**Using the Matlab code**

1) All the data used is in directory **\data** which contains three directories:

**\data\all\_dieStimmenDerHeuschrecken**: All the wav files (965 recordings for 89 species) of the CD from the book “Die Stimmen der Heuschrecken” (originally in three directories “Allgemeiner Teil\_wav”, “Bestimmungsteil\_wav” and “Oszillogramme\_wav”) grouped into one unique folder.

**\data\enregistrements**: data as in Table 1 in section 4 of the report. The seven species present in the filed recordings.

data\_albopunctata: 4 files

data\_biguttulus: 6 files

data\_lineatus: 4 files

data\_parallelus: 4 files

data\_roeselii: 4 files

data\_rufus: 3 files

data\_scalaris: 3 files

**\data\enregistrements\_6:** data as in Table 3 in section 4 of the report. The eight species that have more than 6 files, only six files per species were kept.

data\_biguttulus: 6 files

data\_cantans: 6 files

data\_denticauda: 6 files

data\_gryllotalpa: 6 files

data\_heydenii: 6 files

data\_parallelus: 6 files

data\_stridulus: 6 files

data\_viridulus: 6 files

**\data\field\_sounds: field recordings** (see section 3.1.1 of the report)

2) run the script “**\matlab\_code\test\_results.m**” to produce the results (Tables 2 and 4 in section 4 of the report). The scripts does four runs: each run produce a mat file (“**results1.mat**” to “**results4.mat**”) containing four tables:

- table\_2\_org, table\_4\_org: results with the three validation methods (**crosVad.m, oneVsAll.m, allVsAll.m**) using the features already calculated by William (his messages of 14.01.2016) placed in directory “**\matlab\_code\orgMatFiles**”:

- **data\_fixed.mat**: data as in Table 1 of the report, fixed frame-length method.

- **data\_variable.mat**: data as in Table 1 of the report, variable frame-length method.

- **data\_fixed\_6.mat**: data as in Table 3 of the report, fixed frame-length method.

- **data\_variable\_6.mat:** data as in Table 3 of the report, variable frame-length method.

- table\_2, table\_4: results obtained by first calculating the features (**data\_fixed.mat, data\_variable.mat, data\_fixed\_6.mat, data\_variable\_6.mat)** using “**main\_fixed\_frame.m**” for fixed frame-length method and “**main\_variable\_length.m**” for variable frame-length method, and then applying the three validation methods (**crosVad.m, oneVsAll.m, allVsAll.m**).

Note: in lines 8-9 of “**\matlab\_code\test\_results.m**” we can choose either "seed = 1;" to obtain exactly the same results for the four runs (see “**\matlab\_code\results\_fixed\_seed**”) or "seed = ‘shuffle;" to have each time a different run with slightly different results (see “**\matlab\_code\results\_shuffled\_seed**”).

3) Run the script **\matlab\_code\GMM\_final\_test.m** to test with the field recordings in .\data\field\_sounds (see section 3.1.1 of the report). Note: check that the name of the loaded feature file correspond to the method and species that we want to use (e.g. “load data\_variable.mat;” for the variable frame-length method, with the species of Table 1).

4) Run the script **\matlab\_code\activity\_detector.m** to observe the cutting of the frames by the activity detector. To use it, run activity\_detector(audioFile, showPlot), where “audioFile” is the path to the record that need to be studied and “showPlot” is a parameter to show a plot of the signal (set it to true). For example:

>> audiofile = '..\data\enregistrements\data\_albopunctata\021 Platycleis albopunctata albopunctata\_1\_S.wav';

>> activity\_detector(audiofile,true);

5) Run the script **naive\_bayes.m** an example of implementation to use naïve Bayes matlab function (http://ch.mathworks.com/help/stats/fitcnb.html). It uses “cross-validation” technique.

6) The files **allVsAll\_Resamp.m, oneVsAll\_Resamp.m, crosVad\_Resamp.m** are similar to **crosVad.m, oneVsAll.m, allVsAll.m** but the frames are resampled to have the same amount for each species.

7) Other scripts used by the scripts mentioned above are: **highpass\_filter.m, lfcc\_variable.m, lfcc\_fixed.m, trifbank.m, double\_thresh.m, vec2frames.m, extractLength.m**.

8) The files in the directory **\matlab\_code\new\_fixed** correspond to another version of the fixed frame-length method which automatically search the files in the database (**\data\all\_dieStimmenDerHeuschrecken**) according to a list of wished species (**selectedSpecies.mat** for data of Table 1 and **selectedSpecies\_6.mat** for data in Table 3). Run the script **test\_results\_with\_new\_fixed.m**  to produce the results (Tables 2 and 4 in section 4 of the report) with the features calculated with this alternative version (**calculateFeatures.m**). Note that the files for the three validation methods (**crosVad.m, oneVsAll.m, allVsAll.m**) contained in this directory (**\matlab\_code\new\_fixed**) where slightly modified from the files in directory **\matlab\_code** by uncommenting three lines and commenting three lines (assigning the variables: features, label, id). This is done because the feature vectors are not exactly in the same structure and one needs to adapt the index. Finally, the file **getSpeciesWithMoreData.m** gives an example on how to automatically select all the species with more than 6 files, and then produce datasets with 6 files, 5 files, … with no manual manipulation of the directories and files.

9) Other (document) files are:

**readme.docx**: This file

**presentation\_final.pptx**: final presentation sent by William the 14.01.2016, presented the 14.12.2015.

**rapport\_final.docx, rapport\_final.pdf**: final report sent by William the 14.01.2016

**rapport\_final\_notesSGr.25.01.2016.docx**: Final comments of SGr, for information only, in track mode on document **rapport\_final.docx.**

**confusion\_matrix.xlsx**: the confusion matrices of Table 5 and Table 6 of the report.