SONIFICATION OF EPIGENETIC PROCESSES

By

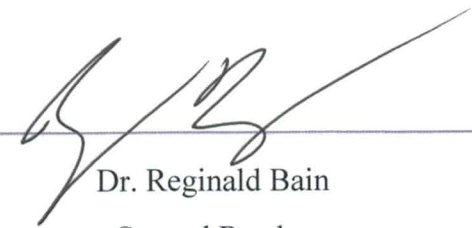
Joelle Strom

Submitted in Partial Fulfillment
of the Requirements for
Graduation with Honors from the
South Carolina Honors College

May, 2019

Approved:

Dr. Jeff Dudycha
Director of Thesis



Dr. Reginald Bain
Second Reader

Steve Lynn, Dean
For South Carolina Honors College

TABLE OF CONTENTS

THESIS SUMMARY

Sonification is the process of creating sound or music from data for investigative or aesthetic purposes. Data sonification has been used in multiple fields to provide an additional layer of analysis to datasets, engage general audiences in a scientific topic, or create an evocative musical piece. Sonification has been extended to biological subjects, including macromolecules such as protein and DNA. Epigenetics, the processes by which gene expression is regulated, is a burgeoning field of molecular biology research. Epigenetic processes have been scarcely sonified, and the process of sonifying DNA base methylation has never been published. DNA methylation is an integral component of gene expression regulation in response to environmental factors, and is partially heritable; methylation research might provide significant clues to the origin of certain diseases. Data sonification has the potential to facilitate detection of methylation patterns in future studies. The goal of this project was to identify the optimal method of sonifying DNA methylation data, retaining a balance of scientific significance and musical interest.

Max/MSP computer music software was used to translate datasets accessed from the Gene Expression Omnibus database. Multiple types of data and methods of sonification were explored before determining the most useful model. This model incorporates and utilizes data to shape notes generated by frequency modulation sound synthesis. Each methylation site becomes a sustained note and the tone quality of the note fluctuates based on the amount of methylation in each sample. This method allows the listener to compare methylation levels across samples with the potential for identifying regions of differential methylation. Future directions of this project include formally testing model efficacy, improving program accessibility for researchers, and exploring the potential for musical enhancement.

INTRODUCTION

Data sonification is a tool used to present complex information in a format that facilitates analysis; it is the process of converting numerical or categorical data into sound. Just as a graph, chart, or diagram would present data in a useful visual format, sonification transforms data into an auditory display. The resulting sound allows for a different perception of the data that may reveal a new pattern or abnormality that went unnoticed in other presentation formats.

Interdisciplinary investigations in various fields of study, including biology, have incorporated sonification. Applications of sonification in this field include the adaptation of electrocardiogram data into biomusic to study the nuances of heartbeats, the amplification of "white noise" produced by cells in an attempt to develop a non-invasive method of detecting cancerous tissue, and music composed to evoke the slow, deliberate life cycle of plants for viewers of a plant biology exhibit (1–3). Some researches even created interactive medical data sonification, in which a user can navigate through a database to select specific sections to listen to, or in which data can be sonified as it is collected (4). Molecular biology in particular presents a rich set of concepts and processes to sonify. With the significant improvement in lab technology, obtaining new DNA sequences has become faster and cheaper than ever, leading to large datasets of protein and DNA sequences. With thorough analysis, these datasets could clarify or illuminate characteristics of our molecular building blocks. Previous studies have developed multiple methods for sonifying DNA and protein sequences and investigated the applications of this tool (5–7).

Epigenetics the study of how gene expression is regulated, is an area of molecular biology that has become increasingly important to our understanding of human development and pathology. Epigenetic processes have been linked to physiological adaptation to stress and other

environmental factors. Studies have shown that stressful life events can leave a lasting imprint on the epigenetic pattern of an individual (8–15). Some research studies suggest that these changes to the epigenome are an underlying mechanism of the development of post-traumatic stress disorder, and may also increase the risk of bipolar disorder (16–18). Epigenetic modifications can be reversible, but they can also be permanent and pervasive enough to be passed on to offspring; significant events in a person's life can create a lasting impact on that person's children (19–21). Researchers studying condensed and expanded regions of DNA have already begun to explore epigenetic sonification. They mapped the various levels of DNA expansion to chords and determined that music generated from this data has detectable patterns (22).

The process of DNA methylation has not yet been sonified. A methyl group, a small chemical compound containing one carbon and three hydrogen atoms, can be substituted onto cytosine bases in the genetic sequence, as shown in Figure 1 (23). When enough methyl groups are added to the promoter region of a gene, the molecular machinery that normally assembles to express that gene is blocked, and the gene becomes silent. This process is shown in Figure 2 (24). This is a normal process that occurs during early development to guide embryonic stem cells toward specialized cells with unique structure and characteristics. Methylation is also one of the processes that occurs throughout life as the body responds to outside stimuli, and is one of the epigenetic factors contributing to the lasting effects of stressful

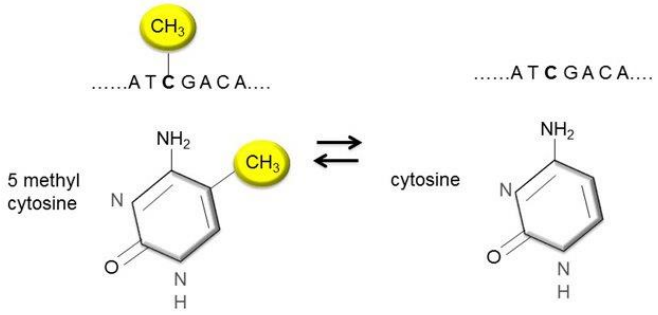


Figure 1: Depiction of a methylated cytosine base

events (9–11,17,25,26). Data sonification might be a useful tool for this burgeoning field because methylation sequencing often involves large datasets and cross-subject analysis to identify similarities and differences in methylation patterns and associate these differences with subject characteristics. Sonification of methylation datasets might help researchers identify areas of unique DNA methylation and connect these areas to the physiological and psychological effects of stress.

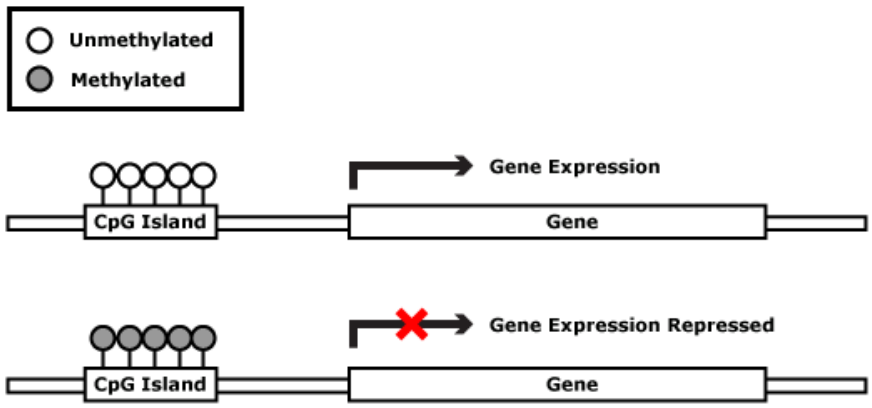


Figure 2: A model of methylation effects on gene expression

PROJECT GOALS

The objective of this project was to take raw data from a research study investigating DNA methylation and transform the data into music to serve musical, educational, and scientific goals. One purpose of this project was to create a unique piece of data-driven music with the potential for novel sounds that may provide inspiration or contemplation to the listener. Another purpose was to create a learning tool that can accompany epigenetics lessons, specifically those explaining DNA methylation. The primary goal, however, was to initiate the development of a research tool that could be used to facilitate data analysis in future methylation studies. This project was designed to explore multiple methods of data sonification and to analyze which method best communicated patterns and abnormalities in the data over time or between data

categories. Future studies in this area could adopt the most effective sonification strategy and develop a standardized method for researchers to sonify and analyze their data.

METHODS

Data Accession

DNA methylation studies were discovered and raw data files were accessed using the National Center for Biotechnology Information's Gene Expression Omnibus search engine. Search terms were entered in the GEO DataSets engine, then study ID numbers were entered in the GEO Accession Viewer to access the supplementary data. Data was downloaded and converted into an Excel spreadsheet for visualization and management. Common variables included methylated and unmethylated signal intensities and methylated and unmethylated counts. Methylation signal intensities were divided by unmethylated signal intensities to calculate a ratio for use in sonification. Methylated and unmethylated counts were totaled and methylated counts were divided by total counts to calculate a percentage. Datasets were then converted into tab-delimited text files for processing by music software.

Two different datasets were used in this project. One was from a study (GSE106379) investigating the influences of genetics and life experience on DNA methylation patterns by cross-breeding mouse strains (27). The other dataset was from a study (GSE94734) determining the correlation of early life adversity in humans and alterations in DNA methylation (28). Both research teams performed whole-genome bisulfite sequencing to measure DNA methylation.

Sonification

Max/MSP 7 was used to convert the data into sound. A patch was developed to split the columns from the data file into separate objects and to establish a read-through functionality that will process each row of the file at a specified tempo. The rest of the patch used Max/MSP

functions to assign the numerical data to parameters that dictate pitch, timbre, amplitude, or other elements of sound.

SONIFICATION MODELS

While searching for datasets that would be suitable for this project, I noticed that there was no standard format for raw data from bisulfite sequencing. Many datasets were organized according to possible methylation sites, each site with an ID number. The type of data used in each dataset, however, varied. Most studies reported "methylated signal intensity" and "unmethylated signal intensity" but each study chose to include different statistical measures in addition to these values. One dataset, used in this project, reported methylated and unmethylated "counts." Because the nature of the reported data varies among researchers, I decided to perform multiple sonification methods to determine how best to present methylation data.

Methylation Ratios

For the dataset from Bush et al. (28), a ratio was calculated within Excel from signal intensities ($\frac{\text{meth}}{\text{unmeth}}$). Percentages were calculated from the counts in the Grimm et al. (27) datasets. Ratios and percentages were used as parameters in frequency modulation (FM) synthesis. Frequency modulation, visualized in Figure 3 (29) includes a carrier frequency, which generates a sine wave that determines the pitch of the note. This sine wave is then modified with a second frequency to rapidly produce overtones that color the

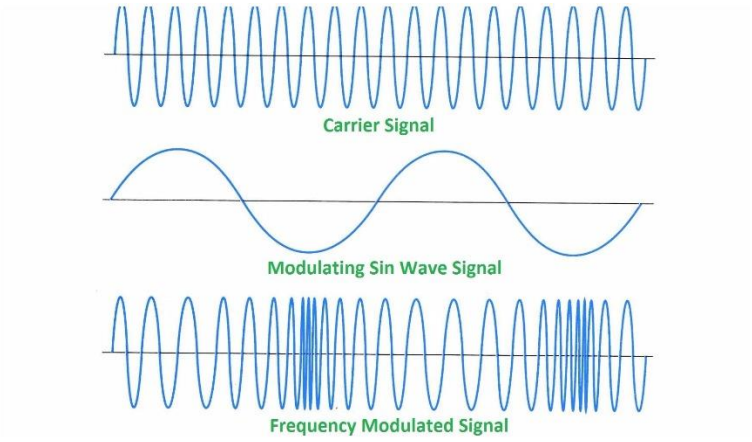


Figure 3: Frequency modulation synthesis

pitch and provide a specific timbre. It is possible to synthesize an infinite number of unique sounds with this method, depending on the ratio of the modulating frequency to the carrier frequency, known as the harmonicity ratio. A harmonicity ratio close to a whole number will generate fuller organ- or voice-like sounds, whereas ratios in between whole numbers will produce unpredictable sounds. The amplitude of the modulating frequency, called the modulation index, determines the prominence of the tone, and can be altered over the duration of a note to provide shape and depth. Ratios from the datasets were used as inputs for FM synthesis, after being scaled to fit the standard range of values for each of the parameters.

In an initial version of this model, the harmonicity ratios were determined by the $\frac{\text{meth}}{\text{unmeth}}$ ratios from the data. Each sample column produced an individual sound, each sound had the same carrier frequency and therefore the same pitch, but the methylation ratio became the harmonicity ratio for the sound produced by each column. Ratios from the dataset ranged from 0 to about 30, and were scaled to a range of 1 to 10 to fit normal harmonicity ratio values. Columns were sonified simultaneously so that each beat of the music contained the sounds of multiple data samples. The goal was to hear differences in timbre across samples at each methylation site, but this proved difficult when samples could not be heard individually.

In a later version of this model, the samples were each given a beat to be heard individually, instead of playing multiple sounds in one beat. One row of the dataset became one long note, with a constant carrier frequency and harmonicity ratio throughout. The samples shaped the tone of the note by changing the modulation index throughout the note's duration. Ratios from the dataset were scaled to a range of 0 to 1, then strung together to create an input function for FM synthesis. The modulation index would start at 0 at the attack of the note, then ramp up to 0.5, then transition to the value from column 1, then column 2, etc., before returning

to 0 at the release of the note. This way the amount of methylation in each sample at the same site could be heard side-by-side.

Methylation Counts

The dataset from Grimm et al. (27) presented methylation at each site as methylation counts, not a signal intensity. These numbers are positive integers within a 0-88 range, which means they can be mapped to MIDI values. MIDI is a standardized numerical notation system that allows computer software and synthesizers to generate and communicate musical information. This sonification scheme does not involve sound synthesis methods but rather relies on pre-set computer music software settings to produce specified pitches based on the number provided at each site by each sample. An interactive function allows the user to select the desired instrumentation. Multiple samples are audible at the same time so the listener hears chords representing the range of methylation levels across samples at each site in the genome.

DISCUSSION

Model Efficacy

The final sonification model based on methylation ratios or percentages provided clear comparison of methylation levels across different samples at a given site. Because each sample was heard individually, its sound could be effectively interpreted and compared to the surrounding sounds from other samples. Each methylation site became a single note in the music, providing continuity and a frame of reference: each new note was a new methylation site. Any changes to the tone throughout the course of that note were due to differences in methylation levels between samples at that site. The carrier frequency (the pitch) was changed every ten

iterations to make the music more interesting and to provide an auditory cue as the program moves through the dataset.

This model provided scientifically relevant sound. At unmethylated sites, the ratio or percentage from the dataset was close to 0, resulting in a modulation index near 0 and a simple, pure tone. High methylation values resulted in modulation indices near 1, which produce strong overtones that catch the listener's attention. This makes intuitive sense to the listener, who might be expecting a more noticeable sound for data points where there is more methylation activity. The changes in tone also allow the listener to identify large discrepancies in the data. Shifts between modulation index values are significantly audible, and can direct the listener to further examine that point in the data.

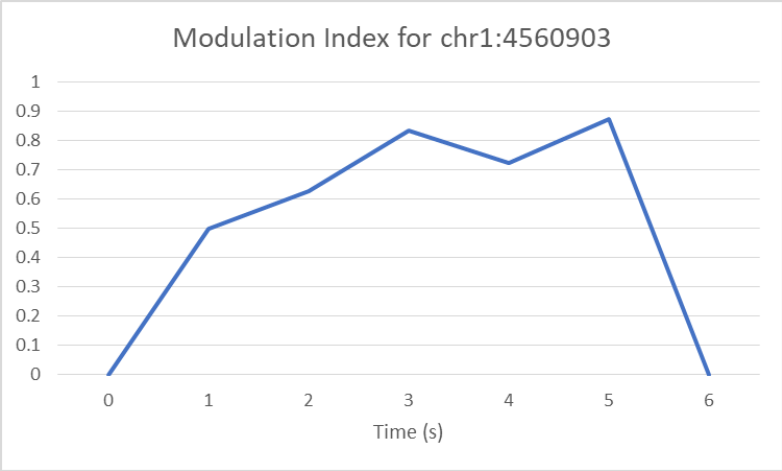


Figure 4: Graphical representation of the shape of the tone color over the duration of one note. Values at seconds 2-5 are methylation percentages from four different samples at one methylation site, based on the dataset from Grimm et al. (27)

The flexibility of this model allows it to be applied to many different studies and datasets. Carrier frequencies and harmonicity ratios can be altered according to the listener's preferences. Lengths of notes and intervals between samples can be adjusted to either speed up the sonification and process large datasets quickly or slow down and identify nuances of a particular section of data. Any number of samples can be audible throughout the notes, and raw values from the dataset can be scaled to fit the expected values of FM synthesis parameters. One shortcoming of this model is that the

small scale of the modulation index (0 to 1) will mask differences in raw data if the range of data is large. A solution for this is to convert the data to a logarithmic function before scaling; this function can be executed in the software and therefore is a relatively simple adjustment. The researchers should be careful to use a function that still allows relevant differences in the data to be audible.

Alternative Approaches

The original model for this project involved sequence-based data instead of numerical data. The goal was to find the bisulfite-converted DNA reads from a methylation study and convert those reads into a nucleotide sequence with notation indicating which specific bases were methylated. This ideally would have been coupled with gene expression data from the same samples to describe how many copies of each gene were produced. When combined these two inputs would create a model of methylation patterns in the DNA and how those patterns affect gene expression. The music would consist of a running melody where each nucleotide would be represented by a note, the presence of a methyl group would be denoted by a percussive hit, and gene expression would be indicated by the presence of chords on top of the melody. This model would provide the listener with an integrated representation of DNA methylation and gene expression and might be able to determine how changes in methylation effect changes in transcription.

There were a couple problems with this model. The biggest hurdle was data accession. Raw sequence reads from bisulfite conversion are not commonly distributed as supplementary data. Most research studies deal with compiled data from multiple sequencing repetitions. Methylation patterns can vary from cell to cell, so experiments are conducted with whole tissue extraction and sequencing of multiple DNA molecules. The reads are then analyzed and

converted into statistical values indicating the probability of methylation at a particular site. While raw sequences from studies are available, they are more difficult to access and lack scientific significance. In addition to this issue, downloading and processing reads requires proficiency in the Linux operating system and the implementation of software with the ability to align downloaded sequences to a reference genome to produce the desired output file. Accomplishing these tasks was time-consuming and required large amounts of troubleshooting and therefore would not be a convenient option for any researcher hoping to analyze raw data. A more viable procedure would be to utilize existing resources for processing and analyzing bisulfite conversion data and to sonify the resulting numerical information.

When investigating the transformation of said numerical data into music, an initial approach was to sonify multiple samples simultaneously, as described in the sonification models section. This would theoretically allow the listener to compare multiple samples side-by-side and identify differences in the sounds during each beat. This could then inform the researcher at which sites the samples diverge in methylation status. Unfortunately when multiple samples are sonified in a single beat, the sounds blend together and obscure any differences. The optimal model therefore features each sample of interest sonified at individual time points, with a common element unifying the data points across all the samples at the same methylation site. This is accomplished in the final model described for datasets with ratios or percentages.

Future Directions

With an optimal model for the sonification of DNA methylation established, the next step would be to test its scientific significance and enhance its musicality. The primary goal of this project is to develop a tool to assist researchers in analyzing data and detecting patterns in DNA methylation. The model must be verified by sonifying multiple datasets, at different points in the

genome, with varying numbers and categories of samples. Listeners of varying academic backgrounds should listen for patterns in the music without prior knowledge of the data, and their inferences from the auditory model should be cross-referenced with statistically-determined patterns. If listeners are able to accurately identify similarly methylated regions and differentially methylated regions, the model might have a role in future research.

The auditory model could be compared to data visualization methods. A potential study could include multiple subject groups; one group would review visual data, another group would review data through the sonification model, and a third group would review both data representations. The study moderator could present both formats to reviewers, then remove visual data, or remove audible data. Analysis would indicate similarities and differences among the groups' abilities to clearly and accurately detect patterns in data. Max/MSP software can generate graphs to supplement the auditory component if visual displays enhance the listener's comprehension and analysis.

Once the model has been verified as a useful research tool, the process of sonification must be streamlined. A software program should be developed to easily allow researchers to convert their data into a format useful for sonification and to input that processed data into the frequency synthesis program. This would involve identifying a standard data format and tailoring a script, likely in a programming language such as Python, to be operable by those with little experience with or no access to Max/MSP software. Alternatively a polished Max/MSP patch can be exported for use without a software license. This exported patcher cannot be edited but would retain all other functionality.

Experimental Music

Another future direction would be to layer additional melodic and percussive tracks and incorporate nuances into the current model to transition from scientific significance to musicality. This is an auxiliary goal of this project, but pursuing the enhancement of musical character of DNA methylation sonification may provide an intriguing contribution to the computer music genre.

The model based on methylation counts did not have use as a research tool but instead served as a study in experimental music. This model utilized data points to generate chords via MIDI notation. The data-driven harmonies were unpredictable and highly varied, creating a unique piece of music. Changes in the imported dataset and the instrumentation choice would alter the listener's experience, allowing for an experimental method of investigating musical possibilities from methylation data.

This model can be further developed as an artistic showcase by altering the method by which methylation count data is translated into musical properties. The current model feeds data directly into a MIDI pitch input, but future models might use data to determine the duration, amplitude, voice (or instrumentation), pitch bend, or other characteristics of the notes. Columns could be played simultaneously or in sequence, or different columns could influence the same note by providing values for different characteristics of that note. Adding multiple layers (percussion tracks, chords, melodies, etc.) will increase the complexity of the piece and may enhance musicality.

WORKS CITED

1. Batsis D, Bitsikas X, Georgaki A, Evaggelou A, Tigas P. Biomusic: The carrier. Technoetic Arts J Specul Res. 2011 Dec;9(2/3):209–16.

2. Roosth S. Screaming Yeast: Sonocytology, Cytoplasmic Milieus, and Cellular Subjectivities. Crit Inq. 2009 Winter;35(2):332–50.

3. Gibson J. Slowlife: Sonification of Plant Study Data. Leonardo Music J. 2006 Dec;16(1):42–4.

4. Pauletto S, Hunt A. Interactive sonification of complex data. Int J Hum-Comput Stud. 2009 Nov 1;67(11):923–33.

5. Temple MD. An auditory display tool for DNA sequence analysis. BMC Bioinformatics. 2017 Apr 24;18:1–11.

6. Shi XJ, Cai YY, Chan CW. Electronic Music for Bio-Molecules Using Short Music Phrases. Leonardo. 2007 Apr;40(2):137–41.

7. Dunn J, Clark MA. Life Music: The Sonification of Proteins. Leonardo. 1999 Feb;32(1):25–32.

8. Burns SB, Szyszkowicz JK, Luheshi GN, Lutz P-E, Turecki G. Plasticity of the epigenome during early-life stress. Semin Cell Dev Biol. 2018 May 1;77:115–32.

9. McCoy CR, Rana S, Stringfellow SA, Day JJ, Wyss JM, Clinton SM, et al. Neonatal maternal separation stress elicits lasting DNA methylation changes in the hippocampus of stress-reactive Wistar Kyoto rats. Eur J Neurosci. 2016 Nov 15;44(10):2829–45.

10. Stankiewicz AM, Swiergiel AH, Lisowski P. Epigenetics of stress adaptations in the brain. Brain Res Bull. 2013 Sep;98:76–92.

11. Essex MJ, Thomas Boyce W, Hertzman C, Lam LL, Armstrong JM, Neumann SMA, et al. Epigenetic Vestiges of Early Developmental Adversity: Childhood Stress Exposure and DNA Methylation in Adolescence. Child Dev. 2013 Feb 1;84(1):58–75.

12. Denhardt DT. Effect of stress on human biology: Epigenetics, adaptation, inheritance, and social significance. J Cell Physiol. 233(3):1975–84.

13. Mulligan CJ. Early Environments, Stress, and the Epigenetics of Human Health. Annu Rev Anthropol. 2016;45(1):233–49.

14. Mirbahai L, Chipman JK. Epigenetic memory of environmental organisms: A reflection of lifetime stressor exposures. Mutat Res Toxicol Environ Mutagen. 2014 Apr;764–765:10–7.

15. Romens SE, McDonald J, Svaren J, Pollak SD. Associations Between Early Life Stress and Gene Methylation in Children. Child Dev. 2015 Feb 1;86(1):303–9.

16. Zannas AS, Provençal N, Binder EB. Epigenetics of Posttraumatic Stress Disorder: Current Evidence, Challenges, and Future Directions. Biol Psychiatry. 2015 Sep 1;78(5):327–35.

17. Marinova Z, Maercker A, Grünblatt E, Wojdacz TK, Walitza S. A pilot investigation on DNA methylation modifications associated with complex posttraumatic symptoms in elderly traumatized in childhood. BMC Res Notes. 2017 Dec 19;10:1–5.

18. The Role of Epigenetics in Early-Life Stress and the Risk of Bipolar Disorder. Bipolar Disord. 20(S1):16–7.

19. Chan JC, Nugent BM, Bale TL. Parental Advisory: Maternal and Paternal Stress Can Impact Offspring Neurodevelopment. Biol Psychiatry. 2018 May 15;83(10):886–94.

20. Zucchi FCR, Yao Y, Ward ID, Ilnytskyy Y, Olson DM, Benzies K, et al. Maternal Stress Induces Epigenetic Signatures of Psychiatric and Neurological Diseases in the Offspring. PLoS ONE. 2013 Feb;8(2):1–9.

21. Manjrekar J. Epigenetic inheritance, prions and evolution. J Genet. 2017 Jul;96(3):445–56.

22. Cittaro D, Lazarevic D, Provero P. Chromas from chromatin: sonification of the epigenome. F1000Research [Internet]. 2016 Mar 3 [cited 2019 Mar 10];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4805159/

23. DNA Methylation in Development and Disease - Epigenetics Core - Environmental Health Sciences Center - University of Rochester Medical Center [Internet]. [cited 2019 Mar 10]. Available from: https://www.urmc.rochester.edu/environmental-health-sciences/cores-facilities/epigenetics-facility/dna-methylation-in-development-and-disease.aspx

24. UCSF School of Medicine: Genes and Genomes Online Learning Module [Internet]. [cited 2019 Mar 10]. Available from: http://missinglink.ucsf.edu/lm/genes_and_genomes/methylation.html

25. Milaniak I, Cecil CAM, Barker ED, Relton CL, Gaunt TR, McArdle W, et al. Variation in DNA methylation of the oxytocin receptor gene predicts children's resilience to prenatal stress. Dev Psychopathol. 2017 Dec;29(5):1663–74.

26. Marzi SJ, Sugden K, Arseneault L, Belsky DW, Burrage J, Corcoran DL, et al. Analysis of DNA Methylation in Young People: Limited Evidence for an Association Between Victimization Stress and Epigenetic Variation in Blood. Am J Psychiatry. 2018 Jan 12;175(6):517–29.

27. Grimm SA, Shimbo T, Takaku M, Thomas JW, Auerbach S, Bennett BD, et al. DNA methylation in mice is influenced by genetics as well as sex and life experience. Nat Commun [Internet]. 2019 Jan 18 [cited 2019 Mar 10];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6338756/

28. Bush NR, Edgar RD, Park M, MacIsaac JL, McEwen LM, Adler NE, et al. The biological embedding of early-life socioeconomic status and family adversity in children's genome-wide DNA methylation. Epigenomics. 2018 Oct 23;10(11):1445–61.

29. JustScience. WHAT ARE THE APPLICATIONS OF FREQUENCY MODULATION? [Internet]. JustScience. 2017 [cited 2019 Mar 10]. Available from: http://www.justscience.in/articles/what-are-the-applications-of-frequency-modulation/2017/06/02

APPENDIX A: DATA SETS

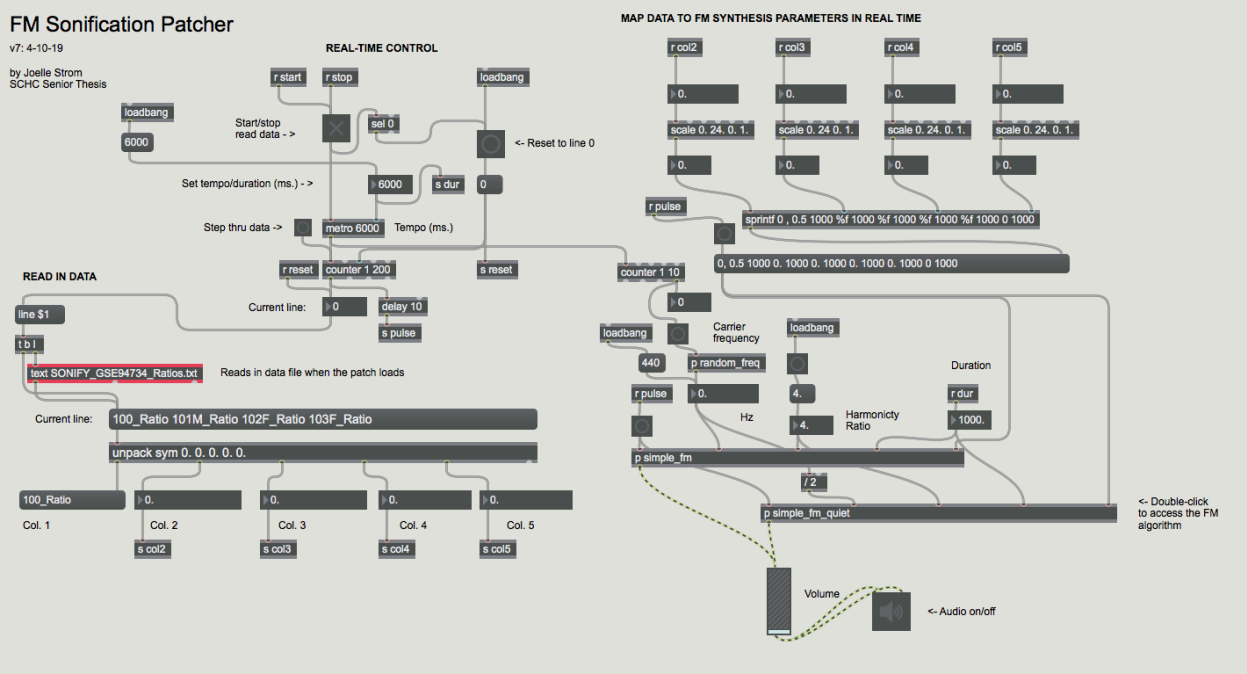
ID	86F_Ratio	119F_Ratio	153F_Ratio	171F_Ratio
cg06193439	9.799893152	10.49782317	11.05072361	3.386559475
cg06193520	1.517718056	1.568189322	1.100878016	0.487841098
cg06193578	0.170267835	0.270378549	0.208387911	0.373011269
cg06193597	3.240138666	3.136315379	13.70257787	0.083048311
cg06193628	1.664956874	2.450886301	1.287765426	0.312277571
cg06193633	0.044237483	0.046646757	0.023064635	2.098086036
cg06193668	21.07303967	18.97337597	41.47267756	1.279982227
cg06193723	0.0819276	0.051049253	0.069229055	0.017990515
cg06193766	0.194317353	0.257025038	0.182509019	0.135161587
cg06193775	0.035794183	0.022270426	0.017971468	0.0728621
cg06193838	4.511215541	3.956809634	3.984895053	1.293614582
cg06193958	0.126698665	0.146953045	0.297301567	0.142296396
cg06193988	0.081270356	0.092089	0.100299003	0.313469432
cg06193995	1.409993386	1.336927486	1.945618484	44.46188913
cg06194003	15.52380983	43.64947629	47.86138411	8.312785343
cg06194010	19.62903922	11.07618763	17.95321773	0.687784123
cg06194026	0.025810322	0.02852516	0.061295816	0.056094323
cg06194070	1.842835866	1.561910306	2.023166463	1.147809971
cg06194119	0.033811122	0.033317593	0.036799642	0.433572475
cg06194186	6.679286367	7.874479815	10.36327007	0.238461254
cg06194421	0.083874249	0.078525135	0.029431692	0.096311534
cg06194479	0.016073071	0.02987385	0.015598466	0.015627896
cg06194536	0.067913498	0.065652994	0.031500965	0.548354956
cg06194602	6.555434822	5.681782154	5.253356437	1.950313871
cg06194738	3.681198189	2.963068986	3.158854712	3.001098307
cg06194808	1.156367451	1.05731115	1.140449024	0.542922041
cg06194809	0.576128186	0.656513769	0.697724359	2.580039033
cg06194885	3.002252829	3.030092603	3.040452036	1.650493886
cg06194960	2.226028527	2.772166284	2.142635546	2.042333887

Example of ratio-based methylation data, calculated from dataset from Bush et al. (28)

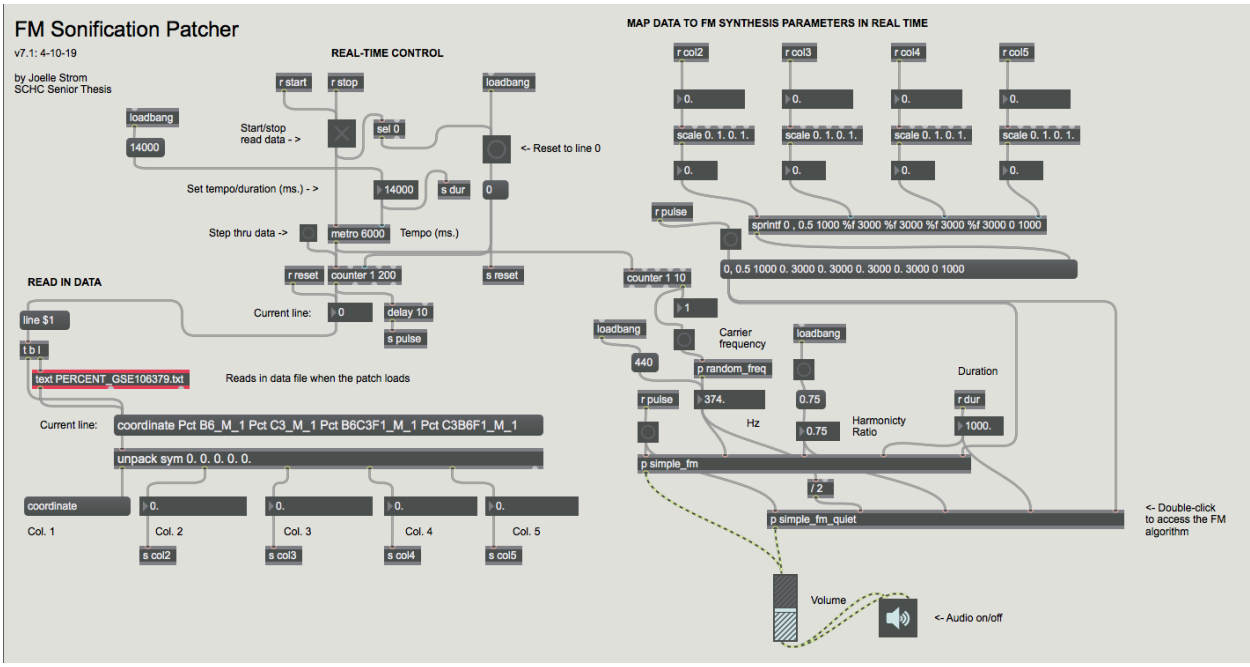
coordinate	B6	C3	B6C3F1	C3B6F1
chr1:3000574	32	25	26	29
chr1:3000726	9	2	7	4
chr1:3000901	9	7	7	5
chr1:3001346	19	15	14	18
chr1:3001394	21	17	17	16
chr1:3001631	16	17	17	18
chr1:3002177	21	7	11	10
chr1:3002338	31	24	20	26
chr1:3002386	27	27	21	22
chr1:3002599	15	17	15	22
chr1:3002921	20	15	13	18
chr1:3003553	15	15	12	12
chr1:3003601	21	24	23	26
chr1:3003660	18	21	21	22
chr1:3004013	8	8	7	11
chr1:3004613	37	33	27	36
chr1:3004636	29	28	22	31
chr1:3004682	23	29	19	27
chr1:3004692	22	28	23	28
chr1:3004780	21	25	16	21
chr1:3004822	17	22	13	19
chr1:3004855	9	8	7	7
chr1:3005055	22	17	19	17
chr1:3005149	27	23	20	20
chr1:3005205	36	34	34	35
chr1:3005535	25	36	27	29
chr1:3005766	30	31	30	32
chr1:3005847	34	31	27	35
chr1:3005928	23	29	22	26

Example of methylation count data, obtained from Grimm et al. (27)

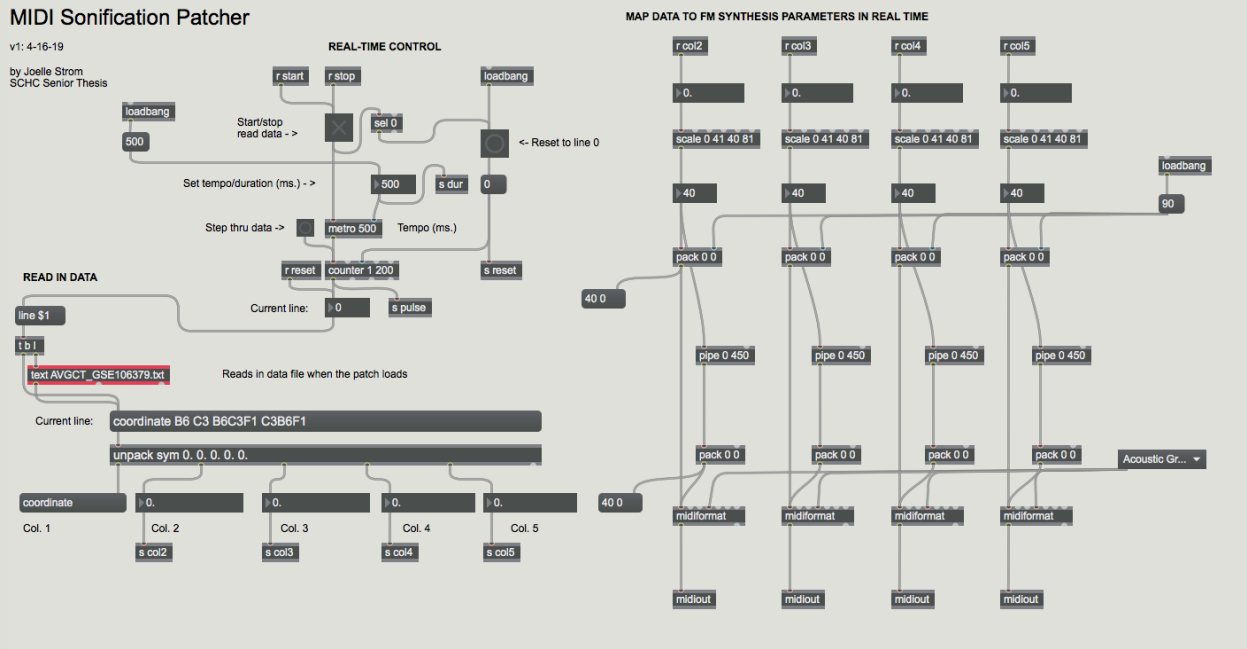
APPENDIX B: MAX/MSP PATCHES



Patch using FM synthesis to sonify ratio-based dataset from Bush et al. (28)



Patch using FM synthesis to sonify percentage-based dataset calculated from Grimm et al. (27)



Patch using MIDI values to sonify methylation count data from Grimm et al. (27)