Speed and Direction as two Overlapping Rhythms: Accuracy and Precision in Interpretation and Performance

Alessio Bellino*
Pontificia Universidad Católica de Chile
Santiago, Chile
abellino@ac.cl

Davide Rocchesso†
Università degli Studi di Milano
Milan, Italy
davide.rocchesso@unimi.it

Chiara Mascetti‡
Università degli Studi di Milano
Milan, Italy
chiara.mascetti@studenti.unimi.it

Giorgio Presti§
Università degli Studi di Milano
Milan, Italy
giorgio.presti@unimi.it

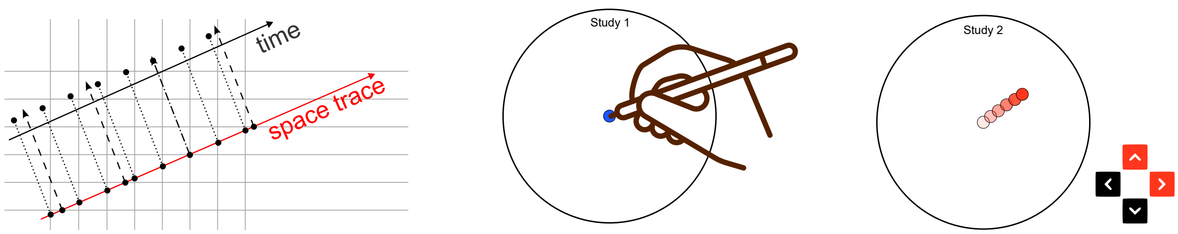

Figure 1: On the left, a physical metaphor: dragging a stick along a grating produces two overlapping rhythms. On the center, study 1 (Interpret): participants listen to two overlapping rhythms, for the vertical and the horizontal components of velocity, and move a blue dot towards the circumference at such velocity. On the right, study 2 (Perform): participants observe a red dot moving towards the circumference and produce two rhythms using the arrow keys, to rhythmically imitate the movement.

Abstract

Human and animal locomotion establishes a tight connection between movement in space and rhythm. Previous studies have explored the connection between rhythmic regularity and locomotor speed. In human-computer interaction, it has been shown that it is possible to control the trajectories of objects in two-dimensional space by producing rhythmic cells. The two studies here reported have the dual aim to investigate how well humans can (i) interpret overlapping rhythms as indications of direction and speed of a dragged moving object, and (ii) perform overlapping rhythms as corresponding to the speed and direction of a visual object. We assessed the accuracy and precision of these rhythmic trajectories using multisensory translation tasks.

CCS Concepts

• **Human-centered computing** → **Auditory feedback**; *Empirical studies in HCI*; • **Hardware** → *Sound-based input / output*.

*conception, study design, code writing, data analysis, report writing
†supervision, conception, study design, study execution, data analysis, report writing
‡study execution, data analysis, report writing
§supervision, report writing

ACM ISBN 979-8-4007-2065-9/25/06
https://doi.org/10.1145/3771594.3771639

Keywords

Multisensory interaction, Rhythmic interaction, Two-dimensional navigation

ACM Reference Format:
Alessio Bellino, Davide Rocchesso, Chiara Mascetti, and Giorgio Presti. 2025. Speed and Direction as two Overlapping Rhythms: Accuracy and Precision in Interpretation and Performance. In *20th International Audio Mostly Conference (AM '25), June 30–July 04, 2025, Coimbra, Portugal*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3771594.3771639

1 Introduction

The connection between rhythm and movement is a fundamental aspect of both human and animal locomotion. For instance, Laffi et al. quantified the rhythmic differences in the temporal structure of horses' walk, trot, and canter, promising to better explore the connection between rhythmic regularity and locomotor speed [4]. In the realm of human-computer interaction, previous research has demonstrated the feasibility of controlling object trajectories in two-dimensional space through the production of rhythmic cells [5]. Each cell is composed of duplets or triplets of pulses, with the Inter-Onset-Interval (IOI) inversely proportional to one of the two orthogonal components of velocity.

The movement of an object in space can be monitored solely by rhythm [5], or by a combination of rhythm, pitch, and spatialization [1]. In data sonification, both pitch and rhythm (or IOI) of sounds can be equally effective for auditorily graphing time-series data [3], and rhythm perception can also effectively occur in haptics [2].

This study aims to demonstrate that humans can interpret two overlapping rhythms and accurately reproduce the direction and speed of a moving object when each rhythm corresponds to one component of the velocity vector. Additionally, it seeks to show that humans can produce two overlapping rhythms that correspond to a visually perceived linear motion trajectory and to measure the accuracy and precision of perception and action of rhythmic trajectories through multisensory translation tasks. The rhythmic cells here considered are duplets that, when iterated, compose two overlapping regular sequences of pulses, one corresponding to the horizontal and the other to the vertical components of velocity. A physical interpretation may be that of dragging a stick at constant speed along a grating, as illustrated in Fig. 1 (left).

2 Experimental Study

A demonstration video of studies 1 and 2 is available at https://youtu.be/2NrITqbthks. The experimental design and procedures were reviewed and approved by the Ethics Committee of the University of Milan.

2.1 Participants

The study involved 20 participants (16 males, 4 females) with an average age of 22.9 years (SD = 2.3). Fourteen participants reported having musical experience, with an average of 10 years (SD = 5.0).

Participants were recruited through advertisements on the research lab's social media channels and internal communications within the Department of Computer Science at the University of Milan. Eligibility criteria required participants to be between 18 and 50 years old. While no specific experience was necessary, participants were asked to report their level of musical practice. Participation was entirely voluntary, with no compensation provided. Each participant attended a single session lasting approximately 30 minutes.

2.2 Apparatus

The experiment was conducted in a soundproof and quiet room. Each participant was seated in front of a tablet-display (Wacom One 13.3", 1920×1080, tilted position) and wore circumaural headphones (Beyerdynamic DT 770 Pro).

For Experiment 1, interaction was facilitated using a stylus on the tablet-display. For Experiment 2, a programmable mechanical mini-keyboard was used, with the directional arrows assigned to four keys arranged as ✣. Both input devices were connected to a computer (Apple iMac). The main display of the computer was visible to and controlled by the experimenter.

The experiment was run on a Chrome browser in full screen, showing a circle having a diameter of 400 pixels, corresponding to 6.13 cm at the chosen resolution. The machine- and human-controlled objects were respectively displayed as red and blue dots having a diameter of 20 pixels.

The experimenter was responsible for configuring all devices and preparing the required software in the web browser.

2.3 Stimuli

The auditory stimuli were made of repetitions of four sound samples, each indicating one of the four directions:

- (left) an isolated click of the recording of the electro-mechanical directional sound of a car[1], presented monaurally to the left ear. Linear-frequency spectral centroid: 3.8 kHz.
- (right) the same as (left), presented monaurally to the right ear;
- (down) the same as (left), transposed to have linear-frequency spectral centroid at 1.5 kHz, and diotically presented to both ears;
- (up) the same as (left), transposed to have linear-frequency spectral centroid at 10.0 kHz, and diotically presented to both ears.

Values for speed and angle were derived from values of IOI for the vertical and horizontal components, randomly generated with a uniform distribution in the range 200 ms − 1600 ms. Consequently, both the horizontal and vertical components of motion always had a corresponding sequence of pulses that is perceived as a rhythm. As the IOI could not approach zero nor infinity, any motion along the main axes was automatically excluded.

2.4 Experimental Procedure

Participants were asked to read and sign an informed consent form. The experimenter ensured that participants understood the consent form and answered any questions they had. Subsequently, participants provided personal data through a paper questionnaire, and the experimenter noted any additional comments made by the participants.

2.4.1 Training. Participants were instructed on how to produce rhythmic patterns to control the horizontal and vertical components of the velocity of a dot on the screen. They were required to produce forty trajectories by tapping on the four directional keys. The horizontal component was set by tapping twice on the left or right keys, while the vertical component was set by tapping twice on the up or down key. During this process, participants listened to a few repetitions of the duplet they produced for each component (horizontal and vertical) and observed the resulting movement of the dot. This was designed to help them learn the correspondence: faster rhythms (shorter IOIs) for a given component (e.g., horizontal) would result in faster movement along that axis, and the combination of the two rhythmic speeds would define the overall speed and direction of the dot.

This training phase lasted approximately ten minutes.

2.4.2 Study 1: Interpret Test. Participants listened to two overlapping rhythms and moved a dot from the center to the circumference of a circle using a pen (see Fig. 1center), based on the perceived direction and speed. Forty trials with random directions and speeds were conducted. This test phase lasted approximately ten minutes.

2.4.3 Study 2: Perform Test. Participants observed a dot moving from the center to the circumference of a circle and produced two rhythms corresponding to the horizontal and vertical components of this movement using directional cursor buttons (see Fig. 1, right). Forty trials with random directions and speeds were conducted. This test phase lasted approximately ten minutes.

[1] https://freesound.org/people/dersuperanton/sounds/434819/

2.5 Data Collection

For each test, we recorded both the target angle and speed, which were randomly generated by the software, as well as the angle and speed produced by the participant. Our goal is to assess the accuracy and precision in reproducing the target's speed and direction under two different stimulus conditions: auditory rhythm and visual motion. Specifically, we compare two conditions: *Interpret*, where participants respond to auditory rhythmic stimulation by dragging a target, and *Perform*, where participants respond to visual movement stimulation by tapping directional buttons.

2.6 Data Analysis

Given that some users' data deviate significantly from normality, as indicated by the Shapiro-Wilk test ($p < 0.05$), we will employ non-parametric tests. To determine the accuracy of a measure, we will use the median, while for precision, we will use the interquartile range (IQR).

Each participant completed 40 trials, which were aggregated into a single value per participant to facilitate statistical analysis and comparisons across conditions.

With a few exceptions, the participants recognized and produced motion in the same quadrant as the stimulus. Among the $20 \times 40 = 800$ trials, in study 1 we recorded 4 horizontal flips (left-right confusion), 32 vertical flips (up-down confusion), and 2 double flips; In study 2 we recorded 2 horizontal flips and 8 vertical flips. We removed all flips from the analysis.

Accuracy in both speed and angle was measured by computing the median error for each participant across their trials, capturing deviations of the response from the ideal stimulus values (stimulus−response). The overall accuracy was then determined as the median of the participants' median errors:

$$\text{accu} = \text{median}(\text{Q2}_{\text{users}})$$

A median significantly different from zero indicates a bias in angle or speed.

Precision (error variability) was quantified using the interquartile range (IQR). For each participant, we calculated the IQR of their errors using the first (Q1) and third (Q3) quartiles. To obtain an overall measure of precision, we computed the median Q1 and Q3 across all participants:

$$\text{prec} = \text{median}(\text{Q3}_{\text{users}}) - \text{median}(\text{Q1}_{\text{users}})$$

This approach captures both systematic deviations (accuracy) and variability (precision) in speed and angle errors.

3 Results

3.1 Speed analysis

We can definitively reject the null hypothesis that the median speed is zero for both *Interpret* and *Perform* studies (see Fig. 2), confirming the significance of the speed bias. The median speed for the *Interpret* test is −9.7 pixels/s and is significantly different from zero (Wilcoxon signed-rank test, $p < 0.01$), whereas for the *Perform* test it is 10.1 pixels/s and is also significantly different from zero (Wilcoxon signed-rank test, $p < 0.01$), providing a robust measure of accuracy. The median of Q1 values and the median of Q3 values across all participants are represented as lower and upper boundaries of the boxes in Fig. 2, and their difference gives a measure of precision.

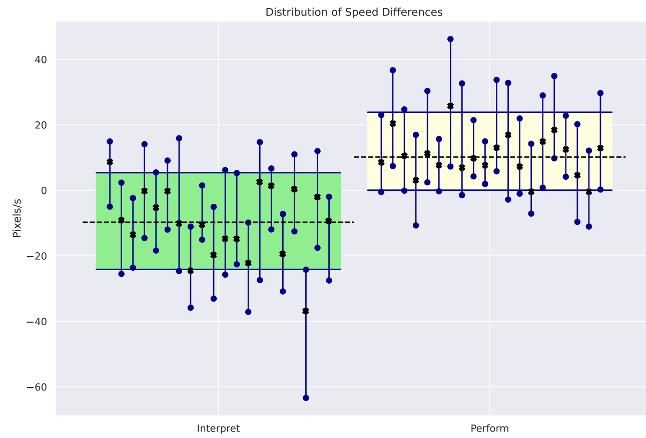

Figure 2: The boxplot displays the data for both the *Interpret* and *Perform* conditions, highlighting accuracy (deviation from zero of the dashed lines) and precision (height of the boxes). The vertically-aligned dots and star represent the quartiles and median for a single user, while the blue line indicates the corresponding IQR.

For the *Interpret* condition, the median of Q1 is −24.1, while for Q3 it is 5.3, resulting in an overall precision of 29.4 pixels/s. In the *Perform* condition, the median of Q1 is 0.0, and for Q3 it is 23.8, yielding an overall precision of 23.8 pixels/s. These precision measures can be understood as differential speed thresholds in the perception-action tasks for the *Interpret* and *Perform* conditions.

From the boxplots, it can be observed that the significant difference between the Interpret and Perform conditions is symmetric in terms of speed deviations from the ideal. In the Interpret condition, the median speed deviation was −9.7 pixels/s (indicating that participants produced a movement faster than the speed specified by the rhythmic auditory cue), while in the Perform condition, the median speed deviation was +10.1 pixels/s (indicating that participants produced a slower rhythm, as compared to the speed of the animated visual stimulus).

3.2 Angle analysis

We fail to reject the null hypothesis that the median angle is zero for both the *Interpret* and *Perform* conditions. The median angular errors were close to zero, with the *Interpret* condition showing a median of −0.10 and the *Perform* condition a median of 0.06. The Wilcoxon signed-rank test revealed no significant difference between the medians of the two conditions and zero ($p > 0.05$), indicating no systematic bias in the angular errors for either condition.

The first (Q1) and third (Q3) quartiles for all users are represented in a comprehensive boxplot (see Fig. 3), where each user's IQR is highlighted with a blue horizontal line linking Q1 and Q3, both for *Interpret* and *Perform* conditions. This visualization allows us to display the variability of IQR (or precision) across users. To quantify

Figure 3: The boxplot displays the data for both the *Interpret* and *Perform* conditions, highlighting accuracy (deviation from zero of the dashed lines) and precision (width of the boxes). The horizontally-aligned dots and star represent the quartiles and median for a single user, while the blue line indicates the corresponding IQR.

the general variability, we calculate the median of Q1 values and the median of Q3 values across all participants. These medians correspond to the left and right boundaries of the boxes in the boxplot, and the difference between the two medians gives us an overall measure of angular precision.

For the *Interpret* condition, the median of Q1 is −6.5, while for Q3 it is 7.3, resulting in a precision of 13.9 degrees. In the *Perform* condition, the median of Q1 is −7.3, and for Q3 it is 7.4, yielding a a precision of 14.8 degrees. Such measures of precision can be interpreted as differential thresholds of angle in the perception-action tasks of the *Interpret* and *Perform* conditions. To assess potential differences in angle between the *Interpret* and *Perform* conditions, a Wilcoxon signed-rank test was performed on the medians of Q1, revealing no significant difference ($p = 0.81$). A similar non-significant result was obtained for Q3 ($p = 0.28$).

4 Discussion and conclusions

After the training phase, participants had already learned to associate rhythm with speed, and the experiment tested how this association was applied in two different tasks. The results reveal a systematic bias in speed when transitioning between sensory modalities, despite prior learning. The symmetry in deviations suggests that the transformation of information between rhythm and movement is not perfectly balanced. In the *Interpret* condition (translating rhythm into movement), participants tended to move faster than expected (−9.7 pixels/s). Conversely, in the *Perform* condition (translating movement into rhythm), they produced a slower beat than the target speed (resulting in difference in speed of +10.1 pixels/s). This indicates that, even though participants had an explicit association between rhythm and speed, their speed reproduction was systematically biased depending on the direction of the transformation.

These findings suggest that even with training, the brain does not apply a perfectly symmetrical transformation between rhythm and speed across different sensory-motor channels. Instead, speed perception and control appear to be influenced by modality-specific constraints, reflecting differences in how timing information is processed and executed in auditory-motor versus visual-motor tasks.

Examining individual user data points in Fig. 2, this asymmetry does not occur systematically. For example, the medians for P4, P6, P14, P15, P17, and P19 (ordered from left to right) appear close to 0 in terms of speed bias in the *Interpret* condition. This indicates that while the bias is a clear overall trend, it is not consistent across all participants. Interestingly, in the *Perform* condition, only P4, P14, and P19 show medians close to 0. This suggests that these participants, whose median is close to 0 in the *Interpret* condition as well, consistently execute the tasks without a significant speed bias when transitioning between modalities.

For the angle, no significant bias is observed in any condition, suggesting that users generally distribute their errors around the target angle. However, this aspect requires further investigation, particularly to determine whether any bias emerges when users interpret or perform angles near the axes. Specifically, near the axes, the proportional difference between horizontal and vertical IOI is greater compared to when performing oblique 45-degree angles (independently of quadrant), where the difference is null, as the 45-degree angle is achieved when the IOIs are identical. Therefore, further analysis is needed to explore how users respond to angles farther from 45 degrees and how approaching angles near the axes influences their behavior.

More generally, it is essential to analyze potential interactions between speed and angle. Specifically, further analysis is needed to determine whether and to what extent speed bias (and precision) vary across different angles, particularly in angles near the axes.

Acknowledgments

The study was funded by the European Union - NextGenerationEU, Mission 4 Component 1 (CUP B53D23013150006).

References

[1] Alessio Bellino, Davide Rocchesso, Rosario Mulè, and Luca D'Arrigo Reitano. 2024. Multisensory trajectory control at one interaction point, with rhythm. In *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*. 399–404.

[2] Corentin Bernard, Jocelyn Monnoyer, Michaël Wiertlewski, and Sølvi Ystad. 2022. Rhythm perception is shared between audio and haptics. *Scientific Reports* 12, 1 (2022), 4188.

[3] L Guiotto Nai Fovino, Anita Zanella, and Massimo Grassi. 2024. Evaluation of the Effectiveness of Sonification for Time-series Data Exploration. *The Astronomical Journal* 167, 4 (2024), 150.

[4] Lia Laffi, Teresa Raimondi, Carola Ferrante, Eleonora Pagliara, Andrea Bertuglia, Elodie Floriane Briefer, Marco Gamba, and Andrea Ravignani. 2025. The rhythm of horse gaits. *Annals of the New York Academy of Sciences* 1543, 1 (2025), 86–93.

[5] Davide Rocchesso, Alessio Bellino, Gabriele Ferrara, and Antonino Perez. 2024. Spacetime trajectories as overlapping rhythms. *International Journal of Human-Computer Studies* 192 (2024), 103358.