

# Toward a Saliency Model for Interactive Audiovisual Applications of Moderate Complexity

Ulrich Reiter, Q2S - NTNU, Trondheim, Norway, reiter@q2s.ntnu.no

---

**Abstract.** To provide users of interactive audiovisual application systems with subjectively high presentation quality of the content (Quality of Experience, QoE), it is usually not effective to increase the simulation depth of the rendering process alone. Instead, by focusing on salient parts of the content, perceived overall quality can be increased without causing additional computational costs. This paper provides the basis for a novel saliency model for interactive audiovisual applications of moderate complexity that is based on influence factors which have been identified in a coordinated series of experimental studies.

---

## 1 Introduction

The question of saliency of objects in audiovisual applications is only recently becoming an issue of examination [1]. Until now, many application systems mainly rely on visual display and feedback to the user, with some kind of “support” in the auditory domain. As the computing power available in home application and consumer electronics systems is constantly increasing, we see a tendency toward integrating more modalities which until now have only been available in specialized Virtual Reality (VR) systems. With this development, users can expect an increased degree of immersion. This is interesting in many aspects, one of them being that applications with higher immersion are generally considered more user-friendly because they provide a feeling of personalization. Additionally, these systems better represent real life and the complexity of real-life experiences by offering information multimodally.

The general problem with this approach is that resources in consumer-oriented application systems are always limited. It is not feasible to perform a fully grown, detailed simulation of multimodal impressions in real-time. Furthermore, the time and investment necessary to develop completely accurate auditory and visual models is as much of a limiting factor for how much detail will be rendered, as is the computational power alone. It is therefore reasonable to focus only on the most important stimuli and leave out those that would go unnoticed in a real world situation. In order to do so, it is necessary to estimate what the most important stimuli or objects in the overall audiovisual percept are.

After giving a definition of saliency in the audiovisual context in section 2, section 3 describes the role of interactivity and presence in the perception of audiovisual quality. Section 4 introduces the saliency model itself. Section 5 summarizes the experiments performed to verify the factors contained in the model. It also dis-

cusses briefly the main results. References to the full papers are given in each subsection. Finally, section 6 gives a short summary and outlook.

## 2 Saliency of Stimuli

In the absence of information about the history of an interactive process, an object can be considered salient when it attracts the user’s visual attention more than the other objects [2]. This definition of saliency originally valid for the visual domain can easily be extended to what might be called “multimodal saliency”, meaning that

- certain properties of an object attract the user’s general attention more than the other properties of that object
- certain objects attract the user’s attention more than other objects in that scene.

Of course, a saliency model requires a user model of perception, as well as it needs a task model. The user model describes familiarity of the user with the objects’ properties, as attention on the properties of an object varies with the user’s background. This correlates to the concept of schemata described in Neisser’s “Perceptual Cycle”, see fig. 1. In Neisser’s model, schemata represent the knowledge about our environment [3]. They are based on previous experiences and are located in the long term memory. Neisser attributes them to generate certain expectations and emotions that steer our attention in the further exploration of our environment. The exploratory process consists, according to Neisser, in the transfer of sensory information (the stimulus) into the short-term memory. In the exploratory process, the entirety of stimuli (the stimulus environment) is compared to the schemata already known. Recognized stimuli are given a meaning, whereas unrecognized stimuli will modify

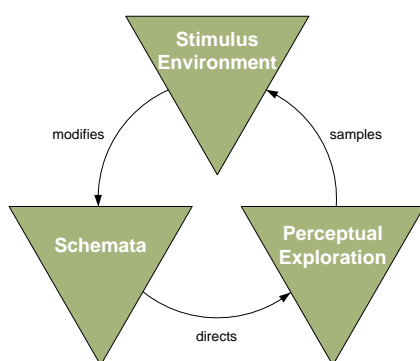


Figure 1: Neisser's Perceptual Cycle, after [3], modified.

the schemata, which will then in turn direct the exploratory process.

Whereas a picture of a human being or a human speech utterance can be considered more or less equally salient to all users, because its significance to humans is embedded genetically, an acoustically trained person might focus more on the reverberation in a virtual room than a visually oriented person. The task model describes the fact that saliency depends on intentionality, so that depending on the task the user is given, his focus will shift accordingly.

Saliency also depends on the physical characteristics of the objects themselves. Following the Gestalt theory introduced by Wertheimer [4], the most salient visual form is the one requiring the minimum of sensory information to be treated. In the auditory domain it is known that certain noises which can be characterized with properties like 'sharpness' or 'roughness' call the attention more than others [5], often by skirting masking effects in the time or frequency domain due to their spectral or temporal characteristics. Adding to this, saliency can be due to spatial or temporal disposition of the objects. Thus a classification of the properties that can make an object salient in a particular context, the so-called *influence factors*, have to be established and verified in order to draw any useful conclusions from the "multimodal saliency" approach.

### 3 Influence of Interactivity and Presence

The concept of interactivity has been defined by Lee et al. based on three major viewpoints: technology oriented, communication-setting oriented, and individual oriented views [6, 7]. Here, the technology-oriented view of interactivity is adopted. The "technology-oriented view of interactivity defines interactivity as a characteristic of new technologies that makes an indi-

vidual's participation in a communication setting possible and efficient" [7]. It is this individual's quality experience that we are interested in.

Steuer holds that interactivity is a stimulus-driven variable which is determined by the technological structure of the medium [8]. According to Steuer, interactivity is "the extent to which users can participate in modifying the form and content of a mediated environment in real time" - in other words, the degree to which users can influence the target environment. He identifies three factors that contribute to interactivity:

- *speed* (the rate at which input can be assimilated into the mediated environment)
- *range* (the number of possibilities for action at any given time)
- *mapping* (the ability of a system to map its controls to changes in the mediated environment in a natural and predictable manner)

These factors are related to technological constraints that come into place when an application is supposed to provide interactivity to the user. Adding to these factors, the perceived quality of the system's feedback to the user - the quality of the audiovisual stimuli generated as a reaction to the user's input - plays an equally important role.

Closely related to interactivity is presence. Presence in interactive audiovisual application systems or Virtual Environments is often described as the feeling of "being there" [9] that generates involvement of the user. Lombard and Ditton define presence in a broader sense as the "perceptual illusion of nonmediation" [10].

According to Steuer, the level of interactivity (degree to which users can influence the target environment) has been found to be one of the key factors for the degree of involvement of a user [8]. Steuer has found vividness (ability to technologically display sensory rich environments) to be the second fundamental component of presence. Along the same lines, Sheridan assumed the quality and extent of sensory information that is fed back to the user as well as exploration and manipulation capabilities to be crucial for the subjective feeling of presence [11].

Interactivity requires attention of the user. Only when both - the system and the user - react to each other, true interactivity is in place. Therefore, interactivity can be regarded as one way of controlling the focus of attention of a user. This is important, as objects in the user's focus will naturally be more salient than others.

### 4 Saliency Model

A saliency model would thus mainly contain (and ideally quantify) the *influence factors* that control the

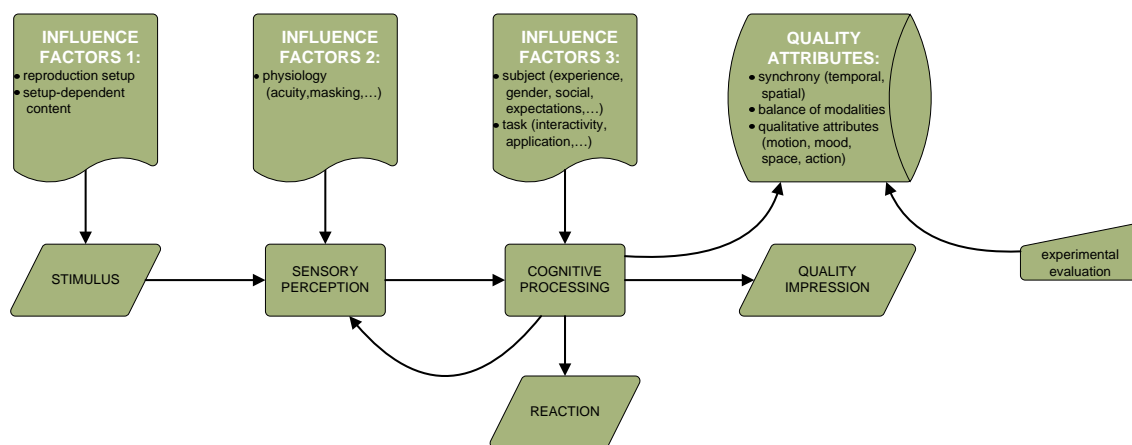


Figure 2: The suggested saliency model for interactive audiovisual applications of moderate complexity.

level of saliency of each perceived object. It is easily seen that a generalized saliency model is too complex and the influence factors too manifold to cope with at the current state of knowledge. Therefore it is necessary to get away from a generalized saliency model. Instead, it is reasonable to focus on a saliency model valid for interactive audiovisual applications of moderate complexity.

Fig. 2 shows how such a saliency model may be structured. The basis of human perception are the *stimuli*. For interactive applications these are generated by the application system itself, so they will depend on a number of factors: the influence factors of level 1. These factors comprise the audiovisual reproduction setup, e.g. (multichannel) loudspeaker setup, headphones, frequency range, panning laws and -algorithms applied, size and resolution of screen, brightness and color distribution, etc. Note that the actual weight of these factors may also depend on the audiovisual content itself: a static, acoustically “dry” sound source in a frontal position will not be critical to different panning laws or number of loudspeaker channels in the back. Influence factors of level 1 also comprise technical input devices for user feedback to the system. As an example, navigation in a 3D scene can be controlled via computer mouse, keyboard, joystick, accelerometer or other advanced technical devices that offer differing degrees of freedom (DOF) and thus differing amount and precision of control. This in turn influences how precisely the system can react by appropriately producing / modifying the stimuli. To summarize, influence factors of level 1 are those related to the generation of stimuli.

The core elements of human perception have been identified to be *sensory perception* on the one hand and

*cognitive processing* on the other hand. Sensory perception can be affected by a number of influence factors of level 2. These involve the physiology of the user (acuity of hearing and vision, masking effects caused by limited resolution of the human sensors, etc.) as well as all other factors directly related to the physical perception of stimuli.

Cognitive processing produces a response by the user. This response can be obvious, like an immediate reaction to a stimulus, or it can be an internal response like re-distributing attention, shifting focus or just entering another turn of the Perceptual Cycle (see [3]). Obviously, the response is governed by another set of influence factors of level 3. These span the widest range of factors, and also the most difficult to quantify: experience, expectations, and background of the user; difficulty and aim of task (if any); degree of interactivity; type of application; etc. Influence factors of level 3 are related to the processing and interpretation of the perceived stimuli.

Cognitive processing will eventually lead to a certain quality impression (Quality of Experience, QoE) that is a function of all influence factors of levels 1-3. This quality impression cannot be directly quantified by humans. It needs additional processing to be uttered in the form of (quantitative) ratings on a quality scale, as (qualitative) semantic identifiers, and so on.

The common way of assessing the overall quality impression is to evaluate single or combined quality attributes. The scientific community has developed a number of attributes that are believed to be relevant for an overall audiovisual quality impression. Among these are audiovisual synchrony (both temporal and spatial), the localization of events, sound as well as video quality by themselves (which, nevertheless, influ-

ence each other), responsiveness to interaction (when applicable), and many more.

Woszczyk et al. have tried to arrange these into a  $4 \times 4$  matrix of “perceptual dimensions” (Action, Mood, Motion, Space) vs. attributes (Quality, Magnitude, Involvement, Balance) within these dimensions [12]. But again, a quantification of their impact is hardly possible as of now. This is because their weight not only depends on the audiovisual content (the stimulus) under assessment, but also on the experimental evaluation (the test methodology) itself. An attribute that is explicitly asked for will probably be assumed to be of higher importance by the test subject (we know from our experience that only important things are asked for in any kind of test). The subject’s attention will be directed toward the attribute currently under assessment, an act that distorts unbiased perception of the overall multimodal stimulus. Therefore, the subject’s reaction might be influenced as well.

## 5 Experimental Evaluation

A number of subjective assessments have been performed to verify the influence factors that were previously identified for a typical prototype reproduction setup of interactive audiovisual content. This exemplary setup makes use of a large projecting screen, a multichannel loudspeaker setup, and real-time room-acoustic simulation rendered on a standard PC. Fig. 3 adumbrates the top view of the reproduction setup with the test subject located in the center. A description of the audio rendering engine can be found in [13]. The results of these experiments reveal a number of interesting points:

### 5.1 Multimodality vs. Unimodality

The first two assessments focused on a possible reduction of algorithmic complexity for the bimodal (audiovisual) case compared to the unimodal (audio only) case. The simplifications assessed were directly related to the computational load that the real-time rendering of audio imposes on the processor.

The first experiment [14] evaluated the number of loudspeakers necessary in interactive audiovisual application systems of moderate complexity using a Vector Base Amplitude Panning (VBAP) approach to position the sound sources. The room acoustic simulation method applied was an image source model simulating the early reflections, combined with uncorrelated diffuse reverberation created separately for each loudspeaker channel. For these situations, the complexity of the algorithm (and thus the computational load) is significantly growing with the number of loudspeakers involved. Thus, it is desirable to keep the number of

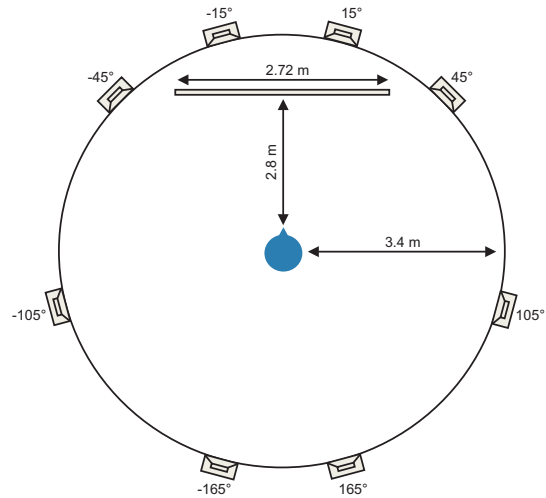


Figure 3: Overview of the test setup: large projecting screen in the front and eight channel loudspeaker setup unevenly distributed between front and back. The higher number of loudspeakers in the screen area (front) increases audiovisual coherence.

loudspeaker channels as low as possible without introducing deteriorations in localization quality.

It has been shown that the necessary number of loudspeaker channels mainly depends on the content itself. Among all tested factors, the different motion paths through the scene that were presented in the assessment (and therefore the directions of incidence of the sound source) had the greatest impact on the perceived subjective quality of the loudspeaker setup used. As a rule of thumb and generalized result, the well-known five-channel setup defined in ITU-R BS.775 [15] is suitable for such systems.

The second experiment [16] evaluated the number of internal workchannels for the MPEG-4 Audio *Perceptual Approach* reverberation algorithm [17]. The MPEG-4 Audio *Perceptual Approach* is also known as the SPAT reverberation algorithm for Cycling’74’s Max/MSP. As opposed to the image source method, it generates reverberation generically, i.e. without a physical/geometrical reference of the room to be simulated. The amount of so-called internal workchannels involved determines the complexity/density of the reverberation pattern. Increasing the number of workchannels therefore generates a higher density reverberation pattern. The density of the reverberation pattern is usually attributed to be the foremost quality criteria of artificial reverberation.

The assessment compared three versions of the *Perceptual Approach* algorithm to each other. The test results make clear that - for the audiovisual case - subjects were not able to identify the three versions of the algorithm under assessment. Increasing the density of the diffuse reverberation part remains without perceivable improvement of the quality in bimodal (audiovisual) perceptual situations. Therefore the *Perceptual Approach* algorithm as specified in MPEG-4 Scene Description can be simplified to use only four internal workchannels without degrading the overall perceived quality in the audiovisual context.

## 5.2 User Interaction

The next three assessments [18, 21, 22] focused on the effect that user interaction with the audiovisual application might have on the perceived overall quality. Here the general assumption was that by offering an attractive interactive content or by assigning the user a challenging task, the user would become more involved and thus experience a subjectively higher overall quality.

The first experiment in this series compared the perceived overall quality of audiovisual scenes under different degrees of interaction [18]. The actual amount of interaction was determined by three different tasks that the test subjects had to fulfill during the assessment. These were:

1. *Listen and watch* task: Test subjects were presented with an automated movement through the virtual scene. No activity on their side was required. The automated movement lasted around 30s, selected from two different predefined motion paths.
2. *Listen and press a button* task: Again, test subjects were asked to experience an automated movement through the virtual scene. This time, an object automatically appeared within the field of view. It was subsequently approached and (again automatically) collected. Then, a new object would appear, and so on. Test subjects were asked to immediately press a button whenever the object appeared.
3. *Listen and collect an object* task: Test subjects were using the computer mouse to navigate freely inside the virtual scene. Their task was to collect the object that was positioned somewhere on the floor. When they had approached the object closely enough, it was collected and re-appeared in another location. The new location was either within the field of view, or the subjects had to turn around to see it again. They were asked to

collect as many objects as possible within a given time limit of 30s.

Interestingly, and contrasting with results obtained by others (e.g. Zielinski, Kassier, Rumsey, Bech et al. in [19] and [20], both based on smaller sample sizes), the different tasks that test subjects had to perform did not have an effect on the quality evaluation (Friedman:  $\chi^2 = 3.3$ ,  $df = 2$ ,  $p > 0.05$ ,  $p = 0.190$ , *ns*).

Two possible explanations exist:

- The subjects' task of navigating through the scene was not demanding enough, making the differences in quality too obvious.
- Whereas user interaction was related to visual and haptic modalities, the quality rating was based on audiovisual percepts. The distraction generated by interaction was not high enough to be significant across modalities.

These possible explanations were examined in the next two experiments [21, 22]. On the one hand, user interaction, rating process and tasks aimed at sharing the same modality. On the other hand, test subjects were confronted with a mentally complex, yet easily scalable task: the n-back working memory paradigm. In this, the subject typically is required to monitor a series of stimuli and to indicate whether or not the stimulus currently presented is the same as the one presented  $n$  steps before.

Here, a sequence of spoken numbers was presented and subjects had to compare the numbers. At the same time, the reverberation time was varied and subjects were subsequently asked to correctly rate the length of reverberation in comparison to a reference reverberation time<sup>1</sup>. Unlike in previously published experiments, both the attribute to be rated and the distracting task were located in the same modality. An analysis of the collected data indicates that the precision with which auditory parameters can be rated / discriminated by humans is dependent on the degree of distraction present in the *same* modality. A highly significant difference in rating accuracy was shown for the "navigation only" condition vs. the "navigation with 2-back task" condition using Wilcoxon's  $T$  test (matched pairs signed ranks,  $T = 20$ ,  $p \leq 0.01$ ).

This result further confirms and specifies the findings of [18, 19, 20]: Whereas cross-modal division of attention only renders a small significant effect and - apart from being listener-specific - depends on the experimental conditions, with inner-modal distraction test subjects would predictably commit errors in their

<sup>1</sup>In [18], an additional semi-structured interview had revealed that reverberation time was regarded as one of the most important attributes for the given type of interactive audiovisual content by all test subjects.

ratings. Apparently, inner-modal influence is significantly greater than cross-modal influence. This is also supported by some of the theories of capacity limits in human attention [23].

### 5.3 Cross-Modal Interaction

Finally, the last assessment [24] in this series investigated the possibility of cross-modal influence of interaction upon perceived quality. Whereas in the previous two assessments the influence of interaction within the same modality was investigated, here the influence of a visual (-motion) task upon the perceived audio quality was evaluated. This experiment is borrowing from what Zielinski et al. [19] and Kassier et al. [20] have described, but the test panel was significantly larger (31 test subjects opposed to 6 and 7, respectively), thus allowing a profound statistical analysis.

For this experiment, a computer game was designed to assess the effect of divided attention in the evaluation of audio quality during involvement in a visual task. Subjects had to collect selected flying objects (donuts) by running into them and avoid the collision with other objects (snowballs). For the navigation, test subjects used the left and right arrow keys of a computer keyboard. Movement was only possible to the sides, at a fixed distance from the source of the flying objects.

A game score was recorded for each subject to verify subjects' involvement in the game and to prod the subjects to actively play the game. By collecting the right object (donut) the score was increased by one point, whereas a collision with a snowball decreased the score by one point.

For the experiment, each subject carried out a passive and an active session. The active session consisted in playing the computer game and evaluating the audio quality. This session was designed to cause a division of attention between the process of rating the audio quality and the involvement in the computer game. In the passive session, subjects were asked to evaluate the audio quality while only watching a game demo. The audio quality degradations were realized by modifying the tonal quality. The original music signal ( $16kHz$ ) was low-pass filtered using three different cut-off frequencies  $f_c = 11kHz$ ,  $12kHz$  and  $13kHz$ . Additionally, an anchor with a low-pass filtering at the cut-off frequency  $f_c = 4kHz$  was created.

The Wilcoxon  $T$  test showed that the quality ratings of the active session varied significantly from the ratings of the passive session for cut-off frequencies up to  $12kHz$ . A significant decrease in rating correctness was shown for the *Game* condition in comparison to the *No Game* condition for the anchor item ( $T = 37$ ,  $p \leq 0.01$ ), the cut-off frequency  $f_c = 11kHz$

( $T = 452.50$ ,  $p \leq 0.01$ ), and the cut-off frequency  $f_c = 12kHz$  ( $T = 812$ ,  $p \leq 0.01$ ). The low-pass filtering in the active session (*Game* condition) was rated as being generally less perceptible.

This assessment showed that a cross-modal influence of interaction is possible when stimuli and interaction are carefully balanced. Interaction performed in one modality (e.g. visual-haptic) can dominate the perception of stimuli in another modality (here: auditory). Yet, at this time it is not possible to determine or quantify that balance *a priori*.

### 5.4 Conclusions

The experiments have clearly identified a number of factors that influence the perceived quality of audiovisual content. These are of technical nature, i.e. depending on the reproduction setup and simulation algorithm used, but also of contextual and subjective nature, i.e. depending on user task, on degree and modality of interaction offered, and on individual attention capacity limits.

## 6 Summary and Outlook

The model introduced here identifies and classifies the most important influence factors that determine the saliency of objects in a multimodal perceptual situation. It has been specifically developed to describe the perception of audiovisual content in interactive application systems of moderate complexity, yet it can be extended to include true multimodality. It is based on the experimental evaluation of perceived overall quality (Quality of Experience) tested in a coordinated series of subjective assessments.

The model needs further refinement to be put to use in real-world applications. One of the tasks that remain is the context-dependent quantification of the influence factors: in its current state of development, the model is a purely qualitative one that does not yet allow *a priori* statements (quantified estimations) on the weight of individual factors.

## References

- [1] Reiter, Ulrich. *On the Need for a Saliency Model for Bimodal Perception in Interactive Applications*. IEEE/ISCE'03, International Symposium on Consumer Electronics. Sydney, Australia, December 3-5, 2003.
- [2] Landragin, Frederic; Bellalem, Nadia; Romary, Laurent. *Visual Saliency and Perceptual Grouping in Multimodal Interactivity*. Proc. International Workshop on Information Presentation and Natural Multimodal Dialogue IPNMD. Verona, Italy, December 14-15, 2001.

- [3] Farris, J. Shawn. *The Human Interaction Cycle: A Proposed and Tested Framework of Perception, Cognition, and Action on the Web*. PhD Thesis. Kansas State University, USA, 2003.
- [4] Wertheimer, Max. Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*. 4, 1923, pp 301-350.
- [5] Zwicker, Eberhard; Fastl, Hugo. *Psychoacoustics - Facts and Models*. 2nd updt. ed., Springer Verlag. Berlin, 1999, ISBN 3-540-65063-6.
- [6] Lee, Kwan Min; Jin, S. A.; Park, N.; Kang, S. *Effects of narrative on feelings of presence in computer/video games*, Annual Conference of the Internat. Communication Association (ICA), New York, NY, USA, May 2005.
- [7] Lee, Kwan Min; Jeong, Eui Jun; Park, Namkee; Ryu, SeoungHo. *Effects of Networked Interactivity in Educational Games: Mediating Effects of Social Presence*, PRESENCE2007, 10th Annual International Workshop on Presence, Barcelona, Spain, Oct. 25-27, 2007, pp 179-186.
- [8] Steuer, Jonathan. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*. 42/4, 1992, pp 73-93.
- [9] Larsson, Pontus; Västfjäll, Daniel; Kleiner, Mendel. *On the Quality of Experience: A Multi-Modal Approach to Perceptual Ego-Motion and Sensed Presence in Virtual Environments*. Proceedings First ISCA ITRW on Auditory Quality of Systems AQS-2003. Akademie Mont-Cenis, Germany, April 23-25, 2003, pp 97-100.
- [10] Lombard, Matthew; Ditton, Theresa. *At the Heart of it All: The Concept of Presence*. *Journal of Computer-Mediated Communication*, 3, 1997.
- [11] Sheridan, Thomas B. *Further Musings on the Psychophysics of Presence*. *Presence*, 5/1994, pp 241-246.
- [12] Woszczyk, Wieslaw; Bech, Soren; Hansen, Villy. *Interactions Between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes*. AES 99th Convention, New York, USA, 1995, Preprint 4133.
- [13] Reiter, Ulrich. *TANGA - an Interactive Object-Based Real Time Audio Engine*. Audio Mostly 2007, 2nd Conference on Interaction with Sound, Ilmenau, Germany, September 27-28, 2007.
- [14] Reiter, Ulrich. *Subjective Assessment of the Optimum Number of Loudspeaker Channels in Audio-Visual Applications Using Large Screens*. Proc. AES 28th Internat. Conf., Pitea, Sweden, June 30 - July 2, 2006, pp 102-109.
- [15] Recommendation ITU-R BS.775-1. *Multichannel stereophonic sound system with and without accompanying picture*. International Telecommunication Union, Geneva, Switzerland, 1994.
- [16] Reiter, Ulrich; Partzsch, Andreas; Weitzel, Mandy. *Modifications of the MPEG-4 AAB-IFS Perceptual Approach: Assessed for the Use with Interactive Audio-Visual Application Systems*. Proc. AES 28th Internat. Conf., Pitea, Sweden, June 30 - July 2, 2006, pp 110-117.
- [17] Int. Std. (IS) ISO/IEC 14496-11:2004. *Information technology - Coding of audio-visual objects - Part 11: Scene description and Application engine*. Geneva, Switzerland, 2004.
- [18] Reiter, Ulrich; Jumisko-Pyykkö, Satu. *Watch, Press and Catch - Impact of Divided Attention on Requirements of Audiovisual Quality*. 12th Internat. Conf. on Human-Computer Interaction, HCI2007, Beijing, PR China, July 22-27, 2007.
- [19] Zielinski, Slawomir; Rumsey, Francis; Bech, Soren; de Bruyn, Bart; Kassier, Rafael. *Computer Games and Multichannel Audio Quality - the Effect of Division of Attention Between Auditory and Visual Modalities*. Proc. AES 24th International Conference on Multichannel Audio, Banff, Alberta, Canada, June 2003.
- [20] Kassier, Rafael; Zielinski, Slawomir; Rumsey, Francis. *Computer Games and Multichannel Audio Quality Part 2 - Evaluation of Time-Variant Audio Degradation under Divided and Undivided Attention*. AES 115th Convention, New York, USA, October 2003, Preprint 5856.
- [21] Reiter, Ulrich; Weitzel, Mandy; Cao, Shi. *Influence of Interaction on Perceived Quality in Audio Visual Applications: Subjective Assessment with n-Back Working Memory Task*, Proc. AES 30th International Conference, Saariselkä, Finland, March 15-17, 2007.
- [22] Reiter, Ulrich; Weitzel, Mandy. *Influence of Interaction on Perceived Quality in Audio Visual Applications: Subjective Assessment with n-Back Working Memory Task, II*. AES 122nd Convention, Vienna, Austria, May 5-8, 2007.
- [23] Pashler, Harold. *The Psychology of Attention*. 1st paperback edition, The MIT Press, Cambridge, MA, USA, 1999, ISBN 0-262-66156-X.
- [24] Reiter, Ulrich; Weitzel, Mandy. *Influence of Interaction on Perceived Quality in Audiovisual Applications: Evaluation of Cross-Modal Influence*. Proc. 13th International Conference on Auditory Displays (ICAD), Montreal, Canada, June 26-29, 2007.