TickTacking – Drawing trajectories with two buttons and rhythm

Davide Rocchesso
Università di Palermo, Italy
d vide.rocchesso@unip .it

lessio Bellino
Universidad Diego Portales, Chile
lessio.bellino@m il.udp.cl

ntonino Perez
Università di Palermo, Italy

BSTR CT

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. Tacking back and forth must be used to make progress along certain directions, similarly to sailing a boat upwind. The proposed rhythmic navigation technique is tested with a target-following task, using a boat-racing trace as the target. 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.

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, 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. Clockwork mechanisms are rhythmical and often associated with motion and locomotion, and the pace of ticking is proportional to the resulting progress velocity.

Looking at the literature of rhythmicity for interaction, however, 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 to control motion 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 and a joystick. 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.

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 give time.

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 [1]. Therefore, the ratio of vertical and horizontal tapping rates can not be made too large or too small. s a result, it is essentially impossible to move the object along the main axes. 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. 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 rhythmic patterns and perceptually separated by auditory streaming. In principle, it is possible to guess the speed magnitude, orientation and direction of the object by sound alone.

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 a 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.

[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.

2. B CKGROUND

In interaction design and human-computer interaction, the rhythmic interaction with devices or technology-augmented objects has been studied in a wide range of contexts and scales. The amodality of rhythms [2] has produced studies and solutions for one or more of the senses of touch, hearing, vision, and proprioception. The rhythms of our cities as they are experienced in everyday urban lives, have been explored through the use of sound and music, and used as a key design dimension for urban planning [3]. In sonic interaction design, systems and interfaces that support rhythmicity and afford the development of virtuosity have been proposed [4]. The role of rhythm in multisensory continuous interaction has been investigated with design exercises [5]. Film editing and rhythmic interaction design techniques have been proposed for use in cinematic virtual reality [6]. Rhythmic tutoring has been proposed for interaction by handclapping [7]. Rhythmic patterns, composed of short and long taps and breaks, have been proposed to replace single commands and tested for recall [8]. Rhythmic microgestures have been proposed for non-visual interaction in mobility [9]. Selection by visual rhythmic patterns and motion synchronization has been proposed [10].

lthough there are rhythmically challenged persons [11], rhythmicity is generally found useful in joint action. The rhythmic propensity of autistic individuals has been given positive value through technology, to foster social interaction [12].

Temporal proximity has been recognized as a unification principle for multiple events that are perceptually grouped to form rhythmic patterns, or gestalts [13]. 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 [14]. Horse gait patterns have been used to augment human locomotion by biking, so that one can get the bike to walk, trot or gallop [15]. Rhythmic structures are emotionally expressive [16], and the similarities between musical performance and everyday motor activity have been described [17, 18]. 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 [19].

The proposed rhythmic control of movement through buttons effectively realizes a rate-control input device [20] and 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. lthough the input device does control movement of an object on a surface, it can not be assimilated to a pointing device [21], as its purpose is to control a velocity vector rather than hitting a target. 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- -Sketch, a classical drawing toy that has recently become a research paradigm to investigate inter-limb and inter-individual coordination [22, 23]. While Etch- -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.

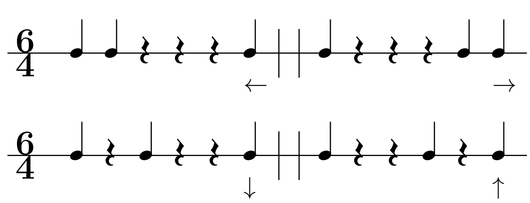
3. DESIGN R TION LE FOR RHYTHMIC CONTROL OF TR JECTORIES

Rhythmic control of movement through buttons would be desirable in a variety of applications and contexts, such as vehicle driving, interfaces for people with special needs, or entertainment. The control would be exerted through tapping rate, that may directly map to speed, similarly to how the galloping rate is related to horse speed. mong the possible ways to control the directional properties of velocity, we map 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.

3.1 Rhythms by tapping

If we are interested in controlling a velocity vector by tapping rhythms, a legitimate question is: 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:



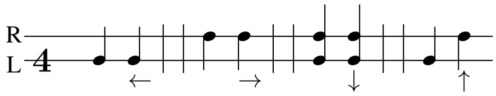
This would require memorization of the sequences, explicit counting, and extensive training [2].

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,

[2] Combining velocity magnitude control by tapping with rotation by holding, we may indeed use a single button with simpler rhythmic cells, and this possibility is being investigated in another study.

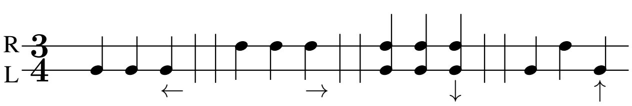
and LR, allowing for RL as well to accomodate for possible left-right inversions:



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. 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, with the X pulse represented by a darker sound and the LR sequence using a couple of pulses, the first darker than the second, as in ascending brightness. For effective auditory monitoring of direction and speed, it is critical to choose sounds that segregate in two streams in all conditions, to make the two (horizontal and vertical) components of velocity clearly discernible [24, 25], 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.

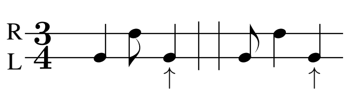
3.1.2 Triplets on two buttons

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.
 mong 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 [16], 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 [17]. To make the vertical semiaxes clearer, we make the XXX sequence correspond to the repetition of a darker sound and the LRL sequence correspond to a triplet of pulses, in ascending order of brightness. Fig. 1 shows an example of polytempo obtained by superposition of triplets LLL and LRL that represent movement in a left-up direction.

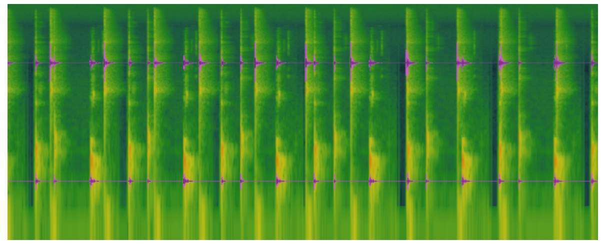


Figure 1. Spectrogram with superimposed stereo waveforms for two overlapping triplets LLL and LRL at different paces, representing motion in a left-up direction. Horizontal linear scale for time, vertical logaritmic scale for frequency.

 nother 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 [1], 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 direction and speed remain almost constant between 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.
 s 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 the two orthogonal directions, but a sequence of contrasting rhythmic cells may produce zig-zag motion around a semi-axis.

In the experiment we are going to describe in section 4, 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

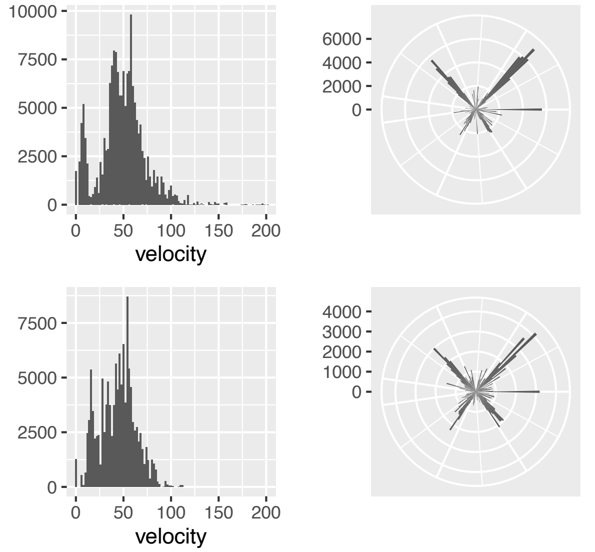


Figure 2. Experimental distribution of velocity magnitudes (in $\mathrm{pixels/s}$) and directions during preliminary practice of participants. Top: two taps, 9 recorded participants (1 not recorded); Bottom: three taps, 7 recorded participants (3 not recorded).

v_x and v_y by discrete differentiation, and to obtain an empirical collection of the distribution of velocity angles ($\arctan \frac{v_y}{v}$). This is illustrated in Fig. 2, where the dead angles are visible, although there are residual directions along the axes that are due to saturation of object position when it reaches the edges of the window. The polar histograms show a tendency, among persons with no prior 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 uditory or tactile rhythms

The rhythmic feedback can be either auditory or tactile. The former can be intimate if delivered through headphones or earplugs, or public if delivered through loudspeakers. Tactile feedback is inherently intimate. The perception of rhythms can be similarly effective with the two senses [26, 27], and the separation between the two domains become blurry if devices based on bone conduction are used.

n application scenario is that of the car, where two sensors and two 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 [28]. Rhythmic tactons have been proposed and used in HCI [2], also for the car driving environment [29]. For the scope of this paper, however, we only consider auditory rhythmic feedback through headphones.

4. EXPERIMENT

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 personalized headphone listening, 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, 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 a rate (reciprocal of inter-onset interval IOI) 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 Hypotheses

The following hypotheses are being tested experimentally:

HP1 It is possible to control the movement of an object in the 2-D space by rhythmic cells on two buttons.

HP2 cquired dexterity with rhythm-based movement control improves performance and increases engagement.

HP3 Three-taps rhythmic cells induce more engaging interactions as compared to two-taps rhythmic cells, although the performance is more difficult.

Hypothesis HP1 is going to be tested through measurements of performance in a target-following task. Hypothesis HP2 is investigated through a questionnaire and the reported experience, after measuring the performance improvement between the two halves of the experimental session. Hypothesis HP3 is investigated by comparing 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.

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 pparatus

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 192 1 8 pixels, 6 frames per second. The active area of the screen was 697 mm 392 mm. The object being controlled on the screen had a comet shape, with a tail that became proportionally longer at higher speed. t 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. 3 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 3, with default settings, were used for rhythmic input. uditory feedback was given through Beyerdynamic DT 770 Pro circumaural headphones driven by a Native Instruments Komplete udio 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 8 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 1 ms between successive repetitions, introduced to enhance perceptual grouping [1]. 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 min 14 s)

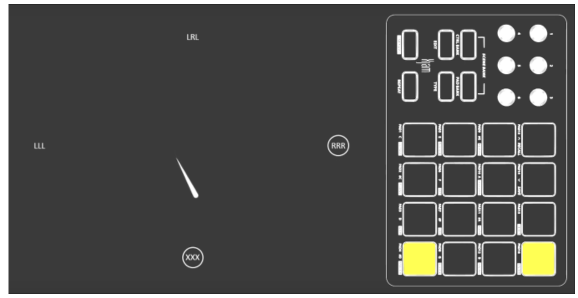


Figure 3. Screenshot of the instruction video. The left part shows a frame of the visual display that the participant would see during interaction, showing downward motion of the controlled object. The right part shows the layout of the Xjam controller, where the two used buttons are highlighted.

video 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. ttention is drawn to the auditory feedback, and an audiovisual example of navigation is given. screenshot of the instruction video is reported in Fig. 3. In the final part of the video, a target-following task is introduced.

fter seeing the video instruction and receiving possible clarifications from the experimenter, the participant was asked to navigate freely on the plane, for about 5 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. 2. It can be observed that participants preferred velocity magnitudes around 5 pixels/s, corresponding to a preferred IOI of about half a second, which is consisted with reported values of preferred tempo [1].

Thereafter, the target-following task was run, as described in section 4.5.1, that lasted about 6 min. The order of exposure to duplets and triplets was counterbalanced among participants, to mitigate carryover effect.

fter both halves of experimentation, the participants were asked to fill two Raw-N S -TLX questionnaires [30], one for the duplet and one for the triplet interface. The paper-and-pencil version with 21 gradations of the rating scales was used, with -10 corresponding to ”very low” and +10 corresponding to ”very high”. 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 4 min.

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

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 mer-ica'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 for about six minutes. Fig. 4 shows the target trace as a thin red line.

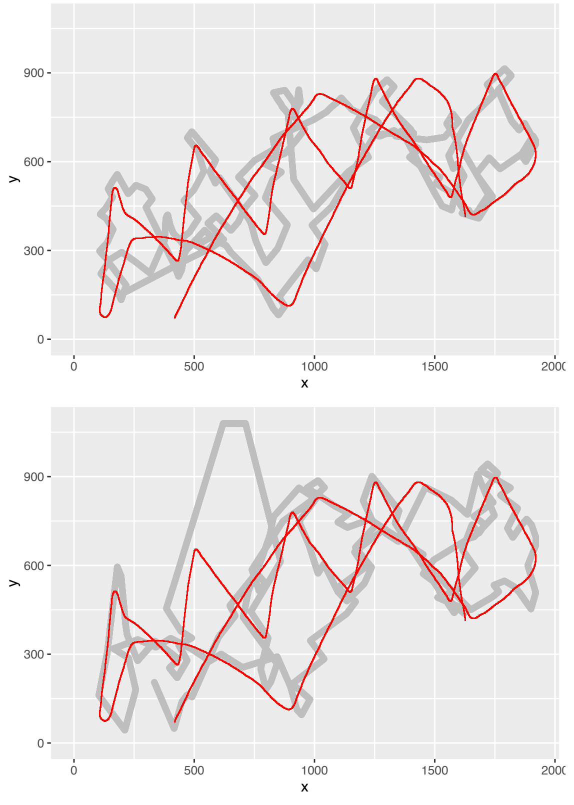


Figure 4. Trajectories of participant number 3. The thin red line is the target trace. Top: duplets; Bottom: triplets.

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

Given that normality assumptions are fulfilled (Shapiro-Wilk test, with $p >$. 5), parametric hypothesis testing was used to assess the significance of the difference of the means. The mean distance for duplets was 1 5.27 pixels, and for triplets it was 139.66 pixels, but the difference of 34.4 pixels was not significant (F $_9 = 4.469$, $p =$. 64). It is well possible that a larger number of participants would let the higher effectiveness of duplets emerge more clearly, but with the collected data the null hypothesis of equality of distances can not be rejected.

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	(87.19)105.27	(113.28)139.66

Table 1. Mean distances from target trajectory, per participant, for duplet and triplet control.

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 pixels, and in the second half it was 1 1. pixels. Given that normality assumptions are fulfilled (Shapiro-Wilk test, with $p >$. 5), parametric hypothesis testing was used to assess the significance of the difference of the means. The difference of 42.93 pixels was significant (F $_9 = 9.639$, $p =$. 13, $^2 =$ .264).

4.6.2 Questionnaire and report

Table 2 reports the median and inter-quartile range of the responses to the six questions of the Raw-N S -TLX questionnaire. Wilcoxon signed-rank test ($z = 36$, $p =$. 14) 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 ($r =$ .845). So, the participants were generally less satisfied with their performance for triplets. 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.

The same ratings of the Raw-N S -TLX questionnaire have been analyzed to check if the subjective task load

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

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.)	n.s.
performance	1.0(8.3)	-5.5(3.5)	0.014 *
effort	2.0(2.8)	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-N S -TLX questionnaire, after having performed with duplets and with triplets.

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.

Wilcoxon signed-rank test shows a significant difference between the two halves for temporal demand and for frustration ($z = 21$, $p =$. 34) and for such questions the Wilcoxon effect size is large ($r =$ .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(5.8)	-5.0(4.5)	n.s.
temporal demand	-7.0(9.3)	-7.5(7.5)	0.034 *
performance	3.0(4.0)	-5.0(9.8)	n.s.
effort	2.0(1.8)	0.0(3.8)	n.s.
frustration	-5.0(11.)	-6.0(4.5)	0.034 *

Table 3. Median(IQR) of the ratings for each of the six questions of the Raw-N S -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 hypotheses listed in section 4.2.

Hypothesis HP1 is confirmed, as users with minimal training could effectively follow a target with a mean distance as low as 75.33 pixels (27.34 mm) for duplets and 76.52 pixels (27.77 mm) for triplets.

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 confirm hypothesis HP2.

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 (HP3). More practice can turn anxiety to engagement, and an increase in difficulty can turn boredom to engagement as well [31]. 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. better key-tap feedback, possibly accompanied by tactile stimulation, may reduce misses and rhythm detection faults as well.

5. CONCLUSIONS

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 are 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. n 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, and drawing trajectories by rhythm proved to be feasible and engaging.

The ability of users to effectively exploit the informative content of the rhythmic auditory display, overcoming possible occasional deprivation of visual feedback, has not been shown yet and is subject of further investigation. Moreover, the tactile rendering of rhythms is going to be considered for applications where auditory display is better supported or replaced by the more intimate sense of touch. Beyond control of a moving object by a single person, interaction by ticking and tacking may also be exploited in an inter-individual coordination perspective for joint action and performance, with applications in art, play, therapy, and training.

cknowledgments

The interaction design, system demonstration, and experiment design were done while lessio Bellino was visiting researcher at the University of Palermo. The presentation has been supported through the FFR 2023 fund of the University of Palermo.

6. REFERENCES

[1] J. D. Mc uley, “Tempo and rhythm,” in *Music Perception*, M. Riess Jones, R. R. Fay, and . N. Popper, Eds. New York, NY: Springer New York, 2010, pp. 165–199.

[2] E. Freeman, G. Wilson, D.-B. Vo, . Ng, I. Politis, and S. Brewster, “Multimodal feedback in HCI: Haptics, non-speech audio, and their applications,” in *Handbook of Multimodal-Multisensor Interfaces*, S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and . Krüger, Eds. ssociation for Computing Machinery and Morgan & Claypool, 2017, p. 277–317.

[3] S. dhitya, *Musical cities*. UCL Press, 2018.

[4] C. Erkut, . Jylhä, and D. Rocchesso, “Heigh ho: Rhythmicity in sonic interaction,” in *Sonic Interaction Design*, K. Franinović and S. Serafin, Eds. MIT press, 2013, pp. 341–350.

[5] D. Rocchesso, P. Polotti, and S. Delle Monache, “Designing continuous sonic interaction,” *International Journal of Design*, vol. 3, no. 3, 2009.

[6] 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.

[7] . Jylhä, I. Ekman, C. Erkut, and K. Tahiroğlu, “Design and evaluation of human-computer rhythmic interaction in a tutoring system,” *Computer Music Journal*, vol. 35, no. 2, pp. 36–48, 2011.

[8] E. Ghomi, G. Faure, S. Huot, O. Chapuis, and 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.

[9] E. Freeman, G. Griffiths, and S. . Brewster, “Rhythmic micro-gestures: Discreet interaction on-the-go,” in *Proceedings of the 19th CM International Conference on Multimodal Interaction*, 2017, p. 115–119.

[10] . Bellino, “Rhythmic-synchronization-based interaction: Effect of interfering auditory stimuli, age and gender on users’ performances,” *pplied Sciences*, vol. 12, no. 6, 2022.

[11] L. Velautham and R. C. S. Yoong, “Can’t clap to a beat? How rhythmically challenged people experience and strategize keeping time to music,” *Psychology of Music*, vol. 50, no. 4, pp. 1254–1266, 2022.

[12] 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.

[13] P. Bozzi and G. B. Vicario, “Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale,” *Rivista di psicologia*, vol. 54, no. 4, pp. 253–258, 1960.

[14] Y. Visell, F. Fontana, B. Giordano, R. Nordahl, S. Serafin, and R. Bresin, “Sound design and perception in walking interactions,” *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 947–959, 2009.

[15] H. Landin, S. Lundgren, and J. Prison, “The iron horse: sound ride,” in *Proceedings of the Second Nordic Conference on Human-Computer Interaction*, 2002, p. 303–306.

[16] . Gabrielsson, “The relationship between musical structure and perceived expression,” *The Oxford handbook of music psychology*, pp. 141–150, 2009.

[17] R. O. Gjerdingen, “ pparent Motion in Music?” *Music Perception*, vol. 11, no. 4, pp. 335–370, 1994.

[18] B. L. Giordano, H. Egermann, and R. Bresin, “The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity,” *PloS one*, vol. 9, no. 12, p. e115587, 2014.

[19] C. Frame, “Ding-dong: Meaningful musical interactions with minimal input,” in *Proceedings of the Sound and Music Computing Conference*, 2023.

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

[21] I. S. MacKenzie, “Interaction elements,” in *Human-computer Interaction*, I. S. MacKenzie, Ed. Boston: Morgan Kaufmann, 2013, pp. 71–120.

[22] C. T. nnand, F. M. Grover, P. L. Silva, J. G. Holden, and M. . Riley, “Early learning differences between intra- and interpersonal interlimb coordination,” *Human Movement Science*, vol. 73, p. 102682, 2020.

[23] D. F. Vanderelst and H. Peremans, “Computational analysis of the Etch- -Sketch task: comment on nnand et. al. 2020,” *Psy rXiv*, 2022. [Online]. vailable: https://psyarxiv.com/7vdhu/

[24] . S. Bregman, *uditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[25] ——, “ uditory scene analysis and the role of phenomenology in experimental psychology.” *Canadian Psychology/Psychologie canadienne*, vol. 46, no. 1, p. 32, 2005.

[26] M. Jokiniemi, R. Raisamo, J. Lylykangas, and V. Surakka, “Crossmodal rhythm perception,” in *Haptic and udio Interaction Design*, . Pirhonen and S. Brewster, Eds. Springer Berlin Heidelberg, 2008, pp. 111–119.

[27] C. Bernard, J. Monnoyer, M. Wiertlewski, and S. Ystad, “Rhythm perception is shared between audio and haptics,” *Scientific Reports*, vol. 12, no. 1, pp. 1–12, 2022.

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

[29] T. Pakkanen, R. Raisamo, and V. Surakka, “ udio-haptic car navigation interface with rhythmic tactons,” in *Haptics: Neuroscience, Devices, Modeling, and pplications*, M. uvray and C. Duriez, Eds. Springer Berlin Heidelberg, 2014, pp. 208–215.

[30] S. G. Hart, “Nasa-task load index (N S -TLX); 20 years later,” *Proceedings of the Human Factors and Ergonomics Society nnual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[31] G. Chanel, C. Rebetez, M. Bétrancourt, and 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.