TRIPHON

A Physical Model Feedback Instrument controlled with Gestures

Emmanouil Dimogerontakis, Marco Fiorini and Vasilena Sotirova Popangelova
Aalborg University Copenhagen, Sound and Music Computing
8th Semester Project Report

***Abstract*—This research explores the possible combinations of interdisciplinary fields of physical modelling and embodied music cognition, through a new digital feedback instrument controlled with gestures. Starting from an existing model of Self-resonating vibrotactile instruments, a string instrument with coupled finite different schemes have been implemented in Faust and ported to Max/MSP. Here, gestures recorded from a smartphone and handled through Open Sound Control have been mapped to the synthesis' parameters, using a set of computed mid-level movement descriptors. After performing an evaluation of the instrument, in the context of a sound installation, qualitative and quantitative data have been reported, showing a strong confirmation of the initial research investigation.**

I. INTRODUCTION

Since when electroacoustic feedback was first characterised by Larsen in 1966 as a convergence to a pure tone in a system with a positive loop gain [1], musicians began to develop different techniques for manipulating it musically, resulting in a wide variety of forms (see [2] for a detailed summary and analytic framework). In the last few years, an international collective of researchers, musicians and designers aiming to exchange and generate strategies, concepts and practices of feedback musicianship, to build community and new musical collaborations, raised under the name of Feedback Musicianship Network [3]. The workshops, concerts and gatherings of this network brought out complex new research challenges, trying to fill the large gaps in new needs for composition, notation and performance techniques and new understandings of virtuosity [4]. Out of this collective, notable research has been released, including ways to shape the behaviour of feedback instruments [5] and the design of *Self-resonating vibrotactile instruments* (SRIs), hybrid feedback instruments, characterised by an electro-mechanical feedback loop that is both the means of sound production and the expressive interface [6], as shown in Figure 1.

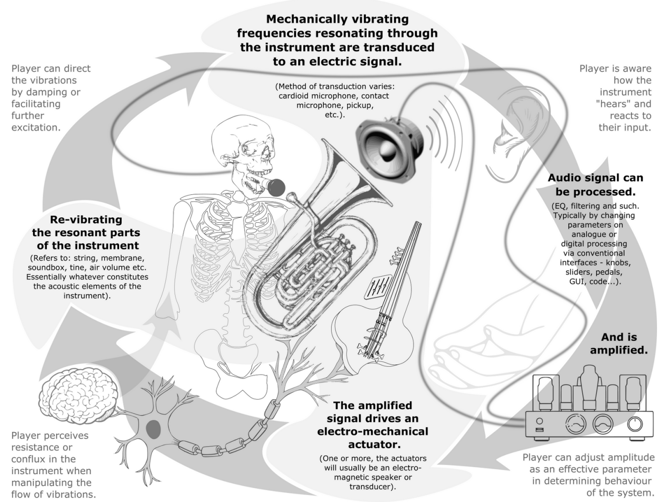

Fig. 1. Overview of the SRI organisation and shared control with human player, from [6]. This figure is present also in the main page of [3] and is shown in this report because of its artistic value and detailed accuracy of the model.

Kiefer gave a summary of what's happening in the world of feedback instruments nowadays [7] and a journal collecting numerous articles on feedback research in different fields has been recently released [8]. Nevertheless, works reporting significant interaction between feedback concepts and physical modelling techniques seem to be lacking, despite the broad range of possibilities of the two areas of study. That's why we decided to focus our research on the development of a digital feedback instrument, where the sound source is synthesized by physical modelling techniques. Furthermore, we considered the work of Visi et al. [9] [10] on embodiment of music cognition and designing of gestural mapping strategies in music performances, and we wanted to investigate the possibilities of controlling this instrument through gestures. Building upon van Dijk's framework of *Designing for Embodied Being-in-the-World* (D4EB) [11], the idea was to design an hybrid artefact conceptually situated in between the inner person and the outer environment, and in between the representational world of information and the purely physical world, operating as a unified whole to support and transform the ways in which the lived body is active in relation to the *"lifeworld"* (the world in which our lived body operates [11]). This resulted in choosing the smartphone, an everyday tool used by all kind of people to make sense of the world in which we are connected, as the object designed to control this instrument.

This led the authors to the following hypothesis/problem statement: *"We will design a new instrument, using gesture tracking from a smartphone to control a digital feedback physical model, in the context of a sound installation"*.

II. Problem Analysis

A. The signal path of a string SRI

Hybrid feedback instruments consist of different parts, mechanisms and architectures. In this specific project the aim is to mimic the mechanisms of a bowed string feedback instrument, thus, a brief explanation of the signal path and the mechanism will be presented in this section. A direct meaningful inspiration from the Feedback Musicianship Network has been given by the FAAB (Feedback-Actuated Augmented Bass), a project by instrument builder Halldór Úlfarsson and double bassist and programmer Adam Pultz Melbye [12]. Therefore, considering a stringed instrument with pick-ups in each string and a speaker attached to the body, the signal path will result as in Figure 2. When a string is excited, the energy travels from the string itself to the body of the instrument through the bridge. Here, the energy activates the resonator (the body) and, after being captured by a microphone, or a pick-up, is routed to a speaker that drives the energy back to the body (resonator). The bridge then transfers again this energy back to the strings, resulting in a self-resonating phenomenon similar to that of sympathetic strings. In the next sections, these interaction dynamics will be presented as two different systems, following the exciter-resonator model (see Section II-C). To mimic these mechanisms, it is first considered a physical model of 3 strings attached to a bar, implemented with FDS schemes (Section II-D). To imitate the resonator, a partitioned convolution reverberator is implemented with a feedback path from the output to the input, using a pitch shifter to create rich harmonic content. A similar architecture can be encountered in the shimmer reverb [13], but here, instead of Feedback Delay Network (FDN) reverberator, convolution is used (more about that in Section II-H).

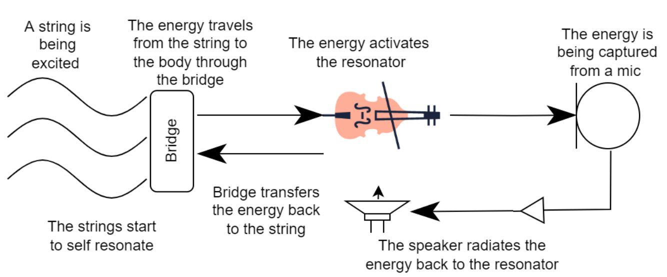

Fig. 2. The signal path of the bowed string SRI.

B. Physical Modelling of Musical Instruments

Physical modeling has quite a long history in the field of sound synthesis. It is based on the idea that when the vibrating structure is simulated in exactly the right way, the sound produced by that model is identical with the sound of the corresponding physical object. Thus, physical modelling of musical instruments simply means that the physical structure of a musical instrument is being modelled with mathematical and physical formulas realized with a computer [14]. A fundamental difference between the physical modelling approach and other synthesis techniques is that the former tries to imitate the properties of the sound source (based on physical reality and controlled by physical parameters, with a specific model for each instrument and synthesis independent from analysis), while the latter focus on the properties of the sound signal heard by the listener (based on perceptual reality and controlled by perceptual parameters, with a general model for all instruments and synthesis starting from simple unit generators), as shown in Figure 3.

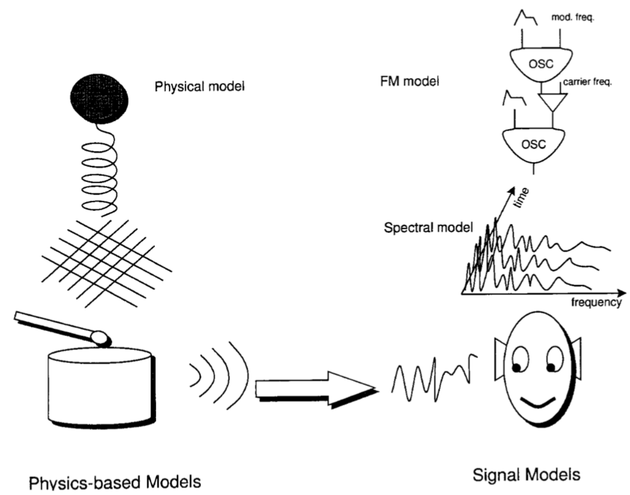

Fig. 3. Physics-based models and signal models, from [15].

Over the years, many different techniques have been proposed [15], such as digital waveguides [16], mass-interaction models [17], modal synthesis [18] [19] or finite difference schemes [20]. Even though sometimes more computationally demanding and difficult to control than other synthesis methods [21], physical modeling techniques offer many advantages. Indeed, creating a model of a vibrating system provides full control of its properties and, as a consequence, the output sound. These approaches theoretically allow us to synthesize natural and realistic sounds, tunable in every detail.

Therefore, as *"there are no theoretical limitations to the performance of a computer as a source of musical sounds, in contrast to the performance of ordinary instruments"* [22], the concrete applications of physical modellings techniques may range between a wide number of usages, including:

- Understanding sound production mechanism of musical instruments and everyday sounding objects (for acousticians),
- Developing algorithms which simulate in real-time sonorities created by such instruments (for computer scientists),
- Using the models to create sound effects and/or sonorities which do not exist in real-world (for sound designers and

composers).

Considering our idea of having a detailed physical model with a wide set of parameters to be mapped to gestures, we decided to focus on finite difference schemes. An addition to this choice is the fact that finite difference schemes present still relatively unexplored possibilities compared to the other physical modelling methods. Moreover, their high computational effort and accurate results seemed very interesting from a research point of view, to achieve new sonic results. Furthermore, the recent and close researches of Willemsen [23], Russo [24] and Sudholt [25], as well as their willingness to guide us in our personal investigation have been a relevant factor in our choice.

C. The Exciter-Resonator model

Nearly any musical instrument can be subdivided into a resonator component and an exciter component, both of which can be simulated individually, as shown in Figure 4. The resonator sustains and controls the oscillation, and is related with sound attributes like pitch and spectral envelope. The exciter is the place where energy is injected into the instrument, and it strongly affects the attack transient of sound, which is fundamental for timbre identification. The interaction of exciter and resonator is the main source of richness and variety of nuances which can be obtained from a musical instrument [15]. In the real world, the interaction between the exciter and the resonator is bi-directional. In other words, the exciter not only affects the state of the resonator, but the resonator affects the exciter as well.

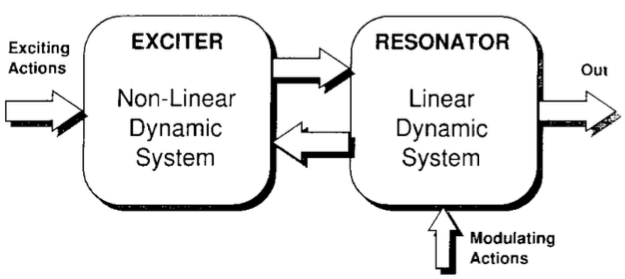

Fig. 4. Exciter–resonator interaction scheme, from [15].

D. Finite-Difference Time Domain Methods

Finite-Difference Time Domain (FDTD) methods aim to solve Partial Differential Equations (PDEs) by approximating them with difference equations, discretizing a continuous system into grid points in space and time. An extensive overview of FDTD methods in the context of sound synthesis is given by Bilbao in [20], while further investigation on the matter has been done recently by Willemsen in [23]. Although computationally expensive, especially when working with higher-dimensional systems, this technique could potentially accurately model any system, whether it is linear or nonlinear, time-invariant or time-variant.

1) Partial Differential Equations: In order to compute physical models with FDTD methods, the Partial Differential Equations (PDEs) in continuous time need to be discretized. As a first step, the system should be described as a PDE in continuous time. For the physical modeling of musical instruments, the most common PDE equation is the 1D wave equation [23]. This can describe transverse vibration in an ideal string, longitudinal vibration in an ideal bar or pressure in an acoustic tube [20] [23]. Its behavior can't be found in the physical world - as the physical realistic systems have damping - although the 1D wave equation is used in more complex models. The mathematical equation is shown in:

$$\partial_t^2 u = c^2 \partial_x^2 u \tag{1}$$

with $u = u(x,t)$ representing the state of the system at position x, in time t and being c equal to the speed of sound. In the first part of the equation, the second order time derivative of u is present, describing the acceleration, while in the right part there is the second order spatial derivative, which describes the curvature (in the case, of the string). It is easily understood that the acceleration of u is equal to the second spatial derivative of u in the same position. The wave speed of an ideal string can be described as $c = \sqrt{T/\rho A}$ with T $[N]$ to be the Tension, ρ $[kg/m^3]$ and A $[m^2]$ to be the cross-sectional area. Knowing the above, we can re-write Equation 1 as:

$$\rho A \partial_t^2 u = T \partial_x^2 u \tag{2}$$

With that we expressed the 1D wave equation in relation to the Newton's second law. In addition, since the system is distributed over space, boundary conditions are needed. The boundary conditions are the end points of the system, when $x = 0$ and $x = L$. In this project, the Dirichlet boundary conditions are used, which can be written as:

$$u(0,t) = u(L,t) = 0 \tag{3}$$

More information about the topic can be found in [20] and [23].

E. Finite Difference Schemes

As it is described above, for computing the wanted model of an instrument, we need to imprint the system as a PDE. As it is known, the systems of PDEs can describe real-world phenomena giving an analytical solution, although it is difficult to solve these equations analytically. Thus, it is possible to use numerical analysis methods, which are discretizing the system into subdivided grid points in both time and space, and solve these systems of equations using linear algebra. The numerical analysis technique which has been used in this project is called the Finite Difference Method, and the notation for both the PDEs and the discretized equations used is taken from [20].

1) Grid Function: To create the time-space grid which is needed for the PDE discretization, $u = u(x,t)$ becomes a grid function u_l^n with $t = nk$ and $x = lh$. Furthermore, h is the spatial step of the discretization and k is the time step, which can be computed from the sample frequency as $k = 1/f_s$. Space and time are indexed respectively by l and n, as integers.

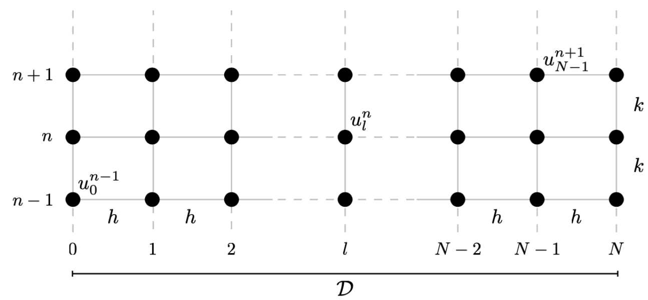

Fig. 5. Spatio-Temporal grid of a 1D system $u(x,t)$ with $x \in D$ discretized to a grid function u_l^n. The spatial domain D is divided into N intervals of length h and spatial range of interest $l = \{0, ..., N\}$. Time is subdivided into time steps of duration k and, together with the discretized spatial domain, forms a grid over space and time. Some grid points are labelled with the appropriate grid function [23]

2) Finite-Difference Operators: The FDS used for solving the PDEs of our model are explicit scheme solutions proposed from [20]. To calculate the discretized system it is needed to introduce the FD operators, which are taken again from [20].

$$\partial_t u \cong \begin{cases} \delta_{t+} u_l^n = \dfrac{1}{k}(u_l^{n+1} - u_l^n), \\ \delta_{t-} u_l^n = \dfrac{1}{k}(u_l^n - u_l^{n-1}), \\ \delta_{t\cdot} u_l^n = \dfrac{1}{2k}(u_l^{n+1} - u_l^{n-1}). \end{cases} \tag{4}$$

$$\partial_x u \cong \begin{cases} \delta_{x+} u_l^n = \dfrac{1}{h}(u_{l+1}^n - u_l^n), \\ \delta_{x-} u_l^n = \dfrac{1}{h}(u_l^n - u_{l-1}^n), \\ \delta_{x\cdot} u_l^n = \dfrac{1}{2h}(u_{l+1}^n - u_{l-1}^n). \end{cases} \tag{5}$$

$$\begin{aligned} \partial_t^2 u &\cong \delta_{t+}\delta_{t-} u_l^n = \delta_{tt} u_l^n = \frac{1}{k^2}(u_l^{n+1} - 2u_l^n + u_l^{n-1}), \\ \partial_x^2 u &\cong \delta_{x+}\delta_{x-} u_l^n = \delta_{xx} u_l^n = \frac{1}{h^2}(u_{l+1}^n - 2u_l^n + u_{l-1}^n). \end{aligned} \tag{6}$$

F. 1D Wave Equation

Above, the continuous model of the 1D wave has been presented in Equation 1. With the discretized methods presented in Equations 4, 5, 6, it is possible to calculate an approximating discrete-time model of the 1D wave equation. The approximation of the derivatives for the 1D wave equation is a straightforward procedure and the FD scheme is formed like this:

$$\delta_{tt} u_l^n = c^2 \delta_{xx} u_l^n \tag{7}$$

Expanding the operators from the given identities in the Equations 4, 5, 6 will result in:

$$\frac{1}{k^2}(u_l^{n+1} - 2u_l^n + u_l^{n-1}) = \frac{c^2}{h^2}(u_{l+1}^n - 2u_l^n + u_{l-1}^n) \tag{8}$$

To have an explicit solution it is needed to solve the previous equation for u_l^{n+1}, resulting in what it is commonly referred to as *update equation*:

$$u_l^{n+1} = 2u_l^n - u_l^{n-1} + \lambda^2(u_{l+1}^n - 2u_l^n + u_{l-1}^n) \tag{9}$$

Where $\lambda = ck/h$ is called Courant number. This number plays a significant role in the stability conditions of the scheme, and it needs to be declared as $\lambda \leq 1$. For more details see [20] [23]. The stability condition of the above scheme can be then written as :

$$h \geq ck \tag{10}$$

as k is a fixed value and the wave speed c is a user-defined value needed to calculate λ.

G. Connected Strings

The model used in this project consists of 3 stiff strings connected to a bar. This model is chosen because it can simulate the behavior of a bowed string feedback instrument. As previously shown in Figure 4, SRI instruments can be divided into the exciter-resonator model. Considering the behavior of such an instrument, presented in Section II-A, we will now implement a model of it using FDTD. In the physical world, the player of an SRI can control the sustain of the energy through continuous excitations (bowing) and amplification fluctuations of the speaker that radiates the energy back to the resonator. Using physical modeling with FDTD it is possible to simulate and manipulate physical parameters of the string and the material properties, such as the damping, radius of the string, density of the material etc. This possibility gives to the user the freedom to explore new parameters that can change dramatically the timbre and the pitch of the sonic outcome. For simplicity, a model of one stiff string attached to a bar will be presented.

1) Continuous Model: Considering a bowed stiff string taken by [20] [23], the transverse displacement of its length L_s in [m] is described as $u(x,t)$, with $\chi \in [0, L_s]$ being the coordinates along the length of the string. The PDE of the model can be seen as:

$$\begin{aligned}&\rho_s A_s \partial_t^2 u = T_s \partial_x^2 u - E_s I_s \partial_x^4 u - \\ &2\rho_s A_s \sigma_{0s} \partial_t \partial_x^2 u - \\ &\delta(\chi - \chi_B) f_B \Phi(\upsilon_{rel}) + \delta(\chi - \chi_C) f_C\end{aligned} \tag{11}$$

With ρ_s $[kg/m^3]$ as the material density, $A_s = \pi r^3$ $[m^2]$ to be the cross-sectional area with r $[m]$ as the radius and the T $[N]$ as the tension of the string. The string also has the material properties E_s $[Pa]$ as the Young's Modulus and $I_s = \pi r^4/4$ $[m^4]$ is the moment of intertia.

The coefficients σ_{0s} $[s^{-1}]$ and σ_{1s} $[m^2/s]$ are describing the frequency-independent and frequency-dependent damping, respectively. Additionally, f_B $[N]$ is the bowing force, which is being applied at a single point χ_B $[m]$ along the string, as specified by the Dirac function $\delta(\chi - \chi_B)$ $[1/m]$. Moreover, f_C $[N]$ is the force of the connection between the string and the bar, considering a rigid connection, applied at the location χ $[m]$, given by $\delta(\chi - \chi_C)$ $[1/m]$.

The relative velocity between the string at the bow position $x_B = x_B(t) \in D$ can be described as:

$$\upsilon_{rel} = \partial u(x_B, t) - \upsilon_B(t) \tag{12}$$

The friction model that is used in this project is proposed by [20] and more about this can be found at [20] [23]. The friction model can be written as:

$$\Phi(\upsilon_r el) = \sqrt{2\alpha \upsilon_{rel}} e^{-\alpha \upsilon_{rel}^2{}^{+1/2}} \tag{13}$$

The string has also simply supported boundary conditions at the edges of it, namely $x = 0$ and $x = L_s$:

$$\begin{aligned}u(0) = \partial_x^2 u(0) = 0, \\ u(L_s) = \partial_x^2 u(L_s) = 0.\end{aligned} \tag{14}$$

For the connection between the stiff string and the bar, it is assumed that they are rigidly connected. The connection force can be considered as the same at the connection points of both elements. The force has an opposite sign in one of the connection points, according to the action/reaction principle (Newton's third law). This means that the displacement of two connected points is always the same, and the distance should be 0 at all times. It is considered as:

$$u(x_c, t) = w(x_c, t) \tag{15}$$

with w representing the continuous state of the bar. Furthermore, the PDE of the bar can be considered as the PDE of the stiff string with 0 tension.

2) Discrete Model: The above PDE Equation 11 can be discretized in the following form with the identities that are given above in the Equations 4, 5, 6. The discretized system yields:

$$\begin{aligned}&\delta_{tt} u_l^n = c^2 \delta_{xx} u_l^n - \kappa^2 \delta_{xxxx} u_l^n - \sigma_{0s} \delta_{x.} u_l^n + \sigma_{1s} \delta_{t-} \delta_{xx} u_l^n \\ &- J_{lB}(x_B^n) F_B^n \Phi(\upsilon_{rel}^n) + J_{lC}(x_C^n) F_C^n\end{aligned} \tag{16}$$

with $F_B^n = f_B^n / \rho_s A$ to be the bowing force, $F_C^n = f_C^n / \rho_s A$ the connection force $J_{lB}(x_B^n)$ and $J_{lC}(x_C^n)$ $[m^{-1}]$ being the first order (or linear) spreading operators as described in [23].

The simply supported boundary conditions are discretized as:

$$\begin{aligned}u_0^n = \delta_{xx} u_0^n = 0, \\ u_0^N = \delta_{xx} u_N^n = 0.\end{aligned} \tag{17}$$

given by [20] [23]. The stability condition of the above discretized system taken by [20] is:

$$h \geq h_{min} = \sqrt{\frac{c^2 k^2 + 4\sigma_1 k + \sqrt{(c^2 k^2 + 4\sigma_1 k)^2 + 16\kappa^2 k^2}}{2}} \tag{18}$$

For the calculation of the force of the rigid connections, the above Equation 17 is considered as true, and according to that $u_l^n = w_l^n$. Knowing that, it is possible to solve the system explicitly for the force f_C^n as:

$$f_C^n = \frac{u^* - w^*}{\frac{k^2}{\rho_s A_s h_s (1 + \sigma_0 k)} + \frac{k^2}{\rho_b A_b h_b (1 + \sigma_{0b} k)}} \tag{19}$$

With u^* and w^* being the intermediate states of the stiff string and the bar models. This can be realized as the discrete Equation 11 without the forces.

Furthermore, the bowing force can be calculated from the given equation:

$$f_B^n(\upsilon_{rel}) = sgn(\upsilon_{rel})[\mu_C + (\mu_s - \mu_C e^{-(\upsilon_{rel}/\upsilon_s)^2}] + s_{2rel} + s_3 w \tag{20}$$

where $s_3 w$ is the random force term and μ the friction coefficient. Additionally, the relative velocity between the bow and the string will be:

$$\upsilon_{rel} = I_B(x_B^n) \delta_{x.} u_l^n - \upsilon_B^n \tag{21}$$

With the above equation, the scheme becomes implicit due to the centered difference operator. Then, an iterative root-finding algorithm is required, such as Newton-Raphson to solve it. More information about the implicit solution can be found in [20] and [23].

H. Partitioned Convolution

In the previous section, a mechanism to mimic the behavior of a self-resonating string feedback instrument have been described. In this section an approach to create a digital resonator (body of an instrument) with an audio feedback will be presented. The most computationally efficient way to represent realistic resonant spaces (rooms, instrument bodies etc) in the digital domain is the convolution with Impulse Responses (IRs). Since the instrument bodies are Linear Time Invariant (LTI) systems and the IRs are known, the Convolution Theorem can be used. In time domain the equation yields:

$$y[n] = x[n] \circledast h[n] = \sum_{m=-\infty}^{+\infty} (x[m]h[n-m]) \tag{22}$$

where $y[n]$ is the output of the convolution, $x[n]$ is the input signal and $h[n]$ is the impulse response.

Although, since it is known that convolution in time domain is not efficient for real time applications, convolution in frequency domain should be implemented, with the operation becoming a simple multiplication. This operation can be realized as:

$$\begin{aligned} &Y(\omega) = X(\omega)H(\omega), \\ &y[n] = IFFT(FFT(x[n])FFT(h[n])). \end{aligned} \tag{23}$$

where $Y(\omega)$ is the frequency response of the output, $X(\omega)$ is the frequency response of the input and $H(\omega)$ is the frequency response of the IR, while FFT is the abstraction of Fast Fourier Transform and IFFT is the abstraction of Inverse Fast Fourier Transform.

With this practice, the latency of the system is equal to the duration of the impulse response. To solve that problem, we need to introduce the partitioned convolution method. The process involves taking chunks of the impulse response h, into equally sized blocks of length K. Consequently, these blocks are convolved individually using overlap-and-save method [26]. These segments, that have a length of $L = 2K$, which is zero padded, are being transferred to the frequency domain with the FFT process, creating a number of filters S. After that, these filters are multiplied with the input blocks with length of L and are then overlapped with the latest multiplication. The whole process of the partitioned convolution is shown in Figure 6. The real-time implementation will be presented later in Section III-E.

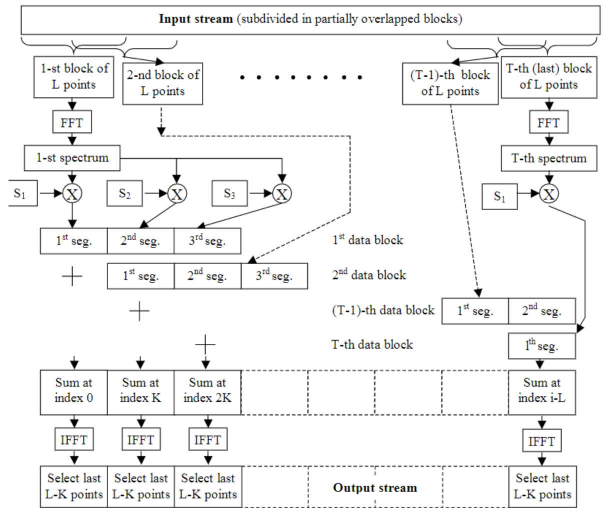

Fig. 6. A representation of partitioned convolution, from [26].

I. Embodied Interaction

Through the embodied music cognition paradigm it can be derived that the involvement of the body has a significant role in human interaction with music, and consequently on our understanding of that interaction [27]. The fundamentals surrounding embodied music cognition is that our knowledge about the perceived world (in this case music) is simplified through interaction. Thus applying embodied cognition theory to human interaction with technological systems has given rise to interaction as an embodied phenomenon [28]. The embodied interaction model is built around the question of how coupling of sensory faculties and physical actions convolved with social and environmental contexts can interpolate and influence the creation and manipulation of meaningful interactions with technology [29]. Having a mediating function, the body is able to build a collection of movements (gesture repertoire) achieving a particular goal (actions) linked with the experiences and sensations resulting from such actions [9]. Through forming a mechanism of action-perception couplings [27], the listener associates physical aspects of movement in space to expressive qualities and intentions, leading to the formation of musical intentional communication and expressiveness [9].

J. Gestures

The demand for new tools in terms of interactive possibilities offered by digital media technology has stimulated interest in gestural foundations of musical involvement [28]. Observing human motion in detail has brought up movement descriptors through which gestures in musical performance can be assessed [30]. Meaning, style and expressiveness can be communicated through a combination of multiple elements under human movement exploration [30]. The concept around gestures can be explained as a way of bridging movement and meaning, going further over the boundaries set between physical world and mental experiences [9].

As presented by Leman and Cirotteau, physical modelling has also put an emphasis on gesture modelling, since it is based on the movements of the physical components that make up the musical instrument [28]. Thus, generating sound is a question of controlling the articulatory parameters of the moving components [28]. It then makes sense also that a person would most likely add expressiveness to their interpretation, and this expressiveness would be based on the flow of body movements that take certain forms and shapes.

K. Movement Descriptors

Following several researches in the field of computable motion descriptors like Larboulette and Gibet [30] and Federico Visi [10], we were able to extract meaningful representations of human body motion. A bit of confusion exists, however, especially when talking about categorizing movement descriptors. Low-level motion descriptors represent dynamic or kinematic quantities directly derived from motion representations. Mid-level descriptors such as Quantity of Motion and Contraction index are often represented as high-level ones. This is not exactly correct since high-level descriptors are based on semantic components. These are structural notations like Laban Movement Analysis (LMA), and they can describe the structural, dynamic and geometric properties of human motion. As cited in Hackney [31], LMA is defined by four basic effort factors: *flow, weight, time, space*, where each one holds two opposing dimensions described by effort qualities. Sustained effort qualities, for example, are expected to have low level of jerkiness, while sudden quick characteristics would have a higher rate of change in acceleration [10]. Therefore, jerkiness and fluidity are valuable for analysing expressive movement qualities. Nevertheless, since Laban effort elements are "qualitative inner attitudes of a person moving towards the effort factor" [10], the use of computable descriptors should not be seen as an attempt to measure effort qualities quantitatively, but more so as a helpful approach for the design of computational models discerning movements of an expressive nature.

1) Velocity: Computes for one joint the rate of change of its position [30]:

$$v^k(t_i) = \frac{x^k(t_i+1) - x^k(t_i-1)}{2\delta t} \tag{24}$$

The speed of one joint is represented as the magnitude of its velocity [30]:

$$v^k(t_i) = \sqrt{v_x^k(t_i)^2 + v_y^k(t_i)^2 + v_z^k(t_i)^2} \tag{25}$$

2) Acceleration: In physics the definition of acceleration would be the rate of which velocity changes with time. As described in [30] by Larboulette and Gibet, it computes the instantaneous acceleration for one joint k and can be estimated by the following equation:

$$a^k(t_i) = \frac{x^k(t_i+1) - 2x^k(t_i) + x^k(t_i-1)}{\delta t^2} \tag{26}$$

3) Fluidity and Jerkiness: In the process of measuring kinematic quantities used to describe motion, "jerk" represents the variation of acceleration over time and is the third-order derivative of movement position [10]. In [32], Flash & Hogan define "jerk index" as the magnitude of the jerk averaged over the entire movement. Thus, jerkiness can be seen as the inverse of fluidity since it relates to the smoothness of the movement [10]. For one joint k it can be computed [30] as:

$$j^k(t_i) = \frac{x^k(t_i+2) - 2x^k(t_i+1) + 2x^k(t_i-1) - x^k(t_i-2)}{2\delta t^3} \tag{27}$$

From the definition for fluidity index, it can be derived that higher values of jerk correspond to lower fluidity [10]:

$$f^k(t_i) = \frac{1}{j^k(t_i) + 1\delta t} \tag{28}$$

4) Quantity Of Motion: Defined as the sum of speeds of a set of points multiplied by their mass by Fenza et al. [33], denoted as overall motion energy by Glowinski et al. [34] and computed by Visi in his modosc implementation as group of points taking into consideration the weight of each point [35], QoM is expressed by the equation [30]:

$$QoM^k(t_i) = \frac{\sum_{k \epsilon K} w_k . v^k(t_i)}{\sum_k w_k} \tag{29}$$

Proposing a straightforward implementation QoM can be quite useful when detecting the presence of the body movements during a real-time performance.

5) Periodic Quantity of Motion: Inspired by Quantity of Motion, in the need of a suitable descriptor that is able to describe multiple periodic gestures, PQoM was proposed as a way to measure a temporal quality in the movement [10]. It is expressed as:

$$PQoM^k(T) = \frac{1}{T} \sum_{i=1}^{T} QoM^k(t_i) \tag{30}$$

6) Time Effort: Defined by the sense of urgency, this high-level effort descriptor has two opposing dimensions - Sudden (quick) and Sustained (streched, steady) [30].

$$Time^k(T) = \frac{1}{T} \sum_{i=1}^{T} a^k(t_i) \tag{31}$$

7) Flow Effort: Explained as describing the continuity of the movement, it is denoted by the two opposite dimensions - Free (fluid) and Bound (restrained) [30]. The computation comes from the combined jerk over time. For the kth part of the body and a movement of length T, it is expressed as:

$$Flow^k(T) = \frac{1}{T}\sum_{i=1}^{T} j^k(t_i) \tag{32}$$

III. Implementation

The architecture of the whole system is presented in Figure 7. The user holds and move the smartphone, as the gestures are tracked by the phone sensors (gyroscope, accelerometer and magnetometer). The data are packed and sent to a PC using the OSC protocol [36] and mapped, using movement descriptors, to the parameters of the physical model in a Max/MSP patch [37]. The model is then connected to an Impulse Response of an instrument's body through partitioned convolution. A Force-Sensing Resistor (FSR) is attached to the phone to hold the pitch of the instrument, as two different kinds of internal digital feedback systems have been implemented to control the damping of the sympathetic strings of the model and the amount of feedback in the partitioned convolution of the impulse response.

The following sections will give a detailed description of each part of the implementation of the system.

A. FDS in Faust

Faust [38] is a high-level, domain-specific, functional programming language, with a strong focus on the development of digital signal processing algorithms for sound and music An intuitive implementation of a FDS algorithm in an imperative language would typically involve representing a discrete grid in a mutable data structure and updating it time step by time step. This approach is not possible in Faust, for the lack of arrays or similar data structures, but recently an `fds.lib` library [24] was added to the official Faust distribution, allowing the implementation of FDS algorithms by employing an approach similar to cellular automata.

1) Cellular Automata: A cellular automaton (CA) is a model of a system of *cell* objects with the following characteristics [39]:

- The cells live on a *grid*
- Each cell has a *state*. The number of state possibilities is typically finite.
- Each cell has a *neighborhood*. This can be defined in any number of ways, but it is typically a list of adjacent cells.

For each cell, a set of cells is defined and called neighbourhood: at each time step t, the next state of a cell is determined by its present state and the state of its neighbours. The rule determining the new state is called transition rule and can be linear or nonlinear. The number of neighbours is defined by a coefficient r, called the neighbourhood radius; this indicates the number of cells at each side of the current cell that are taken into account [24]. It is then straightforward to identify a connection between FDS and CA.

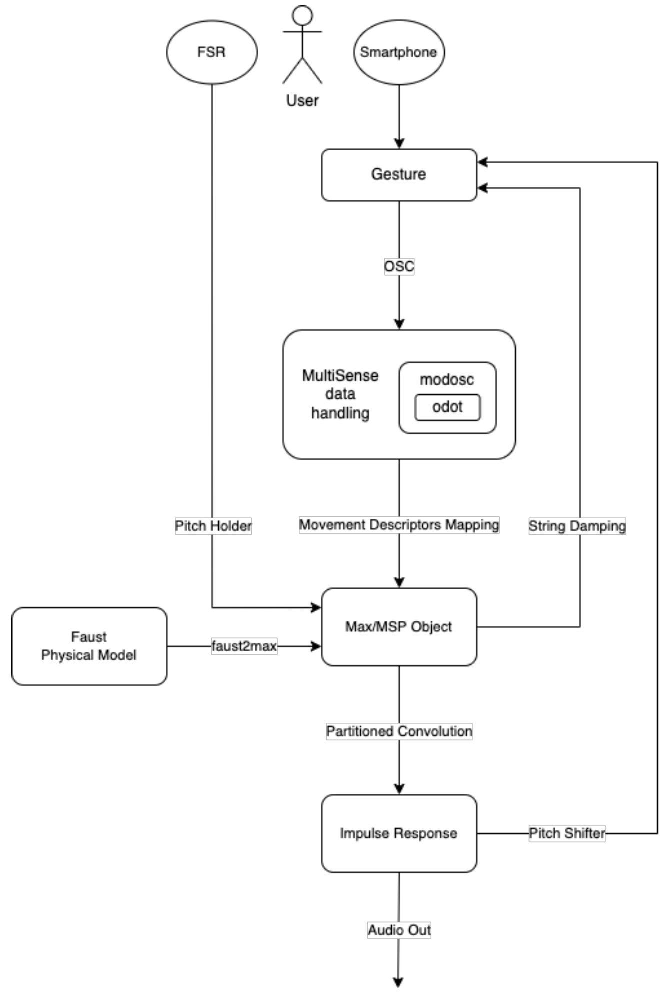

Fig. 7. Flow diagram representing the architecture of the whole system.

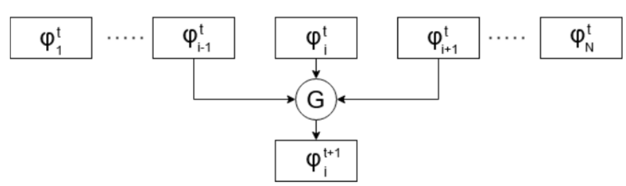

Fig. 8. Scheme of a 1-D CA algorithm with transition rule G

Both FDSs and CA deal with an evolution of state variables on a discrete space-time grid, with the only difference being the fact that cellular automata operate on discrete states, while differential equations look for a continuous domain solution.

However, in computer simulations numerical solutions are always discrete, since the processors' resolution is limited by the number of bits.

2) Coupled FDS in Faust: A detailed explanation of the concepts behind the Faust FDS library can be found in [24] or in the official documentation[1]. In the research presented in this paper, multiple one-dimensional FDS based on the Faust FDS library approach are coupled together. The coupling is achieved by composing the various FDS algorithms in parallel and modifying the Faust FDS library routing to calculate the connection forces, as described in [25].

The basic building block of an FDS algorithm in Faust is a *scheme point*. In a general, uncoupled FDS algorithm, a scheme point calculates u_l^{n+1} as a linear combination of its neighboring grid points in space and time. A scheme point takes as input a force signal (e.g. an excitation), some coefficients and the grid points at time step n, and outputs the value of the grid at u_l^{n+1} as a linear combination of the grid points using the supplied coefficients.

Instead of `model1D`, Sudholt [25] introduced a function `system1D` (see Figure 9) that puts together multiple schemes constructed by `buildScheme1D` and couples them according to supplied coupling information. The individual schemes first calculate the coefficients which are then routed into a function `forceUpdate`; where, for each coupling, the affected grid points are interpolated and the resulting forces are calculated and spread back to the grid points, as shown in Figure 9.

The model proposed in this paper implements three strings coupled to a connected bridge, resulting in 4 different FDS schemes, as described in [25], but adds a selector for the excitation mode of the strings, that can be either plucked or bowed, as shown in Figure 11. The schemes for the stiff strings and the bowed strings are defined in Section II-G, while the scheme for the ideal bar (the bridge) can be obtained from the stiff string by setting the tension to 0. The resulting block diagram is shown in Figure 10. The method works with explicit, one-dimensional FDS algorithms and makes use of rigid connections and linear interpolation.

B. From Faust to Max/MSP

After being implemented in Faust, the model was compiled for Max/MSP [37] using the embedded `faust2max` tool compiler. This feature generates a native standalone application, namely a Max object, a Max patcher and a `js` object that implements the GUI parameters previously described in the Faust code (i.e. faders, buttons, dials). The motivation for this has been given by the authors' decision to control the model using OpenSoundControl (OSC) [36], a data transport specification (an encoding) for realtime message communication among applications and hardware. OSC can be understood as a more flexible alternative to MIDI, as it clears away many of the ideological and hardware constraints inherent to MIDI in favor of a open-ended, user-defined address-space model that provides arbitrary parametric control via standard networking hardware [40] [41].

Due to these characteristics, OSC has already been used in two prior research projects which resulted in Max/MSP libraries for handling OSC data, namely *odot* [42] and *modosc* [35]:

- *odot* is a framework for writing dynamic programs using C-like language inside a host environment such as Max. Compared to more conventional Max objects, it provides access to advanced formatting and parsing of OSC data bundles, allowing for greater control over timing and synchronization of multiple data streams. In addition, it allows the evaluation of functions that would be difficult to implement using standard objects.

- *modosc* is a "set of Max abstractions designed for computing motion descriptors from raw motion capture data in real time. The library contains methods for extracting descriptors useful for expressive movement analysis and sonic interaction design. Moreover, modosc is designed to address the data handling and synchronization issues that often arise when working with complex marker sets. This is achieved by adopting a multi paradigm approach facilitated by odot and OSC to overcome some of the limitations of conventional Max programming, and structure incoming and outgoing data streams in a meaningful and easily accessible manner" [43].

C. OSC Data handling

To send OSC data from a smartphone to a computer running Max/MSP, MultiSense OSC was used. MultiSense OSC is an app which allows the user to send sensor data wireless via OSC protocol from their Android device[2]. It was originally developed to perform head tracking for sound engineers combined with binaural VST plugin. Part of the code is derived from Sebastian O. H. Madgwick, open-source gradient descent angle estimation algorithm [44].

The Attitude And Heading Reference System (AHRS) algorithm combines gyroscope, accelerometer, and magnetometer data into a single measurement of orientation relative to the Earth. The algorithm also supports systems that use only a gyroscope and accelerometer, and systems that use a gyroscope and accelerometer combined with an external source of heading measurement such as GPS. The algorithm, commonly referred to as the Madgwick algorithm, calculates

[1] https://faustlibraries.grame.fr/libs/fds/

[2] https://play.google.com/store/apps/developer?id=MultiSense+OSC

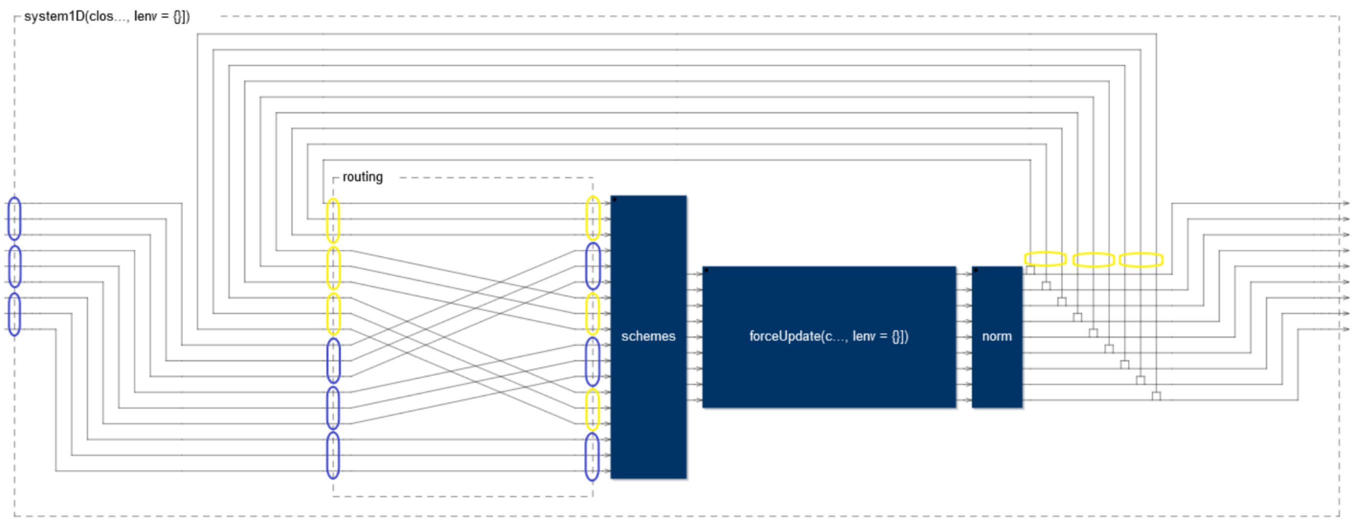

Fig. 9. An example of system1D combining three FDS with three points each into a coupled system, from [25]. The inputs of the entire system are the excitation forces for each point of each system in order, marked in blue. The forces are then routed to be interleaved with the grid points from the previous time step, marked in yellow, so that each individual scheme can receive its proper input. The block labeled here as "schemes" performs the calculation of the coefficients for all schemes. These values are then passed on to add and subtract the appropriate coupling forces in the `forceUpdate` block, and finally passed to the `norm` block, where is finalized the update calculation.

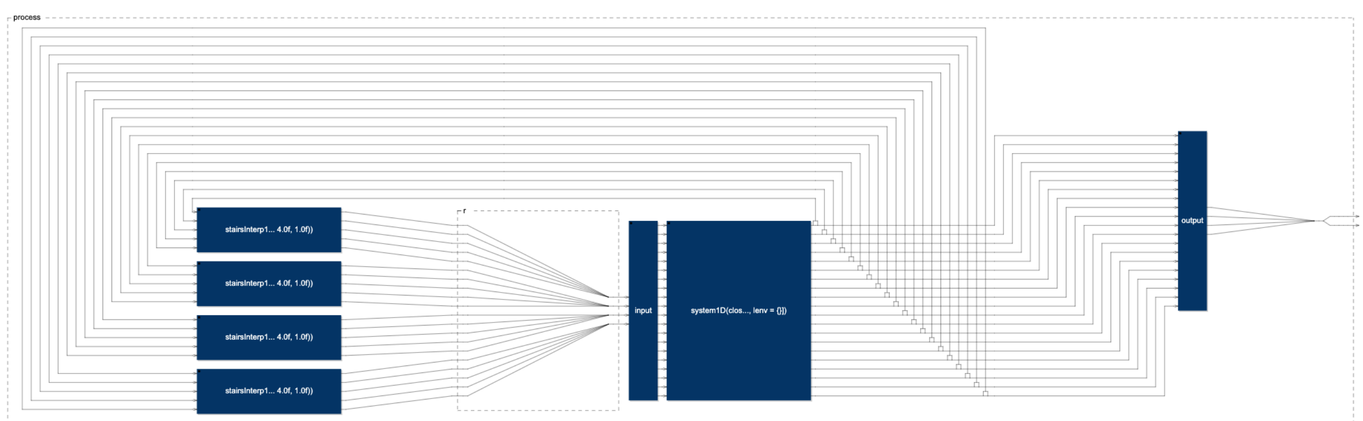

Fig. 10. Faust block diagram for the complete model. For every coupled scheme in the model, a `stairsInterp1D` is defined, to drive the input signals to the correct mesh points. After some routing, defined by the variable r, the `input` block consist of four `linInterp1D` in parallel, individually driven by signals selecting the two different force models (bow and pluck), as described in figure 11. Finally, `system1D` is the block diagram for the coupling of the system, as described in [25]. The feedback loop at the end of it drives back the mesh to the series of `stairsInterp1D` to feed the continuous movement of the bow model. The `output` uses four parallel `linInterp1DOut` to get the output signal from the desired points of the strings.

the orientation as the integration of the gyroscope summed with a feedback term. The feedback term is equal to the error in the current measurement of orientation as determined by the other sensors, multiplied by a gain. The algorithm therefore functions as a complementary filter that combines high-pass filtered gyroscope measurements with low-pass filtered measurements from the other two sensors with a corner frequency determined by the gain. A low gain will 'trust' the gyroscope more and so be more susceptible to drift. A high gain will increase the influence of other sensors and the errors that result from accelerations and magnetic distortions. A gain of zero will ignore the other sensors so that the measurement of orientation is determined by only the gyroscope.

The resulting absolute orientation (quaternion) is then formatted as OSC data and sent to Max/MSP through the User Datagram Protocol (UDP). The requirement for this procedure to succeed is that both the smartphone and the PC must be connected to the same WiFi network, and the smartphone must know the IPv4 address of the PC. After matching the destination port of the smartphone with the source port of the PC, Max/MSP can receive the data packets through the object `udpreceive`, with the number of the selected port as an

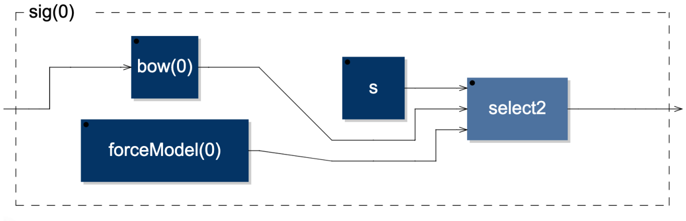

Fig. 11. Faust block diagram for the exciter selector. The `bow` block implements a nonlinear friction based interaction model that induces Helmholtz motion, as described in [20], while the `forceModel` takes care of the plucking. A `select2` is used to change the state of the excitation mode and drive the `linInterp1D` into the `system1D`.

argument. The addresses contained in OSC packets are then matched to extract raw values using the `o.route` function, an odot function that dispatches OSC messages according to an address hierarchy, stripping off the portion of the address that matched. The resulting data are packed, using the `o.pack` object and, using modosc syntax domain, converted with an odot `o.expr.codebox` to a modosc point of data (the phone). A point in modosc consists in data bound to an OSC address. Points can be then collected in groups, as points and groups are the two main data types on which modosc abstractions operate [35].

D. Implementation of Movement Descriptors

As previously stated in Section III-B, modosc was initially designed to work with motion capture (MoCap) data [35] [43]. Visi et al. [45] deepened the research on this field, using different kind of data, taken also from Inertial Measurement Units (IMU). This work culminated in the release of GIMLeT [46], a set of Max patches based on odot and modosc, incorporating neural network, gesture following through PoseNet [47] [48] and handling of OSC data with TouchOSC [49]. Unfortunatelly, TouchOSC is only able to send accelerometer data from the smartphone, thus limiting the number of movement descriptors which could be calculated (see Section II-K). This is why we decided to use MultiSense OSC and its algorithm to gather also data from gyroscope and magnetometer, resulting in orientation values. As shown in Figure 12, we were able to adapt the orientation-based data extracted from the smartphone to the modosc syntax using a o.expr.codebox from the odot library.

As stated by Visi et. al in [10], "The data obtained from IMUs are morphologically very different from positional data returned by optical MoCap, since calculating absolute position from IMU data in real time is technically very difficult if not outright unfeasible, as the operation would require double integration of acceleration data". Nevertheless, since movement descriptors are used instead of raw data, the extracted features can be adapted to work on a cognitive level towards a different meaning of the conveyed movements of the subject. As also presented in [10], movement descriptors most commonly used with positional data can be adapted to work with IMU data.

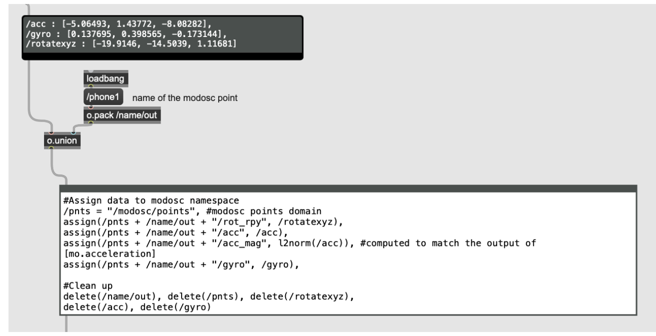

Fig. 12. The data coming from the smartphone's accelerometer, gyroscope and orientation (rotatexyz) are adapted through an odot o.expr.codebox to the modosc syntax, generating a modosc point named /phone1.

Our gestures data are then based on orientation, rather than on position. Furthermore, since we deal with one smartphone, every movement descriptor is calculated as an instantaneous value in the specific moment in time when it's computed. While this is enough for the implementation of most of modosc movement descriptors (such as velocity, acceleration, jerkiness and fluidity), quantity of motion requires the definition of a group of points. We then defined a group of points consisting of all the individual phone movements through time, in order to calculate the quantity of motion as the weighted sum of the speeds of every single point (see Section II-K4). As shown in Figure 13, the selected descriptors are then gathered to be later mapped to the physical model parameters.

In order to compute the periodic quantity of motion and the time and flow effort descriptors, as presented in Section II-K, the computed values over time of the single point movements have been calculated through list operators in Max/MSP, as displayed in Figure 14. Furthermore, a compensation of gravity acceleration has been performed on the z axis.

E. Partitioned Convolution and Pitch Shifter

As it is mentioned above, the resonator has been implemented using a partitioned convolution reverberator. For the real time implementation in Max/MSP we used the HISS library [50], with the `multiconvolve~` object as the main component of this section (see Figure 15). In the feedback path, a pitch shifter has been added to create a wider range of sonic variations, while two different gain sliders are used to control the amount of feedback. In the main Max/MSP patch, two different IRs of violin bodies have been used, with the possibility of switching between them in real time. Additionally, very interesting sonic results have been given by the pitch shifter, with a pseudo-random rate of evolving and changing parameters.

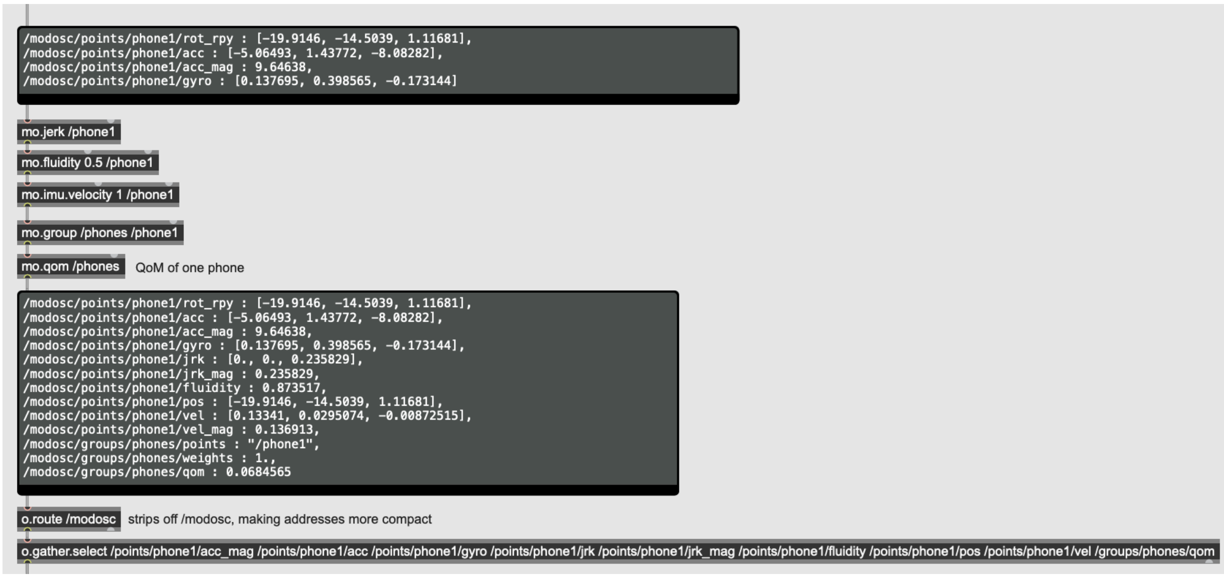

Fig. 13. Movement Descriptors computed on the adapted IMU data using the modosc functions. In the last line of code, the selected descriptors for acceleration magnitude, acceleration, gyroscope, jerkiness, jerkiness magnitude, fluidity, position (orientation), velocity and quantity of motion are selected and gathered to be later mapped to the physical model parameters.

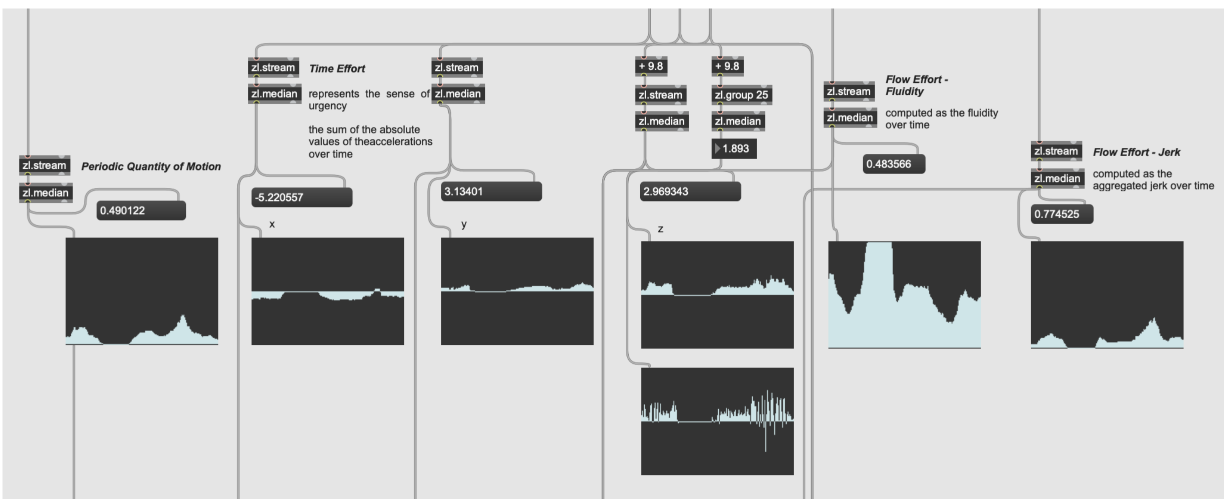

Fig. 14. Curves showing the computed movement descriptors through time. From left to right: periodic quantity of motion, time effort on the x axis, time effort on the y axis, time effort on the z axis, flow effort of fluidity, flow effort of jerkiness. On the second raw, under the time effort on the z axis, impulsive movement on the z axis have been calculated, to control the plucking of the string in the physical model. On the same axis, a compensation of gravity acceleration has been performed.

F. Force-Sensing Resistor

A Force-Sensing Resistor (FSR) sensor was used as a pitch holding mechanism. For the implementation, the FSR was connected to an analog input of an Arduino Uno board. A simple script was used to send the data from the Arduino to Max/MSP in real time, transferring the analog data through a serial port inside Max/MSP with the object `[serial]`. Additionally, a threshold was set to activate and deactivate the pitch holding control.

G. Mapping

Among the reasons for the creation of new instruments are the real-time control of new sound-worlds, and the control of existing timbres through alternative interfaces to enable individuals in the spontaneous creation of music [51]. The term *mapping* is used widely to indicate the mathematical process of relating the elements of one data set onto another. In computer music, mapping is often used in relation to algorithmic composition, where a parameter with a particular set of values is scaled or transformed so that it can be used to control another parameter [52]. As defined by Hunt and

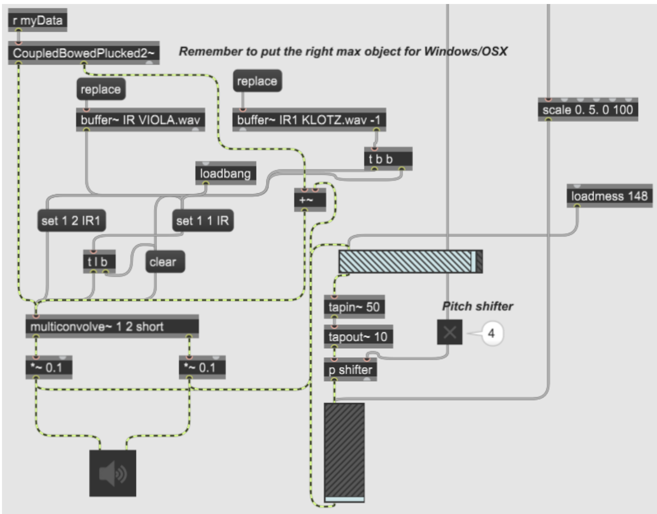

Fig. 15. Real time implementation of the feedback partitioned convolution in Max/MSP .

Wanderley in [53], in this paper we consider mapping as the act of taking real-time performance data from an input device and using it to control the parameters of a synthesis engine, as shown in Figure 16.

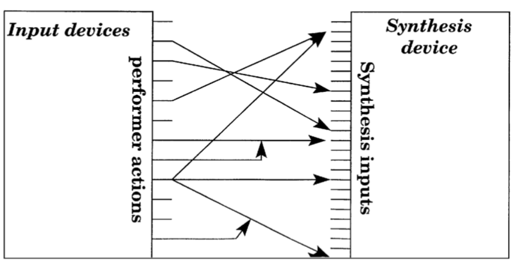

Fig. 16. A representation of the mapping layer (arrows) between performance data and the input variables of a synthesis engine, from [53].

1) Different ways of mapping: The main question to be solved was related to the actual choice of which mapping strategy to implement. The ultimate goal in designing new Digital Musical Instruments (DMI) is to be able to obtain similar levels of control subtlety as those available in acoustic instruments, but at the same time extrapolating the capabilities of existing instruments [54]. Considering mapping as part of an instrument, two main directions could be deduced from the analysis of the existing literature:

- The use of generative mechanisms, such as neural networks.
- The use of explicitly defined mapping strategies.

Although we initially considered a possible mapping with a neural network, in the end our decision was to use explicit mapping strategies, presenting the advantage of keeping the designer in control of the implementation of each of the instrument's component parts, therefore providing an understanding of the effectiveness of mapping choices in each context [53].

2) Explicit mapping strategies: The available literature generally considers mapping of performer actions to sound synthesis parameters as a *few-to-many relationship* [55]. Considering two general sets of parameters, three intuitive strategies relating the parameters of one set to the other can be devised as [53]:

- *one-to-one*, where one synthesis parameter is driven by one performance parameter,
- *one-to-many*, where one performance parameter may influence several synthesis parameters at the same time, and
- *many-to-one*, where one synthesis parameter is driven by two or more performance parameters.

Concerning explicit mappings between two sets of parameters, many ways of abstraction of the performance parameters have been proposed, from perceptual parameters [56] to focusing on continuous parameter changes represented by gestures produced by the user [54] [57].

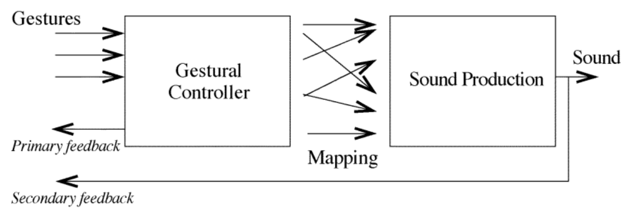

Fig. 17. A symbolic representation of a Digital Musical Instrument (DMI), from [54].

3) Three-layer mapping: In designing an explicit mapping for our system, we decided to adopt a three-layer mapping model. In this model, the first layer is interface-specific, since it converts the incoming sensors information into a set of chosen (intermediate or abstract) parameters that could be perceptually relevant or derived from other forms of interaction, like gesture. These are then mapped – in a second independent mapping layer – onto the specific controls needed for a particular synthesis engine. The advantages of this model is that the first mapping layer is a function of the given input device and the chosen abstract parameters, while the synthesis engine could be changed by changing the second mapping layer, e.g. from physical modelling to FM synthesis or additive synthesis, since the second mapping layer is not dependent on the control parameters directly. The proposed mapping layers (see Figure 18) are thus:

- Extraction of meaningful performance parameters.
- Connection of performer's (meaningful) parameters to some intermediate representation set of parameters (for instance, perceptual or abstract).

- Decoding of intermediate parameters into system-specific controls.

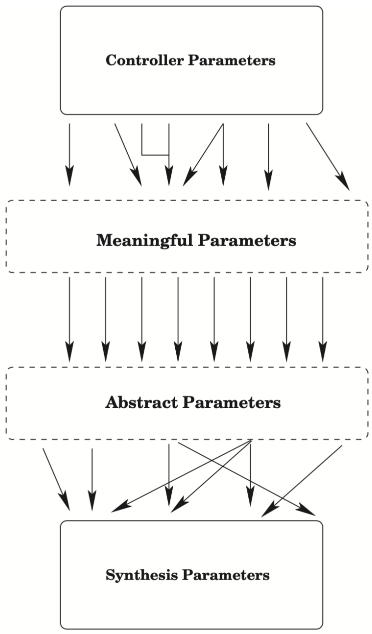

Fig. 18. Three-layer mapping model, as presented in [58].

Specifically, as shown in the full mapping model presented in Figure 19, the first mapping layer extracts meaningful parameters from the control gesture variables of the smartphones and computes the Movement Descriptors, as described in Sections II-K and III-D. The second mapping layer then connects the abstract parameters derived by the first layer to the physical model's synthesis parameters, using the following strategies:

- Bow Velocity: Y axis acceleration (one-to-one and one-to-many)
- Bow/Pluck Selector: boolean comparison between Y axis acceleration and Z axis acceleration (many-to-one)
- Plucking Force: Z axis acceleration (one-to-one and one-to-many)
- Bow Force: Quantity of Motion (one-to-many)
- Amount of Feedback: Quantity of Motion (one-to-many)
- Damping of sympathetic strings: Quantity of Motion (one-to-many)
- Bow Friction: Jerk (one-to-many)
- Pitch: Jerk (one-to-many)
- Pitch Holder: FSR (one-to-one direct mapping)

IV. Evaluation

A. Participants

For the experiment, 10 people (mean age 28, 7 males, 3 females), were recruited on a voluntary basis. All were students from Aalborg University, half of which studying MSc in Sound and Music Computing and the rest Lighting Design, Biomedicine and Medialogy. The participants were presented with the goal of the experiment and a written consent was obtained prior to each participation. The experiment was conducted at Aalborg University (campus of Copenhagen), in the Augmented Performance Lab.

B. Set up

The technical set up consisted of two laptops, two studio monitors, an Android phone with a FSR sensor attached to it, and a camera used for video recording. One laptop was used for running the model in Max/Msp while receiving and recording data from the phone. The other laptop was used for obtaining personal data and questionnaire data from the participants.

C. Procedure

The experiment was carried out individually for each participant and constituted of tree phases. Prior to the start, each person was presented with a personal survey aimed of assessing: name, age, gender, if they have any background in dance and music, preference over musical instruments and personal patterns of mobile phone usage (as shown in Figure 20). The first phase consisted of a short introduction to the system and what is required from the user. They were given an overall explanation on the movements that they are encouraged to explore, while still leaving them freedom around their way of interacting with the system. The choice of conducting the experiment in a large and comfortable space was consciously made so that the individual could have the feeling of taking part in a sound installation and thus provide him/her with the correct experience of fully indulging one-self in the interaction. During the second phase, participants were given 5 minutes of interaction with the system in which they explored different ways of playing the model through movements made with the phone. The third and last phase constituted of presenting the participants with a series of customized questions based on The Questionnaire for User Interaction Satisfaction (QUIS) developed to assess a user's subjective satisfaction with a human-computer interaction system [59] and Likert 7-point scale system offering 7 different answer options (with a neutral midpoint) related to an agreement that would be distinct enough for the respondents, without throwing them into confusion [60]. The questionnaire was developed with a focus on assessing the user satisfaction of the interaction with the system, and evaluating responsiveness, intuitiveness, level of engagement and utility/practical value. Once completed, a series of open oral questions were asked and recorded in audio .m4a format using an iPhone. Participants were presented with a chance to give an overall feedback and elaborate on their opinions regarding the system, explaining their personal preferences and the performance outcome.

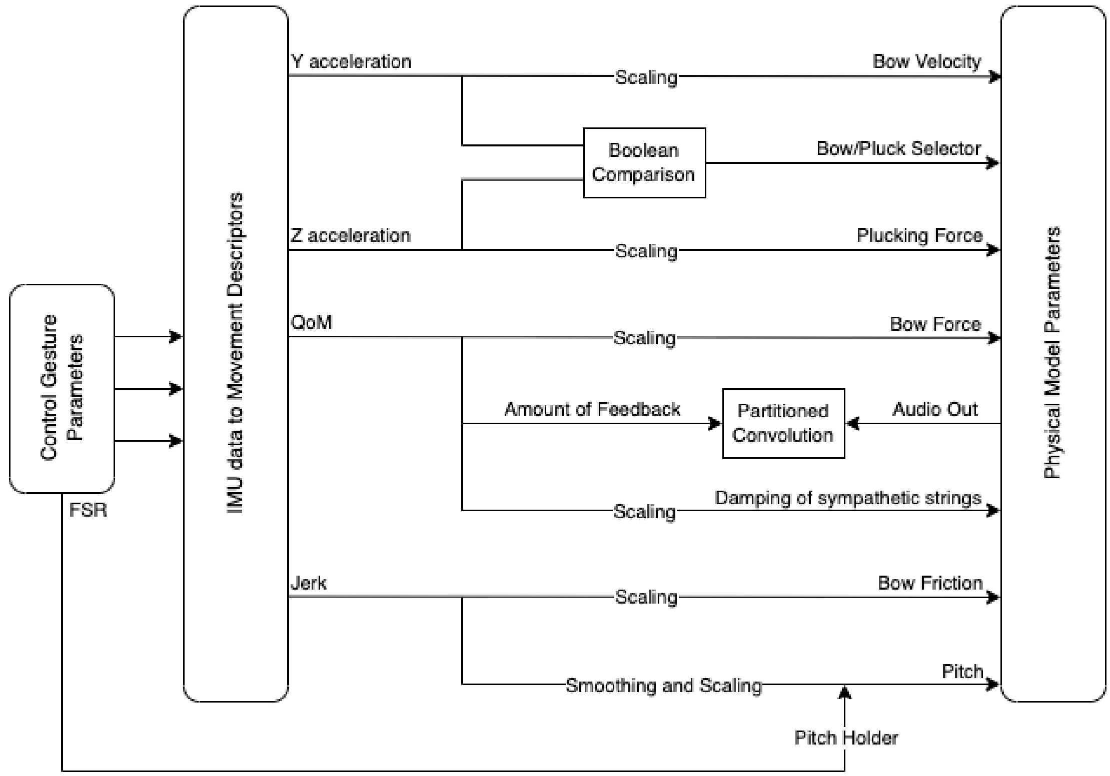

Fig. 19. Three-layer mapping model implemented in this paper. The control gesture variables from the smartphone were mapped to a first layer, extracting meaningful parameters and computing Movement Descriptors. The abstract parameters derived by the descriptors were mapped to the synthesis parameters of the physical model. The FSR were directly mapped to a switcher to hold a desired pitch.

D. Data

1) Sources: The data was collected through qualitative and quantitative methods listed below:

- Sampling of Sensor Data
- Questionnaire / Evaluation sheet
- Interview
- Survey
- Video

2) Analysis: The following methods were used for the analysis and interpretation of the data. A data recording environment was created in Max/MSP, receiving sampled data from the user movement descriptors mapped to the model parameters. Flow, effort, fluidity, and jerkiness, Periodic Quantity of Motion, FSR, Time effort on the x,y,z axis were all respectfully plotted onto a Cartesian Plane to display the oscillatory motion of each interactive component over time (x-axis: time; y-axis: data range). The time indexes and the data changes over time were stored into two column vectors and plotted using MATLAB's 'plot' function. A second order Butterworth IIR Lowpass filter with a cutoff frequency of 0.75 Hz was implemented to reduce the noise of the data. Successively, a cross comparison between the various performances was conducted to assess the most active participant. Furthermore, the plots were compared to the profile of each participant in order to analyse the differences between each participant and their own with the actual relative duration of interaction observed during the data analysis stage.

After the completion of the interaction, the answers of each participant's evaluation sheet and interview were used to derive the mean ratings across all experiments. The results were later organized through stacked bar chart and used to assess the viability of the system and served as a basis for future improvements.

The information gathered from the survey was later compared to the data plots in order to establish a connection between the rate of interaction and the quality of the movements.

The considerations made during the observation of the videos served as a source of information regarding the participant's physical engagement with the system and the quality of their movements.

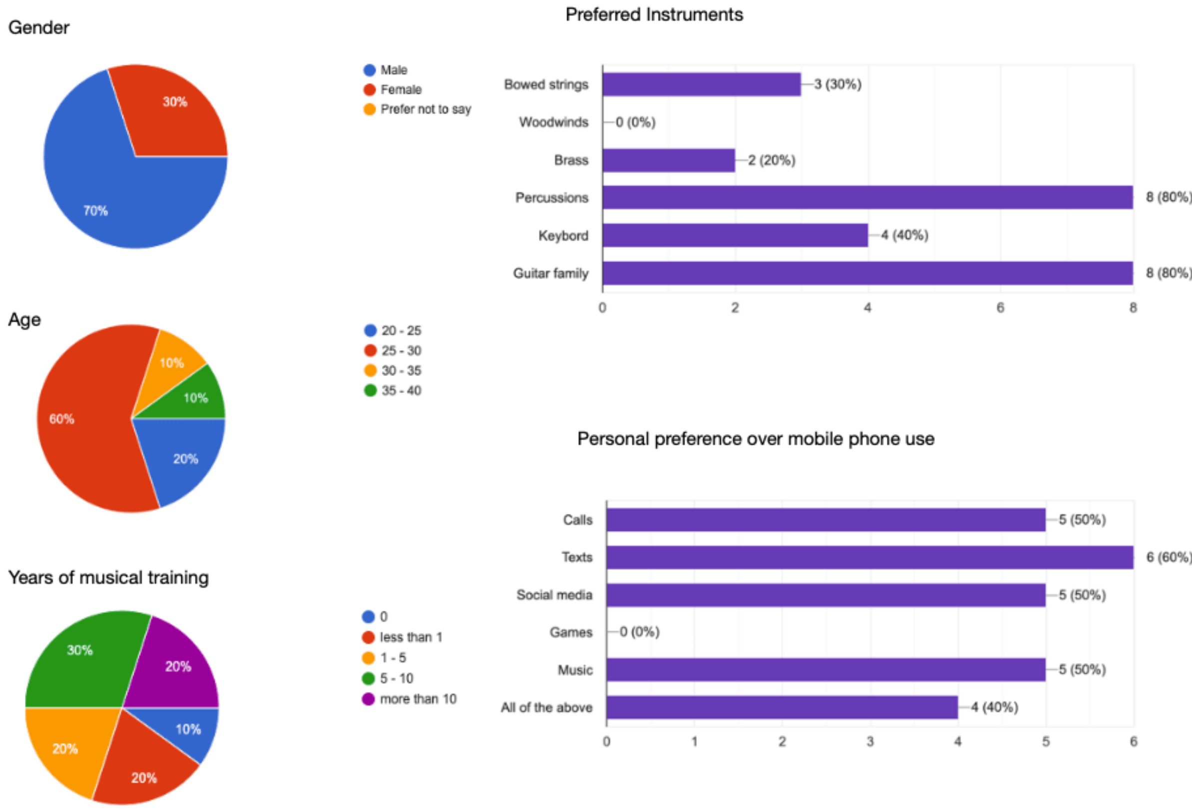

Fig. 20. Display of overall information given by participants through a personal survey.

V. Results

A. *Participant's Subjective Impressions derived from the questionnaire and spoken interview*

Subjective opinions of each participant were assessed through a questionnaire and a brief interview on the interaction with the system was completed. What follows is an overall summary of the answers given by each participant, answering through a 7 point Likert scale on customized questionnaire based on QUIS. For detailed information on the results, see Figure 21.

- All participants stated that before their interaction the overall impression was to expect new music experience from the installation,
- 60 % agreed that their first impression of the system was positive, while the rest 40 percent stayed neutral on that statement (mean rating 5),
- 7 out of 10 participants were satisfied with their experience with the system (mean rating 5),
- The majority of the participants stated that understanding the interaction with the system predisposed slight difficulties for them. (mean rating 3.6),
- 30 % stated they disagree with the statement that they felt in control of the sound output, 40 percent agreed and the rest stayed neutral (mean rating 4.3),
- Half of the participants agreed that the level of interaction is fairly high (mean rating 5),
- The majority acknowledged that their movements made sense to the sounds they were triggering (mean rating 4.8),
- Most of them also stated they were aware of the changes their motions were making (mean rating 5.5),
- 60 % thought that the sounds they experienced were pleasing (mean rating 4.6),
- 80 % understood the type of sounds related to certain movements (mean rating 4.9),
- 60 % agreed that the movements are well mapped to the parameters of the model (mean rating 4.2),
- 90 % agreed that the system is highly engaging (mean rating 5.8) and they were very pleased with their performance (mean rating 5.6),
- 80 % confirmed that their interest was maintained throughout the whole interaction (mean rating 5.6).
- 90 % agreed that the installation can have future contribution to the field of art, design and technology (mean rating 5.3).

When asked about the overall impression of the interaction, one participant stated that he was presented with an "exploratory experience" in which he was "engaged by the way of interaction by movements that trigger sounds", and that he was "trying to visualise the instrument, and associating movement with it like plucking and bowing was the most obvious interaction". Another one appreciated the movements that required more energy since the sounds were becoming more complex, thus bringing him a reward-like sensation. Several participants experienced delay between the movements and the sounds they were supposed to trigger, which they stated predisposed frustration and confusion.

Almost all of the participants were not satisfied with the FSR sensor and proved by the data recording plots several participants choose to ignore its use. Some also complained about the connection of the sensor to the laptop, saying it pushed for distraction and did not allow them to move properly. When asked about their preference on the device, all of them stated that in theory it is an interesting choice since it augments something that we use every day, but in practice they would prefer another object that is more ergonomic, safe to use and because of the phone's fragile nature, they would feel more comfortable having something that would not break easily.

For a full transcribed version of the interview questions and participant's responses, refer to Appendix.

B. Recordings of Motion Data

Several considerations can be made by observing the data displaying the relative duration of interaction. After thorough assessment of all participants data recordings, we isolated Participant number 6 data as being highly active across all control gesture parameters and compare it with Participant number 4, who was was significantly less active (see Figure 22).

An interesting observation can be made between the oscillatory motion of the jerk index, pressure index relating to the FSR sensor and the pitch plot in Figure 22. It is clear that participant number 6 experienced fast paced movements (jerk index), leading to frequent interaction with the pitch shifter which led him to recurring activity towards the pitch holder (FSR). On the contrary, Participant number 4 has been holding a high and steady fluidity index almost throughout the full interaction, keeping him away from jerky movements, resulting in a steady position of the frequency and a minimal interaction with the FSR sensor. We can also observe that halfway through his interaction (150 seconds), Participant number 6 reached high peaks of Periodic Quantity of Motion, resulting in strong levels of feedback in the system, given by both the damping of the strings and the direct amount of feedback mapped to the partitioned convolution of the IR.

C. Additional observations derived from video recordings

A critical analysis of the videos was conducted, based on the observation of the level of activity of each of the participants, the quality of the movements, the level of engagement with the system and novel, creative approaches to the interaction. In particular, we observed that 7 out of 10 people experienced a very active and engaging interaction, with full exploration of different paced movements, while the rest adopted more rigid and "robotic" motion, depicting low level of engagement.

VI. Discussion

The goal of this project was to design a new feedback instrument, based on physical modelling synthesis and played through gesture interaction using a smartphone. The resulted work was evaluated in the context of a sound installation, as reported in section V. In this section we will go through all the steps of the research, discussing about the effectiveness of our choices and feasible alternatives, as well as possibilities for future research.

A. Implementation

While the physical modelling techniques proposed in this research have been widely analysed by the authors, ideas for creating Impulse Responses remain unexplored in the project. As better IRs are needed to enhance the sonic outcome of the application, future work can consider either the recording of IRs from wooden instruments or the implementation of resonators with FDS or other physical modelling techniques. An interesting approach could be the creation of resonators of different sizes and shapes using the FDS approach, creating a control that could fade between the different IRs. This would result in combinations of exciters and resonators that would not be possible in the physical domain. A similar idea and implementation can be found in [61].

B. Mapping

In thinking about the specific mapping for this project, it became immediately clear that mapping constitutes a significant part of the final instrument itself, especially when dealing with complex solutions and multiple ways of coupling gestures and data. The work showed that complex mappings are more effective at conveying a good performance from a human player when multi-parametric tasks are required, than a series of one-to-one mappings. Furthermore, the usage of mid-level descriptors introduced more useful abstractions on the cognitive level, letting the user interacting with the interface on with complex gestures. Nevertheless, not all the computed descriptors have been used in the mapping (i.e. fluidity), as we soon noticed that too many parameters could result in a not good responsive application. This confirms the

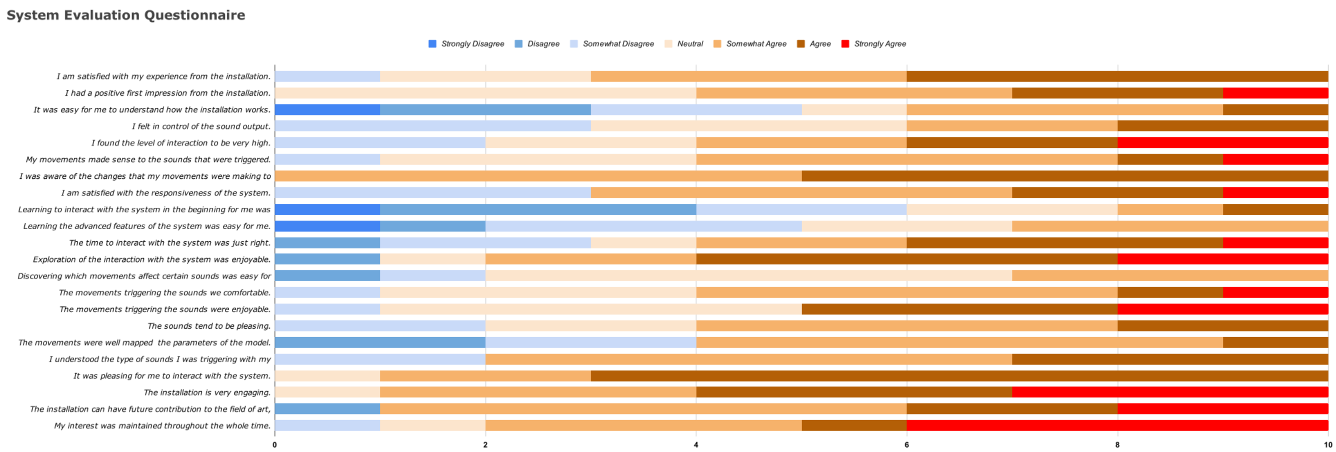

Fig. 21. Stacked bar chart based on 7 point Likert scale type answers (strongly disagree to strongly agree with a neutral midpoint) to customized system evaluation questionnaire (vertical axis) based on QUIS. It represents the averaged answers across 10 participants (horizontal axis)

importance of mapping in the final sounding and interacting result, leaving open more possibilities and alternatives for further investigations.

Concerning the usage of low and mid level descriptors to control a sound synthesis, we came across interesting conclusions, such as the duality of the hierarchical relationship between sound and movement. It can be concluded that when the mid-level descriptors are used, after the initial excitation given by the first gesture, in the interaction context the movement can be considered as representing a higher level of hierarchy, driving the resulting sonic outcome. Here, a feedback perceptive loop between the user and the system enhances the possibilities of the system itself, resulting in an unpredictable and evolving sonification of gestures. This relation can be altered when low-level descriptors are used. In this case, the narrow sound vocabulary derived by the lower gesture possibilities would dictate the results of the movements, thus reversing the hierarchical structure of interaction. This system would in fact be very predictable, with every gesture corresponding to the same sound outcome and no active perception feedback loop between the sounds and the gestures.

C. Data and questionnaire

The purpose of the experimental evaluation was the assessment towards the responsiveness and intuitiveness of the interaction through a questionnaire and a spoken interview. From a conceptual and interactive perspective, the questionnaire was a particularly useful source of subjective opinions. It was also of great importance to understand each participant's level of satisfaction towards the 'movement to sound' mapping. During the oral interview, the participants were given a chance to elaborate on their opinions on the interaction with the system and thus provide analogy on the temporary unity that mediates the interaction both in the real and virtual world, by the fusion between the performer and a DMI [62]. In practice the experimental evaluation was proved to work predominantly as intended amongst multiple individuals. One of the main concerns was towards the interaction with the system. While some people, concretely those with more musical experience, felt comfortable with discovering and exploring the movements and sounds on their own, the participants having less knowledge towards instruments and music in general were having trouble getting at ease with their movements and needed further guidance and more time of interaction to explore the advanced features of the model. There were also slight disagreements on the feeling of control regarding the sound output. By observing the mean rating and percentage of negative, positive and neutral responses, we arrive at inconclusive results which show that further experiments need to be done in order to rate the system as responsive and liable. There were some remarks made on the delay between gestures and sound and some participants considered that this is a result of the device being not responsive enough, others stated that it is a question of mapping more gestures to the parameters of the model. This led us to the conclusion that there could be a problem with the UDP packets transmission and further attention and development is required towards both factors.

However, high ratings were given on acknowledging mapping choices and high activity of oscillatory motion regarding movement on the Y (bow velocity) and Z axis (plucking force), proving that the extracted motion descriptors were intuitively mapped to the parameters of the model. A proposition made by one participant about having a visual reference about the movements was noted , but since at this stage of the prototype we decided to focus explicitly on the movements and the device, this consideration was left for future additions. Although the decision of adding an FSR allowing the users to hold a pitch during the performance seemed as an intuitive

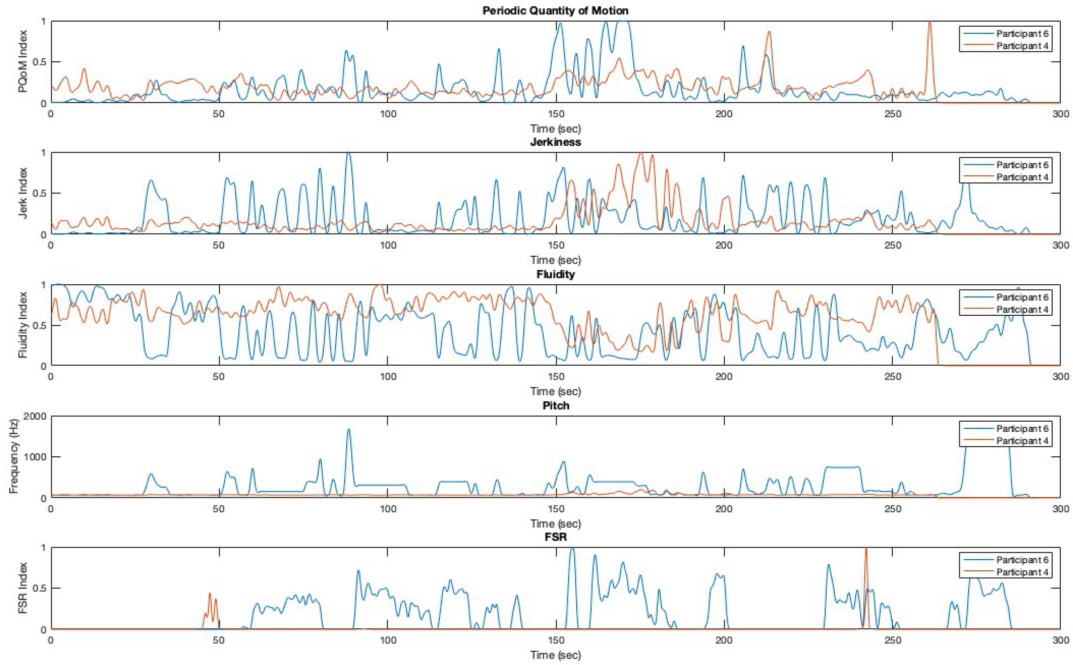

Fig. 22. Oscillatory motion of PQoM Index, Jerk Index, Fluidity Index, Frequency and FSR Index over time during Participant numer 4 and 6's interaction. The plot highlights differences in activity and personal preference on the use of the FSR sensor.

choice, it was reported by the majority of participants that its use was not comfortable. This implies that another approach is needed for providing the participants with this option. A notable result can be observed by coupling Participant 6 data analysis and response to the first question of the interview (see Appendix), referencing his feeling of reward towards the effort of the gestures performed during the evaluation. He managed to reach high levels of feedback in the system through peaks of Periodic Quantity of Motion and Jerkiness (see Figure 22). His interaction revealed that the chosen mapping and the implementation of the movement descriptors could provide a system where higher level descriptors are not directly related to a low-level gesture (as previously stated in Section VI-B). On the contrary, lower lever gestures contribute to build higher lever descriptors through time and make sense of them within independent layers of motion interaction. This can be a validation towards some of the notices given by other participants on the fact that they were trying to repeat certain movements and not obtain always the same results. These proves that multiple layers of interaction were involved, resulting in different values of movement descriptors and a wide range of sonic possibilities.

A lot of the participants suggested that the installation would be more interesting if it was presented in a social context, meaning more phones and more people interacting between each other. This was actually considered but due to technical problems in transmitting data through modosc using multiple devices, this proposition was left to further future exploration.

D. Future research

Despite the initial idea of using a smartphone, results from the evaluation highlighted the evidence that this could not be the best solution for this application, in terms of affordance. Thus, more exploration is needed for the Human-Computer Interaction (HCI) of the model. Different combinations of mappings can be easily implemented due to the architecture of the patch. A problem that we came across was the lack of affordance for the interaction part. It was observed that many participants were biased about how they could interact with the smartphone, and they were locked to specific movements. Part of participants reported that they were afraid to explore gestures using a smartphone because they didn't want to break it. With that being said, different objects with embedded IMUs are needed, with a design focused towards higher orders of affordance. As stated by Visi et al. in [10], IMUs provide data morphologically different from motion capture analysis. However, it is possible to obtain similar meaningful information if the data are interpreted correctly. That's why a further investigation could involve the use of wearable and multiple IMUs, to focus deeper on the possibilities of the movement descriptors codification. Multiple IMUs controlled by multiple users could potentially culminate in a participatory sensemaking action, where social feedback (introduced as the shared responsibility of interacting together with other users towards a common sonification of multiple people's gestures) can be investigated. This can be related also the the FSR which, as stated before, resulted in not being the optimal solution for pitch holding. Dealing with multiple IMUs and multiple users could allow a complex mapping with parameters spread along the whole range of available IMUs, avoiding the usage of external sensors, like in this specific case.

VII. Conclusions

This work aimed at designing a new instrument, using gesture tracking from a smartphone to control a digital feedback physical model, in the context of a sound installation. First, a model for a string instrument, with strings coupled to a bridge has been implemented using FDS synthesis in Faust, providing two kinds of excitation, namely plucking and bowing. The model has then been ported to the Max/MSP environment, where we used the OSC protocol to transfer and handle gesture data from a smartphone to a PC. Then, computation of mid-level movement descriptors have been performed to map the raw data from the smartphone IMUs to the physical model parameters. An evaluation consisting of both qualitative and quantitative methods, in the form of a sound installation, gave us a significant range of considerations, previously reported in Section VI. The hypothesis presented at the beginning of our research has been fulfilled with encouraging results, although ground for future improvement of this project have been highlighted by the analysis of the evaluation. The present study contributes then to a relatively new interdisciplinary research field, covering the areas of physical modelling, NIME and embodied interaction and embodied music cognition. It is hoped that the research done during this study can work as a foundation for future investigations in these fields and could create new engaging musical experiences, resulting in new perspectives in musical expression.

VIII. Acknowledgement

We would like to thank Dan Overholt for his supervision and his guidance throughout the whole project and for his involvement in the Feedback Musicianship Network. Thanks to Silvin Willemsen and Marius Onofrei for their great support and their hints for the realization of physical models. Also, we are grateful for the support that we received from Riccardo Russo and David Südholt regarding the Faust implementation and Fedrico Visi for his precious work in the field of embodied interaction and gesture tracking. We would also like to acknowledge all the volunteers that took part in the evaluation of TRIPHON. This project could not have been realized without the support of the above mentioned people.

References

[1] C. P. Boner and C. Boner, "Behavior of sound system response immediately below feedback," 1966. [Online]. Available: https://www.aes.org/e-lib/online/browse.cfm?elib=1149

[2] D. Sanfilippo and A. Valle, "Feedback systems: An analytical framework," *Computer Music Journal*, vol. 37, pp. 12–27, 2013.

[3] "Feedback musicianship network." [Online]. Available: https://feedback-musicianship.pubpub.org/

[4] D. Overholt, A. Eldridge, and C. Kiefer, "Feedback Musicianship Workshop @ NIME 2020," *Feedback Musicianship Network*, 10 2020, https://feedback-musicianship.pubpub.org/pub/kefjnk2y. [Online]. Available: https://feedback-musicianship.pubpub.org/pub/kefjnk2y

[5] C. Kiefer, D. Overholt, and A. Eldridge, "Shaping the behaviour of feedback instruments with complexity-controlled gain dynamics." [Online]. Available: https://github.com/chriskiefer/

[6] A. Eldridge, C. Kiefer, D. Overholt, and H. Ulfarsson, "Self-resonating vibrotactile feedback instruments ||: Making, playing, conceptualising :||," *Feedback Musicianship Network*, 4 2021, https://feedback-musicianship.pubpub.org/pub/kl8m5o5y. [Online]. Available: https://feedback-musicianship.pubpub.org/pub/kl8m5o5y

[7] C. Kiefer, "What's happening in the world of feedback instruments?" [Online]. Available: https://www.emutelab.org/blog/feedback_insts

[8] "Echo #3." [Online]. Available: https://echo.orpheusinstituut.be/issue/3-feedback

[9] F. Visi, R. Schramm, and E. Miranda, "Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies," *ACM International Conference Proceeding Series*, pp. 100–105, 2014.

[10] F. Visi, E. Coorevits, R. Schramm, and E. R. Miranda, "Musical instruments, body movement, space, and motion data: Music as an emergent multimodal choreography," *Human Technology*, vol. 13, pp. 58–81, 2017.

[11] J. van Dijk, "Designing for embodied being-in-the-world: A critical analysis of the concept of embodiment in the design of hybrids," *Multimodal Technologies and Interaction 2018, Vol. 2, Page 7*, vol. 2, p. 7, 2 2018. [Online]. Available: https://www.mdpi.com/2414-4088/2/1/7/htm https://www.mdpi.com/2414-4088/2/1/7

[12] A. P. Melbye and H. Úlfarsson, "Augmenting the Double Bass." [Online]. Available: https://www.adampultz.com/augmenting-the-double-bass/

[13] J. Zhang, "Music 421a: Audio applications of the fft-final project report-spring 2018 "shimmer" audio effect: A harmonic reverberator."

[14] V. Valimaki and T. Takala, "Virtual musical instruments - natural sound using physical models," *Organised Sound*, vol. 1, p. 75, 1996.

[15] G. D. Poli and D. Rocchesso, "Physically based sound modelling," *Organised Sound*, vol. 3, pp. 61–76, 4 1998.

[16] J. O. Smith, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, p. 74, 24 1992.

[17] C. Cadoz, A. Luciani, and J. L. Florens, "Cordis-anima. a modeling and simulation system for sound and image synthesis. the general formalism," *Computer Music Journal*, vol. 17, pp. 19–29, 1993.

[18] P. R. Cook, "Real sound synthesis for interactive applications."

[19] J. M. Adrien, "The missing link: modal synthesis — representations of musical signals." [Online]. Available: https://dl.acm.org/doi/10.5555/131150.131158

[20] S. D. Bilbao, *Numerical sound synthesis : finite difference schemes and simulation in musical acoustics*. Wiley Publishing, 2009.

[21] S. Bilbao, C. Desvages, M. Ducceschi, B. Hamilton, R. Harrison-Harsley, A. Torin, and C. Webb, "Physical modeling, algorithms, and sound synthesis: The ness project," vol. 43, pp. 15–30, 2019. [Online]. Available: www.ness-music.eu.

[22] M. V. Mathews, "The digital computer as a musical instrument," *Science*, vol. 142, pp. 553–557, 1963.

[23] S. Willemsen, *The Emulated Ensemble, Real-Time Simulation of Musical Instruments using Finite-Difference Time-Domain Methods*. Aalborg Universitet, 2021.

[24] R. Russo, S. Serafin, R. Michon, Y. Orlarey, and S. Letz, "Introducing finite difference schemes synthesis in Faust: A cellular automata approach," 2021.

[25] D. Südholt, R. Russo, and S. Serafin, "A faust implementation of coupled finite difference schemes," 2021.

[26] E. Armelloni, C. Giottoli, and A. Farina, "Implementation of real-time partitioned convolution on a dsp board."

[27] M. Leman, "Musical gestures and embodied cognition," 2012.

[28] E. M. Leman and D. Cirotteau, "Sound to sense, sense to sound: A state-of-the-art," 2005.

[29] S. Lee-Cultura and M. Giannakos, "Embodied Interaction and Spatial Skills: A Systematic Review of Empirical Studies," *Interacting with Computers*, vol. 32, no. 4, pp. 331–366, 11 2020. [Online]. Available: https://doi.org/10.1093/iwcomp/iwaa023

[30] "A review of computable expressive descriptors of human motion," *ACM International Conference Proceeding Series*, vol. 14-15-August-2015, pp. 21–28, 8 2015.

[31] P. Hackney, "Making connections: Total body integration through bartenieff fundamentals," 09 2003.

[32] T. Flash and N. Hogans3, "The coordination of arm movements: An experimentally confirmed mathematical model'," *The Journal of Neuroscience*, vol. 5, pp. 1688–1703, 1985.

[33] D. Fenza, S. Canazza, and A. Rodà, "Physical movement and musical gestures: A multilevel mapping strategy." [Online]. Available: www.eyesweb.org

[34] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, pp. 106–118, 4 2011.

[35] F. Visi and L. Dahl, "Real-time motion capture analysis and music interaction with the modosc descriptor library." [Online]. Available: https://github.com/motiondescriptors/modosc

[36] "OpenSoundControl." [Online]. Available: https://ccrma.stanford.edu/groups/osc/index.html

[37] "Max/MSP, Cycling '74." [Online]. Available: https://cycling74.com/

[38] Y. Orlarey, D. Fober, and S. Letz, "Faust : an efficient functional approach to dsp programming - archive ouverte hal," *New Computational Paradigms for Computer Music*, pp. 65–96, 2009.

[39] D. Shiffman, "The nature of code." [Online]. Available: https://natureofcode.com/book/chapter-7-cellular-automata/

[40] M. Wright and A. Freed, "Open soundcontrol: A new protocol for communicating with sound synthesizers." [Online]. Available: http://www.cnmat.berkeley.edu/People

[41] A. Freed and A. Schmeder, "Features and future of open sound control version 1.1 for nime."

[42] J. Maccallum, R. Gottfried, I. Rostovtsev, J. Bresson, and A. Freed, "Dynamic message-oriented middleware with open sound control and odot - archive ouverte hal." [Online]. Available: https://hal.archives-ouvertes.fr/hal-01165775

[43] L. Dahl and F. Visi, "Modosc: A library of real-time movement descriptors for marker-based motion capture," *ACM International Conference Proceeding Series*, 6 2018.

[44] S. O. H. Madgwick, "AHRS algorithms and calibration solutions to facilitate new applications using low-cost mems." [Online]. Available: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.681552

[45] F. G. Visi and A. Tanaka, "Interactive machine learning of musical gesture," *Handbook of Artificial Intelligence for Music*, pp. 771–798, 11 2020. [Online]. Available: https://arxiv.org/abs/2011.13487v1

[46] F. Visi, "GIMleT – Gestural Interaction Machine Learning Toolkitk," 2020. [Online]. Available: https://github.com/federicoVisi/GIMLeT

[47] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization."

[48] "Real-time human pose estimation in the browser with tensorflow.js — by tensorflow — tensorflow — medium." [Online]. Available: https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5

[49] "TouchOSC, Fully modular touch control surface for OSC and MIDI." [Online]. Available: https://hexler.net/touchosc-mk1

[50] M. Marolt, M. Kaltenbrunner, and M. Ciglar, *The HISSTools Impulse Response Toolbox: Convolution for the Masses*.

[51] R. Kirk, M. Abbotson, R. Abbotson, A. Hunt, and A. Cleaton, "Computer music in the service of music therapy: the midigrid and midicreator systems," *Medical Engineering Physics*, vol. 16, pp. 253–258, 5 1994.

[52] T. Winkler, "Composing interactive music : techniques and ideas using max," p. 350, 1998.

[53] A. Hunt and M. M. Wanderley, "Mapping performer parameters to synthesis engines," *Organised Sound*, vol. 7, pp. 97–108, 2002.

[54] M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proceedings of the IEEE*, vol. 92, pp. 632–644, 2004.

[55] M. Lee and D. Wessel, "Connectionist models for real-time control of synthesis and compositional algorithms." *Proc. of the 1992 Int. Computer Music Conference*, pp. 277–280, 1992.

[56] G. Garnett and C. Goudeseune, "Performance factors in control of high-dimensional spaces," *Proc. of the 1999 Int. Computer Music Conf.*, pp. 268–271, 1999.

[57] A. Mulder, S. Fels, and K. Mase, "Empty-handed gesture analysis in max/fts," *Kansei, The Technology of Emotion. Proc. of the AIMI Int. Workshop*, 1997.

[58] A. Hunt, M. M. Wanderley, and M. Paradis, "The importance of parameter mapping in electronic instrument design," *International Journal of Phytoremediation*, vol. 21, pp. 429–440, 2003.

[59] P. Harper and K. Norman, "Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5," 01 1993.

[60] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert scale: Explored and explained," *British Journal of Applied Science Technology*, vol. 7, pp. 396–403, 01 2015.

[61] M. G. Onofrei, S. . Willemsen, and S. Serafin, "Implementing physical models in real-time using partitioned convolution: An adjustable spring reverb," pp. 108–114, 2021.

[62] G. Rimoldi and J. Manzolli, "Enactive framework for design of digital music interfaces."

Appendix

A. Transcribed Interviews

Question 1: How would you describe your overall experience from the installation and what made the biggest impact on you during your interaction?

Participant 1: It was an exploratory experience, I was engaged by the way of interaction by movements that trigger sounds. Maybe because I knew it is based on physical models and connecting interactivity I was trying to visualise the instrument and associating movement with it. For example plucking and bowing was the most obvious interaction. I realised straight away that I am dealing with a string, so I tried to find a way to play it and work around it.

Participant 2: It was a nice experience, the most impactful part for me was when I discovered that the faster I shake the phone the crazier the sounds got and that made it really entertaining.

Participant 3: What I found nice was that movement was required and agency from my side which made it interesting and engaging. I figured fast that my movements make the sounds and it was nice that there wasn't continuous sound and I was completely in charge of making the sounds , not only manipulating them.

Participant 4: I felt I am interacting with a completely new type of instrument , so at first I felt quite lost and didn't know what to do. Slowly I started making sense of it , although I though the way of interacting is where you place your hand in space, which confused me a bit. My overall experience was quite engaging but still I don't think I was able to explore to full potential of the system.

Participant 5: I think it took a lot of time for me to learn, because I didn't know at first what I was doing and I also felt that I don't have any control over the system. It didn't feel like I was affecting and doing the sound and everything felt a bit random. Then halfway through I started making sense of the sounds relating to the movements. After that realisation I started to have more fun and tried repeating some movements that related to certain pattern of sounds.

Participant 6: Overall very sensory engaging experience. I had a lot of fun figuring out on my own the movements triggering the sounds while overtime getting to very interesting results. It felt kind of rewarding to discover the multiple possibilities of the system. I almost could imagine going even more into depth of exploration, I could almost continue for 20, 30 minutes since I was quite engaged throughout the whole interaction. I was captivated by the sounds and I almost felt that I am creating a piece. I felt I am playing something like a tibetian bowl with some percussive timbres. Mapping wise with movements I feel that jerkiness for example made my hand tired but that also was related to more complex sounds so that made it feel almost as a reward after an exercise. Overall I am very satisfied with the interaction and it was very interesting experience.

Participant 7: The interaction since the beginning was quite interesting, since I was free to explore movements on my own. I wasn't sure how fast I can move but once I realised what I can do I imagined the vast field that you can explore through the interaction. I was always getting audio response with my movement. Overall I found the experience to be extremely nice and engaging.

Participant 8: First I was trying to figure out how the movements that I was doing were affecting the sounds and the more I moved , the more I started mapping in my head where the sounds were coming from and I started imagined the instrument that I was playing. Overall I had a pleasant interaction. I was engaged , I was intrigued, I wanted to start creating something.

Participant 9: I had a bit of mixed feelings. The environment setting was a bit uncomfortable for me. I felt a bit shy to do certain movements. I was a bit self aware . Other than that it was an interesting experience. The more I shook the phone I got very interesting sounds and I was pleased to discover that.

Participant 10: I think I had a hard time in the beginning to figure out which movement trigger the sounds. For 3 minutes I had no idea how to work with it and I felt that things were kind of happening on their own and I had no control over the system. But when I connected the dots it all started to make sense and I was able to figure out the interaction. I think that answers the question about a system that is not there to be figured out but more likely for the purpose of embodying the music that you are making.

Question 2: In which moments during the interaction would you consider that you experienced pleasant engagement or frustration?

Participant 1: Satisfaction is when you expect the system to behave in some way and it does or does something unexpected but interesting. I had several moment with both.I experienced significant latency during the interaction, so that made me a bit frustrated and led to trying out different ways to make it work, but it seemed to be not very responsive. It is hard to associate movement with sound when you have that in the way. I enjoyed the FSR , because every time I found a nice pitch I was able to keep it. Also I would suggest instead of a FSR , to maybe use a button. It seems like a more intuitive choice.

Participant 2: Messing with the pitch was very fun for me. The sensor on the back of the phone was hard to press and sometimes I wasn't able to maintain the pitch that I wanted.

That was a bit frustrating for me. I was also trying to see how the device responds on the different axis, and I thought there is something there but there wasn't.

Participant 3: I found a bit frustrated by the mappings. I didn't get the feeling that I am getting the same sound with the same movement. Or at least it took a while to understand which sounds I am triggering. At the beginning it felt a bit random.Also I was clicking by mistake the buttons on the side of the phone and that annoyed me a lot. I was very pleased to find new sounds and what made me happy was discovering the potential of the sounds that I can reach through my movements, like the higher pithes and different timbres. I got engaged in trying to find new sounds through interesting movements.

Participant 4: I kind of tried to make a rhythm out of it and since I couldn't I felt a bit of frustration, also there was a bit of a latency between my movements and the sound.

Participant 5: There were moments where i felt the system wasn't responsive enough, it didn't feel that the pace of the movements were doing big difference and that frustrated me a bit. Although at some point towards the end I found a really nice sound with a gesture interaction that reminded me of bouncing a ball and this pushed me to start feeling engaged with the interaction.

Participant 6: I guess frustration was when I was unable to lock a note . I was engaged by getting the element of surprise of doing some movement and getting an interesting sound. But then I kind of wished to be able to keep that state and build upon it more likely than getting back from the start.

Participant 7: I experienced pleasant engagement when I was moving slow, the movements and the sounds were delicate and very engaging and I felt in control. I felt I was playing a fictional drums and it made sense how my movements were producing certain sounds.The moments where I was making fast movement I was getting a disturbing sound that I felt I cannot control. Also there were two times were there was a significant delay between my movements and the sound. I expected more continuous reaction.

Participant 8: I was happy to discover that the acceleration was changing the pitch. I was looking for more dynamics in the sound. I was a bit in my head trying to imagine what I am working with so I can control it better.

Participant 9: There were several moments where I didn't get any audio feedback while I was doing some movements and I got a bit frustrated and I thought I am doing something wrong.

Participant 10: When I started I was moving my body more than the phone itself, but that also comes from the fact that I am a mediology student so I am expecting sensors that would catch my whole body movement , but then I focused on the device and started getting the image of like playing an instrument.

Question 3: What affected your behavior the most during the performance?

Participant 1: There was few time where I heard by making a certain movement a specific scratchy sound which made me try to repeat the movement. I was also experimenting with the plucking and what exactly triggers it and trying to find ways to alter that.

Participant 2: Definitely the bowing movement made me realise I can play it like a bowing instrument. I also discovered interesting sounds while doing figure 8 movements , which made me explore more moving in a more fluid way.

Participant 3: Probably it was mostly on me trying to find cool timbres and my interaction with the system was mostly driven by the fact that I want to find cool sounds.

Participant 4: When I moved faster and the pitch went up, it showed me that I have to make more effort so I can get more interesting results out of the interaction.

Participant 5: In the beginning I heard some kind of bowing sound that I didn't find very pleasant , but then when I matched it to a sound mimicking the interaction with an instrument I started to search for more movements that reminded me of the interaction with a real instrument.

Participant 6: Not necessarily . Maybe only my need to keep certain sounds there.

Participant 7: Those times were there was a lag made me search for a way to trigger sounds again.

Participant 8: I discovered some bell sounds that I wanted to obtain again but I wasn't able to find again, so I was constantly searching for them. The acceleration was also very engaging for me .

Participant 9: There were couple of things. I was striving to not repeat movements to try to understand the full potential of the system. I was trying to make faster motions so I make more interesting sounds. I had a feeling that I need to keep varying things. My knowledge of IMU made me realise that it wont make use if I move high or low, so I guess I can say that affected my way of moving.

Participant 10: It took me some time to realise the movement so I guess that affected me the most . For example I got very excited to get an interesting sound and then relate it to the movement that I made.

Question 4: What was your focus on during the performance?

Participant 1: My focus was on finding interesting sounds.

Participant 2: Initially my focus was on the device, because since it is a phone I was being careful to not be very agressive with it so I don't break it. But once I figured the extend to which I can use it I started focusing more on the movements.

Participant 3: I felt that the level of embodiment of the instrument is very high. I could find a lot of expressivity with my movements quite easily and at some point I stopped caring about the way I was moving and focused more on listening and being guided by that. I completely skipped on the sensor since it felt quite uncomfortable and difficult to press.

Participant 4: I was probably mostly listening to the sounds, but also on the device since I thought the placement in space would make a difference.

Participant 5: I was focused mostly on the sound and how to obtain it.

Participant 6: I would say 65 percent on the sound and the rest on the movements. They kind of depend on each other. Most of the times I was getting a sound and then thinking about the movement that I made to trigger it so it was kind of like a feedback. But my attention was mostly on the sound.

Participant 7: In the beginning my focus was mostly on the device. I just wanted to make sure I don't drop it, but after a while I was pretty attracted to the movements. I was absolutely absorbed by them and by the end I really pushed them further and my interest grew exponentionally.

Participant 8: I was mostly focused on the movements. The sounds were pretty similar and repetitive so I was mostly trying to discover weirder movements that would maybe trigger something else. The sensor was the least apparent to me. I was also not interested in holding the pitch but most likely understand what changes it.

Participant 9: It's obviously a combination of all movement , sounds and the device itself. Mostly my focus was on how to explore the full possibilities of the sounds through the movements.

Participant 10: It was kind of intertwined but I had to funnel down my desire of multiple sensors to focus on the phone itself.

Question 5: What do you think about the device you were using- mobile phone or another object?

Participant 1: A phone is not a bad option, but something designed to be more comfortable to hold a grip onto would be better.

Participant 2: I would have liked a different object. Because it is a phone I would tend to be careful with it, but if it is another object specifically designed for that purpose I would go crazier with it .

Participant 3: I would prefer to have another object. The size of the phone big and hard to be held comfortably. The idea about using your phone on the other side is very nice since it is an object that we interact with every day. So it being transformed into a controller to make music with seems very intriguing. Theoretically and conceptually it is a great idea to be integrated in that way.

Participant 4: The mobile phone is a good choice , since it is available to everyone, but it also can. be weird for this exact interaction , because people are used to look at the screen and it is not very comfortable to be held in a strong grip and moved around in a fast pace.

Participant 5: I would prefer to have a different object.

Participant 6: The phone is a super nice choice since it is an every day object. So the idea of giving it such an extraordinary purpose makes it quite an attractive choice.

Participant 7: The device is good, since it is handy, something that we use every day. But I was a bit annoyed with the cable connecting the sensor, since it affected my movements and predisposed some distraction to me. Also if the phone is smaller maybe it would be better.

Participant 8: Since everyone has a mobile phone it makes sense to make an interesting use of it. But because of the ravenous movements people would be too cautious and not be able to explore its full potential. Maybe if you have something on the back of the phone that you can slide your hand through would make it more sturdy.

Participant 9: I wanted to shake and take a looser grip on it , but then I held myself back , since I was afraid to break it. So I guess a sturdier object would make more sense to use since it would make people not be very careful and go a bit crazier.

Participant 10: A phone is a quite accessible tool , but quite frankly I would prefer to not have a device at all. More likely to have a space that I can move in and create sounds.

Question 6: What new ideas would you say were triggered by your experience with the installation?

Participant 1: I would imagine this setting but with an addition that would allow you to keep certain sounds and build upon them.

Participant 2: I can definately see that in an art installation , like a walk through exhibition where you can grab and play with it. I also see a dance piece being made with it.

Participant 3: The model sounds very nice, it has a lot of different timbres, envelopes. The sounds were very diverse and interesting.I would probably add a custom controller or a different sized phone. I would imagine this being set up as an art installation with bunch of phones and people being able to interact with each other.

Participant 4: I would not imagine to sit at home and play this, but for an art installation it would be a nice way to show the endless possibilities that a phone in our pocket is able to execute.

Participant 5: I would add some haptic feedback, something to be more helpful understanding the interaction. Also introducing another device would be a nice implementation.

Participant 6: I imagine this being used by dancers, but also on more compositional lense. It feels nice to bring movement into generating sound. Instead of being infront of your computer composing or pressing buttons you can just grab that and create something really cool. It would be also interesting to introduce some kind of spatialisation into it.

Participant 7: I have never encountered an experience like that. To be able to create sounds with my movements was very engaging and fun experience. I was literally transported in another world.I would imagine this being set up as an art installation where multiple people could interact between each other through multiple phones. This would show how the mix of technology and art can create something beautiful.

Participant 8: I would imagine some kind of social interaction where people have multiple phone and can interact and collaborate between each other and creating a whole piece.

Participant 9: I guess I would fill the dead audio spots when I was moving some movements but wasn't getting any sounds. I would also be happy if I can have some visual feedback of the actions I am doing. I am also thinking about some machine learning, gesture recognition way of mapping the parameters.

Participant 10: I think it would be really cool to add some visualisations of the sounds. In that way it would be more accessible and easy to play with in my opinion.

B. Gestures Data Plots

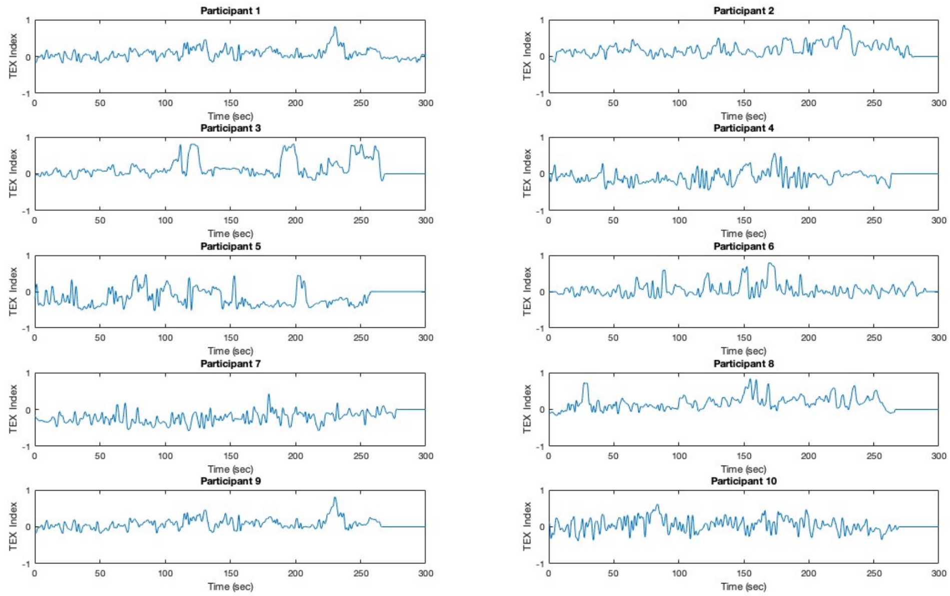

Fig. 23. Time effort on the X axis for all participants.

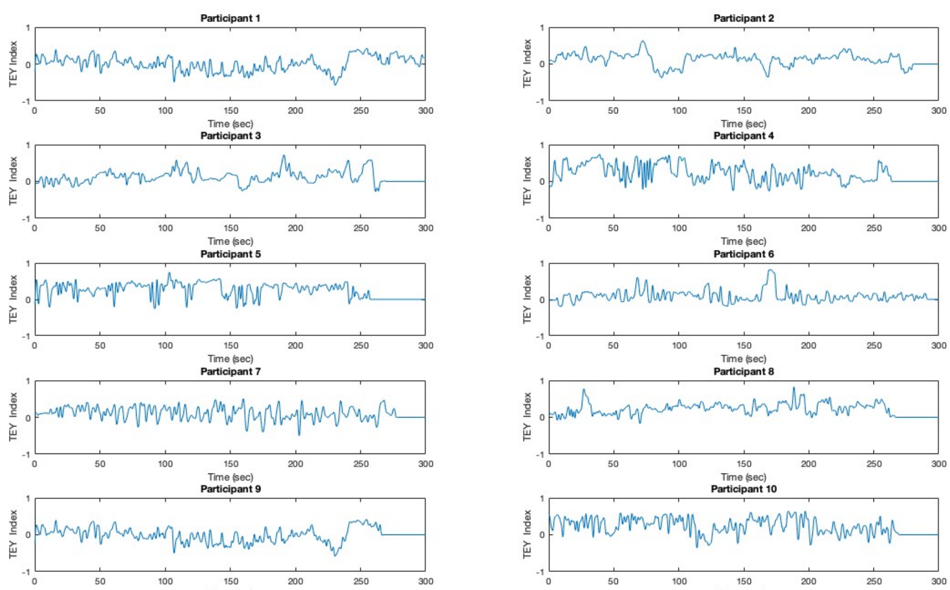

Fig. 24. Time effort on the Y axis for all participants.

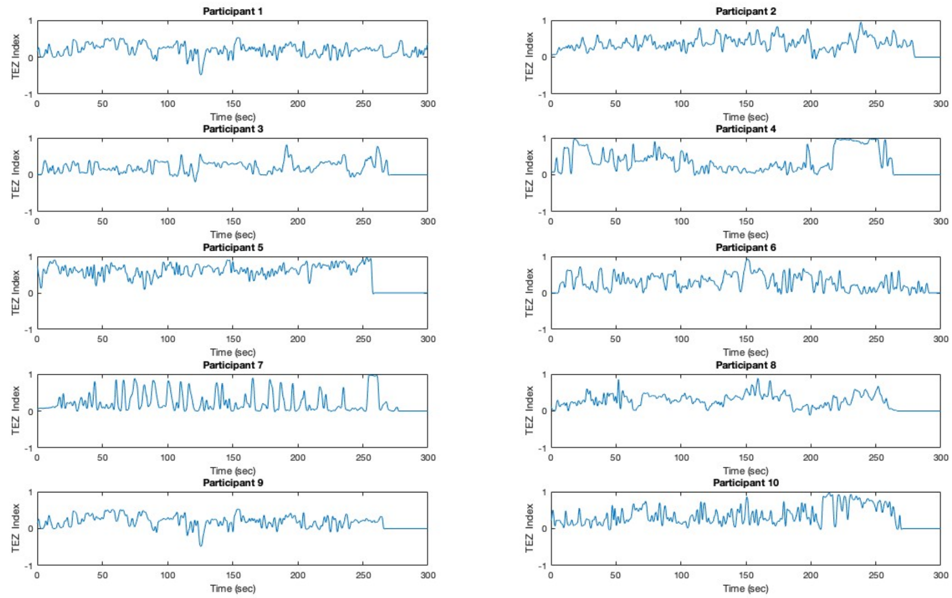

Fig. 25. Time effort on the Z axis for all participants.

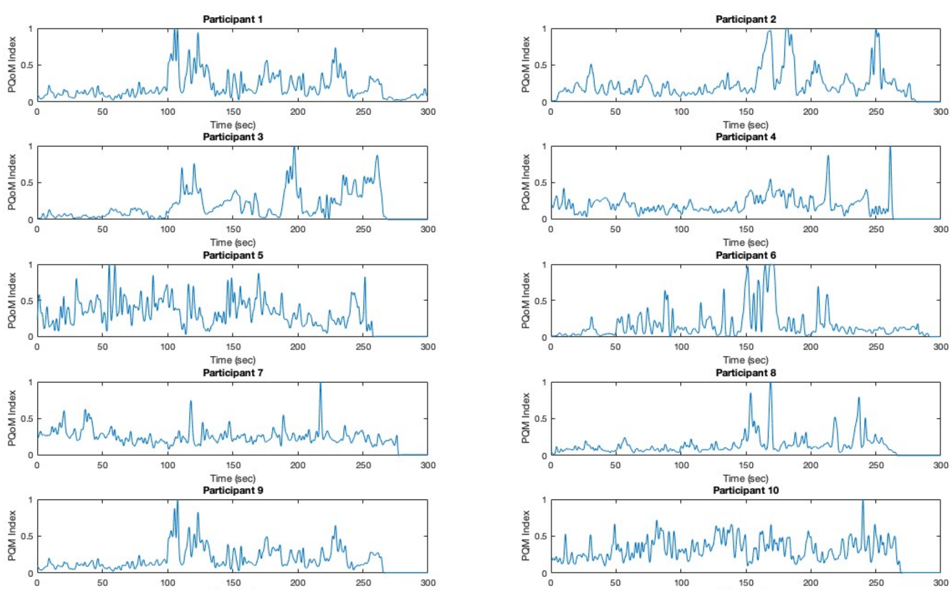

Fig. 26. Periodic Quantity of Motion for all participants.

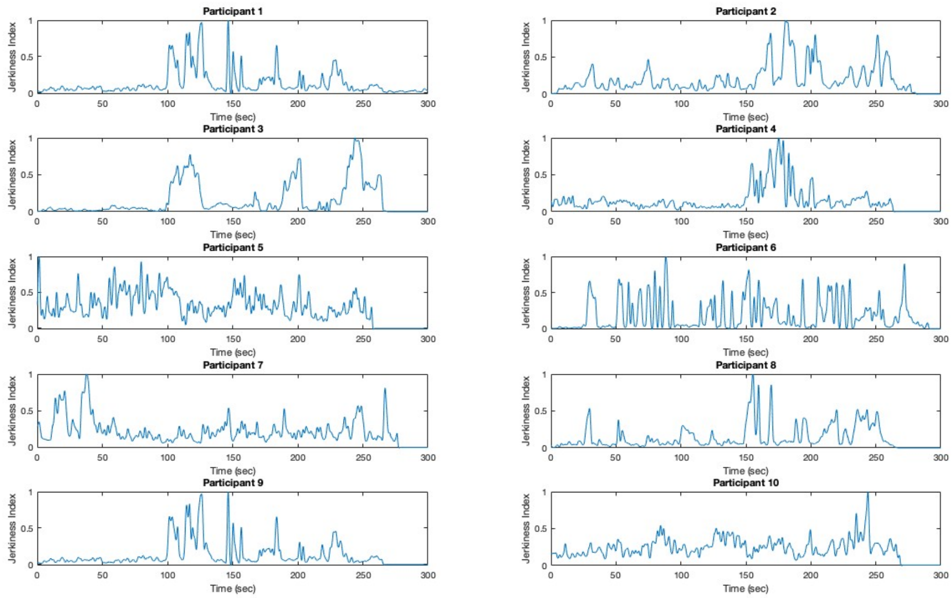

Fig. 27. Jerkiness for all participants.

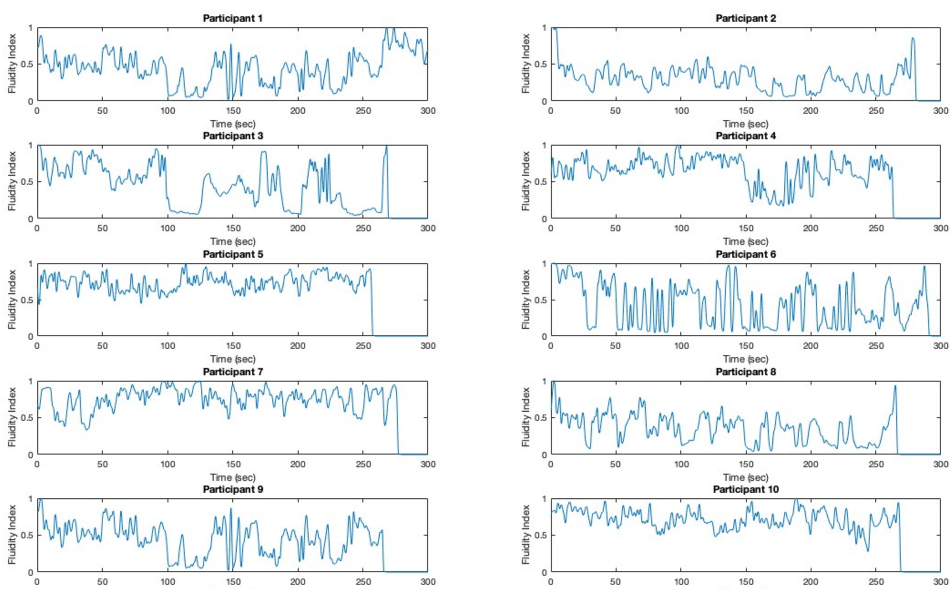

Fig. 28. Fluidity for all participants.

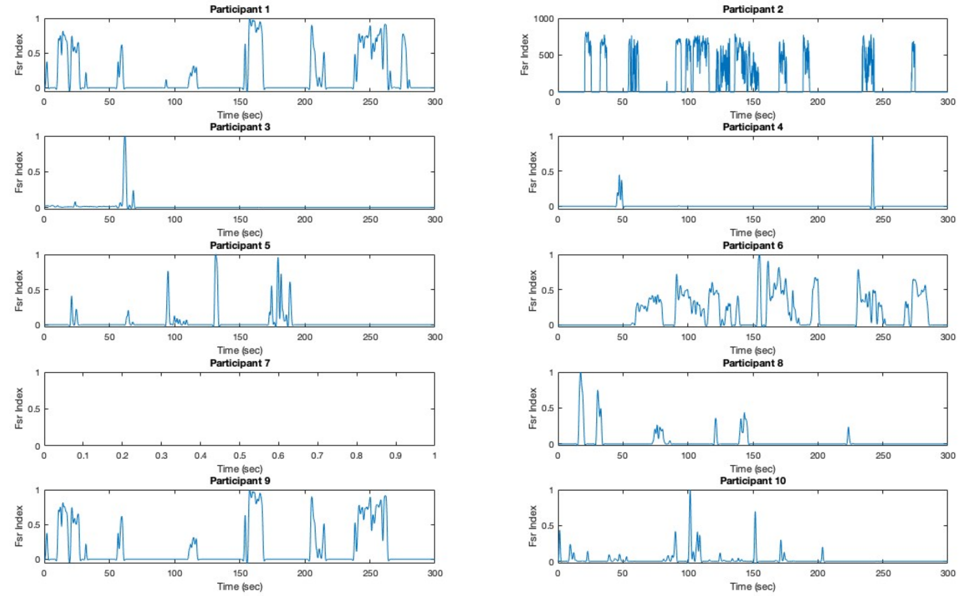

Fig. 29. FSR usage for all participants.

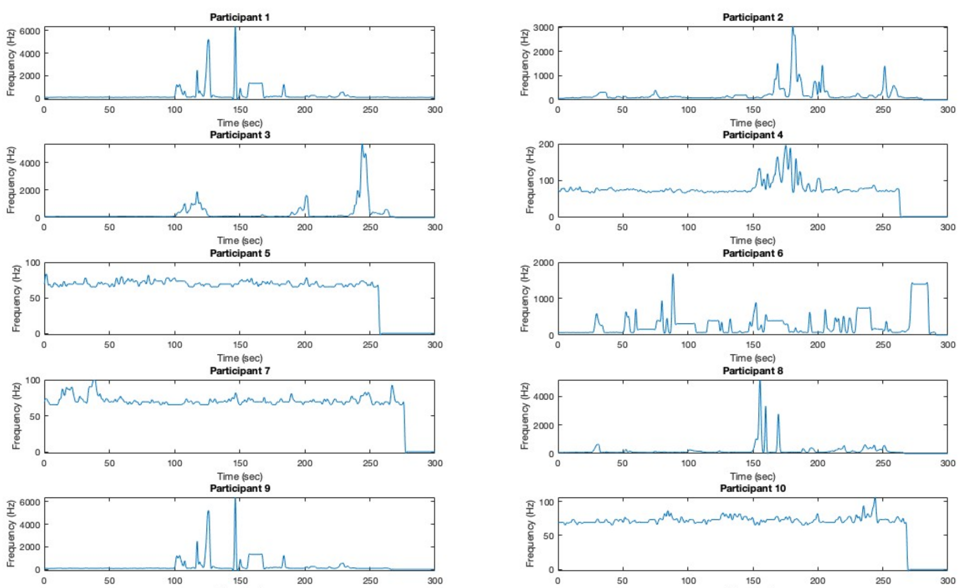

Fig. 30. Pitch (in Hz) played by all participants.