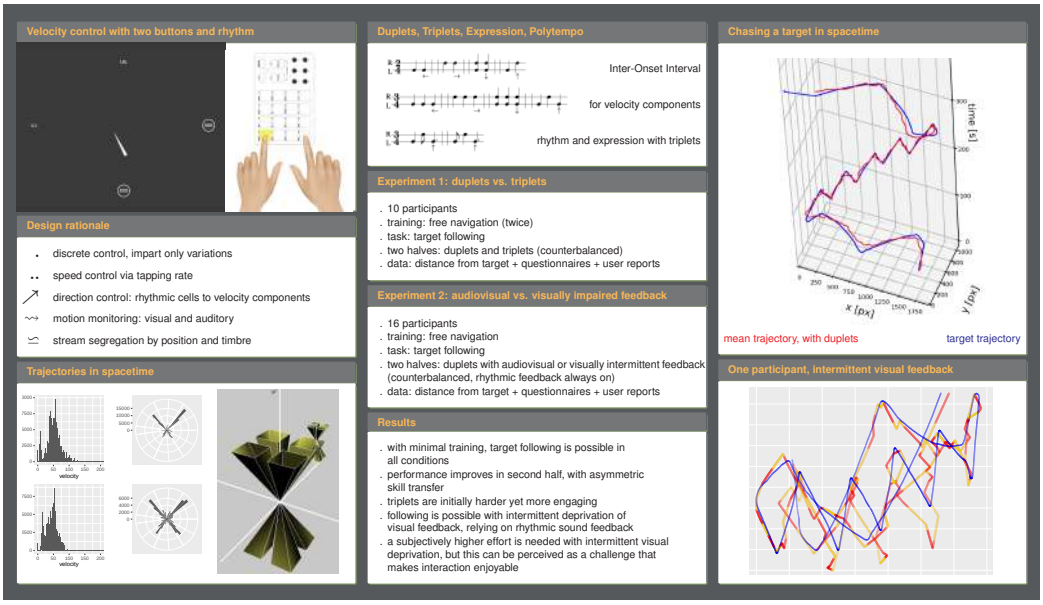
Graphical Abstract

Spacetime trajectories as overlapping rhythms

Davide Rocchesso, Alessio Bellino, Gabriele Ferrara, Antonino Perez



Highlights

Spacetime trajectories as overlapping rhythms

Davide Rocchesso, Alessio Bellino, Gabriele Ferrara, Antonino Perez

- Speed and direction of a moving object can be controlled by discrete short tapping sequences
- Speed and direction of a moving object can be monitored through overlapping rhythmic streams
- Non-visual rhythmic feedback can compensate for temporary deprivation of visual feedback in object motion control

Spacetime trajectories as overlapping rhythms

Davide Rocchesso[a,*], Alessio Bellino[b], Gabriele Ferrara[c], Antonino Perez[c]

[a] *Università degli Studi di Milano, Milan, Italy*
[b] *Pontificia Universidad Católica de Chile, Santiago, Chile*
[c] *Università degli Studi di Palermo, Palermo, Italy*

Abstract

The navigation of two-dimensional spaces by rhythmic patterns on two buttons is investigated. It is shown how direction and speed of a moving object can be controlled with discrete commands consisting of duplets or triplets of taps, whose rate is proportional to one of two orthogonal velocity components. The imparted commands generate polyrhythms and polytempi that can be used to monitor the object movement by perceptual streaming. Tacking back and forth must be used to make progress along certain directions, similarly to sailing a boat upwind. The proposed rhythmic velocity-control technique is tested with a target-following task. Users effectively learn the tapping control actions, and they can keep a relatively small distance from a moving target. They can potentially rely on overlapping auditory rhythmic streams to compensate for temporary deprivation of visual position of the controlled object. The interface is minimal and symmetric, and can be adapted to different sensing and display devices, exploiting the symmetry of the human body and the ability to follow two concurrent rhythmic streams.

Keywords:
Rhythmic interaction, Bimanual movement control, Sonic interaction design, Multisensory motion feedback

⋆A preliminary and partial version of this article appeared in the proceedings of the Sound and Music Computing conference, Stockholm, Sweden, 2023 [1]. The content has been largely extended and the reported experiment 2 is unpublished.

*Corresponding author
Email addresses: davide.rocchesso@unimi.it (Davide Rocchesso), abellino@uc.cl (Alessio Bellino)

1. Introduction

Rhythm and movement in space are tightly connected, as animal locomotion is almost invariantly rhythmical. We can deduce several features of a walking person from perceived patterns of footsteps [2], including the speed of that person in the environment. Similarly, we have the intuitive feeling about how fast a horse is moving, by the pace and rhythm of gait patterns, that differ for walking, trotting, or galloping. Inanimate clockwork mechanisms are also rhythmical and often associated with motion and locomotion, as the pace of ticking is proportional to the resulting velocity. When there is a small number of distinguishable animate or inanimate agents producing rhythmic streams, we can separate them perceptually, and selectively direct our attention to one of them.

Given the deep and tight relation between trajectories of motion in space and rhythmic patterns, one might expect that rhythms have been exploited for controlling as well as for monitoring moving objects. Rhythm-based control of direction and speed of motion, besides its mechanistic purpose, would also allow to differentiate different kinds of motion by articulation (e.g., galloping or trotting) as well as by expressive content (e.g., aggressive or relaxed). Looking at the literature of rhythmicity for interaction (see section 2) the navigation of spaces by generation and adjustment of rhythmic patterns seems to be largely unexplored, despite the control of virtual objects by a few buttons that was ubiquitous in classic arcade games.

In games such as Pong, two buttons are indeed used for positional control of an object on screen, although along a single dimension. For moving along a planar path, as in Pac-Man, the classic choice is between four buttons (as in the directional pad, or D-Pad [3]) and a rate-controlled pointing stick [4]. While rhythmic actions would definitely be possible on a four-buttons controller, in particular to control the directional components of velocity, we are interested in control minimality, with one or two points of action. In particular, two buttons or two sensors would be naturally associated to rhythmic actions as found in walking of bipeds, and would be convenient in a wide range of applications where a human can produce rhythmic patterns via two hands or feet, two fingers, or two controllable symmetric parts of the body. As a drawback, reducing the number of buttons implies complicating the encoding of rhythmic patterns, hence increasing the cognitive load of players.

We propose a technique to move an object over a two-dimensional surface by bimanual tapping of a left button and a right button, where specific rhythmic patterns are used to impart leftward, rightward, upward, and downward components

of motion, and the velocity magnitude is controlled by the tapping rate. In particular, rhythmic cells of two or three taps, respectively called duplets or triplets, can be used to move the object. Tapping a duplet or a triplet sets a new direction and magnitude for a component of velocity along one of the main axes. The commands are discrete, as only one of the two orthogonal components of velocity can be changed at a given time. When the imparted rhythmic cells are iterated and fed back as acoustic stimuli, two rhythmic streams are generated, each representing movement along one of the orthogonal axes.

In the proposed velocity-vector control technique, the magnitude of each velocity component is set as inversely proportional to the temporal interval between taps. For hand tapping, producing audible ticking, rhythmic cells can be reliably produced within a range between a few tenths of a second to a few seconds [5]. Therefore, the ratio of vertical and horizontal tapping rates can not be made too large or too small. As a result, it is essentially impossible to move the object along the main axes, and speed is limited in its magnitude and possible directions. The discrete commands imparted to adjust orientation and speed, and the fact that not all directions are feasible, make it necessary to advance by zig-zaging or tacking[1]. The resulting trajectories are indeed similar to those of a sailing boat. In two verbs, the object speed and trajectory are controlled by ticking and tacking, or TickTacking [1].

While the adjustment commands are discrete, the object moves continuously on the plane and can be auditorily monitored through repeated playback of the rhythmic cells of the two orthogonal components of velocity. The horizontal and vertical components of velocity can be heard as overlapping politemporal rhythmic patterns. The resulting rhythmical flow can be perceptually decomposed into its constituent orthogonal components by auditory streaming. The two streams can be segregated if they are made sufficiently distinct, for example by timbre, brightness, or spatial location [6]. Similarly, when observing an object moving on the plane, the two orthogonal components of its velocity can be determined by visual perception, although speed and orientation are more easily separated perceptually [7]. Just as motion components can be visually found by projection and segregation of velocity along two (or three) orthogonal directions, we can use two (or three) concurrent acoustic (i.e., rhythmic) streams to represent these components. Given our perceptual capability to separate overlapping acoustic streams,

[1]In sailing, tacking means to turn a boat's head into and through the wind. Here, it means to turn the movement direction into and through one of the orthogonal main axes.

we may effectively segregate up to three overlapping streams [8]. Similarly, we may visually represent rates of change in three dimensions or less.

The task of representing and perceptualizing speed by its orthogonal components is relatively easy in two dimensions, where concurrent "orthogonal" rhythms can be used and effectively separated by the listener. Integral listening may also be possible, as different directions would determine different politemporal rhythms, that a listener may learn to recognize. Still, the ability to detect changes in one stream does not depend on the tempo of the other stream, temporarily working as a background [9]. We argue that, by perceptual processes of segregation or integration of rhythmic streams, it is possible to guess the speed magnitude, orientation and direction of an object by sound alone, thus making it possible to control its motion in space.

Given a point in spacetime where a command is imparted, all possible directions of arrival lay within pyramidal volumes with an apex at such point, and all possible outgoing directions lay within reversed pyramidal volumes. Fig. 1 (left) depicts space-time with two points of discrete velocity control. Given a vertical section of the space, orthogonal to one of the two spatial axes, the widest fan corresponds to the maximum tapping rate, and it may be constrained by device or by human motor limits. The narrowest fan corresponds to the minimum tapping rate that is allowed by the input device to be recognized as a rhythmic cell.

From the perspective of sonification of trajectories with rhythm, nothing forbids to display directions close to the main spatial axes, as one of the two concurrent tempi can be made to converge to zero. The limits would only be related to the fastest and to the slowest tempi of rhythm perception. In such case, for pure auditory display of velocity, spacetime can be represented as pyramids with apices at points of velocity change, as represented in Fig. 1 (right). As compared to the spatio-temporal cone of special relativity, where the speed of light is the limit, here we have pyramids instead, as the limits apply separately to the two orthogonal spatial axes.

To assess if and how a user with minimal training can effectively go anywhere on the plane by TickTacking, we ran an experiment where participants had to follow the recorded race trajectory of a sailing boat, with the goal of staying as close as possible to the target. We tested the two conditions of control by duplets or by triplets and compared the mean performance in the two cases. We also collected subjective impressions and the responses to a questionnaire, to assess if and how the two control conditions were differently engaging the participants. We then tested if users can effectively exploit sonic rhythmic feedback when visual feedback becomes temporarily unavailable.

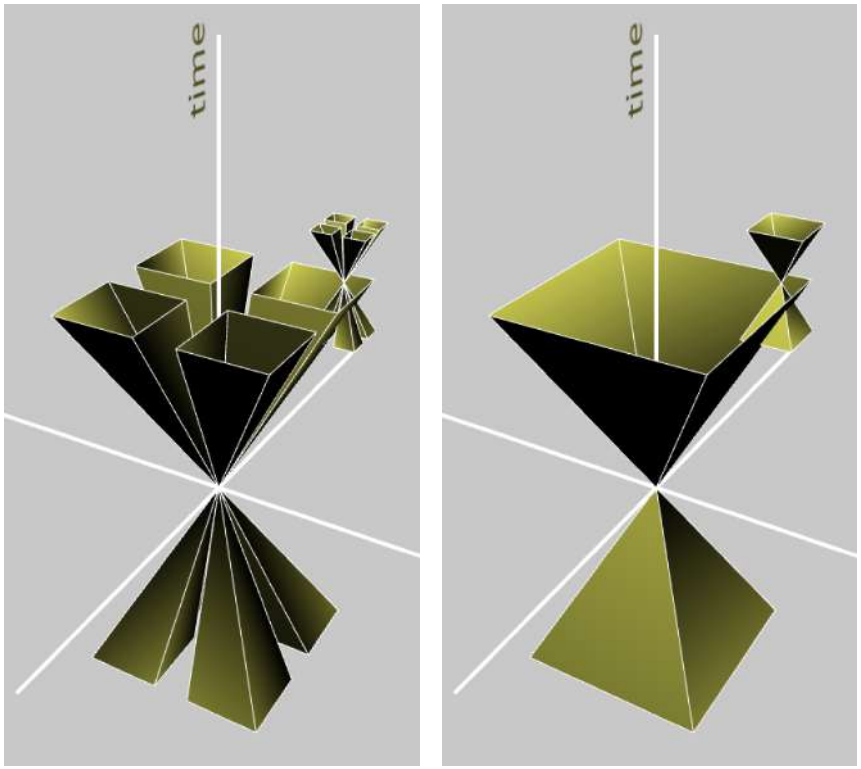


Figure 1: Spacetime of rhythmic velocity control (left) and display (right) with two points of speed-change.

The proposed interaction for velocity control is of interest for applications where the symmetry of the human body can be exploited. The input device can consist of two buttons, keys or batons, as well as two pedals or two myographic sensors. The rhythmic feedback, on the other hand, is not bound to be delivered through audio devices, as it might as well be tactile.

The interaction technique, although being based on human sensorymotor abilities, is not found in everyday activities and it does not belong to the experience of engaging with the physical world. As such, TickTacking may be described as a kind of *non-natural interaction* [10] that has the potential to reflectively engage users.

The task of chasing a moving object with discrete rhythmic control and multisensory feedback poses important challenges to the user, who might have to learn new sensory-motor patterns, with the possibility of developing performative skills and even virtuosity. For this reason, the proposed interaction may open the way

to further studies on human neuro-motor capabilities, including integration and segregation of sensory streams, and learning of sensory-motor tasks.

The article is structured as follows. In section 2, we look at how rhythm and bimanual input have been previously used in interaction design, to draw trajectories in space and time. Section 3 explains the rationale for controlling movement through discrete rhythmic sequences. Experiment 1 is reported in section 4, where sequences of two or three taps to control the orthogonal components of velocity are compared in a target-following task. Section 5 focuses on two-taps sequences and investigates, through experiment 2, how humans can rely on overlapping auditory rhythms to overcome temporary deprivation of visual feedback of the controlled object. Section 6 highlights the informative limits of the two experiments in collecting evidence in support of the effectiveness of multisensory rhythmic interaction for trajectory control. Section 7 concludes the contribution, by summarizing the findings and indicating how the proposed interaction method may be exploited in human-object interaction and in human-human interaction, and how it may be a useful tool to investigate human neuro-motor control and improve motor performance.

2. Interaction Background

The rhythmic interaction with devices or technology-augmented objects has been studied in a wide range of contexts and scales. The amodality of rhythms [11] has produced studies and solutions for one or more of the senses of touch, hearing, vision, and proprioception. An analysis of the rhythms of our cities, as they are experienced in everyday urban lives, was proposed to better understand the relationship between human bodies and the space they inhabit [12]. Urban rhythms have been explored through the use of sound and music, and used as a key design dimension for urban planning [13].

In sonic interaction design, systems and interfaces that support rhythmicity and afford the development of virtuosity have been proposed [14]. The role of rhythm in multisensory continuous interaction has been investigated with design exercises [15], where different kinds of rhythmic feedback have been shown to elicit different behaviors in mundane tasks, such as cutting vegetables in the kitchen. Cutting rhythms are also relevant for cinematic virtual reality, and the kinaesthetic affinity between film editing and rhythmic interaction has been highlighted to focus attention and increase engagement [16]. Rhythmic tutoring has been proposed for interaction by handclapping [17]. Rhythmic patterns, composed of short and long taps and breaks, have been proposed as an input method,

to replace single commands and tested for recall [18]. Rhythmic microgestures have been proposed for non-visual interaction in mobility [19]. Selection by visual rhythmic patterns and motion synchronization has been proposed [20], for environments populated by several interactive objects. Although there are rhythmically challenged persons [21], rhythmicity is generally found useful in joint action. Temporal coordination, body movement and interpersonal interaction can be elicited through designed interfaces that require rhythmic synchronization [22]. The rhythmic propensity of autistic individuals has been given positive value through technology, to foster social interaction [23].

Temporal proximity has been recognized as a unification principle for multiple events that are perceptually grouped to form rhythmic patterns, or gestalts [24]. Rhythm and motion have been extensively investigated for human walking, especially for the purpose of recreating and manipulating the experience of virtual locomotion, as well as to augment walking experiences [2, 25]. Horse gait patterns have been used to augment human locomotion by biking, so that one can get the bike to walk, trot or gallop [26]. Rhythmic structures are emotionally expressive [27], and the similarities between music performance and everyday motor activity have been described [28, 29]. In this respect, the question on how minimal the interface can be for a satisfying musical experience has been addressed, and the single button represents the lower bound of gestural complexity, yet affording expressive interaction [30]. Navigation in two dimensions with velocity control, using a single-button interface, has been investigated [31], where users adjust the controlled-object speed through rhythmic tapping, and its direction by pressing and tilting, with pitch-based auditory feedback.

In this study we propose a rhythmic control of movement by means of two buttons, effectively realizing a rate-control input device [32], where the natural controlled property is velocity, as the tapping rate directly maps to speed, similarly to how the galloping rate is related to horse speed. Although the input device does control movement of an object on a surface, it can not be assimilated to a pointing device [33], as its purpose is to control a velocity vector rather than hitting a target. A target can actually be hit, although the trajectory to reach it would generally be non-rectilinear. Drawing trajectories with two buttons may recall the act of drawing with two knobs, as in Etch-A-Sketch, a classical drawing toy that has recently become a research paradigm to investigate inter-limb and inter-individual coordination [34, 35]. While Etch-A-Sketch is based on continuous manipulation, the proposed interface is based on discrete commands as patterns of discrete taps. While in Etch-a-Sketch the absence of control action implies no motion of the drawing point, in the proposed interface motion is kept at constant

speed and direction between discrete acts of motion adjustment. Since the object keeps its movement between control acts, regardless of absolute positioning, and considering the constraints in spacetime (Fig. 1), the proposed interaction could be described as inertial and relativistic.

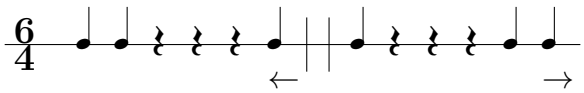
3. Design rationale for rhythmic control of trajectories

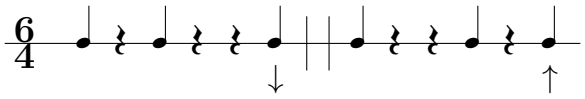
Rhythmic control of movement through buttons, or other kinds of simple sensors, would be desirable in a variety of applications and contexts, such as vehicle driving, interfaces for people with special needs, or entertainment. In the spirit of subtle interaction, or "a way to do less" [36], we should aim at minimizing the number of control points, to make it possible for the motion-control interface to coexist with other input and output devices, and with other activities that may be carried out concurrently. The control would be exerted through tapping rate, that may directly map to speed, similarly to how the ground-hitting rate is related to walking speed. Among the possible ways to control the directional properties of velocity, we propose mapping different rhythmic cells to the velocity components along the four semi-axes, that is the two orthogonal axes, each in positive or negative direction. We want the control to be discrete, that is to impart only variations from Galilean steady motion of the controlled object. Between any couple of imparted commands, it should be possible to monitor the object constant-speed motion through visual motion or auditory (or tactile) rhythmic feedback, while being possibly involved in other activities.

3.1. Rhythms by tapping

If we are interested in controlling a velocity vector by tapping rhythms, a sensible aim is minimal and subtle interaction, and we should try to minimize the number of control points: What is the minimal number of buttons?

In principle, if we design a rhythmic cell for each of the four semi-axes, one button is enough. The mean Inter-Onset-Interval (IOI), or alternatively the time interval between the first and last tap of the sequence, would set the absolute value of velocity. The minimal number of taps per rhythmic cell is three, as we could, for example, count in six and make the following assignment of patterns to directions:

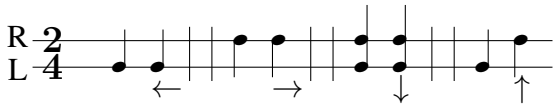




This would require memorization of the sequences, explicit counting, and extensive training.

3.1.1. Duplets on two buttons

With two buttons, physically disposed according to the left-right symmetry of the human body, we have the possibility to tap the left button (L), the right button (R), or both simultaneously (X). We can use rhythmic cells of just two taps, and have a total of 3^2 possible assignments of L, R, and X to the first and second tap of the sequence. Of these 9 possibilities, we choose the rhythmic cells LL, RR, XX, and LR, allowing for RL as well to accomodate for possible left-right inversions:

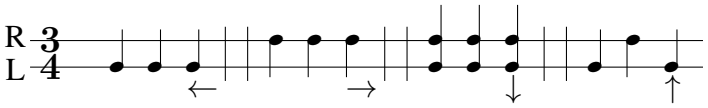


These cells can be mentally represented with reference to the physical layout (left and right buttons) or to physical dynamics (simultaneous or alternate hitting evoking sinking down or bubbling up, respectively). One IOI is sufficient to determine the rate along one of the four directional semi-axes. A consistent sonic output would repeat the horizontal and vertical rhythmic cell, or duplet, with insertion of a pause between each couple of duplets. Two overlapping sequences of duplets would produce two streams that may be segregated by position (L pulse played on the left channel, R pulse played on the right channel) and by timbre (e.g., X pulse represented by a dull sound and LR sequence using a couple of pulses, the second brighter than the first, as in ascending brightness). For effective auditory monitoring of direction and speed, it is critical to choose sounds that make the streams segregate [37, 38, 6], to make the two (horizontal and vertical) components of velocity clearly discernible, so that the user may replicate one of the two concurrent duplets at a higher or lower rate, respectively to increase or decrease one component of velocity.

3.1.2. Triplets on two buttons

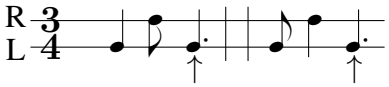
With two buttons, we can move from duplets to triplets, and the L, R, and X assignments to the three taps would give a total of $3^3 = 27$ possible different tapping sequences. Among these, we choose a set of four that seems the most intuitively robust, i.e., LLL for left, RRR for right, XXX for down, and LRL for

up (permitting the left-right inversion RLR):



3.1.3. Expression

Triplets introduce an extra degree of freedom, that is the relative length of the two IOIs, which do not have to be set equal. Two different possible triplets governing the upward vertical component of velocity may be, for example:



The absolute value of the speed component would be given by the sum or the average of the two IOIs, with no apparent change in the resulting motion. However, if the relative timing of the taps is maintained during playback of the rhythm, the user has the possibility to act expressively on the rhythmic display [27], as a range of polyrhythms and polytempi can emerge from the two concurrent streams of triplets. This is a window open to creative abuse of the tool. If the two adjacent IOIs are about equal, the resulting sonic output would be a superposition of two galloping rhythms, that should be segregable by timbre and spatial location. Such rhythmic cells evoke motion [28]. To make the vertical semiaxes clearer, we make the XXX sequence correspond to the repetition of a dull sound and the LRL sequence correspond to a triplet of pulses, in ascending order of brightness. Fig. 2 shows an example of polytempo obtained by superposition of triplets LLL and LRL that represent movement in a left-up direction.

Another possible source of expressiveness, for both duplets and triplets, is the dynamics of taps, or accents. This implies having buttons that can distinguish soft from hard pushes, as it is the case in musical keyboards or keypads that are sensitive to key velocity. This additional dimension does not affect motion but can make perceptual isolation of rhythmic cells easier [5], and the auditory display more engaging and open to expressive action.

3.2. Trajectories

The control of speed components through rhythmic cells is discrete, as each duplet or triplet corresponds to a discrete change of direction and speed of the controlled object. The control action is similar to that of a sailing boat, where

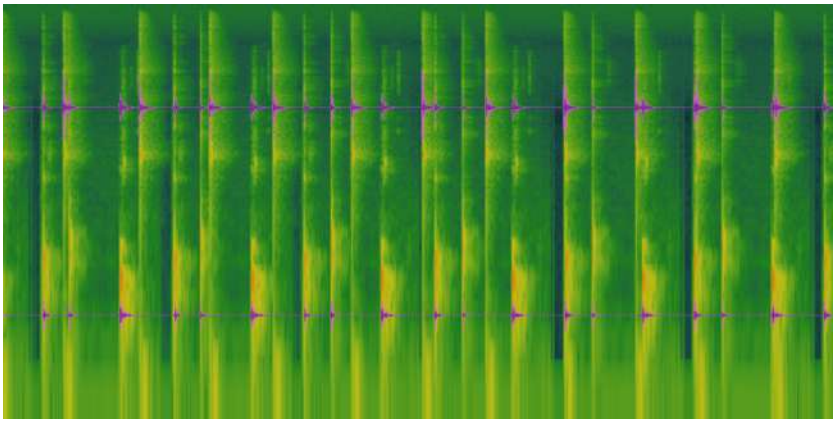


Figure 2: Spectrogram with superimposed stereo waveforms for two overlapping triplets LLL and LRL at different paces, representing motion in a left-up direction. Horizontal linear scale for time, vertical logaritmic scale for frequency.

direction and speed remain almost constant between (almost) discrete turns or adjustments. This inspired us to use traces of sailing regattas as target trajectories in the experiment described in section 4, to test the effectiveness of control. As in sailing, any point on the surface can be reached, although not all directions are actually affordable. In particular, we can not sail straight against the wind, but we effectively go up the wind by a sequence of tacks. Similarly, we can not produce sequences of taps that would move the controlled object exactly along the orthogonal axes, as this would correspond to an infinitely-long or to a zero-approaching IOI. There is a sort of "dead angle" around each of the axes, as represented by the intersteces between the pyramids of Fig. 1, but a sequence of contrasting rhythmic cells may produce zig-zag motion around a semi-axis.

In the experiments we are going to describe in sections 4 and 5, most participants were logged during an explorative phase of free navigation. We logged the coordinates of the controlled object on screen, so to derive velocity components v_x and v_y by discrete differentiation, and to obtain an empirical collection of the distribution of velocity angles ($\arctan \frac{v_y}{v_x}$). This is illustrated in Figs. 3 and 9, where the dead angles are visible as missing bins along the main axes of the polar histograms[2]. The histograms show a tendency, among persons with no prior

[2]Actually, there may be residual directions along the axes that are due to saturation of object position when it reaches the edges of the window. These have been removed from counting in the histograms.

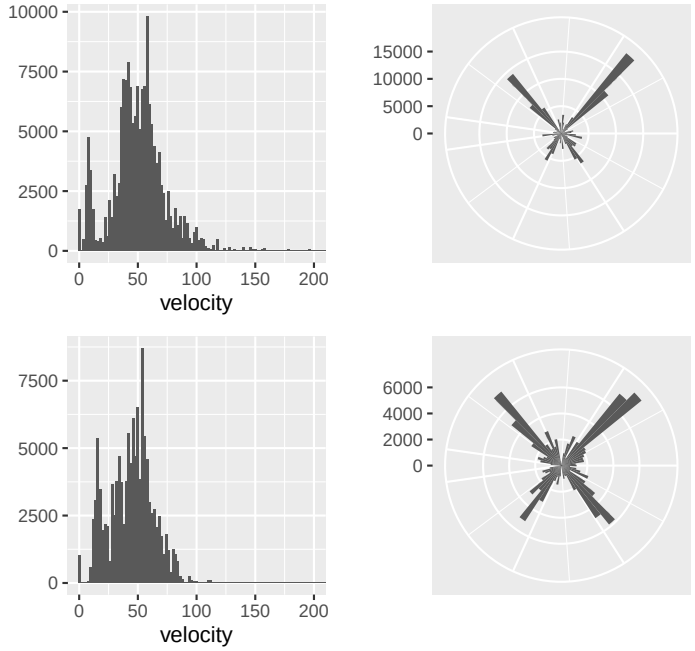


Figure 3: Experimental distribution of velocity magnitudes (in $\mathrm{pixels/s}$) and directions during preliminary practice of participants. Top: two taps, 9 recorded participants; Bottom: three taps, 7 recorded participants.

experience of the interface, to give similar rates to the horizontal and vertical components of velocity, thus preferring movements along the diagonals of the screen window.

3.3. Auditory or tactile rhythms

The rhythmic feedback can be either auditory or tactile. The former can be intimate if delivered through headphones or earbuds, or public if delivered through loudspeakers. Tactile feedback is inherently intimate. The perception of rhythms can be similarly effective with the two senses [39, 40], and the separation between the two domains become blurry if devices based on bone conduction are used.

An application scenario is that of the car, where two sensors and some actuators can be easily applied on the steering wheel, so that the rhythmic cells can be detected and repeatedly reproduced where the action is, for a truly embodied experience [41]. Rhythmic tactons have been proposed and used in HCI [11], also

for the car driving environment [42]. Although frequency and timbre discrimination is much poorer with touch as compared with audition, it should be possible to perceptually segregate concurrent rhythmic streams that are being emitted by a point-like vibrotactile actuator, based on bright/dull timbre differences [43]. The design of the tactile display may compensate for the difficulty of rendering and segregating tactile rhythmic streams, by using a larger number of (four) actuators to differentiate between the components of velocity.

For the scope of this paper, we only consider auditory rhythmic feedback through headphones.

4. Experiment 1

A pilot implementation of the proposed rhythmic interaction was demonstrated at the European Researchers' Night in Palermo on September 30, 2022. Dozens of visitors of our booth tried out the navigation by rhythmic tapping, with headphone- and screen-based audiovisual feedback, and also performed a target-following task. We had the possibility to record some trajectories and to see how the proposed interaction could be learnt and used in reasonable time.

To see, in a controlled setting, if the proposed rhythmic interaction is understandable and effective, and to compare the two input modes of duplets and triplets, then we designed and ran an experiment.

4.1. Objectives

The objectives of the experiment are:

- To measure the performance in following a target moving on the 2-D plane, by an interface with only two buttons, and rhythmic cells of two or three taps, associated to the four coordinate semi-axes, and produced with an inter-onset interval (IOI) inversely proportional to the respective component of velocity;

- To investigate the performative aspect of interaction, in terms of flow and engagement, for the two cases of binary and ternary rhythmic cells.

4.2. Research questions

The following research questions are being addressed experimentally:

RQ1.1 Is it possible to control the movement of an object in the 2-D space by rhythmic cells on two buttons?

RQ1.2 Does the acquired dexterity with rhythm-based movement control improve performance and increase engagement?

RQ1.3 Do more complex rhythmic cells induce more engaging experiences, possibly at the expense of a harder-to-learn interaction?

Research question RQ1.1 is going to be addressed through measurements of performance in a target-following task: A novice user with minimal training should be able to maintain proximity of a controlled object to a moving target. Question RQ1.2 is investigated through a questionnaire and the reported experience, after measuring the performance improvement between the two halves of the experimental session. Question RQ1.3 is investigated by considering three-taps vs. two-taps rhythmic cells, looking for asymmetries in learning, and comparing the measured performance data with user reports.

4.3. Participants

Participants were recruited among students of computer science of the University of Palermo with a call for volunteers. The 10 participants (1 female) reported normal or corrected to normal vision, and normal hearing. Their median age was 22 years, with interquartile range of 4.5 years. Four participants declared some kind of musical practice. The participants were all native Italian-language speakers, and all oral and written interaction with them occurred in Italian.

Given the exploratory nature of the study and practical limitations in recruiting participants and running the experiment, we accept a small sample size. The study will be under-powered yet acceptable in interaction design [44], as the interest is in highlighting an effect that is "grossly perceptible" [45], being comparable to the variability in performance across participants. With 10 participants, and significance level $\alpha = 0.05$, a large effect size (d between 0.8 and 1.0) determines a power ranging between 0.62 and 0.80.

The participants gave their informed consent before the experiment. The experimental protocol was approved by the ethical committee of the University of Palermo.

4.4. Apparatus

A custom audio-visual software was developed in the Processing 4 language and environment, with `themidibus` library and JSyn-based `sound` library. The visual display was a Wacom Cintiq Pro DTH 3220, and the application was run full screen at 1920×1080 pixels, 60 frames per second. The active area of the

screen was $697\,\mathrm{mm} \times 392\,\mathrm{mm}$. The object being controlled on the screen had a comet shape, with a tail that became proportionally longer at higher speed. At the four ends of the semi-axes, short sequences of two or three L, R, or X letters were shown, to help the user recalling the tapping commands that govern the velocity component in the corresponding direction. On the semi-axes delimiting the quadrant of current velocity direction, the letter marks were highlighted with a circle. The left part of Fig. 4 shows a navigation snapshot in the described playground, with the controlled object moving in south-south-east direction with a combination of XXX and RRR triplets.

Two buttons of the ESI Xjam MIDI controller, as highlighted in the right part of Fig 4, with default settings, were used for rhythmic input. Auditory feedback was given through Beyerdynamic DT 770 Pro circumaural headphones driven by a Native Instruments Komplete Audio 6 interface, whose level was set comfortable and constant for all participants. The custom software application was run under Windows 10 with a reported default JSyn audio latency of $80\,\mathrm{ms}$. The sounds for the auditory display were vocal imitations of percussive sounds performed, recorded and edited by the first author. The auditory display was repeatedly playing rhythmic cells corresponding to the velocity components along the two orthogonal axes. Once a duplet or a triplet was acquired it was repeatedly played back to rhythmically display the corresponding velocity component, with a pause of $100\,\mathrm{ms}$ between successive repetitions, introduced to enhance perceptual grouping [5]. Since there are two components of velocity on a surface, two overlapping rhythms were being played during interaction. The key velocity messages sent by the Xjam controller were used to modulate the intensity of the pulses composing the rhythms. In addition to sounds forming the polytemporal texture of the auditory display, a percussive sound, steered to left, right, or left+right channels, was used as immediate (within the latency) feedback of button press, for the left and right button, as well as for simultaneous taps. Being the control based on discrete commands that change the velocity magnitude and direction, with the rhythmic patterns consequently affected, the latency was only perceivable in the feedback of button presses, and did not affect the repeated auditory display of the rhythmic cells.

4.5. Procedure

Each participant was exposed to two versions of the interface, one with duplets and the other with triplets, thus dividing the experimental session into two halves. In each half, the participant was exposed to a short ($3\,\mathrm{min}\ 14\,\mathrm{s}$) video

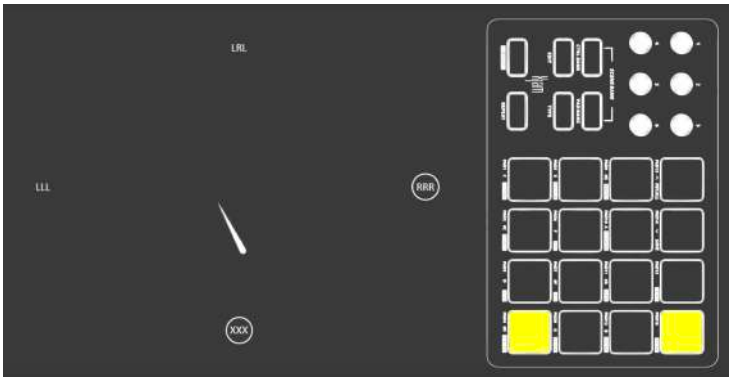


Figure 4: Screenshot of the instruction video. The left part shows a frame of the visual display that the participant would see during interaction, showing downward motion of the controlled object. The right part shows the layout of the Xjam controller, where the two used buttons are highlighted.

specific for duplets[3] or for triplets[4]. The video illustrates the interface, including the controller with the two buttons to be used, and explains how to control the velocity components by tapping. Attention is drawn to the auditory feedback, and an audiovisual example of navigation is given. A screenshot of the instruction video is reported in Fig. 4. In the final part of the video, a target-following task is introduced.

After seeing the video instruction and receiving possible clarifications from the experimenter, the participant was asked to navigate freely on the plane, for about $5\,\mathrm{min}$ by tapping the proposed duplet or triplet rhythmic cells. This free navigation acted as the training phase for that specific rhythmic cell. This free training was preferred to a more constrained training because we observed in the pilot public demonstration that users gradually become familiar with the interface by randomly moving around and experimenting with it. For most training sessions, the object position was logged, so that distributions of the velocity vector could be collected, and they are shown in Fig. 3 for a large subset of participants. It can be observed that, for duplets, participants in free navigation preferred velocity magnitudes around $50\,\mathrm{pixels/s}$. This corresponds, assuming diagonal motion,

[3] https://www.youtube.com/watch?v=ASxdLamllWQ
[4] https://www.youtube.com/watch?v=die9Dz513m8

to the quite short IOI of $170\,\mathrm{ms}$, being the velocity magnitude

$$|v| = k\sqrt{\frac{1}{IOI_x{}^2} + \frac{1}{IOI_y{}^2}}, \tag{1}$$

where the constant k is programmatically set to 100 times the framerate. Very low values of velocity (lower than about $20\,\mathrm{pixels/s}$) may be attributed to missed taps or erroneous input sequences, corresponding to large inter-tap intervals. The system was programmed to restart detecting the first pulse in the rhythmic cell when a tap arrived after more than two seconds from the previously detected one.

After free training, the target-following task was run, as described in section 4.5.1, that lasted slightly less than $6\,\mathrm{min}$. The order of exposure to duplets and triplets was counterbalanced among participants, to mitigate carryover effect.

After both halves of experimentation, the participants were asked to fill two Raw-NASA-TLX questionnaires [46], one for the duplet and one for the triplet interface. The Raw-NASA-TLX questionnaire is aimed at giving a six-fold assessment of perceived workload, along the scales of mental demand, physical demand, temporal demand, performance, effort, and frustration. The paper-and-pencil version with 21 gradations of the rating scales was used, with -10 corresponding to "very low" demand and +10 corresponding to "very high" demand. Moreover, the participants were asked to leave written comments about the learning process, any tactics they may have followed, any sensation of engagement, or any other thoughts they may want to report.

Overall, each participant session lasted about $40\,\mathrm{min}$.

4.5.1. Target following

Given the observed similarity between motion by discrete rhythmic commands and sailing, the trace of a sailing regatta was used as the trajectory of the target to be followed. Namely, the trace of Oracle boat was taken from the America's Cup Final of year 2013[5]. The trace is available as a sequence of 1,710 timestamped observations of (x,y) coordinates that were fit to the screen size and interpolated for smooth display at the chosen frame rate and playback speed. The screen-reconstructed regatta lasted slightly less than six minutes. Fig. 5 shows the target trace as a thin blue line.

[5]https://archive.nytimes.com/www.nytimes.com/interactive/2013/09/25/sports/americas-cup-course.html

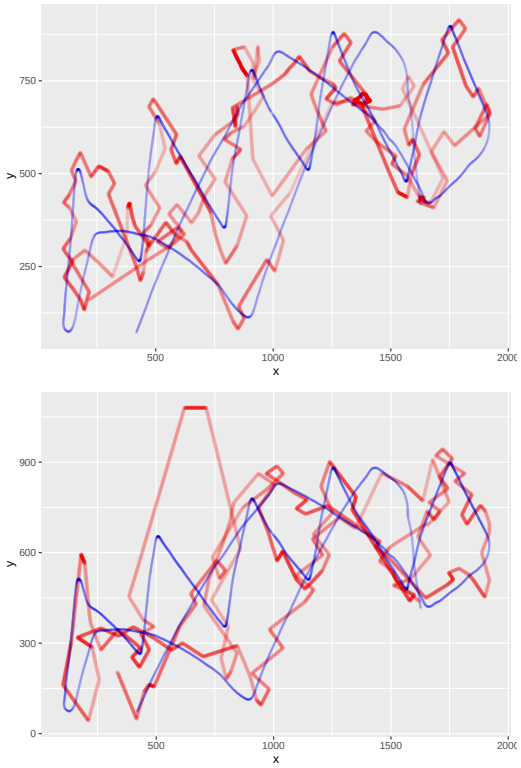


Figure 5: Trajectories of participant number 3. The thin blue line is the target trace. Top: duplets; Bottom: triplets. Transparency of the line is proportional to speed.

Participants were asked to keep the controlled object as close as possible to the moving target.

4.6. Results

4.6.1. Target following

To qualitatively analyze the performance of participants in following a target, we displayed the performed static trace, superimposed to the target trajectory. A quantitative measure was obtained by computing the Euclidean distance of the controlled object from the target, in pixels, at each frame. The istantaneous distances were then averaged over the whole trajectory, to give a mean distance per subject, that is reported in Table 1. The overall mean and standard deviation of the mean distances are also reported in Table 1.

Fig. 5 shows the trajectories of participant number 3, who obtained the smallest mean distance for duplets (exposed first) and the second to smallest distance

Part.	distance/duplets	distance/triplets
1	120.58	100.97
2	161.45	181.19
3	75.33	78.28
4	96.53	184.06
5	97.34	110.62
6	112.67	205.25
7	132.87	103.57
8	75.32	173.79
9	86.56	76.52
10	94.05	182.34
(SD)mean	(27.11)105.27	(49.87)139.66

Table 1: Mean distances (in pixels) from target trajectory, per participant, for duplet and triplet control.

for triplets. Fig. 6 shows the average of all spacetime trajectories of all participants, together with the target trajectory, for duplets-based control.

Given that normality assumptions are fulfilled (Shapiro-Wilk test, with $p > 0.05$), parametric hypothesis testing was used to assess the significance of the difference of the means. The mean distance for duplets was $105.27\,\text{pixels}$ (4.78% of diagonal length), and for triplets it was $139.66\,\text{pixels}$ (6.34% of diagonal length), but the difference of $34.4\,\text{pixels}$ was not significant ($F_{1,9} = 4.469$, $p = 0.064$, $\eta^2 = 0.169$). With the collected data the null hypothesis of equality of distances can not be rejected.

It is also interesting to compare the performances on the first and second halves of the experimental sessions, to see if there has been a learning effect, regardless of the order of exposure to duplets or triplets. Overall, in the first half the mean distance was $143.93\,\text{pixels}$ (6.53% of diagonal length), and in the second half it was $101.00\,\text{pixels}$ (4.58% of diagonal length). Given that normality assumptions are fulfilled (Shapiro-Wilk test, with $p > 0.05$), parametric hypothesis testing was used to assess the significance of the difference of the means. The difference of $42.93\,\text{pixels}$ was large and significant ($F_{1,9} = 9.639$, $p = 0.013$, $\eta^2 = 0.264$).

If an overall dependence of mean distance on the number of taps failed to emerge from the one-way anova, the picture emerges more clearly from a two-way mixed anova, with number of taps as a within-subject factor, and kind of first exposure as a between-subject factor. The normality (Shapiro-Wilk, $p >$

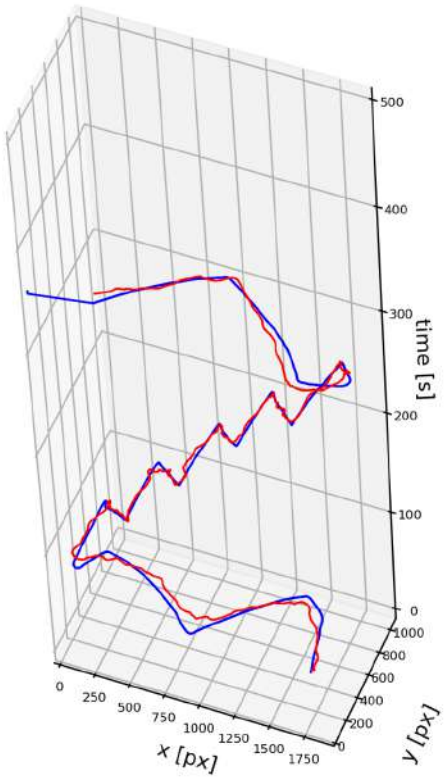


Figure 6: Average of all spacetime trajectories of all participants (red line), together with the target trajectory (blue line), in the case of duplets.

0.5) and homogeneity of variance (Levene, $p > 0.05$) assumptions were fulfilled. In explaining the mean distance from target, there was a statistically significant and large interaction between group of first exposure and number of taps ($F_{1,8} = 27.39$, $p < 0.001, \eta^2 = 0.533$). The simple main effect of group of first exposure was significant for the 3-taps condition ($p < 0.0001$) but not for 2 taps. The simple main effect of number of taps was significant for first exposure to 3 taps ($p < 0.01$) but not for first exposure to 2 taps. In the first half of the experiment, the between-subjects difference of the means was significant and large ($F_{1,8} = 48.51$, $p < 0.001, \eta^2 = 0.86$), but that was not the case for the second half (n.s. difference).

Figure 7 shows the distributions of mean distances for the first and second halves of the experiment, grouped by first exposure. Three-taps interaction pro-

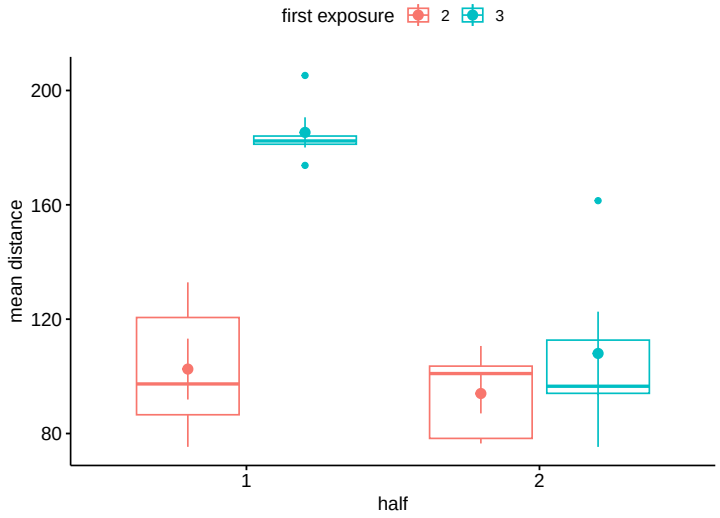


Figure 7: Performances for the first and second half of experiment 1, grouped by kind of first exposure.

duces a significantly worse performance when used in the first half, thus indicating asymmetric skill transfer. In other words, a gentle introduction to rhythmic interaction by TickTacking would better be achieved by using duplets, that proved to be more effective at first exposure, before moving to more complex rhythmic cells.

4.6.2. Questionnaire and report

Participants were asked to rate their mental demand, physical demand, temporal demand, performance, effort and frustration on ordinal scales. We considered each individual scale and refrained from extracting a summary load index. For these reasons [47], non-parametric testing is appropriate to compare the ratings between conditions.

Table 2 reports the median and inter-quartile range of the responses to the six questions of the Raw-NASA-TLX questionnaire. A Wilcoxon signed-rank test ($z = 36$, $p = 0.014$) shows a significant difference between duplets and triplets only for the question on overall satisfaction with performance, and for such question the Wilcoxon effect size is large ($r = 0.845$). The participants were generally more satisfied with their performance for triplets, as a negative value in this scale indicates a higher perceived value of success in performing the task. While the medians of physical and temporal demands are low, the mental demand

and level of effort tend to be moderate. The frustration level is low.

Question	duplets	triplets	p-value
mental demand	2.5(3.5)	3.5(4.0)	n.s.
physical demand	-6.0(4.5)	-6.0(6.0)	n.s.
temporal demand	-7.0(5.5)	-7.0(11.0)	n.s.
performance	1.0(8.25)	-5.5(3.5)	0.014 *
effort	2.0(2.75)	2.5(5.5)	n.s.
frustration	-6.0(4.5)	-5.0(7.5)	n.s.

Table 2: Median (IQR) of the ratings for each of the six questions of the Raw-NASA-TLX questionnaire, after having performed with duplets and with triplets.

The same ratings of the Raw-NASA-TLX questionnaire have been analyzed to check if the subjective task load changed between the first and second half of the sessions. Table 3 reports the median and inter-quartile range of the responses to the six questions, for the two session halves. A Wilcoxon signed-rank test shows a significant difference between the two halves for temporal demand and for frustration ($z = 21$, $p = 0.034$) and for such questions the Wilcoxon effect size is large ($r = 0.758$). The participants generally felt a lower time pressure and lower frustration in the second half, as compared to the first.

Question	first	second	p-value
mental demand	3.0(4.0)	2.5(5.5)	n.s.
physical demand	-6.5(5.75)	-5.0(4.5)	n.s.
temporal demand	-7.0(9.25)	-7.5(7.5)	0.034 *
performance	-3.0(4.0)	-5.0(9.75)	n.s.
effort	2.0(1.75)	0.0(3.75)	n.s.
frustration	-5.0(11.2)	-6.0(4.5)	0.034 *

Table 3: Median(IQR) of the ratings for each of the six questions of the Raw-NASA-TLX questionnaire, after the first and second half of the session.

Reading the comments that were left by participants at the end of the experimental session, a few subjective experiences are worth reporting:

Practice improves performance: Seven participants out of ten reported higher confidence and ease in the second half, regardless of the order of presentation of the number of taps;

It is an engaging game: Half of the participants reported a high level of engagement. Some felt challenged and developed some tactics. Some mentioned flow and concentration;

Three taps are more difficult yet more engaging: One participant reported a greater freedom of movement with three taps, and another one reported a higher difficulty;

The role of sound is not clear: Four participants reported doubts on the usefulness of sound for the target-following task, and two described it as unnerving. Two participants mentioned that they may have been helped by sound;

Control glitches impair the performance: Three participants reported that often the system failed to detect the imparted commands, due to the mechanical compliance of buttons or to misalignment of the rhythmic cells.

4.7. Discussion

Based on the measured performance in target-following tasks and on questionnaires and free reports, we can look back at the research questions listed in section 4.2.

Question RQ1.1 admits a positive answer, as users with minimal training could effectively follow a target with a mean distance as low as $75.33\,\mathrm{pixels}$ ($27.34\,\mathrm{mm}$, 3.42% of diagonal length, participant 3) for duplets and $76.52\,\mathrm{pixels}$ ($27.77\,\mathrm{mm}$, 3.47% of diagonal length, participant 9) for triplets (see Table 1). To have a measure for how good the target-following trajectory is, we can compare it to random navigation by a human. Namely, we can consider the five-minute free-navigation training of one of the participants who exhibited good navigation ability. For example, taken the trajectory produced by participant 3 during training with duplets, and computing the instantaneous distance from the target trajectory in the same time span, we get a mean distance of $637.20\,\mathrm{pixels}$, that is 28.93% of diagonal length, more than eight times the mean distance the participant achieved during actual target following. By noting how a user can keep the controlled object relatively close to a moving target, we can say that movement can be controlled by rhythmic cells on two buttons.

The performance clearly improved in the second half of the experimental session, thus showing acquired dexterity through practice. Participants felt a lower temporal pressure and less frustration with more practice. The reports of increased confidence and development of a sense of engagement positively answer to research question RQ1.2.

An overall performance difference at the edge of significance, the different subjective ratings of performance, as well as some individual comments, point to three-taps-based control as initially less easy yet potentially more engaging (RQ1.3). The performance with triplets was significantly worse at first exposure, but the participants previously exposed to duplets performed comparably well when later exposed to triplets. More practice can turn anxiety to engagement, and an increase in difficulty can turn boredom to engagement as well [48]. More complex rhythms, such as those obtained with triplets, offer a performative potential that may be developed through practice, to turn frustration to engagement.

Even though the proposed interaction technique is based on rhythm, the role of the polyrhythmic and polytemporal auditory display has only been occasionally appreciated by participants. The focus of attention was mainly visual, so the sounds could be unattended without impairing the task. It is expected that, for tasks where the visual display becomes temporarily unavailable, an auditory polytempo that can be interpreted as a velocity vector would reveal its effectiveness. The role of auditory display in determining the level of engagement remains to be assessed, although the interface was found to be an engaging audio-visual whole.

The implementation details of the proposed interaction technique are not irrelevant. The quality of buttons plays a role, as keys for fingerdrumming (as the ones being tested) require a different attitude and physical effort than keys for typing. In particular, key-velocity detection has been used in the experiment to modulate sound intensity, but its expressive role and contribution to engagement have not been investigated yet. There are inherent difficulties to achieve a faultless detection of rhythmic cells, as pauses between cells may be mistaken as within-cell IOIs. A better key-tap feedback, possibly accompanied by tactile stimulation, may reduce misses and rhythm detection faults as well.

5. Experiment 2

A second experiment was designed and run to investigate if rhythmic sound feedback may actually be exploited for controlling movement on a plane under temporary deprivation of visual feedback. Attention was focused on the simplest rhythmic cells, those made of duplets, and the target-following task was modified to hide the controlled object for half of its lifetime, with blank visual feedback on every other time segment of $5\,\mathrm{s}$.

5.1. Objective

The objective of the experiment is:

- Controlling movement by rhythmic cells of two taps on two buttons, to compare the performances in following a target moving on the 2-D plane under the two conditions: (i) full visual and full rhythmic sound feedback, (ii) intermittently deprived visual feedback and full rhythmic sound feedback.

5.2. Research questions

The following research questions are being tested experimentally:

RQ2.1 Is it possible to control the movement of an object in the 2-D space by rhythmic cells on two buttons, relying on audiovisual feedback as well as on rhythmic sound feedback, when visual feedback is intermittently available?

RQ2.2 Is there an asymmetry in skill transfer in acquired dexterity with rhythm-based movement control, with and without impairment of the visual feedback?

RQ2.3 Does the partial deprivation of visual feedback induce focusing on auditory feedback and an increased appreciation of the multisensory experience?

Research questions RQ2.1 and RQ2.2 are going to be addressed through measurements of performance in a target-following task. Question RQ2.3 is investigated through a questionnaire and the reported experience, after measuring the performance in the two halves of the experimental session.

5.3. Participants

Participants were recruited among students and researchers of computer science of the University of Palermo with a call for volunteers. The 17 participants (4 female) reported normal or corrected to normal vision. They all reported normal hearing, except for participant n. 5, who reported wearing a hearing aid. The performance of this participant was analyzed separately and excluded from aggregate analysis. A perfect balancing in the sequence of conditions was organized for the remaining 16 participants. The median age of the 16 retained participants was 28.5 years, with interquartile range of 7.25 years. Five participants on 16 declared some kind of musical practice. Fourteen participants on 16 were native Italian-language speakers, and for them all oral and written interaction occurred

in Italian. For the remaining two retained participants oral and written interaction occurred in English. None of the participants was previously selected for experiment 1.

Expecting an effect that is perceptible and comparable to the variability in performance across participants, we assume a large effect size. With 16 participants, significance level $\alpha = 0.05$, and d between 0.8 and 1.0, the power is ranging between 0.85 and 0.96.

The participants gave their informed consent before the experiment. The experimental protocol was approved by the ethical committee of the University of Palermo.

5.4. Apparatus

A custom audio-visual software was developed in the Processing 4 language and environment, with `themidibus` library and JSyn-based `sound` library. The experiment was run on a MacBook Pro (2.4 GHz 8-core Intel Core i9) with its built-in 16-inch retina display, and the application was run full screen at 1792×1120 pixels, 60 frames per second. The active area of the screen was $345\,\mathrm{mm} \times 215\,\mathrm{mm}$. The experimental apparatus is depicted in figure 8. The visual appearance and interaction were the same as in experiment 1, described in section 4, and depicted in Fig. 4. Two buttons of the ESI Xjam MIDI controller, as highlighted in the right part of Fig 4, with default settings, were used for rhythmic input. Auditory feedback was given through Beyerdynamic DT 770 Pro circumaural headphones driven by a Motu M4 audio interface, whose level was set comfortable and constant for all participants. The custom software application was run under MacOS 14 Sonoma. The round trip audio latency was measured by playing a click sound through the headphones and capturing it back through a microphone, and it amounted to $64\,\mathrm{ms}$. The sounds for the auditory display were the same vocal imitations of percussive sounds as in experiment 1. The auditory display was repeatedly playing rhythmic cells corresponding to the velocity components along the two orthogonal axes. Once a duplet was acquired, it was repeatedly played back to rhythmically display the corresponding velocity component, with a pause of $100\,\mathrm{ms}$ between successive repetitions. Since there are two components of velocity on a surface, two overlapping rhythms were being played during interaction. The key velocity messages sent by the Xjam controller were used to modulate the intensity of the pulses composing the rhythms. In addition to sounds forming the polytemporal texture of the auditory display, a percussive sound, steered to left, right, or left+right channels, was used as immediate (within the latency) feedback of button press, for the left and right button, as well as for simultaneous taps.

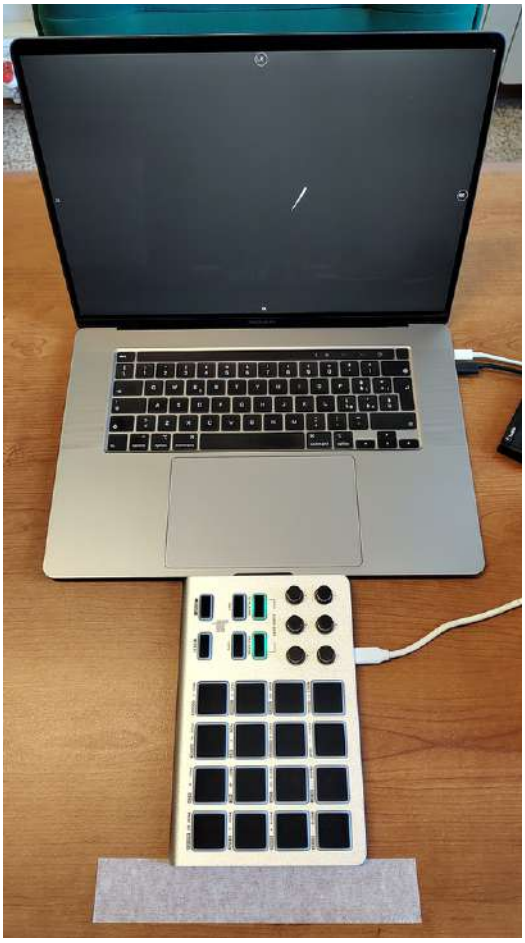


Figure 8: The apparatus of experiment 2

5.5. Procedure

Each participant was exposed to two versions of the interface, one (no-hiding) with complete audiovisual feedback and one (hiding) with intermittent visual hiding of the controlled object, thus dividing the experimental session into two halves. Before starting with the first experimental half, the participant was exposed to a short ($3\,\mathrm{min}\ 14\,\mathrm{s}$) video specific for duplets-based control[6]. The video illustrates the interface, including the controller with the two buttons to be used, and explains how to control the velocity components by tapping. Attention is drawn to the auditory feedback, and an audiovisual example of navigation is given. A screenshot of the instruction video is reported in Fig. 4. In the final part of the video, a target-following task is introduced.

After seeing the video instruction and receiving possible clarifications from the experimenter, the participant was asked to navigate freely on the plane, for about $5\,\mathrm{min}$ by tapping the proposed duplet cells. For most training sessions, the object position was logged, so that distribution of the velocity vector could be collected, and it is shown in Fig. 9. In velocity-magnitude distribution, a value of $50\,\mathrm{pixels/s}$ corresponds, assuming diagonal motion, to the quite short IOI of $170\,\mathrm{ms}$. Lower values of speed correspond to larger IOIs, and higher values of speed correspond to even shorter IOIs. Thereafter, the target-following task was run, as described in section 5.5.1, that lasted about $6\,\mathrm{min}$. The training session and the target-following tasks were repeated in the second half of the experiment. The instruction video was played only once, before the first half. The order of hiding and no-hiding was counterbalanced among participants, to mitigate carryover effect, as well as to measure any asymmetric skill transfer.

As in Experiment 1, the participants were asked to fill two Raw-NASA-TLX questionnaires, one for the first half and one for the second half of the experiment.

Overall, each participant session lasted about $40\,\mathrm{min}$.

5.5.1. Target following

As in experiment 1, the trace of a sailing regatta was used as the trajectory of the target to be followed. In the second half of the experiment, the trace of Oracle boat was specularly inverted about the horizontal and vertical axes, to avoid memorization of the trajectory. The 1,710 timestamped observations of (x,y) coordinates were fit to the screen size and interpolated for smooth display at the chosen frame rate and playback speed. The screen-reconstructed regatta lasted for

[6] https://www.youtube.com/watch?v=ASxdLamllWQ

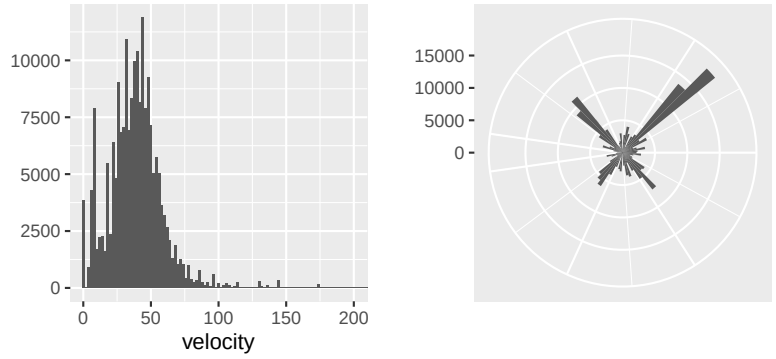


Figure 9: Experimental distribution of velocity magnitudes (in $\mathrm{pixels/s}$) and directions during preliminary practice of 11 recorded participants in condition of full audio-visual feedback.

slightly less than six minutes. Fig. 10 shows the target trace as a thin blue line, in the performances of participant number 13, who was the best performing subject.

Participants were asked to keep the controlled object as close as possible to the moving target.

5.6. Results

5.6.1. Target following

To qualitatively analyze the performance of participants in following a target, we displayed the performed static trace, superimposed to the target trajectory. In the case of intermittent visual hiding of the target, different colors are given to the segments where it was visible (red) and to the segments where it was invisible (yellow), as in Fig. 10. This makes it possible to observe that, indeed, directional adjustments are imparted also when the target is visually hidden, and the user can only rely on auditory feedback.

A quantitative measure was obtained by computing the Euclidean distance of the controlled object from the target, in pixels, at each frame. The istantaneous distances were then averaged over the whole trajectory, to give a mean distance per subject, that is reported in Table 4. The overall mean and standard deviation of the mean distances are also reported in Table 4. Fig. 10 shows the trajectories of participant number 13, who obtained the smallest mean distance for no-hiding (exposed first) and the second to smallest distance for hiding.

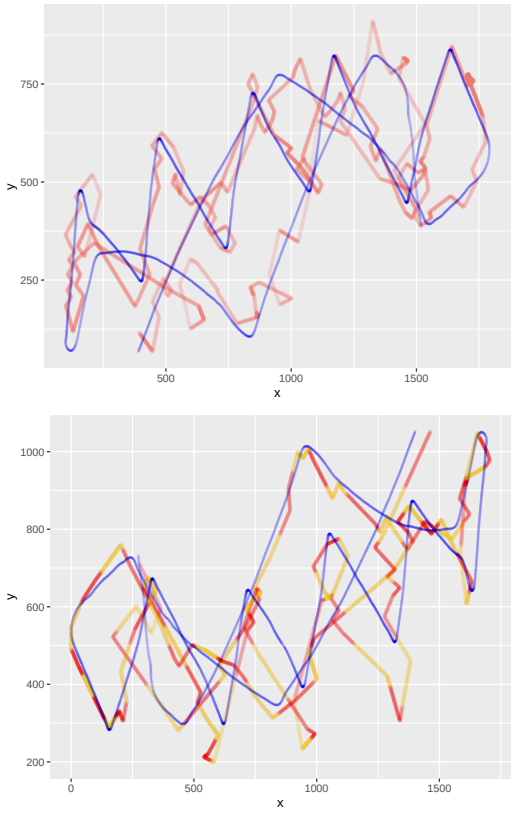


Figure 10: Trajectories of participant number 13 of experiment 2. The thin blue line is the target trace. Top: full audiovisual feedback; Bottom: intermittent hiding of the controlled object (yellow segments). Transparency of the line is proportional to speed.

Parametric hypothesis testing was used to assess the significance of the difference of the means. The mean distance for hiding was $131.48\,\text{pixels}$ (6.2% of diagonal length), and for no-hiding it was $123.81\,\text{pixels}$ (5.9% of diagonal length), but the difference of $7.67\,\text{pixels}$ was not significant ($F_{1,15} = 0.698$, $p = n.s.$). The normality assumptions were fulfilled (Shapiro-Wilk, $p > 0.5$). The null hypothesis of equality of distances can not be rejected.

It is also interesting to compare the performances on the first and second halves of the experimental sessions, to see if there has been a learning effect, regardless of the order of exposure to hiding or no hiding. Overall, in the first half the mean distance was $137.06\,\text{pixels}$ (6.5% of diagonal length), and in the second half it was $118.23\,\text{pixels}$ (5.6% of diagonal length). Given that normality assumptions are fulfilled (Shapiro-Wilk test, with $p > 0.05$), parametric hypothesis testing was

Part.	distance/no-hide	distance/hide
1	178.19	115.04
2	105.88	119.20
3	73.92	60.82
4	135.94	182.29
~~5~~	~~181.52~~	~~224.56~~
6	107.74	120.67
7	169.38	164.98
8	84.16	187.04
9	110.90	101.59
10	166.26	173.30
11	145.93	120.41
12	142.07	135.13
13	51.85	79.05
14	166.70	154.78
15	71.65	82.60
16	80.37	125.16
17	190.05	181.67
(SD)mean	(43.68)123.81	(39.45)131.48

Table 4: Mean distances (in pixels) from target trajectory, per participant, for no-hiding and hiding. Participant 5 has been excluded from the analysis.

used to assess the significance of the difference of the means. The difference of the mean distances of $18.83\,\mathrm{pixels}$ was of medium size and significant ($F_{1,15} = 5.489$, $p = 0.033, \eta^2 = 0.054$). A non-parametric Wilcoxon signed-rank test ($z = 112$, $p = 0.021$) also shows a significant difference between the two halves, and the Wilcoxon effect size is large ($r = 0.569$).

If an overall dependence of mean distance on hiding failed to emerge from the one-way anova, the picture may emerge more clearly from a two-way mixed anova, with hiding as a within-subject factor, and kind of first exposure as a between-subject factor. The assumption of normality was not strictly fulfilled for the distribution of performances at first exposure in hiding condition (Shapiro-Wilk, $p = 0.03$). The assumption of homogeneity of variance (Levene, $p > 0.05$) was fulfilled. In explaining the mean distance from target, there was a statistically significant, medium-size interaction between group of first exposure and the hiding factor ($F_{1,14} = 5.454$, $p = 0.035, \eta^2 = 0.06$). The simple main effect of

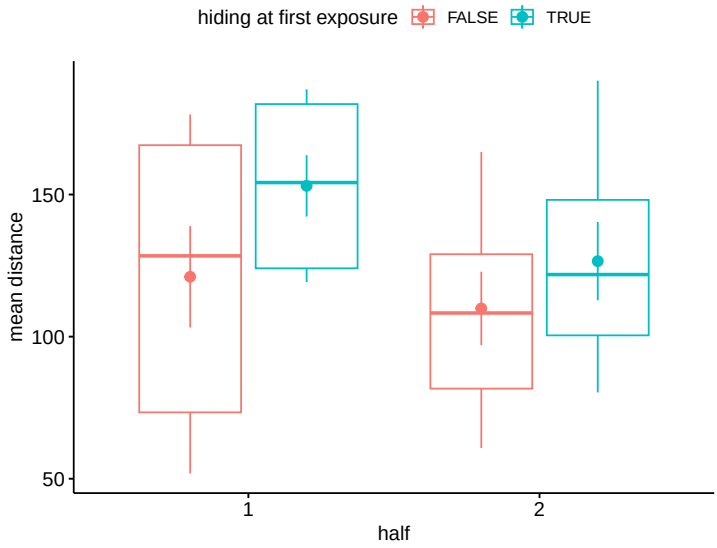


Figure 11: Performances for hiding and no hiding, grouped by first exposure.

group of first exposure was significant for the hide condition ($p < 0.05$) but not for no hiding. The simple main effect of hiding was not significant for neither kind of first exposure. In both the first and second half of the experiment, the between-subjects difference of the means was not significant.

Figure 11 shows the distributions of mean distances for hiding and no hiding, grouped by first exposure. Interaction in hiding condition produces a slightly worse performance when used in the first half of the experiment, thus indicating asymmetric skill transfer, or how learning is more effective when starting without visual impairment.

5.6.2. Questionnaire and report

Table 5 reports the median and inter-quartile range of the responses of the 16 retained participants to the six questions of the Raw-NASA-TLX questionnaire. A Wilcoxon signed-rank test shows a significant difference between hide and no-hide for the question on mental demand, overall satisfaction with performance, effort, and frustration. For such questions the Wilcoxon effect size is moderate to large ($r > 0.49$). So, a significantly larger effort was required when the participants had to rely on sound during visual hiding of the object being controlled. The satisfaction with their performance was smaller with intermittent visual deprivation, and higher was their sense of frustration.

Question	no-hide	hide	p-value
mental demand	-3.5(8.0)	1.5(8.25)	0.01 **
physical demand	-8.0(3.25)	-7.0(4.75)	n.s.
temporal demand	-3.0(11.2)	-1.5(9.75)	n.s.
performance	-3.5(5.25)	-1.5(3.75)	0.022 *
effort	2.0(5.25)	5.0(3.0)	0.0005 ***
frustration	-6.5(5.25)	-3.0(7.0)	0.046 *

Table 5: Median (IQR) of the ratings for each of the six questions of the Raw-NASA-TLX questionnaire, after having performed with duplets and with triplets.

The same ratings of the Raw-NASA-TLX questionnaire have been analyzed to check if the subjective task load changed between the first and second half of the sessions. Table 6 reports the median and inter-quartile range of the responses to the six questions, for the two session halves. A Wilcoxon signed-rank test shows no significant difference between the two halves, for any of the asked questions. This confirms that the first training session was enough to develop the required dexterity for the task.

Question	first	second	p-value
mental demand	-1.0(6.25)	-0.5(8.75)	n.s.
physical demand	-7.5(3.25)	7.5(4.25)	n.s.
temporal demand	-0.5(12.0)	-3.0(10.8)	n.s.
performance	-3.0(4.0)	-3.0(6.0)	n.s.
effort	3.5(4.75)	3.5(7.0)	n.s.
frustration	-5.0(7.5)	-4.0(6.75)	n.s.

Table 6: Median(IQR) of the ratings for each of the six questions of the Raw-NASA-TLX questionnaire, after the first and second half of the session.

The participants left some comments at the end of the experimental session, and the most relevant are here reported:

Practice improves performance: Nine participants out of sixteen commented on the effectiveness of learning through practice. Three of these, exposed to intermittent object hiding in the first half of the experiment, described how the added initial difficulty impaired the learning process, making it difficult to take full advantage of the available feedback while learning complex control patterns. One

participant, first exposed to full audiovisual feedback, noticed a little difference between the first and the second half of the experiment;

Learning the coordination of tapping commands: Five participants tried to express some tactics they used to impart the desired speed and direction by coordination of different tapping sequences;

It is an engaging game: Four participants established a direct link between the level of engagement and how challenging the task was, especially in the condition of intermittent hiding. The task was often described as a game, and four participants proposed software and hardware variations that may render such game even more enjoyable to play;

Sound is appreciated: Eight participants commented on how they relied on sound, especially in the condition with intermittent visual hiding. One of these added that sound makes the experience more fun. A couple of other participants found the sound rhythms annoying, one commenting that this is not the kind of sound that is found in games, aiming at relaxing the player and increasing flow;

Control glitches impair the performance: Three participants reported that often the system failed to detect the imparted commands, due to the mechanical compliance of buttons or to difficulties in coping with quick multiple taps. One of these proposed to add a mechanical clicking feedback to the buttons.

5.7. Discussion

Based on the measured performance in target-following tasks and on questionnaires and free reports, we can look back at the research questions listed in section 5.2.

Question RQ2.1 is positively answered, as users with minimal training could effectively follow a target with a mean distance as low as $51.85\,\mathrm{pixels}$ ($9.98\,\mathrm{mm}$, 2.5% of diagonal length) for full audio-visual feedback and $60.82\,\mathrm{pixels}$ ($11.7\,\mathrm{mm}$, 2.9% of diagonal length) when the controlled object was visually hidden for half of its lifetime (see Table 4). According to the results and to the reported comments on the role of sound, participants were able to rely on rhythmic sound feedback when visual feedback was not available. Only one participant reported relying on imagined visual trajectory rather than on sound. Indeed, the absence of a "deaf-blind" control condition does not allow us to rule out the possibility to achieve a comparable level of performance even without sound feedback during the hiding

segments. We have noticed that adjustments in the direction of motion are often performed in the temporal segments of visual hiding, when only sound feedback is available, but we can not exclude that participants would make turns, or tacks, even without any sensory feedback, just based on memory, mental visualization and proprioception (dead reckoning). However, human locomotion has been shown to suffer from random, systematic and idiosyncratic angle estimation errors in absence of external directional cues [49], and similar drifting is expected to occur for the proposed navigation by tapping, if the controlled object is neither visible nor audible.

A learning process clearly emerged from the performance data, the task load reports, and the participants' comments. However, when the intermittent visual hiding of the controlled object was presented in the first half of the experiment, the learning curve became steeper. In such case, the participant had to quickly learn, at the same time, how to control the velocity of the object and how to listen to the concurrent rhythms to have feedback on speed and direction. Participants who were first exposed to the full audiovisual condition, on the other hand, had a smooth learning process and found little or no problem to face intermittent hiding in the second half, when they could only rely on sound for half of the object lifetime. As from the analysis of section 5.6.1, question RQ2.2 admits a positive answer.

The analysis of the questionnaires and some comments left by the participants give evidence to positively answer the research question RQ2.3. In fact, a subjectively higher effort and mental demand were needed in the condition of intermittent visual hiding. When the controlled object disappeared, participants had to focus on auditory feedback, following and interpreting the concurrent auditory streams to deduce the speed and direction of the object. The additional required effort, however, is often perceived as a challenge that makes interaction more engaging and enjoyable.

6. Limitations

In proposing a new kind of non-natural interaction based on rhythm, a number of questions were raised, that were addressed only partially in the two studies here reported.

The ability to control speed and direction has been verified through a target-following task, where it has been shown that a controlled object can be kept relatively close to the moving target. However, no measure of accuracy of velocity control has been taken. This would imply measuring just-noticeable differences

in tempo perception, and variability in tempo production. Although such perceptual and motor variations have been measured [5], their impact on TickTacking remains to be ascertained.

The number of participants to the two experiments is small, although the sample size meets local standards in experiment 2, and is close to local standards in experiment 1 [44]. In particular, although a relatively large difference in performance between two- and three-taps rhythmic cells was measured in experiment 1, it failed to reach significance, and this may well be a type II error due to low power.

In experiment 2, the difference of measured performance between the two conditions (hiding and no-hiding) was small and not significant. However, some caution should be exercised while stating that auditory feedback allows one to maintain the same level of performance in presence of visual deprivation. The temporal hidden/visible ratio of the controlled object was probably too high, and the intervals of visual deprivation too short, to rule out the possibility of feedback by imagination only [10]. That is, users may internally visualize the continuation of a trajectory without relying on any kind of feedback, and neglecting the available auditory feedback.

7. Conclusions

A rhythm-based technique to control the velocity of a moving object on a surface has been proposed. It is based on two points of action, that could be two buttons that get tapped, or other kinds of sensors that can be controlled by two symmetric parts of the human body. The rhythmic commands (duplets or triplets of taps) trigger discrete changes in magnitude, orientation, and direction of the velocity vector. The moving-object velocity can be auditorily displayed as a polytemporal pattern, that is obtained by iteration of the imparted rhythmic cells.

An implementation of the proposed interaction, using minimal two-taps sequences or more complex three-taps sequences, was tested in a target-following task, similar to chasing a boat in a race. The interaction technique could be understood and learnt in a relatively-short time. A target moving object could be chased at a relatively-small distance by TickTacking, and drawing trajectories by rhythm proved to be feasible and engaging.

Three-taps sequences introduce a degree of freedom in the internal structure of the rhythmic cell, that can be exploited to vary the rhythmic feedback expressively, thus making velocity control a goal to be achieved through a creative activity of rhythm improvisation. This larger space for creative performance is obtained by

making interaction more difficult to learn and to use proficiently. With duplets, on the other hand, the only expressive degree of freedom is found in the possibility to accentuate the taps differently, if the input device can capture key velocity or similar variations of action.

The ability of users to effectively exploit the informative content of the rhythmic auditory display, overcoming possible occasional deprivation of visual feedback, has been tested for duplets-based control. However, it is possible that users fill in the blanks of visual feedback with their own imagination, regardless of auditory feedback. Although several participants reported about their learning to exploit auditory information, further experimentation, to stress the role of auditory rhythmic feedback in absence of visual information, would be necessary.

For the proposed interaction design and exploratory studies, the combined reading of performance measurements and subjective reports indicate that users learn to listen and to interpret the polytemporal rhythms as a display of speed and direction. The reliance on audition becomes relevant in all contexts where visual attention can not be diverted, as in driving, or to make velocity control accessible to the visually impaired.

Although not explored and tested in this study, reliance on the sense of touch is also possible. The tactile rendering of rhythms should be considered for applications where auditory display is better supported or replaced by the more intimate sense of touch. The perceptual segregation of tactile rhythmic streams may be more difficult to achieve from pointlike stimulations, due to technological and sensory limitations, but more research is needed to define a design space for tactile stimuli. Still, the mapping of four tactile feedback stimulation points to the four semiaxes of the velocity space may be practical in some contexts and applications, even keeping only two points of action.

Most people find the proposed velocity-control technique, and the associated target-following task, quite weird or non-natural at first try. However, the experiments have shown that, in a relatively short time, they can learn how to change speed and direction by tapping, and to monitor their directional motion by listening to polytemporal rhythmic feedback. These results make the technique and task suitable for further studies in sensory-motor learning and control, e.g., to investigate the *de-novo* learning processes and how they may be affected by multisensory rhythmic feedback. The proposed interface is indeed being used in studies of motor learning, and a model of automatic control mimicking human behavior is also being developed.

The presence of two control points makes it possible to assign the input devices (buttons or other sensors) to different persons. Going beyond control of

a moving object by a single person, interaction by ticking and tacking may be exploited in an inter-individual coordination perspective for joint action and performance, to investigate cooperative motor control, and with applications in art, play, therapy, and training.

Acknowledgement

The interaction design, system demonstration, and experiment design were done while Alessio Bellino was visiting researcher at the University of Palermo. Antonino Perez and Gabriele Ferrara contributed to experiments 1 and 2, respectively, as master students of the University of Palermo. The work was partially supported through the FFR 2023 fund of the University of Palermo. The first author is supported by the project "Multiscale Analysis of Human and Artificial Trajectories: Models and Applications" funded by the MUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022 - grant 2022RB939W.

Appendix A. Data and software

The software used to run the experiments described in sections 4 and 5, the analysis scripts, and the collected data are available at
https://github.com/d-rocchesso/TickTack

References

[1] D. Rocchesso, A. Bellino, A. Perez, et al., Ticktacking–drawing trajectories with two buttons and rhythm, in: Proceedings of the Sound and Music Computing Conference, KTH, 2023, pp. 63–71.
URL https://smcsweden.se/proceedings/SMC2023_proceedings.pdf

[2] Y. Visell, F. Fontana, B. Giordano, R. Nordahl, S. Serafin, R. Bresin, Sound design and perception in walking interactions, International Journal of Human-Computer Studies 67 (11) (2009) 947–959. doi:10.1016/j.ijhcs.2009.07.007.

[3] I. Shirai, Multi-directional switch (Aug. 1987).
URL https://patents.google.com/patent/US4687200

[4] B. A. Myers, Pick, Click, Flick! The Story of Interaction Techniques, 1st Edition, Vol. 57, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3617448.

[5] J. D. McAuley, Tempo and rhythm, in: M. Riess Jones, R. R. Fay, A. N. Popper (Eds.), Music Perception, Springer New York, New York, NY, 2010, pp. 165–199. doi:10.1007/978-1-4419-6114-3_6.

[6] D. J. Hermes, Auditory-Stream Formation, Springer International Publishing, Cham, 2023, pp. 559–784. doi:10.1007/978-3-031-25566-3_10.

[7] C. Ware, Information visualization: perception for design, Morgan Kaufmann, 2019.

[8] D. Huron, Voice Denumerability in Polyphonic Music of Homogeneous Timbres, Music Perception 6 (4) (1989) 361–382. doi:10.2307/40285438.

[9] R. Brochard, C. Drake, M.-C. Botte, S. McAdams, Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation, Journal of Experimental Psychology: Human Perception and Performance 25 (6) (1999) 1742.

[10] R.-D. Vatavu, From natural to non-natural interaction: Embracing interaction design beyond the accepted convention of natural, in: Proceedings of the 25th International Conference on Multimodal Interaction, ICMI '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 684–688. doi:10.1145/3577190.3616122.

[11] E. Freeman, G. Wilson, D.-B. Vo, A. Ng, I. Politis, S. Brewster, Multimodal feedback in HCI: Haptics, non-speech audio, and their applications, in: S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, A. Krüger (Eds.), The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1, Association for Computing Machinery and Morgan & Claypool, 2017, p. 277–317. doi:10.1145/3015783.301579.

[12] H. Lefebvre, Rhythmanalysis: Space, time and everyday life, Continuum, London, UK, 2004.

[13] S. Adhitya, Musical cities, UCL Press, 2018. doi:10.14324/111.9781911576563.

[14] C. Erkut, A. Jylhä, D. Rocchesso, Heigh Ho: Rhythmicity in Sonic Interaction, in: Sonic Interaction Design, The MIT Press, 2013. doi:10.7551/mitpress/8555.003.0023.

[15] D. Rocchesso, P. Polotti, S. Delle Monache, Designing continuous sonic interaction, International Journal of Design 3 (3) (2009).

[16] C. Erkut, Rhythmic interaction in VR: interplay between sound design and editing, in: IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), 2017, pp. 1–4. doi:10.1109/SIVE.2017.7901611.

[17] A. Jylhä, I. Ekman, C. Erkut, K. Tahiroğlu, Design and evaluation of human-computer rhythmic interaction in a tutoring system, Computer Music Journal 35 (2) (2011) 36–48.

[18] E. Ghomi, G. Faure, S. Huot, O. Chapuis, M. Beaudouin-Lafon, Using rhythmic patterns as an input method, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, p. 1253–1262. doi:10.1145/2207676.2208579.

[19] E. Freeman, G. Griffiths, S. A. Brewster, Rhythmic micro-gestures: Discreet interaction on-the-go, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017, p. 115–119. doi:10.1145/3136755.3136815.

[20] A. Bellino, Rhythmic-synchronization-based interaction: Effect of interfering auditory stimuli, age and gender on users' performances, Applied Sciences 12 (6) (2022). doi:10.3390/app12063053.

[21] L. Velautham, R. S. Y. Chen, Can't clap to a beat? How rhythmically challenged people experience and strategize keeping time to music, Psychology of Music 50 (4) (2022) 1254–1266. doi:10.1177/03057356211049564.

[22] M. Rinott, N. Tractinsky, Designing for interpersonal motor synchronization, Human–Computer Interaction 37 (1) (2022) 69–116. doi:10.1080/07370024.2021.1912608.

[23] R. S. Y. Chen, Embodied design for non-speaking autistic children: The emergence of rhythmical joint action, in: Interaction Design and Children, 2021, p. 648–651. doi:10.1145/3459990.3463396.

[24] P. Bozzi, G. B. Vicario, Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale, Rivista di psicologia 54 (4) (1960) 253–258.

[25] J. Maculewicz, A. Jylha, S. Serafin, C. Erkut, The effects of ecological auditory feedback on rhythmic walking interaction, IEEE MultiMedia 22 (01) (2015) 24–31. doi:10.1109/MMUL.2015.17.

[26] H. Landin, S. Lundgren, J. Prison, The iron horse: A sound ride, in: Proceedings of the Second Nordic Conference on Human-Computer Interaction, 2002, p. 303–306. doi:10.1145/572020.572075.

[27] A. Gabrielsson, The relationship between musical structure and perceived expression, in: Oxford Handbook of Music Psychology, Oxford University Press, 2008. doi:10.1093/oxfordhb/9780199298457.013.0013.

[28] R. O. Gjerdingen, Apparent Motion in Music?, Music Perception 11 (4) (1994) 335–370. doi:10.2307/40285631.

[29] B. L. Giordano, H. Egermann, R. Bresin, The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity, PLOS ONE 9 (12) (2015) 1–23. doi:10.1371/journal.pone.0115587.

[30] C. Frame, Ding-dong: Meaningful musical interactions with minimal input, in: Proceedings of the Sound and Music Computing Conference, 2023. URL https://smcsweden.se/proceedings/SMC2023_proceedings.pdf

[31] A. Bellino, D. Rocchesso, R. Mulé, L. D'Arrigo Reitano, Multisensory trajectory control at one interaction point, with rhythm, in: Proceedings of the 19th International Audio Mostly Conference, AM '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3678299.3678340.

[32] W. Buxton, M. Billinghurst, Y. Guiard, A. Sellen, S. Zhai, Human input to computer systems: theories, techniques and technology, 2011, unpublished book manuscript.
URL https://www.billbuxton.com/inputManuscript.html

[33] I. S. MacKenzie, Interaction elements, in: I. S. MacKenzie (Ed.), Human-computer Interaction, Morgan Kaufmann, Boston, 2013, pp. 71–120. doi:10.1016/B978-0-12-405865-1.00003-0.

[34] C. T. Annand, F. M. Grover, P. L. Silva, J. G. Holden, M. A. Riley, Early learning differences between intra- and interpersonal interlimb coordination, Human Movement Science 73 (2020) 102682. doi:10.1016/j.humov.2020.102682.

[35] D. F. Vanderelst, H. Peremans, Computational analysis of the Etch-A-Sketch task: A comment on Annand et. al. 2020, PsyArXiv (2022). doi:10.31234/osf.io/7vdhu.

[36] H. Pohl, A. Muresan, K. Hornbæk, Charting subtle interaction in the HCI literature, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–15. doi:10.1145/3290605.3300648.

[37] A. S. Bregman, Auditory scene analysis: The perceptual organization of sound, MIT press, 1994.

[38] A. S. Bregman, Auditory scene analysis and the role of phenomenology in experimental psychology., Canadian Psychology/Psychologie canadienne 46 (1) (2005) 32.

[39] M. Jokiniemi, R. Raisamo, J. Lylykangas, V. Surakka, Crossmodal rhythm perception, in: A. Pirhonen, S. Brewster (Eds.), Haptic and Audio Interaction Design, Springer Berlin Heidelberg, 2008, pp. 111–119.

[40] C. Bernard, J. Monnoyer, M. Wiertlewski, S. Ystad, Rhythm perception is shared between audio and haptics, Scientific Reports 12 (1) (2022) 4188. doi:10.1038/s41598-022-08152-w.

[41] P. Dourish, Where the action is: the foundations of embodied interaction, MIT press, 2001.

[42] T. Pakkanen, R. Raisamo, V. Surakka, Audio-haptic car navigation interface with rhythmic tactons, in: M. Auvray, C. Duriez (Eds.), Haptics: Neuroscience, Devices, Modeling, and Applications, Springer Berlin Heidelberg, 2014, pp. 208–215.

[43] F. A. Russo, P. Ammirante, D. I. Fels, Vibrotactile discrimination of musical timbre, Journal of Experimental Psychology: Human Perception and Performance 38 (4) (2012) 822. doi:10.1037/a0029046.

[44] K. Caine, Local standards for sample size at CHI, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 981–992. doi:10.1145/2858036.2858498.

[45] J. Cohen, Statistical power analysis for the behavioral sciences, 2nd Edition, Routledge, New York, 1988.

[46] S. G. Hart, Nasa-task load index (NASA-TLX); 20 years later, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50 (9) (2006) 904–908. doi:10.1177/154193120605000909.

[47] G. Norman, Likert scales, levels of measurement and the "laws"of statistics, Advances in Health Sciences Education 15 (5) (2010) 625–632. doi:10.1007/s10459-010-9222-y.

[48] G. Chanel, C. Rebetez, M. Bétrancourt, T. Pun, Boredom, engagement and anxiety as indicators for adaptation to difficulty in games, in: Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era, 2008, p. 13–17. doi:10.1145/1457199.1457203.

[49] S. Jetzschke, M. O. Ernst, A. Moscatelli, N. Boeddeker, Going round the bend: Persistent personal biases in walked angles, Neuroscience Letters 617 (2016) 72–75. doi:10.1016/j.neulet.2016.01.026.