



LEMOS Lucas
Informatique et Electronique des Systèmes Embarqués
POLYTECH GRENOBLE
Rapport de stage IESE5

**Study of an echo-location method and system for
device-to-device positioning**

Tome principal

2017/2018
Stage du 05 février au 31 juillet 2018

Contents

1 Résumé	2
2 TDK Invensense	3
2.1 Sustainable development	4
3 The internship	5
3.1 Subject	5
3.2 Initial state	5
3.3 Plan	7
4 Personal contribution	7
4.1 Literature review	7
4.2 TOF approach	8
4.2.1 Evaluation	9
4.3 TDOA approach	11
4.3.1 Evaluation	19
4.4 VICON system	22
4.5 User interface	22
4.6 Results	26
4.6.1 Local experiments	27
4.6.2 Anechoic chamber experiments	29
5 Conclusion	32
A Deduction of the 3D positioning equations	34

1 Résumé

Le marché de la réalité virtuelle se présente encore comme le futur. Même si les solutions ne sont pas encore devenu grand public, une partie important des ressources sont utilisé dans ce domaine. Dans ce genre de système, la précision entre la position relatif de la manette et le casque est fondamentale pour expérience utilisateur agréable. Les solutions d'aujourd'hui utilisent une caméra pour avoir cette positionnement de manière précise. Le stage se focalise dans le remplacement du rôle de la caméra dans ce genre d'application pour une matrice de microphones dans le casque plus un émetteur ultrason dans la manette.

Les premiers mois ont été complètement dédiés à la recherche des méthodes pour le positionnement relatif en 3D. Le sujet est assez riche, il y une vaste littérature disponible et le but c'était des trouver, parmi les différentes approches, quels étaient le plus pertinent au limitation du cadre du stage. Deux familles des méthodes ont été étudiés en détaillées: l'approche Temps de Vol (*Time of Flight* (TOF) en anglais) et l'approche Différence de Temps d'arrivée (*Time Difference of Arrival* (TDOA) en anglais).

Les méthodes TOF résolvent l'équation de base qui lie la distance, la vitesse et le temps. Elles partent du principe que le temps de vol ou le temps qui écoule entre l'émission et la réception du signal est connu. Comme la vitesse du son est constant (ou peut être considérer comme constant pour ce genre d'application), la distance entre l'objet qui a émis et l'objet qui a reçu peut être calculée. De cette façon, chaque microphone définit une sphère dans l'espace, dont la surface peut contenir la source du son.

En revanche, les méthodes TDOA utilisent une paire de microphones pour calculer le délais l'un par rapport à l'autre. Avec quelques prémisses qui seront détaillées dans la suite, chaque paire définit un cone dans l'espace, dont la surface peut contenir la source du son.

Les deux méthodes ont été implémentés en Matlab, le principe et résultats théorique seront discutés. En plus, le TDOA a été implémenté en *hardware*, dans une salle normale et dans une chambre anéchoïque.

2 TDK Invensense

Let's get started by having an overview of how the Grenoble site became part of the TDK Group. Back in 2007, a start-up named Movea was founded by researchers coming from a laboratory of CEA. Their idea was to use their experience in the motion domain to create some products and applications, integrating data coming from multiple data sources (accelerometers, gyroscopes, pressure sensors). Movea became a provider of motion sensing and data fusion software, firmware, and IP for the consumer electronics, particularly smart phones and tablets, sports and fitness and eHealth industries. Movea's proprietary SmartMotion™ technology - unique motion processing capabilities - enabled customers and partners to quickly add motion intelligence to their products, providing reduced risk, cost, and time-to-market advantages for delivering compelling new motion-based features that create more value and a more exciting user experience for the consumer.

Seven years after its foundation, the American Invensense acquired Movea. InvenSense is a world leading provider of Micro-Electro-Mechanical Systems (MEMS) sensor platforms found in mobile, wearables, smart home, industrial, and automotive products. More specifically, it produces most of the hardware used by Movea (pressure sensors, magnetometers, accelerometers, gyroscopes, microphones). Since Movea was essentially a software site, this acquisition merged software with hardware.

In 2017, however, the Japanese TDK Group acquired Invensense, as they saw the american company as a key part in their plan to invest in the IoT domain. Invensense catalogue of products was seen as an IoT-enabling business. Founded in 1935, TDK is a company whose main goal was to explore the possibilities to commercialize the ferrite, a recent finding at the time made by Tokyo Institute of Technology. Nowadays, TDK has more than 100 thousand employees around the world, with 90 % of them coming from companies acquired by TDK. The figure 1 shows the number of sales by product group.

In this context, my internship in Grenoble has been in one of the newly acquired companies: Invensense. I worked in the Algorithm & Motion team that develops software and firmware for all the products on its portfolio.

Consolidated sales by product group (unit: 100 million yen)

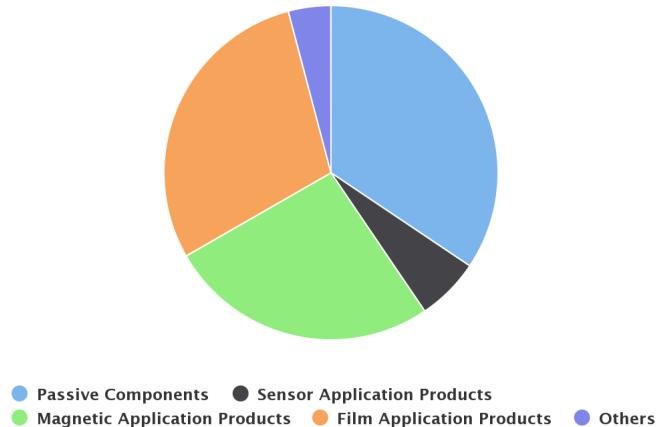


Figure 1: TDK sales by product group.

Source: <http://bit.ly/2zSxUSL>

2.1 Sustainable development

The TDK Group considers ensuring a safe, healthy and comfortable work environment as its first priority in the process of any of its business activities.

Compliance with applicable environmental laws is a social duty of any corporation. The TDK Group shall make use of internal company policies concerning environmental protection, established and adopted by each member organization of the TDK Group, in order to assure that the environment of the community in which each member organization of the TDK Group operates is properly protected. Each TDK Member is expected to fully understand and use internal company manuals or guidebooks for crisis management, established and adopted by each member organization of the TDK Group: (i) to prevent disastrous events or accidents that may be fatal to the continuous existence or development of the TDK Group; and (ii) once such a serious event occurs, to minimize and control the damage.

Regarding the reduction of CO₂ emissions in production activities (environmental load), TDK placed priority on strengthening energy-saving measures in China, which accounts for about 50% of TDK's CO₂ emissions, as well as establishing a setup to promote independent activities and coordinating energy-saving activities with the cost-reduction efforts of business sites in China. As a result, the TDK Group's emissions in fiscal 2014 amounted to 1.063 million tons. Regarding the expansion of CO₂ emission reductions by products (environmental contribution), as a result of expanding the calculable scope of offset environmental contributions by product group and field, emissions in fiscal 2014 amounted to 886,000 tons.

3 The internship

In order to provide an insight of the scenario in the beginning of project, the following sections will be presented chronologically. Firstly, the internship subject and the motivation for its existence will be introduced in the section 3.1. Next, the initial state as well as the available tools will be shown in section 3.2. Lastly, after having a good sense of why this internship was meaningful and to what extent it was developed, the section 3.3 will present how the activities were planned.

Also, throughout the report, vectors will be represented with an underlined variable (Ex: \underline{k}), whereas matrices will be represented with a double underlined variable (Ex: $\underline{\underline{A}}$).

3.1 Subject

One range of products presented in TDK InvenSense catalogue are microphones. With the arrival of products such as *Amazon Echo* (from Amazon), *Homepod* (from Apple) and *Google Home* (from Google), there is an expectation for an increase in the demand for microphones in the next few years. These applications compose a new category that has been called earable [1].

A second motivational reason is the fact that Virtual Reality (VR) headsets like *Oculus Rift*, *HTC Vive* and *PlayStation VR*, they all rely on cameras to track the relative position between headset and controllers. This precise tracking is one of the key factors for a pleasant user experience . This solution, however, has its drawbacks: it is sensible to ambient lighting conditions and occlusions, it drains lots of hardware resources (both computer and energy-wise). Also, *TDK InvenSense* does not produce cameras, but microphones.

With this in mind, the internship subject stimulates the research and development focused on VR market. More directly, the idea is to replace the role of the camera in such systems by a microphone array. Hence, it can be summarized as research, development and implementation of methods to locate a 3D sound source.

3.2 Initial state

The first step taken to have a feeling about the feasibility of this idea was to conceive a prototype. Given the broad range of possibilities to explore in only 6 months of a internship, some assumptions needed to be validated to ensure that it would worth an intern dedicated to this subject. Hence, when I arrived, I had access to the validation prototype and also to plenty of scientific papers regarding this subject.

The prototype was made using, basically, 4 digital microphones and a sound source. In a regular room, the prototype was fixed while the source kept moving. Two phase shift measures were computed - between U and D, and between L and R microphones. The goal was to translate the phase shift to an azimuth angle and,

consequently, to detect from which angle the sound was coming from. The figure 2 shows the prototype.

Thus, the idea is to evolve the algorithm to estimate the position, in the 3D space, of the sound source from a reference coordinate system. To accomplish such a task, I had access to a soldering station, electronics components (resistors, capacitors, PCB pins, wire, solder, etc) and more sophisticated devices (function generator and oscilloscope) according to their availability.

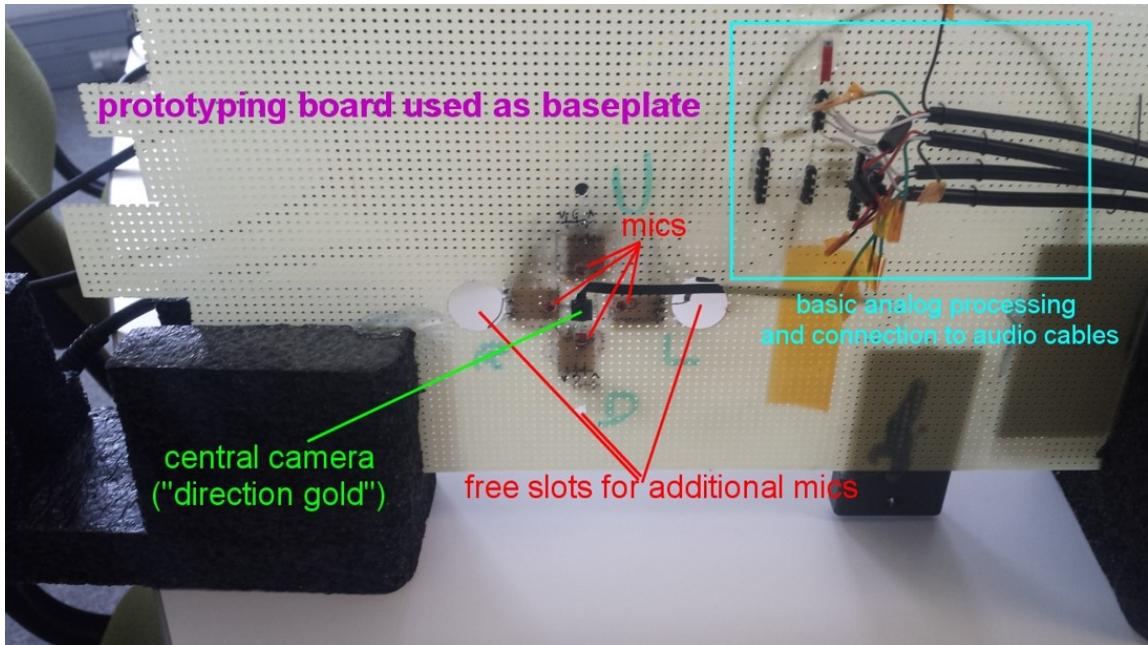


Figure 2: Prototype used to validate the assumptions about audio location methods. It can be seen 4 microphones (U, L, D, R), a camera mounted in the center of these 4 microphones. Also, the cables in the upper-right part connect each microphone to a sound recorder - in this case, a Zoom H6.

Source: The author

3.3 Plan

To reach the established goals, a planning was done to highlight the parts of the project that would be developed. Therefore, 25 weeks of internships were divided into small chunks, considering week 1, from 02/05 to 02/09 and week 25, from 07/30 to 07/31. Table 1 gives an idea of the activities to be done and their deadlines.

Table 1: Activities planned in the beginning of the internship.

Week	Activity
1 - 5	Research of spatial localization methods: Time of Flight (TOF), Time Delay of Arrival (TDOA) and Beamforming (BF).
1 - 5	Specification of a bill of materials to prototype the system.
5 - 15	Definition of the number of microphones, their spatial distribution and the waveform to be emitted.
15 - 18	Definition of the limitations of the chosen method.
18 - 20	Comparison of the real versus theoretical system.
20 - 25	Analysis about the possibility to implement in a real product and elaboration of a final report regarding the internship.

4 Personal contribution

This section will be dedicated to the work developed during the internship. Section 4.1 starts with the literature review, showing the most relevant papers used to develop the solutions proposed. Next, two different techniques to find a 3D point in space are presented. First, section 4.2 demonstrates how the TOF method was applied to this case scenario and, lastly, section 4.3 shows the TDOA approach, its theoretical and practical results.

4.1 Literature review

The beginning of the internship was all about searching papers and articles that talk about 3D audio location methods. In [2], [3] e [4], the authors use minimization techniques of the Cramer-Rao criteria as a way to find the optimal solution to distribute microphones in a room. However, there are no spatial restriction in their cases, which is a crucial factor in the internship.

In [5], similar principles of what will be presented are used, but it relies on iterative methods, which is not ideal in a real time case. Methods based on beamforming, as those presented in [6], [7] and [8], use variations of the optimum Maximum likelihood (ML) estimator that require the solution of non-linear problems.

Finally, the solution that most suits the needs and constraints of this project is the one presented in [9], where a closed-form method for microphones restricted to a plane is presented.

4.2 TOF approach

Methods based on TOF are the most straightforward ones to be understood from the physics standpoint, since they are derived from basic concepts. However, their implementation in a robust fashion can be cumbersome. The figure 3 depicts its basic scheme.

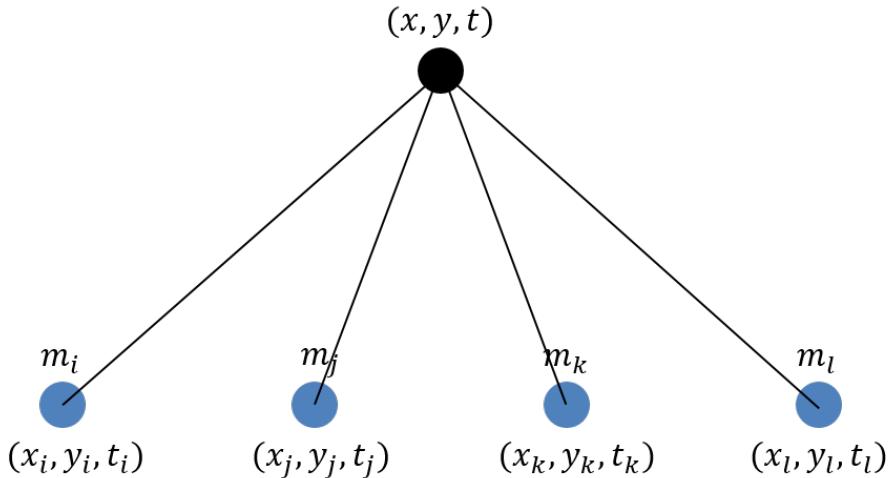


Figure 3: Representation of TOF approach in 2 dimensions, in which the sound source is depicted with a black circle, it has (x, y) coordinates and it emits a signal at time t . The microphones, blue circles, from m_i up to m_l are placed in the positions from (x_i, y_i) to (x_l, y_l) and they receive the emitted signal of the source at the times t_i up to t_l , respectively. The line joining the sound source and each of the microphones represents the straight line distance among them.

Starting from the simplest equation that links the distance d between two objects, the constant speed v which one approaches the other and the time elapsed δt to accomplish such a task, the equation 4.1 highlights the relation between these physical quantities.

$$v = \frac{d}{\Delta t} \quad (4.1)$$

Adapting to the notation shown in the figure 4.2, the equation 4.1 taking the sound source as a reference, rewritten for each of the four microphones, becomes the equation 4.2.

$$\begin{aligned} v(t_i - t) &= \sqrt{(x_i - x)^2 + (y_i - y)^2} \\ v(t_j - t) &= \sqrt{(x_j - x)^2 + (y_j - y)^2} \\ v(t_k - t) &= \sqrt{(x_k - x)^2 + (y_k - y)^2} \\ v(t_l - t) &= \sqrt{(x_l - x)^2 + (y_l - y)^2} \end{aligned} \quad (4.2)$$

It is worth noting that, in the real scenario, the positions of the microphones will be known as well as the sound speed. Besides, it is considered that the instants in which a microphone receives the signal emitted by the source will be estimated. Hence, for each of the equations, there will be 3 unknowns, which are: x, y, t (position and instant that the signal was emitted by the source). To solve the system, it is necessary at least 3 equations, consequently, at least 3 microphones. Nevertheless, as the equations are non linear, it was chosen to add a fourth microphone to allow the system to be linearized. The appendix A details the step by step adopted to linearize the system. The equation 4.2, after some manipulation, becomes the system below, where $r_i = x_i^2 + y_i^2$.

$$\begin{bmatrix} x_i - x_j & y_i - y_j & -v^2(t_i - t_j) \\ x_i - x_k & y_i - y_k & -v^2(t_i - t_k) \\ x_i - x_l & y_i - y_l & -v^2(t_i - t_l) \end{bmatrix} \begin{bmatrix} x \\ y \\ t \end{bmatrix} = \begin{bmatrix} 0.5 \{r_i - r_j - v^2(t_i^2 - t_j^2)\} \\ 0.5 \{r_i - r_k - v^2(t_i^2 - t_k^2)\} \\ 0.5 \{r_i - r_l - v^2(t_i^2 - t_l^2)\} \end{bmatrix} \quad (4.3)$$

The system form is $\underline{\underline{A}} \underline{p} = \underline{k}$, whose solution goes by a matrix inversion: $\underline{p} = \underline{\underline{A}}^{-1} \underline{k}$. In the process, though, it was necessary to adopt a microphone as a reference to the linearization process and, in this case, m_i was chosen. The precision of the solution, then, depends on two factors: first, the matrix $\underline{\underline{A}}$ must be invertible and, because of this, the microphone can't lie on the same plane. Second, the precision of the estimation of $t_i \dots t_l$ will be determinant to estimate the source position, since the microphones will be fixed and their position measured. Lastly, the extension to the 3D case is direct, the same principle is used, needing 5 microphones to linearize the system (more details on appendix A).

4.2.1 Evaluation

It was implemented, in Matlab, a system in which n microphones can be placed in any position in the 3D space, as well as a sound source, whose position is calculated using equation 4.3. From there, it was possible to study the case where there are errors in the estimations of $t_i \dots t_l$, having a emitting frequency (f), a sampling frequency (f_s).

It was known that estimations were the most difficult part of this approach, since it's not obvious to locate an event in time, specially when tweeters are used to generate mechanical waves. These devices are transducers that convert electrical signals into kinetic movement, and, by doing so, its intrinsic inertia comes into play. In practice, a fade-in and a fade-out effect is observed. Hence, when generating short period signals with such a hardware, these transient states have to be taken into account.

To evaluate the errors, it was simulated several scenarios where 5 microphones were placed in a place, respecting a rectangle with 115 mm x 70 mm. These dimensions are related to the area available in the VR headsets. One of these scenarios are shown in the figure 4. The m_1 microphone is used as a reference and from m_2 up to m_5 , a 1 μs error is introduced. Also, $f = 40$ kHz, $f_s = 1$ MHz and 2087 source positions 1 meter away from m_1 were simulated. This distance is a bit above the average length of a straight arm from a human adult.

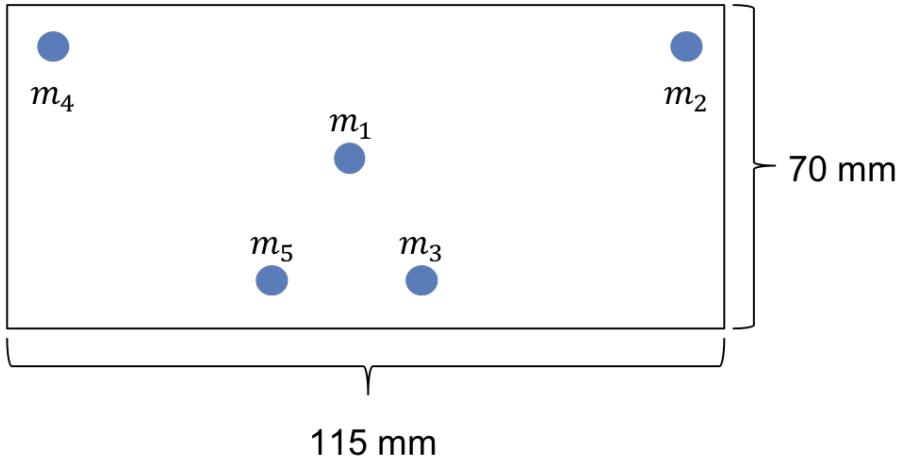


Figure 4: Spatial distribution of 5 microphones in one of the scenarios simulated.

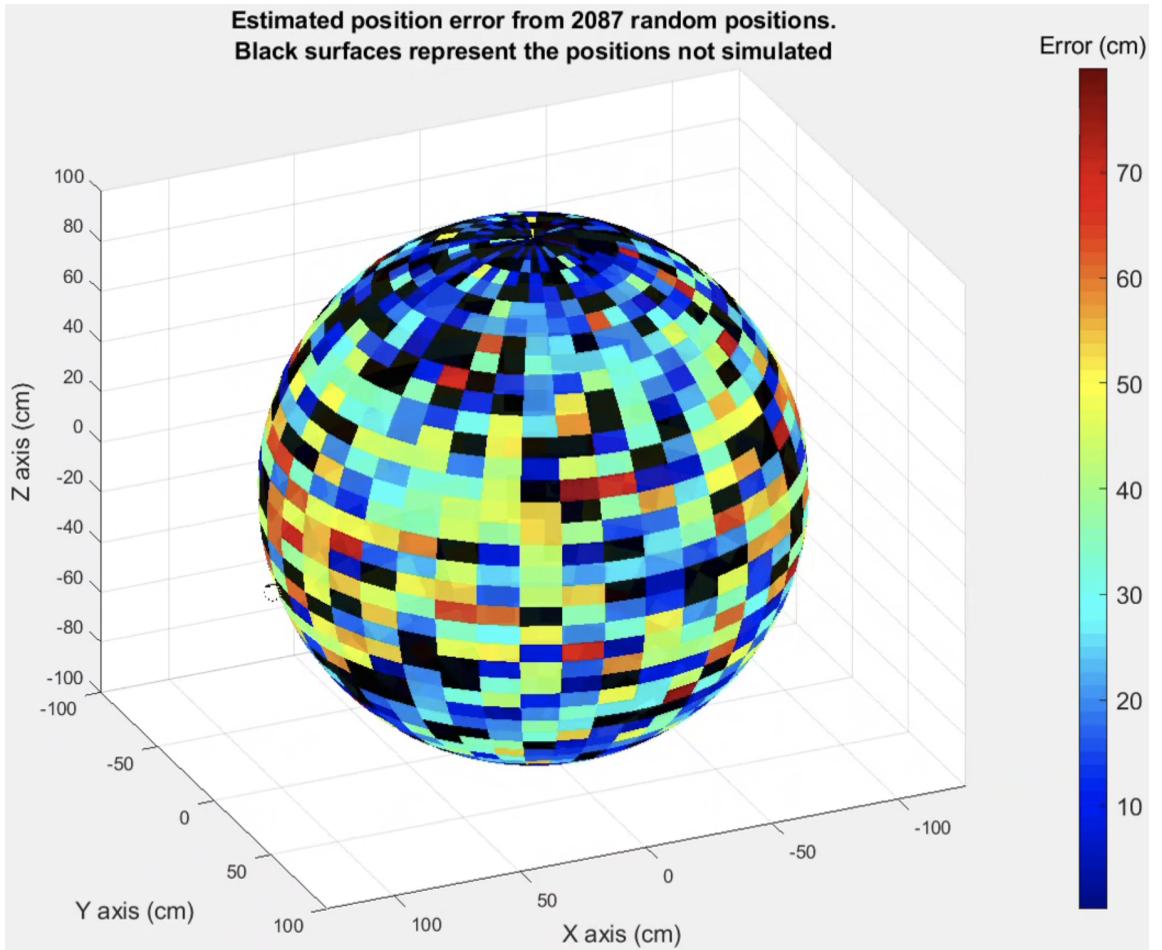


Figure 5: Sphere with radius of 1 m, centered in the origin. Each one of the colored regions depicts a source position simulated, except the black ones. The colorbar in the right part of the figure shows the error, in cm, in the estimated position.

The figure 5 shows the tendency of the error, according to the microphones' spatial distribution. It can be inferred that the accuracy in the polar region is, in general, better than the equatorial region. But, for most of the cases, as the scale in the colorbar shows, the error far above 1 cm, which is the desired accuracy. These results made as take a step back and deviate the attention to other methods, since it would need a high sampling frequency to reach such an accuracy and, with our basic prototypes, it would be almost impossible to reproduce such a scenario.

4.3 TDOA approach

For the TDOA based methods, the interest lies on the delay analysis between the signals from a pair of microphones, which can be seen as a relative timescale. Whereas, comparing to TOF based methods, where the analysis are made individually for each microphone, it can be seen as an absolute timescale. The figure 6 shows the scheme used to clarify the assumptions accepted to implement this method.

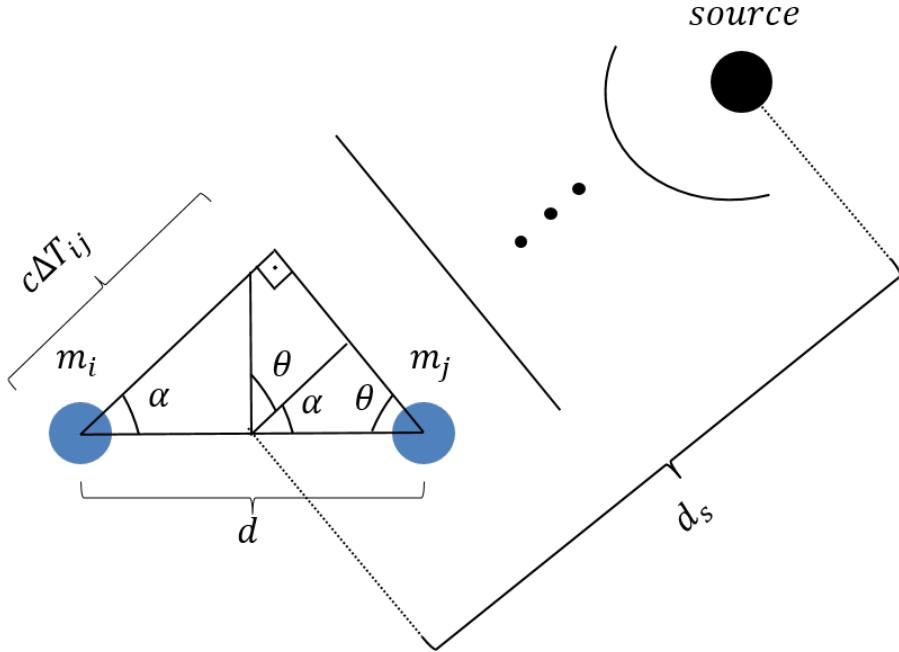


Figure 6: Representation of TDOA approach in 2 dimensions, with the sound source being represented by a black circle and a pair of microphones (m_i, m_j). The distance between the microphones is d , the distance between the source and center of the line that passes through the microphones is d_s and $c \approx 343$ m/s is the speed of sound.

Firstly, it is assumed that $d_s \gg d$, consequently, it is considered that the source is in the far-field region, where the spherical wavefront that propagates in space can be approximated by a planar one, as shown in the figure 6. In this case, when computing the delay (ΔT_{ij}) between m_i and m_j microphones and, also, knowing that the sound speed can be considered constant for the purposes of this application, the α angle can be computed as shown in the equation 4.4:

$$\alpha = \arccos\left(\frac{c\Delta T_{ij}}{d}\right) \quad (4.4)$$

This angle is known in the literature as the Angle of Arrival (AOA) [5], which reflects the direction from where the sound is coming from. In two dimensions, the AOA defines a line, whose origin is the midpoint of the segment that passes through m_i and m_j microphones and whose slope is $\tan(\alpha)$. The reason why it is a line is the following: let's take a closer look at equation 4.4. The only parameter that changes according to the source position is the delay (ΔT_{ij}), since d (distance between the microphones) and c (speed of sound) are constant. Hence, if there are different positions where the sound source, depicted in the figure 6, is seen with the same delay, there will be ambiguities from where the source actually is. Such a scenario is realistic: since $d_s \gg d$, for a given α , the distance between the source and each of the

microphones is roughly the same. In other words, it defines a line, with an α slope, where the source could be.

There are two factors that come into play in this ambiguity issue. One of them is the distance between the microphones (d). Among the methods to estimate the delay, there is the phase shift computation, which was the one chosen in this project. In order to not introduce additional ambiguities, it's necessary that $d = \lambda/2$, where λ is the wavelength. If $d > \lambda/2$, the phase will wrap around the interval $[-180^\circ, 180^\circ]$, causing a $2k\pi$ ambiguity. Whereas if $d < \lambda/2$, the phase shift range would be shortened. Also, as c is constant, this relation can be expressed as a function of f because $c = \lambda f$. The equation 4.5 shows that, in practice, the choice of emitting frequency determines the distance that the microphones should have to avoid this ambiguity.

$$d = \frac{c}{2f} \quad (4.5)$$

Secondly, as it was discussed previously, just one pair of microphones is not enough to determine the source position. Actually, in the 2D scenario, each α defines 2 lines where the source lies on. The figure 7 illustrates this idea. With the source being in a mirrored position with respect to m_{ij} axis, the delay seen by the microphones would be exactly the same, so, it is not possible to know if α is positive (anti clockwise rotation from m_{ij}) or negative (clockwise rotation from m_{ij}).

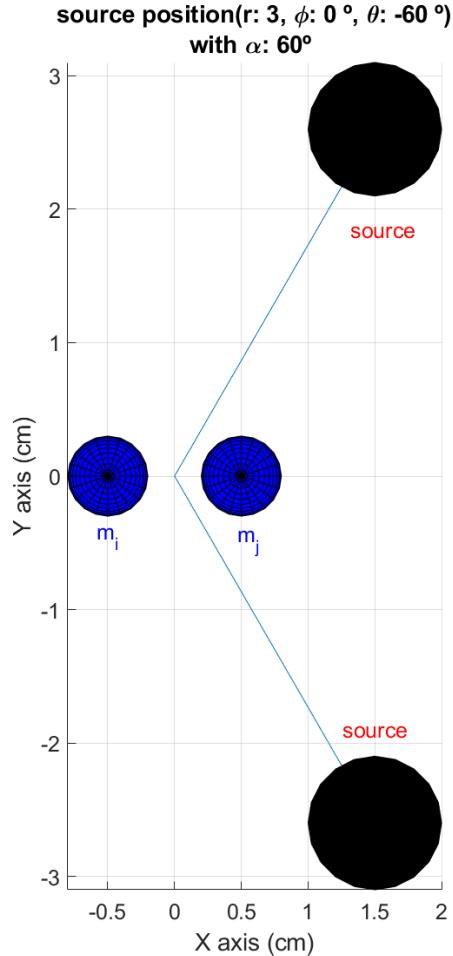


Figure 7: Example, in two dimensions, highlighting that, with just a pair of microphones, there will always be two lines where the sound source could lie. In this case, $\alpha = 60^\circ$, the source could be anywhere along the two blue lines since the delay seen by the microphones will be the same. Distance d is 1 cm. The source position is depicted in spherical coordinates (r, ϕ, θ) , where r is the radius in cm, ϕ is the elevation angle and θ the azimuth angle, both in degrees.

Before introducing a solution to the ambiguity issue, it is interesting to visualize how this approach extends to the 3D case scenario. The angle α will continue to be calculated in the same fashion. However, the two lines in 2 dimensions are part of a cone generator, formed by their revolution around the $\underline{m_{ij}}$ axis, with an α angle of aperture. Thus, generalizing for the 3D scenario, the source could be anywhere along the surface of a cone that the microphones would see the same delay. The figure 8 illustrates this concept.

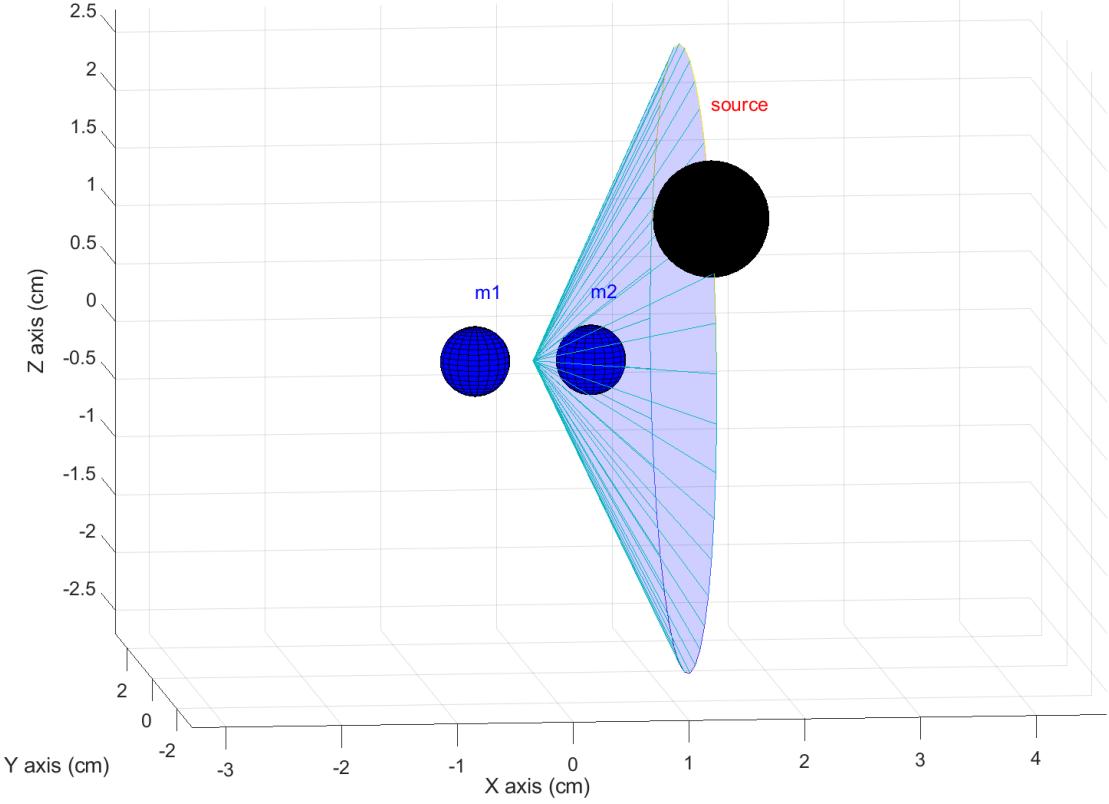


Figure 8: Example, in three dimensions, showing the valid position of a sound source for a given delay. It is important to note that this cone is infinite and it is being represented only up to the point it touches the source for visualization purposes. In this case, α is still 60° , $d = 1$ cm and the source position is $(r : 3\text{cm}, \phi : 30^\circ, \theta : -60^\circ)$

To explain the solution adopted to solve these ambiguities, it's necessary, first, to take up some initial concepts of the project. In the real scenario, the system will use ultrasound frequencies ($f \geq 40$ kHz) and, as expressed in the equation 4.5, the distance between the microphones will be in the millimeters order ($d \leq 4, 2$ mm). Besides, the microphones will be placed into the headset, thus, the available area for it is mostly flat. In this regard, it was opted to the adoption of a method in which a second pair of microphones will be introduced, distant d from each other, but, with the line connecting the second pair being perpendicular to the line connecting the first one [9]. The figure 9 clarifies how these pairs are distributed.

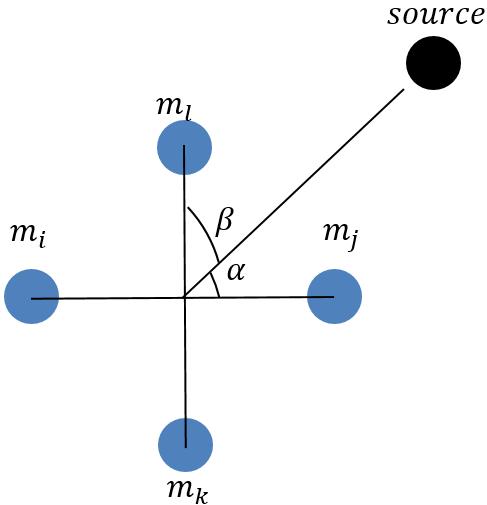
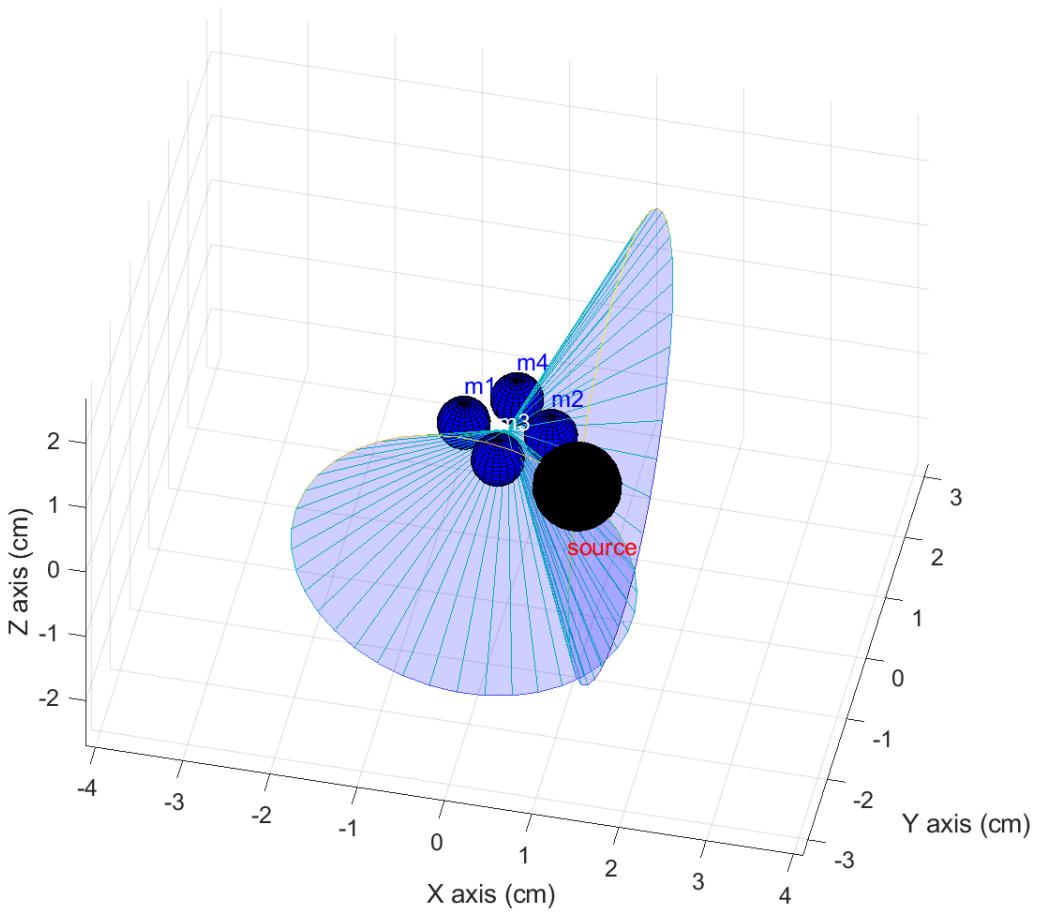


Figure 9: Representation, in two dimensions, of the solution adopted to solve the cone ambiguity. α is the AOA for the first pair $(m_i - m_j)$ and β for the second one $(m_k - m_l)$. This system is called a quadruple.

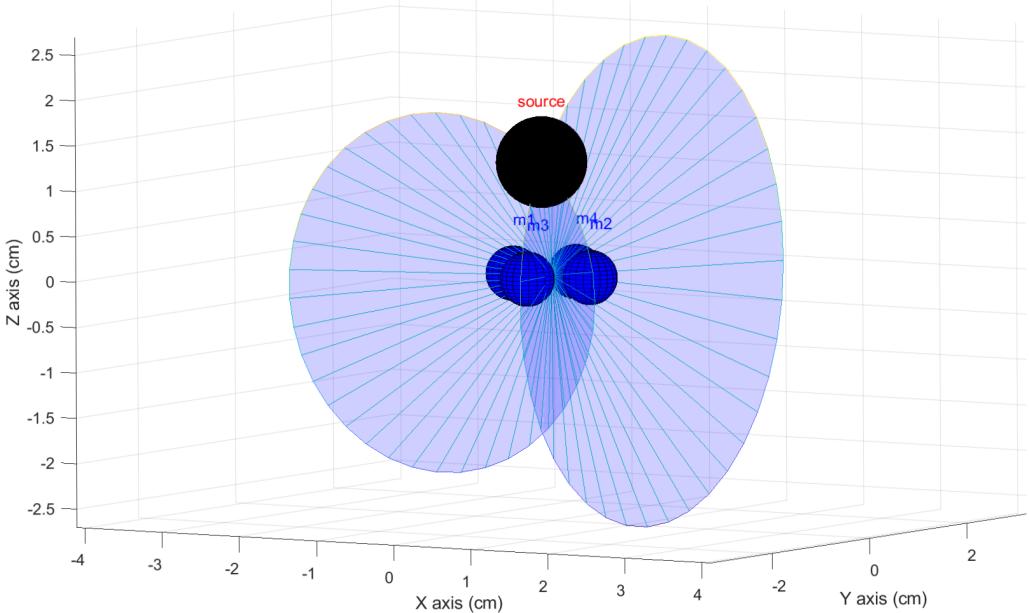
The interest in the use of quadruples is due to the ambiguity in the source position be reduced to a 3D line. The figure 10 shows two points of view from the scenario that has been discussed throughout the report. As there are two cones, the source must lie in the intersection of both and, consequently, the positions in the 3D space where it can be are limited by the intersection of cones, which, in practice, is a line in 3D space.

It is worth noting that, in fact, there are two regions where the cones intersect each other, as shown in the figure 10b. One of the lines is generated in the $z > 0$ plane and the other in the $z < 0$. In practice, going back to the application, as the microphones will be placed into the headset, the positive z plane would depict the region in space where the user is looking at, whereas, the negative z plane would depict the region behind the user's back, where it is not necessary to track the source position. For this reason, only the line in the positive z plane will be considered.

Hence, each quadruple is capable of generating one and only one 3D line where the source could lie on. In order to find out the exact range the source is along this 3D line, it will be used a second quadruple. This way, the intersection between two 3D lines would result in a single point in space.



(a) Quadrupole's view from above.



(b) Quadrupole's view from the microphones' plane.

Figure 10: 3D representation of the cones generated by the quadrupole $m_1 - m_2, m_3 - m_4$. The source position is ($r : 3$ cm, $\phi : 30^\circ$, $\theta : -60^\circ$).

Taking a look a bit more carefully on how the quadruple was composed, the 3D line from the intersection of the two cones can be calculated as follows: α and β angles, from figure 9, represent the direction cosines in a local coordinate system, whose origin is the center of the quadruple and whose axes are composed by m_{ij} and m_{kl} . With this mind, the third direction cosine (γ) can be deduced according to the equation 4.6. The figure 11 gives an idea of how this line looks like in three dimensions.

This approach, therefore, is suited to the scenario of the internship itself. First, due to the requirement of the microphones' proximity in a quadruple, they will be in the same plane (condition that would make the TOF approach unfeasible). Second, even having microphones constrained in a plane enable this solution to find a 3D point in space. Ultimately, the use of the positive square root in the equation 4.6 is explained by the fact that only positive γ values are valid, since $\gamma < 0$ would mean source behind the user ($z < 0$ plane).

$$\gamma = \sqrt{1 - \cos^2(\alpha) - \cos^2(\beta)} \quad (4.6)$$

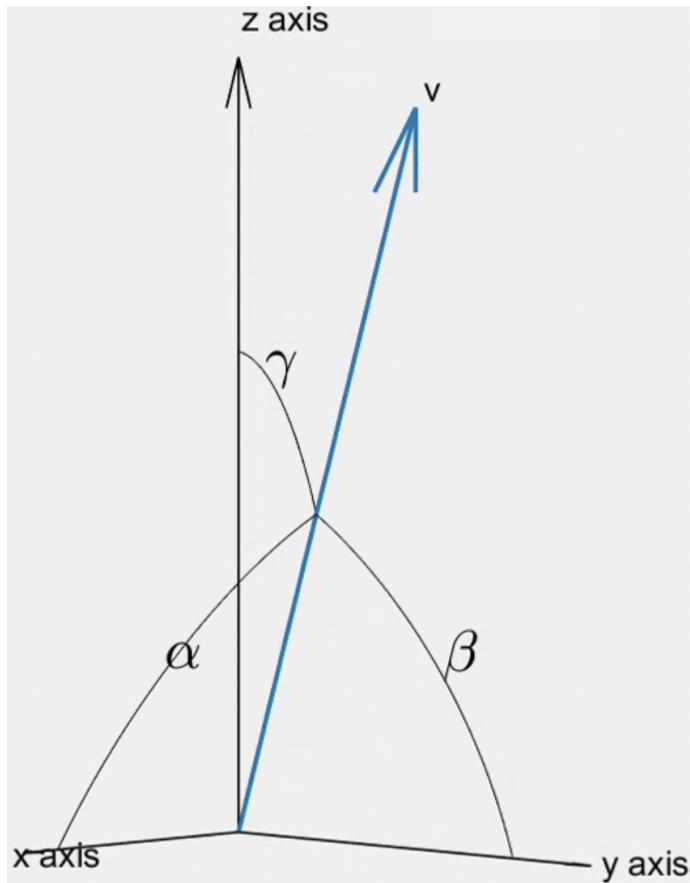


Figure 11: Representation of the 3D line generated by a quadruple with (α, β, γ) as direction cosines.

4.3.1 Evaluation

Following the same procedure made for the TOF case, it was implemented a system, in Matlab, to validate this approach. Before introducing the system itself, it's worth talking about some questions that were used as a guideline throughout the internship: what is the number of microphones needed so that the estimation error be ≤ 1 cm? How the microphones should be spatially distributed? What is the waveform to be emitted?

Treating the first two questions at once, one way to minimize the number of microphones and maximizing the number of quadruples is to distribute them in a matrix arrangement. If, for example, 16 microphones are distributed in a 4×4 matrix, 9 quadruples will be composed and, therefore, 9 lines in three dimensions would point to the source and, in theory, they would intersect in just one point. This 9 quadruples comes from the fact that one microphone can be used to compose more than a quadruple. In the figure 12, the m_3 microphone, for example, is used in the quadruples 1, 2, 4 and 5. To obtain the same number of quadruples without using this matrix arrangement, it would be necessary 36 (4×9) microphones.

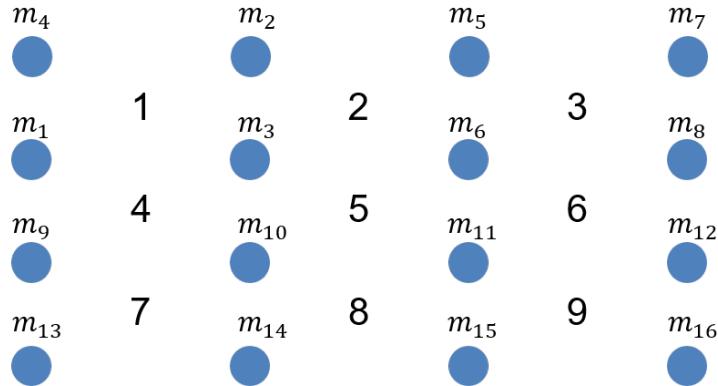


Figure 12: 4×4 microphone array. Numbers from 1 to 9 represent each one of the quadruples

The utilization of this matrix distribution raises another question: how to take advantage of the number of quadruples, in other words, how to increase the accuracy of the estimated source position based on the redundancy brought by having more than 2 lines in 3D? In practice, it is known that they won't intersect in a single point due to errors in the estimations of α and β for the 3D lines. Hence, one criteria that gave the best results in our theoretical system was the Angle of Intersection (AOI).

This criteria is related to the 3D lines generated by each of the quadruples. In a 4×4 microphone array scenario (figure 12), the number of intersections combining 9 lines, 2 by 2, there are: $9!/(9-2)!2! = 36$ points that should result in the source position. With the AOI criteria, it is chosen only the pair of lines whose intersection is the closest to 90° . The idea is: the more perpendicular two lines are, the lesser will be

the impact of the estimated error position. On the other hand, the more parallel two lines are, the more impact a small error in α or β will cause. The figure 14 illustrates this concept. Also, the figure 15 shows an example of the system implemented in Matlab.

A third question raised in the beginning of the section was about the waveform to be emitted. Lacking a convincing reason to use a complex signal, it was opted to transmit a sinusoidal wave. An important factor is the number of periods to emit in order that the phase shift computations of α, β, γ be precise enough so that the estimation error be less than 1 cm. Using the theoretical system, made in Matlab, it was possible to conclude that, with $f = 2.88$ kHz (or $d \approx 5.95$ cm), $f_s = 48$ kHz, 6 periods are enough to reach the desired precision. These values for f and f_s are due to the hardware acquired to validate the system: the UMA16¹ is 4 x 4 microphone array made by miniDSP. The figure 13 shows how it looks like.

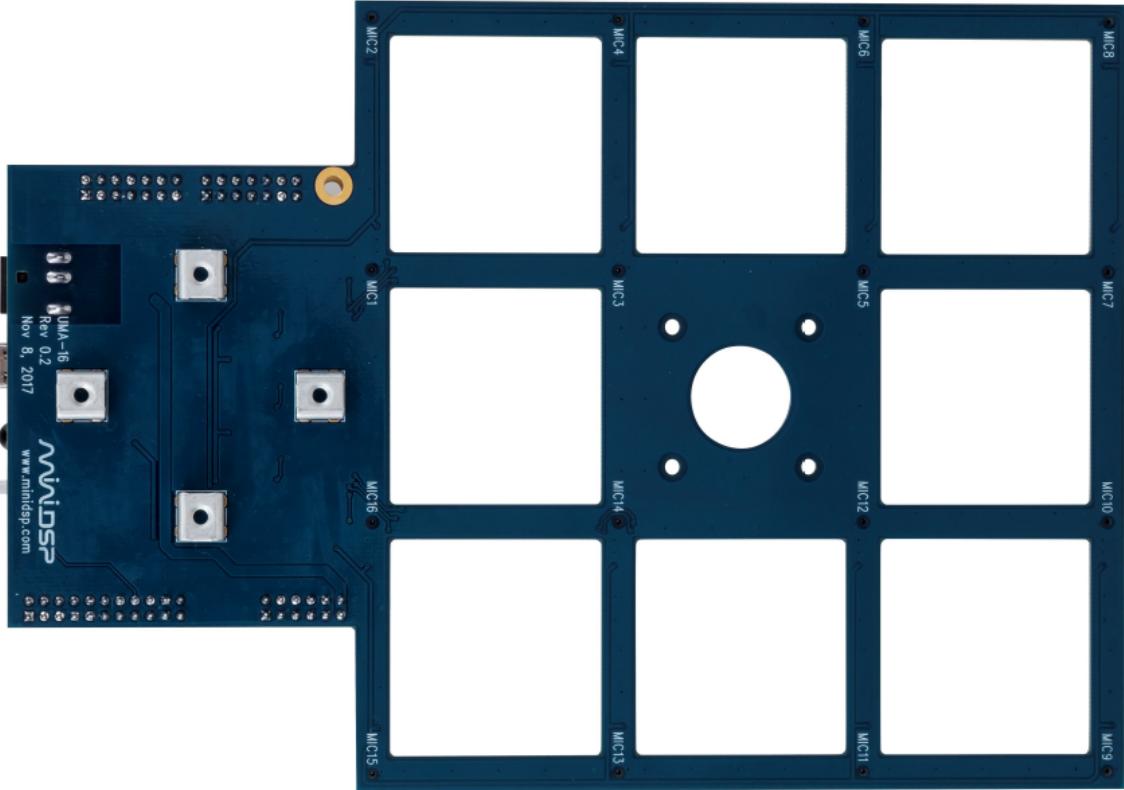


Figure 13: Microphone array acquired to validate the Matlab system. It has 16 microphones and 4 DACs capable of sampling at 48 kHz. Signals are taken via USB connection and it is possible to make live applications with Matlab.

¹<https://www.minidsp.com/products/usb-audio-interface/uma-16-microphone-array>



(a) Intersection of quasi-parallel lines
(AOI $\approx 180^\circ$).

(b) Intersection of perpendicular lines
(AOI $\approx 90^\circ$).

Figure 14: Representation, in 2D, of intersections that are close to being parallel (figure 14a) and close to being perpendicular (figure 14b).

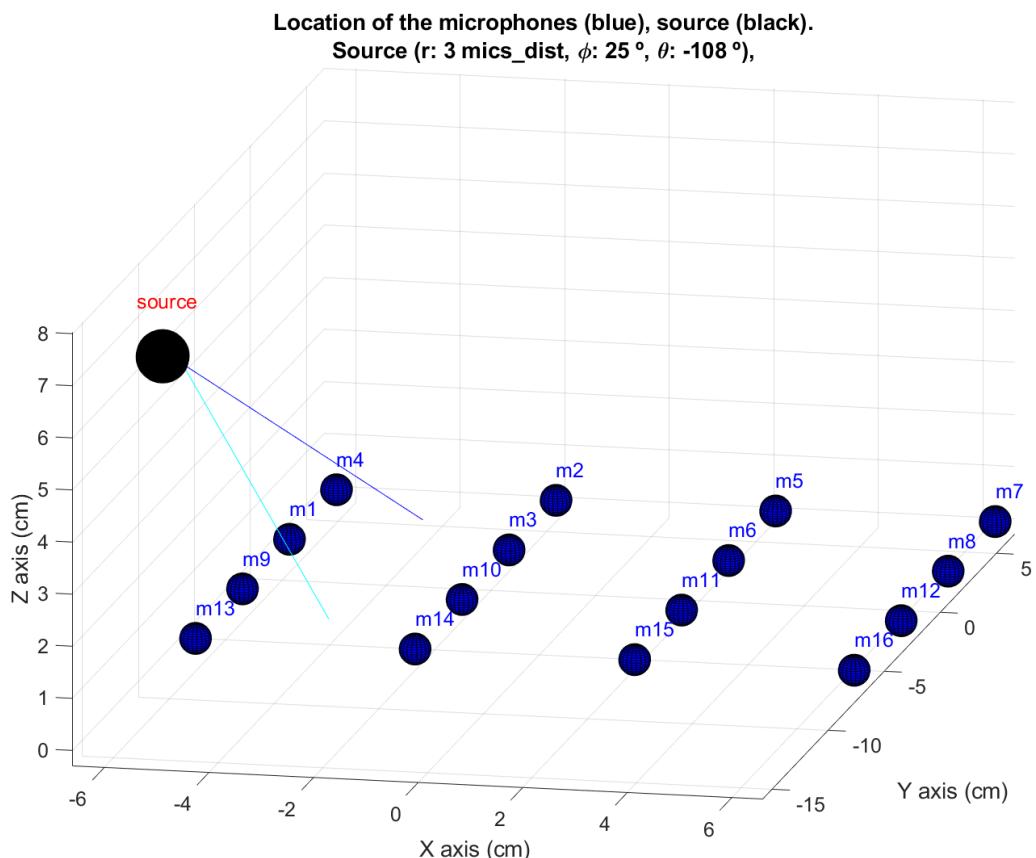


Figure 15: Example of the intersection between two 3D lines, generated by the quadruples $\{(m_1, m_2) - (m_3, m_4)\}$ e $\{(m_{13}, m_{10}) - (m_{14}, m_9)\}$. In this scenario, there are 9 quadruples available due to the 4×4 matrix array. Also, the following parameters were used: $f = 2.88 \text{ kHz}$, $f_s = 48 \text{ kHz}$, 12 periods of a sinusoidal signal were emitted. For sampling, windows of 6 periods wide were taken, with 80 % overlap.

4.4 VICON system

A second factor to enable the system validation is the ground truth provider. The company had a room equipped with 6 infrared cameras made by VICON². These cameras are capable of tracking an object's position and rotation provided that it's equipped with special markers that reflect the infrared waves. It guarantees an error less than 1 mm in the position estimation. Since it is desired to have less than 1 cm of error to the system conceived during the internship, VICON can be used as a ground truth provider.

Hence, UMA16 matrix microphone was mounted in a gyratory axis, where it can be rotated with a 2 ° angular precision. The figure 16 shows the setup made.

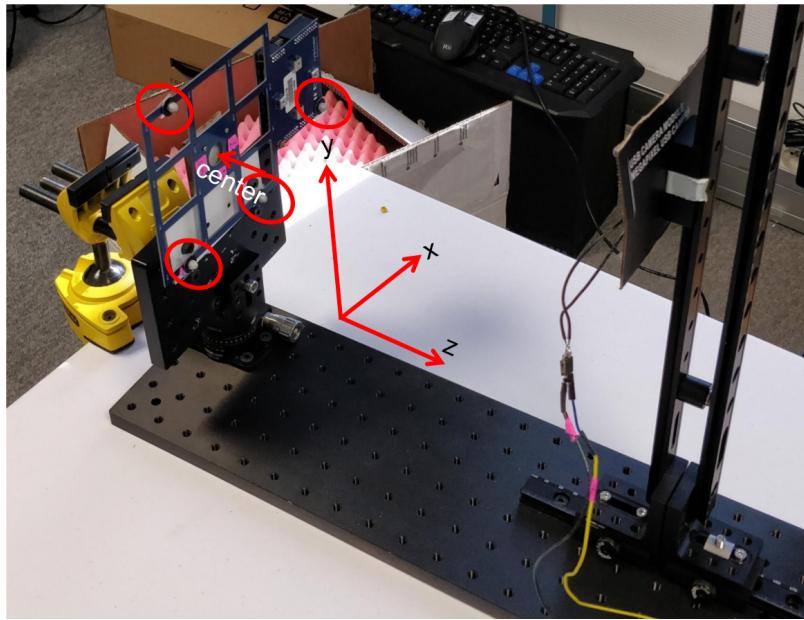


Figure 16: Experimental environment mounted in the VICON room. It can be seen the x , y and z axes, the origin of the coordinate system and the 4 markers used to track the UMA16 position.

4.5 User interface

One last thing implemented to facilitate the utilization of the software to users was the design of an User interface (UI). Actually, four UIs were made to give the user the possibility to tweak several parameters. The figure 17 shows the main UI, where several controls can be adjusted. Some of the most important are:

- upscale factor: number of new samples that will be added between two samples in order to interpolate the input signal.

²<https://www.vicon.com/>

- Intersection method: options to compute the source position. By default, the closest to 90° presented in this documentation.
- Power Thresh Phase Precision: number between 0 and 1 that represents the acceptance threshold. Zero means that all windowed signals will be accepted and one means only a pure sine wave will be accepted

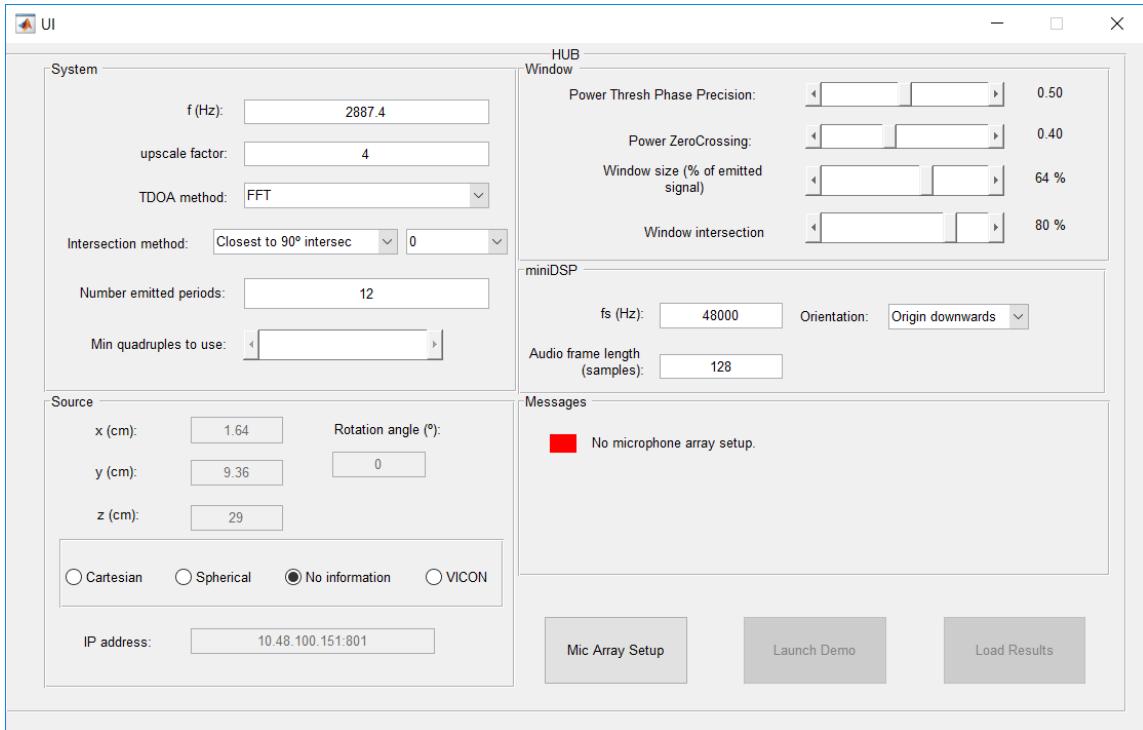


Figure 17: Main user interfaced designed to allow a better interaction with the system.

Second, a dedicated UI for the microphone array setup was built. On it, a microphone array is created with a certain number of rows and columns, an origin and a rotation angle. It allows the user to visualize the matrix disposition before moving forward. The figure 18 shows how it looks like.

Third, it was conceived a UI for live acquisitions. Since the UMA16 is easily interfaced with Matlab, it is useful to have a real time feedback of how the system works. Before performing the computations to estimate the source position, it is necessary to have a calibration phase first, because the phase shift between a given pair of microphones is not exactly what it was expected. MiniDSP was contacted to clarify the subject, but there was no feedback for this issue. Therefore, the software compares the calculated phase shift versus the expected one, given a known source position. This way, the offsets can be compensated. The figure 19 illustrates the calibration process.

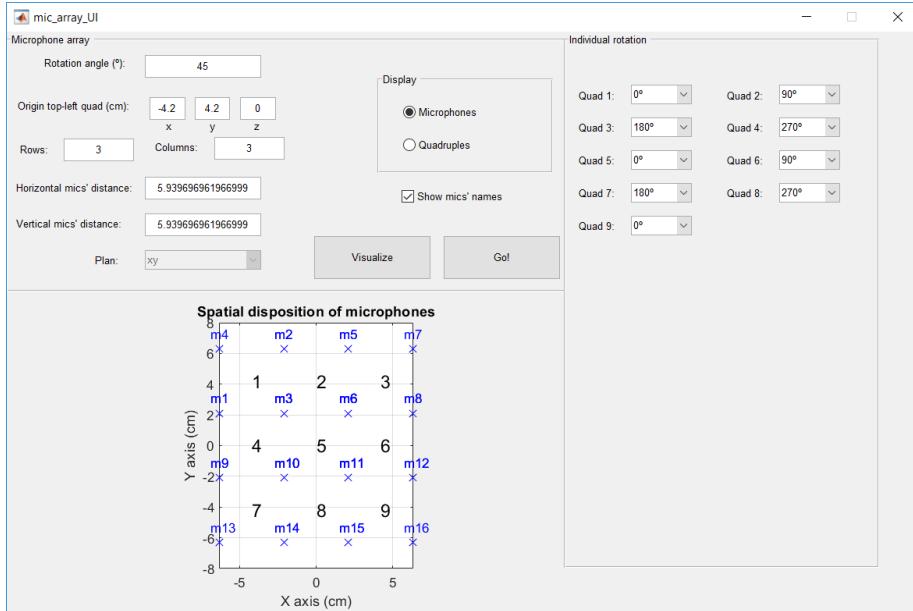


Figure 18: Microphone array UI.

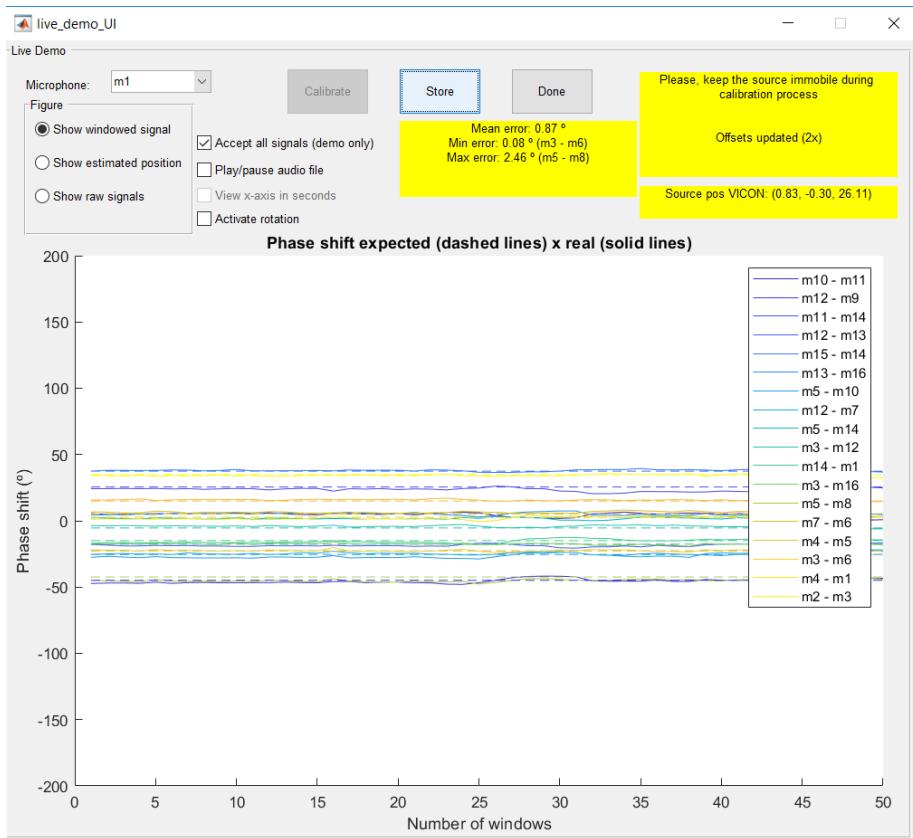


Figure 19: UI dedicated to live acquisitions where the calibration part is being performed

After the calibration process is done, it is possible to choose between 3 modes to analyze the system. The figure 20 shows one of them where the estimated position provided by our system and by VICON along with the microphone array are shown. It can be seen, also, the error between the ground truth provider and the estimated source. It's a very useful feature to have since it allows to observe in real time some factors that are not easy to handle with, such as acoustics reflections.

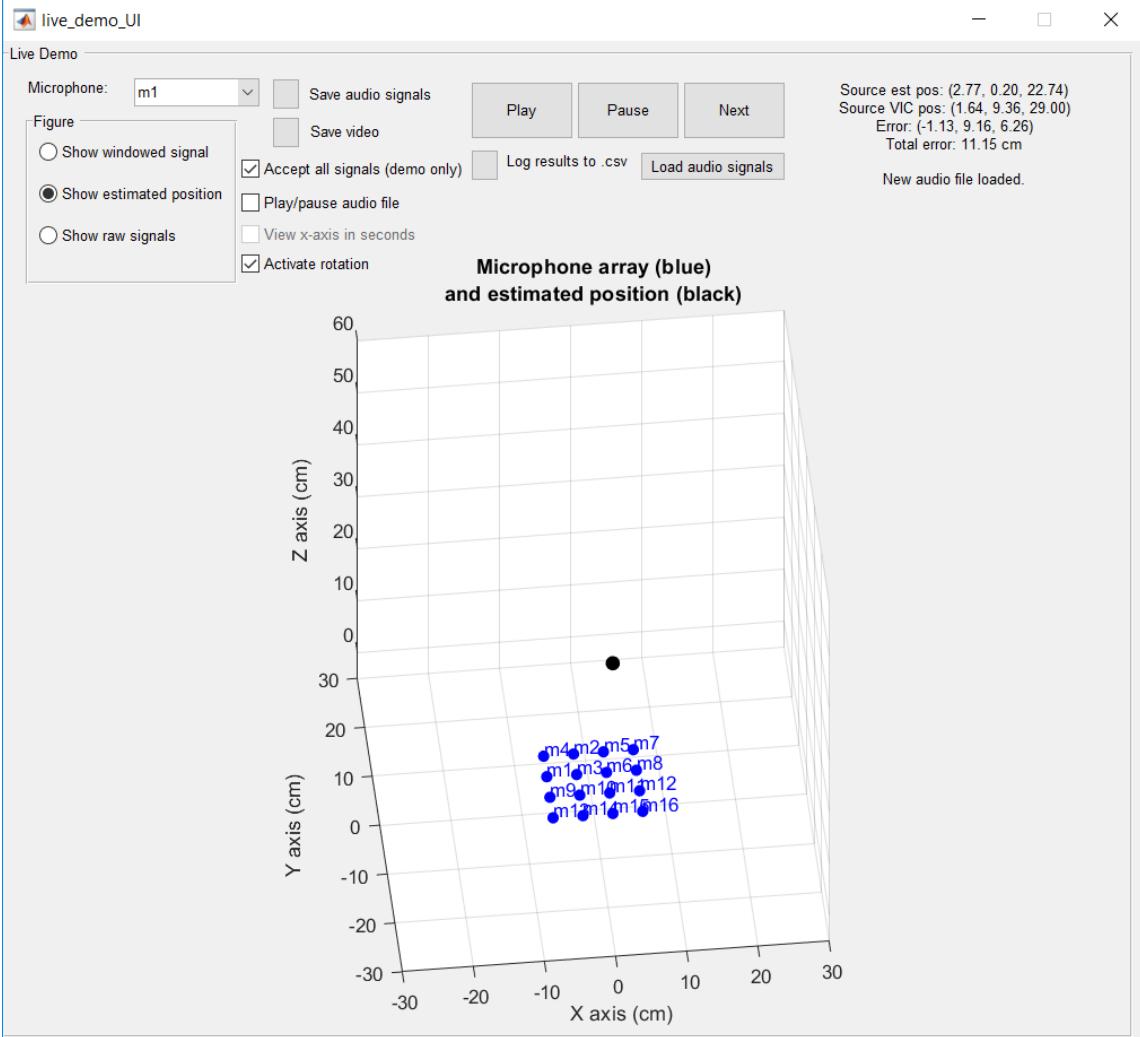


Figure 20: UI dedicated to live acquisitions when the source position is estimated in every new iteration.

The last one is the result analysis UI that loads a .csv file generated in live acquisitions. It's useful to analyze experiments offline, after having performed some acquisitions in different environment configurations. Also, it shows all the intersections with the current microphone array and, when having a ground truth provider, it searches which intersection is the best and highlights it in pink and, in red, it shows the intersection that is the closest to 90 °. The slider allows to move forward and

backward along the acquisition. The selected estimations (pink and red ones) have details on their AOI and the quadruples that generated them. The figure 21 shows the interface.

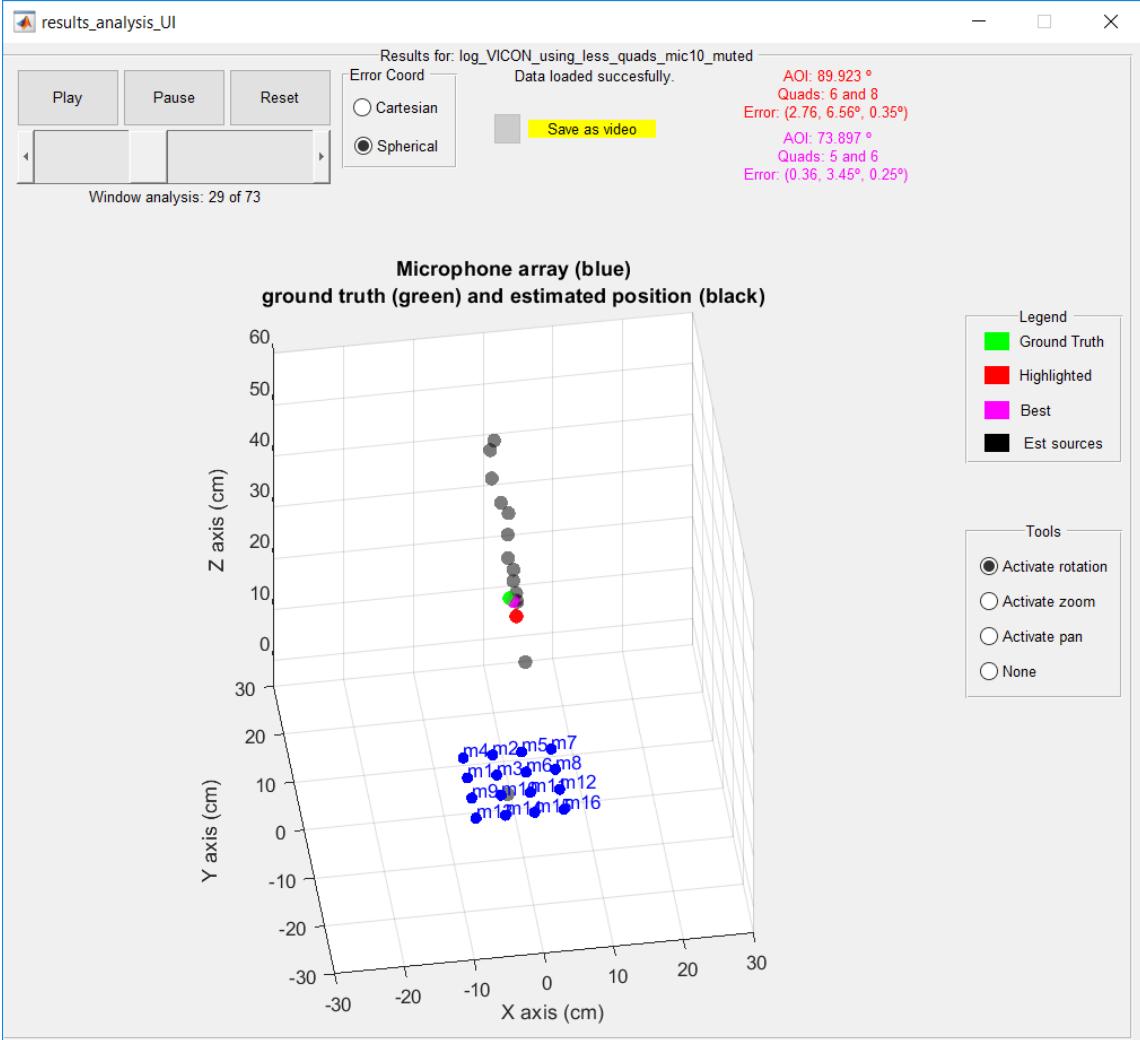


Figure 21: Results analysis UI.

4.6 Results

Evaluating the theoretical system in practice was a key point of the project. The experiments were made in two different scenarios:

1. Static and dynamic analysis in the VICON room, where the experiment was performed locally, inside the company. It was paramount to use this environment to validate the algorithms and to have a sense of how it was behaving.
2. Static and dynamic analysis in an anechoic chamber

4.6.1 Local experiments

The first series of experiments were performed in the VICON room. The sound source used can be seen in the figure 22. The static scenarios tested, based on figure 16, were the following: source in two different positions at (0, 0, 29 cm) and (0, 0, 48 cm) and UMA16 rotated with 0°, 15° and 30°. Also, to attenuate the acoustics effects in the room, an anechoic-like scenario was also tested. The figure 23 shows how it looked like.

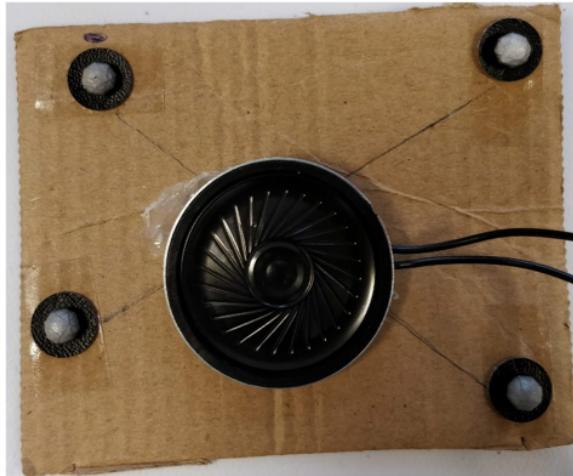


Figure 22: Sound source used to perform all the experiments.

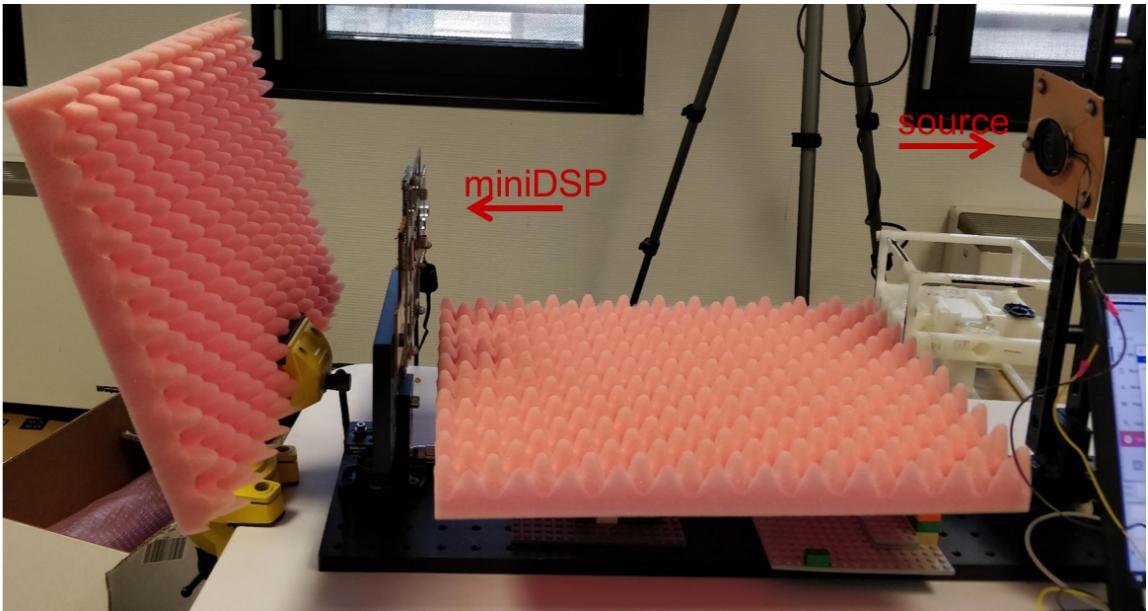


Figure 23: Implementation of the anechoic-like scenario.

The table 2 shows the results obtained. For each of the static scenarios, 360 measures were taken in order to evaluate the stability of the system. Therefore, the

average and the standard deviation are represented in each cell. This metrics show the error in the estimated position, comparing to the VICON system.

First of all, for almost all cases, the standard deviation is very low, which indicates that in these scenarios it is stable. Secondly, a general pattern that could be observed is that as UMA16 is turned around (going from 0° to 30°), the estimation error usually explodes. It could be due to acoustics in the room, since when it's pointing directly to the microphones (0° case in all scenarios), the error was considerably low. By turning UMA16, the indirect path of the sound could play a more important role and disturb the system.

Therefore, it was necessary to test in other environments, but it is also important to have a ground truth provider to be able to analyze what could potentially change.

Table 2: Results for static scenario in VICON room.

Estimated Position Error (cm)	z = 29 cm			z = 48 cm		
	0°	15°	30°	0°	15°	30°
X axis	2.09 0.06	3.30 0.15	89.56 9.61	1.19 0.24	10.38 0.34	26.36 0.09
Y axis	0.11 0.04	0.48 0.08	24.12 2.72	0.19 0.14	2.61 0.06	5.53 0.01
Z axis	3.79 0.24	6.76 0.41	135.66 14.65	2.06 1.57	25.71 0.82	39.35 0.14

(a) Average and standard deviation for non-anechoic foam scenario.

Estimated Position Error (cm)	z = 29 cm			z = 48 cm		
	0°	15°	30°	0°	15°	30°
X axis	0.78 0.07	6.92 0.05	2.08 0.64	0.41 0.16	0.88 0.63	46.36 36.15
Y axis	0.79 0.03	1.29 0.02	1.70 0.11	1.19 0.21	1.78 0.11	14.95 12.90
Z axis	1.44 0.22	14.99 0.12	9.00 0.67	1.76 1.00	8.03 2.81	66.88 59.46

(b) Average and standard deviation for anechoic-like foam scenario.

The second scenario tested was the dynamic one. The big advantage of having a ground truth provider in real time is to explore this capability. Based on the conclusions taken in the static part, the dynamic one was done using UMA16 at 0° rotation. The sound source was moved in front of the microphones with distances similar to previous scenario ($z = 29$ cm and $z = 48$ cm), and for the $x - y$ axes, a pattern was drawn in a way that the source was always kept looking directly to the microphones and it also didn't leave from the microphones' boundaries ($-6.3 \leq x \leq 6.3$ and $-6.3 \leq y \leq 6.3$), in cm.

The table 3 shows the results in the same manner as previously. An average and standard deviation were calculated after performing a complete path. In general, it can be seen that the estimation follows the ground truth, with acceptable errors, and also that the main source of errors was the Z-coordinate. As the matrix is placed in $x - y$ plane, it makes sense that it be the main difficulty.

Table 3: Results for dynamic scenario in VICON room.

Estimated Position Error (cm)	$z = 29$ cm	$z = 48$ cm
X axis	2.05 1.21	3.41 2.37
Y axis	2.48 1.72	2.88 2.05
Z axis	7.04 4.77	14.66 7.29

(a) Average and standard deviation for non-anechoic foam scenario.

Estimated Position Error (cm)	$z = 29$ cm	$z = 48$ cm
X axis	2.15 0.99	3.43 2.38
Y axis	1.16 0.74	2.73 2.15
Z axis	4.50 2.42	19.98 6.61

(b) Average and standard deviation for anechoic-like foam scenario.

4.6.2 Anechoic chamber experiments

The last couple of weeks were spent preparing the system to be tested in an anechoic chamber. One of the important improvements that needed to be done was the addition of a new ground truth provider, because it would be very useful to perform the two scenarios we did in the VICON room.

To accomplish this requirement, one pair of stereo cameras was mounted in behind the UMA16. With it and some image processing, it is possible to associate a pixel in both images with a point in 3D space. Therefore, the sound source was mounted on an acrylic surface, along with some visual markers that can be used by the cameras to retrieve, automatically, the source position.

With this setup, it was no longer possible to have a ground truth in real time, since it was too expensive for the hardware I was working on. So, the results were logged and the analysis had to be done afterwards. Also, due to bandwidth limitation in cables and other problems with the connections available in the room, the experiment was performed with 2 people inside. The figure 24 shows UMA16 and the sound source, along with the acrylic plate inside the anechoic chamber.



Figure 24: Anechoic chamber in GIPSA-lab where the system was tested in.

For the dynamic analysis, the ground truth algorithm using image processing was not finished in time so that the results could be analyzed during the internship. Although, data was acquired and it should be possible to do so afterwards.

For the static part, the scenarios simulated, due mainly to the lack of time to set up, were not the same as the previous one. Also, since there was less than a week after the acquisitions, it was not possible to compile the results in the same manner as previously. It's worth pointing out that it's also interesting to take a look from the results analysis UI. As it shows all quadruples and highlights the closest to 90° and the one with the smallest error, it can be noticed that there are scenarios where these two highlights do not coincide, meaning that the closest to 90° is not always the best estimation. Nevertheless, the figure 25 illustrates a situation, in the anechoic chamber scenario, where AOI close to 90° means the best estimation.

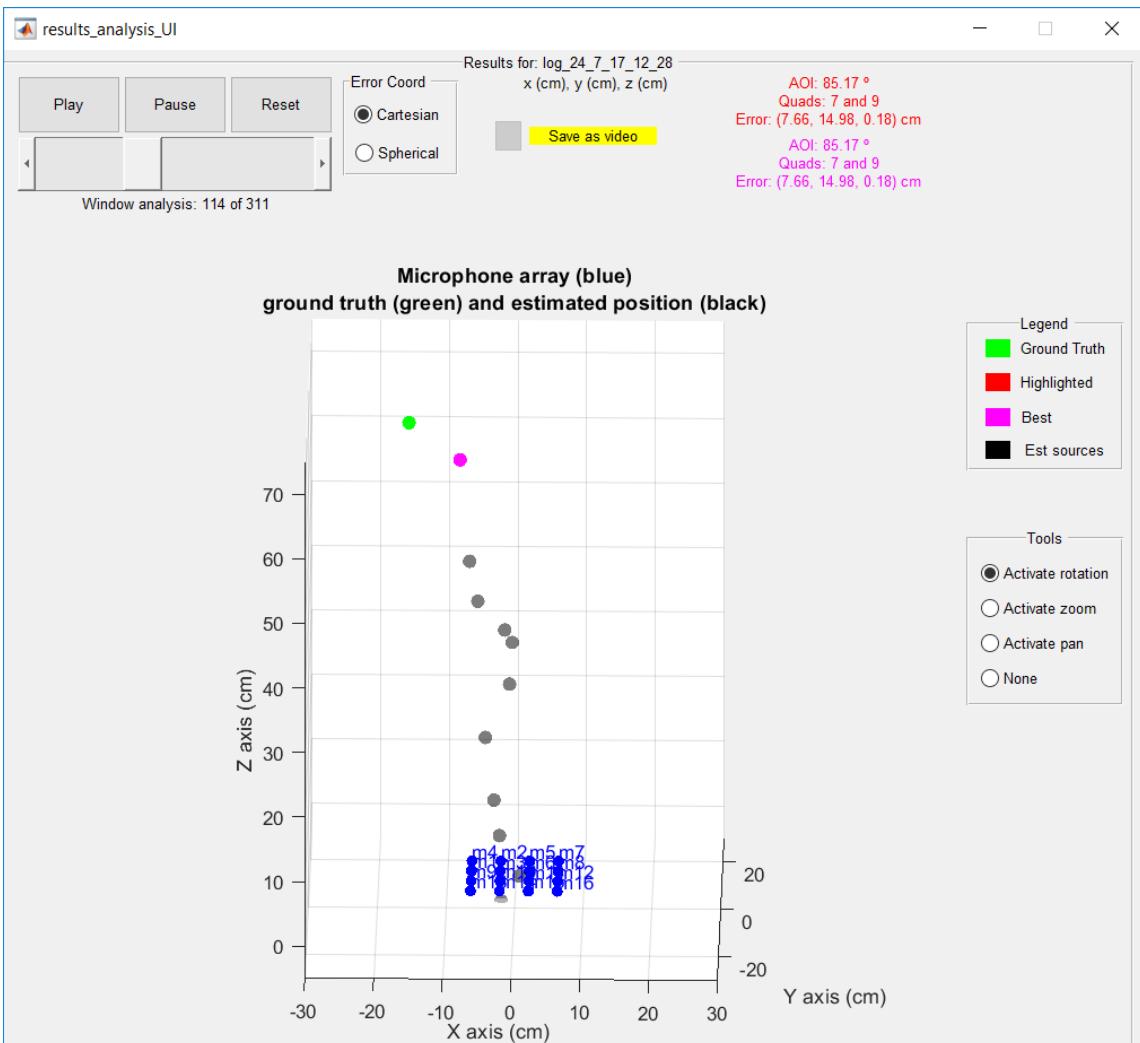


Figure 25: One frame of iteration in one of the acquisition made in the anechoic chamber to exemplify that the best and closest to 90° are the same intersection.

5 Conclusion

The internship was a big challenge in two areas. First, I was immersed in a team whose expertise is software and I had to prototype a solution to validate the research part. Second, the project has been discussed internally for a while, but up until now, there was nobody exclusively dedicated on. In this sense, the freedom to have choice about what paths to explore comes in conjunction with the responsibility to bring relevant results to evolve the project.

The acoustics effects played a huge role in the prototype part, since it's problem difficult to model and also to overcome. There's certainly big rooms for improvement yet: TOF and TDOA could be used together to make a hybrid system, where one could help the other to take the best decision regarding the source estimation; the redundancy of information can be explored better, it's a very rich subject to dive into, trying to use the 36 intersections, in our scenario, to improve robustness; emission of different waveforms, with different kinds of modulations, could also improve the system's robustness to reverberating noises; explore more deeply the emission of burst signals, instead of continuous emission.

The experiments performed in the anechoic chamber have to be analyzed more carefully in order to determine what direction should the research keep going to. Due to lack of time, it was prioritized the system adaptation to be able to acquire data instead of developing other parts.

To top it off, the experience during the internship was fantastic, it was the first time I faced such a problem and looking back to what has been developed, and I don't think that 6 months ago I would thought of reaching this far. I could develop a lot my skills in Matlab, prototyping and mainly how to isolate a phenomenon to further analyze it to take conclusions and keep evolving the concepts.

References

- [1] 3D-Printed “Earable” Sensor Monitors Vital Signs.
<https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/3dprinted-earable-sensor-monitor-vital-signs>. Accessed in 06/18/18.
- [2] Bin Yang and Jan Scheuing. Cramer-rao bound and optimum sensor array for source localization from time differences of arrival. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP'05). IEEE International Conference on*, volume 4, pages iv–961. IEEE, 2005.
- [3] Bin Yang and Jan Scheuing. A theoretical analysis of 2D sensor arrays for tdoa based localization. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 4, pages IV–IV. IEEE, 2006.
- [4] Bin Yang. Different sensor placement strategies for tdoa based localization. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pages II–1093. IEEE, 2007.
- [5] Tarig Ballal and Chris J Bleakley. 3D location and orientation estimation using angle of arrival. In *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, pages 21–26. IEEE, 2009.
- [6] William J Bangs and Peter M Schultheiss. Space-time processing for optimal parameter estimation. *Signal processing*, pages 577–590, 1973.
- [7] G Clifford Carter. Variance bounds for passively locating an acoustic source with a symmetric line array. *The Journal of the Acoustical Society of America*, 62(4):922–926, 1977.
- [8] W Hahn and S Tretter. Optimum processing for delay-vector estimation in passive signal arrays. *IEEE Transactions on Information Theory*, 19(5):608–614, 1973.
- [9] Michael S Brandstein, John E Adcock, and Harvey F Silverman. A closed-form method for finding source locations from microphone-array time-decay estimates. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 5, pages 3019–3022. IEEE, 1995.

A Deduction of the 3D positioning equations

The deductions made below use the notation introduced by the figure 3. This way, (x, y) is the source position and (x_i, y_i) is the position of the i -th microphone. From it, the equations below can be deduced:

$$\begin{aligned} v(t_i - t) &= \sqrt{(x_i - x)^2 + (y_i - y)^2} \\ v(t_j - t) &= \sqrt{(x_j - x)^2 + (y_j - y)^2} \\ v(t_k - t) &= \sqrt{(x_k - x)^2 + (y_k - y)^2} \\ v(t_l - t) &= \sqrt{(x_l - x)^2 + (y_l - y)^2} \end{aligned} \quad (\text{A.1})$$

To take advantage of the fourth microphone, added to linearize the system, all the equations are squared:

$$\begin{aligned} v^2(t_i - t)^2 &= (x_i - x)^2 + (y_i - y)^2 \\ v^2(t_j - t)^2 &= (x_j - x)^2 + (y_j - y)^2 \\ v^2(t_k - t)^2 &= (x_k - x)^2 + (y_k - y)^2 \\ v^2(t_l - t)^2 &= (x_l - x)^2 + (y_l - y)^2 \end{aligned} \quad (\text{A.2})$$

Developing and organizing the terms, there are:

$$v^2(t_i^2 - 2tt_i + t^2) - (x_i^2 - 2xx_i + x^2) - (y_i^2 - 2yy_i + y^2) = 0 \quad (\text{A.3})$$

$$v^2(t_j^2 - 2tt_j + t^2) - (x_j^2 - 2xx_j + x^2) - (y_j^2 - 2yy_j + y^2) = 0 \quad (\text{A.4})$$

$$v^2(t_k^2 - 2tt_k + t^2) - (x_k^2 - 2xx_k + x^2) - (y_k^2 - 2yy_k + y^2) = 0 \quad (\text{A.5})$$

$$v^2(t_l^2 - 2tt_l + t^2) - (x_l^2 - 2xx_l + x^2) - (y_l^2 - 2yy_l + y^2) = 0 \quad (\text{A.6})$$

Next, keeping the same reference made in the section 4.2, the microphone i will be used as a reference and, subtracting the equation A.3 from A.4, it is obtained:

$$x(x_i - x_j) + y(y_i - y_j) - v^2t(t_i - t_j) = \frac{1}{2} [x_i^2 + y_i^2 - (x_j^2 + y_j^2) - v^2(t_i^2 - t_j^2)] \quad (\text{A.7})$$

Repeating the same process, making A.3 - A.5 and A.3 - A.6, a system with 3 equations and 3 unknowns will be obtained, with $r_i = x_i^2 + y_i^2$:

$$\begin{bmatrix} x_i - x_j & y_i - y_j & -v^2(t_i - t_j) \\ x_i - x_k & y_i - y_k & -v^2(t_i - t_k) \\ x_i - x_l & y_i - y_l & -v^2(t_i - t_l) \end{bmatrix} \begin{bmatrix} x \\ y \\ t \end{bmatrix} = \begin{bmatrix} 0.5 [r_i - r_j - v^2(t_i^2 - t_j^2)] \\ 0.5 [r_i - r_k - v^2(t_i^2 - t_k^2)] \\ 0.5 [r_i - r_l - v^2(t_i^2 - t_l^2)] \end{bmatrix} \quad (\text{A.8})$$

The system to be solved is $\underline{\underline{A}} \underline{p} = \underline{k}$, whose solution will be: $\underline{p} = \underline{\underline{A}}^{-1} \cdot \underline{k}$. To deduce the case in 3 dimensions, the exactly same process can be used, needing a fifth microphone. In this case, $r_i = x_i^2 + y_i^2 + z_i^2$, and the system will be:

$$\begin{bmatrix} x_i - x_j & y_i - y_j & z_i - z_j & -v^2(t_i - t_j) \\ x_i - x_k & y_i - y_k & z_i - z_k & -v^2(t_i - t_k) \\ x_i - x_l & y_i - y_l & z_i - z_l & -v^2(t_i - t_l) \\ x_i - x_m & y_i - y_m & z_i - z_m & -v^2(t_i - t_m) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0.5 [r_i - r_j - v^2(t_i^2 - t_j^2)] \\ 0.5 [r_i - r_k - v^2(t_i^2 - t_k^2)] \\ 0.5 [r_i - r_l - v^2(t_i^2 - t_l^2)] \\ 0.5 [r_i - r_m - v^2(t_i^2 - t_m^2)] \end{bmatrix} \quad (\text{A.9})$$

Étudiant: LEMOS Lucas

Année d'étude dans la spécialité:
5ème

Entreprise: TDK Invensense

Adresse complète: 22 avenue Doyen Weil

8ème étage

38000 - Grenoble

Téléphone: +33 04 38 21 19 31

Responsable administratif: BERGONZOLI Chrystel

Téléphone: +33 04 38 21 19 31

Courriel: cbergonzoli@invensense.com

Tuteur de stage: FAYOLLE Romain

Téléphone: +33 07 83 33 79 06

Courriel: rfayolle@invensense.com

Enseignant-référent: TORU Sylvain

Téléphone: +33 04 76 82 79 15

Courriel: sylvain.toru@univ-grenoble-alpes.fr

Titre: Étude d'une méthode et système pour le positionnement entre dispositifs

Résumé: Le marché de la réalité virtuelle se présente encore comme le futur. Même si les solutions ne sont pas encore devenu grand public, une partie important des ressources sont utilisé dans ce domaine. Dans ce genre de système, la précision entre la position relatif de la manette et le casque est fondamentale pour expérience utilisateur agréable. Les solutions d'aujourd'hui utilisent une caméra pour avoir cette positionnement de manière précise. Le stage, donc, se propose à changer le rôle de la caméra par une matrice de microphone dans le casque et des émetteurs ultrasons dans la manette. Deux méthodes de positionnement 3D seront présentés, le TOF et TDOA, bien comme les avantages et inconvénients de chacun. Après, une implémentation en *hardware* du TDOA a été faite et les résultats seront discutés.