Latency Annoyance of Interactive Auditory Virtual Environments

Andreas Silzle, Christian Borß

Institut für Kommunikationsakustik, Ruhr-Universität Bochum, D-44780 Bochum, Email:andreas.silzle@web.de

The influence of the latency of an interactive Auditory Virtual Environment (AVE) on the localisation performance and the user annoyance is investigated. The result of this investigation indicates that an Internet based AVE generator for an localisation adjustment test can be implemented when the overall delay between user input and audio output is less than 600 ms.

Introduction

For the implementation of an interactive Auditory Virtual Environment (AVE), the responsiveness is an important quality feature. The responsiveness describes the perceived time delay between an action or an event and the corresponding change at the output, perceived by the listener (e.g., a change of the head orientation) and the auditory perceived response. The responsiveness is dependent on the corresponding input device (e.g., head tracker or graphical user interface (GUI)), the system latency, the type of output, the type of user action, and the scenario (e.g., chat-room or computer game). Figure 1 shows the set-up of the localisation test with the interactive AVE generator IKA-SIM, [1]. The system latency between the input of the listener on the client side and the audio playback, also on the client side, consists of:

1. Command transmission from the client to the server
2. Audio rendering on the server
3. Audio coding
4. Streaming of the audio data back to the client
5. Jitter buffer
6. Audio decoding
7. Audio playback buffer in the client

The user action could be starting or stopping a sound source, to move a sound source, to change the orientation of the head, etc. Especially the last one is interesting, because, when the listener can change his or her head orientation in the simulated room, the front-back confusion, which often occurs with non-individual HRTF reproduction, is solved (for an early implementation see [2] and later [3]). The threshold for the latency perception of a system with head rotation is about 90 ms, [4] and [5]. The latencies for the system implemented here will be clearly above this threshold. For the user action, which should be performed with the interactive AVE over Internet, the important questions are: How is the task performance influenced by the perceived latency? How is the listener annoyed by this latency? The following experiments deal with these two questions for the first time.

Set-up, Procedure and Subjects

Eleven normal hearing subjects participated in the experiment. The delay between the slider movement and the audio presentation was varied, in steps, between 0 and 1500 ms. The listener could change his or her head orientation in the AVE with the help of a horizontal slider in the GUI. No other visual feedback was provided (i.e. no visual room simulation). The task was to adjust the virtual head orientation in a way, that the listener perceives the presented telephone signal directly in front of him/her. The stimulus in the test was a recorded telephone ring (old telephone with mechanical bells) with a broad-band frequency spectrum. The length of the stimulus was 18 seconds with three rings. Only the direct sound was auralised and there was no further room simulation. Pre-experiments indicated that this is the most critical setting with regard to the annoyance of the delay. The selected non-individual HRTF-catalogue had a measured azimuth resolution of 15° and an interpolated resolution of 5°. There was no scale given on the slider and even the start position of the slider had a random off-set. The listeners were also encouraged to close their eyes during the localisation adjustment. Both steps avoid cross modality effects with proprioceptive or visual feedback and ensure a mono-modal experiment design. After every localisation adjustment task, the subjects were asked how annoying the introduced delay for the execution of the task was. Annoyance is meant here in the sense of the daily usage of the word. The scale is slightly adapted from [6]. The five grading points were expanded by the sixth point *just perceivable*, to expand the resolution a little, see the ordinate in Figure 2.

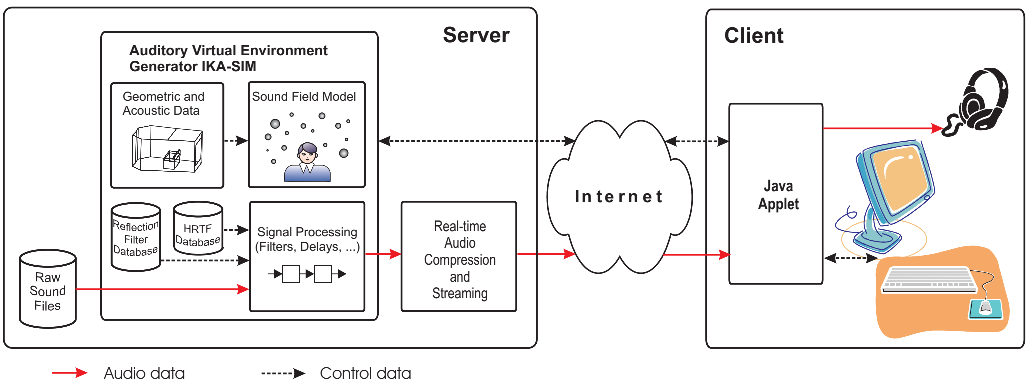

Figure 1: Set-up of the interactive Auditory Virtual Environment generator IKA-SIM over Internet

Result of Annoyance

The result is presented in Figure 2. The annoyance for this kind of task is monotonically increasing, starting from the threshold of perception. The end-results corroborate the decision from the pre-experiment, namely, that the division of the grading point *perceivable but not annoying* into two is useful to represent the fine differences in the answers of the subjects.

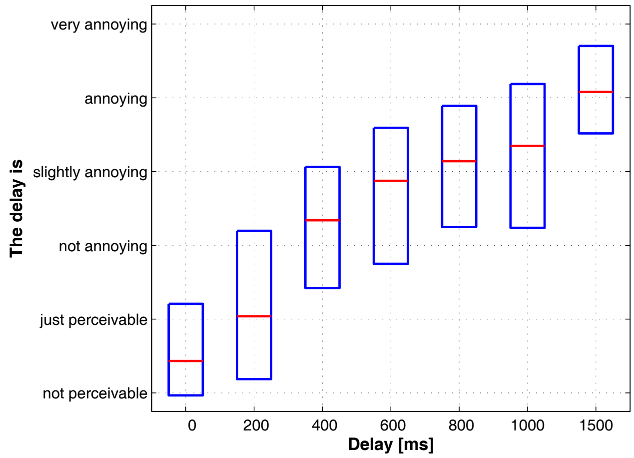

Figure 2: Annoyance of introduced delay between slider input and sound output for the localisation adjustment task, median and quartile values

To avoid for at least 75% of the subjects any annoyance due to the delay, for this kind of task, the delay should be less or equal than 200 ms. In the case that a slight annoyance of the end-user is acceptable, the delay can be increased to around 600 ms.

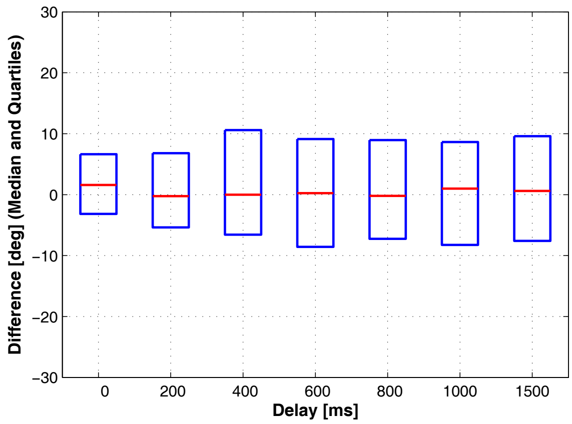

Figure 3: Localisation accuracy (median) and localisation blur (quartiles, 50% of the answers) dependent on the delay between slider input and sound output

Result of Localisation Accuracy

The task performance is indicated by the localisation accuracy and the localisation blur. The localisation accuracy is given here by the azimuthal difference between auditory event and sound event, obtained from the position of the sound source in the simulation. Positive differences in the plot denote a shift to the left, as seen by the listener. The influence of the delay on the localisation accuracy and the blur is presented in Figure 3.

The data are shown with their median and quartile values, because not all of them are normally distributed (Lilliefors test, $\alpha = 0.05$) and the localisation blur is normally defined by the variance of 50% of the answers, given by the quartile values. The 95% confidence intervals would be significantly smaller here, but misleading for comparison with other data. The median values are independent of the delay (Kruskal-Wallis test, $\alpha = 0.05$). The median values for all delays are at a maximum 2° apart from the intended direction. The quartile range for zero delay is ±5°, and increases to about ±8.5°, when the delay is 400 ms and more. The dispersions between the first sample with delay = 0 ms and all other samples are significantly different (Ansari-Bradley test, $\alpha = 0.05$). The dispersion of a delay of 200 ms is different from some of the further distributions and for the higher delays there is no further difference between the dispersions.

Discussion

Although the people are annoyed by the delay, they are able to find strategies to localise and adjust the slider in the limited time of 18 s, and fulfil the task successfully, independent on the delay. This is an experimental result using subjects between the age of 25 and 45, who are trained to concentrate themselves. For another age-group this experiment should be repeated.

These results indicate that an interactive AVE, offering a localisation adjustment test, can be implemented over the Internet when the overall delay between user orientation input and audio output is less than 600 ms. This requirement is below the delay of standard audio streaming solutions, which range from 1 second to over 10 seconds. However, the new audio streaming implementation by Borß [7] has a measured latency in the LAN of 350 ms, which results in a latency of 400 ms including the delay over the Internet within Europe and therefore fulfils the given requirements.

Literatur

[1] Silzle, A., P. Novo, and H. Strauss (2004). IKA-SIM: A System to Generate Auditory Virtual Environments. 116th AES Convention. Berlin, Germany, #6016.

[2] Koenig, W. (1950). Subjective Effects in Binaural Hearing (Letter to the Editor). J. Acoust. Soc. Am. 22(1), p. 61-62;

[3] Boerger, G., P. Laws, and J. Blauert (1977). Stereophone Kopfhörerwiedergabe mit Steuerung bestimmter Übertragungsfaktoren durch Kopfdrehbewegungen. Acustica. 39(21-26);

[4] Felderhoff, U., K. Reichenauer, and G. Theile (1998). Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale (Stability of Localization vs Distorted Reproduction of Binaural Cues). 20th Tonmeistertagung. Karlsruhe, Germany

[5] Mackensen, P. (2004). Auditive Localization. Head movements, an additional cue in Localization. Fakultät I - Geisteswissenschaften. Technische Unversität, Berlin.

[6] ITU-R Rec. BS.562-3 (1990). Subjective Assessment of Sound Quality, Intern. Telecom Union, Geneva, Switzerland

[7] Borß, C. (2006). Audioübertragung mit geringer Latenz für interaktve Web-Services (Audio Transmission with Low Latency for Interactie Web-Services). DAGA. Braunschweig

This work was partly supported by the EU project HEARCOM (IST-2004-004171).