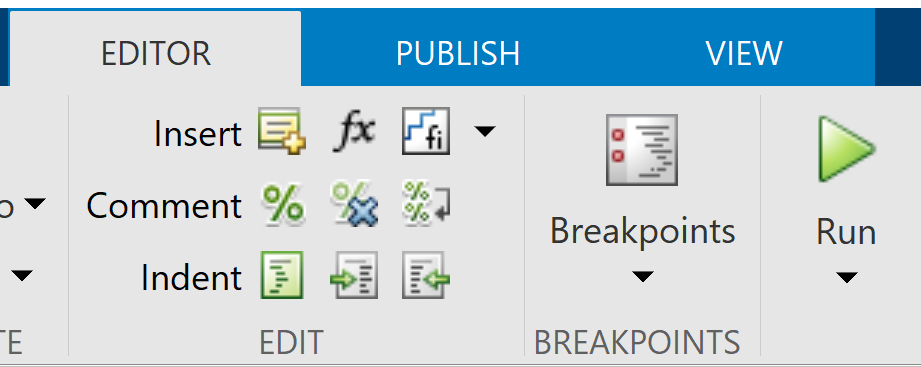
**HAPPE+ER & HAPPILEE Combined Pipeline User Guide**

**Run HAPPE:**

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.”
4. 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.

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

Enter data type:

rest = Resting-State EEG

task = Task-Related EEG

Input task for HAPPE+ER & HAPPILEE.

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

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

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.

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

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

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.