ASA POMA Proceedings of Meetings on Acoustics

Volume 58 http://acousticalsociety.org/

International Symposium on Musical and Room Acoustics

Loyola University
New Orleans, Louisiana

24-28 May 2025

Architectural Acoustics

Comparison of the auditory and auditory-visual perception in four concert halls

Jonas Heck
Institute for Hearing Technology and Acoustics, RWTH Aachen University, Aachen, North Rhine-Westfalia, 52074, GERMANY; jonas.heck@akustik.rwth-aachen.de

Josep Llorca-Bofí
Fraunhofer-Institut für Bauphysik IBP Stuttgart, Baden-Württemberg, GERMANY; josep.llorca-bofi@ibp.fraunhofer.de

Hyeju Park
RWTH Aachen University, Aachen, North Rhine-Westfalia, GERMANY; hyeju.park@rwth-aachen.de

Nils Meyer-Kahlen and Tapio Lokki
Department of Information and Communications Engineering, Helsinki, Uusimaa, FINLAND; nils.meyer-kahlen@aalto.fi; tapio.lokki@aalto.fi

Michael Vorländer
Institute for Hearing Technology and Acoustics, RWTH Aachen University, North Rhine-Westfalia, Aachen, GERMANY; mvo@akustik.rwth-aachen.de

An interesting research question is still to what extent the visual impression of a room influences the auditory perception. Subjective evaluations have been collected for four concert halls based on Individual Vocabulary Profiling (IVP) and MUSHRA-like ratings. The experiment was conducted with two groups, each comprising 20 participants. One group performed the auditory-visual experiment (AV) with 360-degree photographs of the halls being presented in a head-mounted display. The other group performed a blind test (B) evaluating the stimuli purely auditorily without seeing the hall. The acoustic stimuli were identical and reproduced over head-tracked headphones for both groups. The stimuli consisted of anechoic orchestra recordings that have been convolved with measured spatial impulse responses from the concert halls under study. The aggregated comparative ratings allow to study the differences induced by the type of presentation, i.e. auditory versus auditory-visual. For this purpose, a multiple factor analysis has been performed that highlights the dominant perceptual dimensions. The results show that differences between the two presentation types can be mainly seen in the timbre, intimacy, reverberation and spatial perception.

Proceedings of Meetings on Acoustics, Vol. 58, 015014 (2025); https://doi.org/10.1121/2.0002179.

1. INTRODUCTION

While the experience of in-situ listening of concerts presumably leads to more reliable and realistic perceptions than listening using virtual acoustics technologies, the comparison between concert halls is practically challenging and lacks robustness. Therefore, laboratory studies have been established to instantaneously switch between different concert halls and listener locations while listening to the same music piece [1]. Pätynen et al. [2] have introduced spatial impulse response measurements to be able to reproduce the spatial characteristics of the orchestra as a sound source better and to allow for spatial reproduction at the listener/receiver location. This allows a detailed comparison but ignores the influence of vision on the auditory perception.

The aim of this study is to overcome the limitation of a missing visual stimulus by presenting not only acoustic but also visual stimuli. Thus, an experiment has been performed to test the impact of the visual information of the listening position in a hall and to compare the purely auditory evaluation with the auditory-visual perception. The individual vocabulary profiling (IVP) technique has been shown to deliver interesting insights in perceptual dimensions in earlier studies [3]. To do so, we ran an IVP experiment for purely auditory presentation, as done earlier, as well as for audiovisual presentation. The visual information was expected to lead to an emphasis of different perceptual dimensions compared to blind assessments. As in earlier studies, hierarchical multi-factor analysis (HMFA) was used to evaluate the collected data. To ensure good immersion, the test subjects assessed the concert halls in a virtual reality (VR) environment with a head-mounted display, which – unlike a 2D screen – a full panoramic view.

2. SELECTED HALLS

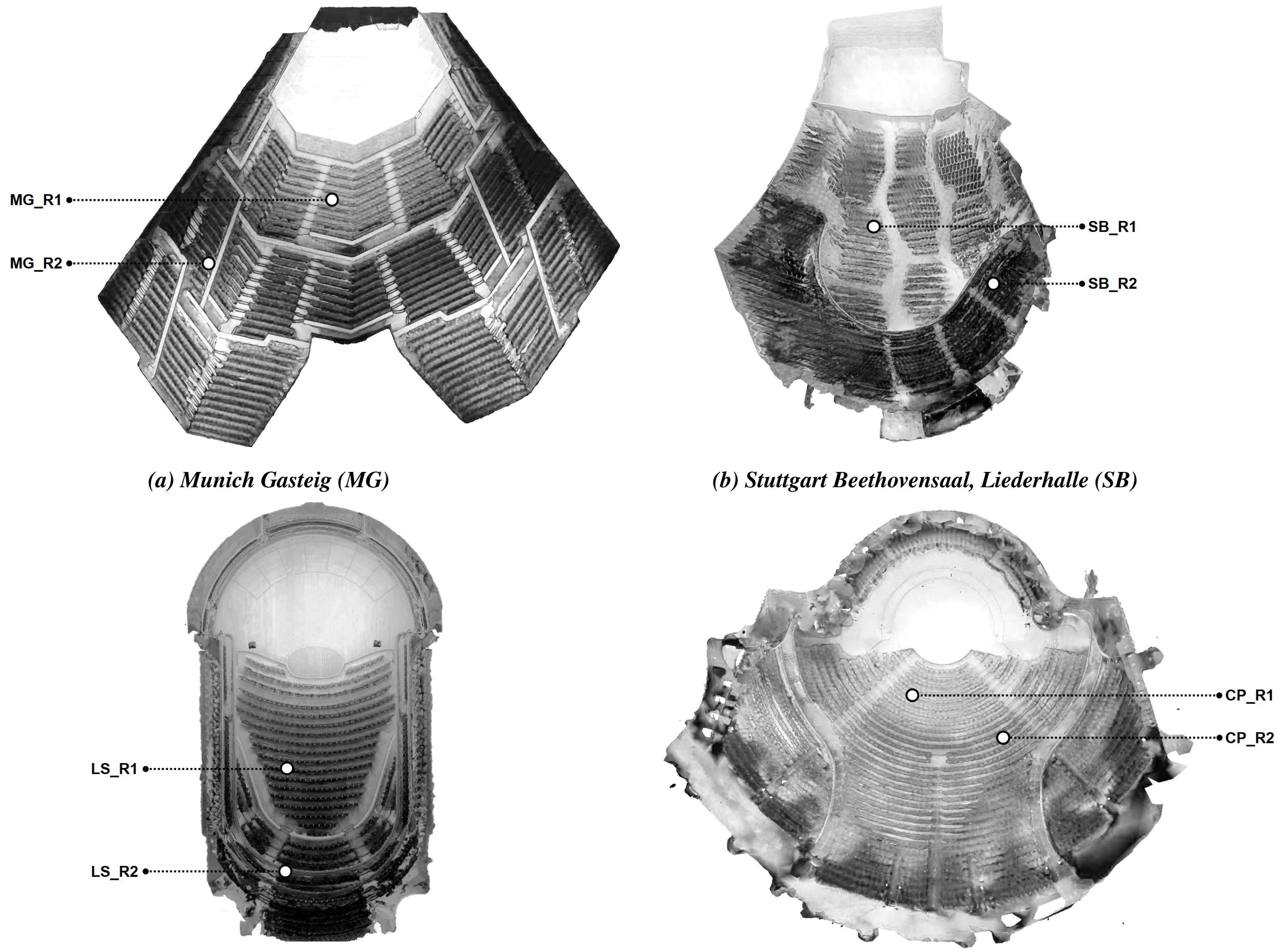
Four halls were selected for the study. The selection was made to ensure variability in both acoustic features and visual appearance. However, the interior design and decoration were not supposed to appear historical. Instead, the halls were chosen to follow a modern aesthetic language as this is assumed to be more relevant regarding the design of new halls. The four halls were the Cologne Philharmonie (CP), the Philharmonic hall in Munich's Gasteig (MG), Lahti Sibelius Hall (LS), and Beethovensaal in Stuttgart's Liederhalle (SB).

As shown in Fig. 1, two receiver locations (= listener positions) have been selected in each hall for evaluation. Receiver 1 (R1) is located closer to the stage while receiver 2 (R2) is always further back. This selection of two receiver locations per hall should allow for some variance in the listener perception not only between halls but also within a room.

A. ACOUSTIC CAPTURE

Spatial impulse responses provide the basis for the analysis in this study. The impulse responses have been measured in the selected halls during a measurement campaign by Lokki et al. in August 2010 (LS) and November 2012 (MG, SB, CP). The loudspeaker orchestra introduced by Pätynen et al. [2] has been used to represent distributed acoustic sources on stage, mimicking the instrument placement of a symphonic orchestra. As receiver, six omnidirectional microphones (a GRAS 50VI intensity probe) have been used to measure spatial room impulse responses. The resulting room impulse response database contains transfer paths between each emitter and each receiver point.

For the purpose of this study, the measured spatial room impulse responses were postprocessed according to the spatial decomposition method. The required directional analysis was performed using a time difference of arrival estimation [4]. Then, a fourth-order Ambisonics B-format room impulse response was synthesized, as in Frank and Zotter [5].



(a) Munich Gasteig (MG)

(b) Stuttgart Beethovensaal, Liederhalle (SB)

(c) Lahti Sibelius Hall (LS)

(d) Cologne Philharmonie (CP)

Figure 1: The selected concert halls in photogrammetry-based top view. In the experiment, 360-degree pictures were used rather than photogrammetry-generated models.

B. VISUAL CAPTURE

For the visual representation of the listener positions, each room has been captured with 360-degree photographs. For this purpose, a Kandao Obsidian S camera was placed on the seat of each listener location under study. The camera position should correspond approximately to the head height and the resulting viewing perspective of an average adult sitting on the chairs in each of the concert halls. Thus, the camera with its six lenses, distributed equally on an horizontal plane, were positioned to 80 cm above the seat level. Due to different chair designs, the camera height from floor level varied within and across rooms between 105 and 120 cm.

During image capture, the camera settings were kept close to a reference setting as follows: ISO = 200; shutter time = 1/2 sec; aperture = F/2.4; temperature = 3800. However, several photographs were taken with slight variations to the reference setting in order to fit the actual dynamic range. While ISO as well as aperture values were fixed, shutter time was adjusted because this parameter had a major impact on the pictures' brightness, without directly affecting the colors [6, 7]. The shutter speed was always below 1/50 sec to avoid sudden changes in color temperature due to flickering lights.

In order to adjust the small differences between photographs, a reference picture was selected for each room that all the other pictures of that room were matched with. The color matching postprocessing included the changes of luminance, color range and neutralization of color cast. For this purpose, the Color Match tool from Adobe Photoshop was used, which performs very similarly to the approach of Reinhard et al. [8] as it maintains the mean and standard deviation of the color distribution of the reference image [9]. When the photographs of the camera's six individual lenses were stitched together to one 360-degree picture, the files were stored as monoscopic pictures in .jpg format. The postprocessed pictures have an image size of 8760×2880 pixels. For the display in the virtual reality environment via head-mounted display, the bitrate was increased from 24 to 32 and the resolution was upscaled to 15360×8640 pixels. Subsequently, it is saved in .hdr format. A database containing the photographs of the receiver locations used here as well as further additional positions in the four concert halls can be found at https://zenodo.org/doi/10.5281/zenodo.15409975 [10].

3. EXPERIMENT

A. EXPERIMENTAL PROCEDURE

The experiment followed a procedure that is known as Individual Vocabulary Profiling (IVP). The technique has been developed from the Flash profile and has been applied to concert hall evaluation [3]. In our case, it consisted of the following elements:

- Introduction to the experiment, demographics questionnaire,
- 1st phase of the experiment consisting of the stimuli exploration in Virtual Reality (VR) and development of an individual vocabulary list,
- 2nd phase to review and revise the vocabulary list,
- 3rd phase in which the items from the vocabulary list are used as attributes to assess the stimuli,
- a short subsequent questionnaire to inquire supplementary information.

During the experiment's 1st phase, the participants have been asked to explore the eight audio-visual scenes, each corresponding to one of the receiver locations introduced in Fig. 1. While exploring the scenes, each participant was asked to find four to six attributes (=individual vocabulary) that can be used to describe and distinguish the scenes by auditory and visual means. An attribute defines an individual perceptual dimension to assess the scenes. For each attribute, a lower and a higher anchor were to be defined to facilitate the assessment/rating in phase 3.

The 2nd phase was used to review the collected vocabulary list together with the experimenter. This has been established to ensure a list of attributes that is as concise as possible. For this reason, a review tool had been implemented which allowed to check for the cosine similarity of a collected attribute with other known attributes from Kuusinen and Lokki [11] as well as Weinzierl et al. [12] using the large language model *word2vec-google-news-300* [13]. The tool suggested a list of ten attributes that were closest to the participants' initial choices, and allowed participants to choose whether they wanted to select a suggestion or stick with the original words.

Eventually, in the 3rd phase, the participants navigated through the different scenes again. But this time, they were asked to assess the differences by entering their ratings on a MUSHRA-like interface, see Fig. 2a. The interface allowed to switch between the different listening locations/concert hall seats while the music was continuously looped. The attributes were assessed one after each other. Once all attributes and all scenes were assessed, the participants could decide to finish the experiment or to continue to re-assess the attributes.

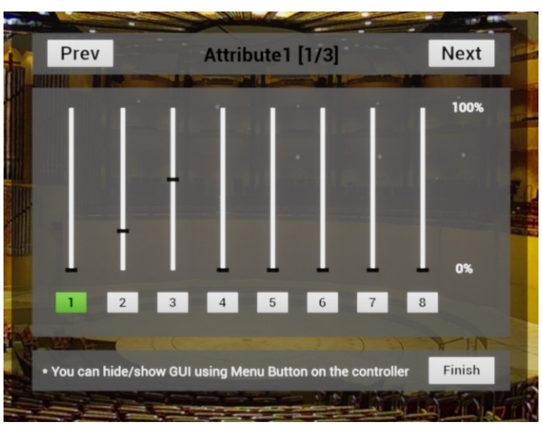
B. VIRTUAL CONCERT HALL ENVIRONMENT

The experiment was conducted in the virtual reality (VR) laboratory at the Institute for Hearing Technology and Acoustics, RWTH Aachen University. The VR laboratory consists of a silent room with low illumination. The participants were seated in an upholstered chair comparable to a typical seat in a concert hall, see Fig. 2b. The following hardware was used:

- A head-mounted display (HMD), HTC Vive Pro, together with
- a corresponding wireless controller for user input,
- a pair of Sennheiser HD650 headphones, and
- a Windows PC with i7-12700K Intel CPU and Nvidia RTX-3070 graphics card.

The framework for the visual reproduction was built in Unreal Engine (UE) to which the eight 360-degree pictures in .hdr format were imported as so-called HDRI backdrop images. These are High Dynamic Range (HDR) images that are spherically projected as a backdrop onto a background mesh. The projection mesh is a sky dome with the virtual actor (= viewer) position at the center of the dome. The actor height was adjusted to the actual floor level around the subject when sitting in the VR laboratory. In the UE project, each photograph and its backdrop was placed in a separate level while all backdrops share the same center point. The switching between listener locations and the display of the corresponding photographs was achieved by hiding or un-hiding the respective level upon user input.

A higher-order, persistent level contained the first-person actor at a fixed position that corresponds to the backdrops' center point. Graphical user interfaces (GUIs) were embedded in this level to guide the participants through the experiment and facilitate the simultaneous comparison between the stimuli. The GUIs provide buttons to select the stimulus/scene to explore. The order in which the stimuli appeared was randomized across participants. During the experiment's 1st phase, the GUI only contained buttons with the numbers 1 to 8 in order to switch between stimuli. In the 3rd phase, the GUI provided sliders to enter ratings on a continuous rating scale (lower to higher anchor) and incorporated two further buttons to change between the attributes, see Fig. 2a. The participants used a wireless controller to navigate through the GUIs and to enter their rating values, see Fig. 2b.



(a) Graphical user interface in Unreal Engine

(b) Participant during the experiment, equipped with head-mounted display, headphones and controller

Figure 2: The listening test environment: Virtual reality (VR) laboratory

C. ACOUSTIC STIMULI

The choice for the piece of music was Beethoven's symphony no. 7, movement 1, bars 14–16 which corresponds to an approximately 11s long excerpt. The anechoic recordings from Pätynen et al. [14] were used. As each loudspeaker position substitutes a specific instrument's position on stage, the anechoic recordings were convolved with the spatial impulse response between the respective loudspeaker and a receiver point. Thus, the sum of all loudspeaker signals convolved with their respective anechoic input signal yield the final stimulus at a receiver location.

The acoustic stimuli were pre-rendered in fourth-order Ambisonics B-format. In the presented setup, we used dynamically adapted binaural playback via headphones which accounts for head rotations. The Ambisonics signal was decoded into binaural signals using head-related transfer functions that were measured from an artificial head [15] in 5×5 degree angular resolution. Eventually, the output was equalized accounting for the headphone characteristics.

The decoding of the Ambisonics audio files into binaural signals was performed in Virtual Acoustics (VA), using the class *BinauralAmbisonicsMixdown* [16]. It incorporates the dynamic head movements that are tracked by the head-mounted display. VA and UE are synchronized sharing the location and orientation of the receiver/rendering point. This is achieved by using the VA Unreal plugin [17].

D. PARTICIPANTS

The experiment was conducted as between-subjects design. Participants were divided into the following two groups:

- Auditory-visual (AV): The participants of this group were presented with both the acoustic stimuli and the visual 360-degree pictures of the concert halls. Thus, they were asked to assess the scenes by auditory and visual descriptors.
- Auditory-only or blind (B): The second group of participants was presented the same set of acoustic stimuli but without visual information. As a result, this group was only asked to provide and assess descriptors of their auditory perception.

In total, a number of 20 participants took part in each group of the experiment. Nobody participated in both parts. Apart from the 360-degree pictures, both groups were exposed to the same conditions. In particular, everyone was assessing the scenes with the same GUI using an HMD and headphones.

The participants cannot be considered as expert assessors/listeners, but most of them were students with experience in acoustics-related fields. As a result, participants were requested to indicate whether they were aware of any hearing loss or deviations from normal hearing. No such case was reported. Thus, nobody was excluded from the experiment.

4. RESULTS AND DISCUSSION

A. INDIVIDUAL VOCABULARY

The first result of the experiment is a vocabulary collected by the assessors to describe their auditory perception in the 1st phase. In Fig. 3a we present a list for the group that assessed the scenes by auditory and visual means (AV) and a second list for the group that performed a blind test (B) without visual information. Both lists are grouped according to the main categories of the "Wheel of Concert Hall Acoustics" proposed by Kuusinen and Lokki [11]. The colors in Fig. 3 indicate the different categories. The distribution of the groups are shown in Figs. 3b and 3c. Only those words that could not be attributed to any category have been removed from the list. For group B, a single mention of "comfort" was removed, while for group AV, two mentions of the attribute "comfort" were removed.

When looking at the distribution across categories, i.e., the occurrence of an attribute of the respective category, both groups AV and B appear to be similar. For most categories, the proportional deviation is below 5%. Although "Loudness" is the most frequently used category in both groups, it shows the largest deviation (8.4%) with an 18.8% portion of AV and 27.2% of B. The only other category that shows a deviation of more than 5% is "Timbre", which was used more frequently in the AV group by 5.3%. Besides "Loudness", the category "Spatial impression" shows a tendency to be more frequently used in B. The tendency for the other categories is the opposite; they show higher contributions in AV. This and the difference in "Timbre" and "Loudness" might suggest that, in the case of audio-visual stimuli, people find it easier to assess a concert hall's qualities in more diverse ways than with regard to loudness. The difference in "Timbre" might be due to lighting and colors influencing the perceived relevance of tone coloration.

B. MULTIPLE FACTOR ANALYSIS

In the next step, we focus on the rating values delivered through the MUSHRA-like assessment during the 3rd phase, see Fig. 2a. For this, we do not analyze each attribute descriptor individually but by applying hierarchical multiple factor analysis (HMFA) [18–20], which allows to get a broader picture. First, a Principal Component Analysis (PCA) is carried out on the dataset of every individual assessor, which is z-score normalized across the attributes. In a second step, the individual datasets are merged into the two groups AV and B. Finally, both groups are combined in the HMFA, with the result that both, AV and B, share the same space for a common representation.

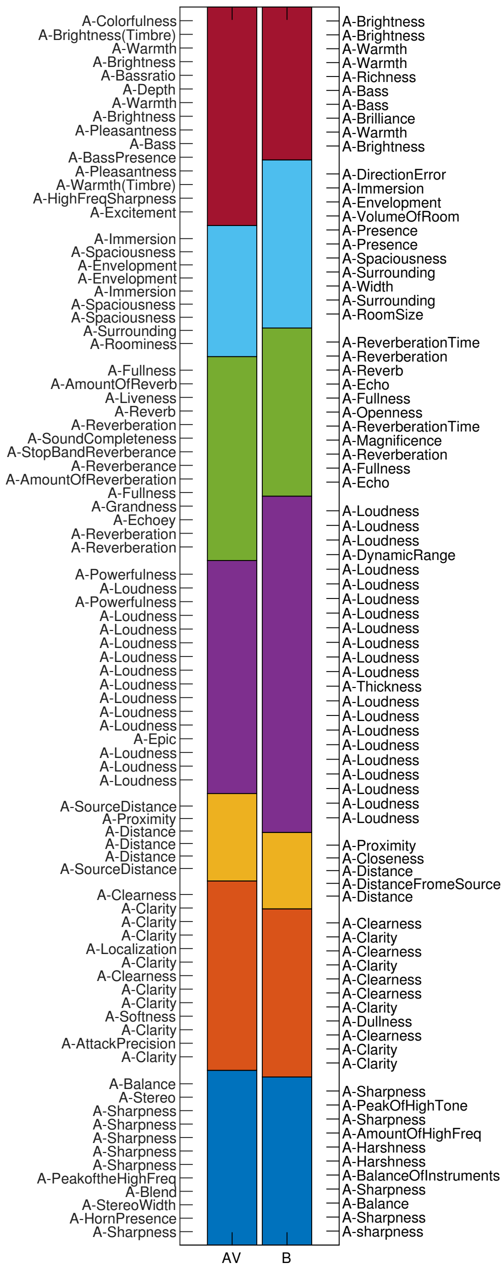
i. Attribute groups

Figure 4 depicts the center points of the categorized attributes' loadings in the HMFA space, which is represented by its first three Principal Components (PCs). The vectors are suggested to be interpreted as the axes of the corresponding perceptual dimensions. Thus, the AV results (left) can be compared with the B results (right) in the 2D-projections of PC1-2 (top) and PC1-3 (bottom). PC1 explains 64.28% of the variance in the underlying data, while PC2 and PC3 explain 9.49% and 8.67% respectively. The line width indicates the number of attributes within a cluster.

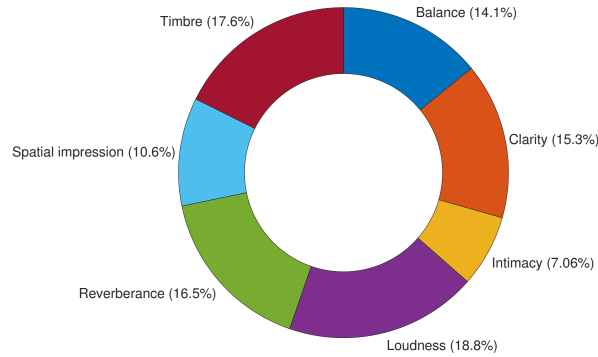
When considering PC1-2 (Fig. 4a and 4b) first, the vector directions of "Loudness", "Clarity" and "Balance" show great agreement with PC1 in both AV and B. For the direction of "Reverberance", PC2 shows more influence, and the distance from the cluster centers to the origin clearly differs along PC1. A strong influence of PC2 can be seen for "Intimacy" and "Timbre", with the direction of "Timbre" aligning well between the two groups, while "Intimacy" does not precisely match in direction. The vector orientation of "Spatial impression" differs most strongly between AV and B, whereby it coincides with PC1 for B and correlates strongly with PC2 for AV.

The representation in PC1-3 visualizes more differences between AV and B. While the vectors "Loudness" and "Reverberance" coincide in both groups, their direction differs. In the case of AV, they correlate closely with PC1. For B, however, both are not pointing in the direction of PC1. Moreover, "Timbre", "Balance" and "Clarity" deviate more from the direction of "Loudness" than in AV. "Spatial impression", like "Reverberance", is located close to the origin in AV. As for PC1-2, "Spatial impression" highly correlates with PC1 in B while it does not in AV. This also holds for "Timbre". The vector of "Intimacy" shows no impact of neither PC2 nor PC3 in AV. In B though, it shows an important contribution of PC2 and PC3.

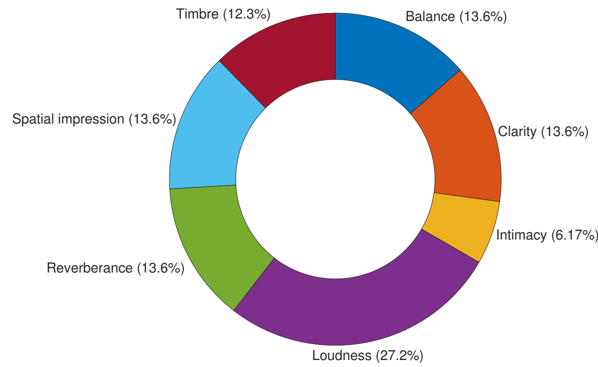
In summary, the comparison of cluster centroids suggests that "Spatial impression" and "Intimacy" are less related to the perceptual dimensions "Loudness" in AV compared to B. Furthermore, the "Spatial impression" seems to be perceived closer to the dimensions of "Timbre" and "Intimacy" when exposed to audio-visual stimuli (AV), compared to the blind stimuli presentation (B) that suggests proximity to "Clarity", "Balance" and "Loudness". Eventually, "Timbre" stands out as PC3 appears to be an important



(a) Auditory-visual (AV) and Blind (B) vocabulary lists



(b) Auditory-visual (AV) vocabulary distribution

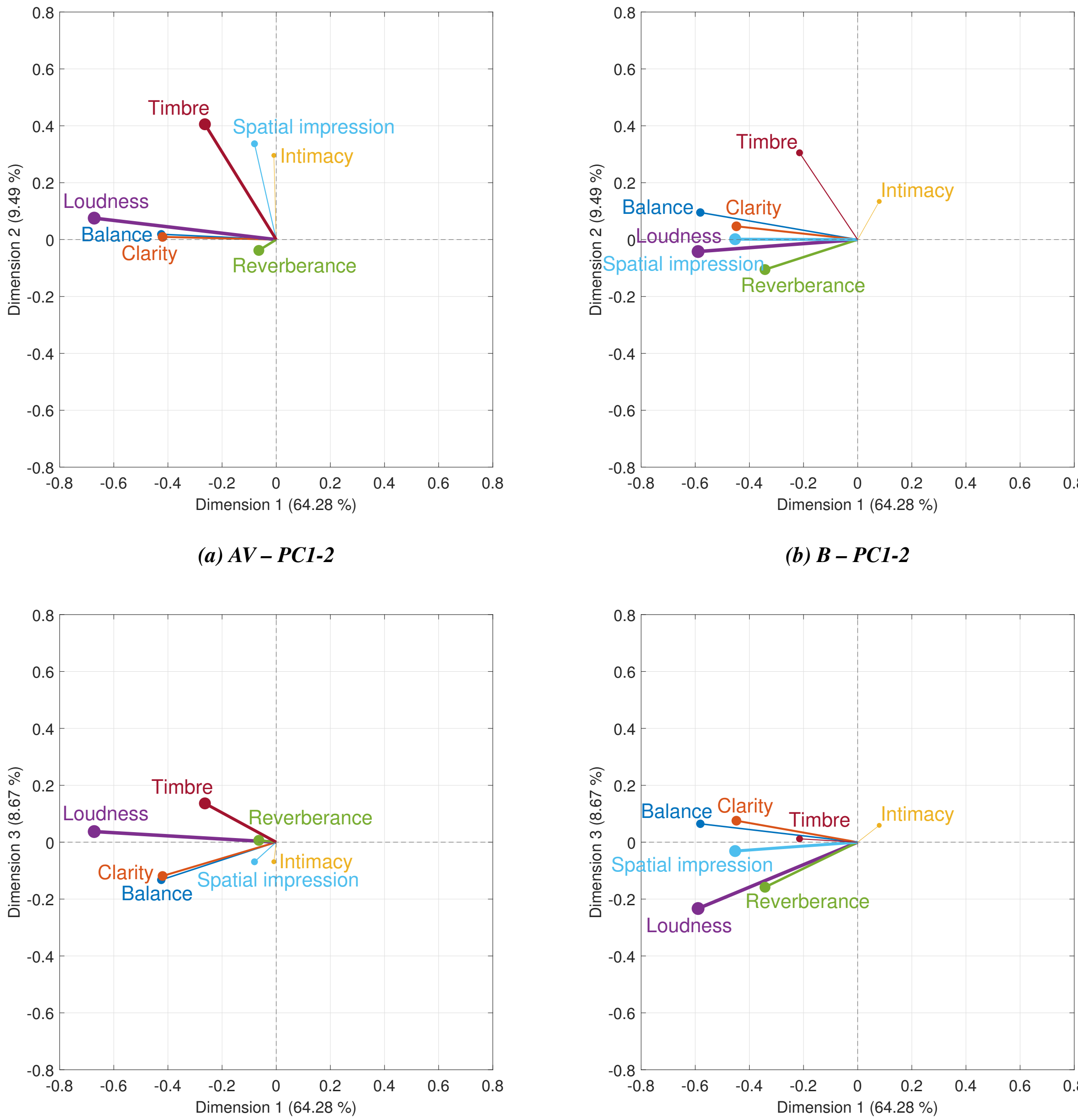


(c) Blind (B) vocabulary distribution

Figure 3: The collected vocabulary grouped according to Kuusinen and Lokki [11]

contributor to it in AV but not in B.

A comparison to earlier studies by Lokki et al. [3, 21, 22] shows that the distribution of attributes across the perceptual clusters is generally similar. Loudness and other perceptual aspects related to loudness or reverberance dominate the discrimination of different concert halls. These studies identify clarity and definition, as well as sometimes balance, to be well distinguishable from loudness-related aspects. However, we found timbre and intimacy to be further separated from loudness than clarity and definition in the HMFA. This is an interesting finding to be further explored in the future since Lokki et al. [21, 22] have reported that concert hall preferences can often be related to timbre and intimacy.



(a) AV – PC1-2 ***(b) B – PC1-2***

(c) AV – PC1-3 ***(d) B – PC1-3***

Figure 4: Representation of the cluster centroids in the principal components' space as a result of the HMFA

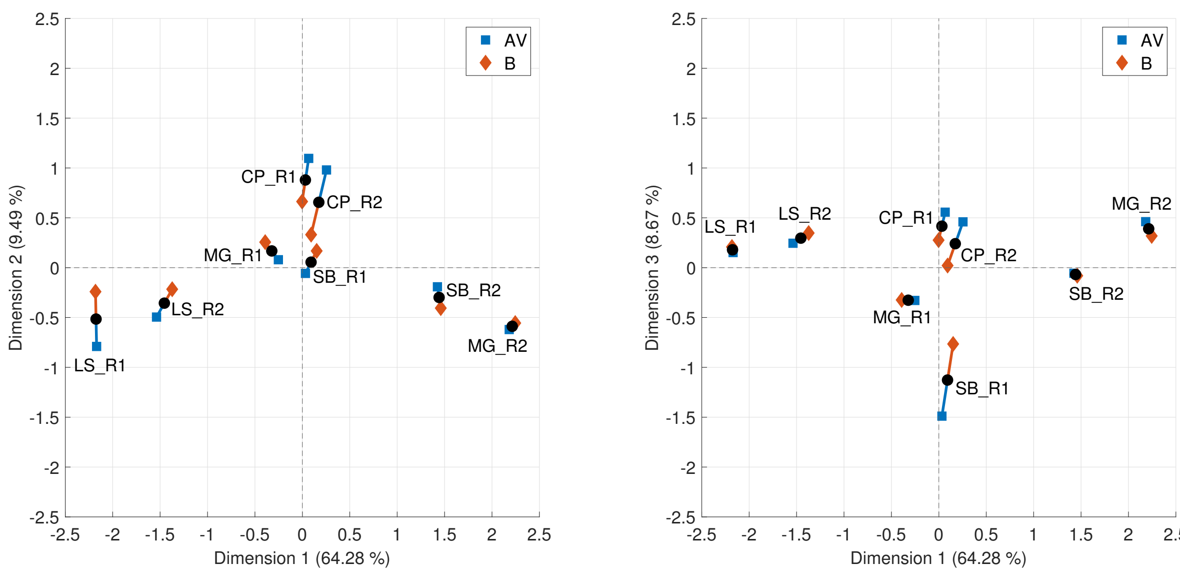
ii. Assessed receiver locations and concert halls

Another way to compare the two modes of stimuli presentation is to contrast the concert halls. Figure 5 depicts the four assessed concert halls in HMFA space. Each hall is represented by two receiver locations as introduced in Fig. 1. All receiver locations are given for the AV group (blue square) and for the B group (red diamond) as well as their average across both groups (black circle).

The first aspect we can observe in Fig. 5 is that both receiver locations R1 (close to the stage) and R2 (further away from stage) in the two concert halls LS and CP are rated consistently close to each other. In contrast, for the halls SB and MG, both receiver locations appear to be more similarly perceived across the halls than within a hall. Also, it can be seen that the spread of values along PC1 is considerably higher than along PC2 or PC3. This suggests that the main dimension of distinction between stimuli is found in loudness. But this does not necessarily mean that this dimension is the main criterion for distinction between the concert halls. Especially those receiver locations centered around the origin of PC1 show higher impact of PC2 or PC3 to differentiate between them.

The probably most important observation, however, is that the difference in perception between AV and B does not appear to lie in PC1, but in PC2 and PC3. As the "Loudness" vectors in Fig. 4 show little contribution of PC2 and PC3, this suggests that it is not the dimensions "Loudness" that drives the difference between B and AV. Instead, the results indicate that bigger differences are perceived for the categories "Timbre" and "Intimacy" as well as potentially for "Spatial impression" and "Reverberance" as they show larger contribution of PC2 and PC3. In consequence, AV results seem to emphasize the differentiation between the halls without completely changing their arrangement.

Finally, it can be observed that the AV and B perceived dimensions do not differ equally in all rooms. From Fig. 5 we can observe that the room ratings of LS and CP spread wider along PC2 than MG and SB. In PC3, a similar spread distribution can still be seen for CP, but not for LS. Instead, SB shows a large variation in PC3 for R1. Presumably, there are room-related visual features triggering these differences. Furthermore, in AV both receiver locations in CP and LS are located closer to each other than in B, which might also be explained by visual similarity.



(a) PC1-2 *(b) PC1-3*

Figure 5: HMFA representation of the concert halls by receiver location and group (AV, B)

5. CONCLUSION AND OUTLOOK

The presented study applied the individual vocabulary profiling (IVP) technique to assess four concert halls by auditory-visual (AV) and by purely auditory, also referred to as blind (B), stimuli evaluation. Respectively, two sets of vocabulary have been collected. The main difference between both attribute lists was found in the categories "Loudness" and "Timbre". While the cluster "Loudness" contains less attributes in

AV than in B, “Timbre” contains more.

MUSHRA-like ratings based on the individually collected attributes allowed for a hierarchical multiple factor analysis (HMFA), which delivered a broader picture of the perceptual dimensions involved when evaluating concert halls. The largest difference between center points of AV and B has been observed in the category “Spatial impression”.

When considering the assessments of receiver locations and concert halls, the main conclusion is that differences between groups AV and B cannot be explained by the loudness-related categories but rather by other dimensions, mainly the perceived “Timbre”, “Intimacy” and “Spatial impression”.

To summarize, we can draw the conclusion that the dominance of “Loudness” reduces in the auditory-visual perception compared to the auditory-only case. This can be seen mainly in the clusters “Spatial impression” and “Reverberance”, which become more distinguishable in AV compared to B with respect to the distance and direction of the “Loudness” vector in Fig. 4. Nevertheless, regarding the comparison of concert hall seats, “Loudness” is governing the largest spread over the perceptual space, which is consistent with the results of previous studies [3, 21]. But at least for certain receiver locations, the audio-visual stimuli presentation increases the importance of other perceptual clusters.

The study has been limited to only four concert halls, one shoebox-shaped (LS) and three fan-/arena-shaped halls (MG, SB, CP), with two receiver locations each. This and the choice of only one short music excerpt might be limitations to the study. Nevertheless, it was shown that the applied techniques are able to identify differences in pure auditory and auditory-visual evaluations of concert halls.

In the future, further analysis of the collected data may provide more detailed results in order to evaluate the underlying mechanisms and for comparison with previous studies. The collection of more case-specific vocabulary for similar scenarios and more participants could be ways to verify the approach’s robustness.

REFERENCES

[1] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse. “Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters”. In: *The Journal of the Acoustical Society of America* 56.4 (Oct. 1, 1974), pp. 1195–1201. ISSN: 0001-4966. DOI: 10.1121/1.1903408.

[2] J. Pätynen, S. Tervo, and T. Lokki. “A LOUDSPEAKER ORCHESTRA FOR CONCERT HALL STUDIES”. In: *Auditorium Acoustics 2008*. Auditorium Acoustics 2008. Vol. Vol. 30. Pt.3 2008. Oslo: Institute of Acoustics, 2008. DOI: 10.25144/17500.

[3] T. Lokki et al. “Concert Hall Acoustics Assessment with Individually Elicited Attributes”. In: *The Journal of the Acoustical Society of America* 130.2 (Aug. 1, 2011), pp. 835–849. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.3607422.

[4] S. Tervo et al. “Spatial Decomposition Method for Room Impulse Responses”. In: *Journal of the Audio Engineering Society* 61 (Jan. 1, 2013), pp. 16–27.

[5] M. Frank and F. Zotter. “Spatial Impression and Directional Resolution in the Reproduction of Reverberation”. In: *Fortschritte Der Akustik - DAGA 2016, Aachen*. 2016.

[6] F. Naccari et al. “Automatic Image Classification by Color Analysis”. In: *Proc. SPIE 5678, Digital Photography*. Electronic Imaging 2005. Ed. by N. Sampat, J. M. DiCarlo, and R. J. Motta. San Jose, CA, Feb. 23, 2005, p. 129. DOI: 10.1117/12.587723.

[7] D. Doeffinger. *Creative Shutter Speed: Master Your Camera’s Most Powerful Control*. Indianapolis: Wiley, 2009. 1 p. ISBN: 978-0-470-45362-9.

[8] E. Reinhard et al. “Color Transfer between Images”. In: *IEEE Computer Graphics and Applications* 21.5 (July 2001), pp. 34–41. ISSN: 1558-1756. DOI: 10.1109/38.946629.

[9] T. Oskam et al. “Fast and Stable Color Balancing for Images and Augmented Reality”. In: *Visualization & Transmission 2012 Second International Conference on 3D Imaging, Modeling, Processing*. Visualization & Transmission 2012 Second International Conference on 3D Imaging, Modeling, Processing. Oct. 2012, pp. 49–56. DOI: 10.1109/3DIMPVT.2012.36.

[10] J. Llorca-Bofí et al. *MUSKLA Database - Muenchen, Stuttgart, Koeln and Lahti Auditoriums Narrow Angle and Spherical Photographs for Perceptual Studies in Architecture*. Zenodo, July 15, 2025. DOI: 10.5281/ZENODO.15409975.

[11] A. Kuusinen and T. Lokki. “Wheel of Concert Hall Acoustics”. In: *Acta Acustica united with Acustica* 103.2 (Mar. 1, 2017), pp. 185–188. ISSN: 1610-1928. DOI: 10.3813/AAA.919046.

[12] S. Weinzierl, S. Lepa, and D. Ackermann. “A Measuring Instrument for the Auditory Perception of Rooms: The Room Acoustical Quality Inventory (RAQI)”. In: *The Journal of the Acoustical Society of America* 144.3 (Sept. 1, 2018), pp. 1245–1257. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.5051453.

[13] T. Mikolov et al. *Efficient Estimation of Word Representations in Vector Space*. Sept. 7, 2013. DOI: 10.48550/arXiv.1301.3781. arXiv: 1301.3781 [cs]. URL: http://arxiv.org/abs/1301.3781 (visited on 07/07/2025). Pre-published.

[14] J. Pätynen, V. Pulkki, and T. Lokki. “Anechoic Recording System for Symphony Orchestra”. In: *Acta Acustica united with Acustica* 94.6 (Nov. 1, 2008), pp. 856–865. ISSN: 16101928. DOI: 10.3813/AAA.918104.

[15] A. Schmitz. “Ein Neues Digitales Kunstkopfmeßsystem”. In: *Acta Acustica* vol. 81 (Number 4) (1995), pp. 416–420.

[16] Institute for Hearing Technology and Acoustics, RWTH Aachen University et al. *Virtual Acoustics - A Real-Time Auralization Framework for Scientific Research*. Version 2022a. Zenodo, 2022. DOI: 10.5281/ZENODO.13744523.

[17] *Unreal Virtual Acoustics Plugin*. Version 4.26. RWTH Aachen University, VR Group.

[18] B. Escofier and J. Pagès. “Multiple Factor Analysis (AFMULT Package)”. In: *Computational Statistics & Data Analysis* 18.1 (Aug. 1994), pp. 121–140. ISSN: 01679473. DOI: 10.1016/0167-9473(94)90135-X.

[19] S. Le Dien and J. Pagès. “Hierarchical Multiple Factor Analysis: Application to the Comparison of Sensory Profiles”. In: *Food Quality and Preference* 14.5–6 (July 2003), pp. 397–403. ISSN: 09503293. DOI: 10.1016/S0950-3293(03)00027-2.

[20] H. Abdi and D. Valentin. *Multiple Factor Analysis (MFA)*. In: *Neil Salkind (Ed.) (2007). Encyclopedia of Measurement and Statistics*. London: Sage Publications, 2007.

[21] T. Lokki et al. “Disentangling Preference Ratings of Concert Hall Acoustics Using Subjective Sensory Profiles”. In: *The Journal of the Acoustical Society of America* 132.5 (Nov. 1, 2012), pp. 3148–3161. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.4756826.

[22] T. Lokki et al. “Concert Hall Acoustics: Repertoire, Listening Position, and Individual Taste of the Listeners Influence the Qualitative Attributes and Preferences”. In: *The Journal of the Acoustical Society of America* 140.1 (July 1, 2016), pp. 551–562. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.4958686.