Towards Sound-based Crowd Management: Investigating Sonification for Pedestrian Steering

Toros Senan∗
toros.senan@gmail.com
Department of Industrial Design,
Eindhoven University of Technology
Eindhoven, The Netherlands

Alessandro Corbetta∗
a.corbetta@tue.nl
Department of Applied Physics,
Eindhoven University of Technology
Eindhoven, The Netherlands

Bart Hengeveld
b.j.hengeveld@tue.nl
Department of Industrial Design,
Eindhoven University of Technology
Eindhoven, The Netherlands

ABSTRACT
Enabling automated and non-intrusive management of pedestrian flows can help unlock intelligent public environments to maximize individual comfort and wayfinding efficiency. An effective option for this purpose is using dynamic nudging stimuli generated through the use of smart technologies. For this purpose, recent experimental studies have considered visual stimuli to steer pedestrians into alternative walking routes. This research lays the basis for using sonification as a pedestrian flow nudging mechanism. We have developed and deployed a spatialized dynamic sonification to nudge between alternative walking patterns in a real-life work environment endowed with highly accurate and automated pedestrian tracking. In this paper, we report on our preliminary nudging results showing that sound can alter the typical walking trajectories, yet counterintuitively.

CCS CONCEPTS
• Human-centered computing → Auditory feedback; • Applied computing → Physics.

KEYWORDS
Automated crowd management, Sonification, Pedestrian tracking, Living-labs

ACM Reference Format:

1 INTRODUCTION
Unresting urbanization, ecological transition, and distancing requirements put pedestrian infrastructures under increasing stress and generate a growing demand for pervasive and effective crowd management. Just-in-time guidance on optimal routing can increase individual efficiency while walking in crowds, thus increasing individual comfort and safety. Field stewards are the typical resource for such guidance, yet, due to cost and labor-intensiveness, this solution has limited applicability to high-risk scenarios. On the other hand, breakthroughs in automated pedestrian tracking emerged in the last few years (e.g. [2, 5, 19]) bring the possibility of developing automated crowd management systems.

Tracking has been used to quantify automated pedestrian management approaches. To date, however, only visual-based stimuli have been investigated, and two studies are available. In [3], researchers used a visual static interface to steer pedestrians towards the left or right when faced with an obstacle. The authors collected data in a living-lab setting and reported an influence of 6 % on pedestrian navigation choices. In [8], the authors report a laboratory experiment in which participants were asked to choose what they perceived as ‘the fastest lane’. The highest flow was achieved when the average speed of each lane was presented visually to participants in real-time.

This paper presents preliminary results of a naturalistic experiment that investigates sonification as a pedestrian steering method: A grid-based sonification design that employs wave-field synthesis (for a review, see [1]) for an enhanced position-sound association, whose output depends on the accurate tracking of nearby pedestrians (with the approach described in [5, 10]). Using sound is motivated by the auditory modality’s unique properties over other modalities. The human hearing system encodes the timing of sound events with remarkable precision [11], which makes us sensitive to acoustic temporal variations. Furthermore, we can process multiple sound streams simultaneously and direct our attention when an unexpected (e.g., a dog barking in an office) event occurs [22]; consequently, we can detect anomalies in our surroundings. Notably, we are highly efficient in sound localization which can be especially beneficial in the context of public space navigation [15]. Thus, our hearing capabilities make us aware of our acoustic environment, and exploiting this capacity using the so-called sonification eases human-computer interaction [9].

Several sonification techniques were developed to represent an existing set of data or information in the auditory modality (for a review, see [7]). Numerous studies report effective use of sonification in guidance tasks [e.g., 14]. Sonification was also proposed as guidance assistant for visually impaired [12, 20, 26], rehabilitation assistant for patients with disabilities [6, 18], positional guidance assistant in surgery [24, 27], and training assistant for improving elite athletes’ movements (for a review, see [17]).

The manuscript is structured as follows: The experiment, with an emphasis on the system, is described in Sec. 2. Then, the preliminary results are presented and discussed together with the next steps in Sec. 3.
Figure 1: (a) Floorplan sketch of the experimental area. The monitored region is framed with a black dashed rectangle. The sound zones (1-4) and silence zone (5) are enclosed by solid blue lines. The six origin/destination regions are denoted with capital letters: A, B, C, D, E, and F. The solid dots at the bottom represent the loudspeakers, and the purple dotted line denotes the reference line for 2.5D WFS. The arrows report the usage modes: The reference trajectory is red, whereas the target trajectory is green. Orange circles are the virtual positions of the sound sources for each sound zone. (b) Photo of the experimental area taken from the staircase above point A.

2 EXPERIMENT

Our audio-based pedestrian steering system is installed in an indoor connection hallway within the authors’ university campus, amidst office spaces and laboratories. Several hundred people walk the hallway during regular working days (which reduce to about one hundred during COVID19-related lockdown). The experimental area, denoted with dashed lines in Fig. 1a, has a rectangular shape of 4.7 m X 3.8 m, while the ceiling is ∼ 9 m high. Fifteen observed walking patterns cross the experimental area connecting any of the six origin/destination regions at the area boundary (regions denoted with letters A–F in Fig. 1a). Among these trajectories, the walking pattern connecting regions A and B is the most frequently used, hence taken as the reference trajectory in the present experiment (red arrow in Fig. 1a).

We designed a grid-based interactive sonification system that generates sound stimuli, piano samples, and aims at bending the reference trajectory by attracting pedestrians towards a target trajectory (green arrows in Fig. 1). The experimental area is divided into four sound zones, denoted as 1, 2, 3, and 4 in Fig. 1a: The zones 3 and 4 are half in width when compared to zones 1 and 2, hence creating a silent zone indicated with number 5 in Fig. 1a. If the pedestrian tracking system detects a pedestrian within the boundaries of the sound zones (zones 1-4), a pre-recorded piano sample is triggered and played back. On the other hand, no sound is played if a pedestrian is detected in zone 5. We hypothesize that sound will steer pedestrians towards the target trajectory by nudging them passing through the sound zones, i.e., avoiding the silent zone (green arrows in Fig. 1), and generating a sequence of ascending or descending piano notes. The system components and the sound stimuli are explained in the following sections.

2.1 Pedestrian tracking

The input component of the system leverages on highly accurate real-time pedestrian tracking, i.e., the estimation of pedestrian positions and trajectories following a regular time sampling. We achieve tracking via a grid of four overhead depth sensors and the HA-HOG localization algorithm [10] (machine learning-based). Depth-based localization is increasingly used in pedestrian dynamics research as it allows real-life pedestrian measurements in a privacy-by-design setting. Overhead depth sensors record “2D distance images”, in which the value of each pixel represents the distance between that point and the camera plane. This approach unlocks more accurate pedestrian localization than traditional color images while never employing privacy-sensitive information. Within images, heads appear as the darkest regions (as the closest to the sensors), shoulders, slightly farther away, are represented in a brighter shade, and, finally, the floor has the highest possible brightness. Notably, depth data is sufficiently rich to allow measurements beyond the mere position, such as shoulder orientation [25]. The juxtaposition of multiple sensors in a grid configuration allows the coverage of potentially limitless regions. We use four sensors (Orbbec Persee) arranged in a 2 x 2 grid. We acquire four raw depth images at 15Hz, which we combine into a unique depth signal [based on the perspective compensation algorithm in 4]. HA-HOG localization and tracking are performed on the real-time streaming data, returning positions, and velocity. All the data from the experiment are stored for statistical analyses.

2.2 Sound engine

The sound engine receives pedestrian position information from the tracking system and triggers short piano samples. There are four
Figure 2: Comparison of the experimental trajectories of pedestrians entering from zone 1 (red frame) and walking across the facility: the case with auditory feedback (left plot), without auditory feedback (middle plot) sharing same y-axis and the probability density function of the two conditions (right plot).

different piano notes, one for each sound zone, and each sound has a virtual position to augment sound zone-piano note association.

One and a half seconds-long piano sounds are generated using a virtual instrument plug-in (Spitfire Audio - Labs) in a Digital Audio Workstation [16]. Each sample consists of two notes that are an octave apart, i.e., a unison interval. Four samples are assigned to four sound zones so that the pedestrians hear a sequence of piano chords ascending in pitch if they walk through the zones in the order of 1-to-4 (or a sequence of piano chords descending in pitch if they walk in the reverse order) and the silent zone, zone 5, generates no sound. The chords assigned to zones 1, 2, 3, and 4 are C4-C5, E4-E5, G4-G5, and C5-C6, respectively.

Each piano sound is pre-processed with a fourth-order Linkwitz-Riley band-pass filter, with 100 and 1000 Hz cut-off frequencies: The low-frequency limit eliminates the differences between the low-frequency responses of the two different sized loudspeakers, and the high-frequency limit reduces the frequency content above the extended aliasing frequency that creates coloration. Then, the band-pass filtered sound is filtered a second time with a pre-equalization filter, up to $f_{\text{alias}} = 1000 \text{Hz}$ [21]. The sound fields of pre-processed monochromatic signals are generated using the driving function for a focused source in a 2.5-dimensional WFS [23, Eq. 2.29b], where the reference line is at 80 cm (horizontal dotted line in Fig. 1a) in the current configuration.

2.3 Apparatus and procedure

Four depth sensors, audio interfaces, and the main single board computer are mounted on a rectangular metal frame, and the metal frame is hung on the ceiling over the experimental area. The sensors are placed at 4.7 m in height. The two audio interfaces transmit 16 channel audio to 16 loudspeakers which are installed at one side of the experimental area. The loudspeakers are spaced with 20.75 cm and placed at the height of 175 cm. The experimental area and the loudspeakers are presented in Fig. 1b.

The sound engine is implemented to the main single-board computer using the Max software [13]. The sound samples are pre-processed by the band-pass and pre-equalization filters. The WFS system and the sound samples are calibrated to produce focused sources at 52 dB$_{\text{Aeq}}$ when measured on the reference line. The virtual positions of sound samples divide the loudspeaker array into four sections, and each sound sample is virtually placed at 30 cm in front of the loudspeaker array (orange circles in Fig. 1a).

Each sound zone would trigger the corresponding piano sound if that zone is currently inactive (no pedestrian). This protocol prevents retriggering of the same piano sound when a pedestrian moves within the same zone and when a second pedestrian enters the active zone. We observed less than five pedestrians passing through the experimental area simultaneously during the pilot trials, and we limited the system to consider a maximum of five pedestrians at the same time.

3 RESULTS AND DISCUSSION

We report in Fig. 2 the trajectories collected during our experimental campaign (14 working days of which eight including auditory feedback). We restrict to trajectories of pedestrians walking from point A to point B and neglect all other trajectories of any other user mode. A typical trajectory in this context is about 3 seconds long. A total of 50 and 170 trajectories were acquired, respectively, with and without auditory feedback. At the right side of Fig. 2, we report the probability density function for the area corresponding to sound zones 3, 4, 5, comparing the steered and control conditions (yellow frames in Fig. 2).

The plots in Fig. 2 show that under control condition, the trajectories have a larger spread in the y-axis. The results suggest that sound has an impact. Counterintuitively, pedestrians seem to be nudged away from zones 3 and 4 into zone 5, where no sound is played. Pedestrians might have interpreted the sounds as alarm sounds instead of acoustic cues intended to attract them towards
sound zones 3 and 4. We regard the preliminary results as positive: The main objectives of the experiment are to understand the impact of sound on walking pedestrians and its effect on steering. We, however, postpone solid conclusions since the sample sizes for the two conditions are more than a factor of three different, and the total sample size (N = 220) is low for reaching statistical significance compared to what is reported in the literature (e.g., ~ 9000 trajectories per condition for 6 % steering ratio in [3]).

The experimental setting and procedure allow continuous data collection in a living-lab setting which provides in-situ observation of the impact of sound conditions on pedestrian trajectories. We plan to continue with the current acoustic condition for two-to-three more months and acquire about 1000 additional trajectories. Afterward, we will introduce a continuous interactive sonification condition that provides feedback to the pedestrians throughout the experimental space so they can anticipate their relative position concerning the target trajectory and modify their routing decisions before reaching a sound zone. Furthermore, we will introduce dynamic variations to both sound source parameters (e.g., pitch, loudness, timbre) and spatial sound parameters (e.g., moving sound sources). After a significant effect of auditory feedback is observed, we will investigate the system’s feasibility for real-life applications. The pedestrian tracking and sound system are currently sized to reach very high performance: sub-centimeter accuracy in pedestrian tracking and state-of-the-art spatial sound reproduction. A forthcoming phase of this research will focus on identifying the minimum technological requirements for a realizable, and scalable sound-based automated pedestrian steering system.

ACKNOWLEDGMENTS
This work is part of the VENI-AES research programme “Understanding and controlling the flow of human crowds” with project number 16771, which is financed by the Dutch Research Council (NWO).

REFERENCES

Torus Senan, Alessandro Corbetta, and Bart Hengeveld