GeKiPe, a gesture-based interface for audiovisual performance

José-Miguel Fernández
IRCAM
Paris, France
jose.miguel.fernandez@ircam.fr

Thomas Köppel
Haute Ecole d'Art et de Design de Genève
Geneva, Switzerland
thomas@werkstadt.ch

Nina Verstraete
Flashback Ensemble
Perpignan, France
nina.verstraete.flashback@gmail.com

Grégoire Lorieux
IRCAM
Paris, France
gregoire.lorieux@ircam.fr

Alexander Vert
Flashback Ensemble
Perpignan, France
alexander.vert.flashback@gmail.com

Philippe Spiesser
Haute Ecole de Musique
Geneva, Switzerland
philippe.spiesser@hesge.ch

ABSTRACT

We present here GeKiPe, a gestural interface for musical expression, combining images and sounds, generated and controlled in real time by a performer. GeKiPe is developed as part of a creation project, exploring the control of virtual instruments through the analysis of gestures specific to instrumentalists, and to percussionists in particular. GeKiPe was used for the creation of a collaborative stage performance (Sculpt), in which the musician and their movements are captured by different methods (infrared Kinect cameras and gesture-sensors on controller gloves). The use of GeKiPe as an alternate sound and image controller allowed us to combine body movement, musical gestures and audiovisual expressions to create challenging collaborative performances.

Author Keywords

audiovisual performance, motion capture, gesture recognition, gestural control, gloves, visualization, spatialization

ACM Classification

H.5.1 [Multimedia information systems], H.5.2 [User Interfaces], H.5.5 [Sound and Music Computing]

1. INTRODUCTION

Most musical activities (e.g. performance, conducting, dancing) involve body movements or gestures. Musical gestures can be studied based on their spatial aspects, functional aspects, their use in performances (as communication or control tools), or for metaphoric artistic purposes. The recent advancements in computing, electronics and sensors technologies resulted in growing interests in new musical interface designs, allowing researchers and artists to address questions about movement and gestures in a musical context. Musical gestures can be interpreted as the intersection between observable actions and mental images [4]. They can be studied at various levels, ranging from the purely functional to the purely symbolic, whether we consider them as effective (sound producing), accompanying (supporting the effective gesture) or for more figurative cues [3]. An analogous definition is suggested by [8] who state that a "gesture is a movement or change in state that becomes marked as significant by an agent. [...] For a movement or sound to be(come) gesture, it must be taken intentionally by an interpreter, who may or may not be involved in the actual sound production of a performance, in such a manner as to donate it with the trappings of human significance." In other words, musical gestures should be meaningful and carry significant information (communication, control, metaphoric).

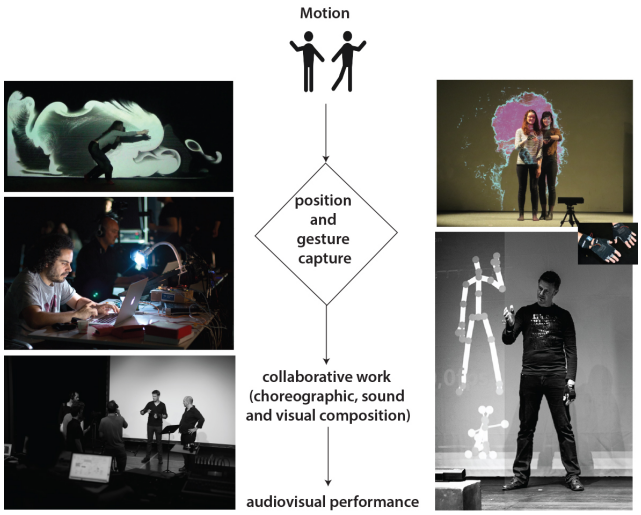

Figure 1: GeKiPe (Geste, Kinect, Percussion).

The GeKiPe (Geste, Kinect, Percussions) interface was developed in 2015 as a creation and research project whose main interest is the exploration and the control of virtual instruments based on the analysis of gestures, specific to percussionists. Built as an interdisciplinary approach involving professional players, composers, music programmers and visual artists, GeKiPe aims at concrete musical and audiovisual applications, with special attention on sound, visual, and gesture qualities. The GeKiPe project was initiated with the objective of improving musician's gesture quality and fineness. Our approach is to achieve this through continuous fine-tuned controls and sound synthesis, rather than using the traditional "on/off" system in which sounds are

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.

mostly triggered by recognized gestures.

The proposed interface allows the performer to trigger sounds and images during performances, and further control them by their body movements and gestures, which are captured by Kinect cameras and sensors. Using their whole body as a musical instrument, performers can draw spatial trajectories that translate as a visual and auditory poetic composition on a virtual stage (Figure 1). Different composers are invited to write original compositions using GeKiPe, impacting its development according to their needs. Research and creation are closely intertwined all along the collaborative process. In particular, research efforts are being directed towards several fields of expertise essential to artistic and musical expression: the control of gesture components by the performer, the relationship between gesture and sound synthesis, the technological development of the interface, data integration and visual design. In this paper we describe how GeKiPe was developed and its recent applications (e.g. performance, education and musical notation).

2. ARCHITECTURE AND DESIGN

2.1 Motion capture and gesture recognition

The way our system combines two motion capture techniques allows for precise spatial detection, with a minimum latency: an infrared Kinect camera (v.2) provides absolute positions (skeletal tracking) and hand glove controllers (R-IoT sensors, Figure 2), sense accelerations, kicks, inclinations, orientation movements and relative angles of the hands (i.e. yaw, pitch and roll). The R-IoT module embeds a ST Microelectronics 9 axis sensor with 3 accelerometers, 3 gyroscopes and 3 magnetometers, all 16 bit precision. The data is sent wirelessly (WiFi) to the audio and video processing interfaces using the OSC protocol. The core of the board is a Texas Instrument WiFi module with a 32 bit Cortex ARM processor that executes the program and deals with the Ethernet / WAN stack. It is compatible with TI's Code Composer and with Energia, a port of the Arduino environment for TI processors (see links). All analysis are carried out directly within the module, the gestural data being received directly from OSC. The use of the glove controllers enabled us to acquire additional motion details and helped improve several dynamic aspects of the performances. Gestural data are processed by different algorithms, for either audio or visual purposes. They can be dynamically configured or personalised according to the compositions.

Programming all the gestural analysis directly into the RiOT sensor allows for a better temporal precision compared to Wireless system. For example, the directionality of the different gestures is recognized instantly within the sensor module, and data is directly sent into the programs that process incoming information in order to create mappings in Antescofo (see Sound Mapping section). The joint positions (from now on, body data) and the images recorded by the Kinect camera are mapped into the physical space and then used in different scenic contexts depending on the audiovisual compositions (Figure 6).

As a template for performance, we produced a dataset of gestures based on the movements executed by percussionist on different instruments such as drums, drum kit, kettledrum, keyboards, digital percussions, accessories (triangle, hit cymbals, claves, castanets). Additionally, gestures were recorded using different hand strikes (rebound, blocked, brushed, shaken) as well as different fingers' positions (straight, folded, virtual stick gripping).

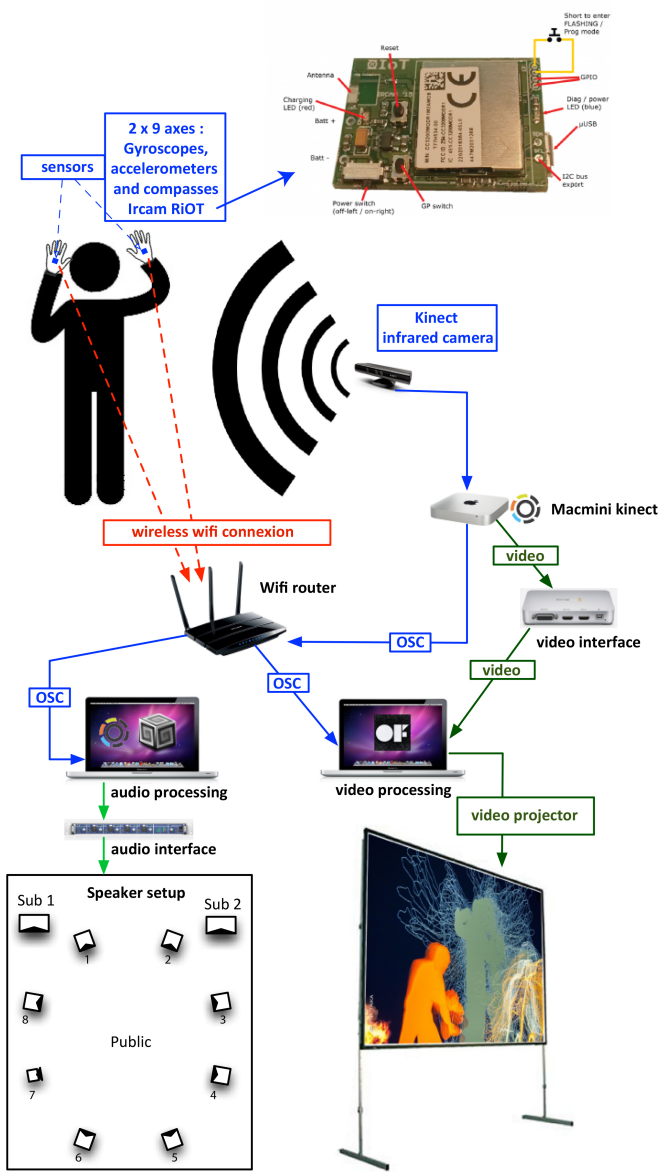

Figure 2: Architecture overview of GeKiPe

2.2 Calibration of the performance area

Six virtual zones are delimited for the performer: right, center, left, and each one distributed in downstage and upstage (Figure 3, lower panel). Three sound control parameters can be mapped on the axes x, y and z, such as height, intensity, effects or panoramics. The performance area is divided in 18 cubes and two different coordinate systems were used (Figure 3, upper panel): a standard one for computer communication (e.g. 331, 121) and a second simplified one, for the performer, based on three planes (A, B, C) and containing six cubes each (e.g. C3, B1).

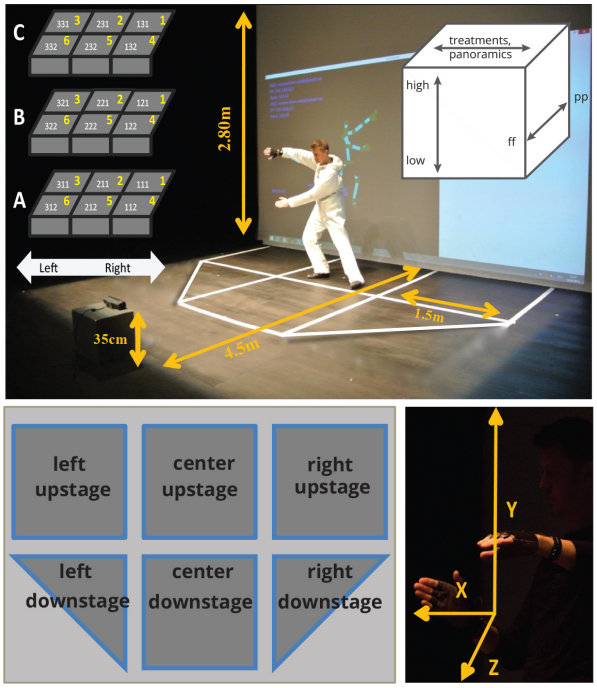

Figure 3: Performance area calibration.

3. INTEGRATION

3.1 Sound mapping

MaxMSP, SuperCollider and Antescofo were used to gather and centralize data coming from the sensors. We also set up a broadcast for constant transmission of data flow. Different types of sound generators were created (additive and substractive synthesis, granular, modulators, phase vocoder, rhythmic patterns) and mapped live with the Kinect device and the sensors based on gesture data from the performer, controlling in a continuous manner the different sound parameters such as height, modulation, panoramic, dynamics, speed of treatments and effects.

Position and movement parameters were captured by the glove controllers and the Kinect camera (Figure 4A), sent via OSC directly to Antescofo, allowing for a dynamic mapping. An example of instantiation for specific gestures is shown on Figure 4B. The sound mappings were done either for synthesis (SuperCollider, continuous control) in the "Hypersphere" composition, or for activation of sound events/effects (MaxMSP, discrete control) in "Le Silence" composition, whether if the composition was fully generative, or mostly based on triggering recorded sounds and live effects. In this latter case, sound events are triggered by specific gestures and Antescofo only allows event n+1 to be triggered if the gesture associated to event n has been captured (Figure 4C).

The advantage of using a text-based programming language such as Antescofo allows events to be dynamically created and destroyed during the performance. This language allows the instantiation of different types of processes, such as continuous controls or rhythmic patterns, and the interpretation of gestural data in order to establish dynamic mappings along the performance. Those mappings may vary according to time, musical sequences, gestural information, or be modified in real time by the programmer (improvisation).

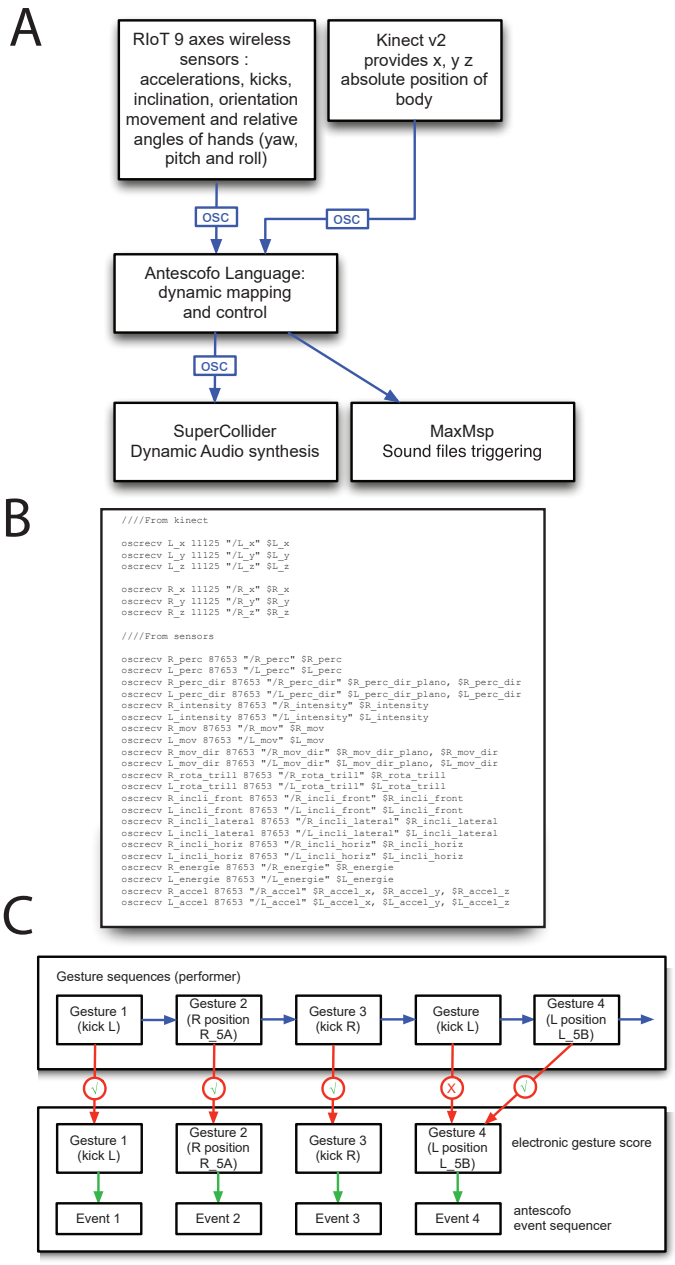

Figure 4: Audio System Integration.

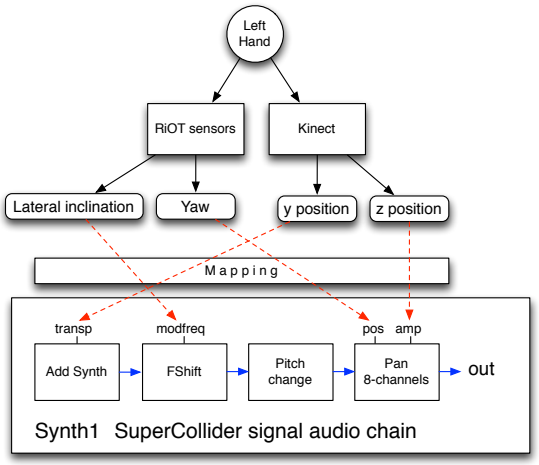

Figure 5: Example of an audio synthesis and mapping programmed in Antescofo. *Antescofo instantiates an audio synthesis in SuperCollider and creates the connections and mappings between the sensors/Kinect (motion capture of the left hand) at a certain point in the score.*

This dynamic mapping method was used in the composition "Hypersphere", comprising 28 sequences, where up to 15 different mappings can coexist. In the first sequence, for example, an additive synthesis (200 Hz, 333 Hz, 427 Hz, 616 Hz, 757 Hz) is generated by SuperCollider. A few mappings were set as follows: 1) inclination of the left hand captured by the controlling gloves : frequency modulation, 2) motion on the x-axis by the left hand, sensed by the kinect : sound spatialization (panoramics), 3) motion on the Z axis by the left hand, sensed by the kinect : amplitude modulation, 4) movement on the Y axis by the left hand, captured by the kinect : transposition. When entering the second sequence, a new additive synthesis sound is generated within SuperCollider (676 Hz, 783 Hz, 898 Hz, 1033 Hz, 1245 Hz), and new mappings are added to control this new synthesis source, which are then controlled by the right hand, in a similar fashion to the previously described mappings (Figure 5).

The audio parameters mapped to the gestures can therefore change as the performance progresses in a written (composed) or improvised manner, since mappings can be created and deleted on the fly. Antescofo also allows to record a series of movements (sequences of gestural data) to generate new materials and sound layers or to control new transformations later in the performance. This process can be an integral part of the composition (previously configured) or be used in real-time (improvisation). This recorded data can then be used to create accelerandos, rallentandos, and generate all kinds of time stretching based on previously recorded gestural data. Antescofo offers the possibility of timing events and the actions relative to the tempo, as in a classical score. This allows for a gradual scaling in real time, and smoother tempo changes. This way, the composer can associate actions to certain events, in absolute time or relative to the tempo, group actions together, define timing behaviors, structure groups hierarchically or in parallel [13].

Our system allows to create a dynamic sequencing of the musical events, directly in the score of Antescofo, according to the gestural score written by the composer. In these events, we write directly the processing chains and sound

Figure 6: Motion capture and image mapping.

syntheses as well as the mappings that will control the syntheses directly from the kinect and RiOT sensors (see example in situ in supplementary figure). The advantage of such a system using Antescofo, over block-diagram languages like Max, is that you no longer need to patch in a static environment to define mappings, you can describe them directly in the Antescofo score which is dynamic. Antescofo serves as a general control at all stages of the composition and allows the connection of the data between them and to realize the mapping between the gestural captured data and the sounds and generative processes of the musical composition.

3.2 Image and video mapping

The visual rendering engine has been programmed with the open source framework openFrameworks. Part of the engine is a fluid simulation realized by calculating a vector field representing the forces in action using movement analysis over multiple captured images (optical flow technique). The behavior of this vector field was programmed according to different parameters sets such as density, gravity or perturbations and different versions were elaborated in order to serve the artistic purpose of the performances.

The captured video images can be used as well as raw visual input and can be affected by different visual effects (e.g. motion blur, time blur, delay). At the same time these images are vectorised (OpenCv library), used for the recognition of contours in the image, for example the performer's body. These vector image data can then be used for algorithmic treatment (perlin noise) for deformation of the initial image.

A data buffer is used to record each frame into a data set containing image vector and movement data, in order to read previously recorded data later in the performance.

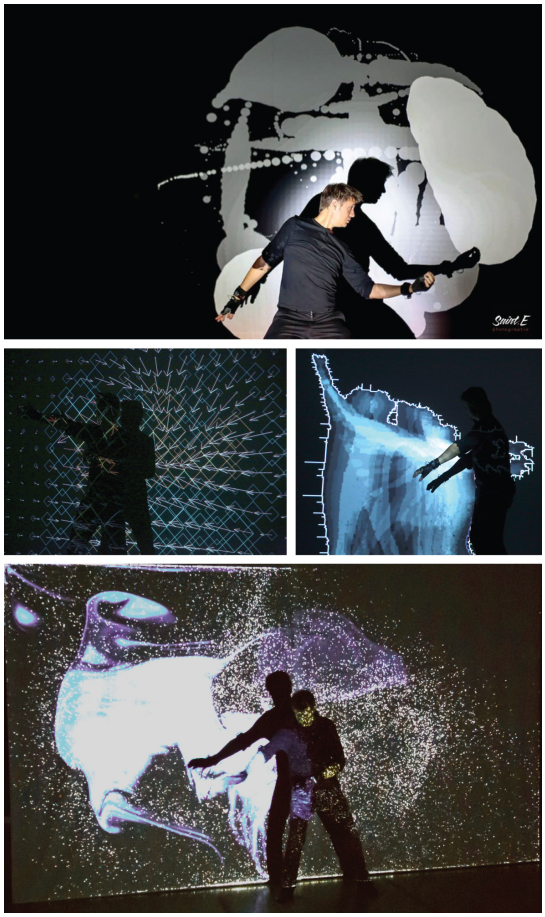

Figure 7: Scuplt performances. *Placed at the center of the system, the performer represents a link between the technological tool, the artwork and the audience.*

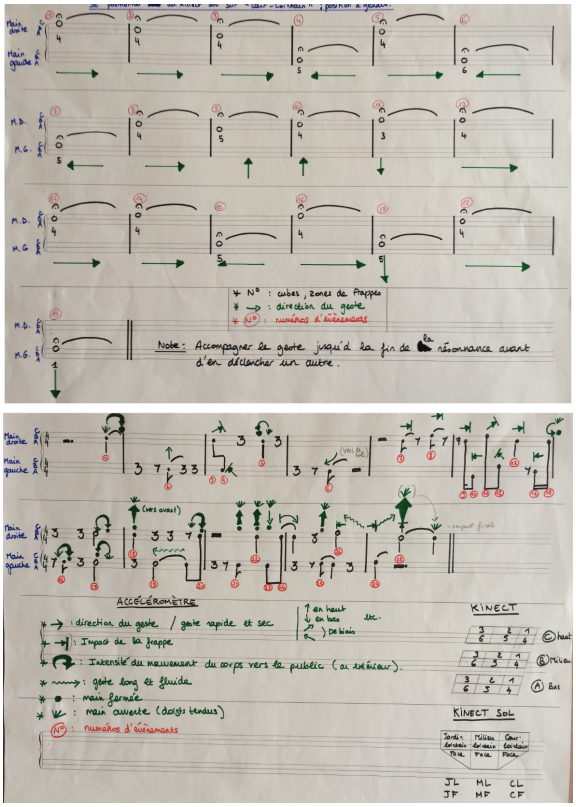

Figure 8: Score transcriptions. *Le Silence, 3rd and 5th movements (top and bottom, respectively).*

Three players can access the buffer, and produce either a direct visual output, or data output for parameter mapping. A routing matrix allows to map any incoming data (sensors data, sound analysis, buffer data) to any of the image rendering parameters (e.g. fluid behavior, visual effects, algorithmic treatments), permitting dynamic data routing. Parameter and mapping data can be stored in presets. By using multiple presets a sequence can be built, in which the parameters can be interpolated from one preset to another, very handy for fluid live applications.

4. APPLICATIONS

4.1 SCULPT Performance

SCULPT is a one-hour multimedia show, mixing music, video and performance, comprising two performances : "Le Silence" by Alexander Vert and "Hypersphere" by José Miguel Fernández (Figure 7). The video is produced by Thomas Köppel, together with both composers and the performer, for the sake of a common writing and poetry. This way, the composers, the visual artist and the percussionist have worked together to produce the show. The audience sees the lead interpreter, Philippe Spiesser, play invisible percussions by drawing trajectories in a three-dimensional space, with his body, his hands and legs. Sounds and images triggered by these trajectories can be further modulated throughout the performance.

In "Hypersphere", uniquely based on generative sound processes, Philippe Spiesser produces all the images and sounds, and modifies them by his movements in real time through algorithms designed by the programmers. Visual intensity of the projected images is in correlation with sound volume, depicting a real interaction between image and sound. In "Le Silence", a more traditional piece, the performer can also improvise through calibrated moves, but he primarily triggers, nearly 400 fixed pre-existing sounds, which he learned and memorized during the elaboration of the performance.

4.2 Score transcription

We elaborated a lexical-based writing system, in reference to standard musical notation. We tried to transcribe musical gestures over time durations in order to facilitate collaboration with written scores, that can be read and played by other interprets (Figure 8). Besides specific gestures belonging to the percussionist's repertoire, works by Thierry de Mey ("Hands", "Musique de Tables", "Light Music") on that topic also served as an essential frame to create our lexicon of musical gestures [7].

4.3 Educational workshops

Along with the research and creation processes, GeKiPe allows a cultural and educational approach, relying on interdisciplinary aspects: musical, visual, choreographic and technologic. GeKiPe workshops have enabled students to relate body motion with programmable sound and music events, in a visual environment scalable as well (Figure 9). On an educational level, GeKiPe allows for some initiation to performing arts through diverse artistic practices: music (sound recording and editing), dance (improvised or guided choreography), visual art (in relation to sound or motion). Direct listening, encouraged by the device form, triggers spontaneously in the students a gestural process. The ex-

Figure 9: Workshops with GeKiPe and students. *Haute Ecole de Musique de Genève, Institut Jaques Dalcroze.*

perimental aspect and the similarities of the approaches used in sound, image and motion design make it possible to build bridges between various art forms (dance, plastic and graphic arts, music, video). By using interactive scenarios leading students to control specific parameters, GeKiPe proved to promote a better understanding of the principles underlying musical and visual composition, and be a powerful educational tool both on artistic and technological standpoints or directly as a medium for learning and teaching.

5. RELATED WORKS

Motion capture and gesture recognition are active fields of research. There are various methods using optical [10], mechanical or magnetic caption [11]. The use of controllers, and more recently glove controllers, have grown in the recent years as well [5], giving rise to an increasing number of applications. Findings in gesture recognition and tracking can be applied to a variety of promising topics such as human-computer interaction, gesture to sound mapping [6], musical creation [1], audiovisual performances [9]. In particular, gesture recognition specific to percussionists have been studied by [2] and [12].

6. CONCLUSIONS AND PERSPECTIVES

GeKiPe is receptive to structural changes. Users should be able to define new gestures and associate them with their own or pre-defined software functions. We hope to suggest new forms of writing, performing and experiencing music, while being accessible to all audience. We work in close collaboration with different composers and artists, mutually sharing their expertise and helping make GeKiPe evolve through original collaborative audio-visual performances. It's important for us that the process require exigent work and relies on new forms of musical expressivity, illustrated by gesture virtuosity from skilled performers. We aim currently at introducing a more important choreographical component, enabling a version for multiple performers in real time, strengthen the sound and image associations and develop their interactions through an image sonification process.

7. ACKNOWLEDGMENTS

GeKiPe, Geste Kinect Percussion, is a research and creation project supported by the Haute Ecole Spécialisée de Suisse Occidentale and developed within the Haute Ecole de Musique de Genève in partnership with the IRCAM - Centre Pompidou de Paris and the association and Ensemble FlashBack. We would like to thank Frédéric Bevilacqua for fruitful discussions and Ignacio Spiousas for critical reading of the manuscript.

8. LINKS

Sculpt Performance teaser: `https://www.youtube.com/watch?v=AM4OwjfhwAs`, Short documentary on GeKiPe:`https://vimeo.com/156279200`, R-iOT sensor module: `http://ismm.ircam.fr/riot/`.

9. REFERENCES

[1] F. Bevilacqua, N. Schnell, N. Rasamimanana, B. Zamborlin, and F. Guédy. Online gesture analysis and control of audio processing. In *Musical Robots and Interactive Multimodal Systems*, pages 127–142. Springer, 2011.

[2] A. Bouënard, M. M. Wanderley, and S. Gibet. Advantages and limitations of simulating percussion gestures for sound synthesis. In *ICMC*, pages 255–261, 2009.

[3] C. Cadoz and M. M. Wanderley. Gesture-music, 2000.

[4] F. Delalande. La gestique de gould: éléments pour une sémiologie du geste musical. *Glenn Gould Pluriel*, pages 85–111, 1988.

[5] R. Fiebrink, P. R. Cook, and D. Trueman. Play-along mapping of musical controllers. In *ICMC*. Citeseer, 2009.

[6] J. Françoise, N. Schnell, and F. Bevilacqua. A multimodal probabilistic model for gesture–based control of sound synthesis. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 705–708. ACM, 2013.

[7] J. Geoffroy. Le geste dans l'oeuvre musicale, la musique et le mouvement. *Le feedback dans la création musicale-Actes des rencontres musicales pluridisciplinaires*, 2006.

[8] A. Gritten and E. King. *Music and gesture*. Ashgate Publishing, Ltd., 2006.

[9] M. Kronlachner and I. m Zmoelnig. The kinect sensor as human-machine-interface in audio-visual art projects. *Live Interfaces: Performance, Art, Music*, 2012.

[10] G. Odowichuk, S. Trail, P. Driessen, W. Nie, and W. Page. Sensor fusion: Towards a fully expressive 3d music control interface. In *Communications, computers and signal processing (pacrim), 2011 IEEE Pacific Rim Conference on*, pages 836–841. IEEE, 2011.

[11] D. Roetenberg, H. J. Luinge, C. T. Baten, and P. H. Veltink. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. *IEEE Transactions on neural systems and rehabilitation engineering*, 13(3):395–405, 2005.

[12] S. Trail, M. Dean, G. Odowichuk, T. F. Tavares, P. F. Driessen, W. A. Schloss, and G. Tzanetakis. Non-invasive sensing and gesture control for pitched percussion hyper-instruments using the kinect. In *NIME*, 2012.

[13] C. Trapani and J. Echeveste. Real time tempo canons with antescofo. In *International Computer Music Conference*, page 207, 2014.