Co-design of a Smart Cajón*

Luca Turchet, Andrew McPherson, Mathieu Barthet,
(luca.turchet@qmul.ac.uk) (a.mcpherson@qmul.ac.uk) (m.barthet@qmul.ac.uk)

Centre for Digital Music, Queen Mary University of London, United Kingdom

Smart Instruments are a novel family of musical instruments that embed sensors, actuators, wireless connectivity, and semantic audio technologies. This paper reports the findings of a participatory design approach to develop a Smart Cajón, a box-shaped percussion instrument with Internet of Musical Things components. Five initial co-design sessions were conducted with different professional cajón player participants. The players were invited to devise tangible mock-ups by placing provided sensors on an acoustic cajón, and to express desirable use cases and interactions. We then designed and implemented a prototype satisfying performers' common requirements. The prototype was assessed using the concurrent think-aloud protocol. On overall, the smart qualities of the prototype and their potential received positive feedback, and areas of improvements related to expressive control and personalization were highlighted.

0 INTRODUCTION

The term "augmented instruments" is used to refer to a class of musical instruments enhancing conventional instruments with sensor and/or actuator technology. Various examples of augmented percussive instruments can be found in both academia (see e.g., [1, 2]) and industrial applications. A percussive instrument that finds uses in a large variety of musical genres and has recently become widely diffused is the cajón [3], a wooden box including a supplemental rattle device. Recently, electronics augmentation has made inroads in the conventional design of acoustic cajones. Examples from the music industry include Roland's Electronic Cajón[1] and De Gregorio's Cajón Centaur[2]. These instruments present sensors that can detect players' hits and map them to audio samples from different percussive instruments thanks to an embedded sound engine. Nevertheless, they are not equipped with a networked system that allows players to update the default sound samples or to change the timbre of the instrument (e.g., using equalization processing). Moreover, the involved technology cannot extract and analyze information related to the musician's playing. It also does not track expressive gestures of musicians (such as fingers' pressure or sliding, instrument tilting) and, therefore, cannot use these for creative control (e.g., synthesizers, sound effects).

Although various augmented cajones exist, the "smartification" of the cajón has not been addressed yet, to the best of our knowledge. By "smartifying" a musical instrument we refer to the process of adding augmented and connected capabilities to a conventional instrument, as described in [4]. Smart Instruments are a novel class of musical instruments that extend the concept of augmented instruments. In addition to sensor and actuator enhancements, typical of augmented instruments, smart instruments integrate embedded intelligence using audio and sensors digital processing, as well as wireless connectivity to join local or remote computer networks. The embedded hardware and software enables the execution of semantic audio applications (e.g., [5]) directly from within the instrument, which is one of the novel aspects to the design proposed in this paper.

The field of Smart Instruments is in its infancy and only recently examples of design products start to emerge (see e.g., MIND Music Labs' Sensus Smart Guitar with use case studies [6]). Smart Instruments are instances of so-called "Musical Things" defined within the emerging field of Internet of Musical Things (IoMusT) [7]. IoMusT refers to an ecosystem of interoperable devices dedicated to the production and/or reception of music, which can lead to novel forms of interactions between performers and audiences.

This paper describes initial and assessment phases of a co-design process for the Smart Cajón, which involved five professional cajón players. We investigate to what extent cajón players' musical interactions can be enhanced with smart technologies and discuss participants' feedback that informs the next steps in the iterative design process. We report results of evaluation tests on a prototype of Smart Cajón, conducted by the participants involved in the co-design sessions. In such evaluations, we focus on how par-

*Corresponding author: L. Turchet (luca.turchet@qmul.ac.uk)
[1] www.roland.com/us/products/el_cajon_ec-10
[2] www.cajondg.com/product/cajon-centaur

ticipants respond creatively to the affordances and constraints offered by the instrument [8, 9, 10, 11].

1 PARTICIPATORY DESIGN

Participatory design processes have proved successful in media and arts technology applications (see e.g., [12]). In this work we follow a participatory design methodology using cooperative prototyping where users are involved actively and creatively in the design and evaluations of early prototypes [13]. This is motivated by the aim to better understand the needs of contemporary cajón players and let them shape how smart technologies may benefit their practice or inspire new creative opportunities.

Guided by participatory practices for the iterative design of software [14], the various phases of the design can be described as follows: 1) problem identification and clarification, 2) requirements, 3) analysis, 4) high-level design, 5) implementation, 6) assessment, and 7) redesign. The phases 1) to 6) are discussed in the next sections. Results of co-design sessions were analyzed and an implementation encompassing recurrent features in the participants' requirements was produced. The prototype was evaluated by four of the users involved in the co-design process.

Our participatory design process does not begin with the gathering of cajón players' general musical requirements and needs, but rather with their reaction to a specific/limited set of sensor technologies and to an example prototype. Although not all musical possibilities are realizable by the particular sensors made available to participants, this methodological choice is motivated by the fact that there is good evidence in the literature on digital musical instruments research to suggest that performers might not be able to imagine hypothetical future capabilities, but that they do respond creatively to specific affordances and constraints that they are given [8, 9, 10, 11].

2 PROBLEM IDENTIFICATION AND CLARIFICATION

2.1 Participants

We addressed problem identification and clarification through individual co-design sessions with five professional cajón players (1 female, 4 males) aged between 28 and 38 (mean = 32.4, SD = 4.03). They had on average 19.5 years of musical experience and were all highly-skilled cajón players. In addition they had substantial knowledge of electronic music instruments, digital audio workstations, as well as analog and digital audio effects.

2.2 Setting and procedure

Three sessions were held in the Media & Arts Technology Performance Lab at Queen Mary University of London, a purpose-built sound proofed room, while the two other sessions were conducted in musicians' respective homes. Each session comprised four stages:

1. An initial explanatory introduction about the Smart Instruments and IoMusT concepts;
2. Technology demonstrations of a preliminary prototype of Smart Cajón;
3. Semi-structured interviews, where participants were asked to envision applications of cajón performance involving smart technologies;
4. Mock-up activities, where each participant produced a tangible mock-up design of a Smart Cajón to address some of the visions identified during the interviews.

Sessions were documented with video recordings. Each session lasted about 1 hour and 30 minutes on average.

2.2.1 Explanatory introduction

The co-design sessions began with a presentation given by the experimenter, which included an introduction of the concepts of IoMusT [7] and of the Smart Instruments [4, 6]. Specifically, the goal was to provide effective understanding of the overall IoMusT framework including existing and envisioned technologies with a particular focus on future musical instruments. Illustrations of IoMusT included short demonstration videos of the Sensus Smart Guitar.

2.2.2 Initial design and demonstrations

The second stage of the sessions involved hands-on demonstrations of a Smart Cajón preliminary prototype developed by the authors. The design builds upon a conventional acoustic cajón and includes the following hardware: two contact microphones using piezoelectric material attached to the interior side of the front panel (Big Twin by K&K); a Bela board for low-latency audio processing [15], based on a Beaglebone Black board; a small wireless router (TL-WR902AC by TP-Link), which features the IEEE 802.11ac Wi-Fi standard as well as a USB port for 4G dongles enabling Internet connectivity; a loudspeaker (Monitor Supreme Center 250 by Magnat) with small preamplifier (SA-36A Pro HIFI Digital Amplifier by SMSL); four vibration motors (307-103 by Precision Microdrives) embedded in a rectangular foam and placed by groups of two on each side; a force sensitive resistor sensor (FSR 406 by Interlink Electronics) to react to finger pressure, placed on the top side of the instrument. Power was supplied externally using AC power plugs.

Software for audio and sensor processing, and tactile stimuli generation have been developed using the Pure Data (PD) visual programming language. In order to demonstrate example of sonic interaction capabilities, a PD patch was implemented to apply delay and reverberation audio effects to sounds captured with the piezo sensor. The pressure sensor was used to demonstrate synthesis functions by mapping its value to the sound level of a synthetic chord. Examples of tactile stimuli were generated using Pulse Width Modulation to produce four types of dynamics and patterns of activations on the four motors: 1) a continuous vibration of strong intensity lasting 2 seconds; 2) rapid intermittent pulses of constant and medium intensity within the span of 3 seconds; 3) intermittent pulses of increasing intensity and duration within the span of 5 seconds; 4) intermittent pulses of decreasing intensity and duration within the span of 5 seconds. Data reception and forwarding over Wi-Fi were achieved using Open Sound Control (OSC) over the User Datagram Protocol. Following the recommendations reported in [16] to optimize the components of a Wi-Fi system for live performance scenarios, in order to reduce latency and increase throughput, the router was configured in access point mode, security was disabled, and support was limited to IEEE 802.11ac standard only.

To demonstrate possible smart qualities of the instrument (e.g., continuous control of sound effects and sound generation, tactile stimuli, connectivity options), we developed a graphical user interface (GUI) enabling control and visual feedback. The GUI, developed in PD, was presented to participants on a laptop connected to the instrument. It included widgets displaying the numerical values of the pressure sensor data wirelessly transmitted. Other widgets could be used to generate tactile stimuli felt on the device. During each session, cajón player participants were invited to explore for about 5 minutes the control and generation of synthetic sounds and tactile stimuli using the GUI.

2.2.3 Semi-structured interviews

Semi-structured interviews were conducted to gather musicians' initial impressions of the Smart Cajón demo prototype presented to them. We also asked them to describe foreseen applications given their music practice. The following topics were covered:

- *Envisioning interactions between a Smart Cajón player and intelligent audio/sensor processing*: how can sound interaction capabilities embedded in the instrument respond to or complement playing?;
- *Envisioning interactions between a Smart Cajón player and audience members*: what type of content may be conveyed and/or received for audience interactions including participatory performance situations?;
- *Envisioning interactions between a Smart Cajón player and other performers of a musical ensemble*: how can tactile signals be used for co-performer communication while playing?

2.2.4 Mock-up activities

During the mock-up activities participants were introduced to and provided with various sensors. These comprised pressure sensors (by Interlink Electronics) of various shapes (square x 5, rounded x 6, small rounded x 6), ribbon sensors (Softpot 100mm by Spectra Symbol x 5), and proximity sensors (by Sharp Microelectronics x 4). The sensor kit also included push buttons with various LED colors (blue and red x 8 each), an inertial measurement unit (IMU), capable of tracking the instrument movements in tri-dimensional space (BNO055 by Bosch). Ten actuators (same model as for the demo prototype), and a smartphone for visual display and control, were also provided.

Participants were asked to achieve a non-working prototype implementing the use cases envisioned during the semi-structured interviews, by optimizing the capabilities of the available sensors, actuators, and visual display. This activity did not involve any hardware or software implementation, but consisted in placing sensors and actuators on a conventional cajón using paper scotch tape. Participants were invited to imagine various layouts and then to select one they judged the best.

3 PARTICIPATORY PROTOTYPING RESULTS

3.1 Interview analysis

The semi-structured interviews were transcribed from video recordings and analyzed to identify common themes.

Embedded intelligence. Concerning the possibility of exploiting embedded intelligence for sonic interaction, three participants suggested that the instrument could trigger different percussive sounds (e.g., snare drum, conga, etc.) based on the position of a hit. Two participants expressed desire to be able to produce classic acoustic cajón sounds but with different timbres (e.g., processed with different types of equalizations), with presets letting them easily switch between sounds (in the same way tone color on electric guitars can be changed using microphone selectors). Three participants were interested in being able to program the sounds or sound effects associated to sensors, via a smartphone, tablet, or personal computer. Moreover, two participants wished to exploit Internet connectivity to upload on the instrument some sound samples from online resources (e.g., other percussive instrument sounds). Two participants also expressed interest in having an instrument capable of playing backing tracks, possibly streamed from the cloud, for practice purposes.

Sensor augmentation. Regarding sensor augmentation, discrete controls such as buttons were envisioned to change different configurations of the instrument (presets). Interaction with sensors for continuous control was deemed interesting by all participants to create modulations of the acoustic sounds via effects, but not to generate additional synthetic or sampled sounds. All participants expressed interest in modulating sounds using back and forth movements in front of a proximity sensor (positioned to the front or side of the instrument) since such gestures were felt to be rather natural while playing.

Audience interaction. Overall, participants were opened to explore novel forms of performances involving audience participation. However, they also clearly pointed out that the nature and structure of such performances should be carefully planned in advance not to be detrimental or annoying for the player. All participants showed a clear preference for audience interventions complementing players' hits (e.g., to produce an accompanying melody or long chord pads), while interventions with shared control over sound production were seen as something that could be funny to try but not for long periods of time as this could be potentially detrimental to the performance. Three participants reported that a shared control of the Smart Cajón sounds could be effected by letting audience members modify "volume", "reverberation", or "equalization". Two participants suggested that audience members could select instrument presets (e.g., congas or bongo), but the sounds should only be triggered by the performer. In such situation, it was considered essential for the instrument to embed a screen to display configurations selected by audience members and be able to adapt to such selections. Three participants also mentioned that real time feedback from audience members could be obtained through the screen e.g., to learn more about their satisfaction about a piece or the performance, whether slower or faster pieces are desired, and if they want more pieces to be played (mediate "encore" requests). However, two participants preferred not to have such feedback information from audience members over the course of a performance. Three participants envisioned to exploit wireless connectivity to control visual projections (including abstract representations, videos, text), while two other participants envisioned control of stage lighting, and another control of smoke machines.

Haptic feedback. The possibility of using touch as a medium of communication between the cajón players and the audience, or between players in an ensemble, was welcomed by all participants. Nevertheless, a recurring comment was the need for stronger tactile sensations than those experienced using the demo prototype. A potential issue highlighted by participants was that other instruments on stage could produce vibrations masking that of the tactile stimuli (e.g., a bass playing at loud level). All the participants mentioned that actuators could be placed in the upper part of the the cajón, but expressed their preference for haptic wearables placed on the body. Two participants reported that tactile stimuli could be used in conjunction with a visual display to notify them about specific information communicated via the screen.

3.2 Mock-up results

Figure 1 presents the different designs produced by the five participants during the mock-up activity. These were analyzed to identify common choices related to sensor types and positions. In Table 1 we report the results of this analysis and what was retained in our design. In the table and in the reminder of the paper the participants are indicated as P1, P2, P3, P4, and P5.

3.3 Implementation

An implementation of a Smart Cajón was produced to support smart functionalities highlighted by participants and further assess the design (see Figure 2; a video is available at www.iomust.eu.). The design aims to satisfy common requirements from cajón players as reported in Table 1. The current iteration focuses on top panel and interior side, while side panels are reserved for future work.

Tactile feedback was addressed by embedding eight motors in a cushion for the top panel. More motors were used compared to the demo prototype to increase sensitivity and the motors were placed along the left and right sides.

A PD patch was developed to handle sensor, motor, and audio processing, as well as process messages received from connected devices. The software allows one to configure the instrument into three mutually exclusive settings, corresponding to the use cases envisioned by participants given technological constraints, namely due to the limited processing power of the Bela board [15]. The three configurations can be selected from a dedicated smartphone app developed using the TouchOSC environment, which allows one to rapidly build modular control surfaces for mobile applications using the OSC protocol. The app acts both as a visual display and as a control interface letting players or audience members program the instrument in real-time.

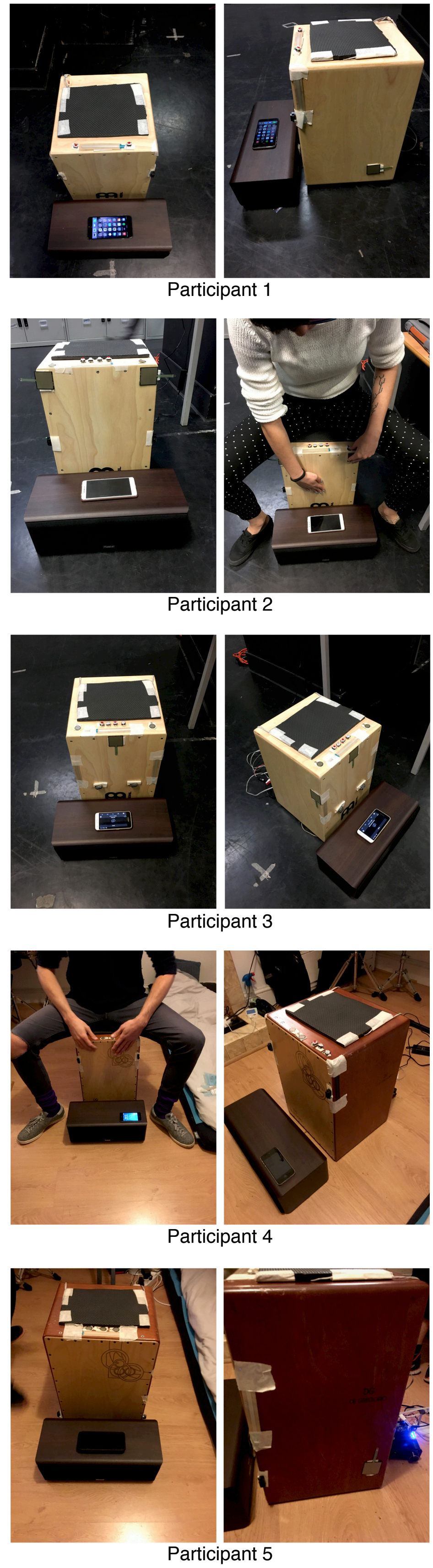
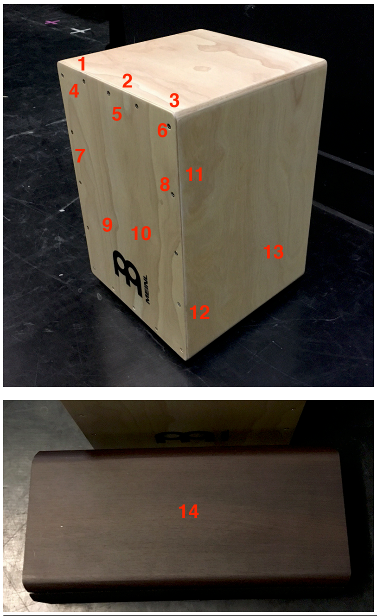


Fig. 1: Mock-ups produced by the five participants.

(i) Virtual instrument configuration. In the first configuration, the Smart Cajón let performers select and play

Table 1: Analysis of the five mock-ups and features selected for the design. Numbers on the photos refer to specific parts of the cajón. Participants' choices are reported in brackets.



Area	Sensor type	Design
Top panel		
1	rounded FSR [P2,P3];	rounded FSR
2	ribbon [P1,P2,P3,P5]; 4 buttons [P2,P3]; 2 buttons [P1]; 2 rounded FSR [P4], 3 rounded FSR [P5]	ribbon, 4 buttons
3	rounded FSR [P2,P3]; 3 buttons [P4]	rounded FSR
Front panel		
4	squared FSR [P2]	-
5	squared FSR [P3]	-
6	squared FSR [P2]	-
7	distance [P1,P2,P4]; ribbon [P3]	distance
8	distance [P1,P2,P4]; ribbon [P3]	distance
9	distance [P3]	-
10	distance [P3]	-
Left and right side panels		
11	ribbon [P1,P4,P5];squared FSR [P3]	ribbon
12	distance [P5]	-
13	squared FSR [P1,P5]	-
Loudspeaker		
14	display [P1,P2,P3,P4,P5]	display
Interior side		
-	IMU [P1,P2,P3,P4,P5]	IMU

different percussive instruments. Two banks of four presets were created to simulate four percussive instruments: drums, congas, bongo, djembe. To trigger the sound samples we detected each hit by using the PD onset detector bonk~ described in [17], configured with a 256-point window, and with low and high thresholds for onset detection set to 5 and 100 respectively (these thresholds are adjustable via wireless messages and were empirically tuned). To detect different hit positions we developed a semantic audio processing technique combining spectral content and amplitude information (see Figure 3). The technique relies on the fact that sounds produced by hitting the top part of the front panel (including the top edges) have a richer high frequency content than sounds produced by hits on the central part for which the lower frequency content is dominant (these are the two regions mostly used by cajón players, as playing below the central part is impractical). To capture these differences, we computed the spectral centroid from the output of the bonk~ object and used a discriminative threshold. Specifically, based on experimental findings for the specific cajón available, we calculated the centroid by utilizing the loudness and bandwidth of the first 5 of the 11 frequency bands produced by bonk~, and set the discriminative threshold to 2.55 (this threshold could be interactively adjusted and was tuned together with a cajón player). The hits associated to these two regions were then mapped to two distinct sound samples of the simulated instrument. The dynamics of each hit tracked by the bonk~ object were mapped to the volume of the triggered samples (specifically, mapping the bonk~ dynamic range [0, 122] to the volume range [0, 9] dB). The latency between the hit detected on the input signal and the triggered sound sample had a mean of 19.92 ms and a SD of 3.62 ms (computed on 40 samples using an oscilloscope).

The banks and presets could be navigated and selected using the smartphone app. Each preset was associated to a button with a specific LED color. The status and color of the buttons were synchronized with information displayed in the app. Another GUI in the app displayed textual content simulating messages conveyed by the audience. For testing purposes such messages could be sent from a laptop running a PD patch. While sensors were not used in this configuration, motors were set to deliver the four types of vibrations described in Section 2.2.2.

Fig. 2: Front and back views of the implemented prototype.

(ii) Audio effects configuration. In the second configuration, the instrument was set to process the sound detected by the contact microphones using a 10-band parametric equalizer chained with a reverberation effect. Eight presets divided into two banks proposed different tunings of the two effects. These presets could be navigated and selected using both the buttons and the smartphone app. The app also displayed the status of the parameters of the two

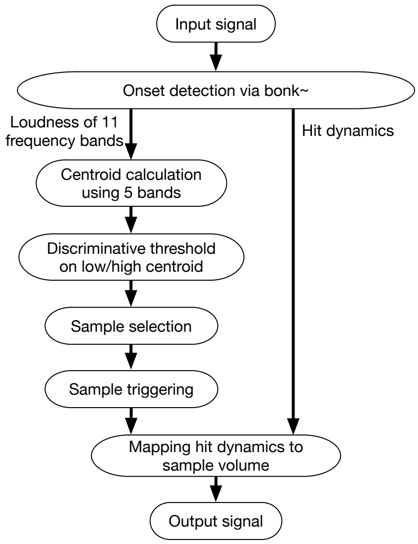


Fig. 3: Diagram illustrating the semantic audio processing technique involved to trigger sound samples of virtual instruments.

effects using faders and rotary knobs widgets. Neither the sensors nor the motors were used in this configuration.

(iii) Interactive sound control configuration. In the third configuration, the sounds captured by contact microphones were processed by audio effects, the parameters of which can be controlled from sensors. The sensors were linearly mapped to effects that were mentioned by participants: the left and right FSRs controlled continuously the input volume of a delay effect with feedback, configured in two different ways so to have fast and slow decaying repetitions (left FSR: volume variation = [0, 4.4] dB, delay time = 400 ms, delay feedback = 0.5; right FSR: volume variation = [0, 4.4] dB, delay time = 150 ms, delay feedback = 0.6); the pitch component of the IMU was mapped to the room size parameter of a reverb (range = [0.01, 0.98]); the roll of the IMU was mapped to a frequency shifter, in such a way that tilting the instrument to the left/right side caused a shift towards low/high frequencies (left tilting variation = [0, -200] Hz; right tilting variation = [0, 500] Hz); the four buttons were used to start or stop four backing tracks in different genres (flamenco, Brazilian jazz, blues, electronic pop); the ribbon sensor (under which a pressure sensor was placed so as to detect both pressure and position) was used to control the overall volume of electronically-generated sounds (volume variation = [-6, 9] dB). The smartphone app displayed the parameters controlled by sensors in real time through knobs and faders. The motors were not used in this configuration.

4 EVALUATION

Traditional human computer interaction user evaluation methods based on task-based usability metrics such as efficiency and effectiveness are ill-adapted for the evaluation of digital music interfaces used in creative performances (see e.g., [18]). Given that the performer-instrument relationship can be “intimate” we opted to conduct in-depth evaluation sessions with a single participant at a time. This was to ensure participants had enough time to explore the instrument and would not be biased by other performers while providing feedback. We report in this section the nature of the procedure and the results obtained with four of the performers who took part in the co-design session.

4.1 Procedure

The evaluation consisted of two stages. In the first stage, we used the concurrent think-aloud protocol (CTA) [19] to gain feedback about interaction issues, the technical quality and interactions made possible with the Smart Cajón prototype in the different configurations described in Section 3.3. For each configuration, the experimenter first conducted a feature walkthrough describing apparatus and functionality. Then the participants were invited to explore and perform with the instrument. The CTA protocol was applied during the explore/play stage. Each configuration was assessed both without and with an accompanying track played by a laptop or by the instrument itself.

The second stage consisted in a semi-structured interview aimed at better understanding the participants’ interaction with the system through reflective feedback, and to guide future developments. The interview included questions about user satisfaction, experience, as well as artistic intent and envisioned applications using the proposed smart technology. The evaluation sessions lasted on average 2 hours and 30 minutes and were video-recorded.

4.2 Evaluation results

Virtual instrument configuration CTA results. All participants wished that the instrument could respond to a larger variety of positions and hit techniques, have a greater sensitivity to dynamics, provide higher sound quality, as well as allow performers to produce subtle timbre nuances.

The responsiveness to the position was the most critical factor. Most of the times the prototype correctly responded to the intention to trigger different sounds when changing hit position. However, correct identification of hits failed in presence of playing techniques producing sounds with frequency content richer in high (low, respecively) frequency even if the hit happened on the bottom (top, respecively) region. Moreover, P2 suggested to associate different sound samples to the regions adjacent to the two top corners. Along the same lines, P3 suggested to *“add to the side panels different sounds belonging to the same instrument, like different snares or cymbals if you are simulating a drum kit.”*. Interestingly, P5 reported *“I would like that the instrument could blend the two sounds: sound A and B trigger together in some zones in between the centre and the edges, like 75% sound A 25% sound B. I would also like to be able to use the two zones [top and bottom] at the same time so to have sound A and B fully triggering simultaneously.”*

Another area of improvement for all participants was the capacity of the instrument to render more accurately the dynamics of the hits (e.g., P3 reported *“I would like to have the same sensitivity of the microphone, even the softest touches should be rendered”*). This was followed by a request for rendering of timbral variations (e.g., P4 reported *"The timbre of the sound samples should better follow my touch, rendering different dynamics of the same sound sample is not enough"*). Related to this, all participants reported that the quality of the sounds of the simulated instruments should be improved (e.g., P3 reported *"I would like to have more realistic sounds: it must really sound like a conga"*). Interestingly, P4 reported *"The [virtual instruments] sounds should be blended with the acoustic sound of the cajon, otherwise why using a cajon? Pads giving only the electronic sound would be enough. I suggest to design [virtual instruments] sounds that can differentiate more from the acoustic sound."*

None of the participants reported to perceive any latency between their actions and the triggered sounds. This is likely to be due to a masking effect in the attack of the acoustic sound that superimposes over the digital one.

Importantly, it was observed that all participants adapted their playing technique to the instrument limits in order to get the most out of it. For instance, P2, P3, and P5 avoided techniques that could be misinterpreted by the system or used types of touches capable of enhancing the sound high frequency content in the top region (e.g., slapping stronger, using the knuckles or the nails) or the sound low frequencies in the bottom region (e.g., using the palm). P3 avoided using very soft touches, while P5 avoided playing the two regions simultaneously. On the other hand, P4 exploited creatively the system failures to trigger, with specific techniques (e.g., using the fingertips), the low frequency sound samples in the top zone.

All participants used creatively the buttons to pass from a virtual instrument to another while staying in the tempo. P4 also used the buttons to combine sounds of different virtual instruments by alternating hits on an instrument and a single hit on another instrument.

Regarding haptic feedback, P4 and P5 reported to be able to distinguish well the four tactile stimuli provided at the backside while playing, while P2 suggested that a maximum of two radically different stimuli should be used in order to not distract too much the player. P5 stressed that the timing of the vibration patterns should not be in conflict with the tempo of the played song, and suggested to use tactile sensations for monitoring purposes: *"Often when you play on stage you can't hear yourself properly, vibrations related to your playing would help."*.

All participants expressed that audience members could be empowered to change the virtual instrument, but that they should not be able to perform such a change whenever they wanted. It was deemed important to let performers decide when to change the sound. P2 reported that performers should be notified in advance of changes induced by the audience to be able to conclude a rhythmic pattern and anticipate another pattern with the new instrument while maintaining musical timing. P2 and P4 suggested that the notifications related to audience control should be displayed on the screen and that a concurrent tactile vibration should be used to let performers know when to look at the screen.

Audio effects configuration CTA results. P3, P4, and P5 reported that the instrument's responsiveness was capable of rendering very well all their touches (e.g., for P4 *"sensitivity and levels of dynamics are great and much better than the first configuration"*). P2 and P4 primarily suggested to improve the sound quality of the eight designed presets. This issue could be overcome using a different type of reverberation and equalizer as well as by designing sounds together with the participant. All participants agreed on the importance of programming the parameters of the effects directly via an app and customize sounds as desired. The possibility of rapidly switching preset using buttons on the instrument was also appreciated by all participants. It was observed that this feature allowed participants not only to extend the range of sonic possibilities compared to a conventional acoustic playing, but also to rapidly and effortlessly switch from one timbre to another within a same piece.

As for sound control, in the case of audience participation, the performers judged important not to let audience members alter audio effect presets or parameters without any notifications to anticipate such interventions. Nevertheless, all participants agreed with the fact that audience control of equalization and reverb parameters was less problematic than that of virtual instruments. Beyond audience participation, P4 suggested to provide some levels of control to sound engineers *"I would prefer granting the possibility of changing the eq to the sound engineer to better mix my sound with that of the band"*.

Interactive sound control configuration CTA results. P3, P4, and P5 showed immediately great confidence in using the sensors in finding ideas on how to exploit the offered expressive potentialities. For P2 at first, it was difficult to make use of continuous sound effect control through gestures. However, after about 10 minutes of practice, P2 started to feel more confident about using those gestures and incorporating them in the normal playing technique.

All participants reported that the front-back tilting was a more appropriate gesture than side tilting and it was used more often in the exploration stage. Controlling sound via left or right tilting the instrument was reported to be less comfortable by all participants. Nevertheless, such gesture was used extensively and participants suggested to find a mechanical system to improve this interaction. P3, P4, and P5 used mostly right-tilting to enhance the high pitch sounds generated playing on the top part of the instrument. Tilting to the left was rarely used by participants because it did not lead to a drastic change in the sound (e.g., for P4 *"it is better to pitch to high than to low frequencies because low frequencies are lost with the acoustic sound. To feel them better you would need a subwoofer"*). Combining front-back and side tilting was found to be even more difficult and impractical, nevertheless all participants used it often while playing. All participants could well integrate pressure sensors generating the delays with front/back tilting activating the reverb. With the exception of P3, the delay was used less than the other effects. All participants suggested that the delay time had to be synchronized with the BPM of the backing track used, and that this feature should be applied to songs uploaded or streamed from the Internet. All participants used less frequently the volume control associated to the ribbon sensor while playing.

Semi-structured interview. For P2 and P3, the most important advantage of the Smart Cajón was the possibility to produce sounds from other percussive instruments. Such a feature was particularly appreciated because it could let cajon player have extra sonic possibilities without having to carry too many other instruments saving time, space, and effort. All participants particularly appreciated the possibility of easily switching timbre or instrument via the buttons while playing. All of them requested to integrate the three configurations so to have effects applied to virtual instruments and controlled via sensors.

Participants also valued very much the possibility of playing over backing tracks generated directly from the instrument. P3 reported *"With this system you give the cajon player the possibility to be more independent: to rehearse, to bask, to compose your own music and then play it alone so to be a one man band, especially if you are able also to sing. This is not possible for cajon players now."*. P4 suggested *"I would also like to have a metronome, with controls, integrated for practicing technique"*.

P2 was the sole participant who stressed on the fact that training and time would be required to learn and master all the novel possibilities enabled by the sensors. All other participants showed confidence from the beginning in using the sensors, and integrating the new gestures in the normal playing technique and in combining them. When participants were asked if they were adapting to the instrument all of them responded positively. P2 reported *"It took me some time to understand how to avoid failing the system but then I tried to express myself with it"*. P3 reported: *"I adapt to the instrument since [in the virtual instruments configuration] there is not the level of dynamics I would like. As a performer you are used to adapt to the instrument."*. P4 was the only performer who understood that the hit tracking system was not truly based on position but on sound frequency: *"I was adapting because I was realizing that the system is frequency based, so I searched the sound. Sometimes I slapped stronger to make sure I could get the high frequency sound. Other times I decided to produce the low pitch sound were it is normally supposed to be high"*.

Several suggestions were made to improve the user experience. P2 and P5 recommended to have a larger screen placed on the loudspeaker to better see the displayed information. P2 and P4 suggested that the system could be used to write in real time a MIDI score of the rhythmic patterns played on the Smart Cajón (including the differentiation of the hits on the instrument regions). P3 proposed that the system could display in real-time the BPM utilized by the player. P2 and P5 suggested to record the sound produced by the instrument and to upload it to a computer for future editing. P4 and P5 asked to have two volumes, one for the instrument sounds and the other for the backing tracks. P2 and P4 suggested to use the tactile stimuli to notify the performer to look at the screen when information was conveyed by the audience. P2 and P3 felt interested to receive information from the audience via the screen such as messages to play louder, faster/slower, change instrument, and messages about the overall emotional status of the audience, for instance via emoticons.

5 DISCUSSION AND CONCLUSIONS

The first noticeable element from the evaluation sessions was that all participants were very enthusiastic about the instrument and its potentialities.

A clear phenomenon of appropriation [10] emerged as the result of both affordances and constraints of the instrument. Firstly, all participants personalized the integration in their normal playing technique of the new gestures afforded by the sensors, generating different ways of expressing themselves. Importantly, these novel pathways for expression are not possible with commercially available cajones that are devoid of the involved sensor types, positions, and mappings. Secondly, participants adapted to the instrument, coping with its limitations and exploiting them in a creative way. In most of the cases participants tried to avoid normal playing techniques that could bring to a malfunctioning of the instrument. In the same vein, the lack of a great sound variation due to a sensor-based gesture (i.e., tilting to the left for pitch down) was felt like a constraint of the instrument and therefore participants focused more on other gestures resulting from sensors. In other cases the flaw of the instrument was instead used in a creative way (e.g., to deliberately trigger low frequency sounds in the top position where they should not be present). It was also observed that participants played faster tempi in presence of sounds of short durations and slower tempi when the instrument was configured with sounds of long duration. Also, participants reacted to the virtual instruments changing their playing styles like if they were actually playing the real instrument being simulated. All these considerations are in line with research finding that show how musicians respond creatively to specific affordances and constraints that they are given [8, 9, 10, 11].

It is worth noticing that this process of appropriation was different for each participant. For P2 the adaptation was initially more problematic than for the others participants, who were immediately capable of coping with and taking advantage of the instrument constraints. P2 also used the new affordances less creatively compared to the other musicians (e.g., using less combination of gestures). This is in line with Cook's statement that some musicians have "spare bandwidth" [20].

Below we summarize findings from the co-design and evaluation sessions which we deem will be useful for designers of smart instruments, in particular cajones. Most of the listed features are not present in commercially available cajones and their implementation would progress the industry state-of-the-art.

Versatility. The simulation of different percussive instruments allows performers to avoid carrying to a show multiple instruments which may reduce setup time. This feature is present also in commercial augmented cajones, which, however, do not offer possibilities to change the timbre-quality of the instrument while playing since they do not provide real-time processing of the acoustic sound of the instrument. Players are attracted by Smart Cajones that allow them to easily switch between presets and different equalizations, even while playing. They also appreciate the possibility of playing and changing backing tracks from the instrument for one-man-band or rehearsing purposes.

Programming, upgrading, and recording. The possibility of fully customizing the configurations of the instrument (e.g., presets for timbres, chain of audio effects, and simulated instruments) has shown to be a fundamental feature for performers involved in the study. Mechanisms to update/upgrade the instrument with banks of sounds downloaded from the Internet is another interesting feature that was requested by some of the participants. The participants also felt interested to practice using backing tracks downloaded or streamed from the Internet, without requiring any external speakers. These features are not offered by commercial augmented cajones, which provide standard and unmodifiable set of sounds simulating percussive instruments and very few tools to select and program them. Furthermore such augmented instruments do not enable the recording of the instrument sounds, which is a feature found useful by players, along with the possibility of wirelessly transferring the recordings to a computer.

Ergonomics and sensor mapping. The addition of sensors tracking gestures different from supplementing hit detection provided in augmented cajones entails a radical rethinking of the instrument and of its practice. The most natural gesture for continuous sound control is the back-forth tilting, as this is a movement commonly done by cajón players while playing. The use of buttons placed on the instrument to enable rapid change of presets while playing was greatly appreciated. Such control gestures can be easily integrated into the conventional playing technique, while pressing regions of the instrument with the fingers is found to disrupt much more the natural interaction of the performer. An embedded screen proved to be a desirable feature, namely to see preset information and, in case of audience creative participation, to display information received from audience members. Participants were also interested by having access to an smartphone or tablet app to complement the display of such information.

Sound quality and touch fidelity. Participants reported on the importance to have sounds of high quality and to be able to render subtle nuances that follow different types of percussive hits. A general dissatisfaction about the sound quality and responsiveness of commercially available augmented cajones emerged from players' feedback, and this was expressed too for the presented prototype. Improvements in expressive control and gesture-to-sound mapping can be addressed using semantic audio techniques. In future work we will investigate how a richer knowledge about the musical gestures can be gained through multimodal audio and signal analyses and machine learning.

Music information retrieval. The intelligence embedded in a Smart Cajón can be exploited to extract information related to music playing (e.g., parameters of each hit, BPM used), which could be exploited for learning purposes or for supporting composition and production. In future work we will exploit the semantic audio process described in Section 3.3 for extracting and displaying in real-time the BPM performed by the player, and for creating a score accounting for the position, timing, and dynamics of the various types of hits.

Embedded tactile feedback. Tactile notifications were judged to be useful. Few, radically different, and strong vibration signals should be used to convey information while not overloading the player with information during performance. Such information could be used to inform players when to start or stop playing, and to look at the information displayed on the embedded screen.

Sharing control with the audience. Creative participation from the audience should be agreed and planned for in advance, and performers should be able to decide when to effect intended interventions from the audience.

6 ACKNOWLEDGMENT

The work of Luca Turchet is supported by a Marie-Curie Individual fellowship of European Union's Horizon 2020 research and innovation programme, under grant agreement No. 749561. Mathieu Barthet also acknowledges support from the EU H2020 Audio Commons grant (688382).

7 REFERENCES

[1] C. Michalakos, "The Augmented Drum Kit: An Intuitive Approach To Live Electronic Percussion Performance," presented at the *Proceedings of the International Computer Music Conference* (2012).

[2] J. Gregorio, P. English, K. Youngmoo, "Sound and Interaction Design of an Augmented Drum System," presented at the *Proceedings of AudioMostly Conference*, pp. 2:1–2:4 (2017), URL https://doi.org/10.1145/3123514.3123521.

[3] D. Ludwigsen, "Acoustic and structural resonances of the cajon," presented at the *Proceedings of Meetings on Acoustics 170 ASA*, vol. 25, p. 035005 (2017).

[4] L. Turchet, A. McPherson, C. Fischione, "Smart Instruments: Towards an Ecosystem of Interoperable Devices Connecting Performers and Audiences," presented at the *Proceedings of the Sound and Music Computing Conference*, pp. 498–505 (2016).

[5] B. Kostek, *Perception-Based Data Processing in Acoustics: Applications to Music Information Retrieval and Psychophysiology of Hearing* (Springer Publishing Company, Incorporated), 1st ed. (2010).

[6] L. Turchet, M. Benincaso, C. Fischione, "Examples of use cases with Smart Instruments," presented at the *Proceedings of AudioMostly Conference*, pp. 47:1–47:5 (2017), URL https://doi.org/10.1145/3123514.3123553.

[7] L. Turchet, C. Fischione, M. Barthet, "Towards the Internet of Musical Things," presented at the *Proceedings of the Sound and Music Computing Conference*, pp. 13–20 (2017).

[8] T. Magnusson, "Designing constraints: Composing and performing with digital musical systems," *Computer Music Journal*, vol. 34, no. 4, pp. 62–73 (2010), URL https://doi.org/10.1162/COMJ_a_00026.

[9] A. P. McPherson, Y. E. Kim, "The problem of the second performer: Building a community around an augmented piano," *Computer Music Journal*, vol. 36, no. 4, pp. 10–27 (2012), URL https://doi.org/10.1162/COMJ_a_00149.

[10] V. Zappi, A. McPherson, "Dimensionality and Appropriation in Digital Musical Instrument Design." presented at the *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 455–460 (2014).

[11] R. H. Jack, T. Stockman, A. McPherson, "Rich gesture, reduced control: the influence of constrained mappings on performance technique," presented at the *Proceedings of the International Conference on Movement Computing*, pp. 15:1–15:8 (2017), URL https://doi.org/10.1145/3077981.3078039.

[12] Y. Wu, L. Zhang, N. Bryan-Kinns, M. Barthet, "Open Symphony: Creative Participation for Audiences of Live Music Performances," *IEEE MultiMedia*, vol. 24, no. 1, pp. 48–62 (2017), URL https://doi.org/10.1109/MMUL.2017.19.

[13] S. Bødker, K. Grønbæk, "Cooperative prototyping: users and designers in mutual activity," *International Journal of Man-Machine Studies*, vol. 34, no. 3, pp. 453 – 478 (1991), URL https://doi.org/10.1016/0020-7373(91)90030-B.

[14] M. J. Muller, J. H. Halswanter, T. Dayton, *Handbook of Human-Computer Interaction*, chap. Participatory Practices in the Software Lifecycle (Elsevier Science B.V.) (1997).

[15] A. McPherson, V. Zappi, "An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black," presented at the *Audio Engineering Society Convention 138* (2015).

[16] T. Mitchell, S. Madgwick, S. Rankine, G. Hilton, A. Freed, A. Nix, "Making the Most of Wi-Fi: Optimisations for Robust Wireless Live Music Performance." presented at the *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 251–256 (2014).

[17] M. Puckette, T. Apel, D. Ziccarelli, "Real-time audio analysis tools for Pd and MSP," presented at the *Proceedings of the International Computer Music Conference* (1998).

[18] D. Stowell, A. Robertson, N. Bryan-Kinns, M. D. Plumbley, "Evaluation of Live Human-computer Music-making: Quantitative and Qualitative Approaches," *International Journal of Human Computer Studies*, vol. 67, no. 11, pp. 960–975 (2009), URL https://doi.org/10.1016/j.ijhcs.2009.05.007.

[19] T. Alshammari, O. Alhadreti, P. J. Mayhew, "When to Ask Participants to Think Aloud: A Comparative Study of Concurrent and Retrospective Think-Aloud Methods," *International Journal of Human Computer Interaction*, vol. 6, no. 3 (2015).

[20] P. R. Cook, "Principles for Designing Computer Music Controllers," presented at the *Proceedings of the International Conference on New Interfaces for Musical Expression* (2001).

THE AUTHORS

Luca Turchet

Andrew McPherson

Mathieu Barthet

Luca Turchet is a Marie-Curie postdoctoral researcher at Queen Mary University of London. He received master degrees in Computer Science from University of Verona, in classical guitar and composition from Music Conservatory of Verona, and in electronic music from the Royal College of Music of Stockholm. He holds a Ph.D. in Media Technology from Aalborg University Copenhagen. His scientific, artistic, and entrepreneurial research has been supported by the European Commission, the Italian Minister of Foreign Affairs, and the Danish Research Council.

•

Andrew McPherson is a Reader in the Centre for Digital Music (C4DM) at Queen Mary University of London. A composer (PhD UPenn 2009) and electronic engineer (MEng MIT 2005) by training, his research focuses on digital musical instruments, especially augmented instruments which extend the capabilities of familiar instruments. Within C4DM he leads the Augmented Instruments Laboratory, a research team exploring musical instrument design, human-computer interaction and embedded audio.

●

Mathieu Barthet Ph.D. is a lecturer in Digital Media and technical director of the Media and Arts Technology Studios at Queen Mary University of London. His research lies at the intersections of music information retrieval, perception and human computer interaction. He currently investigates new interfaces for musical expression, listening and participatory performance, and also conducts fundamental research on musical timbre and emotions. He is Principal Investigator of the EU Internet of Musical Things project and Co-Investigator of the EU Audio Commons project.