Kaplanis, Neofytos; Bech, Søren; Lokki, Tapio; van Waterschoot, Toon; Holdt Jensen, Sören

Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction

Published in:
Journal of the Acoustical Society of America

DOI:
10.1121/1.5135582

Published: 01/11/2019

Document Version
Publisher's PDF, also known as Version of record

Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction

Neofytos Kaplanis, Søren Bech, Tapio Lokki, Toon van Waterschoot, and Søren Holdt Jensen

Citation: The Journal of the Acoustical Society of America **146**, 3562 (2019); doi: 10.1121/1.5135582
View online: https://doi.org/10.1121/1.5135582
View Table of Contents: https://asa.scitation.org/toc/jas/146/5
Published by the Acoustical Society of America

Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction

Neofytos Kaplanis[a)] and Søren Bech
Bang & Olufsen a/s, Bang & Olufsen Allé 1, Struer, DK-7600, Denmark

Tapio Lokki
Department of Computer Science, Aalto University, P.O. Box 13000, FI-00076 Aalto, Finland

Toon van Waterschoot
Department of Electrical Engineering (ESAT-STADIUS/ETC), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

Søren Holdt Jensen
Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

(Received 3 June 2019; revised 1 October 2019; accepted 6 November 2019; published online 27 November 2019)

An experiment was conducted to identify the perceptual effects of acoustical properties of domestic listening environments, in a stereophonic reproduction scenario. Nine sound fields, originating from four rooms, were captured and spatially reproduced over a three-dimensional loudspeaker array. A panel of ten expert assessors identified and quantified the perceived differences of those sound fields using their own perceptual attributes. A multivariate analysis revealed two principal dimensions that could summarize the sound fields of this investigation. Four perceptual constructs seem to characterize the sensory properties of these dimensions, relating to *Reverberance*, *Width & Envelopment*, *Proximity*, and *Bass*. Overall, the results signify the importance of reverberation in residential listening environments on the perceived sensory experience, and as a consequence, the assessors' preferences towards certain decay times.
https://doi.org/10.1121/1.5135582

[BFGK] Pages: 3562–3576

I. INTRODUCTION

Reverberation has been used as an umbrella-term to conjointly describe the acoustical properties of a space and the related effects on the perceived aural experience. The dominant and perceptually pleasing role that reverberation exhibits in performance spaces, such as concert halls and auditoria, has naturally steered the primary focus of architectural acoustics research. A plethora of scientific studies[1] sought to better understand reverberation in those spaces, focusing on its key physical properties as well as decomposing its multidimensional perceptual character. In contrast, the effects of reverberation in geometrically smaller spaces, which we experience most often, e.g., residential and ordinary-sized listening rooms, are still not well understood.

The importance that reverberation holds within domestic listening spaces has been highlighted in recent years, as new spatial audio technologies and multi-loudspeaker rendering protocols have been challenged by its effects. Hughes *et al.*[2] indicated that the reverberation of the reproduction room had a major influence on the accurate realization of spatial audio rendering when attempting to recreate divergent kinds of soundscapes in an ordinary listening room. Grosse *et al.*[3] expanded those results and argued that similar behavior occurs in sound field synthesis schemes, i.e., *higher order ambisonics* (HOA) and *wave-field synthesis* (WFS). These results indicate that the perceived aural impression of the reproduced sound scene is distorted, as the recorded signal is superimposed with the spatiotemporal response of the reproduction room. Those effects limit the performance of the latest spatial audio technologies in domestic settings, and they contradict with the primary aim of a sound reproduction system: to deliver a faithful representation of the recorded sound field (creator's intent) to a reproduction space.

The knowledge that currently pertains to sound reproduction is such spaces as complete sound fields are limited. Investigations in this domain have primarily focused on the interaction of single reflections and loudspeakers' properties in simulated[4] or real[5] sound fields. It is still unknown which elements of the reverberant field in domestic listening environments evoke certain sensations and how these relationships operate. This limits our ability to address the above degradation and restricts the further domestication of advanced spatial audio.

Understanding the effects of reverberation in typical residential spaces may enable new reproduction systems to introduce spatial and temporal compensation on the perceived effects, allowing a perceptually accurate spatial audio reproduction, even in domestic listening environments.

Our research interest focuses on the imposed effects of reverberation in domestic enclosures in a sound reproduction scenario. In this initial investigation, we attempt to decompose the fundamental perceptual characteristics of those

a)Electronic mail: neo@bang-olufsen.dk

sound fields and understand more deeply the effects of reverberation in domestic listening environments. Our aims were: (i) investigate the extent to which the properties of reverberant sound field in residential-sized rooms affect the perceived sensory experience of a reproduced sound field, (ii) identify the major perceptual attributes underlying these properties and the relationships between them, and (iii) examine possible influences of the physical and sensory characteristics of these fields on assessors' preferences.

In Sec. I A, the background and the rationale behind the experimental design is discussed. Section II presents the experimental methods, including the experimental conditions, the apparatus, and the evaluation procedures followed. The statistical analysis of the experimental results is then described in Sec. III. The findings are further discussed in Sec. IV along with the limitations of the study. Finally, the study's major findings are summarized in Sec. V.

A. Background and motivation

Over the last decades, the effects of reverberation in performance halls have been well studied, revealing numerous perceptual attributes that describe the properties of such sound fields. Those findings have been instrumental for architects and designers of concert halls, and form the anticipated characteristics for enhancing the sound of an orchestra or an act.[6]

Naturally, it has been speculated that the prosperous research on reverberation in halls and auditoria could be utilized for smaller spaces, such as domestic-sized rooms. A recent review[1] on the majority of published literature on reverberation evaluated this hypothesis. It was shown that the current literature in psychological and acoustical research suggests that the direct link between findings in one domain, e.g., in performance halls, cannot be assumed for another, e.g., in ordinary-sized listening rooms. The divergence between the purpose-made performance spaces and residential rooms, has significant influence both on the physical characterization of the fields, i.e., the standard objective metrics, and the perceived sensations of human listeners.

In architectural acoustics, the resultant sound field is known to be dependent on (i) the features of the emitted sound signal, (ii) the properties of the excitation source, and (iii) the physical characteristics of the enclosure.

In domestic listening rooms the restricted physical dimensions of the enclosure, compared to the larger performance halls, result in dissimilar reflection patterns. The temporal and spatial distribution of the reflections within the field seem to affect our auditory processing schemata and the weighting of perceptual cues;[1] for example, reducing the influence of the overall loudness and *direct to reverb ratio* (DRR) as a cue for estimating the perceived room size in small rooms compared to large spaces.[7]

Moreover, the excitation source is typically a loudspeaker, encompassing its own properties, which interact heavily with the room. A wealth of scientific investigations in sound reproduction have shown that the interactions between loudspeakers characteristics and the boundaries influence both the timbral and spatial characteristics of the perceived listening experience,[5] as well as listeners' preference.[8] In an attempt to better understand the perceptual effects of the temporal integration of distinct early reflections in such reproduction scenarios, later studies identified the audibility thresholds[9] and perceptual relevance of distinct, early reflections[4] on the perceived qualities of loudspeakers.

Those studies have highlighted the implications of reproduced sound in domestic-sized listening environments and formed the basis of subsequent standards, such as ITU-R BS.1116[10] and IEC 60268–13.[11] Aiming to provide an invariable reference for critical evaluation tasks, those standards define several constraints on the physical properties of the reproduction rooms. Nonetheless, the effects of complete sound fields on the perceived experience are still unknown, as much as the ways in which they relate to the fields' physical properties.

Identifying the perceptual attributes that characterize the experience within those acoustical fields, and understanding their underlying relationships to the physical characteristics of the sound field may encourage the development of perceptually-relevant compensation algorithms capable of controlling both the spatial and the temporal response of the perceived sound fields. This could improve the capacity of reproduction systems to recreate divergent kinds of sound fields and soundscapes in domestic environments, and enable a more faithful reproduction of the recorded signals.

II. METHODS

The common approach in perceptual assessment of architectural acoustics is based on the identification of co-occurring characteristics of reverberant fields, by comparing the features of real spaces.[1] That is, the perceived sensations of complete sound fields are directly compared to each other, so that common trends and dissimilarities between the two are identified, and conjointly quantify both the perceptual and physical aspects of the fields. Naturally, such evaluation follows *in situ* protocols, requiring long time intervals between comparisons. The limited auditory memory that humans exhibit[12] and the low level multi-modal processes associated with *room adaptation*,[13] where the assessors' psychological state and perceptual sensitivity is modified,[1] have been found to alter assessors' judgments in auditory tasks[14] limiting the robustness and generalization of such investigations.

Recently, it became possible to conduct such evaluation schemes in a controlled laboratory setting by capturing and reproducing the sound fields using spatial audio rendering schemes.[15] Those schemes enable controlled, blind, and instant comparisons between diverse acoustical conditions, which are required when scientific experimental frameworks are to be followed. Nonetheless, human assessors are still exposed to complete sound fields, which are, by nature, highly complex stimuli that pose a limitation for standard audio evaluation schemes that are designed to assess specific signal properties. Attempting to overcome the limitations that an evaluation of complex stimuli encompasses, *sensory analysis* (SA)[16] techniques originating from food science have been applied to several domains of audio assessment with promising results, including spatial audio reproduction on loudspeakers[17] and headphones,[18] concert hall acoustics,[19] and automotive audio.[20,21]

In this work, the evaluation of the sound fields was conducted in a laboratory. Human assessors were presented with the auralized three-dimensional (3D) sound fields, based on *in situ* measurements in real enclosures. As common SA techniques tend to be time consuming, this study uses a rapid method, namely, the *flash profile* (FP).[22] FP is believed to be suitable for initial studies on room acoustic properties. It consists of both attribute elicitation and quantification tasks, and the obtained perceptual data can be statistically combined to objective metrics. In previous work, the rationale[23] and the practical implementation[21] of the experimental methodology was discussed in detail. In this section, the applied methodology is presented briefly.

A. Experimental conditions

Four listening rooms were used in the experiment aiming to create a range of possible acoustical scenarios of a domestic environment. All four rooms, labeled as Rooms A–D, comply with the size and structural requirements of IEC:60268-13[11] that depicts the "average" residential listening environment, such as a typical living room. The structural and acoustical characteristics details are given in Table I.

Rooms A, B, and D, comprise of wooden floors and ceiling, and acoustically treated rigid boundaries.[24] They all include ordinary furniture and fully comply with IEC:60268-13[11] specifications. Room C is a critical listening space[10] built to host multichannel reproduction layouts.[25] It is characterized by larger volumetric size, highly absorptive boundaries, no furniture, and a lower *reverberation time* (RT) than a typical room.

In order to capture a wider range of possible acoustical fields than Rooms A–D, the interior of Room D was physically modified. Room D has a modular structure with adaptable acoustic panels which allowed the successive modification of the boundaries' properties, such as the addition/removal of absorptive, diffusive, or reflective acoustic panels in all boundaries, including the ceiling. The modifications varied the RT_{30} as uniformly as possible across frequency. Modifications in the lateral plane were completed in symmetry. The reproduction system was fixed and the direct paths between the source, ceiling, floor, side walls, and the receiver were not modified. This ensured constant contribution of the first reflection points in these measurements, as shown in Table I.

Aiming to primarily focus on the perceived effects of reverberation in a complex acoustical scenario, certain steps were required to minimize known influencing factors that affect the sound field. It is well known that the loudspeaker's characteristics, e.g., its size, directivity, and placement, interact with the acoustical properties of the reproduction room and modify the aural experience.[5] Here, the reproduction system was identical and it was positioned at a common relative point in all rooms. The ratio between the loudspeaker position and the boundaries was common across rooms.[26] This process was followed to neutralize the effect of the rooms' dissimilar modal regions as much as possible,[5] and maintain the contribution of first order reflections, to avoid timbral alterations by comb filtering. Figure 1 depicts the overlaid diagrams of the setup during the acquisition of sound fields in all rooms.

B. Acquisition of room impulse responses and analysis

In situ measurements were conducted to obtain spatial *room impulse responses* (RIR). The excitation sources, two full-range loudspeakers (Genelec 1031 A with 7041 A subwoofer), were placed in a 2-channel stereophonic configuration.[11,25] The sources were level-matched individually and the total system was calibrated at the listening position, at 82 dB (C-weighted) sound pressure level, using 10 s of pink-noise at −37 dBFS.

The acoustical field was captured by a 3D vector intensity probe (50VI-1, G.R.A.S, Denmark), comprising of two capsules on each axis, separated by 25 mm. The RIR of each loudspeaker was measured separately at the reference listening position, using a 5 s logarithmic sine sweep[27] at 192 kHz sampling rate using RME UFX+ soundcard.

The captured RIR were analyzed using *spatial decomposition method* (SDM).[28] SDM provides a spatiotemporal analysis, where the instantaneous direction of each discrete sample is estimated using time difference of arrivals in short time windows. The estimated directions are then combined

TABLE I. Physical properties of the acoustical conditions used in the experiment. Acoustical parameters were calculated based on ISO:3382 (Refs. 31 and 32). Side, Floor, Ceiling, indicate the difference between direct sound and the corresponding first reflection point.

Condition	V(m^3)	RT30 $_{0.5-1k}$ (s)	$EDT_{0.5-1k}$(s)	DRR[a]	TS(s)	C50(dB)	Side (dB)	Floor (dB)	Ceiling (dB)	Compliance
Room A	80	0.32	0.34	−7.49	0.28	0.55	−9	−6	−5	Fully (11)
Room B	62.4	0.36	0.37	−8.90	0.24	1.22	−12	−8	−6	Fully (11)
Room C	172	0.17	0.11	3.64	0.05	21.90	−25	−7	−14	Fully (10) & Partial (11)[b]
Room D-1	90	0.30	0.29	−4.33	0.17	5.85	−9	−8	−10	Fully (11)
Room D-2	90	0.43	0.39	−6.22	0.24	3.09	−9	−8	−10	Fully (11)
Room D-3	90	0.50	0.46	−8.92	0.33	−0.37	−9	−8	−10	Fully (11)
Room D-4	90	0.65	0.56	−9.79	0.38	−1.49	−9	−8	−10	RT $_{500Hz}$ > (11)
Room D-5	90	0.67	0.61	−12.25	0.45	−2.66	−9	−8	−10	RT $_{250-4k Hz}$ > (11)
Room D-6	90	0.83	0.72	−13.05	0.53	−3.99	−9	−7	−9	RT $_{63-8k Hz}$ > (11)

[a]DRR.

[b]Size and geometry only.

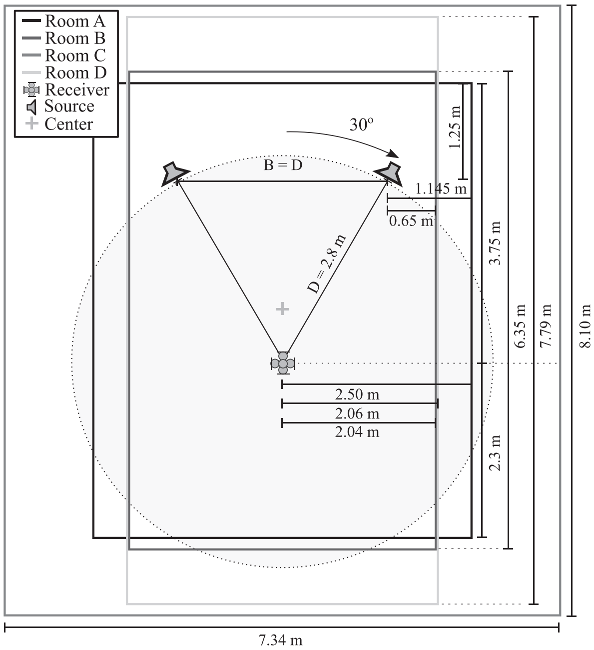


FIG. 1. Experimental setup used in the measurements, including the dimensions of the four rooms used in this study, and the relative position of the reproduction system. The overlayed diagrams are in scale.

with the pressure signal and they are distributed to a set of discrete reproduction points, i.e., a loudspeaker layout, using a parametric rendering scheme. In this study, lower frequencies (<75 Hz) were analyzed and synthesized using a recent development of SDM[29] where the analysis occurs in frequency bands and adaptable window sizes, providing more accurate results at that frequency range. Finally, the nearest neighbor[30] scheme was applied for a purpose-designed symmetrical array of forty loudspeakers positioned in a spherical orientation. The sound field acquisition, analysis, and reproduction schemes are graphically summarized in Fig. 2.

C. Room acoustic parameters

During the acquisition of the RIR described above, an additional set of measurements was performed for the calculation of standard acoustical parameters.[31,32] The acoustical energy was emitted by an *OmniSource* 4295 (Brüel & Kjær) and a subwoofer YST-W300 (Yamaha) at three positions and captured by four arbitrarily placed microphones (1/2 in. 40AZ, G.R.A.S., Denmark), as recommended for precision measurements. The acquisition of the RIR was achieved by the sine-sweep method[27] sampled at 48 kHz. The estimation of RT_{30} was based on Schroeder's backwards integration method[32] on the RIR, averaged across ten repeats for each source-receiver combination to reduce the noise floor at lower frequencies.[33]

D. Stimuli

Nine acoustical conditions were used in this listening experiment: the unmodified sound fields of Rooms A, B, and C, and a set of six acoustical modifications of Room D. They are labeled from Rooms D-1 to D-6, ranging from the lowest measured RT_{30} to the highest, respectively. This set provides

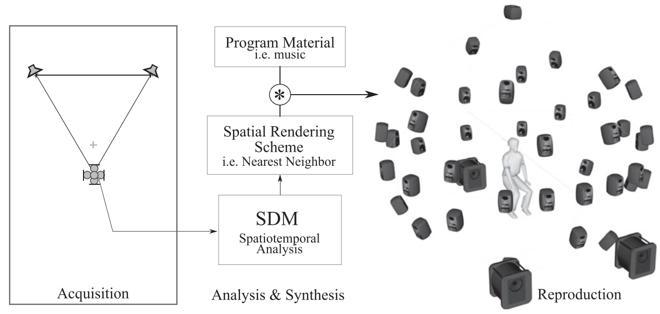


FIG. 2. A schematic describing the processes followed during acquisition, analysis, synthesis, and presentation of the sound fields. Loudspeakers are placed in symmetry at Azimuth and Elevation, respectively: [0,−20],[0,0],[15,0], [0,90], [15,20], [30,−35], [30,0], [30,45], [50,−10], [50,10], [70,−20], [70,0], [70,25], [90,0], [110,−15], [110,10], [130,0], [130,30], [150,−10], [150,20], [180,−10], [180,10], [180,40], [−150,−10], [−150,20], [−130,0], [−130,30], [−110,−15], [−110,10], [−90,0], [−70,−20], [−70,0], [−70,25], [−50,−10], [−50,10], [−30,−35], [−30,0], [−30,45], [−15,0], [−15,20], where 0,0 indicates the forward-facing median

a range of possible RT of an average living room as described in IEC:60268–13,[11] including the extreme cases of Rooms C and D-6. The resulting sound fields formed the nine levels of the first *independent variable* (IV), the acoustical conditions. The details of the acoustical conditions are summarized in Table I. The measured sound pressure at the listening positions is shown in Fig. 3)(a) and the calculated RT_{30} in Fig. 4.

To reproduce the acoustical conditions in the laboratory, the analyzed spatial RIR were convolved with the three program materials, loudness-matched before convolution at -15 dB_{LUFS}.[34] The program materials included: (1) *Shola Ama—You might need somebody, 0:09–0:25*, (2) *Anechoic Kongas African Rhythms—B&O Music For Archimedes, 0:24–0:33*, (3) *Female Speech English—EBU SQAM, 0:00–0:15*. Their spectral properties are shown in Fig. 3(b). These three excerpts were selected based on the relevance given for the experimental scenario, their dissimilar spectral, dynamic, and spatial properties, and their ability to

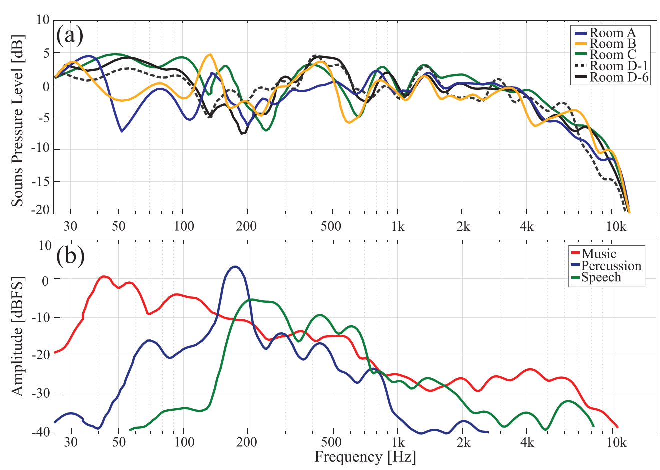


FIG. 3. The magnitude response across frequency for: (a) the pressure at the listening position, and (b) the program types used in the experiment. Levels were normalized and 1/3 octave averaging was applied for visualization purposes.

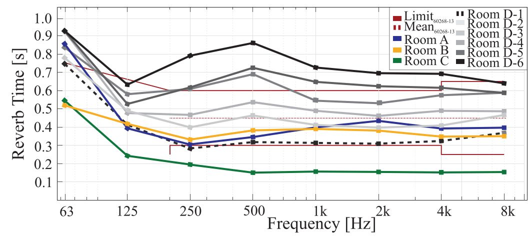


FIG. 4. The reverberation time (RT_{30}) of the sound fields used as the experimental conditions in the study, calculated in octave bands. The recommended tolerances of IEC60268-13 (Ref. 11) are shown in red.

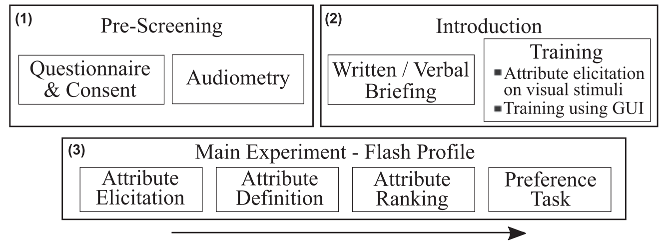


FIG. 5. Experimental procedure. Arrow depicts the timeline.

complementary excite the acoustical conditions under investigation, as typically recommended for audio evaluations.[35] These excerpts formed the three levels of program and are further referred to as music, percussion, and speech, respectively.

E. Experimental apparatus

Forty full-range loudspeakers (Genelec 8020 C) and three subwoofers (Genelec 7050B) were placed in a spherical orientation of 1.55 m radius, in the anechoic chamber at Aalborg University, Denmark. The physical placement of the reproduction loudspeakers was based on a spatiotemporal analysis[36] of the RIR captured in four standard listening rooms. The placement ensured that the spatial energy distribution between the reproduced and the original was minimized, according to the human localization acuity.[37] The reproduction of direct and first early reflections were maintained.

The reproduction array is depicted in Fig. 2. The tolerance between the individual loudspeakers, *in situ*, was ± 1.5 dB at 25–16k Hz. All 43 loudspeakers were temporally matched at the listening position electronically using delays to account for any inconsistencies of their physical positions.

To avoid possible visual biases, the assessors were guided in the anechoic chamber by the experimenter under dark conditions. The experimental setup included an acoustic curtain that separated physically and visually the assessors and the experimental apparatus. During the experiment, the reproduction level was set to $75 \pm 0.5\ \mathrm{dB}_{\mathrm{LAeq(15s)}}$ at the listening position across all stimuli.

F. Experimental procedure

The experiment was completed in a single session and comprised of three phases: pre-screening, introduction, and the main experiment. The experimental procedure is summarized in Fig. 5.

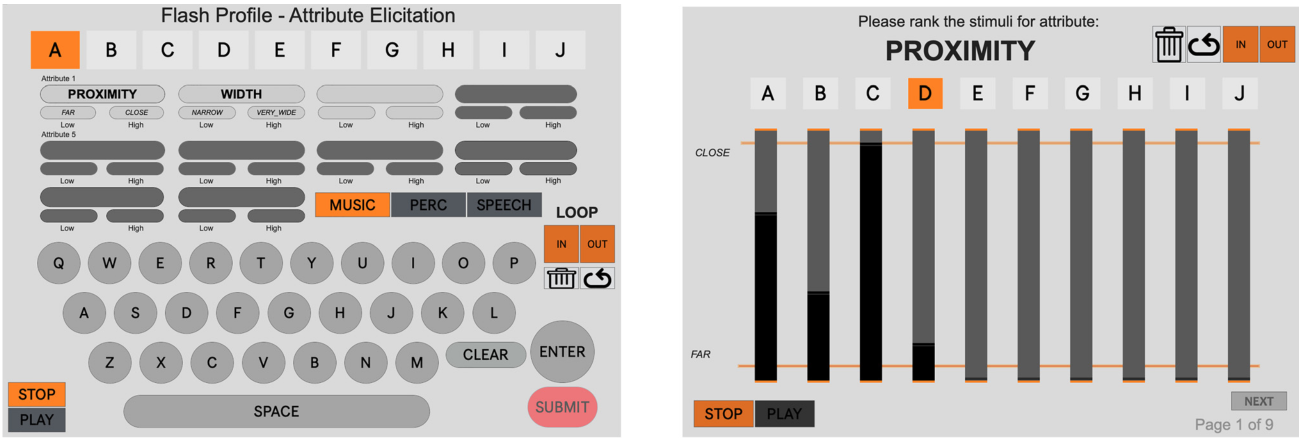
First, the assessors completed a questionnaire where background information was collected and consent was given. During pre-screening, the assessors' hearing sensitivity was tested.[38] They were then introduced to the experimental methodology, including the principles behind FP, in verbal and written form. As eliciting attributes requires assessors to be able to verbalize their perceived differences using uni-dimensional, scalar, and non-hedonic descriptors, a training session was followed on such task. The elicitation task was conducted for a selection of visual stimuli to avoid aural biases. A set of dissimilar visual objects was presented on screen and the assessors were asked to characterize the differences. At the end, the assessors familiarized with the user interfaces.

The main listening experiment included four parts: the attribute elicitation, the attribute definition, the ranking, and the preference task. These tasks were conducted in the experimental apparatus described in Sec. II E, under dark conditions.

During the attribute elicitation phase, the assessors were asked to provide as many verbal descriptors as necessary to characterize the perceived differences between the presented sound stimuli. They were allowed to label the extreme intensities of that attribute, e.g., for the attribute "width" one could set high intensity to "very wide" and low intensity to "narrow." All stimuli were available on screen, including the nine acoustical conditions and the three program types. The graphical user interface is shown in Fig. 6(a). Following the attribute elicitation, the assessor provided definitions for the given attributes and ordered them based on the perceived audibility and perceptual importance.

Then, the assessors performed a ranking task. A block design was followed where each block evaluated one attribute in three trials, one for each program type. At each trial, the assessors were exposed to ten stimuli, the nine acoustical conditions, and a hidden repeat. The scales were presented as a continuous vertical slider, 15 cm long,[16] as shown in Fig. 6(b). Assessors were able to loop any time segment of the presented stimulus, as in standard evaluation procedures.[39] The presentation order of attributes, program types, and acoustical conditions were randomized.

After successful completion of the FP procedure, assessors were asked to indicate their personal preference to the presented stimuli. Prior to this phase, assessors were not informed about the context of the presented sound fields to allow faithful and unbiased judgments between the stimuli.[35] For the preference task, however, it was deemed necessary to inform the assessors about the experimental details so that an ecologically valid scenario could be realized and the appropriate hedonic response could be evoked. Literature pertaining the evaluation of real scenarios in the laboratory[17,40] recommends that it is possible to envisage assessors to a given situation. In the current study, the assessors were given the instruction of: "Imagine that you are in a typical residential room, listening to a 2-ch stereophonic reproduction over loudspeakers. Please rank the presented stimuli in a way that expresses your preference from 'highly dislike' to



(a) (b)

FIG. 6. (Color online) Graphical user interfaces during the two phases of the experiment. Buttons labeled as A–J provide switching between the stimuli. In elicitation, all programs × acoustical conditions were available on screen. In ranking, each trial assessed the acoustical conditions for one attribute and one program type, in a block-design. (a) Interface during elicitation. (b) Interface during ranking.

'highly like'." The attribute "preference" and its anchors were presented on screen, and assessors performed three trials, one for each program type, and all acoustical conditions.

During all the experimental phases, assessors followed self-paced and self-controlled procedures using a tablet. No time limits were set to complete the tasks, but short breaks were regularly followed to avoid possible listening fatigue. The whole experimental procedure, including breaks, was completed in 55–125 min (Mean = 84 min).

G. Assessors

Ten assessors participated in the experiment as volunteers. The assessors are considered as expert listeners, with an average experience of 11.4 years [standard deviation (*s.d.*) ± 5.5] in acoustics research and development, spatial audio reproduction, and critical listening as part of their profession. All assessors reported proficiency in standard audio evaluation procedures. Seven assessors were familiar with standard sensory analysis protocols for evaluation of audio material, of which six had performed attribute elicitation procedures before. They were aged between 27 and 47 years old (M = 34.1, *s.d.* ± 6.2). Their hearing sensitivity was above 20 dBHL between 125 and 8 kHz, confirmed by standard audiometric evaluation[38] at the time of the experiment.

III. RESULTS

In this experiment, nine acoustical conditions were evaluated in terms of their sensory characteristics and preference, for three program types. All ten assessors correctly identified the hidden repeats[41] in both tasks and no data points were eliminated for the statistical analysis. The experimental *dependent variables* (DV) included a total of 48 individually elicited attributes, quantifying the perceptual differences between the stimuli. These observations are further referred to as *sensory data*. The assessor's preferences were also collected in a free ranking task, and they are further referred to as *hedonic data*. The physical properties of the sound fields will be referred to as *physical data*. The datasets are summarized in Fig. 7.

In standard implementations of FP, i.e., in food products,[16] the statistical analysis of such data is based on multivariate techniques such as *generalized procrustes analysis* (GPA)[42] and *multiple factor analysis* (MFA).[43] When evaluating audio material, however, the stimulus-under-test is always the product of an acoustical condition, e.g.,

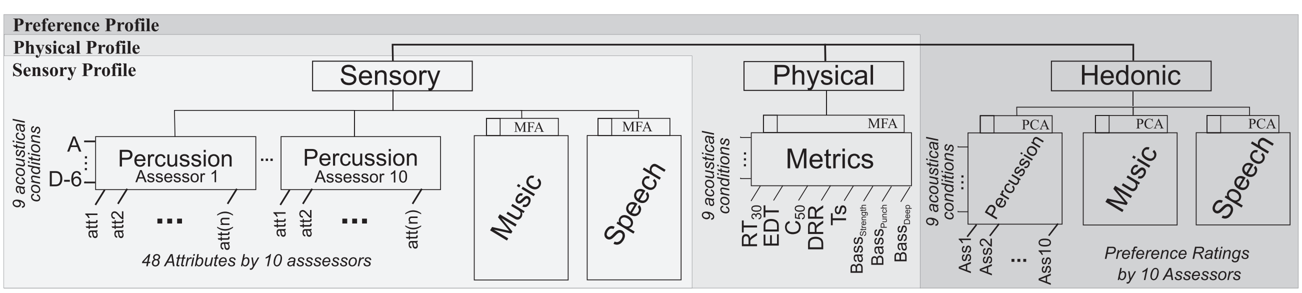


FIG. 7. Data structure used in the statistical analysis. Each hierarchical node, denoted as sensory, physical, and hedonic, describe the observations of a given dataset. The horizontal lines identify the hierarchical nodes for a given hierarchy tree. Sensory data included the individually elicited attributes, used for calculating the sensory profile in Sec. III A. Physical data included RT_{30} and EDT for each of the 31.5–8 kHz octave bands, C_{50}, DRR, and three metrics for the perceived bass, namely the Bass -Strength, -Punch, and DeepBass. The sensory and physical data were combined for calculating the physical profile, in Sec. III D. Hedonic data include the preference ratings of each assessor grouped by program type. All three datasets were combined to provide a global preference profile, presented in Sec. III E.

spectral modification, and an excitation program, e.g., speech signal.[35] This poses a challenge in the analysis, as possible interactions between the two factors, the acoustical conditions and the excitation programs, may not be easily identified by those statistical methods. To overcome these challenges, an extension of MFA was followed, the *hierarchical multiple factor analysis* (HMFA).[44]

Similar to MFA, HMFA provides a multidimensional solution of a given dataset which is typically shown as a graphical ordination of the observations in a common factorial space, known as *factor map*. HMFA applies a multivariate analysis to a group of variables that are hierarchically inter-related. In a given dataset, HMFA performs a factorial analysis at each hierarchical node of a defined hierarchy. The contribution of each node is equal to the global factorial space.[43] This enables a balanced statistical analysis of several multi-table datasets, aiming to identify the most prominent components between them and their underlying relationships. In an audio evaluation protocol, this allows the analysis of the perceptual dimensions for all assessors in a common factorial space, while the within-inertia (within-variance), of each excitation program and each assessor, is individually computed and preserved. Yet, their contribution to the global solution is equal.

In the following sections the data analysis is described. In Sec. III A, the ordination of sensory data is presented for each program level. This is followed by a classification of the perceptual attributes underlying the identified factorial solution in Sec. III B, which are then projected to the sensory profile of the acoustical conditions in Sec. III C. Following a comparison of the perceptual and physical data on the sensory profile in Sec. III D, the sensory, hedonic, and physical data are combined in Sec. III E.

A. Ordination of sensory data using HMFA

Initially, an HMFA[45] was conducted on the sensory data to evaluate the main effects of the design's IVs, the acoustical condition and program. The three hierarchical nodes of the sensory data included the observations of the nine acoustical conditions, for each program type separately, i.e., speech, music, percussion, as shown in Fig. 7.

Figure 8 depicts the factor map of the sensory data, including the partial contributions of the hierarchical nodes, the program levels. The partial points for each condition denote the direction of the variance, explained by the three data sets, i.e., the observations when listening to music, speech, and percussion. The barycenter of these points gives the mean ordination of the individual acoustical condition, across all programs.

The first two dimensions represent adequately the data, accounting for almost 80% of the total variance. All acoustical conditions are well separated, and no overlaps can be seen. On the first dimension, Rooms C and D-6 hold the extreme points, and the remaining conditions are ordered in between these. It can be seen that the acoustical conditions with low RT_{30} values are positioned negatively on this dimension, and the more reverberant conditions, on the positive side.

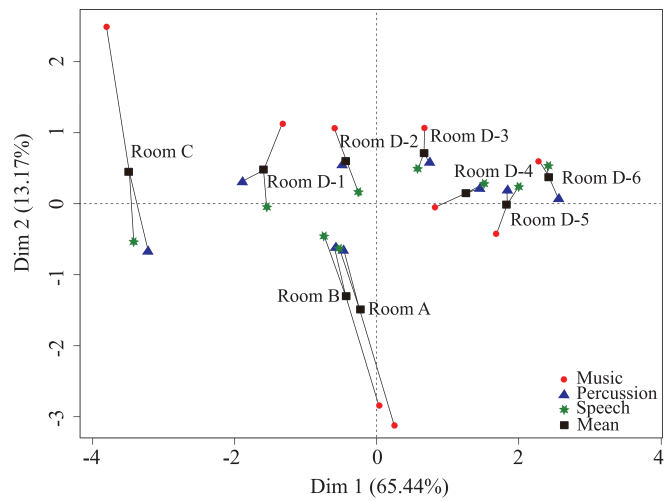


FIG. 8. Factor map showing the sensory profile of the experimental conditions. The profile is a result of three separate analyses, one for each program type. The mean is then calculated based on equally weighted contributions via the HMFA analysis, and the individual components are visualized in a common factorial space as partial points.

The second dimension is mainly driven by Rooms A and B, while the remaining acoustical conditions share similar vertical coordinates. In some conditions, e.g., Rooms A, B, and C, a strong opposition between the different program types is apparent, especially on the second dimension. To further analyze this, the axial inertia[46] for each of the hierarchical nodes for the first five principal components of the HMFA solution was calculated, shown in Table II. The analyzed axial inertia in Table II and the spread of programs in Fig. 8 indicates that music ratings were more sensitive to the perceptual property underlying the second dimension compared to when percussion and speech were used as the excitation signal. That is, although the solution of HMFA is balanced across program types, the second dimension in this analysis is driven by the assessors' ratings for program type music, compared to ratings when assessors listen to speech, and percussion, on the same acoustical conditions.

B. Clustering of individually elicited attributes

In the above analysis, the sensory data were summarized in global factorial solution. This factorial solution depicts the ordination of the experimental stimuli based on the perceived differences and similarities of all assessors and the effect of program levels to this solution was explored. It is possible that the perceptual attributes underlying these factorial relationships are included in the same analysis, suggesting the perceptual factors that drive this ordination. This is a

TABLE II. Decomposition of inertia of the HMFA analysis, in eigenvalues, split by each program type.

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
Percussion	0.99	0.09	0.13	0.05	0.04
Music	0.95	0.43	0.09	0.09	0.07
Speech	0.99	0.07	0.05	0.06	0.05
Total Inertia	2.93	0.59	0.28	0.20	0.17
Total Inertia %	65.44	13.17	6.16	4.48	3.70

trivial matter when assessors use a common list of attributes,[47] known as a consensus vocabulary, as each attribute is a common variable across all assessors.[16]

In individual vocabulary methods, such as the FP, each assessor uniquely labels, defines, and rates the perceived differences with no guidance or intensity training. Naturally, this limits the extent to which one could summarize the perceptual dimensions of all assessors by assuming semantic equivalence. In order to identify the common *perceptual constructs* underlying the individually elicited attributes across all assessors, a *post hoc* analysis is required, whereby attributes are classified into collective categories based on the data structures between them.

Here, the grouping of the collected attributes followed *agglomerative hierarchical clustering* (AHC), similar to previous individual vocabulary studies in audio.[19,21] An AHC based on the Euclidean distances of the MFA coordinates of each attribute was calculated in conjunction with Ward's criterion.[48] As AHC assigns equal weight to all attributes, independent of the contribution to the principal components, any individual attribute that did not correlate well to any of the first two principal components was excluded,[1] i.e., $|r|_{Dim1/Dim2} < 0.65$. A total of 43 attributes were used in the AHC analysis[49] as the best representative drivers to the dimensions of interest. All elicited attributes are given in the Appendix.

Figure 9 shows the two main clusters of the analysis, comprising of four main groups of variables. The first cluster splits into two branches. The first branch highlights attributes relating to the perceived effects of reverberance, e.g., Reverberation$_{\mathrm{A01,08}}$, Room Size$_{\mathrm{A03,09}}$. The second branch consists of attributes relating to the main components of spatial impression[50] of the sound field, i.e., the perceived width and envelopment. Last, the second cluster clearly identifies attributes relating to the perceived bass content on its first branch. Its second branch includes attributes relating to the perceived distance, closeness, and proximity. For consistency with previous studies,[19,51] this cluster will be referred to as *proximity*, indicating how close the auditory event is perceived, as defined by the assessors.

These four clusters form the perceptual constructs of this study and they will be further referred to as *reverberance*, *width and envelopment*, *bass*, and *proximity*.

It is noted that AHC seeks to classify interrelated attributes into a group. The attributes whose ratings were similar across the conditions are classified as homologous. This is also evident in our data. For example, Bass$_{A10}$ appears in the same cluster as Proximity$_{A10}$ and Distance$_{A06,07,09,10}$. The semantic equivalence between the attributes within certain clusters may not be strongly inferred at this point. This is a common observation in free verbal elicitation experiments,[47] where assessors label and quantify their perceived sensations without guidance. The clusters were therefore labeled based on the given attribute definitions by the assessors and the related literature.[1] Further analyses of these hypotheses follows in Sec. III D.

C. Sensory profile of acoustical conditions

Figure 10 shows the ordination of the acoustical conditions, i.e., the mean coordinates calculated in Sec. III A, and the projections of the perceptual constructs, identified in Sec. III B, in the form of vectors. The direction of each vector indicates the direction of inertia within the stimuli that is driven by the ratings of the clustered attributes, i.e., the perceptual construct. In effect, the projected vectors provide a perceptual explanation for the positioning of the acoustical conditions in this two-dimensional factorial space. The length of each vector indicates the quality of representation of the perceptual construct to the factorial space, i.e., its cumulative correlation to the principal components. In consequence, the projections with low correlation to the solution will be poorly represented due to their low contribution to the explained variance.

The analysis summarized in Fig. 10 shows that the first dimension is driven positively by the perceived *reverberance* and *width and envelopment*. *Proximity* vector faces the opposite direction, which indicates its negative correlation to the dimension. This relationship is well known to room acoustics, as the perceived distance increases as a function of reverberation, commonly explained by the lower values of DRR.[52]

The second dimension is described by the perceptual construct of *bass*. All variants of Room D share similar positions on this dimension, indicating perceptual equivalence to the perceived *bass* content. This is an expected result as the loudspeaker position was constant in Room D variations; thus, the coupling to room modes was not altered. Room A and B are perceived as less bassy compared to the remaining acoustical conditions. One should expect such results, as low frequencies in small enclosures are mainly driven by the modal behavior of the enclosure, in consequence, their physical dimensions. Based on the current evidence, one could infer that this dimension describes the perceived low frequency content in terms of spectral dissimilarity between the different rooms. This is also evident when contrasting the magnitude responses between the rooms, in Fig. 3(a), as Rooms A and B indicate the least energy levels below 100 Hz, compared to the other rooms.

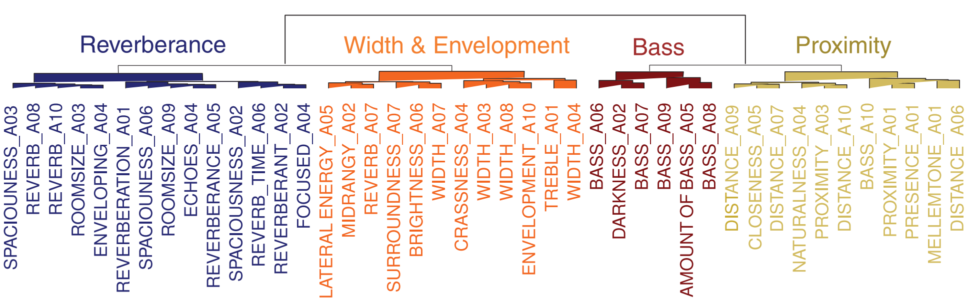


FIG. 9. Clustering dendrogram of individually elicited attributes of each assessor, based on the MFA coordinates of each attribute. The assessor's number is also included for each attribute, denoted as "A01-10."

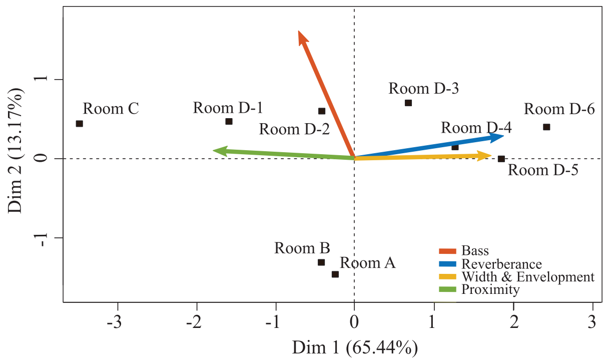


FIG. 10. Biplot showing the sensory profile of the acoustical conditions. Vectors indicate the direction of inertia, given by the mean coordinate each cluster, i.e., the perceptual construct, and colored based on the identified clusters.

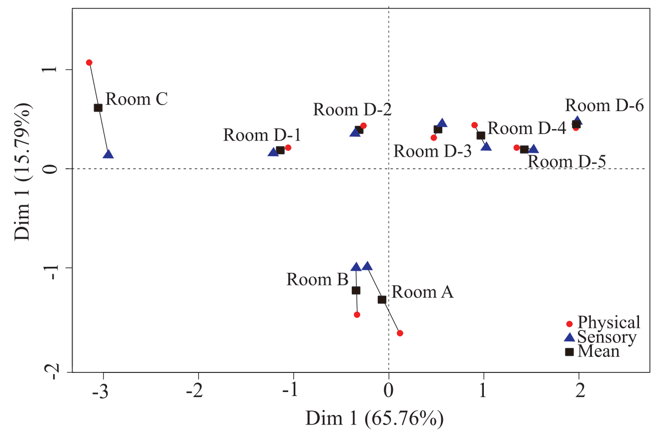


FIG. 11. HMFA factor map depicting the resulting profile of sensory and physical data. Partial points identify the ordination of the stimuli for the two hierarchical nodes of this analysis.

D. Physical profile

In the previous analysis, the sensory profile was constructed based on the assessors' sensory quantification and tangible explanations of the profile were identified and projected to a factorial solution. This analysis showed the sensory characteristics of the presented acoustical conditions, e.g., their perceived differences and similarities, and constructed a factorial plane describing these relationships based on solely perceptual data.

Using HMFA, it is possible to assess the *construct validity* of the applied experimental methodology. That is, the extent to which the identified perceptual constructs and sensory profiling relate to the physical characteristics of these fields. To evaluate this, the sensory data (the perceptual responses given by human assessors) and the physical data (the acoustical parameters of the sound fields) formed the hierarchical nodes of the HMFA. This analysis compared the profile of the stimuli based on the sensory characteristics alone, as given by the assessors' responses, and the profile based on the physical properties of the acoustical conditions—for example, the RT_{30} of an acoustical condition.

Figure 11 shows the ordination of the acoustical conditions based on this analysis in the first two dimensions. The explained variance of the analysis is 80%, suggesting a well represented dataset. Moreover, the close positioning of the partial points indicates a good agreement between the physical profile and the sensory profile as they hold similar coordinates in both dimensions. The partial points on Rooms A, B, and C, seem to differ in the second dimension, denoting a slight disagreement of the two datasets for these conditions.

To further investigate this, it is possible to project both the physical metrics and the perceptual constructs into the factorial space in a biplot, as shown in Fig. 12. The acoustical conditions are positioned in the factorial plane based on the mean coordinates of the sensory and physical ordination. The projected vectors indicate the directions of variance explained by the variables.

On the first dimension, the variance explained by the measured RT_{30} and the *early decay time* (EDT) of the acoustical conditions indicate an excellent relation to the perceptual construct of *reverberance*. The metrics relating to the temporal distribution of energy in the rooms, i.e., the *clarity index* (*50 ms*) (C50) and DRR, indicate strong correlation to the perceived *proximity*, opposing the perceived *width & envelopment*. *Center time* (TS)[32] correlates well to the first dimension, indicating its inversely proportional relationship to perceived *proximity* of the source, as well as its direct relation to the physical measure of RT_{30}.

Based on the sensory data in Sec. III C, it was argued that the second dimension relates to low frequency content of the conditions, as suggested by the perceived construct of *bass*. To objectively verify this hypothesis, a physical metric for the perceived bass content was included in the analysis. Recently, Volk *et al.* proposed a series of perceptually-based metrics, the BassPunch,[53] BassStrength,[54] and DeepBass,[55] aiming to assess the perceptual properties of broadband signals at low frequencies.[56] As shown in Fig. 12, the direction of inertia of all metrics suggests that the second dimension relates to spectral differences at low frequencies between acoustical conditions.

The previously identified interaction between the programs seen in Sec. III A is also apparent here. The projection of the partial vector of *bass* when assessors evaluated music content suggests a good correlation to the second dimension and the factorial solution. In contrast, the relatively short lengths of the partial vectors of *bass* relating to the other two program levels confirm a low quality of representation. This is an expected result, as the low frequency content of percussion and speech is limited, indicated in Fig. 3(b). It could be inferred that when these programs were used as an excitation signal, the audibility of these spectral differences between conditions was reduced, and in consequence the assessors' discrimination ability was affected.

In the previous analysis, described in Sec. III B, a hierarchical clustering was employed to identify the common perceptual constructs. The clusters were then labeled based on the semantic definitions of assessors' own attributes and previous studies. In Sec. III C, the clusters were projected into the factorial space, aiming to understand the perceptual relevance of the dimensions of the sensory profile. In this section, the analysis suggests that the factorial space constructed from the assessors' quantification of the perceived differences relates to the physical characteristics of the sound fields used in this investigation, supporting the

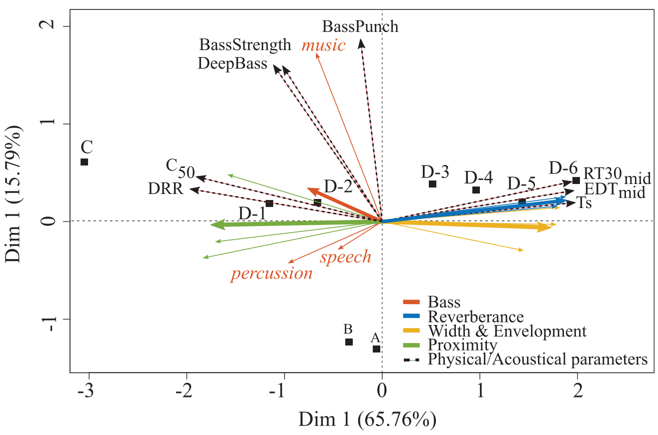


FIG. 12. Biplot depicting the common profile for sensory and physical data and the directions of inertia driven by variables in the sensory and physical data. EDT_{mid} and $RT30_{mid}$, refer to the average values between 500 Hz and 1 kHz as recommended (Ref. 32) for perceptual relevance.

hypotheses described in Sec. III A, and the observed construct validity of the experimental design. The identified perceptual constructs using AHC and the proposed labels come in agreement with the physical characteristics of the sound field, as the projections of the identified perceptual constructs share high similarities to the physical metrics that are known to excite such sensations.

E. Sensory, hedonic, and physical data—Global profile

In the final part of the experiment, the assessors' hedonic responses were also collected in a preference task. In order to understand the relations between the sensory, physical, and hedonic responses of assessors, *preference-mapping* techniques are typically followed. Such analysis allows the direct identification of each acoustical conditions' sensory and physical characteristics, driving the assessors' preferences. For complex multi-table datasets, as the data in this study, preference mapping is achieved by computing the factorial solution of all the datasets simultaneously, and projecting the driving variables in that space.[19]

The common factorial solution of sensory, hedonic, and physical data of this experiment are graphically shown in Fig. 13, indicating the drivers of the inertia in this factorial plane from all three data structures.

The position of Room D-6 opposed the construct of *preference*, suggesting that this condition was the least favorite across assessors, whereas Rooms C, D-1, and D-2 were the most preferred. The perceived sound in these sound fields exhibit high *proximity* and they are identified as less reverberant, wide, and enveloping. Assessors' preference is explained well by lower decay times given by RT_{30} and EDT, higher levels of DRR and C50, and in consequence lower values of TS.

The perceived bass does not seem to explain the preference ratings adequately; yet, it indicates a contribution to the variance due to its position on the first dimension. The depicted partial vectors, indicating the direction of inertia for a perceptual construct split by program type, have similar directions, including the preference vector. That confirms that the mean directional vectors of the perceptual constructs could explain adequately the factorial space, and no interactions between these partial programs can be seen in this analysis.

IV. DISCUSSION

The primary findings suggest that the aural experience within domestic listening environments is highly affected by its acoustical properties and reverberation's characteristics. Both timbral and spatial characteristics of the reproduced sound fields were found to be altered, even when the differences of the rooms are subtle and within the proposed recommendations of audio evaluation standards, such as the ITU-R 60268-13.[11]

In the statistical analysis described above, it is apparent that two factorial dimensions seem to characterize the perceptual space evoked by the presented sound fields; the explained variance in those dimensions was $\approx 80\%$.

Berkley and Allen[57] have previously argued upon a two-dimensional perceptual space, which humans seem to follow when the perceptually evaluated simulated small rooms. In that study, listeners judged the room similarities based on the decay times and the differences in stimuli' spectrum. Zahorik *et al.*[58] reported analogous findings for a larger number of synthesized fields. Our results are consistent with the previous findings, whilst the method provided a more realistic sound field presentation to assessors and followed a controlled sensory analysis protocol.

The followed methodology allowed for the identification of the perceptual sensations as well as to quantify those numerically. This has been achieved by HMFA, where it statistically combined the quantitative sensory data with the perceptual attributes describing those, and contrast the results with the physical properties of the presented sound fields, all of which seem to support a perceptual space of two major dimensions. In our study, the first and dominant dimension relates to the decay times and reverberation's cognate percepts, whilst the second dimension relates to spectral characteristics. This was supported by analyzing both the sensory data, i.e., the perceived spatial and timbral characteristics given by the assessors, as well as the physical characteristics of the presented sound fields.

The main aim of the study was to decompose the factorial dimensions to the individual perceived sensations that they underline. The data indicate that assessors were able to identify four perceptual constructs,[59] and appropriately quantify the differences within the stimuli. In the first dimension, the factorial drivers relate to the perceived *reverberance*, *width and envelopment*, and *proximity*. The second dimension relates to the perceived *bass* as described by the assessors and the perceptually-relevant metrics.

The claimed multi-dimensionality of reverberation and its interdependent perceptual components found in concert hall research is also apparent in this study where the focus was smaller listening environments. For example, although assessors reported attributes relating to perceived width and the perceived envelopment independently during the elicitation, those have been identified as components of one

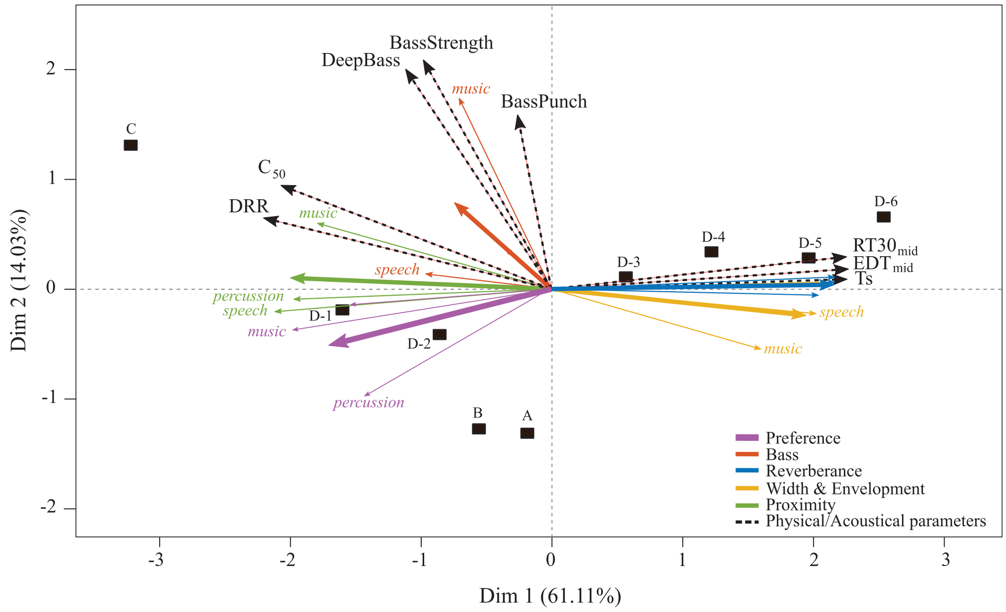


FIG. 13. Biplot summarizing the common factorial space for sensory, physical, and hedonic data. Vectors indicate the direction of inertia for the presented variables. Thinner vectors decompose the averaged inertia, per program, labeled as speech, music, and percussion, accordingly, indicating possible program interactions. The lilac vector shows the direction of preference for the presented acoustical conditions; D-1 and D-2 are the most preferred in contrast to D-5 and D-6.

perceptual construct when AHC was performed on the assessors ranking data. This is a known challenge in evaluations of realistically-complex soundfields, and several studies[1] have shown that the perceived width and envelopment are attributes that listeners show difficulty in separating.

Nonetheless, it is generally agreed that whilst perceived width and envelopment relate to the earlier energy of the sound field, reverberance relates to the later energy. In our data, *reverberance* and *width and envelopment* have been identified as two separate perceptual constructs, yet the direction of the inertia explained by them on the major dimensions is similar. This suggests that a condition that has been described and quantified as more reverberant by the assessors has also elicited a more wide and enveloping feeling and vice versa. This relates to the low levels of RT and early reflection points within the stimuli set, for example, Room C. As the assessors' task was to identify and quantify the perceived differences between the stimuli, the perceived *reverberance* and *width and envelopment* were determined and quantified as separate percepts. Moreover, in this study, there was no intention to systematically vary the early reflection patterns in a specific manner, but rather compare complete sound fields of typical listening environments. Those inherently include several variations of the acoustic sound field and a number of uniform spatial alterations of RT.

It is therefore suggested that the four perceptual constructs identified in the study form major characteristics of perceived aural experience in the studied scenarios.

Finally, the preference profile (Sec. III E) combined the sensory, physical, and hedonic observations. This analysis supports previous findings,[60] that assessors systematically preferred the sound fields with lower RT. In our study, the most preferred acoustical conditions presented fields that evoked the sense of being less reverberant and less wide and enveloping. The sources were perceived as closer to the listener, exhibiting high levels of *proximity*. It is also important to note that the current results suggested that a negative preference is apparent for acoustical conditions with RT higher than 0.4 s, the proposed mean value in the IEC recommendation.[11] Differences relating to the low frequency content within the presented conditions have also influenced assessors' preference, but at a lower degree than one could expect; this may relate to the specific programs used in the study and the controlled experimental apparatus followed.

Summarizing, the results signify the effects of acoustical properties on the perceived sound and indicate the importance of strict RT limits in standard listening rooms, when loudspeakers[11] and impairments of audio signals[10] are evaluated. It was shown that alterations of the RT, e.g., ≈0.1 s between Rooms D-2 and D-3, evoked different hedonic and sensory percepts, as well as the perceived intensities between them. This is an important element when room-in-room effects are observed in the domain of spatial audio and soundscape rendering, in small listening spaces.

A. Limitations and future work

Following the described experimental paradigm, it was possible to identify and accurately quantify the human sensations using a rapid sensory analysis method, the FP. The method followed seems to have overcome multiple challenges relating to evaluation of room acoustics properties. Spatial aspects of the sound fields that one typically perceives in a real environment, i.e., distance from the source and apparent source width, were evoked and appropriately quantified in a consistent way when using spatial reproduction over loudspeakers compared to previous investigations that failed to reproduce such attributes.[23,58] Similarly, temporal, and dynamic elements of the fields were also evoked and quantified in an expected manner.

The current approach has several methodological limitations. The study utilized a panel of expert assessors, trained in the domain of audio evaluation. The robust and consistent quantification of the perceived sensations in individual vocabulary methods without training depends highly on the experience of the assessors, which is a fundamental principle of FP protocol. By conducting FP with expert listeners, the experimental time required was significantly reduced[23] and only a few assessors were needed to complete the task, as no aural or vocabulary-based training was required.[61]

Inexperienced assessors are known to have difficulty verbalizing the perceived sensations and clustering a consensus lexicon using statistical methods is therefore expected to be more challenging. Nonetheless, the professional experience of the assessors in this study is likely to have implications on hedonic responses, such as preference and likeness. All assessors were accustomed with standard listening rooms and critical listening environments due to their profession. In consequence, their internal inference and reference of a typical reproduction system may differ from an average listener. It is therefore noted, that the current findings relating to assessors' preferences may not reveal the judgments and sentiments of an average listener.

Further work is required to expand the limited scope of the study and generalize the results. In this experiment, the stimuli set comprised of nine sound fields excited by a stereophonic setup in acoustical settings defined by IEC:60268–13 recommendation,[11] at a single listening position. The main scope was to verify the extent to which alterations within those standards affect the listening experience and attempt to decompose those into the perceptual attributes that describe those basic reproduction scenarios. Those results stipulate possible relationships to physical quantities, such as standard acoustical metrics.

The extension of this work should follow an extensive objective analysis of such fields, such as spatial and temporal analysis, that could improve our understanding of the perceptual aspects of the fields, as identified here. This may include the investigation of the effects of early reflection patterns on the perceived sound from different directions, and the effects of more geometrically varied rooms, including irregular shapes, as well as asymmetric and multichannel reproduction setups.

In the current experimental design, a controlled acquisition procedure was followed aiming to avoid strong modal behavior at low frequencies, a well known influencing factor in those spaces, and focus on the spatiotemporal properties of the sound field. Still, the effects of low frequency differences between Rooms A and D seem to be an important perceptual aspect; here labeled as *bass*, which explained $\sim$13% of the variance of the sensory data. One could infer that variance of low frequency modal resonance would have impacted the perceived sensation of bass, both in terms of temporal notion, as a longer decay, as well as perceived bass level. This should postulate further research in the domain where the temporal and spectral content at those frequencies is evaluated.

Finally, future work could investigate the sensory sensitivities and hedonic responses of non-trained assessors for such stimuli set in order to identify the extent to which of the findings can be generalized, as the current contextual paradigm cannot support this.

V. CONCLUSIONS

This experiment investigated the perceptual effects of reverberation in residential listening rooms in a basic multi-loudspeaker sound reproduction scenario. Nine sound fields were used in the study, representing a range of possible decay times according to the IEC:60268–13 recommendation,[11] including maximal anchors. The sound fields were captured *in situ* and reproduced over a spherical loudspeaker array in an anechoic chamber. A blind experimental protocol was followed, where ten expert assessors identified their own perceptual attributes and quantified their perceived sensations for the presented sound fields. Using multidimensional factorial analysis, HMFA, a sensory profile for the acoustical conditions was calculated depicting the perceived differences between the acoustical conditions. The main drivers of this profiling, i.e., the common perceptual constructs, were identified and further validated against physical properties of these sound fields. The assessors' hedonic responses were collected and analyzed indicating the assessors' preferences for a typical reproduction setup, in a domestic listening room.

Three research aims were the basis of this investigation: (i) explore the extent to which reverberation affects the perceived sound experience, (ii) identify the related perceptual attributes that characterize this, and (iii) attempt to link those to objective metrics.

Overall, the current results suggest the significance of reverberation on human perception in residential-sized rooms, supplemented by spectral modifications at the low frequencies. This supports previous studies[4,5] and suggests that the perceived sound field can be significantly altered by the acoustical properties of the reproduction room. The previously argued two-dimensional character of small room acoustics[57,58] was also apparent here. In our study, those dimensions seem to be driven by four perceptual constructs that underline the factorial space in this investigation. The perceived *reverberance*, *width and envelopment*, and *proximity* seem to explain the majority of perceptual differences between the sound fields used in the experiment, in addition to alterations of the perceived *bass* at a lower degree.

A global analysis of the experimental conditions combining the sensory, hedonic, and physical properties of the presented sound fields allowed the identification of the main drivers of assessors' preferences. The analysis indicated that rooms described by lower RT are preferred. It is evident that a critical value seems to exist, close to the recommended mean value of 0.4 s,[11] above which assessors' preference degrades. This is an interesting result and follow up studies should investigate the influence of that from assessors' background, the program material, and the listener-loudspeaker positioning within the rooms.

It is our belief that further work in the domain would be highly beneficial for future domestication of spatial audio at home. Understanding the acoustic influence of these environments on the reproduced sound field will enhance the system's ability to recreate a sonic experience in acoustically-dissimilar enclosures,[62] in a more accurate and perceptually relevant way. For example, one could attempt to alter the DRR within a field by means of directivity control in the loudspeakers, aiming to evoke certain perceptual aspects that would otherwise be dominated by the room's natural acoustical field. This desire is evident, as recent advances in spatial audio reproduction over loudspeakers[63] target domestic reproduction environments where the current knowledge is limited.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Program under Grant No. ITN-GA-2012-316969.

APPENDIX: ELICITED ATTRIBUTES & DEFINITIONS

Table III.

TABLE III. Individual Elicited attributes grouped by AHC.

Group	Attribute	Low-High Anchors	Definition
Reverberance	Reverberation$_{A01}$	Little-much	Two effects, decay and transient smearing (danish Efterklang)
	Reverb$_{A08}$	Low-high	Reverberance, its tail length
	Room Size$_{A03}$	Small-big	Size of the room due to reverb
	Room Size$_{A09}$	Dry-reverby	How big the room is, small is dry, big is not.
	Reverb$_{A10}$	Short-long	Roominess, especially on vocals, dry or not?
	Reverberance$_{A05}$	Low-high	Reverb time, tail length in time
	Reverb Time$_{A06}$	Short-long	Room size, damping or not
	Reverberant$_{A02}$	Dry-reverberant	Perceived reverb time, decay but also spatial effects, room type
	Enveloping$_{A04}$	Less-more	Feeling of being surrounded by the field
	Focused$_{A04}$	Less-more	Sound comes from single point or not
	Spaciousness$_{A02}$	Small-large	The amount of directions the sound/reflections appears from
	Spaciousness$_{A03}$	Focused-diffused	Centre sound comes from single point or diffused due to reflections
	Spaciousness$_{A06}$	Small-big	Band in front of you or surrounding
	Echoes$_{A04}$	Dry-echoy	Audible discrete reflections or not
Width & Envelopment	Lateral Energy$_{A05}$	Low-high	Ability to detect discrete reflections coming from side or above
	Midrangy$_{A02}$	Weak-dominant	Mid frequencies dominance, spectral balance difference related
	Reverb$_{A07}$	Dry-wet	Point source or not, from damped room to church feeling
	Surroundness$_{A07}$	Point-diffuse	Width plus reverb, from a point to diffused, all around sound
	Brightness$_{A06}$	Dark-light	Level of high frequency sound
	Width$_{A07}$	Narrow-wide	Location of source in front or wider, still horizontal front.
	Crassness$_{A04}$	Low-high	Midrange aggressiveness, sound annoying due to balance
	Width$_{A03}$	Point-wide	From single point source to diffused, all around sound
	Width$_{A04}$	Narrow-wide	The source width at the front horizontally
	Width$_{A08}$	Narrow-wide	Sound stage, from localized to all over sound
	Envelopment$_{A10}$	Front-surround	How surrounding the sound is, is it frontal or surrounding?
	Treble$_{A01}$	Little-much	High frequency level (danish Diskant)
Bass	Amount of Bass$_{A05}$	Low-high	Bass level and higher-bass i.e, speech low content
	Bass$_{A06}$	Weak-strong	Bass level aural and feel, mainly on music track
	Bass$_{A07}$	Low-high	Amount of bass, both aural and feel
	Bass$_{A08}$	Nobass-rumble	Energy of (any) low end, level related
	Bass$_{A09}$	Light-heavy	Transient of bass, ringing sometimes-big room feeling
	Darkness$_{A02}$	Thin-full	Bass level-related. the full has more bass
Proximity	Proximity$_{A01}$	Little-great	Closeness, distance of source
	Presence$_{A01}$	Little-much	Feeling that I am there (danish Naevaer)
	Closeness$_{A05}$	Near-far	How close/far is the source
	Distance$_{A10}$	Close-far	How close/far is the source
	Proximity$_{A03}$	Close-far	Feeling of source being close to your face
	Distance$_{A06}$	Close-far	Singer in front of you or far away
	Distance$_{A09}$	Near-far	Sound 'in your face' or far away
	Distance$_{A07}$	Near-far	How close/far I perceive the sound
	Bass$_{A10}$	Weak-strong	Bass level, noticeable on music mainly
	Naturalness$_{A04}$	Artificial-natural	Is the source properties (i.e,. voice) as you expect?
	Mellemtone$_{A01}$	Little-much	Midrange, strong vocal dominance on image, all tracks
Excluded	Bass$_{A01}$	Little-much	Bass level, dominance, but not in all tracks
	Punch$_{A04}$	Not-punchy	Feel-vibration related at mid-bass
	Resonance$_{A02}$	Clear-ringing	Feel of resonating frequencies, masking vocals
	Airness$_{A02}$	Dull-shy	Similar to treble, if not there sound is shy
	Height$_{A08}$	Low-high	Sound perception at elevated positions

[1]N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: A literature study," in *Proceedings of the 55th International Conference on Spatial Audio*, Helsinki, Finland (August 27–29, 2014).
[2]R. Hughes, T. Cox, B. Shirley, and P. Power, "The room-in-room effect and its influence on perceived room size in spatial audio reproduction," in *Proceedings of the 141st Convention of the Audio Engineering Society*, Los Angeles, CA (September 29–October 2, 2016).
[3]J. Grosse and S. van de Par, "Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers," IEEE J. Selected Topics Signal Process. **9**(5), 867–880 (2015).
[4]S. Bech, "Timbral aspects of reproduced sound in small rooms: Part II," J. Acoust. Soc. Am. **99**(6), 3539–3549 (1996).
[5]F. E. Toole, *Sound Reproduction: Loudspeakers and Rooms* (Focal Press, Boston, 2008), p. 550.
[6]A. Giménez, R. M. Cibrián, S. Cerdá, S. Girón, and T. Zamarre, "Mismatches between objective parameters and measured perception assessment in room acoustics: A holistic approach," Build. Environ. **74**, 119–131 (2014).
[7]S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, "Psychoacoustic cues in room size perception," in *Proceedings of the 116th Audio Engineering Society Convention*, Berlin, Germany (May 8–11, 2004).
[8]S. Olive, T. Welti, and W. L. Martens, "Interaction between loudspeakers and room acoustics influences loudspeakers preferences in multichannel reproduction," in *Proceedings of the 123rd Convention of the Audio Engineering Society*, New York (October 5–8, 2007), Paper No. 7176.
[9]S. Bech, "Spatial aspects of reproduced sound in small rooms," J. Acoust. Soc. Am. **103**(1), 434–445 (1998).
[10]International Telecommunication Union, *BS. 1116-3: Methods For the Subjective Assessment of Small Impairments in Audio Systems* (ITU-R, Geneva, Switzerland, 1997).
[11]International Electrotechnical Commission, *IEC:60268, Sound System Equipment Part 13: Listening Tests on Loudspeakers* (IEC, Geneva, Switzerland, 1998).
[12]M. Sams, R. Hari, J. Rif, and J. Knuutila, "The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence," J. Cogn. Neurosci. **5**(3), 363–368 (1993).
[13]B. Shinn-Cunningham, "Learning reverberation: Considerations for spatial auditory displays," in *Proceedings of the 6th International Conference on Auditory Displays (ICAD2000)*, Atlanta, GA (April 2–5, 2000).
[14]C. Pike, T. Brookes, and R. Mason, "Auditory adaptation to loudspeakers and listening room acoustics," in *Proceedings of the 135th Convention of the Audio Engineering Society*, New York (October 17–20, 2013), Paper No. 8971.
[15]H. Hacihabiboglu, E. D. Sena, Z. Cvetkovic, J. Johnston, and J. O. Smith III, "Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics," IEEE Signal Process. Mag. **34**(3), 36–54 (2017).
[16]H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices* (Springer, New York, 1999), p. 471.
[17]J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Elicitation of attributes for the evaluation of audio-on-audio interference," J. Acoust. Soc. Am. **136**(5), 2630–2641 (2014).
[18]G. Lorho, "Perceived quality evaluation—An application to sound reproduction over headphones," Ph.D. thesis, Aalto University, Finland (2010).
[19]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).
[20]G. Martin and S. Bech, "Attribute identification and quantification in automotive audio—Part 1: Introduction to the descriptive analysis technique," in *Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Spain (May 28–31, 2005).
[21]N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S. Jensen, "Perceptual aspects of reproduced sound in car cabin acoustics," J. Acoust. Soc. Am. **141**(3), 1459–1469 (2017).
[22]V. Dairou and J.-M. Sieffermann, "A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile," J. Food Sci. **67**, 826–834 (2002).
[23]N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S. Jensen, "A rapid sensory analysis method for perceptual assessment of automotive audio," J. Audio Eng. Soc **65**(2), 130–146 (2017).
[24]Combination of reflections, absorbers, and diffusers as recommended.
[25]International Telecommunication Union, *Rec. 775-3* Multichannel Stereophonic Sound System With and Without Accompanying Picture (ITU-R, Geneva, Switzerland, 2012).
[26]The ratio between the loudspeaker position and the boundaries was common across rooms.
[27]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proceedings of the 108th Convention of the Audio Engineering Society*, Paris, France (February 19–20, 2000), Paper No 5093.
[28]S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**(1/2), 17 (2013).
[29]S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, and T. Lokki, "Spatial analysis and synthesis of car audio system and car-cabin acoustics with a compact microphone array," J. Audio Eng. Soc. **63**(11), 914–925 (2015).
[30]J. Pätynen, S. Tervo, and T. Lokki, "Amplitude panning decreases spectral brightness with concert hall auralizations," in *Proceedings of the 55th International Conference on Spatial Audio*, Helsinki, Finland (August 24–27, 2014).
[31]International Organization for Standardization, *ISO 3382-1: Acoustics—Measurement of Room Acoustic Parameters, Part 1: Performance Spaces* (ISO, Geneva, Switzerland, 2009).
[32]International Organization for Standardization, *ISO 3382−2: Acoustics—Measurement of Room Acoustic Parameters*, *Part 2: Reverberation Time in Ordinary Rooms* (ISO, Geneva, Switzerland, 2008).
[33]G. Vairetti, N. Kaplanis, E. D. Sena, S. H. Jensen, S. B. M. Moonen, and T. van Waterschoot, "The subwoofer room impulse response database (SUBRIR)," J. Audio Eng. Soc. **65**(5), 389 (2017).
[34]International Telecommunication Union, *Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level* (ITU-R, Geneva, Switzerland, 2015).
[35]S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application* (Wiley, England, 2006), p. 442.
[36]J. Pätynen, S. Tervo, and T. Lokki, "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," J. Acoust. Soc. Am. **133**(2), 842–857 (2013).
[37]J. Blauert and J. Allen, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, 1997), pp. 35–40.
[38]International Organization for Standarization, *ISO:8253-1 Audiometric Test Methods Part 1:Pure-Tone Air and Bone Conduction* (IEC, Geneva, Switzerland, 2010).
[39]International Telecommunication Union, *Rec. BS. 1534-2 Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems* (ITU-R, Geneva, Switzerland, 2014).
[40]J. Fields, R. D. T. Gjestland, I. Flindell, R. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. G. U. Felscher-Suhr, and R. Schumer, "Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation," J. Sound Vib. **242**, 641–679 (2001).
[41]The ratings for the relative difference between the repeated stimuli that was over 10% occurred less than 15% of all trials, as proposed in Ref. 39.
[42]J. Gower, "Generalized procrustes analysis," Psychometrika **40**, 33 (1975).
[43]J. Pagès, *Multiple Factor Analysis by Example Using R* (CRC Press, Boca Raton, FL, 2014), p. 247.
[44]S. Le Dien and J. Pagès, "Hierarchical multiple factor analysis: Application to the comparison of sensory profiles," Food Qual. Prefer. **14**, 397–403 (2003).
[45]S. Lê, J. Josse, and F. Husson, "FactoMineR: An R package for multivariate analysis," J. Stat. Softw. **25**, 1–18 (2008).
[46]The inertia of each dimensional axis.
[47]S. Lê and T. Worch, *Analyzing Sensory Data with R* (CRC Press, Boca Raton, FL, 2015).
[48]A. Lucas, "amap: Another Multidimensional Analysis Package (R-package version 0.8-14)," (2014), https://cran.rproject.org/web/packages/amap/index.html (Last viewed June 1, 2019).
[49]5 were removed due to limited correlation to the dimensions.
[50]T. H. Pedersen and N. Zacharov, "The development of a sound wheel for reproduced sound," in *Proceedings of 138th Convention of the Audio Engineering Society*, Warsaw, Poland (May 7–10, 2015).
[51]A. Sotiropoulou, "Concert hall acoustic evaluations by ordinary concert-goers: I Multi-dimensional description of evaluations," Acta Acust. **81**, 10–19 (1995).
[52]A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," Nature **397**, 517–520 (1999).

[53]P. Volk, S. Bech, T. Pedersen, and F. Christensen, "Modelling perceptual characteristics of loudspeaker reproduction in a stereo setup," J. Aud. Eng. Soc. **65**(5), 356–366 (2017).
[54]P. Volk, S. Bech, T. Pedersen, and F. Christensen, "Modelling perceptual characteristics of prototype headphones," in *Proceedings of the International Conference on Headphone Technology*, Aalborg, Denmark (August 24–26, 2016).
[55]P. Volk, M. Lavandie, S. Bech, and F. Christensen, "Identifying the dominating perceptual differences in headphone reproduction," J. Acoust. Soc. Am. **140**(5), 3664–3674 (2016).
[56]Those metrics expect binaural input. To accommodate this, the SDM field was synthesized using the CIPIC database.
[57]D. A. Berkley and J. B. Allen, "Normal listening in typical rooms: The physical and psychophysical correlates of reverberation," in *Acoustical Factors Affecting Hearing aid Performance*, edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Boston, 1993), pp. 3–14.
[58]P. Zahorik, "Perceptually relevant parameters for virtual listening simulation of small room acoustics," J. Acoust. Soc. Am. **126**, 776–791 (2009).
[59]The perceptual attributes were rated consistently and reliably, as hidden repeats were identified and the factorial analysis depicted these constructs as separate groups of variables.
[60]S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, "Preferences of critical listening environments among sound engineers," J. Audio Eng. Soc. **62**(5), 300–314 (2014).
[61]J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (Woodhead Publishing, Elsevier, Cambridge, UK, 2014), p. 555.
[62]C. Hak and R. Wenmaekers, "The impact of sound control room acoustics on the perceived acoustics of a diffuse field recording," WSEAS Trans. Signal Process. **6**, 175–185 (2010).
[63]S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial sound with loudspeakers and its perception: A review of the current state," Proc. IEEE **101**(9), 1920–1938 (2013).