

EMPLOYING SPATIAL SONIFICATION OF TARGET MOTION IN TRACKING EXERCISES

Simone Spagnol

IUAV - University of Venice
sspagnol@iuav.it

Michele Geronazzo, Federico Avanzini

Dept. of Information Engineering
University of Padova
avanzini@dei.unipd.it

Fabio Oscari, Giulio Rosati

Dept. of Management and Engineering
University of Padova
giulio.rosati@unipd.it

ABSTRACT

This paper presents the results of an experiment in which the effect of spatial sonification of a moving target on the user's performance during the execution of basic tracking exercises was investigated. Our starting hypothesis is that a properly designed multimodal continuous feedback could be used to represent temporal and spatial information that can in turn improve performance and motor learning of simple target following tasks. Sixteen subjects were asked to track the horizontal movement of a circular visual target by controlling an input device with their hand. Two different continuous task-related auditory feedback modalities were considered, both simulating the sound of a rolling ball, the only difference between them being the presence or absence of binaural spatialization of the target's position. Results demonstrate how spatial auditory feedback significantly decreases the average tracking error with respect to visual feedback alone, contrarily to monophonic feedback. It was thus found how spatial information provided through sound in addition to visual feedback helps subjects improving their performance.

1. INTRODUCTION

Understanding how the human central nervous system combines different kinds of simultaneous information such as proprioceptive, visual, or auditory feedback is today an open issue. The main goal of our research is to investigate the role of sound in motor learning and motor control as an additional or substitutive sensory information to the visual and proprioceptive modalities, with the final aim of incorporating optimized real-time auditory displays related to one or more variables (e.g., target velocity or position error) in augmented-feedback robotic rehabilitation systems.

Unfortunately, the consistent use of auditory feedback in robot-assisted rehabilitation has been largely overlooked in recent related literature. Despite the evidence that a proper sound may help individuals in learning a motor task [1, 2], the precise ways in which mental engagement, repetition, kinematic error and sensory information in general translate into a pattern of recovery is not well defined for rehabilitation [3]. Audio is used in many rehabilitation sys-

tems with the purpose of motivating patients in their performance; nevertheless, in the majority of these systems the audio component plays mostly a marginal role, for instance by offering a positive or negative feedback if the patient completes or fails a task, or by reinforcing the realism of a virtual reality environment [4].

However, the use of auditory feedback could contribute to overcome some of the current main limitations of rehabilitation systems in terms of user engagement, acute phase rehabilitation, standardization of the rehabilitation process, and development of home rehabilitation devices [4]. In particular, sound is thought to be effective in the recovery of activities of daily living (ADLs). As a matter of fact, ADLs rely on an essentially continuous and multimodal interaction with the world, which involves visual, kinesthetic, haptic, and auditory cues. Such cues integrate and complement each other in providing information about the environment and the interaction itself, both in complex tasks (as walking) and in relatively simpler ones (as a reach and grasp movement). To this regard, in order to effectively represent the environment and/or the user's movements, continuous forms of auditory feedback ought to be used in conjunction with other sensory modalities. An incentive to the present research is offered by the observation that audio, just like video, is more direct and requires less attention than proprioception as input modality [5]. Hence, auditory feedback can be potentially relevant not only as a stimulation to augment patient's engagement and motivation, but also as an additional or substitutive straightforward information towards the improvement of performance and learning.

A previous work on robot-assisted upper limb tracking movements revealed that providing subjects with auditory feedback of tracking error could effectively increase subjects' effort and reduce the effects of visual distraction [6]. Similarly, in a group of related experiments performed on healthy subjects and with no robotic assistance the authors argued that auditory feedback can also be effective in reducing tracking error [7]. In particular, continuous task-related information provided through sound in addition to visual feedback can improve not only performance but also learning of a novel visuomotor perturbation. As a new work along this research thread, the here presented experiment can be seen as a further missing tile towards the definition of an effective auditory display for conveying informative content to the user during target following exercises. In particular, the aim of the experiment presented

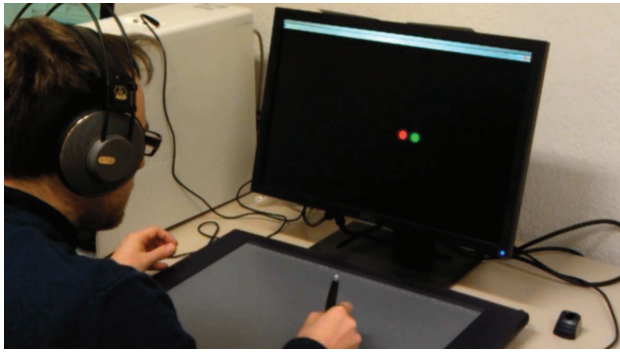


Figure 1. Subject performing a trial of the experiment.

here is to investigate whether the information given to the user by spatial task-related auditory feedback helps the subject improving his or her performance more than monophonic task-related feedback, or not.

2. THE EXPERIMENT

2.1 Experimental setup

As pictured in Figure 1, each participant in turn was provided with a pair of common headphones that presented auditory feedback and a Wacom pen as controller. During the whole experiment, the subject was sitting in front of a Full HD screen and a Wacom pen tablet suitably calibrated in order to match the screen size, both positioned on a common desk and close to each other. The screen was backed by a blank wall.

A simplified scheme of the system's architecture is reported in Figure 2. The graphical user interface for the experiment was implemented in MATLAB and consists of two color-filled, 25-pixel-radius dots displayed on the screen, one representing the controller's position (green dot) and one the target's position (red dot). The target performs a continuous horizontal movement (left to right and *vice versa*) with a minimum-jerk velocity profile. Each task has a specific target motion profile, which can be either

- a *fixed-length* profile, where the length of every left-to-right or right-to-left segment is set to 60% of the screen width (corresponding to a range of motion for the subject's hand of approximately 30 cm) at each iteration; or
- a *random-length* profile, where the length of each segment pseudo-randomly varies from 20% to 90% of the screen width. At the end of the task, the total distance spanned by the target is equal to that travelled in the former case.

Auditory feedback was developed in Pure Data (PD) [8]. Target motion data (i.e., velocity in the x direction) is sent in real time to PD through the OSC (Open Sound Control) protocol. Basically, one sort of auditory feedback was designed for this experiment, a *task-related* sonification of the target roughly simulating the sound of a rolling ball. In order to efficiently obtain such feedback, the instantaneous

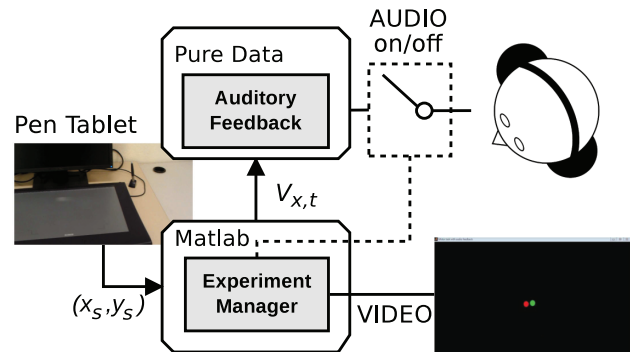


Figure 2. A simplified scheme of the experimental setup.

velocity of the target was applied as a simple gain factor onto the output of a pink noise generator filtered through a bandpass filter with center frequency $f_c = 300$ Hz and Q-factor equal to $Q = 10$, as Figure 3 sketches.

What makes the difference between the two audio modalities for the experiment is the exploitation (or not) of head-related transfer functions (HRTFs) [9] for sound spatialization. To this end, the previously described task-related audio signal can be fed to a binaural spatialization filter (provided by the earplug~ PD external) which renders the angular position of the sound source relative to the subject in the horizontal plane by convolving the incoming signal with left and right HRTFs of a KEMAR manikin. In order to minimize inconsistencies arising from the absence of a head tracking system during spatial audio listening, subjects were told to head towards the center of the screen during the whole task.¹

2.2 Experimental protocol

A total of 16 healthy subjects participated to the experiment. They were aged 19 to 42 (mean age 26.31 ± 6.46), 50% male and 50% female, caucasian, and right-handed. All the participants self-reported normal vision, no color blindness, and no hearing problems.

Each participant was asked to complete six different tasks. During each task, the subject had to draw a trajectory onto the tablet with the pen in order to follow as closely as possible the target presented on the screen. The six tasks, presented in a random order, were:

- task A: fixed-length trajectory, no auditory feedback;
- task Br: random-length trajectory, no auditory feedback;
- task C: fixed-length trajectory, monophonic feedback;
- task Dr: random-length trajectory, monophonic feedback;
- task E: fixed-length trajectory, spatial auditory feedback;
- task Fr: random-length trajectory, spatial auditory feedback.

¹ Admittedly, head movements were seen to be extremely limited, probably also thanks to the “frontal” configuration of the task.

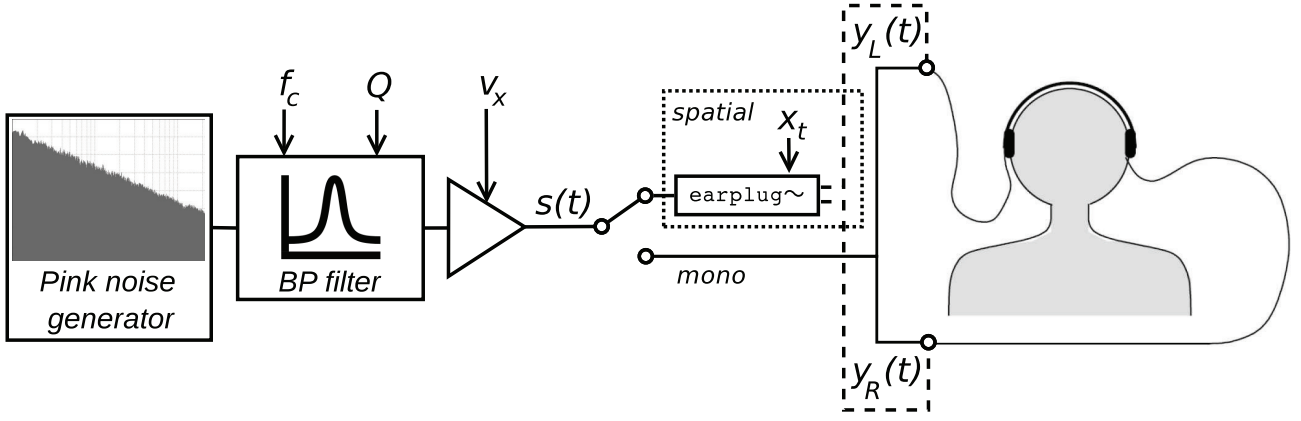


Figure 3. Auditory feedback generation process. The switch, controlled by the experiment manager, selects one of the two audio modalities, i.e. spatial auditory feedback or monophonic feedback.

Each task lasted 80 seconds and consisted of 14 repetitions of the left-right-left cycle. During each task, target position x_t , target velocity $v_{x,t}$, and 2-D subject position (x_s, y_s) were sampled at a frequency $f_s = 50$ Hz. After a first warm-up task showing no target, during which the subject could get acquainted with the tablet, she or he executed the six tasks in a random order. During the three seconds preceding the beginning of each task, a countdown was simulated through a sequence of three tonal beeps.

2.3 Data analysis

All the just mentioned data for a single task was recorded in real time and stored in a matrix. Hence, the full dataset comprised 96 matrices (6 tasks per 16 subjects). Prior to the analysis of such dataset, the subject's velocities in the two axes $(v_{x,s}, v_{y,s})$ were first calculated on the basis of the difference between the current $(x_s(k), y_s(k))$ and the previous $(x_s(k-1), y_s(k-1))$ indicator positions, and then smoothed with a fifth-order moving average filter.

For each task, the integral of relative velocity (i.e., the difference between subject's and target's velocities) and the weighted position error along the x -axis were measured. Each measure was calculated for every left-to-right and right-to-left segment, and then averaged over the whole task. A small number of segments (11 over all subjects and tasks, i.e. 11 over 2688) in which the participant clearly failed to follow the task, suddenly moving to the opposite direction (due to losing control of the pen or occasional distraction), were excluded from the analysis.

The *integral of relative velocity* for the k^{th} segment is formally defined as

$$R_v(k) = \frac{1}{L_k} \int_{t_k}^{t_{k+1}} |\vec{v}_r| dt, \quad (1)$$

where $|\vec{v}_r| = |\vec{v}_s - \vec{v}_t|$ is the norm of the relative velocity vector, L_k is the length of segment k , whereas t_k and t_{k+1} are the start and end times of the segment, respectively. R_v

was calculated using the classical rectangle method:

$$\sum_{h=1}^N \frac{\sqrt{(v_{x,s}(h) - v_{x,t}(h))^2 + (v_{y,s}(h) - v_{y,t}(h))^2} \cdot dt}{L_k} \quad (2)$$

where N is the number of samples in the segment. The R_v parameter measures the extra distance travelled by the subject while following the target, accounting for the movements made to correct tracking errors. A null value of this metric indicates that the velocity profile of the target was exactly reproduced by the subject, even though the average position error (in terms of a constant offset) may have been not null.

The position error along the x -axis was weighted with the target velocity sign and normalized to the target radius R . The *average weighted position error* for segment k is formally defined as

$$e_x(k) = \frac{1}{N} \sum_{h=1}^N \frac{(x_s(h) - x_t(h)) \cdot \text{sign}(v_{x,t}(h))}{R}. \quad (3)$$

This formula takes into account the direction of motion of the target, thus showing whether the subject leads (positive error) or lags (negative error) the target during the exercise. To this regard, *lead error* can be defined as the tracking error when the subject indicator anticipates the target (i.e., leads the target motion), while *lag error* is the tracking error when the subject indicator follows the target. Formally, positive terms in the summation in Eq. (3) contribute to lead error calculation, while negative terms contribute to lag error calculation. A null value in the position error metric indicates that the subject had an average null delay with respect to target motion, even though the distance travelled around the target may have been not null.

A comparison between paired data (D'Agostino and Pearson omnibus normality test [10]) was performed, resulting in a Gaussian distribution for tasks Br-C-Dr-E-Fr (integral of relative velocity), A-Br-Dr-E-Fr (weighted position error and lead error), and A-Br-Dr-E (lag error). Consequently, either parametric or non-parametric (Wilcoxon) paired t-tests were performed in order to compare performance parameters among different tasks. The significance

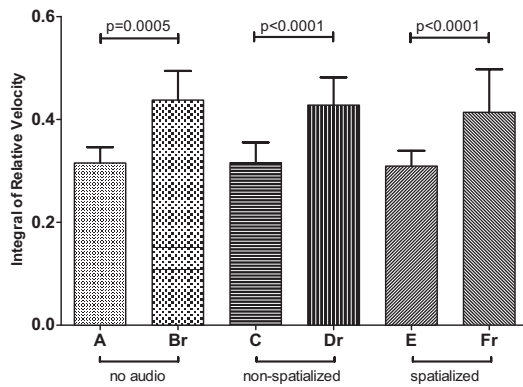


Figure 4. Statistical analysis on the integral of relative velocity.

level for the statistical analysis was set to $p = 0.05$.

3. RESULTS AND DISCUSSION

The only relevant result of the statistical analysis on the integral of relative velocity, reported in Figure 4, is that - as one may expect - the fixed-length task is always much better executed than the corresponding random-length task: subjects made significantly greater corrections in the latter, independently of the audio modality. This result confirms those found in [7].

Conversely, no significant difference between fixed- and random-length tasks within the same audio modality is evidenced by the statistical analysis on the average weighted position error, as Figure 5 points out. In this case, it is the auditory feedback modality that makes the difference. Both the fixed-length audio tasks C and E present a smaller negative error with respect to task A, and the same applies to random-length audio tasks Dr and Fr with respect to task Br. However, only the spatial audio tasks report significant difference with respect to the no-audio tasks, while monophonic ones do not. In other words, only spatial task-related auditory feedback (tasks E and Fr) helps subjects to significantly reduce average tracking delay with respect to having no auditory feedback, both in the fixed-length and in the random-length tasks.

Monophonic feedback lies between the other two modalities in such terms, even though not reporting significant statistical differences with respect to both. It has however to be pointed out that, for fixed-length tasks, the number of outlier cycles (with respect to the weighted position error metric) in the executions of task C is much larger than that of the executions of tasks A and E: this could indicate that the sensory integration of video and audio was more difficult in the monophonic audio condition, especially during the very first cycles of the task.

When comparing these results to the related ones reported in [7]², the keen observer will note that Figure 5 exhibits smaller differences in average tracking error values between

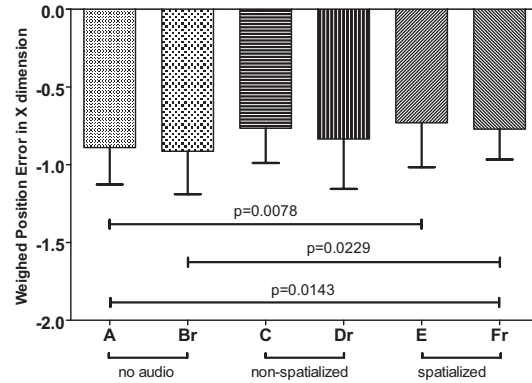


Figure 5. Statistical analysis on weighted position error.

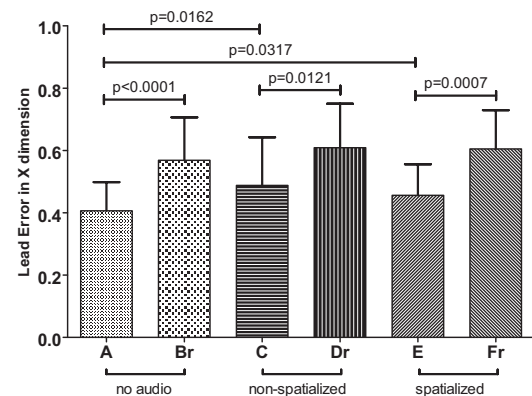


Figure 6. Statistical analysis on lead error.

tasks A-E and Br-Fr with respect to the equivalent feedback couples A-C and Br-Dr in [7]. This may be partly due to the slightly different settings of the rolling sound. However, the statistically significant upgrade given by the spatial task-related auditory feedback is preserved.

While the analysis of lag error does not add much with respect to the previous measure (means, standard deviations, and significance levels are similar to those reported in Figure 5), lead error (reported in Figure 6) is found to be statistically different both between fixed-length and equivalent random-length tasks and among fixed-length tasks themselves. In particular, lead error in task A is significantly lower than in tasks C and E. This result is harder to interpret than the previous ones; still, it could be suggested that the lead error component is greater in random-length tasks because of the sudden, unpredictable deceleration phase for short segments, whereas in fixed-length tasks lead error is lower but tends to increase in presence of consistent auditory feedback. Probably, the additional information allowed subjects to feel more confident while executing the task, tending sometimes to lead the target's movement. It is thus found that task-related auditory feedback involves actions that aim at increasing lead error.

To sum up, the effect of spatialization applied to task-related auditory feedback is found to be overall benefi-

² see Fig. 5 in [7]

cial in the performance of tracking movements. However, the use of generalized HRTFs together with the absence of headphone compensation or reverberation could surely have limited the realism of the spatialization in a number of subjects, psychoacoustically resulting in a trivially panned, non-externalized version of the monophonic feedback [11]. It has indeed to be mentioned that half of the subjects (8 out of 16) informally reported no significant difference between the two audio modalities, and that 4 among them explicitly felt that the rolling auditory feedback was confusing, preferring the condition with no audio. Still, the other half peremptorily affirmed that spatialization added useful information to the task, by helping in particular during the most sudden acceleration and deceleration phases and by letting the subject better concentrate on the task.

4. CONCLUSIONS AND ONGOING WORK

Whereas differences between spatial and monophonic task-related auditory feedback were not seen to be particularly marked, spatial feedback led to a statistically better performance than with no audio that could not be attested for monophonic feedback, even though the information provided by generalized HRTFs could not be unanimously appreciated. In light of this, along with improvements in the monophonic signal, a required step towards a better rendering of the used feedback is the exploitation of customized HRTFs [12]. We are currently focusing on how to improve the spatial sound rendering through a customized HRTF model not involving any cumbersome measurement [13]. The use of customized HRTFs will be expected to ultimately positively augment the gap between performances in the no-audio and spatial audio conditions.

The influence of auditory feedback was studied on healthy subjects first to characterize the normative response of the human motor system to auditory information, yet the experiment should be adapted to a rehabilitative scenario in order to attest the absolute effectiveness of spatial sonification in tracking tasks. However, the presented results definitely provide a basis for a future comparison with impaired subjects.

Acknowledgments

The authors wish to thank all the people who took part to the experiment for their precious collaboration.

5. REFERENCES

- [1] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE Multimedia*, vol. 12, no. 2, pp. 60–69, April-June 2005.
- [2] G. Lemaitre, O. Houix, Y. Visell, K. Franinovic, N. Misdariis, and P. Susini, "Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron," *Int. J. Human-Comp. Studies, special issue on Sonic Interaction Design*, vol. 67, pp. 976–993, 2009.
- [3] G. Rosati, "The place of robotics in post-stroke rehabilitation," *Exp. Rev. Med. Devices*, vol. 7, no. 6, pp. 753–758, 2010.
- [4] F. Avanzini, A. De Götzen, S. Spagnol, and A. Rodá, "Integrating auditory feedback in motor rehabilitation systems," in *Proc. Int. Conf. Multimodal Interfaces for Skills Transfer (SKILLS09)*, Bilbao, Spain, December 2009.
- [5] T. Seizova-Cajic and R. Azzi, "A visual distracter task during adaptation reduces the proprioceptive movement aftereffect," *Exp. Brain Res.*, vol. 203, pp. 213–219, 2010.
- [6] R. Secoli, M.-H. Milot, G. Rosati, and D. J. Reinkensmeyer, "Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke," *J. NeuroEng. Rehab.*, vol. 8, no. 21, 2011.
- [7] G. Rosati, F. Oscari, D. J. Reinkensmeyer, R. Secoli, F. Avanzini, S. Spagnol, and S. Masiero, "Improving robotics for neurorehabilitation: Enhancing engagement, performance, and learning with auditory feedback," in *Proc. IEEE 12th Int. Conf. Rehab. Rob. (ICORR2011)*, Zurich, Switzerland, June-July 2011, pp. 341–346.
- [8] M. Puckette, "Pure Data: Another integrated computer music environment," in *Proc. Int. Computer Music Conf.*, 1996, pp. 37–41.
- [9] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 231–249, April 2001.
- [10] R. B. D'Agostino, "Tests for normal distribution," in *Goodness-of-fit Techniques*, R. B. D'Agostino and M. A. Stephens, Eds. Marcel Decker, New York, 1986.
- [11] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, October 2001.
- [12] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?" *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, 1996.
- [13] M. Geronazzo, S. Spagnol, and F. Avanzini, "A head-related transfer function model for real-time customized 3-D sound rendering," in *Proc. INTERPRET Vork., SITIS 2011 Conf.*, Dijon, France, November-December 2011, pp. 174–179.