**Workflow for FFRClassifiers Scripts**

Here are the steps to follow relative to scripts of the folder “FFRClassifiers”:

a.       Place the matlab matrices in the folder “ffrs”

(The matlab matrices are already there for this tutorial, but this you be where to put them if you use another dataset)

b.       Run “createInputDatabase.m” to get these matrices ready for the hmm/svm/xcorr processing

c.       Run “counting.m” in folder “input1” if you want to check the number of clean ffrs per subject

d.       To visualize individual/grand ffr waveforms/spectrograms, run “exploreInput.m”. Figures will be automatically saved in folder “figures”

e.       Go to the folder of the classifier you want to use and run “main.m”

f.        Once “main.m” is done, run “analysesCorrectRates.m”, “analysesLogProbs.m”, and/or “analysesXcorrs.m” to visualize the results. Plots will be automatically saved in folder “figures” within the corresponding classifier folder

**Observation from Author of the Scripts (Fernando Llanos)**

(1)    In the excel files (attached) you will find the HMM log probabilities for syllable (“logprobs1”) and tone (“logprobs2”). Columns in each of these files contain the following information (from 1st to last column): subjectNumber, groupNumber, trialSize, logProbability. These files are automatically created and saved in the “logprobs” folder by my scripts, so you will be able to get them too if you want to change the models or your subject database in the future

(2)    Subject #210 (not #209) was excluded from the modeling because its number of ffrs per class was smaller than 1500. Some trials for this subject were labeled as “0” rather than “1/2”.This may explain why. Looks like something that can be easily fixed

(3)    This time I applied a standard baseline correction preprocessing step, followed by an artifact rejection step of baseline corrected ffrs with maximal-absolute-amplitude higher than 35uV. The number of ffrs rejected varied across subjects. I kept only the subjects with equal or more than 2300 clean ffrs (11 in musicians: 101,102,103,104,105,106,107,109,110,112,114; and 10 in no-musicians:201,202,204,205,207,208,209,211,213,214). 2300 looked like the maximum number of trials that gave me a decent number subjects in each group. Since I am attaching the scripts here, you will be able to run the scripts over different selection of subjects.

(4)    You will find the results in the pdf attached here. In summary: overall HMM accuracy (correct rate) increased reaching values closed around 0.9 of accuracy for 2300 trials. Musician logprobs for tone were statistically higher than no-musician logprobs for the highest sample sizes (e.g., 2300), and the trend across sizes suggests that between-group differences will increase for larger sizes.

(5)    Overall SVM accuracy (correct rate) is lower than HMM accuracy. As far as I know, SVM does not provide logprobs (or something similar) for categories because it focuses on whether vectors are inside/outside of hyper-volumes (kind of “yes” or “no” metric). I think it would be possible to define a continuous svm metric by computing orthogonal distances from svm vectors to hyper-volumes, but since overall SVM accuracy is lower than overall HMM accuracy, not sure we want to invest time on that.