Universiteit Leiden
The Netherlands

Leiden Institute of
Advanced Computer Science

Master Media Technology

Pragya Jain
S3663868

Towards an Auditory Observatory

A Psychoacoustic Investigation of Gravitational Wave Data through Sonification and Spatialisation Experimental Studies

Under the supervision of:

Edwin F. van der Heide
Leiden University, The Netherlands

Wanda L. Diaz Merced
Universidad del Sagrado Corazón, Puerto Rico

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

March 3rd 2025

Abstract

The application of data sonification for Gravitational Wave (GW) analysis is expanding, yet current GW sonification models are acoustically limited in perceptibility. This limitation is addressed in this research by studying the effect of timbre by varying noise levels and the influence of masking on signal detectability through designed and executed participatory experiments. Two GW data sonification paradigms are proposed in this study. The first approach involves timbre-based GW sonification analysis, which explored the effect of timbre by varying noise levels on simulated GW signal detection in experiment 1. Combinations of 3 timbres and 6 noise levels were displayed under two conditions: the signal is either present or it is absent. Pseudo-GW150819 signals were synthesised in sonifications containing the stimuli. The assessment involved detecting its presence and absence across trials. The second approach was developed by audifying three instrumental noise categories of real GW data derived from LIGO to examine the effect of masking of concurrent noise on GW signal perception and whether spatialisation of these noise sources may contribute to sensitivity to the stimulus. The second experiment examined the effect of auditory masking through real-time interaction by defining the lowest threshold at which the GW170817 signal was perceptible alongside the concurrent display of the three specified noise categories spatialised in both monophonic and quadrophonic installations. The results from experiment 1 remained limited in identifying the timbre-NF combination that allows for signal detectability. Furthermore, observations from experiment 2 also indicated no evident effects of spatialisation on signal perceptibility and distinguishing between diverse sound sources. Although the results from the presented analysis revealed no significant effect, this study identified and reflected on the discovered limitations of the experiment setup and provided promising opportunities to facilitate sound-based GW data analysis using the proposed GW data sonification paradigms, paving the ground for further research.

Keywords: *data sonification, gravitational waves, signal-to-noise, psychoacoustics, human-data interaction, multisensorial astronomy*

Table of Contents

1. Introduction

The domain of astronomical data sonification currently remains limited for scientific exploration due to misrepresentation or inaccurate interpretations of data, potentially caused by the complexity and creativity involved in the sonification design (Quinton, M. et al., 2020), as a result, sonification of astronomical data currently remains largely limited to the purposes of outreach (Enge, K. et al., 2023). Researchers emerging in the field of astronomical data sonification are pursuing sonification studies based on techniques such as parameter mapping and model-based sonification, demonstrating that the theory of data sonification already has an articulated set of design principles. However, more strategic formulations and methodologies of data sonification, such as data-to-audio parameters, are required to examine the feasibility of employing sonification as a mode of analysis, further contributing to scientific investigations. Astronomy requires going beyond the visually stimulating ways of analysis, representation, and demonstration. By undertaking sound as an analogy for gravitational wave data, this study explores a non-visual mode of analysis and proposes a sound-based human-data interaction paradigm for gravitational wave data analysis.

Gravitational waves (GW) propagate through the universe in both temporal and spatial domains, contributing to the multivariate nature of the GW data. Similar to gravitational waves, sound is intrinsically intertwined with space and time. The multidimensional qualities of sound, such as pitch, duration, loudness, and timbre, render sonification to be a relevant methodology for multivariate data analysis (Zanella, A. et al., 2022). To implement sonification as an effective technique for data exploration and contribute to existing analysis methods, listening skills are required to be trained (Zanella, A. et al., 2022). This can provide an alternative approach for GW data analysis in interpreting and identifying signals in noisy data. Adding another experiential layer to the auditory perceptual exploration of GW data, in this research, the proposed approach employs the spatialisation of sonification designs.

Alongside the GW signal, the data consists of environmental as well as instrumental noise, which is categorised based on frequency domain and source of origin. Over 200 different categories of noise sources are characterised by the GW astronomers. Masking of the GW strain by these noise categories is one of the major limitations present in strain detection. As the auditory processing modality facilitates enhanced sensitivity to temporal characteristics compared to visual analysis (Schöpper, L. M., & Frings, C., 2023; Bizley, J. K., & Cohen, Y. E., 2013; Bregman, A. S., et al., 1990; Kramer, A. F., et al., 1994; Flowers, J. H., & Hauer, T. A., 1995; Kramer, A. F., et al., 1999), sonification of GW's time series data may facilitate recognition of diverse noise features and detection of the signal. Diaz, M. W. (2011) emphasised the development of methodologies to improve the perceptibility of the signal from noise as it is generally processed and filtered out from the data to reveal the signal, which can be transient and inconsistent as it may be masked by noise. To develop spatial-acoustic methodologies for GW, it is relevant to acoustically examine noise data signatures that are retrieved alongside the GW signal. These data features, such as varied categories of noise present in the GW data, are investigated in this study, which generally remain unexplored.

Two approaches are developed to investigate whether the GW data-specific sonification and spatialisation approach could facilitate exploration, interpretation, and identification of various GW data characteristics. The assessment is based on psychoacoustic evaluation of timbre perception and the effect of masking through a participatory, task-oriented perceptual examination. The perception experiment I is developed to address to what extent timbre perception may influence signal detectability when displayed in noise. The second approach outlined in perception experiment II uses the method of audification to develop acoustic representations of the noise categories to contextualise the GW non-signal data and investigate the effect of masking of these noise streams on GW signal. For the purpose of this study, three instrumental noise channels were chosen based on a) noise data availability, b) sample rate, and c) the (relative) location of their origin in the instrument. The study concludes with summarising the findings from the participatory experiments and reflecting on future directions of this research in the final section.

2. Background

2.1. Astronomical Data Sonification

Data Sonification is an emerging approach in the fields of astronomy and space science, due to which, its effectiveness and formulation is yet to be collectively determined, especially when intending to employ sonification as a tool for scientific exploration. Sonification tools have been developed to produce auditory representations of astronomical data through interdisciplinary collaborations (De La Vega, G., et al., 2022) that have yielded interfaces such as XSonify, Sonipy, Astronify, SonoUno HighCharts, Sonifyer, Sonification Sandbox, MathTrax, and EarthPlus.

Astronify and SonoUno are unique applications of sonification tools that offer a user-centred approach to astronomical data sonification. Astronify is Python-based software that enables the sonification of astronomical data, including time-series data such as light curves, stellar flares, and exoplanet transients. The application operates by associating data points to auditory parameters to produce sonification (Grond, F., & Berger, J., 2011) through parameter mapping technique in which one column from the data table is mapped to time and other column to pitch. In addition, Astronify also provides a simulation to create synthetic data to sonify, enabling the investigation of the correlation between sound perception and data-to-sound parameters. Astronify projects explore data from the Kepler Space Telescope and Transiting Exoplanet Survey Satellite (TESS) missions, which are designed to discover exoplanets like Earth. The projects translate datasets from these astronomical surveys of light curves, in the sonification the light intensity is represented in brightness through changes in pitch, in which a higher pitch denotes more light, and a deeper pitch represents less light.
SonoUno is another tool used to produce astronomical data sonification and is designed with a user-centred approach and an accessible interface for its human-computer interaction (HCI) based design (Casado, J., et al., 2019). Like Astronify, SonoUno is also developed with Python and enables modular design synthesis. It allows for the uploading of txt/csv extensions and sonify plotted data. It produces sonification on pitch variation, volume, and timbre. SonoUno is a suitable sonification tool for beginners, providing an extensive user manual for sonification of astronomical data with SonoUno software. This application-based projects include the sonification of variable stars, galaxies, and gravitational waves.

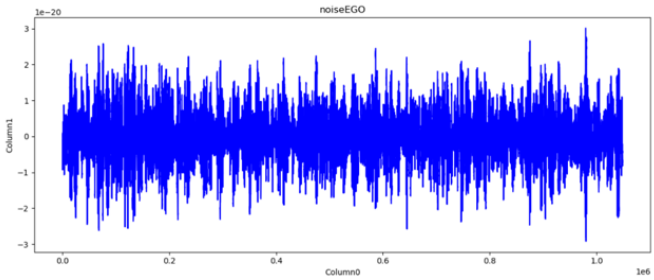


Fig. 1: Plot of noise pattern caused due to instrumental and environmental noise, by European Gravitational Observatory (EGO). Sonification of this graph (a sample) can be found here.

The interferometers detecting the gravitational waves are affected by noise, masking the detection of real astrophysical signals, figure 1 showcases graph of this noise pattern in the sonification application. Transient noise can affect the data quality, researchers have devised this instrumental and environmental noise as glitches as described in section 2.2.C. Figure 2 showcases graph of Glitch 1126409678.84375. SonoUno is part of the REINFORCE project, LIGO, in the framework of citizen science initiatives alongside GravitySpy, stimulating citizen scientists to classify and catalogue these glitches. The sonification of the glitch is conducted with both unprocessed and whitened data. It is interesting to compare both sonification samples as the gravitational waves in the raw dataset appear to be propagating slower than the cleaned data. In the whitened sample, the first dip in the pitch is more apparent than in the sonification of the raw data.

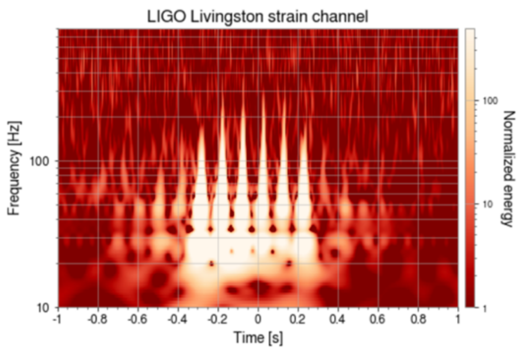


Fig 2: Graph of the Glitch 1126409678.84375, (Frequency and normalised energy vs time). Visualisation and sonification can be found here.

Further investigation is required to sonically assess the differences, similarities, and correlations between raw data and processed data. Sonification is especially useful for interpreting any measurement that changes over time, therefore, research into the sonification of gravitational waves can offer an interesting contribution to the analysis of noise-signal ratios present in data derived from LIGO and Virgo observatories. Sound experts in their own respective domains emphasise taking into consideration sound perception, sound design, psychoacoustics, and experimental psychology to develop well-informed methodologies and design processes to enable a sense of grounding to develop relevant and reliable sonification interfaces for sound-based space data exploration. As the existing sonification approaches to GW data exploration remain limited in perceptibility and quasi-scientific due to the lack of empirical insights in the domain of human psychophysics demonstrating effects on perceptibility, augmentation, as well as sensitivity to the stimulus/signal detection in (concurrent) noise are yet to be formed, this study intends to outline an evaluation of the psychoacoustic effects and their influence on identification and distinguishing between GW data features.

2.2. Gravitational Wave Case Study

2.2.1. Brief introduction to Gravitational waves

GW are caused by compact concentrations of energy, for example, black holes and neutron stars, which travel at the speed of light, immensely warping spacetime upon their shape change as they propagate through the universe (Einstein, A., 1916). While electromagnetic waves are the oscillations of the electromagnetic field, gravitational waves are the oscillations of the 'fabric' of spacetime (Thorne, K. S., 1995). After a century of being predicted in 1916 by Albert Einstein in the theory of relativity, the first direct evidence of these waves was detected on September 14, 2015, by the two detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) in Livingston and Hanford, in USA. Where a signal of a transient gravitational wave was detected with a frequency progressing from 30-250 Hz (Abbott, B. P., et al., 2016) the detected signal GW150914 presented features of inspiral and merger of a black hole binary system and their ringdown resulting in a single black hole, exemplifying Einstein's prediction of the merger of two blackholes. This detection posed a great revolution in the development of the field of gravitational-wave astronomy and cosmology at large. Further research into GW may continue to offer novel insights into the mysteries of our universe and its origin.

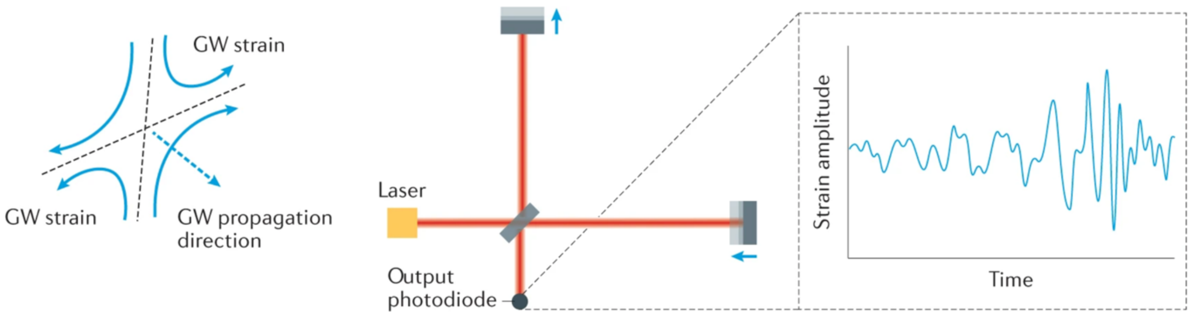


Fig. 3: A simplified illustration of a Michelson Interferometer showcasing the gravitational wave propagation trajectory. As the strain reaches the interferometer one arm squeezes and the stretches resulting in phase change of the laser beam reaching the photodiode, the detected light intensity is then used to reconstruct the GW strain. As derived from Bailes, M., et, al (2021) [accessed via https://doi.org/10.1038/s42254-021-00303-8].

The absence of technology had hindered the ability to gain a more successful detection of gravitational waves, which are ever-present and transient in nature, travel through spacetime, and are extremely faint by the time they arrive at earth with weak signals (Ishak, B., 2018). These waves are detected through a modified Michelson interferometer, whose optical assembly (as illustrated in figure 3), typically consists of two mirrors with highly reflective coating, a beam splitter, a point source of laser, and a screen/photon detector. The interferometer operates as the beam of the laser is projected onto the beam splitter, which splits the beam in half (50/50), enabling one half of the beam to orthogonally reflect from one mirror and the other half to reflect from the other mirror. As the reflected beams recombine in the beam splitter, they project interference patterns that can be studied to determine whether the beams interfere constructively or destructively. When both light beams return to the beam splitter and transmit an optical signal proportionate to the gravitational-wave strain to the output photodetector, the interference pattern and arm lengths shift, indicating the presence of gravitational-wave signals, which are measured by complex augmentation of a Michelson interferometer (Abbott, B. P., et al., 2016). The GW signal originating from the binary black hole merger or neutron stars collisions is detected based on measuring the squeezing and stretching of the interferometer arms (Nuttall, L. K., 2018) the gravitational wave alters the shape of the interferometer's arms as it warps the fabric of space and time!

2.2.2. Gravitational Wave Strain

One of the methods employed to confirm gravitational wave signals as devised by LIGO & Virgo observatories is mapping the strain signature in inquiry with the GW strain template, a simulated GW strain waveform defined by the mass and spin of each stellar object in the binary system during ringdown. The total mass of the binary system is defined as ''chirp mass'', the constant in general relativity characterising the gravitational wave signal, further discussion can be found in B P Abbott et al (2018). The sonifications of the strain are generating the chirp sound as the chirp mass of the binary system is sonified. In the case of the first GW detection, the GW strain was confirmed when the detections from both LIGO interferometers were compared and the time-around the event detection coincides under 15ms, meaning the waves travel from one interferometer to the other under this time window (Usman, S. A., et al., 2016). This method was employed to detect a compact binary coalescence (CBC) such as GW150914 signal as described in figure 4. GW150914 signal was first detected in Livingston observatory and 6.9ms after in Hansford by mapping the strain signature with the GW strain template. The strains detected from both observatories were compared with the strain template of a CBC waveform to confirm the event as a GW event.

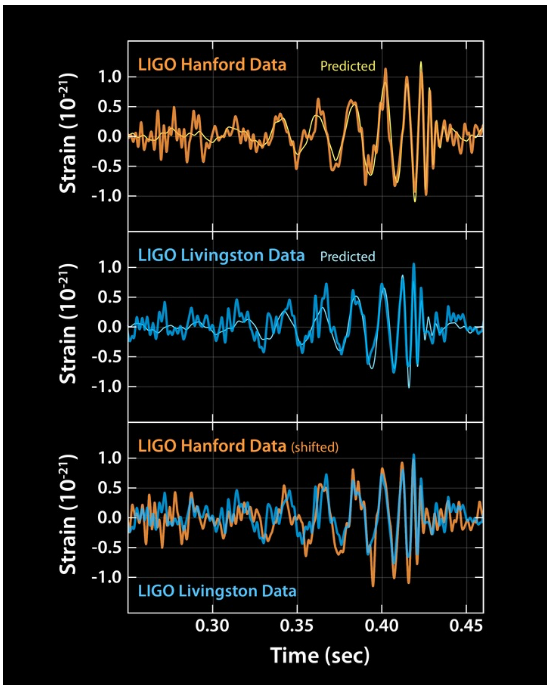


Fig. 4: Series of three plots depicting GW150914 signal as a function of strain over time. The first two plots showcase the GW150914 signal along with the GW template derived from Albert Einstein's equations predicting a binary black hole merger in the theory of relativity. The third plot showcases a comparison

between both observatories' detection superimposed. (CalTech, MIT, LIGO Lab: https://www.ligo.caltech.edu/image/ligo20160211a. accessed on_6th sep'24).

2.2.3. Gravitational Wave Data: Detector Noise Characterisation

The boundary between 'noise' and 'signal' in GW data is determined by GW community as the gravitational wave strain being the 'signal' and the instrumental as well as environmental information/distortion as 'noise'. Data from ground-based observatories possess unavoidable noise features such as thermal noise from detectors, erratic atmospheric conditions or unidentified origin. Data detected from unknown sources may potentially be astrophysical signatures, however the adoption of what constitutes as 'noise' or 'signal' is dependent on the objective (Czesla, S., et al., 2018). The detector noise is one of the major limiting factors in detecting transient GW signals. In Razzano, M., & Cuoco, E. (2018), more detailed specifications are presented as seismic noise being the largest contributor at low frequencies below 10Hz, thermal noise as a significant feature at higher frequencies originating from subsystems like suspension of mirrors and at frequencies higher than 200Hz is the quantum (shot) noise from the laser beam. Such 'disturbances or non-astrophysical contributions from the interferometers' subsystems and atmospheric anomalies are defined as 'glitches', which can mimic the waveform of an astronomical event, therefore the auxiliary channels are required to be studied from each detector to derive relevant information regarding the atmosphere of the interferometer. Furthermore, Razzano, M., & Cuoco, E. (2018) and Zevin, M., et. al., (2024) pose value in thorough investigation of the glitches and develop significant classifications of their frequency-time domain in order to produce a collection of these glitches conforming of similar morphologies.
To explore the dominant effect of noise on the GW strain, both 'noise' and 'signal' features of the data are essential to consider in developing sound-based GW data analysis methods. Instrumental (stationary) noise is studied in this research, as it is one of the most dominating noise categories. To perceive signal from noise as well as interpret varies noise categories through aural perception, spatialised sonification approaches are methodologically outlined in the following sections.

3. Perception Experiment I

3.1. Research question

To what extent does the timbre used for sonifying simulated gravitational wave data influence signal detectability in the context of noise?
In order to address the first question examining the extent to which the timbre used for sonifying simulated gravitational wave data can influence signal detectability in noise, a timbre-based GW simulation-sonification (TGSS) approach is developed.

3.2. Methodology

Frequency Modulation (FM) synthesis is chosen for producing the timbre-based simulated GW data sonifications, as it enables synthesis of complex timbres by using a limited number of oscillators. The FM algorithm is used for the proposed TGSS paradigm uses two oscillators: a carrier and a modulator. The carrier frequency varies based on the modulation signal. The number of harmonics is influenced by the modulation depth; if it increases, the number of overtones around the carrier frequency rises, producing more complex sound. The carrier-modulator frequency ratios contribute to the harmonic and inharmonic spectra. The ratios are chosen on a perceptual basis, while the base carrier frequency, modulation signal, and the modulation depth are defined individually for each synthesised timbre in the algorithm. A schematic of the algorithm is shown in figure 5, depicting modulating and carrier oscillators, the carrier-modulator frequency ratio, and the white noise and simulated GW wave operators, which are further described in this section.

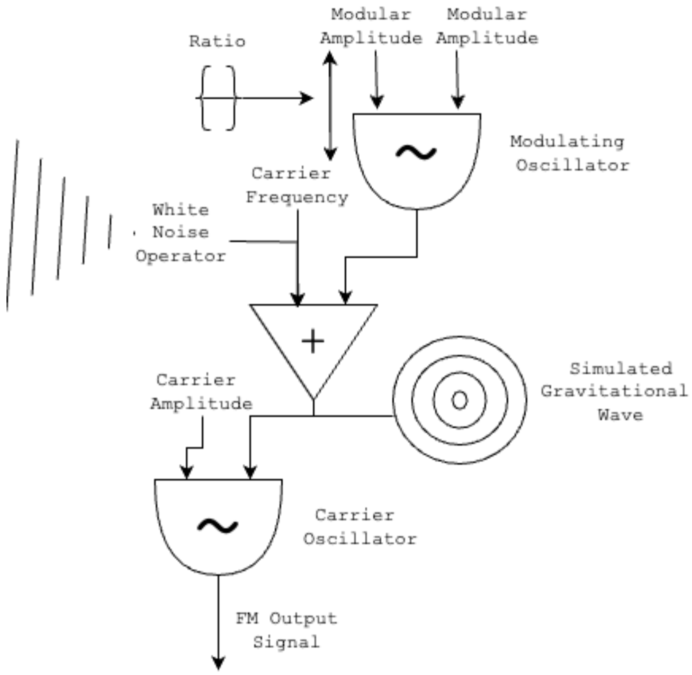


Figure 5: A schematic of Frequency modulation technique with embedded simulated GW signal and white noise operators assembling the complete algorithm for sonifications developed for Experiment I.

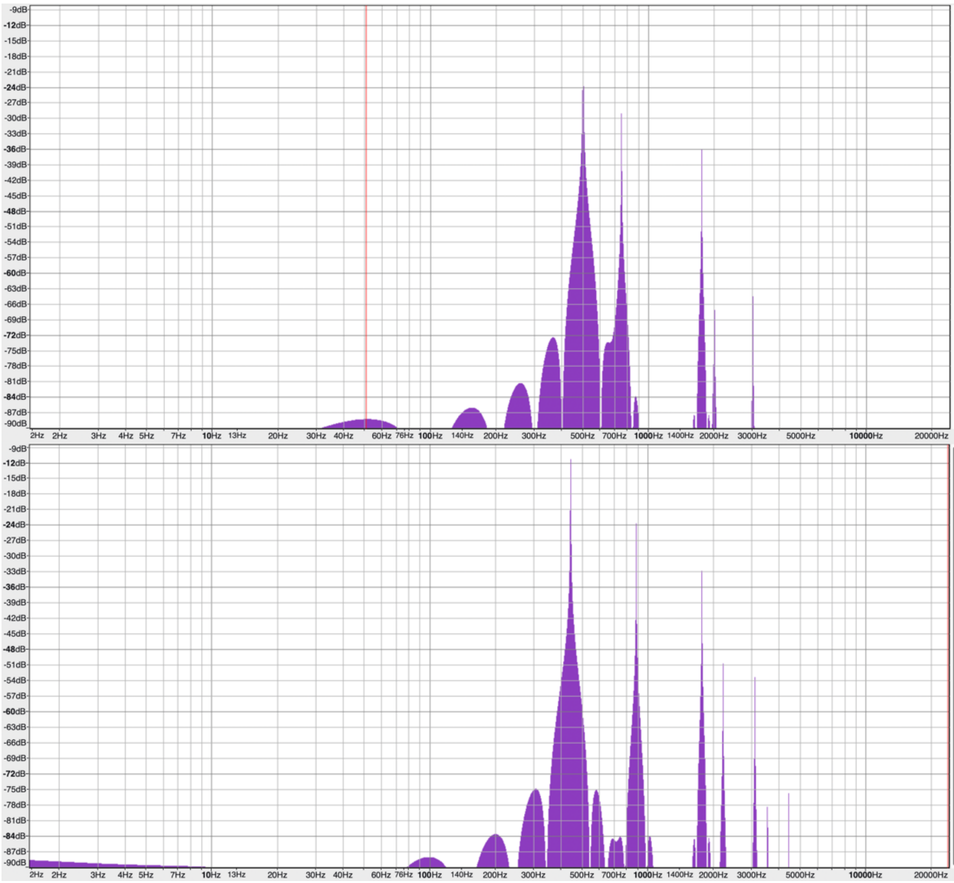
3.2.1. Timbres

Timbres of instruments such as bells (percussive), violin (string), and horn (brass) are selected to develop the sonifications to assess which timbre(s) result in a higher detectability rate when displayed in a noise context. These chosen timbres are synthesised due to their variability in harmonic content, covering a brief yet diverse timbral range, allowing for a comparative analysis in this pilot study to explore the extent to which timbre influences signal detectability in noise. The GW simulated sonifications involve bells' inharmonic while violins' and horns' harmonic timbres. FM-based synthesis for bells-timbre incorporates 500Hz as the carrier frequency and 1250Hz as the modulation frequency, resulting in a 2.5 ratio, and 450 as the modulation depth. For violin-timbre synthesis, 440 Hz is defined as the carrier frequency, and 1320Hz as the modulation frequency, with 3 as the resulting ratio and 900 as the modulation depth. Horn-timbre synthesis involves a carrier frequency of 200 Hz and a modulation frequency of 100 Hz, producing a ratio of 0.5 and 500 modulation depth. The synthesis parameters of each timbre are summarised in Table 1, and the sonifications are made accessible here.

Instrument	Timbre	Harmonic Content	Carrier Frequency (Hz)	Modulation Frequency (Hz)	Ratio	Modulation Depth
Bells (Percurssive)	Resonant, Metallic	Inharmonic Overtones	500	1250	2.5	450
Violin (String)	Bright, Warm	Strong overtones, rich harmonics	440	1320	3	900
Horn (Brass)	Mellow	Smooth harmonics	200	100	0.5	500

Table.1: Synthesis parameters for each timbral synthesis.

Figure 6 illustrates the spectral analysis of all three timbral syntheses, with the vertical axis covering the loudness in decibels from 3 to 90dB and the horizontal axis demonstrating the frequency range from 0 to 20kHz. The first plot (from the top) represents the frequency analysis of bells-timbre synthesis, with Fast Fourier transform (FFT) analysis depicting peaks at 500Hz, 750Hz, 1750Hz, 2000Hz and 3000Hz, indicating inharmonic partials, which perceptually may be perceived as producing a metallic sound. Spectral analysis of violin-timbral synthesis is observed in the second plot, with peaks at 440Hz, 880Hz, 1760Hz, 2200Hz, 3080Hz, 3520Hz, 4400Hz. These peaks correspond to harmonic partials due to evenly spaced spectral lines over the frequency range and are acoustically perceived as bright timbre. Alternatively, the last plot depicts the horn-timbre synthesis, indicating peaks at 100Hz, 200Hz, 300Hz, 400Hz, 500Hz, 600Hz, 700Hz, 800Hz, 900Hz, 1000Hz, 1100Hz and 1200Hz, indicating periodicity resulting in harmonic overtones, and perceived as a mellow timbre.



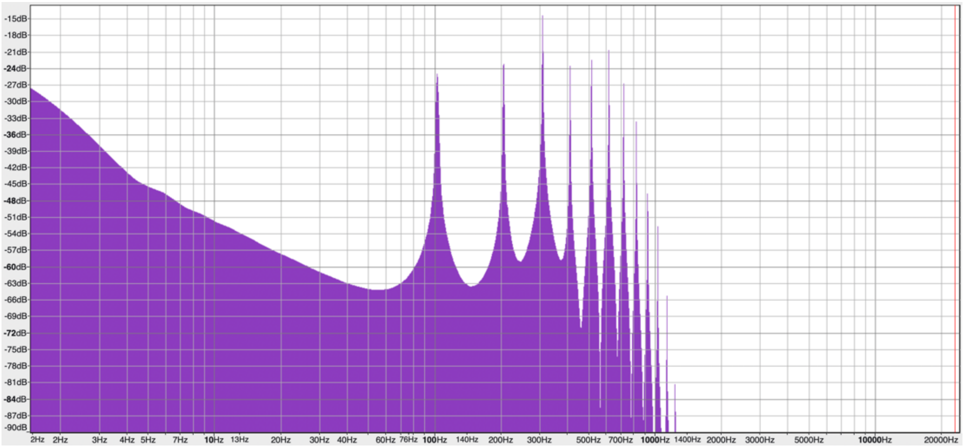


Figure 6: Three spectral analyses plots of synthesised timbres: bells, violin and horn presented respectively.

3.2.2. Noise Factors

To compare the rate of signal detectability in the context of noise, the synthesised timbres are combined with noise at different levels. We speculate that above a certain noise level, it will become impossible to detect the signal and that the threshold at which this occurs would vary for each of the three timbres. In GW signal-detection task it is crucial to optimally represent GW detector noise profile. Existing GW simulation initiatives assumed white noise as the background for GW search algorithms (Summerscales, T.Z., 2006) and has been used in GW simulation initiatives studying signal sensitivity in non-gaussian background noise. Similarly, in our paradigm, white noise is used as an operator in three separate FM algorithms, each per synthesised timbre. Noise factors (NF) 2, 3, 4, 5, 6, and 7 are displayed with all three timbres, ranging from 2 as the lowest and 7 as the highest noise level, hypothesising that under the lowest noise level, the signal would always be perceptually detectable while in higher NF like 7, it would be impossible to detect it for all three timbres.

3.2.3. GW Simulation Stimulus Generation

A pseudo-GW signal is synthesised with the developed synthetic timbres and noise level variance to generate the GW data simulation, in order to assess the extent to which auditory analysis may contribute to signal detectability. The pseudo-GW signal with a duration of 15ms, simulates the chirp sound of the GW150914 strain, which varies in frequency over time. Figure 7 illustrates the spectral analysis of the generated signal, demonstrating variance in its frequency characteristics with the highest intensity observed at 300Hz and lowest at 3000Hz.

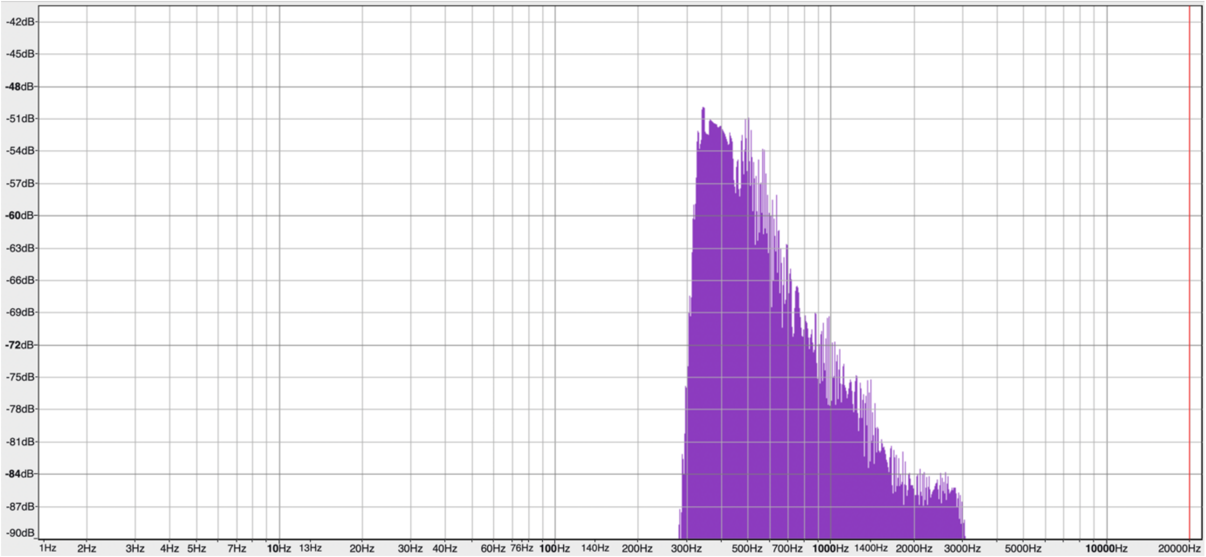


Figure 7: Spectral analysis of generated GW signal.

Each sonification is presented as a trial and is displayed for a period of 1000 ms. The GW stimuli are programmed to occur at different points in the duration of the sonification, meaning the signals' location are randomised over time to maintain the unpredictability factor for the participants and support them in focusing on stimulus detection in each trial rather than primarily during a specific interval period. This is achieved by using the *random* function in the algorithm, which outputs values within a given range at equal probability.

3.2.4. Experiment Description

In the experiment design, the sonifications are characterised in 3 blocks: block I, bells-based synthesis; block II, violin-based synthesis; and block III, horn-based synthesis. A total of 36 sonifications (trials) are presented, with each block consisting of 12 trials. The NFs and the presence of the simulated GW signal are randomised across trials, meaning NFs 2, 3, 4, 5, 6, and 7 are displayed in combination with conditions 1) signal present and 2) signal absent as summarised in table 2. Every NF occurs twice per block and is displayed in both conditions. An uninterrupted recording of 2,21 minutes (accessible here) will be played in a stereo installation, displaying sonifications from three blocks in the following order: first bells, second violin, and third horn-based timbres, presented with the six defined noise levels. The variations in sonifications occur across blocks with varying timbres, whereas within the blocks, trials differ in whether the signal is present or not, as well as alternating NFs. The order of the presented timbres, volume of all trials, and the signals remain constant across blocks.

	NOISE FACTORS					
TIMBRES	2	3	4	5	6	7
Block I: Bells	SIGNAL PRESENT/ SIGNAL ABSENT					
Block II: Violin						
Block III: Horn						

Table.2: Timbres and noise levels configuration categorised in three blocks for Experiment I.

The experiment design of the proposed TGSS paradigm incorporates stereophonic presentation of the developed sonifications as depicted in figure 8. This experiment's setup explores whether spatialisation may have an effect and potentially contributes to signal detection. The beginning of the sonifications start from channel 1 (left side of the listener) and end at channel 2 (right side) (Bonebright, T. L., 2001; Brown, L. M., 2003; Nees, M. A., & Walker, B. N., 2009), creating a panning effect.

The defined sound parameters in the experiment design and their contributions to the auditory perceptibility of the signal in noise will be tested through participatory-experimental analysis using the developed TGSS paradigm. The experiment will begin with participants reviewing the information sheet, followed by the completion and signing of the consent form. Each participant will be screened based on age range and whether they had training in sound/music and/or astronomical data processing and will record/report their responses on the consent form. These parameters are taken into consideration to assess whether age and experience in sound and/or astronomical data processing may have a potential effect on signal detectability. Variables such as audiometry test results, participants' neurodivergent conditions as well as any limitation in their hearing such as tinnitus, were intentionally excluded and not processed during the experiment run for this pilot study. However, they are acknowledged as significant factors influencing the data and are aimed to be included in future iterations of this research.

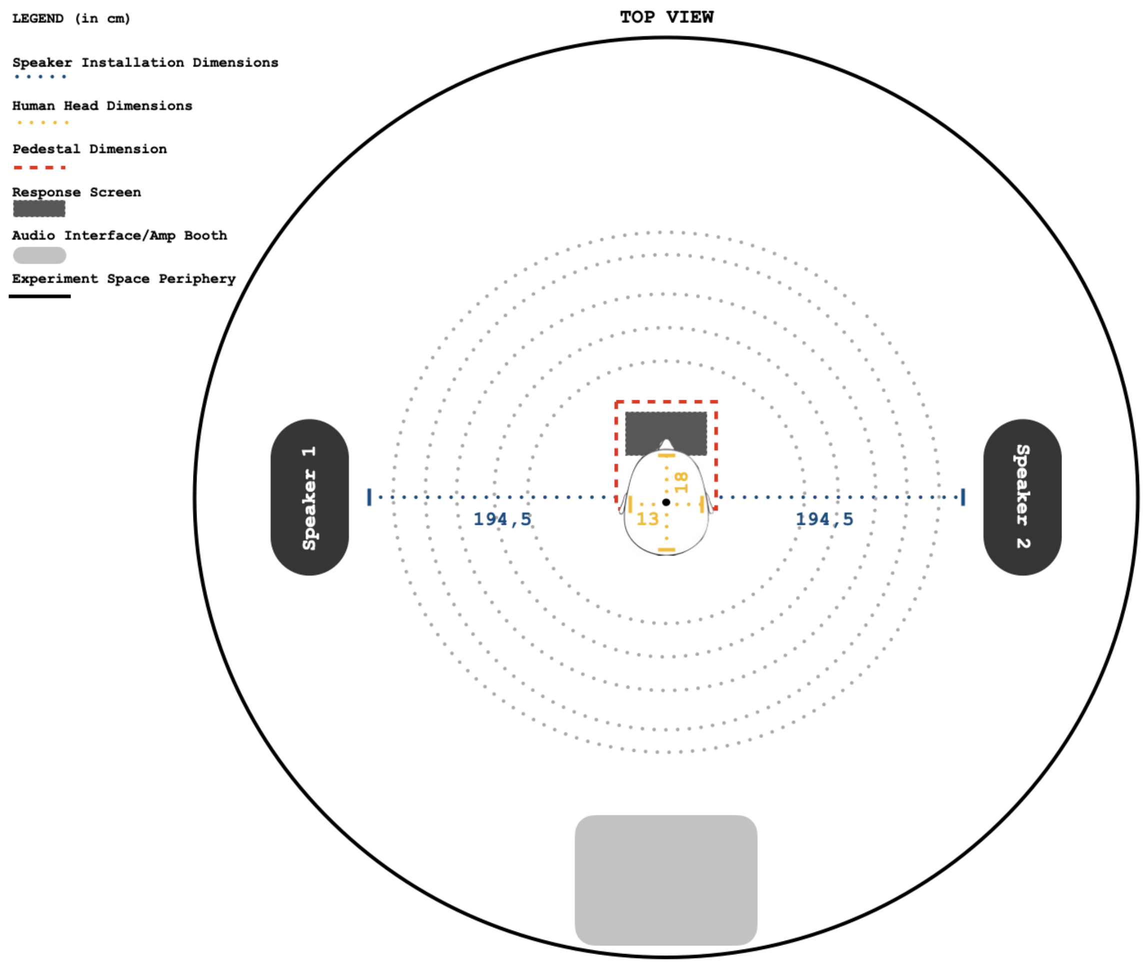


Fig 8: A schematic of stereophonic installation for Experiment I. The two-dimensional representation depicts an average adult human's head in the centre of the circle (the experiment space), a response screen, and the distance between the participant'.

Participants' performance will be measurement based on correct response rate, which will be calculated across conditions varying in timbres, noise levels, and whether the signal is present or absent. Participants will be asked to select 'YES' if perceived the signal and 'NO' if otherwise. They score 1 for a correct answer or 0 for an incorrect answer.

3.2.5. Analysis Method

The binary-response data collected from the participatory experiment of the TGSS paradigm will be analysed using the true positive, false positive and false negative scores, which measure occurrences of correct and incorrect signal detections. This analysis can also highlight potential trends (participants detecting signals in a specific timbre more than others) and biases (a listener selecting a response too often). To conduct the analysis, we will calculate:

- True Positive (TP) – Correct response 'YES' when the signal is present
- False Negative (FN) – Incorrect response 'NO' when the signal is present
- False Positive (FP) – Incorrect response 'YES' when the signal is absent

Based on confusion matrices developed with these measures, the F-scores will be calculated in order to examine whether this experiment has recorded more true signal detections and minimum false alarms and assess the effectiveness of the proposed approach. F-score is a harmonic mean of *precision* and *recall* matrices reflecting on the proportion of YES responses that are actually correct (precision) and the proportion of signal-present trials that were correctly identified as YES (recall). This measure ranges between 0 and 1, with (values closer to) 0 indicating poor detection (FN/FP>TP) and 1 (potential detection TP>FN/FP). We speculate that detectability will be higher for lower NFs with a high signal-to-noise ratio (SNR) and lower in higher NFs (with a low SNR). This will be indicated by comparatively greater F-scores in timbre-NF combinations with lower noise levels, meaning that the closer the score is to 1, the more successful the timbre-NF combination would be in allowing for signal detectability, further suggesting that some participants started to detect the signal. Alternatively, timbre-NF combinations with higher noise levels will record lower F-scores, indicating poor to no detectability of the signal. With the results derived from this analysis, we will be able to obtain the F-score–NF curves for all timbres and determine to what extent participants successfully detected the signal as well as examine the effects of timbre-NF combinations on signal detectability. These F-score–NF curves are expected to be different for each timbre. The analysis will be performed by examining all participants' responses as well as two subgroups, which will include a comparative analysis of musically trained participants and nonmusicians as well as an examination of signal detectability threshold across age groups.

McPherson et al. (2022) evidenced that detectability is higher for harmonic stimuli than for inharmonic by using pitch detection to examine the perceptibility of harmonic and inharmonic signals in noise. Their study incorporated white Gaussian noise as the background, based on which they also demonstrated that harmonicity facilitates the discrimination of sounds in noise. In contrast with their study, we incorporated white noise, and following their results, we anticipate that signal perceptibility will be higher in harmonic timbres of instruments such as violin and horn. With the outcomes, we aim to determine whether, which, and to what extent the timbres used in this investigation effect signal detectability when displayed in the specified noise levels.

Initially, the discriminative index d', which is a measure of correct identification of stimulus in both conditions, was calculated by deducting the z-score of FPR from the z-score of TPR per timbre and NF to compute the accuracy rate (d')—NF curves. However, given the binary classification nature of our data, the d' was determined to be ineffective for this experiment's analysis. As a result, the F-score measure was implemented following the critics/reviewers' recommendation.

3.3. Experiment Setup

The experiment was conducted in a publicly accessible space of Het Nieuwe Instituute's -1 Lab, allowing for a diverse range of participants across ages and professions to contribute. As depicted in figure 9, the experiments were held inside the glass cylindrical space with a diameter of 435 cm and a height of 310 cm. A UV light moderately illuminated the space during the experiments, while the rest of the space was darkened to minimise hindereance due to external visual stimulation for the participants during the experiment. The sound installation setup included four Cornered Audio-C5 speakers, a DAP 4150 amplifier, and a Cymatic LP16 audio interface. The interface was connected via a USB-C cable to a MacBook laptop running the Plug Data patch in real-time during the experiments. The speakers were mounted on the cylindrical space's ceiling and were oriented towards the centre of the glass circle. The participants were standing in the middle of the space, at the 'sweet spot,' and recording their responses on a laptop placed on a relatively waist-height pedestal of 75 cm.

Figure 9: Cylindrical glass space where the experiments were held.

3.4. Participants

The experiment was conducted with a total of 37 unspecialised participants; however, one participant modified their responses in the digital form, therefore, their data was omitted from the analysis, and the final number of participants for this study is therefore, n = 36. The participant population (M_{age} = 28.7) was separated into two groups: group 1 (n = 23) consisted of participants aged between 18 and 28 years, while group 2 (n = 13) included participants from 29 to 63 years. In our sample group, 44.5% of the population was recorded to be musically trained (n = 16) and are identified in this study as 'musicians', while non-musically trained participants (n = 20) are referred to as 'nonmusicians'. Since only 8.3% of the participants' population reported to have training in astronomical data processing, the assessment of whether their experience has any effect on signal detectability was removed from the analysis due to a limited population sample. None of the participants reported having undergone a comparable experiment previously. All the experiments in this study followed the MSc Media Technology framework for research ethics and were approved under the code FWN2022-008 by the Leiden University Faculty of Science Ethics Review Committees.

3.5. Results

In addition to the designed sonifications, an external wideband noise source from the DAP amplifier (located inside the cylindrical space) and noise sources originating from an air duct in the ceiling of the experiment space acoustically contributed to the signal detection task. Additionally, excessive reverberation was one of the acoustic properties of the glass cylindrical space that may have significantly influenced stimuli perceptibility. Furthermore, the spatialisation feature of the model was misrepresented following an error, as only the beginnings of trials were displayed in accordance with the sound design, while the end was faded out, resulting in obscured signals at the end of the trials. Due to this inaccuracy, 5 trials (Block I = 1 trial, Block II and Block III = 2 trials each) with stimuli appearing at the end of

the sonification were less perceptive and detectable than the remaining 31 trials with signals occurring at the beginning or in the middle of the sonifications. Signals occurring over 1000ms of the sonification defined whether they are in the beginning (t_b= 0 to 333ms), in the middle (t_m = 333 to 666ms) or at the end (t_e = 666 to 1000ms).
These limitations of the experimental setup were discovered after the experiment run was completed. The results reported in this section are presented considering these limitations. To account for discrepancies due to the misrepresented trials', the response rates observed in these conditions are indicated in the analysis. Since most of the trials were appropriately and consistently presented to all participants, this pilot study may still be able to draw preliminary inferences from the derived results.

An overall analysis of accuracy was performed to examine all participants' responses in bells, violin, and horn timbres across all noise factors in both conditions. The following results summarise accuracy rates under condition 1. Figure 10 displays the distribution of correct and incorrect responses from participants in the chosen timbres across all noise factors. The analysis for bell timbre recorded a 30% and 51% accuracy rates in NFs 2 and 3, respectively. For NF4, participants scored 16%, whereas in NF5 an accuracy of 24% was recorded. Higher NFs, such as NF6, observed 13.5% and NF7 8% accuracy. Violin timbre recorded a 16% accuracy rate in NF2, while both NFs3 and 5 recorded an accuracy of 13.5%. NF observed only 5% accuracy. Whereas NF6 recorded a 19% correct response rate and NF7 scored 8%. Horn timbre's results indicate 43% accuracy in NF2 and 24% in NF3. While NFs4 and 6 scored equal rates of 11% accuracy. NF5 obtained 22% detectability and NF7 27%.
The results suggest that overall accuracy rates remained minimal under condition 1, with 51% recorded as the highest detectability in the bell-NF3 combination, while responses across all timbres and noise levels remained under 50%. This indicates that the signal detection remained limited in most of the trials. Furthermore, no significant effect of varying noise factors is observed as the accuracy rates in all timbres were distributed in the low range of 5% to 51%.

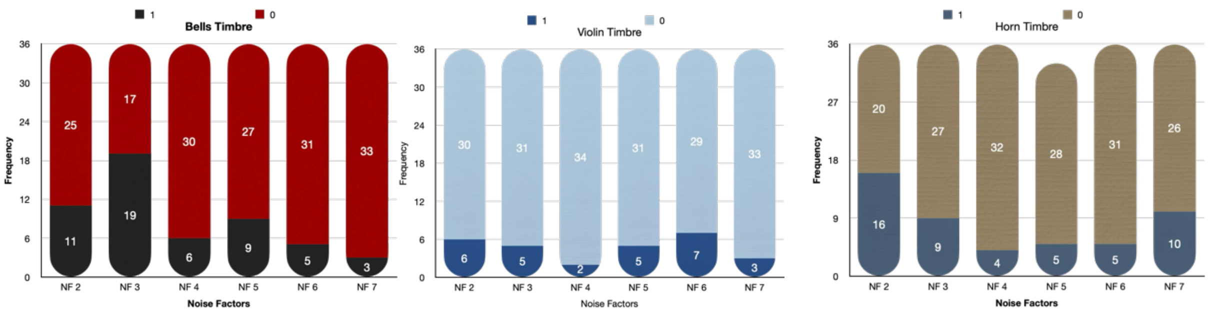


Figure 10: Frequency distribution of all participants' responses under condition 1 in bells, violin and horn across all noise factors.

The following analysis demonstrates the results of participants' responses under condition 2 in all timbres across all 6 noise factors. As illustrated in figure 11, bells timbre in NF2 recorded an accuracy of 57% and in NF3, 59%. NF4 obtained 73% detectability, while NF5 recorded 84%, followed by NF6 at 78% accuracy and NF7 recorded 99%. Participants' responses in violin timbre scored 62% in both highest and lowest noise levels NF2 and NF7. Equal accuracy rate of 84% was recorded in both NF3 and NF5 as well. Whereas in NF4 participants scored 73%, and 76% in NF6. Comparatively in horn timbre, NF2 recorded 55%, and NF3 70%. In NF4, a rate of 67.5 % is observed whereas 73% is recorded in NF5. Both NFs6 and NF 7 scored equal detectability of 84%. The results from condition 2 in all three timbres across all noise levels remain higher than 50%, as the accuracy ranges from 55% to 99% accuracy. In bells and horn timbres, the accuracy rates increase as the noise level increases, while in the violin's timbre, it remains more variable across NFs.

Comparing both conditions, condition 2 obtained higher accuracy rates, implying that participants were relatively proper in detecting the absence of the signal compared to when the signal was present. As illustrated in figure 12, with all timbres combined, in condition 1, both NF2 and NF3 obtained an average of 25% accuracy, while NF4 scored an accuracy of 9%. NF5 recorded 14%, whereas NF6 recorded 13%, and NF7 obtained 12% detectability. Alternatively, results from condition 2 indicate an average of 13% accuracy in NF2, followed by NF3 and NF4, both recording equal detectability of 16%. While NFs5, 6, and 7 obtained an equal mean accuracy rate of 18%. Comparatively, with all timbres combined and across all noise factors, condition 1 detected an average of 16%, whereas condition 2 detected 16.5% detectability. Even though the average recorded detectability rate in condition 2 is slightly higher, limited conclusions can be obtained from this analysis, and further investigation is required to examine to what extent participants were truly detecting the signal as well as its absence and were not responding only 'NO' at each trial independently from the condition.

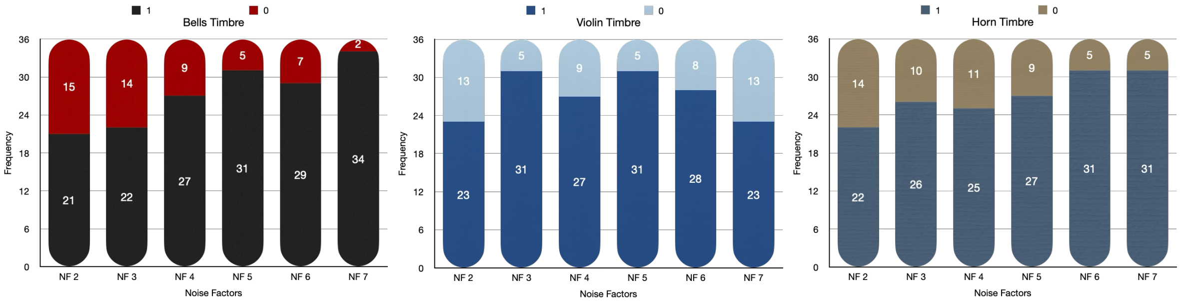


Figure 11: Frequency distribution of all participants' responses under condition 2 in bells, violin and horn across all noise factors.

Furthermore, accuracy rates with combined timbres indicate minimal effects of (varying) noise factors on signal perceptibility as indicated in figure 12. In condition 1, the highest response rate of 25% was recorded in lower NFs—NF2 and 3—followed by a significant decrease in detectability observed from NF4 onwards with responses ranging between 9% and 14% in higher noise levels, i.e., NF4– 7, indicating that few participants were able to perceive the stimuli properly in the highest SNR. Alternatively, in condition 2, the lowest average response rate is observed in the lowest NF2 (13%), followed by an incremental increase across NFs, with a slightly higher signal detection average (18%) recorded in higher NFs from NF 5 to 7, suggesting that participants were better at perceiving the absence of the signal. The results from condition 1 appear to support our hypothesis, however, outcomes from condition 2 lay in contrast as higher signal detectability is recorded in higher NFs, while perceptibility to the signal is expected to be observed in lower NFs and would incrementally decline as the NFs increase, demonstrating an inverse correlation between NFs and accuracy rates. In order to examine to what extent varying timbres as well as noise factors affected stimulus detectability, further analysis is presented in the following subsections.

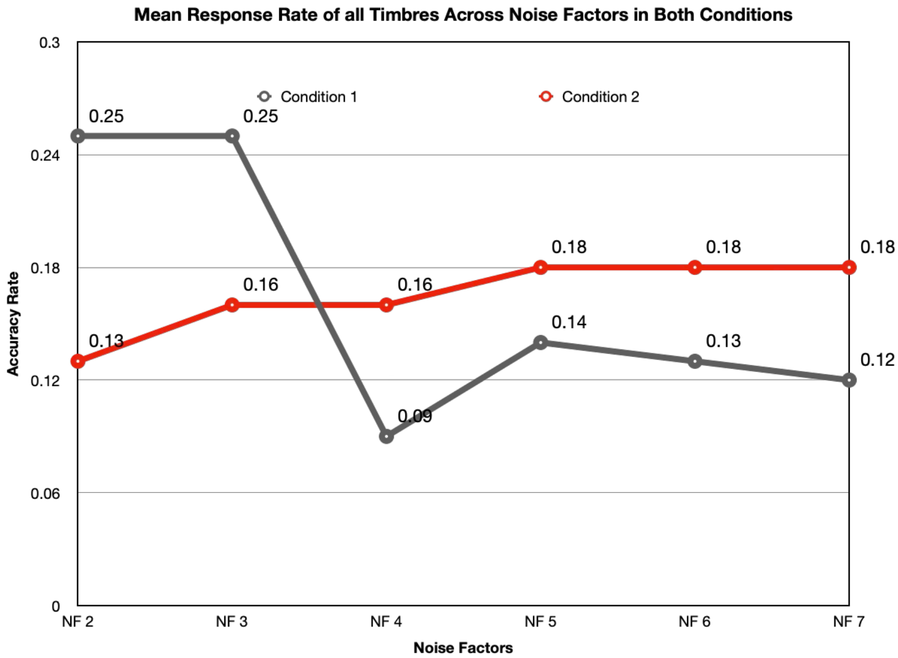


Figure 12: Average response rate of all timbres across noise factors in both conditions.

3.5.1. False Positive & False Negative Analysis | Overall Analysis

To assess the effects of timbre-NF combinations on signal detectability, accuracy was analysed using all participants' responses (n = 36). Accuracy (F-score) closer to 0 indicates poor detection and measures closer to 1 indicate potential detection in the corresponding timbre-NF pairing. Figure 13 illustrates that overall, the accuracy remained below a score of 0.5. Results from NF2 suggest that horn timbre recorded the minimal accuracy of 0.48, followed by bells timbre with 0.35 and least detectability recorded in violin (0.21). This suggests that even at the highest SNR presented during the experiment, the signal was not easily detectable across timbres. In NF3, both horn and violin timbres scored a low accuracy of 0.28 and 0.22, respectively. Alternatively, bells-timbre recorded detectability of 0.54, however this timbre-NF condition was subject to error (as described in section 3.1.5) and highlighted with a black dot in figure 13. The relatively higher accuracy rate in this pairing suggests that some participants may have responded based on expectation rather than perception. All three timbres in noise level NF4 remained below the score of 0.2, suggesting that signals were potentially undetectable in any timbres at this SNR. Comparatively, in NF5, bells-timbre achieved an accuracy of 0.46. Horn-timbre (0.29) and violin-timbre (0.22) were recorded under error conditions and slightly follow the same trend as NF3 results. In NF6, violin timbre scored 0.24, followed by horn timbre (0.2 under error), and bells acquired the lowest accuracy (0.17). In the highest NF7 and lowest SNR, horn-timbre recorded a minimal rate of 0.38, while bells and violin timbres recorded 0.14 and 0.11, respectively.

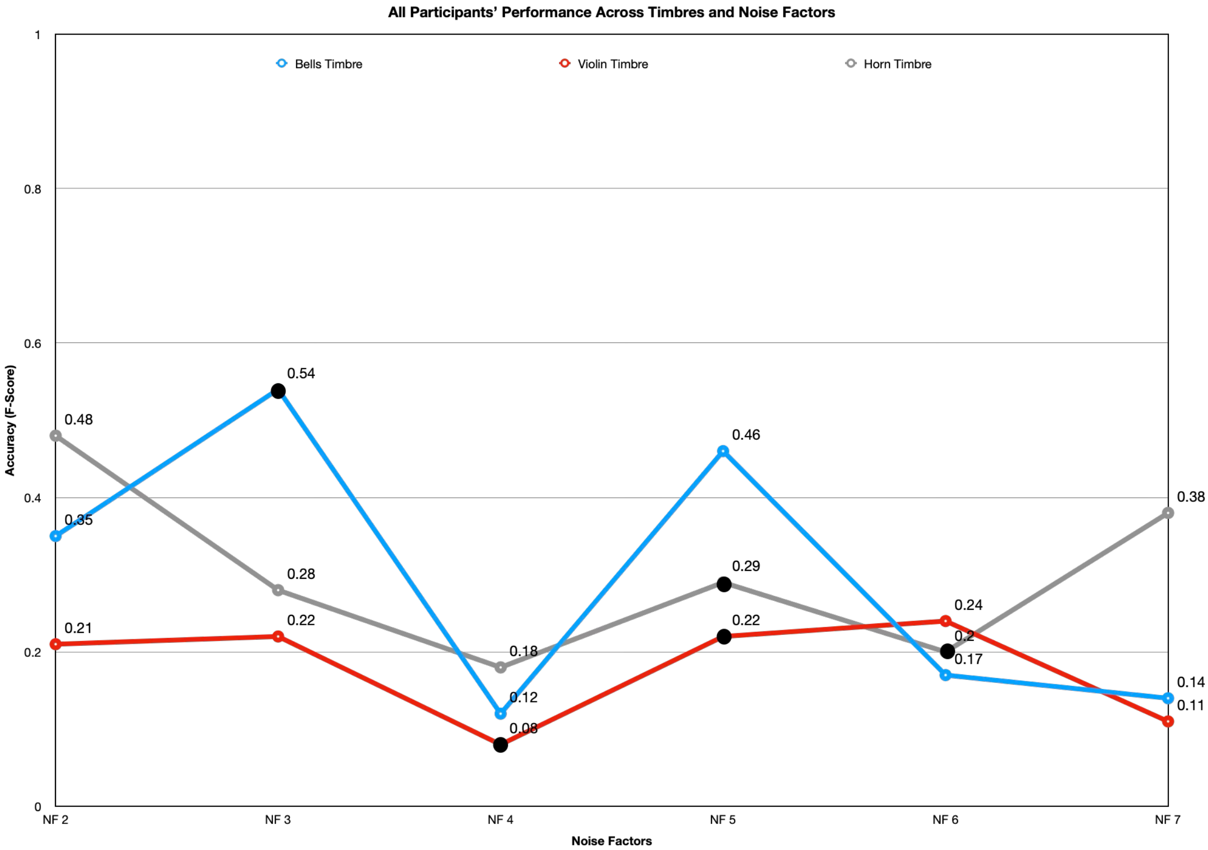


Figure 13: All participants' performance across timbres and noise factors

The outcomes summarised in figure 13 suggest that the signals remain undetectable even at low noise levels, as indicated by the results primarily centred below 0.5 accuracy. As a result, successful timbre-NF combinations that enable signal detection could not be identified. Relatively higher F-scores observed in NF5 and NF7 indicate potential detection, which remains in contrast with our hypothesis, proposing that the accuracy will be higher in lower NFs and incrementally decline as the SNR decreases and NFs increase, resulting in less signal perceptibility due to the signal masked by noise. Furthermore, given the error conditions, no conclusions may be drawn from this analysis.

3.5.2. Subgroup based Analysis | Musicians and Nonmusicians

In consideration of the limitations of the experimental setup, the results from musically trained participants (n = 16), as summarised in figure 14, indicate that horn timbre in NF2 scored 0.53, followed by timbres of bells (0.50) and violin (0.31). This suggests that musically trained participants may be slightly more sensitive to the stimuli in horn timbre at the highest SNR. In NF3, bells timbre obtained a 0.60 score (under the error condition of faded stimuli), whereas the horn timbre recorded minimal detectability with 0.38 accuracy, while no detectability was recorded in the violin (0.00). NF4 acquired the lowest accuracy across timbres, as the horn timbre obtained a score of 0.13, followed by bells (0.12) and violin (0.06, under error), implying that neither horn nor bells' timbres facilitated signal perceptibility in this SNR. Furthermore, the accuracy rate from NF4 onwards remains below the score of 0.30, indicating limited to no signal detectability. As in NF5, horn recorded 0.27, while bells (0.16) and violin (0.17) were slightly lower. Violin obtained 0.24 accuracy in NF6, horn recorded 0.13 (under error) and bells 0.06. In the lowest SNR (NF7), horn timbre acquired an accuracy of 0.33, followed by violin (0.13) and bells (0.12).

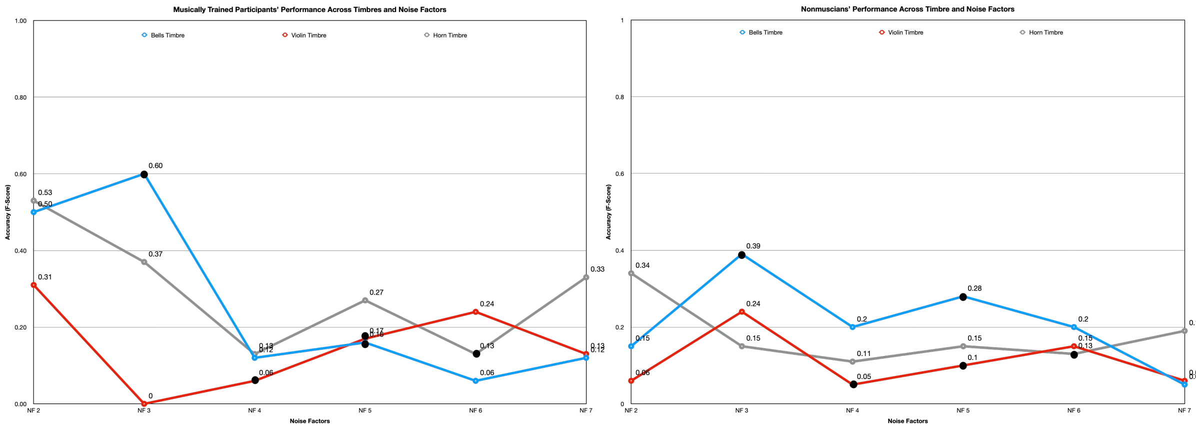


Figure 14: Musically trained participants' performance (left) and nonmusicians' detectability rate (right) across all timbres and noise factors.

Comparatively, response rates recorded from nonmusicians (n = 20) demonstrate all accuracy scores remaining below 0.40 across all timbres and noise factors. Overall, results indicate the lowest detectability in violin timbre, followed by horn and bell timbres. In the highest SNR (NF2), horn timbre recorded 0.34, while 0.15 accuracy was recorded in bells and 0.06 in violin. In NF3, bells obtained 0.39 (given faded stimuli), whereas violin scored 0.24 and horn 0.15. Bells timbre recorded 0.28 accuracy in NF4, followed by horn 0.11 and violin 0.05. The results from NF5 follow a similar trend, with bells recording 0.28 accuracy, horns 0.15 and violins 0.1. In NF6, bells obtained a low accuracy of 0.20 and violin 0.15, while horn scored 0.13 under error. Results from NF7 showcase 0.19 accuracy in horn, followed by violin (0.06) and bells (0.05). As observed in this analysis, all response rates are low, indicating limited detectability of the stimuli across timbres and noise factors for nonmusicians, suggesting that none of the timbre-NF combinations allowed for the stimuli detection to participants with no musical training.

3.5.3. Subgroup based Analysis | Age Group 1 and Age Group 2

The participant population (M_{age} = 28.7) was separated into two groups to assess whether signal detectability varies across age ranges. Group 1 (n_{G1} = 23) consisted of participants aged between 18 and 28 years, while Group 2 (n_{G2} =13) included participants from 29 to 63 years. Analysis from G1, depicted in figure 15, indicates that accuracy across timbres and noise factors remained low. In NF2, bells recorded an accuracy of 0.39, horns slightly lower at 0.38, and violin timbre obtained the lowest score of 0.19. In NF3, bells acquired a minimum accuracy of 0.5. However, this score is redundant due to an error caused in this timbre-NF condition that led to faded and perceptually undetectable stimuli. Alternatively, horn recorded an accuracy of 0.26 and violin of 0.04 at this SNR. In NF4, bells obtained a score of 0.30, whereas horns scored 0.15 and violins 0.04 (under error). Similarly, in NF5, bells recorded 0.28, followed by horn 0.14 and violin 0.13. Violin recorded an accuracy of 0.21, bells of 0.17 and horn of 0.12 in NF6. In NF7, horn recorded 0.27, bells recorded an accuracy of 0.07 and violin acquired a score of 0.05. Comparatively, higher scores are observed in lower NFs, as highlighted by both bells and horn timbre curves following the inverse correlation of accuracy rates and noise level, as hypothesised in this study. However, the analysis demonstrates below-minimum detectability threshold, indicating limited to no detection in the presented conditions for this subgroup.

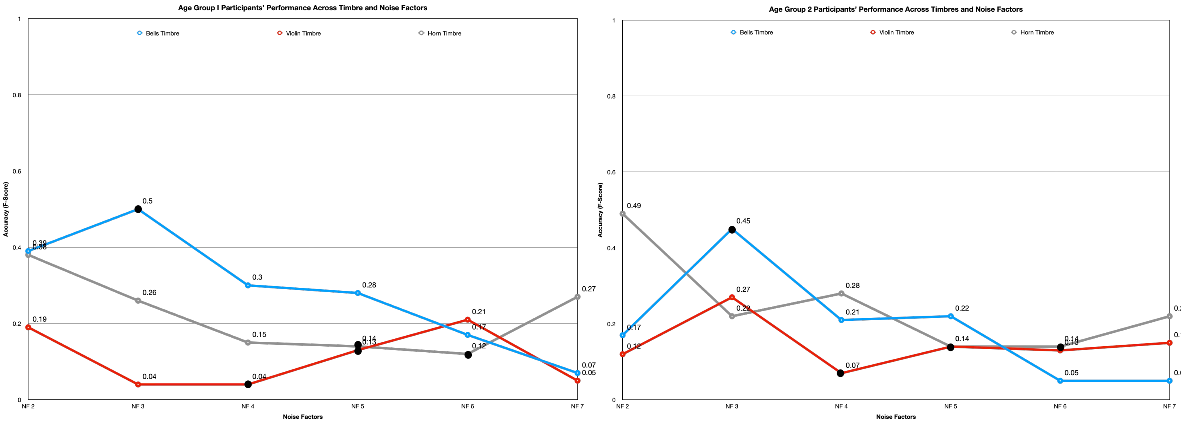


Figure 15: Participants' performance in age group 1 (left) and age group 2 detectability rate (right) across all timbres and noise factors.

Analysis of participants' responses from G2 indicates equally limited accuracy across all timbres and NFs as observed in results from age group 1. In NF2, horn timbre recorded minimal accuracy of 0.49, followed by bells (0.17) and violin (0.12) timbres. In NF3, bells recorded 0.45 (under error), followed by violin with 0.27 accuracy and horn with 0.22. Low accuracy rates were recorded in NF4, with horn timbre scoring 0.28, followed by bells 0.21 and violin with 0.07 (under error). Results from NF5 show a 0.22 accuracy score in bells, while both horn and violin obtained an equal score of 0.14. In NF6, horn timbre recorded 0.14 (faded stimuli error), violin 0.13 and bells 0.05. In NF7, horn recorded 0.22, violin 0.15 and bells 0.05 accuracy rate.

3.6. Conclusions and Future Directions

The analysis from this pilot experiment reflects that the TGSS paradigm remained limited in determining whether and which timbres allow for signal detection in the specified noise levels, suggesting that further investigation is required to draw any conclusions about implementing the TGSS approach to study the effects of timbres and noise levels on signal detection. As observed in the overall analysis, the F-score across all timbre and noise levels remained below the minimal rate of 0.5 due to FPR/FNR>TPR, indicating minimal detectability of the stimuli in presented noise levels. Combinations such as horn timbre-NF2, bells-NF5, and horn-NF7 indicate potential detectability due to relatively higher accuracy scores, indicating that these timbre-NF pairings may offer stimulus detectability in noise. However, comparatively higher accuracy recorded in the lowest SNR makes combinations with NF7 redundant.

Subgroup-based analysis for non/musicians also reflected a limited detectability rate for musicians, whereas no detectability was observed in responses from nonmusicians. This comparative analysis suggests that musical training may have a positive effect on signal detection, as indicated by outcomes observed in horn- NF2 and bells- NF2 combinations. Additionally, comparatively much lower scores in higher noise levels observed in this analysis follow our hypothesis– higher accuracy in lower NFs, with the scores declining as the NFs increase. Bells and horn timbres' curves indicate a potential effect on signal detection for participants with musical training as compared to nonmusicians. Furthermore, results obtained from subgroup analysis based on age range indicate minimal variance in accuracy responses between accuracy scores of G1 and G2, since both groups' detectability remained below 0.5 score. Results from G1 seem to indicate the inverse correlation of accuracy and noise level following bells and horn timbres' curves, whereas responses from G2 do not clearly indicate this trend, suggesting potential influence of varying noise levels on signal detection for the G1 subgroup.

Reflecting on the potential impact of the error conditions on signal detection, the limitations of the experimental setup involving external wideband noise and additional noise sources in the space may have obscured the signal, leading to higher misses (FNR), hence a lower recall score. Additionally, random noise fluctuations may have been mistaken as signals, resulting in higher false detections (FPR) and decreased score of precision. This combined reduction in both precision and recall scores may have resulted in an overall low F-score. The second error condition caused due to the excessive reverberation property of the experiment space may have perceptually reduced the clarity of the signal. Potentially resulting in increased uncertainty in detectability and leading to expectation-based responses rather than perceptual. Further leading to reduced precision (higher FPR) and recall (higher FNR), as a result, an overall low F1 score across all timbre and NF conditions. The third error condition of this experiment involved misrepresentation of the spatialisation feature, due to which 5 trials containing stimuli towards the end of the sonifications were faded, leading to 31 appropriately presented trials/stimuli. This resulted in no findings for the timbre-NF combinations that were subjected to this error, resulting in no indication of signal detectability under these conditions.

Since the experiment aimed to investigate the effects of timbre combined with varying noise levels on signal detection, the role of pitch in potentially influencing signal perceptibility was not addressed. However, it is acknowledged that the adopted approach of FM-based timbre synthesis may have induced pitch characteristics. In sonifications containing signals, distinct pitch may have been a perceptual indicator of the signals' presence, particularly in low noise levels. Furthermore, synthesised timbres such as horn and violin may have perceptually masked the signal more, as indicated by their spectral analysis (shown in figure 6), demonstrating peaks that coincide with the generated signal's pitch (figure 7). The timbral variations may have also been misinterpreted as signals, leading to an increase in false alarms. This study did not investigate the effect of pitch on signal detection; however its potential effects are acknowledged, which demand further investigation in future studies to determine their influence on signal perceptibility with varying noise levels.

The results from this pilot study are modest and remain limited in addressing whether and which timbres allow signal detectability and in which noise levels. Findings indicate a minimal effect of varying noise levels in the subgroup-based analysis; however, assessment is restricted in demonstrating the significant effect of timbres on signal detectability. Furthermore, given the error conditions of the experimental setup, no conclusions may be drawn from the presented experiment analysis. Future experimental studies of the TGSS paradigm are encouraged to conduct the experiment in an acoustically well-isolated space to minimise the reverberation of the room contributing to the signal detection task. The spatialisation feature is required to be presented in accordance with the design to achieve the effect of panning, which may contribute to signal detectability by allowing participants to 'follow along' as the sonification is played. Furthermore, subgroup-based inquiry may involve delving into the influences of musical training and the role of cultural background in timbral perceptibility and differentiation, as well as incorporate analysis based on subgroup with experience in astronomical data analysis. To identify which timbre-NF combinations allow for detecting stimulus in noise, further investigation using the TGSS paradigm is required. The analysis demonstrates a higher FPR and FNR than TPR, suggesting that overall, the presented range of noise levels was inappropriate, this limitation can be approached by redefining the NFs to a smaller range of noise levels. Since 53% of the overall responses across timbres correctly identified the stimuli in NFs2 and 3 with the highest SNR, the NF range can be calibrated around these NFs. Therefore, 6 noise levels such as NF0.5, 1, 1.5, 2, 2.5, and 3 are proposed for future iterations of this study to identify exactly at which level the signal is perceptually masked by the noise. A re-examination of this proposed NF range may contribute to identifying exactly at which threshold the signal is perceptually more masked by the noise.

4. Perception Experiment II

4.1. Research Question

The second research question in this study inquires: *to what extent can the spatialisation of the instrumental noises present in a gravitational wave detector augment the detectability of the gravitational wave signal?*
In order to address this question, an audification-based GW spatialisation sonification (AGSS) paradigm is developed by audifying the specified GW noise channels of real GW data and spatialising them alongside the GW strain in multichannel speaker installation(s) to assess whether spatialisation may contribute to signal sensitivity and identification of GW data features.

4.2. Methodology

Audification, the technique of shifting the waveform of time series data to the human auditory range with nonspeech audio for data exploration (Nees, M.A., & Walker, B.N. 2008; Vogt, K., 2018; Dombois, F., 2001; Necciari, T. et al., 2012).), is used to translate GW noise channels' data points to a one-dimensional data stream. With this process, the series of data points are interpreted as sound samples, with each value corresponding to the level of amplitude of the produced sound at a specific time interval. During processing of the data, the sample rate is specified, which indicates the number of samples played per second. This playback speed defines both the frequency (range) of the data and the duration of the audio file. Audification also ensures data translation without losing or 'manipulating' data, maintaining 'legitimate' sonic representations (Connell, B. R., et al., 1997; McGuire, J. M., et. al., 2006; Nees, M. A., & Walker, B. N., 2009) and has been employed in various disciplines, including seismology, experimental physics, and microbiology (Dombois, F., 2001). Accordingly, this method is considered to develop a sonification-based approach for real GW data to retain the data's behaviour and its intrinsic characteristics, which can be perceived in its auditory representation.

4.2.1. Audification based Stimulus Generation

To develop the pilot study of the AGSS approach, real data of the GW noise channels is derived from the LIGO consortium. The noise categories incorporated in this study are two channels from the Hanford Observatory: frequency stabilisation servo channel H1: PSL-FSS-FAST-MON-OUT-DQ recording the 'distortion' from the frequency stabilisation servo, and H1: OMC-PZT1-MON-AC-OUT-DQ channel monitoring the dithering of a mirror in the output mode cleaner. From the Livingston Observatory: L1: CAL-PCALY-RX-PD-OUT-DQ channel is used, which conducts calibration measurements, containing injections from the Y-end photon calibrator used to subtract lines injected from the photon calibrator, glitch subtraction channel measuring the light transmitted through Fabry-Perot cavities. These noise channels are characterised as instrumental noise, was recorded in 2017 near the detection of the GW170817 signal. Therefore, alongside these noise channels, the GW170817 signal template is presented, which is obtained from the Gravitational Wave Open Science Centre (GWOSC) portal. In this study, these four data sources are referred to as 'GW170817 signal,' 'PSL noise channel, 'Dithering Mirrors noise channel,' and 'Photon Calibrator noise channel. The data from the noise channels are audified using the Audacity application by importing their raw data points into the application. For each noise category, the following parameters were defined:

Sample Rate: 16384 (rate at which they were all recorded)
Sample size: 32-bit float numbers
Channels: Mono

As all noise files are recorded at a sample rate of 16384 Hz, once imported in Audacity, the sample rate is set to the original sample rate of the data file instead of the default 44.1 kHz, defining the duration of the file, with a sample size of 32-bit floats in a mono channel. After defining these parameters, a 'normalised' effect is applied in order to level out all the peaks with maximum peak amplitude to -1 dB. The option of ''remove DC offset'' is an additional feature that has been applied to the dithering mirrors and photon calibrator noise channels as they were off centre. The audifications of the noise categories are made accessible here.

Fig. 16.a represents the spectral analysis of the PSL noise channel, demonstrating a dense and nonuniform energy distribution across frequencies, with peaks at 60Hz, 120Hz, 151Hz, 180Hz, 242Hz. These intensities are represented as a series of equidistant faint lines across the temporal domain in the spectrogram as illustrated in figure 17.a. Spectral analysis of the dithering mirror's noise channel demonstrates relatively uniform intensities across frequencies, and peaks at 60Hz and 144Hz, the noise's spectrogram (figure 17.b.) delineates a clear white line across the temporal domain at 4115 Hz, with a reddish vertical gradient showing lower intensity from 6599 Hz to 8000 Hz. The photon calibrator noise channel's spectral analysis illustrates intensities at 16Hz, 435Hz, and 1092Hz, and the spectrogram as shown in 17.c depicts these intensities as faint equidistant lines across the vertical axis.

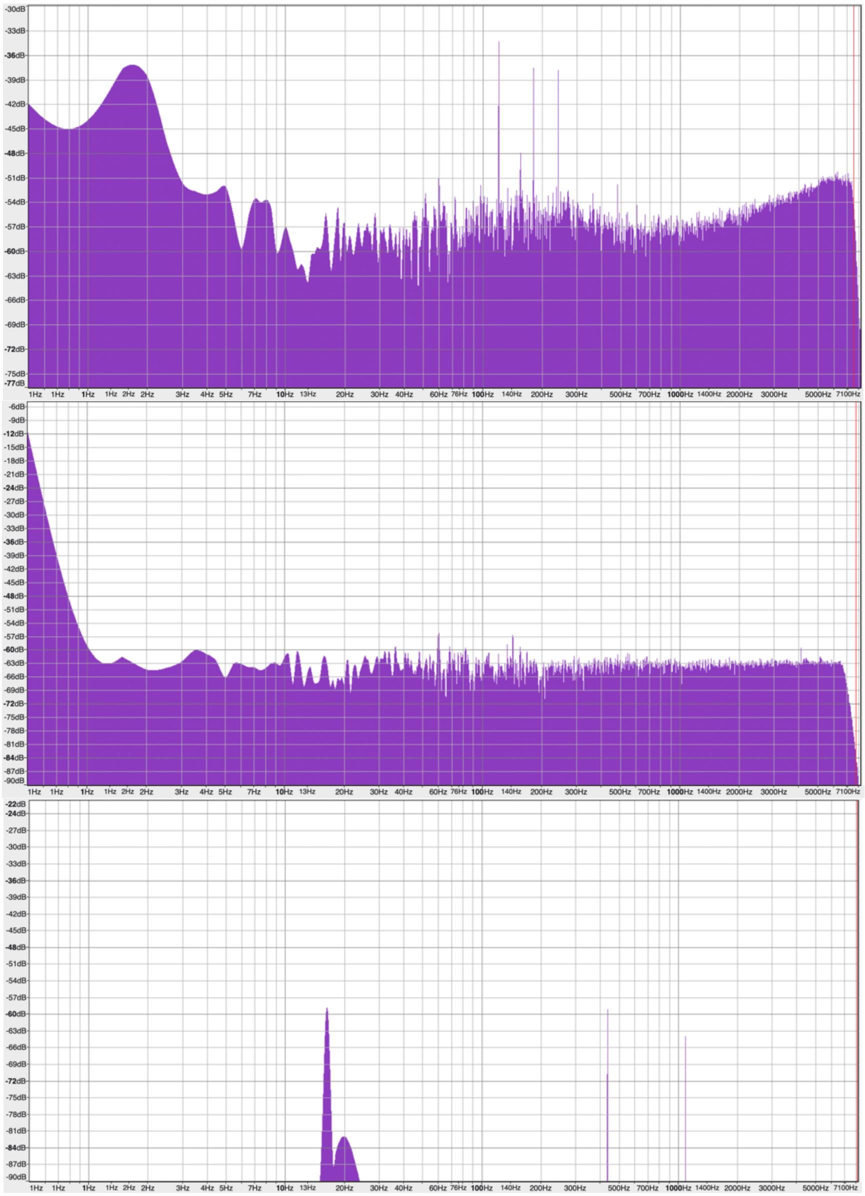


Fig 16: Spectral analysis plots of a) Pre-Stabilised Laser, b) Dithering Mirrors and c) Photon Calibrator noise channels (from top to bottom).

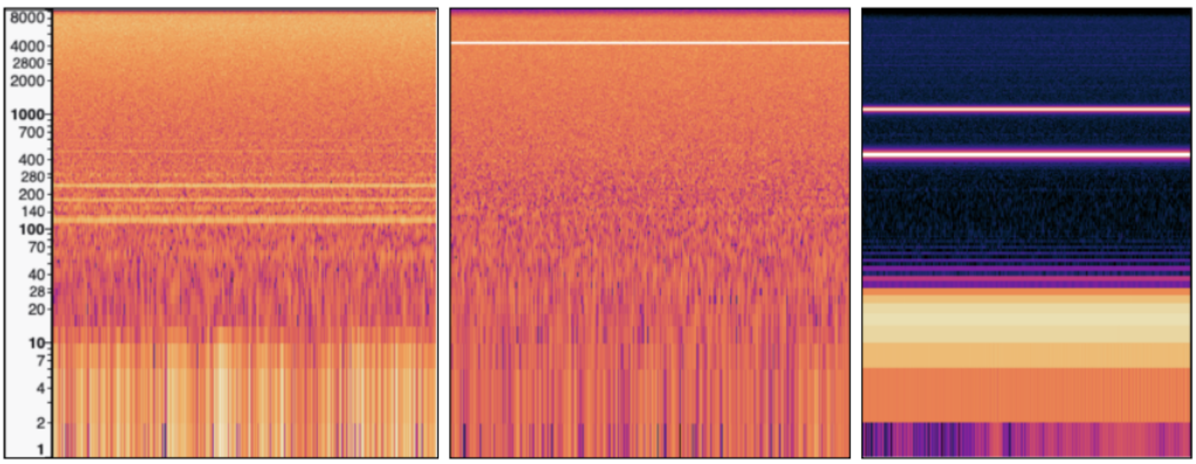


Fig 17: Spectrograms of a) Pre-Stabilised Laser, b) Dithering Mirrors and c) Photon Calibrator noise channels (from left to right).

The GW170817 signal is a result of two neutron star mergers, one of the first detections observed from such a binary system. For the purposes of this study, the GW170817 signal's template is chosen due to the complexities involved in signal-from-noise extraction processes. Figure 18.a. represents the spectral analysis of the GW170817 signal, with highest intensity observed at 72Hz and 75Hz and incrementally decreasing with lowest observed at 1065Hz. In figure 18.b. the signal's spectrogram illustrates the vertical axis ranging from 0 to 8000 Hz and the horizontal axis spanning from 0 to 2 seconds, demonstrating a gradual increase in the frequency domain over time.

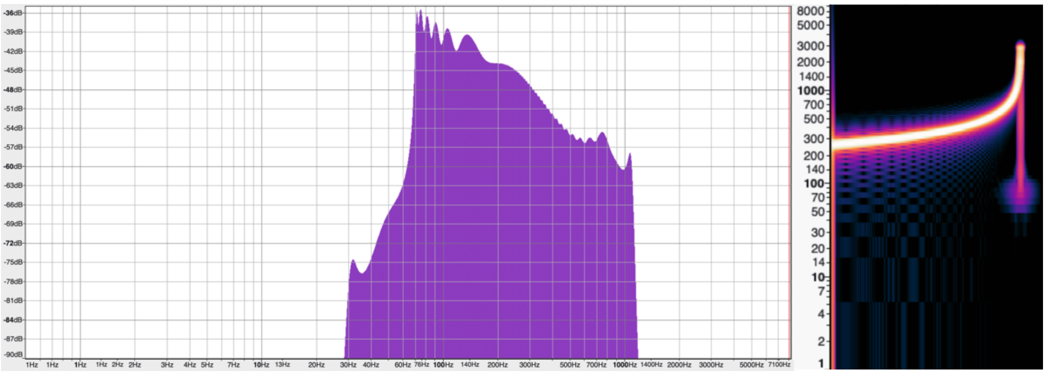


Fig. 18: a) Spectral analysis plot representing the GW170817 signal template (left) and b) spectrogram representing the signal over frequency and temporal domains (right).

4.2.2. The Masking Effect

Masking is an auditory phenomenon altering the perception of concurrent sounds (Zwicker, E., & Fastl, H., 2013) occurring when a higher frequency tone superimposes a tone with a lower frequency, making it perceptually inaudible (McGuire, J. M., et al., 2006; Walker, B. N., & Kramer, G., 2004). Zwicker, E., & Fastl, H. (2013) described the phenomenon of spectral masking as that noise can perceptually superimpose certain features of the auditory domain, resulting in diminished hearing, implying that auditory events closer in frequency domain are masked by each other more than those that are largely separated in frequency bandwidth. Due to spectral masking, a tone can be made inaudible, even if it is heard individually. The effect of masking is investigated in this study by identifying which noise category masks the signal as well as other noise categories the most. To examine signal sensitivity in concurrent sounds, the effect of masking based on varying noise intensities is conducted in a spatialised and non-spatialised speaker configuration to assess whether spatialisation influences signal perceptibility with concurrent sounds.

4.2.3. Spatialisation

Spatialisation is used as an additional sound parameter in combination with audification to explore the effect of masking on signal detectability in concurrent noise sources. Childs, E., & Pulkki, V. (2003) and Nasir, T., & Roberts, J. C. (2007) highlighted the influence of sound spatialisation enabling enhanced data perceptibility when comparing spatialised information display in contrast with monophonic transmission. This further suggests that data consisting of spatial characteristics are valuable to be brought into the spatial domain as an additional contribution to the perceptibility through sonification of the concerned data at hand. This data can be acoustically and spatially mapped, highlighting the data's temporal as well as its spatial behaviour, rendering a way to further discover the relationship between sound, specific data feature(s), and space.

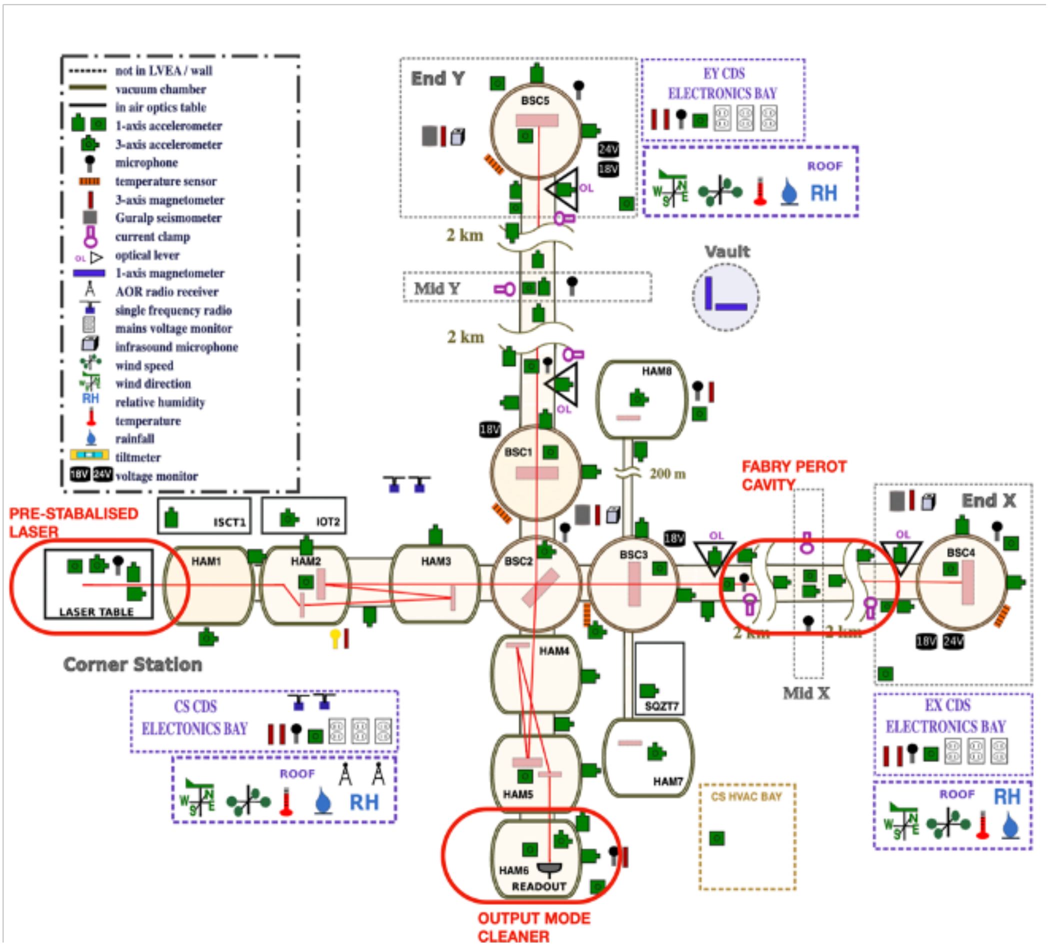


Fig 19: A modified schematic of LIGO Interferometer denoting the locations of the noise sources as employed in this research. Original adapted from http://pem.ligo.org/channelinfo/index.php, last accessed on 2nd Sep'24.

The AGSS approach proposed in this study explores real GW data by spatially mapping the acoustic representation of the data. This is obtained by spatialising the audifications of GW data, the GW170817 signal, and three noise channels. The audifications represent the data's temporal features and the spatialisation indicates the locations of the noise sources. The audifications of the GW noise channels are mapped to the auditory space of a quadrophonic installation, in which the noise channels are mapped to their origin in the Michelson Interferometer, a simplified version of the LIGO interferometer. The noise of the pre-stabilised laser will be displayed by speaker 4, the dithering of mirrors by speaker 3, the photon calibrator will be displayed by speaker 2, and the GW170817 signal template will be displayed by speaker 1. Figure 19 specifies the localisations of these noise sources in the interferometer, contextualising the GW noise information and contributing to the understanding of detector categorisation analysis.

4.3.4. Experiment Description

To assess whether spatialisation may influence signal perceptibility, the task-oriented experiment design is separated into two blocks. In block I, all four audifications will be concurrently displayed in a monophonic setup, with figure 20.a showcasing a schematic of the monophonic setup. Whereas in block II, each sound source will be presented in four separate channels in a quadraphonic environment as specified in Fig. 20.b. In total 8 trials will be displayed with 4 trials per block, with each trial lasting for a period of 45000 ms (unlike the 1000 ms stimulus display in experiment I).

In order to achieve a safe loudness level and prevent clipping, 0.25 (88 dB) is determined as the maximum volume for this experiment to ensure that the combined amplitudes of 4 audifications displayed simultaneously remain below 1. A trial-dependent configuration is developed to assess which noise category masks the GW170817 signal and/or is perceptually more dominant than other noise channels. Table 3 summarises the loudness configuration per trial and noise channel. The first trial has equal loudness levels of 0.15 (83.5 dB) for each noise channel. In trial 2, the PSL channel is set to 0.25 (88 dB), whereas the other noise channels are at 0.15 (83.5 dB). Trial 3 displays the photon calibrator channel at 88 dB, while other noise channels are set to 83.5 dB. In the last trial, all noise channels are set to the maximum volume of 88 dB. The loudness configurations for each noise channel per trial remain consistent across blocks to examine the effect of spatialisation on signal sensitivity.

This AGSS paradigm is explored through a participatory experimental analysis to investigate the effect of masking by spatialising four different concurrent audifications in a monophonic and quadrophonic setup. The participants are asked to define the lowest threshold at which they begin to perceive the signal in each trial in both blocks. This second pilot experiment of this study consists of three phases, in which the same participants from the first experiment will be introduced to the real GW data. This experiment was also conducted in the same space as Experiment 1, i.e., in the glass cylindrical space with the specified dimensions in section 3 of experiment 1.

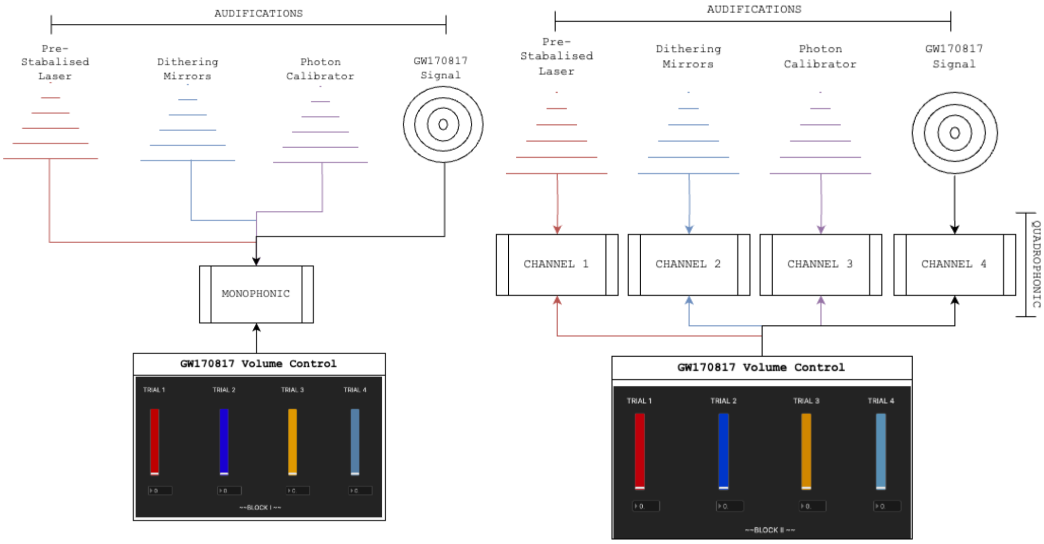


Fig. 20: Schematics of audio configurations in Experiment II for a) Monophonic and b) Quadrophonic displays.

In the introductory phase, participants will develop familiarity with the audifications of the three noise channels and the GW170817 signal template they would be experiencing during the experiment. In the second phase, participants will be presented with the four concurrent sounds in both Block I (monophonic setup) and Block II (quadraphonic installation). The participants will stand in the middle of the experiment space (glass circular space same as in experiment 1) and will be exposed to auditory displays for 45000 ms per trial, providing ample time to interact with the sound sources. In each trial, they are asked to record their perceived loudness threshold on a designed interactive graphical user interface (GUI), which will be provided on a screen. The interface, as illustrated in Fig. 21, depicts 4 sliders, each corresponding to one trial. The sliders volume levels range from 0 to 1, where 0 corresponds to 0 dB and 1 to 100 dB. Participants will be asked to define the loudness level of the GW170817 signal and determine the threshold of the volume of the GW signal by perceptually defining the lowest threshold at which they perceive the signal in both monophonic and quadrophonic setups.

	Trial 1	Trial 2	Trial 3	Trial 4
Pre-Stabalised Laser	**0.15**	**0.25**	**0.15**	**0.25**
Dithering Mirrors	**0.15**	**0.15**	**0.15**	**0.25**
Photon Calibrator	**0.15**	**0.15**	**0.25**	**0.25**

Table 3: Loudness level configuration across trials in both blocks of Experiment II.

Post trials, the participants will be invited to the final stage of the experiment series for an interview; their responses will be textually recorded by the experimenter. After the completion of both experiments, the volunteers will be provided with phosphorescent, space-themed stickers. In order to remain under the auditory attention timespan and reduce chances of repose-collection in fatigue, both experiments for this research will last for a period of 20-25 minutes per participant, under the time-limit mark of 30 minutes (Walker, B. N., & Kramer, G., 2004; Nees, M. A., & Walker, B. N. 2009).

Fig.21: Graphical user interface for Experiment II interaction sonification paradigm.

4.3.5. Analysis Method

Varied data information is required to be displayed separately to ensure the perceptibility of differences between data points for the perceiver (Nees, M. A., & Walker, B. N., 2009). This will be explored in our noise category intensity-dependent analysis. Following Barrett, N., & Crispino, M. (2018), the presentation of multiple data features through a monophonic setup can potentially mask and interfere with the perceptibility of data. Therefore, we anticipate greater sensitivity and the ability to differentiate between concurrent sound sources in block II with quadrophonic installation, recording higher sensitivity to the signal as opposed to the monophonic setup. This will be indicated by a comparatively low detection threshold in block II, falling in line with the existing literature and indicating a positive effect of spatialisation on signal perceptibility.

The recorded responses will be based on the decibel level of the digital signal and will be analysed based on a comparative analysis between the signal detection threshold recorded in monophonic and quadrophonic installations. The average detection threshold will be calculated in every presented condition (4 trials) using all participants' responses from both installations. A higher detection threshold in a particular condition may indicate that the noise category set to the maximum in that specific condition is more dominant than others, suggesting a potential effect of masking. This evaluation will be conducted for the overall participant population as well as two distinct subgroups, including analysis based on musical training and age range to examine any potential influence of these factors on signal sensitivity.

We hypothesise that, in condition 1 with all three noise channels presented at 83.5 dB, the noise sources will be equally perceptible and will record a relatively low signal detection threshold, indicating higher sensitivity to the signal. In condition 2, with PSL at 88 dB and other noise channels at 83.5 dB, a relatively high detection threshold may be observed, suggesting masking of the GW170817 signal and potentially other noise channels by the PSL category. For trial 3, the photon calibrator will be displayed at 88 dB and is anticipated to record lower detection threshold compared to trial 2. In the final trial, all noise channels will be presented at 88 dB, with the PSL channel possibly masking the most. This trial is expected to record the highest signal detection threshold due to all noise channels presented at the maximum loudness. This hypothesis will be examined in a monophonic and quadrophonic installations to assess the effect of spatialisation on signal sensitivity and masking.

4.3. Experiment Setup

The experiment analysis of the AGSS paradigm was conducted in the same experimental setup and with the same volunteers as experiment 1 (detailed in section 3.3). The limitations of the experiment setup were present during experiment 2, resulting in additional noise source(s) contributing to the signal-detection task. A potential approach to resolving these limitations can involve increasing the loudness of all trials. The audifications of the noise channels were developed in Audacity 3.7.0 software, an open-source audio processing application. The GW170817 signal template sound file was obtained from the GWOSC portal via https://gwosc.org/audiogwtc1/. The sonification design of the multichannel output was developed in Plug Data (PD). Two PD patches were developed, for monophonic display and for quadraphonic display, to allow for real-time audio presentation as well as an interaction-based feedback system via the embedded GUI. Both the monophonic and quadrophonic patches are accessible via the corresponding links. The orientation of the speakers corresponded to a simplified version of the Michelson Interferometer, in which each speaker's position "mapped'' the origins of the noise sources used in this study. A detailed illustration is presented in figure 22. Figure 23 depicts an anonymous participant undergoing the experiment.

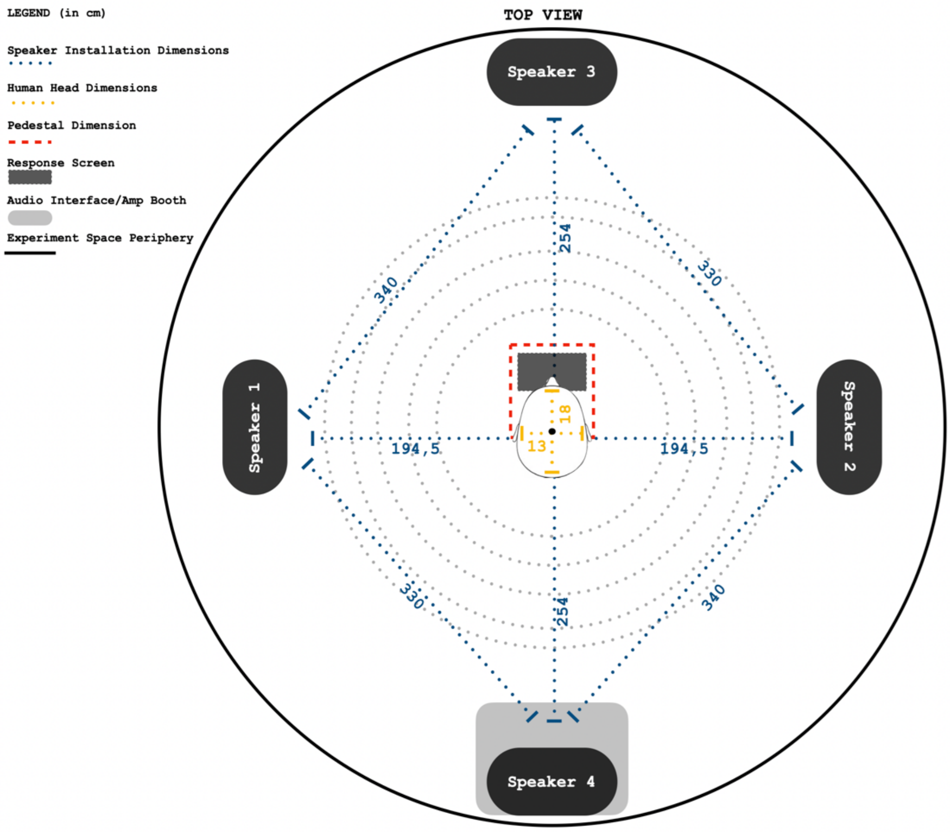


Fig.22: Schematic of Monophonic/Quadrophonic installation for Experiment II. Speaker 3 was the sound source during block I (monophonic display), while the four-speaker setup presented the auditory display for block II.

Fig. 23: Participant positioned at the centre of the experiment space and interacting with the GUI interface to during Experiment II.

4.4. Results

The following results are presented in consideration of the limitations of the experiment setup as described in section 3.1.5.
To examine the effects of spatialisation and masking on GW170817 signal sensitivity and three noise channels, participants' (n = 37) signal detection thresholds were recorded in monophonic and quadrophonic installations involving noise category intensity-dependent analysis.
Analysis summarised in figure 24, indicated mean sensitivity levels to be minimally higher in the quadrophonic setup for trial 1 with an average of 64.6 dB as compared to monophonic with 65 dB. This negligible variance highlights the limited effect of spatialisation on signal detectability observed under the condition that all noise channels are equally loud at 83.5 dB. In trial 2, higher signal sensitivity is observed in the quadrophonic setup (M = 64.4 dB), whereas monophonic recorded an average of 66.4 dB, suggesting that under the condition that the PSL channel is at the maximum loudness (88 dB), the quadrophonic setup allows for slightly higher sensitivity to the signal. In trial 3, the average signal sensitivity threshold was observed at M= 63.5 dB for monophonic presentation, while an average of 65 dB for quadrophonic, suggesting that a monophonic setup comparatively enables minimally higher signal perceptibility when the photon calibrator channel is at 88 dB. Under the condition all noise sources are displayed at the loudest level in trial 4, the monophonic setup recorded the higher sensitivity (M= 65.5 dB), while the quadrophonic setup recorded (M= 67.6).

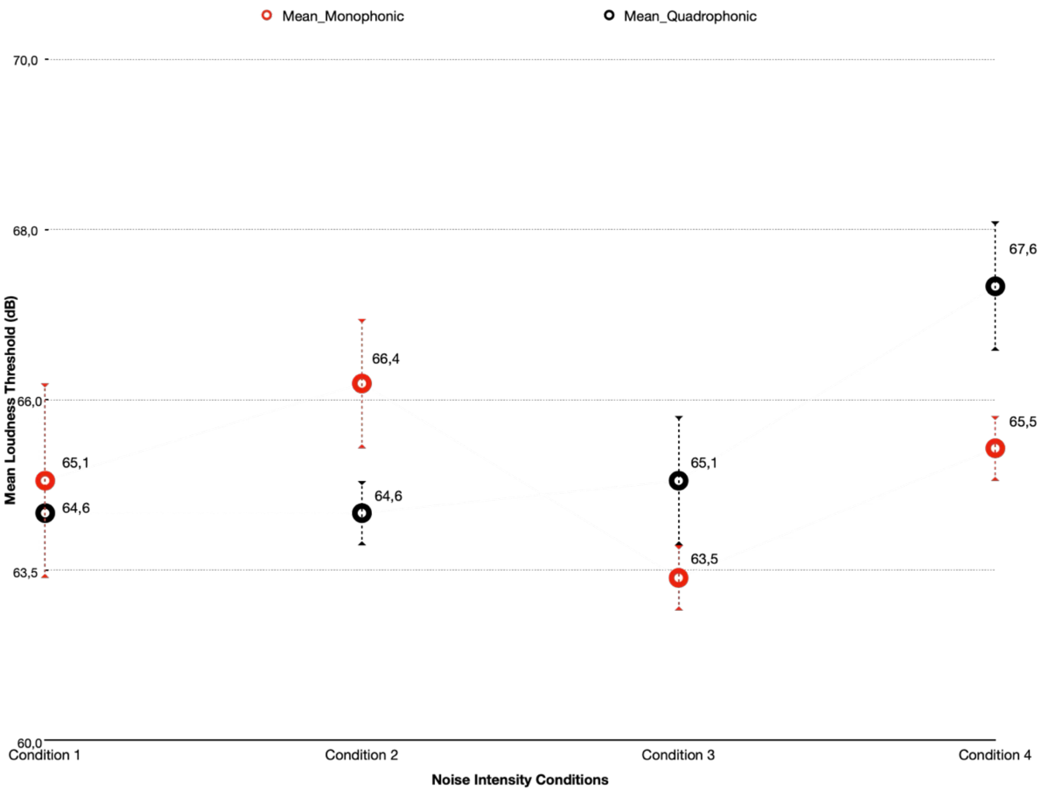


Fig. 24: Signal detection sensitivity across trials in both monophonic and quadrophonic installations. Mean loudness threshold (dB) depicted on the vertical axis corresponding to each trial on the horizontal axis.

This analysis indicates that the signal sensitivity is comparatively slightly higher in quadrophonic installation when all noises are displayed at equal loudness in the condition of the lowest level (83.5 dB). Consequently, the monophonic setup allowed for higher sensitivity than quadrophonic under the condition all noises displayed at the highest volume level (88 dB).

Comparative analysis of signal perception threshold between PSL and Photon Calibrator noise channels suggests minimal influence of spatialisation under the condition the PSL noise channel is at the loudest level (88 dB). While results from the monophonic setup suggest higher sensitivity with the Photon Calibrator at the loudest level. The analysis suggests that spatialisation facilitates sensitivity to the signal when the noise category (PSL) with continuous spectral distribution across the frequency-time domain and intensities ranging from 7500 to 8000 Hz, which exceeds the frequency bandwidth of the GW170817 strain (ranging from 0 Hz to 2865 Hz presented in this study), is displayed at the highest loudness alongside the other 2 noise sources. Whereas, in the monophonic setup, this noise category is perceptually more dominant and masks the signal more in the monophonic setup.

4.4.1. Subgroup-based Analysis Musicians and Nonmusicians

Results obtained from subgroup-based analysis with musicians (n = 17) and nonmusicians (n = 20) are illustrated in Fig. 25. The analysis of the results observed in the monophonic setup, as summarised in Fig. 25.a, suggests that in trial 1, non-musicians (M = 64.6 dB) are more sensitive to the signal compared to the musicians (66.4 dB). Under the condition that the PSL laser is at the highest loudness, musicians demonstrated slightly higher sensitivity (M= 66 dB) as compared to non-musicians (M= 66.4 dB), however, perceptually the variance is negligible. In trial 3, with the photon calibrator at the highest loudness, both subgroups, musicians and non-musicians, performed equally (M = 63.5), the highest level of sensitivity recorded in this analysis. Under the condition all noise categories were equally loud at 88 dB, both musicians (M= 65 dB) and nonmusicians (M= 66 dB) were equally perceptive to the signal with minimal variance.

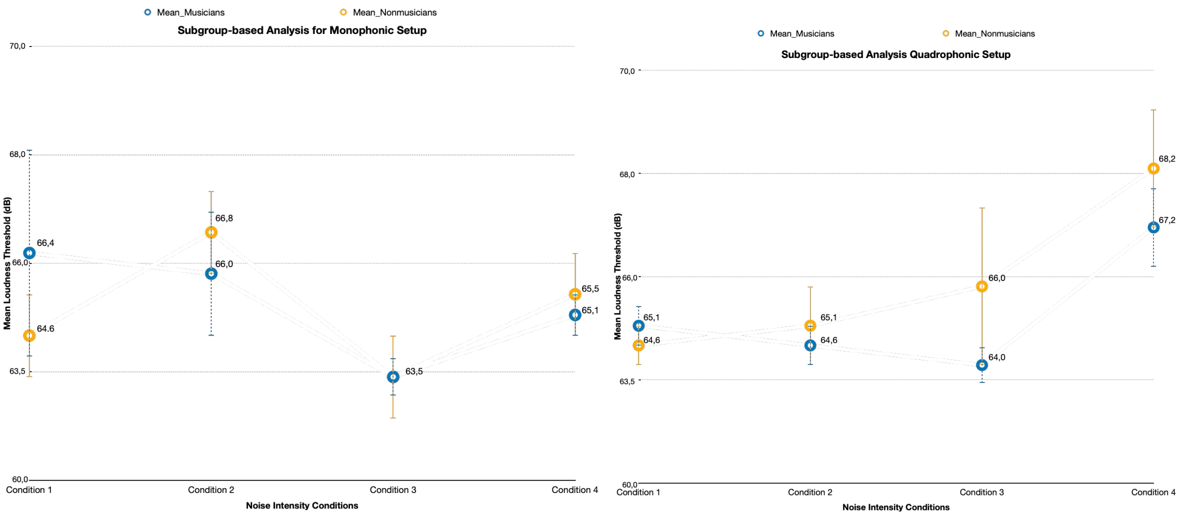


Fig. 25: Signal detection sensitivity of musicians and non-musicians' subgroup in a) non-spatialised monophonic installation (left) across trials depicting mean loudness threshold (dB) on the vertical axis corresponding to each trial on the horizontal axis corresponding to each trial on the horizontal axis and in b) spatialised quadrophonic installation (right).

Findings from the spatialised quadrophonic setup, illustrated in figure 25.b., indicate limited variability in signal detectability between nonmusicians (M = 64.5 dB) and participants with musical training (M = 65 dB) in trial 1. Similar results are observed in trial 2, with the PSL channel at 88 dB, in which musicians recorded a 64.6 dB average while non-musicians (M = 65). Outcomes from trial 3 present a moderate variance as musicians recorded a higher signal sensitivity threshold (M = 64 dB) while nonmusicians recorded an average of 66 dB. The last trial, which included concurrent display of all three noise sources at equal loudness with the highest volume, recorded 68 dB as the mean sensitivity for the nonmusician subgroup, while musicians recorded 67.2 dB. The subgroup-based analysis suggests that participants' performance remained relatively neutral in a non-spatialised setup. However, in a quadrophonic setup, musicians were comparatively more sensitive to signal detection. Limited conclusions can be derived from this analysis, as in both monophonic and quadrophonic installations the subgroups' responses do not significantly differ.

4.4.2. Subgroup-based Analysis Age Group 1 and 2

Analysis of signal sensitivity across age groups evidently demonstrated that participants from age group 1 (n = 23; 18 to 28 years old) recorded a much higher signal sensitivity threshold across blocks and trials as compared to the participants from age group 2 (n = 14; 29 to 63 age range), the results are summarised in figure 26. In the non-spatialised monophonic setup, the results are illustrated in Figure 26.a. Group 1 recorded 66 dB as the signal detection threshold, whereas age group 2 recorded 63.5 dB for trial 1. Under the condition the PSL channel is loudest, group 2 was more sensitive to the signal (M = 65 dB), while group 2 reported 66.8 dB. In trial 3, with the photon calibrator at the loudest, group 1 recorded 62.2 dB as the average sensitivity threshold, while group 2 recorded M = 64 dB. In the last trial with all noises playing at equal loudness, group 1 noted 65.5 dB as the signal detection threshold, whereas group 2 recorded 64.6 dB. These results showcase an almost parallel trend of signal perceptibility amongst participants in both age ranges, indicating higher signal perceptibility in participants from group 2 across all trials in a non-spatialised setup. Consequently, in a spatialised quadrophonic setup with results summarised in figure 26.b, group 2 defined 63.75 dB as the average sensitivity, while group 1 recorded a slightly higher signal detection threshold (M = 65 dB) in both trials 1 with equal lowest loudness and 2 with the highest PSL. In trial 3 with the photon calibrator channel at the highest level, group 2 recorded higher sensitivity (M = 62.9 dB), while group 1 reported 66 dB. Lastly, in trial 4 with all noises being displayed at equal loudness, group 2 reported slightly higher sensitivity (M = 66.4) compared to group 1 (M = 67.9).

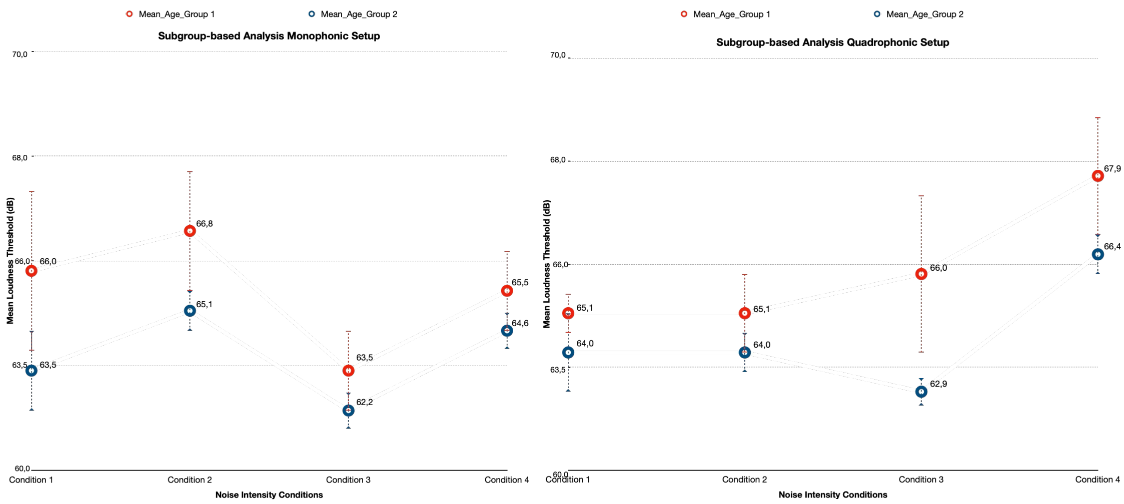


Fig. 26: Signal detection sensitivity of participants in Age Group 1 and Age Group 2 in a) non-spatialised monophonic installation (left) across trials depicting mean loudness threshold (dB) on the vertical axis corresponding to each trial on the horizontal axis and b) spatialised quadrophonic installation.

4.5. Conclusions and Future Directions

The outcomes from experiment II highlighted minimal variance in signal detection thresholds across conditions. As a result, our inquiry – to what extent spatialisation may contribute to signal detection – remains unfulfilled and requires further examination to determine how the frequency characteristics of noise channels and spatialisation affect signal sensitivity when their audifications are presented concurrently. In consideration of the error conditions, our analysis indicates that Y-end photon calibrator noise channel perceptually masks the signal the least, whereas pre-stabilised laser channel masks the most. Participant responses from the interview phase correspond with this analysis, as 45% of the participants (14 responses) noted having perceived the PSL noise channel masking other noise channels and the signal the most. Participant 18 specified that "*Noise 2 was taking over the lower part of the signal, [whereas] noise 1 and 2 were masking over the higher pitch of the signal.* 16% of the participant population noted for each category, Y-end photon calibrator and dithering mirrors channels, masking other noise channels and the signal. It is interesting to note that in the monophonic display of both trials 1 and 4, when all the noises are at equal volume levels, the sensitivity is recorded to be around the same threshold of 65 dB, indicating that participants were equally perceptive to the signal in a non-spatialised setup when all concurrent noise sources are presented at equal volume. In the quadrophonic setup, the signal detection threshold remained moderately similar across conditions. However, it increased relatively significantly in condition 4. Participants reported during the interview that the quadrophonic setup allows for better distinguishing between the concurrent sounds (20 responses), further emphasising that it allows for better signal perceptibility. Participant 8 highlighted during the interview, ''*The quadrophonic setup allowed for better distinguishing between the sounds, more spatial awareness from the sounds, so I could locate them better. And allowed me to perceive the signal easier*. Regarding the monophonic setup, 13 participants suggested having perceived single-speaker installation better in distinguishing between sounds and perceiving the signal and ''*was less tiring''* (participant 10 reported), additionally, participant 26 highlighted that '*'the monophonic setup allowed for better perceptibility; either I distinguish by pitch or* frequency. Lastly, 5 participants reported to have perceived no difference.

Furthermore, results from screening-based participant groups of recording signal detection threshold based on musical training demonstrated that spatialisation of GW data augments signal sensitivity slightly more for musicians compared to nonmusicians. Similarly, in monophonic setup, musicians recorded higher detection threshold compared to nonmusicians, apart from the conditions (1 and 4) with all noise categories being played at equal loudness, nonmusicians recorded higher sensitivity. Assessment across the age range indicated that G 2 recorded lower detection threshold compared to G 1, suggesting that participants in the age range 29 to 63 were more sensitive to the signal in both monophonic as well as spatialised speaker configurations. This analysis indicated minimal, however notable observations from the pilot experiment run. However, due to the limitations of the experimental setup, limited conclusions can be drawn from this analysis. To address whether spatialisation and sonification may contribute to signal sensitivity and/or facilitate in identification of diverse concurrent sounds, further investigation is required to assess whether spatialisation of GW audifications may effect sensitivity to the signal and aid in distinguishing between GW data's acoustic representations. Eventhough the results from this pilot remained limited, the developed AGSS paradigm shows promise for further exploration

In reflection, the experiment setup potentially had a significant impact during the second experiment, resulting in a greater detection threshold (low sensitivity). The wideband noise may have partially dominated the photon calibrator channel, while potentially masking certain spectral features of the dithering mirror noise channel at 60 Hz and 144 Hz, primarily in conditions 1, 2 and 3, where this channel was set to a minimum volume of 83.5 dB. These external noise sources, wideband noise emerging from the amplifier and noise originating from the air duct situated in the ceiling, were both directional sources and were localised within the enclosed experiment space, further resulting in less uniform acoustic distribution. Additionally, the air duct noise source may have influenced the auditory scene both spatially and temporally due to its fluctuation over time, potentially contributing to inconsistent detection performance across conditions. Furthermore, the reverberant property of the experiment space may have perceptually reduced the clarity of the signal, leading to some responses being recorded at a higher threshold. These external unregulated auditory contributions to the developed sonification configurations might have rendered perceptual classification of concurrent noise channels challenging as well as reduced sensitivity to the signal in the concurrent auditory scene. For future iterations, conducting this investigation in a well-controlled environment is crucial. Furthermore, in this study, loudness was measured based on the decibel level of the digital signal, which may have resulted in an inaccurate representation of the listeners' experience, as the recorded value does not consider the acoustic properties of the space. To approach this, prior to the experiments, in-situ measurements can be performed using a sound level meter, and playback systems such as speakers and amplifiers can be calibrated to verify that output in the space closely matches the digital signal.

5. Alternative Outputs

The development of this study yielded alternative outputs, complementary to the process of this research. These experiences included an interactive installation piece exhibited at the Het Nieuwe Instituut, Rotterdam, The Netherlands, as a part of a six-month artist residency. The residency initiated with exploring the boundary between a scientific experiment and an interactive artistic installation, investigating the role of the observer and the operator/the experimentee. Further examining these agents' positionalities and modes of interaction as well as perception in relation to the curated auditory environment. In the second half of the residency, the research transitioned beyond the paradigm of multisensorial to cultural astronomy, leading to compelling exchanges cross-pollinating notions about non/human sensorial experiences in relation to astronomical phenomena. Examining this emergence through the lens of traditional astronomical knowledge cultures such as Inca Cosmovision and Jain Cosmology.

Fig.27: Images from the artist residency, Auditory Observatory installation (left) and interactive lecture on Jain Cosmology at the concluding event of the residency(right).

Our proposed GW data sonification paradigms, Timbre-based GW simulation-sonification and Audification-based GW spatialisation sonification, were demonstrated at the 75th International Astronautical Congress in Milan, Italy, hosted by the Italian Association of Aeronautics and Astronautics (AIDAA). Our demonstration, *Towards an Auditory Observatory*, exhibited developed sound-based GW data analysis paradigms, alongside the presentation on multisensorial astronomy by Prof. Wanda L. Diaz Merced. Both sonification approaches were presented as an interactive installation in a dedicated space at the event, inviting experts in space sciences to participate and share insights. Our demonstration was well received by both experts as well as non-experts, fostering discussions about the ever-present challenge of signal-to-noise in astronomical data analysis, as well as exploring the potential of sound and its multidimensional features in examining other astrophysical data, such as the composition of galaxies and dark matter from the Euclid mission. Furthermore, our work was invited to the Breakfast Seminar series at ESTEC, European Space Agency, in the Netherlands. This opportunity offered a variety of critical aspects involved in GW data analysis, addressed by the mission specialists from ESA's Laser Interferometer Space Antenna (LISA) mission, which will be the first space-based GW observatory. This transition from ground to space-based instrumentation presents enormous challenges for data analysis but also numerous opportunities, including the prospect of incorporating the paradigm of data sonification as a potential approach towards the GW analysis in adjunct with existing visual-based models.

Fig.28: Auditory Observatory demonstration at the IAC, with experts' participation shown on left and young enthusiasts' on the right.

6. General Conclusions and Discussion

Even though both experiments are inherently different, the task-oriented paradigm of signal detection from noise through auditory information processing remained the common denominator. The proposed paradigms aimed to present potential approaches in exploring GW data through data sonification and spatialisation. In the presented analysis, the experiments currently remain limited in assessing the effect of timbre perception and varying noise levels on signal detectability as well as exploring the effect of masking and spatialisation on sensitivity to the stimulus, facilitating a space for further inquiry.

Observations from experiment 1 derived from stimulus detection in noise based on the TGSS paradigm indicated negligible influence of timbres on signal detectability; nonetheless, a minimal effect of varying noise factors is observed in the analysis. The low accuracy rates in both conditions indicate that the task-oriented paradigm of the TGSS model is challenging and offers minimal sensitivity to signal detection. Furthermore, in reflection with the limitations present in the experiment setup and outcomes lying in contrast with existing evidenced literature, limited findings can be drawn from this experiment. Previous studies highlight that harmonic sounds facilitate stimulus sensitivity. However, it remains yet to be uncovered which specific timbral characteristics provide higher sensitivity in auditory processing of stimulus in noise. Our proposed TGSS approach may potentially contribute to this discourse by assessing harmonic and inharmonic timbres' influence on the effect of masking of the stimulus. Further studies are required to determine the effect of noise factors in signal recognition, as well as a deeper inquiry of timbre-NF combinations is required to examine the effects of such correlations that augment or hinder signal perceptibility.

Continued examination of the paradigm of the timber-based synthesis approach in signal detection can involve other timbres, such as brass and keys, alongside the presented timbres in this study that can be applied for developing a diverse and richer timbre-based synthesis sound palette to examine the effect of timbre on signal detection. Additionally, a smaller scale of variance in noise factors, such as NF0.5, 1, 1.5, 2, 2.5, and 3, may enable further exploration of the effects of variability in noise factors and potentially support in identifying the specific threshold at which the signal becomes imperceptible and in which the corresponding timbre-NF combination. Moreover, as summarised in section 3.5, presenting sonifications from the same channel in a stereophonic environment contributed to the limitation in signal perceptibility in the first experiment, resulting in more challenging signal detectability. In future iterations of the TGSS paradigm, its spatialisation feature can be presented as per the outlined methodology and assess whether the effect of panning may contribute to higher perceptibility of the signal across eccentricity levels, and the panning effect may allow participants to 'follow along' as the sonification plays, potentially enhancing higher attention mechanisms in this task-oriented experiment. Consequently, subgroup-based inquiry may involve delving into the influences of musical training and the role of cultural background in timbral perceptibility and differentiation.

The examination from the perception experiment 2 indicated greater perceptibility for the signal at the lowest thresholds in a monophonic setup, suggesting non-spatialised auditory representation can potentially support higher signal detection comparatively. Noise channel(s) composed of the similar frequency bandwidth as the signal when spatialised are observed to hinder perceptibility of signal detection, masking the GW signal over time, and support the constructed priori hypothesis for this assessment. Results showcase limited variance in signal detection threshold. However, relatively higher sensitivity is recorded in quadrophonic setup, indicating that spatialisation can augment signal perceptibility in concurrent sounds compared to non-spatialised, monophonic setup. The minimal effect of masking is observed as findings indicate that the photon calibrator noise channel perceptually masks the least, whereas the pre-stabilised laser channel masks the most, with insights from the interview phase corresponding to the analysis. This suggests that signal sensitivity can be noise-frequency and spatialisation dependent, laying in line with Zwicker, E., & Fastl, H. (2013). Noise channel(s) with continuous spectral distribution and composing of intensities higher than the frequency bandwidth of the GW strain may augment sensitivity to the GW signal detection through spatialisation. The results suggest that signal detection threshold across trials in quadrophonic setup remain consistent and only peaks in the final trial 4 (with highest volume of all noise channels), suggesting that quadraphonic setup facilitates more perceptive ‘parsing’ of number of sounds presented in an environment. The identification of the differences in sounds takes place as the listener learns to identify the signal and localise the ‘source’ where it is emerging from in the environment, resulting in the perceptivity path of identification > distinguishing > comprehension. The analysis presents signal sensitivity slightly lower in monophonic as the mean value between blocks differs by 0.4, representing that the monophonic setup may allow for higher signal perceptibility however, it is much lower in distinguishing between different noise sources. This is also noted by participants as their primary focus being signal detection.

The continuation of exploration of GW data analysis through the AGSS approach may include mapping the GW strain on the representation of the localisations of the GW observatories where the strain is being detected, with the audifications occurring across the speaker configuration based on the time difference of the strain’s travel (propagating at light’s velocity) and may also be a potential future application of the data spatialisation feature. Further investigation of the data’s temporal characteristics may involve acoustic exploration and detection of transient sources in the data that are localised in time and evaluate the evolution of an observed pattern. Furthermore, this approach can be used for glitch classification based on acoustic interfaces, as glitches are erratic occurrences in the data that are nonstationary noise and tend to dominate the strain. Assessment of whether sonification and spatialisation could contribute to the identification and categorisation of glitches may facilitate GW data exploration in conjunction with existing GW data processing methods. Additionally, this approach may potentially also contribute to tracing the glitch pattern in the signal by using its acoustic representation as an indicator for recognising where the glitch originates. Section 2.2 outlines the audification of the GW150914 signal chirp, describing that the sound of the chirp is dependent on the chirp mass of the binary system. For instance, black holes with smaller solar mass $M_o = 14$ (only 14 times bigger than our sun) would generate a gravitational wave with smaller spacing between the amplitudes (smaller ripples) and a shorter wavelength, resulting in a higher-pitched sound. This offers intriguing perspective on employing sound-based approach for GW analysis and investigate whether sonifciation of GW data may contribute to identification of parameters of the binary system such as distance between the two bodies, the system’s distance from us, the mass of the system and even the orbital trajectory of one or both bodies in the system.

The results from both participatory experimental investigations provide limited outcomes in observing whether timbre may have contributed to signal detectability and assessing the effects of masking on signal detection in concurrent sounds. Furthermore, due to the effects of the error, no conclusions may be drawn from the experimental analysis presented in this study. The limitations of the experiment setup may have significantly contributed to the decrease in signal sensitivity due to external noise sources in the space. In subgroup-based analysis involving musicians and nonmusicians, minimal variability is observed, leaving room for further research to assess the effect of musical training on signal detections. Alternatively, age range-based subgroup analysis indicates higher signal detection observed in group 2, results lying in contrast with existing literature. Deeper exploration of data characteristics may provide additional knowledge and aid in the production of relevant GW data-acoustic interaction interfaces. Future iterations of this research are encouraged to address and further explore the developed approaches to support the advancement of an interactive spatio-acoustic GW analysis paradigm.

7. Acknowledgements

This work is an outcome of innumerable conversations, thought experiments, and learning gained from multiple individuals. This section intends to celebrate these serendipitous interactions that profoundly contributed to the realisation of this work.
I am sincerely grateful to invaluable contributions by my first advisor Edwin F. van der Heide and to my second advisor, Wanda L. Diaz Merced, whose curiosity-driven research and depth of knowledge in multisensorial astronomy inspired me to take on this exploration. I extend deepest gratitude to my supervisors whose guidance has been instrumental in shaping this work, their mentorship not only deepened my interest in research, but also greatly contributed to developing my academic vision.

I would also like to express my appreciation to the members of my thesis committee, Dan Xu and Martyna Chruślińska, as well as to the attendees of my presentations, for their valuable feedback and constructive recommendations, which allowed me to identify critical gaps in my work that I otherwise overlooked.

This research represents a small, emerging contribution to the growing cross-disciplinary field of multisensory astronomy. This preliminary work resonated with diverse research paradigms. I am honoured to have collaborated with Het Nieuwe Instituut, and extend my heartfelt gratitude to Ramon Amaro, lead curator of the -1 Lab, and the team for their collaboration in realising experimental studies of this research and deepening the dialogue on the role of sensory perception through artistic research practices.

This work also owes greatly to Geetika Jain, for her constant encouragement and curiosity and always reminding me reach for the stars. I also extend heartfelt gratitude to my friends and colleagues, for their support and motivation throughout this journey.

Lastly, I am honoured to have received valuable feedback on this preliminary work from both experts and non-experts. I am thankful to Alfonso Pagani and the Italian Association of Aeronautics and Astronautics (AIDAA) for their support in presenting this research publicly. I also wish to express my deep appreciation to Willi Exner and Lorenzo Speti, Postdoctoral Fellows at ESTEC, European Space Agency, for their generous support and for fostering a space of dialogue around sound-based astronomical data analysis within expert communities.

8. Bibliography

1. Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., ... & Cao, J. (2018). Constraints on cosmic strings using data from the first Advanced LIGO observing run. *Physical Review D*, *97*(10), 102002.
2. Abbott, B. P., Abbott, R., Abbott, T., Abernathy, M. R., Acernese, F., Ackley, K., ... & Cavalieri, R. (2016). Observation of gravitational waves from a binary black hole merger. *Physical review letters*, *116*(6), 061102.
3. Barrett, N., & Crispino, M. (2018). The impact of 3-D sound spatialisation on listeners' understanding of human agency in acousmatic music. *Journal of New Music Research*, *47*(5), 399-415.

4. Brown, L. M., Brewster, S., Ramloll, R., Burton, M., & Riedel, B. (2003). Design guidelines for audio presentation of graphs and tables. Proceedings of the International Conference on Auditory Display (ICAD2003) (pp. 284-287), Boston, MA.
5. Casado, J., De La Vega, G., Díaz-Merced, W., Gandhi, P., & García, B. (2019). SonoUno: a user-centred approach to sonification. *Proceedings of the International Astronomical Union*, *15*(S367), 120-123.
6. Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., et al. (1997). The Principles of Universal Design, Version 2.0. NC State University, Raleigh, NC: The Center for Universal Design.
7. De La Vega, G., Dominguez, L. M. E., Casado, J., & García, B. (2022, June). SonoUno web: an innovative user centred web interface. In *International Conference on Human-Computer Interaction* (pp. 628-633). Cham: Springer Nature Switzerland.
8. Diaz-Merced, W. L., Candey, R. M., Brickhouse, N., Schneps, M., Mannone, J. C., Brewster, S., & Kolenberg, K. (2011). Sonification of astronomical data. *Proceedings of the International Astronomical Union*, *7*(S285), 133-136.
9. Kramer, A. F., Hahn, S., Cohen, N. J., Banich, M. T., McAuley, E., Harrison, C. R., ... & Colcombe, A. (1999). Ageing, fitness and neurocognitive function. *Nature*, *400*(6743), 418-419.
10. Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychology and aging*, *9*(4), 491.
11. Necciari, T., Balazs, P., Kronland-Martinet, R., Ystad, S., Laback, B., Savel, S., & Meunier, S. (2012, March). Perceptual optimization of audio representations based on time-frequency masking data for maximally-compact stimuli. In *Proceedings of the 45th AES conference on Applications of Time-Frequency Processing in Audio. Helsinki, Finland* (pp. 103-112).
12. Usman, S. A., Nitz, A. H., Harry, I. W., Biwer, C. M., Brown, D. A., Cabero, M., ... & Willis, J. L. (2016). The PyCBC search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, *33*(21), 215004.
13. Zanella, A., Harrison, C. M., Lenzi, S., Cooke, J., Damsma, P., & Fleming, S. W. (2022). Sonification and sound design for astronomy research, education and public engagement. *Nature Astronomy*, *6*(11), 1241-1248.
14. Zevin, M., Jackson, C. B., Doctor, Z., Wu, Y., Østerlund, C., Johnson, L. C., ... & Téglás, B. (2024). Gravity Spy: lessons learned and a path forward. *The European Physical Journal Plus*, *139*(1), 100.
15. Enge, K., Rind, A., Iber, M., Höldrich, R., & Aigner, W. (2023). Towards a unified terminology for sonification and visualization. *Personal and Ubiquitous Computing*, *27*(5), 1949-1963.
16. Schöpper, L. M., & Frings, C. (2023). Same, but different: Binding effects in auditory, but not visual detection performance. *Attention, Perception, & Psychophysics*, *85*(2), 438-451.
17. McPherson, M. J., Grace, R. C., & McDermott, J. H. (2022). Harmonicity aids hearing in noise. *Attention, Perception, & Psychophysics*, *84*(3), 1016-1042.
18. Czesla, S., Molle, T., & Schmitt, J. H. M. M. (2018). A posteriori noise estimation in variable data sets-with applications to spectra and light curves. *Astronomy & Astrophysics*, *609*, A39.

19. Nuttall, L. K. (2018). Characterizing transient noise in the LIGO detectors. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2120), 20170286.
20. Razzano, M., & Cuoco, E. (2018). Image-based deep learning for classification of noise transients in gravitational wave detectors. *Classical and Quantum Gravity*, *35*(9), 095016.
21. Vogt, K. (2018). Sonification and particle physics. *Department of theoretical physics, Institute for physics, Karl–Franzens–University Graz Austria, Institute for Electronic Music and Acoustics, University for Music and Dramatic Arts Graz, Austria, katharina. vogt@ uni-graz. at, Retrieved 16th April*.
22. Ishak, B. (2018). Gravitational waves volume 2: astrophysics and cosmology. *Contemporary Physics*.
23. Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693-707.
24. Zwicker, E., & Fastl, H. (2013). *Psychoacoustics: Facts and models* (Vol. 22). Springer Science & Business Media.
25. Grond, F., & Berger, J. (2011). Parameter mapping sonification. In *The sonification handbook*.
26. Nees, M. A., & Walker, B. N. (2009). Auditory Interfaces and Sonification.
27. Nasir, T., & Roberts, J. C. (2007). Sonification of spatial data.
28. McGuire, J. M., Scott, S. S., & Shaw, S. F. (2006). Universal design and its applications in educational environments. Remedial and Special Education, 3, 166-175.
29. Summerscales, T. Z. (2006). *Gravitational wave astronomy with LIGO: From data to science*. The Pennsylvania State University.
30. Childs, E., & Pulkki, V. (2003). Using multi-channel spatialization in sonification: a case study with meteorological data. Georgia Institute of Technology.
31. Dombois, F. (2001). Using audification in planetary seismology. Georgia Institute of Technology.
32. Flowers, J. H., & Hauer, T. A. (1995). Musical versus visual graphs: Cross-modal equivalence in perception of time series data. *Human factors*, *37*(3), 553-569.
33. Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *44*(3), 400.
34. Einstein, A. (1916). Näherungsweise integration der feldgleichungen der gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 688-696.
35. Thorne, K. S. (1995). Gravitational waves. *arXiv preprint gr-qc/9506086*.
36. Bonebright, T. L. (2001, July). Perceptual structure of everyday sounds: A multidimensional scaling approach. In *Proceedings of the 2001 International Conference on Auditory Display* (Vol. 35). Espoo, Finland: Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology.
37. Quinton, M., McGregor, I., & Benyon, D. (2020, September). Sonification of an exoplanetary atmosphere. In *Proceedings of the 15th International Audio Mostly Conference* (pp. 191-198).