Dynamic virtual acoustical reproduction system for hearing prosthesis evaluation

Grämer, T ; Müller, M F ; Schimmel, S M ; Kegel, A ; Dillier, N

Abstract: Algorithms for the improvement of speech intelligibility in hearing prostheses can degrade the spatial quality of the audio signal. To investigate the influence on distance perception and localization of such algorithms, a system to virtually render arbitrary static acoustical scenes has been developed (Müller et al., 2010). With this system, the localization of sounds processed by a hearing aid algorithm can be compared to unprocessed sound sources. The existing virtual acoustics system has been extended to present more realistic dynamic scenes, and it can also compensate for head movements of test subjects.

Dynamic virtual acoustical reproduction system for hearing prosthesis evaluation

Tobias Grämer, Martin F. Müller, Steven Schimmel, Andrea Kegel, Norbert Dillier

Laboratory for Experimental Audiology, University Hospital Zürich

Keywords: Virtual Acoustics, Head Movements, Sound Localization

Introduction

Algorithms for the improvement of speech intelligibility in hearing prostheses can degrade the spatial quality of the audio signal. To investigate the influence on distance perception and localization of such algorithms, a system to virtually render arbitrary static acoustical scenes has been developed (Müller et al., 2010). With this system, the localization of sounds processed by a hearing aid algorithm can be compared to unprocessed sound sources. The existing virtual acoustics system has been extended to present more realistic dynamic scenes, and it can also compensate for head movements of test subjects.

Dynamic virtual acoustics system

To render a virtual sound source located in a room, the sound is filtered with an impulse response that contains the properties of the source, the room and the receiver. The sound source is modelled as a point source with subcardioid radiation characteristics. The room impulse response is created using a "shoebox" room acoustics simulator which generates a binaural room impulse response (BRIR) (Schimmel et al., 2009). Finally, the receiver characteristics correspond to the subject's individual head-related transfer function (HRTF). The individual HRTFs are measured at 12 positions in the horizontal plane. Between the 12 measured positions, the HRTFs are interpolated by means of time-aligned piecewise cubic interpolation (Park, 2007). Incident room reflections outside the horizontal plane are filtered with generic MIT KEMAR HRTFs (Gardner, 1994).

The real-time generation of the full BRIR is a computationally intensive task; therefore it is computed offline with a spatial resolution of 1°. The head movements of the listener are measured by an acceleration sensor (Xsens MTx) at a rate of 86 Hz. With the knowledge of the head orientation, these movements can be compensated for. The audio signal filtering itself is executed on a standard PC with a worst-case overall processing delay of 23.2 ms. The filtered sound is played over an open micro-speaker prototype (open CIC), which ensures a very natural sound reproduction.
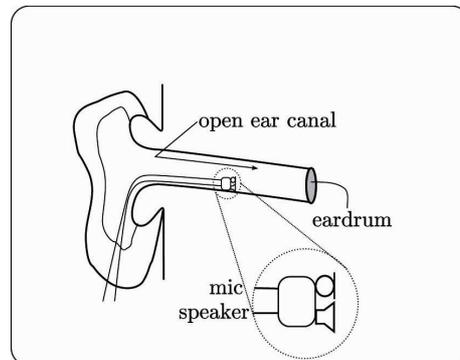

Figure 1: Head-Tracker

Figure 2: open CIC prototype [1]

Evaluation of the system

Experiment Design

To evaluate the accuracy and the limitations of the system, we compared the dynamic virtual acoustics system with real scenes where the sound is played over loudspeakers. The test subject was located in the centre of a circle of 12 loudspeakers. Sound was randomly presented over the loudspeakers or over the simulation, and head movements were encouraged. Seven normal hearing and two hearing impaired individuals served as volunteers for this listening test. A speech signal and a noise signal presented at 60 dB SPL were used as stimuli.

Previous tests with the virtual acoustics system showed that the externalization of sound sources is often not fully convincing, and sound sources are perceived in the head. To test if the motion-tracking sensor and the signal processing is fast enough to render a convincing scene where listeners cannot hear any sluggishness when they move their head, the test subject had to assess the stability of the sound source.

Finally, they had to decide whether the source was simulated or played over the loudspeaker. The more confusions they make, the better the virtual acoustics system is.

After each sound presentation, the test subjects had to answer the following questions:

1. On a scale from 1-5, do you hear the sound source in your head or from the loudspeaker?
2. On a scale from 1-5, does the sound source remain stable if you turn your head?
3. Where does the sound come from: Loudspeaker or simulation?
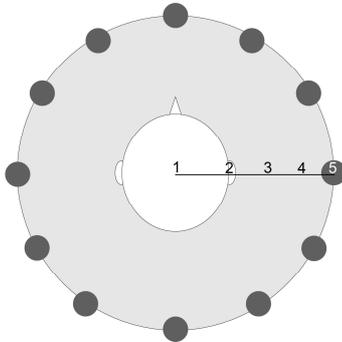4. (If the answer to question 3 was "simulation"): Why? Any further remarks?

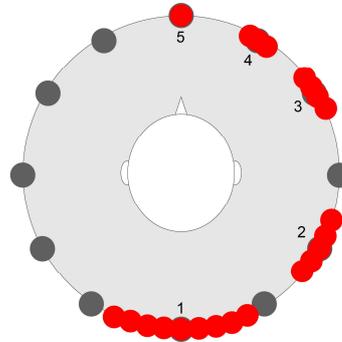*Figure 3*: Answer map for question about externalization

Figure 4: Answer map for question about stability

Results

The answers to the questions are depicted in Figures 5-7 and Table 1, respectively. Statistical significance was tested by means of a Wilcoxon-Test. The reasons for the classification as simulation are grouped: Sensor imperfectness means either a dynamic sensor error, resulting in an unstable source, or a static error, resulting in an off-position sound source. The "other" category includes unnatural frequency response, artefacts, front-back uncertainties and guessing.
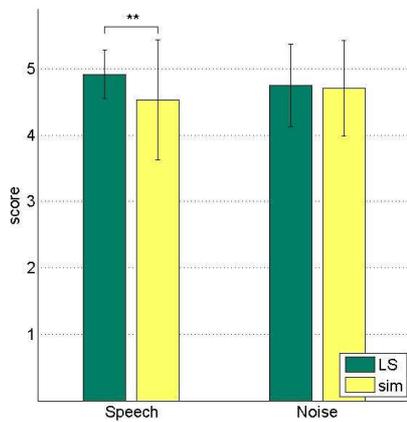

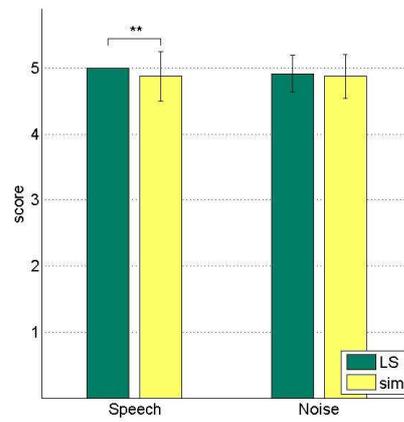*Figure 5*: rating of externalization

Figure 6: rating of stability

Speech		playback over	
		LS	sim
classified as	LS	63	22
	sim	9	50

Noise		playback over	
		LS	sim
classified as	LS	57	41
	sim	15	31

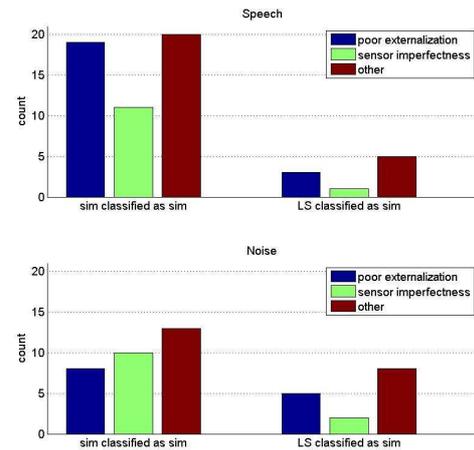***Table 1**: confusion matrix*

***Figure 7**: reasons for classification as simulation*

Localization of moving sounds

Experiment Design

In a second experiment, we compared the virtual acoustics system with real scenes in the case of a moving sound source. The subject was located in the centre of a circle of 12 loudspeakers. Sound was randomly presented over the loudspeakers or over the simulation. Five normal hearing and one hearing impaired individuals served as volunteers for this listening test. A noise signal presented at 60 dB SPL was used as stimulus. The task of the test subjects was to always face the source. The sound source was located randomly somewhere in the frontal hemisphere for 4-6 seconds so that the subject had enough time to face the source. Then, the source began to move with a constant speed of 20 °/s for a certain time. In the end, the source stood still again for 4 seconds. We used amplitude panning (Pulkki, 1997) to simulate moving sources when sound was played over loudspeakers. The head-tracker was used for the compensation of head movements and as an answer device.

Two performance measures were used: The time until the direction of movement is correctly detected, and the difference between the source and the head position during the moving phase (angular rms error). The reaction time corresponds to the minimum audible movement angle, whereas the localization error was – to the best of our knowledge – never investigated in previous studies.

Results and Discussion

The results are depicted in Figure 8. Statistical significance was tested by means of an analysis of variance. The limited dynamic accuracy of the sensor (2° rms error) leads to a higher error in the loudspeaker condition, since the simulated sources are always positioned relative to the measured ead position and therefore not affected by the sensor error.
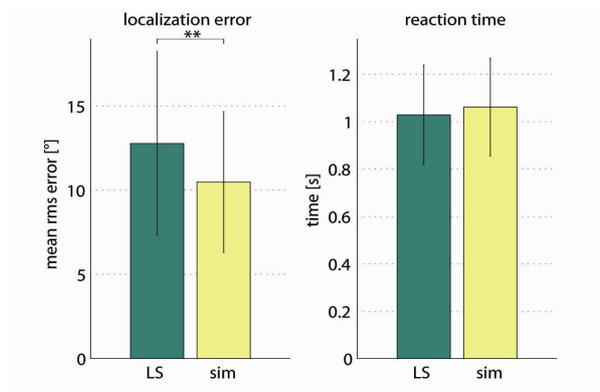

***Figure 8**: localization of moving sources*

Conclusions and future work

The results from the first listening test with static scenes indicate significant differences between loudspeaker and simulation with a speech signal, but the simulation shows a good absolute performance, both in terms of externalization as well as for stability of the spatial impression. The subjective impression of most subjects was that the head movements resolve front-back confusions, improve the externalization compared to the old system without support for head movements, and that they sound very natural. For dynamic scenes, the localization error of simulated sources is comparable to real loudspeaker sources.

The system currently supports only one sound source, since the computational complexity increases linearly with the number of sources. If the system is extended to render multiple sound sources, it could be used to evaluate different, adaptive hearing-aid algorithms also in dynamic scenes. Furthermore, a more accurate motion tracking sensor would further improve the reliability and performance of the system.

References

Müller M F, Kegel A, Schimmel S, Dillier N (2010) Localization of virtual sound sources in realistic and complex scenes: how much do hearing aids alter localization? DGA March 2010

Schimmel S, Müller M, Dillier N (2009) A fast and accurate "shoebox" room acoustics simulator. International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

Park M (2007) Models of binaural hearing for sound localization. Ph.D. thesis, University of Southampton.

Gardner B, Martin K (1994) HRTF Measurements of a KEMAR Dummy-Head Microphone. MIT MediaLab.

Pulkki V (1997) Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society 45(6), 456-466