Audio-visual room perception

Dr. Hans-Joachim Maempel

Scientia

www.scientiapublications.com

Audio-visual room perception

Acoustic scientist Dr. Hans-Joachim Maempel has long been interested in the interplay between hearing and sight. Here he describes his efforts to lay the groundwork for a unified theory of audio-visual room perception.

To start with, please could you describe what inspired you to start looking into the relationship between auditory and visual perception?

After receiving my diploma to become a tonmeister – a profession that combines the work of a music producer and a balance engineer – I composed or produced soundtracks for film and TV for some time. In so doing, you get a feeling for the effect of sound and vision on a semantic and aesthetic level. Within the framework of my PhD thesis I investigated the effect of the mixdown as a production stage of pop music, and as part of that I varied the presence of the corresponding video clip too. In the nineties music television was of special importance. Amongst other things I found that people prefer different mixdowns depending on whether or not there is a video clip. During my time as a postdoc researcher at the TU Berlin I was asked to develop and conduct a course on audio-visual perception, and I was invited to write a lexicon article. So I started to review empirical findings on different audio-visual tasks more systematically.

Why do you think acoustic research has historically failed to properly investigate the interdependence of the auditory and the visual systems?

Since acoustic research is not primarily interested in the investigation of audio-visual perception I think it is quite natural that it has first of all applied itself to its key subjects: for example sound propagation under certain conditions, auditory physiology, or basal perceptual abilities with particular regard to intensity, timbre, and localisation. There are several sub-disciplines in acoustics such as technical acoustics, room acoustics, electroacoustics or psychoacoustics. This is conducive to further differentiation rather than an extension of the scope of research questions. Particularly in room acoustics a physical perspective was predominant for a long time, though actually human perception is always the ultimate criterion for the evaluation of sound in rooms. Though emphasizing the psychological over the physical perspective was desirable, it still disregards the fact that hearing and seeing generally occur at the same time. So acoustic research should eventually start to investigate the optical and also visual conditions under which its diverse results are obtained. The localisation of sound sources is a good example. Auditorily it is quite accurate and reliable, but an additional corresponding optical stimulus in a slightly different position will result in the localisation of a fused percept close to this optical position. Experiments on time determination tasks show in turn that hearing may dominate seeing under certain conditions.

Why is there a need for an empirically founded theory of audio-visual room perception? What are some of the practical applications?

Generally, questions of empirical research are related to present knowledge in order to achieve connectable results. This knowledge is usually formulated in terms of theories that hold a certain degree of generality. But if a theory does not exist, its formation is the priority. Specifically, previous research into audio-visual perception of room acoustics can be characterised by a highly technical approach. This involved the use of room models and the use of artificial sound fields, for example presented by loudspeakers in anechoic rooms; the consideration of few, often highly specific independent and dependent variables; and by the application of different empirical paradigms and methods. Apparently this contributes to results that lack consistency and connectability.

Nowadays research is often required to legitimate itself by its immediate applicability. This criterion may, however, not apply to fundamental research. In fact, applications of our research are not really 'practical' at the moment. Results contribute to the understanding of multi- and intra-modal processes of speech and music perception in rooms; they provide criteria for the validation and/or interpretation of previous studies dealing with the perception of room acoustics under unimodal conditions; they add to the extension and the refinement of simulation techniques; and they are useful for the design of content for opto-acoustic environments.

Why is it so important to use consistent terminology when carrying out this kind of research?

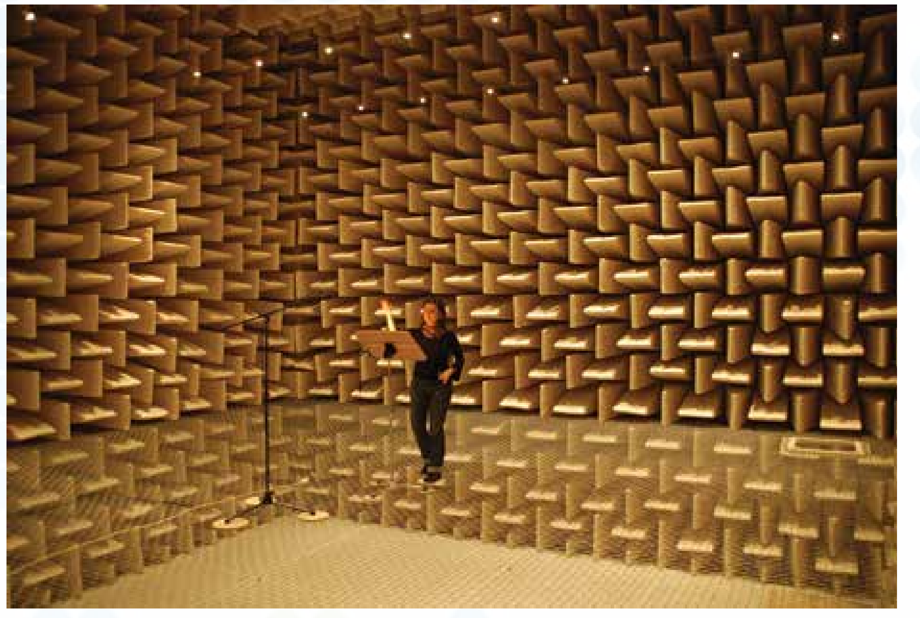
The physical and the psychological realms must be properly differentiated. They need to be clearly differentiated in publications too, which has so far rarely been done. I try to introduce the terms 'acoustical' and 'optical' for denoting a physical representation and 'auditory' and 'visual' for denoting a psychological representation. Particularly with the acoustical terminology, I always struggle with the fact that in English both the physical and the psychological representation of a sonic event is called a 'sound'. It is easier in German – there are two terms, 'Schall' and 'Klang'. Furthermore, unimodal and multimodal perceptual and cognitive aspects should be differentiated. This would enable us to more clearly categorise the direction of effects, for example as 'cross-modal'.

Splitting the senses: creating a virtual environment to test audio-visual perception

Untangling the interactions between auditory and visual perception is a challenging task in the real world. That's why acoustic scientist Dr. Hans-Joachim Maempel turned to cutting-edge simulation technology to tackle the topic.

TACKLING ASSUMPTIONS

It seems intuitive that our perception of the world relies on a complex interplay between the senses, in particular sight and hearing. But in the science of acoustics, the impact of optical stimuli on how we perceive sounds has been largely ignored.



When reviewing the literature, however, acoustic researcher Dr. Hans-Joachim Maempel discovered that many studies assume such an interaction exists with little experimental evidence. This realisation encouraged him to set about developing a comprehensive experimental study of the interaction between auditory and optical stimuli on human's perception of rooms.

What research has been done on this topic has focussed on very specific questions and as a consequence results are not easily translated from one study to another. Inconsistent terminology has only exacerbated this problem, leading Maempel to the realisation that the field required a solid research strategy that encompassed easily reproducible methodology as well as a clear and consistent vocabulary. Under the framework of the German Research Foundation unit Simulation and Evaluation of Acoustical Environments (SEACEN) he commenced the project "Audio-visual perception of acoustical environments" in 2011.

APPLES AND ORANGES

As a starting point he realised there needed to be an obvious delineation between the physical and psychological aspects of both hearing and sight. This means distinguishing between acoustical and optical measures and auditory and visual ones, respectively. A problem with carrying out research on two very different sensory modalities, however, is that their measurements are not really comparable.

"Comparing the effect of the acoustical variation of sound pressure with the effect of the optical variation of illuminance is comparing apples with oranges," says Maempel. In addition, to properly tease out any interaction between the two modalities it is necessary to be able to vary them independently. In the real world this is clearly unfeasible – trying to make the sound from a speaker emerge from anywhere other than where it is optically located is very difficult.

To overcome these limitations, Maempel turned to cutting-edge simulation technology to present a virtual environment in which all the relevant audio-visual parameters could be tweaked independently. Using a combination of auralisation technology and stereoscopic projectors he and his fellow researchers created a virtual concert hall that allowed them to create the conflicting optical and acoustical stimulus that would help them reveal how the two modalities interact.

Solving the lack of comparability between the two modalities relied on a surprisingly simple trick. The team decided to record the optical and acoustic properties of six rooms, which could then be combined independently to create six environments with congruent stimuli and 30 with conflicting stimuli. This was no simple task though. "It was very hard to setup a low-cost optical virtual environment, every imaginable technical problem actually occurred and had to be solved," says Maempel.

IMMERSIVE ENVIRONMENTS

To start with the team had to take photographs for 360 possible head orientations to create panoramic stereoscopic images of each room. These were then projected onto a semi-cylindrical screen using five stereoscopic projectors and warping and edge-blending software that resulted in a nearly 180° field of view for the test subject. Mapping the acoustical environment was even more complicated though, as not only did the representation have to be highly precise, the simulation also needed to be able to respond to the subject moving their head to explore the panoramic visual display.

This required the team to rely on dynamic binaural synthesis – a method whereby acoustic fingerprints are recorded using a 'dummy' head and torso with microphones in the position of each ear. By playing sine sweeps through loudspeakers located exactly where a performer would be positioned in each of the six rooms, the microphones on the dummy were used to build up the properties of the acoustical environment a human subject sat in the same seat would experience.

To cope with the fact that a human subject is likely to move their head, acoustic fingerprints were taken at various head orientations. Electromagnetic tracking technology built into the headphones human subjects subsequently wore for experiments, allowed their head orientation to be traced and the acoustical simulation to be compensatorily altered in real time.

Following the construction of the virtual environments Maempel and his team then made recordings in an anechoic room of four professional musicians playing Claude Debussy's 1893 string quartet and a professional actress performing the first paragraph of Rainer Maria Rilke's Duino Elegies. Both performances were then repeated in a green screen studio and recorded by a stereoscopic video camera. These optical and acoustical recordings were then blended with the virtual environment using chroma key compositing and dynamic binaural synthesis respectively.

This set up allowed human subjects to experience a fully immersive simulation of each room. The ability to turn their head without turning the virtual acoustic environment with it meant they could explore the virtual concert halls during each performance as they wished. It also allowed Maempel to independently vary acoustical and optical stimuli for the first time. To see what this could tell him about how the two modalities interact, he carried out a series of experiments where participants were exposed to each of the congruent and conflicting stimuli. They were then asked to evaluate their experience via an electronic questionnaire displayed on a tablet computer.

A SURPRISING CONCLUSION

Despite the long standing assumption of an interaction between auditory and visual perception, the group's first results indicate that there is neither significant interaction nor cross-modal effects between the two modalities.

To put it simply auditory perception relied on acoustic information and visual perception on optical information. An interesting additional finding was that the way the human brain perceives a room is relatively flexible. In a preliminary experiment with lower quality optical simulations distance and room size judgments relied predominantly on acoustic information. But in later experiments with the more sophisticated optical simulation the visual modality became dominant in making such judgements.

"Since I do not support a naïve-realistic conception of human perception I personally was not surprised about the flexible exploitation of available physical cues by the two modalities," says Maempel. "It took me, however, a little by surprise that we did not observe significant and practically relevant interaction or cross-modal effects. They have often been assumed, though one should take into account that thereby the term 'interaction' has not always been used in a strict methodological sense. Our findings indicate, anyway, that the modalities contribute to a percept (something that is perceived) in a straightforward and more or less additive way."

Refuting this commonplace assumption in acoustic science not only has major implications for research going forward, but will also allow academics to revisit previous studies to understand how this conjecture may have affected results. But as well as solving the question he initially set out to answer, Maempel's work has provided a valuable by-product. "The virtual concert hall is a quite flexible research tool allowing also for the investigation of more specific questions," he says.

The next phase of the project will look into a variety of more detailed research questions including whether humans can audio-visually match optical and acoustic room shapes, both when they are perceived and memorised. The team also wants to look at the role of participant's musical and acoustic expertise as well as their musical preferences in how they perceive audio-visual space.

Researcher Profile

Dr. Hans-Joachim Maempel
State Institute for Music Research in Berlin

Dr. Hans-Joachim Maempel is head of the department for acoustics and music technology / studio facilities and IT at the State Institute for Music Research in Berlin, an agency of the Prussian Cultural Heritage Foundation. He started his career in the creative industry composing and producing soundtracks for film and TV before moving into academia. He received his PhD in Musicology at Berlin Institute of Technology in 2001.

CONTACT
T: +49 30 25481 141
E: maempel@sim.spk-berlin.de
W: http://www.sim.spk-berlin.de/
W: https://www.ak.tu-berlin.de/menue/mitarbeiterinnen/wissenschaftliche_mitarbeiter/dr_hans-joachim_maempel/

KEY COLLABORATORS
Michael Horn, SIM Berlin
Alexander Lindau, MPIEA Frankfurt
Michael Vorländer, RTWH Aachen
Stefan Weinzierl, TU Berlin

GRANTS AND FUNDING
German Tonmeister Association
Federal State Berlin
German Research Foundation

Staatliches Institut für Musikforschung
Preußischer Kulturbesitz

SEACEN