Sonification of large datasets generated from neuroscience in the eXperience Induction Machine

Panagiota Papachristodoulou

Master Thesis UPF / 2013

THESIS DIRECTOR

Paul F.M.J Verschure

THESIS SUPERVISOR

Alberto Betella

DEPARTAMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Universitat Pompeu Fabra Barcelona

To Eleni and Thanasis

Acknowledgments

First of all, I would like to offer my gratitude to my supervisor for his guidance and support during the entire project. Special thanks to Sebastian Mealla for his valuable advice on sonification issues. In addition, I would like to thank a number of people whose help was of great importance for the implementation of this thesis through their technical support and assistance: Enrique Martínez, Armin Duff, Alex Escuredo Chimeno, Xerxes D. Arsiwalla. Finally, I am grateful to Stelios Togias, Kostis Benardis, Frosso Makri, Ana Piñeyro, Ryszard Centarski, the CSIMers and my friends for their support and encouragement.

Abstract

Data deluge has been dubbed the problem of the continuous accumulation of huge amounts of data that is generated nowadays. The analysis of this data has become a great issue in many areas of research. One of these fields is neuroscience. For this reason an application of a 3D visualization of neuronal network models has been previously built in the eXperience Induction Machine (XIM)[1]. In this study we investigate the role of sonification to extract meaning from the connectome datasets. Our hypothesis is that sonification enhances the visualization by giving a further layer of understanding of the network's dynamics during the navigation in the XIM. We conducted an empirical evaluation to test whether sound helps the comprehension of a complex dataset and we exposed the participants to two different conditions; only visualization and both visualization and sonification. Our results revealed that sonification enhances the understanding of the network's characteristic even with the completion of different tasks.

[1] It is located at the laboratory for Synthetic, Perceptive, Emotive and Cognitive Systems (SPECS), at the UPF, Barcelona

Table of Contents

1. INTRODUCTION

1.1 Problem statement

In the last decades there is a continuous growth of data in different fields of research. This data is frequently left unused due to the lack of tools to effectively extract, analyze and understand it (Betella et al., 2012). The graphic presentation of information is a key issue in scientific creative research for the discovery of relations and patterns, as it aids the brain in handling tacit associations and relationships within wide sets of information (Garfield, 1989; Fry, 2004). One of the fields that generates extensive datasets is neuroscience. For this reason a 3D real-time visualization system has been built to graphically represent the massive connectivity of neuronal network models in the eXperience Induction Machine (XIM) (Betella et al., 2013).

Visual display has a long successful history, and it has been employed widely and commonly in most of the traditional HCIs (Fry, 2004). However, the demand of presenting large quantity of information becomes a rising challenge for visual displays. Representations of multidimensional data can become a difficult task for visual displays and may be too noisy to apprehend only with the eye (Eldridge, 2005).

The last few decades there have been made considerable development of representation of complex data through the auditory display and sonification (Pauletto & Hunt, 2009; Scaletti & Craig 1999, Hermann et al., 2001; Grond & Dall' Antonia, 2008; Hermann, 2002).

Humans are able to detect very subtle patterns in acoustic sounds, and this ability has found applications to an impressive degree in the field of music, or in medicine. Traditional tools such the stethoscope and the Geiger counter provide good examples of the use of sound for the comprehension of time-varying structural details and demonstrate that they can be more effective than a visual display (Hermann & Ritter 1999; Eldridge, 2005).

In this way, "auditory data display offers a new and very promising tool to uncover hidden structures and meaning in massive collections of data that would be difficult to scan, explore, or summarize by more conventional means" (Hermann & Ritter, 2004). The use of the sonification of complex data has been developed either in pure auditory displays or with their integration in visual displays as a supportive medium (Kaper et al., 2000; Nesbitt & Barrass, 2002; Walker & Nees, 2011; Hendrix, 1994).

The present thesis project deals with the sonification of a complex network such as the connectome (Hagmann et al., 2008) and is conducted under the framework of the CEEDS[2] project. CEEDS is a European FP7 projects that contributes to addressing the question of how we can advance our understanding

[2] The Collective Experience of Empathic Data Systems Project is founded by the European Union under the 7th Framework Programme, ICT-FET scheme (http://ceeds-project.eu/).

of the world and the data we extract from it, by placing human experience in the center of solution. The CEEDS project develops and deploys new methods to experience and analyze complex high-dimensional data sets by combining mixed reality, pervasive computing, ambient intelligence and interface technologies.

The sonification of the connectome network aims to help the user understand better the relations of the characteristics of the network and discern the changes of their values occurring during the navigation in the 3D visualization in the XIM. The hypothesis under investigation is that sonification will enhance the estimation of the values of the characteristics, make easier and more intuitive the understanding of the changes with the goal to detect differences in different regions of the network.

2. STATE OF THE ART

2.1 The connectome network

The data used for this project come from the study of the connectome network by Hagmann et al. (2008), which is publically available for further research[3]. For this reason it seems appropriate to make an introduction into the nature of this dataset.

The connectome is the complete description of the structural connectivity (the physical wiring) of an organism's nervous system (Sporns et al., 2005). The name of connectome has been given simultaneously and independently by Sporns and Hagmann, giving birth also to the field of science dealing with the assembly, mapping and analysis of data on neural connectoms, named as connectomics.

In the human brain, the significance of the connectome stems from the realization that the structure (connectivity) and function of the human brain are intricately linked, through multiple levels and modes of brain connectivity. The connectome naturally places strong constraints on which neurons or neural populations can interact, or how strong or direct their interactions are.
Structure-function relationships in the brain are unlikely to reduce to simple one-to-one mappings. Despite such complex and variable structure-function mappings, the connectome is an indispensable basis for the mechanistic interpretation of dynamic brain data, from single-cell recordings to functional neuroimaging.

The connectome is the fundamental basis for the mechanistic interpretation of dynamic brain data, from single-cell recordings to functional neuroimaging. Hagmann et al. (2008) constructed a connection matrix from fiber densities measured between homogeneously distributed and equal-sized regions of interest (ROIs) numbering between 500 and 4000. A quantitative analysis of connection matrices obtained for approximately 1000 ROIs and approximately 50,000 fiber pathways from two subjects demonstrated an exponential (one-scale) degree distribution as well as robust small-world attributes for the network. The data sets were derived from diffusion spectrum imaging (DSI).

2.2 Auditory Display and Sonification Background

An auditory display can be broadly defined as any display that uses sound to communicate information. Sonification has been defined as a subtype of auditory displays that use non-speech audio to represent

[3] http://www.cmtk.org

information (Walker & Nees, 2011). Hermann (2008) suggests that the term of auditory display should also encompass the technical system used to create sound waves, or more general: all possible transmissions which finally lead to audible perceptions for the user. Sonification is thereby an integral component within an auditory display system, which addresses the actual rendering of sound signals, which in turn depend on the data and optional interactions.

Sonification is still a relatively new field. Its definition was formally introduced quite recently, in the Sonification Report: Status of the Field and Research Agenda (Kramer et al., 1999) and can be considered as the most agreed definition: "Sonification is the use of non-speech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation."

Another definition proposed later by Hermann (2008) is the following. Sonification is "a technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering) and may be called sonification, if and only if:
(C1) The sound reflects objective properties or relations in the input data.
(C2) The transformation is systematic. This means that there is a precise definition provided of how the data (and optional interactions) cause the sound to change.
(C3) The sonification is reproducible: given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical.
(C4) The system can intentionally be used with different data, and also be used in repetition with the same data".

The last few decades sonification has been broadly used as a tool for the understanding of complex data. It is suggested that auditory display offers a new and very promising tool to uncover hidden structures and meaning in massive collections of data that would be difficult to scan, explore, or summarize by more conventional means (Hermann & Ritter, 1994). One of its aspects is to aid and enhance the currently much wider established techniques of data visualization for the purpose of interactive, or exploratory data analysis. A major reason for this is that the specific properties of sound perception as compared to visual perception make auditory data displays highly suited to offer an additional route to meaning in data that is both synergistic and complementary to visualization. Particular strengths in this regard are: 1) the capability of our auditory system to process several streams of information in parallel 2) to offer a high temporal resolution 3) its high sensitivity for structured motion, in particular, rhythm and 4) its ability to function well even in noisy contexts.

As aforementioned, sonification may have been formally formulated in the late '90s however its history can be tracked back many years ago. It could be suggested that sonification started to get established with the use of alarm signals with the broad use of computers and the appearance of the first user interfaces, in order to focus attention or announce the completion of a task. These would mainly aim in giving cues to the user, such as beeps etc. In the following sections, we will present a brief history background of the techniques developed are described.

2.2.1 Sonification Techniques

a) Audification

Audification was one of the first attempts of representing big quantities of data. Its birth can be found back in 1819 with the invention of the stethoscope and after that we have even more examples of audifying physical data (Laennec, 1830). A rather old example of an auditory display is the well-known Geiger counter, which provides a direct auditory display of the number of registered ions per time caught by the electrodes of the Geiger device. It is used to measure the radioactivity and allows the listener to infer quantitative information. (Hermann & Ritter, 1999; Dombois & Eckel, 2011).

The method of audification translates the data to amplitude values of the waveform and it is applicable if the data itself is a time series, e.g. data from a dynamic system like neural networks or seismic data. Audification has been broadly used in medicine with EEG signals (Dombois & Eckel, 2011; Nadwana, 2012), in seismology (Hayward, 1994; Dombois, 2002) or as a diagnostic tool for heliospheric data analysis (Alexander et al., 2011).

b) Auditory Icons

This sonification technique uses sounds that are used in a metaphorical sense; they are "everyday sounds meant to convey information about computer events by analogy with everyday events" (Gaver, 1989), so that the effort to learn the display is decreased and is mostly applied in GUI. Auditory Icons appeared with the desktop user interfaces and can be considered as the counterparts of visual icons in desktop metaphor. Gaver's work (1993) on Auditory Icons was inspired by Gibson's (1979) ecological theory of visual perception, adapted and applied for the design of auditory user interfaces. Examples of the work on Auditory Icons can be found in the SonicFinder (Gaver, 1989) an auditory interface developed at Apple computer, Soundshark, an application where sounds were used to represent the activity of ongoing processes even when not within a visible window or view on the screen. Other applications of auditory icons can be broadly found in mobile devices (Brazil & Fernstrom, 2011). However, this auditory display could be considered rather unsuited for presentation of general types of data.

c) Earcons

Earcons have been proposed by Blattner et. al. (1989) for navigation/orientation in data trees (like directory trees) with the intention to communicate more complex messages. Earcons were firstly defined as "non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction". This definition was later refined by Brewster, as follows: Earcons are "abstract, synthetic tones that can be used in structured combinations to create auditory messages" (McGookin & Brewster, 2011). Likewise as Auditory Icons, Earcons have been used broadly in mobile devices. The difference with the Auditory Icons is that there is no assumption of an existing relationship between the sound and the information it represents. Earcons are simple tonal combinations or arbitrary acoustic patterns whose meaning must be learned by the user, and which can be combined to build non-verbal messages of a higher complexity.

d) Parameter Mapping Sonification

This is one of the most common used techniques for the auditory representation of large data sets. Parameter Mapping Sonification involves the association of auditory parameters with data for the purposes of display. Given the inherent multidimensionality of sound, Parameter Mapping Sonification is considered to be well suited for sonification of multivariate data. For each data point one or more tones are generated where the parameters of the events, e.g. timestamp, duration, volume, pitch, envelope characteristics, brightness, etc., are controlled by the data vector components. The result can be called a multi-dimensional "sonic scatter plot". A good example used for understanding the Parameter Mapping Sonification is the teapot as described by Grond and Berger (2011). Consider the simple case of a whistling teakettle: the kettle produces a particular sound as the water inside approaches its boiling point. It could be said that such a kettle creates much more sound than necessary considering that it merely represents a binary signal (boiling or not boiling). It would be simpler to use an auditory signal might be achieved by monitoring the output of a thermometer measuring the water temperature in the tea kettle, and mapping the numeric output to a sound synthesis parameter. A simple mapping, for example, would link temperature to frequency, pitch or, perhaps a more obvious and explicit auditory signal. Rather than simply hearing when the target temperature is reached one might wish to listen to the continuous change of the rising water temperature or, perhaps, to hear selective temperatures at various times during the heating process. However, it is important to note the fact that the sound of a whistling teakettle is a broadly understood signal, which carries a positive emotional connotation for some.

Additionally, the progression from noise to unstable frequency to relatively stable frequency can be said to have a musical quality. Thus, Parameter Mapping Sonification may offer something in the way of efficiency, but there are other important considerations, such as intuitive, emotive and aesthetic issues. Previous work on Parameter Mapping Sonification has been implemented with meteorological data, such in Polli's study (2005) or in the Hyperspectral, a diagnostic tool for probing in colon cells where a vocal tract model in which particular data states were anchored to specific phoneme sounds (Cassidy et al., 2004) and the Sonification Sandbox for the sonification of auditory graphs(Walker & Cothran, 2003).

e) Model Based Sonification

Model Based sonification (MBS) can be considered the most recent technique for auditory displays and has been proposed by Hermann (2002). The basic characteristic of the MBS is the concept of interaction. Takes as a paradigm our real life interaction with the environment to apply it in the auditory display. As in nature, normally passive systems are silent and it needs excitation in order to transmit sound.

Model-Based Sonification is a sonification technique that takes a particular look at how acoustic responses are generated in response to the user's actions, and offers a framework to govern how these insights can be carried over to data sonification basis is the imagination of a virtual data "material" for the development of the sonification. Designing a sonification model consists in a "material design" in a data space. The material structure is not only determined by the setup of the elements, but also given by the interactions between the elements. A kind of "virtual physics" must be defined, that permits a vibrational process analogous as in real sounding materials. Thus the data more or less directly becomes the sounding instrument, which is examined, excited, or played by the listener (Hermann & Ritter, 1999).

For example, data points could be conceived as planets and a gravitational force defined. Particles could then be introduced into the data space to probe the gravitational potential at various points, from which the structure of the data set as a whole could be inferred. This approach has proved successful for several data pre-processing tasks such as analyzing clusters in vectorial data, and exploring the separability of a vectorial data set prior to a classification task. MBS has been broadly used for high dimensional data (Hermann & Ritter 2005; Kolbe et al., 2010).

2.3 Issues on Sonification

Sonification is a combination of different research disciplines such as Psychoacoustics, Perceptual Research, Sound Engineering, Data Mining and much more. Taking into account the interdisciplinary character of the field there are various issues that have to be taken into consideration for a successful sonification design.

2.3.1 Task

One of the main and most important issues in sonification is the task that the user will need to complete. In order to create a new sonification model, the first question should be what is the main analysis task, or what type of pattern or structure should become apparent from listening. Taking a task-centered view helps the designer to focus on the relevant features. For example, assume that the goal was to hear whether the data set contains outliers or try to identify patterns in data structure. In each case the sonification design would be very different. Always the designer needs to consider how sonification can best help the listener in order to perform successfully her or his role in the system (Walker & Nees 2011).

2.3.2 Mappings

In sonification, data mapping is the process that determines how conceptual information is translated into auditory displays. It consists of three key aspects – the selection of sound dimension, the choice of polarity and the determination of scaling. The type of data mapping used to sonify data has direct impact on the listener perception (Grond & Berger, 2011).

"Typically, when information to be sonified is of multi-dimensional, two or more variables in the sound dimension are to be used to represent the data. This would further complicate the mapping process due to the possible interactions between different sound dimensions, in which change in one dimension affects the perception of the other. For example, Neuhoff et al. (1999) found that changes in pitch can influence how listener estimate changes in loudness, and vice versa. However, the interaction can also be applied favorably in auditory display" (Shou, 2012).

2.3.3. Human auditory perception

a) Working Memory

Walker and Mauney (2004) completed a specific experiment to study effect of individual differences on comprehension of sonification. Subject cognitive ability, including working memory ("the system which actively holds information in the mind and to make it available for further information processing") the experimental results suggested that individual working memory capacity and gender seemed to have substantial influence on comprehension of sonified data, although the test results were not completely consistent.

b) Training

Unlike visual display, where applications are pervasive and the user cognition is well established, auditory display is relatively unfamiliar to most of the users. Training is identified as one of the factors that can benefit novice users of auditory display. Recent researches focused on investigating the effect of different training methods, mainly divided as conceptual training and perceptual training. Classic perceptual training methods included the use of prompting and feedback. With prompting, a cue of correct response to a stimulus is provided before or during the presentation of the stimulus. With feedback, the correct answer is revealed after user makes a response to a stimulus (Bonebrigth & Flowers, 2011).

2.4 Sonification in Multimodal Displays

Until now we discussed about the sonification and its function as pure auditory display. However, this thesis deals with the integration of the sonification in a visual display. Thus, it would be instructive to review previous studies implemented on the comparison of visual and auditory displays and with their integration in multimodal displays, and expose the findings on the interaction of the two modalities (visual and auditory) and the effectiveness that sonification may have in the representation of complex data, as the data coming of neuroscience that are studied in this thesis.

The auditory information channel can also be used as a complementary input to the visual modality. Audition helps to direct out eyes and can therefore improve our response time to visual stimuli. For example, previous studies have found that auditory cues in addition to visual cues help improve human performance in target

search tasks. Furthermore, the auditory channel is very sensitive to changes in the acoustic signal over time (Bregman, 1994).
In many applications of sonification, it is reasonable to assume that the human listener will likely have other auditory and/or visual tasks to perform in addition to working with the sonification. Surprisingly few studies to date have considered how the addition of a secondary task affects performance with sonifications. The few available studies are promising.

Mezrich et al. (1984) showed that a dynamic sonic and visual representation could assist people in analyzing en eigth-variable economic indicator that a sonic representation may be a more effective aid than a visual representation for identifying and remembering periodic patterns, and that the strong associative memories evoked by music may indicate that the way in which we remember sonic patterns differs from the way we remember visual patterns. Auditory displays, it could be suggested, can offer more than just a companion or an enhancement to visual displays.

One of the most significant contributions to the field came from Sara Bly (1982). Her doctoral thesis was focused on the classification of non-ordered multivariate data sets. In a data set with *n* dimensions, each data point was represented by an audio event in which *n* parameters were controlled by the data. Possible parameters were loudness, pitch, duration, timbre, attack time, and waveform. Bly used a multivariate data set, involving the classification of flower species using four measurements per plant. Using sound, most study subjects were able to correctly classify most of the plants. In the same paper, a logarithmic data set was presented, and the logarithmic relationship between frequency and pitch was used to represent it. The exponential variable of earthquake magnitude was encoded in pure frequency and also in loudness and duration. The result was a positive indication that significant features of seismic data could be represented through sound. Bly conducted formal experiments using multivariate data, which were presented using sound only, graphics only and bimodal displays. Other variables were training methods and the data-to sound mappings. Subjects were asked to classifying a test sample as belonging to one of two possible sets. The results indicated that auditory display was as effective as visual display, and that the combined display outperformed both.

Scaletti and Craig (1991) developed a series of sonifications to compliment visualizations to assist researchers in analyzing and interpreting complex data. The sonifications aimed to enhance the understanding of visualizations of forestry, air pollution and blood diagnostic tools produced in the NCSA University. They found that data-driven sound tracks increased the bandwidth of the scientific visualizations, providing supportive or additional information.

Janata and Childs (2004) conducted an experiment of monitoring task where the participants had to detect changes in the variation of the values of the stock market. The task of the participants consisted in pressing a key for positive changes and another for negative ones. They found that auditory information increases the proportion of correct detections and the helpfulness of sound was even more pronounced when a secondary number-matching task was added.

Peres and Lane (2005) showed in their study that using integral dimensions of sound (pitch and loudness, where interaction between dimensions exists) in data mapping improved listener performance in an auditory monitoring task, in which subjects were asked to determine the status of box plots (on target, off target and skewed) based on sonified data and provided their response visually through buttons. Whereas using separate dimensions (pitch and tempo) in data mapping showed no differences, compared with when only a single auditory dimension was used". They found that while the addition of a visual monitoring task to an auditory monitoring task initially harmed performance of the auditory task, performance soon (i.e., after around 25 dual task trials) returned to pre-dual task levels.

Bonebright and Nees (2009) presented sounds that required a manual response approximately every 6 seconds while participants listened to a passage for verbal comprehension read aloud. The sound used included five types of earcons and also brief speech sounds, and the researchers predicted that speech sounds would interfere most with spoken passage comprehension. Surprisingly, however, only one condition—featuring particularly poorly designed earcons that used a continuous pitch-change mapping—significantly interfered with passage comprehension compared to a control condition involving listening only without the concurrent sound task. Although speech sounds and the spoken passage presumably taxed the same verbal working memory resources, and all stimuli were concurrently delivered to the ears, there was little dual-task effect, presumably because the sound task was not especially hard for participants.

Chang et al. (2010) implemented a sonification for a visual attention task. Although the aim of the study was not necessarily to discover information in the auditory displays that cannot be perceived in the visualizations, the system was considered to provide an engaging and accessible means to explore neural data and extract the main effects in each experiment. The sonified output enhanced and complemented the visualization, providing a multi-sensory means of experiencing and exploring the data.

Lokki and Grohn (2005) tested the navigation in a virtual environment with auditory cues. The results showed that 3D navigation in a virtual environment is possible with auditory cues alone. However, the fastest and most accurate navigation is obtained when both auditory and visual cues are available.

Kaper et al. (2000) implemented a sonification for the exploration and analysis of complex data sets in scientific computing. They suggest that the combination of visual images and sounds provides indeed an extremely powerful tool for uncovering complicated structures. Sometimes, the sounds reveal features that are hidden to the eye; at other times, the visual images illuminate features that are not easily detectable in the sound. The two modes of perception reinforce each other, and both improve with practice.

From the few studies that exist on multimodal displays, it is suggested that audio generally enhances the visual representation of data, referring to the complementary or redundant nature of the use of sound. In addition, there are studies that suggest even the superiority of sonification of data over visualization.

3. METHODS

In this project, based on the data available for the study of the connectome, it was decided to measure whether the addition of sound can enhance the understanding of changes of the values and dynamics of its characteristics during the navigation through the network.

3.1 Data exploration

The neuroscience application represents the connectome networks. For this thesis, we adopted the dataset taken from the study of Hagmann et al. (2008). In particular we used the dataset of the subject B that contains 998 regions of interest (ROIs) and 28000 unidirectional connections. The dataset is stored in graphml format[4] and is publically available for further research. A visual representation of this data has been implemented in the eXperience Induction Machine for the purposes of the CEEDS project in Unity[5]. Specifically, the connectome network that is graphically represented in the XIM consists of 66 anatomical subregions in the left and right hemisphere of the brain (Betella et al., 2013). Each one of the regions presents a certain number of nodes, connections and average strength. Strength refers to the extent a region is connected to the rest of the network. The following figure shows the latest graphical representation of the connectome network built in Unity.

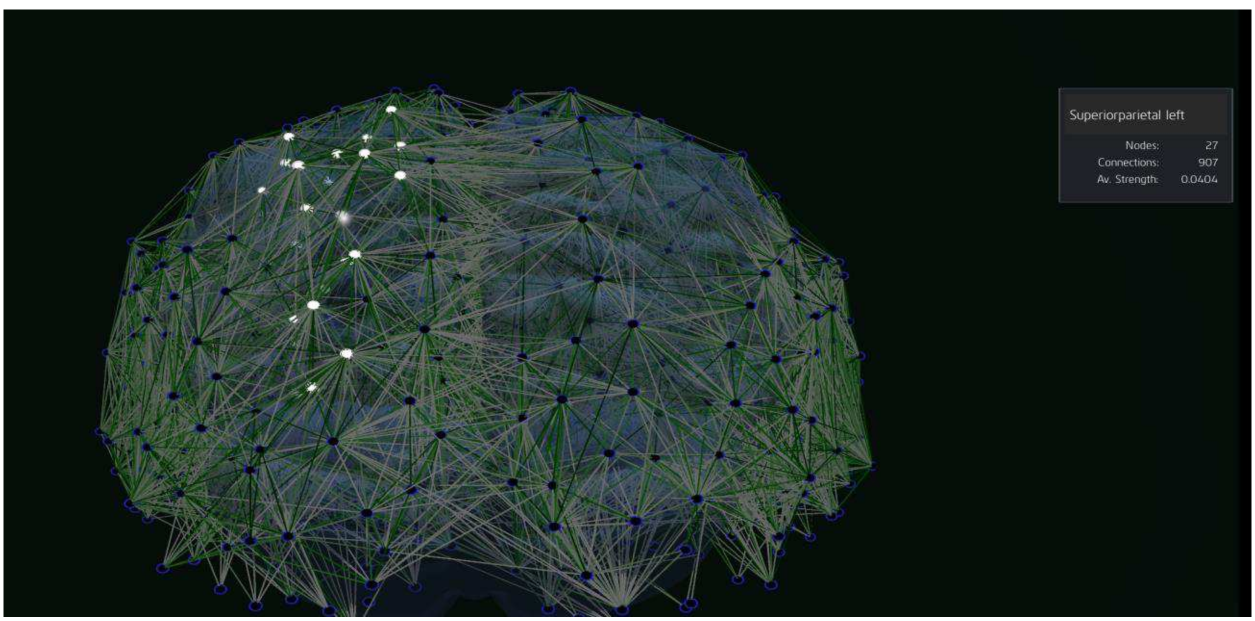


Figure 1. Screenshot of the neuroscience application with the GUI showing the values of the parameters for one of the regions (Superiorparietal Left).

[4] data source http://www.cmtk.org/datasets/homo_sapiens_01.cff
[5] http://unity3d.com

The nodes are represented graphically as spheres (Fig. 1) and are highlighted when a certain region is selected. The connections are the links between the nodes and are represented as tubes that connect the nodes. The average strength is represented with the gradient color on the connections, which varies from white for low av. strength to very dark green for the high av. strength. The application includes a GUI where the values of each region selected are depicted.

Since the sonification is based on the changes of the parameters of the characteristics of the connectome network it was necessary to understand the structure of this data. Before the design of the sonification system a data exploration phase was necessary.

A preliminary examination of the data gave the following distribution of the characteristics.

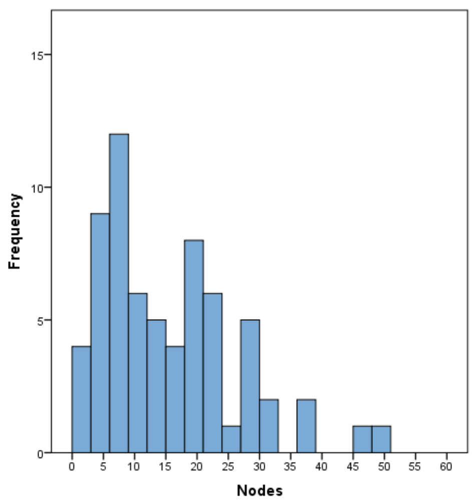


Figure 2. Histogram representing the nodes distribution in the network.

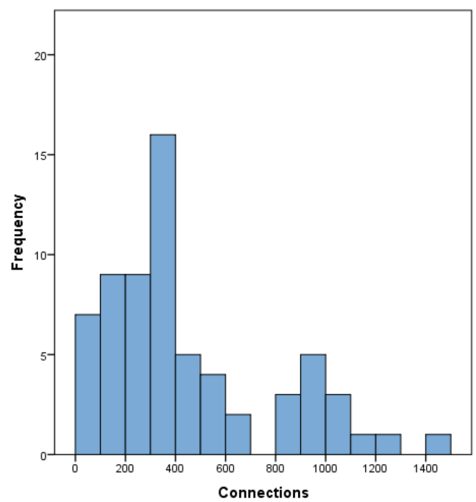


Figure 3. Histogram representing the connections distribution in the network.

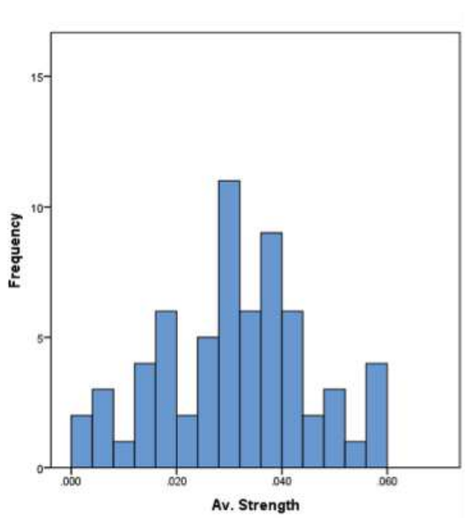


Figure 4. Histogram representing the av. strength of the regions of the network.

The histograms above (Fig. 2,3,4) show that the data presented a varied distribution depending on the characteristic under study. This is important because it is a factor to take into consideration both for the design of the sonification and the design of the experiment, as it will be described in the further sections. Regarding the sonification, the distribution of the characteristics showed that higher resolution was needed in the following cases. In the case of the nodes, as it can be seen from the histogram, most of the nodes are concentrated in the range 0-15, while above the value of 30 there were very few regions. In the case of the connections most of the regions of the network presented values between 0-400, while in the range of 600-800 and above 1200 there were very few regions. Finally, the values of the average strength present a normal distribution with most of the values being concentrated around 0.03 and 0.04.
The following table summarizes the distribution of the parameters in the network and the values that were important for the design of the sonification and the experimental design.

Table 1. Table shows the distribution of the characteristics in the network and the ranges of high and low concentration of values.

Characteristic	**Mean**	**Standard Deviation (*SD*)**	**Ranges of high concentration of values**	**Ranges of low concentration of values**
Number of connections	15.11	± *10.88*	0-15	>30
Number of nodes	438.64	± *343.91*	0-400	600-800 and >1200
Av. Strength	0.041	± *0.078*	0.03 -0.04	<0.03 and >0.04

3.2 Sonification method

Parameter Mapping Sonification is probably the most common used technique for large datasets. The data are mapped into parameters of separate sound events. For each data record an acoustic event is created whose properties are driven by the data values. This technique allows the sound events to superimpose and offers flexibility (Grond & Berger, 2011). For these reasons and due to the nature of the data, Parameter Mapping Sonification has been considered to be the most appropriate for the auditory representation of the network's data. The parameters to be sonified were the four characteristics of the network coming from the neuroscience application: a) number of nodes, b) number of connections, c) average strength of each region and d) the location referring to which hemisphere the region was found.

Two distinct sound sources were chosen whose parameters were mapped to the network's characteristics: a) a grain sound of 16 ms and b) an ambient sound. The sound parameters used for the sonification were: a) the repetition rate of the grain sound, b) the pitch of the ambient sound, c) the loudness of the grain sound and d) panning.

The decision of using two different sounds was essential for the sonification design in order to avoid interaction between the frequency and the amplitude. As it has been suggested in previous studies the pitch of a sound is highly influenced by its loudness (Flowers, 2005). Therefore, changing both of the parameters of a single sound source, could lead to unperceived and confusing differences.

As a solution to this problem it was decided to use two distinct sounds; one representing the number of nodes and the average strength of the regions and the other the number of connections. More specifically, the repetition rate of the grain sound was mapped to the number of nodes of each region of the network while the amplitude to the average strength of each region. The pitch of the second sound was mapped to the number of connections. In the following table the sound mappings are summarized.

Table 2. Table shows the mappings between the networks's characteristics and the sound parameters.

Connectome characteristics	**Mappings**		**Functions**
	Sound source	**Sound parameter**	
Number of connections	Sound grain of 16 ms	Rate of repetition	Linear
Number of Nodes	Ambient sound at 380 Hz enhanced by pure sine wave at the same frequency	Pitch	Linear
Av. Strength	Sound grain of 16 ms	Loudness	Linear

The choice of the sounds is also essential for the effective functioning of the system. On the one hand the sonification needs to give information about the data on the other hand has to be pleasant to the ear (Barass & Vickers 2011). In addition, there is interdependence between these two sounds, which needs to be taken into account. Thus, the two sounds functioned as two different streams that would not be confused but would be easily distinguished even if the first was listened in very low amplitude. Further details on the quality of the sound are given in the following sections regarding the sound mappings of each one of the parameters.

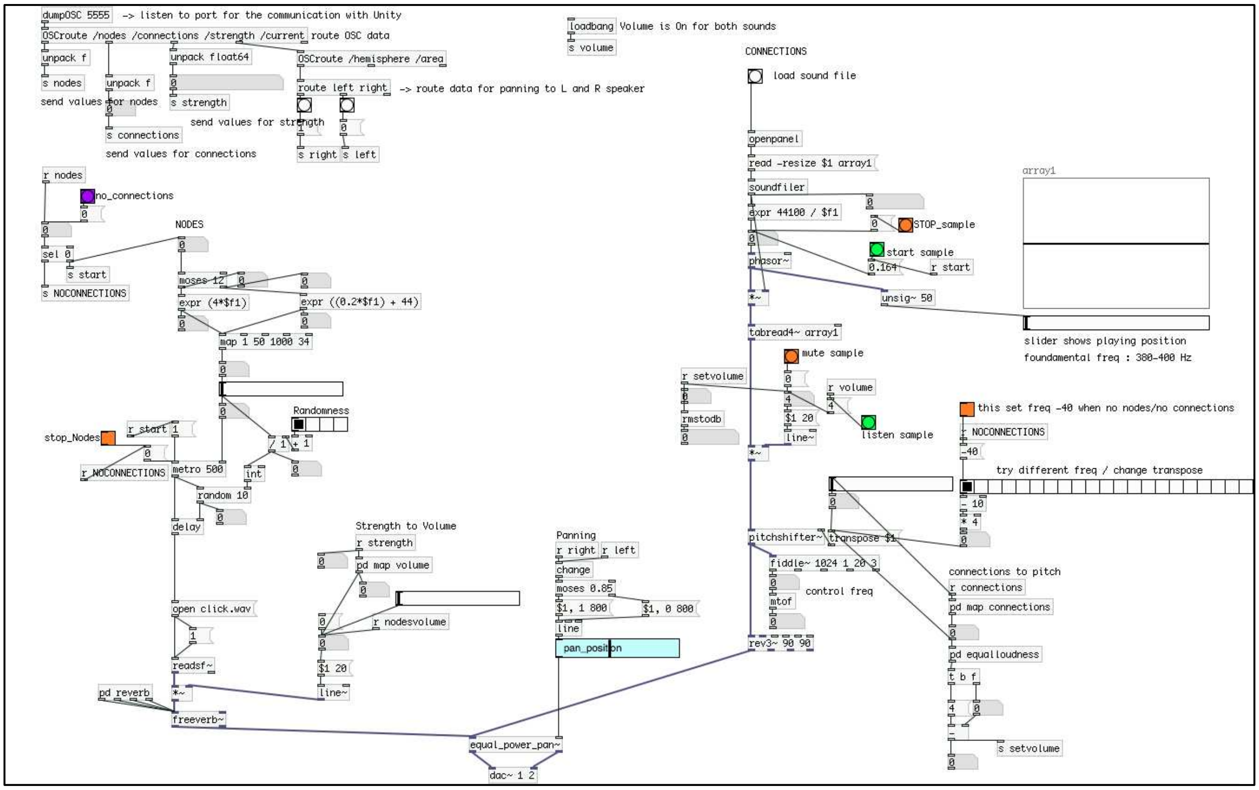


Figure 5. Overview of the sonification engine in Pure Data.

3.2.1 Data Communication between systems

The sonification engine was built in Pure Data[6]. Pure Data is a 'real-time graphical programming environment for audio, video and graphical processing' and has been previously used in various sonification projects while it has been extended by a diverse group of developers, who have contributed additional libraries and functionality. Additionally, Pure Data has been chosen because of its open-source nature and the supportive online community of users.

The communication of the data between the Unity environment, in which the neuroscience application has been built, and Pure Data was established using OpenSoundControl (OSC) messages. In the following figure the subsystem of the sonification engine is shown, receiving the data from Unity.

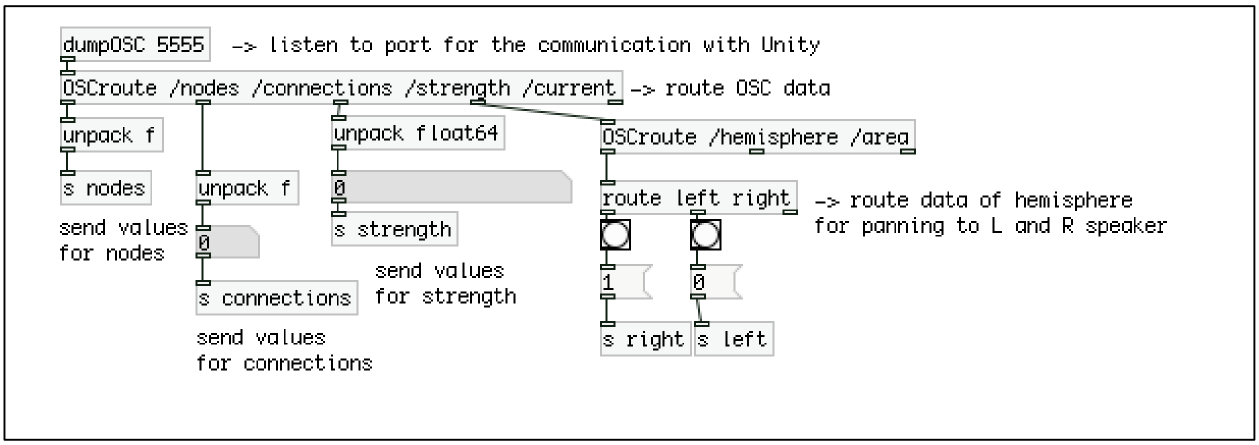


Figure 6. Subsystem in Pure Data responsible for receiving the OSC messages from Unity.

3.2.2 Sonification mappings

a) Nodes

For the nodes, a narrow-band short sound was chosen having an "electric" quality with the objective to simulate laboratory sound recordings of neural activity. As mentioned above the rate of repetition of the sound was mapped to the number of nodes. Since the nodes have a very small range the sample was never played back so fast as to create a continuous sound. On the low end of the range again, the sample was repeated fast enough to remain on the psychological present (Bregman, 1994). Furthermore, it was intended to avoid a clearly mechanical repetitive feeling of the sound. For this, a random factor was added to the

[6] http://puredata.info/

repetition rate, which would simulate the quantity of nodes corresponding to different regions, giving the impression of distinct particles being activated. The tuning of this factor was based on empirical observation, the goal being not to distort the sense of repetition by for example creating too many bursts of sound. A different scaling was required for the low values and the higher values. This was based once again on the distribution of the nodes in the network. Since a high number of regions presented low number of nodes the scaling for these values had to cover a bigger range. Two linear functions served the purposes for these mappings based on two groups below and above the value of 12 nodes, where a more evident difference in the resolution was observed (Fig. 7).

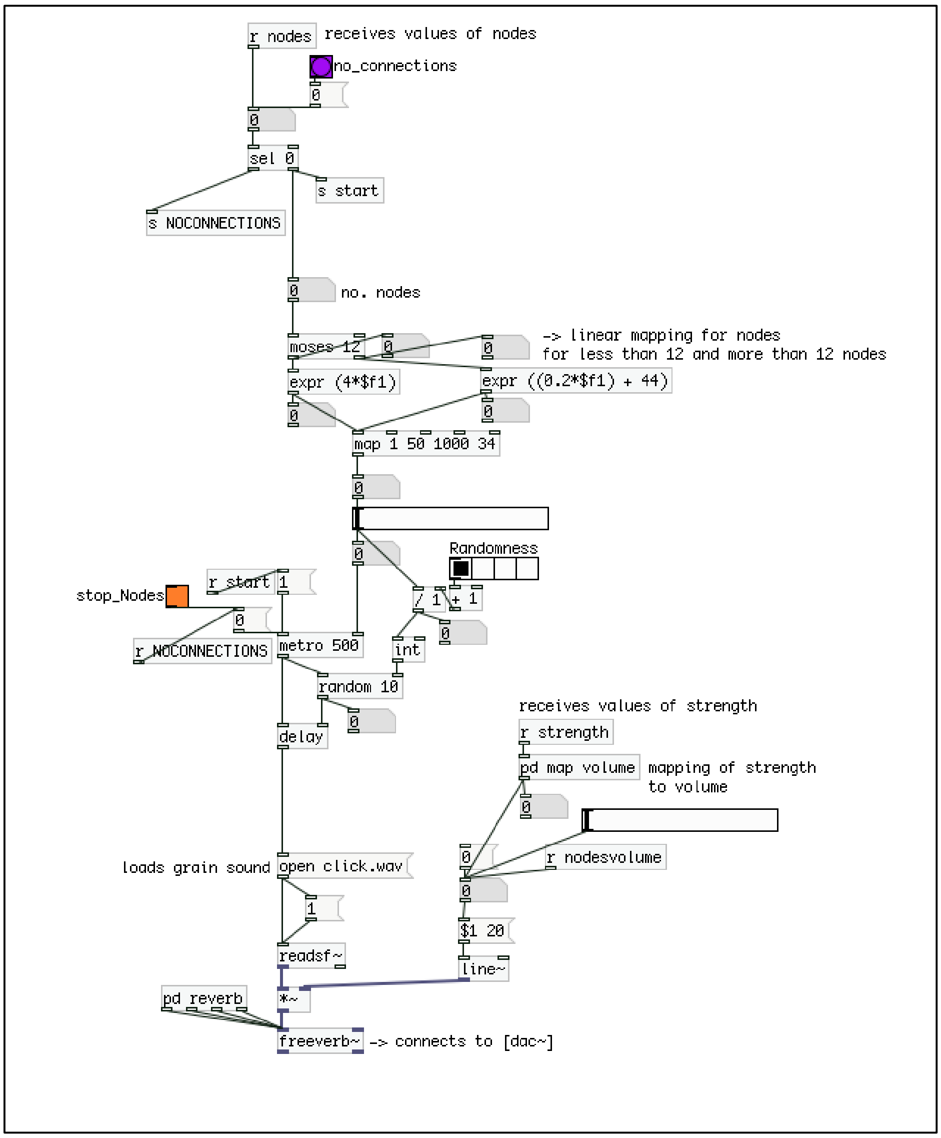


Figure 7. Patch within Pure Data of the mapping of the nodes values to the grain's repetition rate and the strength to the amplitude of the grain sound.

b) Strength

The values of the average strength were mapped to the amplitude of the grain sound as seen in the figure above (Fig. 7). As explained before if the amplitude was manipulated with the frequency of the same sound the interaction of these two parameters would lead to unperceived and confusing changes of the sound. For this reason the amplitude of the grain sound was preferred. The range in which the strength was mapped to was of 30 dB. Low amplitude corresponded to low average strength of a region and high amplitude to higher values of average strength. A linear mapping was implemented in the case of the strength, which was divided in 3 different ranges that were determined by testing the amplitude in relation to its interaction with the other two sound parameters Additionally, a reverb was added to the higher values to enhance the perception of higher strength for the correspondent regions of the network.

c) Connections

The parameter mapped to the number of connections is the pitch. The decision to use pitch was based on two facts. First, it has been used widely with success in previous studies (Walker, 2002; Flowers, 2005). Second, it offers large resolution (Carlile, 2011). The latter is crucial for this system since the number of connections present a very wide range (as shown in Fig. 3 and Table 1). It was decided to use an ambient sound instead of a pure sine wave. The main reasoning behind this is connected to the neuroscience application itself. This sonification should function as an extra channel for the user investigating the network. Since the XIM is an environment where the user is immersed, and taking into account that the connectome is a complex network a user can spend a great amount of time in order to get familiarized with the network. For this reason, a softer, more ambient sound was chosen which would not be annoying to the ear as a pure sine wave would be (Brown et al., 2003). First, it was necessary to analyze the sound. For the frequency analysis the [fiddle] object was used in Pure Data. It revealed a fundamental frequency of around 380-400 Hz although due to inaccuracies further analysis was needed. Spectral analysis with SonicVisualiser[7] confirmed that the most prominent frequency was at 380Hz.

[7] http://www.sonicvisualiser.org/

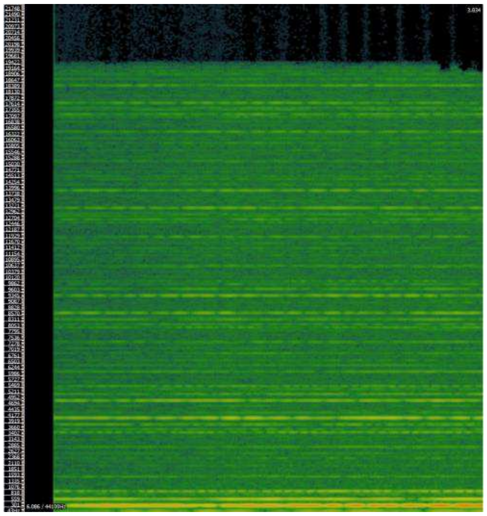


Figure 8. Spectrogram of the sound implemented in the SonicVisualiser. The most prominent frequency found at 380 Hz indicated with the yellow orange color low in the spectrogram.

Furthermore the spectrogram of the sound shows some concentration of energy at around 1200 Hz. This explained why pitch shifting the sample was not working as expected for the whole range. At some frequencies the perception of the pitch was not clear. To rectify this problem and make pitch perception more robust a pure sine was added to the original sound. It was mixed on -12 decibels so it seamlessly blends with the original sound. Finally, the amplitude of the sound was adjusted according to the equal - loudness contours (Carlile, 2011). This correction was important so that all the ranges of the frequencies would be at the same level of loudness.

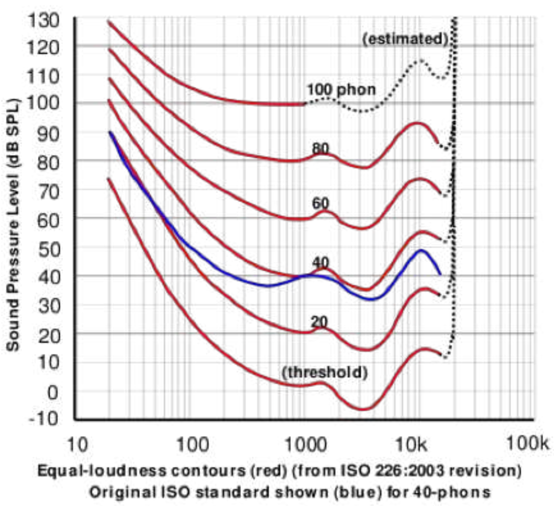


Figure 9. Equal-loudness contours used for the adjustment of the amplitude level of the sound for the mapping of the number of connections.

The [pitchshifter] object was used to change the frequency of the sound sample. This was necessary in order to transpose the original sound's frequency to lower and higher values.

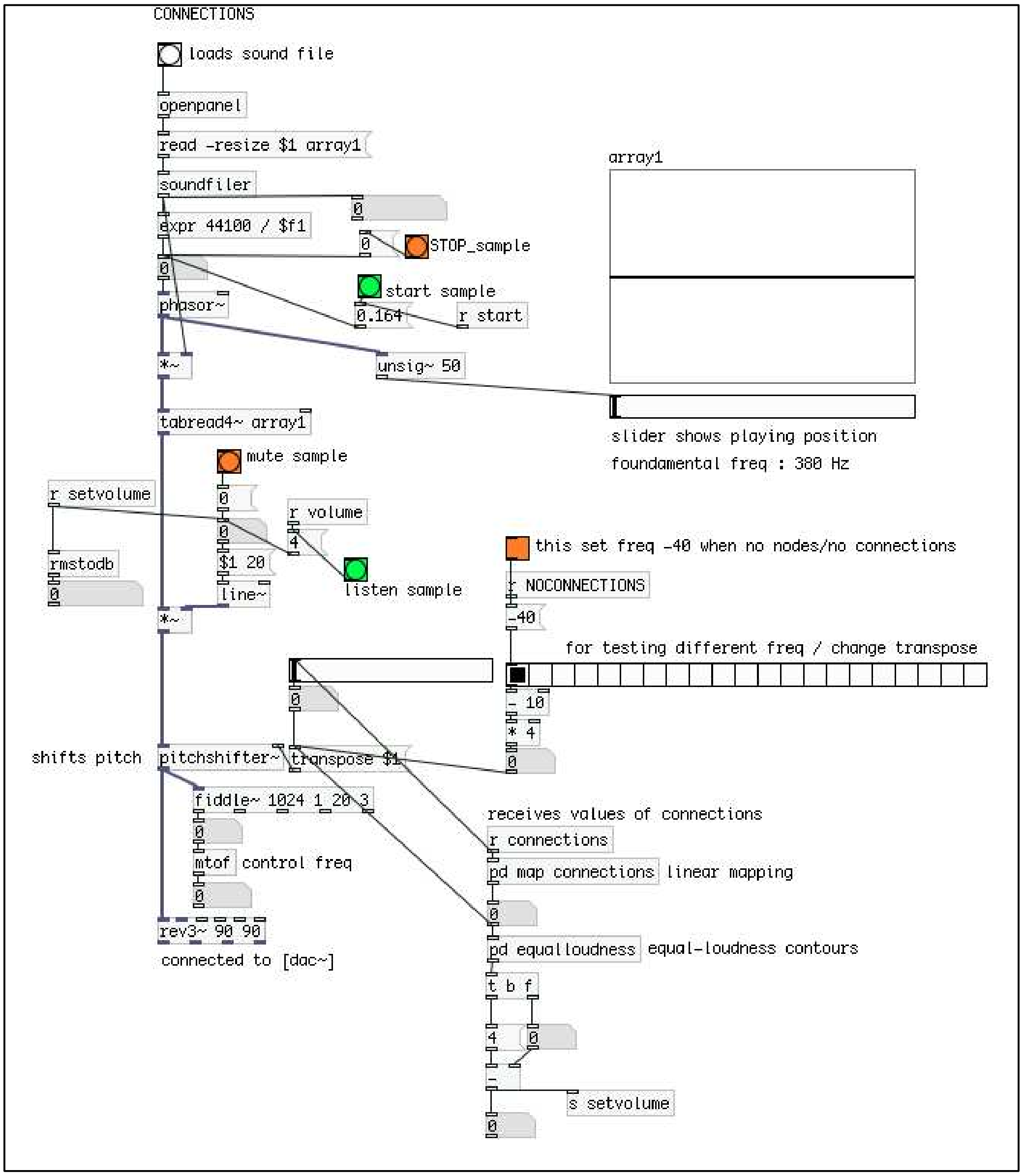


Figure 10. Patch within Pure Data designed for the connections mapping.

The mapping was based on a linear function and the data coming from the number of connections (7-1500) were mapped to frequencies ranging from 100 Hz to 1200 Hz.

d) Brain Hemisphere

The last parameter to be sonified was the brain hemisphere. Panning was thought to be appropriate for the discrimination of the left and right hemisphere as direction of the sound source would make more intuitive and direct the understanding of the location of the regions in the network (Carlile, 2011). Binary signals of 0,1 were received from the Unity application, which were then mapped to the two speakers through the [pan] object.

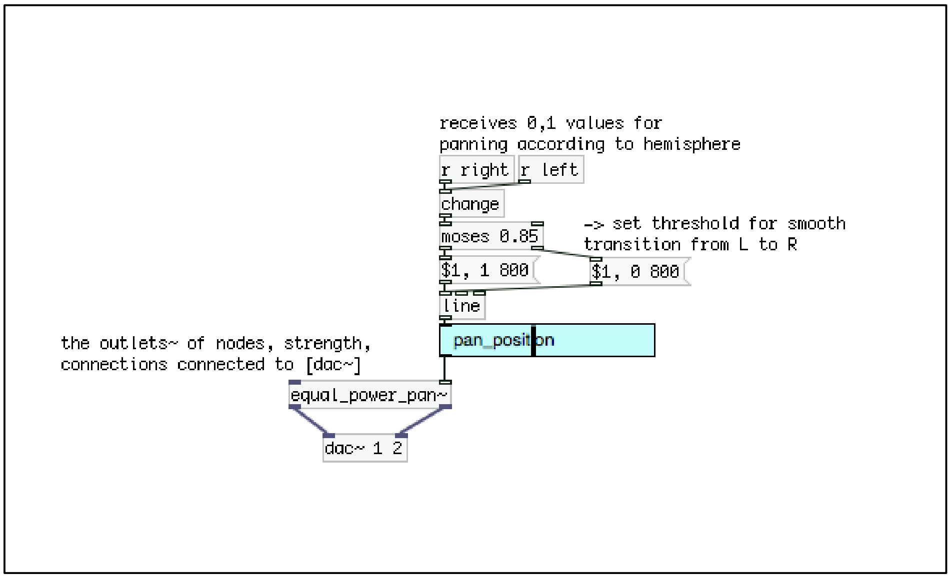


Figure 11. Patch within Pure Data designed for the panning depending on the hemisphere.

3.3 Experimental design and set-up

3.3.1 The eXperience Induction Machine

Before the final experiment two pilot experiments were conducted. The experiments were carried out in the eXperience Induction Machine (XIM) (Eng et al., 2003; Bernardet et al., 2007; Betella et al., 2012). The XIM is a multiuser mixed-reality space covering a surface area of 5.5 x 5.5m equipped with a number of sensors and effectors (Fig. 12). XIM effectors include computer graphics content projected via 8 projectors on 4 separate walls, a luminous interactive floor, movable lights and sonification system. For the purposes of this project four projectors were used as a visual displays of the connectome network on 4 separate walls and two speakers in the left and right corners of the room for the auditory display. A table and a comfortable chair were placed in the central front part of the room so that the participants would have a good auditory perception of the sound coming from the two speakers and additionally they would be able to fill the questionnaires and use the keyboard and mouse to navigate through the network. The lights were dimmed to provide an immersed experience.

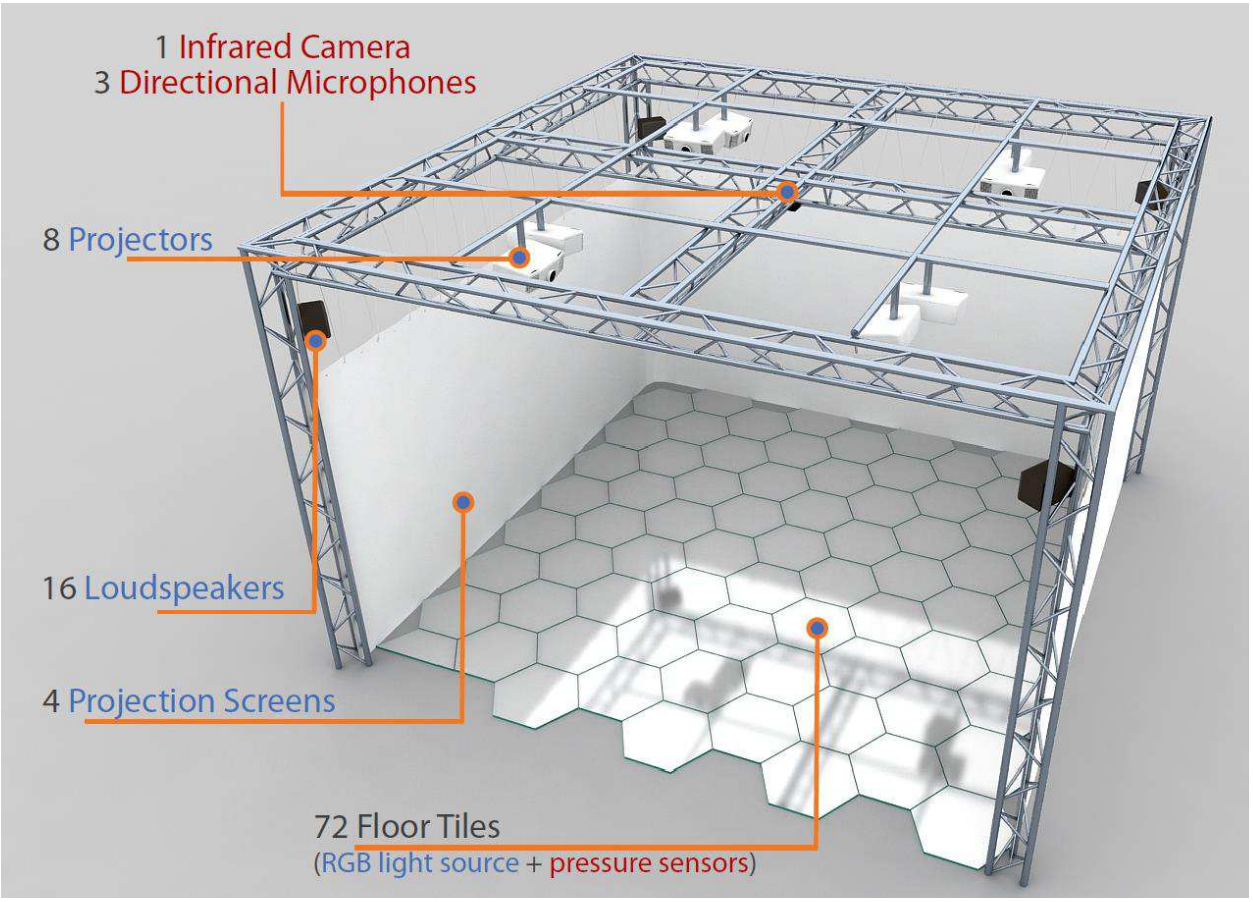


Figure 12. Illustration of the eXperience Induction Machine. The sensors are indicated with red color and the effectors with blue.

3.3.2 Pilot experiment

A first pilot experiment was conducted. The basic objective of this preliminary experiment was to test the design of the experiment and detect defects and problems that could emerge for the final experiment. The pilot experiment consisted of two different sessions. In this experiment the engagement was also measured with the ITC-SOPI questionnaire (Lessiter et al., 2001) that measures spatial presence, engagement, ecological validity/naturalness and negative effects.

The sample for the pilot experiment consisted of 10 healthy adults (4 females, mean age 26.50, SD ± 3.7) that were recruited among the undergraduate students of the Universitat Pompeu Fabra in Barcelona. The subjects were informed that they would be presented various areas of the connectome network and later would be asked questions about its structure and its characteristics.

This pilot experiment followed an independent samples design. The participants were equally divided in two groups and each group participated in a different condition. The first condition was: only visualization and the second one: both visualization with sonification. The procedure followed was the same for both groups.

In the beginning the subjects were presented with the connectome network projected in front of them and different areas were showed to them in order to get familiarized (with or without sound depending on the group). They were also informed of the minimum and maximum values of the characteristics in the network. A previous version of the network application (Fig. 13) was used for the pilot since the latest version with the anatomical atlas and other visual improvements was not available yet.

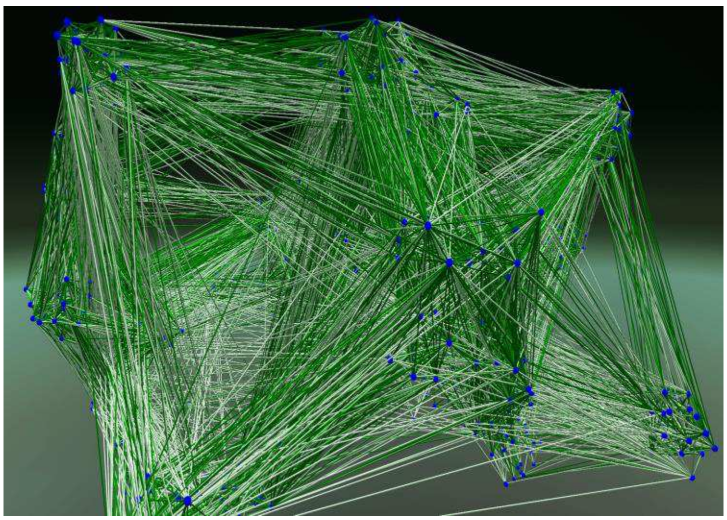


Figure 13. Previous version of the neuroscience application used for the pilot experiment.

In the first session the participants were presented with 5 different consecutive regions (the order was randomized). The GUI of the application was visible which provided the following values: a) name of region including the hemisphere, b) number of nodes c) number of connections and d) av. strength. The participants were also provided with a printed visual guide of the 5 regions in order to facilitate their task and help their orientation in the network. Then they were asked to fill in a questionnaire relative to their characteristics (i.e. which of the five regions presented the highest number of nodes) (Annex I). The participants had to mark their answers in close-ended questions. The measurements for this task was based on the number of correct and wrong answers of the participants. The average score of each participant was calculated.

In the second session they were allowed to navigate freely through the network for 3 min (with or without sound) and explore the network. The participants were informed that they would be given a questionnaire where they would be asked to answer questions concerning their experience. The participants used the keyboard for the navigation and then they were asked to fill the ITC-SOPI questionnaire on engagement (Annex II).

3.3.3 Empirical Validation

For the second experiment 25 healthy adults (15 females, mean age=29.53, SD ± 5.6) with normal or corrected vision and hearing were recruited. The subjects were naïve and had no prior scientific knowledge of neural networks.

In order to avoid differences due to the single subject skills the experiment followed a paired samples design where each participant was exposed to both conditions (in a random order). The two conditions consisted of a) only visualization and b) both visualization and sonification[8]. The independent variable was the presence of sound vs the absence of sound and the dependent variable was the estimation of the values of the changes of the characteristics of the network between different regions with different characteristics. The experiment consisted of 3 sessions that are described in detail in the following sections and the participants were allowed to take breaks between the sessions.

3.3.1 Experimental protocol

a) Demographics

In the beginning the subjects were asked to fill a consent form and a demographic questionnaire. The demographic questionnaire (Annex III) included personal questions concerning the age, gender, musical background, knowledge on neural networks resulting in an average formal or informal music training of 1.84 (SD ± .85) on a scale of from 1 (none) to 4 (expert).

b) Introduction

The subjects were given a short introduction about the experiment and the objective. The task of the subjects was to estimate the number of nodes, connections and the average strength of each region presented and understand the changes of these values during the navigation. They were presented with the network and a brief explanation followed in order to understand the characteristics of the network (nodes, connections, av. strength). In this experiment the latest version of the neuroscience application was used compared to the one

[8] These conditions will be further referred to as a) *visualization* and b) *sonification*

adopted in the pilot experiment. This new version features a better allocation of the nodes on their anatomical X,Y,Z coordinates according to the talaraic atlas. The visualization is improved and there is no cluster effect as in the previous version (Fig. 14).

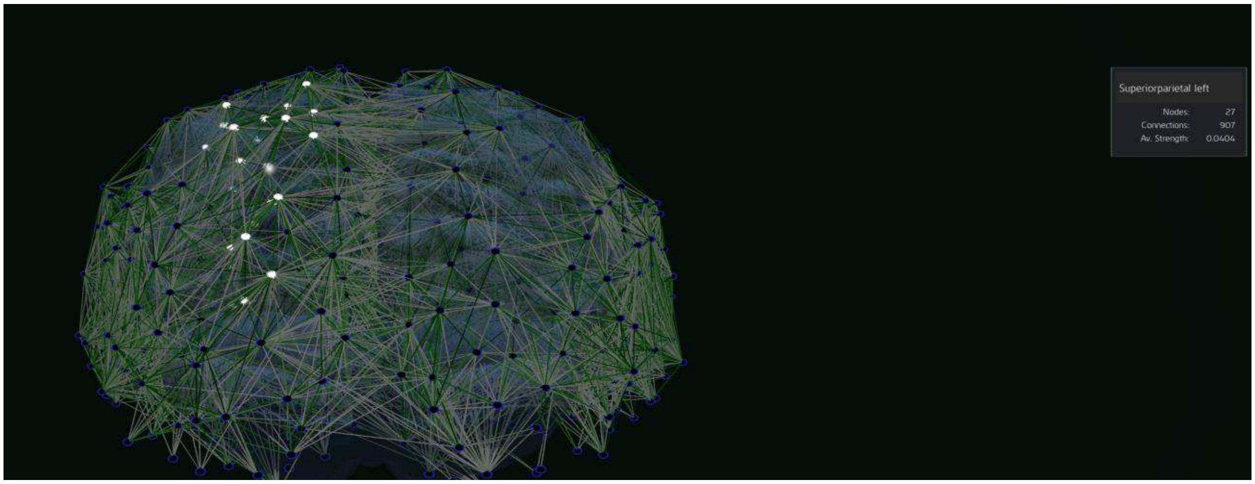


Figure 14. The latest version of the neuroscience application.

The experiment consisted of 3 sessions. Each of the sessions consisted of two parts (one with sound and the other without, in a random order) and the GUI was deactivated. Before the first session the participants were explained how the characteristics were represented visually and acoustically and were informed of the ranges of the scales of the characteristics of the regions in the network. A printed table with the minimum and maximum values of the characteristics was also provided to them. They were shown different regions with the GUI activated and they were informed of the values verbally and visually so that they could control the values and get familiarized with the connectome network. The subjects were informed that the task did not consist in memorizing nor counting, but making estimations about the values. In the following figure an overview of the procedure followed and the timeline of the experiment are presented.

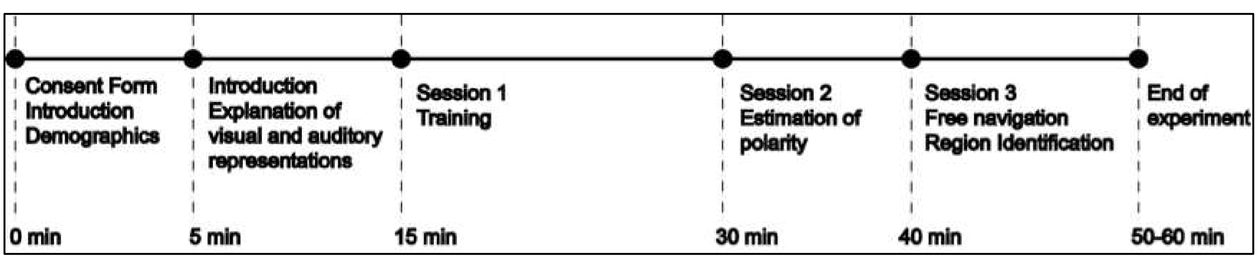


Figure 15. Procedure and timeline of the experiment.

c) Session1 – Training

The first session had a training character so that the participants could get familiarized with the characteristics and their corresponding visual and auditory representations.

Task

The participants were presented with 7 different regions for each of the two conditions (*visualization* and *sonification*) and were asked to estimate the number of nodes, connections and av. strength of each region and in which hemisphere the region was encountered. Then, they were asked to mark their answers in a predetermined scale, which varied depending on the values of each characteristic and their distribution in the network (more specifically for the nodes was a 6-scale, for connections 7-scale, av. Strength 6-scale and 2 options for the hemisphere). After completing the task for each region they were given feedback on their answers so that they could control their possible mistakes.

An example of the questionnaire is shown in the following figure.

Nodes

1.	< 5	☐
2.	6 – 10	☐
3.	11 - 15	☐
4.	16 - 20	☐
5.	21 - 25	☐
6.	> 25	☐

Connections

1.	< 100	☐
2.	100 – 300	☐
3.	300 – 500	☐
4.	500 – 700	☐
5.	700 – 1000	☐
6.	1000 – 1200	☐
7.	> 1200	☐

Av. Strength

1.	< 0.01	☐
2.	~ 0.01	☐
3.	~ 0.02	☐
4.	~ 0.03	☐
5.	~ 0.04	☐
6.	> 0.04	☐

Hemisphere L ☐ R ☐

Figure 16. Example of the questionnaire used in the first session. The scales were determined depending on the distribution of the characteristics in the network.

Regions selection

The regions presented for both of the parts of this session had very similar characteristics and their values entered in the same ranges (Annex IV). This decision was made so that the conditions for both of the conditions were similar and the participants would be exposed in the whole range of values of nodes, connections and av. strength.

In total, the participants were exposed to 14 regions for both conditions, resulting in 7 trials for each condition. After the presentation of each region the participants marked their answers in a questionnaire (for the correspondent region) and then they were given feedback on the correct answers.

Score attribution criterion

In this session the error deviation from the correct answer in a predetermined scale was measured (based on the scale shown in Fig. 16).

The absolute distance from the correct answer was calculated (Walker et al., 2004) and divided by the number of possible answers, which provided the normalization of the scores. The result was subtracted from 1, resulting in scores equal to 1 for the correct answers and 0 for the most distant one in the scale.

The error deviation was calculated according the following formula.

$$\beta_{j=1-\frac{|j-k|}{N}}$$

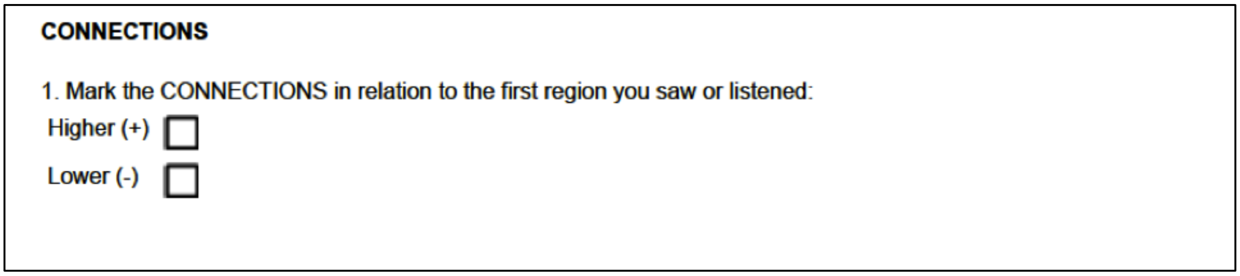
β = normalized score
j = participants answer
k = correct answer
N = total number of probable answers

The average of the normalized scores of all the answers for each participant was calculated, in both conditions.

d) Session 2

Task

In this session the subjects were presented with 18 different pairs of regions (9 for each condition, in a random order). After the presentation of each pair of regions they had to evaluate the polarity of the values. Specifically, the participants were informed of the characteristic they had to evaluate (number of nodes, connections or av. strength) and they were presented two different regions. The participants were asked to decide if the second region presented a higher or lower value of the characteristic asked compared to the first region (this is referred to in previous studies as estimation of polarity (Walker and Lane 2001). Then they were asked to mark with +/- their answer in a questionnaire, as shown in the following figure.


CONNECTIONS

1. Mark the CONNECTIONS in relation to the first region you saw or listened:

Higher (+) ☐

Lower (-) ☐

Figure 17. Example of the questionnaire provided to the participants in the second session.

If the subjects asked to see or listen again the regions they were allowed to ask for repetition. The regions presented with sound did not exceed the duration of 12 seconds since in the literature it is suggested that "this is the most effective duration since auditory sensory memory is an issue for making such comparisons; if displays or stimuli exceed 12 seconds it is likely that memory for events at the beginning of the display will be degraded and the ability of participants to make reliable comparisons will be impaired, but should not presented too rapidly because shortening the duration may run the risk that perception of auditory patterns will be impaired" (Bonebrigth & Flowers, 2011).

Regions selection

For each one of the 3 parameters (nodes, connections, av. strength) at least one pair presents evident difference for the parameters measured. The rest of the regions presented variation in their values of the three characteristics. In Annex V the selected regions are presented.

In total, the participants were exposed to 36 regions for both conditions, resulting in 18 trials (9 pairs of regions for each condition). After the presentation of each region the participants marked their answers in the questionnaire.

Score attribution criterion

The score was based on the number of correct and wrong answers. A score of 1 was assigned to the questions answered correctly and 0 to incorrect answers. A total score from 0 to 9 was calculated for each of the two conditions for each participant.

e) Session 3

Task

In this session the participants were asked to navigate freely through the connectome network. For the navigation they used the keyboard and the mouse. Their task consisted in finding a region with certain values (e.g. a region with low number of connections between 200-400 connections, a region with medium-high number of nodes between 20-30 nodes, etc) (Annex VI). There was no time limitation and they were also given a printed table with the minimum and maximum values of each of the parameters. When they were confident that they found the region that applied to the characteristics asked they were given feedback and they continued with the next one. The task was repeated twice, one for each condition. Each participant completed 6 trials for each condition.

Score attribution criterion

In this session two measurements were made:

1. A score of 1 was assigned to the questions answered correctly and 0 to incorrect answers. A total score from 0 to 9 was calculated for each of the two conditions for each participant.

2. As in the first session the absolute distance from the correct answer was calculated in a predetermined scale designed ad hoc for the evaluation (Annex VI). The distance was divided by the number of possible answers, which provided the normalization of the scores. The result was subtracted from 1 resulting in scores equal to 1 for the correct answers and 0 for the most distant one in the scale. The same formula as in Session 1 was used. The following table summarizes the tasks and the score attribution criteria for each one of the sessions.

Table 3. Table shows the task of each session and the score attribution criterion.

Session	Task	Score attribution criterion
Session 1	Estimation of the values of the network's characteristics	Error deviation from correct answer in predetermined scale
Session 2	Estimation of polarity	Correct answers were attributed with 1, otherwise with 0
Session 3	Finding regions with certain characteristics in predetermined ranges of values	1. Correct answers were attributed with 1, otherwise with 0
		2. Error deviation from correct answer in predetermined scale designed ad hoc for the evaluation

4. RESULTS

4.1. Results - Pilot

The first experiment served as a pilot and had a testing character. The main goal of this experiment was on the one hand to spot potential issues with the protocol and on the other hand acquire some preliminary results that would help us with the experimental design. A statistical analysis was conducted in SPSS.

a) Session 1

No significant results were found in the measurement of the data acquired from the first session between the two conditions. The sample was very small and the statistical analysis did not show any differences that would lead to further analysis of the results. The only observation that could be made is that for regions that presented that presented very high and very low number of nodes we acquired the highest number of correct answers. However, this resulted for both conditions. Nonetheless, the pilot was very useful for the design of the following experiment.

b) Session 2

The data obtained from the engagement questionnaire satisfied the normality criterion as verified using the Shapiro-Wilk test.

An independent T-test was conducted to measure differences in the engagement between the group that navigated without sound (*visualization* condition) and the group that navigated with sound (*sonification* condition). Although there was no significant difference, there was a tendency for higher score assigned to the sonification (M=3.51, SE=0.31) compared to the visualization (M=3.14 , SE=0.25) $t(8)$=-.93, p>.05 (Fig.18)

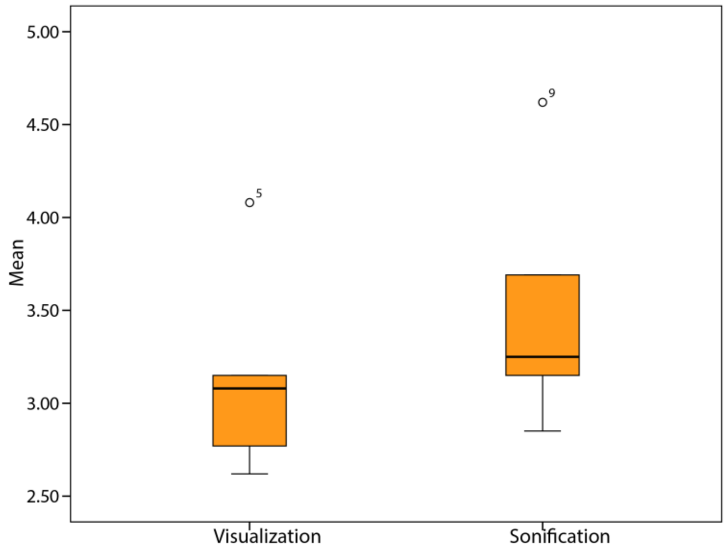


Figure 18. Boxplots for the engagement ratings between the two conditions: *visualization* and *sonification*.

The most important finding was that the independent samples design might not have been the most appropriate one and the second experiment followed a paired samples design.

4.2 Results – Experiment

a) Session 1

Although the first session was considered to be a training session, a statistical analysis for the measurements of the error deviation was conducted. The average values of the normalized scores were calculated for each participant for both conditions (*visualization* and *sonification*).

The means of each of the three characteristics of each region were calculated and the total mean of the scores for both conditions.

Table 4. Means of the normalized scores for the two conditions including the individual scores obtained for each of the three characteristics.

Session1 Estimation of values				
	Visualization		Sonification	
Parameters	Mean	SD	Mean	SD
Nodes	0.90	0.04	0.87	0.05
Connections	0.80	0.06	0.84	0.04
Strength	0.76	0.07	0.79	0.06
Total	0.82	0.04	0.83	0.03

The data satisfied the normality criterion as verified using the Shapiro-Wilk test. A dependent T-test was conducted to evaluate differences on the estimation scorings for the characteristics of the different regions presented in the two conditions. No significance was found and there was no order or gender effect.

b) Session 2

A Wilcoxon test was conducted to evaluate differences on the estimation of polarity of the values of the pairs of regions presented between the two conditions. The correct answers for both conditions were calculated. The *sonification* condition obtained a significant higher score (Mdn=7.00), $z=-2.96$, $p<0.05$, $r=-0.57$ compared to the *visualization* condition (Mdn = 6.00) (Fig. 19).

Since the experiment followed a paired samples design, the data were tested for order effect. There was not found an order nor a gender effect.

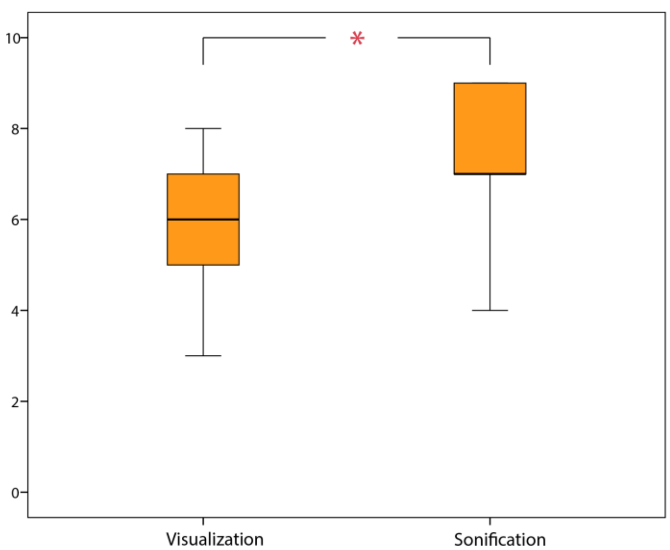


Figure 19. Boxplot shows the significant effect on polarity estimation task between the two conditions.

The means of the correct answers of the participants for each of the parameters of the networks are presented in the following table. Wilcoxon tests were conducted for each of the parameters between the two conditions. The means for the *sonification* were higher compared to the ones of the *visualization*. However, no significance was found in the means between the separate characteristics.

Table 5. Means of the scores for the correct answers for the two conditions including the individual scores obtained for each of the three characteristics.

Session 2 Estimation of polarity of values				
	Visualization		Sonification	
	Mean	SD	Mean	SD
Nodes	2.32	0.80	2.41	0.60
Connections	2.12	0.97	2.48	0.82
Strength	1.76	0.66	2.16	0.99
Total	6.00	1.52	7.00*	1.57

As observed from the boxplot above (Fig. 19) an important amount of the population obtained a low score for the *sonification* condition. For this reason, a correlation was conducted between the musical background and the ratings of the polarity in the *sonification* condition to test if there was any effect.

A significant negative correlation was found between the scores (correct answers) of the participants and their musical background.

Table 6. Table shows the significant negative correlation between the scores obtained from the correct answers and the musical background of the participants.

Correlations

			Sonification_ TOTAL	Former Musical Training
Kendall's tau_b	Sonification_TOTAL	Correlation Coefficient	1.000	-.300*
		Sig. (1-tailed)	.	.042
		N	25	25
	Former Musical Training	Correlation Coefficient	-.300*	1.000
		Sig. (1-tailed)	.042	.
		N	25	25

*. Correlation is significant at the 0.05 level (1-tailed).

c) Session 3

First, a Wilcoxon test was conducted to test for differences in the ratings of correct and incorrect answers between the two conditions. No significant difference was found between the two conditions.

A dependent T-test was conducted for the total scores obtained from the error deviation measurements. The normalized scores for each participant were calculated for each of the regions found by the participant through the navigation in the network. The average score was calculated for each of the two conditions for each participant. A significant higher score was obtained for the *sonification* condition ($M=8,03$, $SE=0.11$) compared to the visualization ($M=7.15$, $SE=0.13$) $t(24) = -5.03$, $p<0.001$, $r=-.12$.

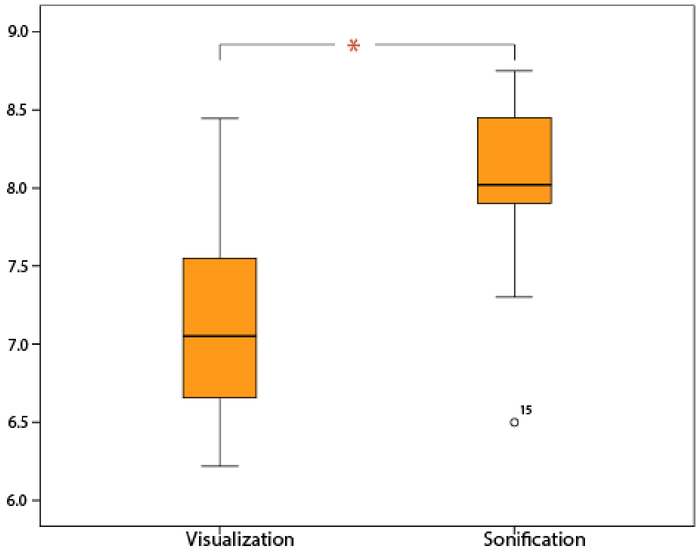


Figure 20. Boxplot shows the significant score obtained for the sonification condition for the task on finding regions with specific parameters.

Wilcoxon tests for the separate parameters were conducted. The scores obtained were higher for the *sonification* for all the separate parameters asked. In addition and significance was found for the regions with low number of connections between the two conditions.

The *sonification* condition obtained a significant higher score (Mdn=1.00), z=-3.42, p<0.05, r=-0.68 compared to the *visualization* (Mdn = 0.20).

Table 7. Means of the scores obtained from the error deviation for the correct answers for both conditions including the individual scores obtained for each of the three characteristics. The significant results are indicated with an asterisk.

Session 3 Correct Answers – Error deviation				
	Visualization		Sonification	
	Mean	SD	Mean	SD
Nodes Low	0.94	0.09	0.96	0.10
Nodes High	0.90	0.14	0.95	0.11
Connections Low	0.48	0.35	0.85*	0.21
Connections High	0.83	0.21	0.87	0.13
Strength Low	0.77	0.29	0.87	0.13
Strength High	0.83	0.16	0.84	0.14
Total	7.15	0.62	8.03*	0.54

5 DISCUSSION AND CONCLUSION

The goal of this study is to explore whether the addition of sound in the neuroscience application enhances the understanding of the relationships of the networks parameters, during the navigation through the connectome network in the eXperience Induction Machine (XIM). The network's parameters (nodes, connections, av. strength and hemisphere) were mapped to different sound parameters (repetition rate, amplitude and pitch) and the estimation of the values in the two conditions (absence and presence of sound) was measured.

Two experiments were conducted in the XIM, the first one having a testing character for the design of the second experiment. The second one consisted of three sessions with the objective to measure whether sound could enhance the understanding of the network's dynamics.

The results of the first session of the pilot experiment played a decisive role in the protocol design of the second experiment. Although no significant results were obtained from this experiment, it provided an insight for the design that would be adopted later.

In the first session five consecutive regions were presented to the participants with their names and the values of their characteristics visible in the GUI. They were asked to compare their characteristics and answer questions, in which they had to select the correspondent region that filled the criteria asked. However, the results and the participants' qualitative comments showed that they were not able to remember the names of the regions and make comparison between them, when asked questions on their values. Even though the participants were provided with a visual printed guide they were not able to answer correctly most of the questions and they commented that they found it difficult to get oriented in the network. These findings showed that the task of comparing many consecutive regions proved to be difficult for the participants and they were taken into consideration for the final experiment. Thus, the experimental protocol was reconsidered and a different design was adopted. During the comparison task of the second experiment pairs of regions were selected to minimize the effect of attention and working memory.

Regarding the measurement of the engagement for the two conditions no significant results were obtained. This led into two considerations. One was related to the nature of the questionnaire, which was considered not to be appropriate for the neuroscience application and the certain task. The ITC-SOPI is merely designed for media, such as films, videos and computer games in which the content usually includes characters and story telling and tests how the participants feel during the experience. Since the nature of the virtual environment of connectome application differs from these media it was decided that engagement would not be further measured in the next experiment.

The second one was related to the experiment protocol. The results showed that a paired samples design would be more adequate for the measurement between the two conditions. Hence, a paired samples designed was adopted in the following experiment.

A second experiment was conducted in the XIM and the estimation of values of the parameters was measured in three separate sessions. The first having a training character, the second one measuring the ability of estimating the polarity of the values of the networks characteristics between pairs of regions and the third one measuring the ability of the participants to find regions with certain values of the three parameters during a free navigation in the network.

The analysis of the data for the three sessions revealed coherence in the results between Session 2 and Session 3. The sonification condition obtained higher scores in two different tasks and thus, the alternative hypothesis was retained. Sonification significantly enhanced the performance of the users in terms of estimation of polarity and finding regions with certain characteristics.

Furthermore, the analysis of the data for the third session showed that participants during the free navigation through the network obtained higher scores for the regions with low number of connections in the *sonification* condition. This shows that pitch enhanced significantly their task compared to the visual task. The network's connections are difficult to discern visually, qualitative comments from the participants revealed that they could intuit the number in relation with the nodes (low number of nodes in the networks corresponds to low number of connections and vice versa in most of the regions) but the addition of sound shows that sonification enhanced the accuracy during the navigation.

In addition, the significant results in favor of the sonification condition shows that the selected sound parameters corresponding to the networks characteristics were effective. In all sessions the *sonification* condition resulted in higher scores revealing that the selection of the sound parameters and the sonification design enhanced the task of the participants.
However, there were some interesting findings to be discussed. In the first session, the analysis of the separate means for the estimation of the three separate network's parameters showed a higher score for the nodes. Examination of the results showed that higher scores were obtained for regions that presented very low number of nodes (between 6-10). This could be explained by the fact that estimation of such low number of elements are especially easy to detect with the eye. In this specific case further experiments would be useful to study whether cross-modal effect interactions may cause interference effects (Eldridge, 2005).

Another interesting finding is that in the third session participants obtained lower scores for the strength in the *sonification* condition for specific regions. This may be due to the interaction of the sound parameters (repetition rate and loudness). When an area presented high number of nodes but low strength the interactions of these two parameters may confuse the perception and these regions maybe perceived as having high number of strength. Further research on the interaction of the sound parameters would provide better understanding of the auditory perception. It would be interesting to examine the effectiveness of the auditory display with different sound parameters and also with different networks exposing the participants to higher number of trials. This would give us insight for the interactions of sound parameters and the auditory perception of the user.

Regarding the negative correlation, in the third session, between the scores obtained of the sonification condition and the musical formal training of the participants our prediction was that those with low musical background would provide lower scores. However, analysis of the results showed the opposite. In the literature, there is no agreement about the relation of the musical background in auditory tasks. The reason underlying these results is not clear and further investigation is required. As suggested by Walker and Nees (2011) "a person could have had many years of musical experience as child, yet that person could be many years removed from their musical training and exhibit no more musical ability than someone who received no formal training". This could explain the results obtained. In addition, a more reliable, and valid measure of musical ability than a filling a questionnaire could probably give a better understanding on the results.

In most of the sonification studies it is suggested that training sessions are very important (Bonebright & Flowers, 2011). Visual information displays owe much of their success to their pervasiveness as well as to users' formal education and informal experience at deciphering their meanings (Walker & Nees, 2011). Visual representations are taught from a young age (Ferguson & Cabrera, 2008) and we are familiarized with graphs and plots. However, complex auditory displays currently are not pervasive, and users are not taught how to comprehend auditory displays as part of a standard education. The training sessions in the sonification research and experiments play an important role with the scope to get the participants familiarized. These sessions usually have duration of 20 minutes. In this experiment the training session for both the conditions were about 10 min and further training would probably enhance the results for the *sonification* condition.

Concluding, the results obtained in this study are consistent with the literature regarding the study of the effectiveness of the auditory display and the addition of sound in visual displays, as cited in the previous sections. Sound enhanced the task of the participants and provided a first step in the sonification of the connectome network. The sonification can be used to help the navigation of the user in the neuroscience

application. Further improvements and experimentation on the sonification models could lead to a powerful tool for the exploration of the neural networks.
The connectome network is a complex dataset and giving a further layer with the sonification we achieved to enrich the neuroscience application helping the user understand better the relations and the dynamics of the network. Large datasets are continuously generated in different research fields and sonification can be used to facilitate the task of extracting important information.

Summary of conclusions

• The alternative hypothesis was retained. Sonification significantly enhanced the performance of the users in terms of estimation of polarity and finding regions with certain characteristics.

• Coherence was found in the results between Session 2 and Session 3. Sonification condition obtained higher score in two different tasks.

• Sonification is an effective method for the visual display of the connectome application.

• The selected sound parameters found to be effective for the certain task.

• The negative correlation between the participants scores and the former musical background is interesting for further research.

• In the third session the scores for av. strength were lower for the sonification condition in certain pairs of regions. Although not significant, interactions between the sound parameters (loudness and rate) should be examined.

• Further training for the auditory condition could enhance the results.

6 FUTURE STEPS

This project can be considered as a first step for the sonification of the connectome network. A first approach related to the estimation and the understanding of the relationships of the networks characteristics was proven to be successful. However, the connectome application has a great potential. Improvements on the current work for extending the sonification model for the neuroscience application are proposed.

Regarding the current sonification design, a different task would be interesting for future experiments, such as presenting different regions or even different networks to the participants and evaluate the ability to detect differences between them. Studying the interchange of the sound parameters correspondent to the network's characteristics would offer further understanding on cognitive abilities of the user and auditory perception.

As aforementioned the visualization used for the final experiment was a newer version of the connectome network. Although the visualization changed aesthetically the dataset remained the same. But, nonetheless the newer version of the neuroscience application has much more possibilities.

In addition, designing a sonification that would be based on the algorithms described in Hagmann's paper for the extraction of patterns and understanding the network's structure would be a challenging work and would enrich the neuroscience application.

This of course would lead to a different task and a new sonification design. This thesis was tested with naïve subjects. Examining the effectiveness of the new proposed multimodal display and testing it with professionals would be of great interest.

Furthermore, creating an adaptive system that would function as a machine learning system for all kinds of networks would add potential to the neuroscience application and would add potential as a powerful tool for uncovering complicated structures.

Finally, the sonification of implicit signals (ECG and EDR) and study the effect of these measurements in a interactive virtual environment such as the eXperience Induction Machine would enhance the neuroscience application and offer the possibility to study cognitive processes on the understanding of big data sets.

7 REFERENCES

Alexander, R.L., Gilbert, J.A., Landi, E., Simoni, M., & Zurbuchen, T. H. (2011). Audification as a diagnostic tool for exploratory heliospheric data analysis. In *Proceedings of the 18th International Conference on Auditory Display*, (pp. 18-21), Atlanta, GA, USA.

Barrass, S. & Vickers, P. (2011). Sonification Design and Aesthetics. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 142 – 171). Logos Publidhing House, Berlin, Germany.

Bernardet, U., Bermúdez i Badia, S., & Verschure P.F.M.J. (2007). The eXperience Induction Machine and its Role in the Research on Presence, *Presence 2007*, 329–333.

Betella, A., Carvalho, R., Sanchez-Palencia, J., Bernardet, U., & Verschure, P.F.M.J. (2012, March). Embodied interaction with complex neuronal data in mixed-reality. In *Proceedings of the 2012 Virtual Reality International Conference* ACM.

Betella et al. (2013) Advanced Interfaces to Stem the Data Deluge in Mixed Reality: Placing Human (un)Consciousness in the Loop, *SIGGRAPH,* Poster, Anaheim, CA

Blattner, M. M., Sumikawa, D. A., & Greenberg, R.M. (1989) Earcons and icons: Their structure and common design principles. *Human Computer Interaction,* 4(1), 11–44.

Bly, S. (1982). Presenting information in sound. In *Proceedings of the 1982 conference on Human factors in computing systems*, (pp. 371-375), ACM, New York, NY.

Eng, K., Klein, D., Baebler, A., Bernadet, U., Blanchard, M., Costa, M. et al., (2003) Design for a brain revisited: The neuromorphic design and functionality of the interactive space Ada. *Reviews in the Neurosciences*. 14, 145-180.

Janata, P. & Childs, E. (2004). Marketbuzz: Sonification of Real-Time Financial Data. In *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display,* Sydney, Australia.

Bonebrigth, T.L & Flowers, J.H. (2011). Evaluation of Auditory Display. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 109 - 144). Logos Publidhing House, Berlin, Germany.

Brazil, E. and Fernstrom, M. Auditory Icons (2011). In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 323 – 338). Logos Publidhing House, Berlin, Germany.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brown, L.M., Brewster, S.A., & Ramloll R. (2003). Design guidelines for audio presentation of graphs and tables, In *Proceedings of the 2003 International Conference on Auditory Display,* Boston, MA, USA.

Carlile, S. (2011), Psychoacoustics. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 39 - 61). Logos Publidhing House, Berlin, Germany.

Cassidy, R. J., Berger, J., Lee, K., Maggioni, M., & Coifman, R. R. (2004). Auditory display of hyperspectral colon tissue images using vocal synthesis models. In *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia.

Chang, M. H., Wang, G., & Berger, J. (2010). Sonification and visualization of neural data. In *Proceedings of the 16th International Conference on Auditory Display*, (pp. 161–163), Washington DC.

Dombois, F. & Eckel, G. (2011). Audification. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 299 - 324.). Logos Publidhing House, Berlin, Germany.

Dombois, F. (2002). Auditory seismology on free oscillations, focal mechanisms, explosions and synthetic seismograms. In *Proceedings of the 2002 International Conference on Auditory Display,* Kyoto, Japan.

Ferguson, S. & Cabrera D. (2008). Exploratory sound analysis: Sonifying data about sound. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France.

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, (pp. 406-409), Limerick, Ireland.

Fry, B.J. (2004). *Computational Information Design*, Doctoral Dissertation, Massachusetts Institute of Technology, 2004.

Garcia-Ruiz, M. A. (2001). Using non-speech sounds to convey molecular properties in a virtual environment. In *International Conference of New Technologies in Science Education (CINTEC), (*pp. 4-6), University of Averio, Portugal.

Garfield, E. (1989). Creativity and Science. Part 2: The Process of Scientific Discovery. Current Contents, 12(46), 314–320.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory source perception. *Ecological Psychology*, 5 (1), 1-29.

Gaver, W.W. (1989). The Sonic Finder: An interface that uses Auditory Icons the use of non-speech audio at the interface. In *ACM CHI '89*, Austin, Texas, USA, 1989. ACM Press.

Gibson. J. J. (1979). *The Ecological Approach to Visual Perception.* Lawrence Erlbaum Associates Inc. Publishers, Hillsdale, NJ, USA, 10th (1986) edition, 1979.

Grond, F., and Berger, J. (2011). Parameter Mapping Sonification. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 363-397). Logos Publidhing House, Berlin, Germany.

Grond, F., Dall´ Antonia, F. (2008). SUMO – A sonification utility for molecules. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS biology*, *6*(7), e159.

Hayward, Ch. (1994). Listening to the Earth Sing. In Kramer G., (Ed) *Auditory Display. Sonification, Audification, and Auditory Interfaces*, (pp. 369–404). Reading: Addison-Wesley.

Hendrix, C.M. (1994). Exploratory Studies on the Sense of Presence in Virtual Environmentsas a Function of Visual and Auditory Display Parameters, Master Thesis, University of Washington.

Hansen, M.D., Charp, E., Lodha, S., Meads, D., & Pang, A. (1999). Promuse: A system for multi-media data presentation of protein structural alignments, In *Pacific Symposium on Biocomputing*, 4, 368-379.

Hermann, T. (2011). Model Based Sonification. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 397 – 427). Logos Publidhing House, Berlin, Germany.

Hermann, T. (2008). Taxonomy and definitions for sonification and auditory display. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France.

Hermann, T. & Ritter, H. (2005). Crystallization sonification of high-dimensional datasets. *ACM Trans. Applied Perception*, 2(4), 550–558.

Hermann, T. & Ritter, H. (2004). Sound and meaning in auditory data display. In *Proceedings of the IEEE*, *92*(4), 730-741.

Hermann, H. (2002). *Sonification for Exploratory Data Analysis*, Doctoral Dissertation, Technischen Fakultat der Universitat Bielefeld, 2002.

Hermann, T., Hansen, M. H., & Ritter, H. (2001). Combining Visual and Auditory Data Exploration for finding structure in high-dimensional data. *Multimedia Systems*.

Hermann, T. & Ritter, H. (1999). Listen to your data: Model-Based Sonification for data analysis. In Lasker, G. E.,(Ed), *Advances in intelligent computing and multimedia systems*, Baden-Baden, Germany, (pp. 189–194). Int. Inst. for Advanced Studies in System research and cybernetics, 1999.

Kaper, H. G., Wiebel, E. & Tipei, S. (1999). Data sonification and sound visualization. *Computing in science & engineering*, *1*(4), 48-58.

Kolbe, L., Tünnermann, R., & Hermann, T. (2010). Growing neural gas sonification model for interactive surfaces. In R. Bresin (Ed.), *Proceedings of the 3rd Interactive Sonification Workshop* (pp. 47–50). KTH, Stockholm, Sweden: KTH School of Computer Science and Communication (CSC).

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., & Neuhoff, J. (1999). Sonification report: Status of the field and research agenda. Tech. Rep., International Community for Auditory Display.

Kolbe, L., Tünnermann, R., & Hermann, T. (2010). Growing neural gas sonification model for interactive surfaces. In *Human Interaction with Auditory Displays–Proceedings of the Interactive Sonification Workshop* (pp. 47-50).

Laennec, RTH & Forbes, SJ. (1830) *A treatise on the diseases of the chest and on the mediate auscultation. :* S. Wood & Sons; 1830.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and virtual environments*, 10(3), 282-297.

Lokki, T., & Grohn, M. (2005). Navigation with auditory cues in a virtual environment. *Multimedia, IEEE*, *12*(2), 80-86.

McGookin, D. & Brewster, S. (2011). Earcons. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 337 - 361). Logos Publidhing House, Berlin, Germany.

Mezrich, J., Frysinger, S., & Slivjanovski, R. (1984). Dynamic representation of multivariate time series data, *Journal of the American Statistical Association*, 79, 34-40.

Nesbitt, K. V., and Barrass, S. (2002). Evaluation of a multimodal sonification and visualisation of depth of market stock data. In *Proceedings of the 18th International Conference on Auditory Display*, Kyoto, Japan.

Pauletto, S. & Hunt, A. (2009). Interactive sonification of complex data. *Int. J. Hum.-Comput. Stud.* 6, 7(11): 923-933.

Polli, A. (2005). Atmospherics/weather works: A spatialized meteorological data sonification project. *Leonardo*, 38(1), 31–36.

Scaletti, C., & Craig, A. B. (1991). Using sound to extract meaning from complex data. In *Electronic Imaging'91, San Jose, CA* (pp. 207-219). International Society for Optics and Photonics.

Nadwana, A. (2012). Auditory Display of Brain Oscillatory Activity with Electroencephalography, Master thesis, UPF.

Shou, A. Y. (2012). Viability of Auditory Display Using Sonifications and Affecting Factors.

Sporns, O., Tononi, G., & Kötter, R. (2005). The Human Connectome: A Structural Description of the Human Brain. *PLoS Comput Biol*, 1(4), e42. Public Library of Science. doi:10.1371/journal.pcbi.0010042

Walker, B. N. & Nees, M. A. (2011). Theory of Sonification. In: Hermann T., Hunt A., Neuhoff J.G., (Eds.), *The Sonification Handbook* (pp. 7-32). Logos Publidhing House, Berlin, Germany.

Walker, B. N., Mauney, L. M., & Godfrey, J. (2004). The Audio Abacus: Representing a Wide Range of Values with Accuracy and Precision. In *Proceedings of the 10th International Conference on Auditory Display*, Sydney, Australia.

Walker, B.N. & Mauney, L.M. (2004). Individual Differences, Cognitive Abilities, and the Interpretation of Auditory Graphs. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, Limerick, Ireland

Walker, B.N & Cothran, J.T, (2003). Sonification sandbox: A graphical toolkit for auditory graphs. In *Proceedings of the 9th International Conference on Auditory Display*, (pp. 161–163), Boston, USA,.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211.

Walker, B. N. & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. In *Proceedings of the 2001 International Conference on Auditory Display*, (pp. 90-94). Espoo, Finland.

Weinberg, G. & Thatcher, T. (2006, June). Interactive sonification of neural activity. In *Proceedings of the 2006 conference on New interfaces for musical expression* (pp. 246-249). IRCAM—Centre Pompidou.

ANNEXES

ANNEX I - Ad hoc Questionnaire for Pilot Experiment

ID:

Questionnaire :

Age ____________

Gender F ☐ M ☐

Nationality ______________

Do you have any vision problems? Yes ☐ No ☐

Do you have any hearing problems? Yes ☐ No ☐

Do you have any formal or informal music training?

None ☐ Basic ☐ Intermediate ☐ Expert ☐

Do you have software knowledge on musical synthesis (e.g. audacity, garage band, pure data etc)? Mark one of the following:

1	2	3	4	5
☐	☐	☐	☐	☐
very low		basic		expert

In which hemisphere were found most of the areas?

Left ☐ Right ☐

Which of the 5 areas had the most nodes? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Which of the 5 areas had the least connections? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Note how many connections do you think the **Caudalmiddlefrontal Right** area had compared to the others? Mark one of the following:

1	2	3	4	5
☐	☐	☐	☐	☐
very few				a lot

Which of the 5 areas presented the most strength? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Note how much strength do you think the **Precuneus Left** area had compared to the others? Mark one of the following:

1	2	3	4	5
☐	☐	☐	☐	☐
low				high

Which of the 5 areas had the least nodes? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Which of the 5 areas had the most connections? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Which of the 5 areas presented the least strength? Mark one of the following:

Caudalmiddlefrontal .... ☐
Ethorinal .... ☐
Lateraloccipital .... ☐
Precuneus .... ☐
Cuneus .... ☐

Note how many nodes do you think the **Lateraloccipital Right** area had compared to the others? Mark one of the following:

1	2	3	4	5
☐	☐	☐	☐	☐
very few				a lot

ANNEX II - Engagement Questionnaire for Pilot Experiment

ID: ________

Questionnaire:

Age_______ Sex F ☐ M ☐ Nationality _____________ Occupation________________

I felt sad that my experience was over.....................................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I had a sense that I had returned from a journey......................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I would have liked the experience to continue...........................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I vividly remember some parts of the experience.......................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I'd recommend the experience to my friends.............................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I felt myself being 'drawn in'...	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I felt involved (in the displayed environment).............................	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I lost track of time..	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
I enjoyed myself	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

My experience was intense	1☐	2☐	3☐	4☐	5☐
I paid more attention to the displayed environment than I did to my own thoughts (e.g., personal preoccupations, daydreams etc.)...	1☐	2☐	3☐	4☐	5☐
I responded emotionally..	1☐	2☐	3☐	4☐	5☐
The content appealed to me...	1☐	2☐	3☐	4☐	5☐

ANNEX III - Demographics Questionnaire for Experiment 2

ID:

Questionnaire 01:

Age ___________

Gender F ☐ M ☐

Nationality _____________

Do you have normal or correct-to normal vision? Yes ☐ No ☐

Are you colour blind? Yes ☐ No ☐

Do you have normal hearing? Yes ☐ No ☐

Do you have any formal or informal music training?

None ☐ Basic ☐ Intermediate ☐ Expert ☐

Do you have software knowledge on musical synthesis (e.g. audacity, garage band, pure data etc)? Mark one of the following:

1	2	3	4	5
☐	☐	☐	☐	☐
very low		basic		expert

Do you have knowledge of neural networks?

1	2	3	4	5
☐	☐	☐	☐	☐
very low		basic		expert

ANNEX IV - Regions selected for the first session of the experiment

Session 1a			
Region	Nodes	Connections	Av. Strength
Lateraloccipital R	19	458	0.019
Fusiform R	22	203	0.0049
Superiortemporal R	28	1213	0.0251
Parsobitalis R	6	145	0.0125
Supramarginal R	16	600	0.0422
Cuneus L	8	392	0.031
Superiorparietal L	27	907	0.0404

Session 1b			
Region	Nodes	Connections	Av. Strength
Superiorparietal R	27	930	0.0441
Fusiform L	22	212	0.0103
Supramarginal L	19	554	0.0406
Cuneus R	10	351	0.0399
Lateraloccipital L	22	507	0.0261
Parsobitalis L	6	166	0.0135
Superiortemporal L	29	1050	0.0314

ANNEX V - Regions selected for the second session of the experiment

Session 2a				
Region	**Nodes**	**Connections**	**Av.**	**Parameter asked**
Middletemporal R	20	624	0.0195	Connections
Superiortemporal R	28	1213	0.0251	
Parsobitalis L	6	166	0.0135	
Parstriangularis L	7	209	0.0321	
Supramarginal L	19	554	0.0406	
Inferioparietal L	25	824	0.0473	
Precentral L	36	830	0.0424	Nodes
Postcentral L	30	864	0.0394	
Postcentral L	30	864	0.0394	
Inferioparietal L	25	824	0.0473	
Cuneus R	10	351	0.0399	
Cuneus L	8	392	0.031	
Parsobitalis L	6	166	0.0135	Av. Strength
Parstriangularis L	7	209	0.0321	
Rostralmiddlefrontal R	22	533	0.0434	
Lateralorbitofrontal R	19	348	0.0146	
Frontalpole R	2	95	0.0132	
Frontalpole L	2	40	0.0558	

Session 2b				
Region	**Nodes**	**Connections**	**Av. Strength**	**Parameter**
Parsopercularis R	10	255	0.0312	Nodes
Lateralorbitifrontal R	19	348	0.0146	
Medialorbitofrontal R	12	308	0.0293	
Lateraloritofrontal R	19	348	0.0146	
Parsopercularis L	11	270	0.0382	
Pastriangularis L	7	209	0.0321	
Supramarginal R	16	600	0.0422	Av. Strength
Superiortemporal R	28	1213	0.0251	
Lateralorbitifrontal R	19	348	0.0146	
Medialorbitofrontal R	12	308	0.0293	

Cuneus L	8	392	0.031	
Lateraloccipital L	22	507	0.0261	
Superiortemporal R	28	1213	0.0251	Connections
Postcentral R	31	850	0.0501	
Superiorparietal L	27	907	0.0404	
Postcentral L	30	864	0.0394	
Inferiortemporal L	17	226	0.0165	
Middletemporal L	19	528	0.0299	

ANNEX VI - Ad hoc scale for the measurement of the error deviation from the ranges asked in the third session (The values in grey are the ranges of values asked).

	Nodes 20-30	Nodes ~8	Connections	Connections 1000-	Av. Strength 0.03 –	Av. Strength <
1.	< 10	<5	< 100	< 700	< 0.01	**<0.01**
2.	10-15	**6-10**	100 – 200	700-800	0.01 – 0.02	0.01 -0.02
3.	15-20	11-15	**200 – 400**	800 – 900	0.02 – 0.03	0.02 – 0.03
4.	**20-30**	16-20	400 – 600	900 -1000	**0.03 – 0.04**	0.03 – 0.04
5.	30-35	21-25	600 – 800	**> 1000**	0.04 -0.05	> 0.04
6.	35-40	>25	> 800		>0.05	
7.	>40					