Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences

Tapio Lokki,[a)] Jukka Pätynen, Antti Kuusinen, and Sakari Tervo
Department of Computer Science, Aalto University School of Science, FI-00076 AALTO, Finland

(Received 1 July 2015; revised 12 May 2016; accepted 30 June 2016; published online 22 July 2016)

Some studies of concert hall acoustics consider the acoustics in a hall as a single entity. Here, it is shown that the acoustics vary between different seats, and the choice of music also influences the perceived acoustics. The presented study compared the acoustics of six unoccupied concert halls with extensive listening tests, applying two different music excerpts on three different seats. Twenty eight assessors rated the halls according to the subjective preference of the assesors and individual attributes with a paired comparison method. Results show that assessors can be classified into two preference groups, which prioritize different perceptual factors. In addition, the individual attributes elicited by assessors were clustered into three latent classes. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4958686]

[MV] Pages: 551–562

I. INTRODUCTION

The public debate on the optimal acoustic conditions for audience in a concert hall is self-evident when a new concert hall is opened. A debate sparks easily on topics that are matters of individual tastes, but which are difficult to study thoroughly and objectively—such as acoustics. The discussion is often dominated by conductors and musicians, who evaluate the acoustic conditions of a new hall mainly from their perspective on stage. However, the audience expects, and likely also prefers, room acoustic conditions different from those on stage. Furthermore, the tastes for optimal acoustics vary greatly among audience members. An all-encompassing understanding about the perceptual factors underlying the individual preferences would provide a basis for recognizing different opinions of the existing venues, and help in the future designs of new halls.

The search for correlations between preferences and both objective and subjective acoustical features started already by Sabine[1] over 100 yr ago, when he presented the ground-breaking analysis of room acoustics and material absorption. However, since the 1960s the best known studies have been made by Beranek,[2] who gathered data from dozens of halls around the world. In the same decade, German research groups from Berlin and Göttingen, summarized in Ref. 3, developed experiment methods for both laboratory and *in situ* evaluations of room acoustics. As outlined by Barron,[4] these results have differences and even contradictions, most probably due to contrasting experimental conditions. However, Schroeder *et al.*[5] found a strong correlation between the consensus preference factor and a reverberation time of 2.0 s: a value often considered optimal. Sufficiently loud sound was also found important for preference. In the 1980s, Barron[4] gathered evaluations in 11 British halls from subjects listening to concerts *in situ*. By a carefully designed questionnaire, Barron confirmed the importance of reverberance and loudness for preferred sound, but also found that some subjects preferred intimate sound over reverberance. The concept of envelopment was, however, interpreted differently by different subjects. In the 1990s, Soulodre and Bradley[6] conducted a series of studies, where they found a correlation between preference and clarity or treble, but not reverberance.

Results reported in new studies often differ from the preceding ones, probably because the choice of music material and the selection of halls have a major impact on results, as identified already by Hawkes and Douglas.[7] Moreover, the differences between listening to live concerts *in situ*, perhaps on different days, and reproduced room acoustics in controlled laboratory conditions may shift the focus of the subjective observations and, thus, influence the results.

Recently, our contribution to concert hall acoustics has been the application of sensory evaluation methods[8] and disentangling the preference ratings with individually elicited attributes.[9] These studies included Finnish concert halls, which are not very well known internationally. However, those studies confirmed many of the results from previous decades. For example, a group of subjects preferred clear and intimate sound while others preferred loud, enveloping, and reverberant sound.[9] The average preference was mostly influenced by proximity,[10] a result also in line with the earlier results.

This article is built on the previous studies[8,9] using the loudspeaker orchestra,[11] a symphony orchestra simulator, to guarantee identical excitation in all halls. The auralizations for listening tests in the laboratory are performed via spatial impulse responses, which are analyzed and reproduced with a novel method as explained in Sec. II. The applied virtual acoustics technology produces very realistic spatial sound reproduction of a measured concert hall. In short, using an identical loudspeaker orchestra on stage in all the halls eliminates the variability of the human performance, and allows

[a)]Electronic mail: Tapio.Lokki@aalto.fi

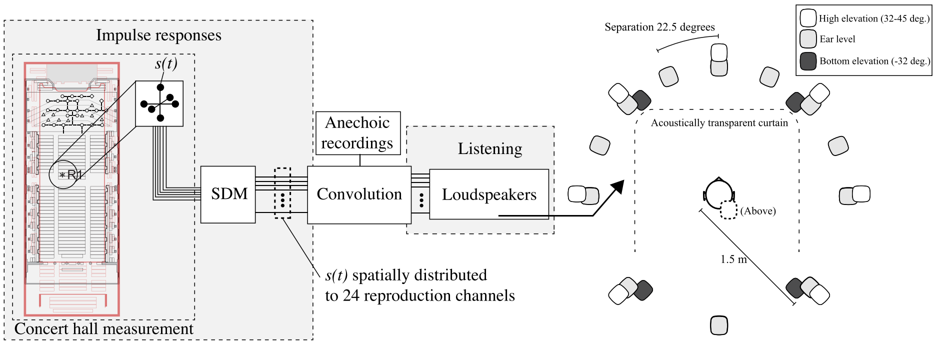

FIG. 1. (Color online) The block diagram of the auralization with the loudspeaker orchestra measurements in the concert halls. The figure shows the method for a single source channel on stage and the process is repeated for all sources for auralizing the entire orchestra.

the listener to focus only on the differences caused by room acoustics.

The purpose of this article is threefold. First, six well known European concert halls are comprehensively compared with different metrics. Three halls are traditional rectangularly shaped, the so-called shoebox halls, while other three are more modern designs in which the audience surrounds the orchestra. Second, all subjective comparisons are performed with two different music excerpts to investigate the possible influence of music style and instrumentation on the acoustic perception. The selected works represent typical orchestral music played in these halls; one music excerpt is a more tranquil passage in *mezzo piano* whereas the other one is more majestic in *forte*. Third, the study contributes to sensory evaluation methodology by applying the paired comparison method for all subjective evaluations, both for preference judgements and for sensory evaluation. The same method has been used earlier to study spatial sound reproduction techniques.[12] Here, the novel contribution is the use of individually developed attributes for the ratings and analysis of individual paired comparison data.

II. CONCERT HALL MEASUREMENTS AND AURALIZATION

The auralization of the concert hall measurements was accomplished using the process illustrated in Fig. 1. The symphony orchestra on stage was simulated with 33 calibrated loudspeakers connected to 24 channels. The details of the loudspeaker orchestra setup can be found in the previous publications.[8,9,11] The room impulse response from each of the loudspeaker channels was measured with a type 50-VI 3D vector intensity probe (G.R.A.S., Denmark) consisting of three co-centric phase-matched pairs of omnidirectional microphones arranged on the x, y, and z axes. The distance between the opposing capsules was 100 mm and the impulse responses were measured with 48 kHz sampling rate using the logarithmic sine sweep technique.[13] The six impulse responses measured at a time were analyzed with the spatial decomposition method (SDM)[14] that estimates the direction of incidence for each sample in an impulse response in short time windows. Based on the spatial information, the impulse response in the topmost omnidirectional microphone was distributed to reproduction loudspeakers as convolution reverberators. The distribution of samples was performed with the nearest loudspeaker technique in order to emphasize the spectral fidelity of the high frequencies[15] at the slight expense of spatial accuracy. Such choice was adopted based on the earlier results, which clearly shows the importance of timbral fidelity over spatial fidelity.[16] Finally, the anechoic recordings[17] were convolved with all reproduction channel responses. The distribution of the instruments to stage loudspeaker channels was the same as earlier[8] and when the process is repeated to all sources on the stage the end result is a realistic reproduction of an orchestra in a concert hall.

III. METHODS

A listening test was organized to explore which concert halls the assessors like and to what kind of differences the assessors pay attention to between halls. This section describes in detail the compared concert halls, the music stimuli, the listening room, the assessors, and the applied evaluation methodology.

A. Concert halls and receiver positions

The measured concert halls include six European concert halls, which are listed in Table I with their generic physical and acoustical parameters. Three of the halls are of rectangular halls with a flat floor, lightweight seats, and parallel side walls. The other three represent more modern designs with terraced seating blocks, or semi-circular amphitheater style. Figure 2 depicts the floor plans of the halls overlaid in the same scale. The plans are aligned with respect to the loudspeaker orchestra on stage and the receiver positions in the audience area.

We included three corresponding receiver positions in each hall in the listening tests. The position at the front stalls (FRO) was at 11 m distance from the front line of the orchestra and 2 m to the left of the midline of the hall. The second

TABLE I. List of European concert halls included in the listening experiment. V, N, G, and EDT denote volume in m^3, number of seats, average strength in dB and average early decay time in seconds, respectively. Measured values for G and EDT are averages from 500 and 1000 Hz octave bands (*estimated).

	Hall	Shape	V	N	G	EDT
VM	Vienna Musikverein	Shoebox	15 000	1680	4.1	3.1
AC	Amsterdam Concertgebouw	Shoebox	18 780	2040	2.8	2.4
BK	Berlin Konzerthaus	Shoebox	15 000	1575	2.7	2.1
BP	Berlin Philharmonie	Vineyard	21 000	2220	2.1	1.9
CP	Cologne Philharmonie	Fan	*19 000	2000	1.9	1.6
HM	Helsinki Music Centre	Vineyard	24 000	1700	1.4	2.0

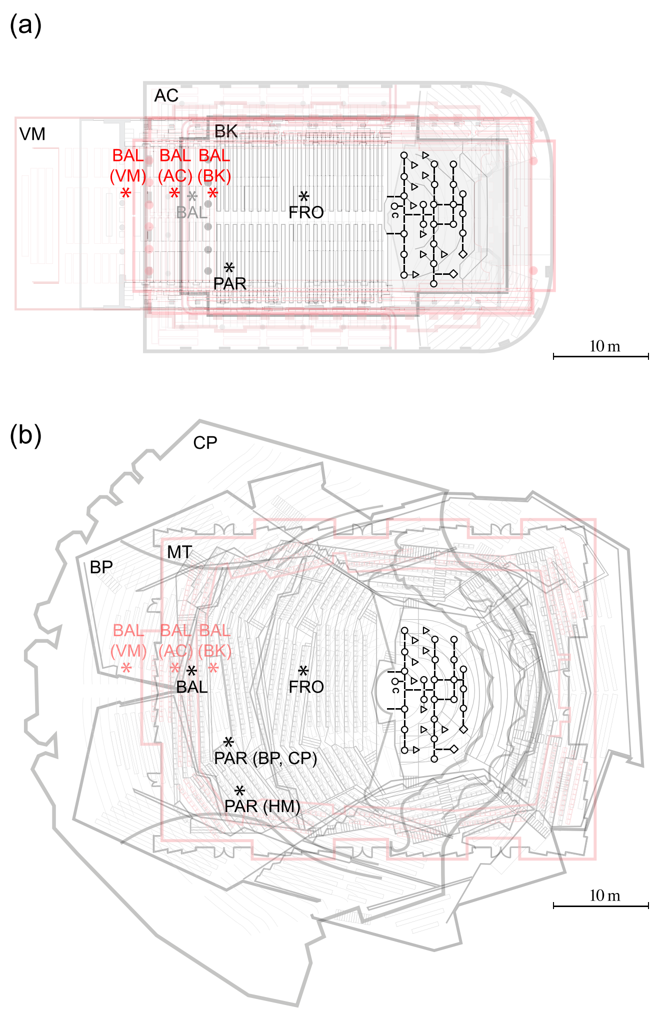

FIG. 2. (Color online) Overlaid floor plans of the included concert halls of two different types in the constant scale. Plans are aligned by the receiver positions (FRO, PAR, BAL) and the loudspeaker orchestra. Balcony areas, where applicable, are displayed with different shade. The abbreviations of the halls are described in Table I.

position further back in the parterre (PAR) was 19 m from the orchestra and 6 m to the right from the midline of the hall, except in HM, where it was 10 m from the midline. The position at the first row of the balcony is abbreviated as BAL. In halls BP and CP without a balcony, a corresponding position at the rear parterre was chosen. The receiver positions are marked in Fig. 2. In the listening tests the halls were compared only on the same seats, meaning that the physical distance between the orchestra and the listening position was constant, except for BAL positions, where small distance differences result from the geometries of the halls.

B. Music stimuli

Two short musical excerpts were convolved with the SDM processed spatial impulse responses. The first one was a tranquil passage in *mezzo piano* from Symphony No. 7, movement I, bars 44–47 by L. v. Beethoven. This 10.5 s long excerpt contains woodwind and string instruments, but no brass. The second music was a more powerful passage in *forte* from Symphony No. 8, movement II, bars 33–40 by A. Bruckner. This sample of 10 s is dominated by brass instruments, mainly trumpets and the French horns. Due to the long exposure by assessors in the listening test, the reproduction levels of the presented stimuli were different than the authentic sound levels experienced in the concert halls. However, the relative sound level differences were left intact. To avoid listening fatigue with repeating music signals and to facilitate hearing the differences between halls, the Beethoven passage was played somewhat louder than would have been *in situ* and the measured L_{Aeq} of the whole 10.5 s sample in the listening room was 65 dB (the minimum L_{AS} was 62 dB). Correspondingly, the 10 s long Bruckner sample was played back at level $L_{Aeq} = 74$ dB (the maximum L_{AS} was 80 dB).

C. Listening room

The listening tests were conducted in a semi-anechoic room with a mean reverberation time of 0.11 s at the mid-frequencies. The walls and the ceiling are treated with sound absorbing materials, which are at least 5 cm thick and leave a varying air space behind the absorption. The rigid floor is covered with a carpet around the listening position. The average peak-to-peak level difference between the direct sound and the strongest reflection on the 1–8 kHz frequency band is 12.8 dB. The ITU-R BS.1116-1 recommendation for subjective audio evaluation systems suggests a peak-to-peak difference of at least 10 dB. Hence, the listening room complies with the recommended values.

The reproduction system for spatial sound comprises 24 loudspeakers surrounding the listening position in three-dimensions (3-D). Figure 1 illustrates the positioning of the loudspeakers. In short, the lateral plane at the ear level has 12 loudspeakers, while eight and four loudspeakers are located in upper and lower elevations, respectively. Most loudspeakers (16) are in the frontal hemisphere, as the spatial resolution of human hearing is the most accurate in front, and in concert halls most of the sound energy arrives from that region. Nominal distance from the listening position to the loudspeakers is 1.5 m. The distances from the listening point to all the loudspeakers are not exactly the same due to simplifications in the installments. The small differences are compensated for by delaying respective loudspeaker signals. Calibrated sound levels from individual loudspeakers are within ±0.5 dB.

D. Assessors

Twenty eight assessors (14 males, 14 females; ages between 22 and 64 yr with average age of 39.6) were recruited for this study. We gathered the participants with a web-based questionnaire and we particularly looked for people who often go to live concerts. The selected assessors were 10 professional musicians (no specific genre), 10 active amateur musicians, and eight active concert goers with varying musical background. Thus, they could all be considered as experts in listening to music, even though not expert assessors nor listening test participants. All assessors were native Finnish speakers.

Assessors were screened with standard audiometry (Madsen Micromate, Otometrics, Denmark). Twelve of the assessors did not have a hearing threshold of more than 20 dB hearing level (HL) in either ear at any frequency between 250 Hz and 8 kHz. Ten of the assessors had a higher threshold on some frequency in one ear and six of them in both ears. For most of these assessors the higher hearing threshold was 25 or 30 dB HL at 6 kHz frequency. Although the higher thresholds suggest a mild hearing loss, none of the candidates were rejected, considering that the average age and occupation represent an average sampling of the typical concert audience.

All assessors visited the listening room five times, 2 h at a time on five different days. The results reported here were obtained from four individual sessions, the results of the fifth listening session are reported in other publications. The assessors were paid 150 Euro each for the active participation in the test sequence.

E. Listening test method

All listening tests were performed with a pairwise comparison methodology. In contrast to earlier tests employing parallel comparisons,[8,9] the intention here was to reduce the number of simultaneously assessed stimuli, and thus to provide the subjects with a simplified listening task. In the tests reported here, all halls were compared against each other in respective receiver positions. Hence, six halls result in 15 comparisons per receiver position. In total, three receiver positions yielded 45 pair comparisons, presented always in fully randomized order.

The subjects completed the pairwise comparisons using a user interface on a touch screen. The screen was positioned on a support in front of the subject at an appropriate height for keeping the subjects looking more forward than down, as in a concert situation. The user interface provided virtual buttons for switching seamlessly between the compared pair of stimuli, and checkboxes to input the selection. The stimuli were played in loops and the assessors could only switch seamlessly between the stimuli, indicate the selection from the present pair, or pause the playback. That is, the system output gain was kept constant, and the subjects could not change the looped segment.

The listening test process for each individual assessor is illustrated in Fig. 3. The first listening session started with a short introduction of all 18 samples (6 halls × 3 positions), which were played in random order. After this, the assessors completed the first preference test. Then, the hearing of the assessors was screened with the audiometry and the first listening session was ended with the attribute elicitation. In practice, the elicitation was performed with an AAB test with 30 randomly selected pairs out of 45 possible pairs, presented in random order. The balanced selection included 10 pairs of halls from each position. The task of an assessor was to find a sample that was different within a triplet (AAB) and to write down the main perceptual difference with an adjective. After completing all the triplets the list of adjectives was discussed with the organizer of the test and two to four most prominent attributes were chosen by the assessor.

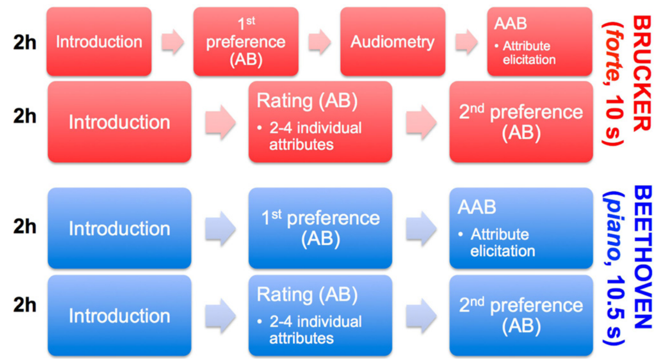

FIG. 3. (Color online) The listening test process for each assessor with four 2 h sessions. AB means paired comparison and AAB means triplet with hidden reference.

Finally, the assessors wrote down one or two sentences that define their attributes.

The second listening session started again with a brief familiarization of all samples. Then the assessors rated the samples with their own two to four attributes, one attribute at a time. There was a short break after each round of attribute rating. Finally, the second session ended with a second preference rating. Third and fourth sessions repeated the whole process (without the audiometry) with another music excerpt, thus new preference ratings and new set of attributes were collected. To balance the dependency of music, half of the assessors started with Bruckner and the other half with Beethoven.

F. Analysis methods

Paired comparisons of a set of stimuli by one assessor yields binary data between the stimulus pairs. The resulting matrix indicates when concert hall X is chosen over concert hall Y. When these matrices are aggregated over all subjects (and/or replications), the result is a matrix of choice frequencies. The analysis of such data have been investigated extensively (see, e.g., David[18] for a review) and while several approaches and model formulations exist they generally tend to yield similar results in terms of interpretation.

In this study, the paired comparison data were analysed with the Bradley–Terry–Luce (BTL) model.[19,20] It estimates the probability of choosing one sample over the other samples. As the comparison was always done with six stimuli, i.e., halls, the chance level was 1/6. Moreover, the analysis is performed with an extension[21] of the latent class segmentation method for paired comparison data presented by Courcoux and Semenou.[22] This method is based on BTL, but it assumes possible group heterogeneity, i.e., a number of latent classes, and incorporates this heterogeneity to the estimation of the scale values. Expectation maximization (EM) algorithm[23] is used to calculate maximum likelihood estimates of the scale values in each class and the probabilities of each assessor/attribute belonging to each class. The relative weights between the classes are calculated as the average of the individual probabilities belonging to each class. The appropriate number of classes is indicated by information criteria, i.e., Akaike (AIC), Bayesian (BIC), and consistent Akaike information criterion (CAIC). More details of the original implementation are given in Refs. 21 and 22.

Since its original formulation, this algorithm has now been extended to allow for a data structure, where the same stimuli have been evaluated by the same assessors/attributes but with different criteria or contexts. For example, in preference testing of consumer products, it is possible to ask assessors to give their preferences in terms of a particular set of criteria, e.g., aroma, odour, colour, design, or usability. Or they can be asked to give their judgments in different contexts or environments, such as at home, on the bus, or on the plane. Both of these scenarios result in a paired comparison data structure with separate data sets for each criterion or context. The latter scenario is the case in the current experiment, where concert halls are compared with two different music and on three different positions, that is in six different contexts.

Given the original implementation of the analysis method, it would be possible to estimate the number of classes and the associated scale values separately per each context. But for instance in preference testing, this could lead to observing a different number of classes for each criterion, i.e., listening position, and the results could be difficult to interpret when the assessors may be associated with a different group for each different criterion. Although, there might be some information in the group configurations between the criteria/contexts, the interpretation is greatly simplified by assuming a single classification across the separate data sets.

In practice, this feature is added to the algorithm by including an extra dimension, that is, the criteria/contexts, and by using the EM algorithm similarly as before, but instead of estimating one set of scale values per each class, each class contains as many sets of scale values as there are criteria/contexts. The analysis was performed in R statistical programming language using the CompR-package.[24]

Besides the differences between concert halls, an additional objective in this study was to see whether the assessors change their opinions between preference ratings collected before and after the descriptive profiling. To this end, McNemar's test[25] is used to investigate the possible changes in opinions as it takes into account that the preferences before and after were collected with the same individuals.

IV. RESULTS

The following sections present first the overall preference ratings, followed by the segmentation of assessors into preference classes. Then the results of the attribute elicitation process are exposed with the clustering of attributes. Finally, both preference and attribute ratings are correlated with the ISO3382-1:2009 room acoustical parameters, computed from the measured impulse responses that were used in the auralizations.

A. Preferences

All concert hall pairs were rated in total of 112 times (28 subjects $\times$ 2 music excerpts $\times$ 2 repetitions) for each listening position. Due to the repeated test before and after the attribute elicitation and rating, it is possible that the evaluations for preference change during the test sequence.

Figure 4 shows the first and second preference ratings for both music excerpts and for all three listening positions. The results clearly illustrate that preferences have large differences between listening positions. An overall observation suggests that at front stalls position, the preference is rather evenly dispersed between halls, in particular, with Beethoven. In contrast, BK is generally preferred over other compared halls on more remote seats.

The rank order of halls does not remain entirely constant before and after the attribute elicitation and rating processes. Figure 4 shows that largest changes in preferences occurred at front stalls. Furthermore, McNemar's test statistics in Table II indicate that halls CP and HM were liked more in the second rating for the disadvantage of VM and AC. Thus, the preferences changed slightly when the assessors had been listening to the same stimuli for an extended period. This change was also observed in the informal discussions after the experiments, where some assessors said that they sometimes got irritated by certain samples, which then manifested in more extreme judgements in the latter preference test. One may speculate on the reasons, but one possibility is the pronounced reverberation in unoccupied VM and AC in comparison to other halls.

The literature review stated that nearly all previous studies have found that different people like different kind of acoustics. As mentioned in Sec. III F, such heterogeneity is assumed in the latent class segmentation method[22] applied here, and allows us to study the effects related to preference groups. The paired comparison data from preference judgments before and after attribute judgments were aggregated to form overall individual preference matrices for each music excerpt and position combination, that is, for each context.

The appropriate number of two latent classes, i.e., preference groups within this data, was decided based on the information criteria as shown Fig. 5. The BTL scales illustrating the probabilities to prefer one hall over the others are depicted in Fig. 6.

The largest assessor group (Class 2, 16 assessors) is clearly in favor of CP for Bruckner at front stalls, i.e., close to the orchestra. At more remote seats these assessors like BK, BP, and CP. For Beethoven, the result is slightly different, at front stalls VM and AC are not preferred while others are found equally good. At parterre, BK stands out from all the others except BP. On balcony, the halls have no difference in preference, except VM, which is disliked. In contrast, the other assessor group (Class 1, 12 assessors) is in favor of VM, AC, and BK at all positions for Beethoven, showing no preference at all for BP, CP, and HM. For Bruckner the preferred halls are different, that is VM, BK, and CP at front stalls, BK at parterre as well as BK and AC at balcony.

To summarize, Fig. 6 shows that preference is highly individual, and on an overall level assessors are grouped into two preference classes. Moreover, the judgements vary between listening positions and for different music. This indicates that it is almost impossible to define which hall is

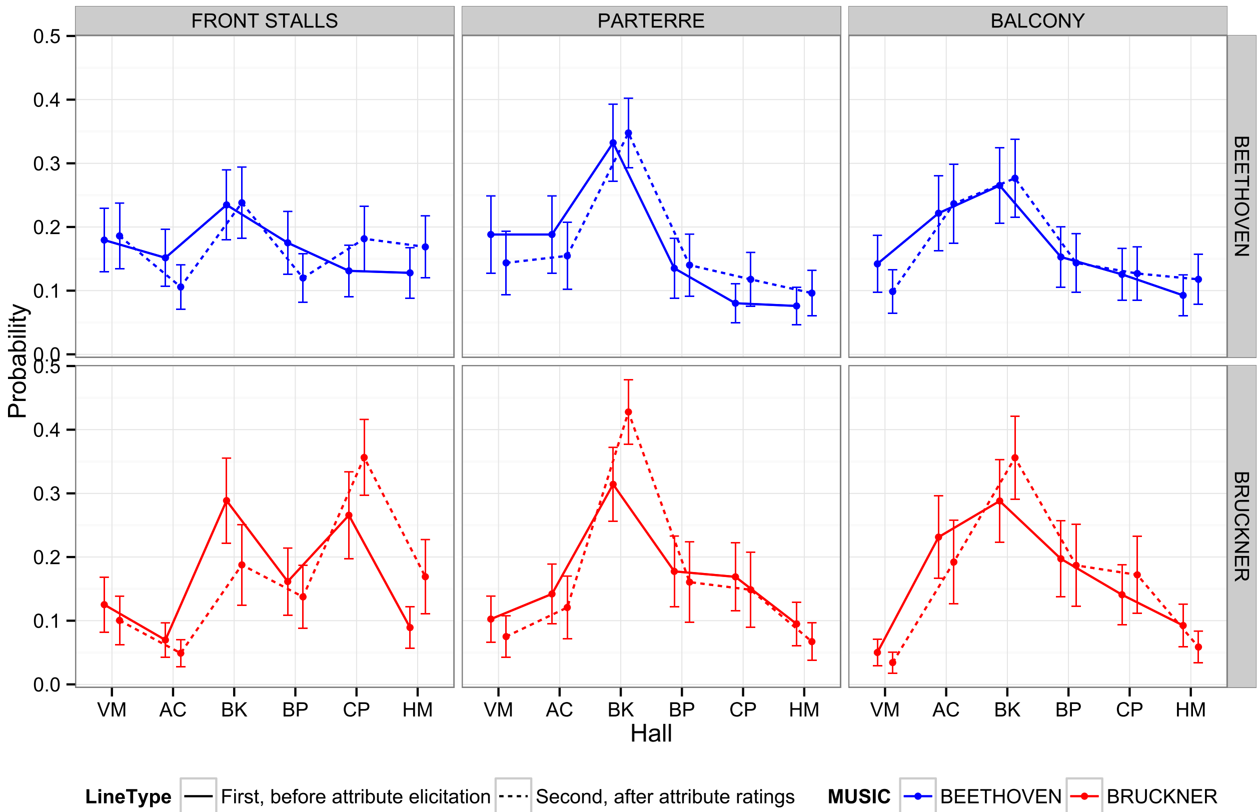

FIG. 4. (Color online) Preferences on the BTL ratio scale, i.e., scale values are parameter estimates of the BTL model fitted to paired-comparison judgments. Error bars show 95% confidence intervals.

the most preferred. However, one may observe that halls BK, CP, and VM stand out in some group and in certain position, while halls AC, BP, and HM never received the highest probability to be rated the best.

B. Attributes

The elicitation of individual attributes for each assessor was conducted with the method explained in Sec. III E. Each of 28 assessors was asked to define four attributes, which would have led to a total of 112 attributes per music. The assessors were encouraged to use whatever attributes they found appropriate. In practice, also two or three attributes were acceptable if an assessor was not confident with all of his attributes. Therefore, the total number of individually rated attributes was 100 for Bruckner and 95 for Beethoven. The same attributes were applied in all three positions.

The AAB test used in the attribute elicitation also allows the evaluation of the performance of the assessors. With Bruckner, the maximum number of discrimination errors in AAB triplets was four out of 30 and the mean over all assessors was 1.1 errors. With Beethoven, the worst assessor made eight errors and the mean was 2.4 errors. This indicates that with Beethoven the differences between halls were slightly more difficult to hear. The Kruskal–Wallis test for the number of errors between music excerpts confirms this observation [$\psi^2(1) = 6.6086$, $p = 0.010$]. However, all 28 assessors were considered capable of hearing differences between halls, thus all assessors were included in all ratings.

TABLE II. Results of McNemar's test for preferences before and after the attribute elicitation and rating processes. Statistics with both music excerpts aggregated together. D means the direction of the change, e.g., + indicates that the particular hall has been liked more in the second rating.

	All positions			Front stalls			Parterre			Balcony		
	χ^2	p	D	χ^2	p	D	χ^2	p	D	χ^2	P	D
VM	13.89	0.00	-	0.34	0.56		4.66	0.03	-	9.03	0.00	-
AC	17.25	0.00	-	13.92	0.00	-	1.84	0.17		0.00	1.00	
BK	1.10	0.29		3.13	0.08		4.97	0.03	+	2.56	0.11	
BP	4.01	0.04	-	7.85	0.01	-	0.00	1.00		0.00	1.00	
CP	21.83	0.00	+	11.29	0.00	+	2.72	0.10		2.33	0.13	
HM	12.83	0.00	+	23.60	0.00	+	0.00	1.00		0.00	1.00	

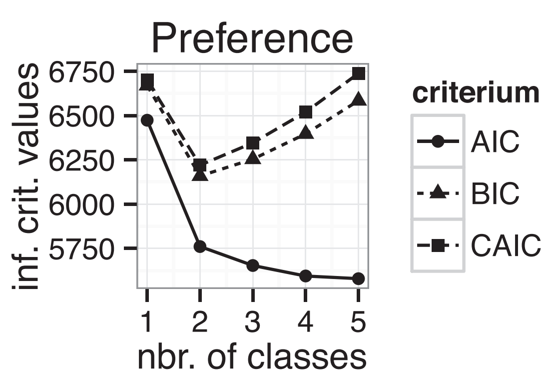

FIG. 5. Information criteria indicating the number of latent classes for preference data. AIC = Akaike information criterion; BIC = Bayesian information criterion; CAIC = consistent Akaike information criterion.

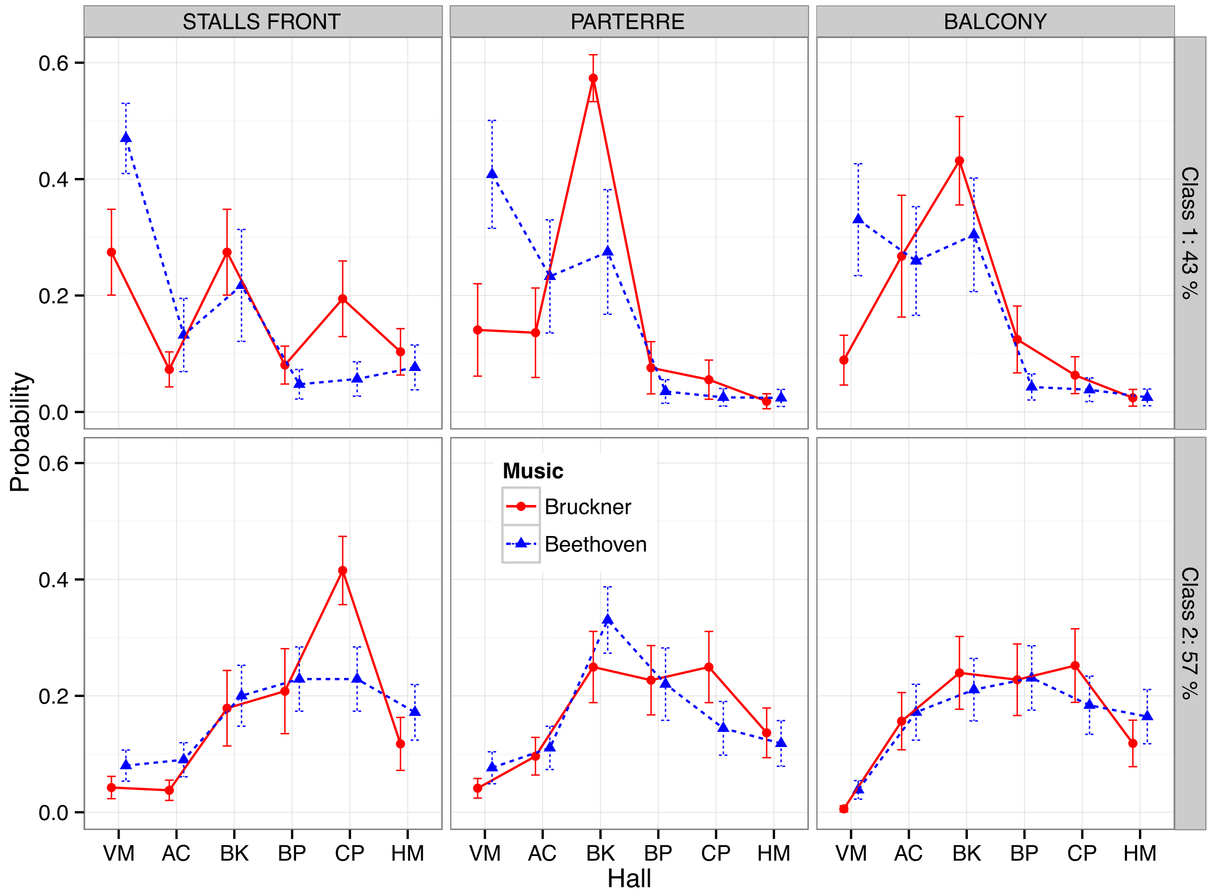

FIG. 6. (Color online) Preference groups for raw data after the segmentation of 28 assessors (both preference ratings aggregated) to latent classes (Ref. 22).

Previously[8,9] we have clustered individual attributes with hierarchical clustering. Here, only the binary data for each attribute is available, thus the attributes are grouped like the preference data using the latent class model.[22] However, as the music excerpts have different number of attributes, both data sets were segmented separately. Again, the decision for the number of classes was based on the information criteria, see Fig. 7, and we found three latent classes for both music. The full attribute list in each class is collected in Fig. 8.

First, the attributes elicited and used for Bruckner are analyzed. Figure 8 reveals that the main latent class consists of reverberance, width, bass, and loudness attributes. This does not necessarily mean that they are the same perceptual factors, but in this case the halls are rated in the same order with all these attributes. Thus, if a hall was perceived to be loud, it was also perceived reverberant with strong bass and wide sound. The BTL scores reveal that VM was clearly the dominant hall in this class, AC and BK being the other reverberant and wide halls with strong bass. The probability of choosing halls BP, CP, or HM the most reverberant and wide was close to zero at all listening positions. A second latent class for Bruckner is dominated by brightness attributes, in addition to many other attributes, such as proximity and sharpness. Clearly, this class consists of attributes that can be associated with the timbre of sound. It is also seen that at front stalls, where the direct sounds dominate the timbre differences are small, only AC and VM differentiating from others significantly. At parterre, BK has almost 0.5 probability and HM is significantly lower than other halls. At balcony the choice frequencies are similar, AC and BK having the highest probabilities and HM clearly the lowest. The last latent class consists of attributes related to clarity. Hall CP has significantly more clarity than the other halls, and VM and AC being the most reverberant halls have the lowest clarity. Interestingly, hall HM has low clarity at parterre and on balcony, despite the lack of reverberance.

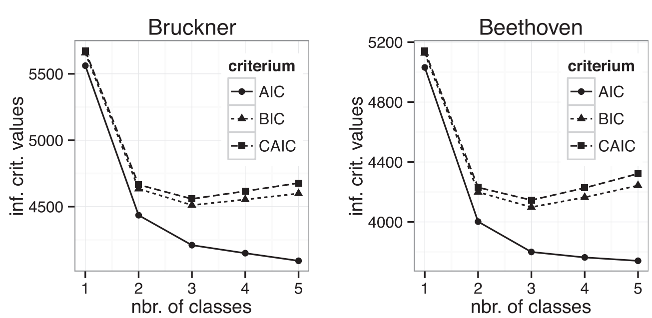

FIG. 7. Information criteria indicating the number of latent classes for attribute data. AIC = Akaike information criterion; BIC = Bayesian information criterion; CAIC = consistent Akaike information criterion.

The attributes elicited with Beethoven were also divided into three latent classes. The main class consists of attributes related to reverberance, loudness, and width, but hardly any

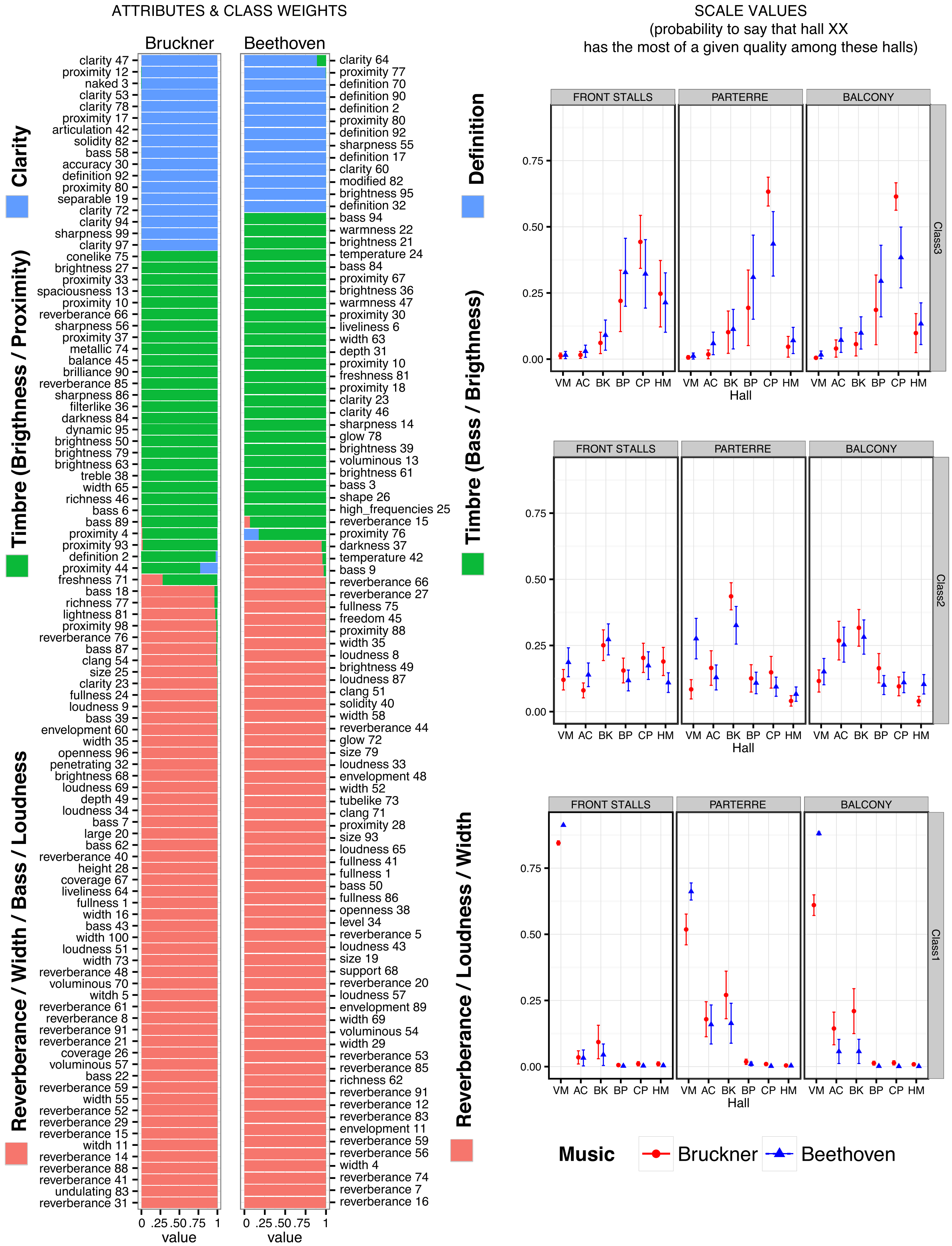

FIG. 8. (Color online) Left: All attributes (originally in Finnish, translated into English) elicited by individual assessors are segmented to different latent classes (Ref. 22). The name of each class is our interpretation of the attribute clusters. Weights represent the probability that a given attribute belongs to a class. Class weights are stacked and colored by class and they add up to one. Right: The parameter estimates of the BTL model are fitted to paired-comparison judgments for each attribute class.

bass attributes. The order of the halls is the same than for Bruckner; VM has a very high probability followed by AC and BK, whereas the rest of the halls have close to zero probability. The second latent class encompasses timbre attributes, but in addition to brightness and proximity, the class has several attributes related to bass and warmth. At front stalls only BK differs from the others, except VM. At parterre, VM and BK have the highest probabilities and on balcony AC and BK have the highest values. Finally, the third latent class is dominated with attribute definition and the order of the halls follows that obtained with Bruckner. However, hall CP does not differ significantly from BP.

C. Objective room acoustical parameters

From the measured impulse responses, we calculated the objective room acoustical parameters in each of the three listening positions with the ISO3382-1:2009 standard equations. It should be noted that all measurements were conducted in unoccupied halls. In addition, the measurement system does not fully comply with the standard, since the sources are active two-way loudspeakers, violating the omni-directionality requirement at high frequencies. The values in Fig. 9 are averages of all 24 source channels on stage. In addition to octave band data, Fig. 9 visualizes the mid and low frequency averages according to the standardized definition for single number values.

D. Correlations between preferences, attribute classes, and objective parameters

Unfolding the relations between perceptual attributes, subjective preferences, and physical factors in concert halls is one of the key objectives in room acoustic research. Our approach here is to search the possible correlations between collected preferences, attributes, and objective metrics by using vectors of 18 values (six halls in all three positions). The data were rank ordered separately in each position before concatenating them in one vector. The Kendall rank order correlation table is presented in Table III. That is, each value in the table corresponds to a correlation coefficient τ between two respective 18-element vectors.

Table III shows that first preference classes are correlated with first attribute classes (reverberance/width/loudness) for both music although for Bruckner the correlation is not significant. In contrast, the second latent preference classes have positive correlation with third attribute classes (clarity/definition). The correlations between attribute classes indicate significant correlations with classes 1 (reverberance/width/loudness) and classes 3 (clarity/definition), but not with classes 2 (timbre). This is in line with the BTL scores (Fig. 8), and could be explained by quite different individual attributes in both timbre classes.

The ISO3382-1:2009 standard includes suggested relationships between perceptual aspects and certain objective parameters at specific octave bands. Here, the correlation between EDT and LJ and the first attribute classes (reverberance/width/loudness/ bass) is high, thus in line with the standard. In addition, clarity and definition classes have high

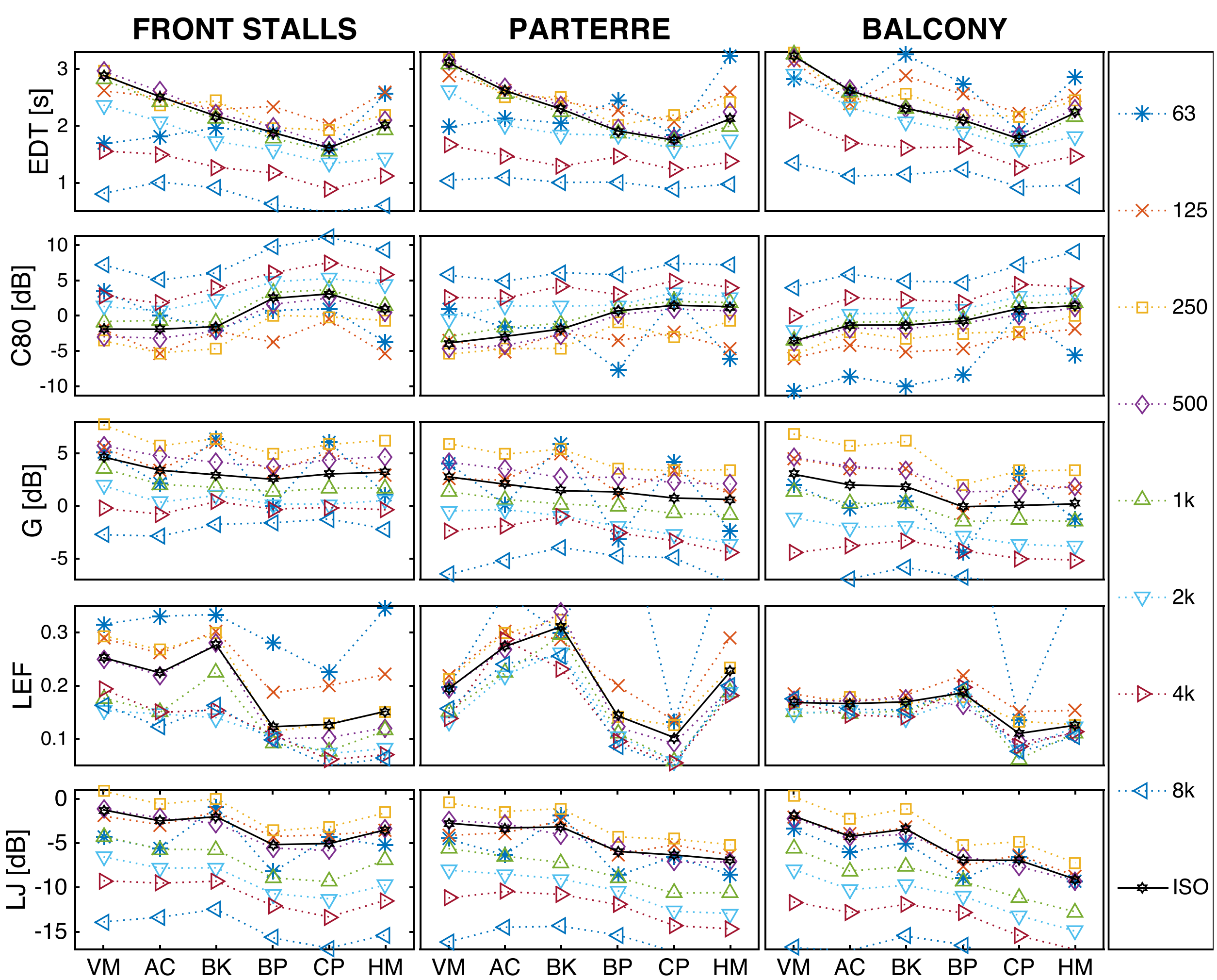

FIG. 9. (Color online) Objective parameters according to ISO3382-1:2009 standard. ISO means the average of 500 Hz and 1 kHz octave bands for EDT, C80, and G; and the average of 125 Hz–1 kHz octave bands for LEF and LJ, as defined in the ISO3382-1:2009 standard.

TABLE III. The Kendall rank correlations of preference ratings (all, see Fig. 4; C1–C2, see Fig. 6), attribute classes (see Fig. 8), and objective parameters (see Fig. 9). Significant correlations ($p < 0.05$) with Holm–Bonferroni correction are bolded.

	Preferences				Attribute Classes					
	Bruckner		Beethoven		Bruckner			Beethoven		
	Class1	Class2	Class1	Class2	R/W/B/L	Timbre	Clarity	R/L/W	Timbre	Definition
Pref_BR_C1										
Pref_BR_C2	0.08									
Pref_BE_C1	0.59	−0.35								
Pref_BE_C2	0.13	0.70	−0.30							
BR1_R/W/B/L	0.56	−0.25								
BR2_Timbre	0.67	0.34			0.27					
BR3_Clarity	−0.17	0.69			−0.47	0.13				
BE1_R/L/W			**0.91**	−0.36	**0.83**	0.21	−0.65			
BE2_Timbre			0.59	−0.08	0.70	0.41	−0.30	0.53		
BE3_Definition			−0.62	0.64	−0.56	0.04	**0.90**	−0.73	−0.33	
ISO_EDT	0.24	−0.70	0.64	−0.56	0.55	−0.04	**−0.92**	**0.75**	0.33	**−0.91**
ISO_C80	−0.33	0.59	−0.73	0.40	−0.65	−0.04	0.74	−0.73	−0.32	0.74
ISO_G	0.27	−0.52	0.65	−0.56	0.65	0.00	−0.61	0.65	0.43	−0.73
ISO_LEF	0.48	−0.21	0.48	−0.05	0.39	0.29	−0.33	0.58	0.37	−0.41
ISO_LJ	0.59	−0.35	**1.00**	−0.30	**0.91**	0.28	−0.56	**0.91**	0.59	−0.62
EDT_63	−0.05	−0.16	−0.05	0.02	−0.14	0.02	−0.07	0.05	−0.12	−0.07
EDT_125	0.13	−0.61	0.47	−0.44	0.37	−0.14	−0.62	0.57	0.06	−0.62
EDT_250	0.26	−0.59	0.65	−0.56	0.56	0.04	**−0.76**	0.74	0.36	−0.74
EDT_500	0.19	**−0.76**	0.59	−0.62	0.50	−0.10	**−0.88**	0.70	0.30	**−0.87**
EDT_1k	0.24	−0.70	0.64	−0.56	0.55	−0.04	**−0.92**	**0.75**	0.33	**−0.91**
EDT_2k	0.33	−0.56	0.69	−0.43	0.61	0.01	**−0.86**	**0.76**	0.39	**−0.78**
EDT_4k	0.20	−0.69	0.56	−0.47	0.48	−0.13	**−0.81**	0.64	0.24	**−0.75**
EDT_8k	0.20	−0.53	0.50	−0.25	0.41	0.03	−0.59	0.58	0.21	−0.53
C80_63	−0.08	0.18	−0.17	−0.08	−0.10	−0.19	0.13	−0.27	−0.03	0.15
C80_125	0.06	0.36	−0.16	0.15	−0.07	−0.01	0.31	−0.19	0.10	0.29
C80_250	−0.42	0.39	**−0.76**	0.30	−0.67	−0.22	0.48	−0.67	−0.40	0.55
C80_500	−0.37	0.56	**−0.77**	0.37	−0.69	−0.08	0.71	**−0.77**	−0.36	0.71
C80_1k	−0.44	0.53	**−0.75**	0.38	−0.66	−0.17	0.73	**−0.84**	−0.44	0.73
C80_2k	−0.39	0.53	**−0.77**	0.34	−0.69	−0.08	0.69	**−0.77**	−0.38	0.69
C80_4k	−0.12	0.64	−0.50	0.35	−0.41	0.08	0.62	−0.60	−0.09	0.62
C80_8k	−0.25	0.51	−0.58	0.25	−0.50	−0.07	0.53	−0.58	−0.21	0.52
G_63	0.31	0.10	0.36	−0.10	0.45	0.30	−0.10	0.28	0.50	−0.17
G_125	0.47	0.01	0.53	−0.15	0.61	0.34	−0.27	0.43	0.73	−0.27
G_250	0.51	−0.34	**0.76**	−0.38	**0.76**	0.22	−0.53	**0.81**	0.60	−0.64
G_500	0.22	−0.54	0.62	−0.59	0.62	−0.05	−0.58	0.62	0.39	−0.69
G_1k	0.37	−0.44	**0.81**	−0.52	**0.83**	0.10	−0.63	0.74	0.56	−0.70
G_2k	0.52	−0.38	**0.81**	−0.30	0.71	0.29	−0.57	0.73	0.44	−0.59
G_4k	**0.87**	0.19	0.44	0.27	0.46	0.73	−0.10	0.39	0.67	−0.11
G_8k	0.40	0.66	−0.07	0.73	−0.05	0.61	0.41	−0.13	0.20	0.41
LEF_63	0.10	−0.15	−0.05	0.14	−0.14	0.05	−0.02	0.05	−0.09	−0.05
LEF_125	0.33	−0.35	0.44	−0.19	0.36	0.18	−0.44	0.54	0.22	−0.52
LEF_250	0.51	−0.21	0.50	−0.10	0.43	0.30	−0.38	0.53	0.41	−0.45
LEF_500	0.57	−0.21	0.57	−0.11	0.48	0.36	−0.44	0.67	0.56	−0.53
LEF_1k	0.44	−0.24	0.44	−0.03	0.34	0.25	−0.38	0.54	0.31	−0.37
LEF_2k	0.29	−0.35	0.37	−0.12	0.27	0.07	−0.54	0.46	0.21	−0.47
LEF_4k	0.28	−0.38	0.42	−0.16	0.33	0.10	−0.56	0.51	0.21	−0.49
LEF_8k	0.33	−0.33	0.44	−0.12	0.35	0.12	−0.53	0.53	0.25	−0.46
LJ_63	0.67	−0.09	0.67	−0.17	0.73	0.48	−0.30	0.60	0.67	−0.41
LJ_125	0.55	−0.17	**0.77**	−0.26	**0.77**	0.40	−0.33	0.67	0.65	−0.41
LJ_250	0.53	−0.31	**0.96**	−0.35	**0.96**	0.24	−0.51	**0.86**	0.63	−0.59
LJ_500	0.42	−0.52	**0.82**	−0.41	0.74	0.10	**−0.75**	**0.82**	0.41	**−0.75**
LJ_1k	0.42	−0.52	**0.82**	−0.41	0.74	0.10	**−0.75**	**0.82**	0.41	**−0.75**
LJ_2k	0.52	−0.43	**0.91**	−0.35	**0.81**	0.21	−0.65	**0.82**	0.50	−0.65
LJ_4k	0.53	−0.34	**0.82**	−0.24	0.73	0.30	−0.55	0.73	0.50	−0.55
LJ_8k	0.56	−0.09	0.59	0.08	0.50	0.43	−0.23	0.50	0.44	−0.21

correlations with C80 although the correlations are not significant. At the same time, EDT has significant negative correlation to these classes. Finally, none of the ISO parameters has significant correlation with either of the timbre classes.

The ranks of objective parameters at octave bands are correlated with both preference and attribute classes. Preference class 1 for Bruckner correlates significantly only with G at 4 kHz octave band. In contrast, for Beethoven there are a high number of significantly correlating parameters, namely, G, LJ, and C80 (negative correlation). Second preference classes with both music excerpts correlate negatively with mid frequency EDT and positively with G at 8 kHz.

The first attribute classes (reverberance/width/loudness) are well explained with LJ at all bands and G at mid frequencies. In addition, for Beethoven the first attribute class has a significant negative correlation with mid frequency C80. In contrast, clarity and definition classes have positive correlations with mid frequency C80 and significant negative correlations with mid frequency EDT and LJ. Timbre class for Brucker was dominated by brightness attributes and this is reflected as high correlations with G at 4 and 8 kHz octave bands. For Beethoven, the timbre class also has warmth and bass attributes, again seen as high correlations for low frequency LJ and G. Finally, an interesting fact is that LEF does not have a single significant correlation with any of the preference or attribute classes.

V. DISCUSSION

The found preference classes are not surprising, as they correspond to the findings of earlier studies.[9,26] However, the current results present the findings in the context of some of the best known concert halls. The assessors who like reverberance, width, bass, and loudness preferred VM, BK, and AC, in particular, BK with Bruckner and VM with Beethoven. Other assessors appreciating clear and defined sound preferred CP, BP, and BK with both music excerpts, but the favorite hall depends on the listening position and music. Moreover, for this preference class unoccupied VM and AC were too reverberant. All in all, the preference results showed the existence of large individual differences.

The preference ratings were collected from all assessors with two music excerpts. This approach enables us to study in detail the influence of music on preference. McNemar's test enlightens the consequences of the change of music from Beethoven to Bruckner for individual preferences. The results are shown in Table IV. The combined results of all positions show that assessors rated three halls more suitable for Beethoven (VM, AC, HM), while the other halls (BK, BP, CP) were ranked higher with Bruckner. Moreover, the overall preference results (Fig. 4) suggest that when the orchestra was playing louder (Bruckner) the overall preference is towards halls that have more clarity, but with more tranquil music (Beethoven) more reverberance is needed for preferred acoustics.

The number of latent attribute classes is small in comparison with all previous studies. The literature of concert hall acoustics clearly states that the fundamental perceptual characteristics and differences between concert halls can be described

TABLE IV. Results of McNemar's test showing the difference (D) in preferences between Beethoven and Bruckner aggregated over both preference tests. Positive (+) difference denotes higher preference for Bruckner than for Beethoven.

	All positions			Stalls front			Parterre			Balcony		
	χ^2	p	D	χ^2	p	D	χ^2	p	D	χ^2	p	D
VM	126.45	0.00	–	13.29	0.00	–	28.93	0.00	–	52.74	0.00	–
AC	39.69	0.00	–	32.66	0.00	–	6.57	0.01	–	0.00	1.00	
BK	14.01	0.00	+	0.45	0.50		0.57	0.45		5.22	0.02	+
BP	33.64	0.00	+	2.75	0.10		9.26	0.00	+	14.79	0.00	+
CP	103.39	0.00	+	52.17	0.00	+	26.23	0.00	+	12.02	0.00	+
HM	6.51	0.01	–	0.14	0.71		0.02	0.88		3.69	0.06	

in terms of reverberance, loudness, spaciousness (width and envelopment), definition or clarity, brightness, proximity/intimacy, and bass (often also called warmth). In this study, all these aspects are listed in the attributes. The reason for only three latent classes here is that within the six selected halls the order is the same with several perceptual factors. For example, if a hall was found reverberant it was also perceived wide and loud, resulting that all these attributes form only one latent class. However, our interpretation is that when an assessors pay attention to, e.g., width, he can separate it from reverberance, i.e., these attributes are indeed different perceptual aspects although they are in the same latent class. On one hand, the selection of six halls led to such a result. On the other hand, the result suggests that some perceptual aspects could be strongly linked together; if one changes, also some other feature changes. Individual perceptual aspects could be controlled to some extent by artificially created reverberation, but here the acoustics of real concert halls were used without any control to individual perceptual characteristics.

The attribute elicitation process was done twice by all assessors, independently with both music excerpts. Therefore, the list of attributes is influenced by musical context. For example, an interesting semantic difference emerged as the assessors more often differentiated Bruckner samples with the word clarity instead of definition, which was used primarily with Beethoven. Another semantic difference was that some assessors used the attribute warmness instead of bass for Beethoven, but for Bruckner none of the attributes are related to the temperature. In Beethoven excerpt, the only low frequency instruments were the double basses playing in *pizzicato* that possibly led assessors use the attribute warmness.

The halls were all measured unoccupied resulting in more reverberant conditions than *in situ* at occupied conditions, which is often the reality in this renowned halls. Although the assessors described the sound samples being very natural and realistic, it is clear that unoccupied conditions might introduce a bias to the presented results, as the difference between occupied and unoccupied conditions varies between halls. Another issue, which might bias the results is that we do not know what the internal reference was for each assessor. In the laboratory the assessors are not aware of the context of a large space (no visual or other cues) that they would have *in situ*. Therefore, it might be that the mental reference for assessors is classical music recordings, which always have high clarity and less reverberation than in reality *in situ*, or in our auralizations.

VI. CONCLUSIONS

This study compared six different well-known unoccupied concert halls on three different positions with two music excerpts. Twenty eight assessors rated the halls according to their preference with the forced-choice paired comparison method. Moreover, the assessors elicited individual attributes to describe perceived differences between halls and rated the halls with their own attributes.

The main results show that listeners can be categorized into two different preference classes. Some listeners prefer clarity over reverberance and the others love strong, reverberant and wide sound. Interestingly, "late Romantic symphonic music" (Bruckner) is appreciated more in less reverberant halls while "Classical period symphonic music" (Beethoven) without dominating brass instruments benefits louder and more reverberant halls. In addition, the preference order of the halls is quite different in different positions in the studied halls and timbre of the sound is an important factor for preference. The individually collected attributes were also different depending on the choice of music although the resulting attribute classes were quite similar with both music.

The results with the six included halls agree with the traditional claims[27] where the rectangular halls with flat floor render music loud, reverberant, and wide with strong bass, while the halls with deeply inclined auditorium are less reverberant with high clarity and definition. However, the studied subset of halls does not encompass all conceived concert hall designs, which may provide other combinations of perceptual room acoustic factors.

Finally, this article incorporated the statistical analysis methods of paired comparison data sets. Both the preferences and the individual attribute data were segmented to latent classes that are reasonably easy to interpret. Such methods are valuable in sensory sciences as they help to disentangle the preference ratings and find fundamental reasons to individual preferences.

ACKNOWLEDGMENTS

This research has received funding from the Academy of Finland, project No. [257099].

[1]W. C. Sabine, "Reverberation: Introduction," reprinted in "Collected papers on Acoustics, No. 1" (Harvard University Press, Cambridge, 1923).
[2]L. Beranek, *Music, Acoustics, and Architecture* (Wiley, New York, 1962), 586 pp.
[3]L. Cremer and H. Müller, *Principles and Applications of Room Acoustics* (Applied Science Publishers, London, England, 1982), Chap. III.3, Vol. 1, pp. 592–605.
[4]M. Barron, "The subjective effects of first reflections in concert halls—the need for lateral reflections," J. Sound Vib. **15**(4), 475–494 (1971).
[5]M. R. Schroeder, G. Gottlob, and K. F. Siebrasse, "Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustics parameters," J. Acoust. Soc. Am. **56**(4), 1195–1201 (1974).
[6]G. A. Soulodre and J. S. Bradley, "Subjective evaluation of new room acoustic measures," J. Acoust. Soc. Am. **98**(1), 294–301 (1995).
[7]R. J. Hawkes and H. Douglas, "Subjective acoustics experience in concert auditoria," Acustica **24**, 235–250 (1971).
[8]T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," J. Acoust. Soc. Am. **130**(2), 835–849 (2011).
[9]T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," J. Acoust. Soc. Am. **132**(5), 3148–3161 (2012).
[10]A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki, "Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls," J. Acoust. Soc. Am. **135**(1), 239–250 (2014).
[11]J. Pätynen, "A virtual symphony orchestra for studies on concert hall acoustics," Ph.D. thesis, Aalto University School of Science, 2011.
[12]S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference," J. Acoust. Soc. Am. **121**(1), 388–400 (2007).
[13]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19–22, 2000, preprint No. 5093.
[14]S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**(1/2),16–27 (2013).
[15]J. Pätynen, S. Tervo, and T. Lokki, "Amplitude panning decreases spectral brightness with concert hall auralizations," in *Proc. 55th Audio Eng. Soc. Conference*, Aug. 27–29 2014, Helsinki, Finland (Audio Engineering Society, New York, 2014), Paper No. 49.
[16]F. Rumsey, S. Zielinski, R. Kassier, and S. Bech, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," J. Acoust. Soc. Am. **118**(2), 968–976 (2005).
[17]J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," Acta Acust. Acust. **94**(6), 856–865 (2008).
[18]H. A. David, *The Method of Paired Comparisons* (Griffin, London, 1963), Vol. 12, 124 pp.
[19]R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," Biometrika **39**, 324–345 (1952).
[20]R. D. Luce, *Individual Choice Behaviour: A Theoretical Analysis* (Wiley, New York, 1959).
[21]M. Semenou, Ph. Courcoux, A. Kuusinen, and T. Lokki, "Segmentation of subjects in multivariate paired comparisons—Application to the preference for concert halls acoustics," Food Qual. Preference (unpublished).
[22]P. Courcoux and M. Semenou, "Preference data analysis using a paired comparison model," Food Qual. Preference **8**(5), 353–358 (1997).
[23]A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. B **39**, 1–38 (1977); available at http://www.jstor.org/stable/2984875.
[24]M. Semenou, CompR: Paired Comparison Data Analysis, 2015, R package version 1.0.
[25]Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," Psychometrika **12**(2), 153–157 (1947).
[26]M. Barron, "Subjective study of British symphony concert halls," Acustica **66**, 1–14 (1988).
[27]L. Beranek, "Subjective rank-orderings and acoustical measurements for fifty-eight concert halls," Acta Acust. Acust. **89**, 494–508 (2003).