Activating *DTWA Consensus and Voting Tool*

This code is supplied as an *“AS-IS”* tool for research with no warranties regarding reliability

For further Information contact *Dr. Mike Smith* at Mike.Smith @ ucalgary.ca

|  |  |
| --- | --- |
| 1. Build a directory *DTWA\_C\_andV\_Tool* 2. Download and expand *“SquiggleStreamData\_As\_txt\_Files.zip”* as a sub-directory 3. Download and expand “*DTWAConsensusAndVotingTool\_Plos1\_June2021.zip”* as a sub-directory” 4. Activate *startup.m* to set immediate Matlab paths 5. Directory information should look like adjacent picture The ghosted directories are not on the Matlab path and are filled with intermediate data to avoid, where possible, recalculation 6. Run *“Build\_Figs1\_3\_Demonstrate\_HeaderRemoval\_Plos1\_June2021”* to check can access data files 7. This code will generate a large version of Fig. 1 to meet PACE requirement 8. I save the small figures for Fig. 3 as .tif files, then join them together inside a Powerpoint presentation and save as combined .tif figure |  |

1. The file *Ensemble\_DefaultSettings\_Plos1\_June2021* in *Ensemble\_DefaultSettings\_Plos1\_June2021* directory contains default settings
2. All the figures in the paper can be generated by activating a Build script. You will have to micro-move the captions A), B) positions for the majority of the pictures depending on your data characteristics

The code runs in four stages, Fig 2.

1. Select *N* streams from squiggle ensemble starting at stream *NSTART* and ending at stream   
   *NSTART* *+ N – 1.* Generate and save cleaned streams if not already stored
2. Generate and save the initial *DBA, SSG* and *MM DTWA* consensus signals built from the cleaned streams if not already stored
3. Prepare warping paths between each ensemble member and consensus and determine the number of ensemble squiggles that agree that this warped path location in the consensus is false.
4. In a loop, generate final consensus variants derived from the initial consensus with entries deleted based on agreement voting levels between 100% to 10%. Choose the final consensus from the variants using a success metric.

To do a quick test (but not the best results) – using 16 squiggles --- then set *doTest = true* in

*Build\_Figs6\_10AB\_14ABCD\_Plos1\_June2021.*m

*Build\_Figs7AB\_Part1\_Plos1\_June2021  
 Build\_Figs11AB\_CompareDTW\_AllGroups\_Plos1\_2021*

*Build\_Figs12AB\_CompareDTW\_MultipleStreamEnolase\_Plos1\_June2021*

Other figures use

*Script\_DTWPathFigs\_Build\_AddCaption\_Plos1\_June2021.m*

*.*

|  |  |
| --- | --- |
|  | *Code\_AccessDataUtilities – Some hard coded paths*  *Code\_DefaultSettings – code sets defaults for various stages of header cleaning, voting and DTWA generation*  *Code\_DTWA -- DTWA code from other authors*  *Code\_PrepareData -- Header cleaning*  *Code\_Utilities – Adjust Matlab dtw( ) pictures and generates new warp-path displays, and does a couple of fancy plotting things, such as multiple colour markers in Figures 11 and 12* |

**Adding your own data**

|  |  |
| --- | --- |
|  | System expects  Data Directory with name as Data file name  DataFileName.txt  GoldStandard\_DataFileName.txt  **And for historical reasons**  TotalNumberStreams\_DataFileName.txt  Is number of streams in file |
| function Build\_Figs4ABC\_MyData1\_Plos1\_June2021()  Script\_DTWPathFigs\_Build\_AddCaption\_Plos1\_June2021(…  'MyData1\_Fig4ABC'); | Example rebuilding Fig 4 using your data |
| Script\_DTWPathFigs\_Build\_AddCaption\_Plos1\_June2021  case {'MyData1\_4ABC'}  numStreams = 128; | Line 18 |
| SetupFiguresParameters\_Plos1\_June2021  case {'MyData1\_4ABC'}  buildFig.buildFig4ABC = true;  ensembleNum = 4; % MyData1  ensembleName = 'MyData1'  posnEnsembleStartStream = 1; | Line 17 |

***One tricky bit of code***

***The code*** *Voting\_DefaultSettings\_Plos1\_June2021.m* sets the parameters for voting – meaning it contains the preset levels for agreement that will be evaluated

I set the ‘Quick test’ mode to use only 2 values for each study group – to generate a fast display to show something working.

However, I also found that I liked the 2 voting values only mode when empirically looking for the best voting value (consensus length = gold length) for each DTWA approach and it became a pain to keep on changing the settings inside *Voting\_DefaultSettings\_Plos1\_June2021.m*

So I did a quick fix and added specific *DTWA* agreement settings for the ‘Quick Test’ at

*DetermineVoteOn\_DTWA\_Plos1\_June2021.m lines 11 on*   
***(Watch out the best SSG vote changes unless you seed the random number routines to do the same thing each run – but is also interesting to see happen)***

**However, since my code looks to minimize recalculating anything, if you change the two vote numbers, they don’t get recognized – so the 2 vote levels never change.**

**I don’t need to permanently solve this problem, as the code must hunt for the best vote not have it preset.**

**But I needed to have a quick fix when I needed to do some empirical hunting for best values to display.**

**The answer was -- delete .mat files in directory *VotedConsensus\_Plos1\_June2021.* There is no great time penalty as I am only calculating 2 votes**