

User Guide

Contents

[What is HAPPE? 4](#_Toc79067257)

[What does HAPPE require? 5](#_Toc79067258)

[How to get HAPPE 6](#_Toc79067259)

[Setting Up HAPPE: 7](#_Toc79067260)

[Adjust Java Heap Settings 7](#_Toc79067261)

[Set MATLAB Paths 8](#_Toc79067262)

[Configure EEGLAB Preferences 9](#_Toc79067263)

[How to Run HAPPE v2 11](#_Toc79067264)

[HAPPE Pipeline Guides 12](#_Toc79067265)

[HAPPE Resting State 12](#_Toc79067266)

[Following Command Line Prompts 12](#_Toc79067267)

[HAPPE Task-Related 22](#_Toc79067268)

[Following Command Line Prompts 22](#_Toc79067269)

[HAPPE ERPs (HAPPE+ER) 32](#_Toc79067270)

[Following Command Line Prompts 32](#_Toc79067271)

[HAPPILEE Resting-State 44](#_Toc79067272)

[Following Command Line Prompts 44](#_Toc79067273)

[HAPPILEE Task-Related 53](#_Toc79067274)

[Following Command Line Prompts 53](#_Toc79067275)

[HAPPILEE ERPs 63](#_Toc79067276)

[Following Command Line Prompts 63](#_Toc79067277)

[Data Quality and Pipeline Quality Outputs 74](#_Toc79067278)

[Tracking Data Quality 74](#_Toc79067279)

[Using Outputs to Exclude Files 74](#_Toc79067280)

[Data Quality Measures 74](#_Toc79067281)

[Pipeline Quality Measures 74](#_Toc79067282)

[Variables as Co-Variates and Reporting Measures 74](#_Toc79067283)

[Data Quality Measures 74](#_Toc79067284)

[Pipeline Quality Measures 74](#_Toc79067285)

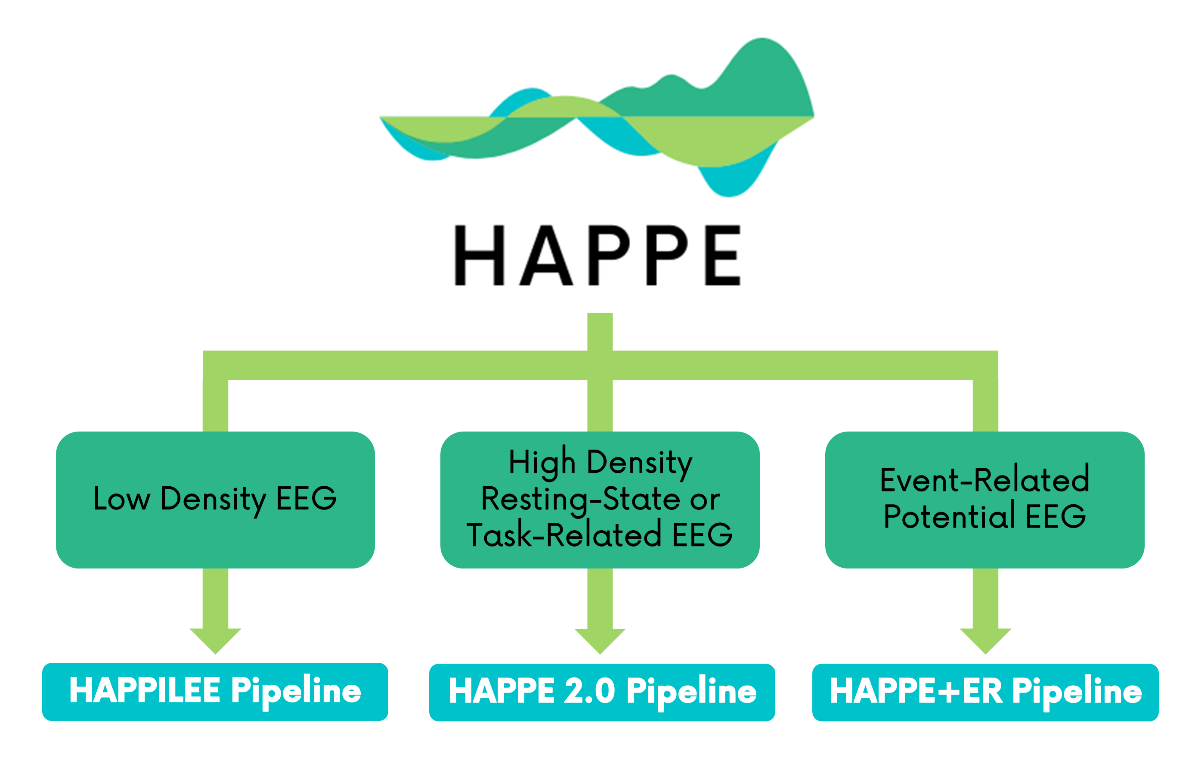
[HAPPE+ER’s generateERPs Add-On 75](#_Toc79067286)

[Run generateERPs 75](#_Toc79067287)

[Following Command Line Prompts: 75](#_Toc79067288)

# What is HAPPE?

* HAPPE is a software for taking unprocessed EEG data and automatically processing it in preparation for analysis. It is a combination of the HAPPE 2.0 pipeline, the HAPPE+ER pipeline, and the HAPPILEE pipelines.
  + HAPPE 2.0: citation pending
  + HAPPE+ER: Monachino, A.D., Lopez, K.L., Pierce, L.J., Gabard-Durnam, L.J. (submitted) The HAPPE plus Event-Related (HAPPE+ER) Software: A Standardized Processing Pipeline for Event-Related Potential Analyses
  + HAPPILEE: Lopez, K.L., Monachino, A.D., Morales, S., Leach, S.C., Bowers, M.E., Gabard-Durnam, L.J. (submitted) HAPPILEE: The Harvard Automated Processing Pipeline in Low Electrode Encephalography, a standardized software for low density EEG and ERP data.
* Translates recent advances in adult EEG processing to developmental data context.
* Implements wavelet-thresholding approaches for EEG artifact removal.
* Agnostic to the program for running analyses afterwards. Compatible with:
  + BEAPP
  + EEGLAB
  + MATLAB
  + Anything that can read a .txt file of EEG data as an input
* Includes processing reports with data and pipeline quality metrics for assessing data and performance, reporting in manuscripts, and setting quality thresholds for removing data from further analysis.



# What does HAPPE require?

* MATLAB version 2014 or newer
  + NOTE: this version is not compatible with NetStation version 4.5, so you cannot run HAPPE on the same computer as NetStation 4.5.
* The following MATLAB toolboxes (all should be standard for academic licenses):
  + Signal Processing Toolbox
  + Optimization Toolbox
  + Statistics Toolbox
  + Wavelet Toolbox
* The following free software (all included in the HAPPE download):
  + EEGLAB
  + CleanLine EEGLAB plugin
  + MARA EEGLAB plugin
  + FASTER functions
* The HAPPE scripts (all included in the HAPPE download)

# How to get HAPPE

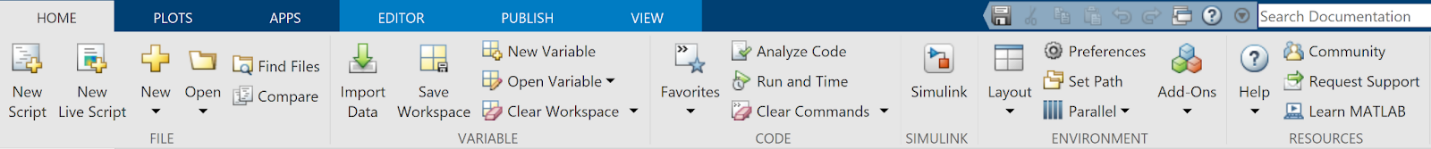
Download HAPPE from the following link: <https://github.com/PINE-Lab/HAPPE>

# Setting Up HAPPE:

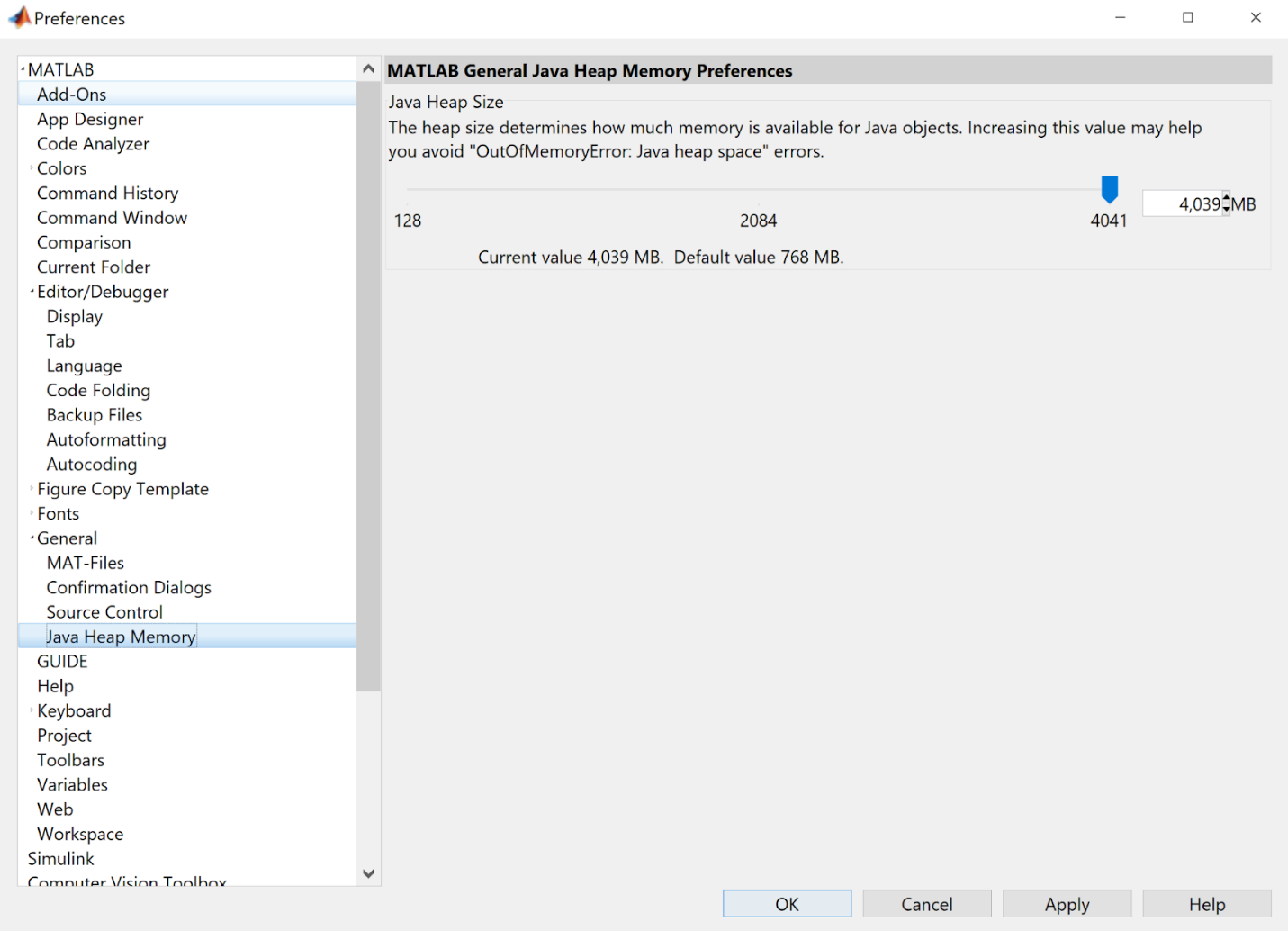
If you do not follow these steps, HAPPE will not run properly.

## Adjust Java Heap Settings

1. In MATLAB under the HOME tab, select “Preferences.”



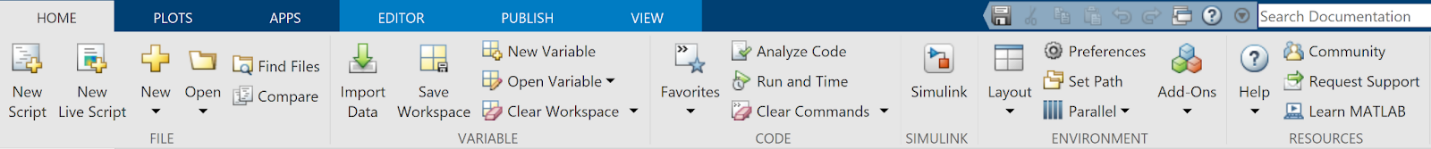
1. In the left-hand sidebar, under General, select “Java Heap Memory.”
2. The default value for the Java Heap is 768 MB. You will want to increase this to the maximum by either dragging the blue arrow all the way to the right or using the text box to the right of the bar.
3. Select “Apply” then “OK.”



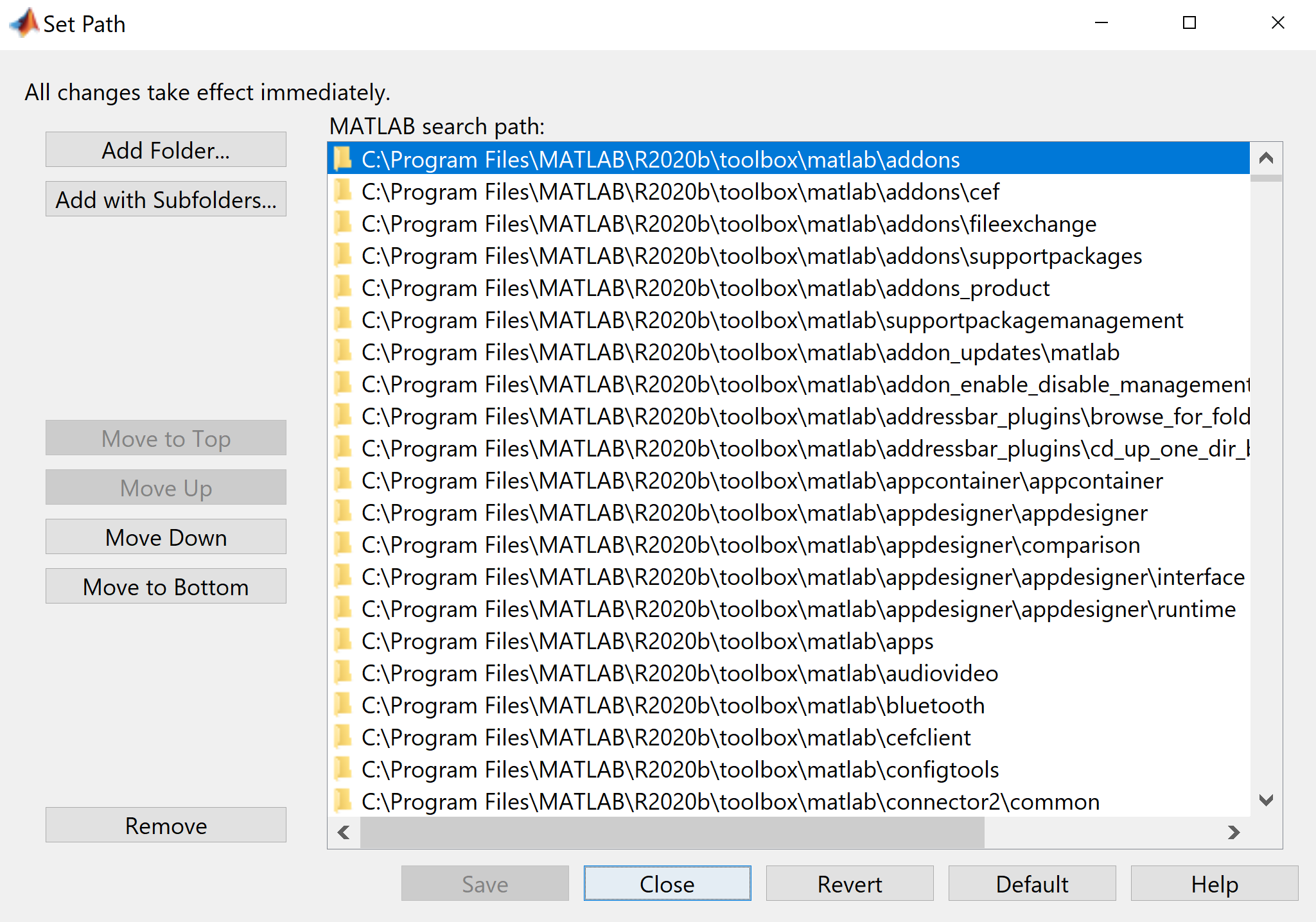
## Set MATLAB Paths

We recommend adding HAPPE and certain subfolders to your path.

1. Under the HOME tab in MATLAB, select “Set Path.”

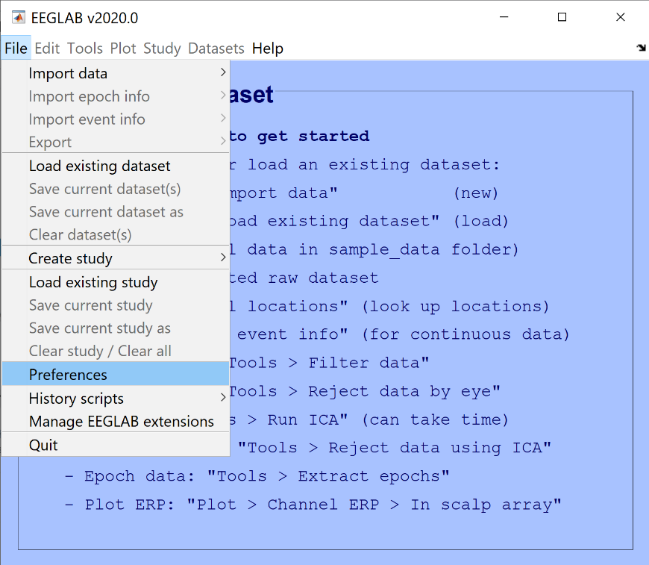


1. In the Set Path window that opens, use the Add Folder button to add the following folders (Please note that if you are using a Mac, \ will be replaced with /):
   * HAPPE
   * HAPPE\scripts
   * HAPPE\scripts\UI\_scripts
   * HAPPE\acquisition\_layout\_information
   * HAPPE\Packages
   * HAPPE\Packages\MARA-master
   * HAPPE\Packages\eeglab2020\_0
2. Click the “Save” button.
3. Click the “Close” button.

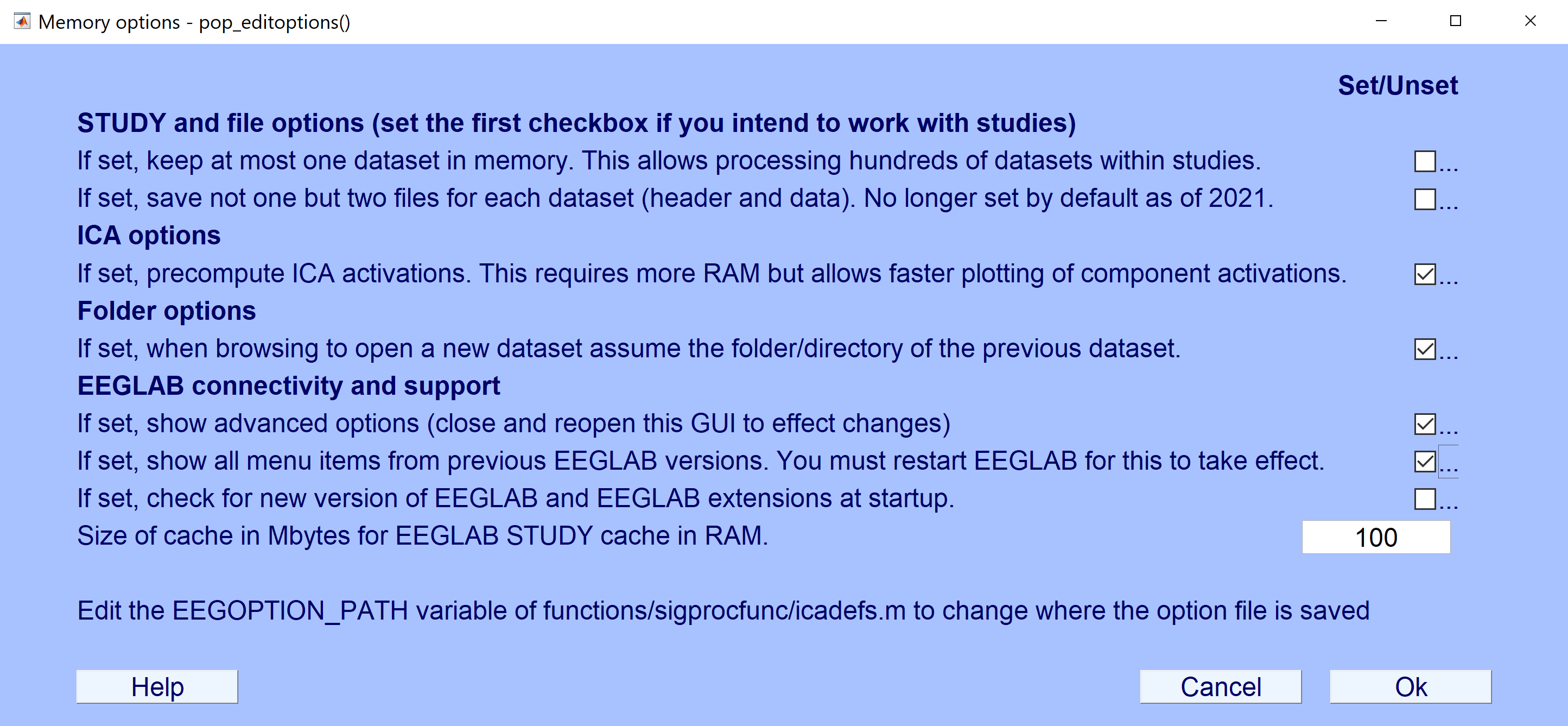


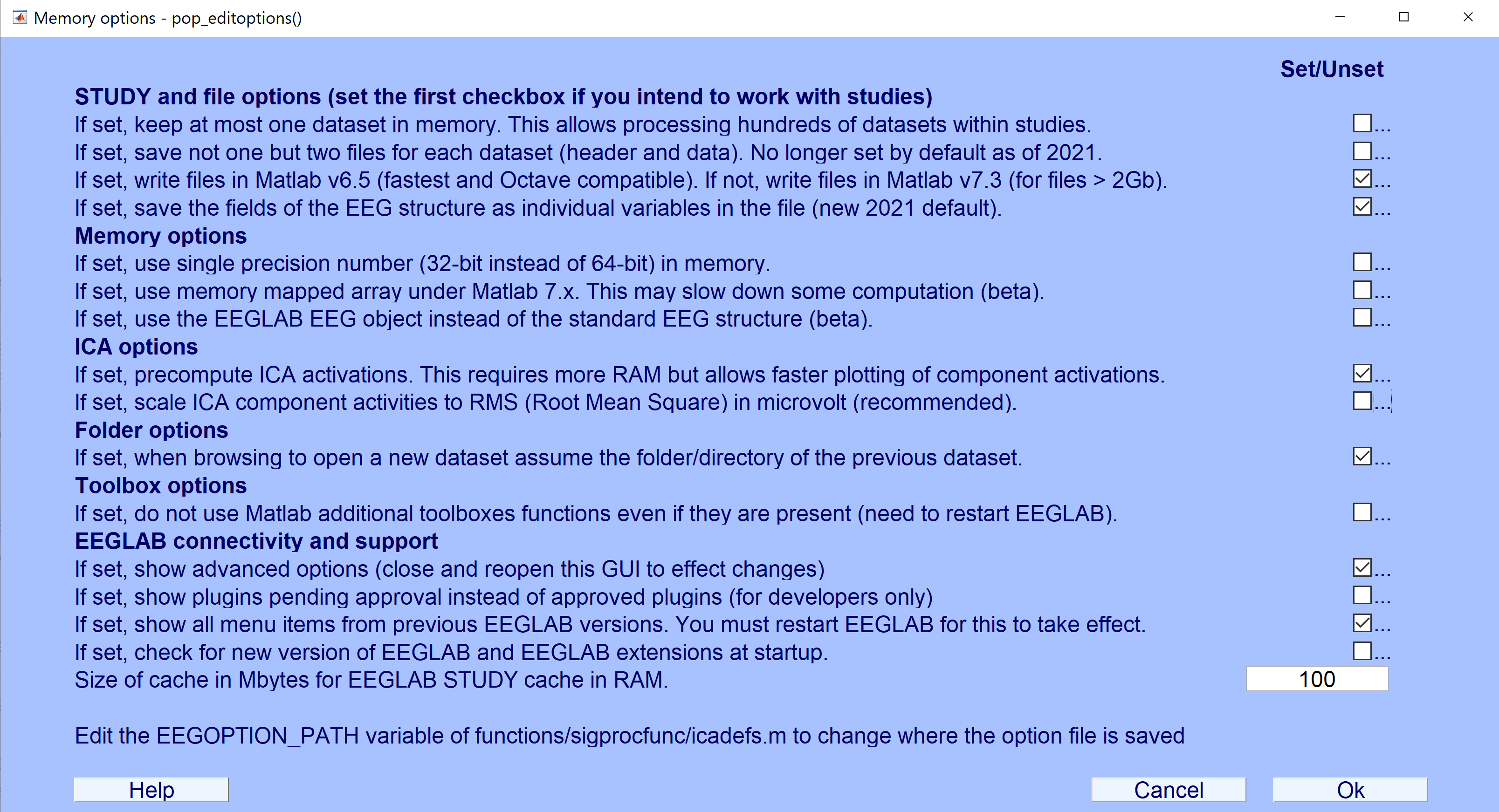
## Configure EEGLAB Preferences

1. Launch EEGLAB by entering eeglab into the command window.
2. In the File menu in the toolbar at the top of the window, select Preferences.

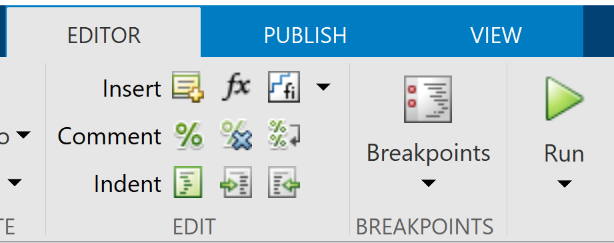


1. Click the checkboxes to enable advanced options and to show all menu items from previous EEGLAB versions. Click OK.



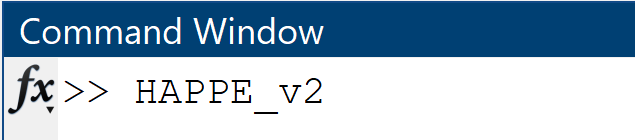
1. Close EEGLAB. Re-Open EEGLAB and the Preferences menu using steps 1 and 2, above.
2. Click the checkboxes so your EEGLAB preferences match those shown below. When done, click OK. 
3. Close EEGLAB.

# How to Run HAPPE v2

1. Navigate to the main HAPPE folder in your file browser.
2. Open HAPPE\_v2.m in MATLAB.
3. In the Editor tab, hit “Run.”

**OR**

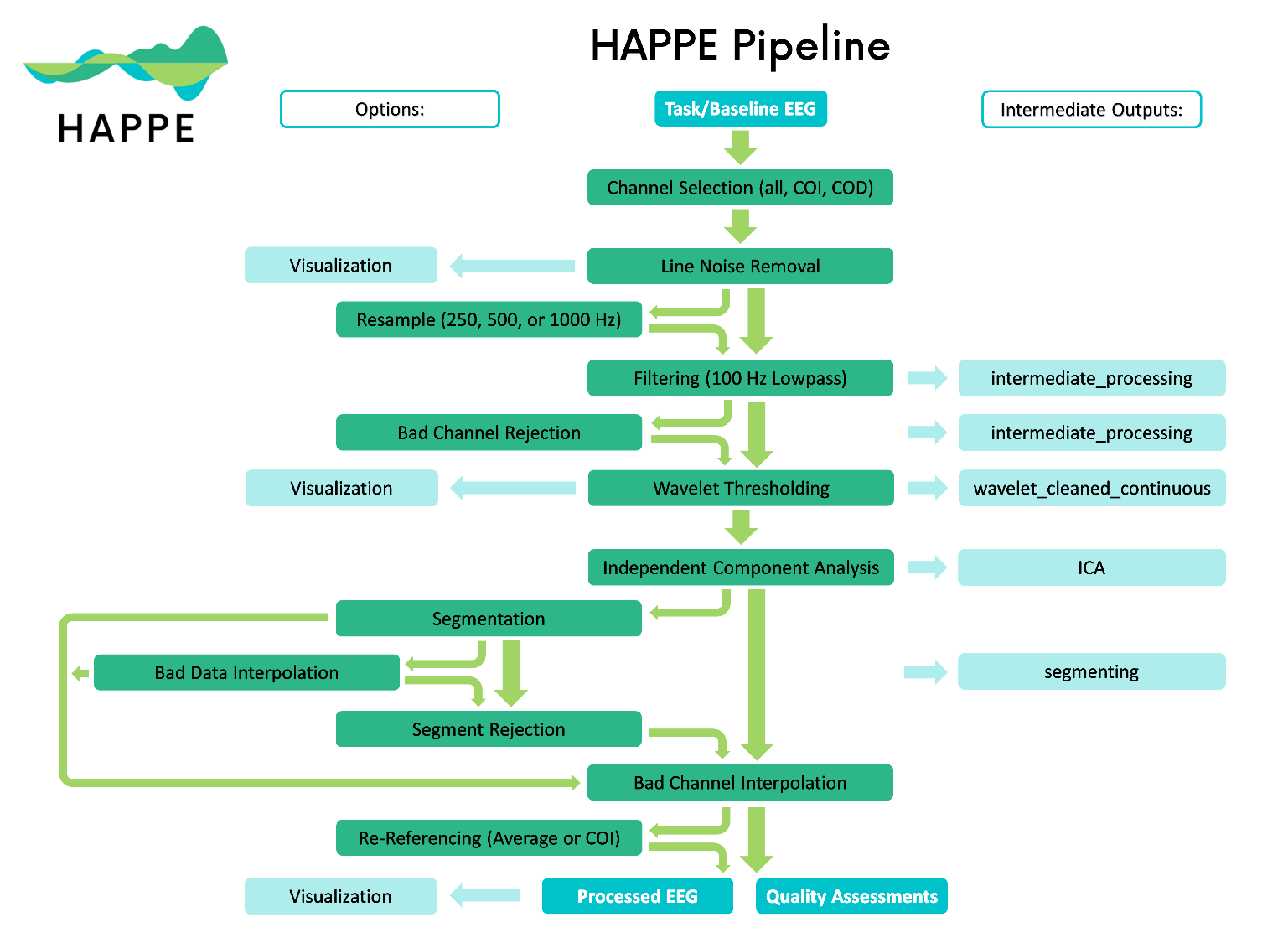
Type HAPPE\_v2 in the Command Window and hit your newline key (enter – Windows; return – Mac)



1. Follow the prompts in the command window of MATLAB
   * For detailed instructions regarding the prompts, see relevant sections below.

# HAPPE Pipeline Guides

## HAPPE Resting State



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post artifact-rejection, input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPE, always enter N (case insensitive).

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input rest for HAPPE resting-state EEG.

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

This is the file format of your raw data. Depending on your choice, you may need to follow a different set of prompts. Note that .raw format is not available for resting-state EEG.

**Example:** 2

Acquisition layout net type:

  1 = EGI Geodesic Sensor Net

  2 = EGI HydroCel Geodesic Sensor Net

  3 = BioSemi

  4 = Brain Products Standard BrainCap (BC)

  5 = Brain Products Wet-Sponge R-Net for actiCHamp Plus (RNP-AC)

  6 = Neuroscan Quik-Cap

  7 = Other

Select the type of cap for your acquisition layout. Different choices will result in different prompts.

**Example:** 1

If you select .mat for your file format, you will get the following prompts:

If you selected an EGI GSN net, an EGI HydroCel GSN net, or a BioSemi net, HAPPE will list the number of electrodes that are compatible with HAPPE for that net.

If you selected a BC net, a RNP-AC net, or a Quik-Cap, you see the following:

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLab with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you do not have channel locations, you will not be able to proceed with HAPPE for high-density data.

Enter the name of the file containing the chanlocs, including the full path and file extension:

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

**Example (Mac):** /Users/laurelg-d/Desktop/chanlocs.sfp

**Example (PC):** C:\Users\laurelg-d\Documents\chanlocs.sfp

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Enter the potential EEG variable names, one at a time.

Press enter/return between each entry.

NOTE: variable names containing "segment" may cause issues.

For .mat files, you need the name of the variable that stores the EEG data. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering potential variable names, enter done.

**Example:** Category\_1

If you select .set for your file format, you will get the following prompts:

Does your file have the 10-20 channels labeled? [Y/N]

HAPPE currently requires the 10-20 channels to be labeled in the dataset to run.

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

If you select .cdt for your file format, you will get the following prompts:

Number of channels:

HAPPE supports 32, 64, and 128 channels.

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

NOTE: 10-20 channels are already included.

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. You do not need to include the names of the 10-20 channels, as they are automatically included. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive).

If you select to perform bad channel detection:

Bad channel detection method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

You can choose to run either the default or the legacy version of bad channel detection. The legacy version was used in the original HAPPE software and is outdated and thus not recommended. The default uses the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a soft Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized on EEG data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment length, in SECONDS:

The desired length for your segments, in seconds.

**Example:** 2

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

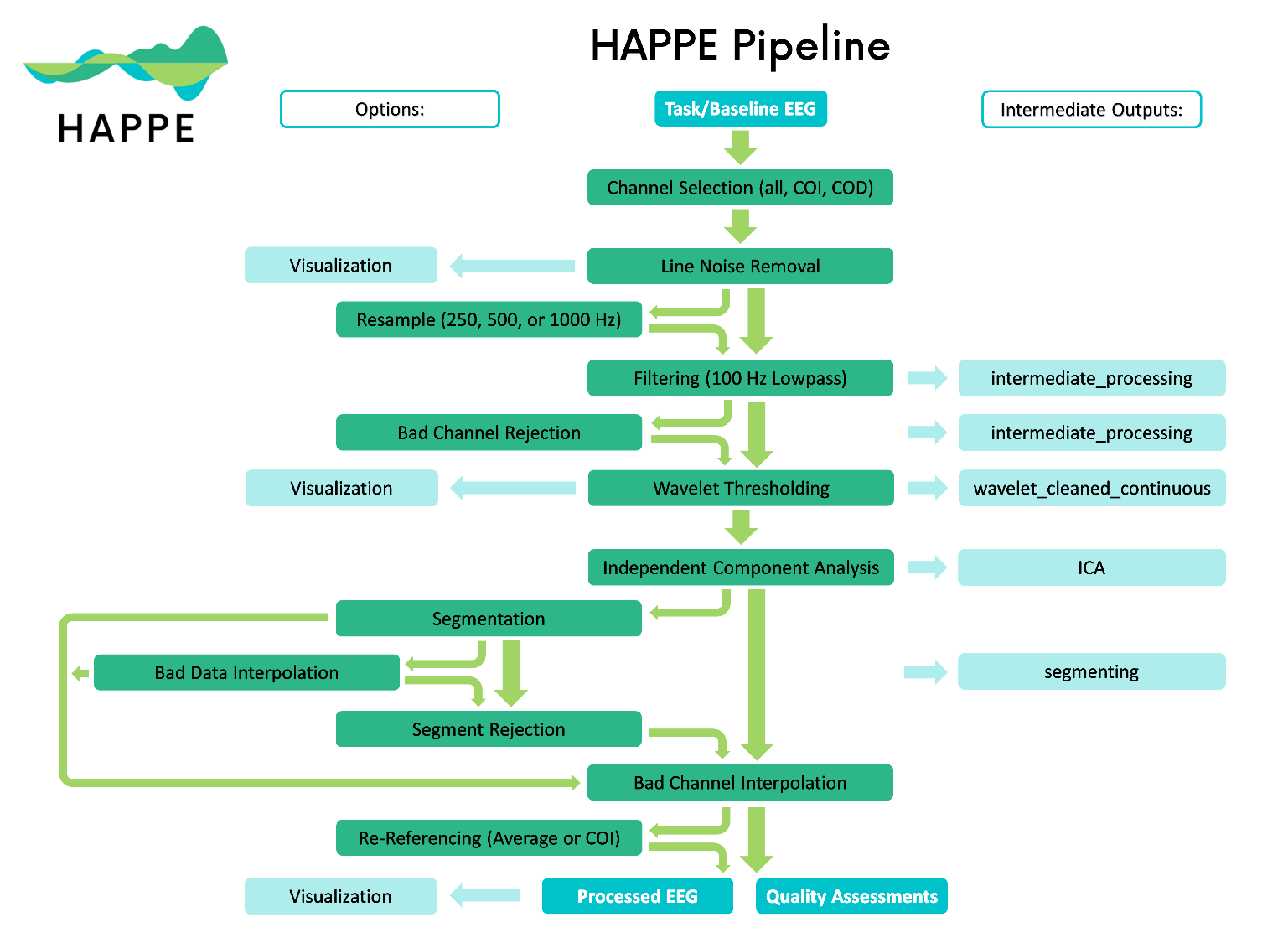
Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

## HAPPE Task-Related



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post artifact-rejection, input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPE, always enter N (case insensitive).

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input task for HAPPE with task-related EEG.

Performing event-related potential (ERP) analysis? [Y/N]

Input N (case insensitive) for HAPPE task-related, non-ERP EEG.

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

This is the file format of your raw data. Depending on your choice, you may need to follow a different set of prompts. Note that .raw format is not available for resting-state EEG.

**Example:** 2

Acquisition layout net type:

  1 = EGI Geodesic Sensor Net

  2 = EGI HydroCel Geodesic Sensor Net

  3 = BioSemi

  4 = Brain Products Standard BrainCap (BC)

  5 = Brain Products Wet-Sponge R-Net for actiCHamp Plus (RNP-AC)

  6 = Neuroscan Quik-Cap

  7 = Other

Select the type of cap for your acquisition layout. Different choices will result in different prompts.

**Example:** 1

If you select .mat for your file format, you will get the following prompts:

If you selected an EGI GSN net, an EGI HydroCel GSN net, or a BioSemi net, HAPPE will list the number of electrodes that are compatible with HAPPE for that net.

If you selected a BC net, a RNP-AC net, or a Quik-Cap, you see the following:

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLAB with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you do not have channel locations, you will not be able to proceed with HAPPE for high-density data.

Enter the name of the file containing the chanlocs, including the full path and file extension:

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

**Example (Mac):** /Users/laurelg-d/Desktop/chanlocs.sfp

**Example (PC):** C:\Users\laurelg-d\Documents\chanlocs.sfp

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Enter the potential EEG variable names, one at a time.

Press enter/return between each entry.

NOTE: variable names containing "segment" may cause issues.

For .mat files, you need the name of the variable that stores the EEG data. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering potential variable names, enter done.

**Example:** Category\_1

Path to .txt files containing task event info:

Enter the path to the .txt files that have your task event info here.

If you select .raw for your file format, you will get the following prompts:

If you selected an EGI GSN net or an EGI HydroCel GSN net, HAPPE will list the number of electrodes that are compatible with HAPPE for that net.

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

If you select .set for your file format, you will get the following prompts:

Does your file have the 10-20 channels labeled? [Y/N]

HAPPE currently requires the 10-20 channels to be labeled in the dataset to run.

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

If you select .cdt for your file format, you will get the following prompts:

Number of channels:

HAPPE supports 32, 64, and 128 channels.

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

NOTE: 10-20 channels are already included.

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. You do not need to include the names of the 10-20 channels, as they are automatically included. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive).

If you select to perform bad channel detection:

Bad channel detection method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

You can choose to run either the default or the legacy version of bad channel detection. The legacy version was used in the original HAPPE software and is outdated and thus not recommended. The default uses the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a soft Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized on EEG data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment start, in MILLISECONDS, relative to stimulus onset:

Example: -500

The starting latency for your segments, relative to the stimulus onset. Including a baseline will result in a negative latency, whereas starting at the stimulus onset would be 0.

Segment end, in MILLISECONDS, relative to stimulus onset:

The ending latency for your segments, relative to the stimulus onset. Stimulus onset is 0, so this number should be greater than 0.

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

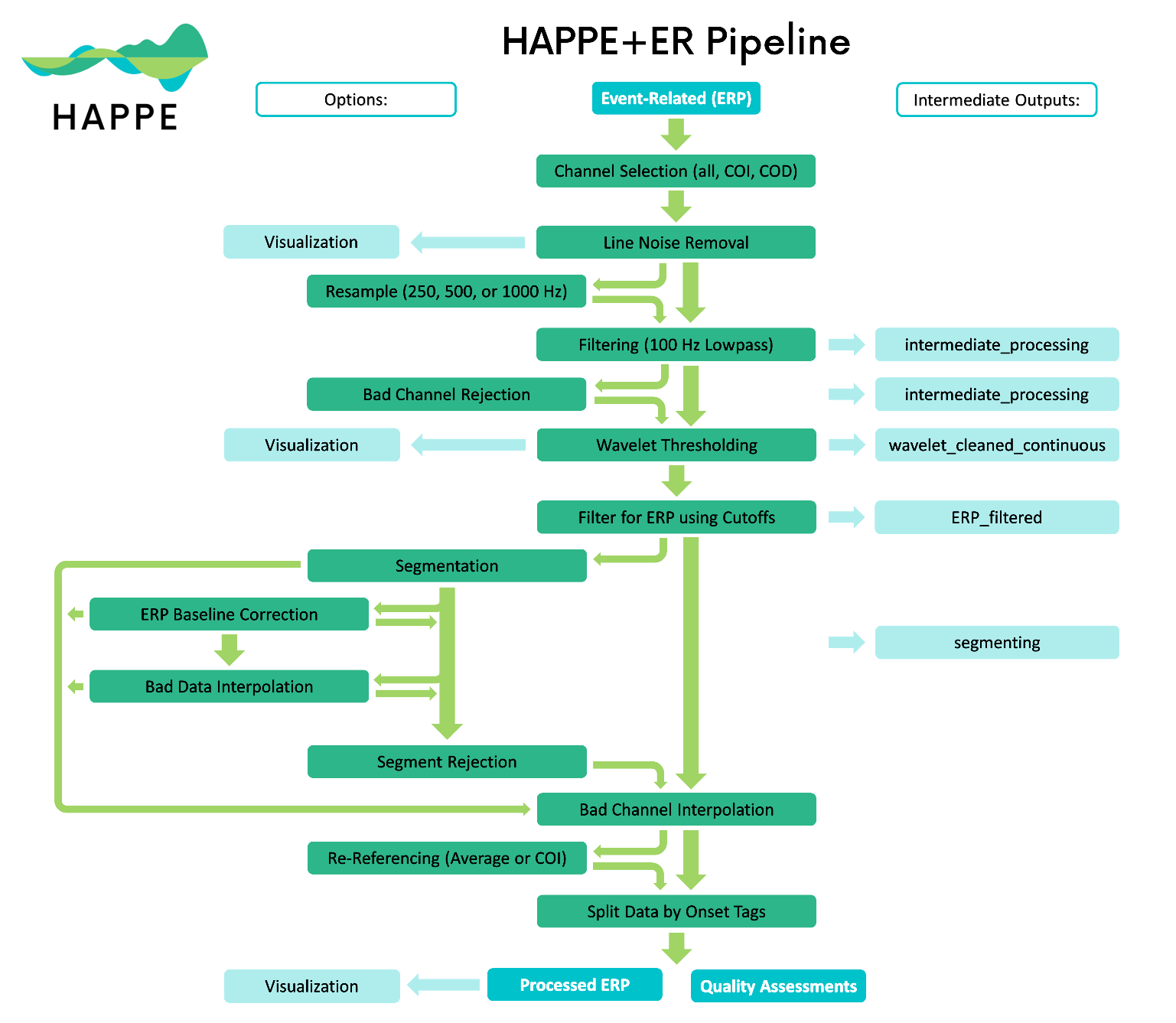
Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

## HAPPE ERPs (HAPPE+ER)



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post artifact-rejection, input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPE+ER, always enter N (case insensitive).

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input task for HAPPE+ER.

Performing event-related potential (ERP) analysis? [Y/N]

Input Y (case insensitive) for HAPPE+ER.

Enter the task onset tags, one at a time, pressing enter/return between each entry.

When you have entered all tags, input "done" (without quotations).

These are the tags present in your dataset that are used to indicate the task onset. Any tags that EEGLAB can read can be used.

**Example:** vep+

Enter low-pass filter, in Hz:

Common low-pass filter is 30 - 45 Hz

Enter the low-pass filter you want to use on your ERP data. Suggestions are provided, but determine what is best for your particular dataset.

**Example:** 35

Enter high-pass filter, in Hz:

Common high-pass filter is 0.1 - 0.3 Hz

Enter the high-pass filter you want to use on your ERP data. Suggestions are provided, but determine what is best for your particular dataset.

**Example:** 0.3

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

This is the file format of your raw data. Depending on your choice, you may need to follow a different set of prompts. Note that .raw format is not available for resting-state EEG.

**Example:** 2

Acquisition layout net type:

  1 = EGI Geodesic Sensor Net

  2 = EGI HydroCel Geodesic Sensor Net

  3 = BioSemi

  4 = Brain Products Standard BrainCap (BC)

  5 = Brain Products Wet-Sponge R-Net for actiCHamp Plus (RNP-AC)

  6 = Neuroscan Quik-Cap

  7 = Other

Select the type of cap for your acquisition layout. Different choices will result in different prompts.

**Example:** 1

If you select .mat for your file format, you will get the following prompts:

If you selected an EGI GSN net, an EGI HydroCel GSN net, or a BioSemi net, HAPPE will list the number of electrodes that are compatible with HAPPE for that net.

If you selected a BC net, a RNP-AC net, or a Quik-Cap, you see the following:

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLab with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you do not have channel locations, you will not be able to proceed with HAPPE for high-density data.

Enter the name of the file containing the chanlocs, including the full path and file extension:

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

**Example (Mac):** /Users/laurelg-d/Desktop/chanlocs.sfp

**Example (PC):** C:\Users\laurelg-d\Documents\chanlocs.sfp

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Enter the potential EEG variable names, one at a time.

Press enter/return between each entry.

NOTE: variable names containing "segment" may cause issues.

For .mat files, you need the name of the variable that stores the EEG data. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering potential variable names, enter done.

**Example:** Category\_1

Path to .txt files containing task event info:

Enter the path to the .txt files that have your task event info here.

If you select .raw for your file format, you will get the following prompts:

If you selected an EGI GSN net or an EGI HydroCel GSN net, HAPPE will list the number of electrodes that are compatible with HAPPE for that net.

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

If you select .set for your file format, you will get the following prompts:

Does your file have the 10-20 channels labeled? [Y/N]

HAPPE currently requires the 10-20 channels to be labeled in the dataset to run.

Number of channels:

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

If you select .cdt for your file format, you will get the following prompts:

Number of channels:

HAPPE supports 32, 64, and 128 channels.

The number of channels in the net used to collect the data. Make sure to use that number, even if your data does not use or include all possible channels in the net.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

NOTE: 10-20 channels are already included.

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. You do not need to include the names of the 10-20 channels, as they are automatically included. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive).

If you select to perform bad channel detection:

Bad channel detection method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

You can choose to run either the default or the legacy version of bad channel detection. The legacy version was used in the original HAPPE software, and is outdated and thus not recommended. The default uses the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Start time, in MILLISECONDS, for the ERP timeseries figure:

The starting latency for a figure of your ERP time series, including all of the channels of interest.

End time, in MILLISECONDS, for the ERP timeseries figure:

NOTE: This should end 1 millisecond before your segmentation parameter ends. (e.g. 299 for 300)

The ending latency for a figure of your ERP time series, including all of your channels of interest. This value must be at least one millisecond before the final latency value present in your dataset.

Enter the latencies, one at a time, to generate spatial topoplots for:

When you have entered all latencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the ERP time series for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done. Note: Ensure that the parameters you input are in milliseconds.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a soft Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized for ERP data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized for ERP EEG, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment start, in MILLISECONDS, relative to stimulus onset:

Example: -500

The starting latency for your segments, relative to the stimulus onset. Including a baseline will result in a negative latency, whereas starting at the stimulus onset would be 0.

Segment end, in MILLISECONDS, relative to stimulus onset:

The ending latency for your segments, relative to the stimulus onset. Stimulus onset is 0, so this number should be greater than 0.

Offset delay, in MILLISECONDS, between stimulus initiation and presentation:

NOTE: Please enter the total offset (combined system and task-specific offsets).

This is dependent on your system and your paradigm.

**Example:** 19

Perform baseline correction (by subtraction)? [Y/N]

If you want to perform baseline correction on your data, select Y, otherwise, choose N.

If you select baseline correction, the following prompts appear:

Enter, in MILLISECONDS, where the baseline segment begins:

Example: -100

This is the start of your baseline segment. It should be a negative number and measured relative to the task onset of 0. -100, for example, would be 100 milliseconds before the stimulus onset.

Enter, in MILLISECONDS, where the baseline segment ends:

NOTE: 0 indicates stimulus onset.

This is the end of your baseline segment. This should be a negative number or 0.

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data. We recommend .txt for HAPPE+ER outputs.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

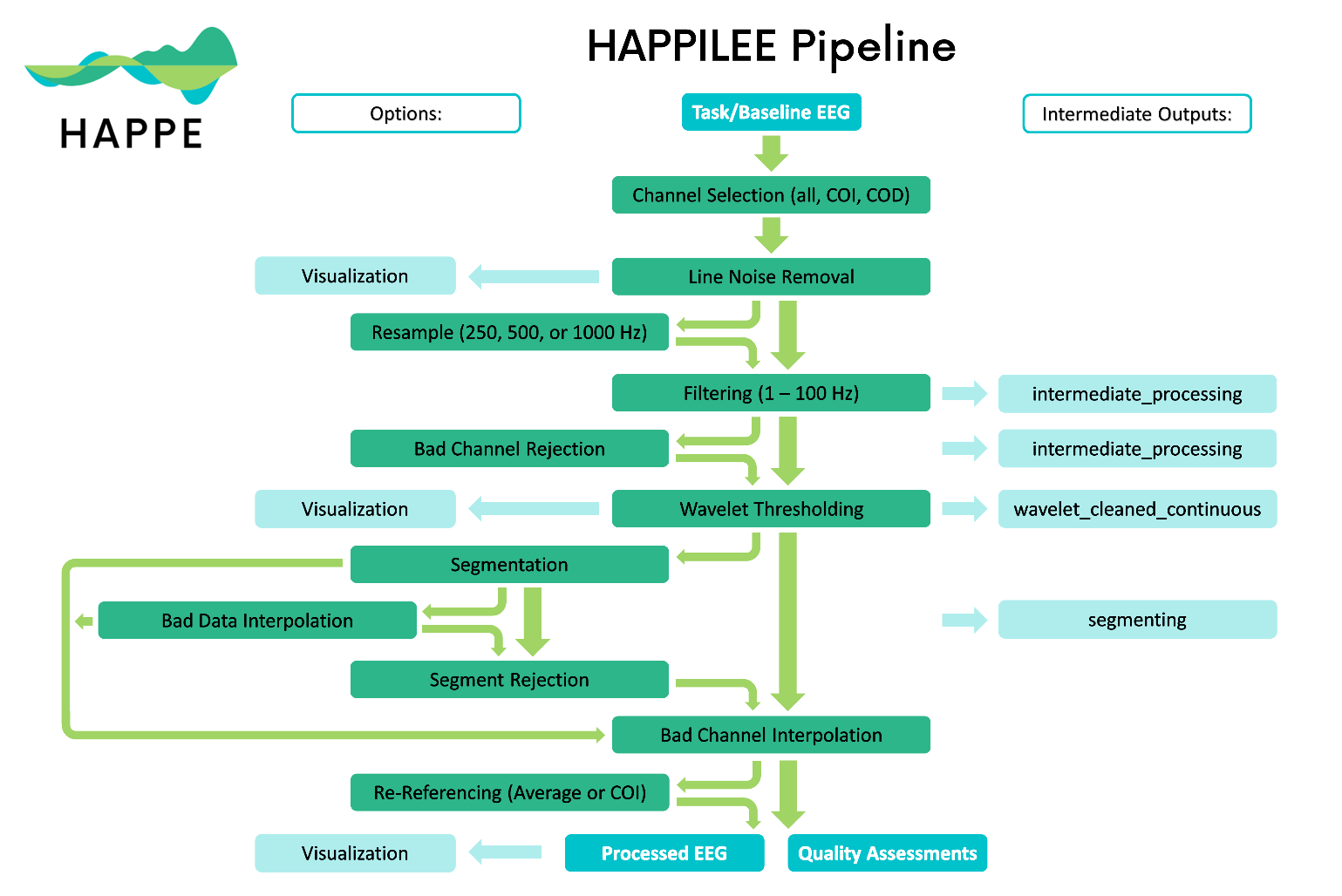
Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

## HAPPILEE Resting-State



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post-waveleting (ICA is not used for low density), input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPILEE, always enter Y (case insensitive).

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input rest for HAPPILEE with resting-state EEG.

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

HAPPILEE currently only supports low density data in .mat and .set formats, so you must input 0 or 2. Depending on your choice, you may need to follow a different set of prompts.

**Example:** 2

If you select .mat for your file format, you will get the following prompts:

Do all your files share the same sampling rate? [Y/N]

If all your files have the same sampling rate, input Y (case insensitive). Otherwise, input N (case insensitive). Depending on your input, you will have to follow different prompts (see below).

If all your files share the same sampling rate:

Sampling rate:

Enter the sampling rate.

**Example:** 250

If your files have different sampling rates:

Enter the name of the file containing the sampling rates for each file, including the path and file extension.

See the HAPPE user guide for how this file should be formatted.

An example for formatting this file is below. The first column should contain the file name while the second column should contain the sampling rate in Hz.

|  |  |
| --- | --- |
| samplefilename1.mat | 250 |
| samplefilename2.mat | 500 |

Include the full path to where the file exists, followed by a slash (forward or back depending on your OS), then the name of the file, including the file extension (e.g., .csv).

**Example (Mac):** /Users/laurelg-d/Desktop/samplingRates.csv

**Example (PC):** C:\Users\laurelg-d\Documents\samplingRates.csv

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLab with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you don’t have a channel locations file, you will still be able to continue running your data. However, by not providing channel locations, you will not be able to filter to channels of interest, perform bad channel detection, interpolate bad channels, or re-reference your data. For this option, enter N (case insensitive). Otherwise, input Y (case insensitive) and answer additional prompts (see below).

If you do have channel locations, you must enter answers to the following prompts:

Enter the name of the file containing the chanlocs, including the full path and file extension.

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of

disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive). HAPPILEE will default to using the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized for low density data.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPILEE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a hard Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized on low density data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized on low density data, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment length, in SECONDS:

The desired length for your segments, in seconds.

**Example:** 2

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

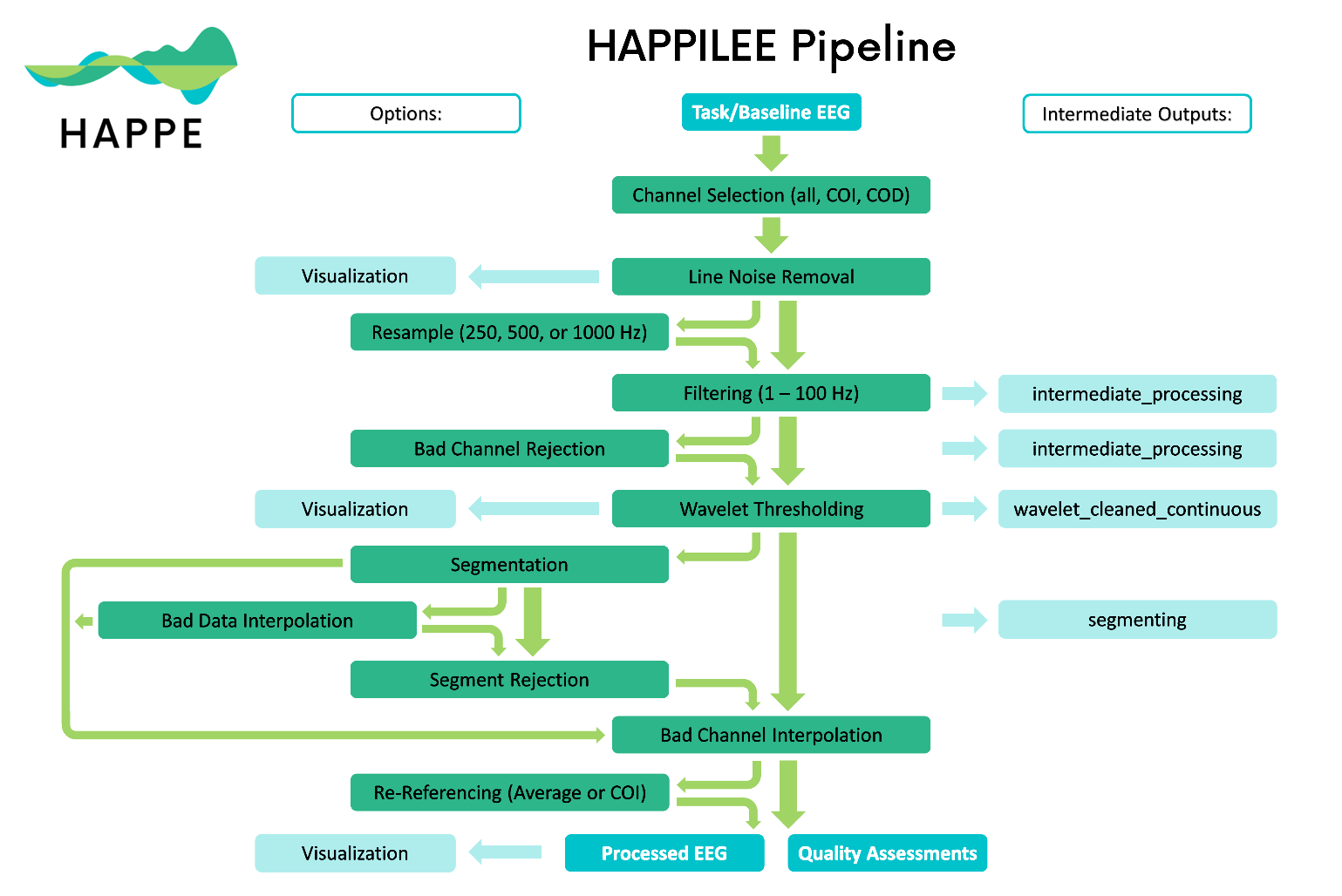
Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

## HAPPILEE Task-Related



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post-waveleting (ICA is not used for low density), input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPILEE, always enter Y (case insensitive).

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input task for HAPPILEE with task-related EEG.

Performing event-related potential (ERP) analysis? [Y/N]

Input N (case insensitive) for HAPPILEE task-related, non-ERP EEG.

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

HAPPILEE currently only supports low density data in .mat and .set formats, so you must input 0 or 2. Depending on your choice, you may need to follow a different set of prompts.

**Example:** 2

If you select .mat for your file format, you will get the following prompts:

Do all your files share the same sampling rate? [Y/N]

If all your files have the same sampling rate, input Y (case insensitive). Otherwise, input N (case insensitive). Depending on your input, you will have to follow different prompts (see below).

If all your files share the same sampling rate:

Sampling rate:

Enter the sampling rate.

**Example:** 250

If your files have different sampling rates:

Enter the name of the file containing the sampling rates for each file, including the path and file extension.

See the HAPPE user guide for how this file should be formatted.

|  |  |
| --- | --- |
| samplefilename1.mat | 250 |
| samplefilename2.mat | 500 |

An example for formatting this file is below. The first column should contain the file name while the second column should contain the sampling rate in Hz.

Include the full path to where the file exists, followed by a slash (forward or back depending on your OS), then the name of the file, including the file extension (e.g., .csv).

**Example (Mac):** /Users/laurelg-d/Desktop/samplingRates.csv

**Example (PC):** C:\Users\laurelg-d\Documents\samplingRates.csv

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLab with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you don’t have a channel locations file, you will still be able to continue running your data. However, by not providing channel locations, you will not be able to filter to channels of interest, perform bad channel detection, interpolate bad channels, or re-reference your data. For this option, enter N (case insensitive). Otherwise, input Y (case insensitive) and answer additional prompts (see below).

If you do have channel locations, you must enter answers to the following prompts:

Enter the name of the file containing the chanlocs, including the full path and file extension.

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

Path to .txt files containing task event info:

Enter the path to the .txt files that have your task event info here.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of

disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive). HAPPILEE will default to using the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized for low density data.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPILEE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a hard Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized on low density data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized on low density data, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment start, in MILLISECONDS, relative to stimulus onset:

Example: -500

The starting latency for your segments, relative to the stimulus onset. Including a baseline will result in a negative latency, whereas starting at the stimulus onset would be 0.

Segment end, in MILLISECONDS, relative to stimulus onset:

The ending latency for your segments, relative to the stimulus onset. Stimulus onset is 0, so this number should be greater than 0.

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

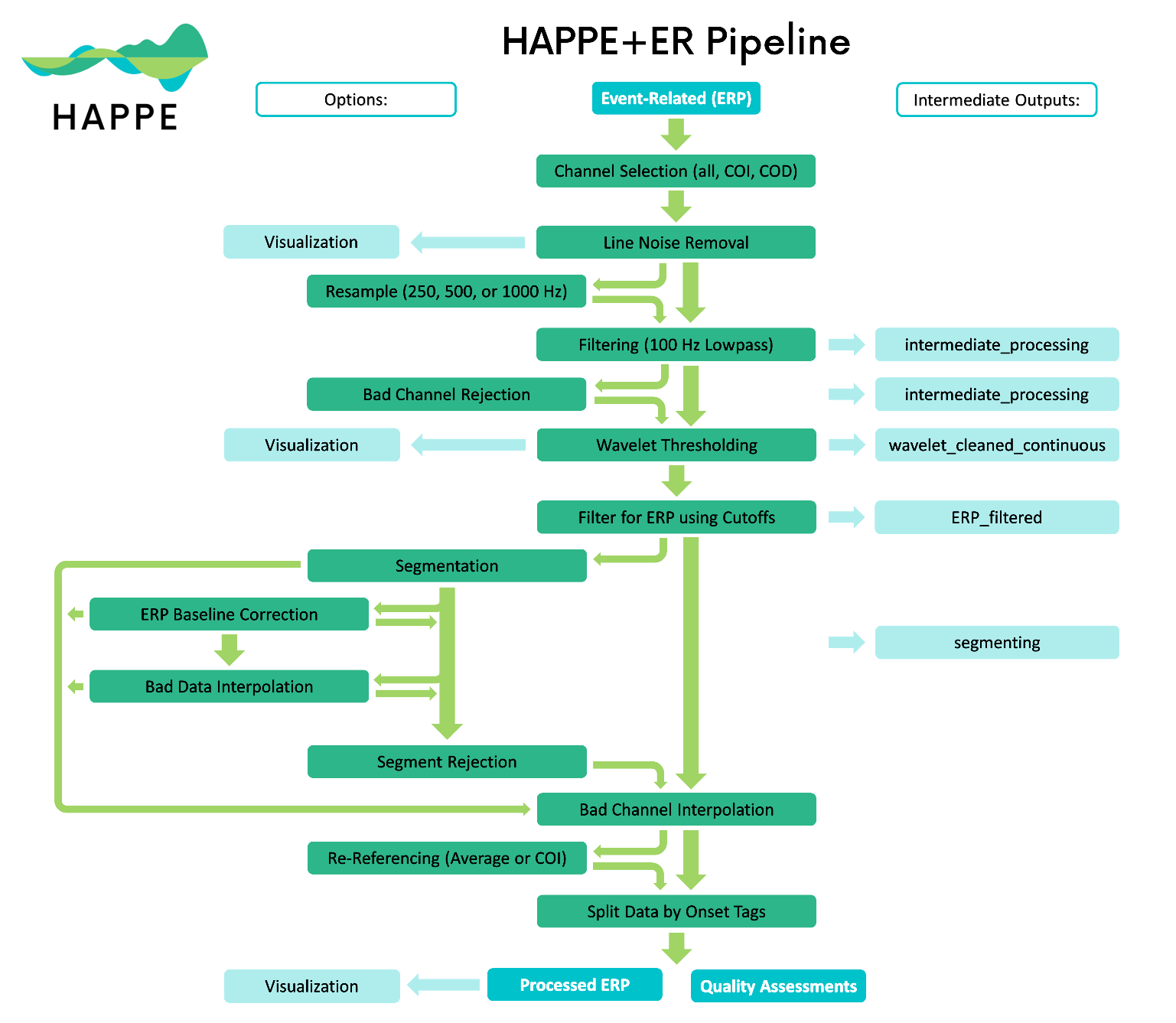
Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

## HAPPILEE ERPs



### Following Command Line Prompts

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Enter the path to the folder containing the dataset(s):

The first step asks you to input the folder where the raw data is located. If running HAPPE for the first time, the folder with the data should ONLY include the raw files you wish to run through the pipeline; no other folders or files should live inside that folder. If reprocessing data, all outputs (folders and documents) should be in the folder in the same file structure as they were created during the original HAPPE run in addition to the raw data.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder

Select an option:

  raw = Run on raw data from the start

  reprocess = Run on HAPPE-processed data starting post-waveleting/ICA

If this is your first time processing the raw data file(s), input raw. If you wish to re-run raw data that has previously been processed, starting with post-waveleting (ICA is not used for low density), input reprocess. Depending on your choice, you may be asked to follow additional prompts (see below).

If you selected to reprocess existing data:

Name of previously created dataQC .csv file:

If no file exists, enter "none" (without quotations).

If you have the dataQC .csv file from the previous run, input the name of the file here. Confirm that it is located in the “quality\_assessment\_outputs” folder before running HAPPE. Otherwise, input noneand HAPPE will still be able to reprocess your data, but some quality control outputs may be missing.

Files, such as processed data and quality metrics, may already exist for this dataset.

  overwrite = Overwrite existing files

  new = Save new files

If you do not want to keep your previous files for this dataset, input overwrite. To keep both the previous files and the current files, input new**.**

Ifyou selected to save new files:

Use default or custom suffix for processed set?

  default = Default name (\_rerun\_dd-mm-yyyy.mat).

  custom = Create your own file name.

To create your own suffix for the processed data, input custom**.** You will then be prompted to enter your custom suffix. We recommend starting your custom suffix with an underscore. Otherwise, input default, which will save your files with a suffix in the format shown above.

Load pre-existing set of input parameters? [Y/N]

If parameters have previously been set through HAPPE for this dataset or another dataset with parameters that support the current dataset, you can load these parameters, if desired. Depending on your choice, you will be asked to follow a different set of prompts (see below).

If you chose to load pre-existing parameters:

Path to the folder containing the input parameters:

The default folder name will be **input\_parameters**. NOTE: the input parameters folder does not have to be in the same path as the datasets you are currently running.

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/input\_parameters

**Example (PC):** C:\Users\laurelg-d\Documents\Data Folder\input\_parameters

Name of file containing pre-existing parameters:

The default file name is **inputParameters\_DD-MM-YYYY.mat**. If you chose a custom file name for the pre-existing parameters, input this name instead.

You will then be presented with a list of your current inputs.

Change an existing parameter? [Y/N]

To change any of the inputs saved in this file, input Y (case insensitive).

If you are not loading a pre-existing set of input parameters:

Low-density data? [Y/N]

For HAPPE, low-density data contains 30 channels or less.

For HAPPILEE & HAPPE+ER, always enter Y (case insensitive)**.**

Enter data type:

  rest = Resting-State EEG

  task = Task-Related EEG

Input task for HAPPILEE & HAPPE+ER.

Performing event-related potential (ERP) analysis? [Y/N]

Input Y (case insensitive) for HAPPILEE & HAPPE+ER.

Enter the task onset tags, one at a time, pressing enter/return between each entry.

When you have entered all tags, input "done" (without quotations).

These are the tags present in your dataset that are used to indicate the task onset. Any tags that EEGLAB can read can be used.

**Example:** vep+

Enter low-pass filter, in Hz:

Common low-pass filter is 30 - 45 Hz

Enter the low-pass filter you want to use on your ERP data. Suggestions are provided, but determine what is best for your particular dataset.

**Example:** 35

Enter high-pass filter, in Hz:

Common high-pass filter is 0.1 - 0.3 Hz

Enter the high-pass filter you want to use on your ERP data. Suggestions are provided, but determine what is best for your particular dataset.

**Example:** 0.3

File Format:

  0 = .mat (MATLAB array)

  1 = .raw (Netstation simple binary)

  2 = .set (EEGLAB format)

  3 = .cdt (Neuroscan)

  4 = .mff (EGI)

HAPPILEE currently only supports low density data in .mat and .set formats, so you must input 0 or 2. Depending on your choice, you may need to follow a different set of prompts.

**Example:** 2

If you select .mat for your file format, you will get the following prompts:

Do all your files share the same sampling rate? [Y/N]

If all your files have the same sampling rate, input Y (case insensitive). Otherwise, input N (case insensitive). Depending on your input, you will have to follow different prompts (see below).

If all your files share the same sampling rate:

Sampling rate:

Enter the sampling rate.

**Example:** 250

If your files have different sampling rates:

Enter the name of the file containing the sampling rates for each file, including the path and file extension.

See the HAPPE user guide for how this file should be formatted.

An example for formatting this file is below. The first column should contain the file name while the second column should contain the sampling rate in Hz.

|  |  |
| --- | --- |
| samplefilename1.mat | 250 |
| samplefilename2.mat | 500 |

Include the full path to where the file exists, followed by a slash (forward or back depending on your OS), then the name of the file, including the file extension (e.g., .csv).

**Example (Mac):** /Users/laurelg-d/Desktop/samplingRates.csv

**Example (PC):** C:\Users\laurelg-d\Documents\samplingRates.csv

Do you have a channel locations file for your data? [Y/N]

NOTE: A list of supported files can be found in the HAPPE user guide.

Accepted EEGLAB channel location file formats (from EEGLAB documentation) include:

1. .loc, .locs, .eloc - EEG polar coordinates
2. .ced - EEGLab with polar, cartesian, and spherical
3. .sph - MATLAB spherical coordinates
4. .elc - Cartesian 3-D from EETrack
5. .elp - Polhemus Cartesian coordinates
6. .elp - BESA spherical coordinates
7. .xyz - MATLAB/EEGLab Cartesian coordinates
8. .asc, .dat - Neuroscan Cartesian polar coordinates
9. .mat - Brainstrom channel locations
10. .sfp - BESA/EGI xyz Cartesian coordinates

If you don’t have a channel locations file, you will still be able to continue running your data. However, by not providing channel locations, you will not be able to filter to channels of interest, perform bad channel detection, interpolate bad channels, or re-reference your data. For this option, enter N (case insensitive). Otherwise, input Y (case insensitive) and answer additional prompts (see below).

If you do have channel locations, you must enter answers to the following prompts:

Enter the name of the file containing the chanlocs, including the full path and file extension.

Enter the name of the file with the chanlocs in a file format that is listed above as supported by HAPPE.

Path to .txt files containing task event info:

Enter the path to the .txt files that have your task event info here.

Examine all channels (all) or only channels of interest (coi)?

To process all possible channels within each EEG file, input all**.** If you wish to only process a subset of channels in each data file, input coi.

If you selected to only examine channels of interest, you will get the following prompts:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of

disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in processing. If you select exclude, the channels you enter will be the only ones not included in processing.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include/exclude one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Perform bad channel detection? [Y/N]

If you wish to have channels that have high impedances, damage to the electrodes, insufficient scalp contact, and excessive movement or electromyographic (EMG) artifact throughout the recording removed, input Y (case insensitive). HAPPILEE will default to using the new method of bad channel detection with the Clean Rawdata function with preset criterion that has been optimized for low density data.

Frequency of electrical (line) noise in Hz:

USA data probably = 60; Otherwise, probably = 50

This input is necessary to help accurately detect line noise in the data, so ensure that you have chosen the correct frequency based on where the data was collected.

Line noise reduction method:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

Legacy line noise reduction method uses the CleanLine plugin for EEGLAB from a far less effective version release. For this reason, we recommend inputting default to process your data with the new line noise reduction method in HAPPILEE that uses an improved version of CleanLine to reduce line noise.

Run HAPPE with visualizations? [Y/N]

By choosing Y(case insensitive) you will run HAPPE in the semi-automated setting, with several visualizations for every file.

If you have selected to run with visualizations, the following prompts will appear:

Minimum value for power spectrum figure:

The minimum value for the plot of the power spectrum.

 Maximum value for power spectrum figure:

The maximum value for the plot of the power spectrum.

Enter the frequencies, one at a time, to generate spatial topoplots for:

When you have entered all frequencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the power spectrum for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Start time, in MILLISECONDS, for the ERP timeseries figure:

The starting latency for a figure of your ERP time series, including all of the channels of interest.

End time, in MILLISECONDS, for the ERP timeseries figure:

NOTE: This should end 1 millisecond before your segmentation parameter ends. (e.g. 299 for 300)

The ending latency for a figure of your ERP time series, including all of your channels of interest. This value must be at least one millisecond before the final latency value present in your dataset.

Enter the latencies, one at a time, to generate spatial topoplots for:

When you have entered all latencies, input 'done' (without quotations).

This input asks you to select any particular frequencies you would like to see topoplots for (spatial distribution of power in that frequency across the scalp) in the same image as the ERP time series for that file. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done. Note: Ensure that the parameters you input are in milliseconds.

Resample data? [Y/N]

NOTE: Resampling is recommended for files <= 60 seconds long.

If you wish to resample your data, input Y, otherwise input N. Either option is case-insensitive.

If you chose to resample your data, the following prompt will appear:

HAPPE supports resampling to 250, 500, and 1000.

Resample frequency:

Choose one of the three possible frequencies above to resample your data to.

Method of wavelet thresholding:

  default = Default method optimized in HAPPE v2.

  legacy = Method from HAPPE v1 (NOT RECOMMENDED).

The new wavelet method uses a soft Empirical Baysian level-dependent threshold for the wavelets. Wavelet family is coiflet (level 4). This method has been optimized on ERP data. The legacy method of waveletting uses a soft, global threshold for the wavelets. The wavelet family is coiflet (level 5). Threshold multiplier is used to remove more high frequency noise. This method has not been optimized on low density data, so it is not recommended.

Segment data? [Y/N]

To segment your data, input Y. Otherwise, input N. If you choose to segment your data for your ERP analysis, you will be asked to answer additional prompts.

If you have selected to segment your data, the following prompts will appear:

Segment start, in MILLISECONDS, relative to stimulus onset:

Example: -500

The starting latency for your segments, relative to the stimulus onset. Including a baseline will result in a negative latency, whereas starting at the stimulus onset would be 0.

Segment end, in MILLISECONDS, relative to stimulus onset:

The ending latency for your segments, relative to the stimulus onset. Stimulus onset is 0, so this number should be greater than 0.

Offset delay, in MILLISECONDS, between stimulus initiation and presentation:

NOTE: Please enter the total offset (combined system and task-specific offsets).

This is dependent on your system and your paradigm.

**Example:** 19

Perform baseline correction (by subtraction)? [Y/N]

If you want to perform baseline correction on your data, select Y, otherwise, choose N.

If you select baseline correction, the following prompts appear:

Enter, in MILLISECONDS, where the baseline segment begins:

Example: -100

This is the start of your baseline segment. It should be a negative number and measured relative to the task onset of 0. -100, for example, would be 100 milliseconds before the stimulus onset.

Enter, in MILLISECONDS, where the baseline segment ends:

NOTE: 0 indicates stimulus onset.

This is the end of your baseline segment. This should be a negative number or 0.

Interpolate the specific channels' data determined to be artifact/bad within each segment? [Y/N]

This option allows you to evaluate within each epoch whether any channels have bad data for that segment by using only the channels that have been marked “good” channels overall from the channel rejection step. Channels flagged with bad data for that segment will then have their data interpolated only for that segment.

Perform segment rejection? [Y/N]

Instead of interpolating data within segments, users can instead select to reject segments that are determined to still be artifact-contaminated. Criteria for rejection include a choice of joint-probability criteria, amplitude-based criteria, or a combined joint-probability criteria with amplitude-based criteria.

If you select to perform segment rejection, the following prompts appear:

Choose a method of segment rejection:

  amplitude = Amplitude criteria only

  similarity = Segment similarity only

  both = Both amplitude criteria and segment similarity

The first option is using amplitude criteria only. Amplitude-based criteria sets a minimum and maximum signal amplitude as the artifact threshold, with segments being removed when their amplitude falls on either side of this threshold. After inputting amplitude,users must set the amplitude to be used for determining artifact-segments. The second option is using segment similarity only. Segment similarity criteria considers how likely a segment’s activity is to be artifact-laden given the activity of other segments for that channel, and also other channels’ activity for the same segment. Outlier segments will be removed. The assumption is that artifact segments should be the rare segments relative to the rest of the data at this point in the processing stream. The third option includes both methods. A combined approach with segment similarity criteria and amplitude-based criteria removes outlier segments based on both standard deviations and a minimum and maximum signal amplitude set by the user. If you input both,you will be prompted to input the minimum and maximum signal amplitude to use as the artifact threshold.

If you select amplitude or both for segment rejection criteria, the following prompts appear:

Minimum signal amplitude to use as the artifact threshold:

This is the minimum signal amplitude used for segment rejection.

**Example:** -200

Maximum signal amplitude to use as the artifact threshold:

This is the maximum signal amplitude used for segment rejection.

**Example:** 200

Use all channels or a region of interest for segment rejection?

  all = all channels

  roi = region of interest

If you plan to analyze all of the user-specified channels in your dataset, input all.If you have a region of interest that you will be analyzing, input roiand you will be prompted to enter the channels in the region of interest.

If you selected to use a region of interest for segment rejection, the following prompts appear:

Enter the channels in the ROI, one at a time.

When you have finished entering all channels, enter 'done' (without quotations).

Enter the channels you wish to include in your region of interest one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Re-reference data? [Y/N]

If you wish to re-reference your data, input Y. Otherwise, input N.

If you selected to re-reference your data, the following prompts appear:

Re-Referencing Type:

  subset = Re-referencing to another channel/subset of channels

  average = Average re-referencing

To re-reference the data to a single (non-reference) channel or a subset of channels, input subset. To re-reference across all the user-input channels, input average.

If you selected to re-reference to a subset, the following prompts appear:

Enter channel/subset of channels to re-reference to, one at a time.

When you have entered all channels, input 'done' (without quotations).

Enter the channels you wish to include in your subset one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. Ensure that quotations are not used when inputting electrodes as well. Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Format to save processed data:

  1 = .txt file (electrodes as columns, time as rows) - Choose this for ERP timeseries

  2 = .mat file (matlab format)

  3 = .set file (EEGLab format)

Select your preferred format to save your processed data. We recommend .txt for HAPPE+ER outputs.

Are the above parameters correct? [Y/N]

Use this tool to check that your inputted parameters are correct. If you would like to change one or more parameters, input N.

If you selected to change a parameter, the following will appear:

Parameter to change: data file format, acquisition layout, channels of interest, bad channel detection, line noise frequency, line noise reduction, visualizations, resampling, wavelet thresholding, segmentation, interpolation, segment rejection, re-referencing, save format.

If you do not see an option, quit (Ctrl+C or Cmd+.) and re-run HAPPE.

Enter "done" (without quotations) when finished changing parameters.

Choose from the above list to change a parameter. You will be prompted with the original command to change the parameter. NOTE: this list may change depending on whether you are reprocessing your data. You may change as many parameters as needed, but must change them one at a time. Some selections may require that you answer multiple prompts.

If you created a new parameter set or changed a pre-existing set, you will be prompted to save your parameters:

Parameter file save name:

  default = Default name (inputParameters\_dd-mm-yyyy.mat).

  custom = Create your own file name.

The default save name for the parameter set is inputParameters\_DD-MM-YYYY.mat, with the current date.For custom save name, input custom and you will be prompted to choose a file name.

# Data Quality and Pipeline Quality Outputs

We include this information to be of use in decided which data files to include in your analyses. Please note that the example thresholds we give are not a global standard for all data, simply recommendations.

## Tracking Data Quality

…

## Using Outputs to Exclude Files

A primary use of the two assessment outputs generated as part of a HAPPE run is to examine the quality of your data and the performance of the HAPPE software on your dataset. You can use the measures to determine which individual files meet the standards needed for further analysis, removing those that do not qualify. Below, how each measure can be used is detailed, including recommended (but not global) thresholds you can use in your quality check.

### Data Quality Measures

File Length: *Does your file have enough data?*

* Remove files that do not have enough data. This could be an insufficient length of time for a particular paradigm, or an insufficient amount of time needed to conduct your analyses. As this threshold varies wildly depending on your experiment and your analyses, we are not able to make a recommendation for this measure.

Percent Good Channels Selected: *Do you have enough channels?*

* A threshold for this measure varies based on the experiment, the analyses, and the density of your acquisition layout. However, we recommend that at least 70% of your channels of interest are good.

Interpolated Channel IDs: *Are your channels of interest good?*

* The interpolated channels are channels that are marked as ‘bad’. If you have less than 2 channels remaining in your channels of interest or your region of interest after removing the listed channels, the file should be excluded.

Channels Interpolated for Each Segment: *How much data has been interpolated?*

* If you have chosen to interpolate your channels, this lists the channels in each segment that have been interpolated. We recommend not having more than 30% of your data interpolated for an individual file, but this threshold can be set at your discretion.

Number of Segments Post Segment Rejection: *Do you have enough trials/segments?*

* For ERP data we recommend having 20 segments or more. This should also apply to the number of segments for each stimulus tag/condition when you have multiple stimulus tags/conditions.
* For data with multiple tags/conditions, we suggest removing files that do not include all your conditions or an insufficient number of segments for any one condition, excluding cases in which the file should not have a stimulus tag/condition present in accordance with your study design.
* For other paradigms…

### Pipeline Quality Measures

## Variables as Co-Variates and Reporting Measures

### Data Quality Measures

…

### Pipeline Quality Measures

### …

# HAPPE+ER’s generateERPs Add-On

## Run generateERPs

1. If ignoring bad channels, create the bad channels spreadsheet.
   * For detailed instructions, read through the following command line prompts section.
2. Navigate to the main HAPPE folder in your file browser.
3. Open generateERPs.m in MATLAB.
4. In the Editor tab, hit “Run.”
5. Follow the prompts in the command window of MATLAB
   * For detailed instructions regarding the prompts, see below.

## Following Command Line Prompts:

Prompts as they appear in the command window are written in Courier New, like this, followed by a brief description of what to enter, with an example, also in Courier New. Sometimes a certain choice will result in a different set of prompts - in which case they will be indented under a heading that describes the choice needed for them to appear.

Preparing HAPPE - ERP ADD-ON...

Enter the path to the folder containing the processed dataset(s):

This is requesting the folder where your processed data is stored. In most cases, this is the “5 - processed” folder created during the HAPPE run. If you enter a non-existent path, you will be prompted to re-enter the correct path. Please note that whether you use a backslash or a forward-slash depends on your operating system (e.g., Mac vs Windows).

**Example (Mac):** /Users/laurelg-d/Desktop/Data Folder/5 - processed

**Example (Windows):** C:\Users\laurelg-d\Documents\Data Folder\5 - processed

Enter the suffix used for this dataset, including stimulus tag (if applicable).

If no extension beyond "AveOverTrials", press enter/return.

Files created by HAPPE as a default are named *filename\_processed\_AveOverTrials*. In the case of a re-run, they may have additional text following “AveOverTrials”, for example, *filename\_processed\_*

*AveOverTrials\_rerun2021*. If there is additional text, enter all the text that follows “AveOverTrials.” In this example, you would enter “\_rerun2021” without the quotations. If there is no additional text, you can simply hit your newline key - “enter” on Windows keyboards and “return” on Mac keyboards. Regardless if additional text follows “AveOverTrials”, you do not need to include “.txt”.

**Example:** \_VEP1

Examine all channels (all) or only channels of interest (coi)?

This question is asking about which channels/electrodes should be included in the ERP. To select all channels present in the dataset, enter all. To select a subset of channels, O1 and O2 for example, enter coi. If you choose to only examine channels of interest, you must follow some additional prompts (see below).

These prompts only appear if you have selected to filter to a set of channels of interest:

Choose an option for entering channels:

  include - Include ONLY the entered channel names.

  exclude - Include every channel EXCEPT the entered channel names.

Choose whether the channels you enter will be the channels of interest or the channels of

disinterest. If you select include, the channel names you enter in the following prompt will be the only channels included in the ERP. If you select exclude, the channels you enter will be the only ones not included in the ERP.

Enter channels, including the preceding letter, one at a time.

Press enter/return between each entry.

Examples: E17

          M1

When you have entered all channels, input "done" (without quotations).

Enter the channels you wish to include/exclude in your ERP one at a time. You should include the preceding letter, if applicable. If you have any questions about your channel names, refer to your acquisition layout. If the 10-20 channel locations are present and named accordingly in your dataset, use those names where applicable (e.g., use O1 instead of E70 for EGI GSN HydroCel 128 layouts). Between each entry, press your newline key (enter - Windows, return - Mac). When you are done entering channels, enter done.

Include bad channels in calculating ERP?

  include = keep bad channels

  exclude = remove bad channels

During HAPPE+ER’s processing, bad channels may have been detected in the dataset. If bad channels were detected, you can choose whether to include these channels (in the case that they were interpolated) or exclude them from the ERP. If you decide to exclude the bad channels, you will need to follow additional prompts (see below).

These prompts only appear if you have selected to exclude bad channels from the ERPs:

Enter the file containing the bad channels, including the complete path.

Refer to the HAPPE User Guide for instructions on creating this file and an example.

Prior to running the generateERPs script, you should make a spreadsheet (.xlsx, .csv, etc.) containing a list of the files and the bad channels associated with each file. The easiest way to do this is to copy the first column (labeled “Row”) and the “Interpolated\_Channel\_IDs” column from the HAPPE\_dataQC output created at the time the data was processed and create a new spreadsheet using those. Make sure to remove any rows for data you are not examining.

|  |  |
| --- | --- |
| Data File Name | Bad Channels |
| file01.raw | F3 F4 T6 CZ E31 |
| file02.raw | T6 O1 O2 P3 P4 CZ E6 |

**Example:**

For this prompt, enter the path to where the spreadsheet is saved followed by a slash (forward-slash or backslash depends on your OS) and the name of the spreadsheet, including the file extension (e.g., .csv).

**Example (Mac):** /Users/laurelg-d/Documents/badChannels.csv

**Example (Windows):** C:\Users\laurelg-d\Documents\badChannels.csv

Calculate ERP values? [Y/N]

This script has the capability to calculate values commonly associated with ERPs, such as peak amplitudes and their latency, and area under the curve. If interested in calculating these values, enter y; otherwise, enter n; selection is case-insensitive. Selecting to calculate ERP values will require following additional prompts (see below).

These prompts only appear if you have selected to calculate ERP values:

Enter latency windows of interest with anticipated peak:

Enter each window as two consecutive numbers followed by "max" or "min" (without quotations).

Press Enter/Return between entries.

When you have entered all windows, input "done" (without quotations).

Example: 10 100 max

Enter your latency periods as two consecutive numbers, with the first number representing the starting latency for your window and the second as the ending latency for your window. Try to use latencies that are included as points in your data or the script will correct your boundaries to the closest latency that exists in your data. Windows are also not allowed to include a negative latency value.

For each window you must also specify whether you anticipate to find a maximum or a minimum amplitude in the provided. You can do so by including max or min, respectively, following the two numbers representing the window’s boundaries. If you want to look for both a maximum and a minimum within the same window, you must enter the latency window twice and alternate specifying max and min.

**Example:** 10 100 min

Choose a method for calculating area under the curve:

  windows = restrict calculations to the specified latency window

  zeros = calculate area under the curve using points where the amplitude = 0

  both = calculate both by windows and by zeros

You can choose whether to calculate area under the curve using the windows you specified as bounds, using zero crossings present in the dataset as bounds, or using both methods. Choosing windows restricts all area under the curve calculations to only the latency windows you specified previously. Choosing zeros allows the script to find the zero-crossings in the dataset and use those points as boundaries to create new windows in which to calculate the area under the curve.

**Errors:**

If at any point, generateERPs is unable to create the ERP or calculate values for a particular file, it will print out “Error in file FILENAME” to the command line and proceed to the next file.