

Master thesis on Sound and Music Computing  
Universitat Pompeu Fabra

# Subjective Evaluation of the Localization Performance of the Spherical Wavelet Format Compared to Ambisonics

Rubén Eguinoa Cabrito

**Supervisors:** Davide Scaini, Ricardo San Martín

31 August 2021





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Spatial audio . . . . .	1
1.2	Motivation . . . . .	2
1.3	Objectives . . . . .	2
1.4	Structure of the report . . . . .	2
<b>2</b>	<b>State of The Art</b>	<b>3</b>
2.1	Localization of sound by humans . . . . .	3
2.2	Techniques to record and playback spatial audio . . . . .	4
2.3	Ambisonics . . . . .	4
2.4	Spherical Wavelet Format (SWF) . . . . .	5
2.5	Subjective listening tests . . . . .	5
2.5.1	Listening tests for spatial audio . . . . .	6
2.5.2	Current tests in spatial audio . . . . .	6
<b>3</b>	<b>Methods</b>	<b>8</b>
3.1	Ambisonics . . . . .	8
3.1.1	Spherical harmonics . . . . .	9
3.1.2	Encoding of a sound field . . . . .	10
3.1.3	Decoding . . . . .	12
3.2	Spherical Wavelet Format (SWF) . . . . .	12
3.2.1	Signal discretization . . . . .	13
3.2.2	Signal decomposition . . . . .	13

3.2.3	Signal reconstruction . . . . .	14
3.2.4	Decoding to loudspeakers . . . . .	15
<b>4</b>	<b>Experimental Setup</b>	<b>16</b>
4.1	Hardware . . . . .	16
4.2	Software . . . . .	18
4.2.1	Specifications . . . . .	19
4.2.2	The application . . . . .	19
<b>5</b>	<b>Listening Tests</b>	<b>23</b>
5.1	Localization test . . . . .	26
5.2	Source width test . . . . .	27
<b>6</b>	<b>Results</b>	<b>29</b>
6.1	Statistical analysis . . . . .	33
<b>7</b>	<b>Discussion</b>	<b>37</b>
7.1	Conclusions . . . . .	37
7.2	Future work . . . . .	38
	<b>List of Figures</b>	<b>39</b>
	<b>List of Tables</b>	<b>41</b>
	<b>Bibliography</b>	<b>42</b>



## Acknowledgement

I would like to thank my supervisors Davide and Ricardo, and also Daniel, for all the help and facilities they have given me during this process; and for always been up to any new idea, no matter how crazy it was. I truly appreciate all the advice you have given me for my future.

Of course, mention my family. Even if sometimes we have been far, they have always been a support, even if they sometimes did not have any idea on what I was working on.

I also want to thank all the people who have taken the time to come to the lab and perform the listening tests in this unusual situation. Without them this thesis would be a white paper. I owe you all a beer.



## **Abstract**

A common goal of most spatial audio techniques is to reproduce the precise location and size of sound sources. Ambisonics is a well-established spatial audio technique that renders sound sources with increasing accuracy as the Ambisonics' order increases. Recently, a novel spatial audio format that replaces spherical harmonics with a set of functions based on wavelets has been proposed. The Spherical Wavelet Format (SWF) aims to improve Ambisonics localization, especially at low orders. In this thesis, a study investigates the perceptual spatial properties of both techniques by means of a set of MUSHRA tests.

Keywords: Ambisonics; Spherical Wavelet Format; localization; listening tests; spatial audio



# Chapter 1

## Introduction

This master thesis aims to compare the spatial properties of the Spherical Wavelet Format (SWF) and Ambisonics. These two techniques provide a complete chain of encoding and decoding for the generation of three dimensional audios.

In section 1.1 the spatial audio is explained and different techniques are described. Section 1.2 outlines the motivation of this thesis, of which the objectives are written in section 1.3. Finally, the structure of the report is detailed in 1.4

### 1.1 Spatial audio

The spatial or 3D audio pretends to obtain a realistic recreation of the surround hearing of humans. With the auditory system a person can describe the position of a forthcoming sound and also the distance of where the source is sounding. A correct recreation should provide the same experience to the subject, providing a completely immersive feeling.

In the last decades, with the development of Virtual Reality systems and 360° videos, the implementation of this field has increased exponentially and new techniques have arose. Ambisonics or VBAP are very well-established techniques to recreate this spatial audio.

## 1.2 Motivation

The development of SWF brought a new complete audio chain for the encoding and decoding to the spatial audio. This technique may improve the outcomes of Ambisonics, and different tests and evaluations have to be carried out to compare both.

During the development of SWF different tests were done, but the loudspeaker layout and the amount of subjects was not representative at all to obtain convincing results. Now there is the possibility to perform the tests in a bigger layout and with a bigger range of subjects, what should return better and more complete data to evaluate the technique and compare it with others.

## 1.3 Objectives

The main objective of this thesis is to evaluate with real subjects the spatial properties of SWF and to compare them with different Ambisonics orders. To evaluate the techniques different subjective listening tests are going to be held in a 24 loudspeaker spherical layout.

## 1.4 Structure of the report

The thesis is structured as follows: Chapter 2 describes the state of the art, with the currently existing technologies and the different subjective listening test types available. In Chapter 3 the two technologies involved in the thesis are explained. In Chapter 4 the developed tools are described. Chapter 5 details the designing of the tests. In Chapter 6 the results are presented. Finally, conclusions and future work are drawn in Chapter 7.

# Chapter 2

## State of The Art

### 2.1 Localization of sound by humans

In human hearing [1], the two most relevant information to decide the localization of a live sound event are the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD). Both help to the brain to decide in which direction a sound source is situated. The ITD is the difference between the amount of time that sound needs to arrive at both ears, it can be measured as the angle of incidence of the sound with the listener, which determines the additional path that needs to travel to arrive at the other ear. The ILD refers to the difference in loudness and frequency distribution between the two ears.

Those two differences are not enough to correctly position a sound source, as front-back and up-down ambiguities appear. To overcome this problem three very helpful mechanisms come into play: the frequency response of monaural signals (the pinnae frequency response), small head turns (which change the ITD) and the human sensitivity to inter-aural signal coherence. All these features are of rather individual nature and depend on the complexion of the person: chest, shoulders, head and pinnae have a big impact in the perception of sounds.

## 2.2 Techniques to record and playback spatial audio

The 3D audio aims to recreate the way sound sources are perceived. Also known as spatial audio or immersive sound, is a research field with growing interest, thanks to the development of Virtual Reality and 360° videos. Its goal is to obtain a completely immersive sound experience, where the users can perceive moving sources around themselves. Many techniques have been developed, this thesis is focused on a comparison between Ambisonics and the Spherical Wavelet Format, being the first the most common (in the research realm) and the second a newly developed technique.

## 2.3 Ambisonics

Ambisonics [2] has been one of the most used technology for three dimensional audio. Starting its development in 1970 with the First Order Ambisonics (FOA), during the years it has evolved with the Higher Order Ambisonics (HOA). The Ambisonics theory is based on the decomposition of the sound field in terms of spherical harmonics, and it covers all the steps of the audio production chain: recording, encoding, transmission and decoding.

The order defines the number of spherical harmonics needed, the higher the better description of the sound space will be obtained, and a bigger sweet spot will be created during the playback. The spherical harmonics permit a series representation with directional functions, from where we can extract a coefficient or weight. The encoding of Ambisonics is obtained by multiplying the incoming signal with the corresponding weight.

For the decoding of Ambisonics signals a decoding matrix is needed, which provides the information about the positions of the loudspeaker layout. *AllRAD* [3] by the Institute of Electronic Music and Acoustics of Austria (IEM) is a widely used plugin. It provides the user the option to design an Ambisonics decoder for the desired loudspeaker layout. The users also have the option to export the configuration to a JSON file.



## 2.4 Spherical Wavelet Format (SWF)

The Spherical Wavelet Format (SWF) [4, 5] is a newly developed spatial audio format. It replaces the Ambisonics' spherical harmonics by an alternative set of functions with compact support (spherical wavelets). It also covers all the steps of audio production, with a complete audio chain from encoding to decoding based on the discrete spherical wavelets built on a multiresolution mesh.

The encoding is done by representing the position of the sound sources on a finest mesh. Then, the wavelet transform is applied and the subdivision mesh is down-sampled recursively with two decomposition filters, obtaining the set of encoded signals.

The decoding algorithm for this technique is the IDHOA decoder [6], an open source software capable to generate decoding matrices to irregular layouts for both Ambisonics and the new wavelet format by optimizing a cost function that weights different perceptual observables (pressure, velocity, energy and intensity). It was initially developed for the decoding of Ambisonics to non-regular loudspeakers layouts. Improvements were made to find an optimal decoding of a given SWF.

## 2.5 Subjective listening tests

Even if sound physical observables can be measured, the human hearing is a complex system that needs a different approach. The best way to assess how the audio is perceived is the performing of listening tests with groups of subjects. The volunteers are exposed to different signals or sounds and, following a guideline on what and how to evaluate, they have to quantify their experience. The test type and used sounds vary depending on the aim of it (quality of sound, width of the source, directivity of loudspeakers...) and the measures are taken asking directly to users.

### 2.5.1 Listening tests for spatial audio

Recommendations and standards give advice on which and how listening tests can be carried for different purposes. For spatial audio the most important aspect is the positioning of the sound sources; but also the width of the source and how smoothly its movement is recreated. The tests must be developed in order to measure how effectively those variables are represented.

MUSHRA (ITU-R BS.1534-3) [7] tests are used to measure the perceived quality in audio. This test is useful to assess the loss of quality (due to codec compression, for example) by making a comparison between a reference signal and a degraded one. The scale used to rate the differences goes from 0 to 100, where 0 is bad and 100 excellent.

The Mean Opinion Score (MOS) (ITU-T P.800.1) [8] is another way to test quality of audio, where the result is obtained from “values on a predefined scale that a subject assigns to his opinion of the performance of a system quality” [8]. In this case the range goes from 1 to 5, where 1 is the lowest perceived quality and 5 the highest. The final result is an arithmetic mean between all the results obtained.

The ABC tests (ITU-R BS.1116-3) [9] are defined to detect small degradations. In the test, three different stimuli are shown, where one is always the reference sound. The subjects have to measure the degradation between the other two stimuli and the reference one, in a scale from 1 to 5, where 1 is very annoying and 5 imperceptible.

Finally, the ABX tests help to determine whether two stimuli are different from each other. In this case, the X sound is the reference one, and the subjects have to decide which of the A or B sounds are more similar to it.

### 2.5.2 Current tests in spatial audio

The both technologies involved in this work have been already tested. Ambisonics counts already with many experiments on its localization properties [10, 11, 12]. It has also been compared with other techniques such as VBAP or MDAP [13, 10].

For example, in the *Phantom Sources using Multiple Loudspeakers in the Horizontal Plane* dissertation by Matthias Frank [13] subjects had to point the position from where they feel the sound was arriving. Four panning methods were evaluated, with 9 directions between  $0^\circ$  and  $-45^\circ$  (to the right) each one. The listening tests used broadband pink noise as stimulus and were evaluated by 41 professional musicians and audio engineers or students in these fields.

The SWF technique has been also tested during its development, and has comparisons with VBAP and the second and third order of Ambisonics [5]. In this thesis a deeper investigation is going to be held to assess the spatial properties of SWF versus Ambisonics.

# Chapter 3

## Methods

This chapter explains Ambisonics and the Spherical Wavelet Format and the theoretical basis behind them.

### 3.1 Ambisonics

Ambisonics [2] is a complete theory to encode, manipulate and decode three dimensional sound scenes. It was developed in the 1970's based on the coincident audio recording and playback principles. Although it was not very successful in its early days, nowadays has reached big commercial success mainly because of the boost of 360° videos and Virtual Reality (VR).

To accurately capture the direction of a sound source, Ambisonics discretizes the space into spherical harmonics (SH). A given region of a sphere, being this the space around a microphone or a head, will be associated to a set of SH with their corresponding weight. Thus, it treats a sound scene as a complete 360° sphere with sound coming from any direction to a central point. This central point is located at the position of the recording microphone or *sweet spot*. For lower orders, which use less SH, once this zone is left the source perception is no longer reliable. This is one of the major problems with First Order Ambisonics (FOA). Together with the need for higher spatial resolution, has led to the development of Higher Order

Ambisonics or HOA.

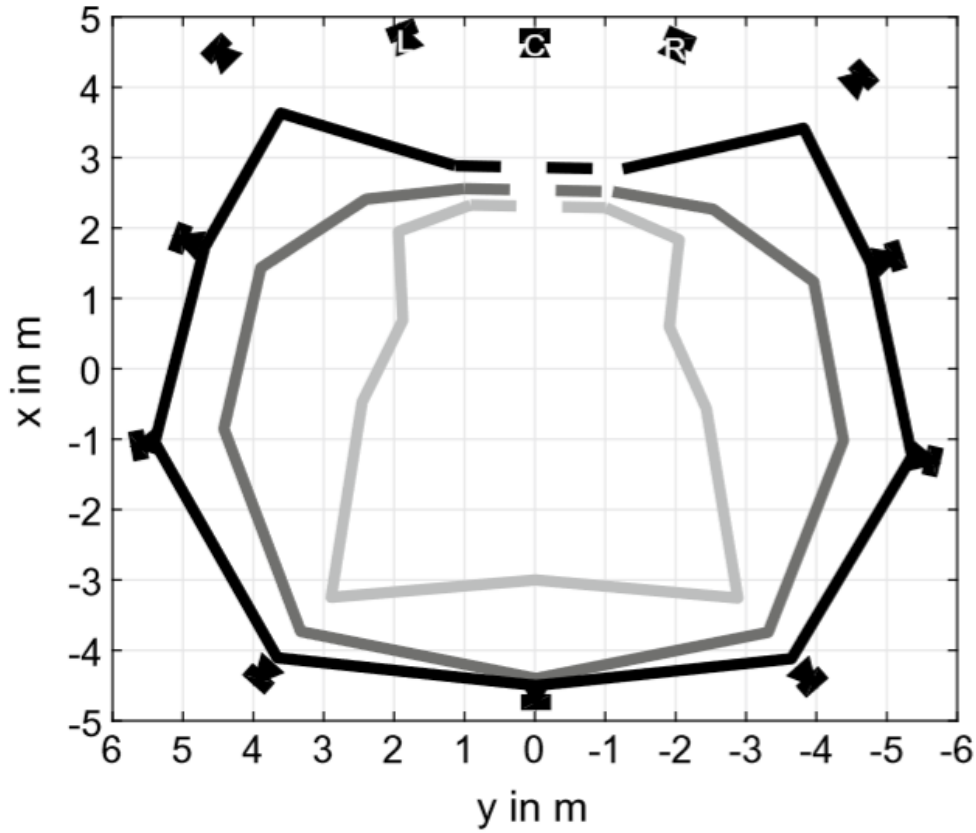


Figure 1: Perceptual sweet spot area for first (light grey), third (dark grey) and fifth (black) Ambisonics orders. Reproduced from Ref. [2]

### 3.1.1 Spherical harmonics

The spherical harmonics are the basis of Ambisonics, since the signal is reconstructed from their combination. The bigger the quantity of harmonics, the better spatial resolution. The number of harmonics needed is given by the working Ambisonics order: the higher the order, the greater the quantity of harmonics. For a given order  $N$ ,  $(N + 1)^2$  harmonics will be used,  $2N + 1$  if the reproduction is in two dimensions.

A First Order Ambisonics (FOA) signal uses the first four SH, these are also collectively called B-Format. For a third order signal 16 harmonics are needed, which are shown in Figure 2

There are two main conventions for the component ordering: *Furse-Malham* (FuMa)

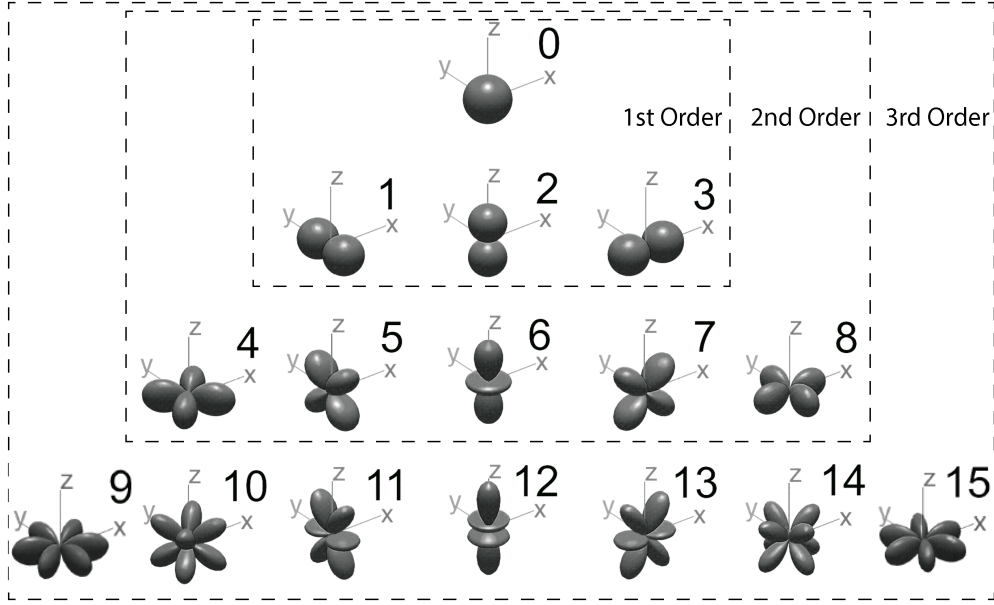


Figure 2: Spherical harmonics for the first three Ambisonics orders plotted as a polar diagram. Reproduced from Ref. [2]

and *Ambisonics Channel Number* (ACN). The first convention names the harmonics with letters, following the B-format character naming (W, X, Y, Z). On the other hand, ACN names the channels numerically ( $ACN = n^2 + n + m$ , where  $n$  is the order and  $m$  the degree).

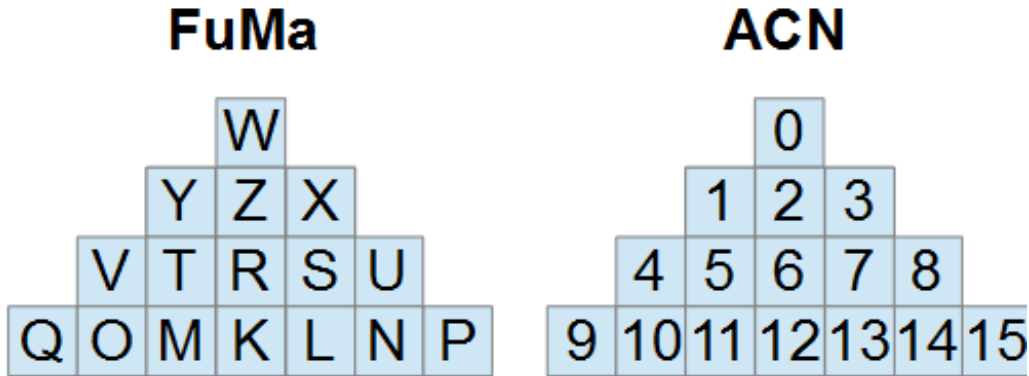


Figure 3: FuMa and ACN channel ordering conventions up to the third order.

### 3.1.2 Encoding of a sound field

To encode any signal  $S$  defined in a  $(\varphi, \delta, t)$  point over the surface of a sphere will need the  $Y_n^m(\varphi, \delta)$  SH and the expansion coefficient  $\mathcal{O}_{n,m}(t)$ , where  $n$  is the order and  $m$  the degree. The scene will be the combination of all the harmonics multiplied

by their associated signal (the signal arriving at the harmonic at a specific instant  $t$  is multiplied and summed for all the degrees and orders):

$$S(\varphi, \delta, t) = \sum_{n=0}^N \sum_{m=-n}^n Y_n^m(\varphi, \delta) \varnothing_{n,m}(t) \quad (3.1)$$

The SH are built with two terms: the normalization term  $N_n^{|m|}$  and its associated Legendre polynomial  $P_n^{|m|}$  (see Equation 3.2). The associated Legendre polynomials are the canonical results to the general Legendre equation, which solves the Laplace equation.

$$Y_n^m(\varphi, \delta) = N_n^{|m|} P_n^{|m|}(\sin(\delta)) \begin{cases} \sin |m|\varphi & m < 0 \\ \cos |m|\varphi & m \geq 0 \end{cases} \quad (3.2)$$

Where  $N$  is the normalization which has different conventions, the best known being full normalization (N3D) and Schmidt semi-normalization (SN3D, most widely used in Ambisonics).

With SN3D normalization, no term will exceed the peak value of order zero. However, N3D normalization ensures equal power in all the encoded components, as long as there is a perfectly diffuse 3D field, which can lead to problems in the decoding and cause distortion at higher orders. Equation 3.3 shows the N3D and SN3D normalizations, together with their relationship:

$$\begin{aligned} N_n^{|m|}(\text{N3D}) &= \sqrt{\frac{(2n+1)(n-|m|)!}{2(n+|m|)!}} \\ N_n^{|m|}(\text{SN3D}) &= \sqrt{\frac{1(n-|m|)!}{2(n+|m|)!}} \\ N_n^{|m|}(\text{N3D}) &= \sqrt{2(n+1)} N_n^{|m|}(\text{SN3D}) \end{aligned} \quad (3.3)$$

### 3.1.3 Decoding

One of the challenges of Ambisonics is that the decoding to a loudspeaker layout is non-trivial, specially for non-regular setups. There are different approaches to the problem, such as decoding to an intermediate layout (AllRAD [3, 14]), mode matching [15], or finding the optimal decoding by solving a non-linear optimization problem [16, 17, 18, 19, 20, 21, 22, 6, 23]. For this work this latter approach is adopted by relying on the IDHOA decoder [6, 23].

## 3.2 Spherical Wavelet Format (SWF)

SWF is constructed in the framework of second generation wavelets [24, 25]. This implies that the wavelets used are defined on a recursive mesh that samples the sphere with increased precision as the SWF order increases. SWF itself is constructed over a discrete polygonal mesh; to go from the continuous points on the sphere to the nodes on the mesh an interpolation method is used.

This mesh is built recursively from a primitive polygonal mesh. Following the original SWF implementation, an octagonal polygonal mesh is considered, subdivided with the so-called Loop scheme [26]. In contrast to Ambisonics, where the encoding and decoding basis functions are the same (the SH), in SWF the encoding and decoding filters are different. Also in contrast to Ambisonics, in SWF the encoding and decoding filters are applied recursively.

A SWF is defined to be each one of the spherical audio encodings determined by:

- i) a recursive subdivision mesh over the sphere, ranging from the coarsest level 0 (the based mesh) to the finest level  $n$ ;
- ii) a set of bi-orthogonal filters  $\{\mathbf{A}^j, \mathbf{B}^j, \mathbf{P}^j, \mathbf{Q}^j\}$ , with  $j = 1, \dots, n$ , defining a wavelet transform, and
- iii) a truncation level  $\ell \in [0, n]$ , defining the level of the wavelet decomposition.



The SWF channels will be composed by the coarser data approximations,  $\mathbf{c}^0$ , in addition to details up to order  $\ell - 1$  ( $\mathbf{d}^0, \dots, \mathbf{d}^{\ell-1}$ ); at level 0, only the coarser approximation  $\mathbf{c}^0$  remains.

Therefore, there is not just one, but actually many possible SWF formats. This thesis relies on the original SWF proposal [5], based on i) a recursive subdivision mesh over the sphere based on the primitive octahedronal mesh; ii) the set of dual interpolating filters (the most important feature of them is that the decomposition filters interpolate on the first neighbours), and iii) truncation levels 0 and 1, having 6 and 18 channels, respectively.

### 3.2.1 Signal discretization

The Spherical Wavelet Format decomposition starts with a continuous source distribution over the sphere. An example of such source distribution could be a delta function at the position of a given point source. First step is to sample the continuous source distribution over the sphere to the nodes on the mesh by using tri-linear interpolation, by means of which a given point source on the sphere will be encoded to a maximum of three nodes on the mesh. This will lead to a set of data defined over the finest level of a mesh, the sampled source distribution or *fine data*  $\mathbf{f} = (f_1 \cdots f_N)^T$ .

In this case, the fine data consists of 66 data elements, corresponding to the vertices of the mesh subdivision at level 2.

### 3.2.2 Signal decomposition

The fine data enters a downsampling process that decomposes it into two signals or sets of data: a *coarse* approximation  $\mathbf{c}^{n-1}$  and the remainder, the *details*  $\mathbf{d}^{n-1}$ . The coarse data vector  $\mathbf{c}^{n-1}$  represents a spatially low-passed and downsampled version of the fine data  $\mathbf{f}$ . This decomposition is carried out with the so called *decomposition*, *encoding* or *analysis filters*  $\mathbf{A}^n$  and  $\mathbf{B}^n$ . The filters connect two levels: from the fine level  $n$  to a coarser level  $n - 1$ . There are as many decomposition filters as mesh

levels minus one.

The decomposition may continue up to the coarsest level available (level 0). From the 66 vertices or data elements of the second level mesh, first level decreases the mesh to 18 points, and zeroth level to 6. This process returns a set of  $n-1$  detail signals or wavelet coefficients,  $\mathbf{d}^0, \dots, \mathbf{d}^{n-1}$ , and one last coarse signal or scaling function coefficients  $\mathbf{c}^0$ . The representation  $\{\mathbf{c}^0, \mathbf{d}^0, \dots, \mathbf{d}^{n-1}\}$  constitutes the wavelet transform. The process is shown in Figure 4.

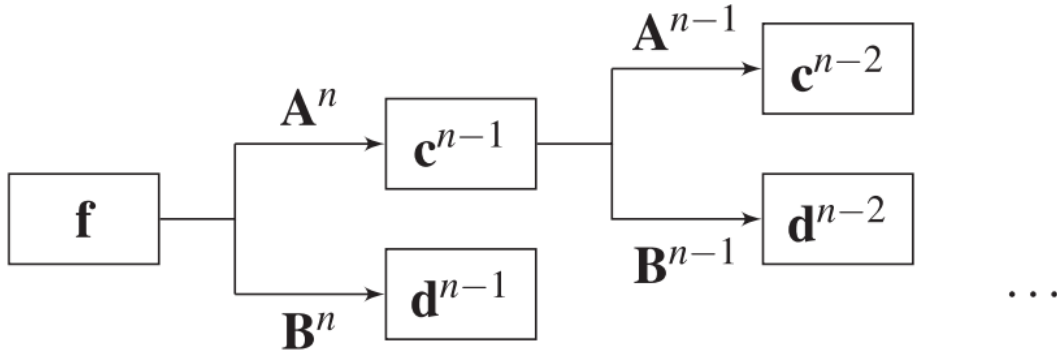


Figure 4: Scheme of the signal decomposition, which illustrates the encoding process to wavelet space. The fine data  $\mathbf{f}$  is decomposed by analysis filters  $\mathbf{A}^n$  and  $\mathbf{B}^n$  into the coarse  $\mathbf{c}^{n-1}$  and details  $\mathbf{d}^{n-1}$  signals respectively. The same operation of decomposition is repeated recursively on the coarse signal  $\mathbf{c}^{n-1}$  up to the desired level. Reproduced from Ref. [5].

SWF is based on the wavelet transform but truncated to a certain level. The coarse data and details will constitute the channels of SWF. At level zero the coarse data  $\mathbf{c}^0$  will constitute the SWF channels (one channel per each node on the base mesh); at level 1 there will be additional channels corresponding to the details  $\mathbf{d}^0$ .

### 3.2.3 Signal reconstruction

A reconstruction of the signal can be done with an upsampling process that increases the spatial resolution of the coarse data  $\mathbf{c}^0$  to the fine data  $\mathbf{f}$ . If the details  $\mathbf{d}^0$  are added, the process will give back the original fine data. This reconstruction is done with the *reconstruction, decoding or synthesis filters*  $\mathbf{P}^n$  and  $\mathbf{Q}^n$  at level  $n$ .

Similarly to the decomposition filters,  $\mathbf{P}^n$  and  $\mathbf{Q}^n$  connect levels, but in the inverse

path: from the coarser level  $n - 1$  to the finest level  $n$ . There are as many reconstruction filters as mesh levels minus one. The reconstruction of the original signal is a recursive procedure that starts from the coarse *level 0* and goes to the finest *level n*. This inverse wavelet transform can be represented as:  $\mathbf{c}^k = \mathbf{P}^k \mathbf{c}^{k-1} + \mathbf{Q}^k \mathbf{d}^{k-1}$ , with  $k = 1, \dots, n$ .

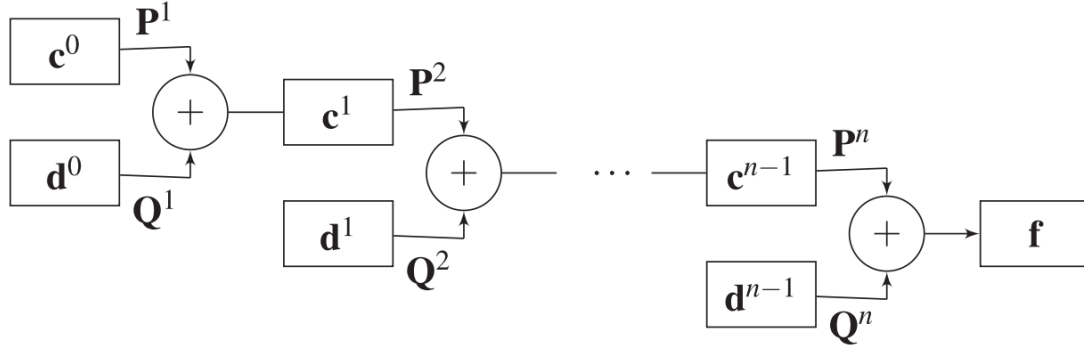


Figure 5: Scheme of the signal reconstruction, which illustrates the decoding process from the wavelet space. The coarse and details signals at the lowest level,  $\mathbf{c}^0$  and  $\mathbf{d}^0$ , are upsampled via the reconstruction filters  $\mathbf{P}$  and  $\mathbf{Q}$  respectively. The resulting signals are summed together to obtain the next level coarse signal  $\mathbf{c}^1$ . The process is repeated at each level with the contribution of the details  $\mathbf{d}^\ell$ , until the original fine data  $\mathbf{f}$  is recovered. Reproduced from Ref. [5].

### 3.2.4 Decoding to loudspeakers

Similarly to Ambisonics, the decoding of SWF to non-regular layouts is non-trivial. Again, the IDHOA decoder is used, which can generate both decodings for Ambisonics and SWF using a non-linear optimization approach. In SWF, IDHOA generates the loudspeaker feeds based on the SWF channels at level 0 ( $\mathbf{c}^0$ ) or 1 ( $\mathbf{c}^0, \mathbf{d}^0$ ).

# Chapter 4

## Experimental Setup

This chapter lists the systems and softwares used for the creation and performing of the experiments.

### 4.1 Hardware

The Acoustics Group of the Universidad Pública de Navarra developed a sound installation to work and experiment with three-dimensional audio.

The structure holding the loudspeakers in place is a sphere, with a diameter of 2.9 meters, built with steel tubes. The five horizontal tubes and the twelve vertical tubes, each at  $30^\circ$ , give a lot of flexibility in the placement of the loudspeakers (Figure 6).

The listening room hosting the structure meets the requirements for reverberation times set out in Recommendation ITU-R



Figure 6: A look into the structure.

BS.1116-3 [9]. For the experiment, a 24 loudspeaker arrangement that corresponds to a spherical 7-design has been used (see positions in Table 1). The so called *t*-*designs* are non-regular polyhedrons designs found by optimization to be regular, approximated with mathematical expressions.

Table 1: Positions of the loudspeakers in the 7-design, in degrees.

Loudspeaker	Azimuth (°)	Elevation (°)
1	-34.4	-25.0
2	-25.5	15.5
3	-19.3	60.0
4	6.2	-60.0
5	12.5	-15.5
6	21.3	25.0
7	55.6	-25.0
8	64.5	15.5
9	70.7	60.0
10	96.2	-60.0
11	102.5	-15.5
12	111.3	25.0
13	145.6	-25.0
14	154.5	15.5
15	160.7	60.0
16	-173.8	-60.0
17	-167.5	-15.5
18	-158.7	25.0
19	-124.4	-25.0
20	-115.5	15.5
21	-109.3	60.0
22	-83.8	-60.0
23	-77.5	-15.5
24	-68.7	25.0

The loudspeakers of the installation have been designed ad hoc for the structure. The loudspeaker boxes are made of PLA and created with the help of a 3D printer, having a spherical shape with a back horn. The transducers are the Dayton Audio PS95-8 full-range woofer fed by two Dayton Audio MA1240A amplifiers and driven with the MOTU 24AO audio interface.

Figure 7 plots the frequency responses of the 24 loudspeakers at their respective

positions on the sphere. They have been obtained with an omnidirectional microphone located in the center of the sphere, using a logarithmic sweep as the excitation signal.

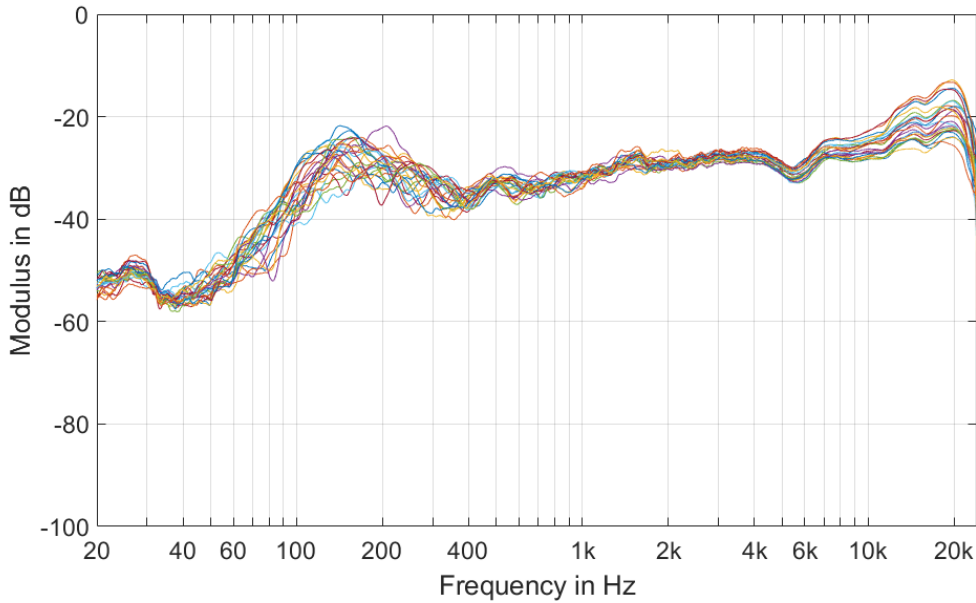


Figure 7: Frequency response for the 24 loudspeakers in the layout. The loudspeakers' frequency response is very consistent in the range between 400 Hz and 7 kHz. At low frequencies the differences in frequency response between different loudspeakers are more evident due to room modes. At frequencies higher than 10 kHz the shadowing effect of the measurement microphone itself may be observed.

## 4.2 Software

To carry out the listening tests a customized desktop application has been created. The application has been developed with the open-source JUCE framework [27]. JUCE provides a wide variety of libraries and it is specifically optimized for developing audio applications and plug-ins. The application is able to perform the two types of tests described in Chapter 5: a MUSHRA test to evaluate the localization, and a modified MUSHRA test to assess the apparent width of an audio source. The code is available as a GitHub repository [28].

### 4.2.1 Specifications

To develop the application important aspects and requirements for the tests had to be taken in account, which are listed below.

- Reproduce multi-channel *wav* type files.
- At least four stimuli plus a reference and an anchor file (six sound files totally).
- A button to stop the playback.
- Random position appearance in each different group of stimuli.
- Continuity of the playback. When there is a change from a stimuli to another it has to be in the same position, the stimuli cannot start from the beginning.
- Only the value for the playing stimuli can be modified, to avoid unwanted changes.
- Connectivity to a MIDI controller.

### 4.2.2 The application

The application understands the evaluations as groups. For each group up to six multi-channel audio stimuli plus the reference and the anchor can be evaluated. All the groups need the same number of stimuli. The playback is done in a synchronous way so, when the stimulus is changed, the new sound will continue to play from the same point in time. Each time a new group is loaded, the stimuli are presented randomly.

To run and customize a test the *testDescription.json* file has to be filled. This file provides basic information to the program, e.g.: the number of audio channels, number of stimuli, name and path to the audio files.

Following the specifications listed in Section 4.2.1 the application is designed so it can be controlled via the AKAI MIDIMIX MIDI controller, which have all the

necessary faders and buttons to perform both tests. This avoids having a computer, a monitor and an input interface inside the structure.



Figure 8: The AKAI MIDIMIX MIDI controller. Each fader and red button is linked to one stimuli, where the button changes the stimulus playback and the fader serves to mark it (**C** in Figure 9). The vertical array of buttons in the right side is used to stop the playback, play the reference stimulus and to pass from one group to the next one (**B** in Figure 9). Stickers were added during the tests to help the subjects.

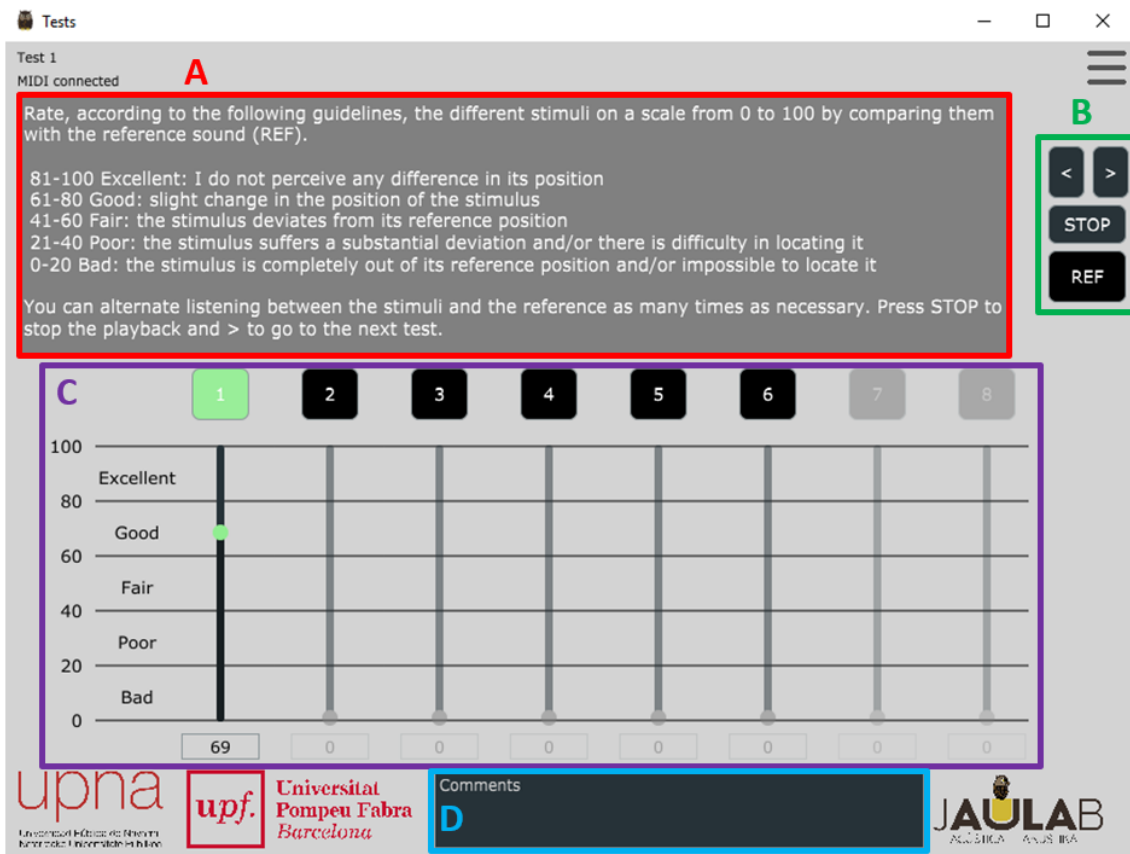
The graphical interface has been designed to be simple and user-friendly. It contains all the information and buttons to present the tests (see Figure 9). Its different zones are listed below, following a letter plus colour scheme:

- A** Help box: guidelines to assist in the grading of the stimulus.
- B** Buttons:  $<$  and  $>$  to move across groups, *STOP* button to stop the playback and the *REF* button to reproduce the reference sound.
- C** Sliders: eight sliders with their respective play button on the top. Out of the eight sliders, only the ones corresponding to an existing stimulus are active, and their value can be modified only if the stimulus is playing.
- D** Observations box: the subjects can leave comments to motivate their evaluation.

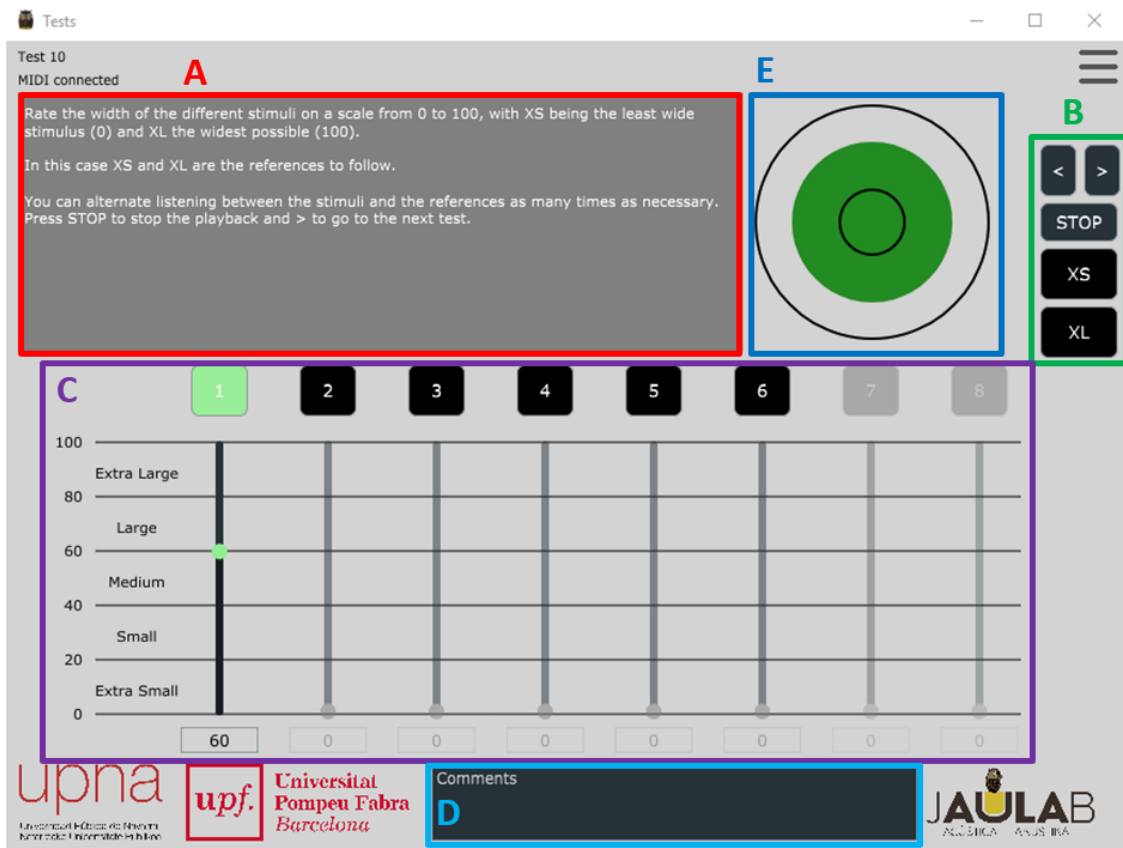


When the current test group running is a source-width test, the reference button is divided in two reference buttons for the narrowest and widest sounds (**B**) and two circles with different size (**E**) are drawn next to the help box (**A**). See Figure 9b.

All the results of the tests are saved in an xml file, where, for each group, the technologies and their marks are listed.



(a) Localization tests interface.



(b) Source width tests interface.

Figure 9: Graphical interfaces for the tests.

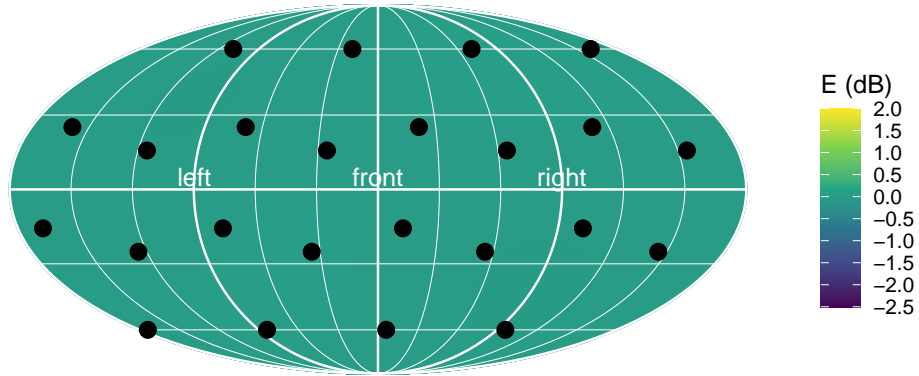
# Chapter 5

## Listening Tests

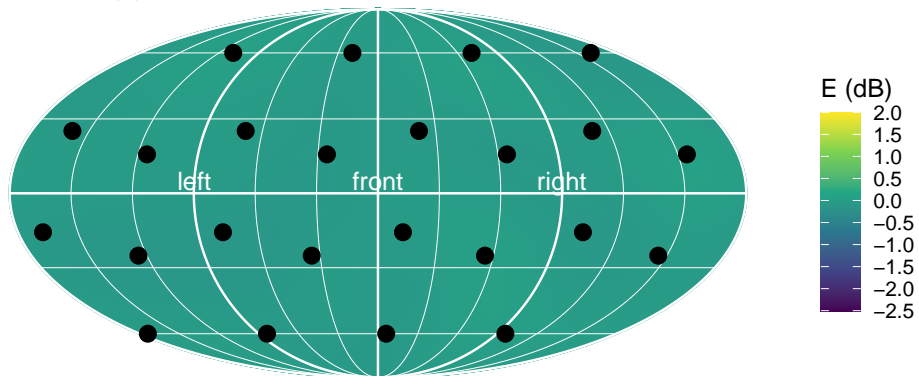
The listening tests target two different characteristics of the sound generated by Ambisonics and SWF: the positioning and the width of the sound sources. All the experiments use pink noise at  $-40$  LUFS, 48 kHz, as stimulus. The resulting sound level at the center of the sphere for all the stimuli is 65 dBA. The stimulus is encoded and decoded in Ambisonics of first and third order and SWF of zeroth and first level.

The encoding for both techniques has been done in MATLAB with two different libraries [29]: in the case of Ambisonics, the `Higher Order Ambisonics (HOA)` library [30] has been used; and for the initial interpolation to the finest mesh of 66 points in SWF, the `Vector Base Amplitude Panning` library [31]. The down-sampling of SWF from the finest mesh at level 2, to level 1 and 0 has been done using the interpolating matrices published here [32].

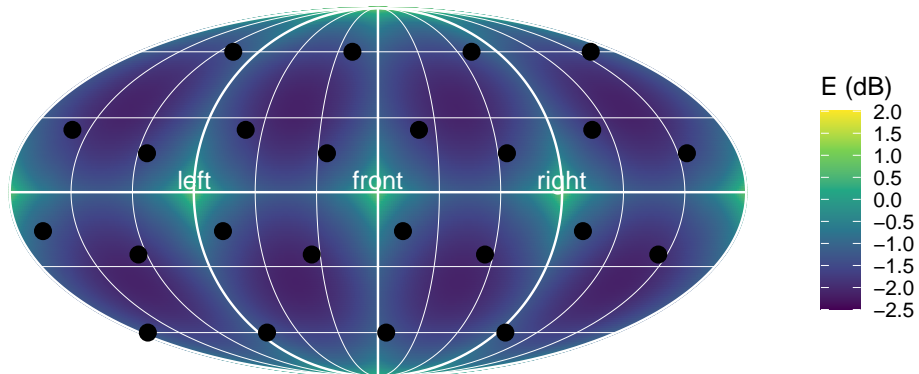
The matrices for the decoding to loudspeakers have been generated with IDHOA [6] for both Ambisonics and SWF. In Figures 10 and 11 the values for the reconstructed energy and radial intensity over the whole sphere are reported for Ambisonics (order 1 and 3) and SWF (levels 0 and 1).



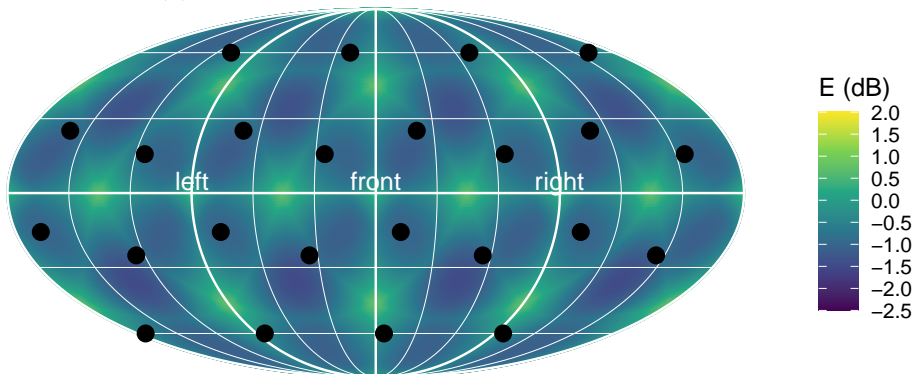
(a) Reconstruction of energy for Ambisonics at order 1.



(b) Reconstruction of energy for Ambisonics at order 3.

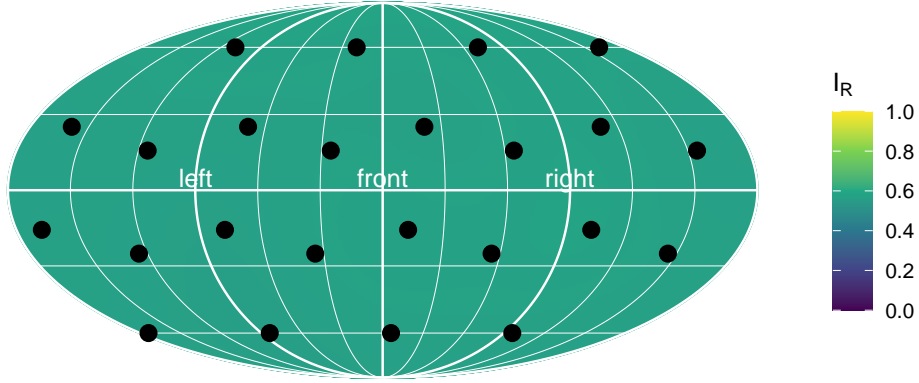


(c) Reconstruction of energy for SWF at level 0.

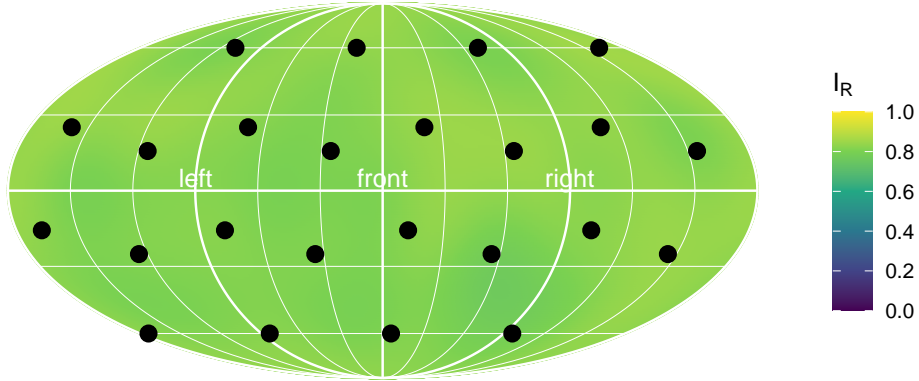


(d) Reconstruction of energy for SWF at level 1.

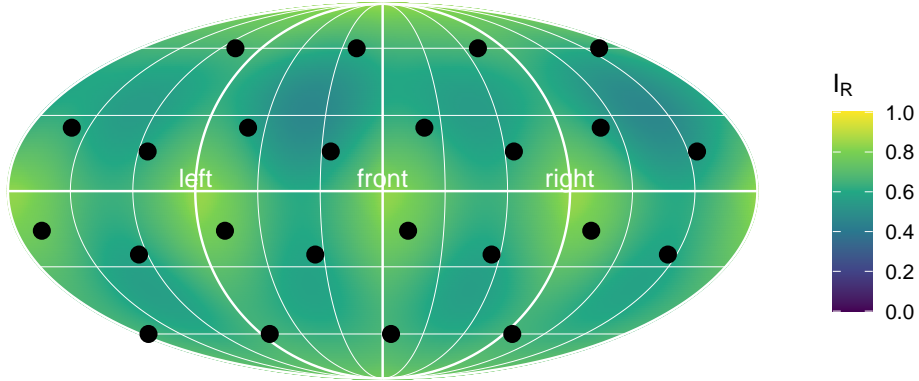
Figure 10: Comparison of energy reconstruction performances over the sphere for Ambisonics and SWF decoders generated with IDHOA. The black dots represent the loudspeakers' location.



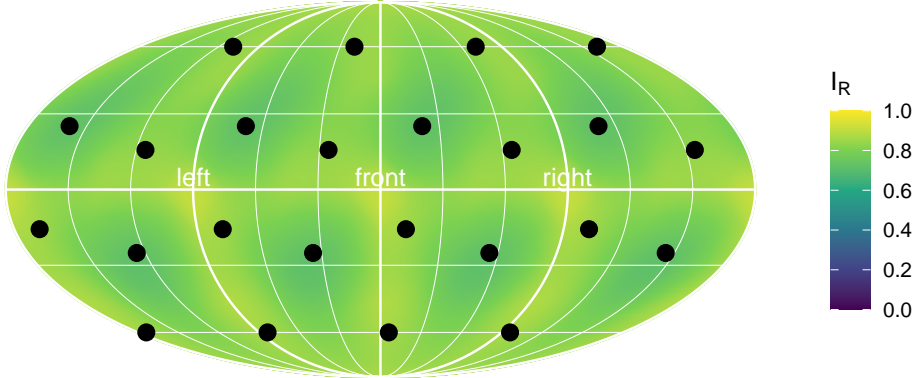
(a) Reconstruction of radial intensity for Ambisonics at order 1.



(b) Reconstruction of radial intensity for Ambisonics at order 3.



(c) Reconstruction of radial intensity for SWF at level 0.



(d) Reconstruction of radial intensity for SWF at level 1.

Figure 11: Comparison of radial intensity reconstruction performances over the sphere for Ambisonics and SWF decoders generated with IDHOA. The black dots represent the loudspeakers' location.

Table 2 exposes the values of reconstructed energy and intensity over the 24 loudspeakers of the setup, together with their mean, maximum and minimum values. Energy is reconstructed with great consistency across the whole sphere by Ambisonics, with  $\Delta E \approx 0.2$  dB. SWF is less uniform in the energy reconstruction but it is still acceptable, with  $\Delta E \approx 2.6$  dB for level 0 and  $\Delta E \approx 1.4$  dB at level 1. Focusing on the values of the reconstructed radial intensity, which correlates well with the apparent source width, SWF at level 1 and Ambisonics at order 3 should perform similarly, and SWF at level 0 perform better than Ambisonics at order 1.

Table 2: Summary of energy and radial intensity reconstruction for Ambisonics and SWF. Mean, maximum and minimum values are calculated over the positions of the loudspeakers.

Technique	Mean <sup>Max</sup> <sub>Min</sub> Energy (dB)	Mean <sup>Max</sup> <sub>Min</sub> Radial Intensity
AMB 1	$-0.02_{-0.03}^{+0.01}$	$0.57_{0.56}^{0.58}$
AMB 3	$-0.04_{-0.12}^{+0.05}$	$0.82_{0.78}^{0.85}$
SWF 0	$-0.84_{-1.97}^{+0.61}$	$0.67_{0.59}^{0.77}$
SWF 1	$-0.82_{-1.07}^{+0.31}$	$0.80_{0.72}^{0.86}$

Finally, in terms of the recruited subjects for the experiment, the listening panel (LP) consists of 18 people (4 female, 14 male) with different musical training. The age of the subjects ranges between 21 and 43. No remarkable hearing problems have been reported.

## 5.1 Localization test

The first carried test has been a MUSHRA [33] localization test. Subjects had to evaluate the accuracy of the source location with respect to a reference and an anchor, which was hidden. They had to evaluate nine different source locations, which are listed in Table 3.

For this type of tests, the stimulus sound was chosen to be a 500 milliseconds pink

noise burst windowed by a raised cosine Hanning window, to create a 50 milliseconds fade-in and fade-out. The burst is followed by 250 milliseconds of silence and it is looped. The subjects could listen to the stimuli, reference and anchor as many times as needed.

The reference is one physical loudspeaker placed in the position of the source, playing the stimulus in isolation. The anchor is built by decorrelating the mono stimulus signal and reproducing it from every loudspeaker, creating 24 decorrelated sources emitting from the loudspeakers [34] and giving the impression of a completely delocalized source.

## 5.2 Source width test

The second test aims at evaluating the apparent source width, with respect to a reference and a non-hidden anchor, which acts as another reference. To measure how the width of the source is perceived, subjects had to evaluate the stimuli in comparison to a completely decorrelated reference (XL) and a single-loudspeaker reference (XS), for the six positions listed in Table 3. In this case the stimuli are continuous pink noise. The subjects could listen to the stimuli and references as many times as needed.

To help visually with the evaluation, the width of the stimuli is represented by circles of different diameter (**E** in Figure 9b). The narrowest reference sound (XS) is compared to the smallest circle, and the widest reference sound (XL) to the biggest circle.

The labeling approach is different from the standard MUSHRA tests. To resemble the width size, labels have been chosen to be: *Extra small*, *Small*, *Medium*, *Large* and *Extra large* (from the narrowest to the largest size possible).

Table 3: Positions (azimuth and elevation) for each group and test type, in degrees.

Group	Localization test		Source width test	
	Azimuth (°)	Elevation (°)	Azimuth (°)	Elevation (°)
g1	-25.5	15.5	-25.5	15.5
g2	-158.7	25.0	-158.7	25.0
g3	6.2	-60.0	6.2	-60.0
g4	-34.4	-25.0	-77.5	-15.5
g5	-77.5	-15.5	12.5	-15.5
g6	12.5	-15.5	-19.3	60.0
g7	-124.4	-25.0		—
g8	21.3	25.0		—
g9	-19.3	60.0		—



# Chapter 6

## Results

With the panel of 18 listeners, 162 evaluations for the localization test and 108 for the source width test (18 evaluations for each source position) have been obtained. This raw data has to be filtered to ensure that the marking have been done correctly. A correct marking in each position will need the reference and anchor been over/under a threshold. For the localization test, only those evaluations where the reference is evaluated above 94 ( $\geq 95$ ) and the anchor below 41 ( $\leq 40$ ) MUSHRA points are kept. On the other hand, for the source width test, the evaluation of the XS reference had to be below 6 ( $\leq 5$ ) MUSHRA points and the XL above 94 ( $\geq 95$ ).

Table 4: Summary of the obtained results, detailing the number of evaluations after post-scanning ( $N$ ), median and interquartile ranges (IQR). Median and IQR expressed in MUSHRA points (0-100).

	Localization tests			Source width tests		
	$N$	Median	IQR	$N$	Median	IQR
Ref/XS	133	100.0	0.0	90	100.0	0.0
Anchor/XL	133	12.0	25.0	90	0.0	0.0
AMB 1	133	36.0	26.0	90	46.5	32.8
AMB 3	133	52.0	35.0	90	65.0	25.5
SWF 0	133	54.0	34.0	90	55.0	33.0
SWF 1	133	64.0	30.0	90	68.5	43.0

After cleaning the raw data, 133 evaluations for the localization and 90 for the source

width are left. This process of filtering is known as post-screening. The median and interquartile range (IQR) for each test and stimuli are shown in Table 4.

To analyse the data a Python notebook has been created. This notebook is able to read all the results files, make the post-screening process, analyse the data and generate different graphics. The Python packages leveraged for the analysis are listed below:

- Pandas open source library for the boxplots [35].
- Researchpy wrapper for the summary table (Table 4) [36].
- Stats module from the Scipy library [37] and Pingouin statistical package for the statisticals [38].

To graphically depict the results, the boxplot graphical method has been chosen. This standardized method represents a series of numerical data through their quartiles and displays its distribution based on a five number summary: the minimum ( $Q_0$ ), the sample median ( $Q_2$ ), the maximum ( $Q_4$ ) and the first and third quartiles ( $Q_1$  and  $Q_3$ ). It is very useful to see how disperse and skewed the obtained data has been.

The boxplots appearing during this chapter follow the next rules. The blue box represents the range between the 25th percentile ( $Q_1$ ) and the 75th percentile ( $Q_3$ ). The median of the data ( $Q_2$ ) is marked with a green line. The lines extending from the box are called whiskers, and they mark the maximum or minimum value of the series as long as they do not exceed 1.5 times the interquartile range (IQR). If so,  $1.5 \cdot IQR$  is added or subtracted to the box limits, respectively, to the 25th percentile and to the 75th percentile. Any data not included between the whiskers is plotted as an outlier with a small circle.

For the localization test (Figure 12), third-order Ambisonics reaches *fair* with a negative difference of 12 MUSHRA points from first-level SWF. In its zeroth level, SWF is rated similarly to third-order Ambisonics, but with a bigger spread of the

data. First-order Ambisonics is the worst rated technology, getting an evaluation of *poor* with a median of 36 MUSHRA points.

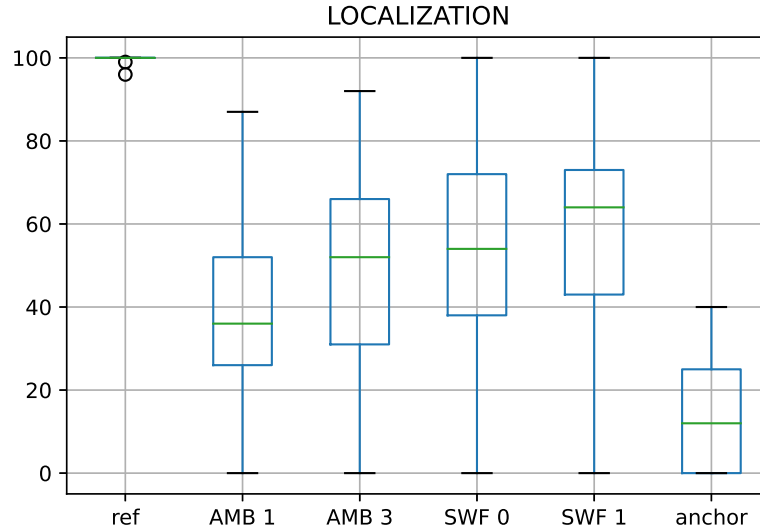


Figure 12: Results for the localization MUSHRA tests.

In the source width test (Figure 13) both third-order Ambisonics and first-level SWF are rated as *good*, with a minimal difference of 3.5 points. Both zeroth-level SWF and first-order Ambisonics get a score of *fair* with a median of 55 and 46.5 MUSHRA points respectively.

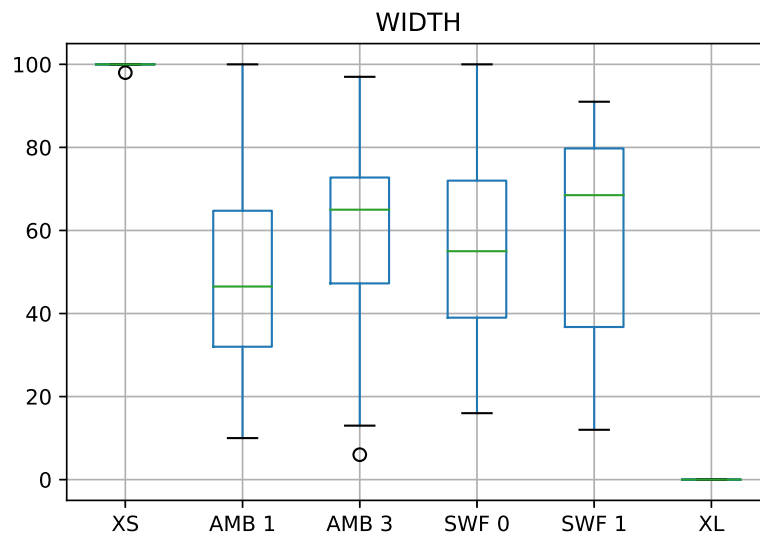
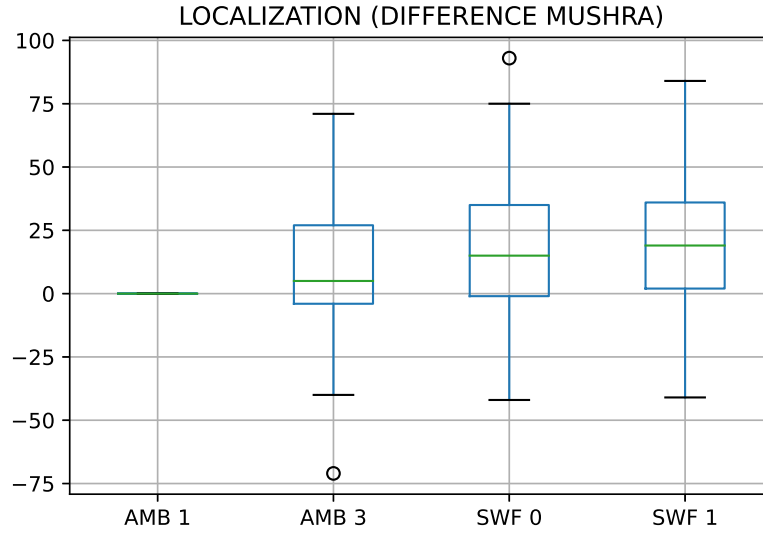
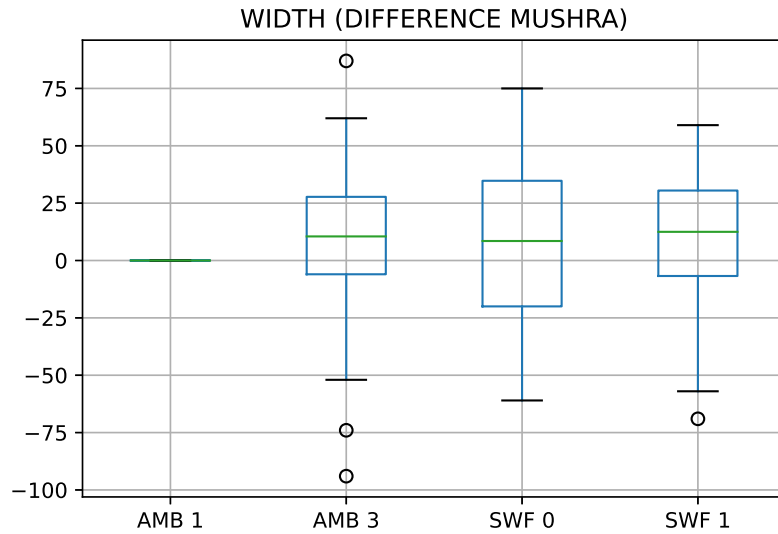


Figure 13: Results for the source width tests.

Subjects have rated first-level SWF as the best technology both for localization and source width reproduction, getting in both cases a score of *good*, with a median value of 64 in the localization and of 68.5 for the source width.



(a) Results for localization tests with difference MUSHRA.



(b) Results for the source width tests with difference MUSHRA.

Figure 14: Results for the tests displayed as difference MUSHRA, with the first order of Ambisonics as reference.

Figure 14 shows the results as a difference MUSHRA using first order Ambisonics as the reference method. This type of plot clearly highlight the relative differences

between the 4 renderings of the stimuli. For the localization test, a gap arises between SWF and third order Ambisonics, with a significant difference of 10 points with level zero and even bigger for first level. The difference between SWF and third order Ambisonics for the source width test is not meaningful enough.

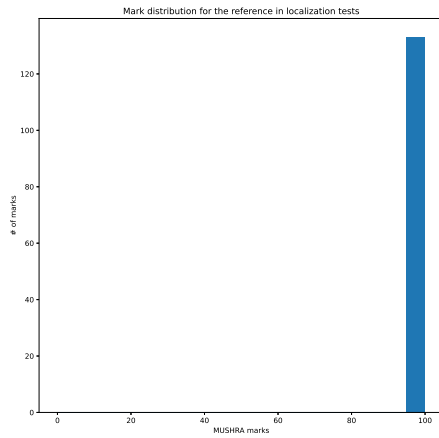
## 6.1 Statistical analysis

When obtaining experimental data, as in this case, it is important to summarize the observations and to conduct statistical tests to ensure that the listening panel is representative and, thus, that the results are conclusive. The analysis of variance (t-tests) are useful to estimate the probability that the underlying phenomena are truly different.

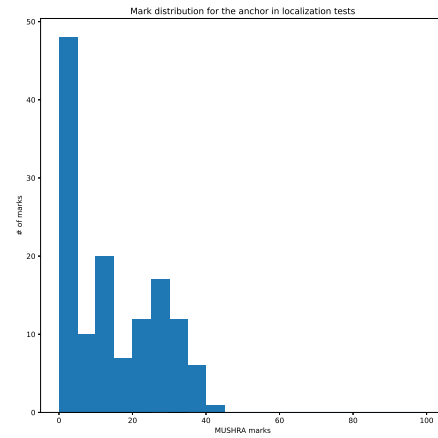
To perform a t-test a normally distributed set of samples is required. As can be seen in Figures 15 and 16 none of the data can be assumed to be normal, what prevent the use of this kind of tests. Therefore, a statistical Wilcoxon test [39] has been performed, to assess if there is significant difference in the evaluation of the different technologies.

The Wilcoxon test can be understood as the non-parametric equivalent to the paired t-test. This statistical hypothesis test is appropriate to compare two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ. It is usually used to determine whether two dependent samples were selected from populations having the same distribution.

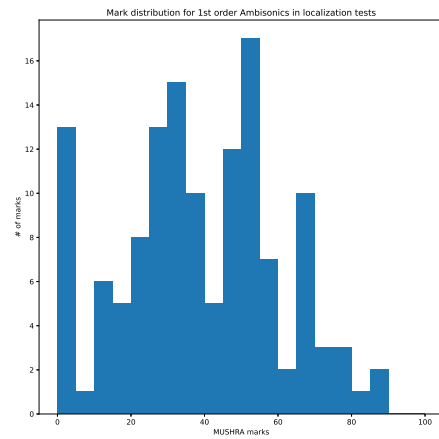
According to the results obtained, three different comparisons have been tested with this method: level 0 SWF vs first order Ambisonics, level 0 SWF vs third order Ambisonics, level 1 SWF vs third order Ambisonics. To handle the multiple comparisons problem, the Holm–Bonferroni method [40] has been applied to the results of the statistical test. This method helps on controlling the family-wise error rate (probability of appearance of Type I errors).



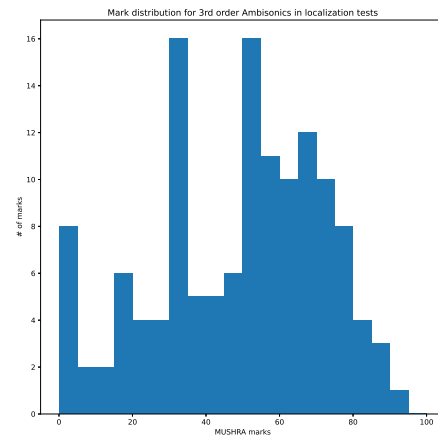
(a) Reference.



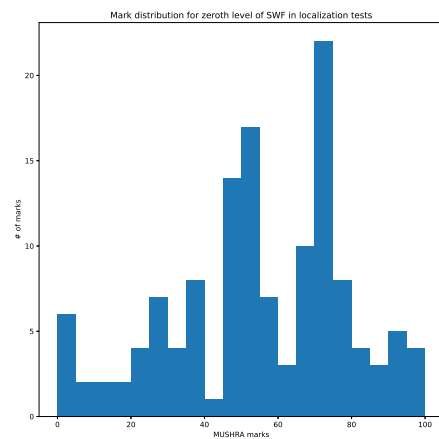
(b) Anchor.



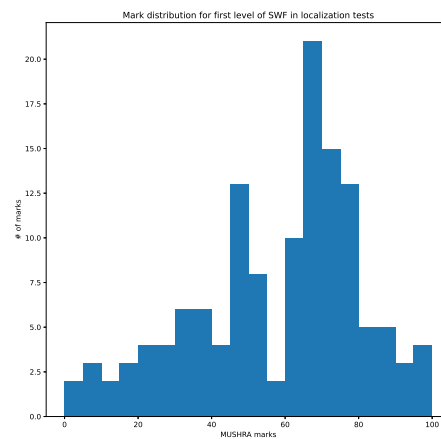
(c) First order Ambisonics.



(d) Third order Ambisonics.

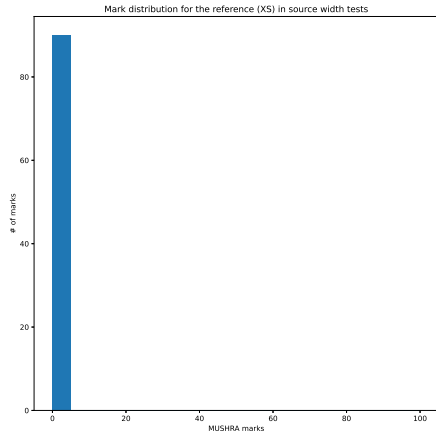


(e) Zeroth level SWF.

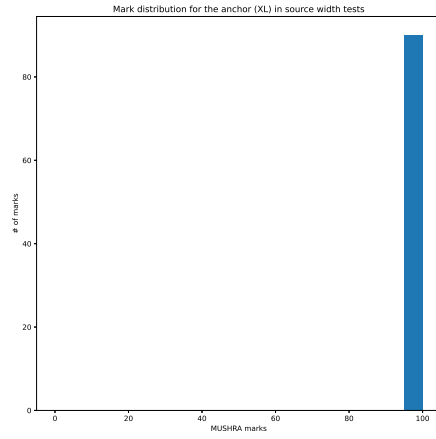


(f) First level SWF.

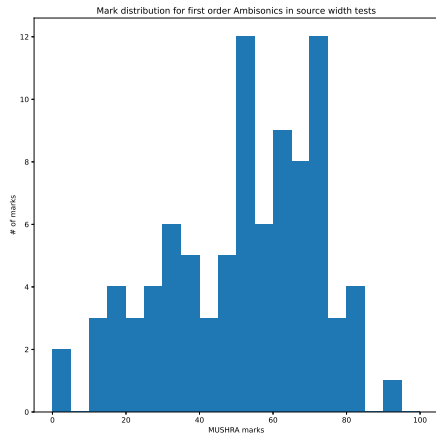
Figure 15: Mark distribution for the obtained results in the localization tests.



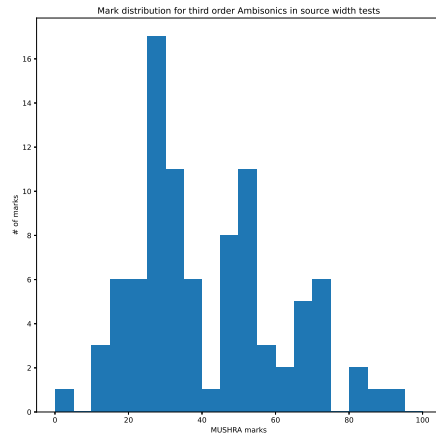
(a) Reference (XS).



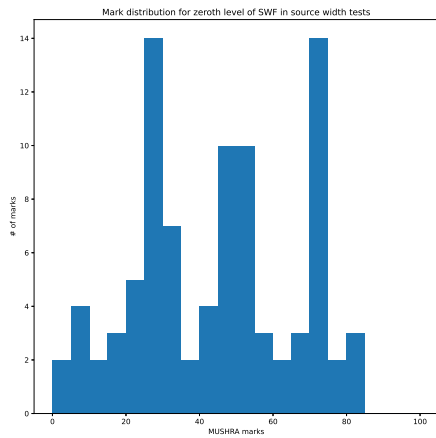
(b) Anchor (XL).



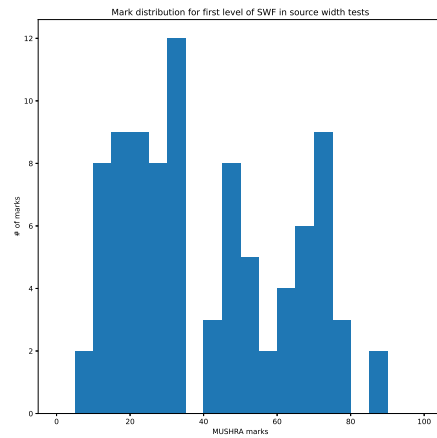
(c) First order Ambisonics.



(d) Third order Ambisonics.



(e) Zeroth level SWF.



(f) First level SWF.

Figure 16: Mark distribution for the obtained results in the source width tests.

Table 5 shows  $p$ -values and the difference in MUSHRA points (diff) for the compared technologies. Setting a significance threshold for the  $p$ -value at 0.05, the Wilcoxon test shows that the difference in favour of SWF for the three comparisons in the localization tests are statistically significant, whereas the difference in the source width tests are not, due to the wider spread of the results in the latter case.

Table 5:  $p$ -values obtained with the Wilcoxon test and median difference (diff) on MUSHRA points (SWF - AMB). An asterisk marks those differences which are statistically significant once the correction for multiple comparisons has been applied.

	Localization test		Source width test	
	$p$ -value	diff	$p$ -value	diff
SWF 0 vs. AMB 1	9.1e-10*	18.0	0.05	8.5
SWF 0 vs. AMB 3	2.2e-2*	2.0	0.38	-10.0
SWF 1 vs. AMB 3	3.5e-4*	12.0	0.99	2.5



# Chapter 7

## Discussion

### 7.1 Conclusions

A set of MUSHRA listening tests have been designed to compare and assess two spatial audio techniques, Ambisonics and SWF, in a common full-sphere setup of 24 loudspeakers. This is the first study addressing the auditory properties of the novel wavelet-based SWF format.

Subjective listening tests have shown that zeroth and first level of SWF performs significantly better than third order Ambisonics in terms of the accuracy in source localization. In terms of source width, there has been a wider variability of the results and the results are statistically inconclusive. In any case, SWF has been rated favorably in all comparisons.

It is remarkable that SWF at level 0, with only 6 channels, has been rated better in the experiments than 3rd order Ambisonics, with 16 channels. However, the points in which the two spatial audio techniques were evaluated coincided with the position of loudspeakers in the setup. Whereas the performance of Ambisonics is largely independent of the presence of a loudspeaker in the region evaluated, SWF has a small performance boost next to a loudspeaker, so SWF may have had a slight advantage in this comparison.

The process and the results have been presented to the International Conference on Immersive and 3D Audio 2021 (I3DA). The paper has been accepted and presented as poster together with a presentation video.

## 7.2 Future work

As future work it is intended to repeat a similar experiment, but considering rendering to points far from loudspeakers intervening in the rendering, and also evaluating the performance of SWF with moving sources, and, more generally, with realistic spatial audio mixes.

Repeating the experiments performing a binauralization process is also interesting to assess how both techniques would arrive to a more normal-life headphone setup. With this, creating an online MUSHRA test will also help to recruit more subjects and validate the results.

# List of Figures

1	Perceptual sweet spot area for first (light grey), third (dark grey) and fifth (black) Ambisonics orders. Reproduced from Ref. [2] . . . . .	9
2	Spherical harmonics for the first three Ambisonics orders plotted as a polar diagram. Reproduced from Ref. [2] . . . . .	10
3	FuMa and ACN channel ordering conventions up to the third order. .	10
4	Scheme of the signal decomposition, which illustrates the encoding process to wavelet space. The fine data $\mathbf{f}$ is decomposed by analysis filters $\mathbf{A}^n$ and $\mathbf{B}^n$ into the coarse $\mathbf{c}^{n-1}$ and details $\mathbf{d}^{n-1}$ signals respectively. The same operation of decomposition is repeated recursively on the coarse signal $\mathbf{c}^{n-1}$ up to the desired level. Reproduced from Ref. [5]. . . . .	14
5	Scheme of the signal reconstruction, which illustrates the decoding process from the wavelet space. The coarse and details signals at the lowest level, $\mathbf{c}^0$ and $\mathbf{d}^0$ , are upsampled via the reconstruction filters $\mathbf{P}$ and $\mathbf{Q}$ respectively. The resulting signals are summed together to obtain the next level coarse signal $\mathbf{c}^1$ . The process is repeated at each level with the contribution of the details $\mathbf{d}^\ell$ , until the original fine data $\mathbf{f}$ is recovered. Reproduced from Ref. [5]. . . . .	15
6	A look into the structure. . . . .	16

7	Frequency response for the 24 loudspeakers in the layout. The loudspeakers' frequency response is very consistent in the range between 400 Hz and 7 kHz. At low frequencies the differences in frequency response between different loudspeakers are more evident due to room modes. At frequencies higher than 10 kHz the shadowing effect of the measurement microphone itself may be observed. . . . .	18
8	The AKAI MIDIMIX MIDI controller. Each fader and red button is linked to one stimuli, where the button changes the stimulus playback and the fader serves to mark it ( <b>C</b> in Figure 9). The vertical array of buttons in the right side is used to stop the playback, play the reference stimulus and to pass from one group to the next one ( <b>B</b> in Figure 9). Stickers were added during the tests to help the subjects. .	20
9	Graphical interfaces for the tests. . . . .	22
10	Comparison of energy reconstruction performances over the sphere for Ambisonics and SWF decoders generated with IDHOA. The black dots represent the loudspeakers' location. . . . .	24
11	Comparison of radial intensity reconstruction performances over the sphere for Ambisonics and SWF decoders generated with IDHOA. The black dots represent the loudspeakers' location. . . . .	25
12	Results for the localization MUSHRA tests. . . . .	31
13	Results for the source width tests. . . . .	31
14	Results for the tests displayed as difference MUSHRA, with the first order of Ambisonics as reference. . . . .	32
15	Mark distribution for the obtained results in the localization tests. . .	34
16	Mark distribution for the obtained results in the source width tests. .	35

# List of Tables

1	Positions of the loudspeakers in the 7-design, in degrees. . . . .	17
2	Summary of energy and radial intensity reconstruction for Ambisonics and SWF. Mean, maximum and minimum values are calculated over the positions of the loudspeakers. . . . .	26
3	Positions (azimuth and elevation) for each group and test type, in degrees. . . . .	28
4	Summary of the obtained results, detailing the number of evaluations after post-scanning ( $N$ ), median and interquartile ranges (IQR). Median and IQR expressed in MUSHRA points (0-100). . . . .	29
5	$p$ -values obtained with the Wilcoxon test and median difference (diff) on MUSHRA points (SWF - AMB). An asterisk marks those differences which are statistically significant once the correction for multiple comparisons has been applied. . . . .	36

# Bibliography

- [1] Pfanzagl-Cardone, E. *Spatial Hearing*, 1–34 (Springer Vienna, Vienna, 2020).  
URL [https://doi.org/10.1007/978-3-7091-4891-4\\_1](https://doi.org/10.1007/978-3-7091-4891-4_1).
- [2] Zotter, F. & Frank, M. *Ambisonics* (Springer International Publishing, 2019).  
URL <https://www.springer.com/gp/book/9783030172060>.
- [3] Zotter, F. & Frank, M. All-round ambisonic panning and decoding. *J. Audio Eng. Soc* **60**, 807–820 (2012). URL <http://www.aes.org/e-lib/browse.cfm?elib=16554>.
- [4] Scaini, D. Wavelet-based spatial audio framework : from ambisonics to wavelets: a novel approach to spatial audio. *TDX (Tesis Doctorals en Xarxa)* (2019). URL <http://www.tdx.cat/handle/10803/668214>.
- [5] Scaini, D. & Arteaga, D. Wavelet-Based Spatial Audio Format. *J. Audio Eng. Soc* **68**, 613–627 (2020). URL <http://www.aes.org/e-lib/browse.cfm?elib=20893>.
- [6] Scaini, D. & Arteaga, D. Decoding of Higher Order Ambisonics to Irregular Periphonic Loudspeaker Arrays. In *Proceedings of the AES International Conference*, vol. 2014 (2014).
- [7] ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems (mushra). *International Telecommunication Union* **3**, 34 (2015).

- [8] ITU-T P.800.1. Mean opinion score (MOS) terminology. *International Telecommunication Union* (2016).
- [9] ITU-R BS.1116-3. Methods for the subjective assessment of small impairments in audio systems BS Series Broadcasting service ( sound ). *International Telecommunication Union* **3** (2015).
- [10] Kearney, G., Bates, E., Boland, F. & Furlong, D. A Comparative Study of the Performance of Spatialization Techniques for a Distributed Audience in a Concert Hall Environment (2007).
- [11] Frank, M. & Zotter, F. Localization Experiments Using Different 2D Ambisonics Decoders (2008).
- [12] Benjamin, E., Lee, R. & Heller, A. Localization in horizontal-only ambisonic systems (2006).
- [13] Matthias Frank. *Phantom Sources using Multiple Loudspeakers in the Horizontal Plane*. Ph.D. thesis, University of Music and Performing Arts Graz, Austria (2013). URL <https://phaidra.kug.ac.at/o:7008>.
- [14] Zotter, F. & Frank, M. Ambisonic decoding with panning-invariant loudness on small layouts (allrad2). In *Audio Engineering Society Convention 144* (2018). URL <http://www.aes.org/e-lib/browse.cfm?elib=19460>.
- [15] Zotter, F., Pomberger, H. & Noisternig, M. Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere. In *2nd Int. Symposium on Ambisonics and Spherical Acoustics*, – (Paris, France, 2010). URL <https://hal.archives-ouvertes.fr/hal-01107091>. Cote interne IRCAM: Zotter10a.
- [16] Wiggins, B., Paterson-Stephens, I., Lowndes, V. & Berry, S. The design and optimisation of surround sound decoders using heuristic methods (2003).
- [17] Wiggins, B. *An Investigation into the Real-Time Manipulation and Control of Three-Dimensional Sound Fields*. Ph.D. thesis (2004).

- [18] Moore, D. & Wakefield, J. The design and detailed analysis of first order ambisonic decoders for the itu layout (2007).
- [19] Moore, D. & Wakefield, J. Designing ambisonic decoders for improved surround sound playback in constrained listening spaces. In *Audio Engineering Society Convention 130* (2011). URL <http://www.aes.org/e-lib/browse.cfm?elib=15892>.
- [20] Cheung, K. & Tsang, P. Development of a re-configurable ambisonic decoder for irregular loudspeaker configuration. *IET Circuits, Devices Systems* **3**, 197–203 (2009). URL <https://digital-library.theiet.org/content/journals/10.1049/iet-cds.2009.0007>.
- [21] Heller, A., Benjamin, E. & Lee, R. Design of ambisonic decoders for irregular arrays of loudspeakers by non-linear optimization **2** (2010).
- [22] Heller, A., Benjamin, E. & Lee, R. A toolkit for the design of ambisonic decoders. *Linux Audio Conference* (2012).
- [23] Scaini, D. & Arteaga, D. An evaluation of the IDHOA Ambisonics decoder in irregular planar layouts. In *AES convention 138* (2015).
- [24] Jansen, M. & Ooninc, P. *Second Generation Wavelets and Applications* (Springer, London, 2005).
- [25] Sweldens, W. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis* **29**, 511–546 (1998).
- [26] Loop, C. *Smooth Subdivision Surfaces Based on Triangles*. Ph.D. thesis (1987). URL <https://www.microsoft.com/en-us/research/publication/smooth-subdivision-surfaces-based-on-triangles/>.
- [27] ROLI. Juce. <https://juce.com/> (2004). Accessed: 2021-05-15.
- [28] Eguinoa, R. mushratests (2021). URL <https://github.com/iRubec/mushraTests>.



- [29] Politis, A. *Microphone array processing for parametric spatial audio techniques*. Doctoral thesis, School of Electrical Engineering (2016). URL <http://urn.fi/URN:ISBN:978-952-60-7037-7>.
- [30] Politis, A. Higher order ambisonics (hoa) library (2015).
- [31] Politis, A. Vector base amplitude panning library (2015).
- [32] Scaini, D. Swf data (2020). URL <https://github.com/davrandom/swf/>.
- [33] ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems (mushra). *International Telecommunication Union* **3**, 34 (2015).
- [34] Kendall, G. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal* **19** (1996).
- [35] McKinney, W. *et al.* Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, vol. 445, 51–56 (Austin, TX, 2010).
- [36] C, B. Researchpy (2018). URL <https://github.com/researchpy/researchpy>.
- [37] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
- [38] Vallat, R. Pingouin: statistics in python. *Journal of Open Source Software* **3**, 1026 (2018). URL <https://doi.org/10.21105/joss.01026>.
- [39] Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin* **1**, 80–83 (1945). URL <http://www.jstor.org/stable/3001968>.
- [40] Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70 (1979). URL <http://www.jstor.org/stable/4615733>.