1) Abstract

This paper describes the implementation of a system used for carrying out the measurement of a sound source's position in a three dimensional space. It is difficult and often imprecise to develop an accurate system for such a positioning task

The approach of this method is based on the use of a pointing device tracked by cameras. The pointing device with a laser diode mounted on it is used by a participant to point in a certain direction. The system is able to measure the straight line between the centre of gravity of the participant's head and the point pointed. This direction is expressed as elevation and azimuth.
One of the aspects that make the method more efficient is the sound localization with an optical cue. By means of a laser diode the participants can see the exact position they are pointing to during the listening tests.

The method is performed in the IEM cube at the Institute for Electronic Music and Acoustics (Graz).This is also where the implementation and the practical evaluation took place.
The accuracy of the method and the error made during the calculation process is explained in the evaluation section. Finally some listening tests with real subjects are performed and the results are shown.

2) Introduction

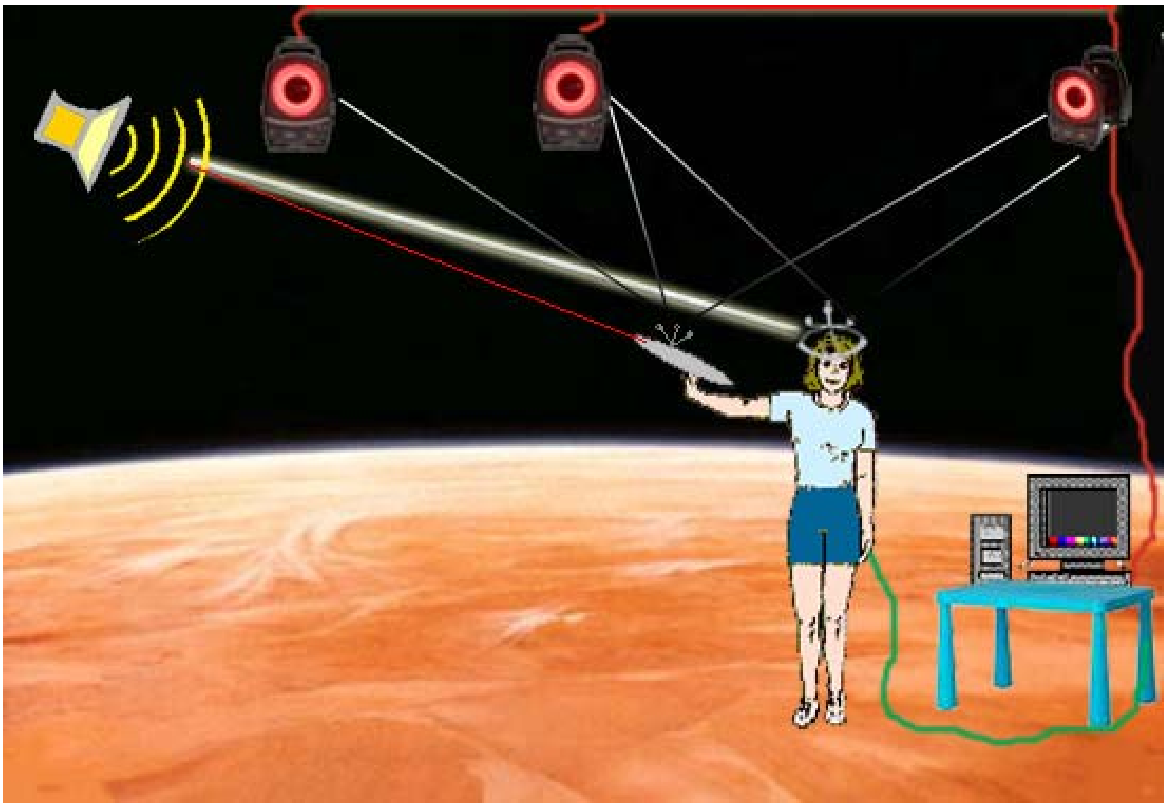
Sound localization is the human beings capacity to find where sounds come from. In many sound localization tests, the manner in which the sound localization is identified is often very imprecise, that is: by choosing the number of loudspeaker emitting the signal, by verbally estimating of localization, by pointing with the finger, etc. However, the method presented obtains accurately the perceived position coordinates where the sound is emitted from. The participants in the listening tests are asked to point (with the pointing device) to the position they think the sound source is located. The system uses this information to work out the straight line in the direction the point pointed and the point in the participants' head between the ears. The elevation and azimuth are the angles used to define this direction. This is the output data.

This article provides a guide describing the present model.
The main component is a tracking system built up by infra red cameras. It can capture the position of almost any arbitrary object formed of retro-reflective markers.

The handheld device is the tool constructed to be hold by the participants during the listening tests. A part of it is the pointing device. This stick is used to point in the direction of the sound source. A laser pointer attached on it helps the listeners in the localization of the sound source emitting the sound. A capture object with five reflective markers is also mounted on the handheld device. It allows the tracking system to detect the localization of the pointing device

Another important component included in the handheld device is a switch-button. This button is to be pressed by the participants when the desired position is being pointed at. At this moment, the corresponding computer operations are worked out in order to get the desired solution. Finally, the output data is printed out.

The following picture represents how the process is carried out:



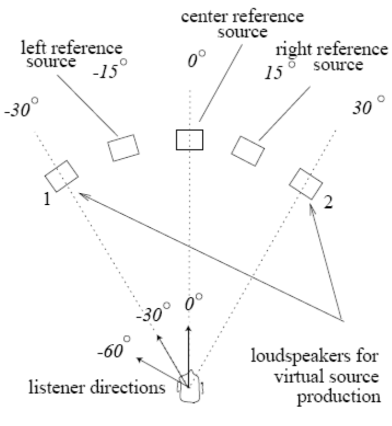
3) Different method utilized in previous studies

To carry out sound localization tests, an accuracy method to provide information about the perceived location may be made up. Researchers have tried to measure the direction pointed using several different direction indication methods.
In this chapter, a brief explanation about methods used in previous localization experiments is shown. Theses experiments can be divided into two groups:

1. Experiments which try to estimate the subject's ability in the sound localization.
2. Experiments which use pointing devices for measuring the coordinates pointed by the subject into a space and analyze the data afterwards.

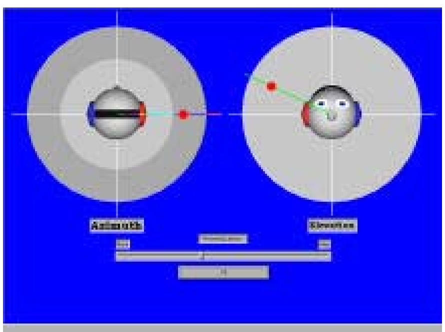
1) Firstly some method examples of the first group are explained:

One way to measure the number of correct choices given by the participants consists in the identification of the loudspeaker emitting signals [1]. The participants are situated in a place where an array of loudspeakers is set up. They are asked for the number of loudspeaker which creates the sound.



Another system to check the participants' ability in the sound localization is the method of adjustment [2]. In this experiment, three loudspeakers whose gain could be modified are set up in a horizontal plane. The gain variability causes the movement of a virtual sound sources between the arcs defined of two adjacent loudspeakers. In this case two virtual sources are created.
The listeners are asked to adjust the gain of the three loudspeakers to situate the virtual sound sources in the same position of real sources (reference sources)

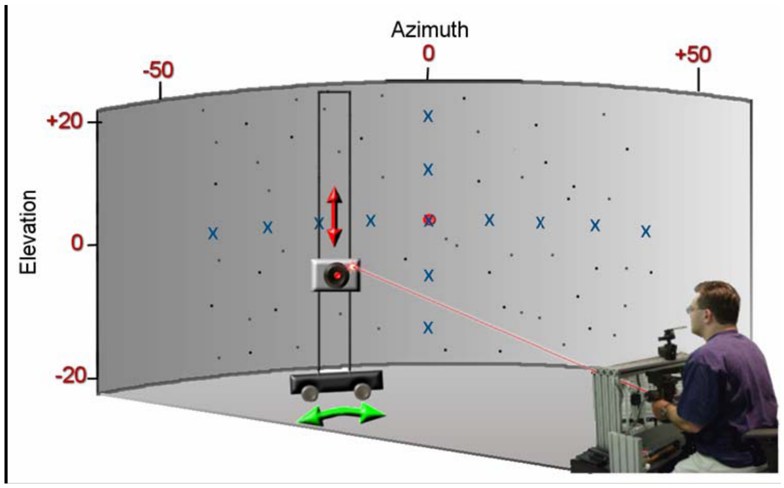
Another solution of the problem localization can be solved by using a graphical response screen [3]. In this experiment each subject has to indicate in a degree calibrated screen the direction they think the sound is coming from.



Moving the two vectors and the red dot in the graphical screen the azimuth, elevation and the sounds distance perceived are stored to obtain the results and made the evaluation afterwards.

2) The other group of methods tries to measure the coordinates where the sound source is located. Different apparatus must be implemented in order to obtain this information. Some advantages and disadvantages of the methods are also commented in comparison with the method described in this paper.

One example that can be mentioned is explained in [4]. In this experiment, the listeners are facing a calibrated screen at 2 m behind which a loudspeaker is hiding. A cylindrical joystick is used to point a red laser- LED on the black calibrated screen to measure the pointed position.

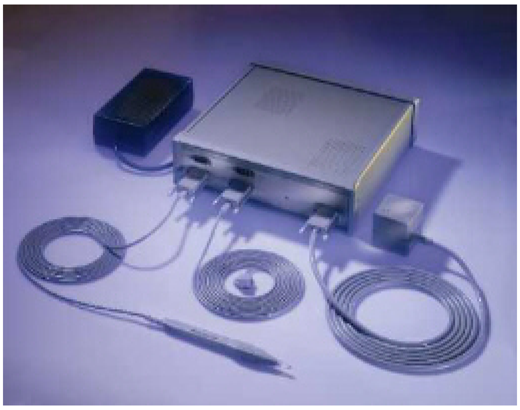


The speaker is moved randomly with a robotic arm behind the acoustically transparent screen. The listeners are asked to align the pointer with their eyes and to point to the perceived direction where the source is located. The sound location is reported with a visually-guided laser pointer.

This method obtains accurately the position pointed by the subject. However the space where the source can move through is only a screen with a define dimensions.
The paper here presented explain a method with a high positional accuracy in a three dimensional space.

A more efficient method is used in [5]. The 3Space Fastrak tracking system is the tool used to achieve the solution. This is an electromagnetic device composite by a system electronics unit, a transmitter and a receiver.
The transmitter emits the magnetic field which is detected by the receiver. The receiver is a small cube whose position and orientation are measured while it is moving. The transmitter is the system's reference in the measurements, therefore the receive motion is evaluated in relation to the transmitter position.

The participants are seated in the middle of an anechoic chamber where a loudspeakers hoop is set up. They are wearing a headband where the magnetic-tracker receiver and a laser pointer are mounted. The listeners are asked to look at the direction where the sound is coming from. The desired position is pointed with the laser situated in their heads. The orientation and localization of the receiver are measured and sent to a computer when each participant presses a button. Then, according with this data, the point pointed is calculated as the intersection of the straight line in the direction pointed and the loudspeakers semi sphere.



This direction indication method can measure the position of sound sources placed in a three dimensional semi sphere. The 3space Fastrak tool provides short latency (4ms) updated at 120 Hz and an accuracy of 0.03" RMS. Nevertheless the method has also some disadvantages. The first one is the interferences problem. Large metallic objects near the device may adversely affect the performance of the system. In addition the subjects can not freely move through the capture scenario because the transceiver and the receiver must be joined to the central electronic unit. Another inconvenient that should be commented is the mobility restrictions to point with the head to the sound perceived.

4) Method

In this section, the method of operation for carrying out sound localization tests and its apparatus are explained. First the required devices for the proposed implementation are listed.

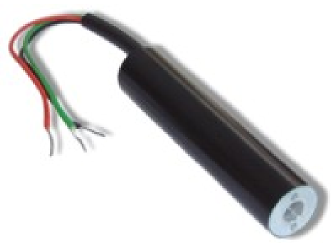
a) Apparatus

The **hand-held device**: This tool is used for the positioning task. Each subject holds it during the listening tests to indicate the desired position perceived where the sound is coming from.
It is made of two parts, the mobile part and the static part. The mobile part, referred in this paper as pointing device or stick, is used for pointing to the desired direction. It is made with a plastic bar in which a laser diode and some reflective markers are embedded. In the static part, all handheld device's connections are made. It is formed of the I-cube system and a prototype board.

Pointing device or stick

Static part

The components of the handheld device are:



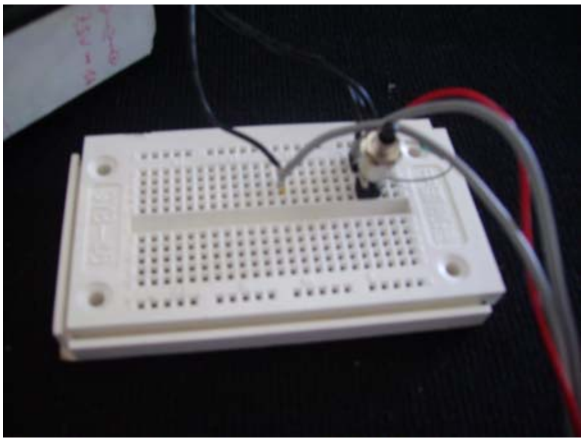
- Standard laser diode: It generates the light signal that helps the participant to indicate optically the perceived source position.

- Markers/Capture objects: The markers are a small reflective spots joined in a specific geometrical manner forming a capture object. The tracking system is able to track the position of each capture object with a certain number of markers.

During the experiment two different capture objects are used. One is included in the pointing device. The other one is required to obtain the reference position of the listeners head.

- I-Cube System: It is an analog (control voltage) to MIDI translator with 32 input channels. It translates sensor signals to MIDI messages and vice versa. It is the link between the participant and the system. For further details can be found in [5].

- Prototype board: It is a small board with holes in which a small electrical circuit is built up. There, it is connected the laser diode, the I-cube system and a button-switch. The button-switch is used to tell the system when the correct direction is being pointed.



The **headgear device**: It is a kind of hat with a capture object mounted on it. The headgear is used to know the localization of the participants' head position during the listening tests.

Tracking system: This is the tool to perform motion measurements. It provides an excellent accuracy and delivers a precise motion tracking of all objects with reflective markers.

Important parts of the tracking systems concerning the proposed implementation are:

- Data station: It is the central unit of the system which controls the operation of all devices included in the system. Its tasks are to provide power to all the cameras and to synchronize their data before it is transferred to a PC.

- Infrared cameras: They capture the motion of the markers walking across the capture volume.

- Vicon system's software: This is the program where the data captured by the tracking system's cameras come in. It is composed of two different applications, ViconIQ 2.5 where the measurement work is performance and QVicon2osc-beta which sends the information to another application. In this case the data is sent to Pure Data.

Pure Data program: The data is manipulated with this program to obtain the desired output. (Which PD Version did you use, which libraries are required…)

Cables: They are used to connect the different devices utilized for the method's implementation. All of them are explained in the section called connections.

Computer: It is the hardware where the software used is running. Two computer are utilized, the Tracker computer and the Rockmore computer.

Loudspeakers: A proper loudspeaker layout must be used to generate sound signals in a three dimensional space.

b) Apparatus relations

Once the tracking system is working, motion measurements of all the objects in the capture volume are carried out. As discussed before, the tracking system is able to obtain the correct Cartesian coordinates and the orientation of the capture objects mounted with reflective markers. A capture object is the association of reflective markers in a certain geometrical arrangement. This shape is known by the tracking software where the cameras' data comes in. All the information is interpreted by the tracking program to achieve the motion measurements.

For the purpose of this study, two capture objects must be built. One is included in the handheld device's stick and the other is situated in the headgear device. The two of them may have at least four reflective markers to ensure accuracy (In general more markers improve the reliability of the tracking data even if some of the markers are invisible to several tracking cameras. However the occultation of markers is not a cruxial case in the proposed system).

During a localization test the participants must wear the headgear on their heads and grip the handheld device. The static part of the handheld device may be held with one hand and the stick (mobile part) with the other.
Each subject must move the stick toward the perceived auditory direction. The laser diode's beam helps the participants in positioning the localized direction.

To identify the straight line between the listeners' head and the point pointed, the coordinates of the participants' head must be known. That is the aim of the second capture object mounted on the headgear device. The tracking system measures the point in the centre of the participants' head between the ears.
The current position and orientation of the two capture objects is provided in real time by the tracking system software. In order to obtain the final solution, this information is sent to another application in which some data calculations are worked out. The program used to perform this task is called PureData (PD).
The switch and the MIDI translator are used to tell PD when the perceived sound localization is selected. In that moment, the data transformations are worked out.

The first calculations are made for the purpose of obtaining the spot situated at the loudspeakers sphere in the direction of the sound perceived. This is the point pointed by the participants with the laser diode.
The last calculations are made to obtain the azimuth and elevation of the straight line between the participants' head and the last spot calculated. This is the output data.

Finally the results are printed out (stored) in a computer.

5) Implementation

In this section the practical implementation of the method and the whole process to obtain the solution is explained. The place where the procedure was developed is the IEM CUBE at the **Institute of Electronic Music and Acoustics (Graz).**

The explanation is divided in two parts. First the physical components employed are discussed and second the software is described.

Hardware:

In this section the handheld device's setup, the Vicon system operating, the headgear device and other components used in the implementation are all examined. Finally the connections between all of them will be commented on.

Vicon System

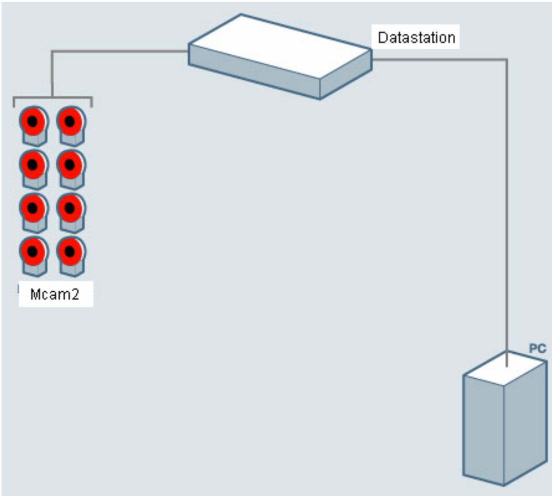
Vicon system is an advanced and accurate optical motion capture system. It is a marker-based solution to capture and analyze motion measurements. An amount of cameras track the position of capture objects mounted with reflective markers moving through the capture scenario.
The operation of the system is very simple. All the cameras emit infrared strobe light through the measurement volume which is reflected back by markers. The reflected strobe light passes through the optical filter with a spectral response which only allows light with the same properties as the strobe to pass into the lens. The lens collects the light and forms the image of the markers. The cameras transform the patter of light to a video signal and transmit it to the Vicon software where all the information is analyzed. All the cameras contribute in the calculation of the object's coordinates and orientation to obtain the best accuracy as possible.

The Vicon system is formed of two parts (hardware and software).The software part is described in the next section.

The Vicon's hardware is a quality tool that provides a high and accurate precision in real time. The data is traced almost instantaneously. Its modular design allows it to be expanded adding extra capability and is therefore easily integrated in any work environment.
The hardware part of the system is composed of the cameras, the data station and a computer where the Vicon software runs.

Firstly the Vicon tracking system must be turned on by pressing a switch in the central unit data station. This device can coordinate up to eight cameras. It distributes the power that all the cameras need to work and synchronizes their data before it is sent to a computer.

Connected to the data station are all the cameras and a computer. The following picture shows how the connection may be put into practice.



Seven Mcam2 cameras operate using the Vicon system in the IEM cube, where the method implementation is being carried out.

The data captured from each camera is sent to a software program. The Vicon software interprets the information received and computes the calculations to generate the output data. In the practical implementation, **the Tracker computer** is the PC used in the IEM-CUBE where the Vicon software is running.

In order to obtain the final solution some calculations of the data obtained from the tracking system have to be made. For that happen, the information must be sent from the Vicon software application to the other application, where these transformations are worked out. In the laboratory, the computer where this software is running is not the Tracker computer (where Vicon software is installed). Another computer called **Rockmore** is employed.

The handheld device's setup

The handheld device consists of two parts, the static part and the mobile part. During the sound localization tests, the participants should hold in one hand the static part and on the other hand the mobile part.

Mobile part of the handheld device

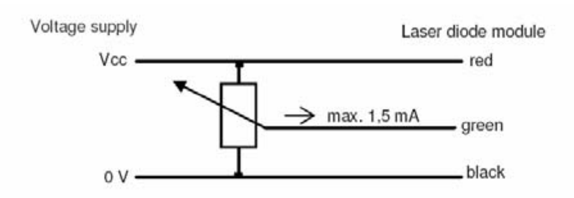
The mobile part is the pointing device. The participants should point with it in the direction of the sound source during the listening tests. The pointing device is a plastic stick that contains a captured object on top and a laser diode in the front.

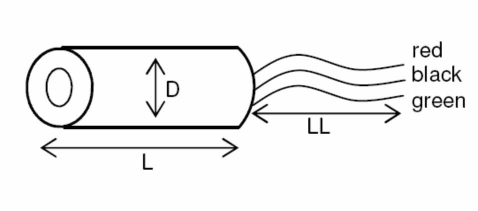
The capture object allows the tracking system to detect the localization of the pointing device. The cameras track the motion of the stick moving through the capture scenario.

A laser diode of the brand IMM was chosen to be built in the handheld device. The laser diode helps the participants to indicate the perceived source position during the listening tests. Its beam diameter on module is 2mm and the wave length measures between 670nm and 678nm. The diameter of the laser diode measure 10mm and the length 40mm. For more features see [7].

To power the laser diode there are three different ways to connect:

1. Continuous wave mode with maximum power: In this case, the red cable is connected to Vcc and the other two (black and green) to 0V.
2. Modulation mode with a control input: In this case red is connected to Vcc, black with 0 V and green must be connected to a frequency generator up to 25 kHz.
3. Power regulation mode with a control input: This is the mode chosen for the method implementation. The red cable must be connected to Vcc and the black to 0V. The green one is the control input used to switch on and off the laser diode

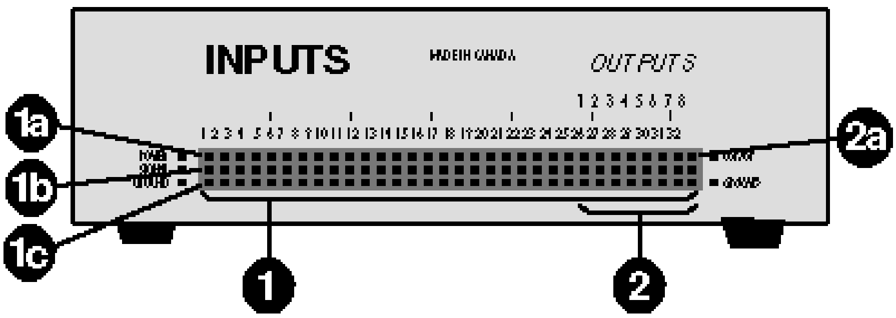




Static part of the handheld device

This is the part where all handheld's device connections are made. It is formed of the I-cube device and the protoboard (pinboard or prototype board).

I-cube system provides all the signals the laser diode needs to work. The picture below shows the connections possibilities of this device.



The number 1 indicates the sensor inputs and the number 2 the digital switch outputs.

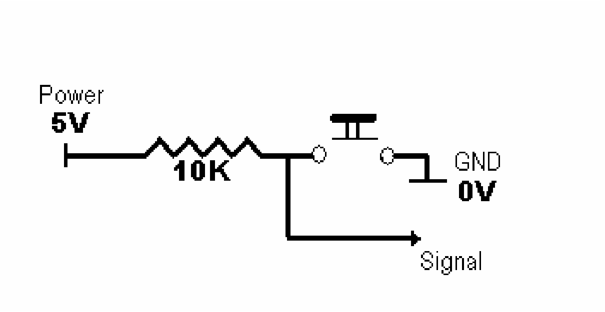
1a: 5VDC, 30mA power supply pin
1b: Sensor signal input in
1c: Sensor ground pin

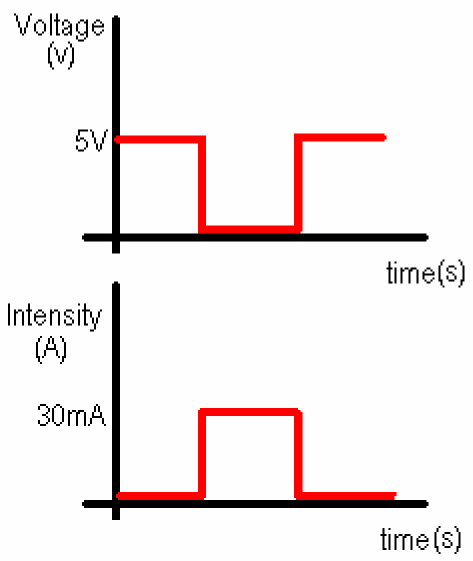
2a:5VDC, 39mA switchable outputs

The I-cube system is the device used for the communication between the proband and the Rockmore computer. In this computer a specific application called PD which is used to manipulate the motion data is running (explained afterwards).The I-cube device allows this application to select the localization data of the two capture objects when the participants press a button. The application also controls the on and off switch for the laser pointer.
The device can transform sensor input signals to Midi message and Midi message to digital trigger outputs. By interchanging Midi signals the device and the software can communicate. Midi signals are created in the I-cube device and sent to the computer and vice versa.
The I-cube can also provide the 5V direct current and the ground (0V) for the laser diode operation.

The sensor signal that creates the Midi signal in the I-cube device is the voltage transition when a participant presses a switch-button. This button is a momentary normally open push switch. It means that the electrical current doesn't go through it, except at the moment when it is pushed.

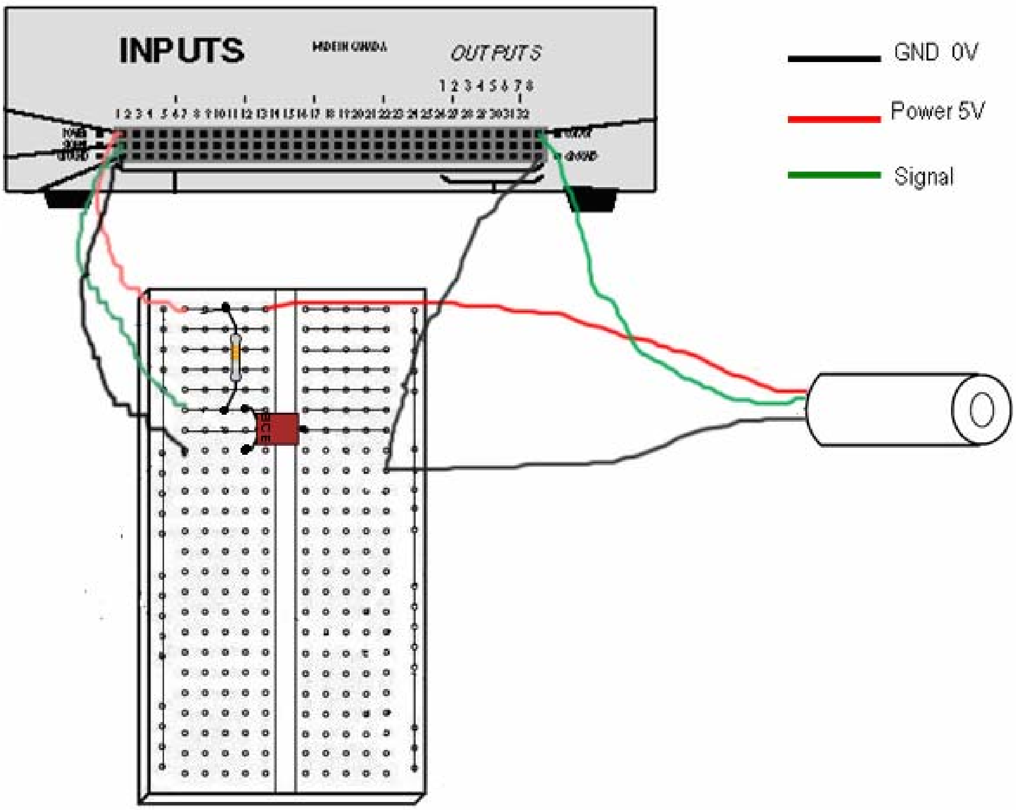
A simple electronic circuit had to be built in order to get the voltage transition from 5VDC to 0VDC. It was made using the button-switch and a very high resistor of 10kΩ connected with solid-core wire.





When the button is at rest the input signal is 5Vand the intensity 0A. However, in the instant the button is pressed, the input signal changes to 0V and the voltage reach to a maximum of 30mA. One of the 32 inputs is used to introduce the voltage translation into the I-cube.
Also one of the eight outputs for making the laser diode work is used. Midi messages are sent from a central computer, which are translated by the I-cube to a digital output. The output signal values are 5V (to turn the laser on) or 0v (to turn the laser off).

The protoboard provides the platform for making all the connections. The three components (laser diode, I-cube tool and switch) forming the handheld device are connected there.
The point-to point construction was the way to interconnect all components in the board.



Other elements to perform the implementation

Due to the rooms where the computers that interpret the data and the IEM-CUBE where the tracking system is built up are different, two more devices were used in the practical implementation (The **P2 connection extension panel** and a **RME multiface**) They were used in order to extend the Midi cables' distance from the handheld device to the Rockmore computer.

The P2 module is a connection panel located inside the IEM CUBE room; it is internally connected to the RME multiface in an adjacent room. The Tracker and Rockmore computers are also situated in this room and connected to the RME multiface.

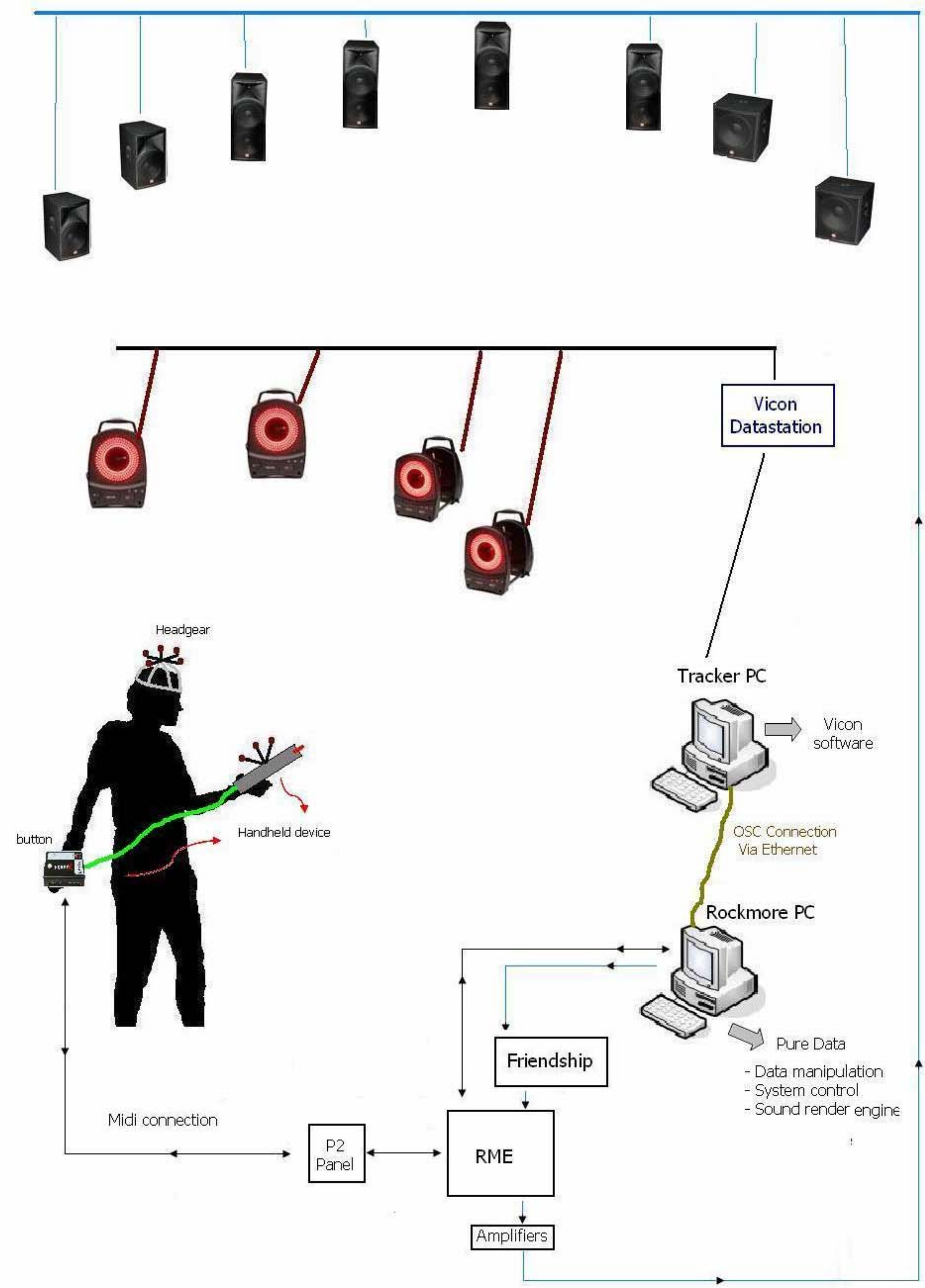
The RME multiface is a multichannel tool that provides different interface formats in an array of channels. In this case it is used to connect the Midi signals between the Rockmore computer and the P2 Connection Extension panel located inside the IEM-CUBE.

The use of these last two devices could be avoided using longer Midi cables between the I-cube tool in the handheld device and the Tracker computer.

Other apparatus required to make evaluations of the method are the loudspeakers. They are used to perform the reproduction of three dimensional sounds. A particular loudspeakers semi sphere layout is set up in the IEM CUBE.
In order to send the correct signal to each loudspeaker, the DMX 32 device is used. It is a programmable matrix to distribute digital input signals from various sources into different outputs. By the mean of different modules it allows to vary the format between the inputs and the outputs. For more information see [10].
By clicking the keys of DMX 32, the patch used to distribute the signals from the Rockmore computer to the loudspeakers must be programmed. Finally the sounds are emitted by the loudspeakers.

Connections

Here, the connections made for the method's implementation in the IEM CUBE are explained. The following picture depicts all these relations:



The participants must wear the headgear and the handheld device during the listening tests. Within the motion capture scenario each subject must point in the direction the sound is coming from. The tracking cameras record the capture object and transfer their data to the data station. The cameras are connected to the data station with distribution cable. This data station collects the data and sends it to the tracking computer where the Vicon software is running. This application analyzes the information and generates the localization parameters of the capture objects. A twisted 100 base T Network Cable is used for the direct connection between the data station and the tracking computer
The position and the orientation of the two capture objects are obtained with the Vicon software and sent to the Rockmore computer.
The Rockmore and Tracker computers share a common communication line within a LAN. The protocol used to send the data is OSC.
The application running in the Rockmore computer is the program called PD. The tasks executed in this program are:

1) The manipulation of the data sent from the PC tracker in order to obtain the final solution
2) The system control by the exchange of Midi signal with the handheld device.
3) The audio processing carried out during the listening tests

The localization parameters of the capture objects generated from the tracking system are sent to PD. This application makes the calculation with the data and prints out the results.

A Midi connection is used for the communication between the handheld device and PD. The Midi messages generated in the handheld device must pass trough the P2 module and the RME tool until they reach the Rockmore computer.

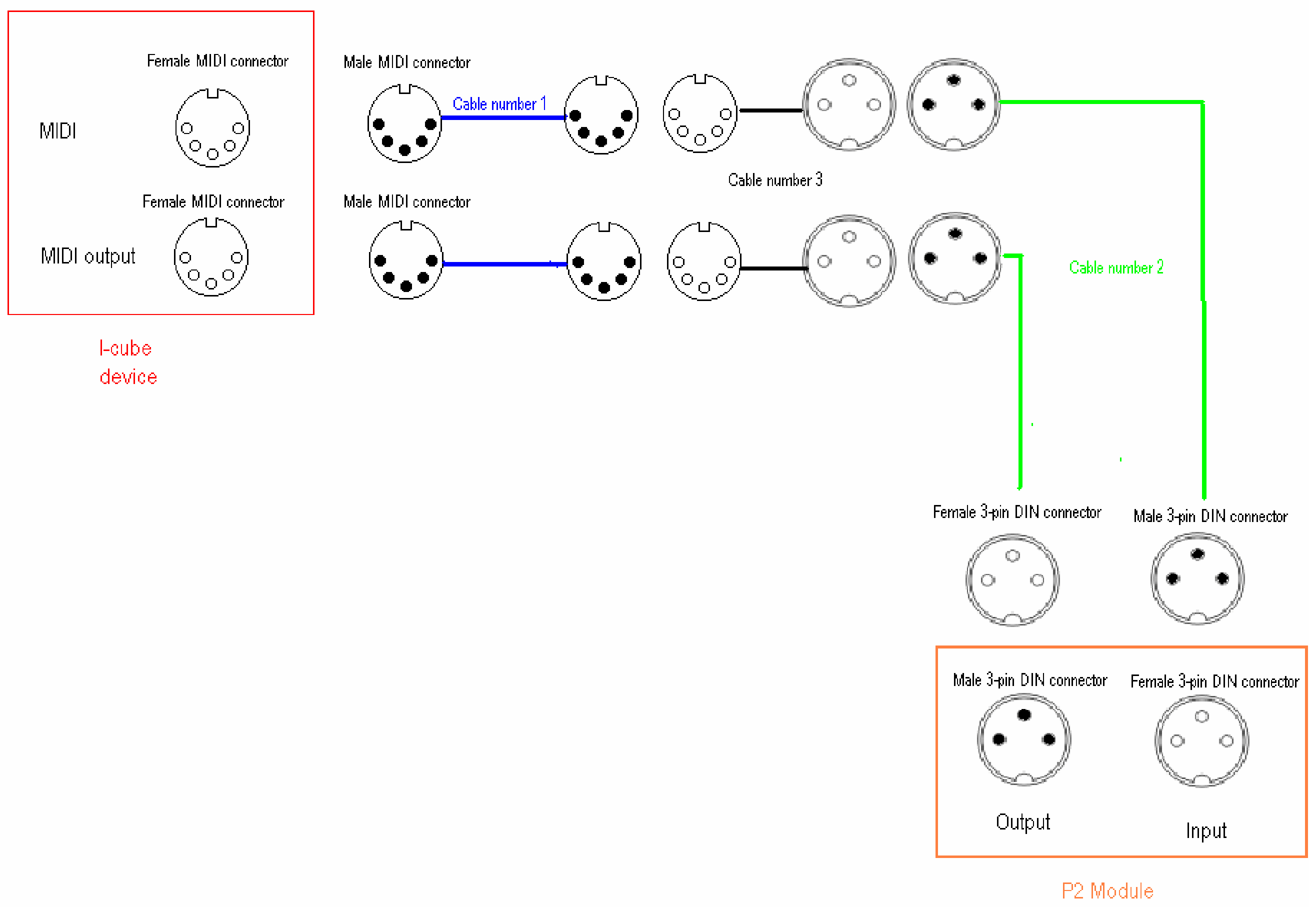
The handheld device is connected to the P2 module inside the Cube room. The RME and the CPU of the Rockmore computer are located in an adjacent room.
The RME multichannel is the device used to interconnect the P2 module and the Rockmore computer. The P2 module and the Rockmore computer are connected at the rear side of the RME multichannel. The signal distribution between P2 and Rockmore must be assigned in the RME device.

Three different cables are used to connect the P2 module and the handheld device. The DIN connectors in the P2 panel and the MIDI connectors in the I-cube device are different. The Midi input and output connectors of the I-cube device are female 5-pin DIN jacks. However, the connector in the P2 module for the output signals is male 3-pin DIN and the connector for the input signals is female 3-pin DIN.
The cables utilized for connecting the two devices are:

1) Standard Midi cable with two male connectors:

2) Cable with a female 3- pin DIN connector on one side and a male 3-pin DIN connector on the other:

3) Cable with a female Midi connector on the right and 3-pin DIN on the left:



The last function of Pure Data is the audio processing. There are some patches implemented in this program in order to render virtual sources onto the speakers layout built in the CUBE. The positional information must be encoded and decoded to recreate the sound in a three dimensional space. Although there is not a speaker in a certain position, the subject perceives the impression that sound is coming from that exact direction.
The patch used to distribute the signal from the Rockmore computer to the loudspeakers is programmed with the Friendship device. Finally the signal is amplified and sent to the loudspeakers.

<u>**Software:**</u>

Two applications are required in order to successfully complete the measurements of a sound source' position:

1. The Vicon software runs in the Tracker computer. The data sent from the cameras is interpreted and analyzed in this computer.
2. PD runs on Rockmore computer. There the Vicon system' data is manipulated to obtain the final solution.

Vicon software

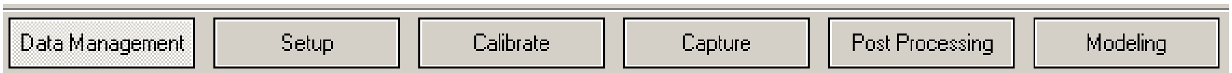
The Vicon software is used for processing the data sent from the cameras. It provides all necessary to manage and perform the motion measurements of the capture objects. There are different kinds of Vicon software depending on the task purpose to be achieved. In this case two Vicon applications are used, Vicon IQ2.5 and QVicon2osc-beta.

Vicon IQ2.5

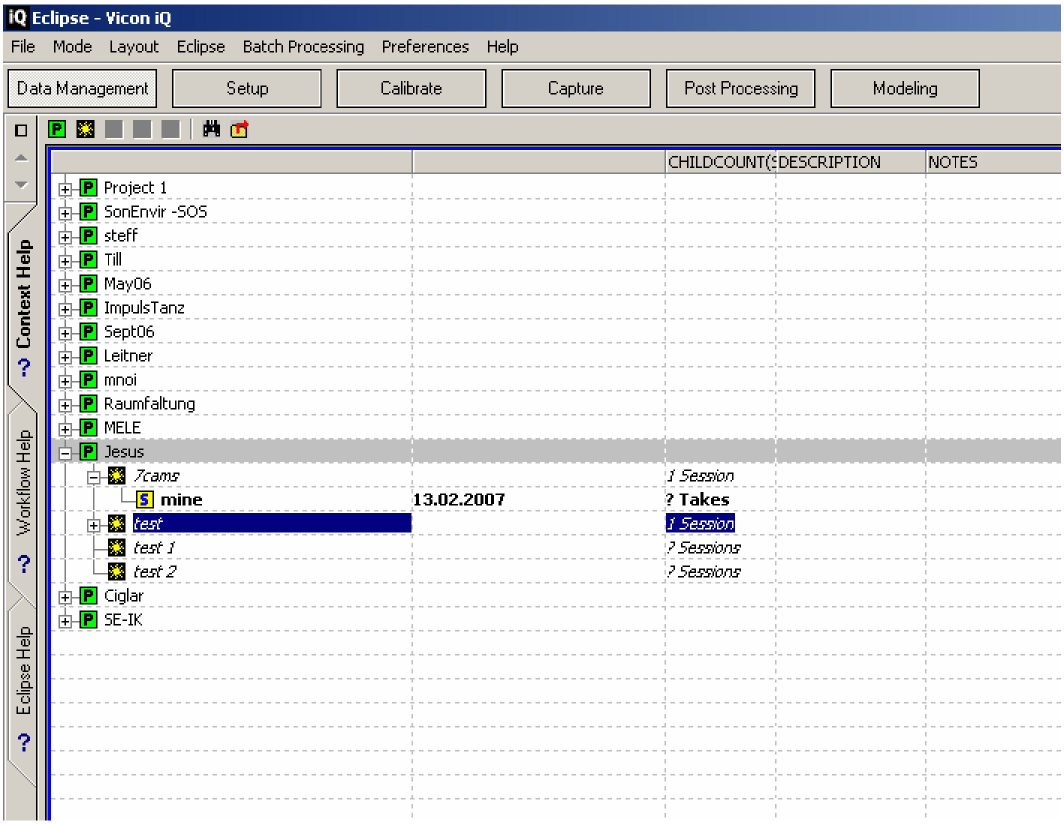
VicoIQ2.5 analyzes the data sent from the tracking cameras and generates the position and the orientation of the two capture objects mounted with reflective markers.

The Vicon IQ software can be initialised from the Windows start menu (by clicking on "All programs" and then clicking on the program ViconiQ 2.5) or from the windows desktop (by double-clicking on the Vicon iQ icon).

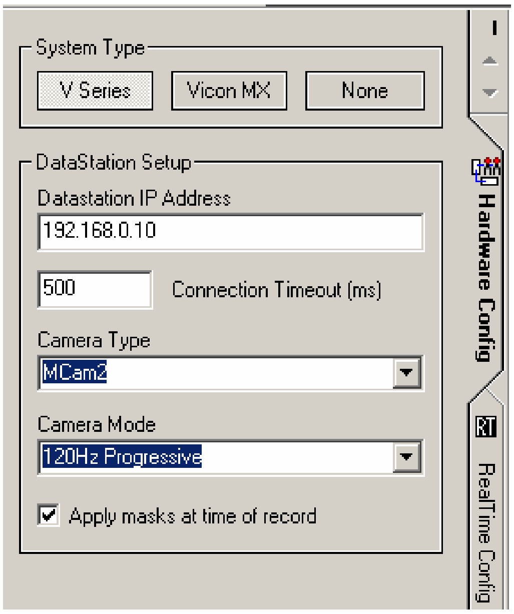
The procedure to control the capture process is divided in six operating modes: Data Management, Setup, Calibrate, Capture, Post Processing and Modeling. These correspond to the six boxes situated in the upper tool bar.



Vicon iQ opens in the **Data Management** operating mode. It allows the creation of a new database to store a new data file or to find saved data files.
By right clicking on the screen and selecting *file- new project* a new database will be created. At that point, a new single clickable button showing a P on a green background is created in the screen.
By the same procedure different "Capture Days" (sun icon) inside a "Project" and different "Sessions" (icon with an S on yellow background) inside the same "Capture day" can be created. This hierarchy allows for a good data classification.
When the program is opened, the database saved the last time is automatically loaded.
If you want to keep on working with a project already started, you must search the file in the browser under **Data Management** operating mode where it was saved.



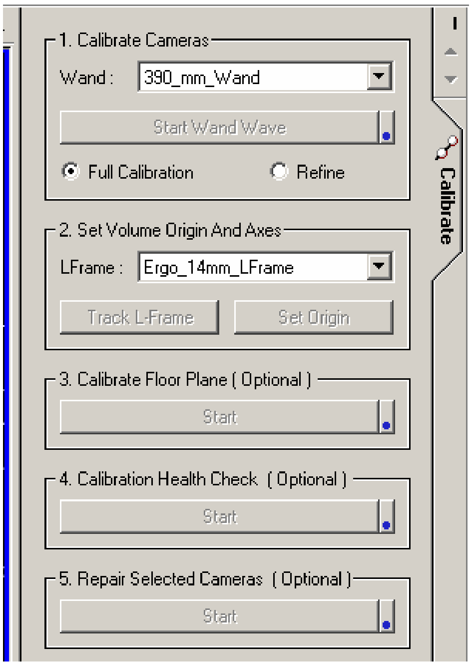
In the *Hardware Config* active bar under the **Setup** operating mode it is necessary to specify the Vicon elements which are installed in the laboratory. The Vicon iQ must be configured so that it can communicate with the other Vicon system components.
In the System Type configuration area you may specify the system category which Vicon iQ will connect to. Then, in the data station Setup area the connection between the data station and the cameras must be configured. It is necessary to specify the database IP address and the type of cameras employed.



In the *Cameras* activity bar, under the **Setup** operating mode it is possible to change the strobe intensity and the sensibility of each camera. Then the current camera settings can be saved to be reloaded the next time the project will be opened.

Before starting the data capture, the system must be calibrated. In the **Calibrate** operating mode you can calibrate the cameras, the capture volume, ensure the status of the current calibration and refine the camera calibration.

There are two types of cameras calibration that can be made, full calibration (in which all the parameters of the cameras are checked) and refined (in which some aspects are post edit).

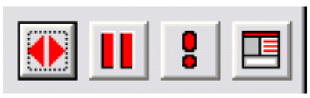


To calibrate the cameras the wand tool utilized in the process must be specified. Then, by clicking the start button, the program obtains the position and orientation of each camera and corrects the necessary camera parameters for an accuracy recording.

In the next area the volume origin and the axis of the capture scenario can be set. To archive this calibration, it is necessary to track a static object which must be introduced in a browser given.

The next three areas that provide calibration options can be useful in other specific situations but not for the aims of this experiment.

A Real Time Engine has been included in ViconiQ 2.5. It enables you to view the data inside the 3D workspace as it is being streamed through the Vicon system, as well as the standard offline capture.
Under the **Capture**, **Setup** and **Calibrate** operating modes, the real-time engine control bar is shown at the bottom of the window.



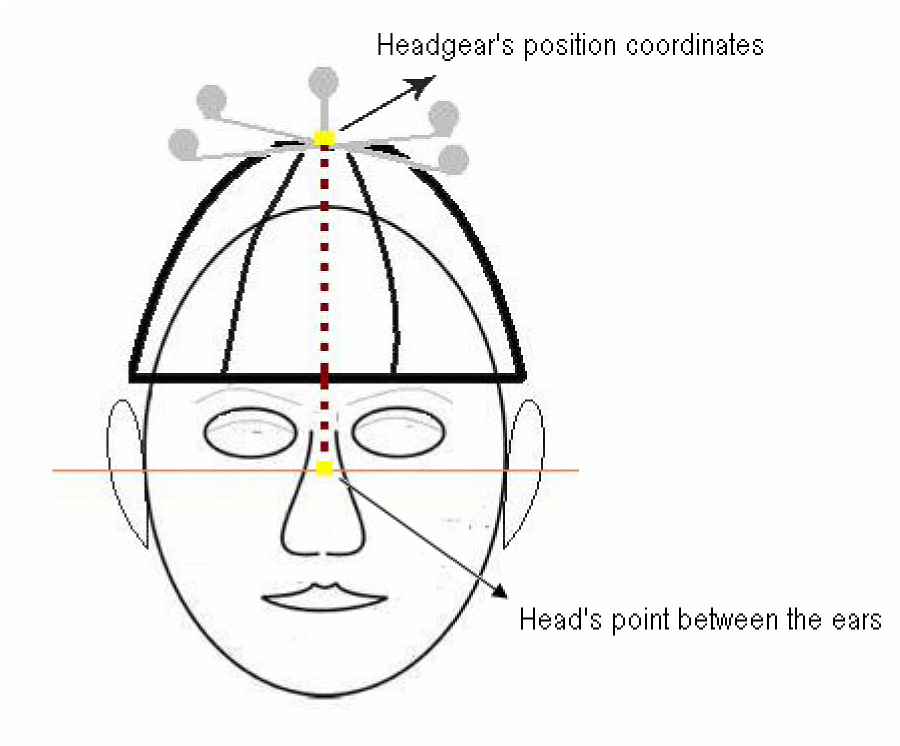
The first button is used to start the data capture. At this moment, the information is processed to reconstruct the three-dimensional markers movement, from the two-dimensional data recorded by each of the cameras. All the markers in the capture volume are displayed in the 3D Workspace window.

The next step in the **Capture** operating mode consists of creating the two objects used in the experiment. The two objects must be introduced inside the capture volume quite a distance from each other. To generate an object, all these actions must be done:

Firstly, the form in which the data is captured must be selected in the *Capture* activity bar.
Secondly, in the *Create/Edit* activity bar, the button Begin Editing must be clicked. In that moment the capture process is paused and all the markers that constitute a capture object must be selected in the 3D workspace. Then, the button Create Object must be clicked.
All the objects created are automatically added to the available objects list in the active object activity bar. Also, they are added in the listing box inside this *Create/Edit* activity bar where different option (change the name, colour, etc) of the exiting objects can be edited.
In the present experiment the name given to the two objects created are:

- *Pointer*
- *Head*.

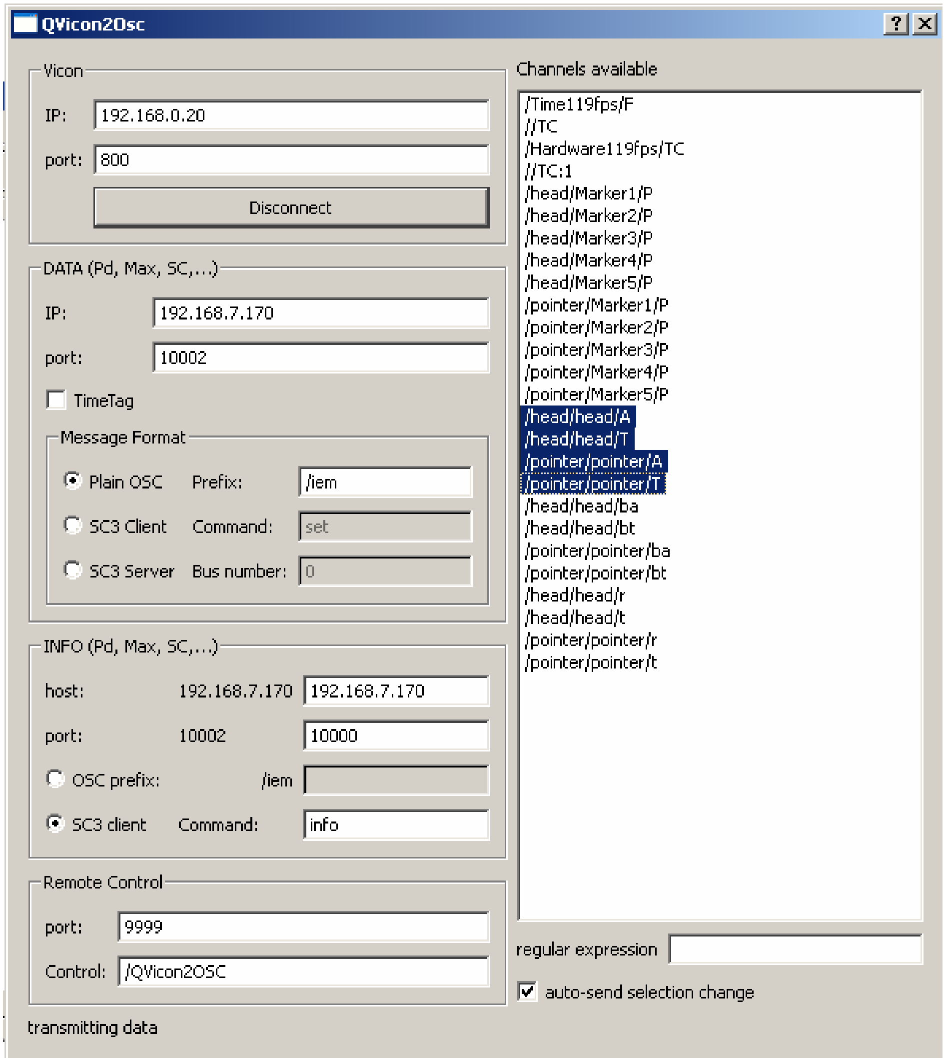
Another important option that had to be modified was the coordinates centre point of the head capture object. The point whose exact position needs to be known is located in the middle of the head between the ears. The output data is the elevation and the azimuth of the straight line that passes through this head point and the point pointed by the listener.



All these modifications are saved in the database session that is active. The name of this session is shown at the top of the window.

QVicon2osc-beta

This software is used to send the information to the application that is waiting for it. The ViconiQ's output data (coordinates position and the orientation of the two objects) must be transmitted from this application to the PD program where the final solution is obtained. The application's window has the following appearance:



First the program asks for the computer's IP address where the ViconiQ program is running, as well as the port which the data is sent though.
In the data area the receiver computer IP' address and the port of application waiting for the data running in that computer must be introduced. Another option that can be modified is the message format to be sent. In this case, it must be plain OSC.
The right landside panel shows a list with all the data available from the Vicon software analysis. By clicking the *auto-send selection change* button it is possible to select the output data that wants to be sent to the receiver application. In this case, the information required is the orientation and the position coordinates of the two capture objects. Finally by pressing the Connect button the connection between the two applications is set up.

Pure Data program

Pure Data (PD) is a free software with a graphical programming environment. It allows to process video, graphics and audio.
PD was developed by Miller Puckette in the 1990s and is quite similar to the Max family programming languages.
PD has a modular code; it is formed of building blocks making easier the programming task. Every element with a specific function has a programming process behind. It is possible to develop new PD documents to extend the program possibilities.There are many developers working on it, whose work is included in new libraries in the new versions of the program.

The PD documents are called patches. Inside a patch the different elements can be added. Connections to one another can be made to obtain a desired solution.
There are four types of elements or boxes that you can add into the PD patches: Object, Message, GUI, and Comment.
The message boxes generate data which will be sent many times while the patch is running. The objects are the boxes which modified the input data to obtain the searched solution. The other elements are the Comments to make annotations into the patch and the GUI boxes, which can have any previous function with a graphical user interface.

Each patch has one main window and any number of sub-windows. The program's way to arrange the boxes is by sub-patches and abstractions. They are new sub-windows which are called from a box in the main patch. When data arrives to these boxes, it is modified in accordance with the code written inside the sub-window.

In the experiment a few patches were created to achieve the solution. A brief explanation about them is now given in order to understand how the measurement process is carried out.

Data stream distribution

A great PD feature used in this experiment is the data load via Open Sound Control (OSC). This is a communication protocol to exchange information in real time. This is the message format used to send information from the Vicon software to the PD program.
The object DumpOSC is the box in which the orientation and coordinates position of the handheld device's stick and the headgear coordinates come in.

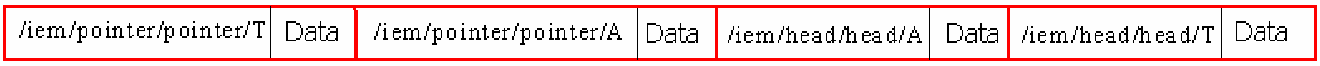
```
[dumpOSC 10010]  port-number depends on Vicon_2_OSC
```

Each localization parameter of the capture objects estimated with the Vicon software, which are sent from the QVicon2osc-beta to PD, is formed of a header and the data.
The header of each parameter is given in the QVicon2osc-beta application.

The header names given for the method's implementation are:

- /iem/pointer/pointer/T (stick's position coordinates)
- /iem/pointer/pointer/A (stick's orientation)
- /iem/head/head/A (head's position coordinates)
- /iem/head/head/T (head's orientation)

The data stream which is sent from the QVicon2osc-beta has the following appearance:



Each piece of information must be transformed in a different manner; therefore the data stream must be divided.
The PD object used to separate the diverse data stream is called *Route*. It checks the headers of the input messages and compares the headers with all of its arguments. If the match is done, the rest of the message appears on the corresponding outlet.

```
route /iem/pointer/pointer/A /iem/pointer/pointer/T
/iem/head/head/A /iem/head/head/T
```

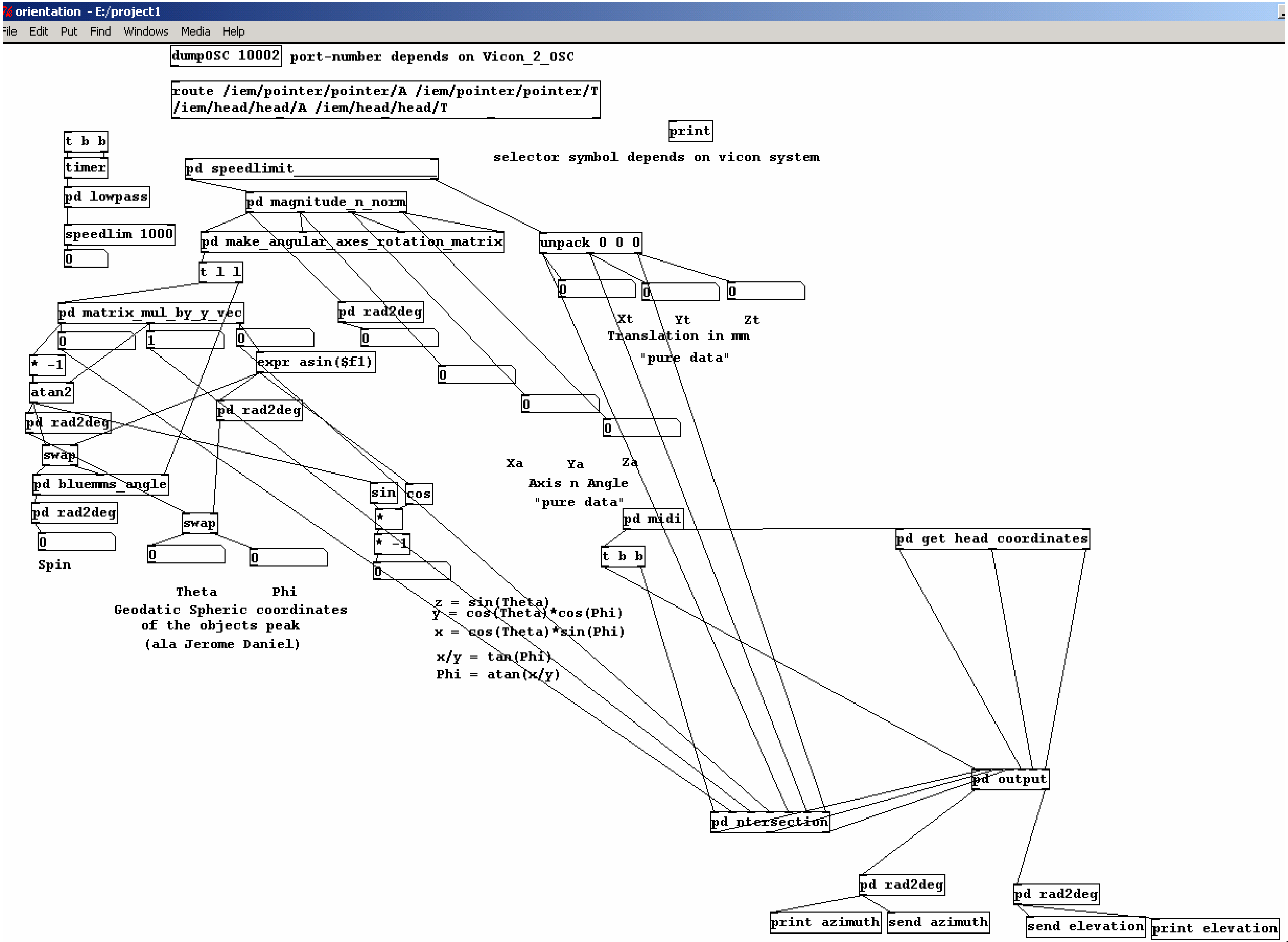
The arguments in the route box must coincide with the data's headers given in QVicon2osc-beta.
From then on, the position coordinates and the orientation of each capture objects follow different ways.

The three values which correspond with the position coordinates of the stick are printed in three number boxes. They are called Xt,Yt,Zt and are expressed in millimetres.
The three values which correspond with the stick's orientation are used to calculate the elevation and azimuth of the direction the stick is pointing to.

The elevation and azimuth are the best way to express directions in a three dimensional space. The elevation is the vertical component and the azimuth is the horizontal component. The convention used in this experiment to describe theses components is given as follows:

- The azimuth sense of rotation is clockwise. It goes from 0° in the positive y axis to 360°.
- The elevation goes from -180º to 180º. It is positive (0º...180º) when the slope of the straight line is positive and it is negative (-180º...0º) when the slope is negative.

Finally, the stick's orientation and the stick's azimuth and elevation are also printed.
The boxes showing the data are labelled with the comments: Xa, Ya, Za (stick orientation), Theta (elevation) and Phi (azimuth).
In this sub-patch called orientation, these equations have been implemented:


orientation - E:/project1
File Edit Put Find Windows Media Help
dumpOSC 10002
port-number depends on Vicon_2_OSC
route /iem/pointer/pointer/A /iem/pointer/pointer/T /iem/head/head/A /iem/head/head/T
print
selector symbol depends on vicon system
t b b
timer
pd lowpass
speedlim 1000
pd speedlimit
pd magnitude_n_norm
pd make_angular_axes_rotation_matrix
unpack 0 0 0
t l l
pd matrix_mul_by_y_vec
pd rad2deg
Xt Yt Zt
Translation in mm
"pure data"
* -1
expr asin($f1)
atan2
pd rad2deg
swap
pd bluemms_angle
pd rad2deg
Xa Ya Za
Axis n Angle
"pure data"
sin
cos
pd midi
pd get head coordinates
t b b
Spin
Theta Phi
Geodatic Spheric coordinates
of the objects peak
(ala Jerome Daniel)
z = sin(Theta)
y = cos(Theta)*cos(Phi)
x = cos(Theta)*sin(Phi)
x/y = tan(Phi)
Phi = atan(x/y)
pd output
pd ntersection
pd rad2deg
print azimuth
send azimuth
send elevation
print elevation

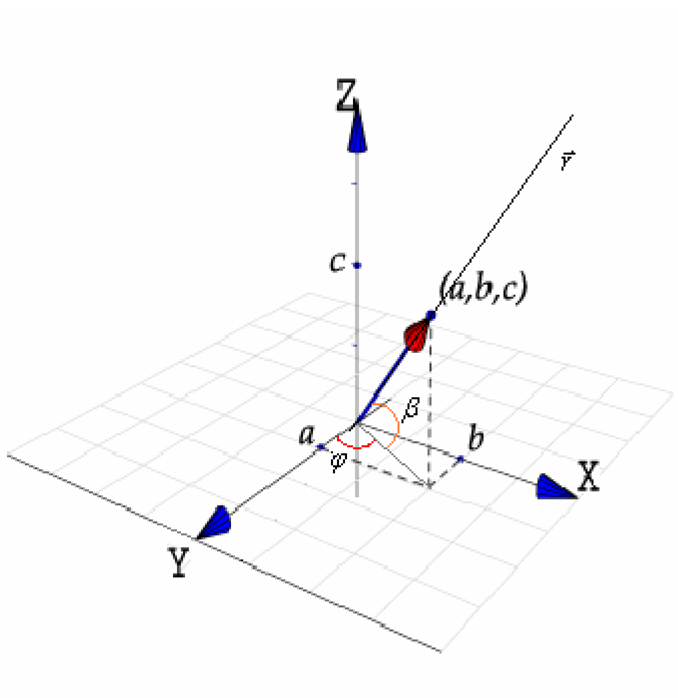
Calculation of the point pointed by the listener

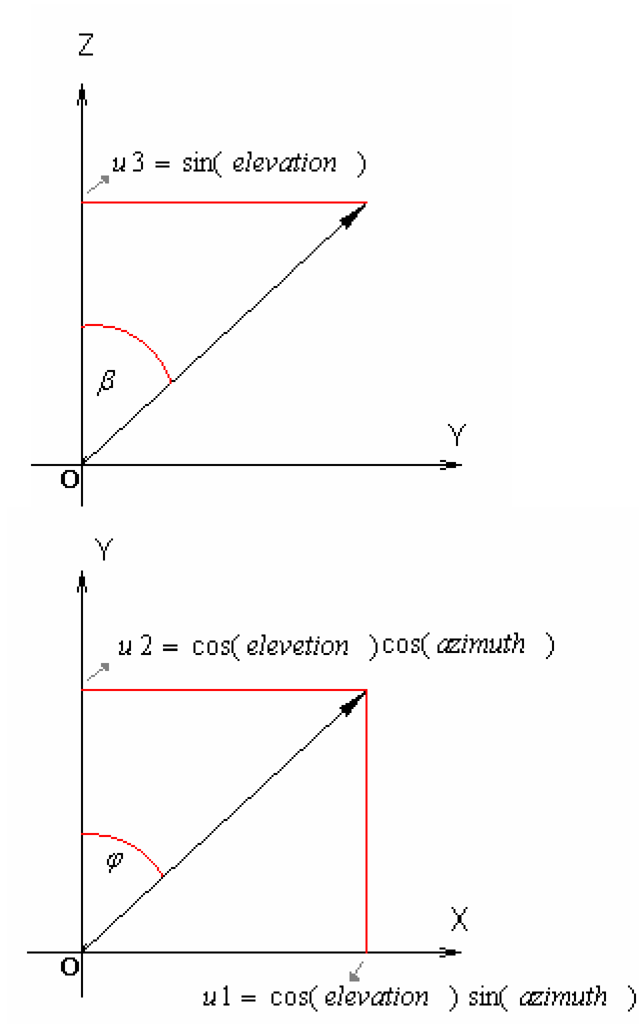
Any straight line might be described if one point and a vector in the direction of this line are identified. Whatever straight line can be expressed as:

$$g(t) = \begin{pmatrix} a1 \\ b1 \\ c1 \end{pmatrix} + t \begin{pmatrix} u1 \\ u2 \\ u3 \end{pmatrix}$$

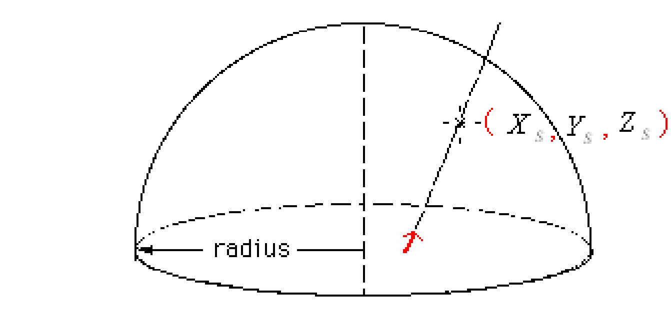
The point inside the straight line is acquired by the position coordinates of the stick in the handheld device. However, the vector in the direction the stick is pointing at should be calculated. It is estimated from the elevation and azimuth previously calculated.

$$u1 = \cos(elevation)\cos(azimuth)$$
$$u2 = \cos(elevetion)\sin(azimuth)$$
$$u3 = \sin(elevation)$$





Following with the process the intersection between the hypothetic loudspeakers sphere and the straight line in the direction the laser is pointing must be worked out.



The calculations implemented in the patches to find the solution are:

$$\left.\begin{aligned} X_S &= a_1 + tu_1 \\ Y_S &= b_1 + tu_2 \\ Z_S &= c_1 + tu_3 \end{aligned}\right\} \quad \text{Parametric equation of a straight line}$$

$$X_S + Y_S + Z_S = R^2 \qquad \text{General equation of a sphere}$$

$$(a_1 + tu_1)^2 + (b_1 + tu_2)^2 + (c_1 + tu_3)^2 = R^2$$

$$(a_1^2 + 2tu_1a_1 + t^2u_1^2 + b_1^2 + 2tu_2b_1 + u_2^2t^2 + c_1^2 + 2tc_1u_3 + t^2u_3^2 = R^2$$

$$(a_1^2 + b_1^2 + c_1^2 - R^2) + 2t(a_1u_1 + a_2b_1 + u_3c_1) + t^2(u_1^2 + u_2^2 + u_3^2)^2 = 0$$

$$(a_1^2 + b_1^2 + c_1^2 - R^2) = q \qquad (a_1u_1 + a_2b_1 + u_3c_1) = p \qquad (u_1^2 + u_2^2 + u_3^2) = 1$$

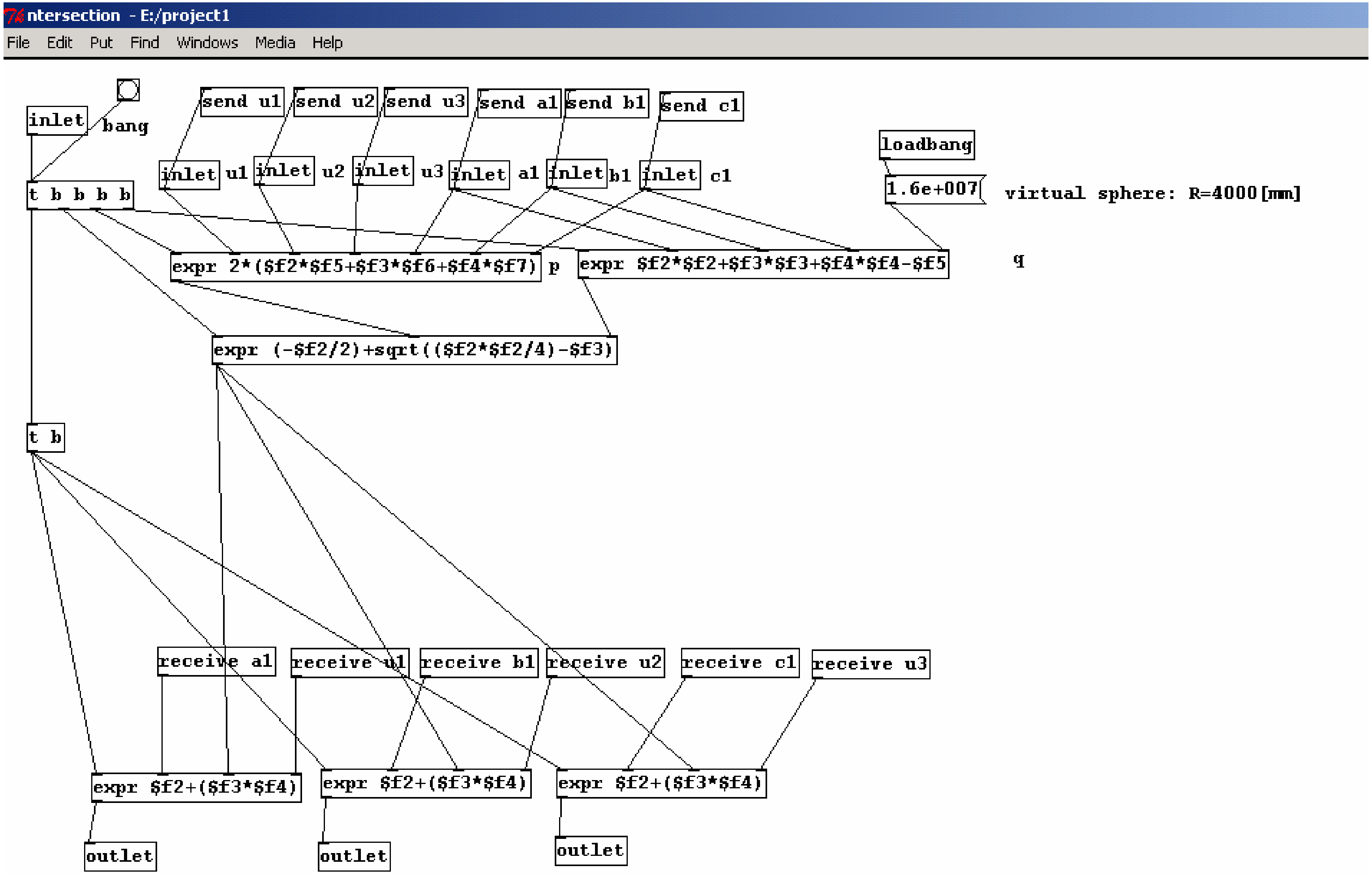
$$t^2 + t(2p) + (q) = 0$$

$$t_1 = \frac{-p + \sqrt{p^2 - 4q}}{2} \qquad \cancel{t_2 = \frac{-p - \sqrt{p^2 - 4q}}{2}}$$

The searched point is defined by a positive value of t, therefore the correct solution is t_1.
If this value is substituted in the first equation, the intersection point is obtained as:

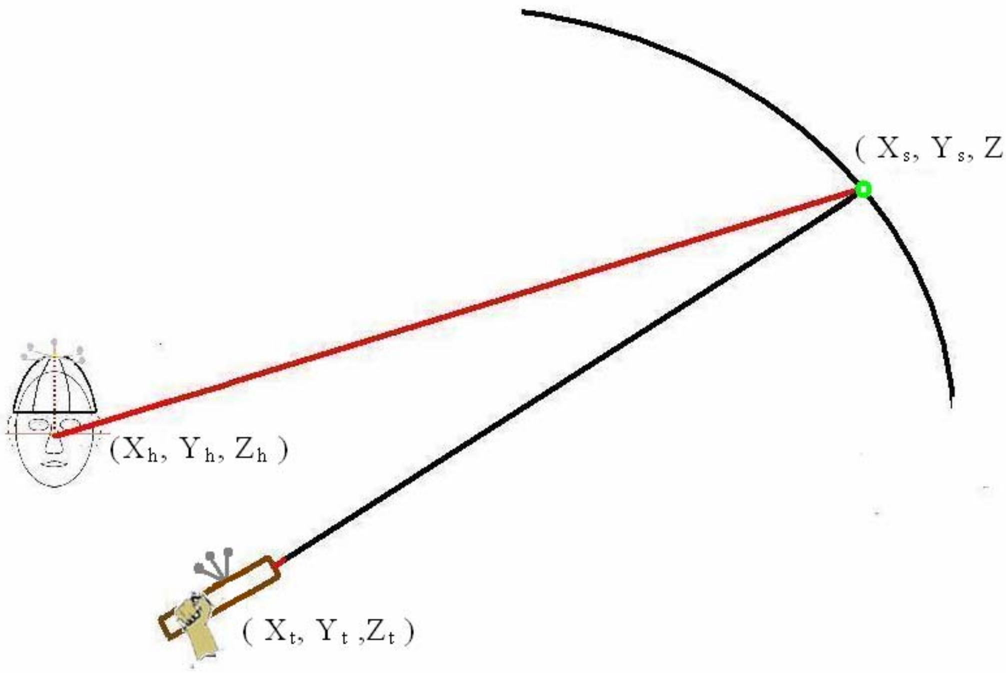
$$\begin{aligned} X_S &= a_1 + tu_1 \\ Y_S &= b_1 + tu_2 \\ Z_S &= c_1 + tu_3 \end{aligned}$$

In the following sub-patch called Intersection all these calculations are made:


intersection - E:/project1
File Edit Put Find Windows Media Help
inlet
bang
send u1
send u2
send u3
send a1
send b1
send c1
inlet u1
inlet u2
inlet u3
inlet a1
inlet b1
inlet c1
loadbang
1.6e+007
virtual sphere: R=4000[mm]
t b b b b
expr 2*($f2*$f5+$f3*$f6+$f4*$f7) p
expr $f2*$f2+$f3*$f3+$f4*$f4-$f5
q
expr (-$f2/2)+sqrt(($f2*$f2/4)-$f3)
t b
receive a1
receive u1
receive b1
receive u2
receive c1
receive u3
expr $f2+($f3*$f4)
expr $f2+($f3*$f4)
expr $f2+($f3*$f4)
outlet
outlet
outlet

Calculation of the output data

The last step consists on calculating the azimuth and elevation of the straight line between the headgear device's coordinates and the sphere intersection point. These are the final solution.



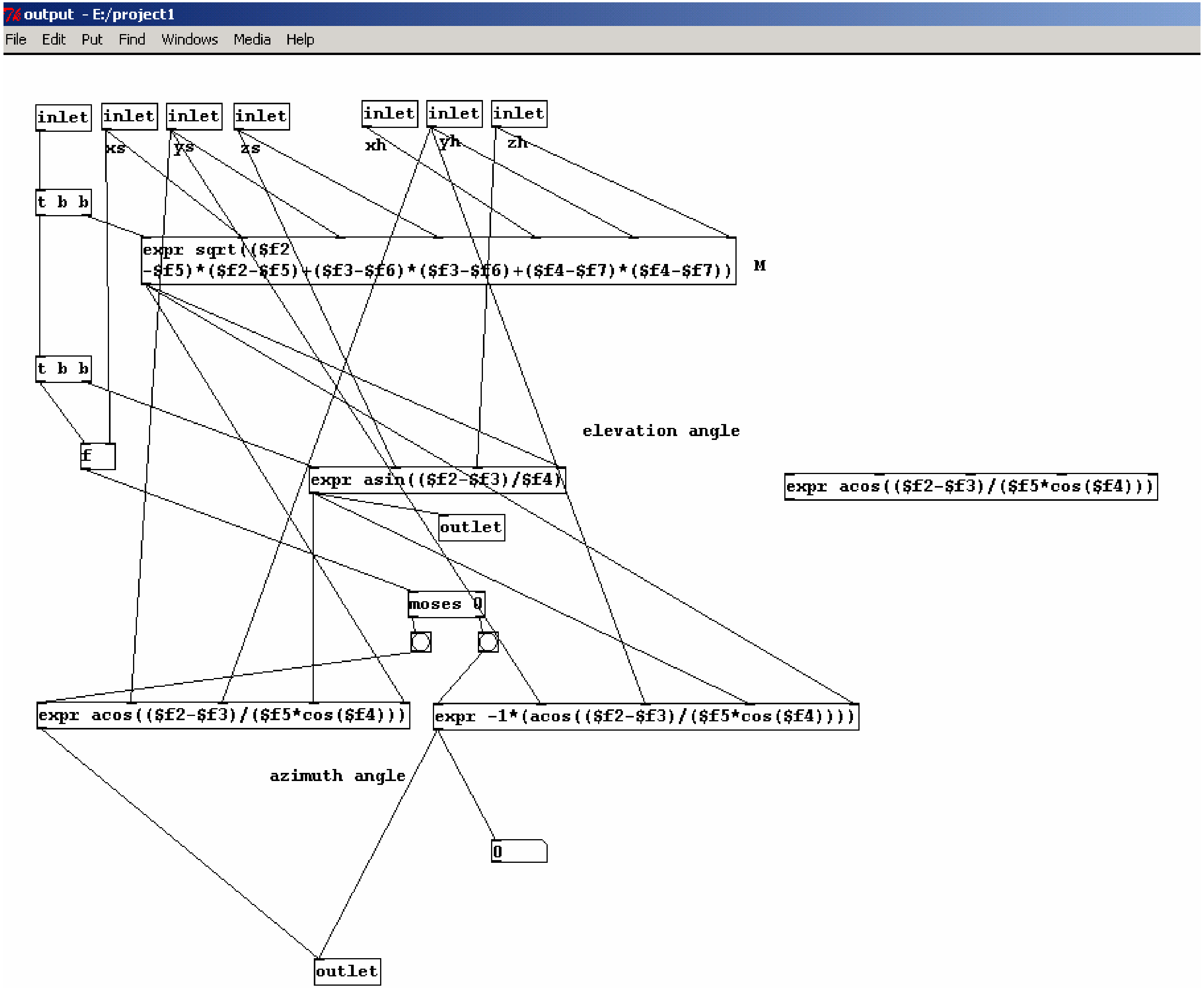
These are the calculations worked out:

$$M = \sqrt{(X_S - X_h)^2 + (Y_S - Y_h)^2 + (Z_S - Z_h)^2}$$ Magnitude of the vector between the two points

$$\sin\beta = \frac{(Z_S - Z_h)}{M} \qquad elevation = \beta = \sin^{-1}\frac{(Z_S - Z_h)}{M}$$

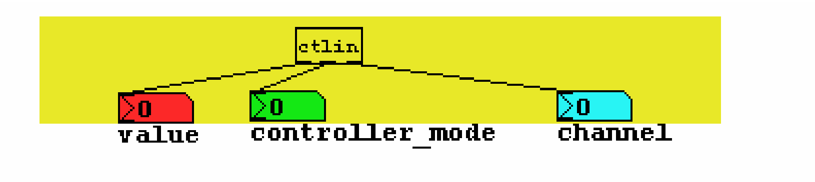
$$\cos\varphi = \frac{(Y_S - Y_h)}{M\cos\beta} \qquad azimuth = \varphi = \cos^{-1}\frac{(Y_S - Y_h)}{M\cos\beta}$$

The elevation and the azimuth are the final solution values. The following patch depicts the programming made in Pure Data to obtain the results.

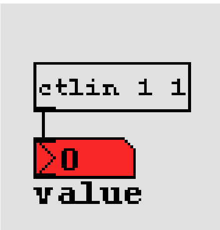

output - E:/project1
File Edit Put Find Windows Media Help
inlet
inlet
inlet
inlet
inlet
inlet
inlet
xs
ys
zs
xh
yh
zh
t b b
expr sqrt(($f2-$f5)*($f2-$f5)+($f3-$f6)*($f3-$f6)+($f4-$f7)*($f4-$f7))
M
t b b
f
elevation angle
expr asin(($f2-$f3)/$f4)
expr acos(($f2-$f3)/($f5*cos($f4)))
outlet
moses 0
expr acos(($f2-$f3)/($f5*cos($f4)))
expr -1*(acos(($f2-$f3)/($f5*cos($f4))))
azimuth angle
0
outlet

Midi signals

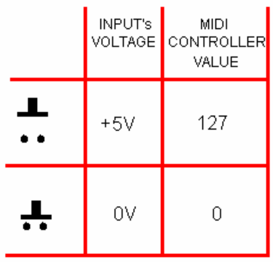
The input data is continuously processing to Pure Data through the OSC object. However, the patches are programmed for executing the calculation only when an expected MIDI signal appears. When the Midi signal reaches the program the localization data sent from the Vicon software is automatically selected. The calculations explained in the section before are made with this selected data.
The PD object used for searching the changes of Midi values sent from a controller is called **ctlin**. A controller is any hardware or software which can generate and transmit Midi messages. In this case the I-cube device is used as controller to send Midi signals to Pure Data.



The information related with the controller mode and the channel is known. The controller mode is omni and the channel which the signals are sent though is the number one. Therefore, the **ctlin** object is modified as:



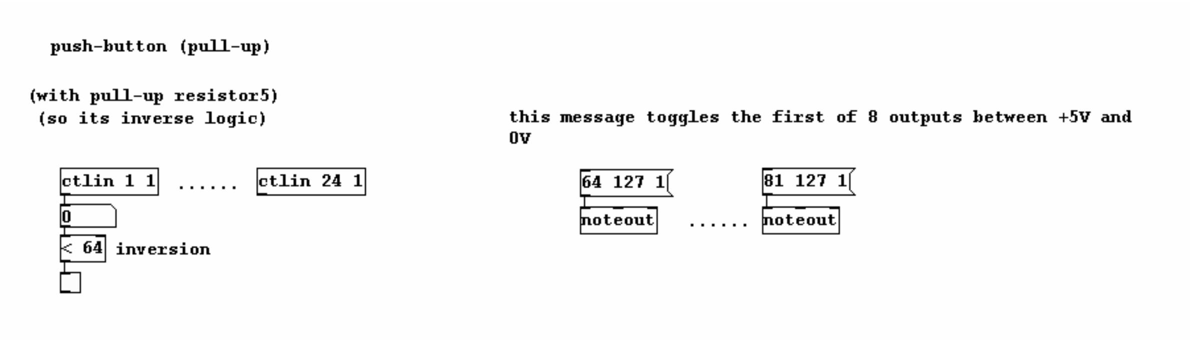
When the switch button is opened (it is not pushed down), the current at the I-cube device input is 5V and the value of the Midi controller sent to PD is 127. At the moment the button is switched, the input current changes to 0V and the controller’s value changes to 0.



A logical operator object is programmed in PD to detect the changes of the Midi controller when the participant presses the button. Then, a special PD message (bang) is created in order to select the localization data of the capture objects in that moment.

Other MiDi signals are also created in PD to switch on and off the laser diode connected to the I-cube tool. The eight digital outputs of this device are controlled via note on and note off messages. The I-cube device changes the output voltage (5V/0V) depending on the Midi signal (Note on/ off) sent from PD.

The patch in which the Midi treatment is carried out is now shown:



Finally, since the motion data is chosen and the calculations are worked out, the final result is printed in the main window of the PD program.

6) Evaluation

In this section some positive characteristics and also some disadvantages of the method implemented are explained. In the last point the accuracy of the method is evaluated.

Advantages

The method described offers several advantages for the solution of the sound localization problem.

Resolution: The system presented deliver great precision. The system provided a six-degrees-of freedom tracking of the capture objects with a resolution up to 0.000001 meters.

Optical cue: As opposed to other models, this one presents an optical cue provided by the laser pointer. It helps the participants to see the position they are pointing to during the sound localization tests. In this way, the visual and auditory systems cooperate together to determinate the sound source position.

Wireless: The capture objects can move freely through the capture scenario. There are problems stem from cable restrictions.

Latency: Another advantage is that the motion measurements of the objects moving through the capture volume are provided in real time. The latency since the participant press the button until the solution is printed out is less than 400 milliseconds. It was measured with the object Timer in Pure Data.

Update rate: The update rate with which the tracking data is sent from the Vicon software to PD can be modified. In the method implementation the data is calculated and sent every 5000ms (really 5seconds ? a bit slow???).

Great handheld design: Since the handheld device is divided in two parts, the hand to push the button does not intercede in the positioning task of the other hand.

No interference problems: unlike other methods, there are no interferences that can modify the tracking measurements.

Disadvantages

The method presented also has some problems that should be commented.

Firstly, the less sunlight there is in the capture scenario, the more accurately the tracking system works. The cameras capture the infra red light reflected by the markers. The sunlight is also composed of infra red light being able to modify a bit the markers localization measurements. Nevertheless these modifications are almost negligible.

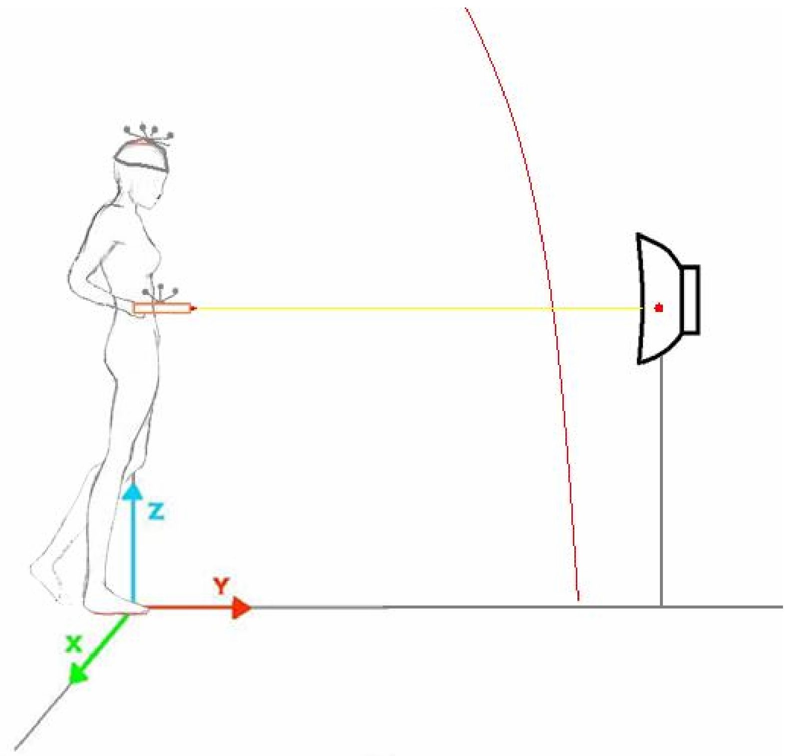
A bigger problem is that the loudspeaker layout in the Cube room is not a real semi sphere. There are some differences in the radius between the loudspeaker position and the volume origin. It makes that the calculation worked out to obtain the final solution have some errors.

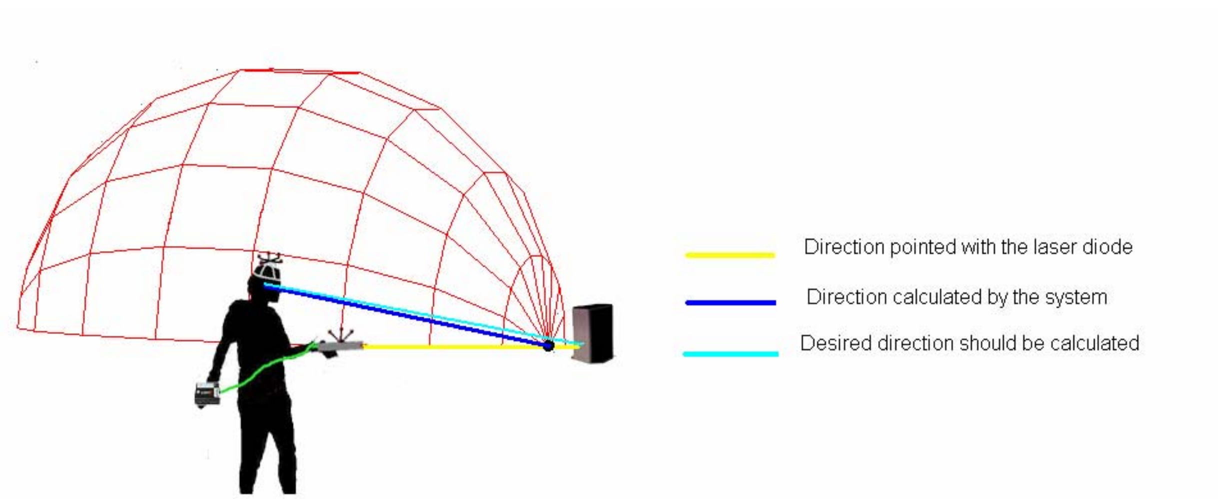
The calculus made in order to determinate the point pointed by the participants is the intersection between the loudspeakers semi sphere and the straight line in the direction pointed. Since the loudspeakers layout is not exactly a semi sphere, some errors are introduced in the final result. The farer the loudspeaker is located from the hypothetic semi sphere the higher the error will be.
The following table shows the distance (radius), the elevation and the azimuth of the direction between the loudspeakers position and the volume origin.

Nr.	Azimuth (Grad)	Elevation (Grad)	Radius (m)
1	0	0	4,67
2	22	0	4,27
3	45	0	3,1
4	71,5	0	4,02
5	102	0	3,59
6	138,5	0	3,4
7	180	0	4,71
8	221	0	3,27
9	258	0	3,63
10	288	0	4,02
11	315	0	2,94
12	338	0	4,22
13	25	28	4,02
14	69	28	3,79
15	116	28	3,19
16	155	28	4,79
17	205	28	4,76
18	244	28	3,3
19	296	28	3,96
20	335	28	4,05
21	45	59,5	5,56
22	135	57	5,89
23	225	57	5,87
24	315	59,5	5,61

The length of the radius varies for each loudspeaker. The radius average worked out as the mean of all the distances is approximately 4.2 meters. Since the majority of the loudspeakers distances with respect to the room's centre are close to 4 m, the radius of the hypothetic semi sphere taken to make the calculation is 4 m. In this way the error made will be lower for most of the cases.

Some examples of the error made in the PD calculations with this semi sphere radius are now shown:
The first situation chosen is for the loudspeaker number one when the participant is pointing in the y axis direction. The loudspeaker is situated 4.67 m from sphere centre.





In this case the pointing device, the loudspeaker position and the intersection point are located in the same straight line.
The positions of the points expressed in millimetres that are supposed for this example are:

Pointing device (0, 0, 1400) Head position (0, 0, 1800)

Loudspeaker position (0, 4670, 1400) Intersection point (0, 4000, 1400)

If a real loudspeaker semi sphere with a radius of 4 meters is assumed, the elevation and azimuth between the intersection point and the participants' head calculated by PD will be:

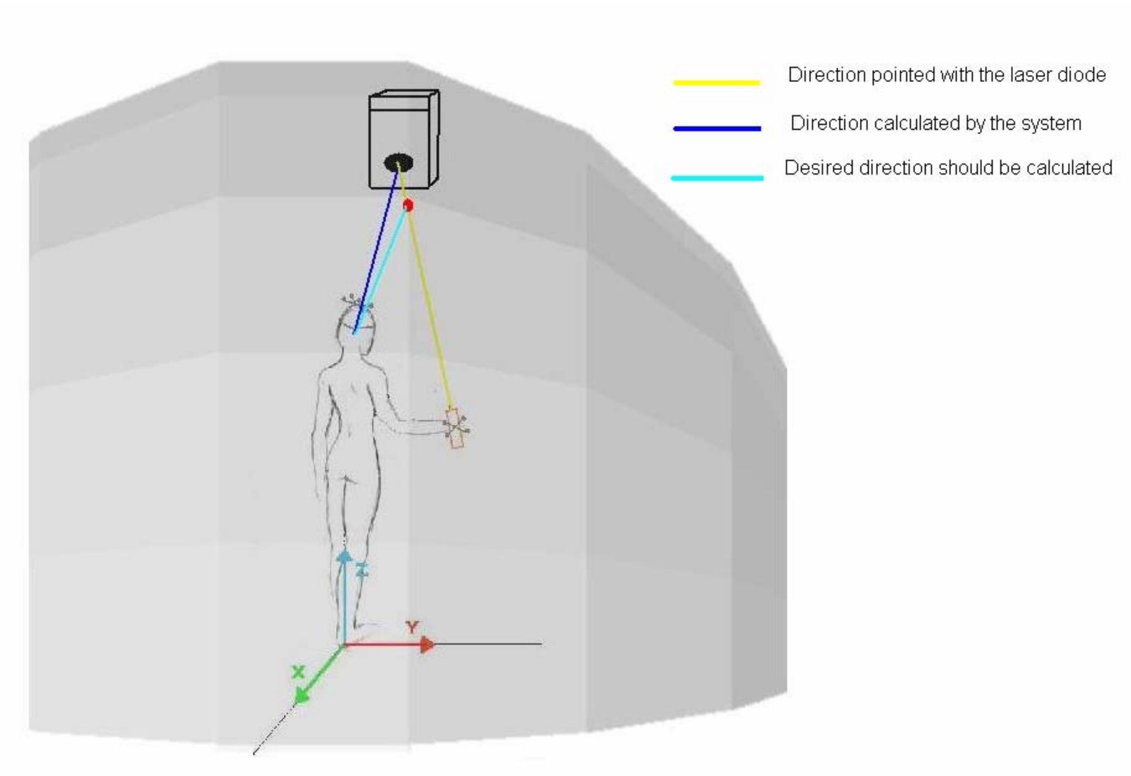
Elevation = 5.7º Azimuth = 0º

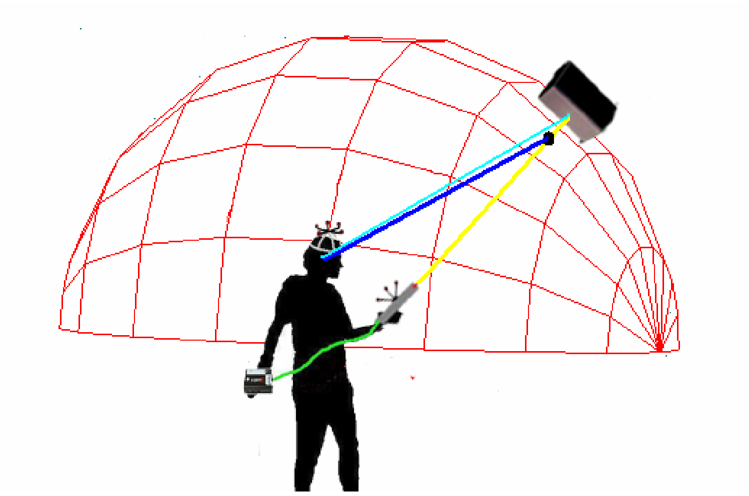
However if the calculations are worked out considering the exact loudspeaker position of 4,67m, the following results are obtained:

Elevation = 4.9º Azimuth = 0º

Since the orientations of the loudspeaker and the pointing device orientation are aligned, there is no error in the azimuth. The error made in the elevation is 0.8 º

In the next example the case in which the loudspeaker orientation and the pointing device are not aligned is supposed.





In this situation, a participant pointing correctly to a certain loudspeaker is supposed.
If this loudspeaker is located forwards or backwards of the hypothetic 4 meters semi sphere radius, the intersection point calculated will not coincide with the loudspeaker position. The intersection point will vary a bit and the azimuth and elevation between this point and the participants' head calculated afterwards will have some errors.
The next example shows how the error increases for pointing to the loudspeaker number one, when the pointing device is located in an arbitrary position. The length of the radius for this loudspeaker is 4,67m.

The positions of the points considered in this situation are:

Pointing device (-200, 0, 400) Head position (0, 0, 1700)

Loudspeaker position (0, 4670, 1400)

If the calculations are worked out for a loudspeakers sphere radius of 4 metres, the point obtained as the intersection between the straight line in the direction pointed and the 4 meters semi sphere radius is:

Intersection point (-36, 3810, 1216) $\approx$ Loudspeaker position (0, 4670, 1400)

Then, the elevation and the azimuth reckoned by Pd between this intersection point and the participants' head point are:

Elevation = -7.2º Azimuth = -0.5°

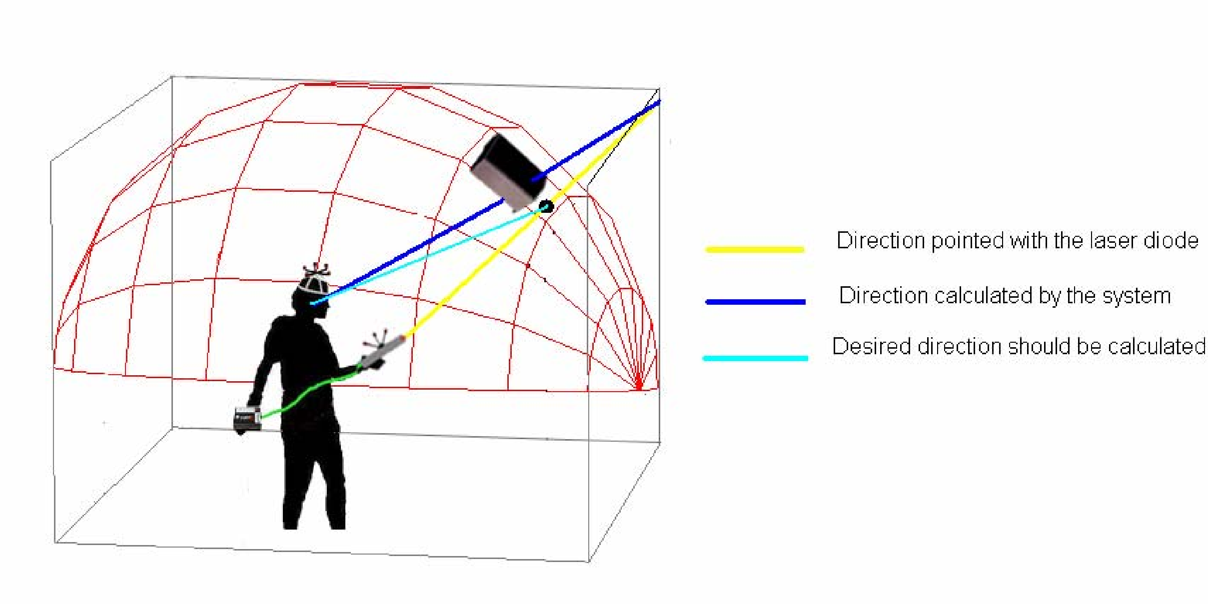
However if the calculation had been made with a radius of 4.67 meters, the intersection point would have been (0, 4670, 1400), where the loudspeaker is placed. The correct elevation and the azimuth considering a sphere radius of 4.67 meters are:

Elevation = -3.7º Azimuth = 0º

The error made in the azimuth is 3.6º and the error made in the elevation is 0.5º.
All these calculations were worked out with the PD program introducing the assumed coordinates as input data. These errors decrease as the pointing device position and the head position are placed near the semi sphere origin.

Another problem that might cause errors is that the IEM CUBE room is not an anechoic chamber. Therefore there is some reverberation left. The reverberation is the time that the sound persists in a particular space after the original sound is removed. If this time is very high, it means that there are a lot of reflections before the sound is extinguished.
The participants in the capture scenario perceive the directional sound and the reflections. The directional sound has the highest energy. However, if the room is very reverberant the first reflections that come from the wall in the same direction the sound is emitted will also have a considerable energy. It can cause the confusion of thinking the sound source is situated near the wall.
The participant can perceive a wrong localization of the sound source and point to another spot near the wall. In this situation the same calculation errors as the previous example occur. The participants point to a sound source position that is situated outside the hypothetic sphere and the calculations are made for a 4 m sphere radius.

The following picture depicts this situation:



Accuracy of the method

In order to determinate the accuracy of the method, an evaluation test was carried out. . Since the loudspeakers elevation and azimuth with respect to the sphere centre is known, the evaluation process consisted on pointing to theses positions and comparing the results.

The headgear object which is the system reference to make the calculations was situated in the sphere centre. No participants wore the headgear device. It was placed on the top of a tripod in order to avoid the inevitable body movements when a subject stays in a position wearing the object. In this way is assured that all the measurements are made with the same reference.

One by one all the loudspeakers' positions were pointed with the laser diode and the results were written down to a file. The exact point which had to be pointed was the centre of the loudspeakers' membrane where the sound was originated. The system measured the direction between the positions of the loudspeakers and the headgear device. The data corresponding to the position (x,y,z) and orientation (elevation and azimuth) of the stick and the headgear device were also registered. In this way, it is possible to know the localization of the two devices and make corporations about how aligned is the stick position and the loudspeaker pointed and also how far is the headgear device from the sphere centre.

The following table shown the output data obtained during the evaluation test:

Loudspeaker number	Elevation	Azimuth	Elevation measured	Azimuth measured	Error (elevation)	Error (azimuth)
1	0	0	0,41	-0,32	0,41	-0,32
2	0	22	0,30	22,79	0,30	0,79
3	0	45	0,16	46,56	0,16	1,56
4	0	71,5	0,16	71,16	0,16	-0,34
5	0	102	0,95	101,27	0,95	-0,73
6	0	138,5	-0,81	135,73	-0,81	-2,77
7	0	180	-1,03	180,45	-1,03	0,45
8	0	221	-1,44	219,25	-1,44	-1,75
9	0	258	-1,15	257,31	-1,15	-0,69
10	0	288	-1,30	289,00	-1,30	1,00
11	0	315	-1,57	314,42	-1,57	-0,58
12	0	338	-0,16	337,29	-0,16	-0,71
13	28	25	28,27	22,91	0,27	-2,09
14	28	69	26,96	68,53	-1,04	-0,47
15	28	116	26,06	114,64	-1,94	-1,36
16	28	155	27,55	155,37	-0,45	0,37
Loudspeaker number	Elevation	Azimuth	Elevation measured	Azimuth measured	Error (elevation)	Error (azimuth)

17	28	205	28,04	202,25	0,04	-2,75
18	28	244	27,49	243,65	-0,51	-0,35
19	28	296	27,10	295,89	-0,90	-0,11
20	28	335	27,49	336,73	-0,51	1,73
21	59,5	45	55,58	47,82	-3,92	2,82
22	57	135	58,20	132,53	1,20	-2,47
23	57	225	58,29	224,44	1,29	-0,56
24	59,5	315	59,05	316,89	-0,45	1,89

Table 1: Loudspeaker position, position pointed with the stick during the evaluation test and error made during the process

speaker number	stick-elevation	stick azimuth	stick-x	stick-y	stick-z	head-elevation	head-azimuth	head-x	head-y	head-z
1	1,55	-1,27	49,39	238,92	1232,17	-0,59	0,41	-7,74	-7,99	1300,33
2	1,70	20,57	234,42	218,17	1216,68	-0,55	0,53	-7,45	-7,39	1300,56
3	2,05	45,03	229,88	86,70	1183,90	-0,65	0,36	-7,46	-8,00	1300,22
4	2,83	69,17	300,70	30,98	1144,35	-0,61	0,42	-7,22	-7,66	1300,77
5	1,56	102,28	472,38	-43,88	1273,71	-0,54	0,53	-7,21	-7,50	1300,59
6	1,99	140,15	432,41	-77,59	1126,92	-0,56	0,52	-7,52	-7,37	1300,82
7	3,15	178,69	77,18	-628,00	1056,82	-0,56	0,29	-7,32	-7,59	1300,61
8	3,57	222,62	218,27	-106,76	962,03	-0,65	0,18	-7,35	-7,60	1300,34
9	3,03	259,77	227,73	-112,16	1011,14	-0,57	0,27	-7,35	-7,46	1300,83
10	3,10	284,81	450,31	195,33	987,59	-0,58	0,42	-7,74	-7,99	1300,12
11	3,91	312,02	177,20	97,55	930,75	-0,60	0,15	-7,35	-7,47	1300,49
12	4,18	333,21	232,79	208,25	1022,67	-0,63	-0,04	-7,57	-7,54	1300,31
13	31,15	18,67	236,56	-10,50	1141,31	-0,63	0,06	-7,34	-7,68	1300,67
14	33,33	65,30	234,61	-62,85	979,05	-0,61	0,18	-7,35	-7,65	1300,29
15	32,08	113,95	210,62	-101,98	1035,81	-0,58	-0,07	-7,88	-7,85	1300,47
16	34,87	160,99	346,00	-47,95	924,47	-0,58	0,08	-7,44	-7,46	1300,14
17	32,77	207,96	362,40	22,60	889,42	-0,55	0,02	-7,76	-7,65	1300,71
18	31,66	246,38	255,56	-31,74	907,32	-0,60	-0,07	-7,74	-7,46	1300,82
19	29,21	291,57	249,25	164,49	1082,14	-0,64	0,07	-7,64	-7,35	1300,55
20	30,74	330,94	263,05	186,59	1115,39	-0,63	0,05	-7,35	-7,35	1300,34
21	66,96	28,73	682,99	120,41	1024,84	-0,62	-0,15	-7,95	-7,35	1300,49
22	62,07	138,47	233,94	17,47	1168,26	-0,55	-0,04	-7,39	-7,23	1300,57
23	55,33	229,14	300,79	92,91	1195,90	-0,59	-0,12	-7,58	-7,99	1300,61
24	57,93	307,20	315,16	106,78	1155,49	-0,61	0,04	-7,55	-7,96	1300,78

Table 2: Orientation and coordinates position of the Stick and the headgear device.

The error calculated for all the cases depends on the position of the loudspeakers and the pointing device. It also depends on the distance each loudspeaker is situated with respect the sphere centre. The farer the loudspeaker is located from the hypothetic 4 meters semi sphere, the higher the error will be.

The lower error is obtained for the loudspeaker number 4 which have an elevation of 71.5º and an azimuth of 0º. In this case the pointing device and the loudspeaker orientation are quite aligned and the distance between its position and the sphere centre is 4.02. The higher error is made for the loudspeaker number 21. It has and elevation and azimuth of 45º and 59.5º respectively and the distance length between this loudspeaker position and the sphere centre is 5.56m, this is considerably higher than 4 metres.

The maximum error in the azimuth is obtained for the following loudspeakers:

Louspeaker number	Azimuth	Azimuth error
3	45	1,56
6	138,5	-2,77
8	221	-1,75
13	25	-2,09
15	116	-1,36
17	205	-2,75
20	335	1,73
21	45	2,82
22	135	-2,47
24	315	1,89

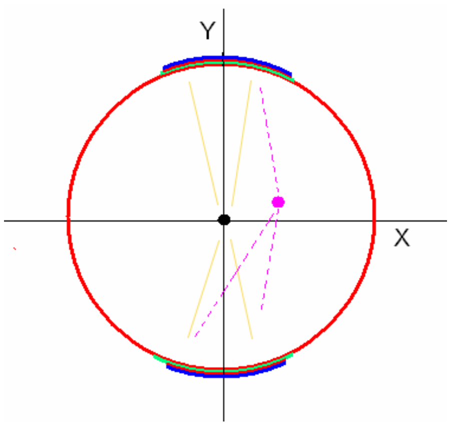


Table 3: Loudspeakers with the highest azimuth error

In table 3 can be seen, that the highest azimuth errors are obtained for loudspeakers which are situated near the y axis. For these positions, the stick and loudspeakers' orientation are the most misaligned, since the pointing device were moving around the area where the x and y position coordinates are positive.
As can be seen in table 1 the loudspeaker number 7 should be located at an azimuth of 180 degree. The loudspeaker is located along the y axis and therefore a bit misaligned with the stick, nevertheless for this case the stick was located near the y axis as shown the table 2. Consequently the loudspeaker and stick orientation are quite aligned and the error is not very high.

The error calculated for the elevation is more or less constant for all the cases. Only in the case of loudspeaker number 21 the error value deviates from the others (see Table 1). The distance from this loudspeaker to the centre is higher than 4 m and it is not aligned with the stick. Even so the reason could be because of the loudspeaker wasn't pointed in the membrane centre or because the loudspeaker position is a bit displaced from the original documented position.

Practical evaluation test in the IEM Cube

Room conditions

The experiment was performed at the IEM CUBE. A loudspeakers layout to reproduce the stimulus and the Vicon tracking system are implanted there.
The loudspeakers layout is approximately a semi sphere sectioned in three horizontal rings.
The lower ring with twelve loudspeakers has the highest number of loudspeakers. It is because the human auditory system is more sensible and can localize better the direction of the sounds in the horizontal plane. The middle ring is built of eight loudspeakers and the top ring hat four loudspeakers. However, not all the loudspeakers are placed at the same distance from the semi sphere centre. Despite all this, there are some irregularities as it was commented before. (The deviation of the loudspeaker positions are compensated by delays and gains after measurements of the impulse responses of each loudspeaker at the reference hearing position).
The room dimensions are (??? 119,88m^2 (11x11 no rectangular shape) and approx. 450m^3) and the reverberation time is not very high (can be found enclosed).

Audio-engine

The audio-engine is the tool implemented in the IEM CUBE used to render virtual sources onto the loudspeakers layout. This process is divided in two parts:

1. The encoding part: In this part, the three-dimensional positional information of the virtual sources is stored in a specific manner on a certain number of channels.
2. The decoding part: In this part, the signal to feed the loudspeakers are calculated and applied. The decoding process depends on the loudspeaker layout where the virtual sound will be rendered.

All these computations were performed with Pure Data. The technique used for recoding and reproducing high-fidelity sound over the loudspeakers array is called Ambisonics. There are some patches programmed with PD in order to manage the virtual sources encoding and decoding.
The Ambisonics technique is based o the idea to decompose the sound field into its spherical harmonic components. Further details can be found e.g. in the IEM Report 15/03 (Reference: Thomas Musil et al., Ambisonic 3D-Beschallungssystem 5.Ordnung für PD, IEM Report 15/03.)

Subjects

Four individuals participated in the experiment. Two of them were male and the other two female between the ages of 20 and a 25. Neither of them had not a previous experience in sound localization experiments and was naive about the kind of the stimulus played. Before the experiment started they were submit a previous training test to get familiar with the operating way. They had a normal hearing although it was not verified by a doctor certificated.

Test signal

The stimulus used to be encoded and decoded onto the loudspeakers array was a pink noise. It was performed as 8 temporal equal spaced pulses of approximately 0.4 seconds duration. The intensity of the pulses varied from 0 to 100%. The value of the higher intensity that the stimulus was presented to the participant could be varied with PD. The full length of the stimulus was 4 seconds and the sampling frequency was 44.1 kHz.

Experimental conditions

The participants were located in the center of the room facing the positive Y coordinate axis. They found the positive x direction to their right and the negative to their left. The Z axis was positive for values above the XY plane and negative for values below.

One by one were seated down in a chair in the centre of the room in order to situate their heads in the centre of the loudspeakers sphere. They were said not to move long distance the position of his head. Nevertheless they could rotate a bit the head so as to improve the accuracy of the sound source localization.
The curtains in the back part of the room were covering the window making the cameras to record precisely the capture objects.
The listening test was made up of a list of 160 points (below you say that there were just 126 ?!). Each point represents a different direction introduced into the audio-engine in order to render the virtual sound source in this exact position.
The procedure for each point during the listening text was:
Firstly the sound was generated from the loudspeaker in a certain direction; secondly the participant had to point with the handheld device's stick to the perceived direction the sound was coming from. Finally when the sound source position was recognized and located with the pointing device, the handheld device's button had to be pressed.
Each sound which corresponds with one point was played only once.

Results obtained

The positions where the virtual sources were rendered were located in three different planes. One is the median plane that passes through the symmetric axis of the participant's body. The sounds of the sources located in this plane run the same distance until they reach the participant's ears. The other two planes are 15º and -15º turned round with respect the median plane.

The number of points played for the median plane is 90, whereas the number of points of the other 2 planes is 18 (see above). The elevation of the sources positions used in the listening test goes from 0° to 90° with increments of 10°.

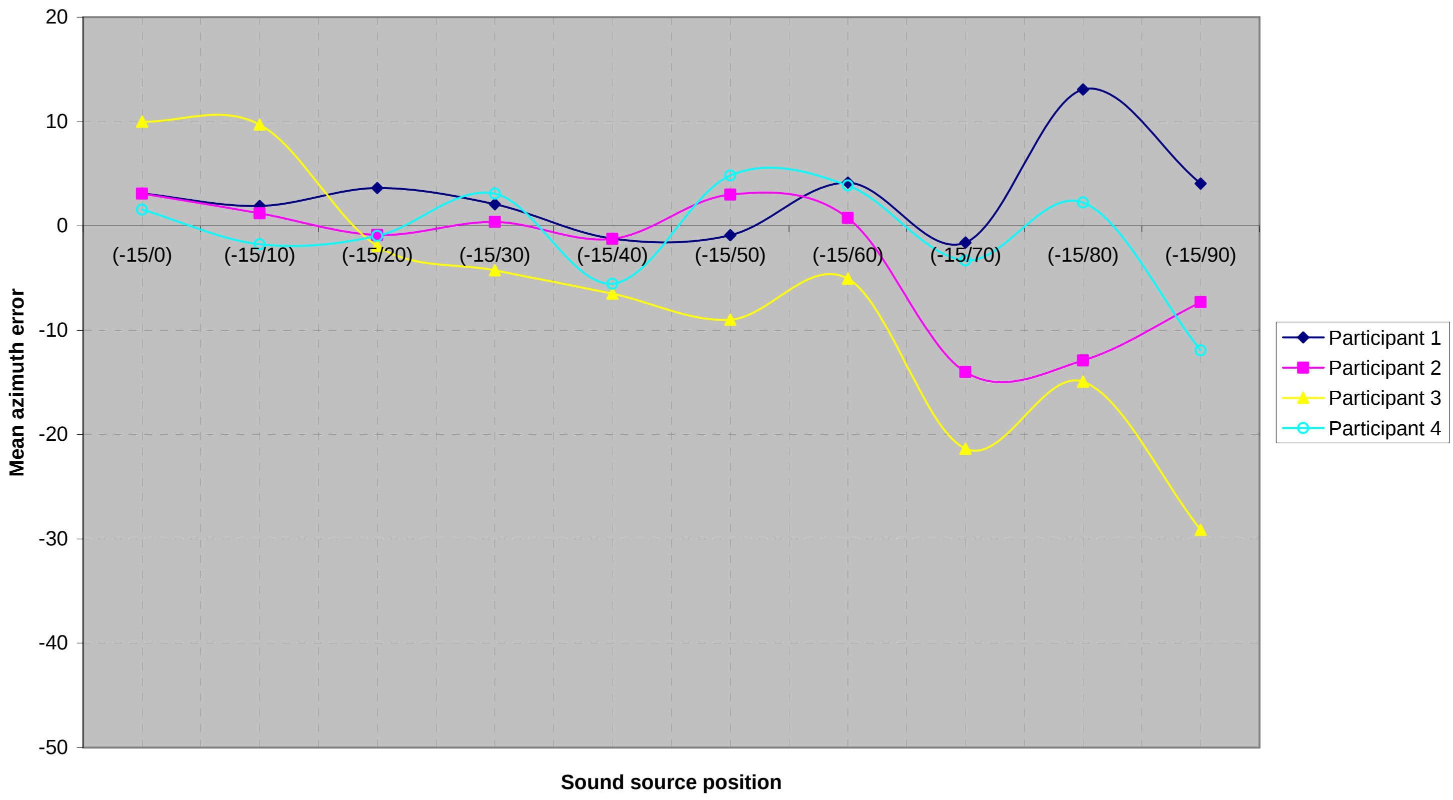
Plane1	Plane 2 (Median Plane)	Plane3
(-15,0)	(0,0)	(15,0)
(-15,10)	(0,10)	(15,10)
(-15,20)	(0,20)	(15,20)
...	…	…
...	…	…
(-15,80)	(0,80)	(15,80)
(-15,90)	(0,90)	(15,90)

whereby the data in brackets represents (azimuth, elevation)

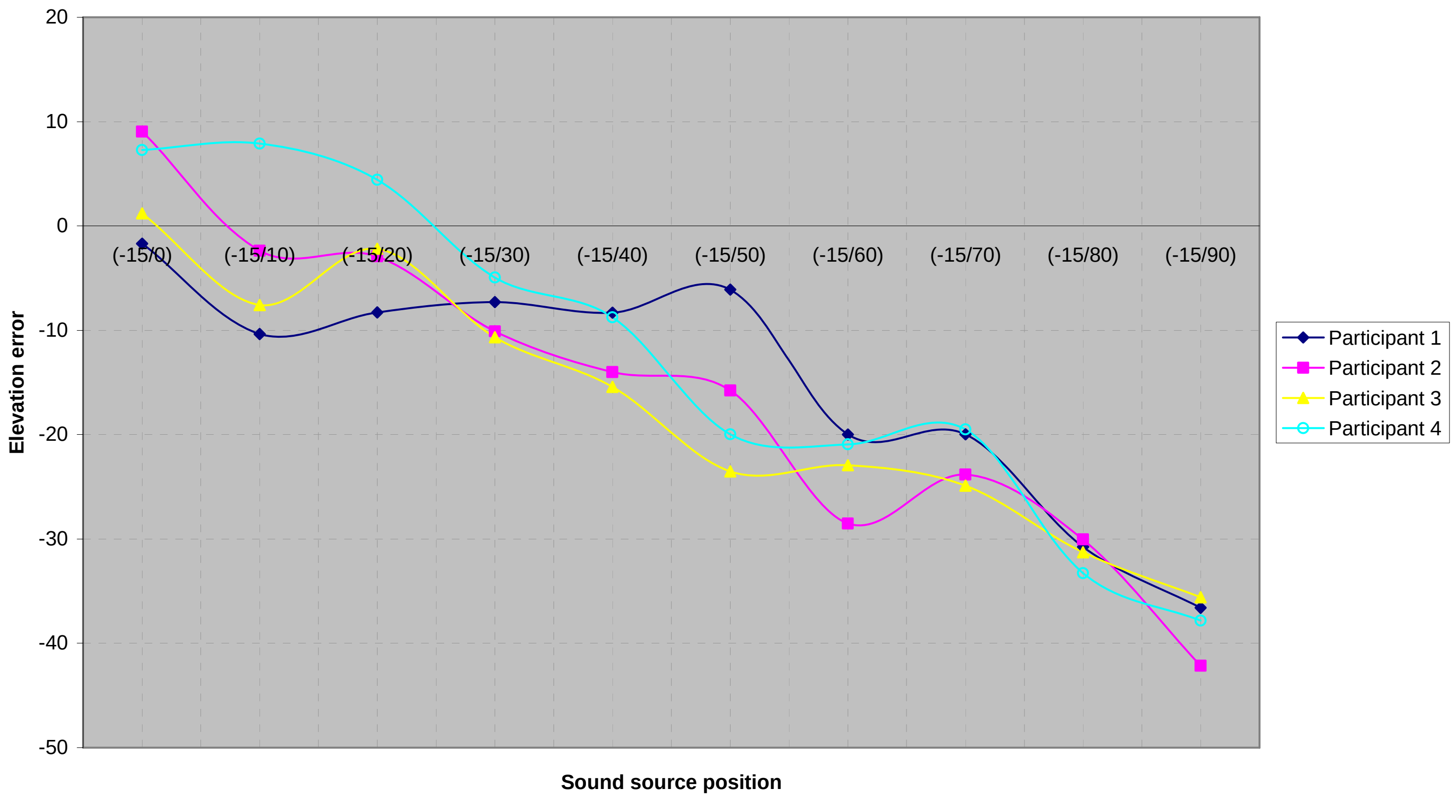
In the analysis of the results the elevation and the azimuth of each plane is separately evaluated. The mean elevation and azimuth error made by the participant in the positioning task for each sound source location is calculated. This error is positive when the point pointed is situated at the left side of the virtual source location and positive when it is situated at the right side.

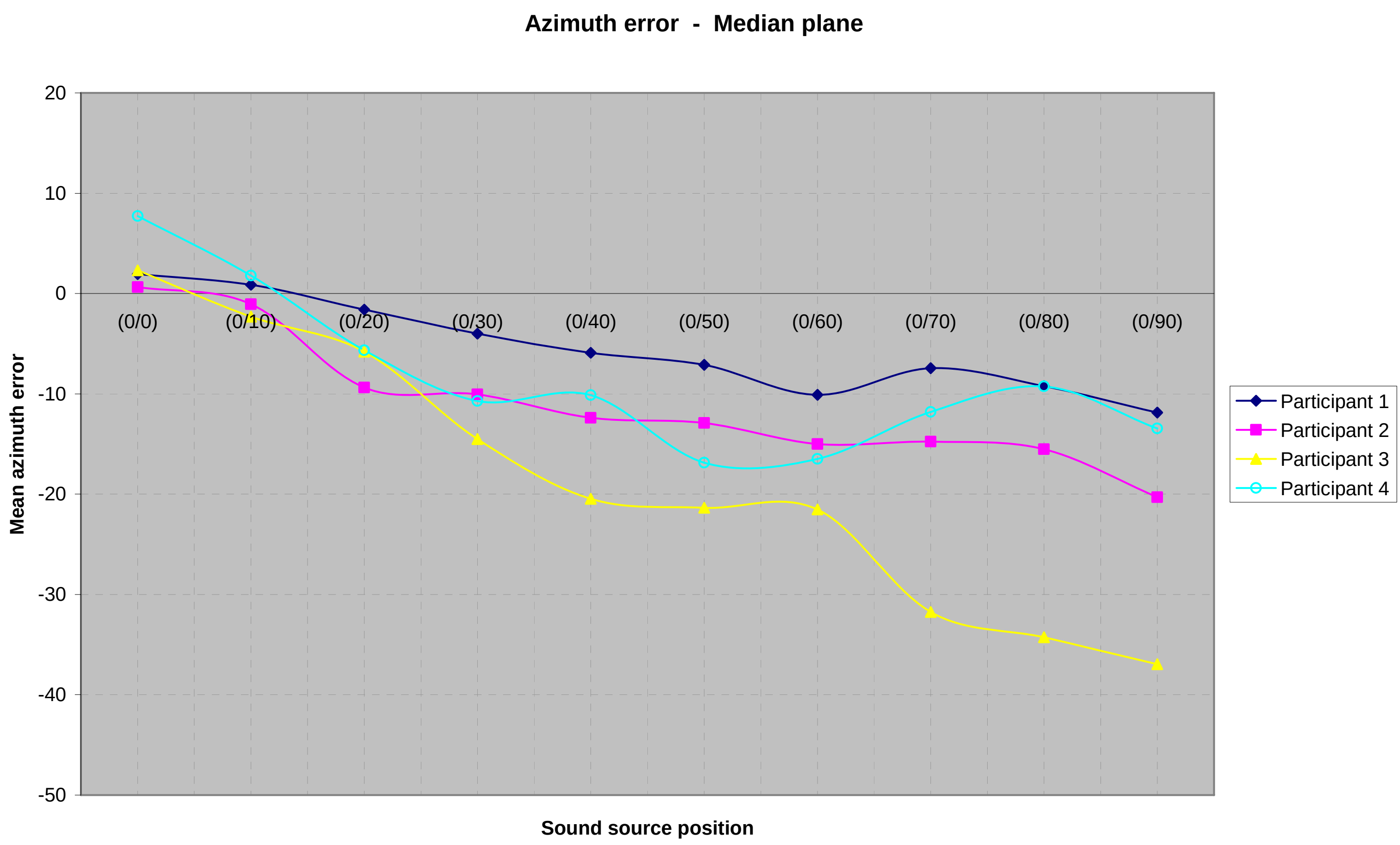
The results obtained are shown in the following graphics:

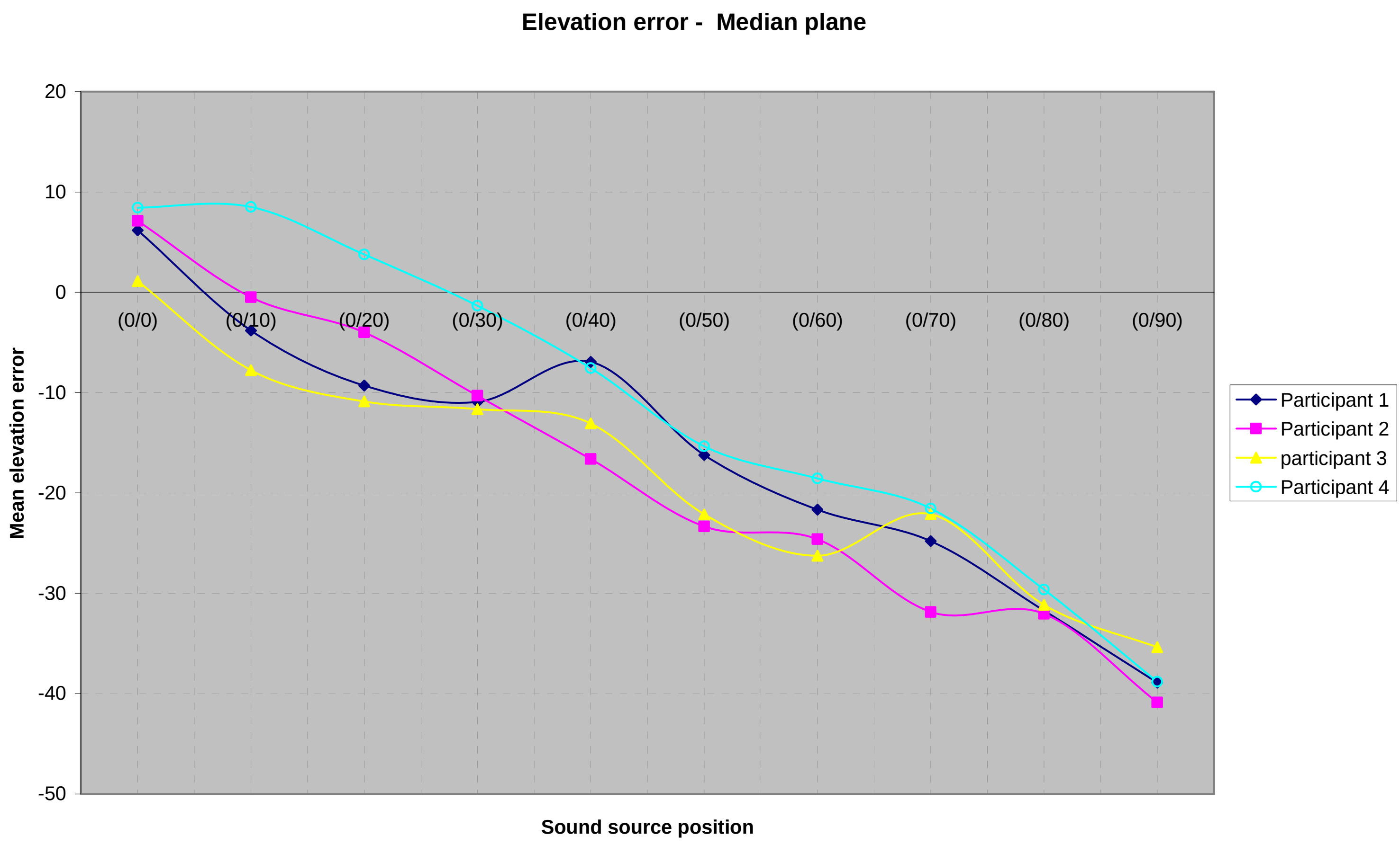
Azimuth error - Plane 1

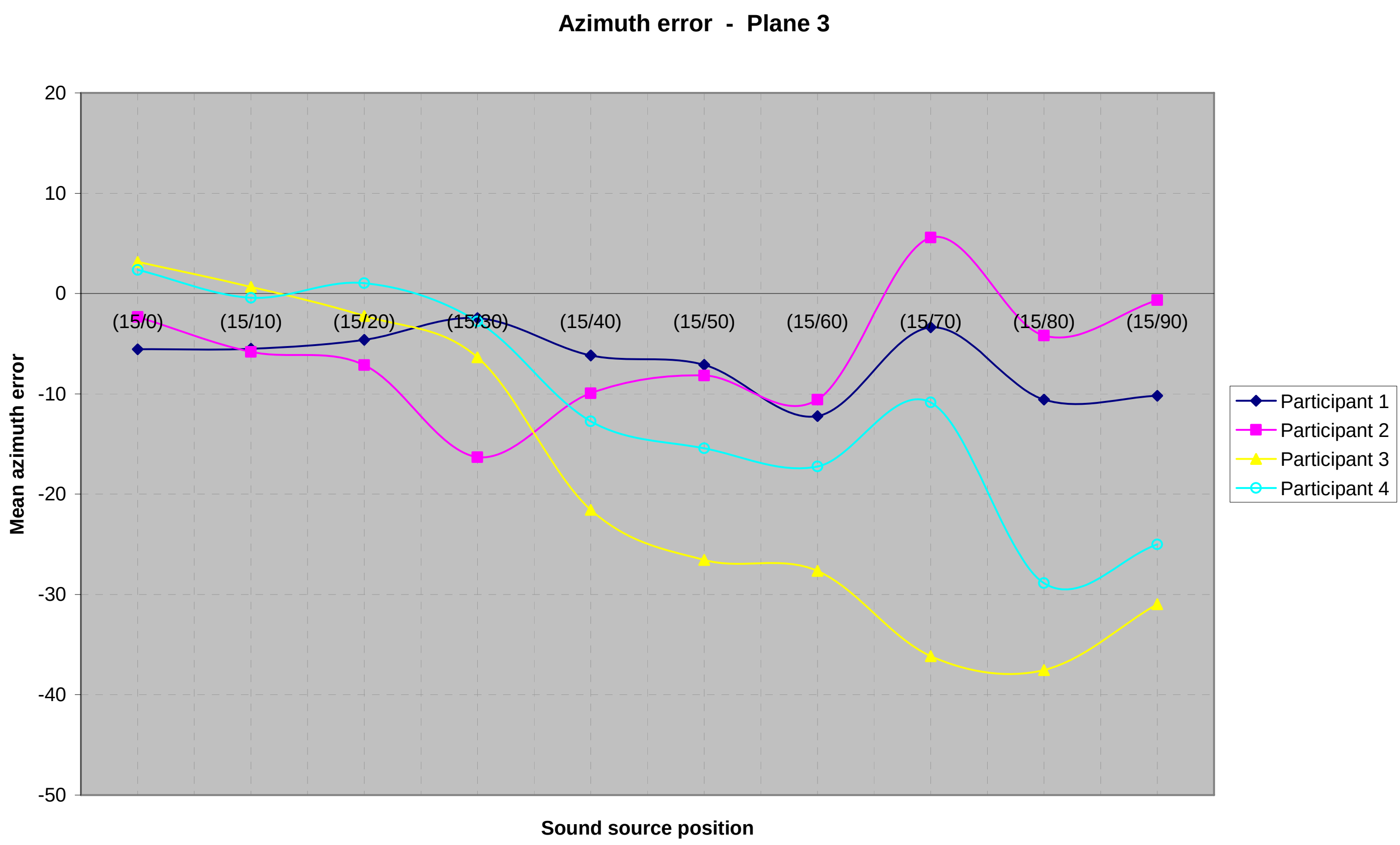


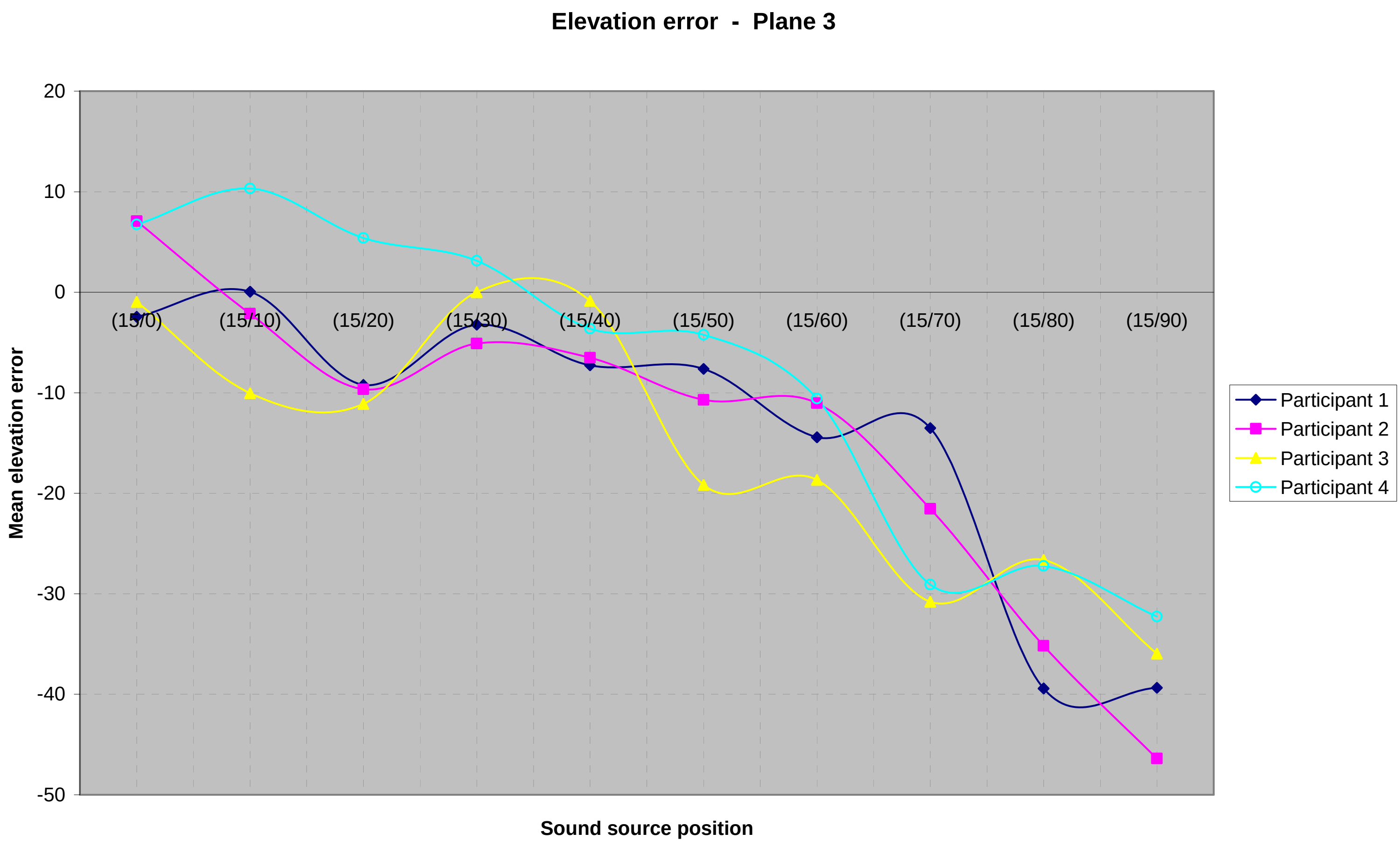
Elevation error - Plane 1




Azimuth error - Median plane
Mean azimuth error
20
10
0
-10
-20
-30
-40
-50
(0/0)
(0/10)
(0/20)
(0/30)
(0/40)
(0/50)
(0/60)
(0/70)
(0/80)
(0/90)
Sound source position
Participant 1
Participant 2
Participant 3
Participant 4


Elevation error - Median plane
Mean elevation error
20
10
0
-10
-20
-30
-40
-50
(0/0)
(0/10)
(0/20)
(0/30)
(0/40)
(0/50)
(0/60)
(0/70)
(0/80)
(0/90)
Sound source position
Participant 1
Participant 2
participant 3
Participant 4


Azimuth error - Plane 3
20
10
0
-10
-20
-30
-40
-50
Mean azimuth error
(15/0)
(15/10)
(15/20)
(15/30)
(15/40)
(15/50)
(15/60)
(15/70)
(15/80)
(15/90)
Sound source position
Participant 1
Participant 2
Participant 3
Participant 4


Elevation error - Plane 3
20
10
0
-10
-20
-30
-40
-50
Mean elevation error
(15/0)
(15/10)
(15/20)
(15/30)
(15/40)
(15/50)
(15/60)
(15/70)
(15/80)
(15/90)
Sound source position
Participant 1
Participant 2
Participant 3
Participant 4

Results analysis

The graphics show a wide variety of results depending of the plane and the participant carrying out the sound localization test. However some similarities in the results can be perceived. For example, in the majority of cases the elevation error is positive and close to 0 when the elevation is 0 degrees. This position is in average perceived around 5º at the right side from the original position.

The elevation error negatively increases as the elevation increases. The justification that explains this error is because of the positions where the participants' head and the pointing device are placed. The participants' head were located around the coordinates (0,0,0) and the pointing device were hold with one hand at the front, for that reason, although the participants point with the stick 90º in the vertical direction, the output elevation obtained is not 90º but less. The point that must be pointed so that the straight line between the intersection point and the participants' head has an elevation of 90º is located in the loudspeakers sphere up to the participant position. Therefore, the stick elevation to point to this spot has to be higher than 90º.

Another noticeable fact that should be commented is related with the azimuth. In the most of the cases the azimuth pointed is placed at the left side of the virtual source position. It makes a negative error as depicts the graphics. This error increases for high elevation. This fact can be explained because the participants were pointing with the right hand to the perceived position. In this pose the arm movement tends to go to the left.

It is also noticeable that the data obtained is considerably acceptable until the elevation reach a value of 30º. In these measurements, both the elevation and the azimuth error for the majority of the cases is less than ±10º.

As the elevation of the rendering sounds increases from 30º, the output elevation obtained follows a descending line until reach a minimum around -40º. In these measurements, the output azimuth obtained varies completely depending on the participant and the plane with a tendency to be negative.

Conclusion (Outlook)

A new method to perform localization measurements of sound sources in a three dimensional space is proposed. All components required and the relations between them to carry out the method implementation are described.
The approach introduced is based on the use of a pointing device tracked by cameras. The pointing device with a laser diode mounted on it is used by the participant to point to the desired direction the sound is coming from. The Vicon tracking system measures the localization of the pointing device and the headgear device. After some calculation the system implemented obtains the straight line between the participants' head and the point pointed.
The characteristic of the method make it more efficient in comparison with other methods. First of all, the optical cue provided by the laser diode helps the participants to know where they are pointing to. There is no ambiguity to know where the position pointed is. Furthermore unlike other method that uses diverse head apparatus to point with the head to the perceived position, here there is no mobility restriction to indicate with the stick the desired direction. Moreover the problem to look with the eyes instead of moving the head to point at the sound is also avoided. The system implemented tracks the motion of the stick and the wireless headgear device so that the positioning task is more comfortable.
The method described estimates the sound localization with high fidelity and precision as it is shown in the system evaluation section. The main problem that introduces errors in the final solution is the assumption of the bounding hemisphere . The calculation worked out by the system taking into account a loudspeaker hemisphere of 4 meters, nonetheless the distance between the loudspeaker and the volume origin varies from 3,1 to 5,89 m. The accuracy can be increased by using a more complex room model which might fit more precise to the real room.
The new localization method is used in the field of virtual acoustics localization. A sound localization experiment using the system was carried out in the IEM CUBE. The procedure and the results are here shown.

References

[1] Pauli Minnaar, Søren Krarup Olesen, Flemming Christensen, Henrik Møller "The importance of the head movements for binaural synthesis." Department of Acoustics, *Aalborg University, Denmark. Presented at Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001.*

[2] Ville Pulkki "Compensating placement of amplitude-panned virtual sources" *Laboratory of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015 HUT, Finland*

[3] Elizabeth L. Wenzel "Effect of increasing system latency on localization of virtual sounds" NASA Ames Research Center, Moffett Field, CA, USA

[4]B. Razavi, W.E. O'Neill, Gary D. Paige " Both Interaural and Spectral Cues Impact Sound localization in Azimuth" *Depts. of Biomed. Eng.1 and Neurobiology & Anat.2 and the Centers for Navigation & Communication Sciences3 and Visual Science4, Univ. of Rochester, Rochester, NY USA.*

[5] Russell L. Martin, Ken I. McAnally, and J.P. Watt "Localization of multiple sounds as a function of the duration of the inter-sound gap" *Defence Science and Technology Organisation ,Australia. Presented at the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.*

[6] http://www.vrlogic.com/html/polhemus/3space_fastrak.html 3space Fastrack webpage

[7] http://www1.produktinfo.conrad.com/datenblaetter/175000-199999/187585-da-01-en-LASERMODUL_670_NM_LAS67_01_L.pdf Laser diode specification

[8] http://www.infusionsystems.com/support/icubex-111-manual.pdf I-cube system manual

[9] http://www.vicon.com/ Vicon motion systems web page

[10] http://iem.at/services/studios/cube/anleitung/ Friendchip device expicifications
http://www.friend-chip.de/index.html

[11] www.puredata.org The Pure Data portal

[12] Florian Hollerweger "An introduction to higher-order Ambisonics" *April 2005*

[13] Johannes Zmölnig, Winfried Ritsch, Alois Sontacchi, Robert Höldrich „The IEM Cube"*Institute for Electronic Music and Acoustics, Graz, Autsria.*

[] add the IEM Report ….