DNA SONIFICATION USING 8-CHANNEL AUDIO FOR DATA ANALYSES AND MUSIC COMPOSITION

Mark D. Temple

School of Science,
Western Sydney University,
Penrith South DC,
NSW, 1797, Australia.
m.temple@westernsydney.edu.au

ABSTRACT

DNA sequences contain vast amounts of biological data and computer algorithms play an important role in processing these data for human inspection. Here we describe an updated computer-generated auditory display tool to be used as stand-alone audio or as a complement to a visual display for DNA sequence inspection. The auditory display uses musical notes to represent the data in relation to the process of gene expression or DNA replication. Given the use of musical notes in the auditory display raises the possibility these may be considered algorithmic music. To pursue this notion further the auditory displays were used in a music studio setting outside of the science laboratory. Musicians were challenged to play in-sync with the audio and to embellish the melodic and harmonic content of the auditory display. New music compositions featuring the auditory displays were recorded and performed live in outreach events to promote a wider understanding of the processes of gene expression and DNA replication and how gene sequence information affects human health conditions.

1. INTRODUCTION

Within the science community the use of audio for data analysis is becoming increasingly popular. This approach is referred to as sonification and the resultant auditory display can be used to supplement a visual display, or it can be used as an analytical tool in its own right. An auditory display of science data may possess rudimentary musical attributes which raises the proposition that it could be designed to produce rudimentary algorhythmic music with a systematic relationship to the data [1].

As is often the case with multidisciplinary approaches, the auditory displays can sometimes confound either the science or arts community. If the auditory display is designed solely on scientific principles, without regard to music theory or artistic performance then it is unlikely to sound particularly musical, but it may be well regarded in the science community as an analytical tool. For instance, an alert sound could be used to distinguish between malignant and benign lesions when analysing skin images [2]. Conversely if the auditory display is mapped to pay strict adherence to musical theory or designed for performance then it may have less discriminatory power as a tool for data analysis [3]. Either of these approaches are completely valid within their discipline [4]. The motivation is to find a middle ground whereby the auditory display can function as a tool for data analysis and have intentional musical attributes This is advantageous to avoid analytical fatigue [5], and enhance user engagement [6] whilst listening. Additionally, intended musical attributes may be of interest to the creative arts and wider community providing an opportunity promote the research more widely [7].

2. SONIFICATION OF BIOLOGICAL DATA

Research into cell biology has been revolutionised by the sequencing of many genomes [8], the development of genome wide experimental approaches [9] and the application of robotics to perform high throughput experimentation. This has led to the accumulation of vast amounts of biological data that requires a computational approach for its analysis. Additionally, the ability of computers to generate audio from various data sources has improved greatly over the past few years. This has encouraged more scientists to sonify data and it is becoming more widely accepted as a legitimate tool for data analyses [10].

One of the many challenges for a cell biologist is to understand biological processes and functions at both an organism and cellular level [11, 12]. Cells organelles can be visualised through a light microscope, and other higher resolution techniques exist to visualise macromolecules within the cell. As a molecular biologist we think about these macromolecules as molecular machines that have a dynamic structure [13] that is related to function [14]. The cellular environment is highly dynamic and one could fancifully imagine that movement and vibration within the cell could be amplified as sound. If it were possible, it would be interesting listen to vibrations of the actual process of transcription, translation, protein folding or the myriad of other cellular processes driven by molecular machines [15]. Currently this is not possible, however, much is known about the logic of how biological sequences are processed in cells [16] and this information can be modeled by algorithms to generate an auditory display [17].

In addition to the sonification of known biological processes, it is more common to sonify research data from biological experiments such as microarray data [18], data relating to the microbiome [10], protein-protein interactions [19] and the complex amino acid sequence and secondary

structure of proteins [20]. These approaches use auditory displays to represent results from matrices, tables or sequence data that are normally visualised using approaches such as graphs, heat-maps or 3D-structure representations. These approaches are often multidisciplinary and bridge the divide between science analyses and creative music production.

3. SONIFICATION OF DNA SEQUENCES

The focus of this research paper is the sonification of DNA sequence data. Biological sequences such as genomic DNA contain vast amounts of information, such as a gene coding regions that contain the instructions for the makeup of proteins. It is a major challenge to understand how variations in DNA sequences can cause human disease [21] and to identify the function of genomic regions such as promoters, introns, exons, enhancers and other yet to be characterised. The motivation of this sonification is to contribute to the analyses and display of DNA sequence information and to better understand the role of the genome in human health.

The question is, how can sonification be applied to generate an auditory display? This question is partly influenced by the fact that the auditory display proceeds over time. A time-based approach adds another dimension to a static representation and exploration of data [22]. An analogy can be made between the way words on a page are read over time and the way information in a DNA sequence can be sonified over time.

At a molecular level the chemical composition of DNA is elegant in its simplicity but it is the functional expression of this sequence information that is complex. DNA is composed of only four chemical building blocks (referred to as nucleotide bases) which consist of guanine (G), adenine (A), thymine (T) and cytosine (C). DNA sequences stored in computer files use these letters to represent the linear sequence of adjacent bases in a biological DNA macromolecule. These four bases are joined together in the human genome to give a sequence of over 6 billion bases.

In the cellular environment there are large macromolecule assemblies (such as polymerases and ribosomes) that act as molecular machines to bind to biological sequences and physically move along the DNA to carry out a function. These processes have an implicit time component. This movement and output can be replicated by computer algorithms to map the DNA sequence into an auditory display [23] in the same way that the cell would use the information for gene expression (protein synthesis) or DNA replication.

It is common in sonification to use musical notes as the building blocks of an auditory display [24, 25]. This approach is modeling the process of translation (part of gene expression). The process of translation in live cells uses a nucleotide sequence to assemble amino acids as building blocks to make a peptide or protein. Translation occurs at an average rate of approximately 20 amino acids [26] per second. If this rate were replicated in the auditory display it would correspond to 60 nucleotide bases per second or 3,600 beats or bases per minute (bpm.) At such a high tempo the individual nucleotides would not be comprehensible to the human ear and typical a slower rate is chosen for base sequence comprehension.

Taken in isolation, each individual base (represented as a letter) does not hold enough information to determine biological function since at any position in the 6 billion base sequence there are only four options. If we continue to develop this analogy between bases and letters, then it is the combination of bases into words (or DNA motifs) that is important to impart biological information. If we consider two adjacent bases on a strand of DNA, there are 16 possible paired combinations (4x4) which are again not typically informative during translation. However, sonification of these is important to provide an audio backdrop upon which the other layers of audio (from other DNA motifs) will sit. An annotated video of the sonification tool is available as Supplementary Media to demonstrate examples (as indicated in Table 1). I use these example in outreach events to explain how biological processes can be sonified and their role in normal cell biology and human health. The audio backdrop is demonstrated in Sonification Video e.g. No.1.

Table 1. Index for five examples in the “Sonification Video”

	Start Time	**DNA sequence characteristics**
No.1	0:01 sec	Without START or STOP codons
No.2	0:28 sec	1 START codon and 1 STOP codon
No.3	1:04 min	Multiple START and STOP codons
No.4	1:42 min	Repetitive Telomeric DNA
No.5	2:20 min	TP53 gene coding sequence

Things get more interesting once combinations of three adjacent bases are considered since these are biologically significant, these are called codons. There are 64 possible combinations of codons (4x4x4) and each of these has special meaning in the process of gene expression (transcription and translation). For instance, the codon sequence ATG often signifies the START of a gene sequence (though it may also code for a Methionine amino acid residue), this and the following codons are used by the cell to determine the order of amino acids to be expressed as a protein (as shown in Sonification Video Example No.2).

Each codon is mapped by the cellular machinery to one of twenty possible amino acids. This mapping is known as the genetic code, which is a unifying principle across all cellular organisms, and is used by biologists to understand how genetic information is used to make over 35,000 different proteins in human cell. There is another important group of codons that are known as STOP codons, and these signify the end of a gene sequence. This level of understanding is rudimentary, yet it helps explain the control of gene expression at tens of thousands of locations within the genome, hence the rules that govern gene expression or DNA replication were chosen for sonification.

When analysing DNA sequences it is important to know that there are three ways (three reading frames) in which the information can be interpreted into codons. This may be explained by thinking about a sequence of numbers from 1 to 9. The codons of the first reading frame would be the three bases in positions 123-456-789, the second frame would be those in position 234-567-89... and the third (last) frame would be those at 345-678-9... Each reading frame produces an entirely different sequence of bases or amino acid residues and typically only one reading frame is functional within human cells.

START codons may occur in either reading frame to turn on sonification of subsequence codons in the sequence whereas STOP codons were used to turn off the sound of subsequent codons. This is demonstrated in Sonification Video Example No.3 where START codons were positioned in each reading frame.

Only less than 5% of the genome codes for proteins and it is important to identify these sequences. An example of a non-coding sequence (which does not code for a protein) is shown in Sonification Video Example No.4. This is a Telomeric DNA sequence consisting of imperfect CCCTAA repeats. This sonification highlights how the embedded STOP (TAA) codons trigger over 50 high pitch alert notes (in the 496 base sequence). The alert sounds are analogous to the use of colour or bold highlighting of features in a visual display as one might see in a genome browser [27]. However, in the absence of the alert, the underlying audio is comparatively musical due to the repetition of a melodic phrase with slight variation. Lastly Sonification Video Example No.5 demonstrates the display of the naturally occurring TP53 gene coding sequence. Predominantly, the sound of this sequence differs from the prior Telomer since in reading frame 1, a START codon triggers a stream of 393 codon pulses with only one STOP codon at the end of the 1222 base sequence. Additionally other START and STOP codons occur in frames 2 and 3 giving rise to additional audio layers.

In these video examples, the rules and logic of gene expression are used to determine the order of notes which is analogous to the order of amino acid residues in a protein (as illustrated in Figure 1). Therefore, by listening to the display it is possible to hear an open reading frame (a term which is used to identify a possible gene sequence) from beginning to end. The order of START and STOP codons determine which regions of the sequence are audible. Lesser motifs such as bases, pairs of bases and two indicator of the GC content are constantly audible to establish the sonification landscape.

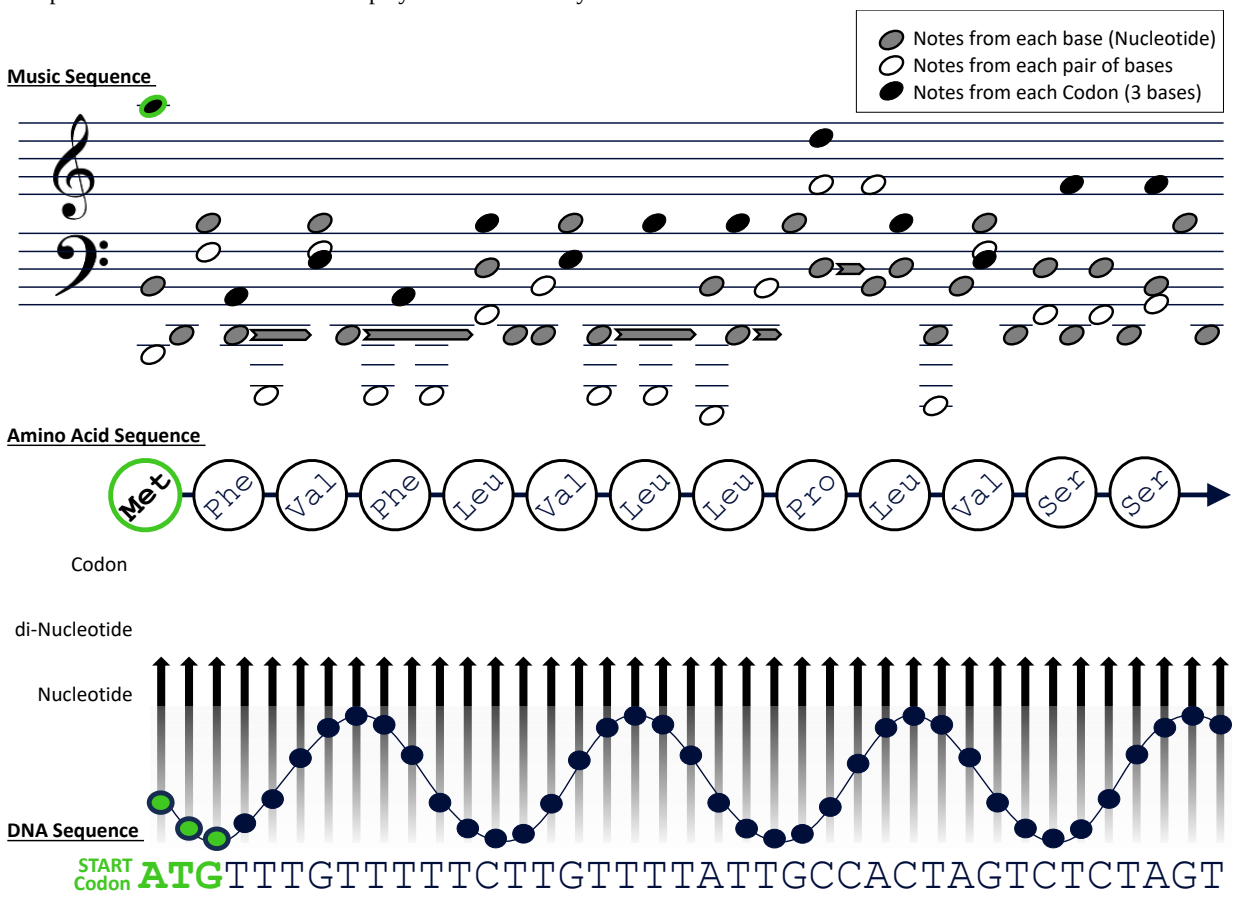

Figure 1: A schematic of how a **DNA sequence** is mapped to layers of a **Musical Sequence** in an auditory display. In this example the sequence contains a START (ATG) codon (shown in green) at position 1 in frame 1. Since no other ATG codons occur in frames 2 or 3, only the sequence in frame 1 is translated as an **Amino Acid Sequence** (peptide) and converted to audio. These codon trigger notes at every third base position. Audio derived from other motifs are also layered in the auditory display. The mapping of individual nucleotide bases will trigger consecutive notes at each base whereas di-nucleotides will trigger notes at every second base position. Data from other attributes were also sonified which, for the sake of clarity, are not shown. This creates a dense auditory display that is information rich and musical due to diatonic constraints in the mapping and timing of notes. Notice how the pitch of the START codon is high to alerts users to the possible beginning of a gene sequence.

4. SONIFICATION METHODOLOGY

DNA sequence files in fastA format were obtained from the National Center for Biotechnology Information (NCBI) database [28]. Up to eight layers of DNA sequence information were systematically mapped to musical notes using code modified from [29], to create a webtool available at https://marktemple.github.io/8ChannelAuditoryDisplay/. The prior code has been substantially updated with the ability to select a variety of DNA sequences, prior sequence-specific metadata has been deleted since many sequence are not annotated. The representation of the DNA polymerase and ribosomes have been updated from protein data bank (PDB) images [30].

Audio is generated within the browser using web audio API [31]. The tool now has the ability to generate a MIDI file in CSV/text format as the sonification is played in the browser. The code used to generate the web audio information was supplemented with additional code to push the same note pitch and timing into a MIDI file. Once the sonification has played through to completion, the CSV file can be downloaded by the users and converted into a binary MIDI file with the UNIX command line program called csvmidi [32]. This is advantageous since the MIDI file can be imported into a digital audio workstation (DAW) for creative composition and recording of further musical instrumentation. Logic Pro (v10.7.7) was used as an example DAW and an imported MIDI file is shown in Figure 2.

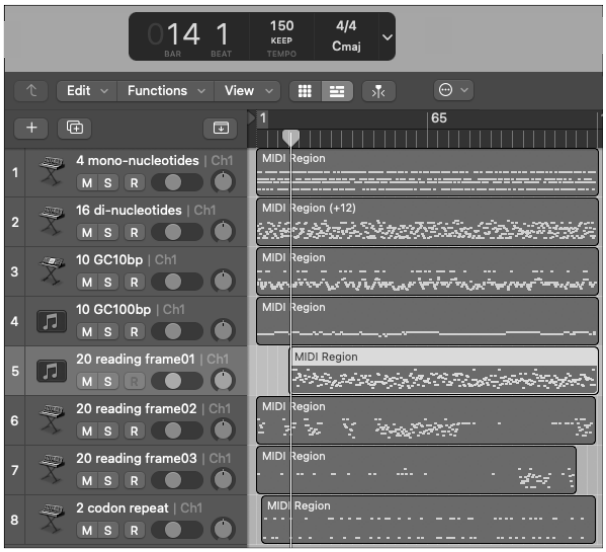

Figure 2: Eight MIDI regions represented in Logic Pro (DAW). Each channel is labelled according to the sonification algorithm used to map the DNA sequence to the note information The prefix number indicates the number of note options of the audio layer, e.g. '20 reading frame01' indicates there are 20 distinct notes used to sonify codons in reading frame 1. The playhead is positioned at the beginning of Bar 14 to highlight the start of the protein coding sequence in Track 5, this is the position of the first ATG START codon in that frame, the audio track was silent until that motif was mapped.

This file was downloaded from the sonification tool and imported into a new software instrument track in Logic using the 'File/Import/MIDI File…' menu command. It is useful to lock the SMPTE (time code synchronization) of all tracks and set the beats per minute (BPM) of the project to 164.0997. This will correctly position the sonified notes to the internal grid of Logic so that the click track will be in-sync. Once this is established, unlock the SMPTE of all tracks. This should ensure that the BPM for the track can be adjusted on demand and the notes will correctly align to the new metronome/click-track of the grid.

The MIDI data contains a header track (not shown) that contains no note information and hence this was deleted in the DAW. For the remaining 8 tracks, an appropriate software instrument was chosen to establish an ensemble of complementary instrument voices. Once these technical procedures have been carried out, further audio tracks can be added to the DAW such as creative performances by live musicians or other recordings of MIDI based instruments.

It is interesting that in the Logic Pro screenshot, the characteristics of the DNA sequence can be observed in the MIDI note information. For instance, in Track 6, clusters of notes can be observed which represent short open reading frames (regions bookended by a START and STOP codons). In Track 1 only four rows of notes are observed, each derived from the one of the four G, A, T or C bases in the sequence. This embedded information is not lost during the musical composition stage and can be used to trigger changes in musical structure.

5. AESTHETICS OF THE WORK

Combining too many layers of audio runs the risk of creating an overly complex auditory display that could be difficult to interpret [33]. To avoid this, care was taken to map the DNA sequence to a diatonic musical scale [34]. For instance, highly repetitive features in the DNA sequence such as each base were mapped to a root note or an interval of a third (both across two octaves) to provide a background tone for subsequent audio layers. This is important since long open reading frames only make up approximately only 1% of the genome and therefore codon based audio is often silent. Less frequent features were mapped to less harmonic intervals, for instance, START codons were pitched higher than other codons making the potential beginning of a gene stand out more clearly.

Sequence information such as the average GC content [35] that spanned either between 10 or 100 bases were mapped long notes with a slower attack time (less percussive in tone). These tweaks to the sonification algorithms were designed so that the listener could distinguish different aspects of the DNA sequence simultaneously. Slower pulses over a longer timeframe add an element of change to re-engage the listener. Additionally, these changes made the auditory display sound more musical which was an advantage when it came to follow up outreach activities to promote biological concepts, and systematic algorithmic music to a wider audience.

Within the DAW different instruments can be assigned to each layer and many other attributes such as the tempo of the audio can be adjusted. The ability to capture the sonification as MIDI data has been an important modification since it has

allowed the reimagining of sonified audio tracks for the purpose of perception, music composition and science outreach events.

Since the MIDI data contains sequences of musical notes it could be considered musical. However, music typically involve more than an apparent random sequence of notes. Music may contains repeating melodies, song structure and dynamics between sections, to say nothing of changes in tempo, key changes or changes in musical instrumentation.

To exemplify the journey to make music from the science audio, four examples will be given, these are available at figshare, https://doi.org/10.6084/m9.figshare.25494223.v1. In the first example, drums were played to the MIDI data to give a clear structure and pattern which helps to establish musical A&B sections, and establish tempo and rhythm, this can be heard in 'Composition Drums with DNA sequence' (mp3). The auditory display contains various syncopated rhythms between the layered audio tracks due to the mapping of different sizes motifs in each layer. Surprisingly 3/3-, 3/4- and 4/4-time signatures can be heard only to become less apparent as the genetic features change.

Once the rhythmic drum patterns are recorded, other melodic and harmonic instruments can be added through creative collaboration with musicians, as heard in 'Composition Musicians with DNA sequence' (mp3). During this procedure the MIDI notes remains unchanged and prominent in the musical composition.

During this process one should be conscious of which sounds are mapped to the motifs of the DNA sequence. In many ways this is an aesthetic or creative choice since the sounds themselves are independent of the algorithms, but they should be carefully chosen e.g. to highlight the start of gene expression. These may also provide triggers for adding melodic and harmonic passages and therefore allow musicians to engage in a scientific narrative. For instance, section A may be played over a gene promoter region whereas section B occurs as the DNA sequence transitions to a gene coding region. An example such as this can be heard at the halfway point (at approximately 2:20 min) into the prior mentioned 'Composition Musicians with DNA sequence' (mp3), the A:B change coincides with the transition to the coding DNA sequence.

6. OUTREACH WITH SONIFIED DNA SEQUENCE DATA

It is here at the intersection of science and art where the interest lies. It's one thing for an educated scientist to perceive the raw auditory display with the intent of informing oneself about a DNA sequence. It's another thing entirely for the public to listen to the auditory display as music or to try to understand the concept behind it [36]. This latter approach lends itself to using the auditory display for outreach to the public to trigger conversations about gene expression, bioinformatics [7] and music composition. The overlap of these two domains in indicated in Figure 3.

When the listener is interested in knowing how and why the auditory display/music was constructed [37] there is an opportunity to discuss the science information on which it is based.

Three general approaches have been used in outreach events based on sonification of DNA sequences. In the first instance the auditory displays is played without further musical accompaniment. An example of this would be the 'Sonification of the Human Mitochondrial Genome' that was displayed at the Powerhouse Museum (Ultimo Sydney) during Science Week 2023 [38].

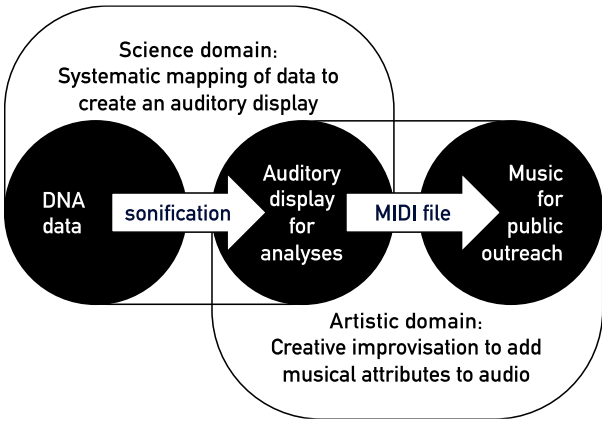

Figure 3: The two domains of science and creative art have in common the auditory display of the sonification analyses. Generation of MIDI data facilitates the intersection of these.

Secondly, the auditory displays is featured with musical accompaniment whereby musicians accompany the sonified data. A minimalist example of this would be a solo rhythmic drum accompaniment as shown in the figshare file 'Outreach Solo Drums with DNA sequence (at Botanical Gardens)' (mp4) from 2023. A more elaborate example is shown in Outreach Ensemble with DNA sequence (at National Art School) (mp4) from 2022. Other performances at City Recital Hall for the 'This Sounds Like Science' outreach program (2023) and Joan Sutherland Performing Arts Centre (2018) indicate a sustained interest at the intersection of science audio and music to blur creative boundaries.

Lastly, events can be held where the auditory display does not itself feature in the performance but the musicians play in the style of the audio or in response to the audio to create a new musical composition [39]. This is facilitated by the generation of musical scores from MIDI data using a DAW.

The final composed musical pieces included introductory passages, outros, middle changes, repetitive passages and changes in tempo and dynamics that are absent in the auditory displays. Having said that the auditory display still took front and centre stage and it can be heard throughout all compositions.

Whilst the discussions in this paper regarding the musicality of the science data may be considered anecdotal it is clear that the intersection between the science and arts domain provides many opportunities for connection and discussion of ideas. The sonification compositions have received good feedback from various media during the journey such as; "this is just amazing" NPR, "beautiful and ethereal music" New Scientist, "turned coronavirus into a musical masterpiece" Science Line, "as cool as it sounds" 3MBS Fine Music, "surprisingly lovely music" The

Conversation, “a unique composition” Limelight magazine, and “surprisingly chirpy, certainly melodic” The Wire (2SER).

7. CONCLUSION

It’s been an interesting and challenging journey to create content that functions in both the scientific and arts domain, this is summarised in Figure 4. The impetuous for this work comes from the scientific domain with an auditory display being used for the display and analysis of biological sequence data. As the iterations of the sonification tool progressed, more attention was given to the sound and mapping of the science data to give it additional musical characteristics, whilst at the same time aiming to improving upon its analytical attributes. Using layers of mapped notes that were harmonious to each other and the use of algorithms that generated notes of different duration and timing also contributed to the musicality. In combination these relatively small changes had a large impact on quality of the auditory display. Additionally, I thought it was important to create musical compositions that can be appreciated by the public without any prior knowledge of how the music was created.

There is a strong interest in music generation from genomic data [40] in addition to the application of sonification as a method of data analysis. It is also clear that music activities can enhance social connections and can improve health and well-being of people [41]. To consider the auditory display of science data as a form of algorithmic music may help foster connections between the science community and the public.

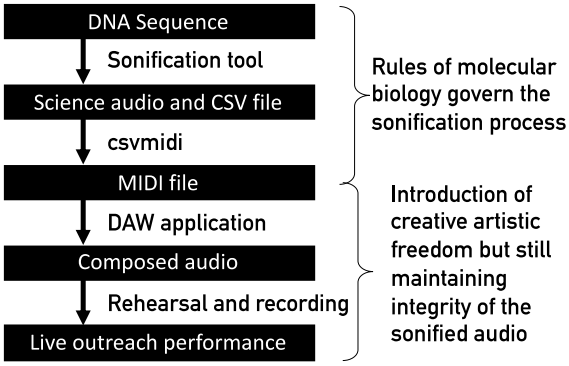

Figure 4: Diagram representing the creative journey and flow of information from the science domain (audio) through to the creative domain (music).

Additionally, music interventions have been shown to be useful in managing symptoms of cancer survivors [42]. In the future it may be possible to listen to algorithmic music generated from one’s own individual genomic sequence. This may be useful to improve one’s mental health and facilitate discussion of how sequence information helps one to diagnose and treat health conditions. This may be particularly relevant for rare genetic diseases where the cause of the condition is to some extent defined by genomic data.

Lastly, it would be interesting from an artistic perspective to edit and write bespoke DNA sequences for the purpose of composition. This is akin to genetic engineering or reverse engineering of sequence data to create a musical composition.

8. ACKNOWLEDGMENT

The outreach events would not have been possible without the contribution of Paul Smith on piano (School of Humanities, Arts and Social Sciences, University of New England), Patricia Young on bass, Peter Veliks on bass, guitar and devices, Mike Anderson on guitar, Mike Bain on synthesizer (School of Computer Science and Engineering, University of NSW), Matthew Tow on guitars, Paul Scott on bass and baritone guitar, Tim Byron on keyboards (School of Psychology, University of Wollongong). Freya Schack-Arnott on Cello, Phillippa Murphy-Haste on Clarinet/Viola, Nick Calligeros on Trumpet, Julian Knowles on Guitar (Macquarie University), Cameron Emerson Elliott on Guitar, Clarence Slockee on Didgeridoo and Bridget Chappell on Cello.

9. REFERENCES

[1] G. I. Mihalas, M. Andor, and A. Tudor, "Adding Sound to Medical Data," in *Stud Health Technol Inform*, vol. 273, no. 1879-8365 (Electronic), pp. 38-53, Sep 4 2020.

[2] A. Dascalu, B. N. Walker, Y. Oron, and E. O. David, "Non-melanoma skin cancer diagnosis: a comparison between dermoscopic and smartphone images by unified visual and sonification deep learning algorithms," in *J Cancer Res Clin Oncol*, pp. 1-9, Sep 21 2021.

[3] F. Rì, R. Masu, M. Graziani, and M. Roncador, "From the Body with the Body: Performing with a Genome-Based Musical Instrument," in *EAI Endorsed Transactions on Creative Technologies*, vol. 7, no. 23, 04/20 2020.

[4] P. Larsen and J. Gilbert, "Microbial bebop: creating music from complex dynamics in microbial ecology," in *PloS one*, vol. 8, no. 3, p. e58119, 2013.

[5] D. Cittaro, D. Lazarevic, and P. Provero, "Chromas from chromatin: sonification of the epigenome," in *F1000Research*, vol. 5, p. 274, 2016.

[6] J. Middleton *et al.*, "Data-to-music sonification and user engagement," in *Front Big Data*, vol. 6, p. 1206081, 2023.

[7] H. Plaisier, T. R. Meagher, and D. Barker, "DNA sonification for public engagement in bioinformatics," in *BMC Res Notes*, vol. 14, no. 1, p. 273, Jul 15 2021.

[8] P. E. Larsen, "More of an Art than a Science: Using Microbial DNA Sequences to Compose Music," in *Journal of microbiology & biology education*, vol. 17, no. 1, pp. 129-32, Mar 2016.

[9] M. S. Staege, "Gene Expression Music Algorithm-Based Characterization of the Ewing Sarcoma Stem Cell Signature," in *Stem Cells Int*, vol. 2016, no. 1687-966X (Print), p. 7674824, 2016.

[10] C. Kim, A. Guo, G. Salhotra, S. Sprinkhuizen, K. Shetty, and D. S. Kong, "Sonifying Data from the Human Microbiota: Biota Beats," in *Computer Music Journal*, vol. 44, no. 1, pp. 51-70, 2020.

[11] J.-L. Boevé and R. Giot, "Chemical composition: Hearing insect defensive volatiles," in *Patterns*, p. 100352, 2021/09/23/ 2021.

[12] C. Gene Ontology *et al.*, "The Gene Ontology knowledgebase in 2023," in *Genetics*, vol. 224, no. 1, May 4 2023.
[13] D. Liu and E. van der Heide, "Interactive Auditory Navigation in the Molecular Structures of Amino Acids: A Case Study Using Multiple Concurrent Sound Sources Representing Nearby Atoms," in *The 25th International Conference on Auditory Display*, pp. 140-146, 2019.
[14] F. Wu *et al.*, "A new coronavirus associated with human respiratory disease in China," in *Nature*, vol. 579, no. 7798, pp. 265-269, Mar 2020.
[15] M. Justin, S. Zdravkovic, M. B. Hubert, G. Betchewe, S. Y. Doka, and T. C. Kofane, "Chaotic vibration of microtubules and biological information processing," in *Biosystems*, vol. 198, no. 1872-8324 (Electronic), p. 104230, Dec 2020.
[16] R. P. Bywater and J. N. Middleton, "Melody discrimination and protein fold classification," in *Heliyon*, vol. 2, no. 10, p. e00175, Oct 2016.
[17] E. R. Miranda, "Genetic Music System with Synthetic Biology," in *Artif Life*, vol. 26, no. 3, pp. 366-390, Summer 2020.
[18] M. S. Staege, "A short treatise concerning a musical approach for the interpretation of gene expression data," in *Sci Rep*, vol. 5, no. 2045-2322 (Electronic), p. 15281, Oct 16 2015.
[19] S. Artigas-Jeronimo *et al.*, "A Novel Combined Scientific and Artistic Approach for the Advanced Characterization of Interactomes: The Akirin/Subolesin Model," in *Vaccines (Basel)*, vol. 8, no. 1, p. 77, Feb 8 2020.
[20] S. L. Franjou, M. Milazzo, C.-H. Yu, and M. J. Buehler, "Sounds interesting: can sonification help us design new proteins?," in *Expert Review of Proteomics*, vol. 16, no. 11-12, pp. 875-879, 2019/12/02 2019.
[21] T. Lappalainen and D. G. MacArthur, "From variant to function in human disease genetics," in *Science*, vol. 373, no. 6562, pp. 1464-1468, Sep 24 2021.
[22] D. Minciacchi, D. Rosenboom, R. Bravi, and E. J. Cohen, "Editorial: Sonification, Perceptualizing Biological Information," in *Front Neurosci*, vol. 14, no. 1662-4548 (Print), p. 550, 2020.
[23] M. D. Temple, "An auditory display tool for DNA sequence analysis," in *BMC Bioinformatics*, vol. 18, no. 1, p. 221, Apr 24 2017.
[24] K. Hayashi and N. Munakata, "Basically musical," in *Nature*, vol. 310, no. 5973, p. 96, Jul 12-18 1984.
[25] E. Riego, A. Silva, and J. De la Fuente, "The sound of the DNA language," in *Biological research*, vol. 28, no. 3, pp. 197-204, 1995.
[26] Z. B. Katz *et al.*, "Mapping translation 'hot-spots' in live cells by tracking single molecules of mRNA and ribosomes," in *Elife*, vol. 5, Jan 13 2016.
[27] M. Goodstadt and M. A. Marti-Renom, "Challenges for visualizing three-dimensional data in genomic browsers," in *FEBS Lett*, vol. 591, no. 17, pp. 2505-2519, Sep 2017.
[28] E. W. Sayers *et al.*, "Database resources of the National Center for Biotechnology Information," in *Nucleic Acids Res*, vol. 52, no. D1, pp. D33-D43, Jan 5 2024.
[29] M. D. Temple, "Real-time audio and visual display of the Coronavirus genome," in *BMC Bioinformatics*, vol. 21, no. 1, p. 431, Oct 2 2020.
[30] S. K. Burley *et al.*, "RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning," in *Nucleic Acids Res*, vol. 51, no. D1, pp. D488-D508, Jan 6 2023.
[31] "Web Audio API." https://webaudio.github.io/web-audio-api/ (accessed March, 2024).
[32] J. Walker. "MIDI to CSV Conversion Tools." https://www.fourmilab.ch/webtools/midicsv/ (accessed March 8, 2024).
[33] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," in *PloS one*, vol. 8, no. 12, p. e82491, 2013.
[34] R. Takahashi and J. H. Miller, "Conversion of amino-acid sequence in proteins to classical music: search for auditory patterns," in *Genome Biol*, vol. 8, no. 5, p. 405, 2007.
[35] B. Berkhout and F. van Hemert, "On the biased nucleotide composition of the human coronavirus RNA genome," in *Virus Res*, vol. 202, pp. 41-7, Apr 16 2015.
[36] E. J. Martin, T. R. Meagher, and D. Barker, "Using sound to understand protein sequence data: new sonification algorithms for protein sequences and multiple sequence alignments," in *BMC Bioinformatics*, vol. 22, no. 1, p. 456, Sep 23 2021.
[37] D. Brocks, "Musical patterns for comparative epigenomics," in *Clinical epigenetics*, vol. 7, p. 94, 2015.
[38] M. D. Temple. "Sonification of the Human Mitochondrial Genome (artistic content)." Powerhouse. https://youtu.be/VgrbRHZzOdk (accessed August, 2023).
[39] M. D. Temple, E. Avery, and S. J. Norman. "Ngiyawaygunhanha (artistic content)." https://www.mai.art/work/adelaide-festival (accessed 2 March, 2024).
[40] B. Wajid *et al.*, "Music of metagenomics—a review of its applications, analysis pipeline, and associated tools," in *Functional & Integrative Genomics*, 2021/10/18 2021.
[41] G. A. Dingle *et al.*, "How Do Music Activities Affect Health and Well-Being? A Scoping Review of Studies Examining Psychosocial Mechanisms," in no. 1664-1078 (Print), 2021.
[42] S. F. Sheikh-Wu, M. A. Kauffman, D. Anglade, F. Shamsaldeen, S. Ahn, and C. A. Downs, "Effectiveness of different music interventions on managing symptoms in cancer survivors: A meta-analysis," in *Eur J Oncol Nurs*, vol. 52, no. 1532-2122 (Electronic), p. 101968, Jun 2021.