SEED, FROM DOUBLE HELIX

Reginald Bain

University of South Carolina
School of Music
813 Assembly St.
Columbia, SC, USA
rbain@mozart.sc.edu

ABSTRACT

"Seed" is the third movement of *Double Helix* (2019), a sonification-inspired algorithmic composition in four movements that utilizes genomic data as source material. Influenced by the early music of Steve Reich and other American minimalist composers who experimented with technology in order to discover new compositional processes, the author explores a variety of musico-genetic themes across the work's four movements. The data employed in the composition is from the human gene CHD8 which has been identified as a leading candidate gene for autism risk. Its DNA sequence and protein sequence are mapped to musical parameters in real time using a custom application written in Cycling '74's Max. Seed is a parameter-mapping musification of the beginning segment of CHD8's protein sequence that is presented in the form of a generative 8-voice canon. The amino acids in the protein sequence are strategically mapped to pitches in a harmonic series. Pitch and intensity values are mapped redundantly so these two dimensions are linked. Other dimensions are similarly linked, but in more complex ways, creating a correlation between all the sonic dimensions.

1. INTRODUCTION

Double Helix (2019) is an 18-minute electronic music composition in four movements: I. The Mapping Problem, II. Biological Clocks, III. Seed, and IV. Time Cycle. It is a sonification-inspired algorithmic composition that utilizes genomic data as source material. The data is from the human gene CHD8 (chromodomain helicase DNA binding protein 8) which has been identified as a leading candidate gene for autism risk.

This work is part of the *Mutational Music Project* – the broader impact component of a scientific research project in evolutionary genetics titled "Mutational variance of the transcriptome and the origins of phenotypic plasticity."[1] As part of this project, the author (an electronic composer) and his collaborator (an evolutionary biologist) have created an interdisciplinary research experience for university-level biologists and composers, and software applications for science education outreach. The sonification designs and musical algorithms that form the backbone of *Double Helix* are being adapted to become more generalized sonification designs that aid in the introduction of auditory display and sonification concepts.

Double Helix received its premiere in a lecture-recital concert for the general public that was given by the author at the University of South Carolina on April 8, 2019. More information is available online at: <www.reginaldbain.com>.

2. GOALS AND AESTHETICS

Genetic data and musical data may both be viewed as a sequence. This creates a common ground where two usually distinct domains may meet. Using a methodology consistent with approaches in the field of algorithmic composition [1], a sub-discipline of electronic music composition [2], the author created *Double Helix*.

Influenced by the early music of Steve Reich and other early American minimalist composers who experimented with technology [3], the author adopted a process-oriented compositional approach [4] in *Double Helix*. A DNA sequence and corresponding protein sequence are strategically mapped in real time to changing musical parameters. What is more, the author explores a variety of musico-genetic themes that help bind the four movements of the work together – including the metaphorical-mathematical bridge that connects the double-helical model of DNA [5] and the helical model of pitch space [6]. The composition is also informed by techniques in generative music [7] – which have deep connections with processes in evolutionary biology [8, 9].

Accepting Edgard Varèse's broad definition of music as *organized sound* [10], and with the goal of improving the communicative ability of sonifications, Vickers and Hogg proposed an *aesthetic perspective space* onto which both musical compositions and sonifications could be mapped and thus compared [11]. This work is carried out in that space.

3. SONIFICATION APPROACH

The sonification approach used in *Double Helix* is *parameter-mapping sonification* (PMSon) [12]. This approach was chosen primarily because of its historical significance in early genetics and music research [13, 14]. After Grond and Berger, the term *musification* is used heretofore to refer specifically to the use of PMSon in artistic applications [12, 390–392], as it clearly acknowledges the connection between PMSon and mathematical, statistical, and stochastic mapping strategies used in algorithmic composition. An introduction to the scientific and musical principles typically involved in early *gene music* experiments may be found in Dunn and Clark [15].

[1] National Science Foundation (NSF) award #1556645, Jeff Dudycha, principal investigator. The author is the other senior person on the grant.

4. GENOMIC DATA

DNA stands for deoxyribose nucleic acid, the so-called molecule of life [16]. DNA's double-helix structure was unveiled by Francis Crick and James D. Watson in 1953 [5]. Their discovery was based on the work of Maurice Wilkins, Rosalind Franklin, and many other scientists [17].

4.1. DNA Sequences

A scientific description of the structure and properties of the DNA molecule and its role in genetics [18, 19] is well beyond the scope of this paper. For the current discussion, it should suffice to say that the classic double helix consists of two intertwined strands of DNA, and each strand may be encoded as a string of four nucleotide bases named Adenine (A), Cytosine (C), Guanine (G), and Thymine (T), which are typically represented using the single-letter codes given above in parentheses. A and G have similar 2-ring chemical structures and are called *purines*. C and T have similar 3-ring chemical structures and are called *pyrimidines*. As shown in Figure 1, these classifications are based on the chemical bonds formed between the two strands: i.e., a double bond (=), or triple bond (≡).

Purines		Pyrimidines
A	=	T
G	≡	C

Figure 1: Classification of the four nucleotide bases as Purines or Pyrimidines.

4.2. Protein Sequences

Figure 2 shows the 20 amino acids that make up protein sequences, along with their single letter codes in parentheses.

Alanine (A)	Leucine (L)
Arginine (R)	Lysine (K)
Asparagine (N)	Methionine (M)
Aspartic acid (D)	Phenylalanine (F)
Cysteine (C)	Proline (P)
Glutamic acid (E)	Serine (S)
Glutamine (Q)	Threonine (T)
Glycine (G)	Tryptophan (W)
Histidine (H)	Tyrosine (Y)
Isoleucine (I)	Valine (V)

Figure 2: The 20 amino acids and their single-letter codes.

Again, a scientific description of the structure and function of proteins and amino acids is well beyond the scope of this paper. For the purpose of understanding the musification design used in Seed (see sections 5 and 6), it should suffice to say that the standard genetic code consists of 64 possible *codons* – three-letter combinations of A, C, G, and T – which encode protein sequences using the 20 amino acids shown in Figure 2. For example, the DNA codons GCT, GCC, GCA, and GCG all code for Alanine, the first amino acid listed in Figure 2. Most amino acids have multiple codons, so it is impossible to determine the DNA sequence from a given protein sequence, yet a protein sequence may be unambiguously encoded from a DNA sequence. Finally, it should also be mentioned that three of the 64 possible codons are "stop" codons that signal where a DNA sequence ends [15].

Protein sequences may be directly downloaded from the National Center for Biotechnology Information (NCBI) [20] and other genetic sequence databases, so the complexities surrounding the conversion of DNA sequences to protein sequences may be avoided in this context. The protein sequence data used in Seed is discussed in the next section.

5. A HARMONIC SERIES MODEL FOR PROTEIN SEQUENCE MUSIFICATION

Seed is a parameter-based musification of the beginning of CHD8's protein sequence. Figure 3 shows the first 160 single-letter amino acid codes in the sequence [21].

MKGESKRITLVLQQPQSGGPQGHRHVVLGSLPGKIVL
QGNQLAALTQAKNAQGQPAKVVTIQLQVQQPQQKIQI
VPQPPSSQPQPQQPPSTQPVTLSSVQQAQIMGPGQSPG
QRLSVPVKVVLQPQAGSSQGASSGLSVVKVLSASEVA
ALSSPASSAPH

Figure 3: CHD8's protein sequence.

This is the data that generates Seed. As shown in Figure 4, the 20 amino acids in Figure 2 were ordered by molecular weight (heavy-to-light) and then mapped to partials 1-20 (low-to-high) of an A1=55 Hz harmonic series [22].

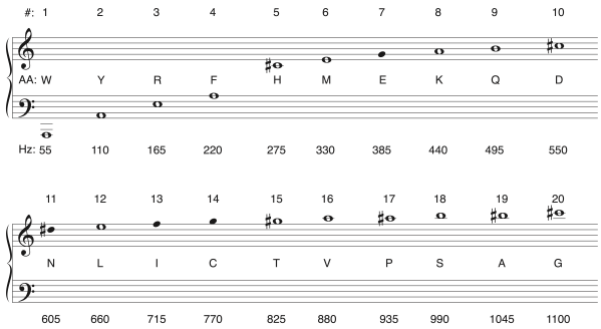

Figure 4: Amino acid to A1 harmonic series partial mapping; partial number (#); single-letter amino acid (AA) code; and frequency (Hz).

6. COMPUTATIONAL ISSUES

Cycling 74's Max [23] is a graphical programming language for music and media that is commonly used by experimental music composers, performers, and sound artists. It is an object-oriented language that is optimized for real-time human computer interaction and device-control mapping in the MIDI and audio domains. In addition to having the power of a general-purpose programming language that specializes in mapping, Max provides a large number of built-in human interface objects – number boxes, dials, sliders, menus, keyboards, notation objects, etc. – that make it easy to quickly prototype a mapping program with human-computer interaction. In Max programming, functional objects are connected together via patch cords, and messages are sent over the patch cords so that objects may communicate with one another.

Figure 5 shows the Max subprogram "aa_to_pitch" that implements Seed's mapping of single-letter amino acid codes to harmonic series partial numbers 1-20, inclusive.

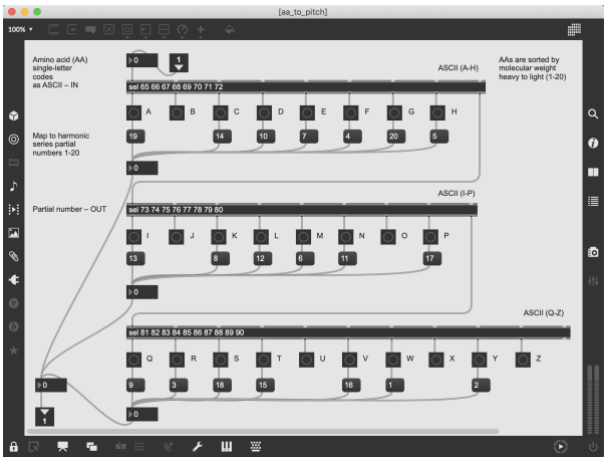

Figure 5: Max subprogram that maps single-letter amino acid codes to harmonic series partial numbers.

The input of this subprogram is the decimal ASCII value for a single-letter amino acid code, and the output is an integer. (It should be mentioned here that any protein sequence encoded in the FASTA bioinformatics format may be read into the program.) To avoid timbral fusion of the partials and achieve computational efficiency, the output of this subprogram was converted to equal-tempered pitch data and stored in a Max **coll** object – a data structure that allows the pitch series generated by the protein sequence to be converted to the duration series discussed below. The Max program encodes all musical events as MIDI information and sends that data out of Max over virtual MIDI ports to Ableton Live [24]. A custom Ableton Live Set receives the MIDI data in real time to realize the movement.

In Seed, pitch and intensity values are mapped redundantly so these two dimensions are linked. The duration dimension is also linked to the pitch dimension, but in a more complex way. In keeping with the work's overarching minimalist aesthetic and experimental music approach, durations are determined by a formula that reflects harmonic series pitch intervals into the rhythmic domain. This approach was informed by a technique employed by American composer James Tenney in *Spectral Canon for Conlon Nancarrow* (1974). In this experimental work for player piano, Tenney creates a 24-voice canon using the first 24 partials of an A1 harmonic series. Voices enter at durational octaves, and all of the voices state a duration series that is derived from successive harmonic series proportions that move forward, then backward, through the duration series. Tenney's process is discussed in detail in Polansky [25], Wannamaker [26], and Santana, Bresson & Andreatta [27].

The duration of Seed is 3 minutes and 36 seconds. It is an 8-voice canon whose leader voice is determined by the mapping described in Figure 4 and whose other voices follow the leader in strict counterpoint. The final audio recording was achieved by running a large number of compositional simulations and setting the tempo at which the data is read into the Max application, and timbral parameters in Ableton Live, to the composer's taste. Analogically, this compositional process is not unlike genetic mutation – another musico-genetic theme that is explored in the work – in that slight variations on a model may generate new forms.

7. ACKNOWLEDGMENTS

This work was done in collaboration with Jeff Dudycha, Professor of Biological Sciences at the University of South Carolina, and with the generous support of the University of South Carolina's School of Music and Department of Biological Sciences.

8. REFERENCES

[1] G. Nierhaus, *Algorithmic Composition: Paradigms of Automated Music Generation*, New York: Springer, 2009.

[2] C. Roads, *Composing Electronic Music: A New Aesthetic*, New York: Oxford University Press, 2015.

[3] R. Glover, "Minimalism, Technology and Electronic Music," in *The Ashgate Research Companion to Minimalist and Postminimalist Music*, edited by K. Potter, K. Gann, and P. A. Siôn, London: Routledge, 2013.

[4] S. Reich, and P. Hillier, ed., "Music as a Gradual Process," in *Writings on Music, 1965-2000*, New York: Oxford University Press, 2004, pp. 34-36.

[5] J. Watson and F. Crick, "A structure for deoxyribose nucleic acid," *Nature* 171 (1953): 737–738.

[6] R. Shepard, "Geometrical Approximations to the Structure of Musical Pitch," *Psychological Review* 89/4 (July 1982): 305–333.

[7] B. Eno, "Generative Music," in *A Year with Swollen Appendices: Brian Eno's Diary*, London: Faber & Faber, 2020, pp. 330–333.

[8] J. McCormack, A. Eldridge, A. Dorin, and P. McIlwain, "Generative Algorithms for Making Music: Emergence, Evolution, and EcoSystems," in *The Oxford Handbook of Computer Music*, edited by R. Dean, New York: Oxford University Press, 2009, pp. 354–379.

[9] A. Eldridge and O. Bown, "Biologically Inspired and Agent-Based Algorithms for Music," in *The Oxford Handbook of Algorithmic Music*, edited by R. Dean and A. McLean, New York: Oxford University Press, 2018, pp. 209-244.

[10] E. Varèse and C. Wen-Chung, "The Liberation of Sound," *Perspectives of New Music*, 5/1 (Autumn - Winter, 1966): 11–19.

[11] P. Vickers and B. Hogg, "Sonification Abstraite/Sonification Concrète: An 'Aesthetic Perspective Space' for Classifying Auditory Displays in the Ars Musica Domain," *Proceedings of the 12th International Conference of Auditory Display* (2006), London, UK.

[12] F. Grond and J. Berger, "Parameter Mapping Sonification," in *The Sonification Handbook*, edited by T. Hermann, A. Hunt, J. G. Neuhoff, Berlin: Logos Verlag, 2011, pp. 363–397.

[13] N. Munakata and K. Hayashi, "Basically musical," *Nature* 310 (July 12, 1984): 96.

[14] K. Hayashi and N. Munakata, "Gene Music: Tonal assignments of Bases and Amino Acids," in Clifford A. Pickover, ed., *Visualizing Biological Information*, Singapore: World Scientific, 1995, pp. 72–83.

[15] J. Dunn and M. A. Clark, “Life Music: The Sonification of Proteins,” *Leonardo* 32/1 (1999): 25–32.

[16] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell*, Cambridge: Cambridge University Press, 1944.

[17] L. A. Pray, “Discovery of DNA Structure and Function: Watson and Crick,” *Nature Education* 1/1, https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/#

[18] H. Lodish, et al., *Molecular Cell Biology*, 6th ed, New York: W. H. Freeman, 2008.

[19] R. J. Brooker, *Genetics: Analysis & Principles*, 3rd ed., New York: McGraw Hill, 2009.

[20] NCBI, https://www.ncbi.nlm.nih.gov

[21] NCBI, CHD8 chromodomain helicase DNA binding protein 8, Homo sapiens (human), NP_065971.2, https://www.ncbi.nlm.nih.gov/gene/57680

[22] K. Gann, “Ch. 2 The Harmonic Series,” in *The Arithmetic of Listening: Tuning Theory & History for the Impractical Musician,* Urbana, IL: University of Illinois Press, 2019, pp. 20–24.

[23] Cycling ’74 Max, https://cycling74.com

[24] Ableton Live, https://www.ableton.com/en/

[25] L. Polansky, “The Early Works of James Tenney,” in *Soundings 13 (1983),* edited by P. Garland, pp. 223–227.

[26] R. Wannamaker, “The Spectral Music of James Tenney,” *Contemporary Music Review* 27/1 (2008): 105–107.

[27] C. Santana, J. Bresson, and M. Andreatta, “Modeling and Simulation: The Spectral Canon for Conlon Nancarrow by James Tenney.” *Proceedings of the Sound and Music Computing Conference* (2013), Stockholm, Sweden.