# THREE-DIMENSIONAL GESTURAL CONTROLLER BASED ON EYECON MOTION CAPTURE SYSTEM

*Bertrand Merlier*

Université Lumière Lyon 2

département Musique

18, quai Claude Bernard

69365 LYON Cedex 07

FRANCE

merlier2@free.fr

## ABSTRACT

This article presents a new gestural computer interface (3DGC = Three-Dimensional Gestural Controller) allowing an instrument-like complex control. Hand gestures are analyzed by means of a camera and by a motion capture software. Contrary to classical instrument interfaces, neither any mechanical organ nor any mark is visible! This concept offers important freedom of gestural control in any 3 dimensions of space. However it requires reconsidering the relation between cause and effect, i.e. between gesture and sound result. We shall present several specific features linked to lack of any visual mark or mechanical constraints. Then we shall underline the interest of programmability and simultaneous control of several parameters in only one simple gesture.

First use of 3DGC is to control generation, transformation and spatialization of live electroacoustic music. Many other artistic applications are imaginable in the multimedia area, such as live video pictures generation and transformation.

The "4 Hands" project - an instrumental multimedia duet - uses two 3DGC for live sound and video creation. It was presented at Ars Electronica Festival (Linz - Austria) in September 2002.

### Keywords

HCI, Human Computer Interface, virtual instrument, computer-based musical instrument, gestural controller, motion capture software, electroacoustic music live performance, hand gesture analysis by means of a camera, Eyecon, MAX/MSP.

## 1. INTRODUCTION

For now 50 years on, composers and performers have new investigation fields at their disposal: timbre and space. Early electroacoustic music experiments were first carried out in studios or research laboratories with machines and without time or performance constraints. Then, instruments appear that allow performing live real-time electroacoustic music. These instruments are now fully based on computer and software.

The critical point in new instrument design is the simultaneous control of the numerous parameters related to timbre, using a simple gesture, reproducible and easy to understand for the audience [1] [5]. The piano keyboard paradigm, that of synthesizers or control surfaces with their buttons and potentiometers panoply, or even that of computer mouse and scrolling menus are very little adapted to an efficient instrumental timbre performance. All these interfaces are limited in virtuosity and not well suited for the sound complexity management: it is seldom possible to simultaneously modify more than 1 or 2 parameters.

An analysis of the functionalities of these interfaces shows that they all constraint the instrument gesture spatiality in only 1 or 2 dimensions: a mouse, a potentiometer or a joystick move over a few centimeters in a X Y plane; a push button moves over a few millimeters along only one Z axis; piano keys are set along 1 horizontal X axis and slide down of a few millimeters along a vertical Z axis.

But human hand(s) or forearm(s) can freely move in a space of about 1 m$^3$ (one cubic meter), with the same easiness and the same precision in the 3 X Y and Z dimensions. It is a pity not to use all this expressive freedom and all this virtuosity.

Building a three-dimension mechanical interface is very often complex. Using a camera motion capture system appears to be a more efficient and easier solution to adjust and run. The lack of any physical constraint, the hand(s) movement detection in the 3 dimensions of space open new horizons for the simultaneous control of the numerous parameters of timbre creation or space trajectories.

Following already established classifications (see for example [10]), our gestural interface can be classified into the Alternate Controllers, as it does not use any instrument-like mechanical organ.

In the following pages, we shall first present this new instrument, then make an inventory of its gestural potentialities (section 2). Section 3 proposes an analysis and a formalization of this new concept, especially making some comparisons with other instruments. Section 4 briefly presents a first live performance use.

Several QuickTime video sequences allow to better understand how the 3DGC works (on the Web at [11]).

## 2. INSTRUMENT DESCRIPTION

### 2.1. Hardware and software

The 3DGC allows real time generation and transformation of sounds (or moving pictures[1]) controlled by one or both hands movements in space.

---

[1] See section 4.

The motion capture is realized by means of the EyeCon system (video motion capture and multimedia production software). This commercial software - developed by Frieder Weiss [12] - is generally used as a stand alone product dedicated to interactive dance performances and multimedia environments: dancers (or audience) movements control sound or multimedia elements generation. This stage situation is really different from our instrumental gesture research.

The Eyecon software is running on a simple PC computer. It requires a video acquisition board[2]. Thus it does not consume a lot of processor resources: a 1 GHz processor is pretty sufficient just for motion capture. The computer monitor displays the hand(s) picture and the detection zones or objects (figures 1 and 2).

The hand(s) gestures are filmed by means of a B/W camera, analyzed by the software, then transmitted to the sound (or pictures) production means. We developed several MAX/MSP patchers running on a Macintosh laptop for gesture parameters management and sound production and transformations.

The instrument will be presented following two main directions: the gesture capture (input) and the control signal generation (output).



**Figure 1: 3DGC overview.**
The performer moves his hands into the camera field. The computer monitor offers a visual look back: the hand position in connection with the sensitive zones. If the sound result is obviously related to the gesture, the performer can rapidly consider forgetting to watch over the monitor.

**2.2. Hand motion capture**

According to our composition needs, we used 6 types of instrumental gestures or movements:

1) a simple triggering, i.e. presence (or absence) of the hand (or the finger) in one space area (figures 3a and 3b);

2) the X-Y position of the hand into a specific area (static mode) or its movement (dynamic mode) (figure 3c and 3d);

3) the size of the lit surface; this parameter can correspond to the closure (clenched fist) or opening (flat

position) of the hand, but also to the measurement of the distance of the hand from the camera (figure 3h and 5);

4) the movement energy or speed inside a specific area allows to add a dynamic factor to the triggering (figure 3e);

5) trajectory following is possible, as well as relative distance between both hands, thanks to a tracking function (figure 3f);

6) last, it is possible to measure the left-right symmetry degree[3], as well as the brightness, width, height or length of a shape (figure 3g);

Several occurrences of these six basic functions can of course be simultaneously used and freely combined together (figure 4), allowing really sophisticated gesture detection. We then have at our disposal an instrument with dynamic sensitive triggering sources and numerous controllers.
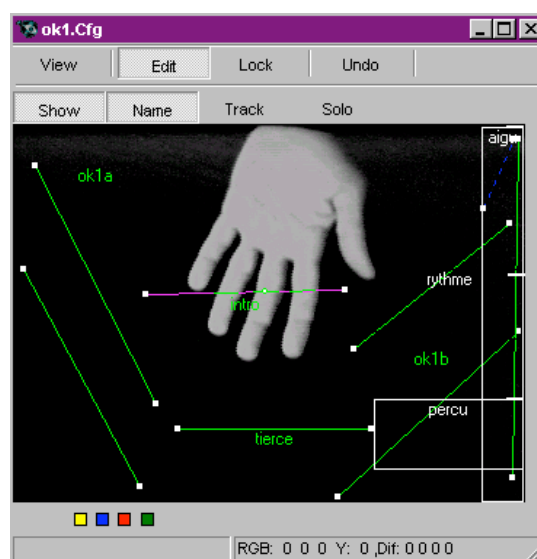


**Figure 2: The display window.**
It first allows the building of the instrumental interface and then it is used as a visual control in the performance time.

It is possible to superimpose a trigger and several controllers, so that one can trig an action with the clenched fist and then modulate it by opening the hand.[4]

Combining several controllers together allows to dispose of a kind of a three-dimensional joystick (X-Y-Z) of about 1 m3 (figures 3d + 3h = figure 4b);

The speed (or energy) detection can be superimposed with any other detection area. With MIDI keyboards, we get used to "Note ON + velocity" coupling. This can easily be done with our 3DGC. A coupling such as "controllers + velocity" is much more original and opens the door to new exploration fields.

---

[2] A new version under finalization will be DirectShow compatible. It will allow using any video device: analog, DV, WebCam…

[3] It is only a surface and symmetry factor measurement but not a real form recognition operation.

[4] In a MIDI keyboard, one needs to use both hands (key + wheel) or 1 hand plus 1 foot (key + pedal) to obtain the same result.
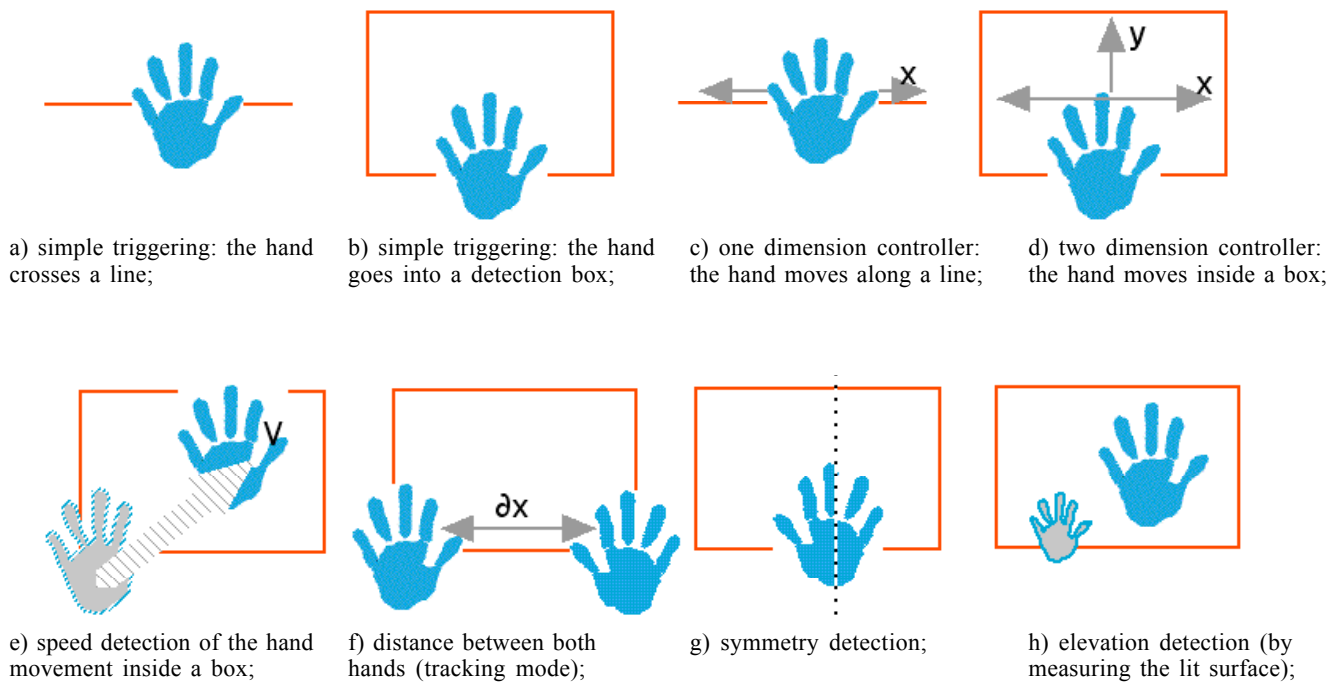
a) simple triggering: the hand crosses a line;

b) simple triggering: the hand goes into a detection box;

c) one dimension controller: the hand moves along a line;

d) two dimension controller: the hand moves inside a box;

e) speed detection of the hand movement inside a box;

f) distance between both hands (tracking mode);

g) symmetry detection;

h) elevation detection (by measuring the lit surface);

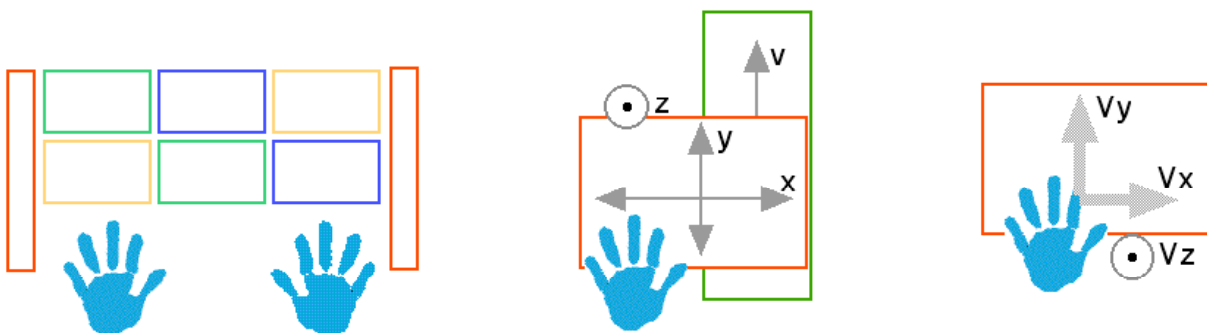**Figure 3: detection of elementary gestures with the 3DGC.**

**Figure 4: instrumental gestures with the 3DGC.**
a) events triggering by means of « on/off switches »;
b) multi-dimensional controllers (X Y Z V...) that can be superimposed or coupled together;
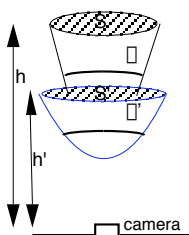c) energy detection (by measuring the speed of the movements in the 3 dimensions X Y Z).

**Figure 5: elevation measurement with the 3DGC.**
On the vertical axis (Z), the detection principle is completely different. The 3DGC measures the hand moving away from the camera by calculating the reduction of the lit surface (see figure 3h). In most cases, this way of proceeding is suitable.

## 2.3. Control signal generation

Internal functions of Eyecon software allow triggering (i.e. read) and controlling (start/end point, loops, speed...) several multimedia elements: sound files, fixed or moving digital pictures, video... For most interactive performance situations, this software can be used stand alone, if the user remains within easy triggerings and manipulations.

For more complex applications, Eyecon can emit MIDI or OSC codes. We used that solution.

## 2.4. Sound designing

The preceding description of the 3DGC instrumental interface has underlined new gestural possibilities: several triggers and controllers coupling, several controllers coupling, velocity-endowed controllers. The MIDI norm is versatile enough to allow coding these gestures without any problem. But, commercial synthesizer or sampler architectures (hardware or software based) are very often unable to efficiently manage all these multiple couplings[5].

Some gestural signals also require a few calculations: filtering, range shifting, (more or less) linear combination of 2 or 3 controllers, transforming controllers into Note On / Note Off data...

The author has developed his own sampler thanks to MAX/MSP software. So any controllers coupling constraint or limitation disappear. Considering the huge potentialities of this new gestural interface, it would have been a pity to accept limitations in sound generation or manipulation.

## 3. INSTRUMENTAL GESTURE ANALYSIS

### 3.1. The gesture spatiality

The 3DGC proposes an active volume of about 1 m$^3$ with a truncated cone shape (figure 5): the activity field yields on about 1 meter in each three directions[6], that is enormous compared to traditional acoustic instruments.

For evident physical reasons, acoustic instruments own a privileged axis (X) generally corresponding to the height control: a guitar neck measures about 90 cm, a piano keyboard measures between 1,20 and 1,50 meter. But, following Y or Z axis, movements are only of a few millimeters or centimeters.

### 3.2. instrument polyphony and gesture types

The analysis software allows a precision of about 1 centimeter (depending on the camera, the lens and the lighting). This precision would permit independent finger movement detection, but, practically speaking, this would happen to be illusory. As the performer no longer has any mechanical or visual mark, trying to reach any precise space position is not at all easy. A melodic or rhythmic play is somewhat difficult.

It seems to be much more reasonable to move away from acoustic instrumental models and to turn towards another paradigm close to an orchestra conductor. The performer gestures will be much more efficient in simple triggerings, as well as in relative movements based on morphologic shapes or energy (leading to velocity detection).

Practically speaking, it is better to consider the 3DGC as a monophonic or "biphonic" instrument (each hand controls one sound).

The specific parameter coupling of the 3DGC - velocity sensitive triggering + velocity sensitive 3D controllers - appears to be very well suited for musical writing or performance focused on timbre and nuances.

Some presentation video excerpts can be seen on the Web site: http://tc2.free.fr/4hands/photos.html [11]

### 3.3. Interface programmability

Computer programming offers the creator a huge flexibility in sound material production and manipulation. With 3DGC, a few instructions allow to immediately change the number, the disposal, the range or any other functionalities of the triggering or control zones. Programming also allows determining which instrumental gesture will be associated to such or such sound or picture, such deformation or such transformation. Intervening at the same time on the cause and effect parameters, for instance it becomes possible to define a specific instrumental "hand language" in tight concordance with the musical writing. Some "relaxing music" would rather accommodate itself with soft gestures such as caresses, when some strongly rhythmic music would better adopt rapid, percussive or jerky movements.

Working with 3DGC requires taking into account both the gesture choice, the capture device definition and the programming of the media generation device. Thus instrument designing and instrumental gesture fully enter into the creation process[7].

### 3.4. Towards new instrumental situations

The lack of any visible instrument is a relatively and even disturbing new situation for the performer as well as for the audience.

The performer loses the visual marks, the ergonomic facilities and the mechanical constraints offered by classical instruments. In exchange, he has at his disposal a great freedom of gesture. Beyond those usual performing functionalities, the gesture can bear other meanings, such as becoming the object of an esthetical preoccupation in itself[8].

The performer's gestures contribute to make sound visible or picture readable. This is especially important as these instrumental gestures produce and control media usually fixed on a support (i.e. recorded on a tape or CD or DVD): electroacoustic music or video.

---

[5] The reason is most often a limitation in simultaneous MIDI controllers!

[6] These parameters are determined by the choice of the camera, of the lens and the lighting mode.

[7] It is a paradox that computer activity allows discovering again more direct relation modes, such as those of traditional societies.

[8] The gesture esthetic is of course present in classical instruments, but the immateriality and invisibility of the instrument here emphasize this phenomenon.

### 3.5. Showing what is usually hidden

The dissociation between gesture and sound production also confers to the gestures a sense that it seldom has in traditional situations. It allows to give a straight understanding of the instrument functioning, linking the performer's intentions to the audience perception. The instrumental gesture is brought to the foreground, as the physical instrument disappeared. This situation is somewhat as if a never seen part of an instrument was revealed. But this point is the main one for understanding how such a music can be produced.

## 4. 4✋ [ 4 HANDS ]

[ 4 Hands ] is a multimedia instrumental duet using 2 3DGC. Two performers on stage are playing each their own invisible instrument. Jean-Marc DUCHENNE - video creator - performs live video sequences, cast on a main screen hanging at the rear of the stage and Bertrand MERLIER - composer - performs live electroacoustic music on 2 or 4 loudspeakers.

Four hands spin, caress, knock, scratch, stretch, describe complex arabesques, strange figures. The gestures of the two performers obviously control images and sounds, but no instrument is visible!

The picture of the two performers'hands are also projected on two lateral screens. Soft lightings may also create hand shadows on the walls or on the ceiling.

The apparent simplicity of the equipment and of the stage design, the use of intimist lightings lead the audience to specially focus on the gestures themselves and on the main elements of the multimedia discourse.

The "4 Hands" project was presented in September 2002 at Ars Electronica Festival (Linz - Autriche), and then in several European electronic arts festivals.

Some presentation elements (music, pictures and video excerpts) can be seen at: http://tc2.free.fr/4hands/ [11]

## 5. CONCLUSION

We have presented the realization of a new gesture interface dedicated to live music or multimedia production and control. The hand(s) gestures are analyzed by a camera and a motion capture software gesture. 3DGC is a virtual instrument as it is fully realized by means of computer programming and no mechanical organs are visible. The lacking of mechanical or visual feedback limits the melodic or rhythmic performances, but it favors a great interpretation freedom and a great virtuosity when working with timbre and dynamic, insofar as one accepts to move away from traditional instruments paradigms. The 3DGC gestural interface presents several similarities with classical acoustic instruments. It offers: triggers, numerous controllers, energy controlled possibilities, all linked together under a unique hand gesture.

Twenty or thirty years ago, the notion of "lutherie electronique" (electronic instrument design) appeared. Sound designing became a part of the composer or performer job.

Today, it is possible to go further on by programming the relationship between an instrumental gesture and the sound result. So, instrument designing may become a creative act.

## 6. ACKNOWLEDGMENTS

## 7. REFERENCES

[1] Arfib D., Couturier J.M., Kessous L., Verfaille V., "Mapping strategies between gesture control parameters and synthesis models parameters using perceptual spaces", *Organised Sound 7(2)*, Cambridge University Press, PP. 135-152, 2002

[2] Duchenne J-M., Merlier B., 4 Hands, "an interactive live audio & video performance", in *Realtime and Presence*, Trans Media Akademie, Dresden, 2003.

[3] Genevois H. and De Vivo R. (sous la dir. de), *Les nouveaux gestes de la musique*, Éd. Parenthèses, 1999.

[4] Kessous L. & Arfib D., "Bimanuality in Alternate Instruments", in *Proceedings of the 2003 Conference on New Instruments for Musical Expression (NIME-03)*, pp. 140-145, Montreal, Canada.

[5] Merlier B., New instruments for playing and spatializing electroacoustic music, Symposium: "Medienkunst - Verknüpfung der Sinne" 27 Nov 98, COMTECart'98 - DRESDEN - RFA

[6] Merlier B., À la conquête de l'espace, *Actes des JIM 1998 (Journées d'Informatique Musicale)*, p. D1-1 à 9, publications du CNRS-LMA, n° 148, MARSEILLE, ISBN : 1159-0947

[7] Vinet H. and Delalande F. (sous la dir. de), *Interfaces homme-machine and creation musicale*, Éd. Hermès, 1999.

[8] Wanderley M. and Orio N., Evaluation of input devices for musical expression: borrowing tools from HCI, *CMJ*, Vol. 26, N°3, Fall 2002.

[9] Wanderley M. and Battier M., *Trends in gestural control of music*, CD-ROM, IRCAM - Centre Georges Pompidou, 2000.

[10] Wanderley M., "Performer-Instrument Interaction: Application to gestural Control of Music", PhD Thesis, Paris, University Pierre and Marie Curie - Paris VI, 2001. http://www.ircam.fr/wanderley/Thesis/Thesis_comp.pdf

[11] Web site URL of "4 Hands" project: http://tc2.free.fr/4hands/

[12] Web site URL of Eyecon and Palindrome: http://www.palindrome.de/pps.htm

[13] Web site URL of Thélème Contemporain: http://tc2.free.fr/

[14] Web site URL of TransMedia Akademie: http://www.tma.de