

COVID-19 GENOMIC NAVIGATOR

Ka Hei Cheng

Louisiana State University - Department of Music
Center for Computation & Technology
102 New Music Building
Baton Rouge, LA, United States
christinecheng1013@gmail.com

1. ABSTRACT

COVID-19 Genomic Navigator is a motion sensing glove that sonifies the protein genomic data of *COVID-19*, which was extracted from *The National Center for Biotechnology Information (NCBI)*¹ and *Protein Data Bank (PDB)*². The PDB files show the structure of the proteins at the atomic and amino acid scale, which contributes to the multi-dimensional parameter mapping of the synthesized sounds. Three sets of data were selected in the project, including *7B3C*, *7D4F*, and *7KRP*, which are all *RNA-dependent RNA polymerase (RdRp)* related. It was formatted into prn or csv files and was read in *Max*³ patches, which sends OSC message to *Kyma (Symbolic Sound)*⁴, which performs audio signal processing in multiple synthesizers through parameter mapping. The three protein data sets distribute evenly in space while the motion sensing glove navigates and sonifies the data in both physical space and time synchronously. As shown in Figure 1, the motion sensing glove captures the data of hand gestures and movement, then sends from the *Arduino* to *Max* patch through OSC messages, which controls the graphical user interface (GUI) in *Max* that manages the parameters of the synthesizers in *Kyma* through OSC messages. The project also incorporate machine learning, which interprets the data from the glove, specifically.

Gesture recognition and *dynamic time warping (DTW)* from *Wekinator* of the motion sensing glove are used to perform machine learning and to catalyze the sonification of the interaction between the different protein data. The motion sensor data is sent from *Max* to *Wekinator* for data analysis (see figure 1). The output result of gesture recognition from *Wekinator* sends back to *Max* and trigger programming events to manifest the musical characteristics of the proteins through the control parameters of the synthesizers in *Max*. Both the protein genomic data and gestural data are reformatted and programmed to run into several synthesizers for audification and sonification. The data of five categorized glove gestures are labelled, recorded, analyzed, and

¹<https://www.ncbi.nlm.nih.gov>

²<https://www.rcsb.org>

³<https://cycling74.com/products/max>

⁴<https://kyma.symbolicsound.com>

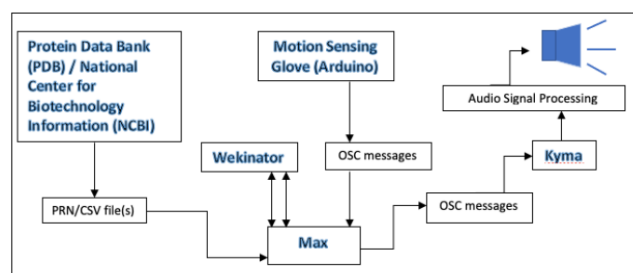


Figure 1: Data Signal Flow

trained in the machine learning model before the performance. When running in real-time, some gestures will not be categorized by the machine learning system. Non-categorized gestures would be used in musical expression alongside with the five categorized gestures to build the growth and development of the piece.

2. PROGRAM DESCRIPTION

COVID-19 Genomic Navigator is an electronic musical instrument that adopts a motion sensing glove to perform the sonification of *COVID-19* protein genomic data. It uses categorized and non-categorized gestures to demonstrate the story of *COVID-19*. The categorized gestures are used in gesture recognition while the non-categorized gestures are used for musical expression and structural development of the project.

*7B3C*⁵ shows the structure of elongating *SARS-CoV-2 RdRp* with *Remdesivir* [1] while *7D4F*⁶ shows the structure of *COVID-19 RdRp* bound to *suramin* [2]. Both *Remdesivir* [3] and *suramin* [4] are drugs that inhibit enzymatic activity of the protein and the replication and transcription of the RNA; whereas *7KRP*⁷ is the structure that shows the backtracking and transcription of RNA in coronavirus. The contrasting role of the protagonist and antagonist is expected to demonstrate through the project because the differences in functions of the proteins builds up climactic contrast.

⁵<https://www.rcsb.org/structure/7B3C>

⁶<https://www.rcsb.org/structure/7D4F>

⁷<https://www.rcsb.org/structure/7KRP>



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

The sonification process of the protagonist and antagonist proteins are used to portray the story of the struggle and reflects the contemporary life of everyone. The contrast of the role would build up musical growth and development and trigger musical events through gesture recognition in machine learning model.

In the demonstration, these three protein genomic data sets are the data source, and a research purpose is reached based on the functions of the proteins. The system built in the project allows different protein genomic data (in *PDB* format) to be loaded in the system as a data input, so as to generate the corresponding music based on the unique data structure. Therefore, users could choose any genomic data of their own preferences and compare them in the sonification system.

PDB file shows the structural significance of the *RdRp* at the atomic and amino acid levels, which contributes to the multi-dimensional parameter mapping of the synthesizing sounds. Three protein data distribute evenly in space while the motion sensing glove navigate and sonify the data in the synchronization of physical space and time dimensions.

The 3D structure of the atoms is mapped with the spherical binaural panner to recreate the 3D images of sounds, while the atoms were concatenated and sonified one by one along the time Index. Therefore, audiences could navigate the position of the atoms in three-dimensional space as if the proteins are painted sonically.

Through biodata sonification, it expresses the genetic mechanisms through data storage, analysis, and manipulation in bioinformatics; and its computational capabilities could be optimized through ubiquitous computing. Through research analysis of the interaction of genetic mechanisms and ecological environment, we could potentially someday contribute to ecological genomics, which shall become one of the criteria evaluating the scientific, sociological, and aesthetic aspects of the environment (i.e. enzymatic responses between the proteins through sonification and its effect in the community). The methodology of designing electronic musical instruments as a medium of data representation and musical expression of the biological and genetic aspects of ecological genomics through sonification shall include and express the conceptual framework of somaesthetics, advocate the theoretical, empirical, and practical principles related to human body perception on ecology, instrumental performance, and presentation.

3. PERFORMANCE

YouTube Link: https://youtu.be/wHHjV_uvobs

4. ACKNOWLEDGEMENT

Thank you all the support from *Carla Scaletti* and *Kurt J. Hebel*.

5. REFERENCES

[1] G. Kokic, H. S. Hillen, D. Tegunov, C. Dienemann, F. Seitz, J. Schmitzova, L. Farnung, A. Siewert, C. Höbartner, and

P. Cramer, “Mechanism of SARS-CoV-2 polymerase stalling by remdesivir,” *Nature communications*, vol. 12, no. 1, pp. 1–7, 2021.

[2] N. Palko, M. Grishina, and V. Potemkin, “Electron Density Analysis of SARS-CoV-2 RNA-Dependent RNA Polymerase Complexes,” *Molecules*, vol. 26, no. 13, p. 3960, 2021.

[3] J. H. Beigel, K. M. Tomashek, L. E. Dodd, A. K. Mehta, B. S. Zingman, A. C. Kalil, E. Hohmann, H. Y. Chu, A. Luetkemeyer, S. Kline, *et al.*, “Remdesivir for the treatment of Covid-19,” *New England Journal of Medicine*, vol. 383, no. 19, pp. 1813–1826, 2020.

[4] C. Salgado-Benvindo, M. Thaler, A. Tas, N. S. Ogando, P. J. Bredenbeek, D. K. Ninaber, Y. Wang, P. S. Hiemstra, E. J. Snijder, and M. J. Van Hemert, “Suramin inhibits SARS-CoV-2 infection in cell culture by interfering with early steps of the replication cycle,” *Antimicrobial Agents and Chemotherapy*, vol. 64, no. 8, pp. e00900–20, 2020.