

Fabio Rovai  
Rov17515544  
Creative Computing Institute (UAL)  
Msc Data Science and AI for the creative  
industries 2020/21  
Supervisor: Rebecca Fiebrink

## **Thesis Project**



A sincere thank you to Umanesimo Artificiale, which broadened the horizons of my creative practice, and Ford Combs, which actively collaborated to develop this project.

## **Abstract**

This project opens an opportunity to examine connections between two apparently different subjects, such as art and science, reflecting how practitioners from both disciplines are interested in considering methods to embrace collaborative experimentations and understanding the importance of transdisciplinarity. There are undoubtedly different ways to promote and explore this link, where we decided to propose sounds as a point of contact. Then, this project draws a rationale on the multiple lines of understanding sounds produce, from forms of performative acts to more informative meaning, showing sound's potential to be a universally perceived form of language. Examining the difference between these two disciplines, we discuss to which extent artists and scientists can choose to sound as a communication tool, their relationship with their recipients, and the significance of their operations. To convey scientific meaning and propose audio as a data enquiry, art and science cooperations need to evaluate to what degree it might be beneficial, accounting for the instrument accuracy against the advantages it takes to literacy. Doing so, this project reflects data sonification as a comprehensive means to grasp information, and it gives it a more creative interpretation. The resulting experience aims to emphasise the role of knowledge and interaction in a world of multiple perspectives, promoting a broader dialogue and enabling different understandings of the practices and establishing a tool to encourage curiosity and creativity. To summarise, this project investigates the need for a new level of knowledge and examination from the general public in order to enable access without a prior perception of data in a determined context, which requires an understanding of a specific discipline, or particular notations, which frequently needs resources not available to a member of the general public. Using as the primary case study the sonification of protein coordinates, this project proposes a new approach for assigning distinct sounds to proteins given their geometrical qualities, intending to develop a tool that enables a creative way to interact with protein-related data by labelling it with sound, in the form of an earcon. To conclude, from an artistic standpoint, this project analyses previous trends in how arts treat information-based data structures, considering their role in our contemporaneity and confronting the evolution of artists' role in scientific inquiries and their relationship, from the sixties to the digital age.

## 1. Introduction

In 2021, in homage to the bicentenary of the fashion brand Louis Vuitton, the London based creative director Ben Ditto has been commissioned what we can call an unusual fashion statement. The beginning of the two thousand twenties has undoubtedly been a time of disruption and change, influenced by Covid.

To stress the role of the pandemic, Ditto teamed up with the virologist Marisa Zuk and Kenneth Robinson, a postdoctoral research associate at the Zoltan Takats laboratory in molecular and structural biology at Imperial College London.

The apparently odd component of this team is justified by the project itself, which consisted in remembering that Louis Vuitton faced and survived two pandemics, Covid and Cholera.

The team was assigned to gain access to a culture of *Vibrio Cholera* bacteria, reverse engineer it to give form bioluminescent properties, and paint it over a suitcase, showing the potential of contamination between art and science to turn disease into an artistic medium (Ditto, Zuk and Robinson 2021).

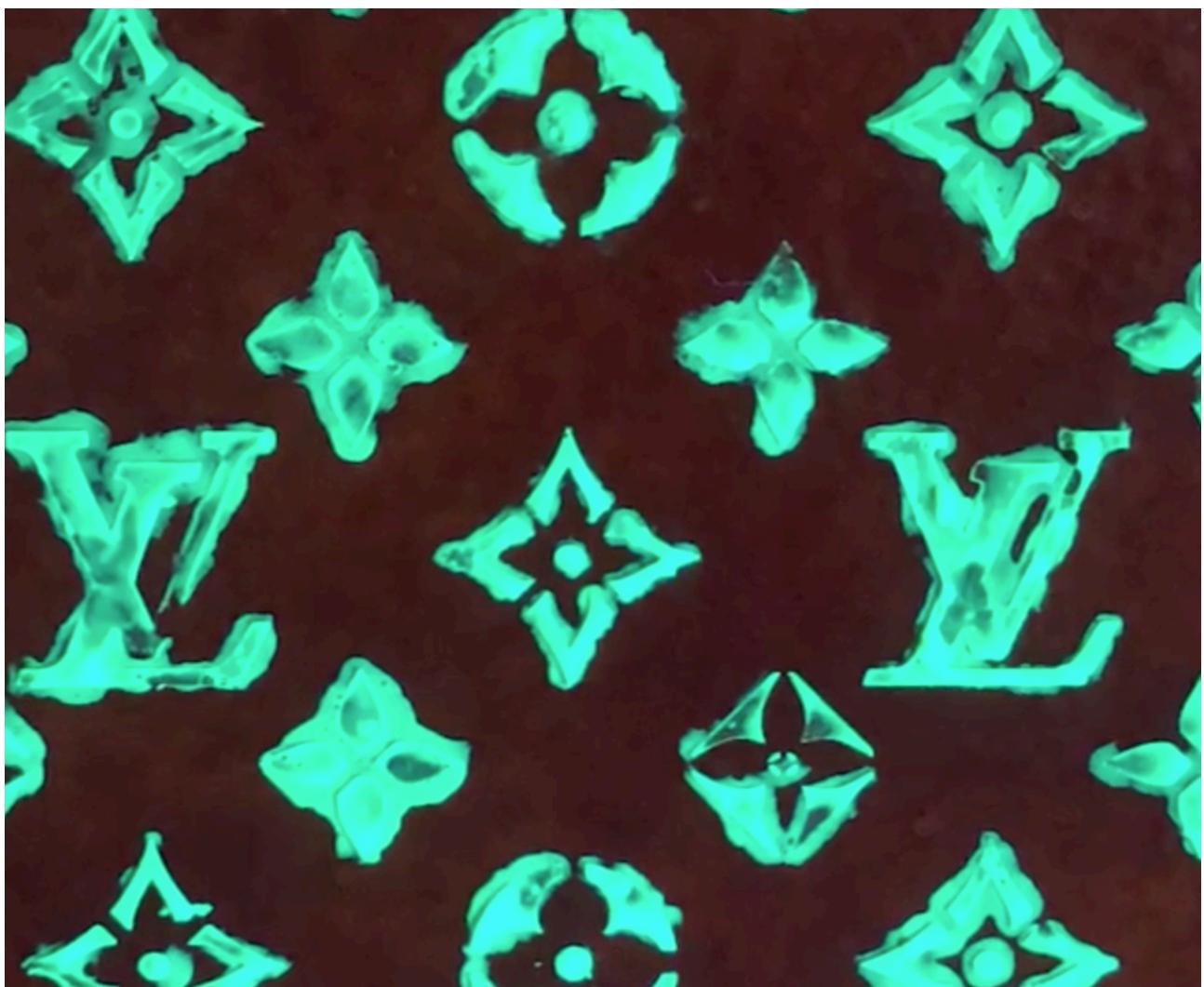


Figure 1.

A likewise peculiar example of projects merging art and science is *On Air*, an exhibition delivered by Tomás Saraceno at the Palais de Tokyo in 2019.

Here, the artist proposed an acoustic exploration of spider web structure, conceived in partnership with the MIT Laboratory for Atomistic and Molecular Mechanics. In the design of this performance, the team assigning each fibre of the spider web to a different sound and frequency, and when these are played together, they generate new and more complex timbres.

The end product is a performative melody that aims to reveal the duality of artistic output while also raising how creative techniques might lead to exploratory data insights, positioning it as another performative piece that confirms efficacious ties in this cross-collaboration between art and science.

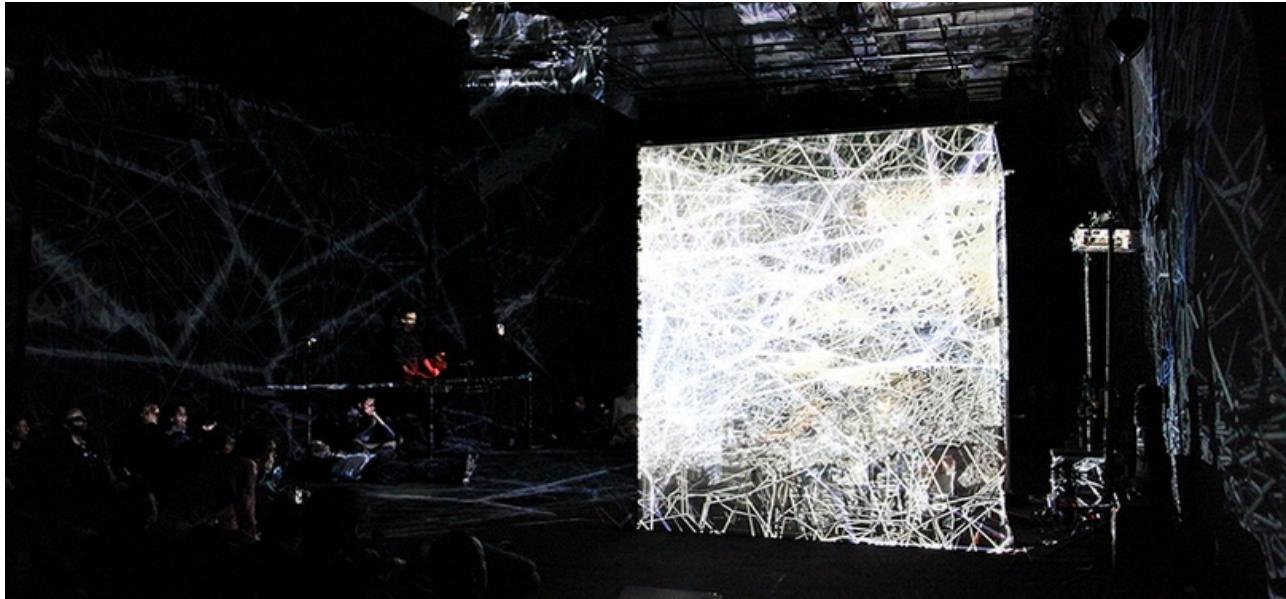


Figure 2.

This performance also directly displays the result of collaborative projects among disciplines, in which the team suggests that the node of a spider web can be seen as a metaphor for the development of an ecosystem of information and ideas, in which individuals have to cross various subjects in order to create and innovate (De Leon, 2019).

These two pieces share the exact origin, riding the wave of a growing interest among artists and biologists to foster cross-pollination between these fields. Furthermore, it should be noted that historically these collaborations of science and performative acts are not unseen happenings, numerous creative movements, such as the Fluxus laboratory in the sixties, encouraged informative artworks and dialogues aimed at the audience.

Fluxus Laboratory was created as a social effort to encourage the spread of domain-specific knowledge without the requirement for prior expertise, prompting the beginning of hybrid disciplines composed of art, technology, and information systems, proposing to produce a fashionable era of parallelism and contemporary interdisciplinarity.

It established a framework of affinity where both disciplines had the capacity to transform history, based on the concept that new ideas in mathematics introduced and drove unique scientific applications, much as newly proposed paradigms in art emerged when the worldview shifts.

Thanks to this definition, art theorist Dick Higgins coined the term *Intermedia* to represent such expansion where subjects needed to cross boundaries and create entirely new forms of practice (Higgins and Higgins, 2001).

In this new understanding of these disciplines, there have been cross contaminations between science and music, philosophical theory, and design, arguably paving the way to UI and UX design theories, creating experiences that priorities discussion with the audience rather than focus on purely technical deliveries.

For science, it is essential to create simple circumstances to explain complicated notions to an audience, placing them to larger applications. It should also be acknowledged that Fluxus artists were not the first to promote collaboration with scientists; however, they were the first art movement to declare it as an entirely new method of exploration rather than a one-time association (Friedman, 1998) where it can be clearly seen how interaction and diversity of interpretation on a broader audience promoted not only new tools for understanding new practices but also an instrument to enquiry, questioning how much art and scientific knowledge shared similar aspects of the same language (Cook, 2016).

The latter works of Entangled Others Studio noted that in addition with this connection about biotechnology and arts, these hybrid artworks can propose new space and focus on more humanistic inquiries.

For the duo, the audience should be made aware of the role of humans in natural ecosystems, wherein their piece *Beneath the Neural Waves* examines biodiversity, creating a speculative depiction of reality. Doing so, they use their artwork as a way to open dialogue, not only to theme strictly related to the discipline but as a way to broader abstract concepts, questioning how we relate and perceive our surrounding (Entangled Others Studio, 2021).



Figure 3.

Thanks to these previous works and investigations, concentrating on how many methods that appear to be in opposition to one another find their equilibrium, it can be argued that exploring information as a medium for art has an entirely new analysis with potential for endless cross-referencing opportunities, make meaningful alternative ways to form information (Hermann, Hunt and Neuhoff, 2011).

Furthermore, artists tend to find motivation in the complexities of learning new disciplines, which may have resulted in the evolution and contamination of the aforementioned areas, embracing experimentation and risk as a core concern in contemporary cultural production (Le Feuvre, 2010), where audience interaction is often prioritised over scientific accuracy.

Based on this idea, this research uses a loose notation specific to the science realm, often simplifying terms and overall structures, rather than using an accurate notation, where the comprehensive meaning and procedure is preferred to scrupulous representation. This project structures an illustrative background of these practices under different points of view, building a narrative that highlights a journey between art and science, testing different approaches, frameworks and proposing future investigations.

## **2. Related Works**

### **2. 1 Information**

The compenetration of art and science propose a different degree of expansion, mutually enriching both disciplines, not only for technical queries but as a critique of our surroundings. Analogously to scientific research, artistic practices stress their examinations on the possibilities and the role of technological innovation, proposing, however, different approaches to conceptually investigate these disciplines. Information produced by scientists is a pivotal component of our culture, and *Information Art* aims to reinterpret some of these topics, questioning the meaning of the relationship between disciplines, artistic contribution to scientific practices and interdisciplinary agendas (Wilson, 2003).

These disciplines found their intermediary frame in several critiques moved by the modern era, which led to the investigation of the importance of rejecting individual dominant cultural streams, deconstructing fixed metanarratives, and introducing a general scepticism to science universal truth and mock art role of elitarian practice. This fertile ground opens to a different form of dialogue and approach for art historians and curators, where scientists are enabled to art and artists are allowed to their version of the truth and interpretation. These investigations, however, are primarily driven by the arts community; they do not yet play the same role in the scientific one, which continues to struggle to see the arts as an independent area of research. Scientific publications frequently fail to see past their disciplinary boundaries, raising questions about what can be the role of an artist in a tech firm and how much they might contribute to their scientific research (Wilson, 2003).

On the other hand, historically, we can find numerous art and science collaborations, from Bell Labs' involvement of artists in sound research to artist Soto's participation in Xerox's Palo Alto Research Center (*PARC*), demonstrating that arts can operate as an autonomous territory of study, bringing in a creative perspective that could drive new insights in the research process, and offering solutions for better approaching general public requirements or perspectives. However, to accomplish so, artists need to make an effort and embrace skills and knowledge that will enable them to effectively participate in these realms and to challenge conventional assumptions of what constitutes an artistic education. To do so, they must consider unexplored research paths and unexpected implications, as well as learn about the information sources utilised by scientists and engineers, such as academic conferences and journals (Wilson, 2003).

Against the critique that scientists might move, questioning if this endeavour to learn from creative practitioners will be sufficient, we can object that different examples in history have demonstrated that topics once thought to be beyond nonspecialists reach become understood and accessible to far larger audiences. For this reason, the central role of artists working with researchers might focus on this demystification of scientific ideas leading to artist-researcher positions. The Wellcome Trust, which proposes a new multidisciplinary function for more biology-related communication, perfectly demonstrates this new interdisciplinary role, funding influential "Sci-Art" initiatives to stimulate cooperation between researchers, and artists, tearing down barriers between the two realms, and encouraging a disciplinary cross-fertilisation, showing to art, the possibility to be inspired by science's insights into the natural world; to science, the ability to see research from a whole other perspective (Wilson, 2003).

Undoubtedly, artists and scientists often have different readings of the world, making collaborative approaches naturally challenging. However, when these figures build tools together, it generates significant outputs (Gewin, 2021), where both mutually develop, interpreting reality differently, proposing new straightforward ways to communicate to various audiences, creating and delivering transformations focused on society (Paterson et al., 2020).

## 2.3 Sound

Sounds in art is traditionally considered as a mode of participation and an instrument to promote cultural awareness and knowledge. This definition comes from the general scepticism that art grew around the sixties, towards the effectiveness of visualisation, or sight, to be a comprehensive medium to represent the complexity of our surroundings.

Visualisation, which traditionally was the artists' preferred method of depiction, left space and opportunities for more comprehensive form of interaction, such as performance and sound, as a possible more effective instrument to interact with the audience (Kelly and Woodley, 2011), or in the specific of this project, to communicate data insights.

Considering this notion, we use sound to create a tool to propose a new creative methodology based on previously existing information and questioning the meaning of interpretation. Interpretation can be subjective, and it heavily relates to the way data are presented or retrieved, so proposing sounds to expand external disciplines, such as biology, aims to produce a more natural output to the audience (Sawe, Chafe and Treviño, 2020).

For example, sounds has a more intuitive approach compared to other forms of data representation, such as graphical plots. This statement finds its evidence in several papers and projects submitted to international events, such as *S+T+ARTS AI and Music Festival*, and the *International Community for Auditory Display*. On the other hand, one critique often moved against sonification, or more generally targeting the representation of data via audible sounds, is that audio often fails to faithfully represent this information, changing or distorting its meaning.

However, artistic sonification seems to embrace this difficulty, shining the aesthetic aspects of the discipline and shifting its aim towards sonic representation to engage with the public and propose a general idea of the data, rather than a meticulous depiction (Neuhoff, 2019), paving the way to new forms of communication and interaction (Hermann, Hunt and Neuhoff, 2011).

## 2.4 DNA Sonification

To have an understanding of the narrative behind artistic approaches to science and resulting collaboration, is pivotal to consider a pre-existing literature, particularly regarding molecular sonification, in order to analyse how this problem has been debated over the last half-century and to see how aims and techniques shifted. Also, we can notice a duality of approaches relating to the auditory display of DNA, where some authors prefer to abstract further the complexity of DNA to generate pleasant music, and others promote this idea as a comprehensive tool to understand these structures (Temple, 2017).

In 1979, philosopher Douglas Hofstadter hypothesised a possible interconnection between music and protein sequences, speculating on how music may aid in identifying their design, fragmenting complicated information into smaller, simpler to recognise pieces (Hofstadter, 1979).

This concept prompted Hayashi and Munakata to conduct one of the first studies on the relationship between DNA and music, where the scientists approach stemmed from the necessity to find a new form of annotation for the increasing amount of available data, which new technologies were producing in the eighties. These data needed to be manually annotated and transferred to computers, resulting in thousands of A, T, G and C, transcript in a complex notation.

As a solution, they proposed sonification as a way to distress the handling of these data, using a tone range of a fifth, similar to what humans encounter in daily speech, with symmetrical half-tone free arrangements (re, mi, sol, la).

The resulting assignment they suggested was re = G, mi=C, sol=T and la=A (Raven, 1984).

Scientists Ohno and Ohno advanced a more creative example of DNA sonification as an evolution of Hayashi and Munakata's project, where instead of aligning DNA sequences to symmetrical half-tone free arrangements, they map them to notes from a fragment of Frederic Chopin's Nocturne 'Opus 55 number 1' (Ohno and Ohno, 1986).

Inspired by these projects, composer John Dunn and biologist Mary Anne Clark created a tool to perform protein sonification, aiming for an audio output thought as an art piece that reflects the aesthetic of natural patterns. The duo began by using DNA sequences to produce music by assigning arbitrary pitches to each amino acid, and then they created a frequency histogram that assigned more consonant intervals to the amino acids that occur more frequently.

After doing so, they arranged the amino acids in a scale based on a randomly given variable, for example, their respective water solubility scale.

Then, the duo ended this project arguing on the importance of inquiring about protein nature using a language that can be easily comprehensible for the public and which relates to the user prior musical experience, wondering where is the art in science and where is the science in art, as an invitation for the public to further explore the topic and finding their own means and interpretations (Dunn and Clark, 1999).

Unlike the earlier mentioned DNA sonification methodologies, Scientist King and musician Angus from the Scottish band 'The Shamen', focused on creating a more scientific output, primarily for data analysis, called *PM* (Garcia-Ruiz and Gutierrez-Pulido, 2006).

They aimed to generate a musical translation of protein sequences, offering sound as a more advantageous method to visualise multivariate data, standing against a colours based methodology which, in accord with the duo, produced tiredness in the viewer. Their project design subsisted in translating a bass line of seven notes into amino acid attributes, such as polarity and size, and then, turning it to a musical bar with notes played following the order of DNA sequences (King and Angus, 1996).

To create a further extension to *PM*'s method and offer a more comprehensive understanding of protein structural alignments, a team of the researcher based at the University of California's Computer Science Department created *PROMUSE*, an integrated hybrid system to analyse of protein structural superposition. Instead of replacing the visualisation of data with an auditory display, as widely discussed in the earlier approaches, the team intended to expand visual depictions of data by enhancing it with audio. *PROMUSE* sonifies protein's polarity, secondary structure, fit goodness and exposure by displaying visual data using *RasMol*, a molecular graphics software package for protein visual analysis, and mapping the resulting outputs on a jazz quartet composed of a drum part, a solo instrument, a rhythmic accompaniment, and a bass line. Using this approach, the researchers observed that acoustic data exhibition performed better than audiovisual or visual displays in their trial outcome, pointing out how this mapping was an effective way to perform molecular visualisation (Hansen et al., 1998).

Further extending the *PROMUSE* idea, artist Grond and bioscientist Dall'Antonia proposed *SUMO* as an open-source sonification plugin to overcome the absence of standards and difficulty in implementing specific auditory display techniques.

Moved by the idea that scientists must continue to work efficiently using the familiar visual representation, the duo built *SUMO* to run inside *PyMOL*, a well established and accepted auditory display for molecules, functioning without the need of any additional hardware requirements, resulting in a simple but scalable solution. In addition, *SUMO* was designed following the idea that

molecular or genetic data should not be used solely as an abstract score to generate musical patterns, where many sonification methodologies are motivated by aesthetic considerations rather than scientific research questions, causing attention to be drawn away from potentially useful data finding.

*SUMO* computed the pairwise distances of all atoms and mapped them to the frequency list of the *Klank filter* inside a software called *SuperCollide*, to turn the protein structure into sound. This technique created a bank of fixed frequency that could be utilised to replicate an object's resonant modes, with pleasant sound and distinct auditory impressions.

However, the model failed to distinguish different type of molecules.

So, they proposed another approach, suggesting an interaction with the PyMOL window by selecting a portion of the display. The plugin operated calculating the mean values for each amino acid residue and communicated them to the responder using the *Open Sound Control protocol (OSC)*. Values were then mapped back and sorted into frequencies ranging from 100 to 2000 Hz in *SuperCollide*. The difference between the values determined the time offset, and clustering in the data was visible in the sonification. The result was an acoustic arpeggio over the targeted region, revealing information about proteins's displacement (Grond and Dall' Antonia, 2008).

*CoRSAIRe* (Combination of Sensorimotor Renderings for Immersive Analysis of Results) offered a more immersive development, suggesting to look at how proteins interact with their environment as a novel parameter for the examination of the sonification process.

This instrument aimed to create new ways to complement the exhibition and the immersive communication of proteins in three-dimensional space, not a mere stand-alone aural experience. By utilising spatial coordinates obtained from a multi-channel virtual reality interaction, and collected via a visual or haptic depiction, sonification was used to provide additional information in regards to the coordinate system.

The programme employed a method known as *parameter mapping* to create biophysical and geometrical dependent audio, where for example, prominent objects reduced the volume and the frequency of the signal (Férey et al., 2009).

The project's goal was to propose modulation as a method for exploring multidimensional data that can express three-dimensional interactions more effectively than other sonification approaches, such the direct translation of audio to a time series, also known as audification (Hermann and Ritter, 1999).

Inspired by *CoRSAIRe*, the work by researchers at the University of Stuttgart proposes an extension of their previously existing product called *MegaMol*, a visualisation middleware for protein prototyping.

In designing this interaction, the researchers employed sound as an auxiliary tool to show results produced by *MegaMol*, leading the analyst's attention to relevant phenomena and proposing an enhanced degree of perception that might be valuable in educational settings, implemented several auditory feedback to distinguish different actions, such as the user interaction with the protein size and position (Rau et al., 2015).

Motivated by the absence of efficient plugins to perform DNA sonification interfacing with popular molecular visualisation tools, a researchers team based at the Northumbria University's department of computer and information sciences presented an open-source sonification plugin for *UCSF*

*Chimera*. It combined molecular structure visualisation and analysis with data, mapping protein properties to pitches, and using phase modulation synthesis, analogously to the *CoRSAIRe* project. As a result, it generated a distinct overview for each protein structure and enabled users to listen to their atomic composition as a form of MIDI pitches (Ballweg, Bronowska and Vickers, 2016).

The Massachusetts Institute of Technology's Laboratory for Atomistic and Molecular Mechanics (*LAMM*) has developed a new method for understanding protein structures in musical space by mixing the resulting audio with a neural network to build new proteins.

According to the researchers, this hybrid approach may accurately reproduce existing sequences and generate entirely new ones, which can be viewed as a way to better comprehend patterns in diverse hierarchical systems. It can also give researchers ideas on making new proteins by promoting diversity as a form of representation and permitting reversibility, retrieving structural information from output sound. They propose a long short-term memory, or *LSTM* model, to treat musical depiction of proteins grouped on arbitrary protein characteristics. They then produce new musical compositions by mapping pitches and the lengths of molecular chains into amino acid sequences. The researchers then utilise a bioinformatic algorithm called *BLAST* (Basic Local Alignment Search Tool) to match estimated sequences of amino acid to existing proteins to evaluate their folding, claiming that their method can be used to build novel protein materials with applications in biology, medicine, and engineering (Yu and Buehler, 2020).

The same group then proposes a similar approach to create sounds out of coronavirus proteins, this time using a type of sonification called materiomusic, which uses the natural oscillations and arrangements of molecules to develop audible outputs that can be employed for biological investigations, proposing access to an accurate representation of protein structures via sound.

The individual vibrational characteristic of the protein's molecule serves as the foundation for translating amino acids into separate musical tones.

Because each amino acid has a three-dimensional vibrational spectrum, we can distinguish them by sound or timbre, described as an overlay of their frequency. Thus, the team applies to any amino acid with a diverse aural appearance defined by its characterisation of the vibration of their molecules assigned via a distinct tone or note.

In this fashion, they suggest assigning protein sounds to a specific sound, such as a bell, or potentially arranging the twenty amino acids to a particular tone in a musical scale, based on their lowest vibrational frequency. Moreover, additional notes could then be executed simultaneously, with the loudness of the latter being dependent on the range between inserted amino acids and their relative primary sequence, generating a distance correlation matrix to manifest their possible similarity (Buehler, 2020).

These MIT projects are arguably the most significant for this thesis project, creating a contact point between my practice and Combs's works.

Combs has a mixed background, with a bachelor's degree in Music from the University of North Carolina, and he is now a PhD student in Bioinformatics and Computational Biology at George Mason University, locating him in the middle of this inquiry where he adopts artistic methodologies into the scientific process of his practice.

I came across Combs' work when he presented an investigation focused on the above mentioned MIT project (Yu and Buehler, 2020), introducing a novel exploratory analysis and transcribing it in a simple Python implementation using a Jupyter notebook. The MIT research he investigated was initially conceived to be a Java-based android app; however, Combs turned it into a more accessible tool, aiming for a broader audience, which also mirrors the purpose of this project. Then, Combs

proposed an outline highlighting sonification as a helpful tool for performing information conversion to non-speech audio in a three-part Medium article.

Also, Combs enters the idea of harmonies to the previous paper, where the author recommends using computation geometry, particularly Delaunay tessellation (Gärtner and Hoffmann, 2014).

With this algorithm, the x, y, and z coordinates of a protein are considered part of one or more simplices, which are made up of four unique points.

The overlaps of these four points are then mapped to an arbitrarily chosen note, resulting in a chord, which we can use to play harmony.

To sum up, introducing this notion, the author proposed a more objective instrument for sonification to interpret data from various perspectives while emphasising that performed sonifications should attempt to stay as faithful as conceivable to the original data set, where proteins are just three-dimensional coordinates with their neighbours in their structure, where melodic sequences can be supported or carried by harmonies made up of adjacent amino acid notes.

(Combs, 2020).

This design is fascinating because it enlightens the possibility of computational geometry to perform sonification, which has not been considered yet on previously mentioned frameworks.

For this reason, Combs supervised the extraction progress and actively collaborated in producing code and audio insights for this project.

### 3 Project Design

This project's design considers the various obstacles users without coding skills might encounter interacting with Python for the first time.

To achieve a more understandable user interface, we disguise the notebook under the appearance of a mock website, resembling an experience more familiar to the user.

To build this clean user interface, we deploy the notebook with the help of the Google application Colab, which in short, is a product that allows browser Python execution.

The key benefit of employing Colab is that it allows forms that hide functions under some markdown visuals, allowing users to interact with code without actually breaking into it.

For example, the form `#@title` allows developers to create a heading similar to `<h1></h1>` in Html, and the more useful `#@param` generates a smaller heading, similar to `<h6></h6>`, combining it with the function `input()` to edit a variable directly.

It is quite fascinating, working similarly to the `argparse` module (Python 3.9.7 documentation, 2021), where, for example we can create a module called `Protein_ID.py`:

```
parser = argparse.ArgumentParser(description='insert protein.')
parser.add_argument('-full_protein_ID', help='protein name', default='')
args = parser.parse_args()
```

needs to be manually edited with the `$` command in the following manner.

```
ID = 3X2MA
! Python Protein_ID.py
-full_protein_ID $ID
```

Using Colab, the process is simplified, allowing the developer to write directly

```
full_protein_ID ="3X2MA" #@param
```

also with the possibility of defining its type to avoid variable format issues, as following:

```
{type:"string"}
```

The user can choose any protein ID, but to make this operation more accessible to an user without knowledge of proteins IDs, we download a reference list of protein initially hosted by the Dunbrack lab, which we currently re-hosted in a Github repository, due to the regularity with which their database is updated.

This list is significant since it provides users with examples that may be used without looking up protein identification codes online. However, it should be noted that the protein data bank contains a higher number of proteins; yet, with over six thousand IDs, it arguably provides a good starting point of enquiry.

To prevent unwanted black spaces, we use the `regular expression` module, deleting them.

```
full_protein_ID=re.sub(r"\t", "full_protein_ID")
full_protein_ID=re.sub(r" ", "full_protein_ID")
```

Using the request Python module to perform an API request, the program connects with the protein data bank repository and scrapes a given protein.

```
requests.get('https://files.rcsb.org/download/{}.pdb'.format(protein_ID.lower())).text
```

During this process, a protein ID is queried in the protein data bank, returning proteins structures distributed in three dimensions, with axes x, y, and z, in a standardised file format (PDB).

To perform this operation, however, we delete the last character, considering the occurrence of different notations.

```
protein_ID = pdb.IDs.loc[pdb.iloc[:,0].str.contains(full_protein_ID)].iloc[0][:-1]
```

To interact with the JSON file, we match lines beginning with the word ‘ATOM’ (Combs, 2020) or as advised by Combs, ’HETATOM’, which stands for hetero atoms, extracting strings that contains the c-alpha coordinates.

```
if 'ATOM' == line[0:4] or 'HETATOM' == line[0:7]:  
    if line[13:15] == 'CA' and line[16] in ['A', ' '] and line[21] == 'A':  
        x = float(line[30:38])  
        y = float(line[38:46])  
        z = float(line[46:54])
```

Once these coordinates are extracted, it is possible to create a specific new attribute to these proteins.

To compute these attributes, we propose two established theorems and an experimental one to show that this project can elaborate arbitrary given values, as long as they somehow depend on the protein structure.

The two established theorems are convex hull to extract protein volume, and rotating callipers to compute the farthest distance between protein points.

The experimental approach we use is the farthest neighbours algorithm, which is adapted to calculate the maximum width a given protein has.

The convex hull is both computed manually, using a graham scan, and with Scipy library with the function `scipy.spatial.ConvexHull`.

This duality happens because, given the first approach, we have a coordinate-based system, which is further used to extract the rotating callipers (Eppstein, 2002), in opposition with the Scipy function, which returns an object challenging to interact with. However, Scipy convex hull is maintained, where, is a pretty straightforward operation extracting its volume (`scipy.spatial.ConvexHull` — SciPy v1.7.1 Manual, 2021).

Rotating callipers are relatively more complex to understand, where we suggest this method to rotate the flatter perimeter of our convex hull against the same shape, finding the farthest distance between its points (Shamos, 1978).

The farthest neighbours algorithm is not a fully documented procedure; however, its result is dependent on the protein structure, so we decided to include it in this documentation,

acknowledging that it works only if the mean value of the protein, minus one, is smaller than the overall structure.

There are two debug for this event, brute-forcing a minus n to the mean, to produce an increasingly smaller value, or as we adopted because it is a relatively rare occurrence, a suppression of the relative errors, inputting a null value instead.

This sometimes happens because the algorithm was proposed to work for a circle, were given a random point, it finds its opposite (Le Bourdais, 2015).

Editing this random point to its mean should find its further opposite point in a three-dimensional space.

```
n=0  
empty_list.append(all_points.pop^((all_points_mean - n)))
```

After we obtained these values, we use them to perform modulation, adding these new conditions to the amplitude of a carrier wave in the following manner.

The carrier wave is defined with a sample rate of 44100; a standardise value that reflects the highest frequency audible by humans.

Then we scale the convex hull volume, dividing it by one thousand, to obtain more meaningful insights.

Time of the carrier wave is given by evenly spaced numbers between zero and ten through the sample rate divided the length of the coordinates of the protein and scaled dividing it by one hundred.

This operation debugs the situation where proteins are too short to produce sounds, making them all of a similar length.

The amplitude of the carrier signal is controlled, which mean that we only consider the max value an int16 can produce, or a range of values from negative 32768 to positive 32767, to avoid signal outside this spectrum.

Its volume is defined by the amplitude, multiplied by a sine wave described in the following code.

```
volume_signal = amplitude * np.sin(2. * np.pi * hull_volume_freq * t)
```

At this point, we modulate or wave shape against the convex hull, the rotating callipers, and the farthest neighbours algorithm.

To do so we, centre the original protein coordinates and label them as positive or negative based on their axis position.

After this operation is performed, we assign positive coordinates incrementing even values in the index; in this way, positive indexes are even, and negative indexes are odd.

For example:

```
pos_neg_xyz[0]  
indicates a positive x value, and  
pos_neg_xyz[3]  
indicates a negative y value.
```

Then we repeat a similar operation for rotating callipers and farthest neighbours algorithm in the following fashion.

```
rot_pos_x = amplitude * np.sin(2. * np.pi * (max_distance+pos_neg_xyz[0]) * t)
rot_neg_ax = amplitude * np.sin(2. * np.pi * (max_distance+pos_neg_xyz[1]) * t)
```

```
far_pos_x = amplitude * np.sin(2. * np.pi * (max_width+pos_neg_xyz[0]) * t)
far_neg_x = amplitude * np.sin(2. * np.pi * (max_width+pos_neg_xyz[1]) * t)
```

Alternatively, if the result was null, as explained before, we query the relative variable in globals, which attribute zero value to the equation.

```
far_pos_x = 0
far_neg_x = 0
```

To conclude, these values are added to the original data.

```
output = volume_signal + hull_pos_x + hull_neg_x + rot_pos_x + rot_neg_ax + far_pos_x +
far_neg_x
```

Performing this modulation, we noted that proteins often express different characteristics, and for this reason, we propose this result to become an earcon or nonverbal audio signal (Blattner, Sumikawa and Greenberg, 1989) to distinguish specific proteins, to be further stored in a sort of dictionary.

## **4 Evaluation**

This project aimed to be a tool to produce creative earcons to be attributed to proteins. Different samples (Appendix 1-6) show various degrees of unicity, where some tend to be complex to distinguish. At the current stage, with the parameters mentioned above, it might be possible to distinguish some proteins, but it is not quite adequate yet to build a proper earcon system. However, this project considers this challenge in its design, allowing the addition of other, protein-dependent functions, such as the earlier mentioned water solubility or considering additional geometrical features. The addition of multiple values should result in effective earcons, which can be further stored and retrieved in a Python dictionary or as a novel dataset.

## **5 Reflection**

As presented from the result section, the creation of sounds using protein-dependent value might produce, in the future, a functioning system of earcons.

We have seen how there has been an eradicated historical interest in the cross-contamination between biology and art, producing results where artists were helping scientists, both to communicate with the public and to design new interdisciplinary researches.

This interest undoubtedly influenced my practice, where this research started with an exhibition at the Maxxi museum in Rome, under the collective name of Umanesimo Artificiale, in collaboration with the art director Filippo Rosati and the sound artist Moisés Horta Valenzuela.

This exhibition initiated my interest in this research, where Rosati, which suffers from a genetic mutation called ABCD1, connected us to create art to generate public awareness.

This happening caused a shift in myself, conjecturing the role of artwork that crosses the barrier between disciplines.

Looking at myself, I would never know of this illness otherwise, which made me reconsider my position as an artist.

After all this evidence, we can take a step back and see such exciting works, which undoubtedly changed the way tools are thought by artists and scientists, leading to increasing numbers of experiments related to today issue, such as the Covid pandemic.

We saw how Fluxes in the sixties paved the way to a new way of thinking and conduct research, and especially now, we can speculate in a resurgence of these contaminations thanks to new technologies and a novel spirit from artists and scientists to promote this cross-contamination.

## 6 Further Works

As we acknowledged before, this project rotates on the idea that proteins are transformed to audio, using modulation of a carrier signal that is sensitive to unique protein structures.

The project takes this idea deeply into account, where here, we propose a small example with diverse parameters.

Instead of considering convex hull, rotating callipers and farthest neighbours algorithm, we consider the one-dimensional discrete Fourier Transformation (RFFT) of the coordinates of the centred protein using the Scipy module `scipy.fft.rfft`

To perform this operation we use something called "Fast Fourier Transform" (FFT), which is an algorithm to retrieve the spectral oscillations of a signal, such as its phase and amplitude (*Fast Fourier Transformation FFT, 2021*).

We use this function to edit amplitude, calculating the distance from both real and imaginary RFFT to the origin, using Pythagoras's theorem, where the square root of the real RFFT (multiplied by the power of two) is summed to the square root of the imaginary RFFT (multiplied by the power of two).

$$\sqrt{rfft(\mathbb{R}x)^2 + rfft(\mathfrak{I}x)^2}$$

For example, the x-axis will be transformed as follows.

```
x_rfft = np.fft.rfft(x)
x_freq = []
x_amp = []
for i in range(len(x_rfft)):
    x_freq.append(i)
    x_amp.append(np.sqrt(x_rfft[i].real**2+x_rfft[i].imag**2))
    x_phase.append(np.arctan2(x_rfft[i].imag,x_rfft[i].real))
```

Considering phasing as an additional parameter, we suggest a similar approach, calculating the arc tangent in the range of the real RFFT and imaginary RFFT.

$$\text{Arctan}(rfft(\mathbb{R}x), rfft(\mathfrak{I}x))$$

```
x_phase = []
for i in range(len(x_rfft)):
    x_freq.append(i)
    x_phase.append(np.arctan2(x_rfft[i].imag,x_rfft[i].real))
```

(Appendix 7a-12a).

From this point, we can create pairs of top recurring frequencies, map them to amplitude, and repeat the previously mentioned modularity with a few adjustments to implement these new parameters.

```
x_highest_freqs = [x for _, x in sorted(zip(x_amp, x_freq), reverse=True, key=lambda pair: pair[0])][:10]
```

This time, we set the base frequency at an arbitrary given value, such as one hundred, and the time is set differently than before, spacing numbers between zero and one without further dividing it. The parameters for amplitude and sample rate are the same as before, resulting in the following:

```
samplerate = 44100
base_frequency = 100
t = np.linspace(0., 1., samplerate)
amplitude = np.iinfo(np.int16).max
```

Given these new inputs, we proceed as before, offering two different outputs: In the first, we scale the frequency for a factor of the twelfth root of two  $\sqrt[12]{2}$  or  $2^{1/12}$ , testing the idea of the musical interval of a semitone in twelve-tone equal temperament (Helmholtz and Ellis, 2009).

```
x_signal += list(amplitude * np.sin(2. * np.pi * (base_frequency*(2***(1/12))**f) * t))
```

(Appendix 7b-12b).

In the second, we multiply the base frequency by its highest frequency.

```
x1 = amplitude * np.sin(2. * np.pi * (base_frequency*x_highest_freqs[0]) * t)
```

(Appendix 7c-12c).

Showing these examples, we aim to introduce and show how modularity can become a quick and scalable system for prototyping and integrate new earcons systems in future works, potentially expanding this concept of sonification to a broader range of scientific problems.

## **Conclusion**

This project looked at the resurgence and evolution of a link between science and the arts, particularly sound and biology, which has been studied since the 1980s, investigating its initial response, which stems from the late 1970s with Hofstadter theories, and newer possibilities, including the MIT Laboratory for Atomistic and Molecular Mechanics, attempt to comprehend and appreciate how respondents' and practitioners' perspectives have changed over the last 50 years. These novel considerations are feasible as a direct consequence of an exhaustive evolution of the concept of art and science, in which, beyond a plausible interconnection between these two disciplines, we find a novel way of understanding information and the audience role allowing a more comprehensive definition of these practices. However, recognising that this kind of awareness often necessitates a thorough examination of the subject, the science-art relationship may not be immediately apparent to all respondents, and assumptions on how and to what extent the procedure can be extended before losing its definition have not been evaluated yet. Furthermore, the effort to which the public is willing to stretch the concept of a discipline has not yet been recognised, and it will surely take further investigation to grasp how, whenever achievable, to establish a creative and widely accepted manner to identify science-art contacts and their social ramifications.



## Bibliography

- Ditto, B. Zuk, M. and Robinson, K., 2021. [online] Louis200.com. Available at: <<https://louis200.com/projects/17>> [Accessed 18 September 2021].
- De Leon, T., 2019. *Spider web music: An inspiring harmony of art and science*. [online] MIT News | Massachusetts Institute of Technology. Available at: <<https://news.mit.edu/2019/spider-web-music-inspiring-harmony-art-and-science-0108>> [Accessed 18 September 2021].
- Higgins, D. and Higgins, H., 2001. Intermedia. *Leonardo*, Volume 34, Number 1, MIT press, pp.49-54.
- Friedman, K., 1998. *The Fluxus Reader*. John Wiley & Son Ltd
- Cook, S., 2016. *Documents of Contemporary Art: Information*. Whitechapel Gallery and MIT press, pp.44-47.
- Entangled Others Studio. 2021. *Beneath the neural waves*. [online] Available at: <<https://entangledothers.studio/beneath-the-neural-waves/>> [Accessed 18 September 2021].
- Hermann, T., Hunt, A. and Neuhoff, J., 2011. *The Sonification Handbook*. Logos Publishing House, p.14.
- Le Feuvre, L., 2010. *Documents of Contemporary Art: Failure*. Whitechapel Gallery and MIT press.
- Wilson, S., 2003. *Intersections of Art, Science, and Technology*. The MIT Press.
- Gewin, V., 2021. *How to shape a productive scientist–artist collaboration*. [online] Nature.com. Available at: <<https://www.nature.com/articles/d41586-021-00397-1>> [Accessed 21 September 2021].
- Paterson, S., Le Tissier, M., Whyte, H., Robinson, L., Thielking, K., Ingram, M. and McCord, J., 2020. Examining the Potential of Art-Science Collaborations in the Anthropocene: A Case Study of Catching a Wave. *Frontiers in Marine Science*, 7.
- Kelly, C. and Woodley, C., 2011. *Documents of Contemporary Art: Sound*. Whitechapel Gallery and MIT press.
- Sawe, N., Chafe, C. and Treviño, J., 2020. Using Data Sonification to Overcome Science Literacy, Numeracy, and Visualization Barriers in Science Communication. *Frontiers in Communication*, 5, p.46.
- Neuhoff, J., 2019. Is Sonification Doomed to Fail?. *The 25th International Conference on Auditory Display (ICAD 2019)*.
- Temple, M., 2017. An auditory display tool for DNA sequence analysis. *BMC Bioinformatics*, 18(1).
- Hofstadter, D., 1979. *Gödel, Escher, Bach : an Eternal Golden Braid*. New York : Basic Books.
- Raven, P., 1984. Bacterial problems. *Nature*, 310(5973), pp.96-96.
- Ohno, S. and Ohno, M., 1986. The all pervasive principle of repetitious recurrence governs not only coding sequence construction but also human endeavor in musical composition. *Immunogenetics*, 24(2), pp.71-78.
- Dunn, J., and Clark, M. A., 1999. “Life Music”: The Sonification of Proteins. *Leonardo*, 32(1), 25–32. [online] Available at: <<http://www.jstor.org/stable/1576622>> [Accessed 18 September 2021].
- Garcia-Ruiz, M. and Gutierrez-Pulido, J., 2006. An overview of auditory display to assist comprehension of molecular information. *Interacting with Computers*, 18(4), pp.853-868.
- King, R. and Angus, C., 1996. PM—Protein music. *Bioinformatics*, 12(3), pp.251-252.

Hansen, M., Charp, E., Lodha, S., Meads, D. and Pang, A., 1998. PROMUSE: A System for Multi-Media Data Presentation of Protein Structural Alignments. *Biocomputing '99*.

Grond, F. and Dall'Antonia, F., 2008. SUMO: A Sonification Utility for Molecules. *The 14th International Conference on Auditory Display (ICAD 2008)*.

Férey, N., Nelson, J., Martin, C., Picinali, L., Bouyer, G., Tek, A., Bourdot, P., Burkhardt, J., Katz, B., Ammi, M., Etchebest, C. and Autin, L., 2009. Multisensory VR interaction for protein-docking in the CoRSAIRe project. *Virtual Reality*, 13(4), pp.273-293.

Hermann, T., and Ritter, H., 1999. Listen to your Data: Model-Based Sonification for Data Analysis. In G. E. Lasker & M. R. Syed, eds. *Advances in intelligent computing and multimedia systems*. Windsor, Ontario: Int. Inst. for Advanced Studies in System research and cybernetics, pp. 189-194.

Rau, B., Friess, F., Krone, M., Muller, C. and Ertl, T., 2015. Enhancing visualization of molecular simulations using sonification. *2015 IEEE 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR)*.

Ballweg, H., Bronowska, A. K., and Vickers, P., 2016. Interactive Sonification for Structural Biology and Structure-based Drug Design. In T. Hermann, R. Bresin, & J. Yang (Eds.), *ISon 2016, 5th Interactive Sonification Workshop* (pp. 41–47). Bielefeld, Germany: CITEC, Bielefeld University.

Yu, C. and Buehler, M., 2020. Sonification based de novo protein design using artificial intelligence, structure prediction, and analysis using molecular modeling. *APL Bioengineering*, 4(1), p.016108.

Buehler, M., 2020. *Nanomechanical sonification of the 2019-nCoV coronavirus spike protein through a materiomusical approach*. [online] arXiv.org. Available at: <<https://arxiv.org/abs/2003.14258>> [Accessed 18 September 2021].

Gärtner, B. and Hoffmann, M., 2014. *Computational Geometry Lecture Notes HS 2013*, pp.83-87.

Combs, F., 2020. *Turning a Protein into Music with Python (Part 2) – Harmony*. [online] Medium. Available at: <[https://levelup.gitconnected.com/turning-a-protein-into-music-with-python-part-2-harmony-a3c2ffa748be?source=user\\_profile-----4-----](https://levelup.gitconnected.com/turning-a-protein-into-music-with-python-part-2-harmony-a3c2ffa748be?source=user_profile-----4-----)> [Accessed 18 September 2021].

Docs.python.org. 2021. *argparse – Parser for command-line options, arguments and sub-commands – Python 3.9.7 documentation*. [online] Available at: <<https://docs.python.org/3/library/argparse.html>> [Accessed 26 September 2021].

Eppstein, D., 2002. *Convex hull and diameter of 2d point sets*. [online] Code.activestate.com. Available at: <<https://code.activestate.com/recipes/117225-convex-hull-and-diameter-of-2d-point-sets/>> [Accessed 22 September 2021].

Docs.scipy.org. 2021. *scipy.spatial.ConvexHull – SciPy v1.7.1 Manual*. [online] Available at: <<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.ConvexHull.html>> [Accessed 22 September 2021].

Shamos, M., 1978. *Computational Geometry*. Yale University.

Le Bourdais, F., 2015. *The Farthest Neighbors Algorithm I*. [online] Frolian's blog. Available at: <<https://flothefof.github.io/farthest-neighbors.html>> [Accessed 22 September 2021].

Blattner, M., Sumikawa, D. and Greenberg, R., 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 4(1), pp.11-44.

Nti-audio.com. 2021. *Fast Fourier Transformation FFT - Basics*. [online] Available at: <<https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft>> [Accessed 25 September 2021].

Helmholtz, H. and Ellis, A., 2009. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*.

## **List of Figures**

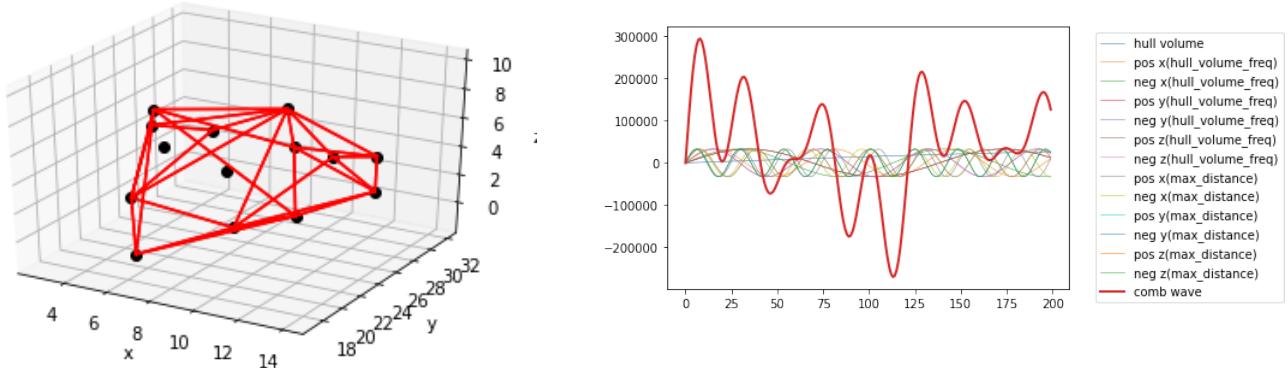
Ditto, Zuk and Robinson 2021, *Louis200*.

Saraceno and MIT Laboratory for Atomistic and Molecular Mechanics, 2018, *On Air*

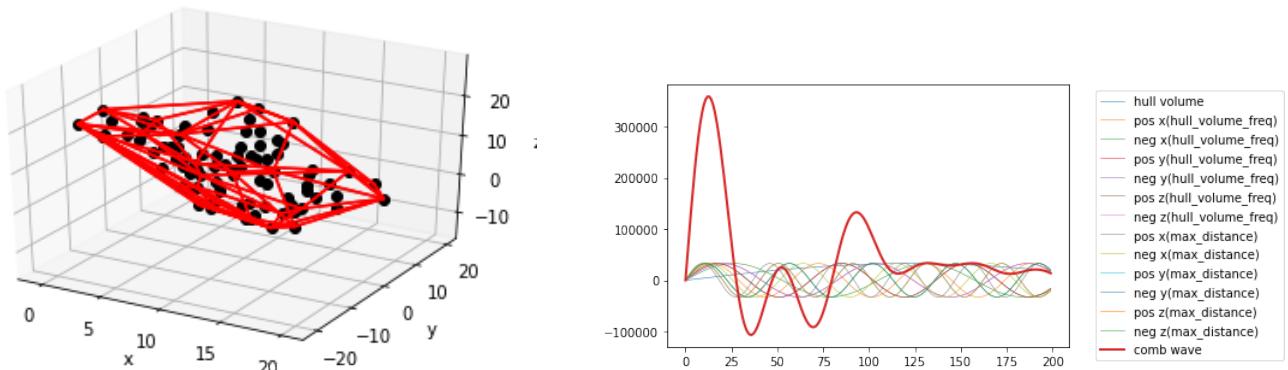
Entangled Others Studio, 2021, *Beneath the neural waves*.

## Appendix

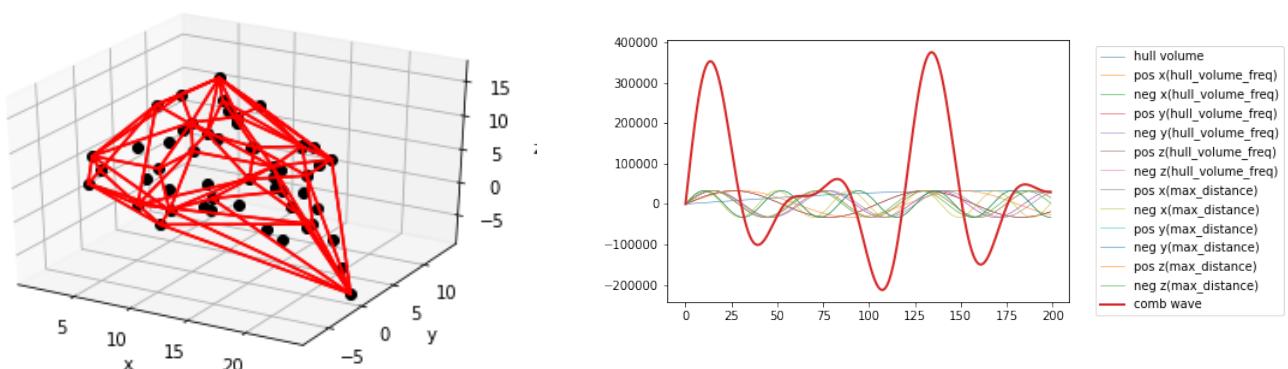
### Appendix 1: 1QOWB



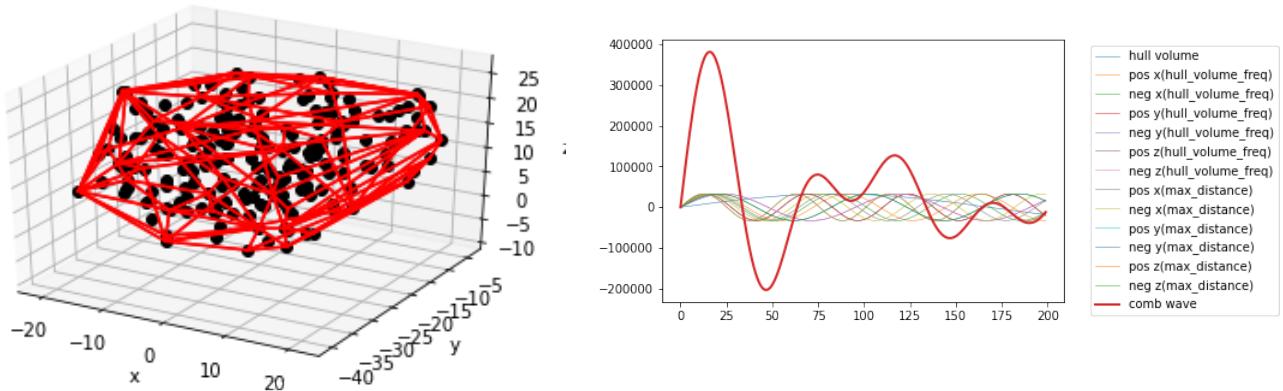
### Appendix 2: 3CALD



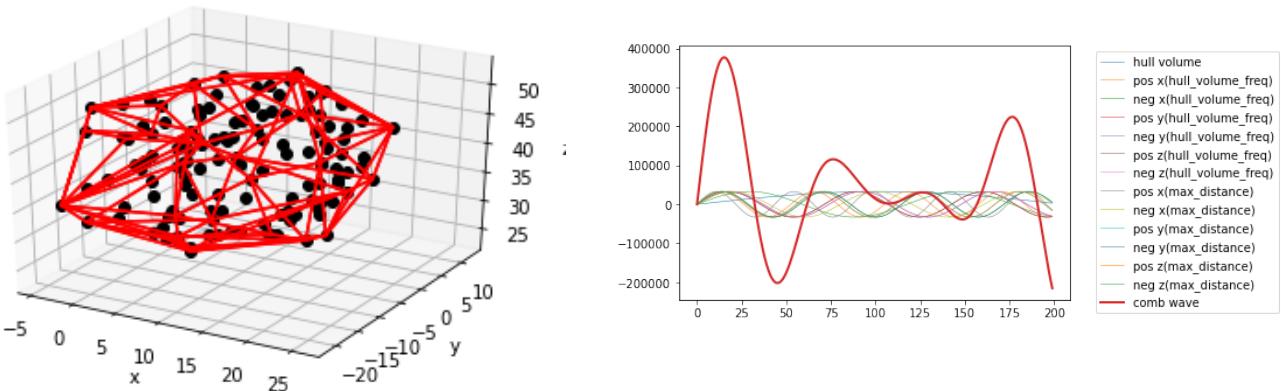
### Appendix 3: 5NW3A



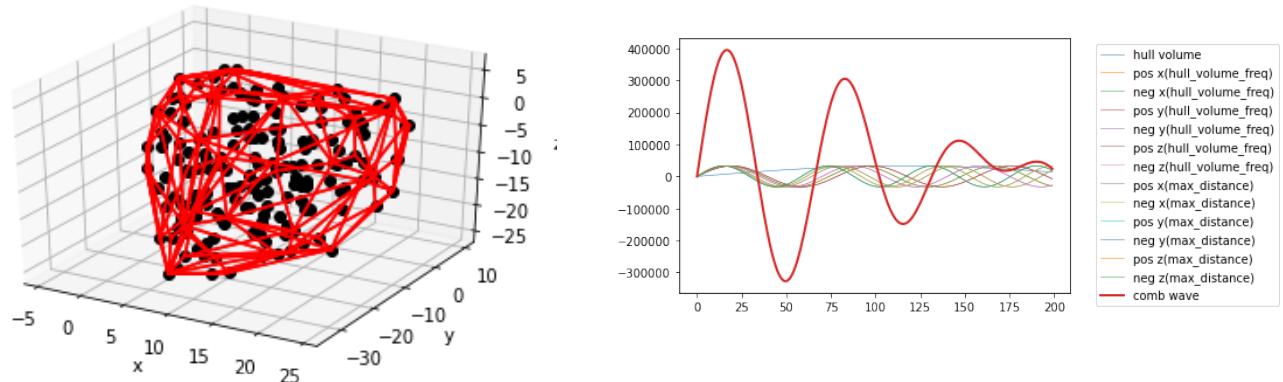
#### Appendix 4: 6SBAB



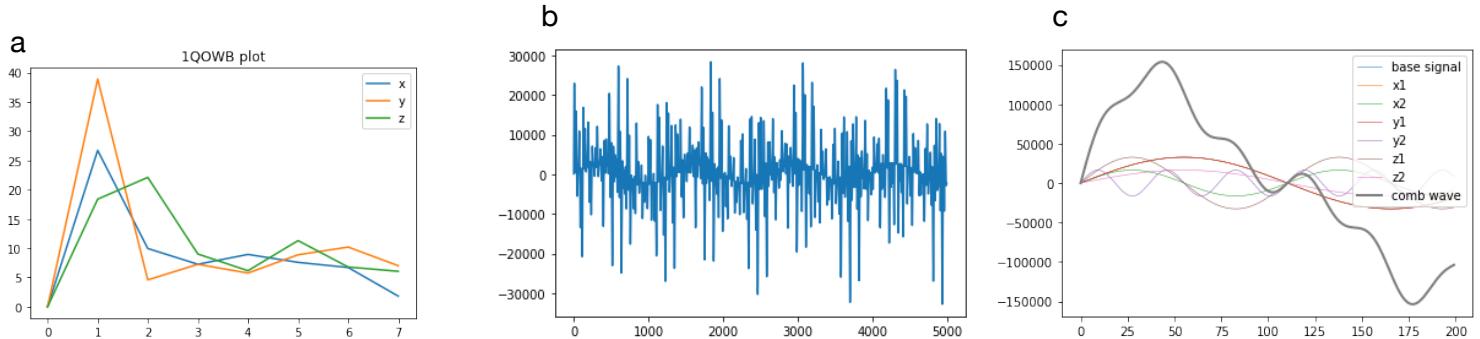
#### Appendix 5: 6S2MA



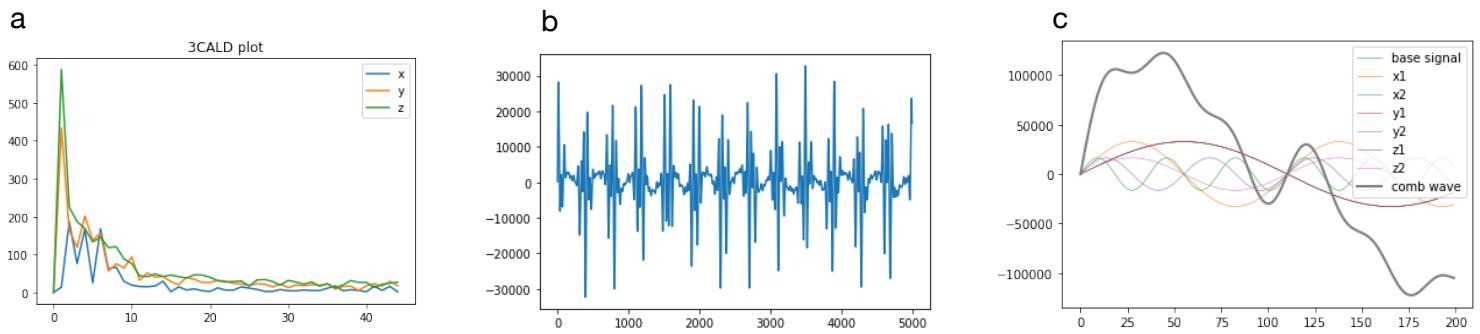
#### Appendix 6: 3X2MA



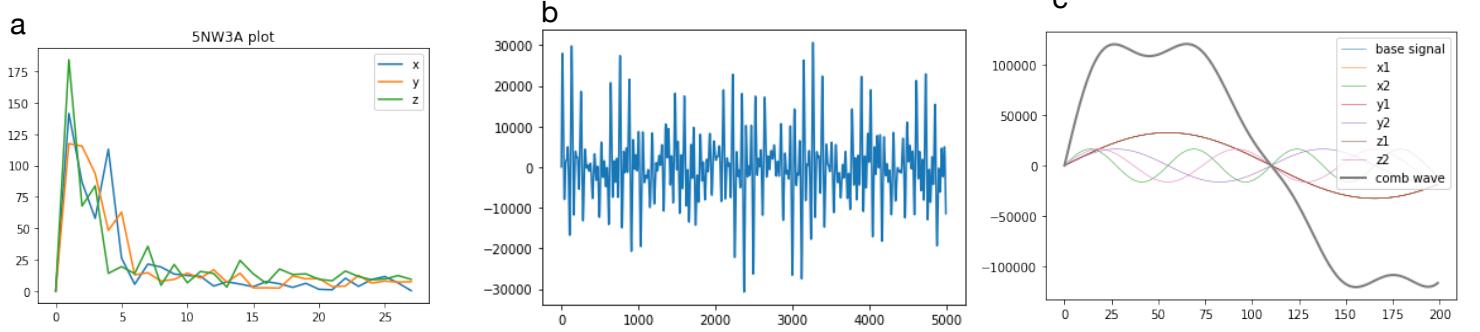
## Appendix 7: 1QOWB



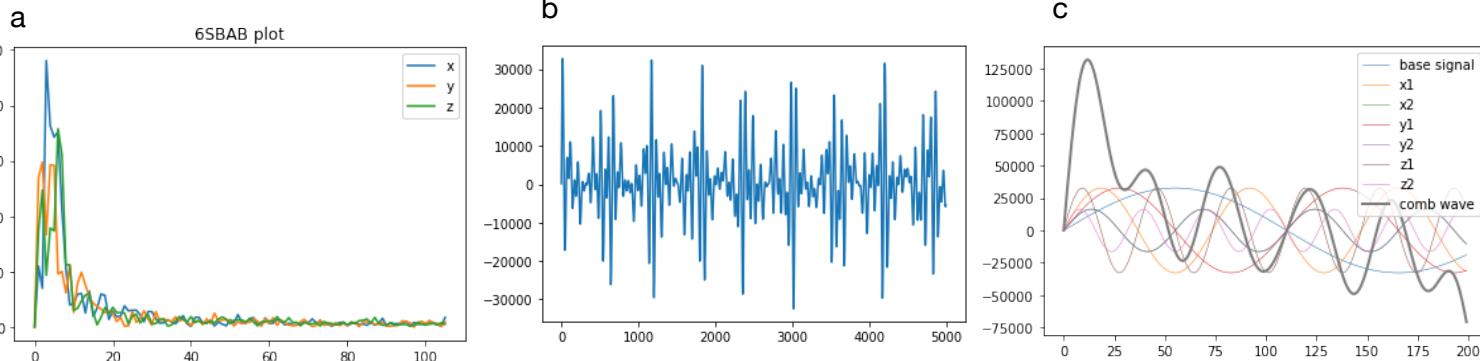
## Appendix 8: 3CALD



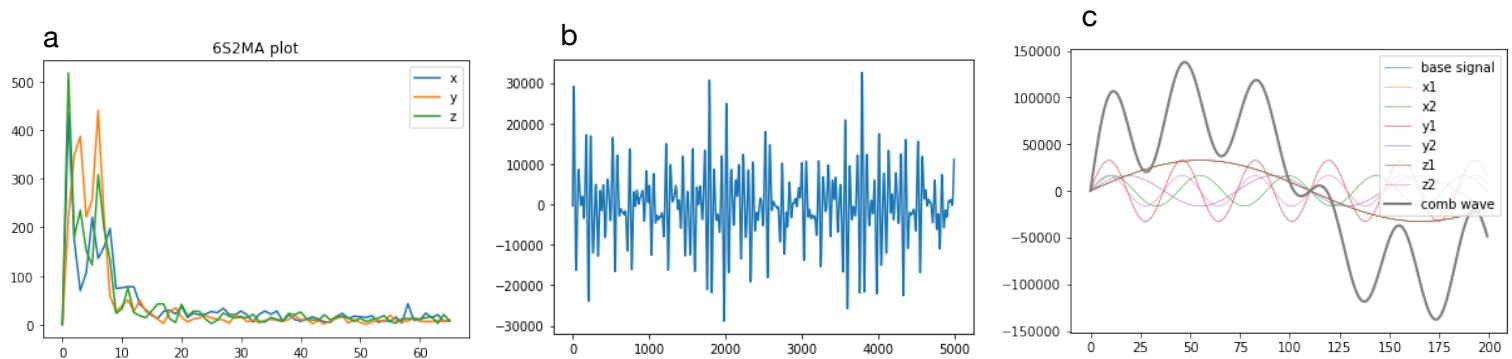
## Appendix 9: 5NW3A



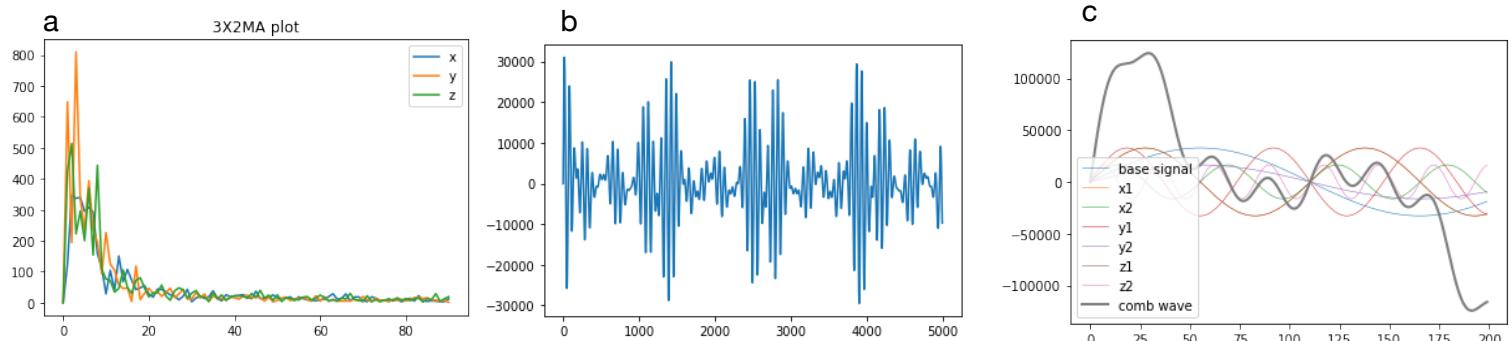
## Appendix 10: 6SBAB



## Appendix 11: 6S2MA



## Appendix 12: 3X2MA



## **Audio Links**

1QOWB

Geometric: <https://soundcloud.com/fabio-rovai/1qowbgeometric/s-bBWgCSxMDht>

Twelfth Root: <https://soundcloud.com/fabio-rovai/1qowb-12/s-OZr18gDybEy>

Highest Frequency: <https://soundcloud.com/fabio-rovai/1qowb-high/s-pdkaoL69HYw>

3CALD

Geometric: <https://soundcloud.com/fabio-rovai/3caldgometric/s-7vljmSNDAeU>

Twelfth Root: <https://soundcloud.com/fabio-rovai/3cald-12/s-W7BZOrD4J1t>

Highest Frequency: <https://soundcloud.com/fabio-rovai/3caldhigh/s-gdanARqXvZE>

5NW3A

Geometric: <https://soundcloud.com/fabio-rovai/5nw3a-geo/s-iaIr4O1UcDw>

Twelfth Root: <https://soundcloud.com/fabio-rovai/5nw3a-12/s-148d19VCjZs>

Highest Frequency: <https://soundcloud.com/fabio-rovai/5nw3a-high/s-U6cYMqdvpdo>

6SBAB

Geometric: <https://soundcloud.com/fabio-rovai/6sbabgeo/s-IHzG8bDMCt>

Twelfth Root: <https://soundcloud.com/fabio-rovai/6sbab-12/s-vyA9keH31Jr>

Highest Frequency: <https://soundcloud.com/fabio-rovai/6sbabhigh/s-42egzQBzIQ0>

6S2MA

Geometric: <https://soundcloud.com/fabio-rovai/6s2ma-geo/s-sA6L78R2cCP>

Twelfth Root: <https://soundcloud.com/fabio-rovai/twelve-tone-comb-wave/s-VOYeCrQQmYH>

Highest Frequency: <https://soundcloud.com/fabio-rovai/6s2mahigh/s-eejsK0SHG9A>

3X2MA

Geometric: <https://soundcloud.com/fabio-rovai/3x2ma-geo/s-jzSefJPldVh>

Twelfth Root: <https://soundcloud.com/fabio-rovai/twelve-tone-comb-wave-1/s-AOd5l0SBaCc>

Highest Frequency: <https://soundcloud.com/fabio-rovai/3x2mahigh/s-ZQq8irD6QEg>