Auditory Locomotion Guidance System For Spatial Localization

Felix Dollack
University of Tsukuba
Tsukuba, Japan
dollack@ai.iit.tsukuba.ac.jp

Monica Perusquía-Hernández
NTT Communication Science Laboratories
Atsugi, Japan
perusquia@ieee.org

Hideki Kadone
University of Tsukuba
Tsukuba, Japan
kadone@ai.iit.tsukuba.ac.jp

Kenji Suzuki
University of Tsukuba
Tsukuba, Japan
kenji@ieee.org

***Abstract*—Head direction has been identified to anticipate trajectory direction during human locomotion, independently of the visual condition. However, experiments so far have explored this phenomenon with visual instructions. To explore head and gaze anticipation without visual influence, we describe a system for instruction and measurement of auditory instructed locomotion. The objective of this work is to describe the system setup and validate its accuracy for scientific investigations in locomotion research. The system is comprised of an auralization server that plays virtual sound sources via wireless headphones. The auralization is fed with head position and orientation angle measured by a Vicon motion tracking system. Communication between the auralization server and the motion tracking host was facilitating the OSC protocol. Auditory instructed locomotion was performed in two visual condition: eyes open and eyes closed for comparison. First, ten sighted participants localized static virtual sound sources at eight positions to confirm the accuracy of the provided virtual sound cues. Afterwards, they listened and actively followed virtual sound sources that were moving along three trajectories: an eight shape, a circle in clockwise direction and a circle in anticlockwise direction. The virtual sound sources could be localized with an accuracy of 0.16 meters in the eyes open condition and 0.12 meters in the eyes closed condition. Participants also were able to follow the moving sound sources.**

***Index Terms*—auralization, virtual sound, sound localization, guided navigation, eyes closed, auditory perception**

I. Introduction

Numerous neurobiological and behavioral studies have demonstrated the close interaction between auditory and motor areas of the brain, and the importance of auditory information for movement execution, control, and learning. Therefore, sonification has gained increasing interest in recent years in research across disciplines. Sonification is the mapping of data onto psychoacoustic parameters like pitch, loudness or rhythm. It has been used for motion sonification in training of posture and gait, to provide access to biomechanical information otherwise not available [5]. It also has been used in the form of rhythmic auditory stimulation for rehabilitation to improve motor performance in patients affected by neurological or movement disorders such as Parkinson's disease [7]. These sonification approaches often utilize wearables to either map sensor information and derivatives (e.g. cadence gait velocity and stride length) to psychoacoustic parameters; or to trigger predefined artificial or recorded natural sounds. Researchers in [8] carried a loudspeaker to instruct participants in walking a trajectory. This approach might be enough for quick and simple pilot studies, but can be significantly improved in terms of movement continuity, position accuracy as well as reproducibility. Hence, we propose an auditory guidance system comprised of motion tracking and a sound synthesis software for cases that require high position accuracy or precise movement and other tasks that need spatial instruction or guidance. We describe the system and evaluate it's accuracy with a localization task. Furthermore, we show results from participants following a moving sound source on different trajectories as used in the investigation of head anticipation [9].

II. Materials and methods

A. System overview

The proposed auditory guidance system (Figure 1) consists of five subsystems, a motion tracking system, a custom commandline[1] tool utilizing the Vicon streaming software development kit[2] (SDK), a sound synthesis server, an open sound control (OSC) server[3] and a control application[4,5]. The motion tracking system (VICON MX, Vicon, UK) captures the position and orientation of tracked objects on an area of 12 m by 4 m. The Vicon software then forwards the captured data to the custom commandline tool to convert the Vicon data stream to UDP messages for the local network. The control application receives the UDP messages and forwards the tracking data together with experimental control commands as OSC messages to the local network as well. An open sound control (OSC) server written in Python receives the tracking and control commands for updating the sound synthesis server. The synthesized sound is presented by a pair of wireless headphones.

B. Apparatus

Vicon is used to record head and torso position and orientation. A total of seven reflective motion tracking markers is placed on the participants. The marker position on the participants' shoulders are right and left acromion and cervical vertebrae 7. Four markers for the head of the participants are

[1] https://github.com/felixdollack/ViconStreamToUdp
[2] https://www.vicon.com/products/software/datastream-sdk
[3] https://github.com/felixdollack/osc_sound_synthesis_control
[4] localization task: https://github.com/felixdollack/phd_experiment_find
[5] guidance task: https://github.com/felixdollack/phd_experiment_follow

978-1-7281-2674-6/19/$31.00 ©2019 IEEE

fixed to a headphone. A MacBook Pro (2,7 GHz Intel Core i7, 16 GB RAM, Mid 2012, OS X El Capitan, Apple, US) is used to run the custom control application, the OSC server and the sound synthesis server. The custom control application is written with the C++ framework openframeworks, the OSC server is written in Python and soundscape renderer [2] is used as sound synthesis server. Head-related transfer functions (HRTFs) are used during the sound synthesis. The HRTFs were recorded in an anechoic chamber from the torso and head simulator FABIAN [1] and equalized for a typical pair of headphones. The synthesized sound stimulus is presented to the participants with a radio frequency (RF) wireless headphone (UHF wireless headphone, Ansee, China). The base station of the RF headphones is connected to the headphone jack of the MacBook. The position of the stimulus and the recording of the participants' responses were recorded.

C. Participants

Ten volunteers (mean age 27.1 years old, $SD = 3.5$), five of which were female, participated in the experiment. The study was carried out at the Tsukuba University Hospital. All participants gave written informed consent in accordance with the Declaration of Helsinki. The experimental protocol was approved by the ethical committee of the University of Tsukuba (2018R259).

D. Stimuli

For a sound to be well localized, it must have a broad bandwidth and a relatively flat spectrum that does not mask monaural spectral cues to location [6]. The basis for the virtual sound source was white noise created with Matlab v2017b (Mathworks, US). The signal had a sampling rate of 44.1 kHz and a duration of five minutes. The duration of five minutes was chosen because each task, the localization as well as the guidance task, could be finished within this time. The stimulus level was set to a comfortable sound level close to normal speech for a distance of 0 meters. The sound is synthesized with respect to the relative position of the virtual sound source to the participants' head. The stimulus level was adapted separately for left and right ear depending on the head orientation of the listener and the distance to the sound source.

E. Task

There were two tasks. A "localization-task", where participants were asked to localize and walk to the origin of a stationary virtual sound source and a "guidance-task", where participants were asked to follow a moving sound source. The localization-task was always done before the guidance-task. Participants performed both tasks with eyes open and eyes closed. During eyes closed conditions participants were asked to close their eyes and wear a pair of swim goggles covered with multiple layers of black electrical insulation tape to avoid any kind of visual stimulation and thus remove all external light. Participants were split into two groups for the experiment. One group started with eyes closed, while the other group started the experiment with eyes open. The

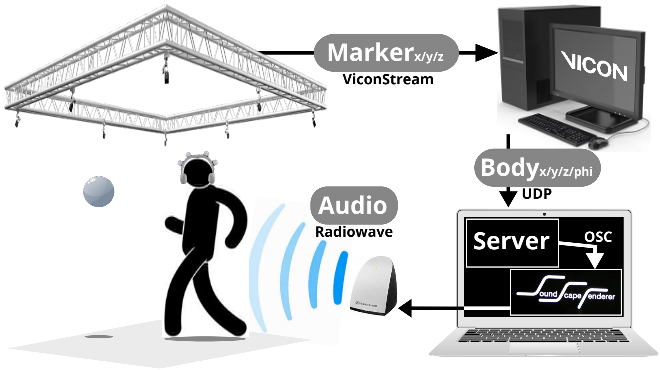

Fig. 1. Schematic overview of the auditory guidance system. Position and orientation of the participant are captured with motion tracking cameras and software. The motion tracking data is transformed to an UDP message and forwarded to the local network. A custom control program receives the UDP message and sends them together with commands to position the sound source as OSC message to an OSC server. The OSC server controls the sound synthesis software. Synthesized sound is played back via a wireless headphone.

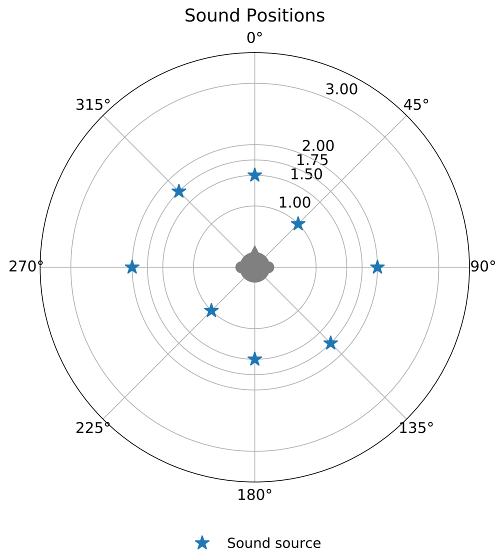

Fig. 2. Target stimuli (black asterisks) and start position (gray head in the center) of the participant as used in the static sound task. Target stimuli are given with distance r and angle ϕ to the start position.

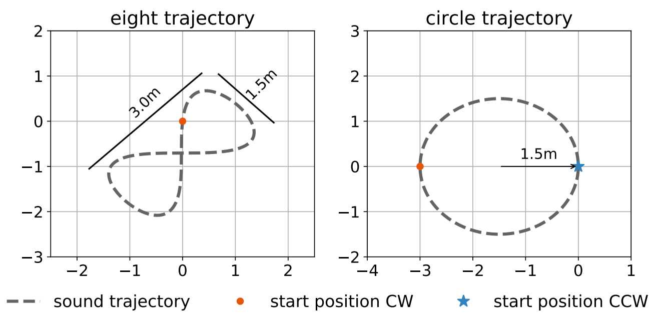

Fig. 3. Target trajectories are drawn with a dashed line for the eight trajectory (Left) and a circle (Right), start positions for clockwise walking direction are marked with an orange dot and start positions for a counter clockwise walking direction are shown as blue star.

experiment was conducted in a therapy room used for gait rehabilitation. During the eyes open condition the room was lightened and therapy tools were visible. The order of the sound targets and trajectories were randomized within two blocks of eyes open and closed trials. These blocks were counterbalanced among participants. The target stimuli in the loclization-task were presented at eight positions (distance in meter r, angle in degrees ϕ) starting at $0°$ with clockwise increments of $45°$ (see Figure 2.During the guidance-task, participants had to follow the sound source on all target trajectories twice. The trajectories were (1) an eight, (2) a circle in clockwise direction, and (3) a circle in counter clockwise direction (see Figure 3).

F. Procedure

After participants were guided to the start position, trials of the localization-task started with the playback of the sound. The trials ended with the participants signaling that they found the target position by saying “stop”. During the guidance-task a virtual sound source started playing in front of the participant and moved along a trajectory. The sound source stopped after 60 seconds regardless of the participants' position. The participants repeated each following-task trial two times.

III. Results

Figure 4 shows the results of all participants during the localization task. Each subplot shows the individual and average locomotion trajectories for both visual conditions, the sound target and a head in the center of the axis, representing the participant. The sound positions are marked by black asterisks. The average trajectory from the start position to the sound target is drawn as solid line. Individual trajectories are drawn shaded. Conditions with eyes open are shown in blue. Eyes closed conditions are shown in red. A general mixed linear model was fitted to the localization error as dependent variable to asses differences between visual conditions, target positions and the interaction between them. The effect of eyes closed is small ($\beta = -0.03$, $SE = 0.04$, 95% CI [-0.1, 0.05], $\eta^2 = 0.12$) and can be considered as significant ($F_{1,72} = 9.73$, $p < 0.01$). The grand mean for accurate determination of the sound source position was 0.14 meters. By visual condition, the means were 0.16 meters (range 0.02 meters to 0.48 meters) with eyes open and 0.12 meters (range 0.01 meters and 0.46 meters) with eyes closed. Performing the task with eyes closed is more accurate by one-third (32%) compared to eyes open. There were no significant effects of target position on localization error ($F_{7,63} = 0.37$, $p = 0.92$), as well as no interactions between visual conditions, target position and localization error ($F_{7,72} = 0.83$, $p = 0.57$). To find the static sound sources, participants needed between 11.56 s and 62.85 s with eyes open and 11.32 s and 79.91 s with eyes closed. There were no significant effects between visual condition and task duration ($F_{1,135} = 1.43$, $p = 0.23$, $\eta^2 = 0.001$) and, between target position and task duration ($F_{7,135} = 0.79$, $p = 0.6$). There was a significant interaction between visual condition, target position and task duration ($F_{7,135} = 2.33$, $p < 0.05$). However, this interaction was only found in the $225°$ direction.

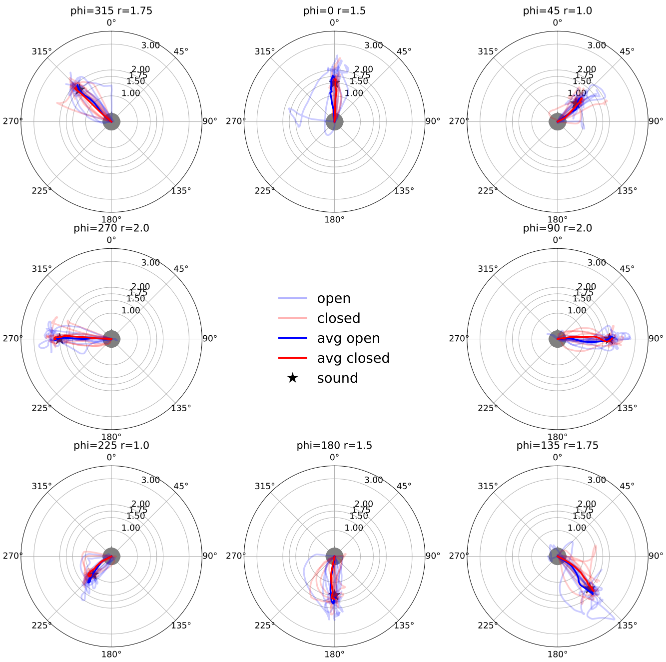

Fig. 4. Locomotion trajectories during the static sound localization task with eyes open (blue) and eyes closed (red). The participants' trajectory is represented with shaded lines, the average path with a solid line, and the sound position with a black asterisk.

Figure 5 shows the locomotion trajectories of all participants during the guidance-task. The individual trajectories are represented with a dashed lightgray line. The average trajectory participants were walking is shown as solid line in blue for the eyes open condition and in red for the eyes closed condition. Target trajectories representing the moving sound source are shown by a dashed black line. Participants were able to follow the moving sound source for all given trajectories. The average distances during the guidance-task were 0.60 m, 0.57 m and 0.57 m in the eyes open condition and 0.58 m, 0.48 m, 0.49 m in the eyes closed condition for the eight, clockwise and counter-clockwise circle trajectories respectively. The averaged reproduced shapes in the eyes open condition are clearly recognizable.

IV. Discussion

The localization accuracy was higher in the eyes closed condition, compared to the eyes open condition. A similar virtual sound localization task was reported in [3]. The error distances between the final position of the participants and the target position was in the range of 0.21 meter (0.15 m to 0.36 m). To reach the final position search times between 9.6 seconds and 25.2 seconds were needed. Our results show slightly larger variations in the localization error than those found in [3]. However, our averaged localization errors are slightly lower than the ones previously reported. Half of the participants in [3] were experienced in tracking or homing

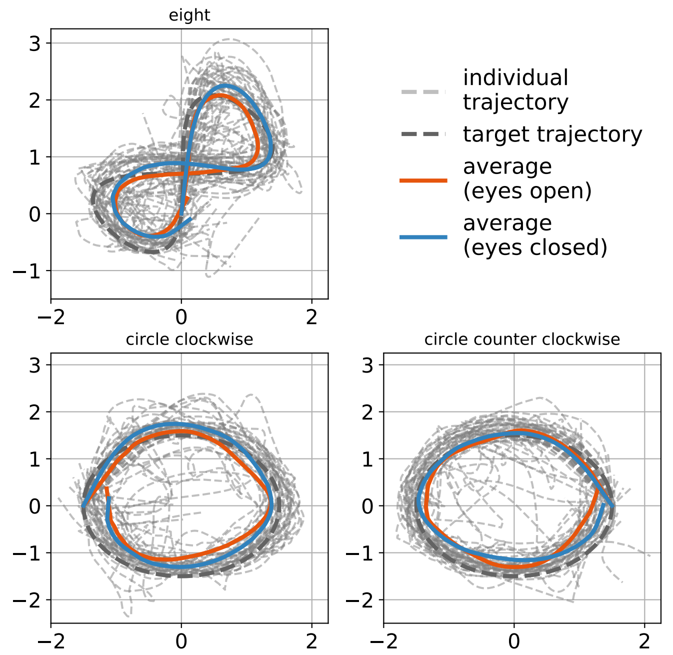

Fig. 5. Locomotion trajectories during the dynamic sound guidance task. The participants' trajectory is represented with a dashed lightgray line. The average path during eyes open condition is shown with a solid blue line, during the eyes closed condition with a solid red line. The target trajectory representing the moving sound source is shown with a dashed black line.

to virtual sound sources, whereas participants in our study had no previous experience and only one trial per visual condition to get familiar with the setup. This might explain the slightly larger range of error distances we measured. The on average smaller localization errors could be credited to the higher update rate of the sound stimulus. Position updates and therefore the sound intensity was adjusted at a speed ten times higher than in their study. Also, instead of rough approximations for the acoustic influence of the head, measured head related transfer functions from an artificial head were used as part of the software soundscape renderer. The large increase in time needed to finish the task compared to [3] could have been emerged from the different room sizes used in the experiments. Since participants have seen the room and were aware of the spatial conditions, the size of the room might have led participants to perform more subtle movement and avoid quick walking out of expectation for sound targets to be closer. [3] conducted their experiment inside a gymnasium on an area of 225 square meters, whereas our study was performed in a medium sized rectangular room within an area of 19.6 square meters.

V. Conclusion

An auditory guidance system based on motion tracking and sound synthesis was proposed. We asked ten participants to localize eight sound sources in eyes open and eyes closed condition, to verify the positioning accuracy of the system. The results confirmed that participants were able to localize the implemented virtual sound sources. The system also showed promising results in guiding people along simple and more complex trajectories. In future applications this setup could easily be extended for example to provide auditory feedback on posture and gait during therapy. An application to guide visually impaired people would also be possible if the outside-in motion tracking is replaced by an inside-out tracking.

References

[1] A. Lindau, and S. Weinzierl, "FABIAN - An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom," Bericht der 24. Tonmeistertagung, pp. 621–625, January 2006.

[2] M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods", In 124 th AES Conv, vol. 8, 2008, pp.1–9, http://www.aes.org/e-lib/browse.cfm?elib=14460.

[3] J. M. Loomis, C. Herbert, and J. G. Cicinelli, "Active localization of virtual sounds," in The Journal of the Acoustical Society of America, vol. 88, 1990, pp. 1757–1764, doi: 10.1121/1.400250.

[4] A. P. Yelnik, S. Tasseel Ponche, C. Andriantsifanetra, C. Provost, A. Calvalido, and P. Rougier, "Walking with eyes closed is easier than walking with eyes open without visual cues: The Romberg task versus the goggle task," in Annals of Physical and Rehabilitation Medicine, vol. 58, 2015, pp. 332–335, doi: 10.1016/j.rehab.2015.08.004.

[5] F. Bevilacqua, E. O. Boyer, J. Françoise, O. Houix, P. Susini, A. Roby-Brami, and S. Hanneton, "Sensori-Motor Learning with Movement Sonification: Perspectives from Recent Interdisciplinary Studies," in Frontiers in Neuroscience, vol. 10, 2016, p. 385, doi: 10.3389/fnins.2016.00385.

[6] R. B. King, and S. R. Oldfield, "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays," in Human Factors, vol. 39(2), 1997, pp. 287-295, doi: 10.1518/001872097778543895.

[7] M.H. Thaut, G. C. McIntosh, R. R. Rice, R. A. Miller, J. Rathbun, and J. M. Brault, "Rhythmic auditory stimulation in gait training for Parkinson's disease patients," in Movement Disorders, vol. 11, 1996, pp. 193-200, doi: 10.1002/mds.870110213.

[8] M. Gori, G. Cappagli, G. Baud-Bovy, and S. Finocchietti, "Shape perception and navigation in blind adults," in Frontiers in Psychology, 2017, vol. 8, pp. 1-12, doi: 10.3389/fpsyg.2017.00010.

[9] F. Dollack, M. Perusquia-Hernandez, H. Kadone, and K. Suzuki, "Head anticipation during locomotion with auditory instruction in the presence and absence of visual input," in Frontiers in Human Neuroscience, 2019, 13:293, doi: 10.3389/fnhum.2019.00293.