






A Cooperative and Interactive Gesture-Based Drumming Interface with Application to the Internet of Musical Things

Azeema Yaseen^(✉), Sutirtha Chakraborty, and Joseph Timoney

Department of Computer Science, Maynooth University, Maynooth,
Co.Kildare, Ireland

{azeema.yaseen.2020,sutirtha.chakraborty.2019}@mumail.ie

Joseph.timoney@mu.ie

Abstract. The Internet of Musical Things (IoMusT) envisions a network of interconnected objects enabling multi-directional communication for novel musical interactions. The requisite demands on the interfacing technologies are significant. Challenges include users with different skills, delivering effective feedback, and signalling adverse conditions in the physical network. To this end, the design and implementation of a cooperative virtual drumming interface is investigated. It supports remote users through gestural interaction with virtual percussion objects. The GUI includes calibration mode (re-positioning objects), a performance window, community rhythm-pattern visualization, and a cooperative synchronicity display. In creative contexts, colour is often used as a metaphorical representation such as music/colour art, provides a second sensory perspective to the user, and here they are applied to highlight evolving user interactions. For evaluation a binary user cohort of students with and without prior musical experience was identified. Data was collected to measure the interactivity response and their rhythmic cohesiveness along with results on the user impressions of the collaborative aspects of this interface.

Keywords: Gesture based musical interaction · Virtual music instruments · Music-color mappings · Virtual musical collaboration · Performance synchronisation

1 Introduction

IoMusT was inspired by the fusion of ubiquitous computing and the Internet of Things (IoT) [2]. Its breadth of things includes single user devices (laptop, computer, iPad or wearables) embedded with computational abilities that have been specifically created for, or can be directed to, performing in-place or virtual musical activities. Networked Music Performance (NMP) is one of the predominant application domains of IoMusT and covers all remote musical interactions between musicians [3].

In the traditional performance environment sonic musical cues used in performer synchronisation are reinforced by the cues generated from the visual, physical connections in the shared composite space. However, in virtual interactions this medium becomes impaired. Thus, the virtual interface must compensate in some way.

Another compromise on the effectiveness of any virtual interactions is the inherent issue of network latency. This is a crucial obstacle as significant sound lags due to latency can completely disrupt the performers' rhythmic synchronisation. Recent papers [3,4] highlighted this challenge and proposed network adaptive metronomes to manage synchrony. Another perspective is to treat latency as a real and ever-present, albeit, unstable feature, and so to think of dealing with it by designing a new musical instrument which exploits latency as being part of the interactions [1].

An interface that could successfully integrate sonic and visual feedback that fits within IoMusT is our motivation for this work. Our approach is to investigate the development of a virtual remote-collaborative drumming interface.

2 Gesture Based Interfaces and Virtual Musical Instruments (VMIs)

Gestures refer to meaningful expressions of the human body when it moves; this may involve the hands, the arms, the head, or whole-body motions [4,5]. More recently, the handy metaphor (Handy hear and Handy see) in [6] was developed for IoMusT; in both modes it enables touchless sonic interactions to manipulate pitch, amplitude and duration of the sound. Another is "Interval Player" [7], an in-air VMI, that allows users to specify the melodic interval using their hands. The vision-based collaborative musical interface "lamoscope" [8] uses image processing to produce music based on the user's body image. [9] presents a web-based interface for virtual performance of an air guitar and an upright bass; collaborative playing of two or more players is enabled. Often for gestural instruments, custom-build devices, e.g., data gloves, or body suits, or off-the shelf products, such as Kinect and Leap motion, are used to collect data. These devices may hinder movement, and have a financial and installation cost. Others use camera-based gesture recognition to avoid this and all laptop computers have a camera as standard. This leads to our use of Human Pose Estimation (HPE) as an advanced manifestation of the computer vision approach.

Pose is an arrangement of human joints and HPE is the localization of human joints from a predefined set of key-points or landmarks in a recording or sequence of live images [10]. These are shown in Fig. 1. These landmarks are tracked continuously using the Mediapipe Pose Estimation (PE) model. It was found to be accurate and so could capture user expressiveness during a musical activity. Figure 2 shows the Region of Interest (ROI) we used for the proposed interface to track, estimate and recognise users' movements during the live performance.

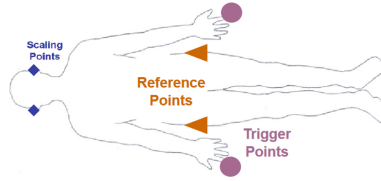


Fig. 1. Key-points extraction from human pose used as referents to determine body, limbs and fingers positions.

3 Design Description: Gestural Drumming Interface with Color Based Musical Objects

3.1 Graphical Interface for Virtual and Collaborative Drumming

The system architecture is designed based on a two-tier client server architecture to facilitate the interactive drumming performances of the remote ensemble. The performance has two levels; 1) Local-level Individual Interaction and 2) Ensemble-level Interaction. The first enables self-interaction where the participant interacts with the drum machine running locally. Here, the camera sensor is placed in front of the player for recording the gestures and detecting the actions of the musical performance. A Deep learning-based PE model [11] is used to detect the body key positions and movement of the user for acquisition of the required gestural information. The algorithm estimates the hands', arms' and lower body 3D key points for each of the video frames captured. A buffer with fifteen frames was created for each arm. To track the movements of each arm we used a queue-replacement approach. A 'speed buffer' calculates the 3D spatial displacements of the key points between two consecutive frames, so whenever an index finger enters the specified boundary on the interface and exceeds a threshold parameter, the 'hit' instance is processed. The boundary is represented as a rectangular color box (see Fig. 2(b) and (c)).

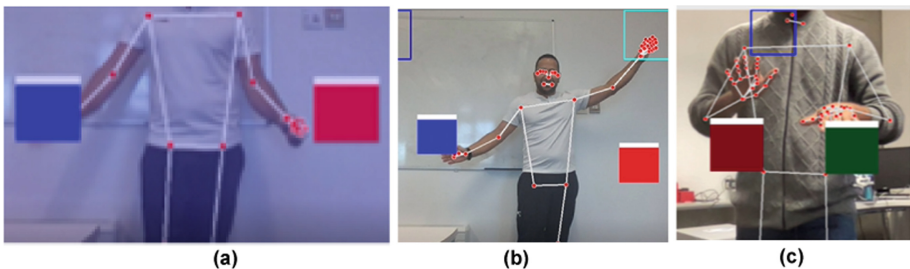


Fig. 2. Gesture recognition using computer camera and interaction with musical objects: hands within the boundary of musical object. (Color figure online)

Color Based Percussion Objects. A group of arranged percussion instruments is called a drum kit. The drum kit can comprise of any number of percussion instruments, but for this study a bass (kick) drum, a snare drum and two cymbals (hi-hat and crash) are included. Due to the variability of drum kit pitches, it was not possible to define a schematic approach for color-pitch mapping so we represented the kick and snare drums with saturated colors (red and blue) and used dark colors (dark green and firebrick) for the cymbals. The general order is observed from left (more saturated) to right (less saturated). There is configuration flexibility of the virtual drums and so re-positioning of the objects is possible (see Fig. 2(a)). When a gesture is performed, the gestural interface executes a “struck” event associated with the particular gesture and the percussion object.

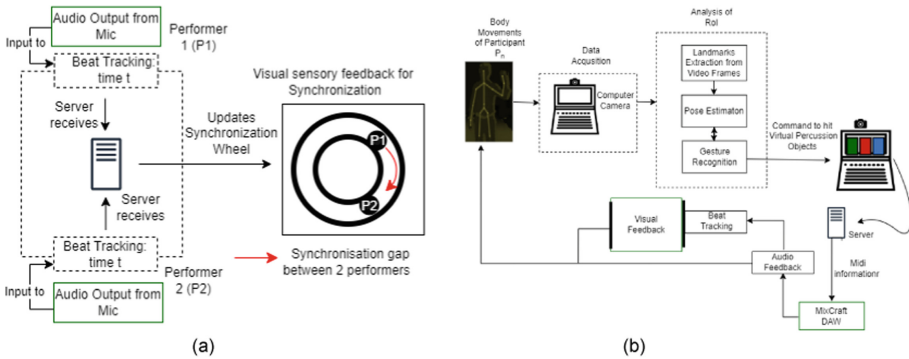


Fig. 3. (a) Representation of the beat tracking for synchronicity and its visualization, (b) the system model.

Performer’s Synchronisation Wheel. For collaborative performance, participants connect to the same server and the ‘hit’ event initiates the virtual collaborative interaction between multiple users. We designed a socket server which handles multiple clients using multi threading. A mathematical model for multiple oscillator coupling, i.e. the Kuramoto oscillator [12,13] is applied to encourage synchronization between the virtual performers. For each participant, each object is considered as an independent oscillator. For example, for two participants P1 and P2, player P1 plays the kick(K) and snare(S), and P2 plays hi-hat(H) and crash(C). This would result in the creation of 4 oscillators OSCK, OSCS, OSCH, and OSCC. On receiving the information, the server calculates the time difference between two consecutive hits of each oscillator. This is fed to the Kuramoto model that informs how to keep these coupled oscillators in synchronization, not just in frequency, i.e. tempo, but also in phase, i.e. that the rhythmic pulse or beats happen at the same time. This is illustrated in Fig. 3(a).

Visual Feedback. Colors for a long time have been used for music visualization and there are many music-color mappings proposed between musical elements (e.g., loudness, timbre, pitch,) and color properties like hue, saturation, and brightness. A very recent survey in [9] gives a detailed overview. In our instrument colours are associated with musical objects and are also used for secondary visual feedback to the performers. Combined auditory and visual feedback supports performers during a collaboration.

We used the same color mappings used for the percussion instruments to give color-based feedback on their rhythmic patterns. Every time “auditory feedback” occurs, the color associated with that specific audio is displayed on the user’s own, and on the virtual performer’s interface. The complete system interface is given in Fig. 4. The interface in (a) and (b) show sample rhythmic patterns and (c) shows the information passed to the synchronization wheel.

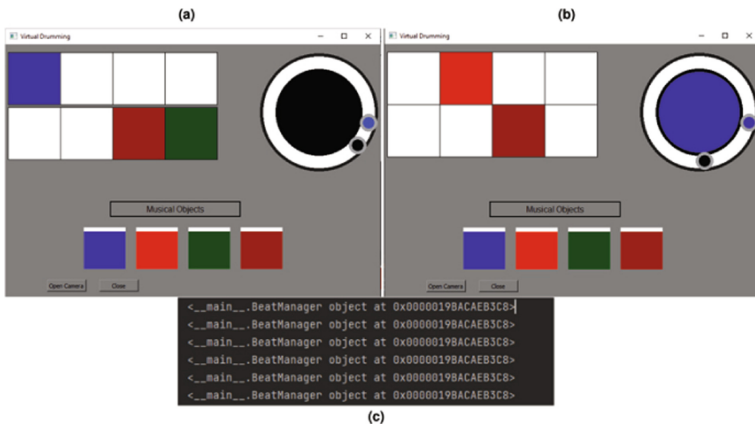


Fig. 4. The GUI of drumming interface.

4 User Experience Study

For this study, 4 participants (2 amateurs and 2 musicians) collaborated in three different settings (Musician-Amateur, Amateur-Musician and Musician-Musician) using this interface. The sessions were conducted for a time-span of approximately 2 h. To get feedback about their experience we used the think-aloud technique [14] and did not interrupt them when they were collaborating. Thus, the performance was let happen, followed by a sharing of their feedback on how they were experiencing the interface, then they played again, and following this commented further until the session ended. Each group spent 15 min on average excluding the time they took to become familiar with the interface elements. We then asked them planned open-ended questions at the end of performance.

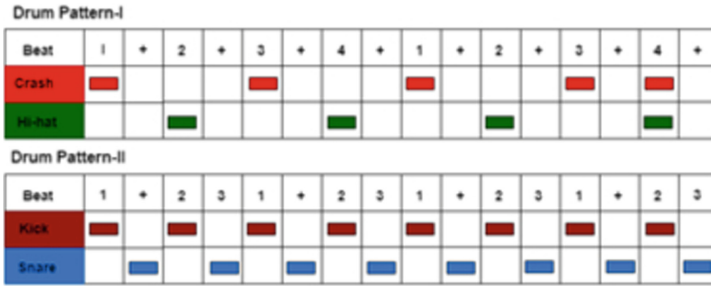


Fig. 5. Rhythmic phrases played by participants during performance.

4.1 Musical Interaction Modes

The participants were given 2 simple rhythmic patterns presented in Fig. 5 to play during the drumming performance. Participants were given time to become familiar with the interface, then they followed the two models of interactions as given below.

The ‘Instrumental’ Mode. An ‘instrumental’ interaction is based on the instrumental metaphors adopted from traditional interaction. Here, the participants tried to synchronize from the auditory feedback only. It was observed that the musicians were able to learn the patterns faster than non-musicians. The amateurs understood the “hitting drums” metaphor but took time to adapt to the sound of each musical object.

The ‘Interactional’ Mode. In ‘Interactional’ mode, they focused on using the interface at its full capability. This mode allowed participants to see the colours of percussion objects that were played by themselves and their virtual participant. Participants tried to adapt their body gestures to the visual feedback before they were able to receive the audio feedback, which was subject to a short delay. Once they found a way to balance between their gestures, with the color and auditory feedback, they tried to understand their partners’ beating patterns. In the earlier stage, the synchronization wheel was not considered. When they realized the position of colors moving in the interface window, they started making gestures at different scales and they expected to achieve quick visual and audio feedback responses. Unfortunately, latency did interfere with the predictive worth of the synchronization wheel and this impacted its value.

Findings. We found that keeping the rhythmic phrases (as shown in Fig. 5) was easily achievable when the tempo ranged between 60 to 125 BPM (for simple sequences). The participants after their involvement offered similar opinions regarding gestural interactions and the drumming interface. They all agreed that the visualization of what the others were doing along with the auditory feedback

in the user interface enhanced the performance. In the “Instrumental” mode the descriptions were flexible, enjoyable, and expressive. In this mode, participants were more engaged to achieve consistency. In the “Interactional” mode the descriptors were “expressive”, “conversational” and “enjoyable”. The synchronization wheel was only helpful when gestures were recognized quickly and the synchronization information was updated prior to the user making the gesture for next beat.

5 Conclusion and Future Work

This paper focused on musically driven efforts to adopt vision-based techniques to capture human gesture and afford non-traditional forms of sonic articulation and creative musical expressions. The intention for this new VMI interface was that it should permit live interactive performances among users from different musical backgrounds and with different levels of skills. This paper described a specially-designed collaborative VMI for drumming that incorporated auditory and visual feedback in its interface that aimed at shaping the performers’ individual and joint interactions. Our experience, based on participant feedback, found that using musical interaction models that were similar to traditional models along with color-based visual stimulus can facilitate engaging and collaborative music making activities. The performances mediated by the color based visual feedback were conversational and conveyed meaningful information among the participants.

For future work, the interaction experience can be expanded by providing more coupling between sound and gestures, e.g. faster movements for high-pitched sounds and slow movements for lower pitched sounds. To achieve this, the gesture recognition model needs modification along with consideration for an enhanced color representation. This should also be easy to learn. Focus should also be placed on solving the time delay issues and their impact on the synchronization wheel. Lastly, a more comprehensive user study must be carried out.

References

1. Wilson, R.: Aesthetic and technical strategies for networked music performance. *AI Soc.* 1–14 (2020). <https://doi.org/10.1007/s00146-020-01099-4>
2. Turchet, L., Fischione, C., Essl, G., Keller, D., Barthet, M.: Internet of musical things: vision and challenges. *IEEE Access* **6**, 61994–62017 (2018)
3. Battello, R., Comanducci, L., Antonacci, F., Cospito, G., Sarti, A.: Experimenting with adaptive metronomes in networked music performances! *J. Audio Eng. Soc.* **69**(10), 737–747 (2021)
4. Battello, R., et al.: An adaptive metronome technique for mitigating the impact of latency in networked music performances. In: 2020 27th Conference of Open Innovations Association (FRUCT), pp. 10–17 (2020)
5. Mitra, S., Acharya, T.: Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **37**(3), 311–324 (2007)

6. Keller, D., Gomes, C., Aliel, L.: The handy metaphor: bimanual, touchless interaction for the internet of musical things. *J. New Music Res.* **48**(4), 385–396 (2019)
7. Lages, W., Nabiyouni, M., Tibau, J., Bowman, D.A.: Interval player: designing a virtual musical instrument using in-air gestures. In: 2015 IEEE Symposium on 3D User Interfaces (3DUI), pp. 203–204 (2015)
8. Fels, S., Mase, K.: Iamascope: a graphical musical instrument. *Comput. Graph.* **23**(2), 277–286 (1999)
9. Lima, H.B., Santos, C.G.D., Meiguins, B.S.: A survey of music visualization techniques. *ACM Comput. Surv. (CSUR)* **54**(7), 1–29 (2021)
10. Josyula, R., Ostadabbas, S.: A Review on Human Pose Estimation. arXiv preprint [arXiv:2110.06877](https://arxiv.org/abs/2110.06877) (2021)
11. Singh, A.K., Kumbhare, V.A., Arthi, K.: Real-time human pose detection and recognition using MediaPipe. In: International Conference on Soft Computing and Signal Processing, pp. 145–154 (2021)
12. Rodrigues, F.A., Peron, T.K.D., Ji, P., Kurths, J.: The Kuramoto model in complex networks. *Phys. Rep.* **610**, 1–98 (2016)
13. Chakraborty, S., Timoney, J.: Robot human synchronization for musical ensemble: progress and challenges. In: 2020 5th International Conference on Robotics and Automation Engineering (ICRAE), pp. 93–99 (2020)
14. Ericsson, K.A., Simon, H.A.: How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind Cult. Act.* **5**(3), 178–186 (1998)