A tabletop instrument for manipulation of sound morphologies with hands, fingertips and upper-body.

Edgar Hemery
Center for Robotics - MINES ParisTech, PSL Research University
60, Bd Saint-Michel
75272 Paris, France
edgar.hemery@mines-paristech.fr

Sotiris Manitsaris
Center for Robotics - MINES ParisTech, PSL Research University
60, Bd Saint-Michel
75272 Paris, France
sotiris.manitsaris@mines-paristech.fr

Fabien Moutarde
Center for Robotics - MINES ParisTech, PSL Research University
60, Bd Saint-Michel
75272 Paris, France
fabien.moutarde@mines-paristech.fr

ABSTRACT

We present a musical instrument, named the *Embodied Musical Instrument* (EMI) which allows musicians to perform free gestures with the upper–body including hands and fingers thanks to 3D vision sensors, arranged around the tabletop. 3D interactive spaces delimit the boundaries in which the player performs metaphorical gestures in order to play with sound synthesis engines. A physical-based sound synthesis engine and a sampler have been integrated in the system in order to manipulate sound morphologies in the context of electro-acoustic and electronic composition.

Author Keywords

Gesture data recording; Gesture Recognition; Physical-based sound synthesis; Morphological transformations; Sound image

ACM Classification Keywords

H.5.m Information Interfaces and Presentation Interaction styles, H.5.5 Information Interfaces and Presentation Sound and Music Computing, J.5 Arts and Humanities Performing arts .: Miscellaneous

INTRODUCTION

Theorising the concept of sound objects in the *Traité des objets Musicaux* [1], Pierre Shaeffer initiated a form of electro-acoustic music, called *musique concrète*, which suggests the listener to identify the sounds individually and to value their sound morphology equally as rules of melody, harmony, rhythm, metre, etc. By the means of new recording and broadcasting technologies, Pierre Shaeffer made tape collages of sound recordings and started exploring music through textures of sounds. This approach encouraged composers to make use of any sound materials they could get a hand on. This idea followed closely the Futurists' manifest, the Art of Noise [2] which first put forward the use of machines sounds for compositional purposes and Edgar Varèse who in 1914, was using "the musical matter itself" in his compositions. Later, with the birth of computer analysis of sound spectrograms, theories of concret music evolved in new musical trends such as spectral music and other musics which focus on timbre as an important element of structure or language. It is worth recalling that before these new perspectives raised in various forms, western music was focused on pitch structures (harmony, modality), construction of musical forms (themes, motives), and rhythm (meter). Timbre was simply used as a matter of colorisation of musical structures and considered in terms of orchestration. Furthermore, electro-acoustic music composition can even be regarded as painting or sculpture [3] where the artist works with shapes and textures.

In 1913, the italian Futurist Luigi Russolo, wrote in *The art of noise*: *'It will be through a fantastic association of the different timbres and rhythms that the new orchestra will obtain the most complex and novel emotions of sound'*. In 1916, Russolo reported police intervention to stop riots at his concert. In the 1950's, Varèse's piece *Deserts*, provoked bad reactions in the audience because of its absence of theme and melodic structures. This facts shows that either the people were not ready for new kinds of experimental sounds or that the music was irritating. Truth is that in the 1930's, composers had only the crudest control over the sounds they were using. Noise music was only at its beginning and artists did not have appropriate tools for controlling it, creating a distance between the composer and the musical manipulation. Varèse and Cage work on percussion music was a natural step in the long process of admitting unpitched sounds into music. It required several generations until people could identify themselves to certain kind of sounds. Nostalgia and melancholia for instance were difficult to convey with these early electronic music, dissociating music to humans' emotions.

In the mean time, the first electronic music instrument appeared in 1928 with the world famous eponymous creation of the russian inventor Leon Theremin. Interestingly, this invention, prior to the computer and motion sensors, was based on *air-gesture* capture as its working principle. However, this latter invention was mostly used to play a classical music repertoire and was not integrated in electro-acoustic compositions at that time. Elektronische music resulted in the 50's

in Cologne from the research of composers such as Stockhausen, Eimert, Beyer and Eppler's on sound synthesis. It was a radically different approach to the concret music since the music and sounds were entirely produced by electronic means. For Eimert, sound synthesis was a real musical control of nature based on the use of sine tones as the fundamental of the art. The main interest of electronic sound synthesis was a desire to control over every aspects of musical compositions. But if this technique changed entirely the course of music, it temporarily lead to total determinism and formalism in the compositional approach.

As mentioned before, one of the main problem in electroacoustic and electronic music is the distance between composers and the composing medium, which later became the computer and the interface. As Pierre Shaeffer wrote: '*The lack of intentional control over musical affect, together with the fact that compositions emanating from such a wide range of compositional aesthetics all produced the same impressions, implicate the common rudimentary sound manipulation technologies*'. This inadequacy for sound manipulation prevents the spontaneity and the emotional intention of the musician. Later in the 80's, improvements over computers CPU capacities allowed for real time control of sound synthesis. This way, the composer had access to a very wide range of sounds and could trigger them spontaneously with the help of keyboards, cursors, mixing desks, buttons and track pads. As Emmerson [4] pointed out, ' *In the 1980s, two types of computer composition emerged from the studio to the performance space, one more interested in event processing, the other in signal processing*'.

Joel Ryan from the Steim Institute in Amsterdam wrote: *In order to narrow this relationship between technology and musicians, it is as much the problem in collaboration to get technologists to respect the thinking of the artists as it is to educate the artists in the methods of the technology* [5]. Signal processing as it is taught to engineers is guided by such goals as optimum linearity, low distortion, and noise. This goals may not be in accordance with musicians wishes. The temptation of programmers is to concentrate on the machine logic rather than the idea of the artist. New suggestions of sound techniques that fit musical expressivity are needed. Several research groups such as the IRCAM in Paris, the STEIM in Amsterdam, the CCRMA in Stanford, and the MTG of Pompei Fabra-Barcelona to name a few, have started to focus on system designs easing interactive manipulation of sounds. This includes thinking about intuitive interfaces, gestural controllers, communications protocols, network designs and an understanding of how they can all be interconnected.

STATE OF THE ART

In the past few years, non-intrusive movement and gesture analysis have been integrated in consumer electronics thanks to the progress made in 3D cameras technology and computer vision algorithms. Computer vision is a branch of computer science interested in acquiring, processing, analysing, and understanding data from images sequences. Non-intrusive gesture tracking systems are ideal for musical performances since they allow freedom in body expression, are not intrusive and are easy to calibrate. Therefore, a musical interface – or instrument – which draws gestural data from vision sensors, feels natural from the user's experience point of view, provided that gesture to sound mapping is intuitive and has a low latency response.

Performing arts have embraced this type of technology from its very beginning, seeing in it an extraordinary springboard for creation of new exciting interactions, highlighting the performer's body and gestures. Starting from 1982 with David Rokeby' series of performances with *Very Nervous System*[1], a new area of embodied interaction making use of cameras and computer vision algorithms was born. A global vision on this scene reveals that the Kinect and the Leap Motion, have already been popular choices among musicians and sound artists for live performances.

Recent advances based on the Leap Motion show its abilities to control high-level music control thanks to a 3D touch-like gesture. GECO is a Leap Motion app (available on the Leap Motion' market space *AirSpace*) which allows to control MIDI, OSC or CopperLan protocols with a simple 3D-gesture vocabulary. Han and Gold [6] use the Leap Motion to create an *air-key* piano and an *air-pad* machine drum while making use of the third dimension to control the sound intensity via the hand' velocity computation. The *BigBang rubette* [7] module uses the Leap Motion to control notes, oscillators, modulators or higher level transformations of sounds and/or musical structures. Alessandro et al. [8] and Silva et al. [9] combined respectively a Leap Motion and transparent sheet of PVC [8] and glass [9] in order to grasp finger movements occurring prior to the touch with the sheet. It is worth noticing that this last example somehow meets with multi-touch tablets and tabletops since the paradigm is based on a finger-screen contact.

Nowadays, multi-touch screens and Omni-Touch wearable interfaces [10] offer tangible interactions that are restricted to a flat surface with finger tapping, scroll, flick, pinch-to-zoom etc. (refer to [11] for extended reference guide of touch gestures). The ReacTable [12] has launched a multi-player tangible interaction of a unprecedented kind with real objects communicating on a multi-touch tabletop. The ReacTable' objects display images that are recognized by an infra–red camera, sending information about the type of sound to be generated to the system. It is striking how this approach falls in with the concept of sketches and shapes of sonic objects described in Schaeffer's typology [1]. A similar work by Thoresen [13] introduced a set of graphical symbols apt for transcribing electro–acoustic music in a concise score, simplifying the sometimes overwhelming complexity of Shaeffer' *Typo–Morphology*.

In the same vein, the use of extra objects such as digital pen in the music production app on Microsoft Surface tablet[2] gives very interesting and intuitive ways for achieving high-level sound control parameters such as drawing amplitude and filter envelopes. This smart tabletop, along with the ReacTable

[1] www.davidrokeby.com/vns.html
[2] http://surfaceproaudio.com/

discussed above, belong to the first generation of devices and instruments to allow embodiment and intuitive manipulation of sound objects.

In line with these latter examples, our instrument, that we are discussing in this article, is an interactive tabletop for playing music in 3D space where the upper–body and fingers' free–movements in mid-air extend the action of the fingers' physical contacts with the table. While Microsoft Research has developed similar technologies and set-ups for grabbing and manipulating 3D virtual objects on and above a tabletop surface with finger gestures ([14] and [15]), the EMI pushes ahead sound mapping strategies in the 3rd dimension. Additionally, the EMI is in line with extended piano-keyboard devices such as the Seaboard by Roli[3] and the TouchKeys[4] by Andrew McPherson, but brings the extended interaction to mid-air with both fingers and upper-body gestures thanks to 3D vision sensors such as the Leap Motions and the Kinect.

We start by describing the structural aspect of our instrument, the sensors that are used and the concepts of *micro* and *macro* bounding boxes articulated around the framework. The Section *Musical embodiment on a tabletop instrument* depicts the metaphors used while designing the interactions with the system. A variety of sound synthesis controls are presented in the section *Sound morphologies manipulation*, showing the musical capacity of our instrument to control sound morphologies. The section *Latency assessment of the EMI* presents a first latency assessment of the system. Then we conclude and give a view of our further works.

DESIGNING A FRAMEWORK

Structure of the instrument

The whole instrument is articulated around an acrylic sheet, which serves as a frame of reference for the fingers. The acrylic sheet is placed 10mm above two Leap Motions, where the sensors' field of view covers the area best and underneath a Kinect placed 1.20mm in front (see figure 1 and 2). The cameras are described later in this section. The sheet also constitutes a threshold of detection for the fingers: one triggers the sound by fingering the tables surface. Gestural interaction is not limited to this surface, but takes part inside a volume above the table. The tracking space serves as a *bounding box*, delimiting the sensors' field of view in which the data are robust and normalized.

The boundary engendered by the table's surface eases the repetition of a type of gesture. This conclusion raised from the difficulties of gesture repetition observed in *air*-instruments, where the movement is done in an environment with no tangible frame of reference. In this regard, it is a profitable constraint to add this surface since it enables the user to intuitively place his/her hands at the right place and helps repeating similar gestures.

As the *Embodied Musical Instrument* is to be used for both performances and learning contexts, it is portable, light, solid and foldable. We have conducted experiments, changing the tilt of the sheet to meet with the literature results concerning the wrist and shoulder posture during touch-screen tablet use [18]. However, movements with the EMI being wide and dynamic, wrist radial deviation was not constant enough to take into consideration optimal tilt angles for specific applications. At last, the sheet supports the arms and allow the user to rest, thus avoiding the *gorilla arm* effect which results in a fatigue while repeating gestures in the air [19].

Vision-based 3D sensors

We present here two types of vision-based sensors, which are used in our research. As this technological field is growing fast, we could not explore all the existing sensors possibilities; however, the sensors we have chosen are well documented, largely spread, low cost and fit our requirements. The first type of sensor is the Microsoft Kinect depth camera. The first version of the Kinect, along with the OpenNI skeleton tracking software delivers a fairly accurate tracking of the head, shoulders, elbows and the hands, but not fingers. It has a 43° vertical field of view, 57° lateral field of view and a ranging limit varying from 0.8 to 3.5 m. Its latency, around 100 ms and its spatial resolution (640x480 pixels) is unpractical for fast and thin gestures at close range (e.g. $<$ 0.50 m). As J.Ballester and C.Pheatt concluded [16], the object size and speed requirements need to be carefully considered when designing an experiment with it. Hence, we will use the Kinect for suitable uses, aware of its limitations and capacities. Typically, the Kinect works sufficiently well between 1.40 and 3m for body tracking, with a 1cm spatial resolution at a 2m distance from an object. Furthermore, latency considerations lead us to use it for higher-level musical structures occurring in the macro space, where temporality is chosen to be loose.

Regarding the small and rapid finger gestures, we are interested in a second type of depth camera, the Leap Motion. This camera works with two monochromatic cameras and three infrared LEDs. Thanks to inverse kinematics, it provides an accurate 3D tracking of the hand skeleton, with more than 20 joints positions and velocities per hand. The Leap Motion has a lateral field of view of 150°, a vertical field of view of 120°. Its effective range extends from approximately 25 to 600mm above the camera center (the camera is oriented upwards). Additionally, the Leap Motion is known for being accurate and fast: processing time for each frame is close to 1ms, which is well below the acceptable upper bound on the computer's audible reaction to gesture fixed by [17] at 10 ms. Although additional latency will be added further with the gesture to sound mapping, the initial latency provides us with a viable starting point.

Micro and Macro bounding boxes

Igor Stravinsky (1970): *The more constraints one imposes, the more one frees one's self... the arbitrariness of the constraint serves only to obtain precision of execution.*

[3] https://www.roli.com/products/seaboard-grand

[4] http://www.eecs.qmul.ac.uk/ andrewm/touchkeys.html

We present here the design of the instrument through the 3D interactive spaces it creates. As presented before, there are three sensors: two Leap motions and one Kinect. Once placed on their slots on the EMI, the Leap Motions' field of view cover the whole surface of the table and a volume up to 30 cm above it. We designate this volume as the *micro bounding box* (figure 1). The Leap motions are centered in the halved parts of the surface while the Kinect is placed in front and above the table as displayed as displayed on figure 2. The position of the Kinect is roughly 1 – 1.20m behind the table and stands roughly 1m above the table. There is no need to place it with great precision as the system auto–calibrates the skeleton with respect to the sensors' field of view each time the software is launched.

The perspective behind the splitting of the two bounding boxes is to differentiate *macro* gestures done with the upper body with *meso/micro* gestures done with the fingers. Hence the macro space deals with the wide movements captured with the Kinect while the micro space deals with finer-grain manipulation captured with the Leap motions. Inspired by Jensenius' terminology [20], we use a unified space for both *micro* gestures happening at a millimeter scale with *meso* sound–producing gestures happening at a centimeter scale.

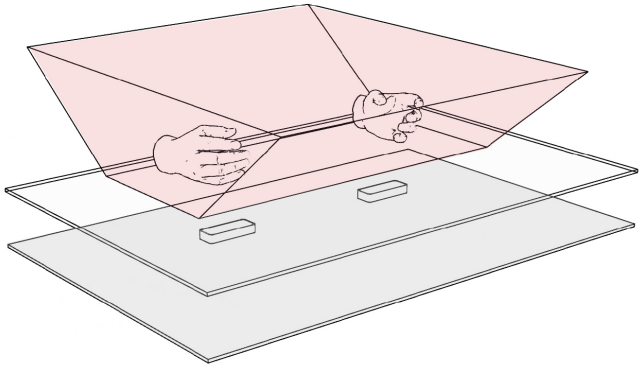

Figure 1. Micro bounding box.

MUSICAL EMBODIMENT ON A TABLETOP INSTRUMENT

One of the objective of the *Embodied Musical Instrument* is to give, through movement, meaning to the sounds thus created. If gestural electronic music performance is technically rendered possible thanks to 3D tracking devices, the coupling of perception and action, however, requires reflections on expressive use of affordance based on practice [21]. The EMI is a framework for gesture tracking and recognition, with its own metaphors and control mappings, unified within an embodied model reducing the cognitive distance between the imaginary imagery of electro-acoustic composers.

Godøy [22] distinguishes among *music imageries* images of acoustic signals, images associated with the performance, images associated with the perception and images associated

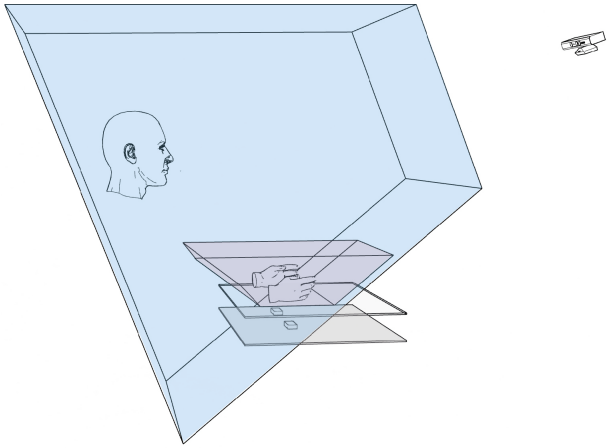

Figure 2. Macro bounding box.

with the emotive experience. He tackles the problem of understanding the nature of these sound images in our mind by drawing sketches of gestural and sonic features in a top-down manner, tracing features and sub-features of sound-morphologies and correlating them with acoustic features of sound objects. We went through a similar thought process, breaking down gestural features in *effective, accompanist and symbolic* gestures (based on Delalande' gestural typology [23]) to some lower-level gestural features enabling very specific sound controls over the articulation, intensity, after-touch and so on. Hence, the interactive space created above the instrument can be seen as a shared gestural space including the *effective, accompanist and symbolic* gestures. In that respect, the EMI inspires an environment analogous to what Tanaka [21] and Graham[24] designate as a *performance gesture ecology*. Basically, we aim with the EMI at capturing the three categories of gestures discussed above and make use of them altogether in order to generate very expressive sounds.

The proposed metaphorical gestures we use are borrowed from keyboard instruments, touch gesture paradigms developed for touchscreen devices [11] and other physically-inspired manipulation metaphors such as an elastic cable, a wheel or a kite. The object metaphor connects to the affordance of a simple object in the mind of the user and thus, leads intuitively to the gesture to be done. We make the same assumption that the simpler the metaphor is, the more intuitive and expressive the result will be. Wessel et al. [25] similarly make this assumption as one of the necessary conditions to get an '*intimate musical control of computers*'. The other conditions being its long term potential for virtuosity, that we believe the EMI also meets with, the clarity of strategies for programming the relationship between gesture & musical results and finally a low latency response of the system.

At last, Young [26] presented how features in electro-acoustic works can be discussed through aural perception of the sound objects in association with an analytical focus based on a common understanding of the way a sound behavioral model operates. This analytical focus is however not always obvious to non initiated listeners and does not solicit the visual un-

derstandings of how things work and are produced. Physical embodiment of music performance, if realistic enough, would convey this additional information, necessary for the spectator to understand the origin of the sounds and reduce the emotional distance between the synthetic sounds and him/herself.

Metaphors in the micro bounding box

First, we were interested in building a model for dynamics, articulation and duration, inherent in the fingering. This led us to the decomposition of the fingering in several phases so as to extract information about the trajectory and the duration of each part. This representation is based on four phases: Rest, Preparation, Attack and Sustain, inspired by a more general gesture segmentation model (Preparation, Attack, Sustain, Release) [27]. Segmenting the fingering into essential phases facilitates the distinction of features for each phase (figure 3). In rest position, the hand and fingertips are relaxed on the table. In preparation, one or several fingers lift upwards. In attack, one or several fingers tap downwards while during a sustain phase, one or several fingertips stay at contact with the surface of the table. At last, the velocity of the fingertip along the z-axis in the few milliseconds prior to its contact with the table during an attack phase is mapped to sound intensity. It is worthy of note that this segmentation, which is articulated around the z-axis is only made possible thanks to the depth finger tracking of the Leap motion.

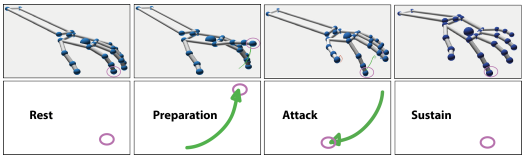

Figure 3. Rest Preparation Attack Sustain segmentation

The EMI makes use of a piano keyboard-paradigm that can be played with the fingers in the *micro–bounding box* (figure 4). The key idea here is to cover a range of notes, without the need to be extremely precise at fingering on the table since the latter is completely flat and transparent. Therefore, the zone (either blue, red or green) corresponds to a set of fives notes (e.g.: EFGAB), where each note corresponds to one finger (see colored areas on figure).

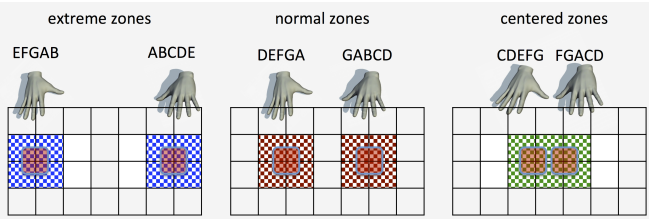

Figure 4. The keyboard paradigm.

We extend the mapping of fingertip positions on the x-axis of the table to the y-axis and attribute this dimension to the timbre space. Hence, the timbre/texture of the sound can be modified continuously by fingering at different locations along the y-axis of the table while keeping the pitch fingering system depicted above. Figure 5 depicts the top-view or avatar of a musician moving arms and hands in the pitch-timbre space.

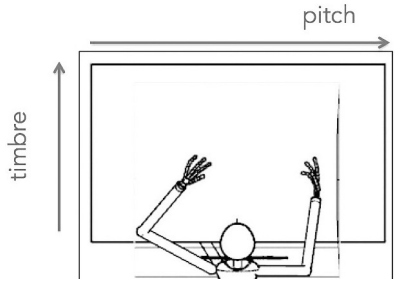

Figure 5. Pitch-timbre space.

Metaphors in the macro bounding box

From the hands' joints provided by the Kinect, we compute the three–dimensional euclidean distance between them and name this feature *elastic control*. The metaphor for lengthening/shortening the 3D euclidean between hands is to imagine that one is stretching/releasing an elastic cable. This gestural metaphor is depicted in figure 6 with the red arrow.

From the three joints *Head – Left Hand – Right Hand*, that we consider as apexes of a triangle, a plane equation is computed. Then, we respectively measure the tilt between this plane and the *xy* plane and *xz* plane of the table. The *xy* vs. *triangle* plane provides a sense of how much left or right your body is rotating, just as if one was pulling the wires of a kite or turning a wheel. Keeping on with the kite–flying metaphor, the *xz* vs. *triangle* plane reacts accordingly if the body is going backward or forward and/or the hands are going higher or lower. These two controls are represented on figure 7 respectively with the red arrow and the yellow arrow.

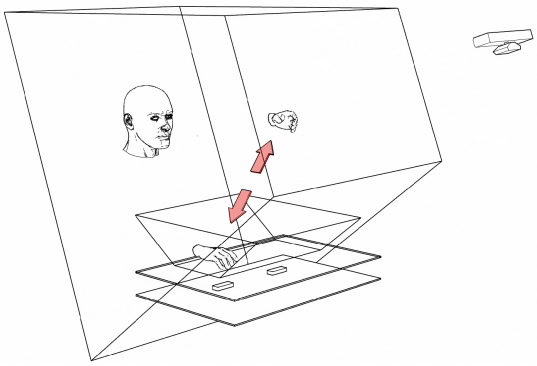

Figure 6. Elastic control: metaphor for lengthening/shortening the 3D euclidean between hands

SOUND MORPHOLOGIES MANIPULATION

In the context of a gesture-based instrument, a necessity for sound morphologies exploration is a repeatable gesture. For this matter, we use the mubu library [28] developed by the STMS team at IRCAM Sound. The mubu library, integrated into the programming language Max, embeds a movement description multi-buffer able to record gestural content in

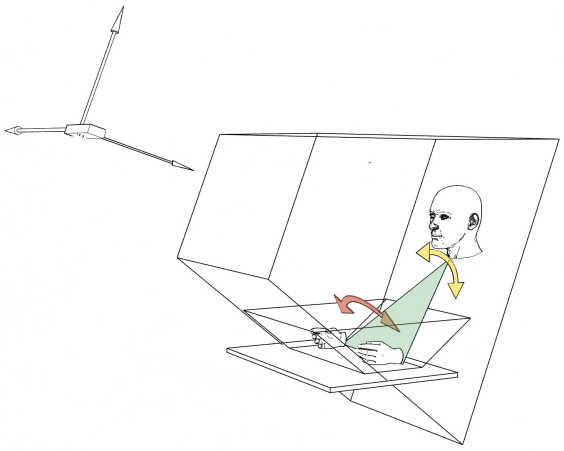

Figure 7. Kite-flying control: The *xy* vs. *triangle* plane provides a sense of how much left or right your body is rotating while the *xz* vs. *triangle* plane reacts accordingly if the body is going backward or forward and/or the hands are going higher or lower

real-time. This buffer can also be replayed, enabling to visualize the gestural data. Combined with the visual programming tool Jitter for Max, it is possible to replay the avatar of the musicians' hands and upper body. This way, one can synthesize a sound in real time with the produced gesture or replay a recorded gesture, at different speeds, forward or backward, and change the sound synthesis parameters in real time.

We discuss here how the timbre dimension is explored thanks to a physical-based sound synthesis engine named *Blotar*. It is a physical modeling synthesizer that is part of PeRColate, an open-source distribution containing a set of synthesis and signal processing algorithms for Max [29] based off the Synthesis Toolkit [30]. Physical-based sound synthesis makes sense for well articulated sounds, which we trigger when the fingers tap onto the table's surface. By changing in turn a mass, a spring or a damper parameter of the Blotar, one can oscillate between a flute and an electric guitar timbre. In our system, the brilliance parameters are mapped with the y-axis of the table's frame of reference. Hence, one can obtain brilliant sounds when the fingers tap close to the edge of the table and rounder sounds when the finger taps in the middle. Finally, the velocity of fingertips when it hits the table is mapped with the attack intensity of the sound taking into account the non-linearities occurring in such events.

Additionally, we have added a virtual piano plug-in [31] simulating physical properties and behaviors of real acoustic pianos. The EMI gesture paradigms being very much inspired by piano-like gestures, this plug-in incorporates well and despite the absence of the spring-keys haptic feedback, provides an intuitive and realistic sensation of piano playing. Finally, we have added an amplitude-convolution functionality to modify the amplitude of the Blotar sounds with the piano plug-in. Hence, one can use the attack and amplitude envelope of piano sounds with the spectral content, transients and effects of the Blotar.

Applying a new morphological frame to various spectral contents, one can reveal and enlarge some aspects such as the transients of the sounds and the sustain phase. For instance, one can imagine a noisy voice with the morphological shape of a bouncing ball or a percussive pitched sound such a vibraphone with long controllable sustain. These enables the composer to select what s/he might be interested in the sound: the shape or the content. Additionally, such physical-based and cross-synthesis techniques, already spread among composers through tools uch as Modalys [32] (IRCAM) could be handled with the EMI.

LATENCY ASSESSMENT OF THE EMI

As we are interested in evaluating the latency of the system, we are looking for the time difference between the moment when one taps onto the table and when the synthesized sound is coming out from the speakers. Therefore the experimental protocol is as follows: a microphone is plugged into a second computer, placed near by the instrument and the speakers. When the player taps on the table in order to produce a sound, the microphone picks up two signals: one is the signal produced by the physical tapping of the fingertip and the other is the synthesized sound coming from the speakers (as can be seen on figure 8). The distance between the respective attacks of the tapping sound and the synthesized sound is measured with a precision of $\pm$2ms, as can bee seen on 9.

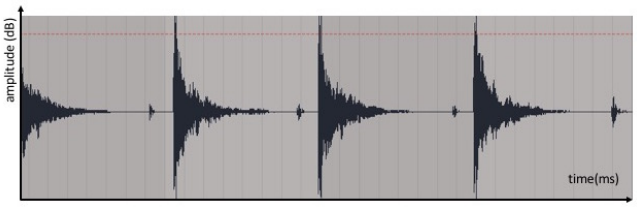

Figure 8. Recording displaying the acoustic signal of the finger tapping the acrylic sheet preceding the acoustic signal of the resulting synthesized sound

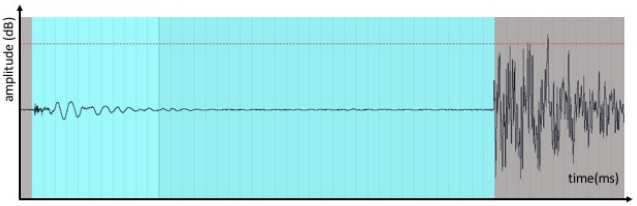

Figure 9. The blue highlighted segment corresponds to the length between the two attacks, corresponding to the latency of the system

We have recorded two sets of gestures in order to evaluate the performance of the system when one single note is repeated several times and when a sequence of notes such as an *arpeggio* is played. The two sets (1) and (2) are the following:

1. Single note repetition with index fingertip – 10 times
2. Arpeggio (thumb-middle-index-index) – 4 times

These two sets of gestures are repeated at various *beat-per-minute* (BPMs) ranging from *Lento* (60 bpm) to *Allegro* (130 bpm) and above. For each series of recording, we compute the average latency in millisecond. Additionally, we compute

the *beat shift* which corresponds to unitary shift of each beat or note (eq. 1). The results are displayed on the table 1.

$$\text{beat shift} = \text{average latency} * (\text{bpm}/60) \quad (1)$$

For instance, a beat shift equal to 0 would be a perfect temporal alignment displaying no latency at all in the system. A 0.5 beat shift musically corresponds to an off-beat and a beat shift greater than 1 occurs when the synthesized sound one hears is produced by the second previous finger tapping.

	Single Note		Arpeggio	
Tempo	Avg (ms)	Beat Shift	Avg (ms)	Beat Shift
60 bpm	154	0.15	167	0.17
70 bpm	174	0.2	184	0.21
90 bpm	230	0.34	232	0.34
110 bpm	245	0.45	273	0.5
120 bpm	304	0.6	314	0.62
130 bpm	301	0.65	345	0.74
140 bpm	369	0.86	364	0.84
160 bpm	463	1.18	419	1.12

Table 1. Average latencies and beat shifts at various BPM's for sets of gestures 1 and 2

From this table, we can see one trend: the average latency increases linearly as the BPM increases. A second observation is that the average latency is slightly greater (about 10ms) for the arpeggio than for the single note repetition.

These results are well above what is considered as acceptable for the computer's audible reaction to gesture fixed at 10 milliseconds (ms) by [17]. The Leap motion processing time per frame being 1 ms, this high audio output latency can only be explained by the typical processing scheduling delays of Max and the limit of our current OS configuration (Mid 2012 MacBook Pro Yosemite, 2.3 GHz INtel Cored i7 with 8 GB 1000 MHz DDR3 RAM). Still, it would be possible to improve the latency problem with this configuration by modifying advanced scheduling parameters in Max, such as increasing the Poll throttle, which sets the number of events processed per servicing of the MIDI scheduler and decreasing accordingly the Queue Throttle which sets the number of events processed per servicing of low-priority event queue such as graphical operations, interface events and reading files from disk. At last, it is possible to decrease the signal vector size and the sampling rate of the sound synthesis, even though this would deteriorate the overall sound quality. Further experiments will aim at finding an optimal compromise with these parameters in order to lower the latency.

CONCLUSION

In electro-acoustic music, the composer has the desire to manipulate sounds in multiple dimensions and to transform, isolate, and remix both natural and digitally created sound objects over time. One aim of the EMI is to reduce the cognitive distance between the imaginary imagery of electro-acoustic composers and the explicitly producing gestures. Embodiment seems necessary in electro-acoustic as it is intrinsic to traditional acoustical instruments and to most people approach to music. Computer music has allowed composers to use all sorts of sounds but the mechanism to produce or trigger them often do not incorporate an adequate physical movement. Realism needs a human form to physically activate processes and to avoid robotic and impenetrable performances. Novel interfaces for musical expression, such as the instrument described here, can significantly change musicians and audiences' perspectives on electronic-based music, putting forward embodied expressions through virtuoso gestures. To our knowledge, the EMI is the first musical instrument based on gesture recognition via 3D vision sensors to put forward finger expert gestures while engaging the upper body in the performance. Its ease of use is also combined with a great potential for virtuosity. The mapping strategies show transparent relationships between gestures and musical results. The latency is currently the main issue we need to solve in order to get what Wessel and Wright designate as an *intimate musical control*.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7-ICT-2011-9) under grant agreement n 600676.

REFERENCES

1. Pierre Schaeffer. Traité des objets musicaux. 1966.
2. Luigi Russolo, Robert Filliou, Francesco Balilla Pratella, and Something Else Press. *The Art of Noise: futurist manifesto, 1913*. Something Else Press, 1967.
3. Gaël Tissot. La notion de morphologie sonore et le developpement des technologies en musiques electroacoustiques: Deux elements complementaires d'une unique esthetique? 2010.
4. Simon Emmerson. Computers and Live Electronic Music: Some Solutions, Many Problems. *International Computer Music Conference Proceedings*, 1991, 1991.
5. Joel Ryan. Some remarks on musical instrument design at steim. *Contemporary music review*, 6(1):3–17, 1991.
6. J Han and N Gold. Lessons Learned in Exploring the Leap Motion TM Sensor for Gesture-based Instrument Design. *Proceedings of the International Conference on New ...*, 2014.
7. Daniel Tormoen, Florian Thalmann, and Guerino Mazzola. The Composing Hand: Musical Creation with Leap Motion and the BigBang Rubette. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 207–212, 2014.
8. Nicolas Alessandro, Joëlle Tilmanne, Ambroise Moreau, and Antonin Puleo. AirPiano : A Multi-Touch Keyboard with Hovering Control. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 255–258, 2015.
9. Eduardo S Silva[1], Jader Anderson O de Abreu[1], Janiel Henrique P de Almeida[1], Veronica Teichrieb, and

Geber L Ramalho. A preliminary evaluation of the leap motion sensor as controller of new digital musical instruments. 2013.

10. Chris Harrison, Hrvoje Benko, and Andrew D Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450. ACM, 2011.

11. Craig Villamor, Dan Willis, and Luke Wroblewski. Touch gesture reference guide. *Touch Gesture Reference Guide*, 2010.

12. Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 139–146. ACM, 2007.

13. Lasse Thoresen and Andreas Hedman. Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer's typomorphology. *Organised Sound*, 12(02):129, jul 2007.

14. Otmar Hilliges. Interactions in the Air : Adding Further Depth to Interactive Tabletops. pages 139–148, 2009.

15. Hrvoje Benko and a Wilson. DepthTouch: Using depth-sensing camera to enable freehand interactions on and above the interactive surface. . . . *on Tabletops and Interactive Surfaces*, (March), 2009.

16. Using the Xbox Kinect sensor for positional data acquisition. *American Journal of Physics*, 81(1):71, dec 2013.

17. Adrian Freed, Amar Chaudhary, and Brian Davila. Operating systems latency measurement and analysis for sound synthesis and processing applications. In *Proceedings of the 1997 International Computer Music Conference*, pages 479–81, 1997.

18. Wrist and shoulder posture and muscle activity during touch-screen tablet use: Effects of usage configuration, tablet type, and interacting hand. *Work*, 45(1):59–71, 2013.

19. Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 1063–1072, 2014.

20. Alexander Refsum Jensenius. Microinteraction in Music / Dance Performance. *Proceedings of the International Conference on New Interfaces for Musical Expression*, (Figure 1):16–19, 2015.

21. Atau Tanaka. Musical performance practice on sensor-based instruments. *Trends in Gestural Control of Music*, 13(389-405):284, 2000.

22. Rolf Inge Godøy. Images of Sonic Objects. *Organised Sound*, 15(01):54, mar 2010.

23. F Delalande. La gestique de gould: éléments pour une sémiologie du geste musical g. *Guertin. G. Gould, ed., Courteau, Louise*, 1988.

24. Richard Graham and Brian Bridges. Managing musical complexity with embodied metaphors. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. Louisiana State University, 2015.

25. David Wessel and Matthew Wright. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*, 26(3):11–22, sep 2002.

26. John Young. Sound morphology and the articulation of structure in electroacoustic music. *Organised sound*, 9(01):7–14, 2004.

27. Jules Françoise, Ianis Lallemand, Thierry Artières, Frédéric Bevilacqua, Norbert Schnell, and Diemo Schwarz. Perspectives pour l'apprentissage interactif du couplage geste-son, may 2013.

28. Norbert Schnell, Axel Röbel, Diemo Schwarz, Geoffroy Peeters, Ricardo Borghesi, et al. *MuBu and friends–Assembling tools for content based real-time interactive audio processing in Max/MSP*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2009.

29. D Trueman and R Luke DuBois. Percolate. *URL: http://music. columbia. edu/PeRColate*, 2002.

30. Perry R Cook and Gary Scavone. The synthesis toolkit (stk). In *Proceedings of the International Computer Music Conference*, pages 164–166, 1999.

31. Jukka Rauhala, Heidi-Maria Lehtonen, and Vesa Välimäki. Toward next-generation digital keyboard instruments. *Signal Processing Magazine, IEEE*, 24(2):12–20, 2007.

32. Gerhard Eckel, Francisco Iovino, and René Caussé. Sound synthesis by physical modelling with modalys. In *Proc. International Symposium on Musical Acoustics*, pages 479–482, 1995.