Watch, Press, and Catch – Impact of Divided Attention on Requirements of Audiovisual Quality

Ulrich Reiter[1] and Satu Jumisko-Pyykkö[2]

[1] Institute of Media Technology, Technische Universität Ilmenau,
Helmholtzplatz 2, 98693 Ilmenau, Germany
ulrich.reiter@tu-ilmenau.de
[2] Institute of Human-Centered Technology, Tampere University of Technology,
P.O. BOX 553, 33101 Tampere, Finland
satu.jumisko-pyykko@tut.fi

Abstract. Many of today's audiovisual application systems offer some kind of interactivity. Yet, quality assessments of these systems are often performed without taking into account the possible effects of divided attention caused by interaction or user task. We present a subjective assessment performed among 40 test subjects to investigate the impact of divided attention on the perception of audiovisual quality in interactive application systems. Test subjects were asked to rate the overall perceived audiovisual quality in an interactive 3D scene with varying degrees of interactive tasks to be performed by the subjects. As a result we found that the experienced overall quality did not vary with the degree of interaction. The results of our study make clear that in the case where interactivity is offered in an audiovisual application, it is not generally possible to technically lower the signal quality without perceptual effects.

Keywords: audiovisual quality, subjective assessment, divided attention, interactivity, task.

1 Introduction

Several multimedia applications offer interactivity between the system and the user. In these applications, the technical constraints such as computing power available or error sensitivity of data transmission require some form of quality optimization. To adapt the quality under these circumstances, subjective quality evaluation tests are conducted to assess the signal or system performance. A typical example for such a quality optimization is the optimum distribution of computing power between auditory and visual rendering processes. However, final applications like games typically demand interaction with users. This interactivity or the user's actual task is usually not taken into account in the subjective quality evaluation studies done for quality optimization purposes. This paper investigates whether the requirements of perceived audiovisual quality varies when the evaluation of quality is performed in parallel with interactive tasks of different complexity.

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 943–952, 2007.

2 Audiovisual Perception, Quality and Attention

Audiovisual perception is more complex than the sum of the two sensory channels, and its processes are not known in depth [2]. However, the goal of many audiovisual application systems is to provide unified perception, like in complex every day life perception [5]. The multimodal perception requires proper synthesis of stimuli, which can be violated by asynchrony of auditory and visual material. Audiovisual perception is also dependent on content, for example human talking heads' cross-modal interaction is very high compared to other content types [18]. Experiments of audiovisual quality in different contexts (from multimodal data compression to virtual environments) have also shown that one modality can enhance and modify the experience derived from another modality. The perceived quality in one modality affects the perceived quality in another modality, especially if the qualities clearly differ [1,18,19]. Stimuli presented in accordance in two modalities also improve the feeling of enjoyment and presence compared to one modality in virtual environments. In these environments, presence as "a feeling of being there in space and time" is assumed to be a goal for multimodality and is reached when auditory and visual information merges [12].

Most of the experiments assessing the audiovisual quality have been studied under passive stimuli viewing by focusing all attention on quality evaluation task. On the other hand, many of these evaluations are conducted for systems with active human-computer interaction. In these systems, user's attention is expected to be focused on tasks relevant to user's goals (gaming as entertainment, watch the story of content) rather than quality. To improve the ecological validity of the experiments some previous studies have tackled effects of focused and divided attention on quality evaluations. The main question in these experiments is do we perceive the quality similarly if we pay attention only on quality than if we divided it to some other task simultaneously to quality evaluation task.

To clarify the concepts, attention as information selection process is characterized by limited information processing recourses (overviews e.g. in [14,21]). Studies of focused attention give several inputs for participants and ask them to follow one. Typically the nature of unfocused stimuli is examined. In the divided attention tasks, also called dual task experiments, several input are given and participant is asked to pay attention several of them at the same time which describes the individuals processing limitations. The similarity, difficulty and training of tasks affect to the ability of processing. Taken together, it could be assumed that in the real use of system the focused attention is on the relevant task and not very detailed information is not extracted from unfocused input of quality. This would give an option to provide the technically lower quality without perceptual effects in the relevant use.

Rimell & Owen [20] have studied the impact of focused attention on audiovisual quality with talking head material. In their experiment, participants paid attention on either auditory or visual stimuli. After the presentation they were asked to rate either audio or video quality. The results showed that the modality to which attention is paid dominates over the perceived quality of the other modality. This phenomenon is symmetrical between the auditory and visual senses. On the other hand, when attention is focused on one modality, the ability to detect errors in another modality is

greatly impaired. This study would support the idea to lower level of produced quality in unattended modality without perceptual costs.

Hands [6] has studied multimodal quality perception when dividing attention also to content simultaneously to quality evaluation task. Overall quality of transmitted audiovisual sequences with severe impairments was evaluated. The experiment was conducted with two samples: One sample was asked to evaluate the quality. The other sample was asked to recall the audiovisual content in parallel to performing the quality evaluation. The results showed no difference between the samples, thus indicating that quality ratings are independent of content recall. Practically, these results would mean that the produced quality cannot be lowered eventhough participants would pay attention on content.

Zielinski et al. [22] studied multi-channel audio quality in a computer game. In their study, six participants assessed the audio quality, firstly with gaming as a parallel task and secondly with simultaneously watching static screen shots of the game, by using the single stimulus method with reference. The audio stimuli presented instrumental jazz music with static changes (low pass filtering). Their study concluded some listener specific, but not any global effects. Later on, Kassier et al. [13] have conducted a similar experiment for time variant audio degradations with seven participants. To involve participants even more in the gaming task ("Tetris") they added a more advanced scoring system suitable for short time playing. The study summarized that involvement in the task decreased the consistency of audio quality grading and therefore may impacted in evaluation of audio impairments.

All these previous studies illustrating inconsistent results show a clear need to further study the effect of divided attention on requirements of perceived multimodal quality.

3 Audiovisual Rendering

Most of today's audio visual application systems aim at simulating an accurate representation of the real world by focusing on the (arguably) most important human sense, vision. Auditory stimuli are used in these systems to enhance the overall impression of realism. Still, the stimuli of the two modalities are rendered and presented mostly independently from the other modality. The level of detail in the respective (visual or auditory) simulation is kept as high as possible with regard to computing power available, independently from the level of detail in the other modality.

In contrast, the MPEG-4 standard ISO/IEC 14496 provides a so-called object oriented approach where objects may have both auditory and visual characteristics [7]. These characteristics are attached to the object at the description level, so that they form an integral part of the object itself. A sound source object may have shape and color attributes and at the same time a certain directional pattern for the sound radiation. An obstructing object may have shape, color and (visual) transparency and at the same time acoustic properties like frequency-dependent reflection and transmission characteristics.

Unfortunately, real time acoustic simulation processes are computationally very expensive. Only recently have we seen personal computers capable of handling the

necessary calculations based on the physical and geometrical characteristics of the virtual room to be rendered audible. Still, a significant number of compromises in the accuracy of the simulation have to be accepted for real time performance.

In geometry-based room acoustic simulations that use the so-called mirror-source method, the main factor for computational load is the maximum order of mirror sources that are rendered audible. The order of mirror sources correlates exponentially to the number of mirror sources computed. Each mirror source represents a single early reflection coming from one of the walls of the (virtual) room. In the simulation algorithm used for this experiment, the number of early reflections (and therefore the order of mirror sources) also influences the total amount of reverberation, its strength and its length. Reverberation is increased with increasing order of mirror sources.

In the work described here we have used an MPEG-4 player (I3D) as a platform for subjective assessments of overall perceived quality. The I3D was developed over the last four years at the Institute for Media Technology (IMT) at Technische Universität Ilmenau / Germany. It can render three-dimensional virtual scenes, it allows users to navigate freely inside of these scenes, and it provides real time rendering of auditory simulation via its modular TANGA audio engine [15].

4 Research Method

The tests were conducted at Technische Universität Ilmenau between May and June 2006. Three pilot tests were done prior to finalizing the test set-up. The average duration of the test was 65 minutes including an interview.

Participants - The experiment was conducted with 40 participants, mostly university students, aged from 23 to 39 years (M: 26, SD: 3.6). Ten participants were female and 30 males. All participants reported to have normal hearing. 30% of the participants could be regarded as experienced assessors.

Test procedure - The experiment consisted of three different parts. In the beginning, demo-/psychographic data (age, gender, professionalism in video and audio handling, attendance to earlier listening experiments, playing computer games and instruments and listening experience with surround sound systems) was collected with a pre-questionnaire.

The actual test contained a quality anchoring and three evaluation tasks including a training prior to each of them, see fig. 1. The anchoring introduced the quality extremes of the test materials with different contents. The quality evaluation included three different parallel tasks: *listen and watch* task, *listen and press the button* task and *listen and catch the ball* task. All tasks had the same evaluation instructions and the order of the tasks was randomized between the experiments.

The Single stimulus method, also known as Absolute Category Rating, is suitable for multimedia performance and system evaluation (e.g. ITU-T BT.500 [8], ITU-R P.910 [10]). The stimuli were viewed one by one, overall quality was rated independently and retrospectively (e.g. ITU-R BT.500-11 [8]) on a continuous and unlabelled scale from 0 to 100 in the randomized presentation order. Even though

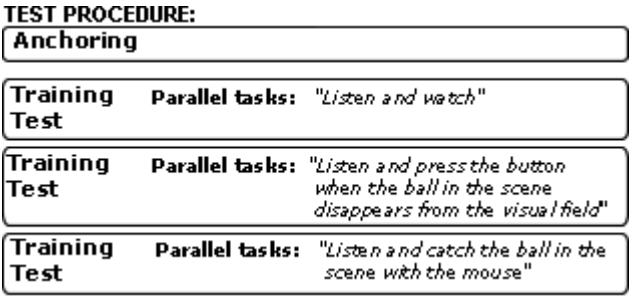

Fig. 1. The actual test procedure was divided into quality anchoring and three different tasks

double and multi stimulus methods are powerful for high quality discrimination, they would have made the quality evaluation with parallel task becoming very complicated for the participants.

The final part of the test session focused on the quality evaluation criteria and the impressions of the evaluation tasks. A semi-structured interview gathered data about the overall quality evaluation criteria with and without parallel task (detailed description in [11]). A post-questionnaire about the experienced easiness of the evaluation tasks and the presented quality in the tasks ended the test session.

Stimulus materials - All test materials were 30 seconds long audio visual contents. Two different audio contents, music (acoustic guitar) and speech (male voice), were presented with three different reverberation strengths: the lowest amount of reverberation was produced by a mirror source algorithm of order one, the highest by an algorithm of order three. Two different audio contents were selected because of the different spectral distribution, familiarity and the preference of reverberation amount [17]. The visual content, a sports gym (see fig. 2, left), was presented with two different motion paths representing a spatial movement within a virtual space (fig. 2, right). These were selected to have an equal number of items with main direction of sound incidence from the left as from the right hand side and they were made as equal as possible between the parallel tasks.

Experimental environment – The experiment was conducted in a laboratory environment in accordance to ITU-R BS.1116 [9] and EBU 3276 [4], suitable for listening tests with wide screen and 8-channel loudspeaker setup. The loudspeaker setup consisted of eight active full-range monitor speakers located in a circular array, with four speakers in the frontal area to increase the precision of localization, and four speakers to the sides and to the back (fig. 3). This particular setup is not standardized, but orientated on the human directional hearing capabilities. The test subject was positioned at the center of the circular loudspeaker array and visual content was displayed on a projecting screen (width 2.7m, viewing distance 2.8m). The sound pressure level within each scene varied depending on the virtual distance between the loudspeaker in the center of the gym and the position of the test subject in the scene (max. SPL 78dB(A)).

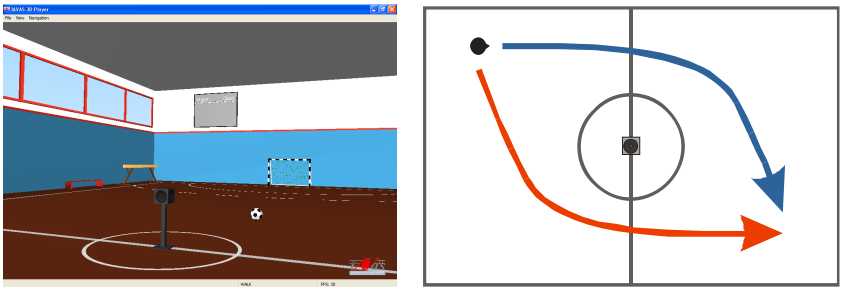

Fig. 2. (left) Visualization of the virtual room (sports gym) as used in the stimulus material. **(right)** Motion paths one and two inside the sports gym.

Data-collection and analysis - During the experiment the data collection was done with the help of an electronic input device especially built for the purpose of audio visual subjective assessments, see [16].

The results were analyzed using SPSS for Windows version 13.0. Non-parametric methods of analysis were applied because the data did not reach the preconditions of normality for parametric methods. Friedman's and Wilcoxon's tests were used to compare the differences between ordinal independent variables in the related design [3]. In the analysis of the questionnaire data, Kuskall-Wallis' and Mann-Whitney U tests were used to compare differences between two groups in the unrelated design [3].

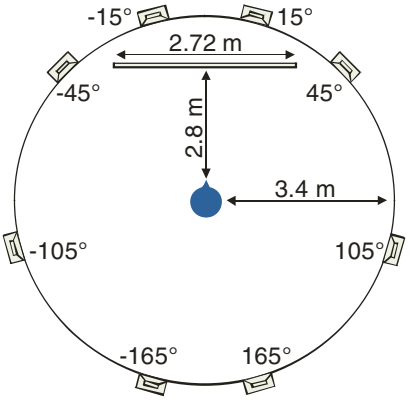

Fig. 3. Loudspeaker and projecting screen setup used in the subjective assessments

5 Results

5.1 Experiment – Tasks, Reverberation Orders, Auditory Content and Visual motion Paths

Tasks did not have effect on the quality evaluation (Friedman: $\chi^2 = 3.3$, df = 2, p>.05, p=.190, ns) when the values were averaged across the reverberation orders, contents and motion paths.

Reverberation strength impacted on the quality evaluation (Friedman: $\chi^2 = 106.6$, df = 2, p>.001). The material presented with the lowest reverberation order was the most pleasant, followed by second order and then third reverberation order. The differences were significant between all reverberation orders when results were averaged over other factors. (Wilcoxon: Order 1 vs. Order 2: Z=-8.16, p<0.001; Order 1 vs. Order 3: Z=-9.87, p<0.001; Order 2 vs. Order 3: Z=-2.43, p<0.05). The results remained the same within task examination, with the exception that there were not significant differences between the second and third reverberation order in any of the parallel tasks (watch: p>.08, press: p>.190, catch: p>.224).

Quality evaluations were not affected by the audio content types or visual motion paths. The music and speech contents were mostly rated into the same level within each task (p>0.05). The exception that the music content was preferred over the speech content appeared with the presentation of the first reverberation order (*watch* task: Wilcoxon Z=-3.01, p>0.01; *press the button* task: Wilcoxon Z=-2.92, p>0.01). When the contents were averaged over the other factors the music content was rated as more pleasant than speech content (Wilcoxon: Z=-2.88, p<0.01).

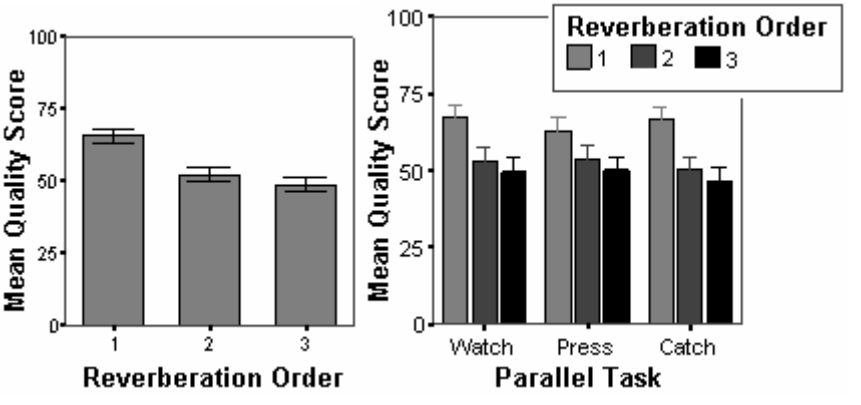

Fig. 3. Error bars show 95% CI of mean

Two different motion paths were equally rated in each task (p>0.05). The only exception appeared in the *listen and press the button* task with music content presented with the second reverberation order (Wilcoxon: Z= -2.2 p>.03).

5.2 Effect of Task and Content Experiences on Quality Evaluation

Evaluation easiness between the tasks: A difference of the evaluation easiness between the tasks was reported by 90% of the participants. The *watch* task was experienced as the easiest, followed by *press the button*, and the hardest *catch the ball* task with significant differences between them (p<0.05). However, the reported evaluation easiness did not impact on the evaluations between the tasks (Krusskal-Wallis: Chi=1.41, df=2, p>0.05).

Quality differences between the tasks: The majority of the participants (62.5%) experienced the presented quality as being the same between the parallel tasks. Within

the group that experienced differences (37.5%), the *watch* task had shown higher quality compared to other tasks ($p<0.001$) which were evaluated being in the same level (Wilcoxon $p>0.05$). There were no differences in the ratings with respect to the level of experienced quality between the tasks (Krusskal-Wallis: Chi= 2.05, df=2, $p>0.05$).

6 Discussion

This study investigated the effects of interaction tasks with different complexity on the requirements of perceived audiovisual quality. Ideally, the goal was to see if the produced quality could be lowered due to interaction without perceptual impact. In the experiment, simultaneously to the overall audiovisual quality evaluation task, participants had to perform three different types of parallel tasks: passive *listen and watch* presentation, *listen and press* the button in the case a visual object appeared, and *listen and catch* the ball tasks. Different reverberation orders, audio and visual contents were varied in the virtual room presentation. Easiness and impressions of presented quality differences between the parallel tasks were gathered with a post-test questionnaire after the experiment.

The results of the experiment showed no differences for the audiovisual quality requirements between parallel visual tasks. This result is supported by Jumisko-Pyykkö & Reiter's [11] earlier results targeting the same problem qualitatively. They concluded the main quality evaluation criteria during the experiment being the different impressions of auditory quality – not the impacts of tasks. The result was the same independently whether the results were drawn from the overall quality evaluation criteria or from the detailed interview material, conducted with different stimuli material and parallel tasks.

Controversially, some previously reported studies have concluded some sporadic changes in evaluations of multi-channel audio when visual gaming was used as a parallel task [22, 13]. These significant results were summarized from very small sample sizes (<7) and with a possibly more involving parallel task than in our study.

Even though in our study participants reported that some evaluation tasks were experienced as being more complicated than others, neither our study nor others' have been able to conclude any real trend in changes of audiovisual quality requirements. Difficulty, similarity and training of tasks are the basic factors affecting dual-task performance [21]. It is possible that the levels of dual-tasks in our experiment were so easy and separate from each other that people were able to divide the attention between the tasks without assumed processing difficulties. In addition, it is possible that these dual-tasks do not distract so much the relatively experienced assessors which we had compared to naïve assessors. Hands' [6] study of quality evaluation and content recall also gives some support for separate processing of content and quality. He concluded that simultaneous content recall did not affect the requirements of multimodal quality in television contents. These results might indicate that the signal quality cannot be technically lowered without perceptual effects in the case where interactivity is involved with the application. Further research conducted with a variety of more complicated and involving tasks, still relevant to user's goals, is needed to confirm this finding.

Acknowledgments

This work is supported by the EC within FP6 under Grant 511568 with the acronym '3DTV'. Satu Jumisko-Pyykkö's work is supported by the Graduate School in User-Centered Information Technology (UCIT) and this publication preparation work by Ulla Tuominen Foundation.

References

1. Beerends, J.G., de Caluwe, F.E.: The influence of video quality on perceived audio quality and vice versa. Journal of the Audio Engineering Society 47(5), 355–362 (1999)
2. Coen, M.: Multimodal Integration - A Biological View. In: Proceedings of IJCAI'01, Seattla, WA (2001)
3. Coolican, H.: Research methods and statistics in psychology, 4th edn. J. W. Arrowsmith Ltd, London (2004)
4. EBU Tech. 3276-E-2nd edn., Listening conditions for the assessment of sound programme material, Geneva (1998)
5. Gibson, J.J.: The Ecological Approach to Visual Perception. Lawrence Eribaum, Houghton Mifflin, Boston (1979)
6. Hands, D.: Multimodal Quality Perception: The Effects of Attending to Content on Subjective Quality Ratings. In: Proceedings of IEEE 3rd Workshop on Multimedia Signal Processing, 1999, Copenhagen, Denmark, pp. 503–508 (1999)
7. ISO/IEC 14496:2001, Coding of audio-visual objects (MPEG-4) (2001)
8. ITU-R BT.500-11 Methodology for the subjective assessment of the quality of television pictures, International Telecommunications Union – Radiocommunication sector (2002)
9. ITU-R BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, International Telecommunication Union, Geneva (1997)
10. ITU-T P.910 Recommendation P.910, Subjective audiovisual quality assessment methods for multimedia application, International Telecommunication Union – Telecommunication sector (1998)
11. Jumisko-Pyykkö, S., Reiter, U.: Produced quality is not the perceived quality – A Qualitative Approach to Overall Audiovisual Quality. In: Proceedings of 3DTV Conference, IEEE (May 2007)
12. Larsson, P., Vastfjall, D., Kleiner, M.: Ecological Acoustics and the Multimodal Perception of Rooms: Real and Unreal Experiences of Auditory-Visual Virtual Environments. In: Proc. 2001 International Conference on Auditory Display, Espoo, Finland (July 29 - August 1, 2001)
13. 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. In: Proceedings of the AES 115th International Conference, New York, USA (October 10-13, 2003)
14. Pashler, H.E.: The psychology of attention. MIT Press, Cambridge, MA (1999)
15. Reiter, U., Schwark, M.: A plug-in based audio rendering concept for an MPEG-4 Audio subset. In: Proc. IEEE/ISCE'04 International Symposium on Consumer Electronics, Reading/UK (September 2004)

16. Reiter, U., Holzhäuser, S.: An Input Device for Subjective Assessments of Bimodal Audio visual Perception. In: IEEE/ISCE'05, International Symposium on Consumer Electronics, Macau SAR/China (June 2005) ISBN 0-7803-8920-4
17. Reiter, U., Großmann, S., Strohmeier, D., Exner, M.: Observations on Bimodal Audio visual Subjective Assessments. In: Proceedings of the 120th AES Convention, Paris, France, Convention Paper 6852 (May 20-23, 2006)
18. Rimell, A.N., Hollier, M.P., Voelcker, R.M.: The influence of cross-modal interaction on audio-visual speech quality perception. In: Presented at the AES Convention, San Francisco, Audio Engineering Society Preprint 4791 (September 26-29, 1998)
19. Rimell, A.N., Hollier, M.P.: The Significance of Cross-Modal Interaction in Audio-Visual Quality Perception. In: e-proceedings of Workshop on Multimedia Signal Processing, September 13-15, 1999, Copenhagen, Denmark. IEEE Signal Processing Society (1999)
20. Rimell, A., Owen, A.: The effect of focused attention on audio-visual quality perception with applications in multi-modal codec design. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP '00, 5-9 June 2000, vol. 6, pp. 2377–2380, vol.4. IEEE (2000)
21. Styles, E.A.: The psychology of attention. Psychology Press, Hove, England (1997)
22. 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. In: presented at the AES 24th International Conference on Multichannel Audio, Banff, Canada (June 26-28, 2003)