Preprint

This is a preprint of an article accepted for publication in *Jahrbuch des Staatlichen Instituts für Musikforschung 2016*, ed. by Simone Hohmaier, Mainz et al.: Schott. Estimated publication date: Spring 2018.


Based on the project »Audio-visual perception of acoustical environments« within the framework of the research unit *Simulation and Evaluation of Acoustical Environments (SEACEN)*, funded by the German Research Foundation (DFG) and coordinated by the Audio Communication Group of the TU Berlin.

© 2017. Licensed under the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE): https://creativecommons.org/licenses/by/3.0/de/legalcode

Correspondence to be sent to Dr. Hans-Joachim Maempel, Department for Acoustics and Music Technology | Studio Facilities and IT, Staatliches Institut für Musikforschung Preußischer Kulturbesitz, Tiergartenstraße 1, Berlin, Germany, maempel@sim.spk-berlin.de
Apples and Oranges
A Methodological Framework for Basic Research into Audiovisual Perception

Abstract
In recent years, increased efforts have been made to investigate the various aspects of audiovisual perception. In general, highly-specific variables are considered, and specific paradigms and methods are applied. Thus, connecting results of various experiments is not always possible in a systematic manner, and several related research fields still lack comprehensive theories and corresponding strategies. Based on prototypical research questions a methodological framework has been developed that differentiates two ontological realms (a physical and a perceptual), two modalities (sound and vision), and the modal specificity of physical properties and perceptual features. It may be further differentiated by introducing several physical and perceptual processing stages and be extended by introducing perceptual feedback loops and the dimension of time. The framework supports the empirical investigation of well-directed effects and interaction effects based on modally-balanced experimental designs in order to progressively model audiovisual perception. Additionally, methodological issues such as operationalization, design paradigms, optoacoustic congruence and optoacoustic commensurability are discussed. As an example of application, a technical research tool is introduced whose development was exclusively geared to the proposed methodological guidelines and that allows for the experimental investigation of fundamental questions on audiovisual perception: The Virtual Concert Hall.

Keywords: Multimodal Perception, Cross-modal Processing, Audiovisual Interaction, Cue Combination, Methodology, Virtual Reality, 3D, Room Simulation, Concert Hall.
I Introduction

Our access to the everyday world and many arts is inherently audiovisual, as far as no hearing- or sight-impairment handicaps perception. This statement may be trivial, but the questions that it raises are fundamental, and answering them by means of the empirical sciences is challenging. To what extent are we influenced by the voice and the facial expression of a person? To what extent are the localization of objects and the temporal determination of events based on hearing and sight? To what extent do the reaction time and the decisions of drivers depend on optical and acoustic information? How are pictorial and musical aspects of a film conducive to suspense? Does the visual aspect of a musical performance contribute to its emotional impact? And what role do the acoustic and optical properties of the performance space play therein? In this paper, several methodological issues raised by the experimental investigation of such questions are discussed and general or question-specific methodological criteria are derived. Finally, a technical research tool adequate to these criteria is presented.

II Subject

Research on audiovisual perception and processing indicate that the mental representation of physical objects is normally based on, amongst others, both the auditory and the visual modality. This may be observed in various fields such as intensity rating\(^1\), localization\(^2\), motion perception\(^3\), event time perception\(^4\), synchrony perception\(^5\), per-


http://dx.doi.org/10.14279/depositonce-6424
ceptual phonetics, quality rating, construction of meaning, and room perception. There is also comprehensive work on the neurophysiological mechanisms of multisensory perception and integration. Despite this, both research on auditory perception and research on visual perception have historically paid little regard to the cooperation of the auditory and visual modalities. Thus, many of the aforementioned fields still lack comprehensive theories of audiovisual perception and corresponding research strategies. Rather, researchers have considered quite specific variables and applied accordingly circumscribed experimental paradigms and methods. As a consequence, it is not always possible to connect the results of various experiments in a systematic manner. Researchers might improve this situation through increased consideration of strategic aspects, methodological reflection, a con-

---


sistent terminology, and the formation of an empirically-founded model as general goals of research projects on audiovisual perception.

### III Towards a general working model

Against this background of patchy findings on audiovisual perception, the primary goal of research efforts in the applicable fields should be the formation of empirically-founded models. A general working model based on the philosophical principle of complementarity – i.e., taking two ontological realms as bases – would be apt to describe fundamental effects of complex (in terms of multidimensional) physical properties on complex perceptual features within and across the modalities. The general working model might be differentiated by considering specific (in terms of unidimensional) physical properties, physiological and neurophysiological processing stages, specific perceptual features, and be extended by introducing feedback loops (e.g. orientational reactions, top-down processes, etc.). The investigation of further factors (contexts, personal features, etc.) could yield supplementary information about the scope of the model.

In order to build such an empirically-founded model efficiently, a funnel-shaped research strategy might be applied by first investigating basic questions, and subsequently investigating specific questions (or, subsequently including existing specific results). Types of research questions (RQs) are:

1. What are the proportional contributions of hearing and sight to (complex or specific) perceptual features?
2. Do the modalities interact, and if so, in which way?
3. What are the effect sizes along effect directions within and between modalities and ontological realms, and – more specifically – between certain physical, physiological, neurophysiological and mental processing stages?
4. What is the scope of the findings about the above questions regarding perceptual conditions (quality of stimulus presentation, context), personal features (socio-demographic features, expertise), stimulus type (speech, music, noise), and semantic content?
From a research-economic point of view, the use of a corpus of interrelated data is desirable. To this end, data might be collected concurrently from different realms, modalities, and processing stages, and different design paradigms involving different bi-modal stimulation principles might be integrated. Integrative data collection provides the opportunity to reveal multi-level and complex relationships between perceptual features, and to retrace the perception process along a larger section of the transmission and processing chain.

IV Methodological considerations

A Ontological realms

As a prerequisite for modeling both two modalities (sound and vision) and two ontological realms (the physical and the perceptual), a clear factual and terminological distinction of their respective categories has to be made. Thus I refer to the relevant physical properties as **acoustic** and **optical** and to the respective perceptual features as **auditory** and **visual**. As a complement to the term **modality** that differentiates between system-specific processes such as hearing and sight in the perceptual realm, I apply the term **domain** in the physical realm in order to differentiate between processes that are based on sound and light (table 1).

TABLE 1. Factual and terminological distinctions between ontological realms, and between domains and modalities, respectively.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Realm</th>
<th>Modality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>physical</td>
<td></td>
</tr>
<tr>
<td>acoustic</td>
<td>properties</td>
<td>features</td>
</tr>
<tr>
<td>visual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>optical</td>
<td>properties</td>
<td>features</td>
</tr>
</tbody>
</table>

Consequently, according to the suggested taxonomy, the term **stimulus** denotes a mere physical condition or event. Within this taxonomy, collocations such as **stimulus modality** or **visual stimulus** do not make sense.
because they blur the ontological boundary between the physical and the perceptual realm unnecessarily and without other advantage.

This blurring is also a problem inherent in classical psychophysics. The expression of sensory quantities by means of physical quantities may be reliable, but does not meet construct validity by definition. For example, we do not hear a frequency spectrum itself, but rather the corresponding perceptual features of timbre and eventually pitch. And in contrast to the psychophysical Mel scale, the perception of pitch is at least two-dimensional along a linear dimension (pitch height) and a cyclic dimension (pitch chroma), and a clear line between pitch and timbre aspects can hardly be drawn.

The English language often does not support such a differentiation of the ontological realms. For example, both the physical and the perceptual representation of a sonic event are called ‘sound’. By contrast, in German, the physical event itself is termed ‘Schall’, while its associated perceptual representation is termed ‘Klang’.

B Processing stages

Both integrative data collection (sec. III) and the differentiation of effect directions (sec. IV D) require the introduction of transmission stages in the physical realm and processing stages in the perceptual realm – irrespective of to what extent each stage is causally determined.

by the previous stage. Naturally, in the physical realm deterministic relations are predominant, whereas in the neurophysiological and perceptual realm probabilistic relations are prevalent, and super-additive or sub-additive effects occur.\textsuperscript{15} A future extension of the outlined model should surely also take into account physiological and neurophysiological processing stages.

C Modality-specificity

The first step in differentiating processing stages is to clearly distinguish between domain-/modality-specific and non-domain-/non-modality-specific properties/features (table 2): Acoustic and optical properties depend on physical properties that are not domain-specific themselves: for example, material and structural properties. Correspondingly, auditory and visual features influence other perceptual features that are not modality-specific themselves: for example, perceived material and structural features or aesthetic impressions.

Modality-specific or unimodal features such as loudness presume information from a specific sensory system, and may at the same time only be applied meaningfully and in a non-metaphoric, denotative manner to the system-specific sensation. Within the processes of extracting increasingly abstract perceptual features, modality-specific features are low-level features as a rule. The terminology low-level, mid-level and high-level feature reflects this hierarchy of abstraction.\textsuperscript{16}

In contrast, non-modality-specific features exploit information from several sensory systems variably, resulting in high-level features. They


may be further differentiated into intermodal and supramodal features: Intermodal features, typically matching features such as the perceived synchrony, result from a comparison of at least two modalities on the ground of their intersecting or coincident unimodal percepts; this process plays an important role in the development of perception in infancy.\textsuperscript{17} Supramodal features such as perceived location, room size, aesthetic impressions or perceived emotions, however, appertain to superordinate areas to which different modalities may or may not contribute.\textsuperscript{18}

**TABLE 2.** Factual and terminological distinction between modality-specificities

<table>
<thead>
<tr>
<th>Specificity</th>
<th>modality-specific</th>
<th>non-modality-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation</td>
<td>unimodal</td>
<td>intermodal</td>
</tr>
<tr>
<td>Condition for the formation of the percept (the signified)</td>
<td>information from a specific modality is available</td>
<td>information from two or more specific modalities is available</td>
</tr>
<tr>
<td>Coverage of the term (the signifier)</td>
<td>specific modality</td>
<td>relation of specific modalities</td>
</tr>
<tr>
<td>Examples of perceptual features</td>
<td>loudness, pitch, timbre, brightness, contrast, color</td>
<td>synchrony, synlocation</td>
</tr>
</tbody>
</table>

Such non-modality-specific features express the mental reconstruction of the physical world and accompany its processing as information. This is why many material and structural properties/features appear in both the physical and the perceptual realm, and may be respectively


described by means of the same terms and units. For example the combination of the properties/features «room, box-shaped, wood-paneled, height 3 m, width 4 m, length 7 m» is valid within the physical and the perceptual realm. In contrast to modality-specific features, perceived material and structural features may be therefore reliably compared with physical material and structural properties. Naturally, this does not hold for the results of the further processing of this information: for example, aesthetic impressions.

D Effect directions

The categorization of properties and features by means of ontological realms, modalities/domains, and processing stages allows for the denotation of several effect directions, leading from independent toward dependent variables that may be empirically tested. I suggest the following denotations for the most important effect directions (figure 1):

- **Intra-modal effects** (B→C, D→E), for instance, the effect of frequency spectrum (B) on the timbre/pitch perception (C) of a violin.
- **Cross-modal effects** ([D]B→E, [B]D→C), for instance, the effect of color (D) on the perceived loudness (C) of a train emitting a constant sound pressure level [B];¹⁹ the squared brackets indicate a required additional stimulus which does not need to be varied; in the special case of the investigation of genuine synaesthesia, however, the additional stimulus is unnecessary.
- **Trans-domain effects** (A→B, A→D), for instance, the effect of surface material (A) on the reverberation time (B) and color spectrum (D) of a concert hall.
- **Trans-modal effects** (C→F, E→F), for instance, the effect of perceived loudness dynamics (C) and colorfulness (E) on the aesthetic impression (F) of a TV program.
- **Supra-modal effects** (A→F), for instance, the effect of physical distance on the perceived distance of a singer.

The modal specificity of perceptual features (sec. IV C) is closely related to the process of multimodal integration raising a crucial question known as the binding problem. Experimental paradigms in this field involve different times of domain-specific stimuli and responses, respectively.20 With regard to those experiments the proposed outlined model might be orthogonally extended by the dimension of time. Thus, multiple sets of basic effect directions according to figure 1 might represent different time layers.

### E Operationalizing independent variables by means of dependent variables

Normally, experimental test conditions are varied by means of the selection or manipulation of (physical) stimuli or sensory organs. However, in some prior experiments, participants have been asked to rate a supra-modal feature, but to take into account information derived from only one modality (e.g. »participants were instructed to judge the emotion perceived auditorily«21; »auditory expressivity«22). Such instruc-

---


tions entrust test participants with the task of dissociating auditory and visual information. Which is to say, test participants are supposed to perceive a feature based on two modalities as if it relied on just one modality, despite the fact that such features rely upon audiovisual integrative processes that occur prior to (or in the course of) evaluation. Test participants might attempt to comply with this challenge by means of directed attention; however, it is not clear whether participants are capable of suppressing both their conscious perception of excluded unimodal features, and the routing of unimodal information through preconscious audiovisual integrative processes. Alternatively, test participants might try ex post to assess the proportion of the non-modality-specific feature that is based on the demanded modality; however, the degree of validity and reliability with which this cognitive task may be effected remains unclear. Moreover, it is not clear which of the two strategies a given participant might pursue. Hence, compared to the experimental variation of (physical) stimuli, such instructions cannot guarantee a proper dissociation of auditory and visual information. They are useful at most for the investigation of directed attention.

F Interaction effects

General audiovisual cooperation is often referred to as »audiovisual interaction. However, this term does not account for the different ways in which the senses cooperate (e.g. mutually supporting, mutually interfering, non-monotonic contributing), and at least from an empirical point of view its use is frequently incorrect. In a methodological rather than a colloquial sense, interaction effects occur whenever the effects of one independent variable depend upon a second independ-

---


http://dx.doi.org/10.14279/depositonce-6424
ent variable (or variables) in a non-additive way.\textsuperscript{24} In order to be able to reveal optoacoustic interaction effects, acoustic and optical variables have to be dissociated. Which is to say, we must be able to isolate acoustic and optical properties as factors rather than just levels of one factor, and we must be able to vary them mutually independently. Only by doing so can the two main effects and the interaction effect on each auditory, visual, and audiovisual variable be independently quantified. The question of whether such a quantification is desired leads to the issue of design paradigms.

\section*{G Design paradigms}

Disregarding the considerations made above, some studies on the so-called audiovisual interaction treat hearing and sight as levels of factors rather than as factors themselves. Typically, such studies apply a factor called \textit{presentation mode} or \textit{modality} based on three levels (optical, acoustic and optoacoustic). At least formally, such designs are not apt to achieve their aim. They apply a principle of variation that might be called the \textit{co-presence (CP) paradigm}, since only the presence of the acoustic and optical stimuli are varied. Because no percepts normally occur in the absence of both acoustic and optical stimuli, in a full-factorial design one cell is empty; which is to say, the acoustic and the optical stimuli have not been mutually independently varied. Hence, the respective influence of the optical and acoustic domains may not be determined, let alone an optoacoustic interaction effect. Moreover, the CP paradigm can only yield data about two conditions: a condition in which one modality completely lacks input, and a condition in which both stimuli are optoacoustically congruent. The first is ecologically invalid in general and might change the mode of perceptual processing, while the second is epistemically unproductive.

In order for the acoustic and optical stimulus components to constitute mutually independent variables, as demanded for the determination of interaction effects, in a full-factorial design not their presence

\footnotesize{\textsuperscript{24} DÖRING & BORTZ, 2016, p. 533.}
but their properties must be varied. This approach can be called the conflicting stimulus (CS) paradigm. A conflicting stimulus apt to investigate audiovisual localization might be realized by the synchronous presentation of the acoustic and optical components of a speaker in different locations. In contrast to the CP paradigm, the CS paradigm provides sensory input to both modalities on all combinations of factor levels. Because the CP and CS paradigms speak to different conditions of optoacoustic experience and information-processing (see above), a comparison of both paradigms is of interest for theory-formation (although the CP paradigm is not useful for the comparative quantification of domain-specific and modality-specific main and interaction effects, nor for the consistent collection of unimodal features). In view of integrative data collection (sec. III), merging both paradigms would be productive, as illustrated in table 3.

**TABLE 3.** Comparison of design paradigms: co-presence paradigm (left), conflicting stimulus paradigm (center), and integrated design (right). Abbreviations: O=optical, OA=optoacoustic, A=acoustic, v=visual, av=audiovisual, a=auditory. Light grey shading indicates uni-domain stimuli, dark grey shading congruent two-domain stimuli, and no shading incongruent (conflicting) two-domain stimuli.

<table>
<thead>
<tr>
<th>C0-presence</th>
<th>Optoacoustic properties</th>
<th>Acoustic properties</th>
<th>Perceptual features</th>
<th>Optical properties</th>
<th>Acoustic properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-presence</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Due to the large number of cells, using within-subject factors might be practical. Analysis of variance (multivariate or univariate, with or without repeated measures) lends itself to data analysis in view of RQs of type 1 to 3 and, if applicable, 4. RQs of type 3 and 4 might, however, require other, more specific designs and other statistical approaches,
such as analysis of covariance, multiple regression analysis, or structural equation modeling.

H Optoacoustic congruence

The quantification of the proportionate contribution of hearing and sight to perceptual features (sec. III RQ 1) in particular requires compliance with further methodological criteria.

In reality, the acoustic properties of objects such as items, persons, rooms, environments, and whole scenes are indissolubly interrelated to their optical properties due to physical laws, and integrated perception processes formed over the course of the corresponding long-termed experience. Under ecologically valid conditions, stimulus objects generally sound as they look: which is to say, they are optoacoustically congruent. The possibility of optoacoustic congruence is thus a prerequisite for ecologically valid applications of the CS paradigm. Accordingly, while the CS paradigm is designed to break down optoacoustic congruence in controlled measure, it still establishes the zero-conflict combination (perfect congruence) as a reference point from which other combinations of factor levels can be made to deviate.

The physical and perceptual determination of the degree of optoacoustic congruence is, however, only possible on the basis of several (i.e. complex) acoustic and optical properties. For example, illuminance and sound pressure level alone are not sufficient for this purpose. Because the determinability of congruence increases with the diversity of available acoustic and optical stimulus properties, perceptually-relevant physical cues must not be removed from the acoustic nor the optical domain when presenting the experimental stimuli. I refer to the maintenance of perceptually-relevant physical cues of stimulus objects as rich cue condition\(^{25}\), or in the ideal case as full cue condition.

Hence, a reliably determinable high optoacoustic congruence, as demanded by the CS paradigm, requires (1) A stimulus that is based on a real, naturally-occurring object, and (2) Its presentation under the rich or full cue condition.

I Simulation

Optoacoustically-conflicting stimuli cannot be practically realized using real stimulus objects (sound and light sources and transmission systems, respectively) due to their natural optoacoustic congruence, as the above example of the speaker indicates (sec. IV G). Conflicting stimuli are incongruent by definition, and must therefore be simulated. In order to maximize the ecological validity of the simulated stimuli in general, and the optoacoustic congruence required by the zero-conflict combination in particular, the simulation has to be data-based (as opposed to numerically modeled) – that is, it has to display real instead of virtually-designed objects. According to the rich or full cue condition, the simulation is furthermore required to be as transparent and immersive as possible. This may be achieved by technical features such as a transmission path with sufficiently high temporal and spatial resolution, the application of 3D audio and 3D video transmission methods, a largely nonrestrictive viewing/listening angle, a correct acoustic and optical projection geometry, the reproduction of real energetic conditions (sound pressure level, illuminance, dynamics), and shielding from distracting information. Of course, the question of whether the applied empirical methods of data collection yield results comparable for real stimuli and stimuli simulated in this way must be empirically validated.

J Commensurability

Ensuring optoacoustic congruence is still not sufficient for an experimental appraisal of RQ 1 (sec. III). An internally valid quantitative comparison of the contributions of the senses demands an identical range of stimuli. In other words, the respective variation of the acoustic and optical experimental stimuli must be quantitatively commen-
surable. A reasonable quantitative comparison of ranges requires their qualitative commensurability in turn. Naturally, this does not apply to the specific acoustic and optical properties expressed by means of different physical measures, for example, illuminance and sound pressure level. To apply the same numerical ranges to these two measures in an experiment would be to compare apples and oranges. Using an identical physical quantity, for example the power $P$, and applying an identical numerical range in both domains would not offer a solution to this principle problem, because the resulting acoustic or optical range would not be in line with power ranges of a real, naturally-occurring stimulus object. Thus, the respective stimulus components were neither ecologically valid nor optoacoustically congruent (sec. IV H). Qualitative commensurability of domain-specific properties may not be achieved in this way.

The commensurability problem may, however, be solved by relying on the qualitative commensurability of non-domain-specific properties, upon which the various domain-specific properties themselves depend. Given an experimental situation in which factor levels comprise the acoustic and optical components of several optoacoustically-congruent stimulus objects, the range of the complex acoustic properties will largely correspond to the range of the complex optical properties. This is because on each level, the optical and acoustic properties derive from the same non-domain-specific properties such as physical materials, structures and dimensions. Even though the complex variation of independent variables does not allow for an internally valid identification of unidimensional factors, it does allow for the methodologically-founded experimental appraisal of the basic RQ 1 (sec. III). Ensuring quantitative and qualitative optoacoustic commensurability likewise requires a transparent simulation in order to meet the rich/full cue criterion (sec. IV H).
V Application: The Virtual Concert Hall

A Thematic background

The author of this paper is currently investigating the above research questions 1 through 4 within the scope of an experimental research project on audiovisual room perception. The stimulus objects utilized in this project are rooms themselves. Presupposing a constant illumination, a room shows its substantial optical properties in the form of light distribution without any further contribution; which is to say, a room itself may be described as an optical stimulus from both a functional and a perceptual point of view. A room is not, however, a self-contained acoustic stimulus. Within the acoustic domain it is just a transmission system. Thus, the project requires sound sources capable of exciting the room's acoustics. So in addition to the above considerations, the investigation of room acoustics requires a differentiation between transmission system (room) and content (performance). In order to ensure the internal validity of the experiments, the content (performance) has to be held constant.

A research tool whose development was exclusively geared to the above-mentioned methodological criteria is the Virtual Concert Hall, an optoacoustic virtual environment for the presentation of artistic renditions in performance rooms. It allows for the mutually independent variation of the optical and acoustic components of both the artistic renditions themselves and the spaces in which they are staged. Thus, although the factual stimulus objects are both renditions and rooms, it is possible to center the rooms themselves as the stimulus objects of primary interest by holding the rendition constant across rooms. On the other hand, it also makes sense to vary the type of rendition (for example, music or speech) independently from the rooms, in order to improve the external validity of the experimental results regarding room perception.
Technical stages towards the realization of the Virtual Concert Hall include the acquisition of room properties, the production of artistic renditions, the merging of rooms and renditions, and the setup of a reproduction system – within both the acoustic and the optical domain, respectively. The acquisition of room acoustic properties was carried out by recording binaural room impulse responses (BRIRs) for different azimuthal head orientations of a head-and-torso simulator. The optical properties of the rooms were acquired in the form of stereoscopic full-panoramic images. The acoustic renditions were taken in an anechoic chamber applying poly-microphony and multitrack recording. The optical renditions were stereoscopically recorded in a greenbox studio applying full playback. At the moment of reproduction, the acoustic renditions were embedded into the rooms by means of dynamic binaural synthesis, originally referred to as binaural room scanning. This compensates for the head movements of listeners, resulting in a constant space-related localization across different head orientations. The resulting audio signal is reproduced by the use of an extra-aural headset and a DSP-driven power amplifier providing a linearized transfer function of the audio reproduction system. The optical renditions were embedded frame-by-frame into the rooms by means of chroma-key compositing, the addition of shadows, and the correction of colors. The stereoscopic semi-panoramic high-resolution videos

generated thereby were projected on a semi-cylindrical screen \((d = 5 \text{ m}; h = 2.8 \text{ m})\). In this manner, a large field of view \((> 160^\circ)\) and a reasonable physical resolution \((4754 \times 1872 \text{ pixels})\) corresponding to about 2.7 times the human eye’s angular resolution could be realized. A secondary reproduction system based on an 85″ flat screen was also designed, providing much better angular resolution close perceptual threshold at a more restricted field of view. For experimental purposes, a test sequence control and an electronic questionnaire were also programmed.

C Features of the Virtual Concert Hall

Due to the combination of acquired room properties, produced content, and a state-of-the-art reproduction system, the Virtual Concert Hall allows for the presentation of identical artistic renditions under rich cue conditions in performance rooms with independently-variable acoustic and optical properties. Thanks to several technical enhancements regarding the spatial resolution of the BRIRs,\(^{30}\) the system latency,\(^{31}\) the compensation of the headphone transfer function,\(^{32}\) and the adaptation of the interaural time differences to the individual listener,\(^{33}\) the applied binaural synthesis system provides a highly plausible three-dimensional reproduction.\(^{34}\) Sound sources and room reflections may be perceived omnidirectionally. Because the virtual environment does

---


not turn with head movements, the recipient may actively explore the virtual scenes. By acquiring further transmission systems and recording further content, the Virtual Concert Hall may be adapted for the data-based simulation of numerous scenes, allowing for the experimental investigation of the diverse research questions outlined in sec. III.

Acknowledgements

This publication is based on the project »Audio-visual perception of acoustical environments« (MA 4343/1-1) within the framework of the research unit *Simulation and Evaluation of Acoustical Environments (SEACEN)*, funded by the German Research Foundation (DFG) and coordinated by the Audio Communication Group of the TU Berlin.

Bibliography


Hans-Joachim Maempel


http://dx.doi.org/10.1525/mp.2003.20.3.241


HORBACH, ULRICH, KARAMUSTAFAOGLU, ATTILA, PELLEGRINI, RENATO, MACKENSEN, PHILIP, & THEILE, GÜNTHER: »Design and applications of a data-based auralization system for surround sound«, in: 106th AES Convention, Munich 1999, Preprint 4976.


LINDAU, ALEXANDER: »The perception of system latency in dynamic binaural synthesis«, in: NAG/DAGA 2009 Int. Conf. on Acoustics, Rotterdam, pp. 1063–1066.


