

Article

VR Drumming Pedagogy: Action Observation, Virtual Co-Embodiment, and Development of Drumming “Halvatar”

James Pinkl *  and Michael Cohen * 

Spatial Media Group, University of Aizu, Tsuruga, Ikkī-Machi, Aizu-Wakamatsu 965-8580, Fukushima, Japan
* Correspondence: jamesjpinkl@gmail.com (J.P.); xilehence@gmail.com (M.C.)

Abstract: Virtual Co-embodiment (VC) is a relatively new field of VR, enabling a user to share control of an avatar with other users or entities. According to a recent study, VC was shown to have the highest motor skill learning efficiency out of three VR-based methods. This contribution expands on these findings, as well as previous work relating to Action Observation (AO) and drumming, to realize a new concept to teach drumming. Users “duet” with an exemplar half in a virtual scene with concurrent feedback to learn rudiments and polyrhythms. We call this puppet avatar controlled by both a user and separate processes a “halvatar”. The development is based on body-part-segmented VC techniques and uses programmed animation, electromechanical drum strike detection, and optical bimanual hand-tracking informed by head-tracking. A pilot study was conducted with primarily non-musicians showing the potential effectiveness of this tool and approach.

Keywords: personal sensing; Virtual Co-embodiment (VC); virtual reality; mixed reality; music pedagogy; rhythm training; Action Observation (AO); motor skill learning; sound and music computing; information interfaces and presentation; sense of agency



Citation: Pinkl, J.; Cohen, M. VR Drumming Pedagogy: Action Observation, Virtual Co-Embodiment, and Development of Drumming “Halvatar”. *Electronics* **2023**, *12*, 3708. <https://doi.org/10.3390/electronics12173708>

Academic Editor: George A. Papakostas

Received: 21 June 2023

Revised: 2 August 2023

Accepted: 18 August 2023

Published: 1 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this introduction, an overview of drumming and general motor skill learning is presented. Polyrhythms and rudimental drumming, musical concepts referenced throughout this research, are described. Virtual reality (VR) fields relevant to this research, such as Action Observation (AO) and virtual co-embodiment (VC), are reviewed. Lastly, rhythm games and other technology-based musical education tools are also surveyed.

1.1. Motor Skill Learning

Motor skill learning is defined as the process of increasing the accuracy of movement in the space and time domains through practice. This learning can be further broken down into cognitive, associative, and autonomous phases [1]. The cognitive phase consists of learners striving to understand movements correctly. The associative phase consists of conscious effort to gain the skill to perform movements correctly. The autonomous phase is the process of gaining the ability to perform movements unconsciously.

1.2. Technology-Accelerated Motor Skill Learning

Recently developed systems have been designed to help teach movement and gestures in new ways. YouMove is a novel augmented reality mirror and system that helps users learn sequences of physical movements [2]. The system displays guidance and feedback at varying levels to help train users. A study using YouMove showed that the system significantly improved learning and short-term retention compared to learning performed via video.

Similarly positive results were found with OctoPocus, a dynamic guide for the teaching of specific gestures with visual feedforward and feedback functionalities [3]. The system helps users achieve proficiency and was observed to teach significantly faster than that of standard Help menus.

1.3. Drumming and Rhythm

Drumming is defined as the action of expressing rhythm by striking a membranophone (such as a drum), idiophone (such as cymbal), or other objects using drum sticks or by directly striking with the hands or fingers. Rhythm, despite having been described countless times, eludes a comprehensive definition [4]. The book "*The Geometry of Rhythm*" compiles many interpretations. "A measuring of time by means of some kind of movement" was declared by Baccheios the Elder. Didymus defined it as "a schematic arrangement of sounds", and R. Parncutt said, "a musical rhythm is an acoustic sequence evoking a sensation of pulse" [4].

When considered from a physical perspective, drumming, like any motor skill, requires practice before successful execution [5]. It entails both fine and gross motor skills, as small muscles in the fingers are used for stick control, while the movement of larger muscles such as those in one's arms and torso are stretched to reach various parts of the drummer's kit.

From a motor control perspective, the varied motions of drumming can be classified as serial, discrete, or continuous [6]. A challenging piece of music may demand combinations of all three movements within a relatively short amount of time.

1.4. Musical Practice Tools

The metronome is a musical tool used since the 19th century, designed to improve musical performance of practicing musicians [7]. The metronome expresses a steady pulse, atop which musicians can practice. It prevents unwanted modulations of tempo and helps a player understand time-related issues with performance. Since then, many electronic and digital versions of metronomes have been released, giving players many more options. PolyNome [8] is a mobile application that lets users program desired rhythms, including drum set grooves and accent patterns. PocketDrum [9] and HyperDrum [10] are examples of portable virtual drum set products that detect drumming motions via motion sensing, rendering air drum-like gestures into audible drum sounds. They feature volume dynamics and haptic feedback capabilities, allowing drummers to practice anywhere.

1.5. Musical Practice

Practice can be described as a "systematic activity with predictable stages and activities" [11]. Musical practice can fall into one of four categories: technical practice, automation and memorization, detail-focused review, and maintenance of a piece [12]. Practice with a metronome is usually considered technical practice. For example, a common technique is to start a metronome at a relatively slow tempo, practice an excerpt until it can be naturally performed, and then, gradually increase the tempo in increments until it reaches a desired speed and level of execution.

1.6. Rudiments

In 1933, drummers in the United States organized a meeting to discuss drum instruction. Considering control, coordination, and endurance, thirteen rudiments were selected and deemed essential for any drummer to know [13]. They remain an important part of the musical language of drumming in most contexts and cultures, regardless of genre. They also remain staples in most drummers' practice routines, as complete mastery of a rudiment is never fully achieved. The speed, clarity, and technique involved in the performance of a rudiment can be improved endlessly. While some rudiments have rhythmic dynamics, others, such as those shown in Figure 1, focus on sticking or the specific assignment of notes to the left or right hand. These exercises build coordination and help players play specific accent patterns or phrases smoothly.

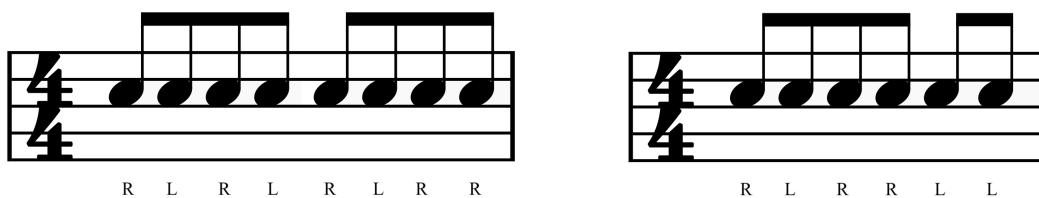


Figure 1. Western notation for two drumming rudiments: triple paradiddle and paradiddle-diddle.

1.7. Polyrhythms

A polyrhythm is “a set of simultaneous rhythms, where each rhythm strongly implies an entirely different meter” [14]. In other words, it is the simultaneous use of mathematically related divisions of pulse in music. For example, fitting a group of three notes in the same amount of time as another musical line fitting two notes creates a 3:2 polyrhythm. More generally, a musical phrase with m subdivisions played at the same time as a musical phrase with n subdivisions (where $n \neq x * m$ and n, m , and x are whole numbers) results in a $m:n$ polyrhythm. Although polyrhythms are often written as different subdivisions of the same tempo when scored or transcribed using Western notation, as shown in Figure 2, polyrhythms can also be thought of as rhythms played in different, but mathematically related tempos.

Peter Magadini defines the elements of a polyrhythm as the juxtaposition of basic pulse (or primary tempo) and counterhythm(s), the accompanying pulse(s) that references a different division of time relative to the basic pulse [15]. A polyrhythmic piece or musical exercise can consist of a primary tempo overlaid by one or more counterhythms. The simultaneous awareness of the primary tempo and counterhythm, an often difficult task to perform just as a listener, has been called “metronome sense” [16].

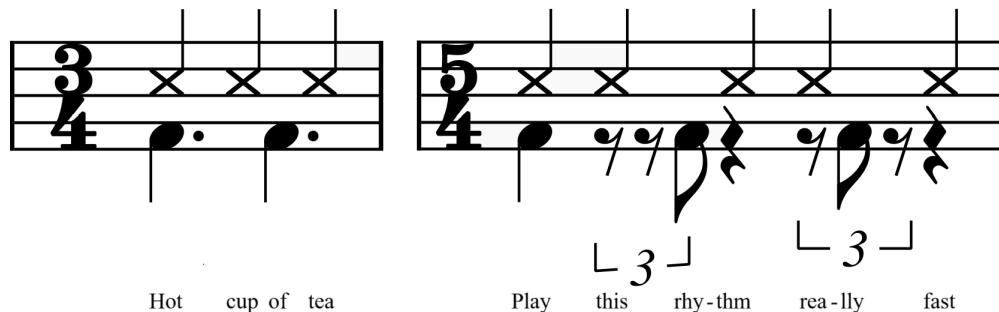


Figure 2. Western notation for 3:2 and 5:3 polyrhythms, with corresponding verbal mnemonics.

Gilbert Rouget said of polyrhythms: “When reduced to a single instrument, the rhythmic pattern is always simple. But the combination of these patterns in an interwoven texture, whose principles remain to be discovered, produces a complicated rhythm that is apparently incomprehensible” [17]. This trait makes learning polyrhythms, particularly by ear, relatively difficult. Due to brain-related limitations of listening to music, humans may have trouble focusing on multiple tempi at once [14]. Therefore, slowing music down with the aid of a metronome while learning, a longstanding learning strategy for simpler rhythms, is not so effective for polyrhythmic practice.

Alternative methods for polyrhythmic learning involve metric patterns in language [18]. For example, the phrase “hot cup of tea” can rhythmically represent the 3:2 polyrhythm as in the first example of Figure 2; “what atrocious weather” can represent the 3:4 polyrhythm; “I’m looking for a home to buy” can represent the 5:4 polyrhythm [19]. Figure 2 shows compatible phrases below the two written polyrhythms. However, these too are limited as there does not exist a helpful mnemonic phrase for every $m:n$ polyrhythm.

1.8. Action Observation (AO)

Action Observation (AO) is an application of the phenomenon in which observing the behavior of another person or avatar produces neural activity similar to that when performed by one's self. Such observational learning can be an effective approach for learning specific movements, including surgical skills [20] or for movements related to physical therapy [21,22].

AO can be implemented via the observation and mimicry of actions by another human, an avatar animated via motion-capture, or a virtual avatar with a combination of programmed animations, rigging, constraints, and models for motion. For example, in the authors' previous pedagogic tool [23], three different 1st-person VR AO perspectives were utilized in the learning phase of a pedagogic tool teaching rhythm: Photospherical Video, Programmed Animation, and Mocap Animation [23].

In a recent study [24], subjects who learned via 1st-person VR exhibited significantly better prosthetic limb control skills than those in a control group. In another study involving Tai Chi novices [25], positive results, specifically limb position error reduction, were observed while practicing using a system that rendered the movement of an instructor's avatar atop the users'.

Limitations of AO have been alluded to or observed in previous studies. For example, in the previously mentioned prosthetic-limb-training application [24], the benefits of AO training were observed for a bow knot test (asking a subject to use an unaccompanied hand and prosthetic arm to hold and tie shoelaces), but not for a simpler-to-conduct test that entailed picking up and moving blocks with the prosthetic limb. It was, therefore, inferred that VR-based AO is likely to be effective in relatively difficult, bimanual tasks of relatively high coordination [24], a description that fits the action of drumming complicated rhythms.

Sense of agency (SoA), the subjective feeling of initiating and controlling an action, is also an area where AO-based approaches may not be adequate [26]. The body schema is the unconscious representation of the self that supports all types of motor actions, and it can be updated via repeated body movements with a high SoA [26]. AO-based learning has a generally lower level of SoA, as it consists solely of following the body movements of another person or avatar. This, in turn, lowers the potential for a user to acquire an optimal body schema, preventing motor skill learning from achieving maximum efficiency.

Embodied simulation is a concept related to Action Observation and explores the possibility of sharing experiential realities. Paint with Me [27] is an application with which art students can embody the experience of a teacher to learn via movement-based instruction in mixed reality. Much like 1st-person AO-based motion learning, the user learns painting techniques by following another's movements via hand-tracking.

1.9. Virtual Co-Embodiment (VC)

Avatar embodiment is most conventionally implemented via a humanoid avatar with limbs controlled by a single user. Virtual Co-embodiment (VC), a relatively new concept and field of research, describes applications of embodiment with multiple users collaboratively controlling "joint actions" of a single avatar [28]. Studies show that the sensation of agency regarding actions not performed by oneself can arise [29], and VC is a means of achieving that effect.

Real-life analogs for joint actions include two-player pinball, when each player controls one flipper of a shared machine, and three-legged racing, when the inner legs of side-by-side runners are lashed together. Also related is the personal psychological attachment to a body part that is not truly attached to one's body, such as a fake limb. A body part assimilated into one's notional self can be called a body ownership illusion.

One implementation of VC includes weighted average-based VC, whereby the positions of multiple users' limbs are tracked and averaged to control an avatar. In a recent study [26], a weighted average-based VC application was developed for motor skill training. When used in a subjective experiment, learning efficiency was found to be higher than that of AO (perspective sharing with a more-skilled user) and solitary practice.

Allowing multiple users to control different sections of an avatar's limbs is a separate approach to VC [30]. By allowing two users to collaborate and control a robotic arm, more rotational freedom and inhuman movements were realized. Similar to weighted average-based VC, the adjustable mixing of movements allows an expert to scalably support a beginner trying to practice stable movements [30].

Another approach is called body-part-segmented VC, whereby multiple users control separate limbs as a single avatar. In a body-part-segmented VC study [28], SoA and ownership towards the avatar arm controlled by a partner were significantly higher when both subjects could see each others' targets, demonstrating the importance of visual information in terms of enhancing embodiment towards limbs controlled by others.

A focus of our project entails actualization of body-part-segmented VC with an autonomous entity. Rather than multiple users sharing control of a virtual avatar, control is shared between a single user and a program. This concept was a part of VC's original definition [31], but to the authors' knowledge, the realization of this concept for motor learning pedagogy did not heretofore exist.

An example of VC where a human shares an avatar with an autonomous agent is in a contribution exploring social presence for robots and conversational agents [32]. Pondering whether or not autonomous entities and robots should have a human-like social presence, that work experimented with social presentation bound to a single body and more extraordinary capabilities where social presence could be migrated across avatars. A study showed that participants felt comfortable with such capable agents and with the comparatively seamless and efficient experiences that resulted.

1.10. Rhythm Games

A genre of music-based video games, so-called "rhythm games", allows players to play along to music on drum pads or other instrument-like interfaces. Using a dynamic "note highway" or "notefall" visual display, players interpret and perform rhythms to play, as running and final scores are displayed and reported. Representative instances include Parappa The Rappa (1996) [33], Guitar Hero Live (2015) [34], and Beat Saber (2018) [35].

While these games might provide experience somewhat similar to physical and mental musical practice, the lack of authentic musical instruments and formal musical notation prevent the genre from being widely considered a serious educational tool. A previous study [36] suggested that players skilled at rhythm games improve performance in terms of perception and action rather than sensitivity to metrical structure. Other music video games such as Rocksmith [37] teach players how to play the guitar, but their interface is based on notefall and guitar tablature and does not incorporate AO or VC techniques.

NeuralDrum is an extended reality interactive game that measures and analyzes the brain signals of two users as they play drums together [38]. Brain synchronization allows people to better work together [39], occurring when two individuals perform the same activity simultaneously. The Phase Locking Value is calculated to measure brain synchronicity, and it is used to alter the audio and visuals within the game in real-time. While subjects reported relatively high enjoyment, synchronization was not easily perceivable between pairs of subjects [38].

2. System

2.1. Hardware and Sensors

We developed a system to explore AO- and VC-informed drumming pedagogy. Our system uses a Roland TD-25 Drum Sound Module sensing user performance on Roland PD-7 electronic drum pads, the integrated sensors on the Meta Quest 2 head-mounted display (HMD), and a workstation or laptop computer running the Unity 2023.1.0b6 game engine. The components and connections are shown in Figure 3. Each PD-7 dual-trigger drum pad is a 7.5" rubber pad with two piezo-electric transducers embedded in the pad's rim and center [40]. When a pad is struck in either area, an analog, impulse-like signal is output via a stereo 1/4" cable to the sound module via the trigger input.

A workstation runs the virtual scene developed in the Unity editor while connected to the sound module and a Quest 2 VR helmet. A high-speed USB 3 cable is used for the HMD connection, which allows it to be used as a thin client display.

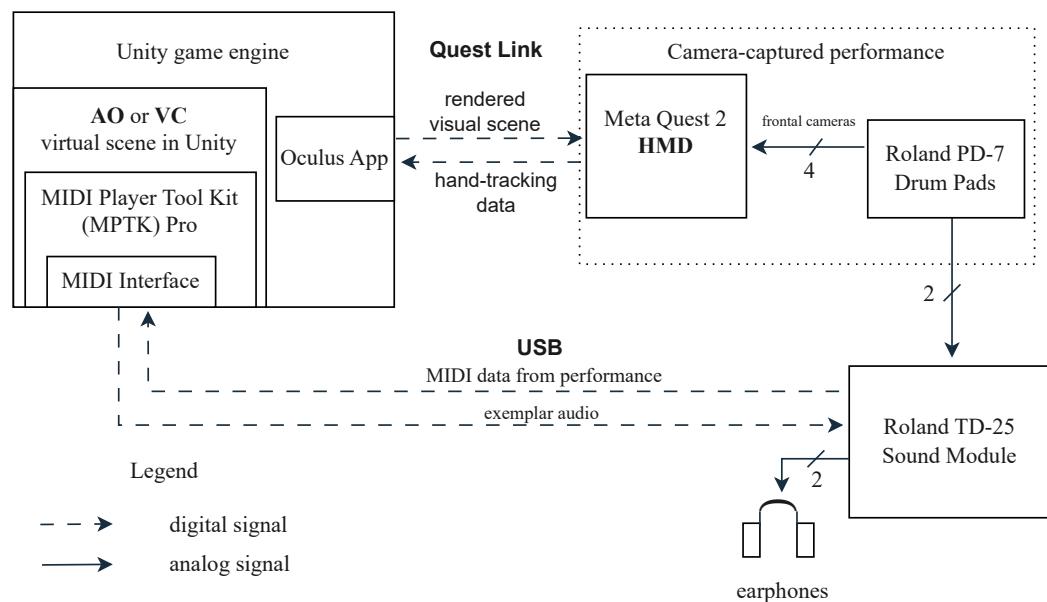


Figure 3. System schematic.

2.2. Development

The Unity virtual environment (VE) for our system utilizes assets including Maestro Midi Player Tool Kit Pro (for real-time MIDI capability) [41] and Open XR (for VR/AR development) [42]. The Oculus desktop application and the Quest Link function on the Quest 2 are used to display the scene through the HMD.

Each virtual scene is populated with drum and drum stick elements. The positions and sizes of these GameObjects correspond to the physical arrangement of the Roland drum pads and the physical drum sticks held by the user. Using programmed keyframe animation, the drum heads atop the drums within the scene momentarily dilate when struck. The exemplar drum sticks similarly use programmed animations that the user is expected to look at while practicing. A short video summary of the system is publicly available: <https://www.youtube.com/watch?v=d-pw7aZpvME>.

The binaural auditory soundscape (a mix of the user performance on the drum pads and the virtual scene's audio) is displayed to the user via stereo earphones driven by the audio output of the Roland sound module. Via built-in MIDI soundfonts, the module can express performance on the drum pads with a variety of virtual percussive digital instruments. The module sends the MIDI data of the user's performance to the computer over USB while also receiving the virtual scene's rendered audio stream from the workstation. Audio displayed to the user via earphones is a perceptually balanced mix of the exemplar avatar's performance in the virtual scene and the user's own performance.

The computer reads the MIDI data of the performance to control concurrent feedback elements in the scene for the user. The timing of the notes from the exemplar avatar's performance and the user's performance are compared in real-time, determining whether each note is played too late, too early, or accurately timed (within a tolerance). These data are used to control illuminating light objects within the scene. Temporary illumination of a red light above the virtual drum positioned towards the drum head indicates the last strike was too late, while a green light signifies the note was played too early. If the pad is struck within the timing tolerance, neither light object is switched on. This is similar to practice with a metronome: a click track can be heard when successive notes are played on

either side of the beat, but often becomes nearly inaudibly masked when a player “locks in” to the metronome’s expressed tempo. Tolerances for this are currently set such that, if a note is played within 25% of its duration, either ahead of or behind the beat, it is considered on-time.

Another aspect of the real-time feedback is the user’s hands being tracked. Through this visual cue, timing of strokes is better understood and potentially easier to compare with those of the exemplar performance, encouraging microadjustments. The user’s perspective is entirely virtual, but hand-tracking allows a virtual representation of the user’s hands to be seen.

2.3. Action Observation (AO) Scene

In the AO scene, the user wears an HMD and experiences a virtual scene while drumming, as shown in Figure 4. The user is expected to drum in-sync with the exemplar sticks of the virtual scene as the concurrent feedback is conveyed. Because of the lack of SoA, it is expected that this virtual scene is not as beneficial as a VC-based scene, but it is predicted to be more effective than a video. The AO scene may not be optimized for the cognitive phase of learning, but it could excel in the associative and autonomous phases. The user’s point of view of the virtual scene set up for AO practice can be seen in Figure 5.



Figure 4. A user interacting with the tool in AO polyrhythm mode: the user performs a polyrhythm on two drum pads in sync with an exemplar avatar observed in the virtual scene.

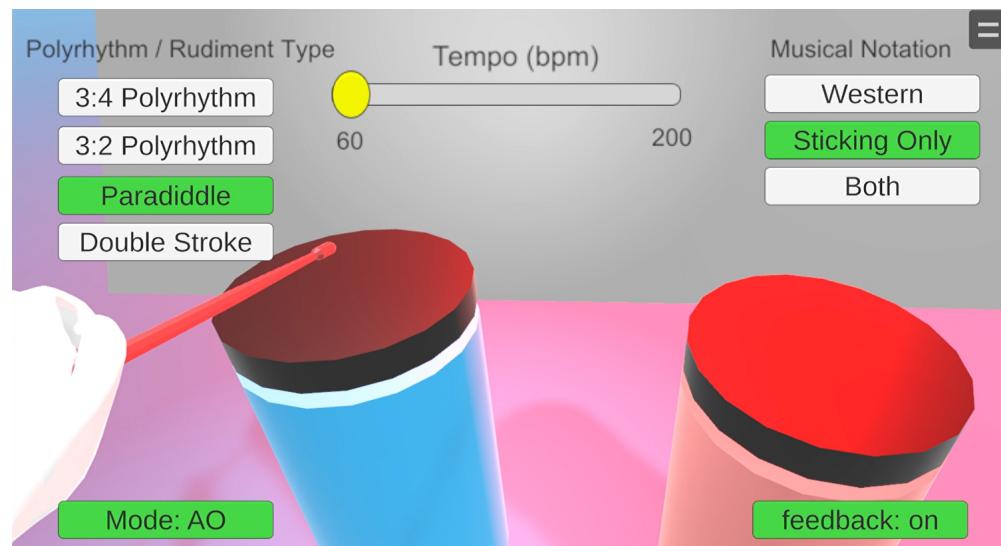


Figure 5. The developed VE can host practice sessions for different rudiments and polyrhythms via AO or VC with real-time hand-tracking. The UI includes controls to achieve a wide range of tempos, feedback options, and other visual aids. This screenshot shows the AO virtual scene at a moment when the user is playing slightly ahead of the beat, as indicated by the red overhead light GameObjects.

2.4. Virtual Co-Embodiment (VC) Scene: The “Halvatar”

While weighted average-based VR-based VC has had observable previous success, foreseeable problems arise when considering this approach for an application purposed for the teaching of drumming. The trajectory of a drum stick in use is inherently full of sharp turns and direction changes that can occur almost instantaneously. While weighted average-based VC techniques have had success in slower actions with smoother trajectories of motion, applying these same techniques to a drumming tool could result in double hits or glitchy motion when multiple users controlling an avatar are at different phases of a drum stroke. Therefore, an approach similar to body-part-segmented VC, as opposed to a method based on weighted averaging, was implemented for this tool.

We call the variation on the body-part-segmented VC concept proposed by this work a “halvatar.” Rather than delegating control of an avatar’s disparate limbs to human users, a portion of the avatar is controlled by a single user, while another portion (in our approach, the opposing hand and corresponding stick) is controlled exemplarily through programmed animations. This is opposite to body ownership illusions, as parts of a self-identified avatar are shared with an external agent. Embodying a halvatar in practice session results in a duet with an idealized half. This approach avoids complications that would otherwise arise through a weighted average-based VC approach while still maintaining higher levels of SoA.

This approach is particularly suited for polyrhythmic applications. Since the primary pulse and counterrhythm elements of a polyrhythm can be thought of as two musical phrases referencing different tempos, the halvatar is similarly split, the primary pulse expressed on one side and its counterrhythm expressed on the other. VC compartmentalizes the polyrhythm to increase focus on its simpler elements, seemingly helping most with the cognitive phase of learning.

The practicality of the halvatar extends beyond polyrhythms as well. Drum lines in marching bands often consist of multiple-sized bass drums, each carried and played by a different drummer. Sometimes, the bass drum rhythm is “split” amongst members playing different parts of the line to take on melodic properties [43]. Only when an entire line’s members are successfully performing their disparate parts together can the bass drum part be heard in full. Hocketing is a similar and more-generalized concept that involves instruments (not only drums) playing interlaced portions of a single musical line [44]. The

halvatar's self-duo concept seems to work well for such call-and-response arrangement styles in addition to polyrhythmic phrases.

3. Pilot Study

3.1. Pilot Study Summary

We conducted a pilot study to broadly explore the quantitative effects of our VE and to gather qualitative feedback from subjects. Specifically, we addressed the two following questions. Do practice sessions for drumming using VR-based AO or VC tools offer a measurable benefit over standard practice methods? How do users respond to our development and to practicing drumming in VR in general?

Based on the previously discussed findings, VR-based AO and VC tools have the potential to yield observable positive effects in motor learning pedagogy applications. When considering these techniques in drumming applications, we hypothesized that practicing with either modality would yield better results than practicing with video demonstrations.

To test this hypothesis and to gather baseline data and user feedback moving forward, we conducted a pilot study in which subjects' execution of drumming exercises was recorded before and after three different types of short practice sessions: Video, VR-based AO, and VR-based VC. Subjects were randomly assigned to a group and used one of these three practice methods between an initial and a final recording. To gather subjective feedback, subjects were asked to complete a user experience questionnaire (UEQ) shortly after experiencing his/her assigned practice mode.

3.2. Participants

There were 15 subjects (12 male, 3 female), students recruited from the university, who participated in the pilot study (age: $M = 23 \pm 4 (SD)$). Before the start of the experiment, subjects self-reported their age, dominant hand, whether or not they were able to read Western musical notation, musical experience, and primary and secondary instruments (where applicable). Of these, 86% were right-handed, 40% had some prior experience with musical instrument practice, and 20% were currently active performing members of university music clubs. Subjects were randomly assigned to one of three groups corresponding to three modalities they were to use when practicing two rudiments and two polyrhythms throughout the experimental procedure, as shown in Figure 6. The group assignments (Video, AO, and VC) came into effect during the practice session phase of each section (tutorial, rudimental, polyrhythmic) of the experiment. Subjects were given information and instructions via a script read aloud before they were given an informed consent form to sign. The experiment then proceeded from the tutorial.

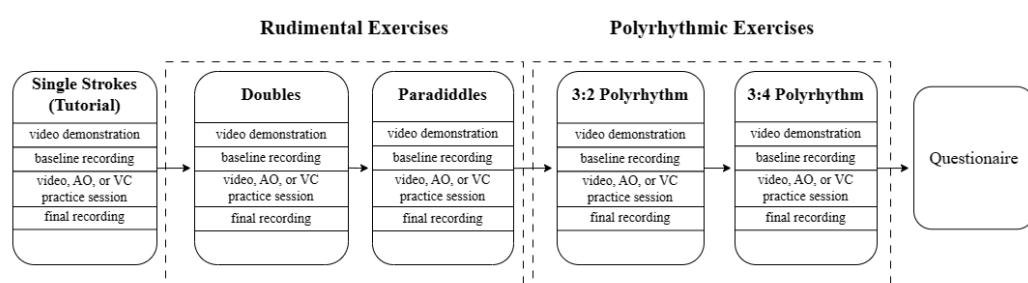


Figure 6. Procedure for the pilot study.

3.3. Tutorial and the Four-Phase Procedure of the Experiment

All subjects first completed a four-phase introductory tutorial to help familiarize themselves with the flow of the experiment and their assigned system. Subjects were told that they could ask questions or adjust the height of the seat and drums any time throughout this tutorial. Simple single-stroke eighth notes were used as the tutorial's lesson, as shown in Figure 7.

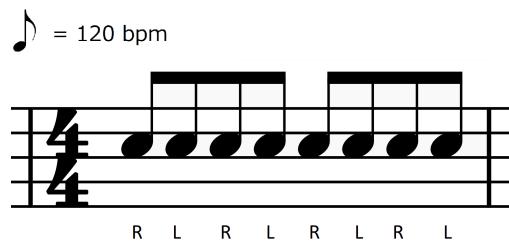


Figure 7. Western notation representation for single-stroke eighth notes, the exercise used for the tutorial.

The first phase of the tutorial consisted of showing a video demonstration to the subject via a computer monitor and stereo loudspeakers. The video was a performance of single-stroke eighth notes filmed with an overhead camera placed behind a drummer, as shown in Figure 8. The video started with a metronome for one measure before the demonstration of the exercise began. The exercise was played for four measures, a 22 s video in total. The audio of the performance in the video was quantized and panned, to rectify timing imprecision and to achieve stereo separation, respectively.

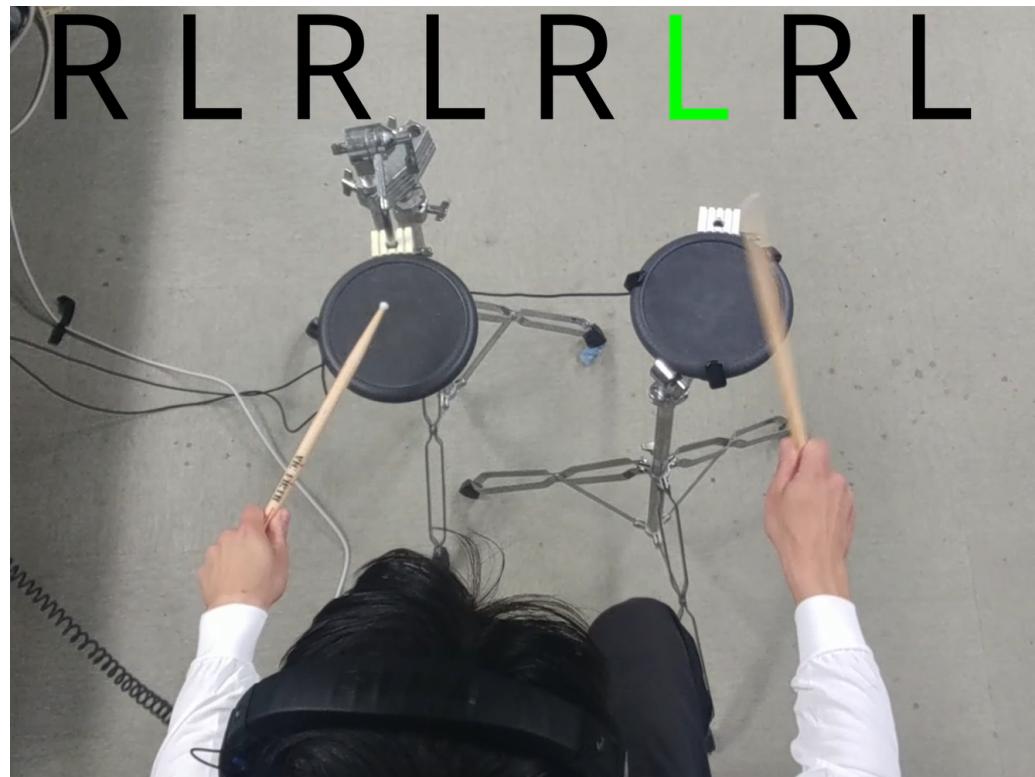


Figure 8. The first phase of all sections of the pilot study was a 3rd-person (overhead view) video demonstration of the corresponding exercise.

Subjects were asked to watch the video attentively before receiving a pair of medium-weight drum sticks and recording their own four-measure take of the previously observed single-stroke eighth note rhythm. As in the overhead demonstration video, the metronome played for the subject at eighth note = 120 beats per minute (bpm), where the metronome expressed every eighth note beat and added a downbeat accent on every fourth beat, or the "1" and "3" of each four-beat measure.

After the baseline recording, subjects were asked to conduct a short practice session with their assigned mode of practice: Video, AO, or VC. If the subject was in either VR group, they were helped with the set-up and fitting of the HMD. Regardless of the assigned group, each practice session had a duration of about two minutes, six repetitions of the exercise, with a small break after each repetition.

In the case of subjects in the Video or AO groups, they were asked to practice along with their assigned exemplar. In the VC case, subjects were given slightly different instructions. Rather than simply playing along with the exemplar, a third of the session was for practicing the left-hand rhythm while the halvatar's right hand drummed autonomously. Another third was to practice with the right hand while the halvatar's left hand autonomously drummed. A final third was for practicing both hands simultaneously.

The last phase of the tutorial was a final recording of the single-stroke eighth notes. Each subject was instructed to once again record a four-measure performance of the previously practiced eighth note exercise. Though this performance was recorded, the data were not analyzed as this portion of the experiment was only a tutorial, intended to get the subject accustomed to the flow of the following four sections. For the proceeding exercises, the pre- and post-practice session recordings were both recorded and analyzed, to measure the improvement of each subject.

3.4. Rudiments

After the tutorial, subjects were asked to complete the same four-phase procedure for four musical exercises, the first two of which were rudiments, specifically doubles and paradiddles, shown in Figure 9. Rudiments were chosen as they are the building blocks of drumming, practiced by players of all levels. Beginners may practice them slowly to learn the sticking, while advanced players may practice them at challenging tempos or with heightened focus on timing accuracy.

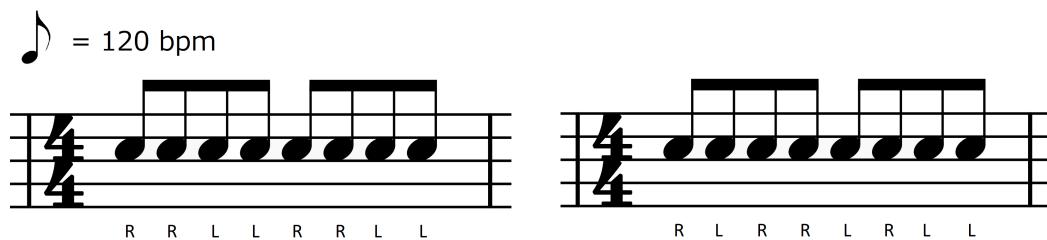


Figure 9. Western notation representation for eighth note doubles and paradiddles, the exercises used for the pilot study's rudimental section.

Much like the tutorial, each subject first viewed an overhead video (3rd-person perspective) of a doubles performance before baseline performance recording. Then, the assigned group-dependent practice session and final recording were conducted. After this, the process repeated for paradiddles.

3.5. Polyrhythms

Much like the doubles and paradiddles sections, the same four-phase process was repeated for two polyrhythms. The first learned polyrhythm was the 3:2 polyrhythm, as it is the most common and may have been heard or performed without necessarily having been recognized as a polyrhythm. Following that was the 3:4 polyrhythm, which, while less common, can also be found in pop music.

It is likely that both of these polyrhythms were considered to be more difficult than the double-stroke eighth notes or paradiddles by most subjects. This is because the interval between the exercise's notes varies, and because some notes of the counterrhythm (bottom lines on Figure 10) fall between divisions of the beat expressed by the metronome.

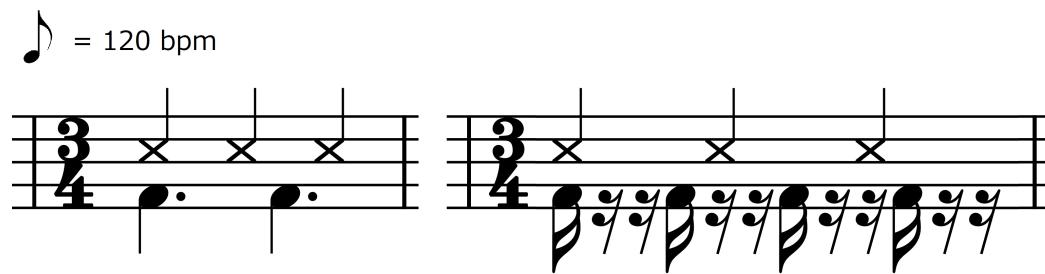


Figure 10. Western notation representation for 3:2 and 3:4 polyrhythms, the exercises used for the polyrhythmic section.

3.6. Practice Session Phase

As previously outlined, each exercise, including the tutorial, contained four phases, corresponding to the rows of the blocks in Figure 6. The third phase was the practice phase and was the only phase dependent on the subjects' assigned group. The first group was the group assigned to use videos shot in 1st-person perspective to practice. Within the videos' frames were the exemplar drummer's hands, sticks, and the two pads being performed on, as shown in Figure 11. Like the overhead 3rd-person perspective demonstration videos, these videos started with one measure of the metronome followed by four measures of performance. Afterward, there was a short pause and the video was looped five additional times.



Figure 11. For subjects of the Video group, the third phase of all sections of the pilot study was a 1st-person ("PoV") video demonstration of the corresponding exercise looped 6 times.

The other two groups, VR-based AO and VR-based VC, similarly presented the subject with a 1st-person perspective and a performance of the same length, but within a virtual scene displayed to the user via HMD. Shown in Figures 12 and 13, the interface of the virtual scene was simplified from that of Figure 5. Rather than present the user with options and drop down menus, the virtual scene featured only the sticking cues, drums, and drum sticks. The real-time feedback lighting GameObjects were similarly disabled, as users of the Video group did not have such aids, and it was desired to keep the variables as consistent as possible across the three groups.

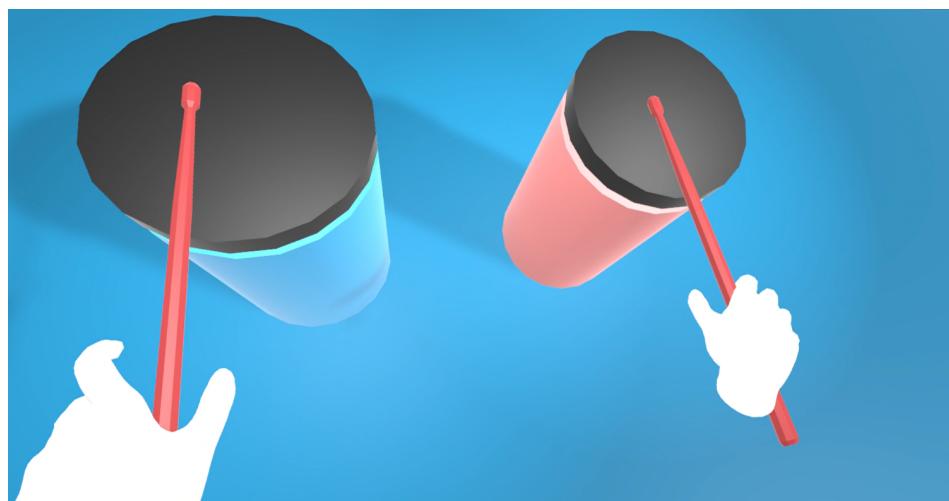


Figure 12. Polyrhythmic AO virtual scene: The user's hands are tracked and displayed in a scene with automated drum sticks performing exemplarily.

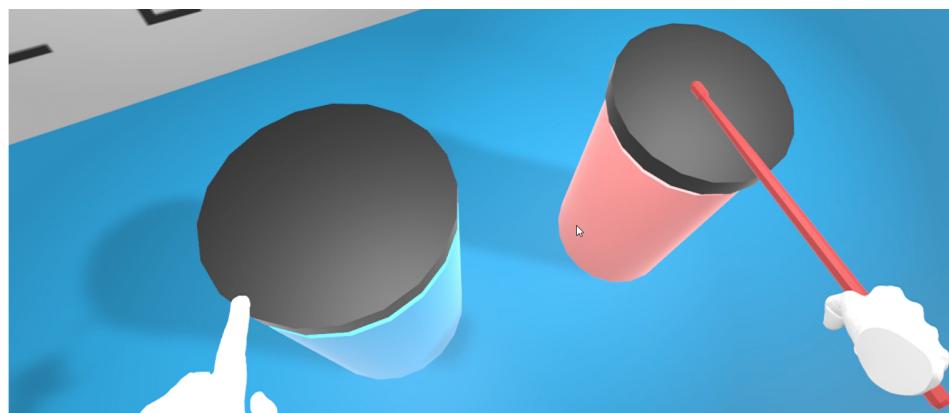


Figure 13. Polyrhythmic VC scene: The left hand of the halvatar is controlled by the user, and movements of the right hand and stick are controlled via programmed animation.

There was no verbal instruction in the video, AO, or VC practice session media. It was also decided to not display Western musical notation of the exercise due to varying familiarity with it among subjects. However, an animated sticking diagram was displayed in all practice modes, as shown in Figure 14. This helped subjects keep track of where they were within the exercise.



Figure 14. In the video and AO and VC virtual scenes, a visual aid to help subjects learn the sticking during practice sessions is included. Similar to the sticking diagram in Figure 8's 3rd-person video, the sticking is displayed at the top of the video screen and placed behind and above the drums in the virtual scenes. This shows the entire sticking pattern while highlighting the current note.

In all the practice session versions, the musical exercise, tempo (eighth note = 120 bpm), and expression of the metronome were kept consistent. The right drum pad corresponded to a floor tom sound, and the left corresponded to a snares-off snare sound. This was

chosen for both sonic and practical reasons. The floor tom and snare occupy significantly different frequency ranges, and the set-up used in the experiment was similar to common placements of a snare and floor tom pair.

Both drum samples were stereo .WAV recordings, where the sustain of the drums' resonance was kept short through acoustic treatment and audio processing. Damper pads and transient shaper audio plug-ins were used to shorten the length of the note, as this articulated each hit and better clarified to the subject the timing of the strike.

For the video playback and AO groups, subjects were asked to study or play along with the teacher or exemplar avatar for all six repetitions of the exercise, whereas in the VC group, subjects performed the right and left hands' rhythm individually twice before combining them in simultaneous performance for an additional two repetitions.

3.7. Post-Experiment Questionnaire

After completing the four-phase procedure for the 3:4 polyrhythm, subjects were asked to complete the user experience questionnaire (UEQ) [45] to gauge the quality of experience of the video, AO, and VC modalities. The scales of the questionnaire covered a comprehensive impression of user experience. Both classical usability aspects (efficiency, perspicuity, dependability) and user experience aspects (attractiveness, stimulation, novelty) were measured. Within the survey, users were prompted with a spectrum between bipolar adjectives. Using a discrete Likert scale from 1 to 7, subjects' evaluations were made. Per the UEQs' standardized methods, scores were later normalized to a -3 to +3 scale for final analysis.

4. Results

4.1. Results of Rudimental Exercises

All subjects of the pilot study successfully performed doubles during both the second and fourth phases, before and after the practice session. Three subjects, one subject in each of the video, AO, and VC groups, were not able to play the paradiddle at the time of the initial recording, but were able to after the practice session. One subject in the VC group was not able to play the paradiddle before or after the Phase 3 practice session. These results are shown in Figure 15.

		Number of subjects (out of 5) who successfully performed Doubles	
		Before practice session	After practice session
Video		5	5
AO		5	5
VC		5	5

		Number of subjects (out of 5) who successfully performed Paradiddles	
		Before practice session	After practice session
Video		4	5
AO		4	5
VC		3	4

Figure 15. Number of subjects who could perform the correct sticking for the rudimental exercises before and after their assigned practice session.

For subjects who successfully played the exercise before and after the practice session, jitter was calculated to determine improvement in each subject's rhythmic accuracy. As shown in Figure 16, unquantized performance's timing nonidealities can be seen and heard

in a digital audio workstation, particularly when there is a metronome click track present for reference.

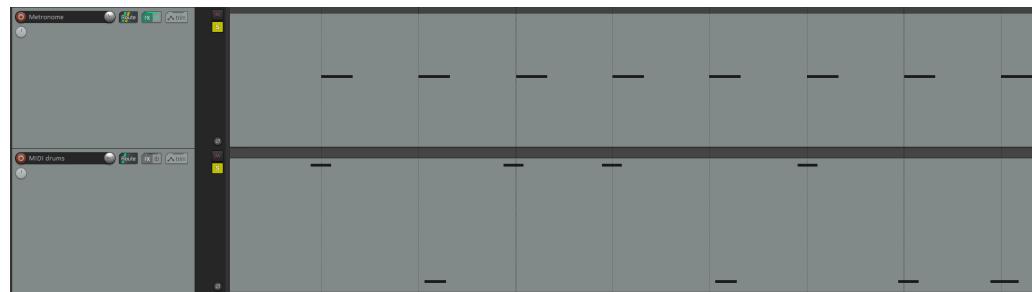


Figure 16. When viewed in a DAW, jitter can be observed between the uniform timing of the metronome (above) and the recording of a paradiddle (below). Relative to the gridline, some strikes of the paradiddle are early or late, whereas the metronome’s timing is uniform throughout.

Recordings made by the subjects were exported as MIDI files and converted into csv files with an online tool [46]. The portion of the MIDI files containing the timing of each stroke was used and compared against the timing of the metronome. Time differentials (in fractions of a beat, relative to the tempo of eighth note = 120 bpm) between each subject’s every stroke and the metronome were averaged for this calculation.

For doubles, all five subjects in each group had before and after data to compare. VC yielded an average decrease of a quarter of a beat (about a thirty-second note at an eighth note = 120 bpm, or about 120 ms), and video had an increase in jitter of about 80 ms. Conversely, in the paradiddle case, subjects had an increase in jitter of about 130 ms after practicing with VC, whereas the average jitter after a practice session with video had a near-zero decrease. Despite a significant decrease in jitter for VC subjects learning doubles, an improvement on the subjects’ averaging timing accuracy, there was an increase in jitter during the paradiddles case, alluding to inconsistency. Changes in jitter for all rhythmic exercises and modes of practice are plotted in Figure 17.

Change in Average Jitter for Rudimental Exercises Before and After Practice Session

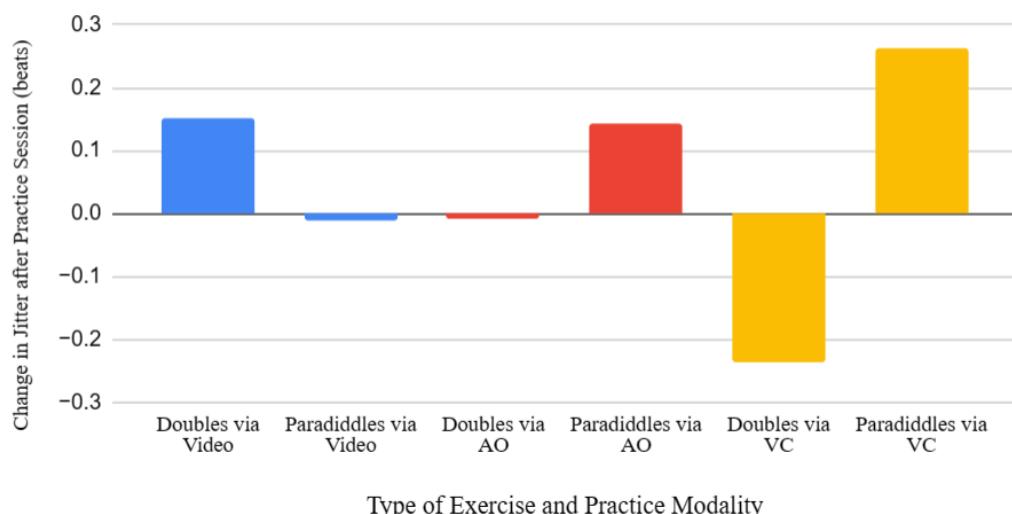


Figure 17. Plot to visualize increases and decreases in average jitter for the two pilot study rudiments after video, AO, and VC practice sessions, respectively.

In many cases, subjects’ technique and grip on the stick resulted in double triggering, meaning one intended stroke would be captured as two hits. Such double triggering was ignored, and only the first note of subjects’ strokes were considered for jitter calculations. It

was also common that subjects' continued to play the rhythmic exercise after the instructed four measures during recording. In those cases, notes after this duration were ignored. Some subjects stopped the exercise a few notes early. In those cases, the notes not played were left out of the averaging calculations.

4.2. Results of Polyrhythmic Exercises

As expected, subjects were not able to perform the two polyrhythmic exercises as proficiently as rudiments. It was decided before the experiment to measure the improvement of polyrhythms via the number of consecutive successful polyrhythmic strikes by a subject. Most subjects were not able to maintain a polyrhythm, especially the 3:4 polyrhythm, for the entirety of the recording. When analyzing the performances, the maximum number of consecutive hits per two-measure phrase (an 8-note phrase for the 3:2 polyrhythm and a 12-note phrase for the 3:4 polyrhythm) during the entire performance was determined for each subject.

In Figure 18, the average numbers of consecutive successful hits for the 3:2 polyrhythm before and after the practice sessions are displayed, with respect to practice mode. All subjects were able to perform eight notes of this polyrhythm correctly with the exception of one subject of the Video group, who scored 5 out of 8 hits.

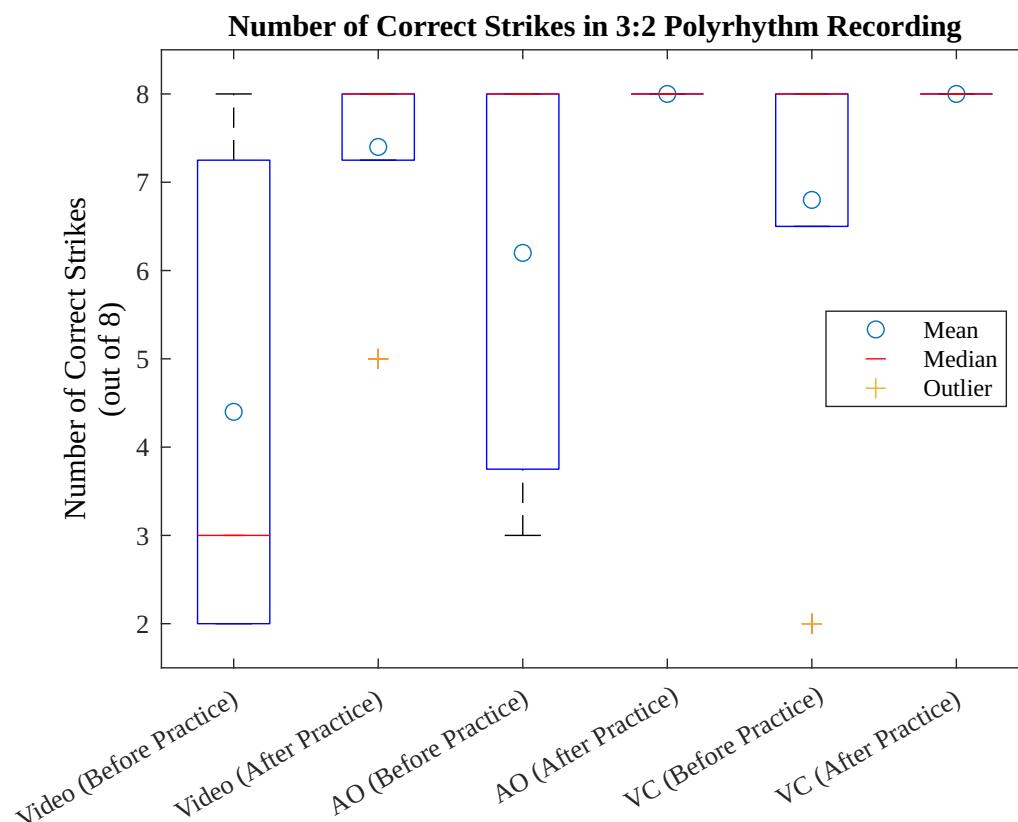


Figure 18. Comparison of successful strikes in 3:2 polyrhythm performance before and after practice session for Video, AO, and VC groups.

In Figure 19, the before and after data for the 3:4 polyrhythm are shown. In both figures, all practice modes show improvement, with some cases reaching the corresponding maximum number of hits (eight or twelve). In terms of the median and mean, the VC group's after-practice session results outperformed the Video group's results, whereas only the AO group's median was higher than the Video group's. The Video group's mean result after the practice session was higher than the AO's after-practice session mean.

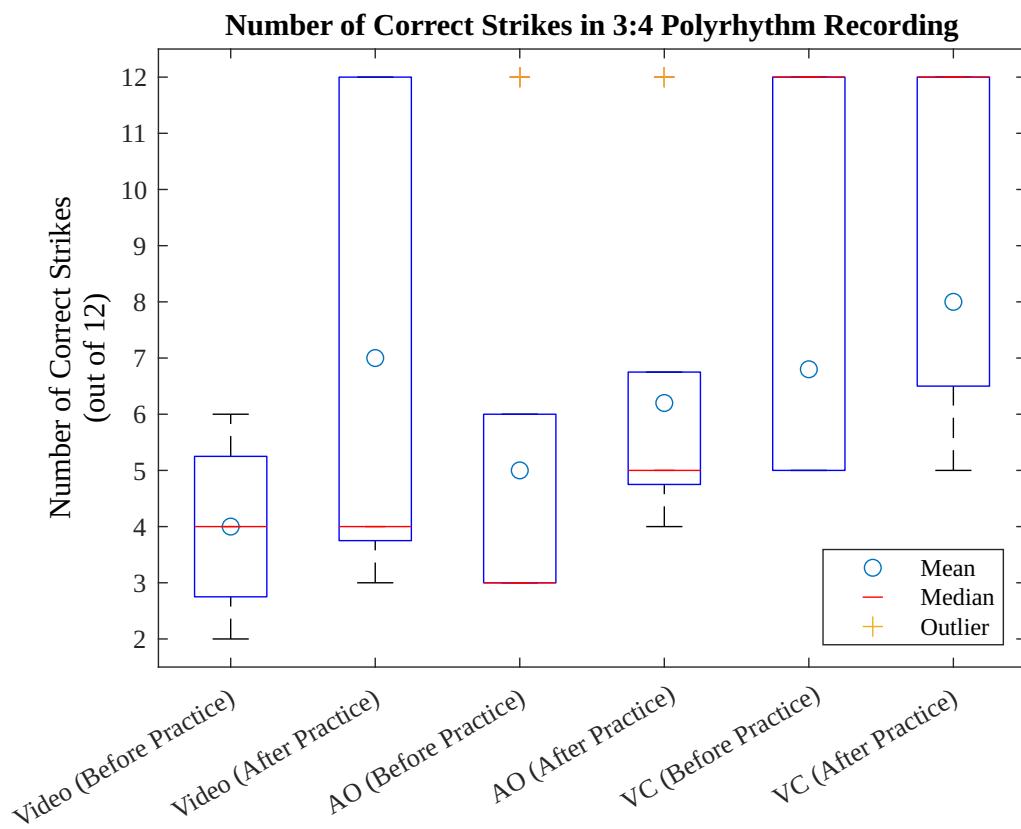


Figure 19. Comparison of successful strikes in 3:4 polyrhythm performance before and after practice session for video, AO, and VC groups.

One unanticipated complication not accounted for before the experiment was subjects' tendencies to invert (mirror-reflect) the sticking, which happened for 10 subjects (3 in video condition, 4 in AO condition, and 3 in VC condition) in the initial recording phase of the 3:4 polyrhythm. After the practice session, only 2 subjects (1 video and 1 VC) performed their final recording with inverted sticking.

4.3. UEQ Questionnaire

The UEQ survey was used to measure each subject's evaluation of the user experience. After each subject used the electronic drum pads and either the video, AO, or VC tools, an experience that lasted about 30 min, the 26-question survey was administered. Through UEQ data analysis tools, the results were organized into six factors, and a high-level comparison can be seen in Figures 20–22. The data's range was scaled from -3 to $+3$, and as shown, the subjects' response to the video was mostly comparable with the two VR experiences. Exceptions included perceived novelty, where the VC group's mean achieved a score 1.3 points higher than that of subjects using video, and efficiency, where the Video group's mean achieved a score 0.55 points higher than the VC group.

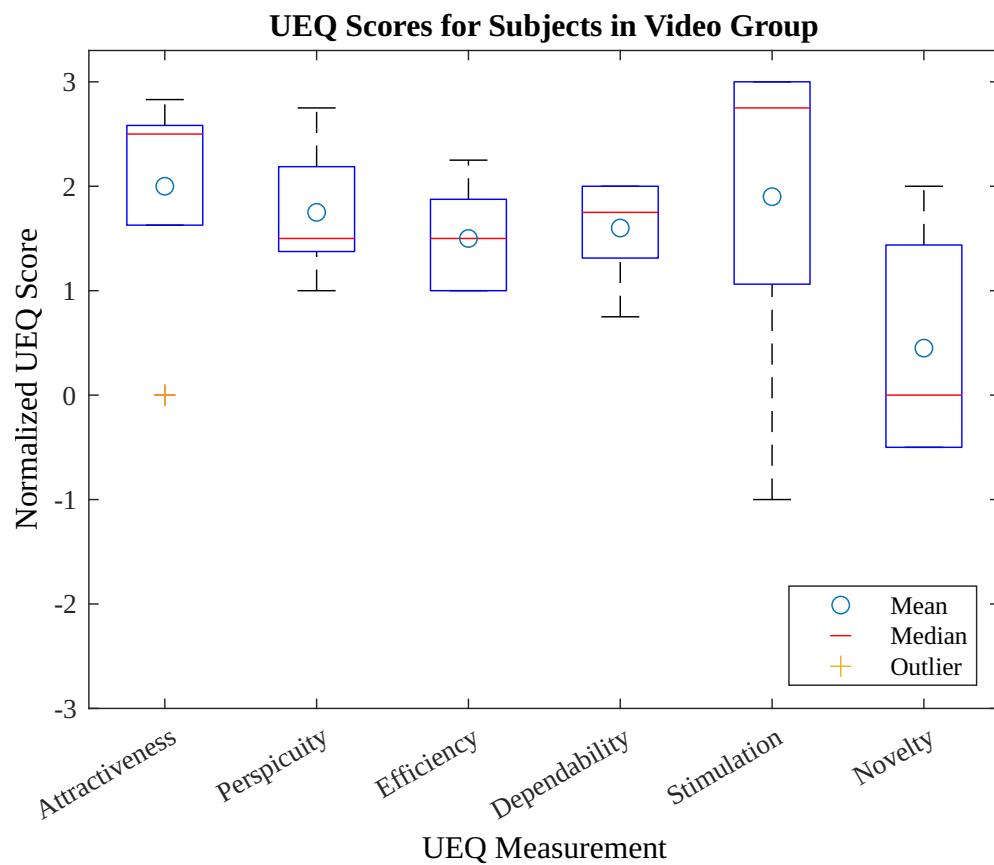


Figure 20. Questionnaire results from Video subjects after the experiment.

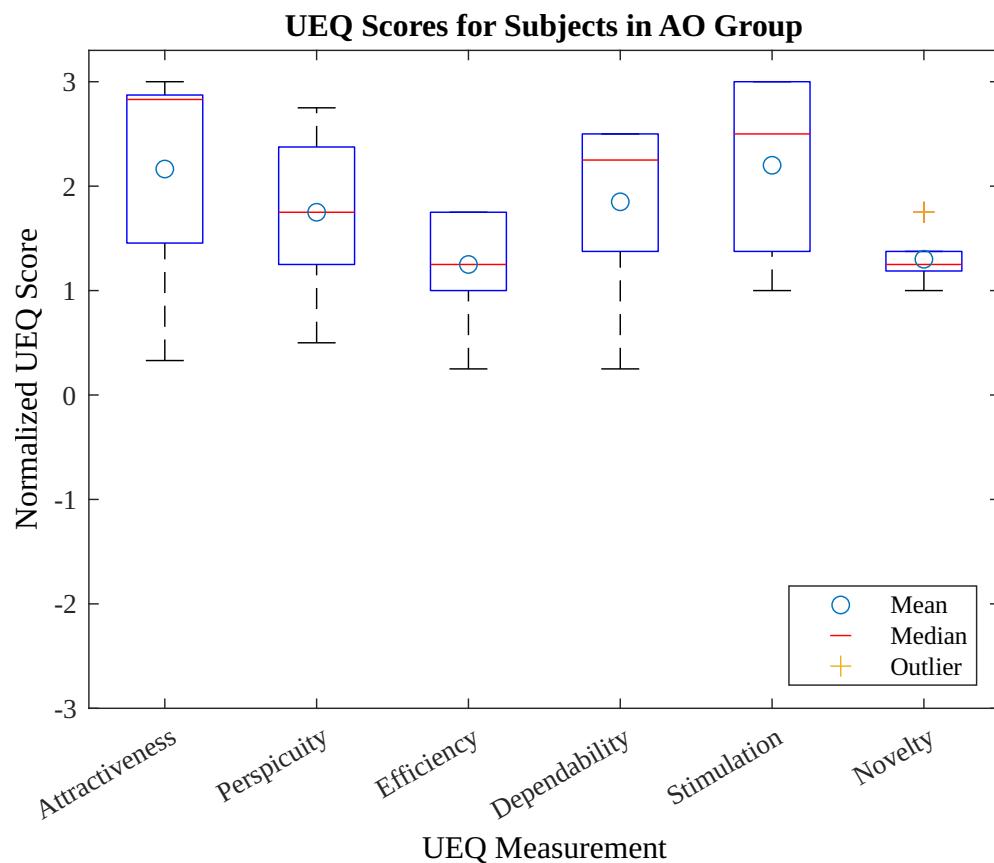


Figure 21. Questionnaire results from AO subjects after the experiment.

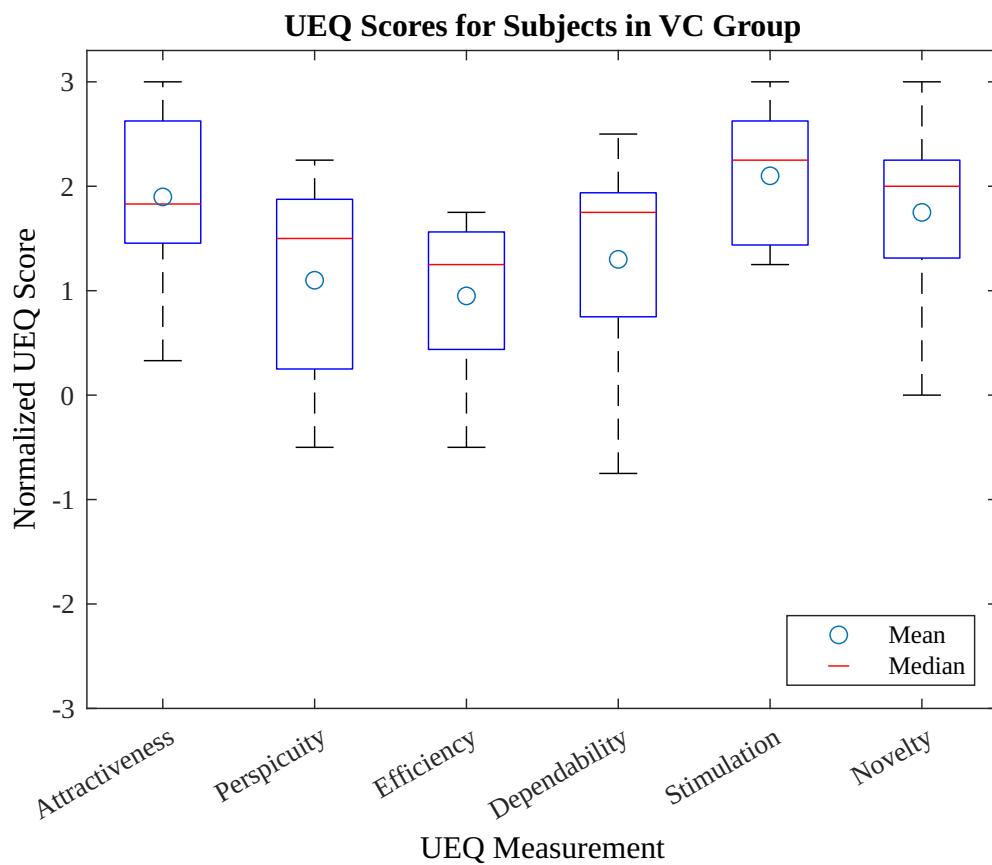


Figure 22. Questionnaire results from VC subjects after the experiment.

4.4. Other Observations

Due to his familiarity with the system, one of the authors conducted the pilot study, including recording the subjects and setting up the demonstration videos and HMD used by the subjects. During the pilot study, several behaviors were observed.

Some of the students within the AO and VC groups did not directly view or interact with the drum stick animations in the practice session. Instead, they were entirely focused on the dynamic sticking charts in front of the drums. Although these were deliberately placed in the VR scene such that both could be viewed simultaneously and subjects were instructed to focus on the movement of the animated sticks, some students left the moving sticks mostly out of their field of view. In previous AO and VC studies, allowing the subjects to clearly see such movements has been correlated with successful results [24,28]; so it is possible that this unanticipated complication had an effect on the AO and VC results in this pilot study.

One student performed a 3:2 polyrhythm perfectly after an AO practice session where he did not perform it correctly once, which was surprising. In most cases, successful performances following the practice session meant that the subject was able to follow the exemplar and show some comfort with the rhythm before final recording.

Feedback regarding general excitement and enjoyment of the pedagogical technique seemed high. Students who learned even with just the video expressed positive final comments in the questionnaire. For subjects with and without musical experience, the process of learning how to drum was perceived as fun.

5. Discussion

This development was successful in creating a tool for rhythmic pedagogy that incorporates both AO and VC techniques. The pilot study, while successful in some respects, did not show promising results in others. The VC group achieved a significantly higher novelty score on the UEQ survey relative to the Video group, but received a lower score in

the efficiency category. Similarly, the average jitter results for subjects using VC to practice doubles showed drastic improvement compared to the other groups, but also a significant increase in average jitter relative to the other groups after practicing paradiddles. The VC condition taught a paradiddle to one of two subjects who could not successfully perform the rudiment prior to the practice session, whereas the other two conditions taught one out of one subject.

The skill and experience level of recruited subjects had wide variation, with many subjects having no musical experience and some being performing drummers in bands. In addition, the exemplar avatar in the AO scene and the autonomous portion of the halvatar in the VC scene were designed with programmed animations rather than imported Mocap data. This could have had an effect on subjects' perception and experience compared with the video, as that practice session modality was based on a human performance. Goals and ideas for future work follow.

5.1. Further Quantitative Experimenting

The pilot study was successful in obtaining initial data and subjective feedback while various, mostly musically novice, users interacted with our system. In the near future, we would like to perform more rigorous testing with subjects.

This includes advanced polyrhythm pedagogy trials with experienced drummers. In past AO studies [24], it was presumed that VR-based AO is likely to be effective in complex, bimanual tasks. Having an experienced drummer learn a seldom-performed, technically difficult pattern such as a 13:9 polyrhythm could yield interesting discoveries.

Retention of learning was also not explored within the scope of the pilot study. VC excels in motor learning because of its ability to update the body schema of users, as opposed to limiting users' involvement to following a teacher. Because a relationship between the body schema and long-term retention of motor skills has been observed in previous studies [26], it seems possible that the VC mode of our developed system has unexplored potential in cases of retention, as opposed to immediate results. Such analysis, however, was beyond the scope of the pilot study. A future study across multiple days might show more-positive effects of VC-based drumming practice.

The pilot study was also limited to a single tempo for all groups. Modulation of tempo, however, is a commonly practiced rehearsal technique [47] and worth considering in future experiments. Comfortable tempos are subject-dependent, so granting subjects of future studies some control of tempo is desirable. Real-time feedback cues comprise another feature of the VE that was not utilized in the pilot study. Including this functionality in a build used in an experiment would similarly be of interest.

5.2. Drumming with Brushes

Brushes started to be used by drummers in the early 1900s as an alternative to sticks. Although still used in contemporary performance, brushes were often used to maintain quieter volumes on drums in acoustic musical settings of the last century. In addition to the staccato articulation that results from playing brushes in the same overhead manner as drum sticks, brushes could also achieve more legato, or smooth-sounding, notes by using lateral strokes. Brushes can also achieve infinite sustain by swirling bristles around a drum head in circular motion, very unlike the dynamic volumes that drums usually produce. This technique, or the combination of this technique with lateral or overhead strokes, makes up the bulk of the musical language of drumming with brushes.

Building off the successes seen during the experimental phase of motor learning as taught by weighted average-based VC [26], a next step of interest is developing a tool to teach the motions of brushes. Because brushes comprise a style of drumming that consists of smoother movement with fewer sharp direction changes, it could match well with a weighted average-based VC approach.

5.3. User-Shared Body-Part-Segmented Virtual Co-Embodiment Drumming Application

As the development of a body-part-segmented VC drumming application that uses programmed animations has been validated, a next step is to develop a tool that allows two drumming users to simultaneously occupy the same body-part-segmented avatar. In a previous motor-learning VC study [26], it was thought to be important that the teacher be a human who could make quick adjustments to adapt to the student's inexperience with the system and motion being taught. If this approach were to be realized in a drumming application, a human teacher could potentially correct for microtiming issues of a student and could provide alternative benefits. Comparing the results and subject feedback from the halvatar approach and this standard body-part-segmented approach could provide interesting findings.

6. Conclusions

The developed VE offers a new way to teach rhythmic exercises and rhythmic techniques. Technology-accelerated drum instruction often entails watching professional teachers' or musicians' demonstrations from a distance or from a 3rd-person perspective. Our project gives students the opportunity to learn through an exemplar drumming avatar via two VR-based 1st-person perspectives: AO and body-part segmented VC.

Based on the quantitative data of the pilot study, our VR-based AO and VC virtual scenes did not result in the observation of a measurable significant benefit over video-based drumming practice in our experiment. Subjects in all three groups were able to learn new rhythms, and the improvements in jitter for subjects in the VC group lacked consistency. The results of the survey showed VC to have a significantly higher perceived novelty score than the other two modes. However, VC scored slightly lower than AO and video in the other rating scales. Aside from its quantitative and qualitative results, this work also inspired ideas for future extensions and experiments.

This project represents a "softening" or generalization of the VC concept. Rather than rigidly configuring the control:display mapping between human pilots and digital puppets, as instantiated by humanoid avatars, we exposed the rigging to not only humans, but also surrogates, namely programmed animation for our teaching system, but generally any kind of exemplar (keyframe animation, IK-informed events, AI-driven dynamics, etc.), prepared offline (as in our system) or derived from live events determined and delivered at runtime. In ordinary scenarios, a single human controls each avatar "solo", but more-flexible arrangements can articulate an avatar, so that its control can be shared by multiple humans or algorithmic agents. Ultimately, mapping between human users and virtual or robotic expression is flexible, and a single human can "fan-out" to multiple displays, and multiple users in conjunction with computational agents can control arbitrarily assigned "fan-in" inputs.

Author Contributions: Conceptualization, J.P.; methodology, J.P. and M.C.; software, J.P.; validation, J.P. and M.C.; writing—original draft preparation, J.P.; writing—review and editing, M.C.; visualization, J.P. and M.C.; supervision, M.C.; project administration, M.C. and J.P.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: The Spatial Media Group at University of Aizu provided some financial support for this project.

Informed Consent Statement: Prior to their participation, the experimental participants provided written consent after being fully informed about the contents of this study. Additionally, explicit written consent was obtained from each subject regarding publication of any potentially identifiable data within this article.

Data Availability Statement: Data acquired and used in this manuscript can be found online: <https://github.com/jPinkl/VRDrumPedagogy>.

Acknowledgments: We thank the anonymous Referees for their useful and thoughtful suggestions. We also thank the subjects of the experiment and fellow members of the Spatial Media Group, especially Julián Villegas, Peter Kudry, and Juanca Arevalo.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Shuell, T.J. Phases of meaningful learning. *Rev. Educ. Res.* **1990**, *60*, 531–547. [CrossRef]
- Anderson, F.; Grossman, T.; Matejka, J.; Fitzmaurice, G. YouMove: Enhancing movement training with an augmented reality mirror. In Proceedings of the ACM Symposium on User Interface Software and Technology, St. Andrews, UK, 8–11 October 2013; pp. 311–320. [CrossRef]
- Bau, O.; Mackay, W.E. OctoPocus: A dynamic guide for learning gesture-based command sets. In Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, Monterey, CA, USA, 19–22 October 2008; pp. 37–46. [CrossRef]
- Toussaint, G.T. *The Geometry of Musical Rhythm: What Makes a “Good” Rhythm Good?*; CRC Press: Boca Raton, FL, USA, 2019.
- Magill, R.A.; Lee, T.D. *Motor Learning: Concepts and Applications: Laboratory Manual*; WCB McGraw-Hill: New York, NY, USA, 1998.
- Schmidt, R.A.; Lee, T.D.; Winstein, C.; Wulf, G.; Zelaznik, H.N. *Motor Control and Learning: A Behavioral Emphasis*; Human Kinetics: Champaign, IL, USA, 2018.
- Caracciolo, M. *The Experientiality of Narrative: An Enactivist Approach*; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2014; Volume 43.
- Crabtree, J. PolyNome. Available online: <https://polynome.net> (accessed on 15 June 2023).
- AeroBand. PocketDrum 2 Plus. Available online: <https://www.aeroband.net/collections/all-products/products/pocketdrum2-plus> (accessed on 15 July 2023).
- Theodots. HyperDrum. Available online: <https://www.theodots.com> (accessed on 17 July 2023).
- Ericsson, K.A.; Hoffman, R.R.; Kozbelt, A. *The Cambridge Handbook of Expertise and Expert Performance*; Cambridge University Press: Cambridge, UK, 2018.
- Chaffin, R.; Imreh, G.; Crawford, M. *Practicing Perfection: Memory and Piano Performance*; Psychology Press: London, UK, 2005.
- Carson, R.; Wanamaker, J.A. *International Drum Rudiments*; Alfred Music Publishing: New York, NY, USA, 1984.
- Frishkopf, M. West African Polyrhythm: Culture, theory, and graphical representation. In Proceedings of the ETLTC: 3rd ACM Chapter Conference on Educational Technology, Language and Technical Communication, Aizu-Wakamatsu, Japan, 27–30 January 2021. [CrossRef]
- Magadini, P. *Polyrhythms: The Musician’s Guide*; Hal Leonard Corporation: Milwaukee, WI, USA, 2001.
- Waterman, R.A.; Tax, S. African Influence on the Music of the Americas. In *Mother Wit from the Laughing Barrel: Readings in the Interpretation of Afro-American Folklore*; Wiley: Hoboken, NJ, USA, 1973; pp. 81–94.
- Arom, S. *African Polyphony and Polyrhythm: Musical Structure and Methodology*; Cambridge University Press: Cambridge, UK, 2004.
- Neely, A. 7:11 Polyrhythms. Available online: <https://youtube.com/watch?v=U9CgR2Y6XO4> (accessed on 15 June 2023).
- Huang, A. Polyrhythms vs. Polymeters. Available online: <https://youtube.com/watch?v=htbRx2jgF-E> (accessed on 12 June 2023).
- Harris, D.; Vine, S.; Wilson, M.; McGrath, J.S.; LeBel, M.; Buckingham, G. Action observation for sensorimotor learning in surgery. *J. Br. Surg.* **2018**, *105*, 1713–1720. [CrossRef] [PubMed]
- Buccino, G.; Arisi, D.; Gough, P.; Aprile, D.; Ferri, C.; Serotti, L.; Tiberti, A.; Fazzi, E. Improving upper limb motor functions through action observation treatment: A pilot study in children with cerebral palsy. *Dev. Med. Child Neurol.* **2012**, *54*, 822–828. [CrossRef] [PubMed]
- Ertelt, D.; Small, S.; Solodkin, A.; Dettmers, C.; McNamara, A.; Binkofski, F.; Buccino, G. Action observation has a positive impact on rehabilitation of motor deficits after stroke. *Neuroimage* **2007**, *36*, T164–T173. [CrossRef] [PubMed]
- Pinkl, J.; Cohen, M. Design of a VR Action Observation Tool for Rhythmic Coordination Training. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, Christchurch, New Zealand, 12–16 March 2022; pp. 762–763. [CrossRef]
- Yoshimura, M.; Kurumadani, H.; Hirata, J.; Osaka, H.; Senoo, K.; Date, S.; Ueda, A.; Ishii, Y.; Kinoshita, S.; Hanayama, K.; et al. Virtual reality-based action observation facilitates the acquisition of body-powered prosthetic control skills. *J. NeuroEng. Rehabil.* **2020**, *17*, 113. [CrossRef] [PubMed]
- Hoang, T.N.; Reinoso, M.; Vetere, F.; Tanin, E. Onebody: Remote posture guidance system using first person view in virtual environment. In Proceedings of the Nordic Conference on Human-Computer Interaction, Gothenburg, Sweden, 23–27 October 2016; pp. 1–10.
- Kodama, D.; Mizuho, T.; Hatada, Y.; Narumi, T.; Hirose, M. Effects of Collaborative Training Using Virtual Co-embodiment on Motor Skill Learning. *IEEE Trans. Vis. Comput. Graph.* **2023**, *29*, 2304–2314. [CrossRef] [PubMed]
- Gerry, L.J. Paint with me: Stimulating creativity and empathy while painting with a painter in virtual reality. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1418–1426. [CrossRef] [PubMed]
- Hapuarachchi, H.; Kitazaki, M. Knowing the intention behind limb movements of a partner increases embodiment towards the limb of joint avatar. *Sci. Rep.* **2022**, *12*, 11453. [CrossRef] [PubMed]

29. Wegner, D.M.; Sparrow, B.; Winerman, L. Vicarious agency: Experiencing control over the movements of others. *J. Personal. Soc. Psychol.* **2004**, *86*, 838. [CrossRef] [PubMed]
30. Hagiwara, T.; Katagiri, T.; Yukawa, H.; Ogura, I.; Tanada, R.; Nishimura, T.; Tanaka, Y.; Minamizawa, K. Collaborative avatar platform for collective human expertise. In Proceedings of the SIGGRAPH Asia Emerging Technologies, Tokyo, Japan, 14–17 December 2021; pp. 1–2. [CrossRef]
31. Fribourg, R.; Ogawa, N.; Hoyet, L.; Argelaguet, F.; Narumi, T.; Hirose, M.; Lécuyer, A. Virtual co-embodiment: Evaluation of the sense of agency while sharing the control of a virtual body among two individuals. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 4023–4038. [CrossRef] [PubMed]
32. Luria, M.; Reig, S.; Tan, X.Z.; Steinfeld, A.; Forlizzi, J.; Zimmerman, J. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In Proceedings of the Designing Interactive Systems Conference, San Diego, CA, USA, 23–28 June 2019; pp. 633–644.
33. Parappa The Rappa Remastered. Available online: <https://rodneyfun.com/parappa> (accessed on 13 June 2023).
34. Activision. Guitar Hero Live. Available online: <https://support.activision.com/guitar-hero-live> (accessed on 16 July 2023).
35. Beat Saber. Available online: <https://beatsaber.com> (accessed on 17 June 2023).
36. Gaydos, M. Rhythm Games and Learning. In Proceedings of the International Society of the Learning Sciences (ISLS), Chicago, IL, USA, 1–2 July 2010.
37. Rocksmith Plus. Available online: <https://ubisoft.com/en-gb/game/rocksmith/plus> (accessed on 11 June 2023).
38. Pai, Y.S.; Hajika, R.; Gupta, K.; Sasikumar, P.; Billinghamurst, M. NeuralDrum: Perceiving brain synchronicity in XR drumming. In Proceedings of the SIGGRAPH Asia Technical Communications, Virtual Event, Republic of Korea, 1–9 December 2020; pp. 1–4. [CrossRef]
39. Toppi, J.; Borghini, G.; Petti, M.; He, E.J.; De Giusti, V.; He, B.; Astolfi, L.; Babiloni, F. Investigating cooperative behavior in ecological settings: An EEG hyperscanning study. *PLoS ONE* **2016**, *11*, e0154236. [CrossRef] [PubMed]
40. PD-7 Dual Trig. Drum Pad-7". Available online: <https://roland.com/us/products/pd-7> (accessed on 29 May 2023).
41. Maestro—Midi Player Tool Kit (MPTK). Available online: <https://paxstellar.fr> (accessed on 17 June 2023).
42. OpenXR Plugin. Available online: <https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.7/manual/index.html> (accessed on 21 May 2023).
43. Weiss, L.V. Marching Percussion in the 20th Century. In *A History of Drum and Bugle Corps*; Sights and Sounds, Inc.: Madison, WI, USA, 2002; pp. 91–96.
44. Ellison, B.; Bailey, T.B.W. *Sonic Phantoms: Composition with Auditory Phantasmatic Presence*; Bloomsbury Publishing: New York, NY, USA, 2020.
45. UEQ. User Experience Questionnaire. Available online: <https://www.ueq-online.org> (accessed on 15 July 2023).
46. Rex Rainbow's Github. Available online: <https://rexrainbow.github.io/C2Demo/MIDI%20to%20CSV/> (accessed on 17 June 2023).
47. Donald, L.S. *The Organization of Rehearsal Tempos and Efficiency of Motor Skill Acquisition in Piano Performance*; University of Texas at Austin: Austin, TX, USA, 1997.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.