# SARS Coronavirus Replicase Sonification

*Edward J. Martin*

## Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh Charlotte Auerbach Road Edinburgh, United Kingdom
**E.j.martin@sms.ed.ac.uk**

**https://vimeo.com/sonifyed/sars-coronavirus-replicase**

## ABSTRACT

The year 2020 will be remembered for one thing so obvious to all of us that naming it seems ridiculous. The Covid-19 coronavirus pandemic will be a chapter in all of our life-stories.

This work was created in February 2020 and premiered at a live in-person concert (harmonic.function) in Edinburgh, Scotland on 5th March [1]. It was created with the context of the developing pandemic - *before we knew*. It was performed in the context of cognitive dissonance – *before we acknowledged*. Both the host and audience – containing a large quotient of research biologists – tittered at the suggestion that anyone had not heard of the virus, though all were still out at a concert unmasked. Amid universal uncertainty, the piece was borne of, and reflecting on, that overload of political and scientific opinion.

The piece is created using a sonification process of the protein sequence of the replicase polyprotein, the method by which coronaviruses replicate in host cells. This sonification process is then live-mixed alongside vocal samples by a performer to create the audio of the piece. A video demonstrating a similar viral process in a retrovirus accompanies the sound.

The piece represents the time before lockdown, mirroring a state of mind of academic (dis-)interest in the unforeseen burgeoning pandemic. It was an attempt to understand what was going on. It is now an artefact of that unique time.

## 1. BACKGROUND

Proteins are an essential part of all organisms, including viruses. They are molecules composed of long strings of amino acids – linked together like links in a chain. The sequence of the amino acids determines the three-dimensional structure and biological role of the protein. In bioinformatics, the 20 amino acids are represented by capital letters which can be seen in Figure 1.

Viruses are very small agents that can only replicate within the cells of living organisms. In order to survive, a virus must penetrate the cell walls of a host and hijack the cellular machinery to reproduce [2]. Coronaviruses do this using a polyprotein called replicase – this is a series of many attached smaller proteins. Each of these proteins is specialised to perform one step of the heist. This allows the virus to reproduce and spread throughout the host. In the video
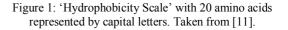
accompanying this piece, the equivalent to the coronavirus replicase is labelled as a 'multi-protein chain' and appears at 03:48.

In the interests of contextualising the creation and first reception of the work, I will outline a short timeline of the outbreak in the UK. The first confirmed cases the UK were reported on the 31st January [3]. The story grew to dominate the news headlines by the 23rd February - with an outbreak on a UK-registered cruise ship, *The Diamond Princess* [4]. By the 2nd March the government's emergency COBRA committee had met and announced to the public that widespread transmission was 'highly likely' [5]. The UK suffered its first fatality on the 5th March [6], with Scotland's first fatality following on the 14th [7]. By the 23rd March, the police were ordered to enforce a ban on gatherings of more than two people and strict limits on exercise, as the British public were told: 'You must stay at home'[8].

At the time of the work's composition in February 2020, researchers were trying to understand the new coronavirus through past outbreaks. Two recent viral outbreaks caused by coronaviruses were the Severe acute respiratory syndrome (SARS) outbreak in 2003 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 [9,10]. Both were seen as models from which we could determine our response.

## 2. METHOD



Figure 1: 'Hydrophobicity Scale' with 20 amino acids represented by capital letters. Taken from [11].

This piece takes the protein sequence of the replicase polyprotein 1ab from the SARS coronavirus (UniProtKB: P0C6X7) [12]. This is the underlying dataset of the sonification work in the piece.

The use of the replicase polyprotein from a different SARS coronavirus to Covid-19 is indicative of the uncertainty of the time. Good quality, corroborated sequence data was not easily

available at the time – so an analogue was used for the piece. This mirrors the public health and news media information at the time – searching for appropriate analogues to explain and understand an unprecedented scenario.

The sound is generative. Other than the vocal samples and a synthesized drone, all parts of the sound are driven by the sequential amino acid residues from the protein sequence of the replicase polyprotein. One amino acid residue is used per beat, and the system increments to use the next amino acid in the protein sequence. The sound ends when the entire sequence has been used.

The specific process of sonification is different for the percussion samples, bass synthesizer, and lead synthesizer. However, each follows the broad schema of using a Perl script to map an amino acid residue to MIDI sound parameter (i.e. pitch). This MIDI data is then triggered sequentially by MaxMSP software and sent to the Digital Audio Workstation (DAW), Ableton Live 10. A pulse is felt through the piece as the MIDI data is sent and triggers sound synthesis sequentially.

The tempo of the piece changes by incremental steps caused by a random seed. The performer is also able to make larger step changes in tempo using the MaxMSP patch.

This work uses the 'hydrophobicity scale' (see Figure 1) mapping developed by the author as part of scientific research into sonification of biological sequence data [11]. This mapping uses experimental data on how each amino acid interacts with water (known as hydrophobicity) to create a unique, 1-to-1 mapping from the amino acids to western musical tones (encoded as MIDI numbers). This is done using a Perl script available from GitHub [13] (Algorithm I).

The lead distorted synthesizer, which can be heard from roughly the 2-minute mark, is generated directly from the protein sequence of the Replicase polyprotein. This uses the 'hydrophobicity scale' mapping seen in Figure 1. When turned up in the mix by the performer, every residue of the polyprotein can be heard as they are triggered.

The bass parts are similarly determined, however the 'hydrophobicity scale' mapped MIDI numbers are put through a mathematical transformation in the MaxMSP patch which creates irregular and long bass notes throughout the piece.

The drums/percussion are triggered from drum loop samples using the same 'hydrophobicity scale' MIDI numbers. Here the numbers are matched onto a different location within the drum sample from which the sound is created. There are several different percussion samples used simultaneously and the performer changes their levels in the mix.

The vocal samples are taken from old news archive interviews about viral outbreaks in the past. These *vox-pops* add depth to the piece, centring the social context of viral outbreak and human interaction. These are triggered live by the performer.

The visual accompaniment is kindly provided by the HHMI's Biointeractive suite of educational materials [14]. The video shows a similar process to the replicase polyprotein in a coronavirus, however the video is created specifically detailing HIV – a retrovirus. The visual aesthetic captures many of the images shown uninterrupted on screens and feeds through the early pandemic at the time of the work's production. This 'scientific aesthetic' was how the unseen virus could be communicated to us all as a threat.

As a whole, the design choices of the piece reflect the zeitgeist of this time of great uncertainty - a time where everyone became epidemiologists, hospitality venues were told to stay open while the public were told not to go to them, and I assumed I would follow the performance in March with other live performances all through the summer after this all died down. Choices were made off the cuff alongside the notion of *this is good enough*.

## 3.  PERFORMANCE

The piece has a live performance aspect. As the MIDI data is sent to the DAW (Ableton Live 10 Suite) the performer is able to manipulate the sound as it is synthesized. The performer controls the different channels driven by changing the relative volumes, turning on/off different channels, and triggering vocal samples. The performer is also able to make large step changes in tempo via the MaxMSP patch. In this way the performer controls the production of the piece from the constant pulsing data created from the replicase polyprotein sequence.

Although this is the method the piece was conceived and first performed, for the ICAD 2021 performance a recording of a live performance has been made. This live-recorded studio performance is available at https://vimeo.com/sonifyed/sars-coronavirus-replicase

## 4.  RECEPTION

The piece first and to date only live performance came at 'harmonic.function' at the University of Edinburgh on 5th March 2020. Anonymous feedback on the performance was positive, with the striking soundscape complimented.

Some members of the original audience have seen the video a year later – having experienced the effects of the coronavirus pandemic in between. This later feedback spoke about the piece in much grander terms, finding the piece moving and emotional.

## 5.  ACKNOWLEDGEMENTS

## 6.  REFERENCES

1.      Harmonic Function - Festival of Creative Learning [Internet]. [cited 2021 Jun 10]. Available from: https://festivalofcreativelearning.wordpress.com/202

0/04/09/harmonic-function/

2. Ziebuhr J. The coronavirus replicase. Vol. 287, Current Topics in Microbiology and Immunology. Springer, Berlin, Heidelberg; 2005. p. 57–94.

3. Coronavirus: Two cases confirmed in UK - BBC News [Internet]. [cited 2021 Jun 10]. Available from: https://www.bbc.co.uk/news/health-51325192

4. Coronavirus: Four new UK cases among ship evacuees - BBC News [Internet]. [cited 2021 Jun 10]. Available from: https://www.bbc.co.uk/news/uk-51606368

5. Coronavirus: Widespread transmission in UK "highly likely" - BBC News [Internet]. [cited 2021 Jun 10]. Available from: https://www.bbc.co.uk/news/uk-51700604

6. Coronavirus: Woman in 70s becomes first virus fatality in UK - BBC News [Internet]. [cited 2021 Jun 10]. Available from: https://www.bbc.co.uk/news/uk-51759602

7. Coronavirus: Death of first Scottish patient with Covid-19 confirmed - BBC News [Internet]. [cited 2021 Jun 10]. Available from: https://www.bbc.co.uk/news/uk-scotland-51874899

8. Boris Johnson orders UK lockdown to be enforced by police | Coronavirus | The Guardian [Internet]. [cited 2021 Jun 10]. Available from: https://www.theguardian.com/world/2020/mar/23/boris-johnson-orders-uk-lockdown-to-be-enforced-by-police

9. Severe Acute Respiratory Syndrome (SARS) [Internet]. [cited 2021 Jun 10]. Available from: https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1

10. Middle East respiratory syndrome coronavirus (MERS-CoV) [Internet]. [cited 2021 Jun 10]. Available from: https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers#tab=tab_1

11. Martin EJ, Meagher TR, Barker D. Using sound to understand protein sequence data: new sonification algorithms for protein sequences and multiple sequence alignments. BMC Bioinformatics, under Rev.

12. UniProtKB. Rep - Replicase polyprotein 1ab - Severe acute respiratory syndrome coronavirus (SARS-CoV) - rep gene & protein [Internet]. [cited 2021 Apr 30]. Available from: https://www.uniprot.org/uniprot/P0C6X7

13. Martin EJ. Protein_Sound [Internet]. GitHub; 2021. Available from: https://doi.org/10.5281/zenodo.4683746

14. HHMI. HIV Life Cycle [Internet]. Biointeractive. 2007 [cited 2021 Apr 30]. Available from: https://www.biointeractive.org/classroom-resources/hiv-life-cycle