MSc Informatics: Sonification of Biological Sequence Data

Page 1: Participant Information

Please read the Participant Information Sheet

https://static.onlinesurveys.ac.uk/media/account/93/survey/494461/question/participant_information_sheet_.pdf

If you have any questions before completing the survey, please contact researcher Edward Martin (E.j.martin@sms.ed.ac.uk).

Project Title: The use of paramater-mapping sonification to facilitate knowledge discovery from protein multiple sequence alignments in bioinformatics

Principal Investigator (PI): Daniel Barker

Researcher: Edward Martin

PI contact details: Daniel.Barker@ed.ac.uk

Please tick yes or no for each of these statements

1. I confirm that I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction. ✽ *Required*

○ Yes	○ No

2. I understand that my participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights. ✽ *Required*

○ Yes	○ No

3. I consent to my anonymised data being used in academic publications and presentations. ✽ *Required*

○ Yes	○ No

4. I understand that my anonymised data can be stored for a minimum of two years. ✽ *Required*

○ Yes	○ No

5. I allow my data to be used in future ethically approved research. ✽ *Required*

○ Yes	○ No

6. I agree to take part in this study. ✽ *Required*

○ Yes	○ No

7. I confirm that I am based in the UK. ✱ *Required*

○ Yes	○ No

8. Name of person giving consent: ✱ *Required*

[   ]

Name of person taking consent: Edward Martin

The link below contains a hard copy of the equivalent printed consent form.

https://static.onlinesurveys.ac.uk/media/account/93/survey/494461/question/190313_participant_consent_for.pdf

○ Yes	○ No

Page 2: Participant Experience

9. What level of experience with biological sequence data do you possess? ✳ *Required*

```
[                    ]
```

10. What level of musical experience do you possess? ✳ *Required*

```
[                    ]
```

Page 3: Protein Task

https://soundcloud.com/sonifyed/major-prion-protein

This sound file contains a sonification of a major human prion protein. Each residue has been sequentially translated into sound. The method has mapped each amino acid to pitch according to a hydrophobicity scale. An increase in pitch corresponds to an increase in hydrophillicity.

This protein contains a short (<20 letters) amino acid motif, or word, repeated four times.

Please listen to the sound file and use the sonification to identify the repeated pattern. When you have identified the pattern, please enter the corresponding sequence in the box below (e.g. MCAPQ...).

To help identify the residues used, please make use of the sequence given below and listen to the sonification as many times as you like.

11. Enter your solution here: ✳ *Required*

Please enter a response that contains only upper case letters.
Your answer should be no more than 20 characters long.

Here is the example protein sequence visualised with MView. The '1' and '2' above the sequence mark the 100th and 200th residues in the alignment.

```
                     cov      pid    1 [      .        .        .        .        :        .        .        . 80
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%     MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGG

                     cov      pid   81      .        1        .        .        .        .        .        : . 160
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%     WGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQ

                     cov      pid  161      .        .        .        2        .        .        .        . 240
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%     VYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV

                     cov      pid  241         :   ] 253
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%     ILLISFLIFLIVG
```

Page 4: Protein Task Feedback

Here is the repeated amino acid motif highlighted on the sequence. The sequence is 'PHGGGWGQ':

```
                        cov    pid   1 [         .         .         .         :         .         .         . 80
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%   MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGG

                        cov    pid  81         .         1         .         .         .         :         . 160
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%   WGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQ

                        cov    pid 161         .         .         .         2         .         .         . 240
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%   VYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV

                        cov    pid 241           :  ] 253
1 sp|P04156|PRIO_HUMAN 100.0% 100.0%   ILLISFLIFLIVG
```

MView 1.63, Copyright © 1997-2018 Nigel P. Brown

12. Did the sonification sound file help you identify the repeated motif? ✳ *Required*

○ Yes
○ No

13. What was the best thing about the protein sonification? ✳ *Required*

14. What was the worst thing about the protein sonification? ✳ *Required*

Page 5: Multiple Sequence Alignment Task

https://soundcloud.com/sonifyed/algorithm-v-questionnaire

This sound file contains a sonification of a multiple sequence alignment of 5 members of a protein family. Each residue has been sequentially translated into sound. The method has mapped each amino acid to pitch according to a hydrophobicity scale. An increase in pitch corresponds to an increase in hydrophillicity. Gaps have been mapped to rests. Conserved amino acids will sound louder.

This protein contains 3 conserved domain (<50 letters).

Please listen to the sound file and use the sonification to identify the repeated pattern.

To help identify the residues used, please make use of the sequence alignment given below and listen to the sonification as many times as you like. Please do not proceed until you have made a prediction as to the location of the conserved domain.

Multiple Sequence Alignment visualised using MView:

```
                                      cov     pid   1 [         .         .         .         :         .         .         . 80
1 tr|A0A087ZXM2|A0A087ZXM2_APIME 100.0% 100.0%     MEAIAKHDFTATAEDELSFRRSQILKILNMEDDMNWYRAELDSREGLIPSNYIEMKNHDWYYGRITRADAERLLMNKH--
2 tr|A0A016SG30|A0A016SG30_9BILA  83.4%  49.0%     MEATAEHDFNANADDELSFRRGQILKVLNKDEDPHWFKAEIDGVEGFVPSNYIRMHDSPWYLGKISRLDAEILLKKQG--
3 tr|A0A044TM65|A0A044TM65_ONCVO  83.4%  52.5%     MEAVAEHDFNATAEDELSFRKNQILKVLNKDEDPHWYKAELDGHEGFIPSNYIRMNEHDWYLGKISRADAEALLLRSG--
4 tr|A0A068Y1U5|A0A068Y1U5_ECHMU  85.3%  39.6%     MEADAKHDFQASAADELPFCKNSVLKILSVEDDPNWYLAEQDGRKGLVPCNYISFRPNPWYMQACRRNTAEECLLETDPY
5 tr|W5Q6L9|W5Q6L9_SHEEP          80.7%  29.8%     MEAIAKFDFMASGEDELSFHAGDVLKILSNQEE--WFKAELGSQEGYVPKNFIDIKFPEWFHEGLSRHQAESLLMGKE--

                                      cov     pid  81         .         1         .         .         :         .         . 160
1 tr|A0A087ZXM2|A0A087ZXM2_APIME 100.0% 100.0%     -------EGAFLIRISESS-PGDFSLSVKCSDGVQHFKVLRDAQGKFFLWVVKFNSLNELVEYHRTASVSRSQDVKLRD-
2 tr|A0A016SG30|A0A016SG30_9BILA  83.4%  49.0%     -----TRDGNFLVRQCESS-PGDFSISVKFEDTIQHFKVLRDNNGKYFLWSVKFKSLNELVRYHRTASVSRTHTILLQD-
3 tr|A0A044TM65|A0A044TM65_ONCVO  83.4%  52.5%     -----NVDGAFLVRQSESS-PGDFSISVRFQDSVQHFKVLRDNNGKYYLWVVKFNSINELITYHRSASVSRSHTILLQN-
4 tr|A0A068Y1U5|A0A068Y1U5_ECHMU  85.3%  39.6%     TGLPVQPDGAFVVRRSESNGPG-FSLSVKDGQKVQHFKVLQDEMGKYFVWLRKFDSINQLIDYHRRTSISRDGFLLLVD-
5 tr|W5Q6L9|W5Q6L9_SHEEP          80.7%  29.8%     -------LGFFIIRASQSS-PGDFSISVRHEDDVQHFKVMRDNKGNYFLWTEKFPSLNKLVDYYRKNSISKQKQIFLRDR

                                      cov     pid 161         .         .         .         2         .         .         . 240
1 tr|A0A087ZXM2|A0A087ZXM2_APIME 100.0% 100.0%     --------------------------------------------------------------------------------
2 tr|A0A016SG30|A0A016SG30_9BILA  83.4%  49.0%     --------------------------------------------------------------------------------
3 tr|A0A044TM65|A0A044TM65_ONCVO  83.4%  52.5%     --------------------------------------------------------------------------------
4 tr|A0A068Y1U5|A0A068Y1U5_ECHMU  85.3%  39.6%     --------------------------------------RQPS-------------------------------R--------
5 tr|W5Q6L9|W5Q6L9_SHEEP          80.7%  29.8%     TREEQGQRGNSLDRRSQGGHPLSGAVGEEMRPSMNRKPSDHPLLPASQYPPAPLQSQQRQHQHHFNQERRGTSLDINDGP

                                      cov     pid 241         :         .         .         .         .         3         . 320
1 tr|A0A087ZXM2|A0A087ZXM2_APIME 100.0% 100.0%     --------MV------PEECLVQ-----------ALYDFQPQEPGELEFKRGDVITVTDRTDQHWWHGEIGN--RRGLFP
2 tr|A0A016SG30|A0A016SG30_9BILA  83.4%  49.0%     --------LD------LETKFVQ-----------ALFDFNPQEEGELPFKRGEIITLINKDDMNWWEGSLNN--KRGVFP
3 tr|A0A044TM65|A0A044TM65_ONCVO  83.4%  52.5%     --------MDSVAA--QGTHLVQ-----------AMFDFKPQEEGELGFKRGDIITVTNREDENWWEGTLNG--KSGMFP
4 tr|A0A068Y1U5|A0A068Y1U5_ECHMU  85.3%  39.6%     --------MISMGRGDQQSQMVQ-----RVI---ARFDFNASEPEELSFHRGDVIEVLGQEDENWWRGRISNTGSTGLFP
5 tr|W5Q6L9|W5Q6L9_SHEEP          80.7%  29.8%     CGLGMGTEMNAALMHRRHTDPVQLQVAGRVRWARALYDFEALEDDELGFRCGEVVEVLDSSNPSWWTGRLHN--KLGLFP

                                      cov     pid 321         .         .         :         .         ] 377
1 tr|A0A087ZXM2|A0A087ZXM2_APIME 100.0% 100.0%     STYASCCAHRGQRHRPHYPTTITYRIRLTAELINNILQTPPIEWATRMYWWSRRRVV
2 tr|A0A016SG30|A0A016SG30_9BILA  83.4%  49.0%     ANYVC-----------------------------------PFNRSPSNT--------
3 tr|A0A044TM65|A0A044TM65_ONCVO  83.4%  52.5%     ATYVC-----------------------------------PFNNSTSAH--------
4 tr|A0A068Y1U5|A0A068Y1U5_ECHMU  85.3%  39.6%     ANYVD---------------TL----------------PPIMP-NAAH----RPTN
5 tr|W5Q6L9|W5Q6L9_SHEEP          80.7%  29.8%     ANYVA-----------------------------------PMIR------------
```
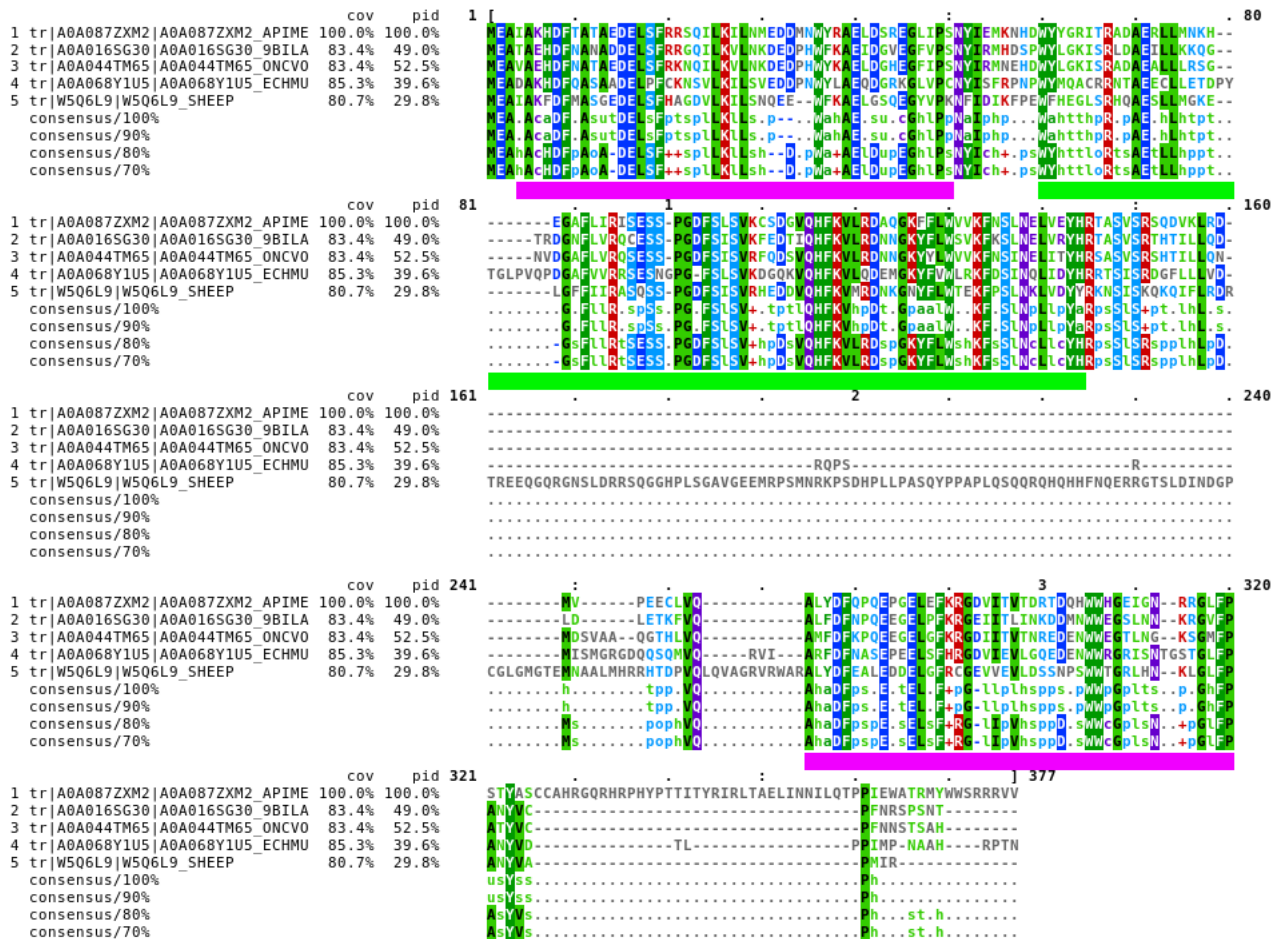
MView 1.63, Copyright © 1997-2018 Nigel P. Brown

Page 6: Multiple Sequence Alignment Task Feedback

Here follows the multiple sequence alignment with the 3 conserved domains highlighted manually with the 3 coloured bars. The 1st and 3rd domain are the same.

The other colouring is from MView:

MView 1.63, Copyright © 1997-2018 Nigel P. Brown

---

15. Did the sonification sound file help you identify the conserved domains in the multiple sequence alignment? ✴ *Required*

○ Yes
○ No

16. What was the best thing about the multiple sequence alignment sonification? ✴ *Required*

17. What was the worst thing about the multiple sequence alignment sonification? * *Required*

Page 7: NASA Task Load Index 1/2

The NASA Task Load Index is a framework of evaluating workload of tasks developed by the famous space agency. It is typically used to evaluate human-computer interaction tasks. It consists of six questions ranked on a 23 point scale, followed by 15 'this-or-that' choices to gauge the aspects most relevant to the task.

In this case the task refers to using the sonification to find the conserved domains in the multiple sequence alignment. Therefore the following questions concern **only the second task with the multiple sequence alignment.**

If required, definitions of the terminology are given at the bottom of the page.

Please note that the scales on this page are all identical, except for the fourth, which has 'good' on the left and 'bad' on the right.

18. Mental Demand: How mentally demanding was the task? ✳ *Required*

	1	2	3	4	5	6	7	8	9	10	11	12
Very Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

19. Physical Demand: How physically demanding was the task?

	1	2	3	4	5	6	7	8	9	10	11	12
Very Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

20. Temporal Demand: How hurried or rushed was the pace of the task?

	1	2	3	4	5	6	7	8	9	10	11	12
Very Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

21. Performance: How successful were you in accomplishing what you were asked to do?

	1	2	3	4	5	6	7	8	9	10	11	12
Perfect	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

22. Effort: How hard did you have to work to accomplish your level of performance?

	1	2	3	4	5	6	7	8	9	10	11	12
Very Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

23. Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

	1	2	3	4	5	6	7	8	9	10	11	12
Very Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	good/poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Source: https://humansystems.arc.nasa.gov/groups/tlx/tlxpaperpencil.php

Page 8: NASA Task Load Index 2/2

24. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Effort   ○ Performance

25. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Temporal Demand   ○ Frustration

26. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Physical Demand   ○ Frustration

27. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Temporal Demand   ○ Effort

28. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Physical Demand   ○ Performance

29. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Frustration   ○ Performance

30. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Physical Demand   ○ Temporal Demand

31. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Temporal Demand   ○ Mental Demand

32. Which is the more important contributor to workload for the task you performed in this experiment? ✳ *Required*

○ Frustration   ○ Effort

33. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Mental Demand	○ Performance

34. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Temporal Demand	○ Performance

35. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Effort	○ Mental Demand

36. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Mental Demand	○ Physical Demand

37. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Effort	○ Physical Demand

38. Which is the more important contributor to workload for the task you performed in this experiment? ✱ *Required*

○ Frustration	○ Mental Demand

Page 9: Further Information

39. Do you have any other comments?

[   ]

40. If you would like to receive further information about this research project, please leave your email address in the box below and indicate which you would be interested in receiving (tick any number of boxes)

☐ MSc Thesis
☐ Any future publications or articles

41. Contact Email Address (REQUIRED if you ticked one or more boxes in Question 31)

[   ]

Page 10: Final page

Thank you for your participation in this survey.

If you have any further questions, please contact Edward Martin (E.j.martin@sms.ed.ac.uk)

Key for selection options

9 - What level of experience with biological sequence data do you possess?
    Little or No Experience (0 years)
    Equivalent to High School Study (0-2 years)
    Equivalent to Undergraduate Study (2-4 years)
    Equivalent to Master's Degree Study (4-6 years)
    Equaivalent to PhD Study (6-8 years)
    Experience beyond PhD Level (>8 years)

10 - What level of musical experience do you possess?
    Little or No Experience (0 years)
    Equivalent to High School Study (0-2 years)
    Equivalent to Undergraduate Study (2-4 years)
    Equivalent to Master's Degree Study (4-6 years)
    Equaivalent to PhD Study (6-8 years)
    Experience beyond PhD Level (>8 years)