Proceedings

M.C.P.A. '97

2nd INTERNATIONAL WORKSHOP on MECHATRONICAL COMPUTER SYSTEMS FOR PERCEPTION AND ACTION

Edited by:
Giorgio Buttazzo
Ettore Ricciardi

Scuola Superiore di Studi Universitari S.Anna
Pisa, Italy
February 10-12, 1997

The Twin Towers: a Remote Sensing Device for Controlling Live-interactive Computer Music

Leonello Tarabella, Graziano Bertini, Tommaso Sabbatini

Computer Music Laboratory of
CNUCE/C.N.R. - Pisa, Italy
via S.Maria 36 - 56126 Pisa
Tel. +39-50-593276 Fax +39-50-904052
email: L.Tarabella@cnuce.cnr.it
http://spcons.cnuce.cnr.it/music/cmd.html

Abstract

In the Computer Music activity area, the design of original man-machine interfaces for controlling interactive live computer music performances is taken into consideration. In this context we designed and carried out the TwinTowers device which detects information from movements of the hands with no kind of physical connection with them: it consists of two sets of four infra-red sensing devices which create two zones of the space (the vertical edges of two square-based parallelepipedon, or towers) where the hands can be detected in terms of distance and front and side rotations with respect to a reference frame, using the values of the four heights of each tower. The system implements a sort of double aerial three-dimensional joy sticks.

Originally developed for on-stage musical performances, the Twin-towers can be considered as a general purpose remote sensing device to be used in a wide range of applications such as games, mouse-like operations and 3-D pointing device for CAD-like applications.

1 Introduction

One of the main concern in contemporary electronic music is the synchronization between musicians and computers, and the possibility of affecting with gesture the synthesized music in order to regain human feeling and sensitivity during the live performance [1]. In the computer music area of research and application, in the last decades many different kinds of special devices have been developed: in this context, the main trend is toward the realization of systems able to detect as much as possible information from the movements of human body or parts of it, tipically the hands.

Strictly speaking, playing a musical instrument entails activating and co-ordinating a specific set of muscles which in turn activates those parts of the body in contact with that instrument. To be sure, making music is something more than merely activating sound generators; infact, man-instrument interaction implies the existence of a global unity which starts at the deepest levels of will and creativity and leads to a set of bio-mechanic events which, transferred to a musical instrument, determines the final musical result.
In traditional music, whatever the genre, playing a piece of music means generating sounds with suitable tools controlled by one or more parts of the body: mouth, arms, hands, feet. This is what is commonly said to be "playing a musical instrument" or even "playing notes with a musical instrument" which basically means making the right actions at the right times.

The way to play a note, i.e. selecting a note and giving expression to that note, may be either *integrated* or *distributed* depending on the instrument used. When playing the piano, each finger executes in an integrated manner both the task of selecting a note and of giving it expression, while with the guitar and the strings these tasks are, basically, distributed between the two hands. In brass instruments fingers contribute to the selection of the note which is mainly determined by the lips which also control the expression; in reed instruments pitch is mainly established by an open/closed holes configuration given by the fingers of both hands, and expression is controlled by the oral and breathing systems.

In the piano and in the other claviers, much more than in wind and string instruments, it is possible to consider two well defined and separated parts: the controller, i.e. the keyboard, and the sound generator, i.e. the harp. When analog electronics was replaced by digital electronics, the functionalities were greatly improved both from the point of view of control and of sound generation, but the same architecture was observed. In fact, current commercial digital keyboards consist of two main parts: the keyboard itself, i.e. the controller, and the sound generator circuitry, i.e. the synthesizer. When in the early '80s the first digital keyboards were built, it was soon realised that it was possible to carry out external synthesizers controlled via MIDI which expanded the timbric capabilities and thus commonly called *expanders.* MIDI stands for Musical Instruments Digital Interface and in worldwide the standard of communication in the realm of Computer Music [2]: the sigle stands for both the physical interface (a start-stop serial communication at 31250 baud) and for the real-time language of commands which regard activation of sounds, changing of sound parameters such as timbre, volume, left-right panning etc.

A number of mute-keyboards (i.e. pure keyboards without the synthesizer part which only issued MIDI messages) and other controllers such as MIDI-saxes, MIDI-guitars, MIDI-trumpets, drum pads, etc., as been so far built. Although there are currently a great number of MIDI products, it is possible to simplify the matter by saying that there are a number of controllers which can get as much information as possible from the gestures of a performer and transform them into MIDI messages; and a number of high quality expanders which can generate and process sound in accordance with MIDI messages coming from controllers, commonly used with the configuration *controller -> synthesizer.*

When three decades ago the activities on computer music began, in a wide sense the typical configuration used was: *computer -> synthesizer*; researchers focused on implementing strategies and/or special devices for producing sound, and made great efforts in order to develop specific musical languages to tackle the compositional activity from the micro level of sound (timbre, effects, spatial movements, etc.) until the macro level of music events. This approach to composition was in some way inherited from the one previously in use in analog electronic studios where the audio tape was considered to be the final product.

Later, when computers became more and more powerful and available at a personal level, and MIDI spread all over the world, the computer music people realised they could regain control with gesture of the overall musical result and started to use a new configuration of the elements involved: *controller -> computer -> synthesizer.* The term *interaction* was re-introduced, but with a new meaning which referred to the activity practised on an musical instrument which always responds in the same manner at the same stimula; from that moment on the term *interaction* has referred to the activity of a performer on a controller which detects gesture and translates them into data that a computer uses for controlling synthesizers.

The computer thus assumes a new role during the performance, as it generates events in accordance with predefined music/acoustic material combined with information from the controllers a performer is acting on. So, the simple one-to-one mapping rule valid for traditional instruments leaves room for a theoretically infinite range of mapping rules definable by the composer for a specific piece and even for each part of that piece. The mapping as part of the composition.

1.1 Paradigm of interactive computer music

For defining the mapping, i.e. the composition, many music language has been developed and also proposed on the market: MAX [3] is with no doubt the language which has had the best success for it graphic approach to design patches which links Midi message regarding manual activity of a performer to sound synthesizers. Other languages are much more algorithmic in traditional sense and allows what in computer music is known as "algorithmic compositional approach" that is the composition of a piece of music in terms of processes which define the behaviour of the various part and the various aspect of music: melody, timbre, spatial movement, loudness, etc..

Moreover, the behaviour can be controlled in real-time by a performer who, with new dimensionality can controllon the overall developing of music. This approach opens a complete new and wide territory to explore for composition, and especially, for live performance. It is no longer a matter of playing an instrument in the traditional sense, but rather playing a specific piece of music in terms of activating and controlling during the live performance musical/acoustic material prepared during the compositional phase. Furthermore, the compositional approach too should be reconsidered because it becomes particularly necessary to balance *what is to be stated* and *what will be done.*

During the last decade many researchers and musicians have focused on the possibilities offered by this new approach to make music by computers. This is being done in two main fields: designing and carrying out original controllers, and designing and carrying out music languages which can facilitate both the tasks of algorithmic composition and defining mapping rules.

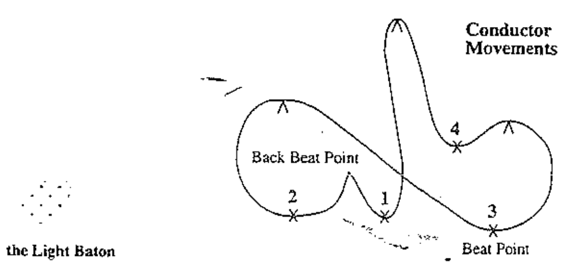
At the moment there exist on the market a number of MIDI controllers which can get information from the gestures of a performer and transform them into MIDI messages (keyboard, drum-pads, pitch-to-midi converters microphones, etc.); and there exist a number of high quality synthesizers which can generate and process sound in accordance with MIDI messages coming from controllers. Great effort has been made by electronic instruments companies for putting on the market controllers and synthesizers able to repropose the same facilities offered by traditional music instrument in terms of both "gestural interface" and sound response.

All that, basically, for satisfying the requests of commercial music world. But in a wider creative context, where the 12-tones music leaves room, more generally, to the 'sound', new and different gesture capturing devices may be (and must be) invented for detecting gesture of a performer. So, many original new man-machine interfaces have been designed: sensors typically used in robotics are taken into consideration for developing them: infra-red beams and real time processing of video captured images.

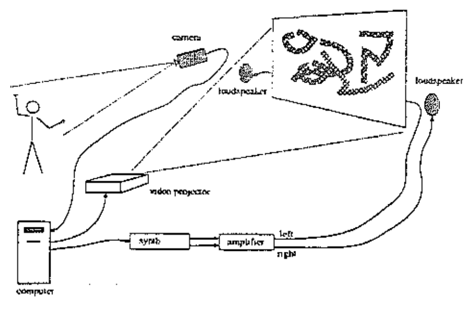
1.2 The approach of the Computer Music Lab

Our research activity led us to realize a number of orignal devices for detecting human gesture following the basic idea of remote sensing moving objects handled by performers and/or remote sensing gesture of the human body (or parts of it). They are:

- the Light Baton System that is a conductor baton with a light source placed at its tip which traces the movements in the air; it consists of - a thin, light and strong glass-resin rod thinning towards the tip, with a cork-handle very similar to an ordinary conductor baton - a CCD camera that detects the baton movements - an image acquisition board which converts images into an array of values associated to the levels of luminosity and - a computer which performs all the processing of the movements for identifying the main characteristics of gesture in terms of order number of the movement, amplitude and duration.

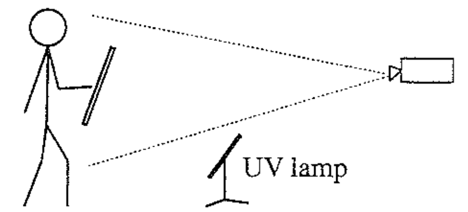


- the Aerial-Painting-Hands system which detects positions and movements of a performer's hands: the performer wears two ordinary cloth gloves of different colours and moves the hands in a CCD camera video area.

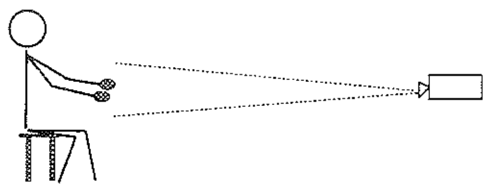


As before the signal is sent to a computer which processes the images and recognizes the x-y positions and open/close status of the hands: the computer then synthesizes in real time coloured images projected on a large video-screen in accordance to the movements of the performers's hands. The performers, while *painting in the air* also controls in realtime computer generated music.

- the UV-stick system uses the system configuration just described (CCD camera+video acquisition board+computer) and consists of a short (50 cm circa) straight stick lighted by a UV lamp whose 2-d position within the video area of the camera and 3-d rotation are recognized; data are then used for controlling in real-time music and sound synthesis algorithms. The stick is handled by the performer with movements emphasis.



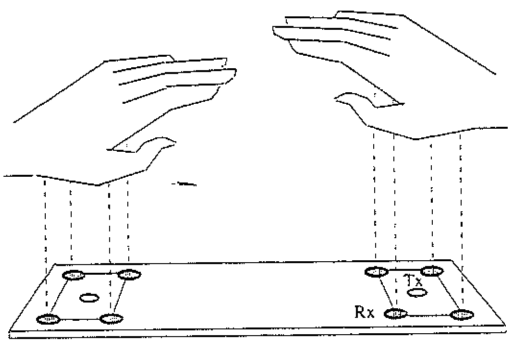
- the Imaginary Piano; here, with the same hardware, the nake hands of a performer are recognized while playing an non-existent piano keyboard. The hands are moved in the air mimicking the typical gesture of a piano player with his/her typical posture; the system issues information when fingers cross a horizontal line (the imaginary keyboard): how fast the zone is crossed states the velocity; where (from left to right) the zone is crossed states the pitch.



The four system just described are base on standard hardware (camera, digitizing board and computer) and the work is, finally, only a matter of programming [4] . The last device realized, the object of this paper, is based on infrared technology and is the result of lot of time spent in developing special circuitry, looking for appropriate electronic components, months of labor of specialized people in the electronic lab. The idea of the Twin Towers device was born after the prototype of an Infra-red based device developed by Genove et alii [5] at ARTS Lab in Pisa: we declared them our great interest to use the starting idea for designing a device with proper characteristics to be introduced in the real world of computer music. Not interested to continue their original project, they gave us very friendly and very kindly the details of that project, on the base of which we designed the device now called TwinTowers, with a new geometry, different electronic components and protocol for communicating with the computer, and new characteristics.

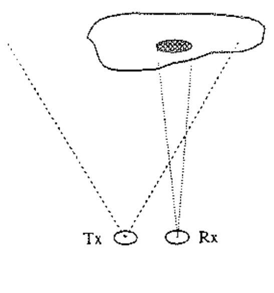
2 The Twin Towers system

This device gets information from positions and movements of the hands for controlling interactive computer music. It consists of two sets of four sensing elements which create two zones of the space (the vertical edges of two square-based parallelepipedon, or *towers*) where an object (i.e. a hand) can be detected in terms of height and front and side rotations with respect to a reference frame. Physically the Twin Towers device consists of a tablet of 20x50 cms held in a horizontal position at 70 cm circa from the ground by a stand very similar to a music stand: control electronics is on board, and a cable with power and signal wires, connects it to a power supply and the computer



Basic layout of the Twin Towers

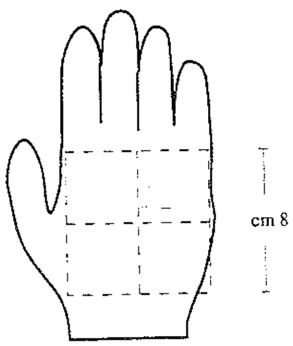
Infrared technology has been used with two kinds of IR components: a wide-angle transmitting diode (Tx) and a narrow-angle receiving (Rx) phototransistor both working in the 800÷950 nm near-infrared range. Tx's and Rx's are placed very close (2 cms circa) on the same plane and with the same upward orientation. An object placed in front of them, reflects on the Rx's the light coming from the Tx diode and its distance from the reference plane is measured in terms of quantity of the reflected light.



The term *irradiance* is introduced as the quantity per time unit present on a surface shed by electromagnetic radiation.

Since the reflecting object is actually an hand, measurement is greatly made easy just because human skin reflects up to 98% of near-infrared light. According to Hobbie [6] and Duck [7] this property of the skin depends on two main reasons: - the high degree of epidermis wrinkledness which is of the same order of greatness of near-infrared light wavelenght; - the presence of high quantity of water in the layers of derma very close to external surface of skin.

For detecting the mere distance of the hand from the tablet, a single Tx-Rx couple could be enough; but for detecting more information such as rotations and different kinds of movements, it is necessary to measure the irradiance regarding separate portions of the hand. For that, we consider the palm as a square of about 8 centimeters divided in four parts as shown in the following figure, each one seen by a separate Rx phototransistor.



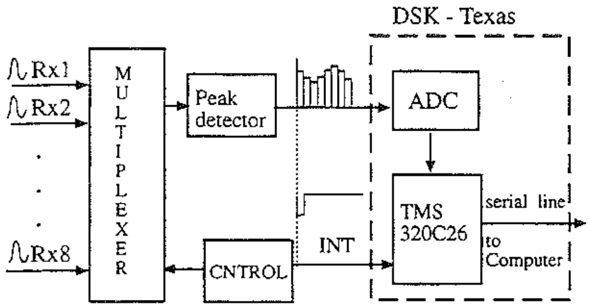
Considering an active working range of 30 cms starting from the tablet surface, in order to prevent from overlapping the four parts to be deteced, receiving sensors with narrow active angles (< 10 degrees) have been used. Hence comes the basic layout with a single wide angle (45 degrees) transmitting photodiode (Tx) at the center which serves 4 phototransistors (Rx) placed at the corner of a 4 cms square.

Since we measure the distance of the hands in terms of quantity of reflected light in the near-infrared range of spectrum, it's necessary to filter unwanted sources of light which may seriously affect the expected results: the sunlight and bulb lamps and, mainly, neon lamps.

Solar radiation affects the base of phototransistors, issuing a base current which appears as a continuos voltage level at the phototransistor output. Regarding the second source of light noise, the AC powered lamps have a spectral component very close to the working range of phototransistors and present at the output as a component centered at 100Hz. For both the sources of noises we realized proper filters. However a good choice is to use the TwinTowers after the sunset or at least in places where the sunlight cannot enter: studios, concert-stage...

3 Control electronics and data generation

Once filtered, signals issued by the Rx phototransistor are sent to a multipexer and a peak detector circuitry which creates a pulsed train of analog values properly converted in digital values sent, at the end, to the computer via a serial line at a sampling rate fast enough (40 times/sec) for giving meaning to human movements.



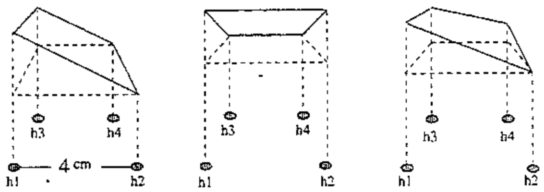
Our original circuitry ends with the line which carries the multiplexed analog values. Digital conversion is performed by the Texas Instruments DSK (Dsp Starter Kit) based on the TMS320C26 dsp processor: The DSK consists of a very small and inexpensive board with the microprocessor and a TLC AD/DA 14bits-19Khz converter. Basic tools (Editor-Assembler-Loader) running on the host computer make it possible the development of programs with necessary functionality such as that of reading analog values, digital convertion and serial transmission of values toward the computer itself.

The CONTROL part issues a strobe signal connected to the microprocessor and scans the eight analog values coming from the Rx phototransistors. The basic functionality of the convertion program resembles the start-stop protocol of serial transmission: here, the strobe signal starts the process which converts the analog values present at the ADC at specific time intervals. For what concern the range of values it must be noticed that the relation between voltage versus distance is not linear going with the inverse of the squared distance; but the convertion is performed on 14bits resolution, high enough to allow a precise linearization of values. The accuracy of measurement so resulted to be of 1 millimeter circa on the full range of 30 centimeters.

4 Gesture reconstruction

Since the values are always updated inside the host computer, it is possible to give meaning to data and recognize different kinds of positions and gesture: we decided to perform the operation of gesture reconstruction at computer level rather than at microprocessor level for better flexibility.

For each tower it is so possible to compute the height of the hand by adding to 4 values h1, h2, h3, h4; also it is possible to compute side and front rotations:



H=h1+h2+h3+h4 FrRot=h1+h2-h3-h4 SdRot=h1+h3-h2-h4

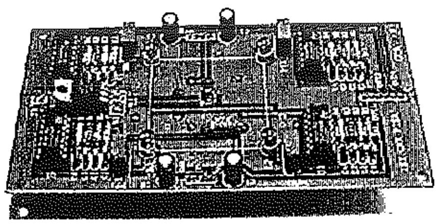
It is also possible to compute the velocity of movement by simply computing time elapsed between two different positions.

A second gesture we can recognize is the *flying hand* over a tower at different heights, speeds and directions, computing the order of on-off transitions of all four values of each tower.

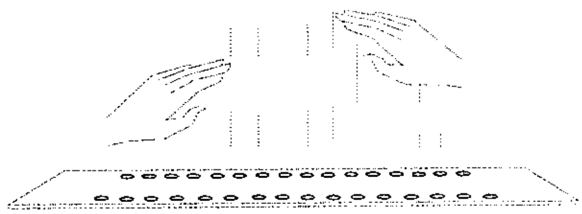
5 Enhanced configurations

The multiplexing circuitery has been designed for controlling up to 8 Tx and 32 Rx phototransistors, which makes it possible to implement larger and differently shaped configurations.

For a single tower, one Tx diode at the center and four Rx phototransistors in the corners of the square, we realized a 8cm wide board as shown in the photo;



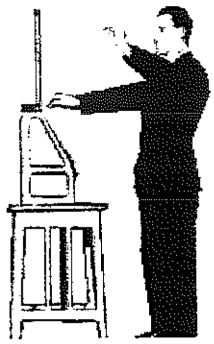
it is so possible to assembly up to 8 boards side by side which yields a double array of 16 equispaced receivers, the two arrays being spaced 4 centimeters.



This configuration will allow to recognize also lateral positions and movements of both hands.

6 Comparison

There exist mainly two devices the TwinTower may be compared with: the Theremin and The Hands. The theremin is a device invented by russian physicist Lev Sergeivitch Termen (anglicized to Leon Theremin) in 1919 while working for the Russian governmenton alarm devices. The theremin uses an electronic oscillator as a stable reference tone of a very high frequency and another electronic oscillator, initially in tune with the reference, which has a variable frequency controlled by the proximity of the hand to a capacitive sensing element (usually an antenna of some sort). The difference between the two frequencies is a pitch in the audible range which is detected and amplified. Move your hand near and away from the sensing element and get musical pitches. The scenic impact is absolutely impressive, but the



"analogic" working approach and the fixed one-to-one relationship between gesture and the acoustic result, make this device an instrument "closed" in its timbric characteristics and in the modality of execution. Nevertheless there exists so great interest toward this device that the Moog Company put on the market several hundreds of theremins [7].

The second device, the Hands, has been developed by Michael Waiswicz in Amsterdam about ten years ago, and since then, with some improvements, he uses the device giving concerts and demos around the world. The device consists of two paddles Mr. Waiswicz attaches at his hands, which sense relative position and orientation in space and contain switches and potentiometers for controlling pitch, timbre, and selection of instruments via the MIDI protocol. The device has been reproduced in very few copies for the researchers and musician of STEIM, the Institute in Amsterdam where it was designed and carried out. While this device can be considered as a gesture recognition device, it is too much similar to commercial data-glove commonly used in virtual reality application[8].

7 Conclusion

The Twin Towers device can be considered a general purpose gesture recognition device and it is not difficult to think it can simulate the functionalities of the Theremin and/ or of the Hands. The device has been used for many different concerto and demo situations with great success. Also, a prototype of theTwinTower as here described is installed as permanet realtime working exhibit at the International Museum of Science and Technology of Bagnoli, Naples. Due to the requests of people working in the computer music area we are now thinking to a version compatible to the MIDI standard which will allow to put on the market our device.

8 Acknowledgements

We want to thank Giovanni Chiparo and Massimo Magrini who greatly contributed to develop the AD convertion part of the project and CarloAntonio Giorgi who physically realized both the electronics and the cabinet part of the Twin Towers.

9 References

[1] Tarabella, L., guest editor (1993) *Special issue on Man-Machine Interaction in live performance* - INTERFACE. vol.22 n.3 - Swets & Zeitlinger B.V.

[2] Perotti, G., (1990) *MIDI Computer e Musica* - Jackson Libri - 853

[3] Zicarelli, D., [1991] *MAX 2.5 manual* - Opcode System, Inc.

[4] Tarabella, L., Bertini,G., Carosi, P. [1994] *A language and some interfaces for computer music live performances*, in Proceedings of the Aaurhus Symposium on Generative Grammars for Music Performance, Joahn Sundberg ed.

[5] Genovese, V., et alii [1991] *InfraredBased MIDI Event Generator*, in Proceedings of the International Workshop on Man-Machine Interaction in Live Performance, Servizio Tecnografico, Area di Ricerca, Pisa

[6] Hobbie, R.K., [1992] *Intermediate Physics for Medicine and Biology* - Second ed., J.Wiley & sons

[7] Duck, F.A., [1990] *Physical properties of tissue, a comprensive refence book*, Academic Press, H.Broce Javanovich Pub.

[8] *The Theremin home page* - http://www.Nashville.Net/~theremin/

[9] Waisvisz, M., [1985] *The hands, a set of remote midi controller*, in International Computer Music Conferrence 1984 Proceedings, CMA P.O.Box 1634, S.Francisco, CA