# Teach Me Drums: Learning Rhythms through the Embodiment of a Drumming Teacher in Virtual Reality

**Mie Moth-Poulsen**
Aalborg University Copenhagen
miemp123@gmail.com

**Tomasz Bednarz, Volker Kuchelmeister**
UNSW Art and Design, Sydney
t.bednarz,Kuchel@unsw.edu.au

**Stefania Serafin**
Aalborg University Copenhagen
sts@create.aau.dk

## ABSTRACT

This paper investigates how to design an embodied learning experience of a drumming teacher playing hand drums, to aid higher rhythm understanding and accuracy. By providing novices the first-person perspective of a drumming teacher while learning to play a West-African djembe drum, participants' learning was measured objectively by their ability to follow the drumming teachers rhythms.

Participants subjective learning was assessed through a self assessment questionnaire measuring aspects of flow, user-experience, oneness, and presence. Two test iterations were conducted. In both there was found no significance difference in participants' ability to follow the drumming teacher' s tempo for the experimental group exposed to the first-person perspective of the teacher in a Virtual Reality (VR) drum lesson, versus the control group exposed to a 2D version of the stereoscopic drum lesson. There was found a significant difference in the experimental group' s presence scores in the first test iteration, and a significant difference in experimental group' s oneness scores in the second test iteration. Participants' subjective feelings indicated enjoyment and motivation to the presented learning technique in both groups.

Figure 1. First drumming recordings.

Figure 2. Second drumming recordings.

## 1. INTRODUCTION

Several studies have shown the potential of Virtual Reality (VR) as an alternative training and learning platform for acquiring new skill sets and improving existing ones. Amongst others seen in the field of music learning, training rhythmical skills and musical expression [1], along training physical movements [2]. The multidimensional nature of VR provides a unique possibility to take the perspective of another person than one self. This quality provides a strong tool to facilitate learning and communication between individuals, which has been examined by expert-novice mentorship simulations in the art of painting [3]. One advantage of using VR for education is the ability to present abstract topics in a tangible way. For example, teaching mathematics through collaborative environments for learning geometrical concepts, against traditional pen and paper. Moreover, VR supports doing, instead of observing, as the user participates in the virtual world instead of using it, compared to other types of human-computer interfaces [4]. The focus on incorporating the body in designing movement-based interfaces has been fueled by the advances in sensor technology [5]. The affordance of VR technology leverages interactivity, not seen in medias such as video and text, based on the vast possibilities for behavioral tracking. This essentially allows users to participate in a VR instead of using it, through embodied interactions. This project seeks to investigate how to teach music more effectively, by incorporating VR to communicate somatic knowledge of a drumming teacher, providing a first-person perspective of the playing teacher.

Figure 3. A screenshot from the experiment. Top: experimental condition: wearing the VR headset. Bottom: control condition where the visual feedback is given by the screen. In both cases auditory feedback is given by headphones, while tactile feedback is given by the real drum.

## 2. DESIGN AND IMPLEMENTATION

This section presents the design of a VR setup for teaching novices drums through an embodied first-person perspective of a drumming teacher. The underlying motivation is to aid higher rhythm understanding and accuracy through the participants embodiment of the playing teacher. The design consists of two parts. First, the development of the teaching material constituting the test stimuli of a prerecorded drum lesson, and a discussion of the suitable hand drum for the rhythm teaching of novices. Secondly, the recording setup with a stereoscopic 3D camera is discussed, for capturing a reliable first-person perspective of the drumming teacher, to be viewed through a Head Mounted Display (HMD) while playing along the teacher.

### 2.1 Design of Teaching Material

The design of the teaching material for the drumming recording was revised and discussed in two iterations, with two professional drummers respectively. A prerequisite to the drummers' qualifications was teaching experience and skills on hand drums of a West-African Djembe and a bongo drum. Hand drums were chosen based on perceived affordances and signifiers on how to be operated by the hands. Secondly, allowing immediately tactile feedback, to provide a direct sense of the drum when viewing the stereoscopic footage of the hand drum through a HMD.

Before each of the two individual drumming recordings, the drummers were explained motivation for the research of the study. Additionally, the test objective of comparing two learning medias (VR and video), and their effectiveness on rhythm learning. The requirements for the teaching material included content directed to novices who had no to little experience with drum lessons before. This was to ensure an equal level of drumming experience, and skill level between the test participants. Furthermore, the drum recording was structured with drumming sequences that left enough time for the novice to play along. This is based on the objective measure, to test the effect of the two learning medias, which involves a comparison of the drumming rhythms produced by the mentor and the novice. The drummers' knowledge was incorporated into structuring the suitable rhythms for teaching hand drums to novices.

The first drummer had over 40 years experience of music practicing. The teachers experience included teaching, performing drumming shows with a west African drumming group, along skills in other instruments. A bongo drum (size 6 and 7 inch) was used for the first recording (see Figure 1), due to affordability. An initial instruction and trial phase was dedicated to familiarization with the bongo before the rhythm training. It was incorporated in the first part of the drum lesson, allowing the participants to get familiar with the bongo. This included, how to place the bongo between the knees, the hit method, and how to produce a pulse on the bongo. The general structure of the drumming lesson was composed of 4 sequences of rhythm patterns. The teaching material for both drumming recording followed the general structure of a trial phase and four rhythm patterns. The trial phase included how to hit the djembe. The teacher instructed how to play two different djembe tones in the trial phase, to match the skill level for a novice. The base tone (centre drum skin), and the tone (edge of drum skin). Halfway in each play along sequence, the teacher increased the tempo, to challenge the novice.

### 2.2 Recording setup

The recording was filmed with the stereoscopic 3D LucidCam. For the second drumming recording, the static rig was improved from the first recording. The rig was altered to be positioned from the side, without the two fish-eye lenses capturing the extended rig arm, allowing more flexibility for adjusting to the drummers height (see Figure 2).

### 2.3 Implementation

The game engine Unity3D 2017 was used as the software to implement the 3D stereoscopic viewing for the HTC Vive HMD. The digital hands from a Leap Motion device were incorporated to support the position of novices own hands within the physical drumskin, and the matching of their own hands to the playing teachers.

## 3. FIRST ITERATION

The first test iteration included a between-group design, comparing two viewing conditions of VR and video. The test stimuli for both test groups were the same pre-recorded video of a drum lesson instructed by a drumming teacher, from the second drum recording. The test stimuli consisted of 5 phases. An initial trial phases taught the participants how to hit the djembe drum, to get comfortable with it. Next phases four rhythm patterns were taught. In each rhythm pattern, the teacher demonstrated the rhythm three times on the djembe drum, before playing along. Each play along rhythm sequence was on average 49.5 seconds long. The total length of the test stimuli was 5 minutes and 53 seconds. The first 5 seconds of the recordings was a black screen. 11 seconds was left in the end, after the teacher finished the final and fourth rhythm. In the experimental group, participants were taught to play drums by taking the first-person perspective of the drumming teacher in VR, viewed through a HMD, projecting the stereo-scopic recording of the teacher. Participants were presented with a physical djembe drum, matched to the position where it was located in the virtual world (test stimuli). In the control group, participants viewed the same test stimuli on a 2D monitor placed in front of them. The audio of the test stimuli was recorded with the built in stereo microphones in the LucidCam, recording audio at 48Hz, uncompressed 16-bit audio, from the teachers visual perspective. The viewed test stimuli was recorded in the same room as the participants were seated, with the participants sitting on the same position as the teacher on a stationary chair.

### 3.1 Participants and Recruitment

35 participants were recruited at the University of New South Wales Art Design (Sydney). The data of five participants was not usable and discarded, producing a final sample of 30 (male=12, female=18). The groups ages ranged from 20 to 40; the majority accounted 25-34 years (50%). The majority were students (86.6%). Before conducting the test, participants were handed a consent form along with a participant information sheet for the test. Participants' data was assured confidentiality, along with their right to withdraw from the test at any time. Participants were recruited based on the criterias of being novices to drumming, not having any hearing disabilities, and fully functional limbs.

### 3.2 Setup

A djembe hand drum of height 60cm and diameter 30cm was used. Participants were seated on a stationary stool of fixed height 43.5 cm, and a diameter of 33 cm. The drumming lesson took place in a section of a closed room, partly covered by a black curtain. Each participant's sound from the drumming, was recorded with a lavalier microphone clipped to their clothes, centre at chest. On-ear headphones were used to play the sound of the viewed test-stimuli. The HTC Vive HMD was used to display the test stimuli for the experimental group. The control group sat at a distance of one meter from the 2D monitor.

### 3.3 Procedure

Participants were randomly assigned to each of the two test groups, and all naive to the purpose of the test. Participants were asked demographic questions before the drumming lesson in each of the two test groups (experimental and control), in an online self-assessment questionnaire. Both groups were asked about their experiences with music practising. The experimental group was further asked about their previous experience with VR, by a rating scale from 1 (never tried it before) to 7 (uses it daily). Participants were recorded to compare the rhythm accuracy between the drumming teachers recording and the participant. At the beginning of the viewed test footage, an audible synchronisation clap from the test footage, was outputted through a pair of speakers, before switching the sound output to the worn headphones by the participant. This was done to generate the same synchronization point in the participants audio files to the teacher, for the rhythm comparison. After the pre-recorded drum lesson, participants completed post questions to their subjective learning experience, filled out in the online self-assessment questionnaire.

### 3.4 Rhythm accuracy

The variable of rhythm accuracy was chosen to objectively quantify the novices ability to follow the teachers djembe rhythms. Each participant produced a unique audio file, capturing their rhythmic performance during the experiment. The learning efficiency of both test groups was quantified by comparing participants individual audio files analysed in beats by minute, with the drumming teachers.

### 3.5 Self-assessment questionnaire

Participants answered a self-assessment questionnaire post to their participation in the pre-recorded drumming lesson. The questionnaire was designed to measure 4 aspects of the participants subjective experience of the given learning media, including: Flow, User-experience, Oneness, and Presence. Flow was measured using the Flow Short Scale [6].

The Inclusion of Other in the Self Scale was originally developed in [7]. The scale was used to measure the extent to which the participants felt bodily in sync with the mentors location and rhythmic movements, related to the concept of body-syntonicity. The measure of presence targeted presence seen through the Plausibility Illusion (Psi). Psi relates to the fact that the scenario presented is felt as actually happening [4].

In the experimental group, the audio recordings of five participants was discarded. Participants noticed that the audio and video played through the Unity application was out of sync. The synchronization of the audio and video played through Unity was therefore monitored during each test, by the participants wearing the HMD at their drumming position before starting and playing of the unity application. Participants soundfiles were synchronized with the teachers in post-processing. The drumming lesson consisted of four rhythm patterns, both including a normal and fast tempo. Eight soundfiles were generated for each participant, containing the normal and fast tempo of each rhythm. It was prioritized to avoid the teachers voice in the generated soundfiles, not to cause a further difference in the teachers and participants audio recordings. Eight sequence markers of label tracks were created related to the teachers starting points. The sequence markers contained both the normal and fast tempo of the rhythm play along periods. The sequence markers ensured the generation of the eight individual wav files for each participant, to match precisely the teachers soundfiles files respectively. The MIRtoolbox 1.7 for Matlab, was used to analyse participants rhythm performances [8].

An independent t-test was performed with participants final tempo score in the two test groups. The reported results showed no significance difference for the experimental group exposed to the VR drum lesson (M=35.311, SE=2.956), than for the control group exposed to the 2D drum lesson (M= 30.821, SE=2.1043), t(28)= 1.236, p= 0.226

Though there was found no significant difference among the two test groups, the control group performed better with 4.49 BMP less in difference from the drumming teacher, than the experimental group. Inspecting the data further, the total average of the fast tempo difference scores of the two test groups, produced almost equal results. The control group produced a fast tempo difference score with a mean value of 20.29 BPM. While the experimental group produced a mean value of 20.49 BPM. The total average of the normal tempo difference scores produced 41.36 BPM in the control group, and 50.13 BMP in the experimental group, indicating the control group was 8.77 BMP closer to following the teachers tempo, than the experimental group, in these sequence parts. An independent t-test was also performed among the two groups total average of the normal tempo scores. The result also indicated no significance among the two groups ability to follow the teachers tempo in normal pace. An independent t-test was performed with participants average flow score from the FSS in the two test groups. The result of the FFS flow scores was found not significantly different for the experimental group exposed to the VR drum lesson (M=5.580, SE=0.152) than for the control group exposed to the 2D drum lesson (M= 5.1067, SE=0.254), t(28)= 1.601, p= 0.121.

Participants' ratings in the first seven user-experience items produced a total mean value for the experimental group at 5.73, and a mean value at 5.70 for the control.

There was found no significance difference between the experimental and control group, in the reported oneness ratings, with the exact same mean of 4.73, and median of 5 in both groups. The Oneness scores indicated that the first-person perspective rendered in the HMD experienced by the experimental group did not provide a stronger sense of feeling in tune and synchronizing with the teachers movements in this experimental setup.

An independent t-test was performed with participants average presence score in the two test groups. The result of the presence scores was found significantly different for the experimental group exposed to the VR drum lesson (M=5.617, SE=0.251) than for the control group exposed to the 2D drum lesson (M= 4.7333, SE=0.263), t(28)= 2.4282, p= 0.0219,  = 0.05. Table 1 summarizes the results.

| Measurement | Mean | Std. dev. | Std. err. |
|---|---|---|---|
| Flow Exp. | 5.580 | 0.588 | 0.152 |
| Flow Con. | 5.1064 | 0.982 | 0.254 |
| User-experience Exp. | 0.833 | | 0.215 |
| User-experience Con. | 0.827 | | 0.213 |
| Oneness Exp. | 4.733 | 1.792 | 0.463 |
| Oneness Con. | 4.733 | 1.624 | 0.419 |
| Presence Exp. | 5.617 | 0.972 | 0.251 |
| Presence Con. | 4.733 | 1.019 | 0.263 |

Table 1. Descriptive statistics of the Subjective Self-assessment Questionnaire for each measurement for the control (con) and experimental (exp) group for the first test iteration.

## 4. SECOND ITERATION

The experimental design followed the same as the first test iteration. However, the test's stimuli was revised with a third drumming recording, to produce a final pre-recorded drum lesson. The hired drumming teacher was the same used for the test stimuli in the first test iteration. The structure of the test stimuli followed the same as the first test iteration, with the same four rhythms. The total length of the recorded test stimuli was 7 minutes and 9 seconds long. The test stimuli consisted of a longer trial phase with more deliberate instructions. This was to ensure that the participants got a sense of how to hit the djembe properly before the first rhythm instruction. The four rhythm patterns taught was on average 82 seconds long. The audio recording of the mentor in the first experiment was recorded with the built-in stereo microphones in the LucidCam. In the first experiment, the participants could see what the teacher saw, but not hear a reliable version of what the teacher heard. This experiment revised the audio capture to be recorded from a binaural point-of-view with the Roland CS-10EM binaural microphones. The microphones are electret and omnidirectional, capturing a frequency range of 20 Hz to 20,000 Hz.

### 4.1 Participants and recruitment

41 participants were recruited at the University of New South Wales, Art  Design (Sydney). It was ensured that none of the subjects had participated in the first test iteration. One participants data was not usable and discarded, giving a final sample of 40 (male=14, female=26). The groups ages ranged from less than 20 to 55; the majority accounted less than 20 years (35%) and $21 - 25$ years (30%). The majority were students (80%). Participants were recruited by the same conditions as in the first test iteration.

### 4.2 Setup and procedure

The setup and procedure followed the general structure of the first test iteration. The sound of the participants was captured with a Zoom H4n Pro, placed in front of the drum. The changes to the procedure included the test conductor demonstrating how to hold and position the djembe cor-

| Measurement | Mean | Std. dev. | Std. err. |
|---|---|---|---|
| Flow Exp. | 5.333 | 1.159 | 0.259 |
| Flow Con. | 5.105 | 0.881 | 0.197 |
| User-experience Exp. | 5.714 | 1.061 | 0.237 |
| User-experience Con. | 5.728 | 0.828 | 0.185 |
| Oneness Exp. | 5.250 | 1.564 | 0.336 |
| Oneness Con. | 4.4 | 1.353 | 0.303 |
| Presence Exp. | 5.575 | 1.162 | 0.259 |
| Presence Con. | 4.975 | 1.186 | 0.265 |

Table 2. Descriptive statistics of the Subjective Self-assessment Questionnaire for each measurement for the control (con) and experimental (exp) group for the second test iteration.

rectly, with the right angle tilting the djembe from the floor away from the participant. Furthermore, the test conductor explained in the brief about the drum lesson content, that the participant would first be instructed by the teacher in the drum lesson on how to hit the djembe. Next, following a count in on four beats, to hit along the teacher. The evaluation of the second test iteration relied on the same measurements methods used in the first test iteration.

### 4.3 Results

Similarly to the first test iteration, there was found no significant difference between the experimental group exposed to the VR drum lesson ($M = 24.567, SE = 2.282$), and the control group exposed to the 2D drum lesson (M= 21.739, SE=1.932), t(38)= 0.946, p= 0.350.

### 4.4 Subjective learning

There was found no significance in the subjective ratings of flow, user-experience, and presence. The Mann-Whitney test was used as a significance test of the oneness ratings, due to non-parametric data. The feeling of oneness with the teacher showed significance difference between the experimental group exposed to the VR drum lesson (M=5.250, SE=0.336), than for the control group exposed to the 2D drum lesson (M= 4.400, SE=0.303), t(38)= 1.558, p= 0.048. Table 2 summarizes the results.

## 5. CONCLUSIONS

This paper investigated if learning hands drums through an embodied first-person perspective mediated in VR leads to better rhythmic understanding than learning through a 2D video. The results of the rhythm comparison in the two test iterations found no significant difference between the experimental and control group learning of rhythms, evaluated in the ability to follow the teachers tempo in BPM. The majority of the participants described their experience as enjoyable in both test iterations. Additionally, indicating motivation towards the given learning technique in both test groups. The results can situate the question whether the given musical instrument and task was a motivation, along the given teaching material. The two test conditions were designed to detect the effect of a first-person perspective of a drumming teacher, on a novices rhythm accuracy and learning. Thus, the two conditions differed in visual display, the control group had the ability to watch their own hands playing on the physical djembe. This could produce an advantage in terms of acquiring a better sense of the edge when the hitting drum skin. However, a restrains in this scenario was the shifting of attention between the participants hands and the playing teacher viewed on the 2D monitor in front of them. In a first iteration, we tried to project the hands of the player of top of the VR experience, using the Leap Motion's tracking. However, the inconsistent tracking of the participants hands caused the attention to be directed to the quality of the 3D rendered hands more than the experience. As the viewed drum lesson was a pre-recorded video, the possibility of corrections to participants rhythm performance real-time from a teacher was not available. Participants were not given any feedback upon how well they performed the rhythm in the first and second test iteration - from an objective source. To optimize the participants learning, future studies could explore real-time feedback of participants rhythm performances.

## 6. REFERENCES


[1] S. Serafin, A. Adjorlu, N. Nilsson, L. Thomsen, and R. Nordahl, "Considerations on the use of virtual and augmented reality technologies in music education," in *K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR), 2017 IEEE Virtual Reality Workshop on*. IEEE, 2017, pp. 1–4.

[2] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S.-H. Jung, and G. Kurillo, "The effect of interactivity on learning physical actions in virtual reality," *Media Psychology*, vol. 11, no. 3, pp. 354–376, 2008.

[3] L. J. Gerry, "Paint with me: Stimulating creativity and empathy while painting with a painter in virtual reality," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 4, pp. 1418–1426, 2017.

[4] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality," *Frontiers in Robotics and AI*, vol. 3, p. 74, 2016.

[5] K. Höök, *Designing with the Body: Somaesthetic Interaction Design*. MIT Press, 2018.

[6] S. Engeser and F. Rheinberg, "Flow, performance and moderators of challenge-skill balance," *Motivation and Emotion*, vol. 32, no. 3, pp. 158–172, 2008.

[7] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness." *Journal of personality and social psychology*, vol. 63, no. 4, p. 596, 1992.

[8] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International conference on digital audio effects*. Bordeaux, 2007, pp. 237–244.



ISBN 978-84-09-08518-7 ISSN 2518-3672