



OPEN ACCESS

EDITED BY

Fang Jiang,
University of Nevada, Reno, United States

REVIEWED BY

Aaron Seitz,
Northeastern University, United States
Simon Andrew Whitton,
Cleveland Clinic, United States

*CORRESPONDENCE

Hamza Bayd
✉ hamza.bayd@mines-ales.fr

RECEIVED 18 March 2025

ACCEPTED 11 July 2025

PUBLISHED 30 July 2025

CITATION

Bayd H, Guyot P, Bardy B and Slangen P
(2025) Influence of rhythm features on
beat/movement synchronization using a
low-cost vision system.
Front. Comput. Sci. 7:1595939.
doi: 10.3389/fcomp.2025.1595939

COPYRIGHT

© 2025 Bayd, Guyot, Bardy and Slangen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Influence of rhythm features on beat/movement synchronization using a low-cost vision system

Hamza Bayd^{1*}, Patrice Guyot¹, Benoît Bardy² and Pierre Slangen¹

¹EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France, ²EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Montpellier, France

This study examines how musical expertise, tempo, and beat division influence synchronization accuracy and regularity in two movement tasks: finger tapping (discrete movements) and arm swing (continuous movements). Using a markerless motion capture system, we analyzed synchronization metrics across different rhythmic conditions. Motion data were extracted via AI-based pose estimation, and synchronization was computed by aligning movement peaks with beat times detected from audio stimuli. Results show that musicians exhibit higher synchronization accuracy and consistency than non-musicians, particularly in finger tapping tasks. Furthermore, simpler beat structures (binary rhythms) and moderate tempos facilitate better synchronization, whereas increased rhythmic complexity and tempo variability reduce performance. Interestingly, finger tapping leads to more precise synchronization than arm swing, suggesting that movement type significantly impacts rhythmic alignment. These findings support applications in therapy, training, and interactive systems, and demonstrate the value of AI-based motion tracking for scalable rhythm analysis.

KEYWORDS

sensorimotor synchronization, music information retrieval, beat tracking, multi-scale rhythmic, human pose tracking, motion analysis, motor coordination

1 Introduction

Why do humans instinctively tap their feet, nod their heads, or synchronize their body movements to music? This natural ability, known as sensorimotor synchronization (SMS), is essential in dancing, clapping, and music performance. SMS is a well-studied phenomenon that has attracted attention across cognitive science, psychology, neuroscience, musicology, and movement science (Repp, 2005; Large et al., 2023).

Understanding rhythm synchronization has practical implications in both artistic and athletic settings. Accurate synchronization ensures cohesive ensemble performance in music and enhances coordination in sports such as rowing or synchronized swimming (Large and Jones, 1999; Schiphof-Godart and Hettinga, 2017). SMS also has important therapeutic applications, including rhythmic auditory stimulation (RAS), which has been shown to improve motor coordination in individuals with Parkinson's disease and support post-stroke rehabilitation (Cochen De Cock et al., 2021; Thaut and Abiru, 2010).

In human-computer interaction, SMS supports more immersive and responsive systems, such as music-based gaming and interactive virtual environments (Barbancho et al., 2013).

Traditionally, SMS studies have relied on simplified paradigms—often involving finger tapping with isochronous tones or metronomes—to ensure strong experimental control. These paradigms have yielded robust findings on synchronization accuracy and regularity,

such as negative mean asynchrony (NMA), tempo sensitivity, and timing variability (Repp, 2005; Repp and Su, 2013; Aschersleben, 2002). For example slower tempi, which are typically defined by a low number of beats per minute (BPM), generally improve synchronization, while faster ones increase variability. Finger tapping has consistently been shown to yield more precise synchronization compared to gross motor tasks like arm swing (Chen et al., 2008).

However, many prior studies focus on simple, metronomic rhythms (Repp, 2005), which may not fully reflect the complexity of real-world musical contexts. While sensorimotor synchronization (SMS) is often investigated through finger tapping paradigms, the daily interactions with music typically involve more complex and continuous movements, making it difficult to generalize findings from tapping tasks to naturalistic settings. Additionally, while finger tapping tasks are usually inexpensive and easy to implement, the study of full-body or naturalistic movement typically requires high-cost equipment (e.g., 3D motion capture systems) and time-consuming processing pipelines. These factors can limit ecological validity and scalability.

To address these challenges, we adopt a more naturalistic paradigm using complex audio stimuli (real music) and motion capture. This approach allows us to investigate whether the findings of traditional SMS research hold in more musically and physically realistic settings. We also reflect on how genre, instrumentation, and familiarity may influence synchronization performance—factors often overlooked in controlled laboratory studies.

From a technical point of view, recent developments in artificial intelligence (AI), computer vision, and signal processing offer promising alternatives to traditional setups. Tools like Madmom, Librosa, and Essentia enable beat and tempo detection from audio signals using supervised learning (Bock et al., 2016; McFee et al., 2015), while deep learning-based pose estimation frameworks (e.g., MediaPipe, PoseEstNet) now enable accurate human motion tracking using standard cameras (Mao et al., 2022; Panteleris and Argyros, 2022).

Finally, this study leverages these tools to analyze SMS in an ecologically valid yet scalable framework. Our pipeline combines neural network-based audio beat extraction, video-based motion tracking, and synchronization scoring metrics to assess how individuals align with musical beats across tasks, tempi, and rhythmic structures.

2 Materials and methods

2.1 Participants

The study involved 24 participants (13 females, 11 males; $M_{\text{age}} = 22.1$ years). Participants were divided into two groups based on musical expertise: the musician group (4 females, 4 males; $M_{\text{age}} = 21$) and the non-musician group (9 females, 7 males; $M_{\text{age}} = 23$).

All participants were students or staff members from IMT Mines Alès, France, with a Western cultural background. Inclusion criteria required participants to be between 18 and 50 years old, with no reported hearing impairments or motor disorders that could affect rhythmic movement. Participants were recruited

through internal communication and volunteered without financial compensation. Each participant provided informed consent prior to the experiment.

Musicians were recruited from the school's orchestra, a student- and staff-run ensemble composed of individuals who regularly play and rehearse music. All musicians had a minimum of two years of formal musical training. Non-musicians were individuals who reported no formal music education, had never attended a music school or conservatory, and did not actively play any musical instrument. Although group assignment was based on self-report, we ensured that participants in the non-musician group clearly lacked musical background. Musicians were also asked to report the instrument they played; the most frequently cited instruments were percussion, guitar, and piano. Information about participants' prior dance or movement training was not collected. Similarly, no standardized tools such as the Goldsmiths Musical Sophistication Index (MSI) or Ollen MSI were used for group classification.

2.2 General procedure and experimental design

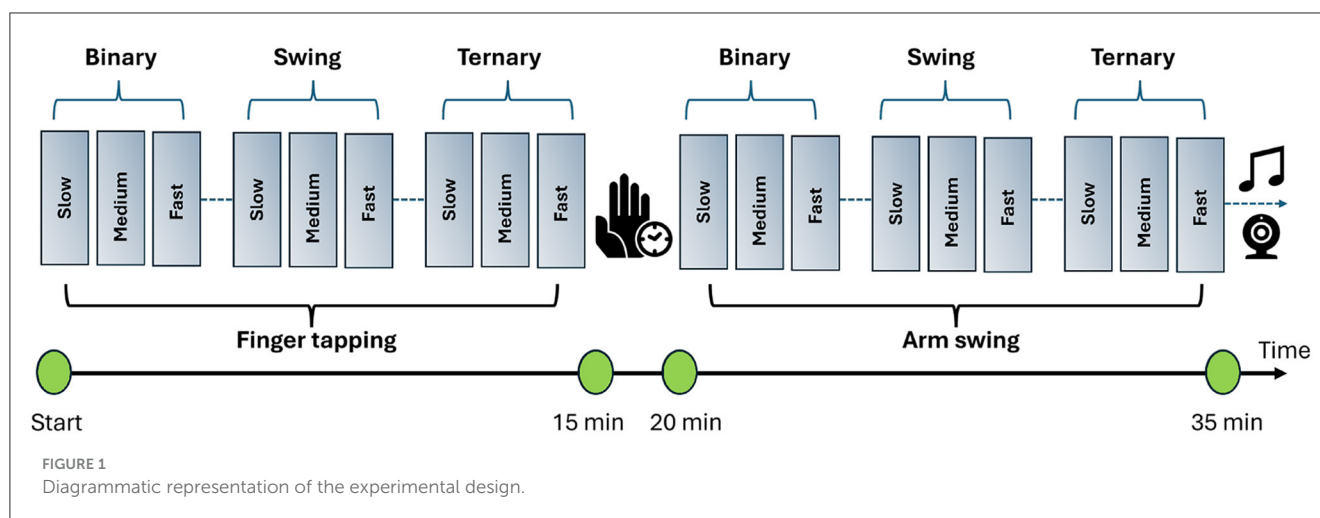
This study aimed to evaluate the impact of tempo and multi-scale beat division on participants' ability to synchronize their movements—specifically, finger tapping and arm swing—with musical rhythms. Additionally, it examined the influence of movement type and musical expertise on synchronization ability.

Finger tapping and arm swing were chosen for their distinct motor properties. Finger tapping is a fine-motor task involving precise, repetitive movements with tactile feedback, making it suitable for analyzing synchronization accuracy. In contrast, arm swing involves broader, continuous movements requiring coordination across multiple degrees of freedom, thus presenting a greater challenge. Including both movement types enabled a broader assessment of synchronization across fine and gross motor control.

Participants were given task-specific instructions. In the tapping condition, they were instructed to tap their thumb and index finger together in time with the beat, as precisely as possible. In the arm swing condition, they were asked to swing their right arm—which is typically the dominant arm—left to right in synchrony with the beat, matching each full swing with one beat. Importantly, left-handed participants were not excluded from the study. Participants were explicitly instructed to follow the beat and not to move freely or interpretively.

Each trial began with a 3-s introductory segment during which participants heard the music but were not recorded. This segment allowed them to identify the beat and begin moving in synchrony when the recording phase started.

Each session consisted of 18 trials, evenly split between the two tasks: nine trials of finger tapping and 9 of arm swing. The trials followed a fixed order—tapping first, then finger tapping and arm swing were presented in blocks grouped by beat division: first the three binary tracks, then the three swing tracks, and finally the three ternary tracks. Participants were not informed about the beat division or tempo beforehand, and none of the tracks were previewed prior to the experiment. Each trial lasted 30 seconds,



with short breaks between blocks to reduce fatigue and ensure performance consistency (see Figure 1).

We used a Logitech Meetup webcam, which recorded at $1,920 \times 1,080$ resolution and 30 frames per second (fps). During the finger tapping task, participants were seated 2 meters from the webcam, focusing on their hands and forearms. For the arm swing task, participants stood approximately 3 meters from the webcam to ensure full upper-body visibility. All participants used their preferred hand or arm for both tasks. Audio was delivered via headphones at 75 dB SPL to ensure consistent sound perception. The experiment was conducted in a well-lit room with a black-and-white background to improve video processing. Figure 2 illustrates the physical setup.

2.3 Multimedia synchronization and recording software

To capture and analyze movement-to-beat synchronization, we developed a Python-based system that integrates audio-video recording and playback. The software, called SYNC Beat M4S, was built using OpenCV, Pygame, GStreamer, and Tkinter, providing an interactive interface for experiment control.

The software provides an interface, allowing control of various aspects of the experiment, such as task selection, beat division, rhythm music, tempo ranges, and specific music tracks, as shown in Figure 3.

Each trial began when the experimenter pressed “Play”, triggering video recording. The musical stimulus started precisely on the 200th video frame, ensuring audio-video alignment. This mechanism functioned as a clapperboard reference in film production. After each trial, the system automatically saved and sorted recordings by participant ID and task, enabling efficient and reproducible data management.

2.4 Auditory stimuli and beat structures

The musical stimuli were selected from the GTZAN Rhythm Corpus (Marchand et al., 2015), a dataset commonly used in

rhythm perception studies. We selected tracks to compare the effects of tempo and rhythmic complexity on synchronization. Each of the nine tracks represented a unique combination of tempo (slow = 80 BPM, medium = 100 BPM, fast = 120 BPM) and beat division structure (binary, swing, or ternary). The corresponding tracks are listed in Table 1, including the song title, artist, BPM, and time signature (TS). Notably, binary and swing tracks were mostly in 4/4 or 3/4 meters, while ternary stimuli used compound time signatures like 9/8 or 12/8.

Musical rhythms were categorized into three beat structures. Binary division involves dividing each beat into two equal parts. This structure dominates most Western music and supports strong perceptual anchoring. Ternary division, by contrast, divides each beat into three equal parts and appears in compound meters such as 6/8, 9/8, and 12/8. This structure allows greater rhythmic flexibility but can challenge synchronization. Swing represents a stylized variation of binary subdivision, where the first note in a pair is lengthened and the second shortened. This timing asymmetry gives jazz and blues their characteristic “swing feel”.

Figure 4 offers a visual comparison of these rhythmic categories.

Finally, the tracks were also selected to avoid globally recognizable tracks and aimed to balance cultural diversity with rhythmic clarity. Participants were not shown or informed about the audio material prior to the experiment.

3 Data analysis

Audio and video data were collected for each participant using our SYNC Beat M4S software, enabling precise temporal synchronization measurements. The analysis process, illustrated in Figure 5, follows a structured pipeline divided into three interconnected blocks. The first block extracts musical rhythmic features from the audio signal, primarily focusing on beat and tempo detection. These elements define the temporal structure necessary for synchronization analysis. In the human movement feature extraction block, video recordings are analyzed to track body movements over time using computer vision techniques. This process identifies motion beats and generates a movement



FIGURE 2

Illustrations of the experimental setup for both task modalities. Here the participant is performing the arm swing task.

vector, capturing both temporal and spatial characteristics for subsequent analysis. Finally, in the synchronization analysis block, the extracted music and movement features are compared as discrete time series to compute synchronization scores. This block provides a comprehensive assessment of participants' synchronization performance. The following sections detail each step of this process.

3.1 Audio beat detection

To ensure that the selected beat detection method provided reliable ground truth for synchronization scoring, we conducted a comparative benchmarking analysis of several commonly used algorithms, including Librosa (McFee et al., 2015), Essentia (Bogdanov et al., 2013), and Madmom (Bock et al., 2016). Each algorithm was evaluated on a custom corpus of 500 audio clips

spanning four musical genres (disco, hip-hop, blues, and classical), using beat annotations from the GTZAN (Marchand et al., 2015) dataset as ground truth.

Madmom consistently achieved the highest alignment with the annotated beat positions across genres and tempo ranges. As a result, it was chosen for integration into our synchronization analysis pipeline.

Madmom leverages recurrent neural networks (RNNs) trained on annotated audio datasets to predict beat positions with high temporal precision. The model processes spectral features and outputs a beat activation function, assigning to each audio frame a probability of containing a beat (Figure 6).

Beat positions and tempo estimates were extracted for each of the nine musical stimuli used in the experiment and stored for synchronization scoring. Beat division categories (binary, swing, ternary) were determined using GTZAN rhythm annotations to ensure accurate classification and track selection.

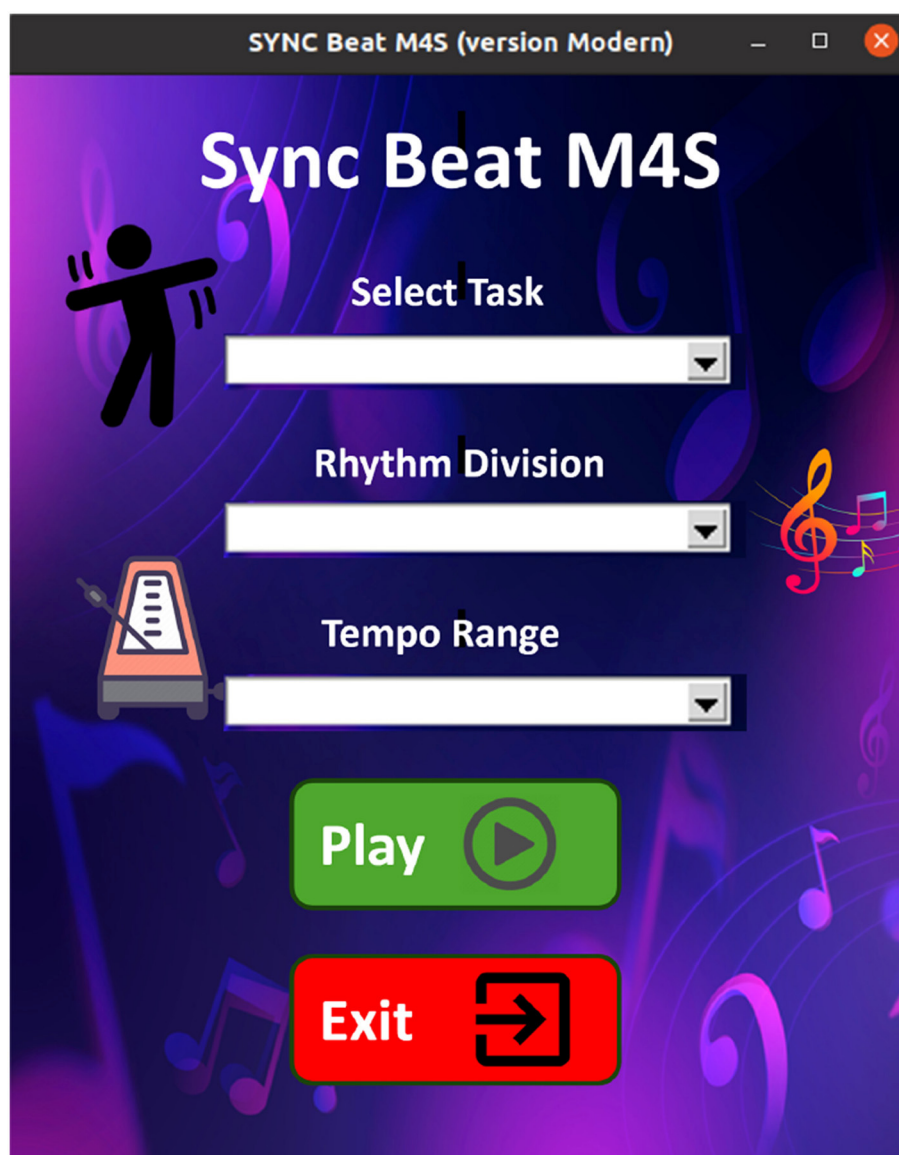


FIGURE 3
User interface of the “SYNC Beat M4S” software.

3.2 Motion capture and motion beat estimation

3.2.1 Motion capture

Recent advances in deep learning have enabled markerless motion capture, removing the need for traditional systems that require physical markers. Using only a standard webcam, state-of-the-art frameworks such as MediaPipe (Zhang et al., 2020) now allow for real-time and highly accurate tracking of human movement.

To validate the accuracy and reliability of MediaPipe for synchronization analysis, we conducted a separate validation experiment comparing its outputs to a high-precision optical motion capture system (Qualisys), which is considered a gold standard in biomechanical research. Both systems simultaneously

recorded the same set of movements under controlled conditions. The comparison confirmed that MediaPipe provides sufficiently accurate temporal and spatial data to be used in sensorimotor synchronization (SMS) studies, particularly for large-scale and naturalistic experimental designs where marker-based systems may be impractical or cost-prohibitive.

MediaPipe was used to extract kinematic data from video recordings captured at 30 fps. The pose estimation process followed a two-stage pipeline. First, a detector module localized regions of interest containing hands or body parts. Then, a convolutional neural network (CNN)-based estimator predicted two-dimensional (x, y) keypoints for each relevant anatomical feature. This model, trained on heatmaps and offset regressions, provides dense keypoint information from monocular video input.

MediaPipe outputs landmark coordinates that represent key points on the human body. For hand tracking, the system identifies 21 keypoints per hand, enabling precise analysis of finger tapping movements. For whole-body tracking, the model provides 33 keypoints, which are essential for analyzing arm swing trajectories.

TABLE 1 Musical tracks used in the experiment, organized by beat division, beat per minute (BPM), and time signature (TS).

Beat division	Song title	Artist	BPM	TS
Binary	Back off Boogaloo	Donna summer	89.46	4/4
	Heartless	Evelyn Thomas	104.94	4/4
	Sing a song	Earth, Wind and fire	122.00	4/4
Swing	Pocket full of gold	Vince Gill	81.84	3/4
	I've got the world on a string	Coleman Hawkins	106.61	4/4
	Is this love	Bob Marley	122.25	4/4
Ternary	Piano Sonata No. 21 in B flat	Franz Schubert	83.17	9/8
	Lonesome trouble	Magic Slim and The Teardrops	105.74	12/8
	Symphony No. 35 in B flat major	Franz Joseph Haydn	129.00	9/8

After extracting the keypoints, a series of custom Python scripts processed the data to detect motion beats. These motion beats correspond to timestamps where body movement aligns with the musical beat, enabling an objective evaluation of synchronization performance.

Figure 7 illustrates the full architecture of the video capture and processing pipeline used in this study.

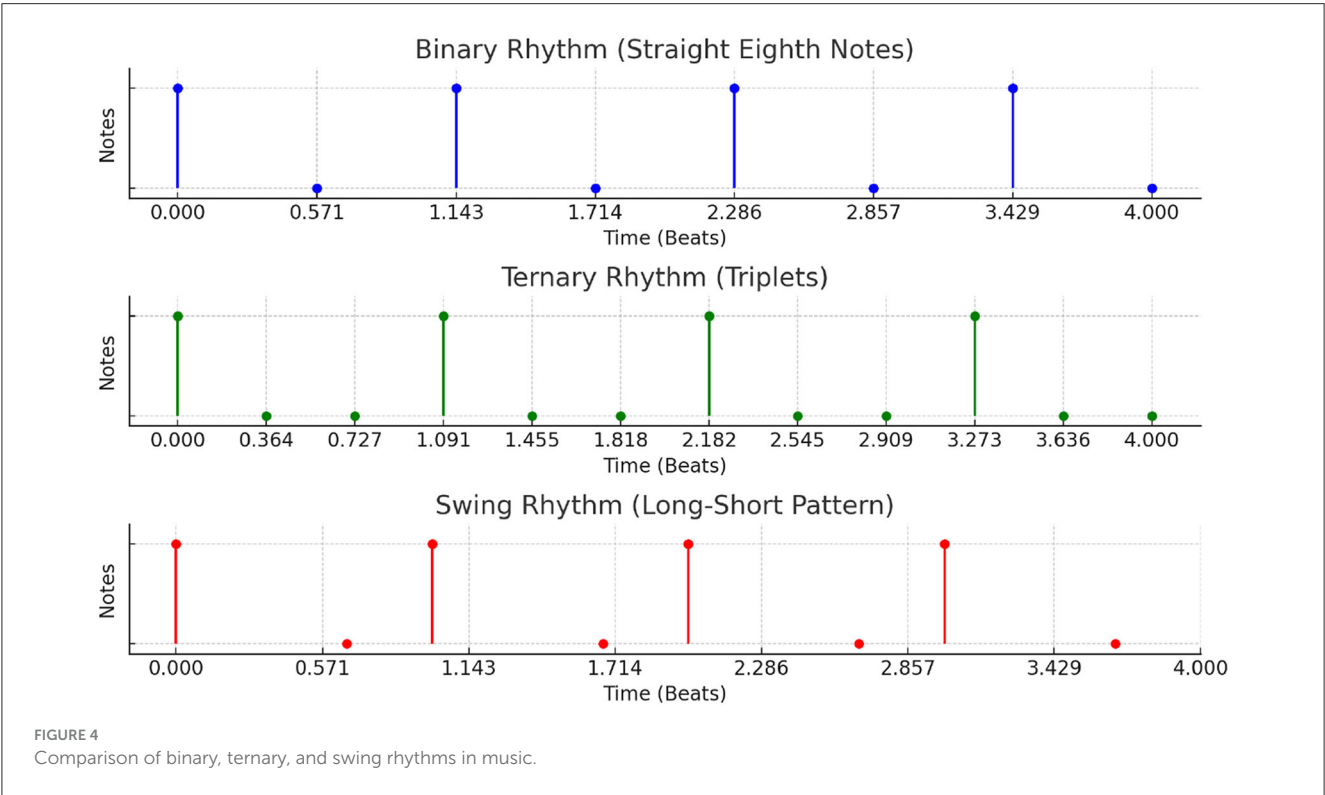
3.2.2 Finger tapping

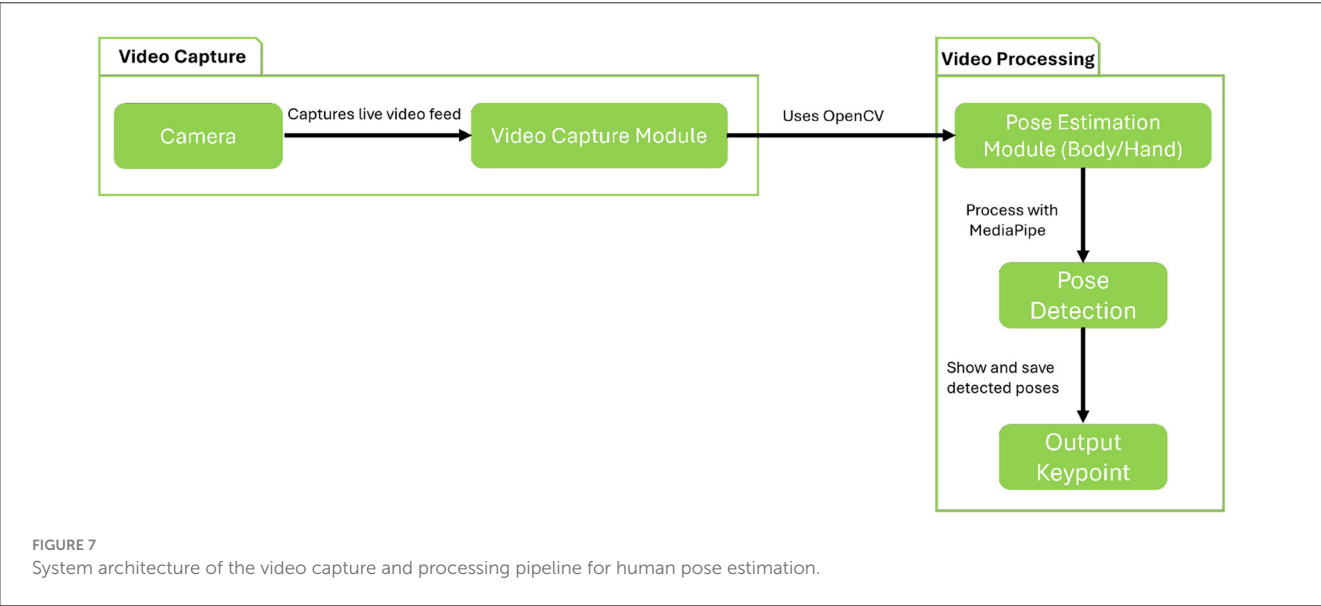
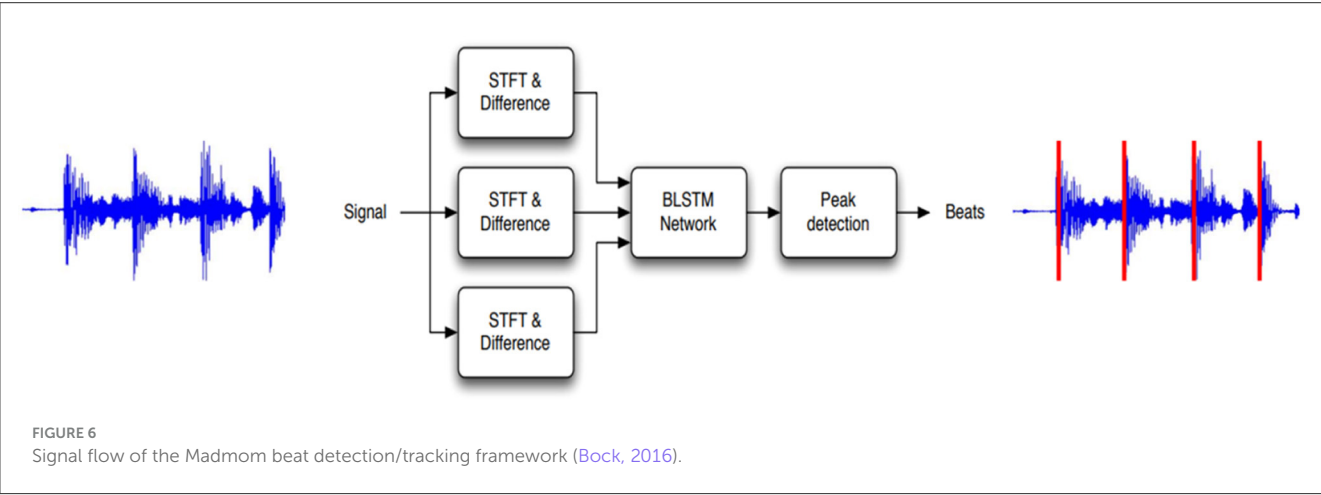
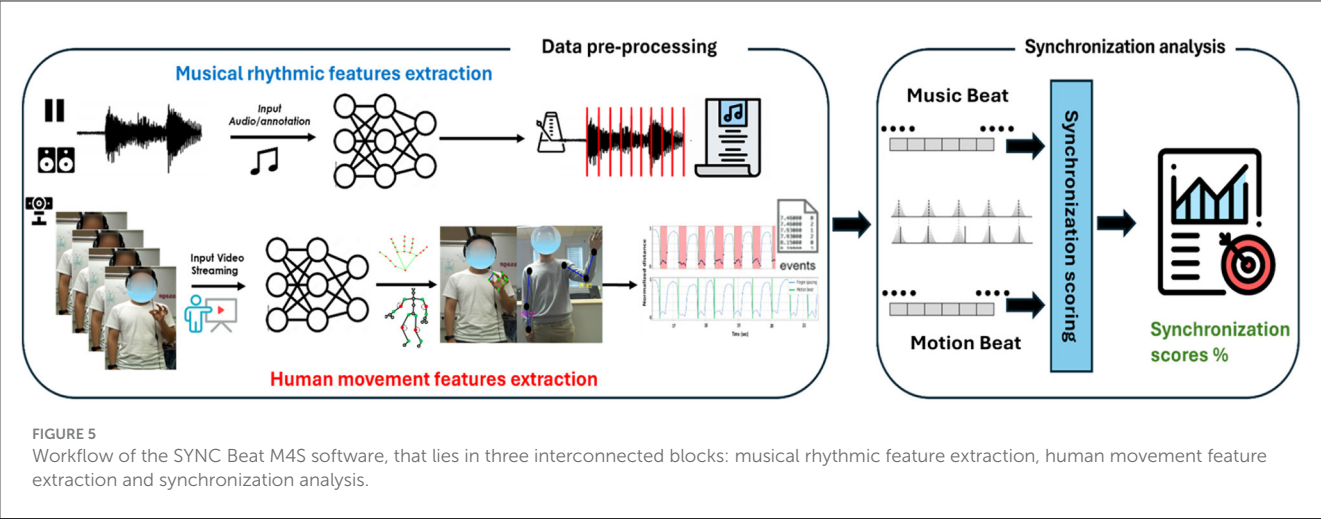
In the finger tapping task, motion beats were identified by measuring the Euclidean distance between the thumb and index finger. A beat was detected when this distance approached zero, indicating physical contact between the fingers. The detection process involved two main steps. First, the system identified the contact time by locating local minima in the distance signal. Then, a peak-finding algorithm was used to automatically extract candidate motion beats from the time series data.

Figure 8 presents the analysis of finger spacing and the corresponding motion beats, with visual annotations highlighting keyframes that mark detected contacts.

3.2.3 Arm swing

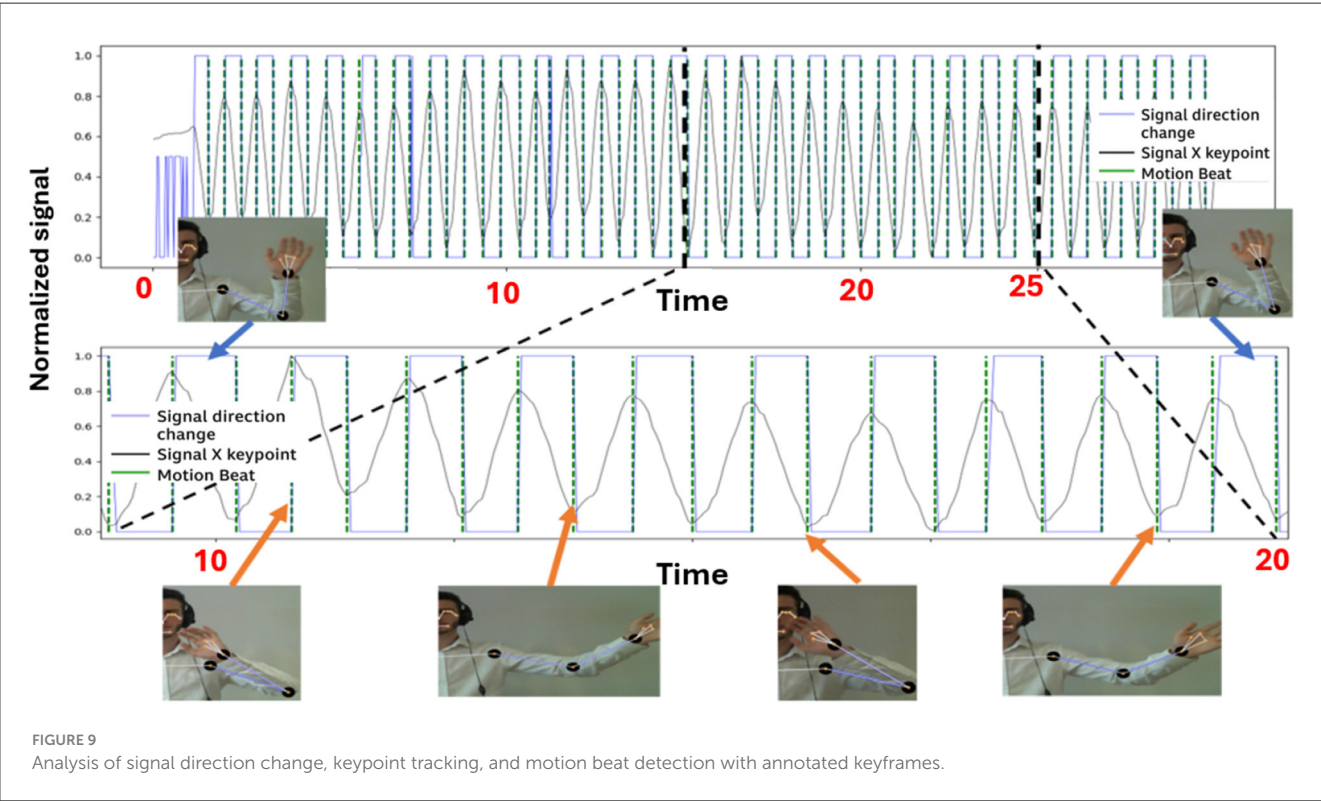
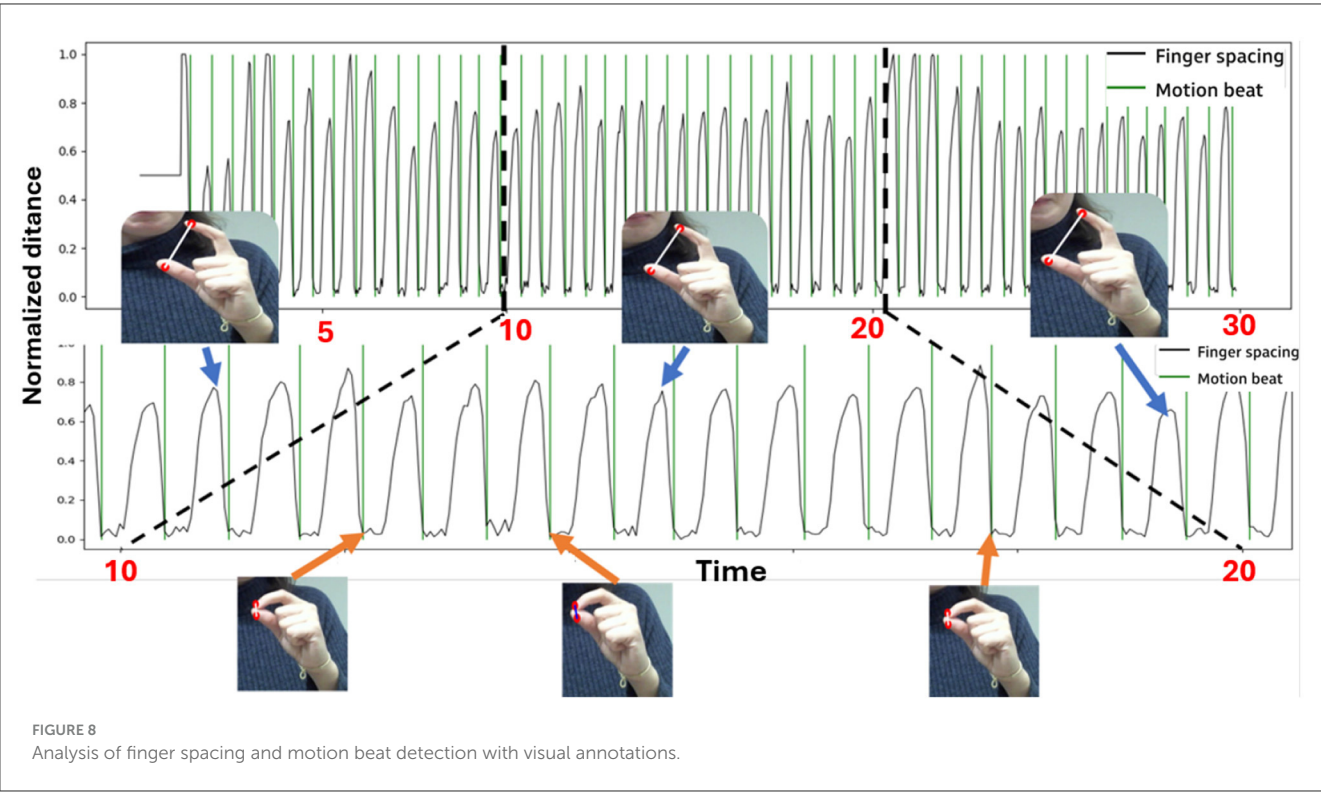
In the arm swing task, motion beats were detected by tracking changes in the direction of wrist movement. Each reversal in movement direction was treated as a beat. To achieve this, a peak detection algorithm monitored consecutive frames for stable changes in movement direction, marking the transition points.





The frame index just before each directional change was recorded as a motion beat candidate. Finally, the extracted motion beats were aligned with the corresponding musical beats to evaluate synchronization performance.

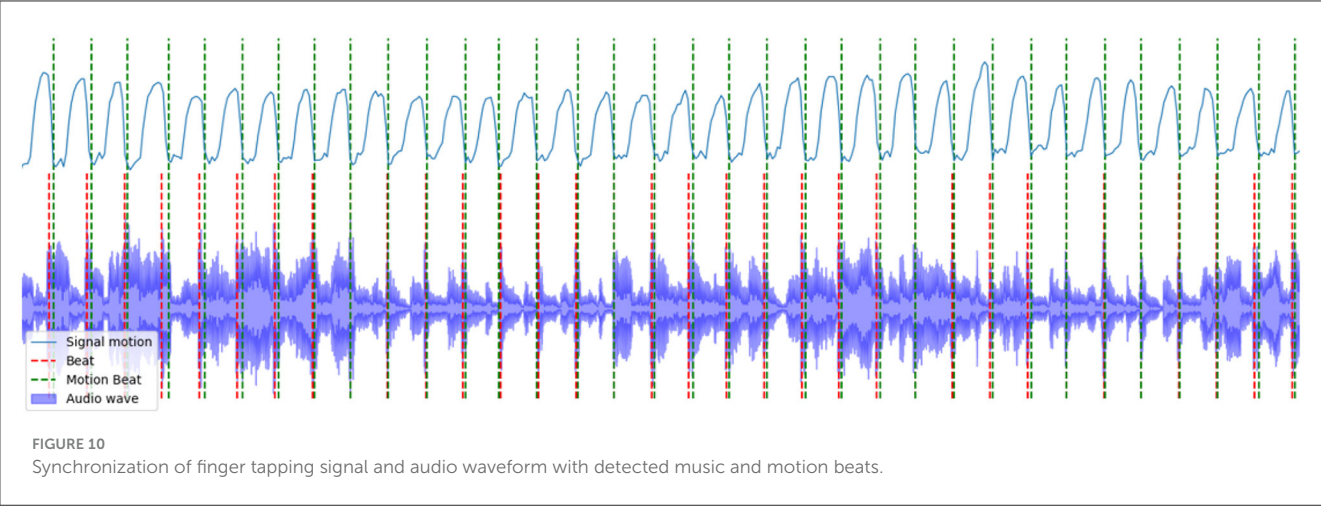
Figure 9 illustrates the wrist’s x-coordinate trajectory (black line), showing directional changes (in blue) and the identified motion beats (in green), which correspond to the reversal points in the arm’s swing cycle.



3.3 Scoring synchronization

Synchronization between participants' movements and musical beats was assessed by aligning the detected motion beat positions

with the corresponding musical beat positions. The analysis was restricted to the interval between 3 and 28 seconds of each trial. This aimed to avoid transient effects at the beginning and end of the audio, and to provide participants a short adaptation period before



their synchronization performance was evaluated and ensured that the extracted measures reflected more stable and representative movement patterns.

The number of beats analyzed varied depending on the tempo condition. For instance, fast-tempo stimuli yielded more beat events than slow-tempo stimuli. However, synchronization scores were computed as normalized metrics, ensuring that the results remain comparable across tempo conditions. An example of beat alignment during tapping tasks is shown in Figure 10.

Two primary metrics were used to quantify synchronization Bayd et al. (2024). The Gaussian Score measured synchronization accuracy by computing the temporal difference between motion beats and musical beats. This score applied a Gaussian function, where the parameter σ controlled the tolerance for deviation – allowing the evaluation to be more or less strict depending on the selected value. Higher scores indicated closer alignment to the target beats.

In addition, Mean Vector Length (R) was used to assess synchronization regularity. This measure was based on the relative phase between motion and beat, capturing the consistency of movement timing. Higher R values indicated that the participant maintained a more stable rhythm over time.

For each participant, synchronization scores were compiled along with task type, tempo, beat division, and musical expertise. This structured dataset enabled a detailed evaluation of performance across experimental conditions and provided insights into how tempo, rhythm structure, and prior musical training influence sensorimotor synchronization with music.

4 Results

To ensure the reliability of the analysis, five participants were excluded based on predefined performance criteria. Specifically, participants with a synchronization score below 0.3 on more than five trials were removed from the dataset. Among these, one belonged to the musician group and four to the non-musician group. This exclusion aimed to retain only participants who demonstrated a minimum level of engagement with the task and produced usable data for analysis.

TABLE 2 Summary of significant effects from the ANOVA model testing the influence of musical expertise, task type, tempo, and beat division on synchronization performance.

Source	Sum Sq	Df	F	PR(> F)
Expertise	0.97	1	24.39	1.44e-06 < 0.001
Task	0.81	1	20.26	1.04e-05 < 0.001
Tempo	1.95	2	24.48	1.96e-10 < 0.001
Beat division	0.64	2	8.00	4.28e-04 < 0.001
Expertise × task	0.41	1	10.38	1.50e-03 < 0.001

Only effects with $p < 0.01$ are reported in this main table.

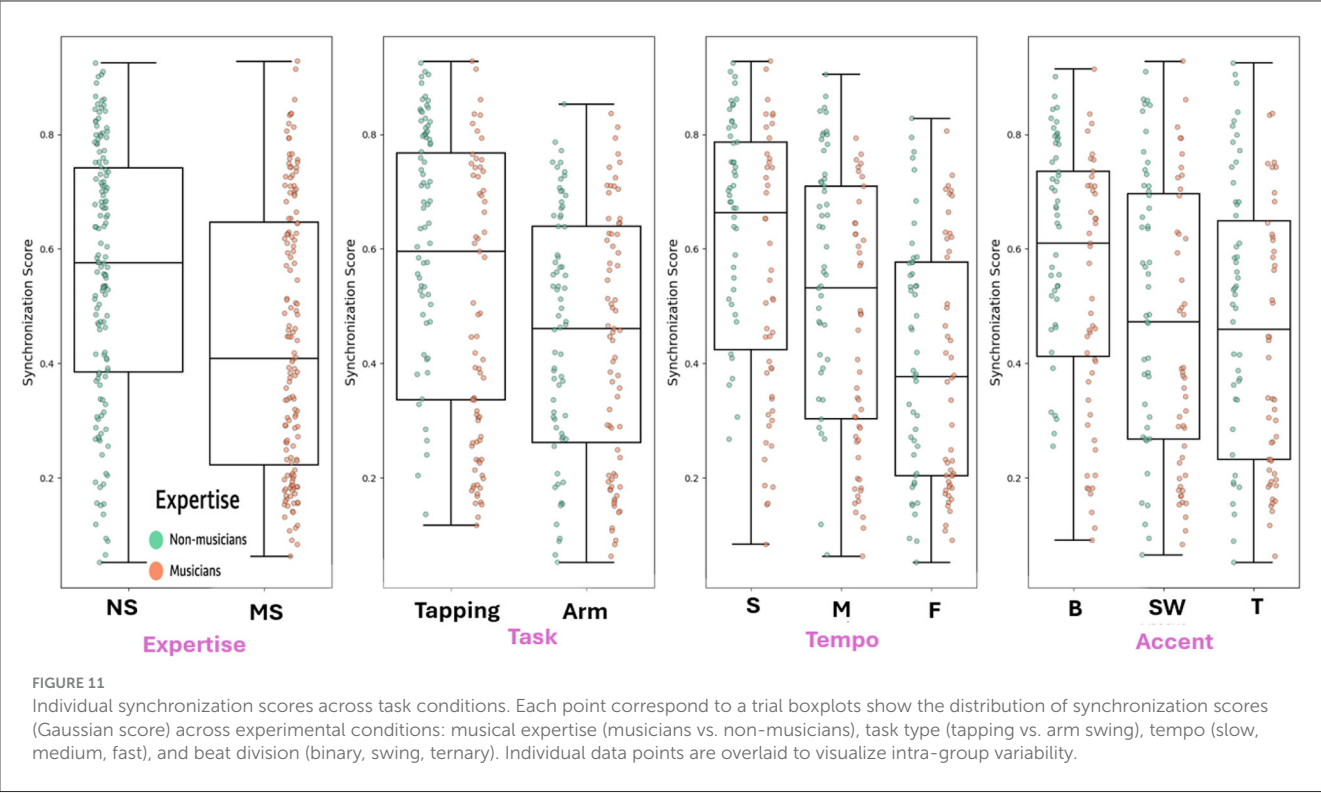
We tested whether the data met the assumptions for parametric analysis. The normality of the dependent variables was assessed using the *Shapiro – Wilkstat* = 0.98. This test confirmed that the data met the assumptions required for conducting an ANOVA.

A single, comprehensive ANOVA was conducted to assess the effects of musical expertise (musicians vs. non-musicians), task type (tapping vs. arm swing), tempo (slow, medium, fast), and beat division (binary, swing, ternary) on synchronization performance.¹ The main findings are reported in the reduced ANOVA Table 2, which includes only significant main effects (Expertise, Task, Tempo, Beat Division) and interactions (Expertise × Task).

Figure 11 presents corresponding boxplots with overlaid individual data points for each condition. These visualizations allow direct comparison between musicians and non-musicians and highlight variability across tasks and rhythmic conditions.

Expertise had a significant impact on synchronization performance, $F_{(1,250)} = 24.39, p < 0.001$, indicating that musicians consistently outperformed non-musicians across both tasks. The main effect of task was also significant, $F_{(1,250)} = 20.26, p < 0.001$, showing that the type of movement (tapping vs. arm swing) influenced synchronization accuracy. Participants generally

1 ANOVA: Score ~ Expertise × Tempo × Beat Division × Task.



performed better in the tapping task, as reflected in higher Gaussian scores.

Tempo had a substantial effect as well, $F_{(2,250)} = 24.48$, $p < 0.001$, with synchronization accuracy declining as the tempo increased from slow to fast. The beat division factor was significant, $F_{(2,250)} = 8.00$, $p < 0.01$, with synchronization accuracy decreasing in the following order: binary, swing, and ternary. This shows a clear preference for binary division in terms of better synchronization outcomes. These effects were consistent across both tasks and expertise levels, as shown by the absence of significant interactions involving tempo and beat division with other factors.

Interestingly, the interaction between expertise and task was significant, $F_{(1,250)} = 10.31$, $p < 0.01$, indicating that the advantage of musicians over non-musicians was more pronounced in one task compared to the other. The boxplots in Figure 11 suggest that this difference was particularly evident in the tapping task, where musicians showed higher synchronization accuracy and regularity, as indicated by both the Gaussian score and mean vector length.

4.1 Post-hoc analysis of expertise and task interaction

To further explore and clarify the interaction between expertise and task, a *post-hoc* Tukey HSD test was conducted. In the remainder of this section, we refer to non-musicians as NS and musicians as SM.

The results revealed significant differences specifically between the tapping task for musicians (SM-Tapping) and non-musicians (NS-Tapping), with a mean difference of 0.2261($p < 0.01$),

TABLE 3 Post-hoc Tukey HSD test results for expertise x task interaction, showing mean differences, adjusted p-values (p-adj), and whether the null hypothesis was rejected (reject).

Group 1	Group 2	Mean	p-adj	Reject
NS_Arm	NS_Tapping	0.0336	7.95e-01	False
NS_Arm	SM_Arm	0.0382	7.26e-01	False
NS_Arm	SM_Tapping	0.2261	1.00e-04	True
NS_Tapping	SM_Arm	0.0046	9.99e-01	False
NS_Tapping	SM_Tapping	0.1924	1.00e-04	True
SM_Arm	SM_Tapping	0.1879	1.00e-04	True

confirming that musicians performed significantly better in the tapping task. Additionally, significant differences were observed between SM-Tapping and SM-Arm (mean difference = 0.1924, $p < 0.01$), as well as between SM-Tapping and NS-Arm (mean difference = 0.1879, $p < 0.01$). These findings are summarized in Table 3, which shows the mean differences, adjusted p-values, and whether the null hypothesis was rejected.

The *post-hoc* analysis did not find significant differences between musicians and non-musicians in the arm swing task nor between the arm swing tasks of both groups. This indicates that the advantage of musical expertise is more task-dependent and most pronounced in the tapping task.

These results highlight the importance of task type in moderating the effect of musical expertise on synchronization performance. The tapping task, which requires fine motor control, may better leverage the rhythmic skills developed through musical

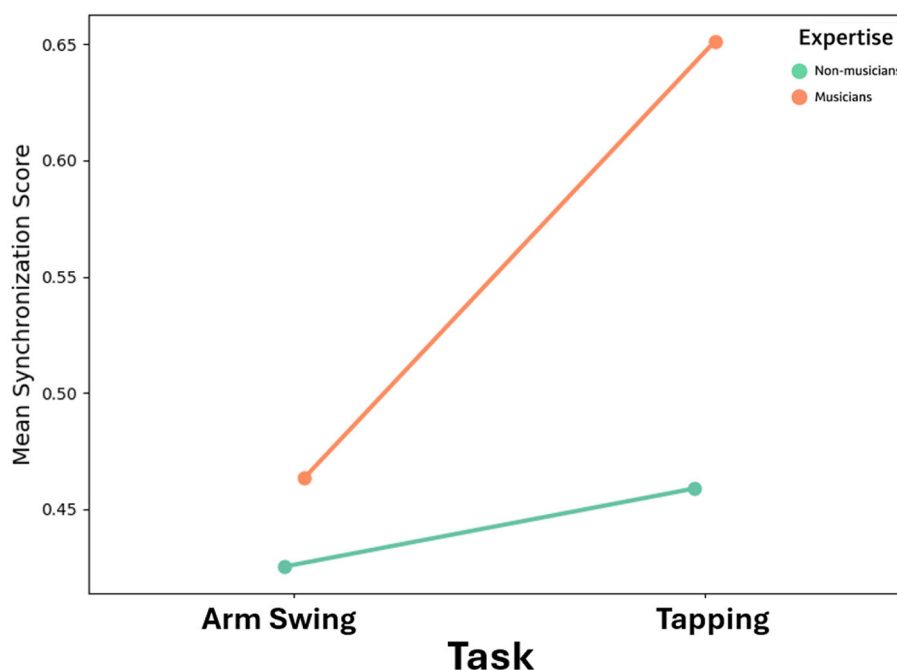


FIGURE 12
Interaction plot showing synchronization performance across tasks and expertise levels.

training compared to the arm swing task, which involves larger, less precise movements.

Figure 11 illustrates this interaction between musical expertise and task type, showing mean synchronization scores for each group (musicians vs. non-musicians) across both tasks (tapping and arm swing). As shown in Figure 12, both groups performed better in the tapping task than in the arm swing task. However, the improvement from arm swing to tapping was substantially greater for musicians than for non-musicians.

This interaction suggests that musical expertise had a particularly strong effect in the tapping condition, which requires fine motor control and precise timing. In contrast, the benefit of expertise was less pronounced in the arm swing task, which involves broader and more variable movements. These findings indicate that the impact of musical training on synchronization may depend on the motor demands of the task.

5 Discussion

One of the primary aims of this study was to bridge the gap between controlled laboratory-based sensorimotor synchronization (SMS) research and more naturalistic movement contexts. While traditional SMS studies have typically relied on simple and controlled tasks—such as finger tapping to metronomes or isochronous tones—to maximize internal validity and isolate specific timing mechanisms (Repp, 2005; Repp and Su, 2013), our approach prioritized ecological validity by using full-body movements and real-world musical stimuli.

We recognize that using real musical excerpts introduces greater variability in terms of rhythm, instrumentation, and genre characteristics, potentially influencing synchronization behavior in ways that are less predictable. For instance, genre-specific features such as the use of syncopation in jazz or the polyrhythmic textures in classical or swing pieces may modulate beat perception and motor alignment differently. While we controlled for beat division and tempo, future studies could explore more systematically how specific musical properties—such as instrumentation, groove, or harmonic complexity contribute to synchronization performance.

In comparing our findings with classical SMS research, we note both convergence and divergence. As with many traditional studies, we observed that synchronization accuracy decreases with increased rhythmic complexity (e.g., ternary and swing divisions vs. binary). However, the additional influence of motor modality (tapping vs. arm swing) and musical expertise highlights the complexity of synchronization in real-world settings. Our results suggest that ecological paradigms can still replicate key SMS findings while revealing new interaction effects not typically observed in simplified tasks.

The findings of this study strongly support our initial hypothesis that both musical expertise and the inherent properties of rhythmic stimuli (such as tempo and beat division) significantly influence synchronization performance. Across both finger tapping and arm swing tasks, musicians consistently outperformed non-musicians in synchronization accuracy and regularity. This aligns with prior research indicating that musical training enhances sensorimotor synchronization abilities (Repp, 2010; Repp and Doggett, 2007; Krause et al., 2010).

5.1 Effects of tempo and rhythmic complexity

Our results further demonstrate that tempo and rhythmic complexity play a crucial role in synchronization. Slower tempos and simpler rhythmic structures (e.g., binary rhythms) facilitated more accurate and consistent synchronization. This finding is consistent with prior studies (Mathias et al., 2020; Rose et al., 2021; Hammerschmidt and Wollner, 2020), which show that increasing rhythmic complexity negatively impacts synchronization. In particular, (Møller et al., 2021) demonstrated that listeners naturally perceive and synchronize better with binary rhythms, reinforcing our observation that binary structures provide a perceptual and motor advantage over ternary or swing rhythms.

Moreover, the nature of movement itself influenced synchronization performance. Finger tapping, a discrete and precisely controlled action, yielded higher synchronization accuracy compared to arm swing, which involves larger and more continuous movements. This echoes findings from Peckel et al. (2014), suggesting that movement precision influences synchronization efficacy. However, this raises an open question: do musicians excel in tapping tasks simply due to their training, or does the complexity of larger motor movements inherently challenge synchronization? Future research could investigate whether musicians trained in movement-based disciplines (e.g., dancers or percussionists) exhibit advantages in synchronizing broader limb motions.

5.2 Cultural considerations in synchronization

All participants in this study were from Western cultural backgrounds, where binary rhythms and duple meters (e.g., 4/4 time) are dominant in everyday musical exposure. This cultural familiarity may have contributed to the observed preference for binary rhythms. Prior research has shown that rhythmic perception and synchronization abilities can vary across cultural contexts due to differences in exposure to specific metrical structures (Cameron et al., 2015).

For example, west African music—particularly from cultural groups such as the Ewe and Yoruba—features complex polyrhythmic drumming practices (Dor, 2014; Tchetgen, 2024). These rhythmic systems are structurally different from Western binary meters and may shape sensorimotor synchronization in distinct ways. Future studies should aim to compare synchronization performance across culturally diverse populations to better understand how rhythmic enculturation influences movement coordination.

5.3 Practical implications

5.3.1 Music therapy and rehabilitation

Understanding how tempo and rhythmic structure influence synchronization has direct implications for music therapy and

rehabilitation. Rhythmic auditory stimulation has been shown to improve motor coordination in patients with Parkinson's disease (Cochen De Cock et al., 2018) and stroke survivors (Raglio et al., 2017). Our findings suggest that using slower tempos and binary rhythms may optimize therapeutic interventions by enhancing movement synchronization.

5.3.2 Dance and athletic training

In dance, synchronization with musical rhythm is fundamental. Our results highlight that rhythmic complexity affects movement coordination, which has implications for dance instruction. Novice dancers may benefit from starting with slower, simpler rhythms before progressing to more complex patterns (Karageorghis et al., 2019). Similarly, in sports training, structured rhythmic timing has been shown to enhance motor performance (Ronnqvist et al., 2018). The integration of rhythm-based training could be beneficial in disciplines requiring precise timing, such as gymnastics or martial arts.

5.3.3 Human-computer interaction and AI applications

In human-computer interaction, rhythm synchronization insights can guide the development of interactive systems such as rhythm-based games or AI-driven music applications. Games like Beat Workers (Beat Workers Team, 2023) leverage rhythm-based challenges to enhance user engagement. Future AI-driven models could refine real-time movement-to-music synchronization, extending applications to interactive dance training, virtual reality, and motion-controlled gaming.

5.4 Future directions and AI-driven synchronization

Beyond traditional signal processing approaches, AI-based methods hold promise for advancing synchronization research. Deep learning architectures—such as Transformers or Temporal Convolutional Networks (TCNs)—could model movement-to-music relationships in a data-driven manner, bypassing manual beat detection. A multimodal AI model integrating computer vision (e.g., CNNs or Vision Transformers) with audio processing could enable real-time synchronization between human movement and musical rhythm. However, such advancements require extensive annotated datasets and highly optimized real-time performance models.

6 Limitations

While the present study provides novel insights into movement synchronization across rhythmic structures and expertise levels, several limitations must be acknowledged.

First, the participant sample was relatively small ($n = 24$) and drawn entirely from the same academic institution. This introduces potential selection bias and limits the generalizability

of the findings. Moreover, group assignment (musicians vs. non-musicians) was based on self-report without the use of standardized musical sophistication indices such as the Goldsmiths MSI.

Second, the study did not include detailed assessments of participants' prior dance or movement experience, which could influence synchronization ability independently of musical training. Similarly, although efforts were made to minimize recognition effects, we cannot entirely exclude the influence of stimulus familiarity, particularly for genre-specific tracks.

Third, genre distribution across beat divisions was not balanced due to constraints in the available database. For example, binary stimuli were predominantly from the disco genre, while ternary examples included mainly classical music. This imbalance may have affected participants' responses and should be addressed in future studies using more controlled stimuli.

Finally, although age was relatively homogeneous in our sample (mean age 22), it was not matched across groups, and no age-specific analyses were performed. Future research should explore potential age-related effects more systematically.

Despite these limitations, the study lays important groundwork for future research in embodied music cognition and rhythm perception using accessible motion tracking tools.

7 Conclusion

This study has demonstrated that both musical expertise and the inherent properties of rhythmic stimuli—specifically tempo and beat division—play crucial roles in influencing synchronization between body motion and musical rhythm. Musicians consistently exhibited higher synchronization accuracy and regularity compared to non-musicians, particularly in tasks involving finger tapping and arm swing. Additionally, our findings highlight that slower tempos and simpler rhythmic structures, such as binary rhythms, facilitate more precise and consistent synchronization. These results align with prior research on sensorimotor synchronization and contribute new insights into how rhythmic complexity affects movement timing.

Beyond its theoretical contributions, this study has practical implications across multiple fields, including music therapy, dance, and human-computer interaction. Understanding how rhythmic structure and tempo influence synchronization can aid in the development of targeted interventions for motor rehabilitation, optimize choreographic designs, and enhance user experiences in interactive systems such as rhythm-based games and virtual training environments.

A key strength of this study is the methodological framework, which enables precise control over experimental conditions while ensuring efficient data collection and management. Using only a simple webcam and our custom software, this system provides a portable and accessible solution for researchers investigating sensorimotor synchronization. This approach facilitates studies in diverse environments, particularly for experiments involving simple movement tasks.

Despite these advancements, certain limitations should be acknowledged. MediaPipe's markerless tracking technology,

while effective, is occasionally prone to misclassification errors, particularly under challenging lighting conditions or when hands are partially occluded. Additionally, the beat-tracking models used, though generally reliable, exhibited difficulties in processing complex musical structures and tempo fluctuations. Lastly, while the proposed system is scalable, further optimization is necessary to ensure robustness across a wider range of musical genres and movement types.

Future research should aim to address these limitations by integrating more advanced motion-tracking techniques and improving beat-tracking algorithms to accommodate diverse musical styles. Moreover, expanding investigations to include culturally diverse populations and more complex movement patterns will provide deeper insights into the universality and variability of synchronization mechanisms. With continued refinement, this line of research holds significant potential for advancing our understanding of the intricate relationship between body motion and musical rhythm, paving the way for novel applications in both scientific and applied domains.

Data availability statement

The datasets generated during and/or analyzed in the current study are not publicly available due to participant confidentiality and institutional restrictions, but they are available from the corresponding author on reasonable request. Requests to access the datasets should be directed to baydhamza99@gmail.com.

Ethics statement

Ethical approval was not required for the studies involving humans because all participants were volunteers from among the lab's students and staff. IMT Mines Alès does not have an Ethics Committee, so we were unable to obtain local ethical approval. However, participants provided written informed consent to participate in this study. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. PG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing. BB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing. PS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration,

Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain Cogn.* 48, 66–79. doi: 10.1006/brcg.2001.1304
- Barbancho, I., Rosa-Pujazon, A., Tardon, L. J., and Barbancho, A. M. (2013). “Human-computer interaction and music,” in *Sound-Perception-Performance* (Springer International Publishing), 367–389.
- Bayd, H., Guyot, P., Bardy, B. G., and Slangen, P. (2024). “Scoring synchronization between music and motion: local vs global approaches,” in *EUSIPCO 2024-32nd European Signal Processing Conference (IEEE)*, 636–640.
- Beat Workers Team (2023). *Beat Workers*.
- Bock, S. (2016). *Event Detection in Musical Audio: Beyond Simple Feature Design/Submitted by Sebastian Bock*.
- Bock, S., Korzeniowski, F., Schluter, J., Krebs, F., and Widmer, G. (2016). “Madmom: a new python audio and music signal processing library,” in *Proceedings of the 24th ACM International Conference on Multimedia (ACM)*, 1174–1178.
- Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., et al. (2013). Essentia: an audio analysis library for music information retrieval. *ISMIR*. 13, 493–498. doi: 10.1145/2502081.2502229
- Cameron, D. J., Bentley, J., and Grahn, J. A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. *Front. Psychol.* 6:366. doi: 10.3389/fpsyg.2015.00366
- Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* 20, 226–239. doi: 10.1162/jocn.2008.2.0018
- Cochen De Cock, V., Dotov, D., Damm, L., Lacombe, S., Ihalainen, P., Picot, M. C., et al. (2021). Beatwalk: personalized music-based gait rehabilitation in Parkinson’s disease. *Front. Psychol.* 12:655121. doi: 10.3389/fpsyg.2021.655121
- Cochen De Cock, V., Dotov, D., Ihalainen, P., Begel, V., Galtier, F., Lebrun, C., et al. (2018). Rhythmic abilities and musical training in parkinson’s disease: do they help? *NPJ Parkinson’s Dis.* 4:8. doi: 10.1038/s41531-018-0043-7
- Dor, G. W. K. (2014). *West African Drumming and Dance in North American Universities: An Ethnomusical Perspective*. Jackson, Mississippi: Univ. Press of Mississippi.
- Hammerschmidt, D., and Wollner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time. *Music Percept.* 37, 263–277. doi: 10.1525/mp.2020.37.4.263
- Karageorghis, C. I., Lyne, L. P., Bigliassi, M., and Vuust, P. (2019). Effects of auditory rhythm on movement accuracy in dance performance. *Hum. Mov. Sci.* 67:102511. doi: 10.1016/j.humov.2019.102511
- Krause, V., Pollok, B., and Schnitzler, A. (2010). Perception in action: the impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychol.* 133:28–37. doi: 10.1016/j.actpsy.2009.08.003
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106:119. doi: 10.1037//0033-295X.106.1.119
- Large, E. W., Roman, I., Kim, J. C., Cannon, J., Pazdera, J. K., Trainor, L. J., et al. (2023). Dynamic models for musical rhythm perception and coordination. *Front. Comput. Neurosci.* 17:1151895. doi: 10.3389/fncom.2023.1151895
- Mao, W., Ge, Y., Shen, C., Tian, Z., Wang, X., Wang, Z., et al. (2022). “Poseur: direct human pose regression with transformers,” in *European Conference on Computer Vision (Cham: Springer)*, 72–88.
- Marchand, U., Fresnel, Q., and Peeters, G. (2015). *GTZAN-rhythm: Extending the GTZAN Test-Set with Beat, Downbeat and Swing Annotations*.
- Mathias, B., Zamm, A., Gianferrara, P. G., Ross, B., and Palmer, C. (2020). Rhythm complexity modulates behavioral and neural dynamics during auditory-motor synchronization. *J. Cogn. Neurosci.* 32, 1864–1880. doi: 10.1162/jocn_a_01601
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., et al. (2015). librosa: audio and music signal analysis in python. *SciPy* 2015, 18–24. doi: 10.25080/Majora-7b98e3ed-003
- Møller, C., Stupacher, J., Celma-Mirallès, A., and Vuust, P. (2021). Beat perception in polyrhythms: time is structured in binary units. *PLoS ONE* 16:e0252174. doi: 10.1371/journal.pone.0252174
- Panteleris, P., and Argyros, A. (2022). “Pe-former: pose estimation transformer,” in *International Conference on Pattern Recognition and Artificial Intelligence (Cham: Springer)*, 3–14.
- Peckel, M., Pozzo, T., and Bigand, E. (2014). The impact of the perception of rhythmic music on self-paced oscillatory movements. *Front. Psychol.* 5:100163. doi: 10.3389/fpsyg.2014.01037
- Raglio, A., Zaliani, A., Baiardi, P., Bossi, D., Sguazzin, C., Capodaglio, E., et al. (2017). Active music therapy approach for stroke patients in the post-acute rehabilitation. *Neurol. Sci.* 38, 893–897. doi: 10.1007/s10072-017-2827-7
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychon. Bulletin Rev.* 12, 969–992. doi: 10.3758/BF03206433
- Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Hum. Mov. Sci.* 29, 200–213. doi: 10.1016/j.humov.2009.08.002
- Repp, B. H., and Doggett, R. (2007). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Percept.* 24, 367–376. doi: 10.1525/mp.2007.24.4.367
- Repp, B. H., and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bulletin Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2
- Ronnqvist, L., McDonald, R., and Sommer, M. (2018). Influences of synchronized metronome training on soccer players’ timing ability, performance accuracy, and lower-limb kinematics. *Front. Psychol.* 9:378920. doi: 10.3389/fpsyg.2018.02469
- Rose, D., Ott, L., Guerin, S. M., Annett, L. E., Lovatt, P., and Delevoe-Turrell, Y. N. (2021). A general procedure to measure the pacing of body movements timed to music and metronome in younger and older adults. *Sci. Rep.* 11:3264. doi: 10.1038/s41598-021-82283-4

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Schiphof-Godart, L., and Hettinga, F. J. (2017). Passion and pacing in endurance performance. *Front. Physiol.* 8:83. doi: 10.3389/fphys.2017.00083

Tchetgen, P.-V. N. (2024). "Can multimodal rhythmic interaction impact the literacy and socio-emotional development of children: the case of the african talking drums," in *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*, 116–129.

Thaut, M. H., and Abiru, M. (2010). Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. *Music Percept.* 27, 263–269. doi: 10.1525/mp.2010.27.4.263

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., et al. (2020). Mediapipe hands: on-device real-time hand tracking. *arXiv [preprint]* arXiv:2006.10214. doi: 10.48550/arXiv.2006.10214