3D Gesture Acquisition Using Ultrasonic Sensors

Emmanuel Fléty
Ircam - Centre Pompidou
flety@ircam.fr

Abstract

Research and musical creation with gestural-oriented interfaces have recently seen a renewal of interest and activity at Ircam. In the course of several musical projects, issued either by from young composers enrolled in the Cursus of Composition and Computer Music or by guests artists, the Pedagogy and Creation departments of Ircam have proposed various solutions for gesture-controlled sound synthesis and processing. In this article, electronic engineer Emmanuel Fléty describes the technical aspects of a specific gestural device designed for acoustic and electronic percussion music. After a brief introduction to the musical context (for further information, see the Roland Auzet article in this issue), the making of a prototype based on ultrasonics technology is presented.

Musical Context

The idea of that new gestural controller was born with a musical project by Roland Auzet called *Le cirque du tambour* ("The Drum Circus"). The percussionist wanted to have a sound controlling system supplied with his own gestures, without any constraint for the instrumental playing. In fact, It was out of the question to think up a gestural acquisition device that would influence somehow the original instrumental playing of the instrumentalist. It meant, in the one hand, to respect the gesture of the percussionist – essentially the mobility of his arms – and in the other hand, not to alter the musical response of the instrument through the instrumental gesture.

With that gestural controller, Roland Auzet wished to control some synthesis and sound processing parameters and also spatialization parameters. The electronic sound material resulting from the synthesis and the processes could have as origin either the acoustic percussion, real-time sampled or pre-recorded sound samples stored on the hard disk of the computer.

After Roland Auzet had established the main line of his musical work[1], we determined with his help the specifications of that new gestural controller. The outcome of this brainstorming session was that the motion capture of one of his hand in a cubic volume of 80 cm by side was appropriate for the musical control that was expected.

Technical Context

We started the development of that new gestural controller by the study of the different methods for 3D motion capture. We listed the following technologies :

- infrared detection;
- ultrasonic ranging;
- magnetic field sensors;
- geopositioning (GPS / Global Positioning System);
- video shape and color recognition.

1. See the article by Roland Auzet in this volume.

Reprint from : *Trends in Gestural Control of Music*, M.M. Wanderley and M. Battier, eds.

Generalities

Before talking about the subject of the technologies for 3D motion capture and gestural measurement, we will first introduce the different methods and algorithms allowing that kind of measurement. In fact, one method can be used with different technologies, as we will see it in the following sections.

Single detection

In that case we just want to detect a movement, without trying to quantify it by its amplitude, its direction, its speed or its acceleration. The method generally used in this situation is to set up a physical or mechanical constant in the environment in which the motion will occur, and then to detect a perturbation or a modification of the constant related with to movement. For instance, it's the way garden automatic lighting systems work, the motion detection being triggered by passive infrared generated by human body activity.

Fig. 1. Passive infrared sensor.

Detection with distinguishing criteria

The idea to detect a movement with a special characteristic, for instance speed or direction. The detection method should only react to that criteria. Two implementations can be used :

- the method can detect several types movements, but with additional measures, it's enable to ignore the movements that don't validate the detection predicate;
- the method is only sensitive to the movements that validate the predicate. To illustrate the case, which is the most frequent, we will think about the light barrier (visible or infrared) of the underground garages automatic doors, which detects the motion of a vehicle in a particular direction, by cutting the ray of light.

Motion measurement through a distinguishing direction

It's the easiest kind of measurement to complete, because most of the sensor technologies are sensitive only with certain directions (unidirectional sensors). For instance, to measure the speed of a unidirectional movement, we can set two light barriers on the route and calculate the time running out between the cut of the first barrier and the cut of the second one. Knowing with accuracy the distance separating the two barriers, we can calculate the speed with the following mathematical expression :

$$speed = \frac{\Delta d}{\Delta t}$$

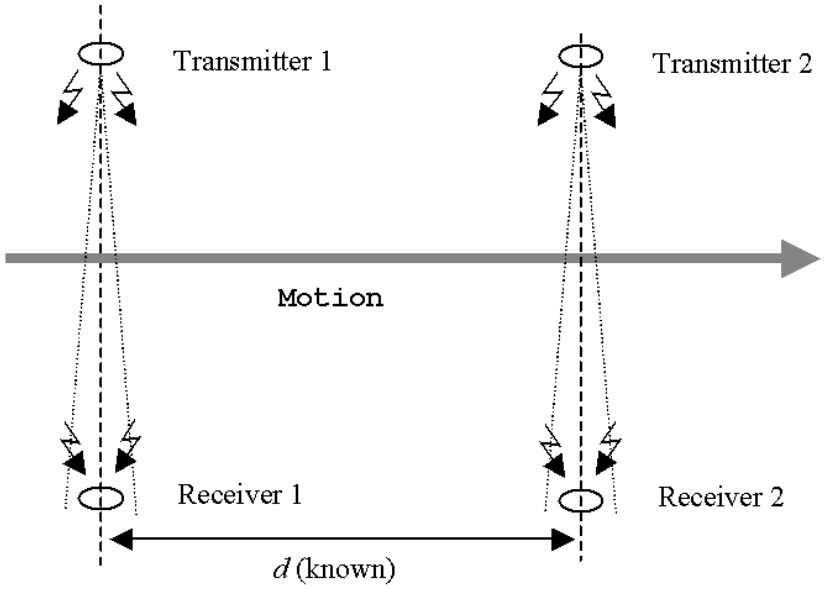

Fig. 2. Speed measurement on a unidirectional motion.

3D Movement measurement

It's the most complicated case, and the one we wish to implement for the gestural controller developed for Roland Auzet. This kind of measurement is difficult because it needs capture in the three dimensions : as mentioned in the section above, the unidirectional detection aspect of the sensor does not make it easy.

The method generally used for movement measurement in 2D or 3D is the *triangulation method*. It's also the method we choose to implement for our gestural controller. The triangulation method was originally used to obtain the geographical position of a radio transmitter. By disposing three radio receivers around the source, and by measuring the signal power for each receiver, it's possible to obtain an accurate position of the transmitter. In this context, the sensors directivity is less a problem, because it's quite easy to build omnidirectional antennas. Unfortunately, with other capture technologies, the sensors directivity characteristic is a real problem because It reduces the detection space to the solid angle of the sensor. In the following sections, we will detail the triangulation method and it's implementation for the chosen sensor technology.

Triangulation method

The triangulation method theory described in this section concerns the position acquisition in three dimensions. As we will see it, the 3D case can easily be restrained to the 2D case.

Generally speaking, the idea is to have three georeferenced points (points with known coordinates in a given coordinates system) forming a marker in space. We will demonstrate that, knowing the three distances separating the mobile point we want to obtain the position and the three georeferenced points, we can calculate the coordinates of the point[2].

2. In our gestural controller, It is the hand of the percussionist.

To measure these distances, that will be noted d_1, d_2 and d_3, several capture ways can be used. Usually, distance measuring uses a physical or mechanical property of either the sensor technology or the environment, the property varying quantitatively with the distance. This property can establish a relationship between the distance and, for instance, an electric current intensity, an electric potential difference, or a phase difference.

The law giving the relationship between the physical property and the distance is usually given by the data sheet of the sensor used, by a mathematical expression, or a graphical representation.

In order to complete measures independent from the environment in which the measurement is done, the distance measure generally uses a transmitter / receiver architecture. The aim of the transmitter is to generate a constant physical phenomenon, punctual if possible, so that it can be used as a reference. Then, the receiver measures the physical phenomenon generated by the transceiver, the measure varying with the distance between the transceiver and the receiver.

Two topologies are possible for distance measurement :

- One transmitter radiates towards several receivers. In this case, the transmitter is the moving point whose position we want to know, and the receivers are the georeferenced points.

- Several transmitters radiate towards one receiver. In this case, the receiver is the moving point, and the transmitters are the georeferenced points. That topology differs from the previous one by the fact that the receiver must be able to distinguish the source of the different signals received, with for instance a coding or frequency predicate.

These two methods work the same way : the aim is to obtain the three distances between the moving point and the georeferenced points. For our gestural controller we chose the first topology. The transmitters are linked to the wrist of the percussionist, thanks to a glove, while the receivers are fixed to a metallic structure above him.

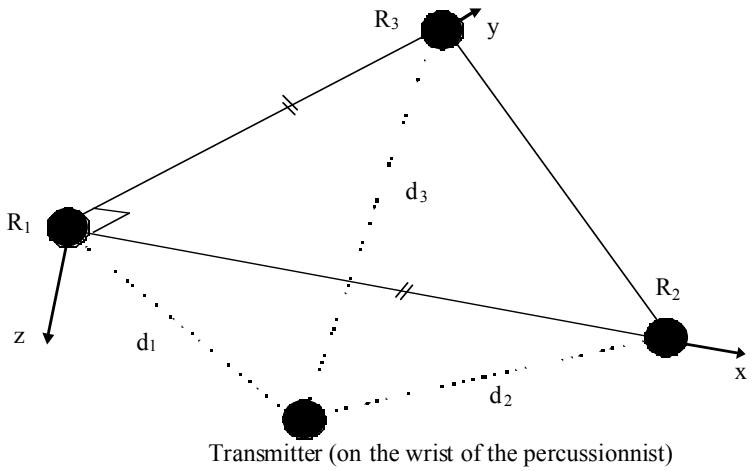

Fig. 3. Implemented topology in our gestural controller.

To make the triangulation calculation easier, the receivers are arranged in an isosceles triangle shape. Thus, the vertex with the right angle is the origin of an orthonormal marker. Such a marker makes the locating of the point easier because the coordinates calculated in an orthonormal marker are cartesian coordinates. Thus, the position of the point is given by an abscissa, an ordinate and an altitude {x,y,z} in the local marker shaped by the triangle composed of the three georeferenced receivers.

Comparison of different technologies

Infrared transceivers (transmitter / receiver)

Infrared light is quite a good technology for the triangulation method, except that the transmitters and the receivers are very directive. However they are invisible and their range can be up to ten straight meters. Unfortunately, they have two major drawbacks that forced us to avoid using them for our gestural controller :

- Infrared receivers are especially sensitive to EMI, notably generated by high brightness spotlights, often used in live performances.
- The distance measurements needed by the triangulation calculation use a complicated calibration. In fact, the mathematical relationship giving the distance between a transmitter and a receiver uses the intensity of received infrared light. This relationship, usually given by the product data sheet, is seldom linear, and logarithmic in most of the cases, which complicates the measurement.

Magnetic sensors

Those sensors can quantify magnetic field disruption, which make it possible to establish the distance between a transmitter and a receiver. The major drawback is that these sensors have a really short range (a few centimeters). Magnetic sensors with longer range exist but they are very expensive.

GPS (Global Positioning System)

That device makes it possible to establish the absolute position of a GPS receiver on the Earth ground. That positioning is completed with a triangulation method, using a minimum of three satellites. Here again, several drawbacks forced us to avoid using that technology :

- Common GPSs present a resolution of tens meters. Specific GPSs have a higher resolution (a few mm), but again they are very expensive.
- The refresh period of the GPS is not compatible with real-time sound controlling motion capture. In fact, the refresh period is about a second, sometimes more, while we need a refresh period under 10 milliseconds.
- GPS needs to receive radio signals from satellites : it's only possible when the GPS receiver is outdoors, without any obstacles, so it's impossible to use it indoors.

Video motion capture

This method does not use the triangulation method. Using one or more video cameras and a shape or color recognition algorithm, it's possible to quantify the 2D or 3D motion of a point. Nevertheless, shape recognition algorithms consume a huge amount of CPU time. The color recognition is attractive, but it is very sensitive to the ambient lighting, as well as the shape recognition, and thus was incompatible with the stage performance imagining by the composer, who would like to use many modulated spotlights.

Ultrasonic ranging

In spite of some drawbacks in this technology, it is the one we chose for our gestural controller, especially for the low cost aspect and the easiness of development and implementation.

A solution for the 3D motion capture of the hand, with a light device, is the ultrasonic ranging based on a transmitter / receiver architecture. As mentioned in section 2., the ultrasonic transmitters are linked to the wrist, thanks to a glove worn by the instrumentalist. The receivers are fixed to a metallic structure placed above him.

The development of our device follows on from the implementation of ultrasonic sonars, designed for the musical work composed by Lucia Ronchetti, during 1996-1997 Ircam Computer Science and Composition Course[3]. These sonars were used to measure distances between a mobile VCR and an instrumentalist.

The working principle of ultrasonic sonars uses the propagation speed of acoustic wave in order to measure the distance between a ultrasound transceiver and a ultrasound receiver. Using this principle, it's possible to measure the three distances d_1, d_2, d_3 used by the triangulation method. In the musical context, we use 40 kHz ultrasonic waves. In fact, these acoustic waves cannot be heard by human beings and do not affect musical sound sources that can exist where the device is used.

Measuring the distance is completed by sending a burst of ultrasonic impulsions during a duration rather short compared with the propagation time of the wave[4], and then by measuring the time elapsed between the emission and the reception. Distance is then given by the multiplication of this time by the wave speed. That operation is executed simultaneously by the three receivers in order to measure the three distances.

To make a connection with the triangulation method theory previously described, the physical property used is the propagation speed of acoustic waves and the measure completed is a time differential measure the reference of which is set by the emission of a burst of ultrasonics

Implementation of the triangulation method

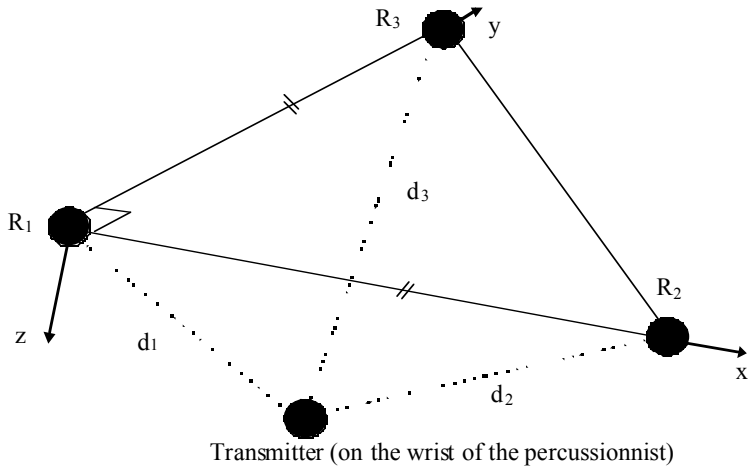

Fig. 4.

The system described measures the distances separating a ultrasonic transmitter from three ultrasonics receivers forming an isosceles triangle.

Knowing the three distances d_1, d_2 and d_3, it's possible to calculate the x, y, z coordinates of the transmitter in the marker determined by the isosceles triangle. The intersection of the two edges which have the same length is the origin of the marker (Receiver R1 on the previous figure). This marker thus composed is orthonormal, which ensures a cubic detection volume.

Calculation of the abscissa of the moving point

The calculation of the abscissa uses the d_1 and d_2 distances. We thus consider the plane containing the points T (Transmitter), R_1 (receiver 1) and R_2 (receiver 2).

3. *Cursus de composition et d'informatique musicale*, a one-year long course organized by the Pedagogy Department of Ircam. The piece by Lucia Ronchetti is *Éluvion-Étude*.
4. And so with the measured distance.

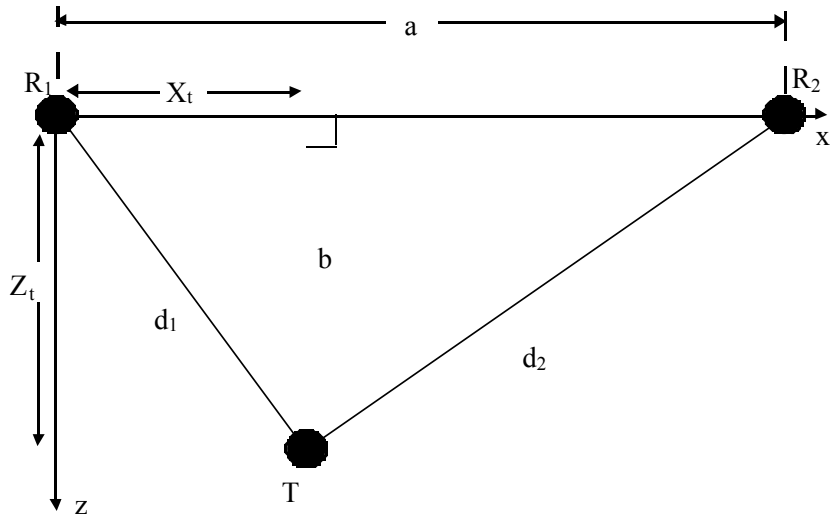

Fig. 5. Calculation of the abscissa of the moving point.

Projecting orthogonally the point T on the straight line (R_1R_2), we get two right-angled triangles on which we can apply the Pythagorean theorem and its converse.

$$\begin{cases} d_1^{\,2} = X_t^{\,2} + b^2 \\ d_2^{\,2} = (a - X_t)^2 + b^2 \end{cases} \Leftrightarrow \begin{cases} b^2 = d_1^{\,2} - X_t^{\,2} \\ b^2 = d_2^{\,2} - (a - X_t)^2 \end{cases} \Leftrightarrow \left\{ d_1^2 - X_t^2 = d_2^2 - (a - X_t)^2 \right.$$

Hence

$$\boxed{X_t = \frac{d_1^{\,2} - d_2^{\,2}}{2a} + \frac{a}{2}}$$

Calculation of the ordinate of the moving point

This time, we consider the plane containing the points T (Transmitter), R_1 (receiver 1) and R_3 (receiver 3). The calculation uses the d_1 and d_3 distances.

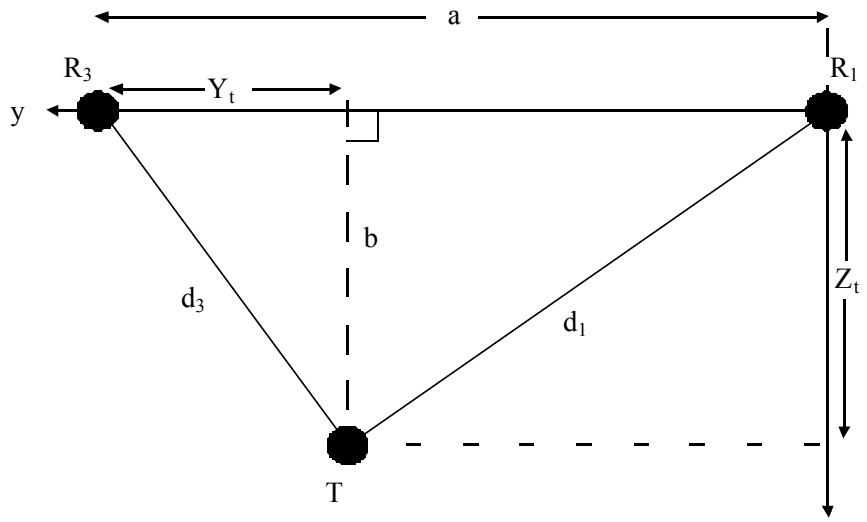

Fig. 6. Calculation of the ordinate of the moving point.

Permuting the subscripts in the previous expressions, we easily obtain :

$$Y_t = \frac{d_1^{\,2} - d_3^{\,2}}{2a} + \frac{a}{2}$$

Calculation of the altitude of the moving point

This calculation is optional when we only want a 2D motion capture, that's to say only in the plane $\{R_1, R_2, R_3\}$.

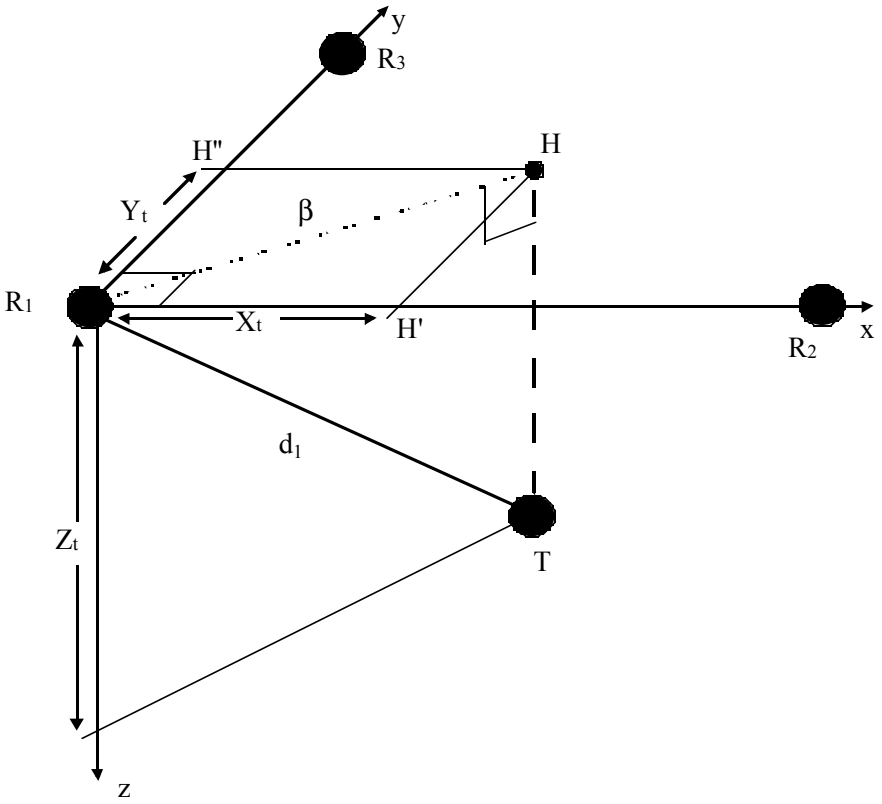

Fig. 7. Calculation of the altitude of the moving point.

The calculation of the altitude uses a quadratic combination of Xt and Yt. In fact, using the Pythagorean theorem on the right-angled triangles {R_1,H,T} and {R_1,H',H}, we get :

$$Zt^2 = d_1^2 - \beta^2$$
$$Zt^2 = d_1^2 - Xt^2 - Yt^2 \qquad \text{Hence} \qquad \boxed{Zt = \sqrt{d_1^2 - Xt^2 - Yt^2}}$$

Global synoptic

In order to implement the triangulation method, we build a electronic device composed of a transmitter and three ultrasound receivers. The following figure highlights the global synoptic of this device :

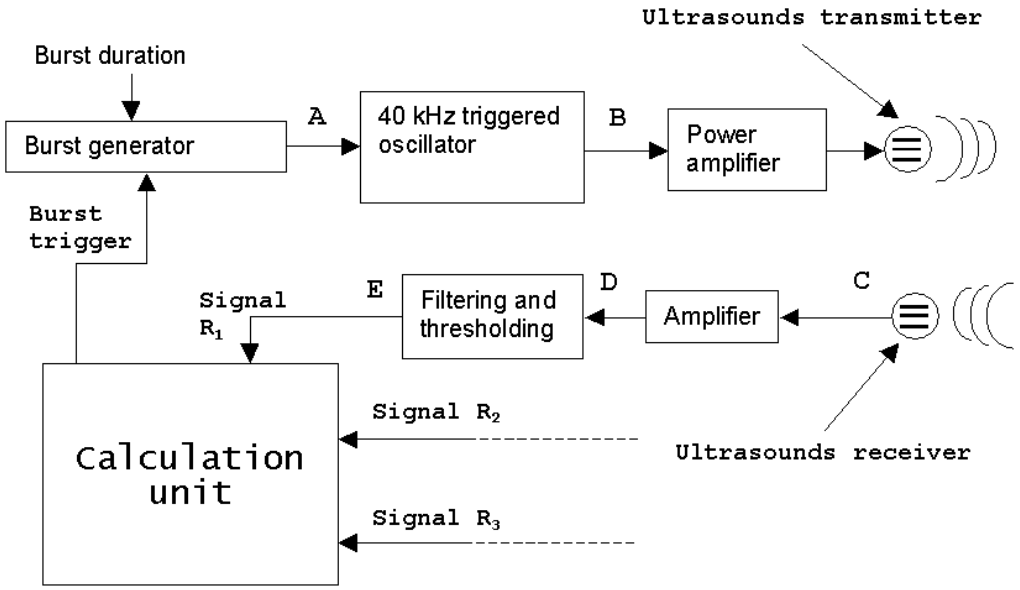

Fig. 8. Global synoptic of the electronic device.

The ultrasound receiver and amplifier section is only represented once on the previous schematic, but it's actually built for each ultrasound receiver used as georeferenced points for the triangulation method.

The role of the calculation unit is measuring the elapsed time between the moment when the burst is triggered and the moment when the ultrasound receivers receive the signal after propagation (signals R_1, R_2 and R_3).

Signals timings and waveforms

The following figure details the timings and the waveforms of different signals in the triangulation system.

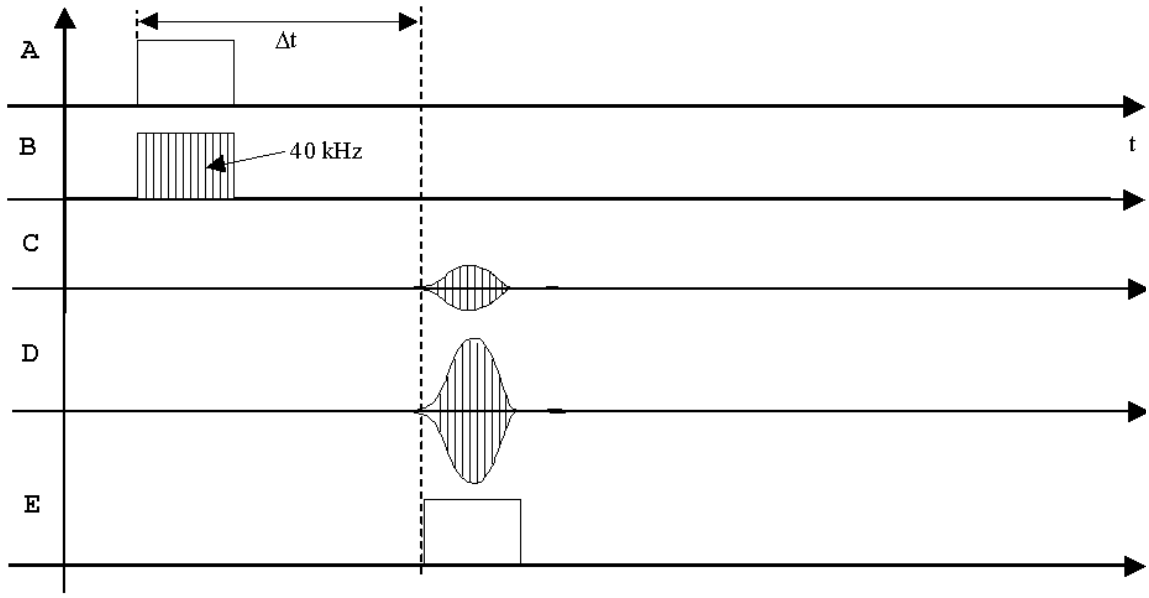

Fig. 9. Signals timings and waveforms.

Gestural controller specifications

Limit of the spatial coordinates x, y, z and the distances d_1, d_2, d_3

After a reflection session with the instrumentalist about the amplitude of his gesture, we finally get that a eighty centimeters edged cubic detection volume was sufficient for the gestural control wished. Therefore, each cartesian coordinates will move between 0 and a (a = 80 centimeters). Knowing this, we can establish the upper boundary of the three distances d1, d2 and d3 : It is the distance a = a. $\sqrt{3}$, i.e. the cube diagonal (a. $\sqrt{3}$ = 138 cm).

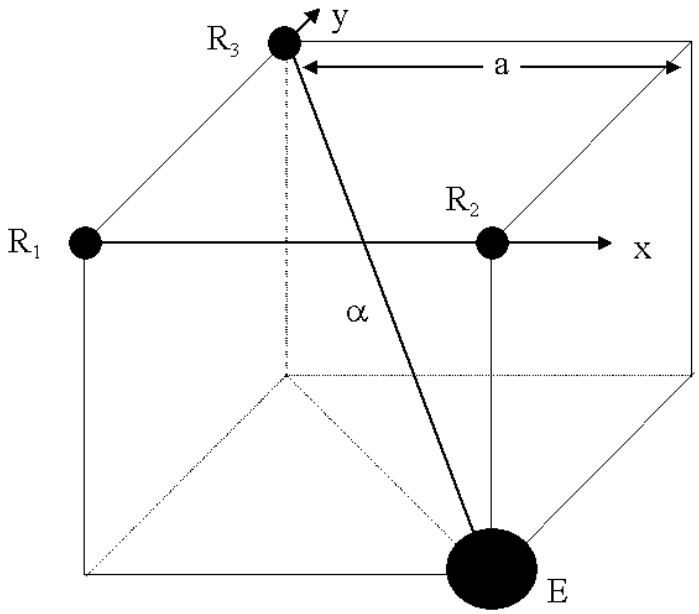

Fig. 10. Highlighting the maximum distance to measure.

Digital coding of the spatial coordinates

The cartesian coordinates will be coded on seven bits maximum, because we wanted to encapsulate them in a MIDI control change value[5].

Considering that the cartesian coordinates move between 0 and ***a***, we associate to ***a*** the maximum value that can be coded on seven bits.

$$2^7 - 1 \Leftrightarrow a$$
$$2^7 - 1 \Leftrightarrow 80cm$$

5. The MIDI standard allows to have data coded on a maximum of seven bits, the eighth bit being used to know the content of the byte (status or data).

$$\begin{cases} 2^7 - 1 \Leftrightarrow a \\ 2^x - 1 \Leftrightarrow a.\sqrt{3} \end{cases} \quad \Rightarrow \; 2^x = \sqrt{3} \,.\, (2^7 - 1) \quad \Rightarrow x.\ln(2) = \ln(\sqrt{3}.\,(2^7 - 1))$$

$$\boxed{x \approx 7.7 \text{ bits.}}$$

However, the number of bit x must be an integer. So, the distances d_1, d_2 and d_3 will be coded on the ceiling integer greater than 7.7, i.e. 8 bits.

Actual detection volume

Coding the cartesian coordinates on 8 bits, they have as maximum value :

$$xmax,\ ymax,\ zmax \ = \ \text{a} \ \Leftrightarrow \frac{2^8}{\sqrt{3}} \approx 147$$

As the MIDI data must be coded on seven bits (0-127), the value 147 is still to high. Actually, we should have scaled the cartesian coordinates x, y and z by dividing them by the ratio 147 / 127. This would have imposed us to implement a fixed point dividing routine, which was quite difficult to complete, due to the RISC architecture of our calculation unit[6]. We chose to shorten to coordinates to 127. This has as consequence a dead zone in the detection volume of :

$$\frac{(147 - 127).a}{147} \approx 10\ cm$$

The detection volume is so shorten to a seventy centimeters edged cube, which still was sufficient according to the percussionist.

Anyway, the dimensions of the cube are floating because the propagation speed of acoustic waves changes with temperature.

Device resolution

Resolution of the d_1, d_2, d_3 distances

As α is coded on 8 bits, we can calculate the spatial resolution of the device :

$$res_{d1,d2,d3} = \frac{138.10^{-2}}{2^8} \approx 5{,}40\,.\,10^{-3}\ m \approx 5\ mm$$

The propagation speed of acoustic wave is about 330 $m.s^{-1}$ at ambient temperature (20-25°C). This physical data allows us to calculate how long the sound take to cover the shortest distance that the device can measure (i.e. the spatial resolution) :

$$\Delta t_{\min} = \frac{res_{d1,d2,d3}}{330} = 16\mu s$$

6. RISC : Reduced Instruction Set Computer. In fact, our calculation unit does not own neither the multiplication operator, nor the division one : we emulated these two operators with some hand written code, but only for integers.

Resolution of the x, y, z coordinates

The actual detection cube has an edged of 70 cm : the cartesian coordinates x, y and z will move between 0 and 70 cm, being coded on 7 bits. Thus, the resolution of the cartesian coordinates are :

$$res_{x,y,z} = \frac{70.10^{-2}}{2^7} \approx 5.46\ .\ 10^{-3} m \approx 5.5mm$$

Filtering and measurement error correction

In spite of the use of several ultrasound receivers for each vertex of the triangle forming the orthonormal markers of our system, it is impossible to ensure a systematic reception by the receiver because of the quick attenuation of ultrasonic signals with the distance, and also because casual contact losses between transmitter and receivers when it goes out of the detection volume. Those contact losses can generate measurement artefacts and so wrong cartesian coordinates. To avoid such a problem, we took three precautions :

An overflow detection has been implemented in the program of our calculation unit. If the ultrasound burst is not received, or received to late, the distance measurement is invalidated.

- The calculation of Xt needs the d_1 and d_2 distances. If one of these two distances has been invalidated, Xt will not be updated. It will be the same for Yt but with the d_1 and d_3 distances. As for Zt, Its calculation needs Xt and Yt (and so d_1, d_2 and d_3). If one of these three distances has been invalidated, Zt will not be updated. Thus, we avoid the generation of wrong cartesian coordinates.

- Some measures can be skewed by parasite reflection on surrounding objects. Of course, It's advisable to place the device as far away as possible from walls and glasses. However, parasite reflections are impossible to avoid but, after several tries, they seem to be punctual. Nevertheless, we implemented a low-pass filter (moving average on four values) for each cartesian coordinates.

The filtering of course generate a latency on the variation of the coordinates, but that reduces the stepping effect on the values. In fact, as the system is based on the propagation speed of acoustic waves, the duration of a measure is not negligible : the measurement of the maximum distance (a. $\sqrt{3}$ = 138 cm) takes about 4 ms. We add to this duration a security delay of 10 ms in order to wait for the parasite reflection to disappear before making another measure. We reach then an efficiency of 50 to 100 measures per second, which is not enough for gestural acquisition (we would like about ten times more). With this low sampling rate, a quick move will generate a stepping effect on the cartesian coordinates, from a measure to the next. The filtering smooths the values, and ensures a better continuity of the data flow.

Analog inputs

Our calculation unit has several analog inputs that can converted into digital values thanks to the embedded A/D converter. Two inputs are use to convert the mechanical pressure of FSR sensors (FSR : Force Sensitive Resistance ; the more the mechanical pressure is, the less is the ohmic resistance). The first FSR is attached to the palm of the glove that owns the ultrasound transmitters (right hand). The output voltage of the sensor is converted into a digital value, then goes through a threshold logic. The goal of that sensor is the On / Off command of the gestural controller. Each time the sensor is pressed and released, the device changes its state (On/Off – Off/On). This allows starting and stopping the MIDI messages generation when the incrementalist know that he is about the move his arm with a critical gesture that can make a contact loss between the transceivers and the receivers[7]. It also allows the user to freeze the coordinates when he wants it.

The other FSR, attached to the palm of the other glove (left hand), generates, after A/D conversion and MIDI adaptation, a control change message proportional to the mechanical pressure applied to the sensor which gives another dimension to the gestural acquisition.

A third analog input is also accessible on the device. It can be used to convert into a MIDI message (again a control change message) a voltage between 0 and 5 volts. In the project of Roland Auzet, this input was connected to and external sonar measuring a distance between itself and an obstacle.

7. Leaving the detection volume or excessive wrist rotation.

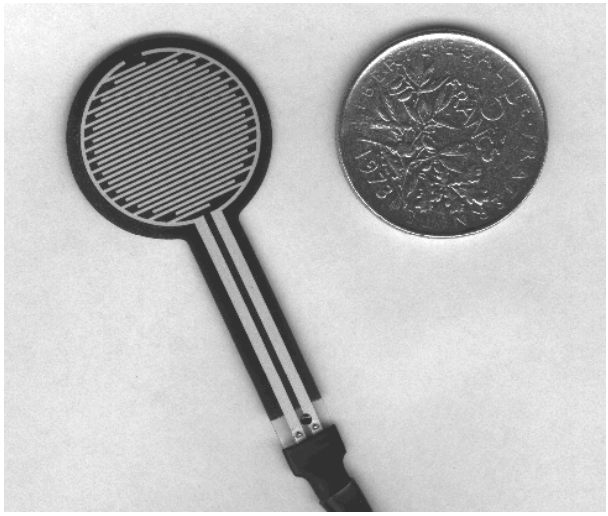

Fig. 11. Force Sensitive Resistance.

MIDI conversion

The MIDI interface ensuring the conversion of the cartesian coordinates into MIDI data is embedded in our device. The messages are serialized by an asynchronous serial unit cadenced to the MIDI baudrate of 31.25 kbits/s, then converted into current signals, in accordance to the MIDI standard.

The MIDI message giving the coordinates are control change messages. From a base address, the calculation units sets controller numbers associated to each field to transmit.

Example :

```
Base address : controller 64
Xt : controller 64
Yt : controller 65
Zt : controller 66
FSR1 : controller 67
FSR2 : controller 68
Aux. input : controller 69
```

The base address and the MIDI channel can be selected with two small coding wheels, on the side of the box containing the measurement device.

Example of application

The 3D tracker was successfully tested for basic operations before Roland Auzet used it as a composition and performance tool, in order to ensure the data flow was really continuous, without too much measurement errors. To complete that task, we created a very simple MAX patch, as shown on the next figure.

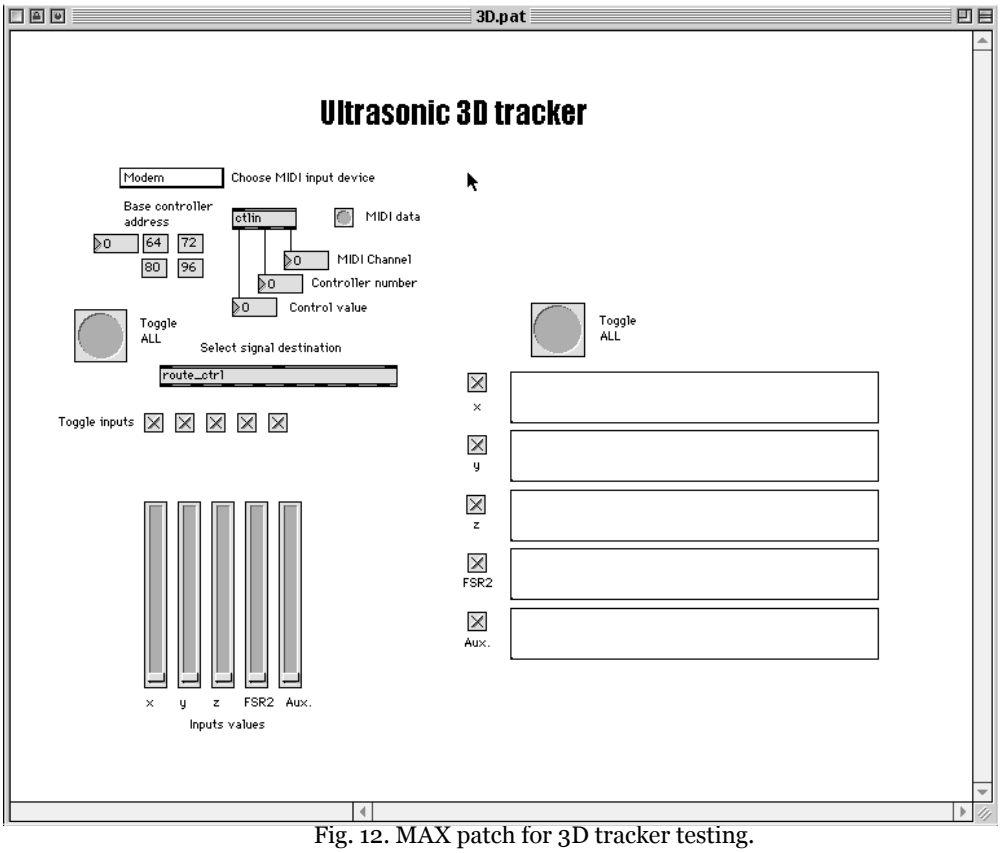

Fig. 12. MAX patch for 3D tracker testing.

By moving the glove in the detection cube, we checked the accuracy of the device and the system's response to quick movements to finally see that the behavior of the system is satisfying. This was next confirmed by Roland Auzet when he used the device for his work.

Conclusion

To conclude on the triangulation method using ultrasonics, we would like to highlight that the main drawback of this sensor technology is the directivity of the ultrasonic transceivers. We succeeded in the conception of a $1m^3$ space movement detection and tracking, but we hardly think that kind of system can be extended to larger working volumes without significant performance losses in this kind of music application.

Nevertheless, this project provided us a new direction for future studies about gesture capture and gestural interfaces in an instrumental musical performance.

