INFLUENCE OF INTERACTION ON PERCEIVED QUALITY IN AUDIOVISUAL APPLICATIONS: EVALUATION OF CROSS-MODAL INFLUENCE

Ulrich Reiter

Technische Universität Ilmenau
Institute for Media Technology
Helmholtzplatz 2, 98693 Ilmenau, Germany
ulrich.reiter@tu-ilmenau.de

Mandy Weitzel

Technische Universität Ilmenau
Institute for Media Technology
Helmholtzplatz 2, 98693 Ilmenau, Germany
mandy.weitzel@stud.tu-ilmenau.de

ABSTRACT

This paper presents a subjective assessment among 32 test subjects performed to investigate the question of possible cross-modal division of attention in interactive audiovisual application systems. We give an overview on recent related research, and we describe in detail the experimental setup, the procedure and the analysis of the data obtained. As a result, the experiment described here verifies that interaction or task can have an influence upon the perceived audio quality, even if the interaction / task is performed in another modality.

1. INTRODUCTION

Perceived overall quality in audiovisual application systems is the benchmark that is set when comparing these systems. Yet, produced quality is not equal to perceived quality. Therefore it is not enough to increase the degree of auditory or visual simulation depth by means of higher computational power applied. Low visual quality decreases overall perceived quality, even if audio quality was superb, and vice versa. Therefore, auditory and visual simulation quality need to be matched. Also, a number of potentially influencing factors like the amount of interactivity offered by an application and the resulting divided attentiveness of the user need to be considered.

The effect of interaction upon the perception of quality is largely unknown. A number of experiments have been performed recently which try to give answers. These experiments will be summed up in the next section. On the basis of the results from these experiments we will describe and motivate another subjective assessment performed in order to evaluate a possible cross-modal influence of interaction. This will be described in section 3. Section 4 will present an analysis of the data obtained and will detail the results of the experiment. Finally, in section 5 we provide a conclusion and an outlook.

2. RECENT RESEARCH

Zielinski et al. [1] and Kassier et al. [2] have both reported studies in which they tested whether time-invariant impairments (filtering) and time-variant impairments (drop-outs) used to provide degradations in audio quality of multi-channel audio were equally well detected by test subjects when visual gaming was performed as a parallel task as when subjects were watching a still picture. It was observed that involvement in a visual task may change the results obtained during the evaluation of audio impairments for some experimental conditions. These results were significant, yet concluded from a very small sample size (≤ 7).

Reiter and Jumisko-Pyykkö [3] conducted an experiment in which they asked test subjects to rate the overall perceived quality of an audiovisual scene under three different levels of interaction: watching the scene while an automated translational and rotational movement was performed in a 3D virtual room, pressing a button whenever a ball appeared with otherwise identical automated movement, and collecting a ball by approaching it using the computer mouse as an input device for self-directed navigation in the virtual scene. Although subjects were asked for an overall quality rating, only one auditory attribute was varied between items (speech and music stimuli). As the room acoustic simulation used for audio rendering was based on a mirror source method, the maximum order of mirror sources computed was varied between 1 and 3, resulting in different diffuse reverberation decay curves. Reiter and Jumisko-Pyykkö used both quantitative as well as qualitative analysis methods on a sample size of 40 participating subjects. The results of the study showed that there were no differences in subjects' ratings of perceived overall audiovisual quality when the quality was estimated with a parallel visual task (pressing a button whenever a ball appeared and collecting the ball) compared to the passive watching only task. Both quantitative and qualitative results supported this.

Reiter, Weitzel and Cao [4] performed an experiment with a sample size of 21 test subjects in which they hypothesized that the lack of influence of the visual task upon the perceived overall quality observed in [3] was related to the variations of quality being present only in the auditory stimuli - and thus in a different modality than the visual task. Therefore they performed an audiovisual subjective assessment in which subjects were asked to rate an auditory parameter (reverberation time) while being distracted with an auditory n-back working memory task. In addition to the n-back task subjects had to move through a virtual scene using a computer mouse. Thus the experiment was divided into three evaluation sessions: 'navigation only' task, '1-back task with navigation' and '2-back task with navigation'. Unlike in previously published experiments, both rating and the predominant task (n-back task) took place in the same modality. The ratings of each subject in the three sessions were analyzed in terms of their correctness using the binomial test. It was tested whether subjects' ratings during '1-back with navigation' and '2-back with navigation' task included less correct answers than in the 'navigation only' condition. The participants' correct percent in the '1-back with navigation' condition was not significantly different from the 'navigation only' condition ($p = 0.125$). Yet, in the '2-back with navigation' condition the participants' correct percent was significantly lower than in the navigation only condition ($p = 0.007$). Thus the analysis of the data obtained indicated that the precision with which auditory parameters can be rated by humans is dependent on the degree of distraction in the same modality.

Reiter and Weitzel [5] reported another experiment which further tried to evaluate the role of auditory distraction for the overall perceived quality. A total of 21 test subjects (sample size) were presented with spoken numbers played back via a virtual loudspeaker in a 3D audiovisual virtual scene. While moving through the scene using a computer mouse, subjects were asked to memorize the numbers and indicate whether they were repeating or not in an n-back task with $n = 1$ and $n = 2$. Each correct answer was awarded with a point, and test subjects tried to reach a score as high as possible in each round. The score was recorded as an evidence and measure for the involvement of the subject into the scene. During the assessment, the reverberation time of the audio stimuli (the spoken numbers) was varied dynamically at arbitrary times in each round, always starting with a reference reverberation time. After each round, subjects were asked to rate the reverberation time in comparison to the reference reverberation time: was it much shorter, shorter, equal, longer or much longer? Reiter and Weitzel hypothesized that with an increasing degree of the interaction task (from 'navigation only', to '1-back task', to '2-back task') subjects would become more unconfident with rating the reverberation differences and make incorrect ratings. Therefore the obtained results were transformed into 'correct' and 'incorrect' answers and were statistically analyzed using nonparametric test procedures. The binomial test delivered a significant result ($p \leq 0.05$). In contrast, the test for several related samples which compared the distributions of correct and incorrect answers of the passive task ('navigation only') and the two active tasks ('1-back task' and '2-back task') did not show a significant result (Cochrans Q: $p > 0.05$).

Summing up the results from [4] and [5], there is evidence of interaction or task influencing the perceived quality whenever this takes place in the same modality as the main variations in quality. Yet, it is not clear whether modality alone is the decisive factor, or whether the amount of involvement / immersion can also play a decisive role, thus making cross-modal influence possible[1]. The experiments described in [1] and [2] could give an answer to this question, but the results are vague and rely on a very small sample size. Therefore, we have devised a similar experiment (but with a sample size of 32) to answer this question, also using the cutoff-frequency of the auditory stimuli as varying attribute.

This experiment is described in detail in the next section. We expected to see that interaction / task actually has an influence, which should manifest itself in test subjects giving higher quality ratings in the interactive scenario than in the passive (listen and watch) one. Also, differentiation between quality degradations should be more difficult for test subjects in the interactive scenario.

3. DESCRIPTION OF EXPERIMENT

3.1. Experimental Platform

The subjective assessments have been performed in the listening lab of Technische Universität Ilmenau. The listening lab is a $64m^2$ room acoustically prepared for listening and audio quality assessment experiments ($T_{60} = 0.34s$). It fulfills recommendations ITU-R BS.1116 [6] and EBU Tech 3276 [7].

Audio reproduction was done via two active, full range monitor loudspeakers[2] located behind an acoustically transparent projection screen at the ±15° positions, see fig. 1. The SPL at the listener's position was measured to be $76dB$ during the experiment.

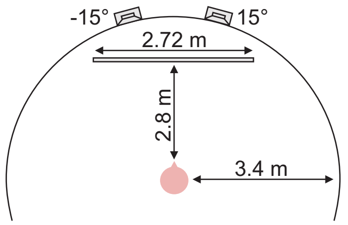

Figure 1: Test setup with the position of the listener, the loudspeakers and the projecting screen.

Visual reproduction was done on a $4 : 3$ projection screen of $2.72m$ of width, located at a distance of $2.80m$ from the test subject, see fig. 1. The picture was produced by an LCD projector mounted in a sound insulated housing, such that the overall background noise level in the listening lab was measured to be below $15dB(A)$. The screen showed a section of the virtual scene that was in accordance to the field of view of the subject.

Subjects were seated in a chair mounted on a platform of $0.40m$ of elevation during the experiment, such that their eyes were at the same height as the center of the screen. Loudspeakers were mounted on loudspeaker stands at the height of the subjects' ears.

The software used for presenting the interactive game was the MPEG-4 player 'I3D'. I3D allows for the reproduction of interactive audiovisual, three dimensional content as defined by the MPEG-4 standard ISO/IEC14496 [8]. I3D provides a modular audio engine which has been used to apply filtering to the audio stimuli. The computer game used in this experiment has been completely written as an MPEG-4 scene. Apart from providing the interactive functionality necessary for the game, the scene description also takes care for the communication with both the input device used for the subject's ratings and the logging tool used to log all subject's activities during the assessment.

Fig. 2 shows a schematic view of the input device used during the assessment. During the anchoring phase of the experiment (see subsection 3.5) subjects could use the two buttons 'imperceptible' and 'very annoying' to familiarize themselves with the quality extremes. During the experiment, the 'start' button was pressed to start a trial. By sliding the fader into the corresponding vertical position, the test subject could make a rating on the perceived quality and transmit it to the system by pressing the 'rating' button. The motorized fader was automatically moved into a neutral position before each trial.

[1] Cross-modal influence means that stimuli perceived in one modality, e.g. visual stimuli, influence the processing and interpretation of stimuli perceived in another modality, e.g. auditory perception. Although cross-modal effects might also happen between other modalities (e.g. tactile stimuli influencing visual perception), in this paper we only look at audiovisual perception.

[2] Genelec 1031 APM

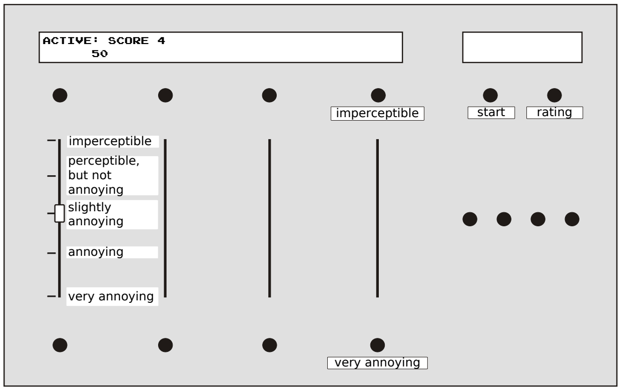

Figure 2: Schematic view of the input device used in the assessment.

The data was collected by a proprietary JAVA tool (SALT[3]) running on a separate computer and connected to I3D and input device via MIDI[4], and subsequently saved into individual XML files along with personal information about the test subjects. After the experiment, all the subjects' ratings were then exported for data analysis in SPSS.

3.2. Test Subjects

A total of 32 test subjects participated in the experiment. The majority of the subjects were students and scientific assistants of Technische Universität Ilmenau. Seven of the participants were females and 25 males (age $M = 25.7$, $SD : 5.36$). Regarding the listening experience, 20 of the subjects belonged to the category of *initiated assessors* and 12 subjects classified as *naive assessors*. The group of initiated assessors had already gained abilities and knowledge in rating the quality of auditory displays in preceding unimodal and bimodal subjective assessments.

3.3. Interactive Scenario

A computer game was created to evaluate the effect of divided attention in the evaluation of audio quality during involvement in a visual task. In this game (for a screenshot see fig. 3) two different types of objects moved through the virtual room in random directions: donuts and snowballs. Subjects had the task to collect selected flying objects (donuts) by running into them and to 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 to the wall on the other end of the room.

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. The actual game score was displayed in the visual scene near the source of the flying objects.

For the experiment, each subject carried out a passive and an active session. The active session involved playing the computer game and evaluating the audio quality. This session was designed to cause a division of attention between the rating of the audio quality and the involvement in a computer game. In the passive session, subjects were asked to evaluate the audio quality while a game demo was presented. Here, the attention of the subjects was directed to the auditory display.

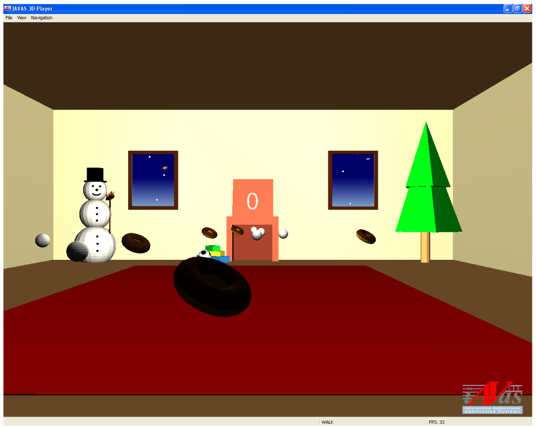

Figure 3: Screenshot: visual scene of the experiment.

3.4. Auditory Scenario

A typical background music for a computer game was chosen for the auditory presentation during the game.

The audio quality degradations were realized by modifying the tonal quality. Therefore the original signal ($20kHz$) 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. Thus three test items, one anchor item and a reference item (corresponding to the original full range signal) were presented to the test subjects in the experiment.

The experiment was performed using a test method (see fig. 4) based on the Degradation Category Rating (DCR), which is standardized in ITU-T P.911 [9]. At the beginning of a trial the reference item was played. During this presentation the tonal quality was changed: the audio presentation was switched to one of the test items, switched to the anchor item, or the reference item was kept unchanged during the transition phase. The exact point in time of the transition between the reference and the test item was changed randomly. Transition took always place in an *area of transition* (see fig. 4) which began after the first eight seconds and ended before the last six seconds of the game. Therefore each trial always began with the reference item and ended with the item to be rated by the subject.

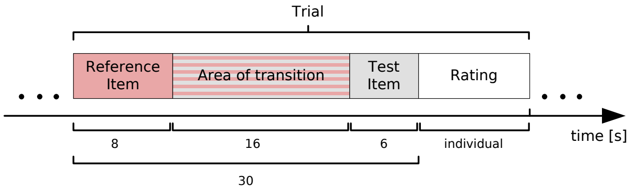

Figure 4: Modified Degradation Category Rating to present the audio material.

After the presentation of the auditory signal (at the end of a

[3] Subjective Assessment Logging Tool
[4] Musical Instruments Digital Interface

gaming session), subjects had to rate the perceived tonal quality degradation using the standardized ITU-T P.911 [9] five-level impairment scale. The scale values and the semantic identifiers of the rating scale as well as the German translations actually used in the experiment are listed in table 1. The ratings of the test items were repeated four times each, those of the anchor and reference item two times each. The order of appearance was changed randomly.

Scale value	Standardized identifier	German identifier
100	imperceptible	nicht wahrnehmbar
75	perceptible, but not annoying	wahrnehmbar, aber nicht störend
50	slightly annoying	etwas störend
25	annoying	störend
0	very annoying	sehr störend

Table 1: Five-level impairment scale with scale values, standardized identifier and German language translated identifier as used in the experiment.

3.5. Procedure

At the beginning of the experiment subjects were presented written instructions which included descriptions of the test procedure, the rating method and the attribute to be evaluated. In case of subjects' questions, additional information was given orally by the experimenters.

Subsequently, subjects could familiarize themselves with the reference item and the anchor item in an anchoring process. The duration of the anchoring process was individually controlled by each subject.

After this the passive and active test sessions followed. They were presented in random order. Instructions and a training session preceded each session. Subjects were briefed in the instructions to concentrate on the game in the active session and to concentrate on the auditory display in the passive session. The training was used to familiarize the subjects with the test method, as well as to practice the navigation with the arrow keys of the keyboard in the gaming session.

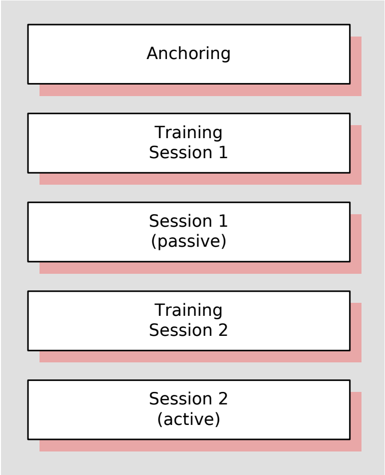

Figure 5: Example procedure of the experiment. The succession of active and passive sessions was determined at random.

4. ANALYSIS AND RESULTS

4.1. Statistical Analysis

The results of all subjects were summarized for the different cut-off frequencies in the passive session (*No Game* condition) and active session (*Game* condition) and are shown in fig. 6.

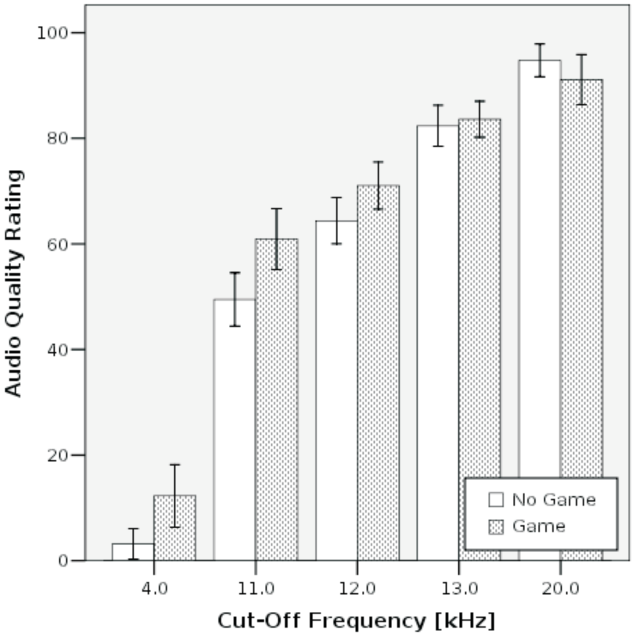

Figure 6: Audio quality ratings for passive session (*No Game* condition) and active session (*Game* condition) for different cut-off frequencies (bars show 95.0% confidence interval of mean).

The bar chart shows that the ratings of the cut-off frequencies $f_c = 4kHz$, $11kHz$, $12kHz$ and $13kHz$ were graded with a better audio quality on average during the active session (*Game* condition) than in the passive session (*No Game* condition).

This tendency was analyzed using tests of significance. Because the Kolmogorov-Smirnov test shows a significant departure from normality ($p < 0.05$), nonparametric tests of analysis were applied. The Wilcoxon test, which compares two dependent samples, shows whether the quality ratings of the active session vary significantly from the ratings of the passive session.

- The difference between the active and passive session ratings for the cut-off frequencies $f_c = 4kHz$ and $12kHz$ was very significant ($p = 0.009$).
- The differences between the active and passive session ratings for the cut-off frequency $f_c = 11kHz$ were highly significant ($p = 0.000$).

For the cut-off frequency $f_c = 12kHz$ the 95% confidence intervals of passive (*No Game* condition) and active (*Game* condition) items overlap. In normally distributed data this is an indicator for not significant variances. As we are dealing with not-normally distributed ratings the confidence intervals presented are not a reliable criterion. Therefore the Wilcoxon test needs to be applied. The results of the Wilcoxon test are shown in table 2.

Fig. 7 presents the rating differences between the active (*Game* condition) and the passive (*No Game* condition) session for differ-

Test Statistics[c]

	Reference (Game) - Reference (No Game)	Anchor (Game) - Anchor (No Game)	Item 12 kHz (Game) - Item 12 kHz (No Game)	Item 13 kHz (Game) - Item 13 kHz (No Game)	Item 11 kHz (Game) - Item 11 kHz (No Game)
Z	-1,161[a]	-2,628[b]	-2,623[b]	-,552[b]	-3,541[b]
Asymp. Sig. (2-tailed)	,246	,009	,009	,581	,000

[a] Based on positive ranks.

[b] Based on negative ranks.

[c] Wilcoxon Signed Ranks Test

Table 2: Results of the Wilcoxon test.

ent cut-off frequencies. It is shown that the subjects graded the audio quality degradation less perceptible in the active session (*Game* condition) than in the passive session (*No Game* condition). Fig. 8 shows that the majority of the subjects graded the audio quality higher during the active session (*Game* condition) than in the passive session (*No Game* condition). Yet, there are also some subjects which graded the audio quality lower while being involved in the game.

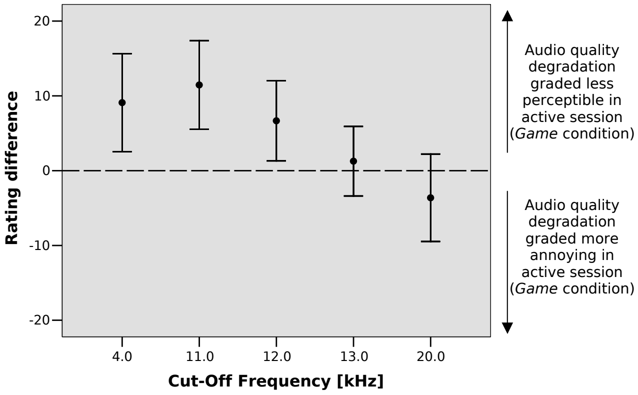

Figure 7: Rating differences between the active and the passive session for different cut-off frequencies (error bars show 95.0% confidence interval of mean).

The score across all subjects in each trial is shown in fig. 9. The relation between the duration of the experiment (trial) and an increasing game score was very significant ($p = 0.002$). The results of the test of correlation are summarized in table 3.

A correlation between the game score and a high rating difference between the active (*Game* condition) and the passive (*No Game* condition) session could not be substantiated (see table 4).

The standard deviation between repetitions of identical items in the passive session (*No Game* condition) served as an indicator for the reliability of the test subjects. A large standard deviation is usually an indicator for an unreliable subject. The mean standard deviation across all subjects was found to be half a step on the five-level impairment scale ($SD = 11.2$ scale values, see table 1).

4.2. Results

The statistical analysis shows that the ratings of the tonal quality degradations in the active session differs from those in the passive

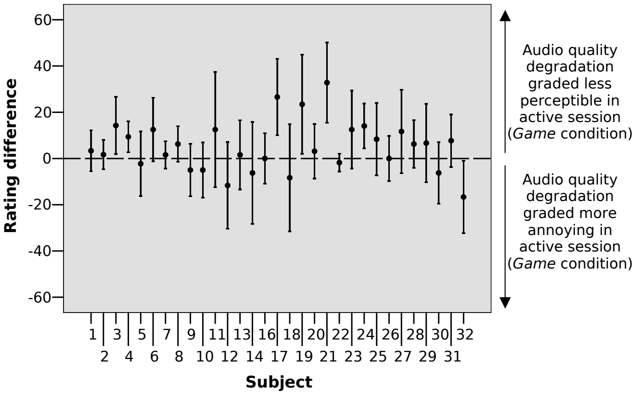

Figure 8: Rating differences between the active and the passive session for all subjects (error bars show 95.0% confidence interval of mean).

Correlations

		Score	Trial
Score	Pearson Correlation	1	,635**
	Sig. (1-tailed)		,002
	N	18	18
Trial	Pearson Correlation	,635**	1
	Sig. (1-tailed)	,002	
	N	18	18

**. Correlation is significant at the 0.01 level (1-tailed).

Table 3: Results of correlation test between the duration of the experiment (Trial) and an increasing game score.

Correlations

		Score	Rating Difference
Score	Pearson Correlation	1	-,019
	Sig. (1-tailed)		,460
	N	32	31
Rating Difference	Pearson Correlation	-,019	1
	Sig. (1-tailed)	,460	
	N	31	31

Table 4: Results of correlation between a high rating difference between the active and the passive session and a high game score.

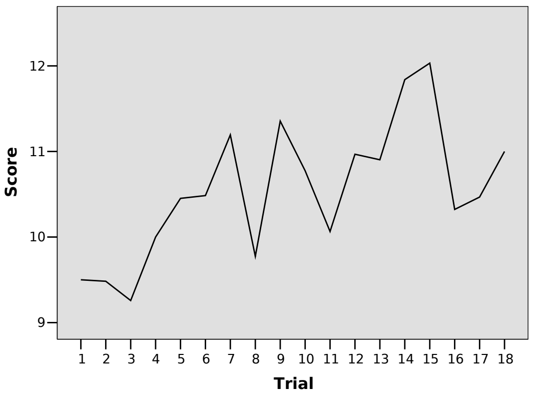

Figure 9: Mean game score against the duration of the experiment.

session. The low-pass filtering in the active session (*Game* condition) was graded as being less perceptible. This effect can be classified as significant at the cut-off frequencies of $f_c = 4kHz$, $11kHz$ and $12kHz$. There is also a difference between the ratings in the active and the passive session at the cut-off frequency of $f_c = 13kHz$, but an effect toward a better rating was not significant. Probably this filtering was not readily discriminable from the original signal.

By recording the game score we were able to verify the active involvement of the subjects in the computer game. The increasing game score over time (duration of the whole experiment) possibly indicates a learning effect of the subjects, see fig. 9.

5. CONCLUSION AND OUTLOOK

In this paper we have described a subjective assessment which investigates the question of possible cross-modal division of attention. Apparently, division of attention not only occurs intramodally as verified in [4] and [5], but also cross-modally. The experiment described here shows that interaction or task can have an influence upon the perceived audio quality in an audiovisual interactive application, even if the interaction / task is performed in another modality. Therefore, when engaged in an interactive game, audio quality degredations are less noticeable than when the same game content is simply displayed in a non-interactive manner. Whether the influence of interactivity is significant or not seems to be related to the amount of quality degradation actually present: slight degradations that would go unnoticed in the unimodal case will not have any influence, whereas large degradations are perceived as being smaller compared to the perception in the non-interactive / no task case.

When comparing this experiment with the findings published in [3], it still remains open what the differences in the results are more related to:

1. Unequal difficulties of the task (complex, free self-movement using a computer mouse in [3] vs. simple left / right movement here) or
2. Varying width of quality steps (unequally easy to detect) between the items (order of reverberation model in [3] vs. low-pass filtering here).

This question will need to be addressed in future experiments. What can be ruled out is the assumption that division of attention might only occur intra-modally. The experiment described here has provided evidence of the potentiality of cross-modal effects.

The effect of cross-modal division of attention might be exploited in future audiovisual applications: By offering attractive and interesting interactivity options equivalent to what in this experiment we called 'task', a user could be distracted from the process of permanently rating the quality of a scene and scanning it for deficiencies in terms of scene realism, thus resulting in a higher overall quality impression.

6. ACKNOWLEDGMENT

This work is supported by the EC within FP6 under Grant 511568 with the acronym '3DTV'.

7. REFERENCES

[1] Zielinski, S.K.; Rumsey, F.; Bech, S.; Bruyn,B.; Kassier, R., "Computer Games And Multichannel Audio Quality - The Effect Of Division Of Attention Between Auditory And Visual Modalities", *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 26-28, 2003.

[2] Kassier, R.; Zielinski, S.K.; Rumsey, F., "Computer Games And Multichannel Audio Quality Part 2- Evaluation Of Time-Variant Audio Degradations Under Divided and Undivided Attention", *Proc. of the AES 115th Convention*, New York, USA, October 10-13, 2003.

[3] Reiter, U. and Jumisko-Pyykkö, S., "Watch, Press and Catch - Impact of Divided Attention on Requirements of Audiovisual Quality", to be published at the *12th International Conference on Human-Computer Interaction*, Beijing, PR China, July 22-27, 2007.

[4] Reiter, U.; Weitzel, M.; Cao, S., "Influence of Interaction on Perceived Quality in Audio Visual Applications: Subjective Assessment with n-Back Working Memory Task", to be presented at AES 30th International Conference on Intelligent Audio Environments, Saariselk, Finland, March 15-17, 2007.

[5] Reiter, U.; Weitzel, M., "Influence of Interaction on Perceived Quality in Audio Visual Applications: Subjective Assessment with n-Back Working Memory Task, II", to be published in *Proc. of the AES 122nd Convention*, Vienna, Austria, May 5-8, 2007.

[6] Recommendation, ITU-R BS.1116-1: Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems. International Telecommunication Union, Radiocommunication Assembly, Geneva, 1997.

[7] EBU Tech 3286, Technical Document: Supplement 1 - Listening conditions for the assessment of sound programme material: multichannel sound. European Broadcasting Union, Geneva, 2004.

[8] ISO/IEC, 14496-1: 2004: Information technology Coding of audio-visual objects. Part 11.1, MPEG-4 Systems Node Semantics, 2004.

[9] Recommendation, ITU-T P.911: Subjective Audiovisual Quality Assessment Methods for Multimedia Applications. International Telecommunication Union, Telecommunication Standardization Sector, Geneva, 1998.