KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Neofytos Kaplanis, Søren Bech, Sakari Tervo, Jukka Pätynen, Tapio Lokki, Toon van Waterschoot, and Søren Holdt Jensen **A rapid sensory analysis method for perceptual assessment of automotive audio** *J. Audio Eng. Soc.*, vol. 65, no. 1/2, pp. 130-146, Jan./Feb. 2017.
Archived version	Author manuscript: the content is identical to the content of the submitted paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.17743/jaes.2016.0056
Journal homepage	http://www.aes.org/journal/
Author contact	toon.vanwaterschoot@esat.kuleuven.be *+ 32 (0)16 321927*
IR	ftp://ftp.esat.kuleuven.be/pub/SISTA/vanwaterschoot/abstracts/16-76.html

(article begins on next page)

Journal of the Audio Engineering Society
Vol. 65, No. 1/2, January/February 2017 (© 2017)
DOI: https://doi.org/10.17743/jaes.2016.0056

A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio*

NEOFYTOS KAPLANIS,[1, 2] *AES Student Member*, **SØREN BECH,**[1, 2] *AES Fellow*,
(neo@bang-olufsen.dk) (sbe@bang-olufsen.dk)

SAKARI TERVO[3], **JUKKA PÄTYNEN**[3], **TAPIO LOKKI,**[3] *AES Member*,
(sakari.tervo@aalto.fi) (jukka.patynen@aalto.fi) (tapio.lokki@aalto.fi)

TOON VAN WATERSCHOOT,[4] *AES Associate Member*, **AND SØREN HOLDT JENSEN**[2]
(toon.vanwaterschoot@esat.kuleuven.be) (shj@es.aau.dk)

[1] *Bang & Olufsen A/S, Peter Bangs Vej 15, 7600 Struer, Denmark*
[2] *Aalborg University, Department of Electronic Systems, 9220 Aalborg, Denmark*
[3] *Aalto University, Department of Computer Science, FI-00076 Aalto, Finland*
[4] *KU Leuven, Department of Electrical Engineering, ESAT-STADIUS/ETC, 3001 Leuven, Belgium*

Today's car audio systems encompass some of the latest sound technologies available capable of delivering unique and novel aural experiences. It remains unclear whether current perceptual assessment protocols follow this trend, and their ability to fully capture the human sensations evoked by such systems is questioned. This paper reviews the applicability of existing assessment protocols in today's systems and draws upon the identified limitations. It further reports the design and implementation of a new method to perceptually investigate the properties of automotive audio. The method uses Spatial Decomposition Method for acquiring, analyzing, and reproducing the sound field in a laboratory over loudspeakers, allowing instant comparisons of automotive audio systems. A Rapid Sensory Analysis protocol, the Flash Profile, is employed for evaluating the perceptual experience using individually elicited attributes in a time-efficient manner. A pilot experiment is presented where experts, experienced, and naive assessors followed the proposed procedure and evaluated three sound fields. The current findings suggest that the method allows the assessment of both spatial and timbral properties of automotive audio. This may form a scientific framework for characterizing the acoustical qualities within the automotive environment and stipulate research paths to better understand these sound fields.

0 INTRODUCTION

Over the last decades the automotive industry has been focusing on identifying and improving the major factors that influence the sensory experience within the vehicles. As a consequence, the study of sound quality in automotive audio systems has been brought into the limelight [1]. Although sound quality research has established and standardized a plethora of assessment procedures [2, 3], protocols for automotive audio are yet to be defined. Here, the published literature on the past and current practices is reviewed, aiming to stipulate new approaches and encourage a general framework for automotive audio assessment.

The highly complex and acoustically hostile environment of a car cabin [4–6] hinders the effectiveness of standard objective metrics [7], lacking robustness, repeatability, and perceptual relevance [8, 9]. This has naturally led to the use of the human auditory system as a major instrument in developing and evaluating car audio. Aiming towards a high quality aural experience, car audio manufacturers normally employ perceptual assessment protocols to characterize and optimize these sound fields [10].

In the lack of standardized evaluation procedures [11], automotive audio assessment protocols adopted paradigms from sound quality research in rooms. As a consequence, the majority of the studies have been focusing on comparative evaluation of system properties. For example, towards the electroacoustic properties of the transducers, signal processing algorithms, and equalization settings [11–17], while attaining a perceptual experience similar to a conventional listening space, such as a mixing studio or a listening room [13, 18].

Yet, the car cabin is far from a common listening environment [13, 14, 19]. Compared to a typical listening room

*Parts of this paper have been presented at the 60th International Conference of Audio Engineering Society, Leuven, Belgium, 2016.

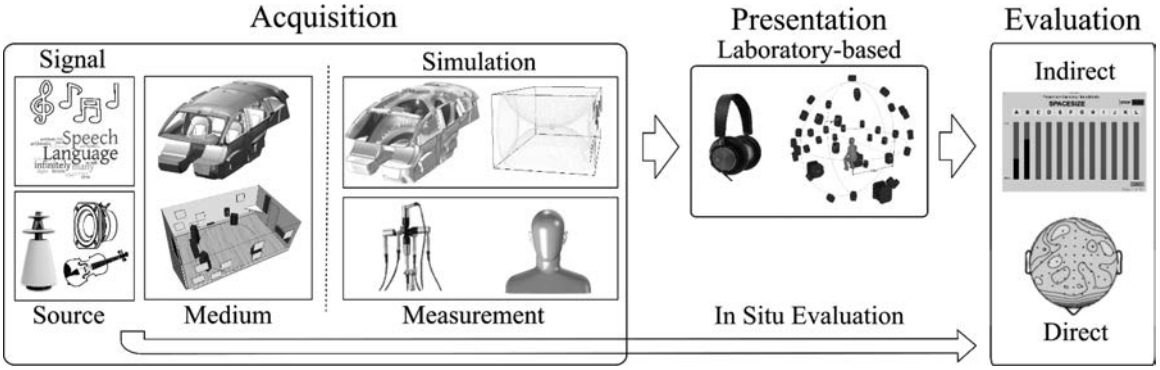


Fig. 1 Basic principles of perceptual audio assessment.

[2], the car cabin is characterized by a small volume, highly reflective surfaces that contrast with the inner absorptive upholstery, complex geometry, limited and often sub-optimal placement of the sound sources, as well as asymmetric and dissimilar acoustical paths.

In addition, current car audio systems are comprised of highly advanced loudspeaker grids [1] capable of delivering novel and environment-specific aural experiences. This trend surpassed the development of sound reproduction in typical rooms, increasing the dissimilarity of the two scenarios.

One could therefore challenge the applicability of the current perceptual assessment protocols for automotive audio evaluation [20]. New experimental frameworks should be developed, with the restrictions and specificity of the automotive environment in mind. That is, methods that would allow the identification and quantification of human perception even when the properties of the sound fields are foreign to the human ear, i.e., percepts that are difficult to estimate a priori. Such approaches may enhance our understanding of these fields and stipulate perceptually relevant ways to address the subsequent degradation.

This paper proposes a research framework for perceptual assessment of automotive audio, targeting the evaluation of both the properties of the reproduction system and the acoustical properties of the car's interior (e.g., cabin). First, a brief literature review is presented in Sec. 1 where past and current protocols are discussed, their limitations are identified, and potential improvements are proposed. A new experimental method is then presented (Sec. 2) including a pilot verification experiment (Sec. 3) where the initial results are shown. Several remarks of the method including limitations and future work are discussed (Sec. 4), followed by the concluding remarks of the study (Sec. 5).

1 BACKGROUND

1.1 Perceptual Assessment in Automotive

The aim of perceptual assessment is to investigate the relationships between measurable physical quantities and the perceived sensations. In the domain of audio these protocols comprise three major components: the *acquisition* of the signal to be investigated, its *presentation* to human assessors, and the *evaluation* protocol where assessors' responses are collected. The principles of audio assessment protocols are summarized in Fig. 1. Typically, each experimental protocol follows the given contextual factors and limitations of the study in a way that the physical data collected during acquisition and the perceptual data collected during evaluation could be combined and a relationship could be established.

Audio research has demonstrated a wealth of protocols for assessing audio material that one could categorize into two major groups: the in-situ and the laboratory-based methods. In-situ evaluations describe the protocols where the assessor evaluates the perceived sound in the intended natural settings, e.g., listening to an orchestra in a concert hall. In the laboratory-based methods, the signals are captured via measurements or simulations, and a presentation scheme is followed to recreate the sound field in a neutral environment, e.g., in a laboratory via headphones or loudspeakers.

In the following section the literature on perceptual assessment of automotive audio will be examined following the three basic elements of audio assessment—the acquisition, presentation, and evaluation. The analysis aims to identify the limitations of current practices and stipulate new approaches that would overcome such constraints.

1.2 Acquisition and Presentation

1.2.1 In-Situ Methods

Considering the complex acoustics and the multi-sensory experience of a car cabin, the in-situ evaluation is the most straightforward approach. In fact, the automotive industry has been conducting in-situ evaluations since the early days of car audio. It was soon realized though that exposing assessors in a car cabin introduces strong biases caused by non-acoustical factors, e.g., size, price, brand, interior materials [11, 21, 22], that are likely to affect assessors' judgments. Shively [14] proposed a blind in-situ procedure where the non-acoustic feedback of both the interior and the exterior of the car was highly controlled. Later, a "placebo" [23] method was introduced aiming to force assessors into evaluating stimuli in random phases, under an in-situ sighted protocol.

A major shortcoming of in-situ assessment protocols is the restricted ability of conducting instant and double-blind comparative evaluations [21]. Cecchi et al. [24] addressed

this by proposing a cabin-based apparatus that allowed the evaluation of signal processing algorithms under simultaneous and controlled comparative protocols, yet within the natural environment. However, the method cannot be extended in investigations where physical alterations are in question, e.g., cabin acoustics or loudspeaker placement.

One could argue that several car audio systems could be compared via in-situ methods, as far as the contextual factors are identical. Yet, the access to prototype cars is limited and such approach has not been followed [12, 13]. Instead, Olive [18] proposed the use of a reference listening scenario in a room, as a common comparison baseline against automotive audio systems.

Nonetheless, these protocols inherently include long test-to-test periods between the different cars or system settings. Based on the restricted auditory memory [25], it is likely that in-situ evaluations may influence the experimental results in an uncontrolled manner, for example following the assessor's mood and expectations [20, 26]. Moreover, comparing multiple systems or cars would require repeated and sequential experiments, hence, longer experimental times.

1.2.2 Laboratory-Based Methods

The direct response to the constraints of in-situ protocols was the development of laboratory-based methods. Laboratory-based methods aim to impose higher control on both auditory and non-auditory parameters, while ensuring repeatability and scientific validity. This is normally achieved by providing standardized source and signals, fixed listening environments and settings, and simulations or measurements to capture the car's sound field.

Although utilizing acoustical simulations in automotive audio assessment is highly anticipated [27, 28], current methods are not able to adequately characterize the cabin's acoustical field [29]. Automotive audio has therefore focused on measurement-based methods. The majority of these methods require the acquisition of the sound field in a car cabin using microphones, typically employing dummy-head recordings. The obtained signals are then used to reproduce the binaural field over headphones. An extension of the method has been also proposed, the so-called Binaural Car Scanning (BCS) [11–13, 21, 30]. BCS allows for natural head-movements during evaluation by dynamically updating the appropriate measurement angle, based on the assessor's head-orientation [31] during the evaluation. BCS overcomes the practical limitations of in-situ evaluations and allows simultaneous comparisons between car cabins or audio systems in a simple and repeatable framework. It therefore remains a valuable tool for car audio assessment, especially when the investigation focuses on a quantification of perceived intensity differences between systems.

The fundamental disadvantages of binaural techniques include the acoustical modification of the acquired signals due to the physical properties of the dummy-head and the inherent necessity of headphone-based playback. That is, the Head-Related Transfer Function (HRTF) of the dummy-head used during acquisition may not fit the anthropometric properties of the human assessor during evaluation. This reduces the degree of externalization [32] of the sound source, leading to what is known as in-the-head perception of sound. This unnatural effect limits the degree of true perception of the investigated field [33]. Hegarty et al. [12] have shown that several spatial alterations in car audio systems were interpreted as timbral, even by expert assessors, questioning the validity of BCS when the spatial properties of the field are investigated.

Moreover, the physical constraints of headphone transducers limit the extent on which low frequencies could be reproduced accurately. This results in timbral and level differences between the acquired and the presented field [34]. Also, it restricts "whole-body" vibrations, a sensation known to influence the listener preferences in binaural reproduction of rooms as well as car cabins [35, 36].

Obtaining Impulse Responses (IRs) with a dummy-head imposes several practical limitations. The analysis of the captured signal is restricted due to the embedded HRTFs, increasing the difficulty of applying standard analysis metrics. Moreover, the requirement for repeated measurements at different dummy-head orientations yields hundreds of measurements per source. Such acquisition may last several hours for a stereophonic setup. Even when it is desired to capture single transducer IRs, the thermal effects of the transducers, and the cabin, would make it rather difficult to perform accurate repeated measurements suitable for BCS. Therefore, BCS methods of individually measured transducers are not followed.

1.3 Evaluation—Experimental Design

The final stage of a perceptual assessment comprises the evaluation of the stimuli. Typically it follows an experimental procedure where assessors' evoked sensations are quantified. The majority of audio evaluation protocols lies upon the indirect collection of human responses in verbal, graphical, or written form [37].

Many studies on automotive audio focused on the evaluation of the global quality of the sound field, i.e., using Basic Audio Quality (BAQ) protocols and Preference scales [38]. Later, studies expanded on parametric evaluations, where the specific sound characteristics that influence the assessors' preference were sought, by evaluating audio material based on perceptual attributes. Initially these investigations adopted attributes from other domains of audio [23, 39, 40], followed by the implementation of elicitation techniques [41, 42] where a number perceptual attributes for car audio systems were identified [10, 12, 43].

1.4 Summary and Motivation

By examining the literature one could realize that automotive audio assessment protocols balance on a trade-off between the requirements of high ecological validity and direct, single-dimensional control of the contextual factors and parameters.

In car cabins, the complex sound field requires stimuli acquisition in the form of measurements, or in-situ evaluation under real conditions. When assessing car audio in-situ, a number of non-auditory features introduce many

restrictions during evaluation. Similar to other streams of audio evaluation [20], following a laboratory-based approach where a captured signal is recreated in neutral conditions seems to provide flexible and well-controlled experimental designs. For automotive audio, such a method would allow the assessment of several automotive audio systems, cabins, and their properties in a comparative double-blind protocol. These benefits can be considered superior to the in-situ methods, which rely on assumptions of flawless auditory memory.

Evaluating the general satisfaction of human perception is useful for benchmarking processes, yet, the specific sound characteristics that influence the subject's preference remain unknown. Using verbal descriptors, known as perceptual attributes, one could focus on singular auditory percepts of the sound field, avoiding cognitive biases relating to a human's expectations, preference, and sentiments [37]. As car audio systems may deliver novel experiences to the human ear and specific to the car environment, such techniques would improve our understanding of the perceived field by providing sensory information normally hidden behind hedonic and affective judgments.

It is therefore desirable to develop new experimental designs that would:

1. Be able to evaluate both the physical and the electrical properties of automotive audio systems, e.g., electroacoustical modifications and acoustical properties of the car's interior, in simultaneous, double-blind comparative protocol;
2. Provide means for physical quantification of such alterations;
3. Employ acquisition and presentation schemes where both the spatial and timbral characteristics of the field are preserved and provide flexibility to commonly used rendering schemes;
4. Allow the evaluation of uncommon and novel sound experience; still, meet the practical limitations of the automotive environment, i.e., fast, flexible, and efficient in both acquisition and evaluation phases.

2 EXPERIMENTAL METHOD

This section describes the implementation of a new methodology for acquiring, presenting, and evaluating automotive sound. It is the paper's intention to provide a general research protocol for the automotive audio, where the limitations identified in Sec. 1.4 are addressed. Here, several steps were taken towards a more adaptable and flexible perceptual assessment method, by incorporating new approaches in both the acquisition and presentation of the sound field, as well as the evaluation procedures.

The acquisition and presentation stages of the method are based on the Spatial Decomposition Method (SDM) [44, 45]. SDM is a spatial analysis and synthesis scheme where the IRs obtained by a compact microphone array are analyzed parametrically, in terms of instantaneous pressure, time, and direction of arrival. Applying SDM in cars may introduce three major advantages over the previously discussed methods. First, the spatiotemporal analysis of the SDM data may enable a better understanding of the behavior of the cabin's sound field. The additional physical quantities and visualization capabilities of SDM [46] could be used as a physical metric when the perceived spatial properties are investigated. Second, it allows reproduction of the analyzed sound field over loudspeakers, addressing several issues of headphone-based playback. The spatial responses can be then reproduced with any rendering scheme, e.g., Vector Based Amplitude Panning (VBAP) and High Order Ambisonics (HOA). Last but not least, SDM synthesis makes use of a single omnidirectional pressure microphone rather than beamforming or directional processing techniques. That is, the pressure used to synthesize the field originates from a single omnidirectional microphone, whereas the DOA calculation uses all six microphones on the array. In consequence, the reproduced sounds are not altered (e.g., colored) by the characteristics of the receiver, as commonly encountered when using directional microphones and dummy-head apparatus.

The requirements identified for the evaluation procedures for automotive audio assessment, seem to depict the need of Descriptive Sensory Analysis (DA) [41] techniques. DA combines the sensory characterization of the stimuli and the quantitative rating of the associated perceptual attributes within the same framework. Such methods have been successfully applied in audio, e.g., in concert halls [47, 48], spatial audio reproduction through loudspeakers [30, 42, 49], and headphones [50], hearing aids [51], and active noise cancellation [52]. DA encompasses attribute elicitation methods where human assessors are able to epitomize and appropriately quantify their sensations for the given set of stimuli, by defining their own perceptual attributes. Therefore, it allows the perceptual assessment of novel experiences, without the need of a priori quantification of the evaluation attributes. That is, the requirement of the experimenter to pre-select possible attributes to be used by assessors as scales during a parametric evaluation.

Although such techniques seem to suit the needs of automotive audio assessment, they require extensive training per product, as well as multiple sessions per assessor [41]. The time restrictions within the automotive environment limit the use of common DA techniques, even if their outcome would be ideal. Addressing this time limitation, one could employ the recently developed Rapid Sensory Profiling techniques [26]. In this paper Flash Profile (FP) [53, 54], the most closely related rapid method to conventional DA profiling [26], is adapted for audio evaluation within the automotive environment. FP allows the listener to quickly elicit new and non-limited attributes, which is a significant advantage compared to lengthy consensus attribute elicitation techniques [41] or fixed attribute lists (e.g., [2]) that may not reveal the full perceptual experience of the presented stimuli.

The methods and experimental procedure in the following sections serve as an example of the proposed framework. The following sections provide the details of each stage of the method, followed by a pilot experiment, for

reasons of completion, as an implementation example of the proposed procedure.

2.1 Acquisition—In-Situ Car Measurements

In order to obtain the acoustic characteristics of a sound reproduction system in a car cabin, in-situ recordings are required. For this study measurements were conducted in a four-door sedan equipped with 17 band-limited transducers (5 tweeters, 7 mid-range transducers, 4 woofers, 1 subwoofer) and a multichannel automotive amplifier. The feed to the individual transducers was post-processed (i.e., compensation delays, equalization) to represent a typical production car equipped with a tuned, premium sound system.

An open spherical microphone array (G.R.A.S VI-50) comprising two coincidental microphones on each axis, separated by 25 mm, was positioned at the driver's seat, at the average seating position [55]. The microphone probe was aligned to match the position of a dummy-head seating in the car—the center point of the head and ears' height. The distance between the microphone array and the headrest was set to 15 cm.

The IRs were measured in a way that the (electrical) input to the amplifier was measured by the electrical output of the microphones in the cabin. This type of measurements is referred to as Vehicle Impulse Response (VIR).

The VIRs were measured for each transducer, using a 5 s logarithmic sine-sweep method [56] at 192 kHz sampling rate using an RME UCX multichannel sound interface. The measurements were performed at 82 dB (C-weighted RMS) estimated using the forward facing microphone of the array, with system default settings. The electrical output of the measurement system was kept constant for all transducers. The car measurements were conducted in a temperature and noise regulated garage at Bang & Olufsen's premises.

2.2 Presentation

2.2.1 Spatial Analysis and Synthesis of VIRs

The spatial analysis and synthesis procedures followed in the current paper are described in detail in a recent report [45], where the applicability of the SDM in cars was discussed and physically evaluated. SDM divides the sound field into spatially discrete elements of a preset analysis time window. As the native SDM assumes a wide-band source, i.e., a typical full-range loudspeaker, it is normally recommended to use as short a window as possible [44]. In this experiment the captured VIRs were band-limited, due to the type/size of each transducer in the cabin (Sec. 2.1). Hence it was possible to implement a custom-length (L) analysis window based on the properties of the transducers. For each transducer type, L was set to span three periods of the shortest wavelength in the reproduced frequency band. This allows a more accurate spatial decomposition of band-limited near-field sources—an important advantage when analyzing such complex sound fields as in car cabins.

2.2.2 Reproduction Protocol

SDM provides a spatial analysis and signal encoding for a given set of VIRs. The resulting data allows auralization of the sound field using a given spatial rendering scheme over loudspeakers, or Binaural Synthesis based on anechoic HRTFs. In this paper the synthesis of the SDM-encoded spatial IRs was implemented using the Nearest Neighbor (NN) loudspeaker approach, similar to [57].

The performance of the system when NN is employed is highly benefited by a physical arrangement where the placement of the loudspeakers is based on the spatial analysis of the sound fields under investigation; automotive audio in this case.

2.2.3 Reproduction Setup

For new types of synthesized acoustic environments, designing the reproduction loudspeaker array can benefit from the spatial analysis with SDM. This approach allows the experimenter to better understand the structure of the original sound field and design an optimal reproduction layout. For this study individual analysis of the captured VIR was employed to ensure that the direct sound as well as reflections from the cabin's surfaces are preserved in the best possible way during the reproduction phase. This included the aforementioned VIRs in addition to a database of 20 car cabins and system types.

The analysis followed a systematic comparison of both objective and perceptual metrics. The spatiotemporal energy distribution in time intervals [46] of each measured VIR was combined with the corresponding weighted energy error estimation. This error term results from assigning instantaneous VIR pressure to the NN loudspeakers instead of their absolute position given by the SDM analysis. The perceptual assessment focused on three perceptual constructs—spectral fidelity, temporal integrity, and accuracy of spatial representation—for two auralization sets: a full car audio system and single transducers (see [45]). Attention was also given to the electroacoustic properties of the loudspeakers and the spatial acuity of the human hearing system, ensuring a high level of detail in the frontal plane while maintaining the perceptual qualities from all directions.

For the final auralizations a 40.3 loudspeaker system was specified (see Fig. 2). The setup comprises of 40 full-range (Genelec 8020C) and 3 subwoofers (Genelec 7050B). One could note a second layer of loudspeakers at lower elevation (-10°) in the frontal plane (-70° to $+70^\circ$). Naturally, the direct sound paths in the car cabin originate from lower elevations, as typically no transducers are placed at the ear level, and this was also evident during the spatial analysis of the VIR.

The magnitude responses of the loudspeakers in the reproduction array were confirmed in-situ, to lie primarily within ± 1.5 dB; including a low-frequency compensation (<200 Hz). Each loudspeaker was level-matched at the listening position, within ± 0.5 dB (C-weighted RMS), using 5 s pink-noise. It is noted that the inherent inconsistencies in the physical placement of the loudspeakers introduce

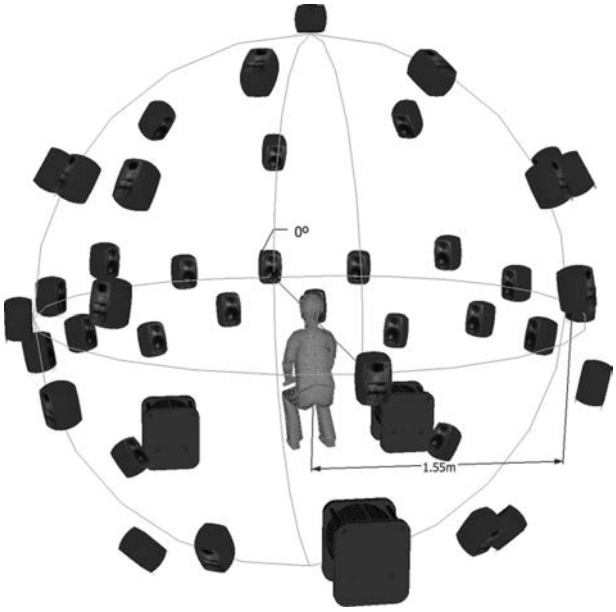


Fig. 2 Diagram of the reproduction loudspeaker positions. The loudspeakers are placed on a spherical surface with a 1.55 m radius. The positions are left-right symmetric and their positions (azimuth, elevation) are: 0,0; 11,–10; 22,0; 32.5,–15; 45,5; 55,–10; 65,0; 75,–10; 90,0; 120,–10; 135,10; 0,30; 40,40; 90,30; 115,30; 135,30; 150,55; 0,90; 55,–40; 120,–45; 150,–35.

different times of arrival at the listening position. It was therefore required to ensure that the acoustic delay between any loudspeaker and a microphone at the listening position was temporally matched.

2.2.4 Reproduction Environment

When assessing spatial audio over loudspeakers it is necessary to limit the acoustic influence of the reproduction room on the reproduced sound field that is intended to be evaluated [37] as it is known to be perceptible by listeners [58, 59]. This is normally achieved by ensuring that the reproduction room is characterized by a lower reverberation time compared to the room that is being reproduced. Due to the nature of the sound field in a car cabin and the very short reverberation time [29], the experimental setup used in the implementation of the current method was installed in an anechoic chamber (B5-104) located at Aalborg University. The chamber is designed and constructed to host simulation setups with human occupancy, and it is treated with absorption wedges that are 0.4 m long. Its free inner dimensions are 5.0 × 4.5 × 4.0 m. The chamber meets the requirements for anechoic performance [60] down to 200 Hz. The experimental apparatus was covered with absorption material to eliminate any reflections from the structural installation.

2.2.5 Visual Influence

When assessing virtual acoustics, it is important to understand and address the cross-modal behavior of the human brain. The relative importance between information within each modality (i.e., vision and hearing) is known to alter our perceptual processing [61]. This has been recognized in room acoustics research [62] in both in-situ and virtual acoustics assessment [63]. In a mismatch between visual and auditory cues, it has been argued that visual information may dominate the auditory sense [64]. In virtual spaces this was found to be a crucial component of user acceptance [65].

In this study we aim to reproduce a highly complex sound field that is naturally unique to car cabins under laboratory settings. To limit possible cross-modal biases, such as visual influences, a number of steps should be followed. These controls should aim to reduce the visual influence of the experimental apparatus to the reproduced sound field.

2.3 Evaluation—Flash Profile

The experimental design follows the principles of Flash Profile (FP). FP comprises two parts: (1) the elicitation of attributes and (2) the ranking of the elicited attributes.

2.3.1 Flash Profile—Elicitation

During elicitation the assessor is asked to provide a non-limited number of verbal descriptors that capture the perceptual characteristics of the stimuli set under investigation. It is vital that the assessor focuses on this procedure and define verbal descriptors that are singular, non-hedonic, and scalable, while they do not exhibit redundancy [37, 41].

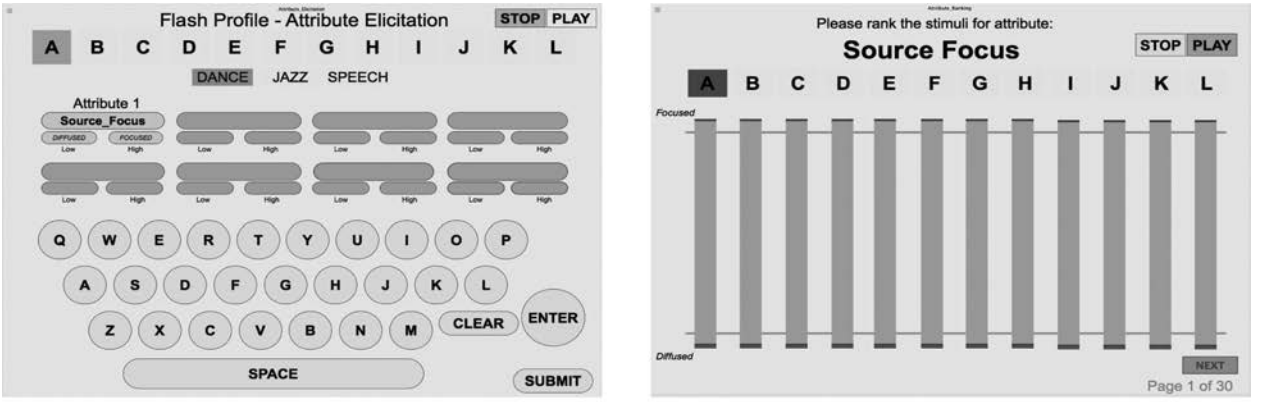
FP requires the whole stimuli set to be available to the assessor at all times. However, when assessing audio material the evaluation of acoustic conditions requires an excitation signal (program), i.e., music, speech, or noise [37], as each type of signal excites the conditions differently. As a consequence the stimulus is a product of a Program type (i.e., music) and a Condition (i.e., timbral alteration). To address this issue, it is recommended that during elicitation, the assessor has the option to change the program material while the order of the presented stimuli is maintained. This would allow assessors to explore subtle perceptual differences between specific conditions over a variety of programs in the minimum time possible. Thus, the interface (Fig. 3a) includes anonymously labeled stimuli (e.g., A-L) and the available program material, which assessors could select at any point during this phase. The order of presentation should be randomized between assessors.

2.3.2 Flash Profile—Ranking

During the ranking phase the assessor is asked to order the stimuli presented based on the perceived intensity differences of the given attribute. Each given perceptual attribute forms a block of n trials (one trial per program material). At this phase, the presentation order of the stimuli, as well as the program material, should be randomized on each trial following typical audio evaluation conventions [37]. The graphical interface is shown in Fig. 3b.

3 PILOT EXPERIMENT

The section above described a method for perceptual assessment of automotive audio in terms of acquiring,



(a) Interface during Elicitation (b) Interface during Ranking

Fig. 3 User interface during the two phases of the experiment. Buttons labeled with letters A-L provide switching between signals. The whole experimental procedure is self-paced and self-controlled.

analyzing, reproducing, and evaluating the sound field. For reasons of completion, a brief description of a pilot study using the proposed method is included. The experiment investigated the ability of the method to assess automotive audio over basic alterations as well as the influence of the assessor's expertise on the experimental procedure.

The example below demonstrates the perceptual effects of modifying the acoustical field in a car cabin when altering the DSP settings as well as the physical properties of the cabin. Three system configurations were presented to the assessors: audio system without DSP processing (*No Equalization*), audio system with DSP processing (*Reference*), and audio system with DSP processing while the front side windows were open (*Front Windows Open*). The DSP processing included equalization of the transducers, delays, and individual tuning of the magnitude responses set by an automotive sound engineer (tonmeister). The condition that included DSP processing and no physical modification of the cabin, labeled as reference, serves as the baseline and represents a premium production car audio system.

3.1 Materials and Apparatus

For this pilot experiment the car measurements were processed with the method described in Sec. 2.1.

Music material was then convolved with the corresponding 40.3-channel SDM responses. Here, only one program is included (*Armin van Buuren feat. Ana Criado - I'll Listen*), for simplicity. In a complete study multiple program types should be included. The playback was based on multichannel 24-bit PCM reproduction sampled at 48 kHz.

The assessor was given a tablet (iPad) controlling MAX 7 via MIRA remote interface, shown in Fig. 3. The reproduction room, setup, and calibration measurements were identical to the aforementioned settings in Sec. 2.

As mentioned in Sec. 2.2.5, when reproducing sound fields in the laboratory, the visual influences of the experimental apparatus should be addressed. Here, the experiment was conducted in dark conditions and any acoustic feedback of the space was controlled. In addition, no information was given to assessors about the experimental room, the loudspeaker setup, and the content of the stimuli as recommended [26].

3.1.1 Assessors

Six assessors participated in the pilot experiment: two expert assessors, two experienced assessors, and two naive assessors [66]. Expert assessors ($As_{1,2}$) had more than 15 years of experience ($Mean = 17.5$) in acoustical development and critical listening; with the last 10 years focusing on automotive audio systems. Experienced assessors ($As_{3,4}$) had between 3 to 6 years of experience ($Mean = 4.5$) in audio evaluation and acoustical research but no experience in automotive audio. Naive assessors ($As_{5,6}$) reported no prior experience with audio evaluation or technical knowledge on the subject. All assessors were male, aged between 26–43 ($Mean = 32$, $s.d = 6.2$). The assessors' hearing sensitivity was confirmed with standard procedures [67] to be above 20 dBHL at 125–8000 Hz.

In the case where assessors were not familiar with verbal elicitation procedures, an introduction was given and the assessors performed verbal elicitation for a set of visual stimuli.

3.2 Results

Fig. 4 presents the perceptual responses, i.e., the sensory profiles, of all assessors for the three acoustical conditions used in the evaluation. Overall, it can be seen that all assessors identified both timbral effects and spatial differences between the presented stimuli and created similar profiles. Remarkable timbral differences can be seen between the conditions of no equalization and reference. Similarly, alterations of spatial properties were identified for the condition where the windows were open. Opening the front windows revealed less prominent timbral effects, compared to no equalization, yet, they were identified and rated accordingly by most assessors.

One should note that in FP, the elicited individual vocabulary of each assessor is used as given by each assessor, in contrast with consensus methods were a common

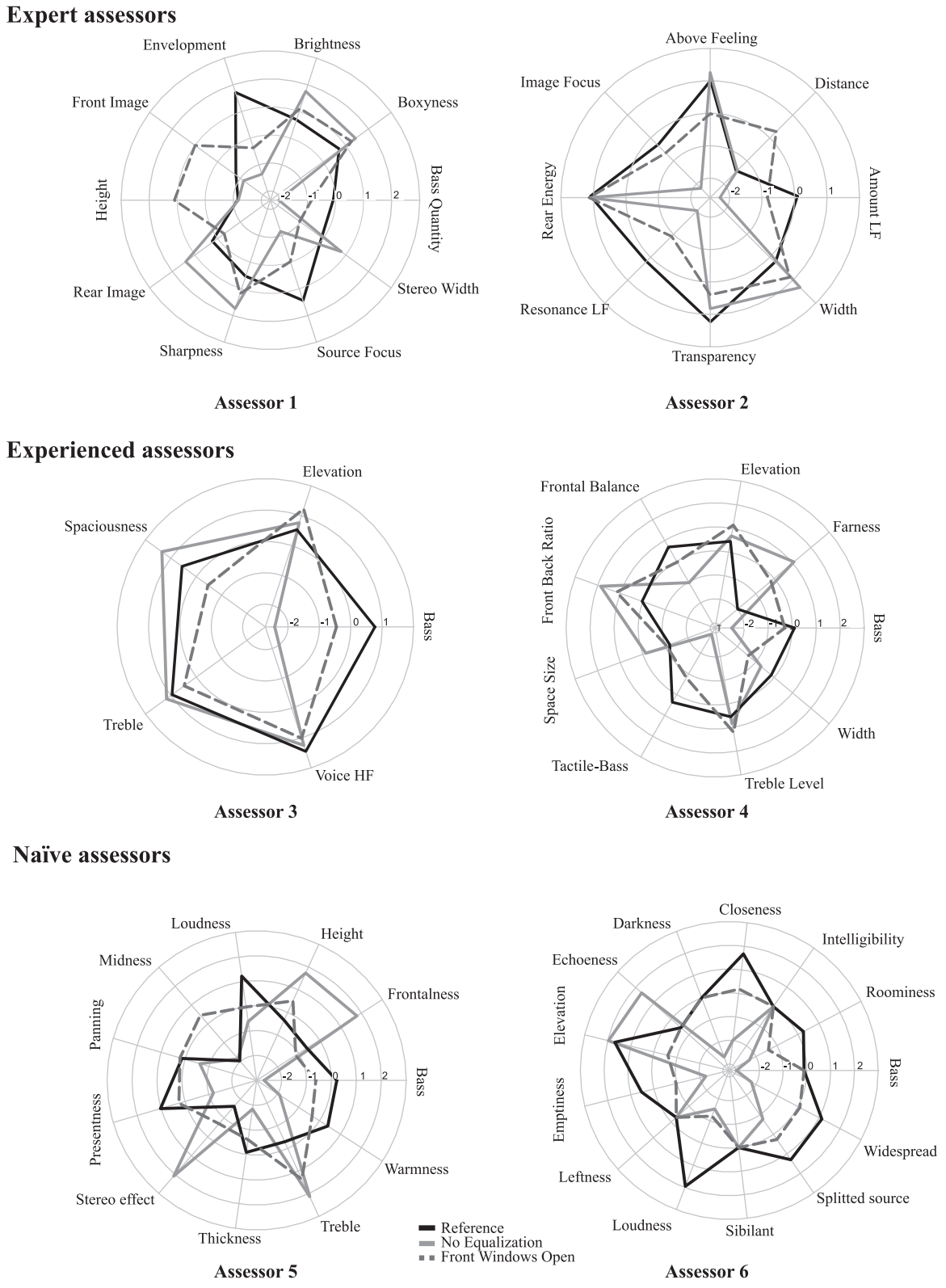


Fig. 4 Sensory profiles depicting the assessors' responses for the three acoustical conditions. To limit scaling effects between the assessors and allow comparisons the raw data were centered and normalized (*Mean* = 0). The attributes' definitions are given in Table 2.

vocabulary is defined and used by all assessors. Thus, the semantic meaning of each attribute used in FP is unique to each assessor. It is expected that the scaling, the extreme intensities (scale anchors), and the perceptual constructs underlying the ratings may differ across individuals. That makes the direct comparison across assessors a non trivial matter, even when the same label was used [68]. To limit possible ambiguities of attributes' labels and identify relationships between descriptors given by different assessors, each assessor provided definitions and anchors for their

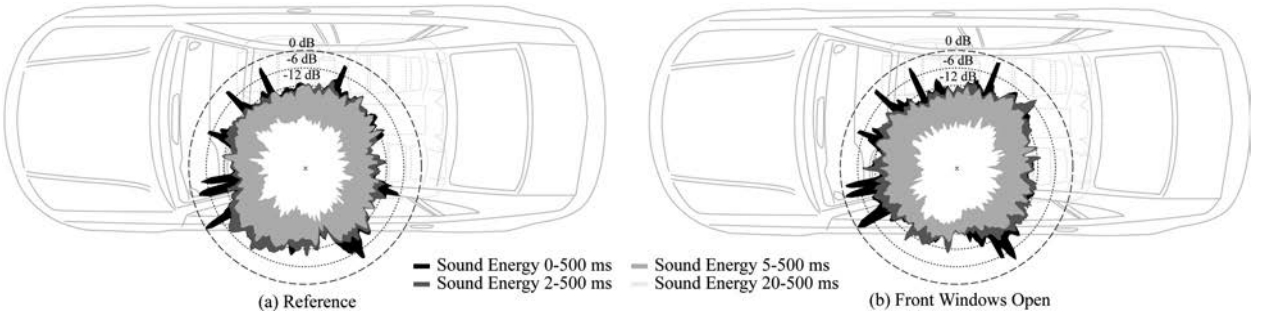


Fig. 5 Spatiotemporal visualization [46] based on the directional energy response of the system for Reference (a) and Front Windows Open settings (b).

attributes. The definitions are summarized in the Appendix (Table 2).

Here, a limited number of attributes is selected for comparison, where their definitions and anchors indicate strong relationships across assessors and tangible relation to physical quantities was possible (i.e., low frequency alterations). This would allow a basic comparison of individually elicited attributes across multiple assessors in order to provide holistic insights within this exploratory investigation.

As shown in Fig. 4, no equalization condition seem to strongly affect the perceived bass content, indicating lower intensities for *Bass Quantity*$_{\text{As1}}$, *Amount LF*$_{\text{As2}}$, *Resonance LF*$_{\text{As2}}$, *Bass*$_{\text{As3,4,5,6}}$, and *Tactile Bass*$_{\text{As4}}$, compared to the reference. Less prominent alterations at low frequencies were identified by assessors for front windows open. All assessors, independent of training experience, have equally identified these differences at the appropriate intensity levels. The perceptual data come in agreement with the known physical changes between these two conditions, shown in Fig. 6; compared to reference, no equalization condition included differences up to 20 dB, while front windows open condition was limited to 3 dB in that frequency range.

Further, assessors indicated that no equalization has a negative effect on attributes related to the feeling of sound arriving from multiple directions e.g., *Envelopment*$_{\text{As1}}$, *Spaciousness*$_{\text{As3}}$, and the ability to locate the sound source, given by *Source Focus*$_{\text{As1}}$, *Image Focus*$_{\text{As2}}$, and *Frontal Balance*$_{\text{As4}}$. The source's positioning and focus was also indicated as alteration of the perceived width, e.g., on *Stereo Width*$_{\text{As1}}$, *Width*$_{\text{As2}}$, *Stereo Effect*$_{\text{As5}}$, *Splitted Source*$_{\text{As6}}$, and *Widespread*$_{\text{As6}}$. Moreover, assessors identified noticeable differences for attributes relating to the ratio between the energy coming from the front and back directions such as *Rear Image*$_{\text{As1}}$, *Front Back Ratio*$_{\text{As4}}$, and *Frontalness*$_{\text{As5}}$. The above perceived differences across the assessors indicate spectral and level imbalances in the no equalization condition, compared to that of the reference. The identified differences are among the common characteristics that sound engineers aim to improve using DSP processing as in the presented reference condition. One should expect such perceptual effects to be altered in the absence of DSP processing.

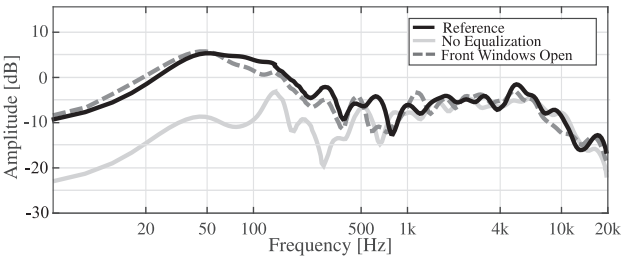


Fig. 6 The total magnitude response of the car audio system at the listening position for (1) Reference, (2) No Equalization, and (3) Front Windows Open conditions.

Based on previous investigations (see [20]), altering the front side windows is likely to affect perceptual attributes relating to Apparent Source Width and Distance. Due to changes in the volumetric properties of the cabin [13], the low frequency behavior should also change. Fig. 5 shows the spatiotemporal analysis of reference condition and when the front windows were opened. It can be seen that the energy originating from sides ($\pm 60^{\circ}$) is decreased when the windows are open for the first three time intervals. It also noted that the overall reproduction symmetry of the system would be affected based on the depicted energy distribution after the first 20 ms.

The perceived differences (Fig. 4) seem to depict similar observations, as opening the front windows reveals high intensity on spatial attributes such as the perceived *Front Image*$_{\text{As6}}$, *Stereo Width*$_{\text{As1}}$, and *Width*$_{\text{As2,4}}$ compared to the reference condition. Timbral effects have been also noted by most assessors indicating decreased bass content. Interestingly, although opening the windows shows changes on the perceived *Height*$_{\text{As1}}$, *Above Feeling*$_{\text{As2}}$, and *Distance*$_{\text{As2}}$, removing the equalization reveals no such differences. This follows the current understanding, as the perceived distance is known to relate to lateral reflections [20], and such differences could not be easily elicited by simple equalization settings.

Table 1 reports the explained variances per dimension for each assessor, using Principal Component Analysis (PCA) [70]. The more trained assessors were able to identify the perceptual differences and rate them with attributes that support both dimensions (Dim. 1 $\approx$ 65%, Dim. 2 $\approx$ 35%). The judgments of less trained assessors were primarily explained by the first dimension ($\approx$ 85%), and much lower

Table 1 Results of Principal Component Analysis for each assessor. Correlation Ratio (CR) refers to the percentage of the well correlated attributes (R >|0.5|) to a dimension, noted as (C), divided by the total number of given attributes of that assessor (T), i.e., $\frac{C}{T} \times 100$.

As	Dimension 1	Dimension 2	CR $_{\text{Dim.1}}$	CR $_{\text{Dim.2}}$
1	67.03	32.97	80	50
2	61.88	38.12	75	50
3	70.74	29.26	80	40
4	68.83	31.17	89	44
5	84.60	15.40	91	9
6	81.90	18.10	69	15

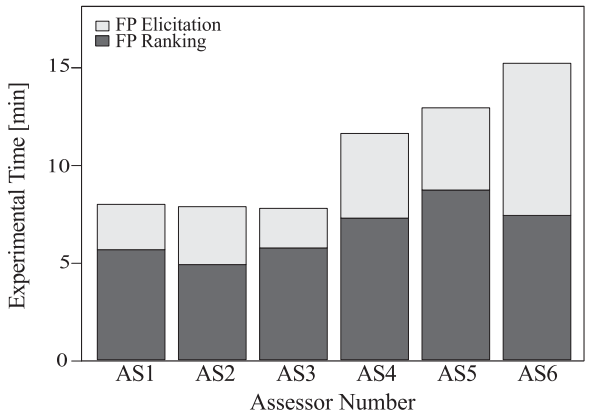


Fig. 7 Time required for each assessor to perform the elicitation and ranking phases of Flash Profile.

second dimension ($\approx 15\%$). To further analyze this finding the correlation of the attributes to the identified dimensions was also calculated, given as the Correlation Ratio (CR). This metric effectively indicates the quality of the given attributes and the ability of the assessors to quantify and differentiate the perceived sensations in the multidimensional space. The CR for the second dimension is marginally lower for naive assessors than the experts and the experienced. That denotes the enhanced ability of the trained assessors to conduct the experiment reliably, with well supported attributes, as noted before [54].

In terms of the assessors efficiency, the current observations suggest that the more trained assessors were able to perform the experiment in less time as shown in Fig. 7. Experts spent less time during the elicitation procedure compared to the less trained assessors. This confirms the advantage of using trained assessors in such protocols, due to their enhanced ability to efficiently identify and verbalize perceived differences, in the auditory domain.

Overall, the current results indicate that the assessors perceived the physical changes in an expected way and they are in close agreement with previous elicitation studies in automotive audio [12] and spatial audio reproduction [20]. It should be noted that the data presented here is an illustrative set to initially assess the method proposed and should not be used to conclude findings about car acoustics, due to the limited contextual factors. An in-depth analysis of car acoustics and the experimental setup should follow this preliminary investigation to further validate and establish the experimental protocol applied.

4 DISCUSSION

Although experiencing automotive sound in-situ is unequivocally a true reference to reality, in practice it could be employed in a limited number of scientific investigations, e.g., comparing different equalization and signal processing algorithms within a cabin. Expanding in-situ protocols to investigations of physical parameters (i.e., loudspeaker placement, cabin acoustics) is impractical and highly laborious. Even if made possible, it would inherently include long test-to-test periods, contrasting the basic requirements of perceptual evaluation of audio material.

Here, a laboratory-based protocol was proposed. SDM was employed for the acquisition and presentation of automotive audio. This approach maintains the benefits of perceptual assessment in the laboratory similar to the aforementioned methods (Sec. 1.2.2). It therefore allows instant, double-blind, and comparative assessment of different sound fields in the laboratory. Yet, it overcomes issues related to non-auditory feedback (i.e., brand) during in-situ evaluation and the limitations imposed from headphone-based playback of BCS.

Dealing with perception requires to employ a scientific evaluation method where the perceptual properties are quantified, while it avoids uncontrolled biases on human's sentiments and judgments, such as individual's taste, beliefs, experience, and needs [37, 69]. Descriptive Sensory Analysis (DA) methodologies, found in the food and wine industry [41], have been instrumental in decoding such complex perceptual constructs. Still, they are laborious and require long experimental procedures. Here, we propose the use of FP, a rapid sensory analysis technique, developed to provide a sensory profiling similar to common DA, still at the least time possible (e.g., typically within a 1–3 h session).

Multiple limitations exist within automotive audio evaluation that FP seem to overcome. First, the time restrictions imposed will be well addressed when using FP, yet the benefits of conducting DA are preserved. Moreover its need of only four-to-five (expert) assessors with general sensory expertise, instead of higher numbers of product-trained assessors [26], its comparative ranking nature that does not require audio reference, as well as the flexibility of the method may well suit the processes within automotive audio.

In fact, the comparative nature of FP combined with direct access to the entire stimuli set simultaneously allows the assessors to adapt their cognitive strategies during the evaluation. When performing FP, the assessors may employ both comparative and short memory-based judgments. That results to a higher number of direct comparisons than traditional forced-choice comparison methods [26, 53]. It also permits the evaluation of very similar, or highly dissimilar stimuli within the same procedure, with no statistical disadvantage, as common repeated measures protocols [37].

In the pilot study presented the performance of experts, experienced, and naive assessors was investigated. The results indicate that experts completed the experimental procedure faster (Fig. 7), providing precise and technically familiar attributes that correspond well to the subsequent ratings of the stimuli (Table 1) as well as to their definitions (Table 2).

The quality and number of non-correlated attributes is a vital requirement for a fast and efficient profiling during FP [26]. Based on the current findings, one could argue that naive listeners failed to meet this requirement, as they provided a high number of inter-correlated attributes, some of which did not contribute to the explained variances, as no differences were identified. The two experts provided clear results even for physically small differences, e.g., the low frequency alteration between reference and front windows open. These alterations were identified also by experienced and naive listeners but not in a consistent manner, as noted in previous investigations of FP in food products [54].

Moreover, the nature of FP highly benefits from a common background of auditory sensitivity across assessors, so that the perceptual constructs driving the ratings could be used in a common factorial space [26]. This is an important element when performing multitable quantitative analysis, as the typical FP statistical analysis processes. Here, the results illustrate similarities between assessors of similar experience level but not as well across them. The PCA analysis (Table 1) indicates that although all assessors adequately performed the task, a strong common statistical relationship across all assessors might be difficult to achieve. It is therefore recommended that further investigations should consider assessors' background when establishing an evaluation panel for FP.

Finally, FP is based on quantitative description. Here, a simple data analysis was presented for completeness. Although such an analysis is informative, the real benefit of FP is realized when it is combined with multivariate statistics, such as Multiple Factor Analysis [71]. This type of analysis allows the investigation of both the individually elicited attributes as well as the given ratings in a common factorial space for all assessors. Effectively, FP merges quantitative and qualitative data using a mathematical approach rather than subjective analysis of the experimenter [70].

4.1 Limitations and Future work

One should note that the SDM provides a faithful and plausible acoustical representation, however, as any spatial reproduction method to date, it has certain limitations. It was shown recently [45] that a post-equalization of the analyzed response is needed when SDM is applied to cars due to high echo density. There it was also shown that the complex geometry of a car cabin and the extreme acoustical conditions may violate the basic assumptions of SDM of plane waves. As an Impulse Response (IR) -based method, several aspects of the field are not captured during acquisition, i.e., the non-linearities or structure-bone vibrations, as well as the acoustical effects of the human body. Thus, one should not expect that the reproduced sound field is an exact physical replica of the real field.

Nevertheless, both objective and perceptual results suggest that SDM preserves the basic perceptual differences between the stimuli set in the pilot experiment. The results of the pilot study support the capacity of the method to address both timbral and spatial properties of the sound fields in cars. Further advantages of the method include the flexible reproduction scheme, the fast acquisition of VIR compared to BCS, as well as the analysis and novel visualization capabilities of the spatial properties of the sound field.

Rapid Sensory Analysis methods aim to remove tedious processes within DA, such as building a consensus vocabulary of well-defined descriptors across assessors, as well as several training and evaluation sessions. As a direct consequence, when employing FP the experimenter cannot argue easily on the semantic meaning of the descriptors. Moreover, FP gains experimental time by not including repeats or hidden anchors. For audio material, this is a serious limitation as it is well known that the context and properties of the program material influence the results [37]. In fact critical listeners often require specific program material for assessing certain attributes (e.g., speech content for intelligibility ratings). Therefore, when conducting FP of audio it is strongly recommended to employ several program materials.

Further work on this topic includes the perceptual assessment of automotive audio systems, and car cabins in detail, aiming to identify the perceptual constructs originating from acoustical alterations in the sound field in question. Additional validation studies of the proposed method should be conducted, for example by comparing the results to in-situ evaluation when possible, as well as contrasting FP to common DA procedures with identical contextual factors. Such approaches would improve the understanding of the method and further validate its applicability in audio evaluation.

Here, the reproduction of the auralized stimuli is conducted over loudspeakers in an anechoic chamber. It is noted that the presentation and evaluation method could be altered to meet further research objectives and the related practical implications. For example, the presentation on headphones is still possible, as well as following standardized audio evaluation methods. Yet, the spatiotemporal analysis of the measured field will still be available. Moreover, the experimental methodology proposed here could be applied in several domains, for example in assessing room acoustics. Future work could perceptually assess a variety of acoustical settings in standard everyday rooms, to better understand the influence of the acoustical properties of a reproduction room that are inherently imposed on the reproduced sound field.

5 CONCLUDING REMARKS

This paper reviewed the past and current practices for perceptual assessment of automotive audio and stipulated new approaches and research paths to address the industry's

future demands. A novel method was described in terms of acquiring, presenting, and evaluating the reproduced sound of automotive systems, targeting a general framework for perceptual studies in automotive audio. Finally, a pilot experiment was presented and the preliminary results were discussed.

The method applied SDM for capturing, analyzing, and presenting the sound field to human assessors. The current results indicate that this approach yields a faithful representation of timbral and spatial properties of automotive sound, while providing novel analysis tools. Employing this method may allow several properties of automotive audio to be assessed in controlled and instantaneous comparative protocols, including both acoustical, as well as electrical changes over the same experimental paradigm.

Conducting the evaluation using rapid sensory analysis such as Flash Profile, allows the assessors to use their own vocabulary to describe and quantify the auditory sensations within a single experimental method. Thus, such approach may permit a more detailed assessment of unique and novel experiences in cars, e.g., 3D sound reproduction and upmixing, noise masking schemes, and individual sound zones within cars.

6 ACKNOWLEDGMENTS

The authors would like to thank Morten Lydolf, Martin Møller, Martin Olsen, Claus Vestergaard Skipper, and their colleagues at Aalborg University for their support and helpful input. The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA-2012-316969.

7 REFERENCES

[1] F. Rumsey, "Automotive Audio," *J. Audio Eng. Soc.*, vol. 60, pp. 1070–1074 (2012 Dec.).

[2] IEC, "60268-13 (BS 6840-13) Sound System equipment—Part 13: Listening Tests on Loudspeakers," International Electrotechnical Commission (1998).

[3] ITU-R Rec., "BS 1534-1: Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," International Telecommunication Union Recommendations, Geneva, Switzerland (2001).

[4] M. Strauss, J. Nowak, and D. de Vries, "Approach to Sound Field Analysis and Simulation Inside a Car Cabin," presented at the *AES 36th International Conference: Automotive Audio* (2009 Jun.), conference paper 10.

[5] M. Kleiner and C. Lindgren, "Objective Characterization of Audio Sound Fields in Automotive Spaces," presented at the *AES 15th International Conference: Audio, Acoustics & Small Spaces* (1998 Oct.), conference paper 15-007.

[6] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations," *J. Audio Eng. Soc.*, vol. 44, pp. 835–849 (1996 Oct.).

[7] ISO, "3382-1: Acoustics Measurement of Room Acoustic Parameters Part 1: Performance Spaces," International Organization for Standardization, Geneva, Switzerland (2009).

[8] C. Choi, L.-H. Kim, S. Doo, Y. Oh, and K.-M. Sung, "Assessment of Sound Field in a Car," presented at the *113rd Convention of the Audio Engineering Society* (2002 Oct.), convention paper 5701.

[9] L. Klapproth, "Acoustic Characteristics of the Vehicle Environment," presented at the *77th Convention of the Audio Engineering Society* (1985 Mar.), convention paper 2185.

[10] G. Martin and S. Bech, "Attribute Identification and Quantification in Automotive Audio—Part 1: Introduction to the Descriptive Analysis Technique," presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6360.

[11] S. Bech, M.-A. Gulbol, G. Martin, J. Ghani, and W. Ellermeier, "A Listening Test System for Automotive Audio Part 2—Initial Verification," presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6359.

[12] P. Hegarty, S. Choisel, and S. Bech, "A Listening Test System for Automotive Audio—Part 3: Comparison of Attribute Ratings Made in a Vehicle with Those Made Using an Auralization System," presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), convention paper 7224.

[13] F. Christensen, G. Martin, M. Lydolf, M. Pauli, S. Woo-Keun, and P. Benjamin, "A Listening Test System for Automotive Audio—Part 1: System Description," presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6358.

[14] R. Shively, "Subjective Evaluation of Reproduced Sound in Automotive Spaces," presented at the *AES 15th International Conference: Audio, Acoustics & Small Spaces* (1998 Oct.), conference paper 15-011.

[15] N. Otto, "Listening Test Methods for Automotive Sound Quality," presented at the *103rd Convention of the Audio Engineering Society* (1997 Sep.), convention paper 4586.

[16] A. Azzali, A. Farina, G. Rovai, G. Boreanaz, and G. Irato, "Construction of a Car Stereo Audio Quality," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6306.

[17] E. Ugolotti, G. Gobbi, and A. Farina, "IPA—A Subjective Assessment Method of Sound Quality of Car Sound Systems," presented at the *110th Convention of the Audio Engineering Society* (2001 Oct.), convention paper 5281.

[18] S. E. Olive, "A New Reference Listening Room for Consumer, Professional and Automotive Audio Research," presented at the *126th Convention of the Audio Engineering Society* (2009 May), convention paper 7677.

[19] W. N. House, "Aspects of the Vehicle Listening Environment," presented at the *87th Convention of the Audio Engineering Society* (1989 Oct.), convention paper 2873.

[20] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of Reverberation in Small Rooms: A Literature Study," presented at the *AES 55th International*

Conference: Spatial Audio (2014 Aug.), conference paper P-3.

[21] S. E. Olive and T. Welti, "Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems," presented at the *AES 36th International Conference: Automotive Audio* (2009 Jun.), conference paper 7.

[22] F. E. Toole and S. E. Olive, "Hearing Is Believing vs. Believing Is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things," presented at the *97th Convention of the Audio Engineering Society* (1994 Nov.), convention paper 3894.

[23] N. House and R. Shively, "The Placebo Method: A Comparison of the In-Situ Subjective Evaluation Methods for Vehicles," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5136.

[24] S. Cecchi, A. Primavera, F. Piazza, F. Bettarelli, E. Ciavattini, R. Toppi, J. G. F. Coutinho, W. Luk, C. Pilato, F. Ferrandi, V. M. Sima, and K. Bertels, "The hArtes CarLab: A New Approach to Advanced Algorithms Development for Automotive Audio," *J. Audio Eng. Soc.*, vol. 59, pp. 858–869 (2011 Nov.).

[25] M. Sams, R. Hari, J. Rif, and J. Knuutila, "The Human Auditory Sensory Memory Trace Persists about 10 Sec: Neuromagnetic Evidence," *J. Cogn. Neurosci.*, vol. 5, no. 3, pp. 363–368 (1993).

[26] J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (Elsevier, 2014).

[27] A. J. Svobodnik and G. Ruiz, "Virtual Development of Mercedes Premium Audio Systems," presented at the *AES 36th International Conference: Automotive Audio* (2009 Jun.), conference paper 21.

[28] N. Harris, "A Comparison of Modelling Techniques for Small Acoustic Spaces such as Car Cabins," presented at the *122nd Convention of the Audio Engineering Society* (2007 May), convention paper 7146.

[29] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations," *J. Audio Eng. Soc.*, vol. 44, pp. 835–849 (1996 Oct.).

[30] S. Choisel, P. Hegarty, F. Christensen, B. Pedersen, and W. Ellermeier, "A Listening Test System for Automotive Audio—Part 4: Comparison of Attribute Ratings Made by Expert and Non-Expert Listeners," presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), convention paper 7225.

[31] P. Mackensen, U. Felderhof, G. Theile, U. Horbach, and R. Pellegrini, "Binaural Room Scanning—A New Tool for Acoustic and Psychoacoustic Research," *J. Acous. Soc. Am.*, vol. 105, no. 2, pp. 1343 (1999). doi: 10.1121/1.426373

[32] A. Roginska, G. H. Wakefield, and T. Santoro, "User Selected HRTFs: Reduced Complexity and Improved Perception," in *Proc. Undersea Hum. Syst. Integr. Symp.*, Providence, IL (2010).

[33] J. Blauert, *The Technology of Binaural Listening*, J. Blauert, Ed. (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013).

[34] W. Rudmose, "The Case of the Missing 6 dB," *J. Acous. Soc. Am.*, vol. 71, no. 3, p. 650 (1982). doi: 10.1121/1.387540

[35] W. L. Martens, H. Sakanashi, and W. Woszczyk, "Whole-Body Vibration Associated with Low-Frequency Audio Reproduction Influences Preferred Equalization," presented at the *AES 36th International Conference: Automotive Audio* (2009 Jun.), conference paper 16.

[36] S. Germain, S. E. Olive, and T. Welti, "The Effect of Whole-Body Vibrations on Preferred Bass Equalization of Automotive Audio Systems," presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), convention paper 7956.

[37] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application* (Wiley, 2007).

[38] P. Dennis, "In-Vehicle Audio System Sound Quality Preference Study," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9393.

[39] A. Farina and E. Ugolotti, "Subjective Comparison of Different Car Audio Systems by the Auralization Technique," presented at the *103rd Convention of the Audio Engineering Society* (1997 Sep.), convention paper 4587.

[40] D. Clark, "Listening Technology for Autmotive Sound Systems," presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), convention paper 5723.

[41] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices* (New York: Springer, 1999).

[42] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Elicitation of Attributes for the Evaluation of Audio-on-Audio Interference," *J. Acous. Soc. Am.*, vol. 136, no. 5, pp. 2630–2641 (2014). doi: 10.1121/1.4898053

[43] A. Melka and P. Pelant, "Psychoacoustic Methods Applied to SKODA Passenger Car Development Programmes: Intentions and First Experiences," *Acta Acust. united with Acust.*, vol. 85, pp. 657–660 (1999).

[44] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, pp. 17–28 (2013 Jan./Feb.).

[45] 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*, vol. 63, pp. 914–925 (2015 Nov.).

[46] J. Pätynen, S. Tervo, and T. Lokki, "Analysis of Concert Hall Acoustics via Visualizations of Time-Frequency and Spatiotemporal Responses." *J. Acous. Soc. Am.*, vol. 133, no. 2, pp. 842–57 (2013). doi: 10.1121/1.4770260

[47] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling Preference Ratings of Concert Hall Acoustics Using Subjective Sensory Profiles," *J. Acous. Soc. Am.*, vol. 132, no. 5, pp. 3148–3161 (2012).

[48] A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki, "Relationships between Preference Ratings, Sensory Profiles, and Acoustical Measurements in Concert Halls." *J. Acous. Soc. Am.*, vol. 135, no. 1, pp. 239–250 (2014). doi: 10.1121/1.4836335

[49] J. Berg and F. Rumsey, "Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods," presented at the *AES 16th International Conference: Spatial Sound Reproduction* (1999 Mar.), conference paper 16-005.

[50] G. Lorho, "Perceived Quality Evaluation—An Application to Sound Reproduction over Headphones," Ph.D. dissertation, Aalto University (2010).

[51] C. Simonsen and S. Legarth, "A Procedure for Sound Quality Evaluation of Hearing Aids," *Hear. Rev.*, vol. 17, no. 13, pp. 32–37 (2010).

[52] N. Zacharov, J. Ramsgaard, G. Le Ray, and C. V. Jorgensen, "The Multidimensional Characterization of Active Noise Cancelation Headphone Perception," *Second Int. Work. Qual. Multimed. Exp.* (2010), pp. 130–135.

[53] 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.*, vol. 67, no. 2, pp. 826–834 (2002). doi: 10.1111/j.1365-2621.2002.tb10685.

[54] J. Delarue and J. M. Sieffermann, "Sensory Mapping Using Flash Profile. Comparison with a Conventional Descriptive Method for the Evaluation of the Flavour of Fruit Dairy Products," *Food Qual. Prefer.*, vol. 15, no. 4, pp. 383–392 (2004). doi: 10.1016/S0950-3293(03)00085-5.

[55] S. Parkin, G. M. Mackay, and A. Cooper, "How Drivers Sit in Cars," *Accid. Anal. Prev.*, vol. 27, no. 6, pp. 777–783 (1995).

[56] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5093.

[57] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, "Preferences of Critical Listening Environments among Sound Engineers," *J. Audio Eng. Soc.*, vol. 62, pp. 300–314 (2014 May).

[58] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *J. Audio Eng. Soc*., vol. 50, pp. 651–666 (2002 Sep.).

[59] F. E. Toole, *Sound Reproduction: Loudspeakers and Rooms* (Boston: Focal Press, 2008).

[60] ISO, "3745 - Determination of Sound Power Levels and Sound Energy Levels of Noise Sources Using Sound Pressure. Precision Methods for Anechoic Rooms and Hemi-Anechoic Rooms," International Organization for Standardization (2003).

[61] B. E. Stein and M. A. Meredith, *The Merging of the Senses* (Boston: MIT Press, 1993).

[62] A. Defays, S. Safin, A. Billon, C. Decaestecker, N. Warzée, P. Leclercq, and A.-S. Nyssen, "Bimodal Interaction: The Role of Visual Information in Performing Acoustic Assessment in Architecture," *Ergon. Open J.*, vol. 7, pp. 13–20 (2014).

[63] C. Nathanail, C. Lavandier, J.-D. Polack, and O. Warusfel, "Influence of Sensory Interactions between Vision and Audition on the Perceptual Characterization of Room Acoustics," *J. Acous. Soc. Am.*, vol. 105, no. 2, p. 1046 (1999). doi: 10.1121/1.424982

[64] P. Larsson and D. Västfjäll, "When What You Hear Is What You See: Presence and Auditory-Visual Integration in Virtual Environments," *Proc. 10th Annu. Int. Work. Presence*, Barcelona, Spain (2007), pp. 11–18.

[65] P. Svensson, "On Reverberation Enhancement in Auditoria," Ph.D. dissertation, Chalmers University of Technology (1994).

[66] ITU-R Rec., "BS 2300-0: Methods for Assessor Screening," International Telecommunication Union, Geneva, Switzerland (2014).

[67] ISO, "8253-1 Audiometric Test Methods Part 1: Pure-Tone Air and Bone Conduction," International Organization for Standardization (2010).

[68] J. Pagès, *Multiple Factor Analysis by Example Using R* (Boca Raton, FL: CRC Press, 2014).

[69] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers, "Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction," *Comput. Graph. Forum*, vol. 31, no. 1, pp. 102–131 (2012). doi: 10.1111/j.1467-8659.2011.02086.x

[70] S. Lê and T. Worch, *Analyzing Sensory Data with R* (Boca Raton, FL: CRC Press, 2015).

[71] H. Abdi, L. J. Williams, and D. Valentin, "Multiple Factor Analysis: Principal Component Analysis for Multitable and Multiblock Data Sets," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 5, no. 2, pp. 149–179 (2013). doi: 10.1002/wics.1246

8 APPENDIX

Table 2 Individually elicited attributes, including the provided anchors and definitions. The attributes shown were elicited using a larger stimuli set.

Attribute	Low anchor	High anchor	Definition
Bass Quantity$_{As1}$	too little	too much	level of bass content
Boxyness$_{As1}$	boxy	open	some sound boxy, in a very small space
Brightness$_{As1}$	dull	bright	how bright is the sound (high frequencies)
Envelopment$_{As1}$	low	high	how enveloping is the sound field, sound from around
Front Image$_{As1}$	side	front	if the frontal image is exactly in front of me
Height$_{As1}$	absent	present	sound appears higher points
Rear Image$_{As1}$	absent	present	sometimes sound appears from the back
Sharpness$_{As1}$	blunt	sharp	sharp/transients at mid-high freq.
Source Focus$_{As1}$	diffuse	focused	sourced located at a point?
Stereo Width$_{As1}$	narrow	wide	how wide the sound image is
Above Feeling$_{As2}$	low	high	feeling that sound comes from above me
Amount LF$_{As2}$	none	a lot	the low frequency energy levels
Distance$_{As2}$	near	far	how distant the image feels
Image Focus$_{As2}$	muffled	focused	image clarity / clearly at a location
Rear Energy$_{As2}$	low	high	the amount of sound energy coming from the back
Resonant LF$_{As2}$	not resonant	resonant	specific low frequency resonance
Transparency$_{As2}$	not transparent	transparent	how transparent the sound is
Width$_{As2}$	narrow	wide	the width of the sound image
Bass $_{As3}$	low	high	how much bass there is (spectrum)
Elevation$_{As3}$	low	high	if the sound appears higher than 0°
Spaciousness$_{As3}$	frontal	lateral	if the sound comes from front or multiple lateral directions
Treble$_{As3}$	low	high	how much treble there is (spectrum)
Voice HF$_{As3}$	muffled	bright	if the vocal is bright, as present
Bass$_{As4}$	low	high	bass levels
Farness$_{As4}$	near	far	distance - the source is not as present
Elevation$_{As4}$	lower	higher	sound perceived as higher (location) as others
Front Back Ratio$_{As4}$	too back	too front	the level (balance) of the sound coming from back and front
Frontal Balance$_{As4}$	too left	too right	the level (balance) of the stereo image
Space Size$_{As4}$	small	big	how big does the room that I am in, feels
Tactile Bass$_{As4}$	not there	there	a feeling of low frequency pressure or body vibration
Treble Level$_{As4}$	little	too much	the high frequency content level
Width$_{As4}$	toowide	focused	the horizontal size of the ensemble
Bass$_{As5}$	a little	a lot	how much bass I hear
Frontalness$_{As5}$	front	back	balance of front-back sound
Height$_{As5}$	little	a lot	sound from above, higher than others
Loudness $_{As5}$	low	high	level differences
Midness$_{As5}$	side	mid	sound from centre (front) or sides
Panning $_{As5}$	left	right	imbalance of left-right sound
Presentness$_{As5}$	not present	present	presence, prominent sound
Stereo Effect$_{As5}$	a little	a lot	sound from sides, like wide stereo
Thickness$_{As5}$	thin	thick	feeling of thick, full sound
Treble$_{As5}$	low	high	higher frequencies level
Warmness $_{As5}$	cold	warm	feeling for sound being warm
Bass$_{As6}$	low	high	bass level
Closeness$_{As6}$	not close	close	does the sound feel close to me?
Darkness$_{As6}$	bright	dark	dark (frequencies) sounds or not
Echoeness$_{As6}$	not much	too much	if I can hear echo
Elevation$_{As6}$	ok	high	playback from higher positions
Emptiness$_{As6}$	empty	full	the room sounds empty
Intelligibility$_{As6}$	not	ok	quality of voice to understand, like bad codec
Leftness$_{As6}$	too left	center	more sound from left
Loudness$_{As6}$	lower	higher	the level of the music
Room Print$_{As6}$	there	not there	how loud is the room in sound
Sibilant$_{As6}$	ok	too much	‘s’ or ‘z’ sound in speech
Splitted source$_{As6}$	split	one point	sound from many directions or not
Widespread$_{As6}$	around	front	sound spread around

THE AUTHORS


Neofytos Kaplanis


Søren Bech


Sakari Tervo


Jukka Pätynen


Tapio Lokki


Toon van Waterschoot


Søren Holdt Jensen

Neofytos Kaplanis holds a B.Mus. Tonmeister (Surrey, UK) and a M.Sc. in auditory neuroscience (London, UK). He has been an R&D acoustic engineer at HARMAN Automotive and a visiting researcher at University of London (2013), KU Leuven (2014), and Aalto University (2015). In 2013 he joined Bang & Olufsen as an acoustic research fellow, pursuing a Ph.D. (EE) from Aalborg University. His research interest centers upon the auditory human perception, aiming to merge physical phenomena with their underlying cognitive and perceptual properties.

•

Søren Bech received a M.Sc. and a Ph.D. from the Department of Acoustic Technology (AT) of the Technical University of Denmark. From 1982 to 1992 he was a research Fellow at AT studying perception and evaluation of reproduced sound in small rooms. In 1992 he joined Bang & Olufsen where he is currently Head of Research. In 2011 he was appointed Professor in audio perception at Aalborg University. His research interest includes human perception of reproduced sound in small and medium sized rooms. experimental procedures, and statistical analysis of data from sensory analysis of audio and video quality. General perception of sound in small rooms is also a major research interest.

•

Sakari Tervo was born in Kuopio, Finland, in 1983. He received a M.Sc. degree in audio signal processing from the Tampere University of Technology in 2006 and a D.Sc. degree in the field of acoustic signal processing from Aalto University in 2012. He has been a visiting researcher in Philips Research, in The Netherlands, in 2007, and in the University of York in 2010. Currently he works as a postdoctoral researcher in the Department of Computer Science in Aalto University and researches spatial sound analysis and reproduction.

•

Jukka Pätynen was born in Espoo, Finland, in 1981. He received a M.Sc. and a D.Sc. (Tech.) degrees from the Helsinki University of Technology, Finland, in 2007, and Aalto University, Finland, in 2011, respectively. He is currently working as an Academy post-doctoral researcher in the Department of Computer Science in Aalto University. His research activities include room and concert hall acoustics, musical acoustics, and signal processing.

•

Prof. Tapio Lokki was born in Helsinki, Finland, in 1971. He has studied acoustics, audio signal processing, and computer science at the Helsinki University of Technology (TKK) and received an M.Sc. degree in electrical engineering in 1997 and a D.Sc. (Tech.) degree in computer science and engineering in 2002. At present Dr. Lokki is an Associate Professor (tenured) with the Department of Computer Science at Aalto University, where he leads the virtual acoustics team jointly with Prof. Lauri Savioja. The passion of Prof. Lokki is to understand how rooms modify sound that we hear. To pursue the encompassing understanding of room acoustics, his team is investigating auralization, spatial sound reproduction, binaural technology, and novel objective and subjective evaluation methods, as well as physically-based room acoustics modeling methods.

•

Toon van Waterschoot received M.Sc. (2001) and Ph.D. (2009) degrees in electrical engineering, both from KU Leuven, Belgium. He is currently a tenure-track Assistant Professor at KU Leuven, Belgium. He has previously held teaching and research positions with the Antwerp Maritime Academy, the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT), and the Research Foundation - Flanders (FWO) in Belgium, with Delft University of Technology in The Netherlands, and with the University of Lugano in Switzerland. His research interests are in acoustic signal enhancement, acoustic modeling, audio analysis, and audio reproduction. He is currently serving as an Associate Editor for the *Journal of the Audio Engineering Society* (AES) and for the *EURASIP Journal on Audio, Music, and Speech Processing*, and has been a Guest Editor for *Signal Processing*. He is a Nominated Officer of the European Association for Signal Processing (EURASIP) and a member of the IEEE Audio and Acoustic Signal Processing Technical Committee (AASP-TC). He has been serving as an Area Chair for Speech Processing at the European Signal Processing Conference (EUSIPCO 2010, 2013–2015) and as General Chair of the 60th AES Conference in Leuven, Belgium (2016).

•

Søren Holdt Jensen received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd., Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNI C), Lyngby; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. He is Full Professor and heading a research team working in the area of numerical algorithms, optimization, and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications. Prof. Jensen was an Associate Editor for the *IEEE Transactions on Signal Processing*, Elsevier *Signal Processing* and *EURASIP Journal on Advances in Signal Processing*, and is currently Associate Editor for the *IEEE Transactions on Audio, Speech and Language Processing*. He is a recipient of a European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and the IEEE Denmark Sections Signal Processing Chapter. He is member of the Danish Academy of Technical Sciences and was in January 2011 appointed as member of the Danish Council for Independent ResearchTechnology and Production Sciences by the Danish Minister for Science, Technology and Innovation.