Hearing Architecture and Seeing Soundscapes: Methodological Approach of Individual Vocabulary Profiling to Evaluate Audiovisual Perception

Jonas Heck, Josep Llorca-Bofí, Michael Vorländer
Institut für Hörtechnik und Akustik, RWTH Aachen University, 52074 Aachen, Deutschland,
Email: jonas.heck@akustik.rwth-aachen.de

Introduction

Architectural and urban environments are exposed to a variety of judgments. Not only those who draw the plans evaluate their design intuitively, but also those who actually use the space will always have a certain mood or prejudice. Despite the diversity of recipients and their perceptions, there is a lack of general agreement on which reproducible methods can be used to evaluate architectural design and space.

Reviewing literature, we found the following two approaches to be promising with regard to people's perception and to evaluate architecture and space: Multidimensional sensory profiling based on auditory, visual and audiovisual attributes with given attributes (cf. [7, 8]) as well as individual vocabulary profiling (IVP) which is found to be used with restriction to one-dimensional attributes, i.e. auditory perception (cf.[4, 5]).

Based on the work performed by Lokki, Lorho, Zacharov et al., we propose an approach that uses individual vocabulary to describe the subjective perception of a space (cf. [4, 5, 6, 9]). We expect this to be a great advantage, as perception varies greatly from individual to individual, yet it promises to extract a *common sense* vocabulary. We plan to use it as the basis for future studies that build upon subjects' judgments and ratings. Here, we will sketch our first concept of the methodology that will gradually be applied to various study cases.

Research Question

Our first aim is to apply techniques known from sensory profiling to reproducibly study architecture and the perception of space. For such sensory evaluations, the corresponding senses are typically addressed separately, divided into auditory or visual attributes. Complementing a third perceptive modality (audiovisual) that combines both auditory and visual perception could allow us to evaluate both modalities at the same time, instead of consecutively. This may increase authenticity as the subject is not concentrating on one modality only but assessing the whole impression.

Furthermore, we are interested in what vocabulary is used to describe audiovisual perception compared to separate auditory and visual evaluation. For this, we will collect case-specific audiovisual vocabulary for several scenarios, such as daily scenes of urban squares and parks, classrooms or music in concert halls (e.g. [2, 3]). Examples of Llorca-Bofí et al. that can be adapted in the level of modeling detail are shown in figure 1. Our approach is to start analyzing recordings of real places that have already been built. But the target is to deliver a method which performs just as well when analyzing architecture in virtual environments. Thus, it could be a useful tool to apply already in an early design process before a building is being constructed.

(a) Classroom (b) Public park (c) Concert hall

Figure 1: Exemplary study cases, cf. [2, 3]

Audiovisual Representation

As we aim for analyzing architecture that is or could be actually built, we search for immersive experiences able to deliver a complete picture of the studied scene. Thus, we consider a 3D-audio and video representation for experimental evaluation. For the recordings in our first studies, we used

- 3D video: Insta 360 Pro2 to capture 7680 x 3840 px (8K) 360° pictures,
- 3D audio: Zoom H3-VR, a first-order Ambisonic microphone.

For reproduction, YouTube can be used which is able to play back 3D video in high resolution and incorporates first-order Ambisonics. When watching these videos on a screen, the degree of immersion is not as high as desirable. But it allows to rotate the view by adjusting the sound accordingly. Moreover, it is easy to implement and can be used for online surveys without additional effort.

However, for future studies we are going to implement a system combining a visual renderer, such as Unreal Engine, with a audio rendering tool, i.e. Virtual Acoustics [1]. This will allow us to study recorded, virtually rendered as well as hybrid environments in cave-like environments or with head-mounted displays.

It will be a further research question how detailed virtual representations have to be modeled and how simulations can be compared to real-world recordings.

Experimental Evaluation

Individual Vocabulary Profiling

The first phase of our experimental procedure is the **development of individual vocabulary**. This means that each participant has to brainstorm words in order to describe his/her perception of the scene with individual vocabulary and assign attributes to what is seen and heard. The participants are asked to not only develop a list of vocabulary for each single scene (e.g. each perspectives on the same object), but one list for all studied scenes (e.g. all perspectives).

This is followed by an **elicitation** in order to sort words out that are not precise enough. This step causes high effort since supervision is needed. Furthermore, it is a trade-off between objectivity and people's personal perception which is not always reflected in objective parameters. It can already be stated that this is a challenge, especially when considering various cultural and linguistic backgrounds where the same word could have a different connotation.

The **definition of scales** is the third part. It comprises the individual definition of upper and lower limits for all resulting attributes. The last part is the **rating** on these self-defined scales. With this, we expect to get authentic ratings from each participant as he/she understands what perception is related to which attribute.

Clustering

In order to make individual ratings comparable, they need to be summarized into clusters. Each cluster represents one attribute and contains those attributes of all participants that correspond to each other. What is called *Clustering I* on the poster of this contribution (see appendix), is the allocation of attributes to clusters of either the visual or the auditory domain. Within clustering *Clustering II*, the attributes are clustered into visual, auditory and audio-visual domains (V, A, VA).

A comparison of both clustering approaches allows to analyze the deviation of the attribute ratings based on the AV domain from ratings based on the separate domains A and V, i.e. when the participants are asked to evaluate both domains separately. Furthermore, it has to be studied whether a purely statistical clustering approach will be sufficient or experts must perform the categorization.

Summary and Outlook

The presented approach is meant to initiate a reproducible methodology of capturing and representing architectural and urban sceneries in order to evaluate them under controlled experimental conditions. For this purpose, we propose to use the individual vocabulary profiling technique (cf. [4, 5, 6, 9]).

This approach seems to be promising for obtaining authentic ratings. On the other hand, it is time consuming, especially because we have so far personally revised each participant's attributes in order to obtain vocabulary that is as objective as possible. In addition, statistical analysis is a challenge, since one cannot necessarily rely on automated pattern recognition, especially when the amount of subject data is small. Moreover, it can be expected that some attributes can be related to both auditory and visual domain. It must be further studied which predominates.

For the future, we plan to apply this approach to a variety of scenarios including indoors and outdoors, real recorded scenes or virtual environment. We will have to investigate whether case-specific attributes need to be developed individually for each particular scene, or a general database of attributes will be sufficient. For outdoors, certainly a number of complex factors play a role, e.g. weather conditions. Our procedure is suited to be performed as online experiment.

Acknowledgment

This work was carried out by the *Person-focused Analysis of Architectural Design – PAAD* research group, a Junior Principal Investigator fellowship at RWTH Aachen University, funded by the German Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the Federal State of North Rhine-Westphalia (NRW) under the Excellence Strategy of the Federal and the State Governments.

References

[1] Institute for Hearing Technology and Acoustics, RWTH Aachen University. Virtual Acoustics – A real-time auralization framework for scientific research, `http://virtualacoustics.org/VA/`, accessed September 3, 2021

[2] Llorca-Bofí, J. et al. (2018). 3D modelling photogrammetry to support acoustic measurements and derive geometries for simulation, Conference: Auditorium Acoustics, October 2018

[3] Llorca-Bofí J. and Vorländer M. (2021). Multi-Detailed 3D Architectural Framework for Sound Perception Research in Virtual Reality, Front. Built Environ. 7:687237.

[4] Lokki, T. et al. (2011). Concert hall acoustics assessment with individually elicited attributes, J. Acoust. Soc. Am., Vol. 130, No. 2, August 2011.

[5] Lokki, T. et al. (2016). Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences, J. Acoust. Soc. Am., Vol. 140, No. 1, July 2016.

[6] Lorho, G. (2005). Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction, Audio Engineering Society Convention Paper 6629, 119th AES Convention, 2005 October 7-10, New York.

[7] Maempel, H.-J. (2012). Experiments on audiovisual room perception: A methodological discussion, DAGA 2012, Darmstadt.

[8] Maempel, H.-J. and Horn, M. (2017). Audiovisual perception of real and virtual rooms, Journal of Virtual Reality and Broadcasting, 14(2017), no. 5, May 2018.

[9] Zacharov, N. (Ed.) (2018). Sensory Evaluation of Sound (1st ed.), CRC Press, 2018, Boca Raton.

Appendix: Poster

The appendix shows the poster as presented during the session *Umgebungslärm* on August 16 at the 47th Annual German Conference on Acoustics 2021 (DAGA 2021) in Vienna, Austria.

DAGA 2021 Wien

IHTA Institute for Hearing Technology and Acoustics | RWTH AACHEN UNIVERSITY

Hearing Architecture and Seeing Soundscapes

Methodological Approach of Individual Vocabulary Profiling to Evaluate Audiovisual Perception

Jonas Heck, Josep Llorca-Bofí, Michael Vorländer

Motivation

Architectural and urban environments are exposed to a variety of judgements: from those who draw the plans to those actually using the space. Despite the diversity of perceivers of such environments, there is a **lack of agreement on which reproducible methods** can be used **to evaluate architectural design and space**.

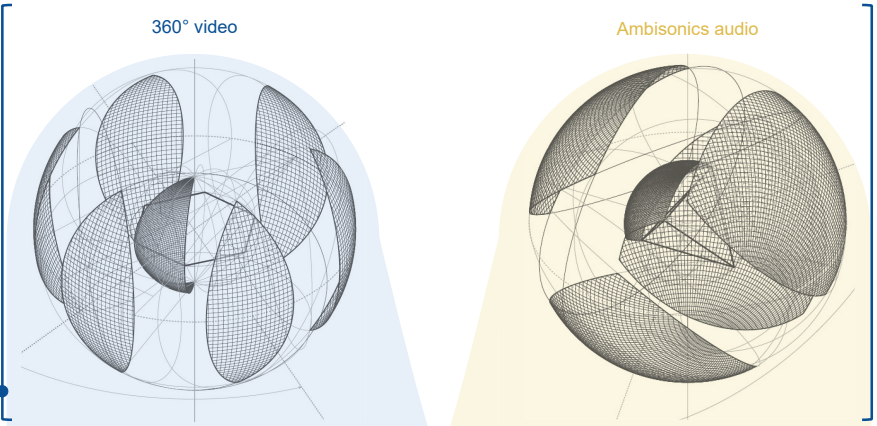
Based on the work performed by Lokki, Lorho, Zacharov et al., we propose an **approach that uses individual vocabulary** to describe the subjective perception of a space.[1, 2, 3, 4] We expect this to be a great advantage, as perception varies greatly from individual to individual, yet it promises to extract a "common sense" vocabulary. We plan to use it as the basis for future studies that build upon subjects' judgements and ratings.

Reviewing available literature, the following two approaches can be found: **multi-dimensional sensory profiling** based on auditory, visual and audiovisual attributes[5, 6] with given attributes as well as **individual vocabulary profiling (IVP)** which is found to be limited to one-dimensional attributes, i.e. auditory perception[1, 2].

Research Question

Complementing a third perceptive modality (audiovisual) that combines both auditory and visual perception could allow us to **evaluate both modalities at the same time**, instead of consecutively. This may **increase authenticity** as the subject is not concentrating on one modality only but assessing the whole impression.

Furthermore, we are interested in **what vocabulary is used to describe audiovisual perception** (audiovisual attributes) compared to separate auditory and visual evaluation. For this, we will collect case-specific **vocabulary for several scenarios**, such as daily scenes of urban squares and parks, classrooms or music in concert halls (e.g. [7, 8]).



Audiovisual Representation

The attributes development: Subjects describe their perception of the scene with **individual vocabulary**. This includes the definition of **scales**.[3]

Rating the scenes based on the self-developed scales.

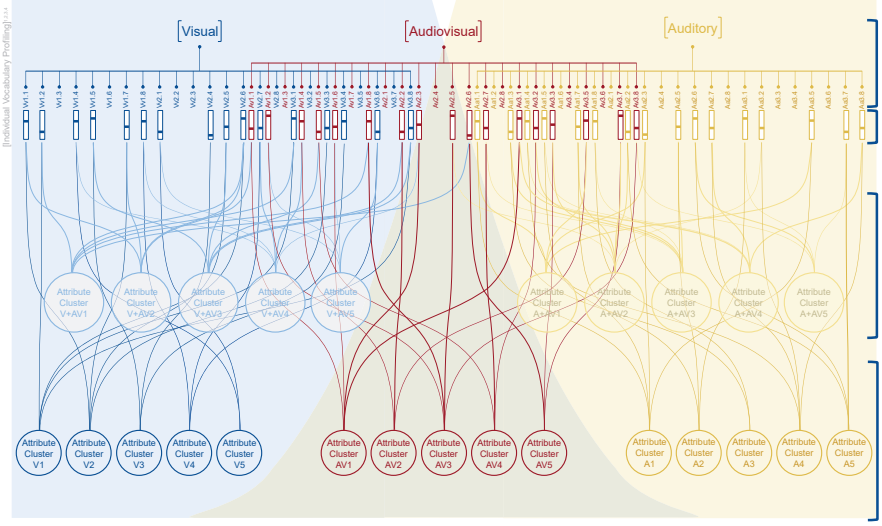
Experimental Evaluation

Clustering I

All rated attributes are summarized into corresponding clusters. There exist clusters for both visual and auditory attributes. They also comprise audiovisual attributes that can be assigned to either one or the other domain.

Clustering II

The attributes are clustered into the visual, auditory and audiovisual domain (V, A, VA). Compared with clustering I this allows to analyze the deviation of the attribute ratings based on the AV domain from ratings based on the separate domains A and V.



Outlook

The presented approach is meant to initiate a reproducible methodology of capturing and representing architectural and urban sceneries in order to evaluate them under **controlled experimental conditions**. For this purpose, we propose to use the **individual vocabulary profiling** technique (see [1, 2, 3, 4]).

This approach seems to be promising for obtaining authentic ratings. On the other hand, it is **time consuming**, especially because we have so far personally revised each participant's attributes in order to obtain vocabulary that is as objective as possible. In addition, **statistical analysis is a challenge**, since one cannot necessarily rely on automated pattern recognition, especially when the amount of subject data is small. Moreover, it can be expected that some attributes can be related to both auditory and visual domain. It must be further studied which predominates.

For the future, we plan to apply this approach to a variety of **scenarios including indoors and outdoors**, **real recorded scenes or virtual environment**. We will have to investigate whether case-specific attributes need to be developed individually for each particular scene, or a general database of attributes will be sufficient. For outdoors, certainly a number of complex factors play a role, e.g. weather conditions. Our procedure is suited to be performed as online experiment.

References

[1] Lokki, T. et al. (2011). Concert hall acoustics assessment with individually elicited attributes, J. Acoust. Soc. Am., Vol. 130, No. 2, August 2011.

[2] Lokki, T. et al. (2016). Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences, J. Acoust. Soc. Am., Vol. 140, No. 1, July 2016.

[3] Lorho, G. (2005). Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction, Audio Engineering Society Convention Paper 6629, 119th AES Convention, 2005 October 7-10, New York.

[4] Zacharov, N. (Ed.) (2018). Sensory Evaluation of Sound (1st ed.), CRC Press, 2018, Boca Raton.

[5] Maempel, H.-J. (2012). Experiments on audio-visual room perception: A methodological discussion, DAGA 2012, Darmstadt.

[6] Maempel, H.-J. and Horn, M. (2017). Audiovisual perception of real and virtual rooms, Journal of Virtual Reality and Broadcasting, 14(2017), no. 5, May 2018.

[7] Llorca-Bofí J. and Vorländer M. (2021). Multi-Detailed 3D Architectural Framework for Sound Perception Research in Virtual Reality, Front. Built Environ. 7:687237.

[8] Llorca-Bofí, J. et al. (2018). 3D modelling photogrammetry to support acoustic measurements and derive geometries for simulation, Conference: Auditorium Acoustics, October 2018

Institute for Hearing Technology and Acoustics
RWTH Aachen University
Kopernikusstraße 5 | 52074 Aachen | GERMANY
www.akustik.rwth-aachen.de
Tel.: +49 241 80-97983
Fax: +49 241 80-92214

Jonas Heck
+49 241 80-97836
jonas.heck@akustik.rwth-aachen.de

Josep Llorca-Bofí

Michael Vorländer

This work was developed by the "Person-focused Analysis of Architectural Design" - PAAD research group, a Junior Principal Investigator fellowship at the RWTH Aachen University, funded by the Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW) under the Excellence Strategy of the Federal Government and the Länder.