Impact of Doppler Effect, Echo and Reverberation on the Externalization and Plausibility of Binaural Rendered Moving Sound Sources Presented via Headphones

Song Li, Sanam Moghaddamnia, Jürgen Peissig

Institut für Kommunikationstechnik, Leibniz Universität Hannover, 30167 Hannover

Email: {song.li, sanam.moghaddamnia, peissig}@ikt.uni-hannover.de

Abstract

In this work, different virtual 3D audio scenarios with various moving sound sources and parameter setups (combinations of Doppler Effect, echo and reverberation) were simulated. The basic principle to create these scenarios presented by headphones is to filter a monophonic sound source with a Head-Related Transfer Function (HRTF). The HRTF, measured in free field, contains acoustic cues (ITD, ILD) for sound localization in horizontal plane. Air- and wall absorption, echoes (1^{st} reflection with two walls) and late reverberation were implemented to simulate the acoustical environment. The inverse-square law was implemented as a cue for the auditory distance perception. The HRTFs were interpolated according to the time-varying incidence angle between the sound source and the receiver in order to render a virtual moving source. The resulting Doppler Effect was realized using a "time-varying delay line". The air absorption and echoes were realized by a distance-dependent low-pass filter and the image-source method, respectively. Exponentially decaying white noise was considered to simulate the diffuse reverberation. A subjective experiment was implemented to investigate the plausibility and the externalization of the simulated virtual 3D audio moving sources for the considered scenarios.

Introduction

The basic principle to create a virtual 3D audio signal via headphones is to filter a monophonic sound source with a Head-Related Transfer Function (HRTF) [1]. However, when audio signals are presented via headphones, the sound image might be localized within the head. This phenomenon is known as "in-head localization" [2]. Therefore, it is reasonable to examine the quality of the rendered sound image in terms of externalisation. Another quality feature for perceptual evaluation of virtual acoustic environments is "plausibility" [3]. To assess the influence of Doppler Effect, echoes and reverberation on the externalisation and plausibility, different virtual 3D audio scenarios with various moving sound sources and parameter setups are simulated. Subsequently, correspondent subjective experiments are conducted to investigate the plausibility and the externalization of the simulated virtual 3D audio moving sources for the considered scenarios.

This paper is organized as follows. First, simulation and modeling of a moving virtual sound source in different scenarios were described. The following sections present subjective experiments and the results. Finally, a conclusion is drawn.

Modeling of Sound Source Trajectories

In this work, two linear trajectories were modeled. A listener stands in the middle, in front of the first trajectory, which spans from the front left to front right. The second trajectory is from the front right to rear right and the listener stands in the middle on the left side of the trajectory. The length of both trajectories is 100 m, the virtual sound source moves with 20 m/s and each trajectory is 10 m away from the listener.

HRTF Interpolation

In order to synthesize a moving virtual sound source HRTFs are interpolated in frequency domain. Each HRTF is represented by its magnitude |HRTF(θ,f)| and unwrapped phase Φ(θ,f) as follows:

$$HRTF\,(\theta, f) = |HRTF\,(\theta, f)\,| e^{j\Phi(\theta,f)}\ . \qquad (1)$$

The magnitude and the phase are linear interpolated separately [4].

Simulation of the Doppler Effect

Doppler Effect is caused by the relative movement between a sound source and a listener. We can implement Doppler Effect using a "time-varying delay line" [5]. The propagation paths between the sound source and the listener corresponds to a delay line. The moving sound source is simulated by controlling the delay line via the read-pointer increment from 1 to 1$\pm v_s$/c. v_s is the relative speed between the sound source and the listener and v_s/c denotes the delay growth rate. Subsequently, the new updated read pointer must be interpolated. If there is no relative movement between the sound source and the listener, the delay growth rate equals zero. Figure 1 illustrates the simulation procedure of the Doppler Effect.

Air- and Wall Absorption

In real acoustic environments, the propagation medium of sound waves is usually air. The sound energy is partially absorbed or attenuated. To reproduce air absorption a distance-dependent low-pass filter is designed [6]. The relative humidity and the temperature amount to 50 % and 20 °C, respectively.

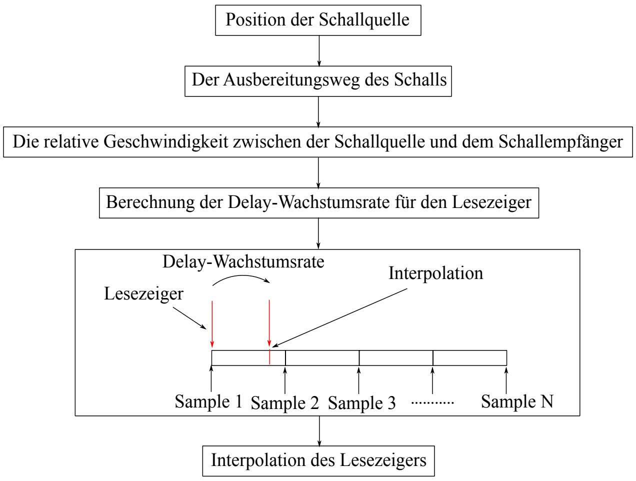

Figure 1: Simulation of Doppler Effect using "time-varying delay line"

If the acoustic environment is not a free-field, sound is not only absorbed by air, but also by other obstacles, such as walls, windows, floors, etc. In this study, we consider only the wall absorption. The degree of this absorption depends on the wall material.

Echoes and Diffuse Reverberation

Sound reaches the listener not only via the direct path, but also via detours, as the sound source also radiates in the direction of walls, floors, windows, etc. The echoes, which reach the ear shortly after the direct sound, amplify the direct sound and provide some room information. In this work, the first echo from each wall was simulated by means of the "image-source method" [7].

The diffuse reverberation is an important parameter to perceive the room size and consists of a large number of reflections. The diffuse reverberation can be generated with an exponentially decaying white noise. This is carried out for the left and the right ear separately, so that the reverberation on both ears is uncorrelated.

Synthesis of a 3D Moving Sound Source

Regarding the length of the trajectory and the velocity of the virtual sound source, the duration of a static audio signal is 5 s. In the first step, the audio signal is segmented by using a hanning-window with 50 % overlap. According to the time-varying incidence angle between the sound source and the receiver, the required HRTF can be interpolated. The attenuation of the audio signal can be calculated by the inverse-square law. Each audio frame is transformed using a FFT (Fast Fourier Transform) in the frequency domain and is then multiplied with the corresponding HRTF. Afterwards, each calculated frame is transformed back into the time domain [8]. The synthesis of a virtual moving sound source is completed by considering the inverse-square law, the air and wall absorption and different combinations of Doppler Effect, echoes and diffuse reverberation. The described implementation concept is shown in Figure 2.

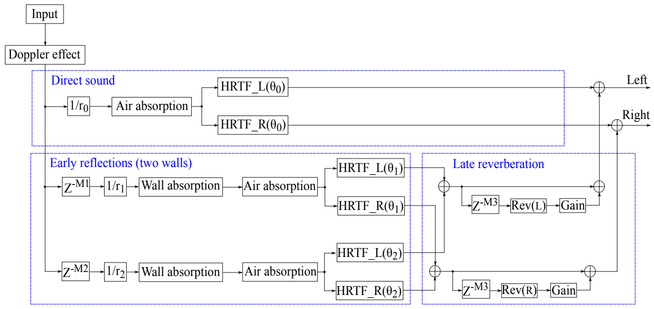

Figure 2: Block diagram of synthesising of a moving virtual sound source

The aforementioned parameter setups were combined to the following modes:

M1: Direct sound without Doppler Effect

M2: Direct sound with echoes and without Doppler Effect

M3: Direct sound with echoes, diffuse reverberation, and without Doppler Effect

M4: Direct sound with Doppler Effect

M5: Direct sound with echoes and Doppler Effect

M6: Direct sound with echoes, diffuse reverberation and Doppler Effect

Experiment Setup and Measurement

The subjective experiment took place at the Institute of Communications Technology (IKT) of the Leibniz Universität Hannover. For the experiment the RME Fireface 400 sound interface, Sennheiser HD 800 headphones and two K&H O100 loudspeakers were used. Three sound-absorbing baffles around the listener were used to prevent strong reflections and to ensure a silent environment.

Engine noise, police siren, fire engine sirens and a train horn are used as audio signals in the experiments. Two kinds of GUIs (Graphical User Interface) were designed to carry out the experiments. The first GUI is used for A-B Test and the second one is for the modified MUSHRA Test based on MUSHRA (Multi-Stimulus Test with Hidden Reference and Anchor) [9].

Seventeen subjects of age 18-35 were involved in the experiment with the first and second audio signals and another seventeen subjects of the same age range were involved in the experiment with the third and fourth audio signals.

For each experiment, there are two linear trajectories of sound sources; one from front left to front right (T1) and the other one from front right to rear right (T2). Before performing the experiment the subjects were explained about the "in-head localization" and "externalization". As an example, "externalization" was demonstrated with two loudspeakers, where the audio signal is generated by the "Vector Base Amplitude Panning" method [10]. The effect of the "in-head localization" is rendered by the headphones via simulated signal moving through the head from the left to the right ear. In A-B Test, the subject hears the audio signals A and B. After the comparison of the two audio signals an answer must be selected on the right side of the GUI as seen in Figure 3. The only difference between the audio signal A and B is the presence or absence of the Doppler Effect.

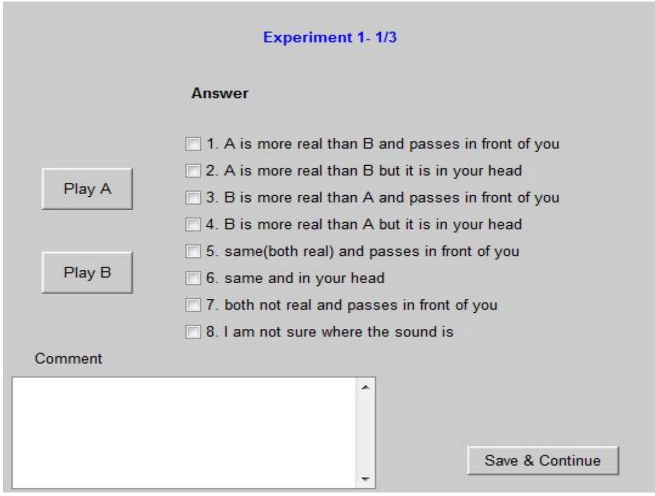

Figure 3: GUI for the A-B Test

The modified MUSHRA Test is illustrated in Figure 4. There are six audio sequences in this test. The first audio signal serves as a reference (rating 0). Each subject listens to other five audio sequences, compares them to the reference and chooses a score between -2 (Bad) to +2 (Excellent). "Excellent" means the signal is plausible and far away from the listener. "Bad" means that the signal is not plausible and inside the head.

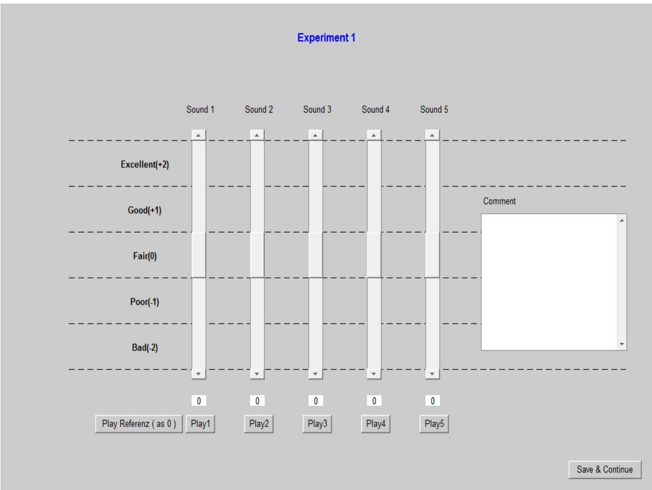

Figure 4: GUI for modified MUSHRA Test

Results

The results of the experiments with all four audio signals put together are shown in Figures 5 - 8 for the A-B Tests as column bar graphs and in Figure 9, 10 for modified MUSHRA Test as box plots.

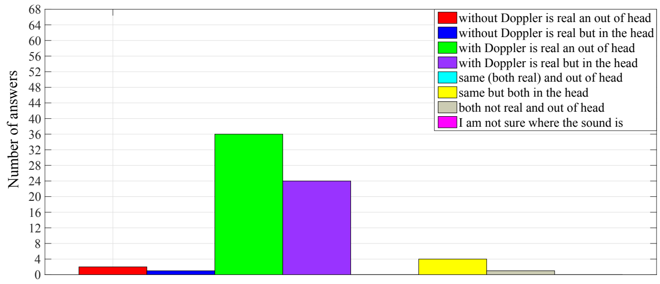

Figure 5: A-B Test (T1): Direct sound compared to the direct sound with Doppler Effect

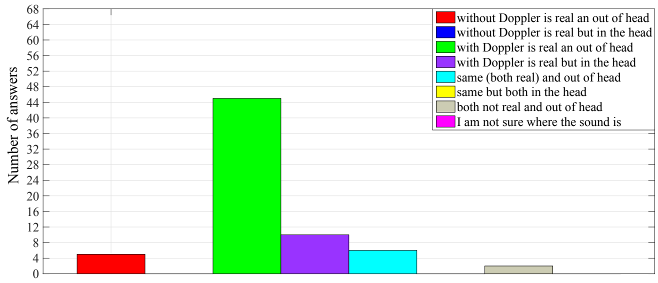

Figure 6: A-B Test (T1): Direct sound with echo compared to the direct sound with echo and Doppler Effect

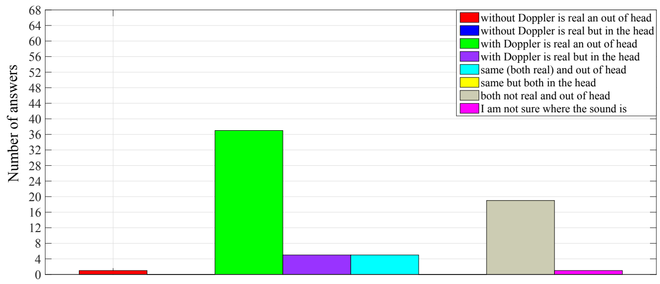

Figure 7: A-B Test (T1): Direct sound with echo and diffuse reverberation compared to the direct sound with echo, diffuse reverberation and Doppler Effect

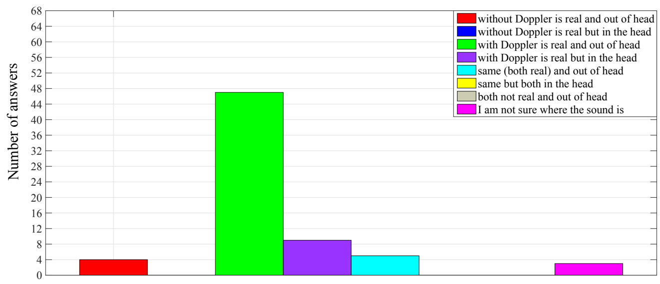

Figure 8: A-B Test (T2): Direct sound compared to the direct sound with Doppler Effect

Figures 5, 6 and 7 show the results for the first trajectory (T1) and Figure 8 shows the result for the second trajectory (T2). The results in Figure 5 show that the simulated Doppler Effect enhances the plausibility of the moving virtual sound source. Figures 6 and 7 demonstrate that the echoes improve the externalization of a moving virtual sound source but the diffuse reverberation reduces the plausibility. Figure 8 illustrates that the T2 improves the externalization in comparison to the T1.

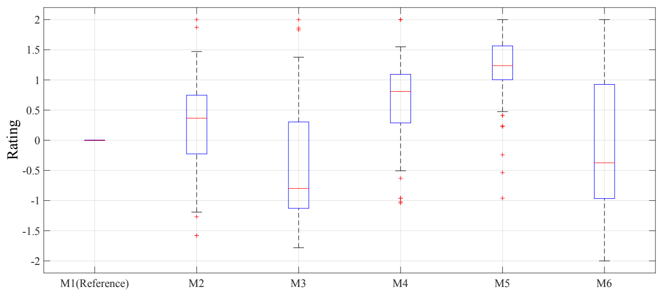

Figure 9: Modified MUSHRA Test (T1)

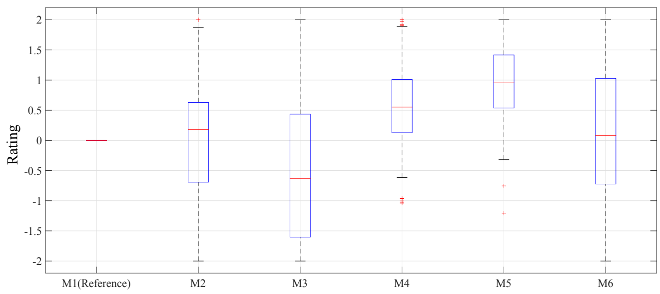

Figure 10: Modified MUSHRA Test (T2)

Figures 9 and 10 show that M5 is the best choice for simulating a moving virtual sound source. This means the interaction of Doppler Effect and echoes improves the plausibility and the externalization of a moving virtual sound source in our considered scenario. However, diffuse reverberation leads to plausibility reduction represented by M6.

Conclusion

Different virtual 3D audio scenarios with various moving sound sources and parameter setups (combinations of Doppler Effect, echo and reverberation) were simulated. To examine the plausibility and the externalization of the simulated virtual 3D audio moving sources for the considered scenarios, different subjective experiments were conducted. According to the experiment results, the Doppler Effect enhances the plausibility of the simulated moving virtual sound source and the echoes improve the externalization. Moreover, the interaction of Doppler Effect and echoes results in further improvement of the plausibility and the externalization. It should be noted that the diffuse reverberation decrease the plausibility of the moving virtual sound source in our considered scenarios. The second trajectory (T2) improves the externalization compared to the first trajectory (T1).

References

[1] Wightman, Frederic L. and Kistler, Doris J.: Headphone simulation of free-field listening. I: Stimulus synthesis, Journal of the Acoustical Society of America: February 1989.

[2] Georg Plenge: On the Problem of In Head Localization, Acta Acustica united with Acustica 26, 241-252, 1972.

[3] Alexander Lindau, Stefan Weinzierl: Assessing the plausibility of virtual acoustic environments, Forum Acusticum, Aalborg Denmark, 2011.

[4] Johannes Zaar: Phase unwrapping for sphericalinterpolation of headrelated transfer functions, Master's thesis, University of Music and Performing Arts, Graz, 2011.

[5] Smith J, Serafin S, Abel J, Berners D: Doppler simulation and the leslie, Proceeding of the 5th International Conference on Digital Audio Effects (DAFx '02), Hamburg 2002

[6] Peter Lessing: A Binaural 3D Sound System Applied to Moving Sources, Master's thesis, University of Music and Dramatic Arts of Graz, 2004.

[7] Jont B. Allen and David A, Berkley: Image method for efficiently simulating small room acoustics, Journal of the Acoustical Society of America, vol. 65, pp. 943-950, April 1979.

[8] J. Proakis and D. Manolakis: Digital Signal Processing: Principals, Algorithms and Applications. Macmillan Publishing Company, 1992.

[9] A.J. Mason: The mushra audio subjective test method. BBC Research & Development, White Paper WHP 038, 2002.

[10] Ville Pulkki: Virtual sound source positioning using vector base amplitude panning. Audio Engineering Society, 1997.