.

Considering the importance of the early development in the frequency response and its apparent connection to the subjective acoustic quality, the present findings call for detailed studies on the lateral reflections and seating geometries to optimize the low-frequency response in concert hall designs. The information from the time-frequency and spatiotemporal analyses can be potentially used as a basis for novel objective indicators of the quality of concert hall acoustics. Moreover, the analyses should be applied to more halls of different types for better understanding their characteristic acoustic features.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the Academy of Finland, Project Nos. 218238 and 140786 and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement No. [203636].

APPENDIX: DIRECTIONAL ANALYSIS METHOD FOR SPATIAL ROOM IMPULSE RESPONSES

For each time step, the directional analysis (DIR) estimates the 3-D location of the arriving sound field from the spatial room impulse response. Then DIR assigns a pressure value from an omnidirectional pressure impulse response for each of these locations. The applied localization function depends at least on the applied microphone array and

acoustic conditions. In this article, we use the unconstrained least squares solution for plane-wave propagation model.[39] It is the efficient estimator for solving the plane-wave direction from the time difference of arrival (TDOA) estimates when independent errors and equal variances for the errors (of the TDOA estimates) are assumed. Other localization functions are listed, for example, by Tervo *et al.*[32]

Given the room impulse responses $\mathbf{h}_l(t) = \{h_{l,n}(t)\}_{n=1}^{N}$ at N microphones, the localization proceeds as follows. For simplicity, l is omitted in the following. For each discrete time step k, i.e., at every $\Delta t = 1/f_s$, where f_s is the sampling frequency, the room impulse responses are windowed with a Hanning window of size 1.3 ms (64 samples at 48 kHz sampling frequency). The window is centered at the sample of interest k. Then generalized correlation method with direct weighting estimates the TDOA between microphones i and j (Ref. 40):

$$\hat{\tau}_{i,j}^{(k)} = \arg\max_{\tau}\{\mathrm{R}_{i,j}^{(k)}(\tau)\}, \tag{A1}$$

where

$$\mathrm{R}_{i,j}^{(k)}(\tau) = \mathcal{F}^{-1}\{H_i^{(k)}(\omega)(H_j^{(k)}(\omega))^*\} \tag{A2}$$

with inverse Fourier-transform $\mathcal{F}^{-1}\{\}$, frequency domain versions of the windowed impulse responses $H_n^{(k)}(\omega)$, and $()^*$ denotes complex conjugate. For subsample accuracy, each TDOA estimate is interpolated with the exponential fit[41] (the superscripts are omitted for simplicity):

$$\hat{\tau} = \hat{\tau}^d + \delta \tag{A3}$$

where $\hat{\tau}^d$ is the original TDOA estimate from Eq. (A1) and

$$\delta = \frac{(\log(R(\Delta t+1)) - \log(R(\Delta t-1)))}{4\log(R(\Delta t)) - 2\log(R(\Delta t-1)) - 2\log(R(\Delta t+1))}. \tag{A4}$$

The set of instantaneous TDOA estimates is denoted with

$$\hat{\boldsymbol{\tau}}_k = [\hat{\tau}_{1,2}^{(k)}, \hat{\tau}_{1,3}^{(k)}, \ldots, \hat{\tau}_{N-1,N}^{(k)}]^T,$$

where N is the number of microphones, and the corresponding microphone position difference vectors with

$$\mathbf{V} = [\mathbf{r}_1 - \mathbf{r}_2, \mathbf{r}_1 - \mathbf{r}_3, \ldots, \mathbf{r}_{N-1} - \mathbf{r}_N]^T.$$

The least squares solution for slowness vector is given as[39]

$$\hat{\mathbf{m}}_k = \mathbf{V}^+\hat{\boldsymbol{\tau}}_k, \tag{A5}$$

where $(\cdot)^+$ is Moore–Penrose pseudo-inverse, and the direction of the arriving sound wave is given as $\hat{\mathbf{n}}_k = -\hat{\mathbf{m}}_k / \| \hat{\mathbf{m}}_k \|$. The distance to the plane-wave k is given directly by $d_k = ck\Delta t$, where c denotes the speed of sound. These four parameters, the normal vector and the distance of each plane-wave, can be translated to azimuth and elevation angles $[\hat{\theta}, \hat{\phi}]$ with standard coordinate transformations. The distance can be omitted from the presentation because it is directly described by the time moment as shown in the preceding text.

The first part of DIR assigned a location for each time step in the spatial room impulse response. In the second step, each location is given a pressure value from an omnidirectional impulse response h_p, which is one of the microphones of the applied array. Ideally, the pressure value is obtained from an omnidirectional microphone that is in the geometric center of the microphone array. In case the pressure microphone is not in the geometric mean of the array, one has to predict the value of the pressure signal according locations of the estimated plane-waves. However, because in this paper the dimensions of the array are small compared to the dimensions of the enclosure and the prediction of the pressure value would require some computational effort due to Fourier-interpolation of the signal, we directly assign pressure at the topmost microphone ($n = 5$) as the pressure signal, i.e., $h_p(\Delta t) = h_5(\Delta t)$. Then each spatial impulse response is presented by three values $[h_p(\Delta tk), \hat{\theta}(\Delta tk), \hat{\phi}(\Delta tk)]$ at each time moment Δtk [see Eq. (5)]. The window length of 1.3 ms limits the sensitivity of the direction estimation. Therefore directions for incident sound energy, i.e., reflections, are estimated fully independently if their separation in time is 1.3 ms or more. This is considered to be sufficient accuracy for most real spaces.

The accuracy of TDOA-based localization methods for spatial room impulse responses has been evaluated in Ref. 32. In short, the TDOA-based methods enable the accurate localization of the most significant features in the room impulse responses. This is clearly illustrated in the principal results of the present paper.

[1]ISO 3382-1:2009; *Acoustics—Measurement of Room Acoustic Parameters. I: Performance Spaces* (International Standards Organization, 2009).
[2]J. S. Bradley, "Review of objective room acoustics measures and future needs," Appl. Acoust. **72**, 713–720 (2011).
[3]T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, and L. Savioja, "Engaging concert hall acoustics is made up of temporal envelope preserving reflections," J. Acoust. Soc. Am. **129**, EL223–EL228 (2011).
[4]R. Lacatis, A. Gimenez, A. B. Sevillano, S. Cerda, J. Romero, and R. Cibrian, "Historical and chronological evolution of the concert hall acoustics parameters," in *Aoustics'08*, Paris, France (2008), pp. 2151–2156.
[5]J. Bradley, "The evolution of newer auditorium acoustics measures," Can. Acoust. **18**, 13–23 (1990).
[6]T. Rossing, "Acoustics in halls for speech and music," in *Handbook of Acoustics* (Springer Verlag, New York, 2007), Chap. 9, pp. 299–360.
[7]A. Abdou and R. W. Guy, "Spatial information of sound fields for room-acoustics evaluation and diagnosis," J. Acoust. Soc. Am. **100**, 3215–3226 (1996).
[8]L. Beranek, *Concert and Opera Halls: Music, Acoustics, and Architecture* (Springer, New York, 2004), 661 pp.
[9]E. Kahle and J.-P. Jullien, "Some new considerations on the subjective impression of reverberance and its correlation with objective criteria," in *Sabine Centennial Meeting*, Cambridge, MA (1994), pp. 91–94.
[10]T. Hidaka, "Supplemental data of dependence of objective room acoustical parameters on source and receiver positions at field measurement," Acoust. Sci. Technol. **26**, 128–135 (2005).
[11]T. Hidaka, Y. Yamada, and T. Nakagawa, "A new definition of boundary point between early reflections and late reverberation in room impulse responses," J. Acoust. Soc. Am. **122**, 326–332 (2007).
[12]M. Barron, "Bass sound in concert auditoria," J. Acoust. Soc. Am. **97**, 1088–1098 (1995).

[13]T. Hidaka and L. Beranek, "Objective and subjective evaluations of 23 opera houses in Europe, Japan, and the Americas," J. Acoust. Soc. Am. **107**, 368–383 (2000).
[14]W. J. Davies and Y. W. Lam, "New attributes of seat dip attenuation," Appl. Acoustics **41**, 1–23 (1994).
[15]J. S. Bradley, "Some further investigations of the seat dip effect," J. Acoust. Soc. Am. **90**, 324–333 (1991).
[16]G. A. Soulodre and J. S. Bradley, "Energy relations in concert auditoriums," J. Acoust. Soc. Am. **98**, 294–301 (1988).
[17]J. Merimaa, T. Lokki, T. Peltonen, and M. Karjalainen, "Measurement, analysis, and visualization of directional room responses," in *111th Convention of Audio Engineering Society*, New York (2001), Paper No. 5449.
[18]A. Bassuet, "New acoustical parameters and visualization techniques to analyze the spatial distribution of sound in music spaces," J. Build. Acoust. **18**, 329–347 (2011).
[19]B. Gover, J. Ryan, and M. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array," J. Acoust. Soc. Am. **116**, 2138–2148 (2004).
[20]S. Tervo, J. Pätynen, and T. Lokki, "Acoustic reflection path tracing using a highly directional loudspeaker," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY (2009), pp. 245–248.
[21]Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by closely located four point microphone method," J. Acoust. Soc. Jpn. **10**, 101–110 (1989).
[22]Y. Fukushima, H. Suzuki, and A. Omoto, "Visualization of reflected sound in enclosed space by sound intensity measurement," Acoust. Sci. Technol. **27**, 187–189 (2006).
[23]T. J. Schultz and B. G. Watters, "Propagation of sound across audience seating," J. Acoust. Soc. Am. **36**, 885–895 (1964).
[24]G. M. Sessler and J. E. West, "Sound transmission over theater seats," J. Acoust. Soc. Am. **36**, 1725–1732 (1964).
[25]D. Takahashi, "Seat dip effect: The phenomena and the mechanism," J. Acoust. Soc. Am. **102**, 1326–1334 (1997).
[26]W. J. Davies, T. J. Cox, and Y. W. Lam, "Subjective perception of seat dip attenuation," Acustica **82**, 784–792 (1996).
[27]J. Bradley, "Using ISO 3382 measures, and their extensions, to evaluate acoustical conditions in concert halls," Acoust. Sci. Technol. **26**, 170–178 (2005).
[28]T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," J. Acoust. Soc. Am. **131**, 3148–3161 (2012).
[29]J. Pätynen, S. Tervo, and T. Lokki, "A loudspeaker orchestra for concert hall studies," Acoust. Bull. **34**, 32–37 (2009).
[30]J. Pätynen, "A virtual symphony orchestra for studies on concert hall acoustics," Ph.D. dissertation, Aalto University School of Science, Espoo, Finland, 2011.
[31]T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," J. Acoust. Soc. Am. **130**, 835–849 (2011).
[32]S. Tervo, J. Pätynen, and T. Lokki, "Acoustic reflection localization from room impulse responses," Acta Acust. Acust. **98**, 418–440 (2012).
[33]M. Barron, "The subjective effects of first reflections in concert halls—the need for lateral reflections," J. Sound Vib. **15**, 475–494 (1971).
[34]J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, 2nd ed. (MIT Press, Cambridge, MA, 1997), pp. 201–287.
[35]M. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[36]T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," Acta Acust. Acust. **79**, 27–41 (1993).
[37]M. Schroeder, "Complementarity of sound buildup and decay," J. Acoust. Soc. Am. **40**, 511–549 (1966).
[38]M. Barron and L.-J. Lee, "Energy relations in concert auditoriums," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[39]T. Pirinen, "Confidence scoring of time delay based direction of arrival estimates and a generalization to difference quantities," Ph.D. thesis, Tampere University of Technology 2009, Publication 854, p. 75.
[40]C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327 (1976).
[41]L. Zhang and X. Wu,"On cross correlation based-discrete time delay estimation," IEEE Int. Conf. Acoust., Speech, Signal Process. **4**, 981–984 (2005).