IMDEA project
Vector Based Amplitude Panning

Author:
LABOURE Valentin

Supervisors:
GAZENGEL Bruno, MELON Manuel, NOVAK Antonin

03 June 2020
International Master's Degree in ElectroAcoustics
Year 2019 - 2020

Contents

List of Figures

List of Tables

Introduction

Spatial audio aims to recreate spatial attributes when reproducing audio over loudspeakers. Some spatial audio methods such as Wave Field Synthesis [1] or High Order Ambisonics [2] focus on synthesize the complete sound field around the listener. While other methods such as Vector Based Amplitude Panning [3] try to recreate the feeling of a spatialized sound by reproducing the human audio spatial cues at the listener ears. Each of this methods have pros and cons. Moreover, the accuracy of the spatialization is mainly dependant to the loudspeaker system.

1 Project presentation

The aim of this project is to propose a plugin able to integrate the Vector Based Amplitude Panning (VBAP) on various multi-loudspeaker systems. This plugin will be coded in Python and the Kivy framework is used to create the graphical user interface. The final form of this project should propose a software and a multi-loudspeakers system to test this software. First, the theoretical context of this project will be presented. Some basic knowledge about human spatial hearing and amplitude panning will be depicted. Then, the software part will be detailed by presenting the graphical user interface of the project and the main algorithm. Finally, the hardware part of the project will be presented. A simplified mindmap of the project is presented on Figure 1.

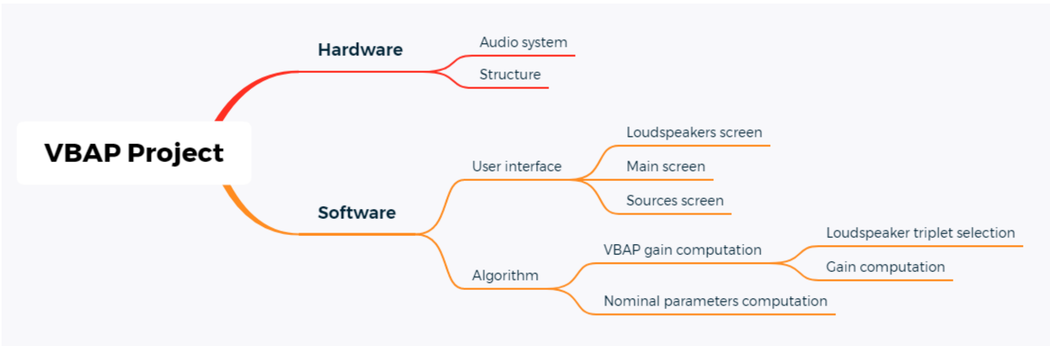

Figure 1: Mindmap

Part I
Theoretical background

2 Spatial hearing theory

The human spatial hearing is based on a set of different cues. Some of this cues are used for vertical, right and left or front and back direction. Other cues are used for motion or distance perception. But the most important are the cues used by the brain to localize a sound source in the horizontal plane.

Since Rayleigh's "duplex theory" [4], it is known that the interaural time difference (ITD) and interaural level ratio (ILR) (or interaural level difference), at the ears are the main cues to locate the horizontal direction of a sound source. This theory states that at high frequencies, the head is larger than the wavelength and casts an acoustic shadow. So the sound level at the ear nearer to the sound source would be greater than at the farthest ear. At lower frequencies this "intensity theory" is not correct because the sound wave "bends" around the head and the sound level difference between ears is negligible. After some experiments Rayleigh determined that "the only alternative to the intensity theory is to suppose that the judgement is founded upon the difference of phases at the 2 ears". Early free-field studies [5], [6] and [7] confirmed that sound location were accurate for frequencies below 1.5 kHz with use of ITD, and above 4 kHz with the use of ILR. However, a recent binaural model [8] suggests that ILR and ITD can be used at all frequency bands.

The last cue to determine the horizontal localization of a sound source is known as the precedence effect, or Haas effect [9]. When a sound is followed by a strongly correlated sound with a sufficiently short time delay, listeners perceive a single auditory event. The spatial location is mainly dominated by the location of the first-arriving sound. Therefore the lagging sound also affect the perceived location. The lagging sounds also allow our brains to understand the acoustic space (reverberation, etc). This effect can be strong in the case of a multi-loudspeaker system. In this project we have to be aware of this effect.

3 Amplitude panning theory

3.1 Derivation of ITD and ILR equations

In the case of a symmetric stereo loudspeaker system, where the listener is in the middle of two loudspeakers positionned with an angle ϕ (see Figure 2), Bernfeld [10] showed that interaural level ratio and interaural time difference between listener ears can be written as

ILR:
$$Q = \frac{L_o}{R_o} = \frac{km+1}{k+m} \tag{1}$$

ITD:
$$T = \frac{mT_{\phi_0}(k^2-1) + kT_s(m^2-1)}{m(k^2+1) + k(m^2+1)} \tag{2}$$

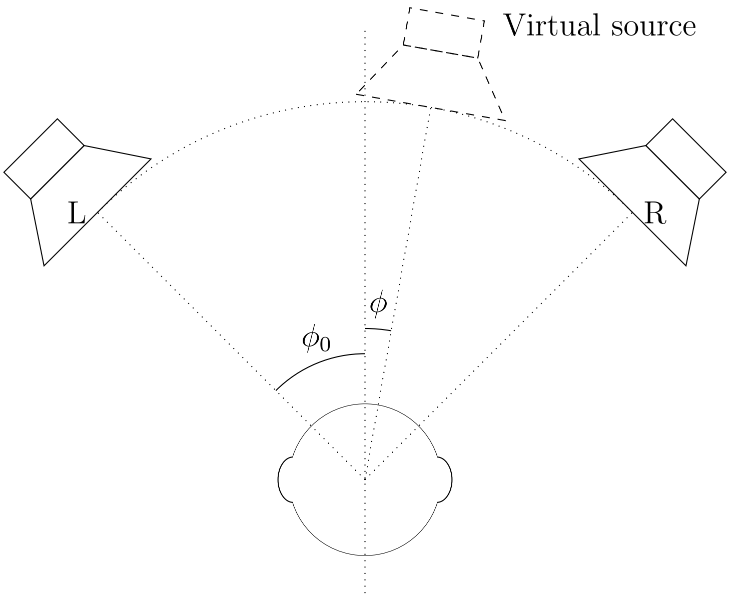

Figure 2: Stereo loudspeaker system

With L_o and R_o the amplitude of sound at left and right ear, $k = \frac{g_1}{g_2}$ the interchannel level ratio, m the head diffraction attenuation (or ILR in the case of a single loudspeaker), T_s the interchannel time difference and T_{ϕ_0} the interaural time difference created by a single loudspeaker. For low frequencies, the shadow effect of the head is neglectible ($m \approx 1$) and if there is no time difference between channel ($T_s = 0$), Equation 1 and Equation 2 become:

ILR:
$$Q = 1 \tag{3}$$

ITD:
$$T_\phi = \frac{T_{\phi_0}(k-1)}{k+1} = \frac{g_1 - g_2}{g_1 + g_2} T_{\phi_0} \tag{4}$$

where g_1 and g_2 are the gain of each loudspeaker.

Also, by using simple trigonometric relation, the ITD created by a single source at angle ϕ can be written as:

$$T_\phi = \frac{h}{c} \sin\phi \tag{5}$$

With h the distance between listener ears, c the speed of sound and ϕ the angle between the source and the median plane.

Equation 5 and Equation 4 can be combined to obtain the ratio of loudspeaker level depending of the virtual sound source angle and the loudspeaker angles. This equation is called the "sine law".

$$\frac{\sin\phi}{\sin\phi_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{6}$$

The sine law is valid only if the listener's head is pointing directly forward. Berfeld [10] showed that if the listener turns his or her head following the virtual source, the tangent law is more correct.

$$\frac{\tan\phi}{\tan\phi_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{7}$$

However, subjective listening tests [11] showed that this effect is negligible.

3.2 Vector based formulation

This section is describing the vector based amplitude panning proposed by Pulkki [3]. The configuration of Figure 2 is reformulated as a 2-dimensional vector base. The base is defined by unit-vectors $\overrightarrow{l_1} = \begin{bmatrix} l_{11} & l_{12} \end{bmatrix}^T$ and $\overrightarrow{l_2} = \begin{bmatrix} l_{21} & l_{22} \end{bmatrix}^T$ which are pointing toward loudspeakers 1 and 2. With l_{11}, l_{12}, l_{21} and l_{22} are the cartesian coordinates of the loudspeakers relative to the listener position. We can now describe as a linear combination of $\overrightarrow{l_1}$ and $\overrightarrow{l_2}$ a unit-vector $\overrightarrow{p} = \begin{bmatrix} p_1 & p_2 \end{bmatrix}^T$ pointing toward the virtual source. This configuration is shown on Figure 3.

$$\overrightarrow{p} = g_1 \overrightarrow{l_1} + g_2 \overrightarrow{l_2} \tag{8}$$

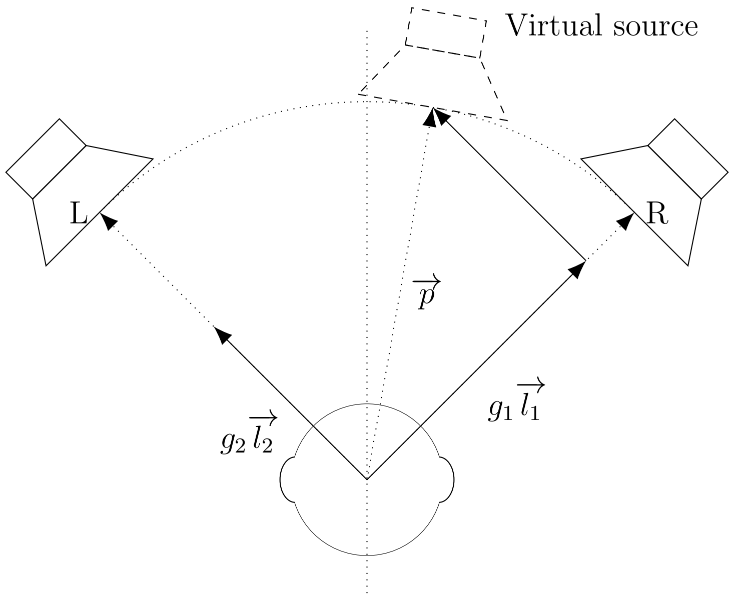

Figure 3: Stereo vector formulation

By considering the basis $L_{12} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}$ and the gain-matrix $\overrightarrow{g} = \begin{bmatrix} g_1 & g_2 \end{bmatrix}$, Equation 8 can be written in matrix form :

$$\overrightarrow{p} = \overrightarrow{g} L_{12} \tag{9}$$

And the gain factor of each loudspeaker can be solved if L_{12}^{-1} exists.

$$\overrightarrow{g} = \overrightarrow{p} L_{12}^{-1} \tag{10}$$

The last step of the vector based amplitude panning is to normalized the gain factors using Equation 11.

$$g^{scaled} = \frac{\sqrt{C}g}{\sqrt{g_1^2 + g_2^2}} \tag{11}$$

where C is the sound power of the source.

$$g_1^2 + g_2^2 = C \tag{12}$$

It can be easily checked that the gain factor computed with the vector formulation agreed the gain factor computed with the tangent law (see Equation 7).

This idea can easily be extended to system with more than 2 loudspeakers by selecting the closest pair of loudspeakers for each position.

And finally, this can also be easily extended to 3-dimensional system by creating a 3D vector basis using a loudspeaker triplet instead of a loudspeaker pair as shown in Figure 4.

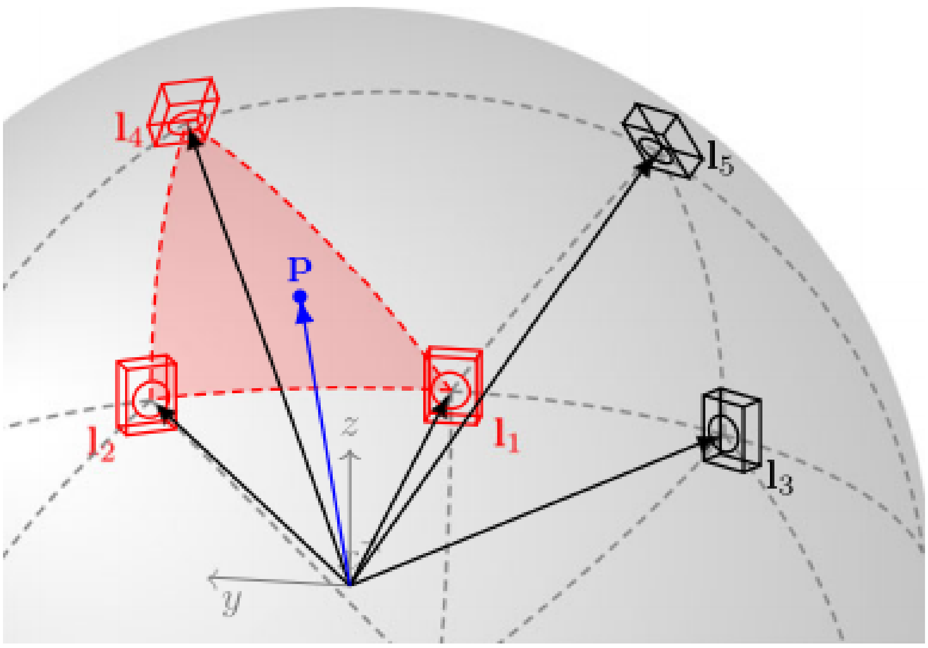

Figure 4: VBAP example in 3D, The active loudspeaker triplet is marked in red. (from [12])

Part II
Software

4 User interface

The user interface is divided in three different screens :

- The main screen, to show the configuration and play the VBAP.
- The loudspeakers screen to input the loudspeaker system configuration.
- The sources screen to input the sources configuration.

4.1 Main screen

The main screen is the first screen shown when the user open the software. This screen is presented on Figure 5. This screen is composed by :

- Two buttons to switch to loudspeakers screen or sources screen.
- A 3D view of the current configuration.
- A button to play the VBAP configuration.

The 3D view allows the user to rotate, pan and zoom on the loudspeaker and source configuration with left click, right click and mouse wheel. The green cube shows the listener position, the blue cubes show the loudspeakers position and the red cubes show the sources position.

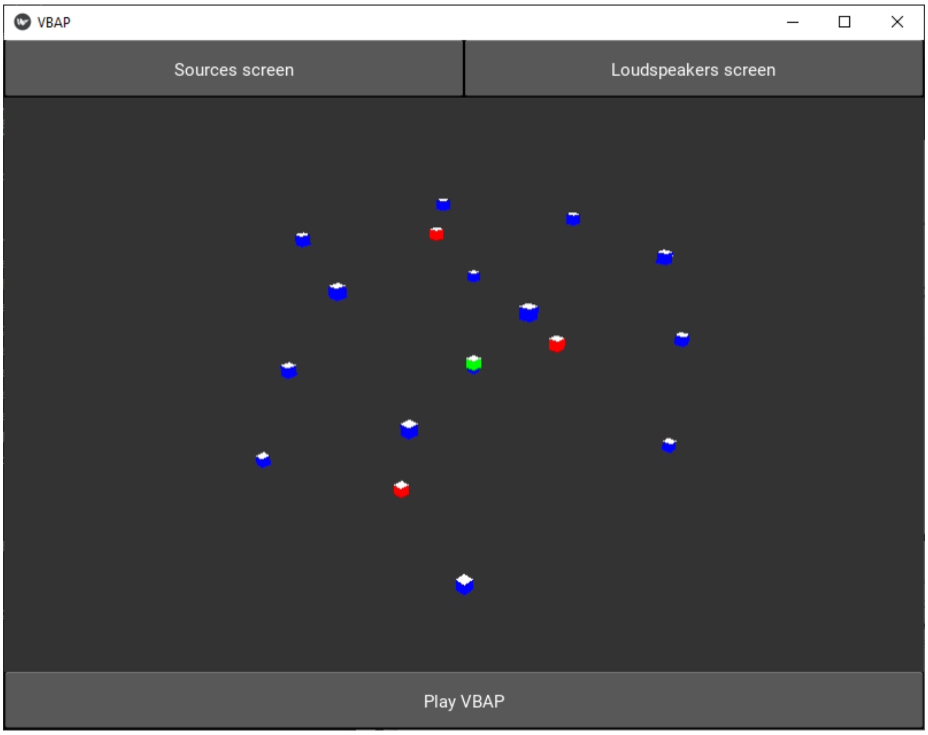

Figure 5: Main screen

4.2 Loudspeakers screen

This screen allows the user to set the loudspeakers configuration. It is presented on Figure 6. This screen is composed by:

- A button to go back to the main screen.
- A button to add a loudspeaker.
- A button to compute the nominal parameters of the loudspeakers.
- Two buttons to open an existing configuration or save the current configuration.

For each loudspeaker, the user can :

- Input a name.
- Input the position (in cartesian coordinates).
- Input the nominal gain and delay.
- Input the soundcard output.
- Delete this loudspeaker.

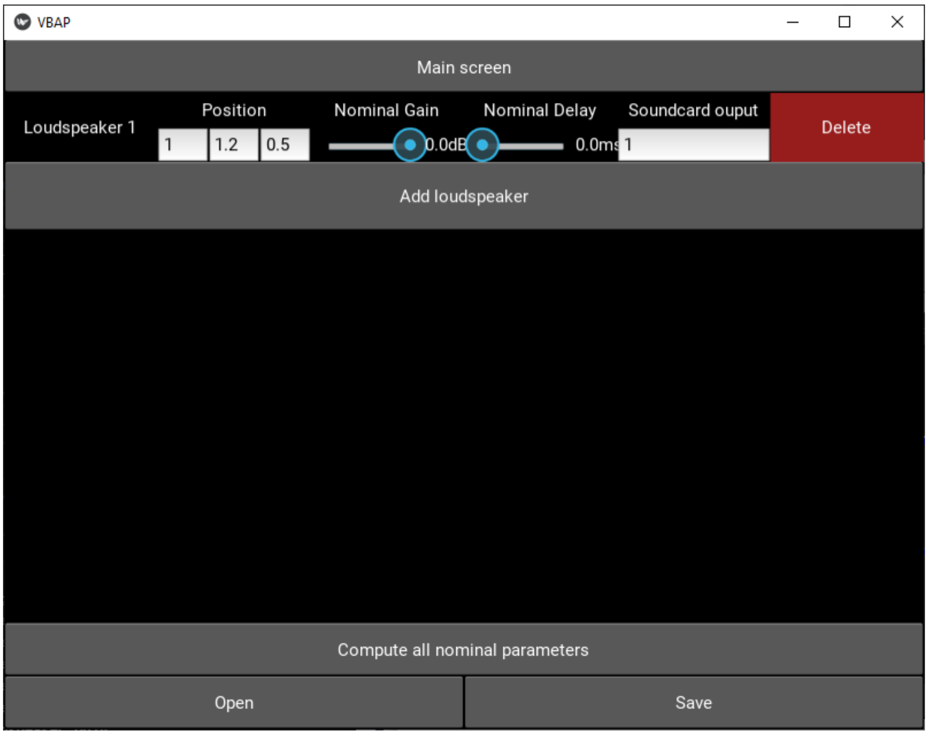

Figure 6: Loudspeakers screen

It can be noticed that the soundcard output can be typed manually by the user or selected from a list of all the sound devices of the loudspeaker. An example of this list is shown on Figure 7 and Figure 8.

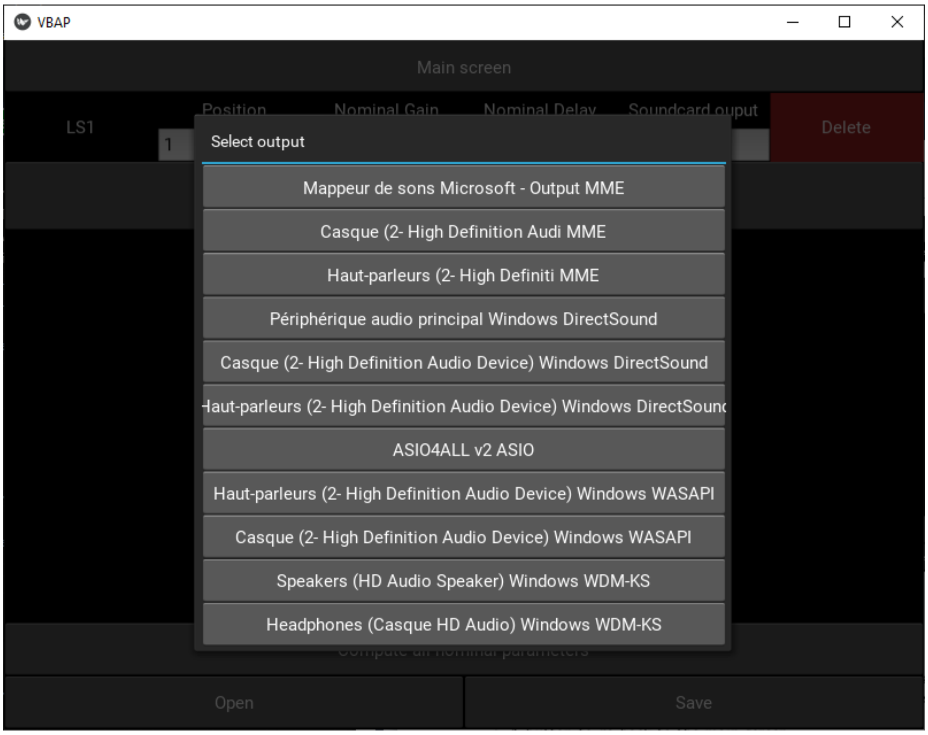

Figure 7: Soundcard selection

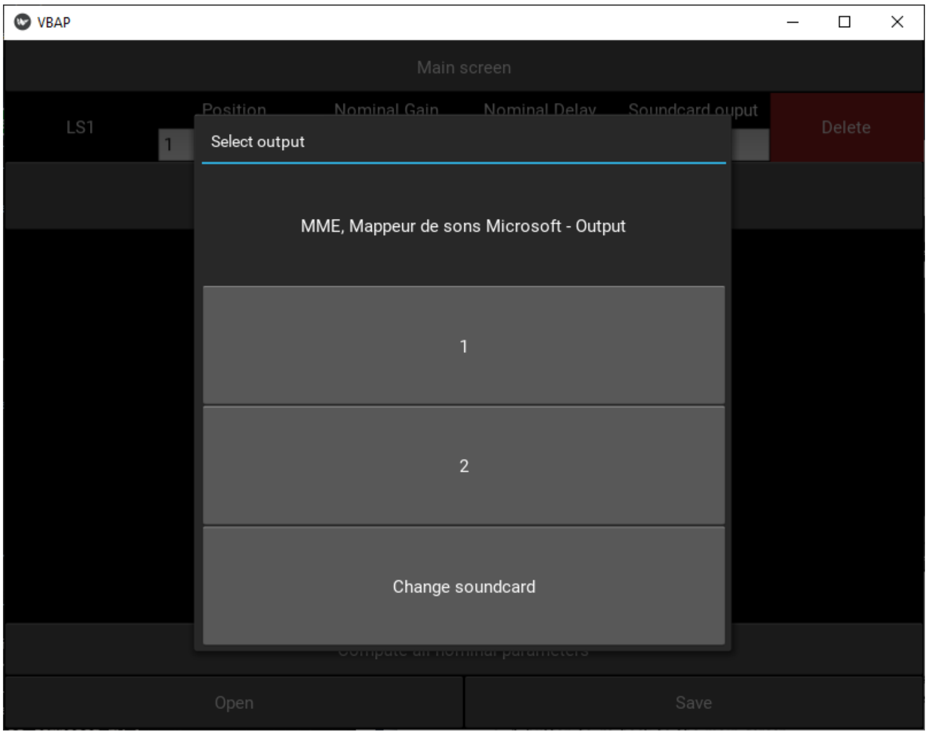

Figure 8: Output selection

4.3 Sources screen

This screen is similar to the loudspeakers screen, but it allows the user to select the sources configuration. This screen is presented on Figure 9. It is composed by :

- A button to go back to the main screen.
- A button to add a source.
- Two buttons to open an existing configuration or save the current configuration.

For each source, the user can:

- Input a name.
- Input the position (in cartesian coordinates).
- Set the level with a slider.
- Select the audio file.
- Delete the source

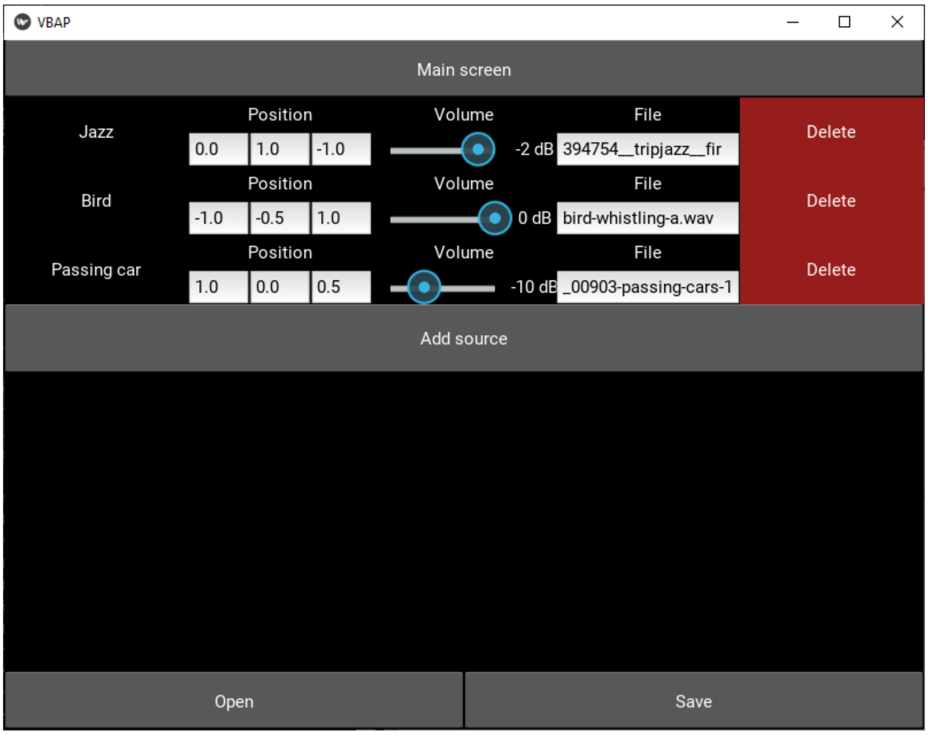

Figure 9: Sources screen

5 Algorithm

5.1 Loudspeakers nominal parameters

The VBAP algorithm assumes that all loudspeakers are at the same distance from the listener. Most of the time this is not the case in a multiloudspeakers system. To compensate that a nominal gain and delay are added to each loudspeaker. First, the farthest loudspeaker have to be find. Then each nominal gain and delay can be computed by considering the loudspeakers as monopoles with Equation 13 and Equation 14

$$G_{nom} = \frac{D_{ls}}{D_{max}} \tag{13}$$

$$\tau_{nom} = \frac{D_{max} - D_{ls}}{c} \tag{14}$$

With D_{ls} the distance between loudspeaker and listener, D_{max} the maximum distance and c the speed of sound.

An example of the effect of nominal gain and nominal delay on loudspeaker transfer functions is shown on Figure 10 and Figure 11. On Figure 10, there is no nominal gain or delay. So, an amplitude and phase difference can be observed at listener position. On figure 11, the same loudspeaker configuration but with nominal gain and delay is used. It can be observed that the amplitude and phase difference between loudspeakers is deleted.

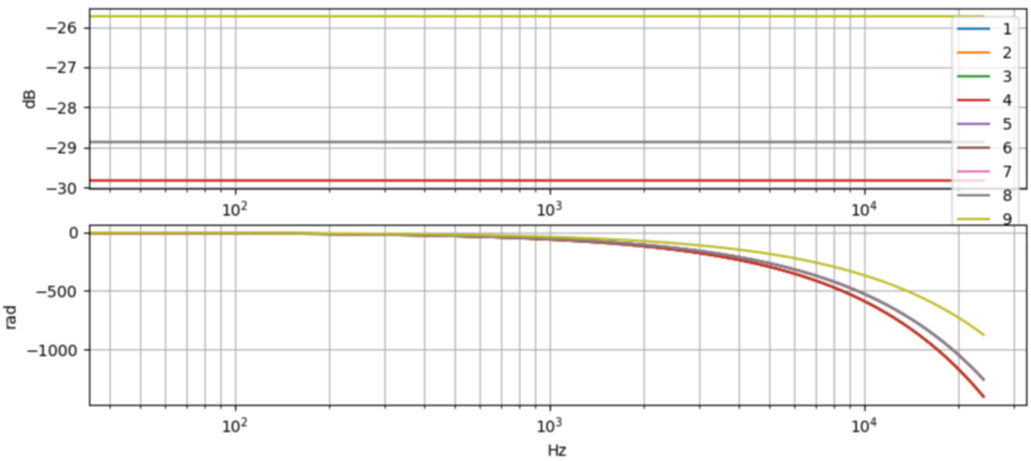

Figure 10: Loudspeaker transfer functions without nominal parameters

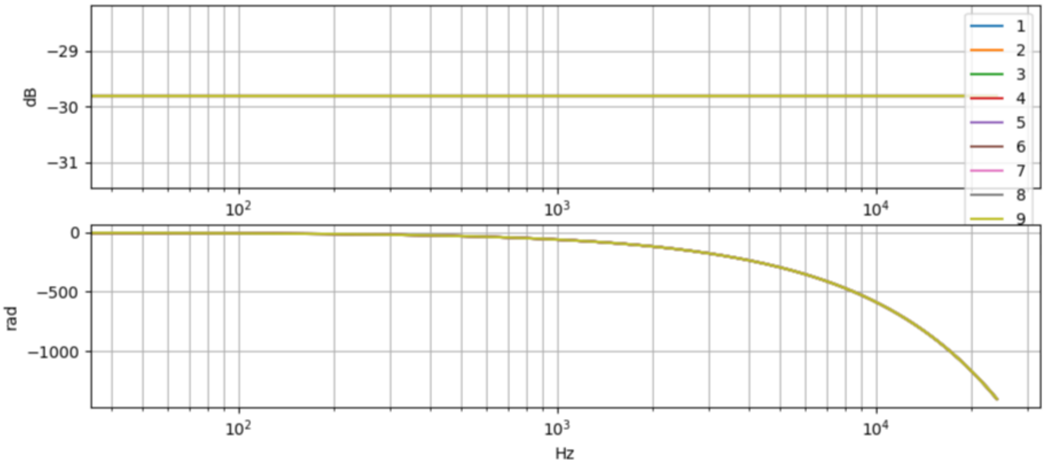

Figure 11: Loudspeaker transfer functions with nominal parameters

5.2 Triplets selection

To use VBAP algorithm, the loudspeakers triplet for each source have to be found. The selection is made by:

- Projecting the sources and the loudspeakers on a sphere around the listener
- Project this spehere on a plane
- Perform a Delaunay triangulation with the loudspeakers
- Check which sources are inside each triplet

5.3 ITD and ILR computation

In order to compute the ITD and ILR, two points standing for the ears of the listener are created at 9 cm on the left and 9 cm on the right of the listener position. The ITD is computed for virtual sources and for VBAP configuration. For virtual sources, the transfer functions between each sources and the right and left ears are computed. Then the ITD and ILR are obtained with Equation 15 and Equation 16.

ILR:
$$Q = \frac{|FRF_{left}|}{|FRF_{right}|} \tag{15}$$

ITD:
$$T = \frac{arg(FRF_{left})}{\omega} - \frac{arg(FRF_{right})}{\omega} \tag{16}$$

With FRF_{left} (or FRF_{right}) the transfer function between source and left (or right) ear, and ω the pulsation.

In the case of VBAP configuration, the same methodology is used but with :

$$FRF_{left} = FRF_{ls1_{left}} + FRF_{ls2_{left}} + FRF_{ls3_{left}} \tag{17}$$

and

$$FRF_{right} = FRF_{ls1_{right}} + FRF_{ls2_{right}} + FRF_{ls3_{right}} \tag{18}$$

With FRF_{ls1}, FRF_{ls2}, FRF_{ls3} the FRF of the first, second and third loudspeakers of the triplet.

This methodology allows to plot the ITD and ILR difference between, real sources position and VBAP. An example is shown on Figure 12.

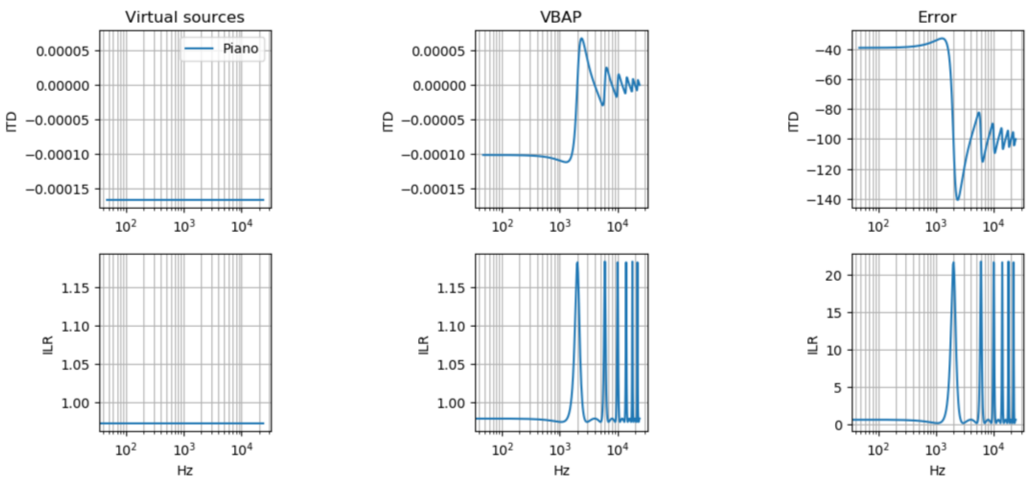

Figure 12: ILR and ITD comparison between real source and VBAP

On figure 12, it can be observed that before 1000 Hz the ILR of real source and VBAP are very close (around 1% of error), but ITD is very different (around 40% of error).

This is perfectly what can be expected from an **amplitude** panning. Also, because in this simulation the shadow effect of the head is neglected (and many other effects), the results over 1000 Hz can't be trusted.

Part III
Hardware

6 Audio system

The synoptic of the hardware is depicted on Figure 13. It can be noticed that this synoptic is simple and classical for a multi-loudspeakers system. The equipment from the University WFS project will be used.

$$\boxed{\text{PC}} \rightarrow \boxed{\text{Audio interface}} \xrightarrow[\times 16]{} \boxed{\text{Digital to analog converter}} \xrightarrow[\times 16]{} \boxed{\text{Amplifier}} \xrightarrow[\times 16]{} \boxed{\text{Loudspeakers}}$$

Figure 13: Hardware synoptic

The audio interface will be a **RME Madiface**. The digital to analog converter will be a **Ferrofish**. The amplifiers are **HPA D604**. And the loudspeakers will be the loudspeakers handmade by Alann Renault.

The real system is shown on Figure 14.

Figure 14: Audio system

7 Structure

The main issue is the structure to handle the loudspeakers. The structure have to fit in the listenning room of the LAUM and have to be easy-to-move and stand alone to be used in different show (such as the IMDEA show). The first idea was to use an aluminium structure commonly used for light show. A draft of a structure that can fit in the listenning room is depicted on Figure 15.

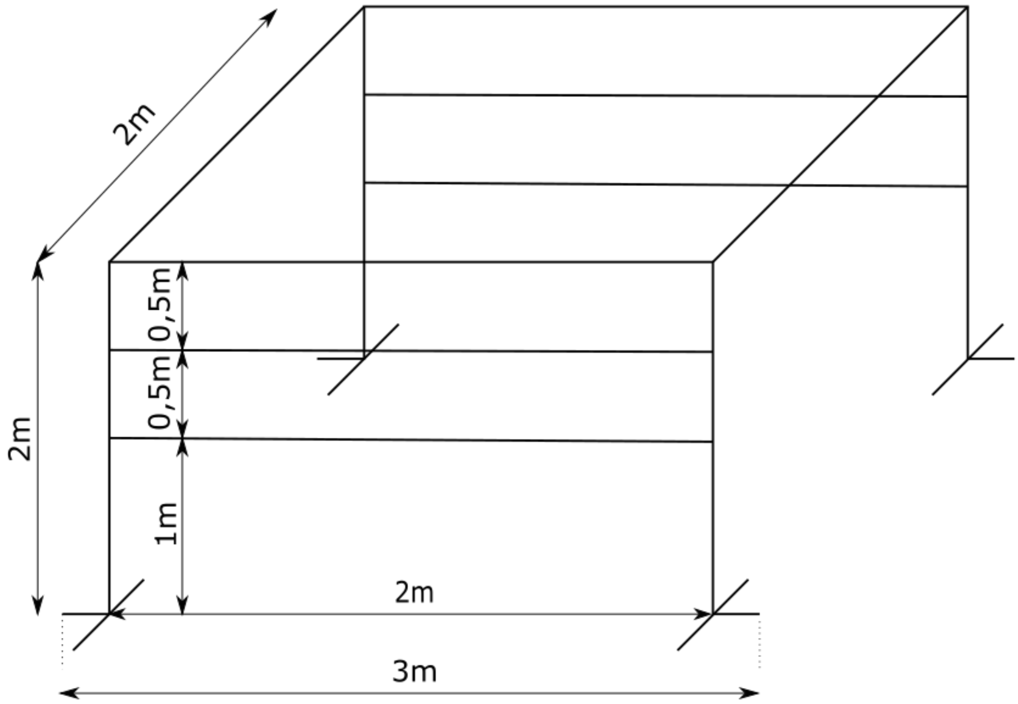

Figure 15: Structure draft

But this type of structure is expensive. The price of the structure of Figure 15 is presented in Table 1. This price is off-budget. Therefore, other solutions should be consider.

Products	**€/u**	**Quantity**	**€**
Tube 50	35	20	700
Tube 100	39	4	156
Tube 200	55	8	440
Angle	16	45	720
Bracket	13,90	16	222,4
Total			2238,4

Table 1: Structure price

The solution of a handmade cube with a Bosch or Norkan system is considered. The main part of the structure is modeled on Figure 16. Then horizontal and vertical bar can be added to add loudspeakers or reinforce the structure.

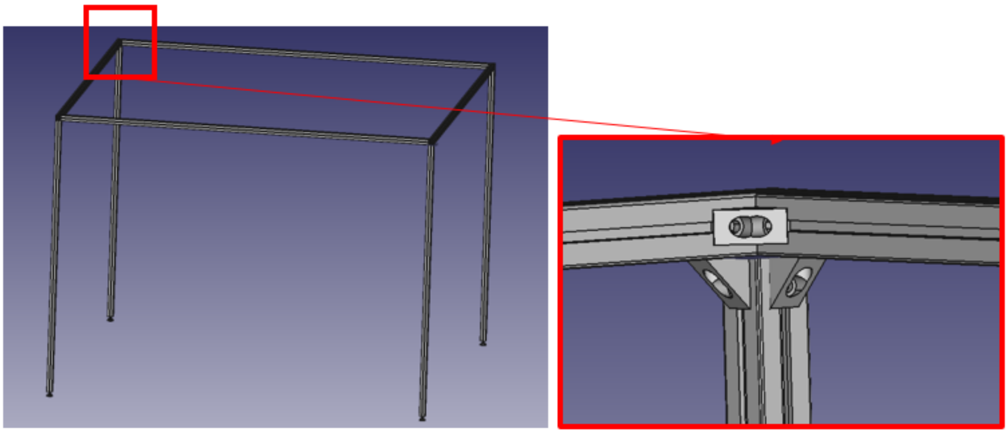

Figure 16: Norkan structure

Some pictures of the final structure in the LAUM listening room are presented on Figure 17 and Figure 18.

Figure 17: Built structure in the listening room

Figure 18: Built structure in the listening room

Part IV
Calibration and validation

8 Calibration

The system still have to be calibrated. Even if the loudspeaker are made with the same drivers and the same enclosure, they aren't perfectly identical. The VBAP algorithm assumes that every loudspeaker are identical. The sensitivy of each loudspeaker have to be calibrate to be identical. This can be done with the volume potentiometers of the amplifiers or the volume slider in the loudspeaker screen of the GUI.

This steps haven't be done yet for the configuration in the LAUM listening room. This can be done following the below steps.

- Define a distance d for the measurement point (for example 20cm).
- Define a measurement frequency f (for example 1kHz).
- Define a desired level P (for example 85 dB)
- For each loudspeaker measure the level at distance d with a sine at frequency f. Tune the potentiometer until reaching the level P.

9 Validation

To validate the project, two types of measurement are proposed: objective and subjective test. The aim of the objective test is to measure the error of physic clues between the VBAP configuration and a point source. But spatial sound is a very complex subject and objective test aren't enough to evaluate the quality of a system. That's why the final assessment should be let to the user evaluation. The subjective test is there to complement the objective test. Those tests haven't been perform yet for the project but they are described in this section.

9.1 Objective test

This test is performed with a recording dummy head such as the one presented on Figure 19.

Figure 19: Recording dummy head

This dummy head is placed at the listener position. First, the signal is generated by a point source. Then the same signal is generated with the VBAP algorithm and multiloudspeaker system.

With the signal recorded by the left and right ear, the ITD and ILd can be computed and compared for the two configurations.

This experiments have to be done for several source positions.

9.2 Subjective test

The aim of this test is to assess if the user is able to localize the sources at the expected position despite the results on the objective tests. Because this is a subjective test, it has to be performed on a large group of user (which can be difficult according to the actual sanitary context). The idea of this test is to ask the user to give a source position using a card board as the one presented in Figure 20 for many source position. The user have to give an azimuth and an elevation angle And then get the average of position error for all the users and for each position.

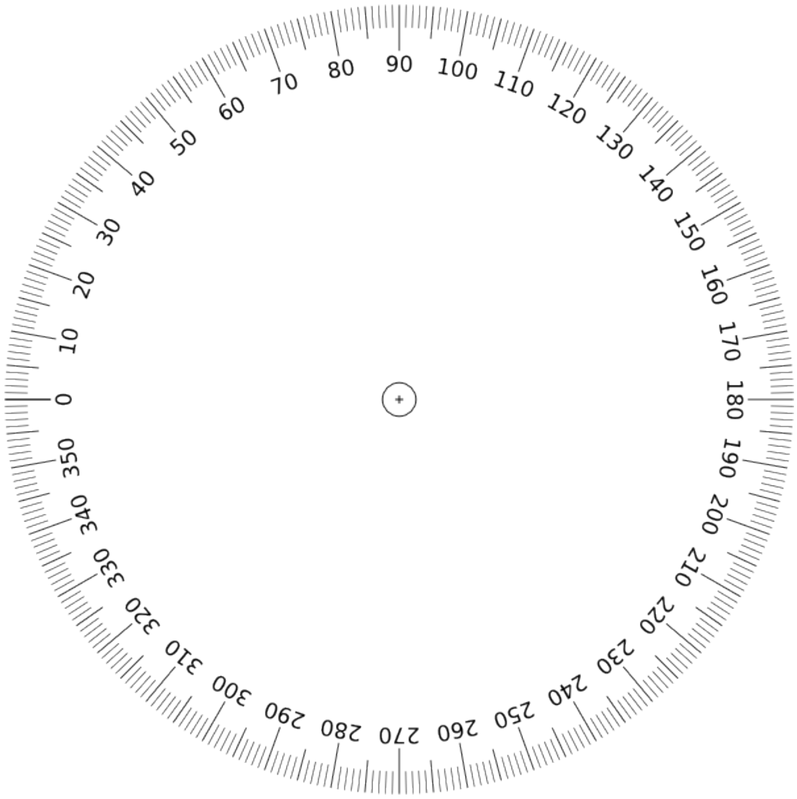

Figure 20: Test cardboard

Part V
Futur work

10 Todo

Some bugs are remaining in the GUI. For example the list of soundcard output can be unreadable when there is too many outputs. Also the 3D view should be improved. The control aren't perfect and adding the name of the sources and loudspeakers in the view can help the user to use the software.

Moreover, the software have a weak error handling. When an unexpected situation happens the software crashes. In the future, a better error handling should be coded with error message for example to avoid this.

Finally, as presented in Part IV, the calibration of the hardware and the objective and subjective test campaign needs to be done.

11 Improvements

A lot of improvements can be imagined for the project. Here is some ideas for the future :

- Be able to make moving sources
- Be able to control source position with a joystick
- Use the hardware with other methodology to compare with VBAP.
- ...

Conclusion

To conclude, this software is ready to be used in real situation. The first subjective tests show that the sounds are mainly located at the loudspeaker position. But this can't be assessed without a full test as describe in section IV.9 The project is available on **GitHub** at `https://github.com/Bouaaah/VBAP_M2`. The main software and some example files are available on this url. Some tasks are remaining and the project can be improved.

References

[1] K. Brandenburg, S. Brix, and T. Sporer. "Wave Field Synthesis". In: *2009 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*. 2009, pp. 1–4.

[2] Jerome Daniel and Sebastien Moreau. "Further Study of Sound Field Coding with Higher Order Ambisonics". In: *Audio Engineering Society Convention 116*. May 2004. URL: `http://www.aes.org/e-lib/browse.cfm?elib=12789`.

[3] Ville Pulkki. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". In: *J. Audio Eng. Soc* 45.6 (1997), pp. 456–466. URL: `http://www.aes.org/e-lib/browse.cfm?elib=7853`.

[4] Lord Rayleigh O.M. Pres. R.S. "XII. On our perception of sound direction". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13.74 (1907), pp. 214–232. DOI: `10.1080/14786440709463595`. eprint: `https://doi.org/10.1080/14786440709463595`. URL: `https://doi.org/10.1080/14786440709463595`.

[5] S. S. Stevens and E. B. Newman. "The Localization of Actual Sources of Sound". In: *The American Journal of Psychology* 48.2 (1936), pp. 297–306. ISSN: 00029556. URL: `http://www.jstor.org/stable/1415748`.

[6] A. W. Mills. "Lateralization of High-Frequency Tones". In: *The Journal of the Acoustical Society of America* 32.1 (1960), pp. 132–134. DOI: `10.1121/1.1907864`. eprint: `https://doi.org/10.1121/1.1907864`. URL: `https://doi.org/10.1121/1.1907864`.

[7] T. T. Sandel et al. "Localization of Sound from Single and Paired Sources". In: *The Journal of the Acoustical Society of America* 27.5 (1955), pp. 842–852. DOI: `10.1121/1.1908052`. eprint: `https://doi.org/10.1121/1.1908052`. URL: `https://doi.org/10.1121/1.1908052`.

[8] Ewan A. Macpherson. "A Computer Model of Binaural Localization for Stereo Imaging Measurement". In: *J. Audio Eng. Soc* 39.9 (1991), pp. 604–622. URL: `http://www.aes.org/e-lib/browse.cfm?elib=5969`.

[9] Helmut Haas. "The Influence of a Single Echo on the Audibility of Speech". In: *J. Audio Eng. Soc* 20.2 (1972), pp. 146–159. URL: `http://www.aes.org/e-lib/browse.cfm?elib=2093`.

[10] Benjamin Bernfeld. "Attempts for Better Understanding of the Directional Stereophonic Listening Mechanism". In: *Audio Engineering Society Convention 44*. Mar. 1973. URL: `http://www.aes.org/e-lib/browse.cfm?elib=1743`.

[11] D. M. Leakey. "Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Sound Systems". In: *The Journal of the Acoustical Society of America* 31.7 (1959), pp. 977–986. DOI: `10.1121/1.1907824`. eprint: `https://doi.org/10.1121/1.1907824`. URL: `https://doi.org/10.1121/1.1907824`.

[12] A. Franck, W. Wang, and F. M. Fazi. "Sparse ℓ_1-Optimal Multiloudspeaker Panning and Its Relation to Vector Base Amplitude Panning". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25.5 (2017), pp. 996–1010.

Appendices