Sound-Guided 2-D Navigation

Effects of Information Concurrency and Coordinate System

Kantan, Prithvi Ravi; Dahl, Sofia; Spaich, Erika G.

Published in:
Participative Computing for Sustainable Futures - Proceedings of the 12th Nordic Conference on Human-Computer Interaction, NordiCHI 2022

DOI (link to publication from Publisher):
10.1145/3546155.3546688

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Kantan, P. R., Dahl, S., & Spaich, E. G. (2022). Sound-Guided 2-D Navigation: Effects of Information Concurrency and Coordinate System. In *Participative Computing for Sustainable Futures - Proceedings of the 12th Nordic Conference on Human-Computer Interaction, NordiCHI 2022* [40] Association for Computing Machinery. https://doi.org/10.1145/3546155.3546688

Sound-Guided 2-D Navigation: Effects of Information Concurrency and Coordinate System

PRITHVI RAVI KANTAN*, Aalborg University, Copenhagen, Denmark
SOFIA DAHL*, Aalborg University, Copenhagen, Denmark
ERIKA G. SPAICH*, Aalborg University, Aalborg, Denmark

Auditory guidance conveying positional information through concurrent variations in properties of synthesized sound has previously been investigated. Auditory guidance may be more effective if multidimensional tasks are divided into unidimensional tasks where the user sequentially tackles each dimension and sound property. User performance may also depend on the coordinate system used for providing guidance. We compared concurrent and sequential guidance presentations in Cartesian and polar coordinate systems in a computer-based 2-D target-finding experiment with 15 participants. Sequential guidance was superior regarding completion time and number of interruptions with less cognitive burden than concurrent guidance. Participants were slower with the polar coordinate system than the Cartesian. These findings can contribute to the development of more efficacious guidance systems.

CCS Concepts: • **Human-centered computing → Empirical studies in interaction design**; *Interaction design theory, concepts and paradigms*; • **Applied computing → Sound and music computing**.

Additional Key Words and Phrases: auditory guidance, sonification, navigation, perceptual test, concurrency

ACM Reference Format:
Prithvi Ravi Kantan, Sofia Dahl, and Erika G. Spaich. 2022. Sound-Guided 2-D Navigation: Effects of Information Concurrency and Coordinate System. In *Nordic Human-Computer Interaction Conference (NordiCHI '22), October 8–12, 2022, Aarhus, Denmark.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3546155.3546688

1 INTRODUCTION

Although computer interfaces normally convey information visually, there are situations where the visual modality is fully utilized, impaired, or unavailable, making it beneficial to exploit other sensory routes [11] for certain types of tasks. A clear-cut example is that of sound-based navigation applications for blind people [25]. Interfaces that communicate through sound often do so by the process of *sonification* (converting data to sound), which is more formally defined as the "transformation of data relations to perceived relations in an audio signal to facilitate communication or interpretation" [23]. Aside from providing usability benefits to traditional interfaces [11], interactive (real-time) sonification [19] has been tested in *auditory guidance (AG)* applications in numerous contexts. Here, a human user is interactively guided through a real-life task by an auditory stimulus; use-cases include navigation [1, 9, 28], spatial orientation tasks [41], surgical interventions and precision tasks [3, 20, 33], movement rehabilitation [15, 24], and aircraft control [34]. AG can also be applied in situations where the visual modality is ineffective or unavailable, such as firefighters moving through smoke-filled buildings, military personnel navigating to a rendezvous in darkness, or police navigating through thick teargas in the midst of a protest [38]. In all these cases, the position of a user is continuously measured and compared to

---

a desired/target position or set of coordinates; the AG signal is the sonified difference between user and target positions. The benefits of the auditory medium lie in its potential to complement vision and reduce the associated cognitive load [41] or replace vision entirely in its absence as a form of sensory substitution [26]. However, the real-life adoption of sonification at large [30, 36] and AG in particular [3, 26, 32] remains very limited. Many plausible reasons for this have been cited, such as challenges with interdisciplinarity in sonification research [30], lacking aesthetic and psychoacoustic motivations in sound design decisions [3, 32], and the general unfamiliarity of target users with the relatively nascent field of auditory display [3].

Any navigation task can generally be seen as the combination of a knowledge problem (i.e. understanding where a target location is relative to one's present location) and a perceptual-motor problem (the bodily specifics of executing the task of reaching the target location) [27]. Although 2/3-D navigation tasks are straightforward when using vision, it is relatively hard to convey multivariate navigation data through sound in an unambiguous manner [10, 14, 41]. Existing AG designs have usually been based on a **concurrent** or simultaneous mapping of each positional coordinate to a distinct sonic property [6, 41]. Past experimental results indicate that the coordinate information may be more effectively conveyed to users by sonifying each spatial dimension **sequentially** instead, breaking up a 2- or 3-D navigation task into a series of 1-D sub-tasks. Moreover, the chosen coordinate system itself may dictate user behavior and strategy, influencing how easily and effectively the user is able to navigate. A point on a 2-D plane can be represented as a pair of positional coordinates (x, y) along two orthogonal axes in the Cartesian system or in the form of an angle (bearing) and a distance from the origin (range) in the polar system (θ, r). Existing applications have employed either the Cartesian [7, 12, 41] or the polar [1, 5, 17] system, but to the best of our knowledge not compared their guidance efficacy (in terms of completion time, trajectory length, cognitive load, etc.) side-by-side when controlling for other parameters. The goal of our present study was to experimentally assess AG guidance efficacy in a computer-based 2-D navigation task, specifically comparing (i) the concurrent and sequential sonification of dimensional coordinates, and (ii) the sonification of Cartesian and polar coordinates.

2 BACKGROUND

2.1 Coordinate Systems Used in AG Applications

Past studies of AG for navigation or spatial orientation purposes have typically sonified either Cartesian or polar information depending on the task as well as the type of user positional data available from measurement apparatus.

Polar System: AG systems for locomotion guidance (e.g. GPS-based navigation) tend to be based on the polar coordinate system as it is readily suited to how humans locomote (by turning to face the desired direction and moving forward until the destination or next turn). Users usually have to 'follow the sound' representing the target location (such as beacon sounds [37, 39], music [17, 34], or noise [9]), whose azimuth position is encoded as the perceived location of the sound in the stereo field [9, 37, 39]. Complex routes are commonly broken down into simpler segments separated by *waypoints* where a direction change usually occurs [37, 39, 40]. The user navigates from waypoint to waypoint in order to reach the final target location, and target sound varies depending on whether the waypoint is in front of or behind the user (e.g. changes in perceived position [37] or timbre [39]).

Cartesian System: For non-locomotion tasks where user position is available as a set of orthogonal coordinates in a plane/space, Cartesian AG schemes have been applied. In a study on audio-guided surgical needle placement [4], left-right sound panning (placement) was used to provide x information, while musical pitch metaphors were used to convey y information to guide an operator to a target point in a 2-D plane. In a dynamic balance training application [12], a force-based measurement system tracked the displacement of the center of mass of the body in the horizontal plane (left-right and front-back - a Cartesian paradigm). The authors built and tested several 2-D AG paradigms, where Cartesian information was sonified in the form of manipulations to synthesized stimuli (both musical and nonmusical). They found that the specific type of sound manipulation used to convey positional information was a strong determinant of user performance and preferences [12]. Ziemer and Schultheis [41] tested a psychoacoustically informed system that guided users to a target on a 2-D plane by sonifying x and y information through perceptually distinct auditory properties (pitch, beating, roughness) of a single synthesized sound. In all these studies, users were able to navigate to target points using the AG, but task performance varied greatly among individuals. Users were also considerably faster and reported a lower cognitive load with visual guidance [4, 12, 41].

It makes little sense to use the Cartesian system in GPS-based locomotion (e.g. comparing user latitude and longitude to that of the destination and sonifying these differences) because (i) real-life navigation paths between two points may be unpredictable based on environmental obstacles, and (ii) Cartesian information is not directly applicable to the locomotion paradigm of turning, facing, and moving forward. However, there is no such constraint impeding the use of the polar system in 2/3-D spatial orientation [41], balance training [12], and surgical needle placement [4] where only Cartesian AG has previously been tested. It is possible that using the polar system is more efficient in these settings as it can allow users to precisely define the angle of movement and traverse direct, short paths to the target. As none of those studies compared Cartesian and polar AG, one of our goals is to explore which coordinate system enables superior navigation performance in tasks where either is practical to apply.

2.2 Sequential Presentation of Dimensionwise Guidance in Multidimensional Tasks

Findings from past studies provide useful insights into user behaviors, which can guide the design of more effective AG schemes for multidimensional navigation. Previous Cartesian AG designs for 2-D navigation simultaneously sonified user position along each axis using a distinct perceptual property of sound (e.g. pitch, volume fluctuations, inharmonicity, etc.) [4, 12, 41]. This, in theory, enabled users to navigate along both dimensions at once to reach the target. In [41], it was observed that users tended to approach the task in an axis-by-axis manner, i.e. by navigating along one dimension at a time until the target was reached - essentially breaking the multidimensional navigation task into a series of 1-D tasks. This behavior was not exhibited when visual guidance was provided; users manipulated multiple dimensions at once and followed more direct navigation paths [41]. Chung et al. [6] observed a similar axis-by-axis tendency where users first navigated horizontally using and then vertically. Due to each axis being given a separate sonic representation in the Cartesian AG schemes [4, 12, 41], we argue that these designs were the auditory equivalent of a visual display with X- and Y-coordinates printed separately on-screen and updated in real-time. Such a visual display was tested as a virtual navigation aid by Darken et al. [8], who observed similar behavior to [6, 41] - the users treated movement along each dimension separately. These findings suggest that that regardless of the sensory modality, even if positional information is concurrently conveyed for each axis through a distinct 'information stream', users tend to sequentially tackle individual dimensions.

Unambiguously conveying multidimensional information through sound is a challenge due to psychoacoustic constraints. In order for users to disambiguate the information streams, they must be independent/orthogonal to each other, meaning that moving along one dimension should not affect the perception of the the target location along other dimensions. While this is trivially easy to achieve in a visual display as in [8], it is well understood that several properties of sound interact/interfere perceptually with one another [29]. For instance, if pitch represents one dimension and loudness represents another, manipulating either dimension causes perceptible changes in the sound property representing the other dimension [29]. Even other sonic properties such as timbre, roughness and fluctuation strength indirectly involve pitch/loudness modification of a sound and/or of its constituent harmonics. But despite this inevitability of perceptual interactions, it often suffices for AG design purposes that the sonified auditory properties are perceptually separable if not fully independent [42]. The key implication of perceptual interactions for multidimensional AG is that manipulating both axes at once is bound to, to a greater or lesser extent, cause confusion in interpreting the resulting changes in the AG signal.

As sonifying both dimensions concurrently is prone to perceptual ambiguity, we propose that a more effective AG approach would be to sonify spatial dimensions *sequentially* (e.g.using 'two-step guidance' as it is termed in [6]). Such a scheme guides the user along each individual dimension serially until the target is reached. *Hence, only one dimension is sonified at a time*, enabling the user to locate each coordinate of the target one-by-one without being confused or distracted by information about other dimensions. Finding the target coordinate along one dimension triggers the sonification of the next dimension, and the process repeats along all dimensions until the target has been found. On the other hand, if the user manipulates a dimension that is not currently being sonified, there is no change in the AG sound. In the present study, we aimed to empirically compare concurrent and sequential AG efficacy in terms of navigation task performance and cognitive load in the Cartesian and polar coordinate systems. Specifically, our research questions were:

(1) How do concurrent and sequential sonification of dimensional coordinates compare in terms of 2-D auditory guidance efficacy (task performance time, trajectory length, etc.)?
(2) How is 2-D AG efficacy influenced by the coordinate system (Cartesian or polar) in which it is provided?

3 METHODS

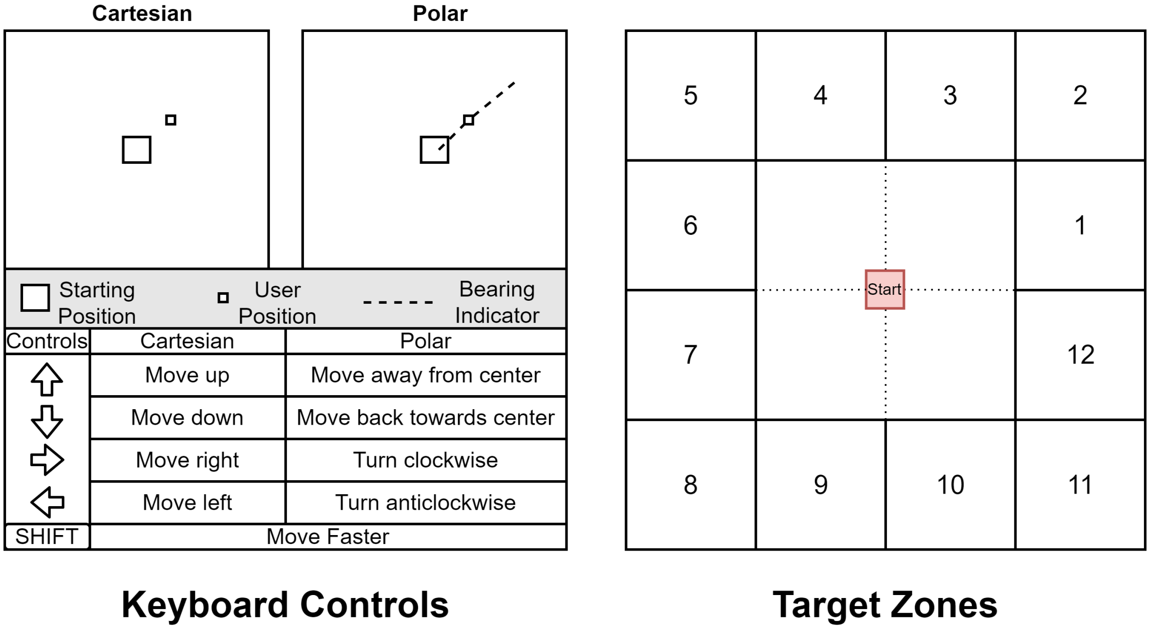
We designed and built a specialized 2-D AG platform in C++ using the JUCE framework[1], with audio functionality implemented using the FAUST programming language[2]. The navigation space was a square black 800 × 800 px area, with the instantaneous position of the user represented by a white 12 × 12 px square. The navigation task was to move the square from the center of the space to an unspecified and invisible target using AG. The interface had additional controls to pseudo-randomly define target coordinates, switch coordinate systems and presentation modes, and record user trajectories during navigation tasks.

Coordinate Systems and Keyboard Controls: The interface supported 2-D navigation in two coordinate systems: Cartesian and polar. The Cartesian coordinate range representing the entire navigation space was defined as (x,y) ∈ [-1, 1]. In the polar system, absolute bearing values $\theta \in [0, 360^\circ)$, with 0 representing the direction of the positive x axis. The range $r \in [0, r_{max}]$, where r_{max} depended on θ. This was due to the shape of the navigation space; the distance between the origin and the corners of the square was 1.414 times its perpendicular distance to each side (see Fig. 2 (right)). Hence the maximum range r_{max} in any direction varied from 1-1.414 depending on θ. In both coordinate systems, the user start position was the origin at the center of the space (0, 0). Keyboard controls were used to manipulate the user position (white square) within the navigation space. Figure 1 (left) shows their respective GUI elements and control schemes. Note that for the polar system, the center of the space was used as the point of reference for both bearing and range, and user position was always locked such that it lay along the direction shown by the bearing indicator. Hence if

---

[1] https://juce.com/
[2] https://faust.grame.fr/

bearing was manipulated whilst keeping range fixed, the white square moved along the circumference of a circular path whose radius was the range.



Controls	Cartesian	Polar
⇧	Move up	Move away from center
⇩	Move down	Move back towards center
⇨	Move right	Turn clockwise
⇦	Move left	Turn anticlockwise
SHIFT	Move Faster	

Fig. 1. LEFT: The navigation space, GUI elements, and keyboard controls for both coordinate systems. RIGHT: A to-scale illustration of the 12 target zones.

Target Area and Randomized Assignment: When a new target-finding trial was initialized in the interface, the target coordinates were defined by a pseudorandom process. Targets could only lie within the 12 zones shown in Fig. 1 (right). To make every zone equiprobable, a random zone order of length 12 was defined at program startup using the JUCE *Random* class functions. This order was repeated indefinitely within that instance of the program (but varied between separate program instances/runs). This guaranteed that (i) targets in consecutive trials lay in different zones, and (ii) targets were defined within all twelve zones before any one zone was repeated. The exact target coordinates within the bounds of the selected zone were then randomly defined. A square target area of 31×31 px was defined around the target coordinates, and the user was considered to have reached the target coordinates if the entirety of the white square lay inside the target area.

AG Variable Computation: The 2-D AG scheme sonified the difference between the instantaneous user coordinates (center of white square) and the target coordinates in the respective coordinate system. Specifically, it sonified positional differences d_1 and d_2 along two dimensions D1 and D2 in either scenario - x/y position in the Cartesian system, and bearing/range in the polar system (see Fig. 2). To allow standardization of units and their scaling across coordinate systems, d_1 and d_2 were normalized to *mapping variables* m_1 and m_2 respectively which ranged from -1 to 1, such that 0 represented the target coordinate along the concerned dimension, and -1/1 represented the extreme positions in either direction. d_1 and d_2, and in turn m_1 and m_2 for each coordinate system were computed as follows:

- Cartesian coordinates: The absolute positional differences were first computed from the Cartesian coordinates of the user u and target t as follows:

  $$d_1 = u_x - t_x \qquad\qquad d_2 = u_y - t_y$$

  m_1 was calculated such that values of -1 and 1 represented the left/right edges of the square navigation space. Similarly, m_2 values of -1 and 1 represented the bottom/top edges of the space. m_1 and m_2 were normalized to scale linearly from 0 to ±1 from target to peripheral edge irrespective of the target coordinate values, and were calculated in terms of the positional differences d_{dim} along each dimension as follows:

  $$\begin{aligned} m_{dim} &= \frac{d_{dim}}{1+t_{dim}} & \forall \quad u_{dim} < t_{dim} \\ m_{dim} &= \frac{d_{dim}}{1-t_{dim}} & \forall \quad u_{dim} > d_{dim} \end{aligned}$$

- Polar coordinates: For any target bearing θ_t and instantaneous user bearing θ_u, it was possible for the user to attain the target bearing by rotating either clockwise or anticlockwise. To guide the user along the shortest possible angular path, d_1 was first calculated as the minimum angular difference ($< 180^\circ$) between θ_t and θ_u. The d_1 sign convention was such that it was positive if a clockwise rotation was the shortest angular path to θ_t and vice versa. m_1 was then obtained by normalizing d_1 through division by 180. Hence,

  $$\begin{aligned} m_1 = 0 &\implies \theta_t = \theta_u && \text{(user bearing matches target bearing)} \\ m_1 = \pm 1 &\implies |\theta_t - \theta_u| = 180 && \text{(user bearing opposite to target bearing)} \end{aligned}$$

  d_2 was then calculated from the user range r_u and target range r_t as:

  $$d_2 = r_u - r_t$$

  The maximum range along the target direction $r_{max}(\theta_t)$ was calculated as the length of the segment along the target bearing θ_t joining the origin and the edge of the navigation square. The normalized variable m_2 was then calculated as follows, such that it was -1 at the origin, 0 at r_t, and +1 at the edge of the navigation square:

  $$\begin{aligned} m_2 &= \frac{d_2}{r_t} & \forall \quad r_u < r_t \\ m_2 &= \frac{d_2}{r_{max}(\theta_t) - r_t} & \forall \quad r_u > r_t \end{aligned}$$

Finally, m_1 and m_2 were forcibly set to zero if the user square lay within the bounds of the target area along the respective dimension in both coordinate systems.

Auditory Guidance Design. Our 2-D AG scheme (summarized in Table 1 and illustrated in Fig. 2) was fundamentally based on the psychoacoustically motivated design in [41], which was shown to effectively communicate spatial information through manipulating a single synthesized sound. In their work, x information was conveyed through the direction (upward/downward) and rate of pitch changes of a continuous tone. y information was represented by auditory dimensions separable from pitch, i.e. beating (amplitude modulation - AM)[3] and roughness (frequency modulation - FM)[4] applied to these tones, depending on whether the user was below or above the target. We generalized this scheme such that the audio engine sonified m_1 and m_2 (normalized d_1 and d_2 values) in a manner agnostic to the chosen coordinate system (see Fig. 2). In terms of auditory parameters, our implementation made several key modifications to that in [41]:

- **Use of Melody:** Instead of using continuous pitch changes, We chose to represent m_1 using a looped major scale melody, one of the simplest musical pitch structures. Similar to [41], the direction of the pitch change indicated

---

[3] This manifested as periodic volume fluctuations in the sound, whose regularity and intensity increased with $|m_2|$

[4] This was applied in the form of periodic frequency fluctuations, manifesting as a perceptual sensation of auditory roughness whose intensity increased with $|m_2|$.

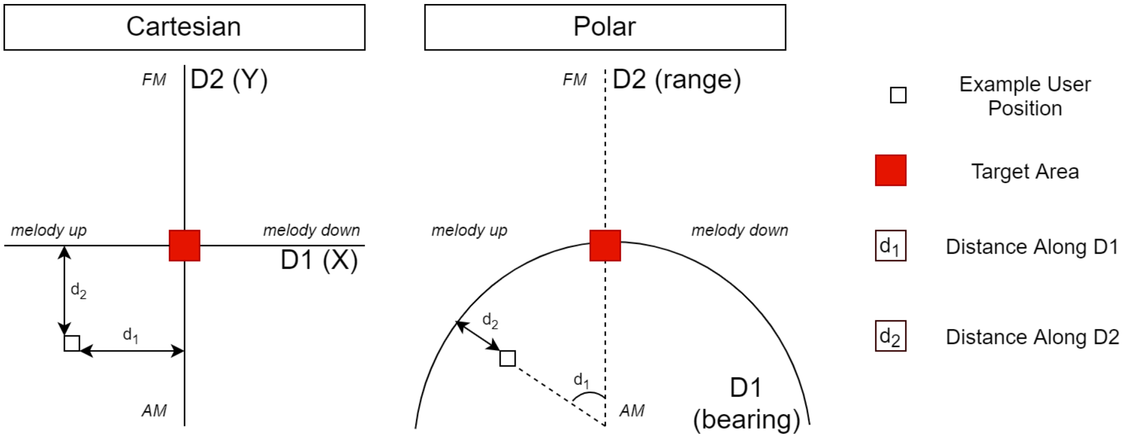


Fig. 2. Our Cartesian and polar AG schemes shown **from the perspective of the target area**. It contrasts how the positional difference variables d_1 and d_2 are calculated for a given position of the user relative to the red target area. The directional sound cues along both D1 and D2 are italicized. Note that d_1 and d_2 were computed from the center of the user square (could not be schematically reproduced).

which direction to move in, whilst the temporal rate of note transitions depended on the absolute value of m_1. In other words, a static note played at $m_1 = 0$, whilst the note transitions were fastest at $m_1 = \pm1$. Our reason for using musical pitches was to make the AG more pleasant to listen to. The melody was synthesized using a pair of octave-separated triangle waves playing in unison and whose overall fundamental frequency spanned the pitch range between the A4 and A5 musical notes.

- **Modulation Intensity:** So as to enhance the perceptual salience of the D2 AG , we opted to amplify the FM and AM effects used in [41], and achieved this by widening the range of *modulation indices* used across the range of m_2 values. For the AM, we used an index range from 0 ($m_2 = 0$) to 1 ($m_2 = -1$) compared to the 0 - 0.5 range used in [41]. This resulted in stronger amplitude fluctuations for negative m_2 values. For the FM, our modulation indices ranged from 0 ($m_2 = 0$) to 500 ($m_2 = 1$). The exact range used by [41] was not specified, but by comparing audio examples, it was clear that ours was far greater.
- **Melody tempo and modulation frequency ranges:** While the tempo of the pitch (chroma) cycle in [41] ranged from 0 Hz to 10 Hz, we kept our maximum melody repetition rate (comprising 8 notes) at 1 Hz instead. For D2, we kept the AM frequency range from 0 Hz ($m_1 = 0$) to 5 Hz ($m_1 = -1$) so as not to interfere severely with the intelligibility of the melody. The FM frequency was kept constant at 40 Hz as in [41].
- **Target Sound:** When the user is within the target area, this was indicated through the sound of a continuously ringing bell (a physics-based synthesis model from FAUST libraries) as opposed to the white noise used in [41].

3.1 AG Presentation Modes

3.1.1 Concurrent: In this mode, the AG manipulations pertaining to both m_1 and m_2 variations in the chosen coordinate system were simultaneously audible irrespective of the user position (as in [41]). Hence, this mode did not put any restriction on the target-finding strategy adopted by the user, who would be free to find the target location by manipulating the dimensions concurrently or individually in any order.

Dim	**Coordinate System**		**Keyboard Controls**	**Auditory Parameter**	**Directional Cues**		**Distance Cues**
	Cartesian	*Polar*			$m \in [-1, 0)$	$m \in (0, 1]$	
D1	X Position	Bearing	← →	Melodic Pitch	Ascending Scale	Descending Scale	Tempo
D2	Y Position	Range	↑ ↓	Modulation	AM (Beating)	FM (Roughness)	Modulation Depth
Notes	*Relative to target*		*Hold SHIFT to raise speed*	-	*Continuous bell sound within target area*		

Table 1. A summary of the dimensional mappings and auditory parameters for both coordinate systems. The *Distance Cues* column indicates how the distance between the user and target coordinates along each axis was communicated to the user.

3.1.2 Sequential: The AG corresponding to each dimension was presented sequentially. Only m_1 was sonified until the user was within the target area bounds along D1. Only when this condition was satisfied was m_2 sonified, allowing the user to then locate the target along D2 whilst keeping the D1 coordinate constant. If D1 was manipulated at this stage and the user moved out of the target area bounds along D1, m_2 sonification was disabled and the m_1 sonification was audible once again until the user once again entered the D1 target area bounds. This mode forced the user to tackle each dimension separately and in a fixed order (D1 ⟶ D2).

A detailed demo video of the AG in all four combinations of coordinate systems and presentation modes is provided in the supplementary material.

4 EXPERIMENTAL PROTOCOL

To investigate the impact of the two coordinate systems (Cartesian (Car) and polar (Pol)) and presentation modes (concurrent (Con) and sequential (Seq)) on user navigation performance and cognitive load, we conducted a within-participant full factorial experiment. At the outset, we framed the following hypotheses:

- **H1: The sequential presentation mode will lead to better performance (smaller task completion time, shorter user trajectories, fewer interruptions and target overshoots) and lower user cognitive load irrespective of the coordinate system.**
- **H2: Users will traverse shorter trajectories when using the polar coordinate system than with the Cartesian irrespective of the presentation mode.**

4.1 Participants

A convenience sample of 16 participants (3 women) aged 26.27 ± 3.83 years volunteered to participate in the experiment. Each of them was briefed about the purpose and length of the experiment beforehand, and informed that they could withdraw at any time. All experimental procedures conformed to the ethics code of the Declaration of Helsinki. Informed consent was obtained prior to participation, and no sensitive or confidential information was collected from the participants. All participants reported normal/corrected to normal vision and all reported normal hearing. The participants had an average of 6.47 years of musical training, and their average self-reported prior experience with sonification and/or auditory guidance on a scale of 1 to 10 was 3.07 (1 = no experience, 10 = highly experienced).

4.2 Experimental Setup

The experiment was conducted in a quiet laboratory using a Dell Windows laptop. The laptop screen was mirrored on a widescreen Samsung monitor (model S34J550WQR) for use by participants, who were also provided an external keyboard (HP KU0316) and mouse (Microsoft Basic Optical Mouse 2.0) to operate the JUCE-built interface and fill questionnaires. Audio Technica ATH-M50X closed-back headphones at a comfortable volume setting (36%) were used throughout.

4.3 Procedure

Participants first underwent a brief tutorial phase where they were given basic information on:

- The coordinate systems and their keyboard controls (where they could also try using the interface)
- The 2-D guidance strategies and their working within each coordinate system (accompanied by audiovisual demos)
- The presentation modes (concurrent/sequential) and their working
- The overall structure of the experiment

A total of four experimental conditions corresponding to all combinations of *Coordinate System* and *Presentation Mode* were then applied in counterbalanced order using a 4×4 Latin square design. For each condition separately *(Car-Con, Car-Seq, Pol-Con, Pol-Seq)*, the participants underwent three phases:

4.3.1 Training Phase. Participants were first given a chance to train in a hands-on manner. Here, they attempted to locate 5-15 targets in the assigned coordinate system and presentation mode until they felt they had gained sufficient familiarity and speed. Once a new target was pseudorandomly defined, the participant initiated the trial by hitting the space bar. This triggered the AG playback, which the participant used to locate the target. When this was done, the participant concluded the trial by hitting the space bar again. During trials, the elapsed time in seconds was displayed on-screen.

4.3.2 Main Testing Phase. This phase comprised a total of 25 consecutive trials, which were conducted without interruption in a manner identical to the training phase. *Participants were expressly instructed to try and navigate to the target as fast as possible.* Here, navigation trajectories corresponding to each trial were recorded at a sampling frequency of 100 Hz and logged for analysis.

4.3.3 Subjective Rating Phase. Upon completing the main testing phase, participants were asked to fill out the NASA-TLX [16] questionnaire to assess their subjective workload. This part of the experiment was carried out using Google Forms. Note that although the original scale has ratings from 0-100 in increments of five, our version used the same range but in increments of ten due to the limitations of Google Forms. Participants were then asked to rate how pleasant and helpful they perceived the AG to be on scales of 1 (very unpleasant/unhelpful)-10 (very pleasant/helpful).

Once this procedure was complete for all four conditions, participants were asked to provide some basic demographic information related to age, gender, years of musical training, and experience with auditory guidance. Finally, they were asked which of the four conditions they preferred most. The entire experiment took approximately an hour.

4.4 Data Analysis and Statistical Analysis

One participant was unable to comprehend the AG scheme and withdrew from participation halfway through the experiment. For the remaining 15, the recorded trajectories were processed in MATLAB to yield a set of performance metrics which were statistically analysed in SPSS 27.0 (using a significance criterion $\alpha = 0.05$ for all tests). For each of the 1500 recorded trials (15 participants × 4 conditions × 25 trials), the following performance metrics were computed:

- **Completion Time:** This was the elapsed duration in seconds between task initiation and completion. It was normalized by the Euclidean distance between the starting point (origin) and center of the target area in order to account for the variability in target location between trials.
- **Trajectory Length:** This represented the total absolute distance traversed by the user (white square) during each trial, normalized in a manner similar to *completion time*.
- **Interruptions:** An interruption was registered if the user paused, i.e. did not manipulate either dimension, for a duration exceeding 400 ms (excluding at the beginning) (duration threshold based on [41]).
- **Target Overshoots:** This was the number of times the center of the user square crossed the target coordinates in either direction (counted separately for D1 and D2).

To minimize the impact of individual differences among participants [2], we normalized the trial-wise absolute value of each metric through division by its grand mean across conditions *for that participant*. The individual NASA-TLX scale items [16] for each condition were averaged (no weighting) to yield *raw scores* representing subjective cognitive load. These, too, were normalized by the participant grand mean. Hence for all normalized metrics, a value of 1.0 represented the average for a participant across all conditions. The data were then checked for normality (Q-Q plots and Shapiro-Wilks tests) and homogeneity of variance (Levene's tests) within each condition, and analysed as follows to test **H1** and **H2**:

Completion Time and Trajectory Length: We first removed a total of 21 outlier values for each metric ($|z| > 3.29$ with respect to overall grand mean). We found that the remaining data were non-normally distributed (strong positive skew) and had unequal variance across conditions. Therefore, we chose to carry out a *repeated measures (RM)* analysis using a *generalized linear mixed model (GLMM)* approach. Here, we first defined the data organization in terms of subjects and repeated measures. Next, we specified *Coordinate System (CoordSys) and Presentation Mode (PresMode),* as fixed factors. Due to the positive skew, we estimated the data distribution as a gamma regression with a logarithmic link function to the linear model. Assuming a diminishing correlation as trial indices grew further apart, we modelled the covariance using a *first-order autoregressive* structure. Using restricted maximum likelihood estimation, the model assessed the effects of *CoordSys* and *PresMode* as well as their interaction, yielding means and 95% confidence intervals for each condition. Summary measures are reported as **mean ± standard error**.

Overshoots: The individual trial data were non-normal with a large proportion of trials with zero values. We therefore averaged across trials within each participant and ran Shapiro-Wilks and Levene tests on the aggregated data, which showed only non-significant deviations from normality and homogeneity of variance. We then carried out two-way RM ANOVA tests for overshoots along D1 and D2 with *CoordSys* and *PresMode* as within-subject factors. If significant main effects were detected, post-hoc comparisons were carried out with Bonferroni correction applied.

Interruptions: Due to a large number of individual trials with zero values, we again carried out within-participant averaging, and the Levene test on the aggregated data revealed non-homogeneous variance across conditions. Therefore,

Metric	Predictor	Coefficient	95% CI	df	F	p-value
Completion Time	CoordSys	0.267	[0.326, 0.209]	1, 433	164.61	<0.001
	PresMode	-0.326	[-0.267, -0.384]	1, 468	223.51	<0.001
	CoordSys * Presmode	0.015	[-0.098, 0.068]	1, 452	0.114	0.723
Trajectory Length	CoordSys	-0.291	[-0.218, -0.364]	1, 349	1.519	0.219
	PresMode	-0.679	[-0.607, -0.752]	1, 407	263.738	<0.001
	CoordSys * Presmode	0.517	[-0.620. -0.414]	1, 380	97.761	<0.001

Table 2. Results of the GLMM analysis of *Completion Time* and *Trajectory Length* presenting model coefficients (logarithmic due to the link function) for each predictor and their interaction along with their 95 % confidence intervals, degrees of freedom (calculated using Satterthwaite's method), F, and *p*-values. Note that *CoordSys* was coded as 1 = Cartesian, 2 = Polar, whilst *PresMode* was coded as 1 = Concurrent, 2 = Sequential, which determined the polarity of the coefficients. Although significant in both cases, the intercept estimate is not shown.

we used a non-parametric approach to compare *Car-Con, Car-Seq, Pol-Con, and Pol-Seq* (we term the factor as *Condition*). We used a Friedman test followed by Bonferroni-corrected Wilcoxon Signed-Rank tests for post-hoc pairwise comparisons.

Cognitive Load: The normalized NASA-TLX scores (collected once per participant per condition) were found to satisfy the assumptions of normality and homogeneous variance, and we therefore performed two-way RM ANOVA tests and Bonferroni-corrected post-hoc comparisons like for *Overshoots*.

Preference Analysis: In order to explore the performance- and experiment-based criteria determining user preference, we checked whether the participants' reported condition of preference matched the condition in which (i) they were fastest (Min Time), (ii) they reported the lowest cognitive load (Min NASA-TLX), (iii) they encountered first in the experimental order (First Condition), and (iv) they encountered last (Last Condition).

5 RESULTS

Overall, the participants were able to find the target area in 99.47% of the 1500 trials, taking an average of 11.55 (SD 6.5) sec per trial across all four conditions. They gave the AG mean pleasantness and helpfulness ratings of 6.81/10 and 8.39/10 respectively. In terms of overall preference there was no clear consensus; six participants (40%) preferred the *Car-Seq* condition, and the remainder were evenly divided among the other three conditions. The preference analysis revealed that nine participants preferred the condition in which their NASA-TLX score was lowest, seven preferred the last condition they encountered, six preferred the one in which they took minimum time (which also coincided with the lowest NASA-TLX score in all instances), and one preferred the first condition they encountered. The preferences of two participants did not match any of the criteria.

The results of the GLMM analyses of normalized completion time and trajectory length are shown in Table 2. For completion time, there were significant fixed effects of both *Coordsys* and *PresMode*. As seen in Fig. 3, participants were significantly slower when (i) using the polar system than the Cartesian for both presentation modes, and (ii) using concurrent than sequential AG for both coordinate systems. There was no significant interaction between the predictors ($p = 0.736$). For trajectory length, there was a significant fixed effect of *CoordSys* but not *PresMode*, but there was a strong interaction between the factors as shown in Fig. 3. Depending on the presentation mode (concurrent/sequential), participants traversed significantly longer or shorter paths respectively in the polar system as compared to the Cartesian.

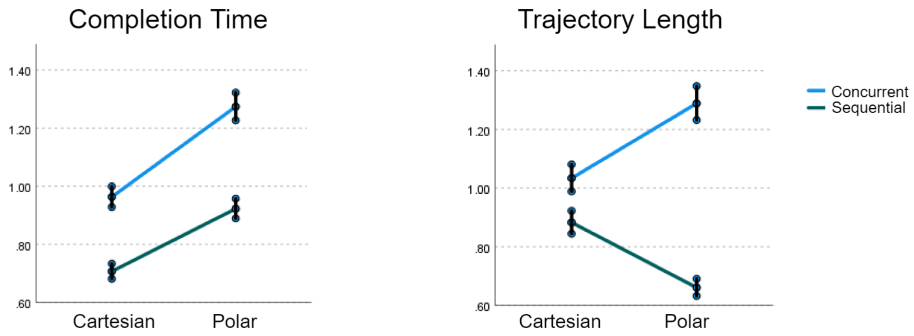


Fig. 3. Interaction plots for *Completion Time* and *Trajectory Length* that illustrate the model-estimated means for each coordinate system and presentation mode. The error bars indicate 95% confidence intervals. A value of 1.0 represents the participant grand mean value. Both outcomes are significantly lower for the *sequential* presentation mode, and trajectory length shows a significant interaction between the factors.

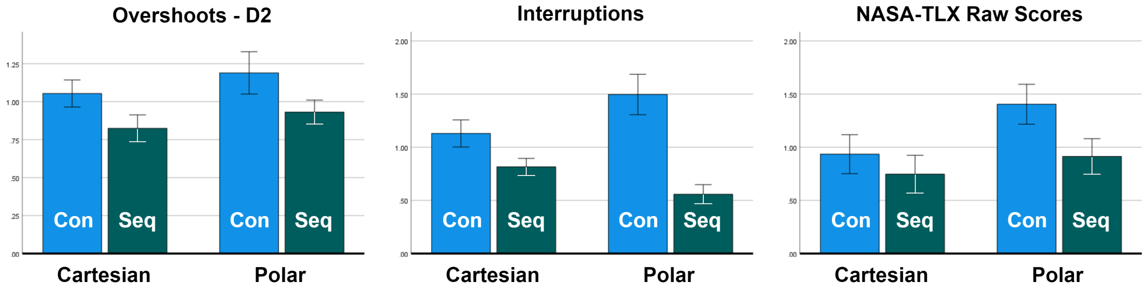


Fig. 4. Bar plots for the participant-aggregated results that showed significant main effects across the coordinate systems and presentation modes. Bar heights indicate means, and the error bars show 95% confidence intervals.

Moving on to the analysis of within-participant aggregated metrics, there was a significant main effect of *PresMode* on *Overshoots (D2)* (RM ANOVA, $F_{(1,14)} = 26.391$, $p < 0.001$, $\eta_p^2 = 0.653$) but not of *CoordSys* (RM ANOVA, $F_{(1,14)} = 4.045$, $p = 0.064$, $\eta_p^2 = 0.224$), and post-hoc tests showed that sequential AG elicited significantly fewer overshoots along D2 than concurrent AG irrespective of coordinate system (see Fig. 4 (left)). For *Overshoots (D1)*, however, there was no main effect of *CoordSys* (RM ANOVA, $F_{(1,14)} = 0.334$, $p = 0.572$, $\eta_p^2 = 0.023$) or *PresMode* (RM ANOVA, $F_{(1,14)} = 4.444$, $p = 0.054$, $\eta_p^2 = 0.241$). For *interruptions*, there was a significant main effect of *Condition* (Friedman test, $\chi^2(3) = 35.08$, $p < 0.001$). Post-hoc tests showed, as seen in Fig. 4 (middle), that *Con* conditions led to more interruptions than their *Seq* counterparts for both coordinate systems. The coordinate system alone also appears to have played a role in the sequential conditions, as *Pol-Seq* elicited significantly fewer interruptions than *Car-Seq*. Moving on to the self-reported cognitive load (NASA-TLX raw scores), there was a significant main effect of *CoordSys* (RM ANOVA, $F_{(1,14)} = 11.315$, $p = 0.005$, $\eta_p^2 = 0.447$) and *PresMode* (RM ANOVA, $F_{(1,14)} = 19.59$, $p = 0.001$, $\eta_p^2 = 0.583$). Pairwise post-hoc comparisons found that the reported cognitive load was significantly higher for the polar coordinate system than the Cartesian across presentation modes, and significantly lower for sequential presentation than concurrent across coordinate systems (see Fig. 4 (right)).

6 DISCUSSION

6.1 Sequential AG Leads to Superior Task Performance

In this study, we experimentally assessed the effects of information concurrency and coordinate system on user performance and cognitive load in a computer-based 2-D navigation task. Our results strongly support our hypothesis that users would perform better and experience a lower cognitive load when presented with sequential AG as compared to concurrent **(H1)**. The participants performed faster (shorter completion times) and more efficiently (shorter trajectories) in *Seq* conditions irrespective of the coordinate system. They also exhibited considerably fewer interruptions, indicating that they had fewer instances per trial of needing to stop moving and make sense of the AG. There were also fewer overshoots along the second dimension D2, which can be explained by the participants being better able to hear subtle FM/AM changes near the target D2 coordinate when the melodic pitch was constant. Lastly, participants expressed a lower cognitive load for *Pol-Seq* than *Pol-Con*, while there were no differences for the Cartesian system. These results suggest that by breaking down the 2-D task into two 1-D tasks, the sequential AG was successful at minimizing the impact of perceptual interactions [29] and promoting a consistently reproducible navigation strategy [18] by encouraging users to focus on one dimension at a time. Also, locating the target coordinate along one dimension led to a perceptually salient change in the sound as the sonification of the next dimension was suddenly triggered. Hence, this intermediate point would have served as a clearly defined waypoint. The waypoint approach is a proven and effective AG method in locomotion guidance for aiding decision-making during locomotion along complex paths [37, 39, 40], and our findings are in line with this. Sequential guidance minimizes perceptual interaction [29] between the sonified dimensions, and this advantage is especially relevant to greater-dimensional navigation tasks where it is difficult to find several perceptually separable sound properties to represent each dimension [42].

6.2 Polar Navigation Behavior Depends on Information Concurrency and Control Schemes

The hypothesis that users would traverse shorter trajectories when using the polar system than the Cartesian **(H2)** was only partially validated by our results. Whilst the participant trajectories were shortest in the *Pol-Seq* condition, they were interestingly longest in *Pol-Con*. It is clear that in *Pol-Seq*, the AG forced participants to first carefully adjust their bearing to the correct value and then adjust the range, which led to them traversing the shortest possible path from origin to target. Had the navigation task been from a first-person perspective (e.g. locomotion) with the coordinate system centered around the user, the approach of adjusting bearing before range would have intuitively made sense (the behavior seen in *Pol-Seq*. In our experiment, the control scheme made it practical for them to adjust range first. During *Pol-Con*, the perceptual interaction between D1 and D2 AG could possibly have obscured the scale melody, prompting users to first adjust the range (D2), and then the bearing (D1). This would have led them to traverse relatively long radial and circumferential paths to the target as opposed to the direct radial path during *Pol-Seq*. The authors of [41] (whose 2-D AG design we took close inspiration from) reported the strategy of adjusting D2 before D1 (modulation-based AG before pitch-based AG) as being the most common user strategy. They estimated that this was because the point of transition from AM (beating) to FM (roughness) was perceptually salient [41]. This reinforces the value of a salient auditory change to signal navigation waypoints, something that the sequential AG scheme integrated by design. One can argue that we could have reduced the perceptual interactions between D1 and D2 AG in the concurrent conditions by attenuating the AM and FM effects, but this would have been detrimental to the perceptual salience of the D2 AG, making the underlying information harder to comprehend. Thus, we assess that for concurrent AG, there is a clear tradeoff between perceptual salience along any one dimension and perceptual interactions between dimensions - this does not exist for sequential AG.

In terms of *CoordSys* effects, we did not anticipate that participants would take longer with the polar system than the Cartesian irrespective of presentation mode. A possible reason for this is that the keyboard-based control scheme for the polar system was not as straightforward as for the Cartesian. While the latter featured a relatively straightforward mapping between the arrow keys and the four cardinal directions, the polar scheme required some mental operations, which may have increased cognitive load and hindered spatial orientation [41], especially when D1 and D2 were concurrently sonified. This notion is corroborated by the higher NASA-TLX scores for polar than Cartesian conditions (see Fig. 4). Future studies should perform the comparison with a controller (e.g. joystick-based) that facilitates more direct input mappings to polar parameters. Another contributor to the result may have been the high degree of angular precision required to correctly pinpoint the target bearing. Angular measurement errors are the main source of uncertainty in polar-coordinate measurement systems [13], and the participants may have spent a relatively long time fine-tuning the bearing to ensure that it fell within the angular bounds subtended by the target area.

Lastly, it is odd that there were no differences between conditions in terms of Overshoots (D1). This may have been because the D1 guidance provided both directional pitch cues (ascending/descending melody) and non-directional tempo cues (note transition rate), causing participants to adopt different navigation strategies along D1, resulting in greater variability in the results and masking any systematic differences between conditions. Also, it has been shown that task instructions have a strong effect on user strategies and trajectories in 1-D tasks [31]. We only instructed participants in relation to speed, and future studies can perhaps include an experimental block where users are asked to minimize overshoots, allowing for a more standardized comparison across AG conditions.

6.3 User Preferences are Possibly Related to Cognitive Load and Familiarity

The analysis of user preferences yielded some interesting insights. Although *Car-Seq* was preferred by a small majority of the participants, our sample was small and this result should therefore be interpreted with caution. Participants preferred the condition where they experienced minimum cognitive load (9 of 15), which in some cases (7 of 15) coincided with the last condition they encountered, and most participants were accounted for by some combination of these criteria (12 of 15). This suggests that irrespective of the condition order (which was counterbalanced), an increasing comfort level with the task over time impacted the participants' cognitive load and performance [35], causing them to prefer conditions they encountered later on.

6.4 Are These Results Generalisable?

The AG design we tested incorporated a very specific set of auditory perceptual parameters that were previously proven to be effective in 2-D navigation [41]. AG designs have been very diverse in terms of their manipulated auditory properties, including timbre, rhythm, harmonicity, synchronicity, physical modelling parameters, and other more complex modulations [1, 4, 12, 31, 33]. Nevertheless, we argue that our finding related to sequential v/s concurrent AG presentation is generalizable to other designs as well. First of all, the tendency of users to navigate axis by axis has been observed not only when using visual displays [8] but also when using very different and unrelated 2-D AG designs [6, 41]. Sequential AG facilitates this approach by minimizing the effect of perceptual interactions [29] and providing salient waypoints, and is therefore likely to enable superior performance when applied to other types of 2-D and 3-D AG designs too. However, there are also certain drawbacks. Because only one dimension is sonified at a time, each point in space does not have a unique auditory representation, and sequential AG cannot be used in applications where multiple spatial variables must be monitored concurrently (e.g. static posture guidance [7] or aircraft orientation [34] where two axes need to be monitored at once - pitch, roll). Here, the only option is a concurrent AG scheme that minimizes the detrimental effects of perceptual interactions.

6.5 Limitations and Future Work

A limitation of our experimental protocol lay in its ecological validity; other studies evaluated their AG paradigms in the context of performed or simulated real-world tasks [1, 4, 6, 9, 27, 33, 34], so it is still unclear how well our findings hold up when the task involves a more complex perceptual-motor component (e.g. movement rehabilitation [15, 24]). In the future, we plan to test sequential Cartesian guidance in the training of the sit-to-stand transition, which involves a clear horizontal component (leaning forward) followed by a vertical one (rising). This will build on past work on providing auditory feedback on this movement [21]. An important limitation of our polar system design was that the bearing was always computed from the perspective of the center of the space. While such scheme would work for an application such as dynamic balance training [12, 22] where the center represents an equilibrium position, it would more realistically suit other applications (e.g. needle placement [4]) if the bearing were computed relative to the instantaneous user position. Our participant sample was relatively small, gender-skewed, and included participants with considerable music training (avg. 6.47 years). The participant who withdrew from the experiment was not replaced, and hence counterbalancing of condition order was not fully achieved. The experiment may have arguably also been more ecologically valid if the interface did not display elapsed time or any form of visual feedback on user position, but this is a relatively minor concern as these aspects did not change across experimental conditions. Although we normalized our performance metrics to highlight the differences between AG conditions and minimize the impact of individual differences, it is still necessary to validate our findings on musically untrained participants, most preferably target users (e.g. visually/physiologically impaired people). It would also be of interest to perform a comparison between sequential AG and visual guidance in multidimensional navigation tasks. A limitation of our analysis is that we did not examine the individual trial trajectories to ascertain and quantify the incidence of different navigation strategies, but this will be part of a future study. Overall, we believe that our key findings related to the effects of information concurrency and coordinate systems are worthy of further investigation.

7 CONCLUSIONS

We experimentally compared concurrent and sequential 2-D auditory guidance based on the Cartesian and polar coordinate systems in a computer-based target-finding task, finding sequential guidance to be superior in terms of task performance with reduced cognitive burden as compared to concurrent guidance. We also found that users were generally slower when using the polar coordinate system, but traversed the shortest trajectories with sequential polar guidance. We believe that these findings can be generalized to higher dimensional tasks and different guidance designs, although future work should include comparisons with visual guidance, aside from evaluating sequential guidance during real-world tasks related to navigation, orientation, and movement rehabilitation with the respective target user groups. We believe that the evidence we provide can contribute to the development of more potent auditory guidance systems in the future.

ACKNOWLEDGMENTS

Author PRK was responsible for the study design, system development, and conducting the experiment and data analysis. Authors SD and EGS supervised the project and assisted with the data analysis. All authors contributed to the writing process and approved of the final manuscript. We would like to thank the participants who took part in the experiment, as well as Janhvi Sampat for helping with the pilot tests. Lastly, we would like to express our gratitude to Henrik Knoche and Milo Marsfeldt Skovfoged for their valuable advice on data analysis.

REFERENCES

[1] Dragan Ahmetovic, Uran Oh, Sergio Mascetti, and Chieko Asakawa. 2018. Turn right: Analysis of rotation errors in turn-by-turn navigation for individuals with visual impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 333–339.

[2] Nuray M Aykin and Turgut Aykin. 1991. Individual differences in human-computer interaction. *Computers & industrial engineering* 20, 3 (1991), 373–379.

[3] David Black, Christian Hansen, Arya Nabavi, Ron Kikinis, and Horst Hahn. 2017. A survey of auditory display in image-guided interventions. *International journal of computer assisted radiology and surgery* 12, 10 (2017), 1665–1676.

[4] David Black, Julian Hettig, Maria Luz, Christian Hansen, Ron Kikinis, and Horst Hahn. 2017. Auditory feedback to support image-guided medical needle placement. *International journal of computer assisted radiology and surgery* 12, 9 (2017), 1655–1663.

[5] Hans-Peter Bruckner, Christopher Bartels, and Holger Blume. 2011. PC-Based Real Time Sonification of Human Motion Captured by Inertial Sensors. In *Proceedings of the 17th International Conference on Auditory Display (ICAD-2011), June 20-24, 2011, Budapest, Hungary*.

[6] SeungA Chung, Kyungyeon Lee, and Uran Oh. 2020. Investigating three-dimensional directional guidance with nonvisual feedback for target pointing task. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 206–210.

[7] Giovanni Costantini, Daniele Casali, Fabio Paolizzo, Marco Alessandrini, Alessandro Micarelli, Andrea Viziano, and Giovanni Saggio. 2018. Towards the Enhancement of Body Standing Balance Recovery by Means of a Wireless Audio-Biofeedback System. *Medical Engineering & Physics* 54 (2018), 74–81.

[8] Rudy P Darken and John L Sibert. 1993. A toolset for navigation in virtual environments. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*. 157–165.

[9] Felix Dollack, Monica Perusquía-Hernández, Hideki Kadone, and Kenji Suzuki. 2019. Auditory Locomotion Guidance System For Spatial Localization. In *2019 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. IEEE, 1–5.

[10] John H Flowers. 2005. Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. In *Proceedings of ICAD'05-Eleventh Meeting of the International Conference on Auditory Display*.

[11] Euan Freeman, Graham Wilson, Dong-Bach Vo, Alex Ng, Ioannis Politis, and Stephen Brewster. 2017. *Multimodal Feedback in HCI: Haptics, Non-Speech Audio, and Their Applications*. Association for Computing Machinery and Morgan Claypool, 277–317. https://doi-org.zorac.aub.aau.dk/10.1145/3015783.3015792

[12] Dominik Fuchs, Martin Knauer, Marion Egger, and Petra Friedrich. 2020. Audio Feedback for Device-Supported Balance Training: Parameter Mapping and Influencing Factors. In *Acoustics*, Vol. 2. Multidisciplinary Digital Publishing Institute, 650–665.

[13] W Gao, Seung-Woo Kim, H Bosse, H Haitjema, YL Chen, XD Lu, W Knapp, A Weckenmann, WT Estler, and H Kunzmann. 2015. Measurement technologies for precision positioning. *CIRP Annals* 64, 2 (2015), 773–796.

[14] Joachim Gossmann. 2012. A perspective on the limited potential for simultaneity in auditory display. In *Proceedings of the 18th International Conference on Auditory Display, Atlanta, GA, USA, June 18-21, 2012*.

[15] Joao Guerra, Lee Smith, Domenico Vicinanza, Brendon Stubbs, Nicola Veronese, and G Williams. 2020. The Use of Sonification for Physiotherapy in Human Movement Tasks: A Scoping Review. *Science & Sports* 35, 3 (2020), 119–129.

[16] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[17] Florian Heller and Johannes Schöning. 2018. Navigatone: Seamlessly embedding navigation cues in mobile music listening. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.

[18] Alycia M Hund and Jennifer L Minarik. 2006. Getting from here to there: Spatial anxiety, wayfinding strategies, direction type, and wayfinding efficiency. *Spatial cognition and computation* 6, 3 (2006), 179–201.

[19] Andy Hunt and Thomas Hermann. 2011. Interactive Sonification. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Verlag Berlin.

[20] Emil Jovanov, Kristen Wegner, Vlada Radivojevic, Dusan Starcevic, Martin S Quinn, and Daniel B. Karron. 1999. Tactical audio and acoustic rendering in biomedical applications. *IEEE Transactions on Information Technology in Biomedicine* 3, 2 (1999), 109–118.

[21] Prithvi Kantan, Erika G Spaich, and Sofia Dahl. 2022. An Embodied Sonification Model for Sit-to-Stand Transfers. *Frontiers in psychology* 13 (2022), 806861–806861.

[22] Prithvi Kantan, Erika G Spaich, and Sofia Dahl. 2022. A Technical Framework for Musical Biofeedback in Stroke Rehabilitation. *IEEE Transactions on Human-Machine Systems* (2022).

[23] Gregory Kramer et al. 1999. The sonification report: Status of the field and research agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. (01 1999).

[24] Dagmar Linnhoff, Shahab Alizadeh, Nina Schaffert, and Klaus Mattes. 2020. Use of Acoustic Feedback to Change Gait Patterns: Implementation and Transfer to Motor Learning Theory—-A Scoping Review. *Journal of Motor Learning and Development* 1, aop (2020), 1–21.

[25] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. 1998. Navigation system for the blind: Auditory display modes and guidance. *Presence* 7, 2 (1998), 193–203.

[26] Shachar Maidenbaum, Sami Abboud, and Amir Amedi. 2014. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews* 41 (2014), 3–15.

[27] Keenan R May, Briana Sobel, Jeff Wilson, and Bruce N Walker. 2019. Auditory displays to facilitate object targeting in 3D space. In *The 25th International Conference on Auditory Display (ICAD 2019), 23–27 June 2019, Northumbria University*.

[28] Angélique Montuwy, Béatrice Cahour, and Aurélie Dommes. 2017. Visual, auditory and haptic navigation feedbacks among older pedestrians. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–8.

[29] John Neuhoff. 2011. Perception, Cognition and Action in Auditory Displays. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Verlag Berlin.

[30] John G. Neuhoff. 2019. Is Sonification Doomed to Fail?. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*. Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, 327–330. https://doi.org/10.21785/icad2019.069

[31] Gaëtan Parseihian, Charles Gondre, Mitsuko Aramaki, Sølvi Ystad, and Richard Kronland-Martinet. 2016. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia* 18, 4 (2016), 674–686.

[32] Gaëtan Parseihian, Sølvi Ystad, Mitsuko Aramaki, and Richard Kronland-Martinet. 2015. The process of sonification design for guidance tasks. *J Mob Med* 9, 2 (2015).

[33] Hessam Roodaki, Navid Navab, Abouzar Eslami, Christopher Stapleton, and Nassir Navab. 2017. Sonifeye: Sonification of Visual Information Using Physical Modeling Sound Synthesis. *IEEE Transactions on Visualization and Computer Graphics* 23, 11 (2017), 2366–2371.

[34] Brian D Simpson, Douglas S Brungart, Ronald C Dallman, Richard J Yasky, and Griffin D Romigh. 2008. Flying by ear: Blind flight with a music-based artificial horizon. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 52. SAGE Publications Sage CA: Los Angeles, CA, 6–10.

[35] Daniel R Smith and Bruce N Walker. 2005. Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 19, 8 (2005), 1065–1087.

[36] Alexandra Supper. 2015. Sound Information: Sonification in the Age of Complex Data and Digital Audio. *Information & Culture* 50, 4 (2015), 441–464.

[37] Antonio Vasilijevic, Kristian Jambrosic, and Zoran Vukic. 2018. Teleoperated path following and trajectory tracking of unmanned vehicles using spatial auditory guidance system. *Applied acoustics* 129 (2018), 72–85.

[38] Bruce N Walker and Jeffrey Lindsay. 2006. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human factors* 48, 2 (2006), 265–278.

[39] Jeff Wilson, Bruce N Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. Swan: System for wearable audio navigation. In *2007 11th IEEE international symposium on wearable computers*. IEEE, 91–98.

[40] Xiaoping Zhou, Qingsheng Xie, Maozu Guo, Jichao Zhao, and Jia Wang. 2020. Accurate and efficient indoor pathfinding based on building information modeling data. *IEEE Transactions on Industrial Informatics* 16, 12 (2020), 7459–7468.

[41] Tim Ziemer and Holger Schultheis. 2019. Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *Journal on Multimodal User Interfaces* 13, 3 (2019), 205–218.

[42] Tim Ziemer and Holger Schultheis. 2019. Three orthogonal dimensions for psychoacoustic sonification. *arXiv preprint arXiv:1912.00766* (2019).