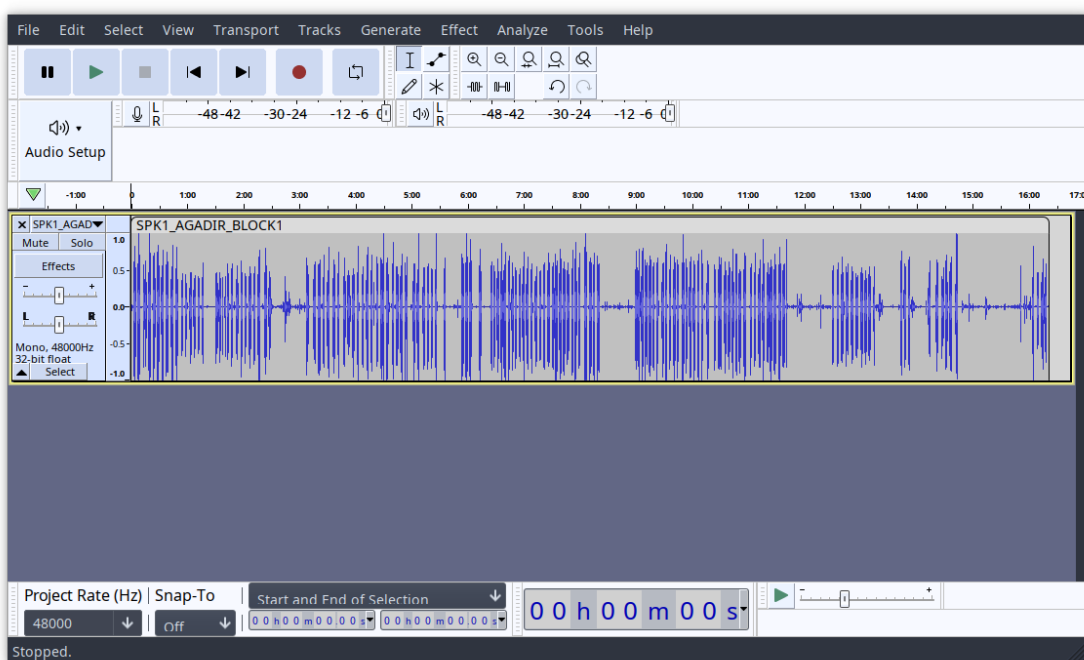


# Audacity & PRAAT: useful functions

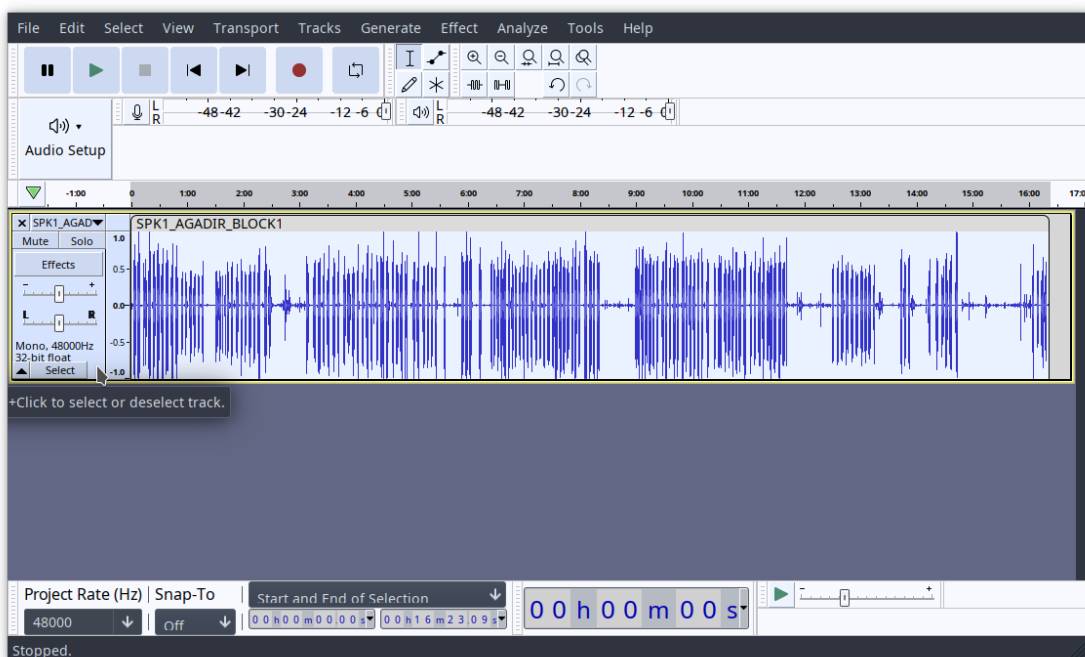
The MFA works best if utterances were pre-segmented instead of using one long audio file and the corresponding transcriptions. Fortunately, Audacity and PRAAT provide functions that allow us to do this task automatically.

## 1 Audacity

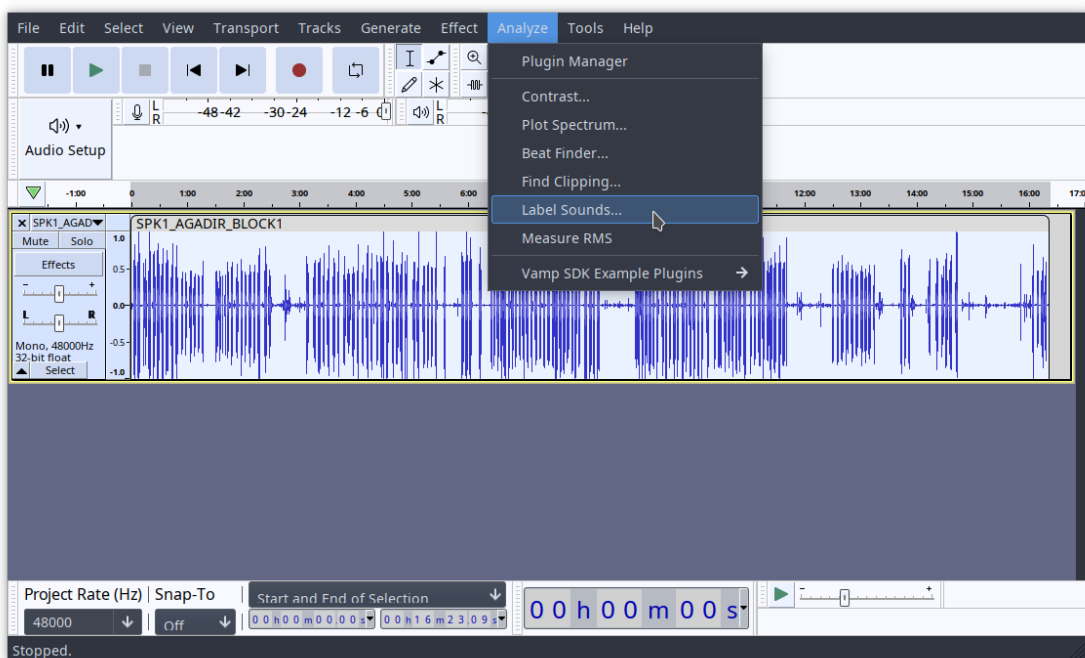
In order to use Audacity for the pre-segmentation, download it [here](#). Start audacity and import an audio file:



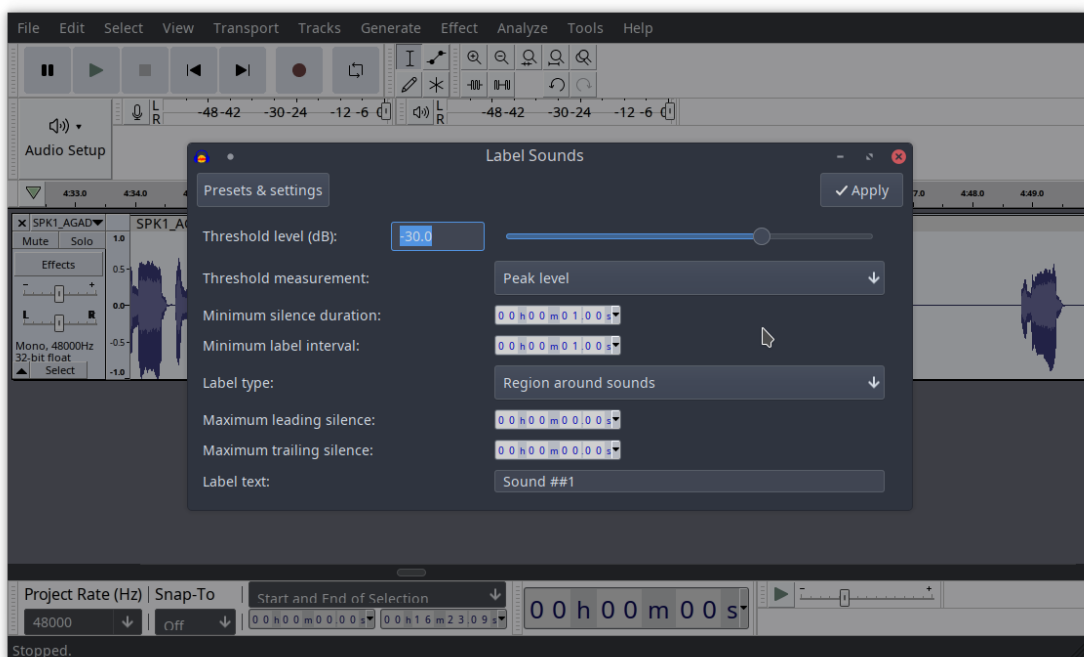
Select the audio by clicking on the track panel to the left.



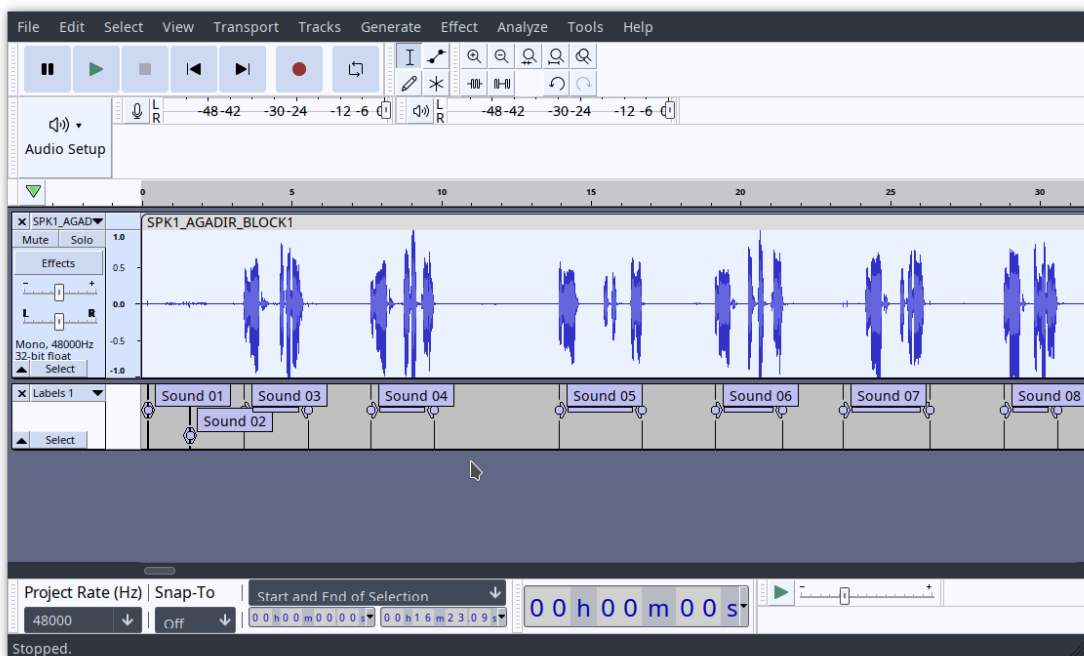
Now go on Analyze and click on Label Sounds...



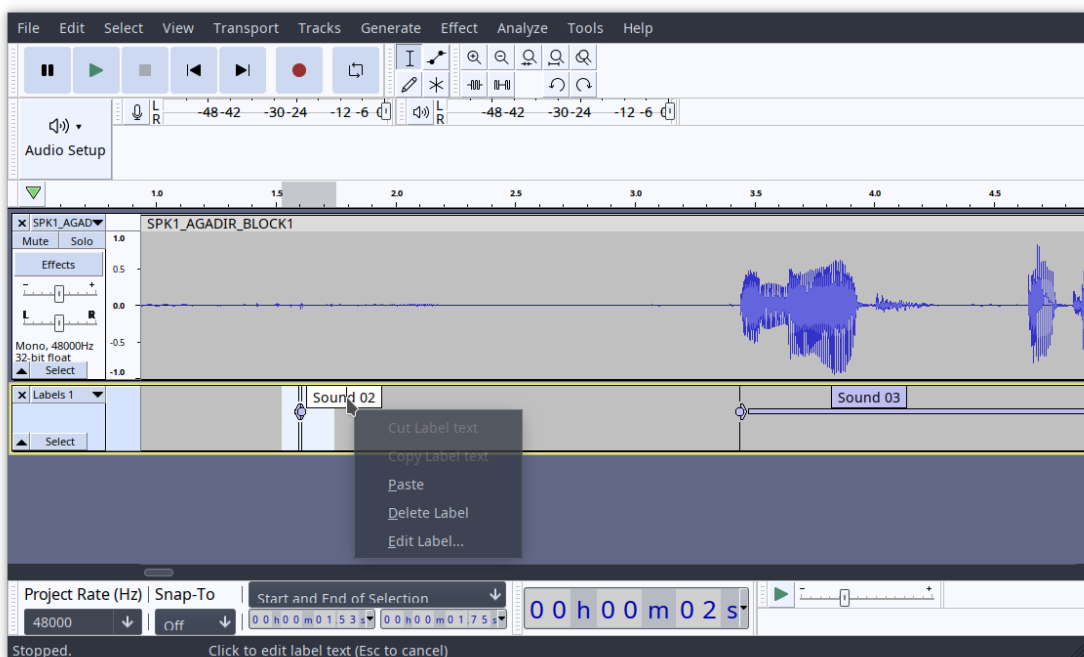
A new window appears, where you can modify some parameter for the detection of voiced regions. This is the Threshold level that is used for this detection, the kind of Threshold measurement, as well as different kinds of durations for silence and label intervals. In this case, I imported an audio file from an experiment and I used the standard settings for the Label sounds function. This should be sufficient, however, you can modify these options according to the needs of your recordings, of course.



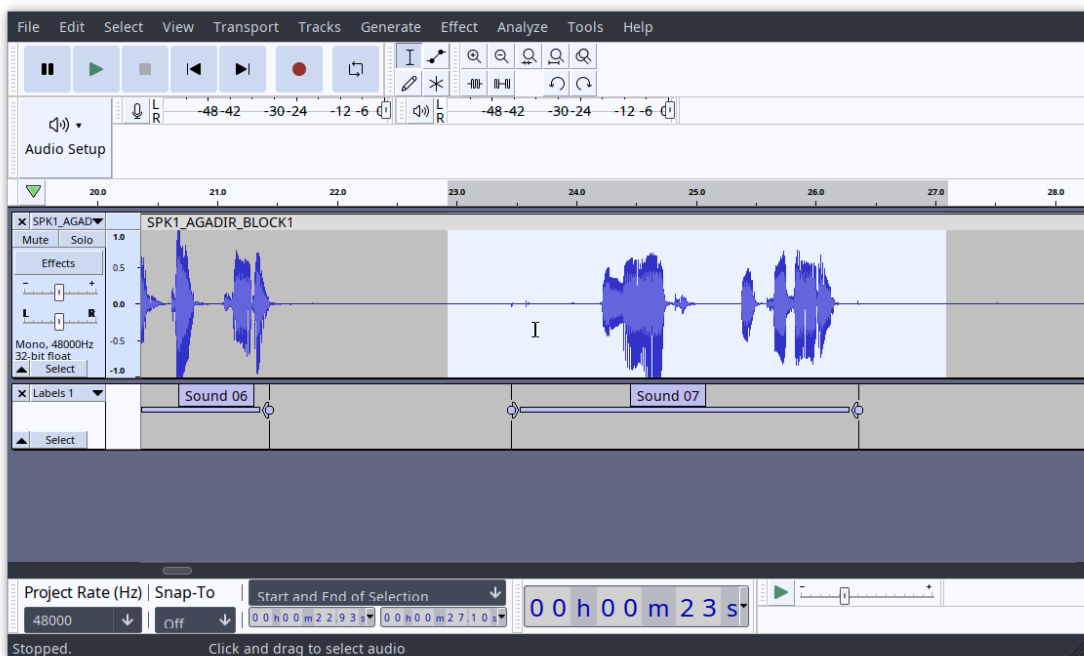
After applying these settings, you will see that a new track appears below your audio, that shows the intervals and generic Sound labels. When you zoom into the track, you will see that all of your utterances are detected.



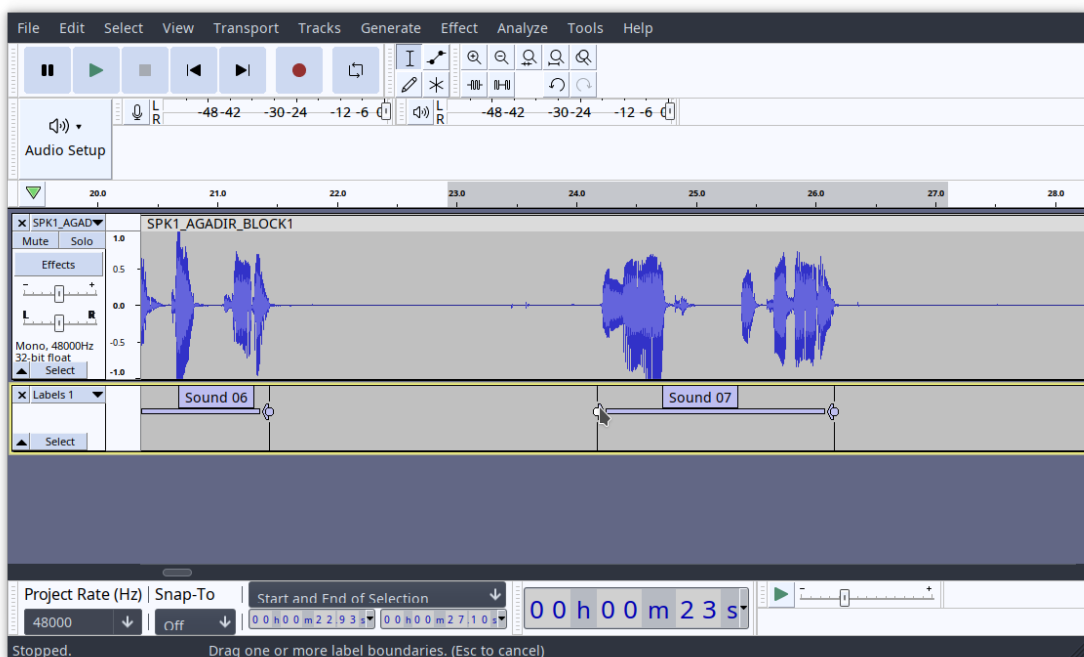
However, you will also see that some non-speech intervals were detected. You can either ignore them or remove them by right-click on the label's text box and then select delete.



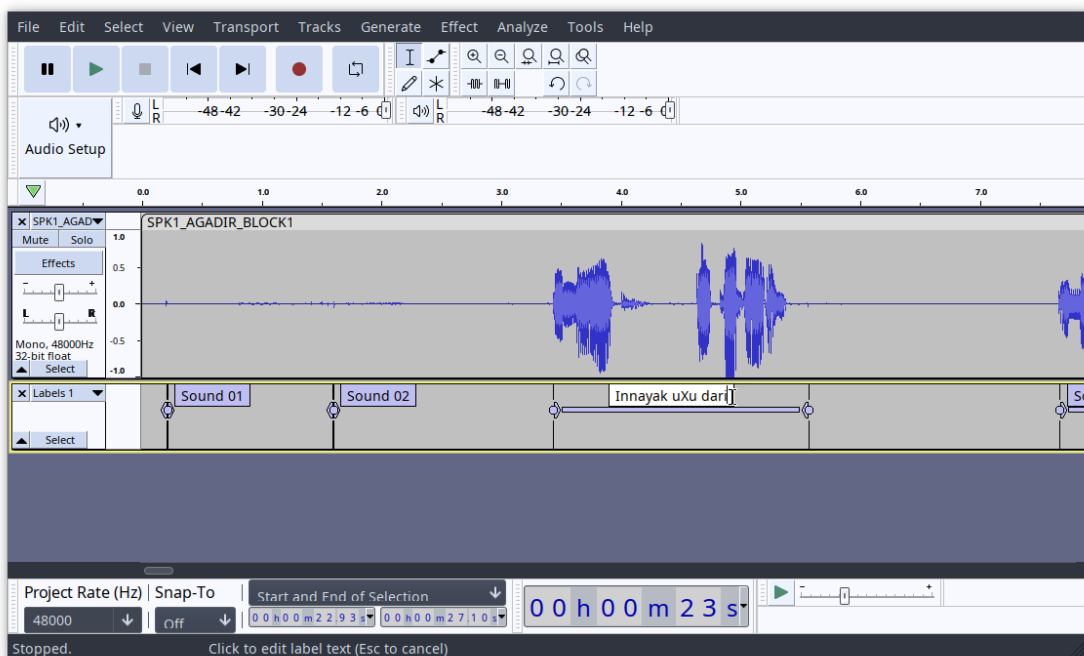
You may also see that in some cases larger intervals occur, e.g. if there are some clicks before the utterance:



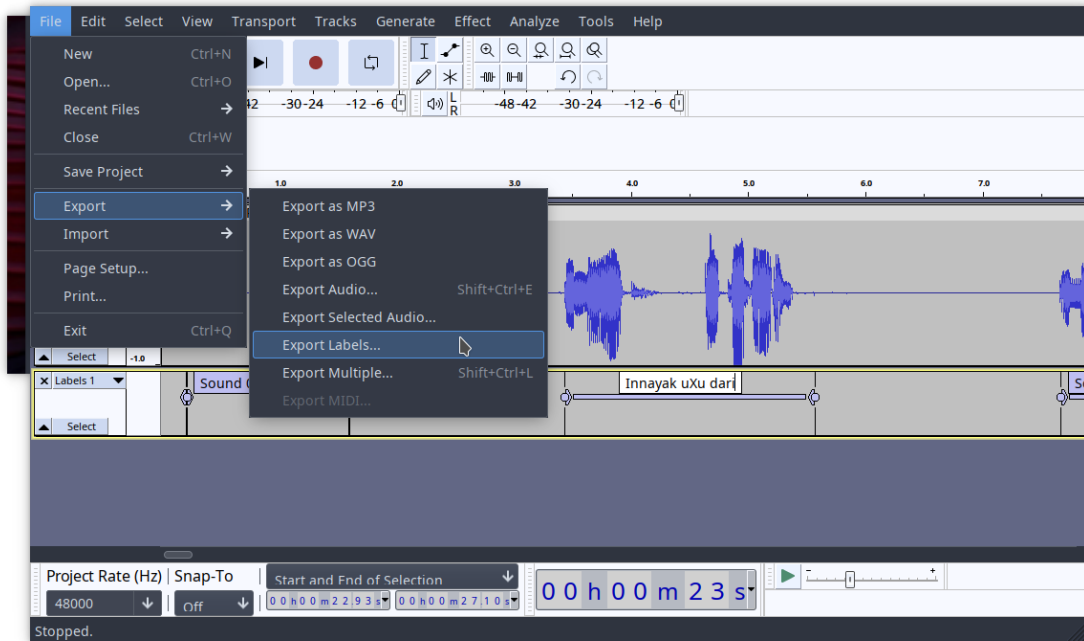
The intervals can be modified by dragging the interval boundaries to the position you want:



Then you can enter your text into the label's text box. This must not necessarily appear in this order. Of course you can also enter the text first and then change the interval.



After you entered the text of all your utterance, you can export the labels by clicking on File > Export > Export Labels. . .



You should now see a text file. In this text file are rows for each interval that consist of the time stamps of the beginning and the end of the interval as well as its label.

1	3.392891	5.580000	Innayak aka dari
2	7.670000	9.770000	Innayak aga dari
3	13.809494	16.730000	Innayam ud_?d_?u bahra
4	19.024934	21.440000	Innayak agg_wa dari
5	24.126167	26.360000	Innayak iki dari
6	28.770573	30.610000	Innayak axa dari
7	33.148497	35.050000	Innayak uRu dari
8	38.020000	39.960000	Innayam ulu bahra
9	43.389031	45.730000	Innayam assa bahra
10	48.318954	50.200000	Innayam ad_?gal bahra
11	54.310000	56.120000	Innayam uzzu bahra
12	58.710000	60.610000	Innayam uzu bahra

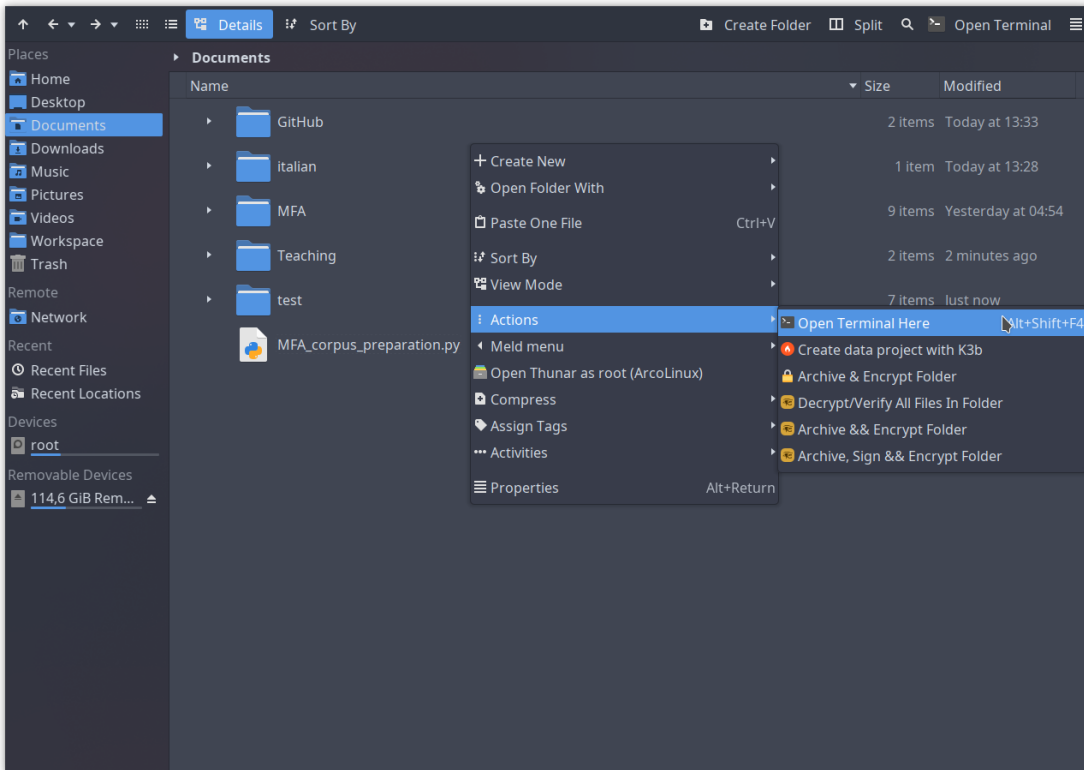
As the next part of the workflow, you can use the script `MFA_corpus_preparation.py`. In order to use this, you have to activate the anaconda base environment. If you installed anaconda and open your terminal/command line, it should be active automatically (indicated by “(base)” somewhere in the current line):

```
(base) [philipp@pxps ~]$
```

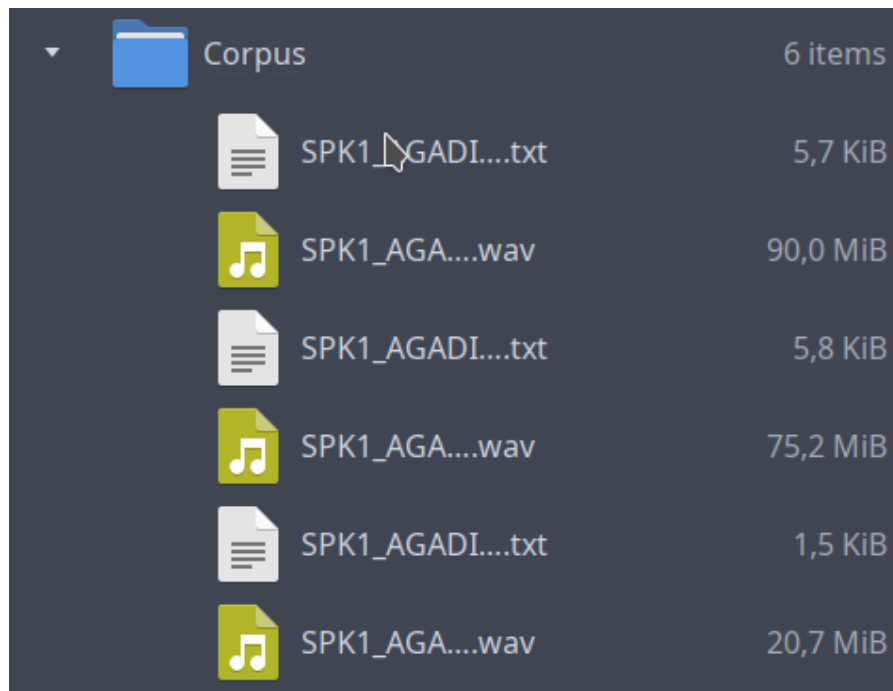
if this is not the case, enter the following command to activate the base environment:

```
[ ]: # conda activate base
```

Navigate to the location of the script by the `cd` command. Alternatively, you can also navigate to the folder of the script in your Windows Explorer/Finder(?)/Window Manager, right-click somewhere and press open terminal here or the equivalent in your OS.

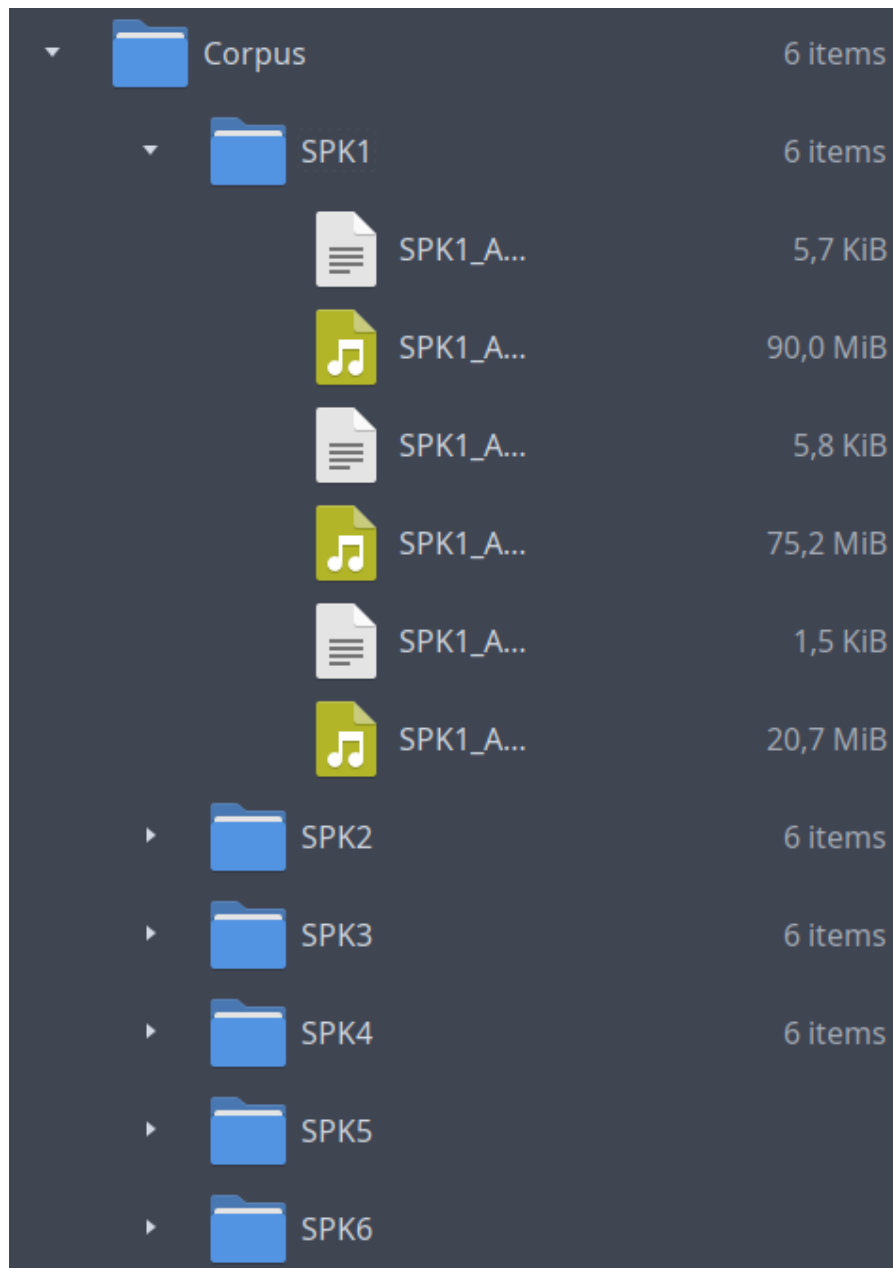


In this case, I moved the `MFA_corpus_preparation.py` script into the Documents folder. When you open your terminal, you should see the current path. If you enter `ls`, you can see that the `MFA_corpus_preparation.py` script is in the current location. We prepare now a corpus for the alignment. The requirement is that you have either - 1) A folder containing wav and text files or



- 2) A folder containing folders for





The script has to be used as follows in your terminal:

```
[ ]: # python MFA_corpus_preparation.py -i INPUT_PATH -o OUTPUT_PATH -mf "yes"/"no"
↳ (OPTIONAL) -el PATH_TO_EXCEPTION_LIST (OPTIONAL)
```

We have two obligatory flags: -i for the input path -o for the output path

additional fals are

-mf multiple folders, enter “yes” if you have multiple folders for each speaker

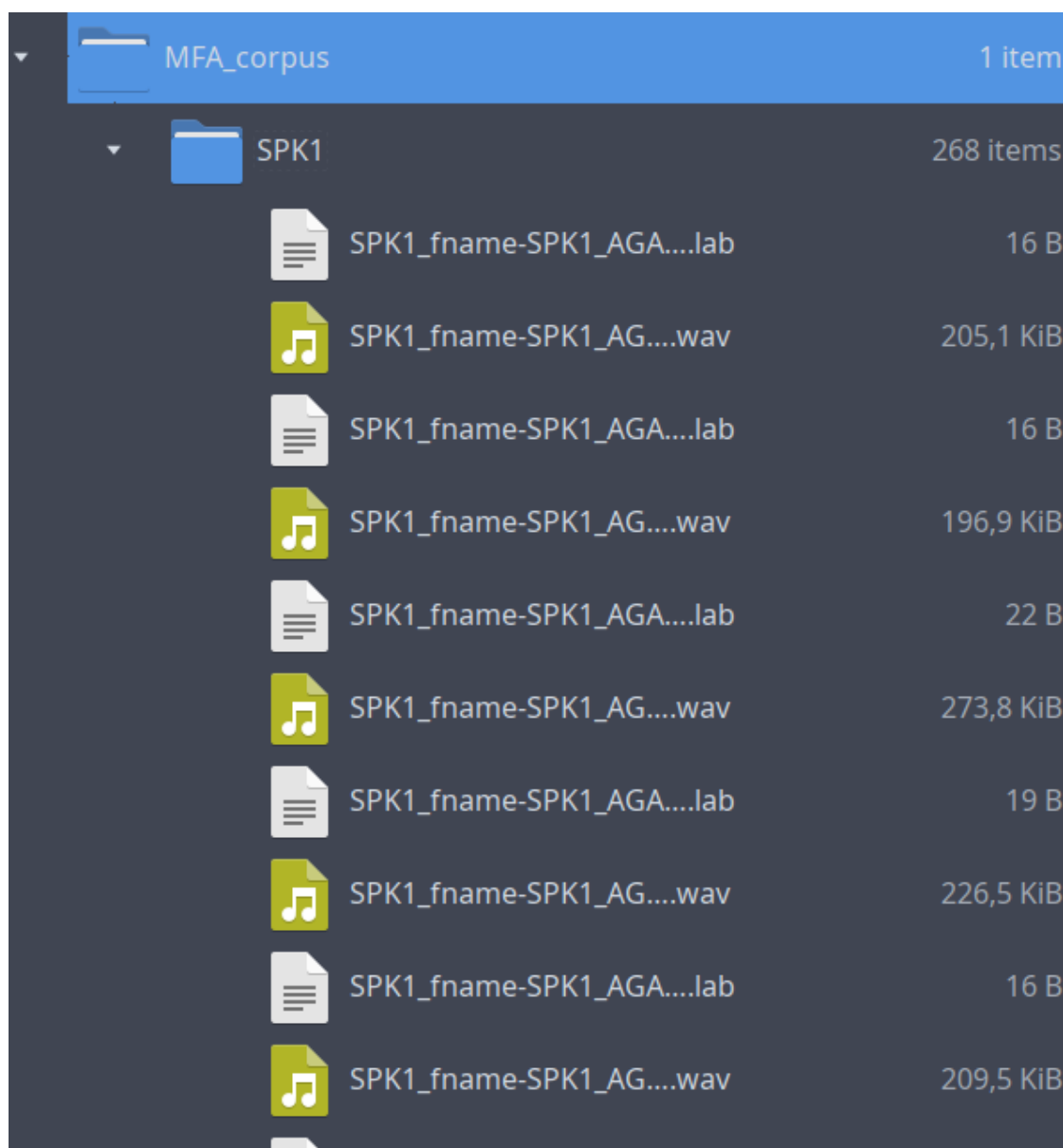
-el exception list. This flag is needed if you have certain labels you want to exclude from the corpus preparation. This is the case if you used the Sound Labels... function of Audacity and you not deleted the some intervals that are not representing speech. These intervals are labeled with “Sound”

+ a number. Simply create a text file and enter the word Sound in it. If you use the -el flag with the path to this text file, all the intervals containing “Sound” as label will be skipped.

In this case, I have the corpus files in a “Corpus” folder in the Documents directory. I want export all the utterances in a new folder called “MFA\_corpus” in Documents. The corpus has the structure as in 2), that is, one folder per speaker, in which the wav and txt files can be found. This is why I used also the flag -mf “yes”. The full command is:

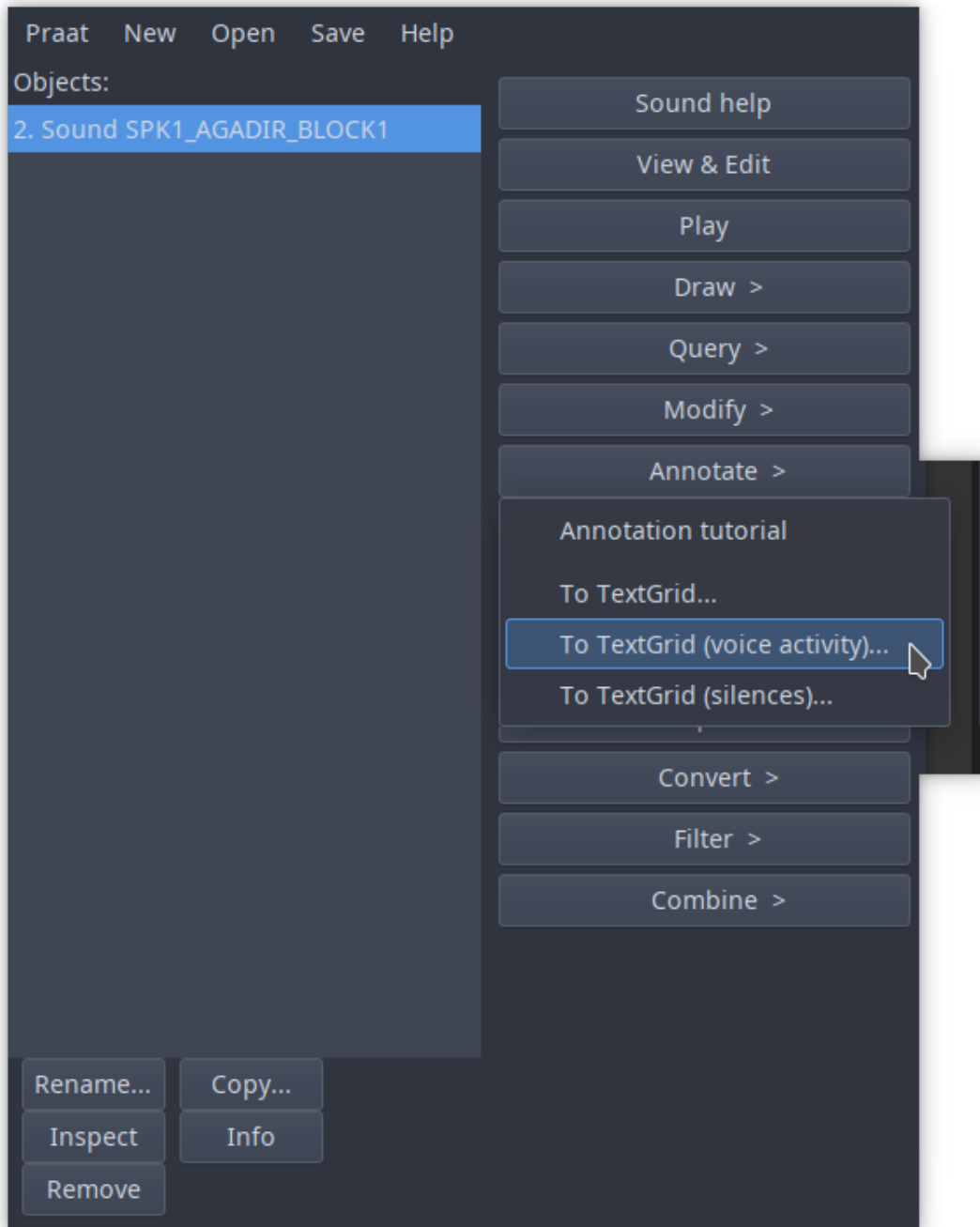
```
[ ]: # python MFA_corpus_preparation.py -i ~/Documents/Corpus/ -o ~/Documents/  
    ↪ MFA_corpus/ -mf "yes"
```

First, the output folder has not been created yet. Enter “yes”. The script starts to extract the data. The resulting MFA\_corpus has then the structure and all files that is used for the training of an acoustic model or aligning the acoustic data in the MFA.

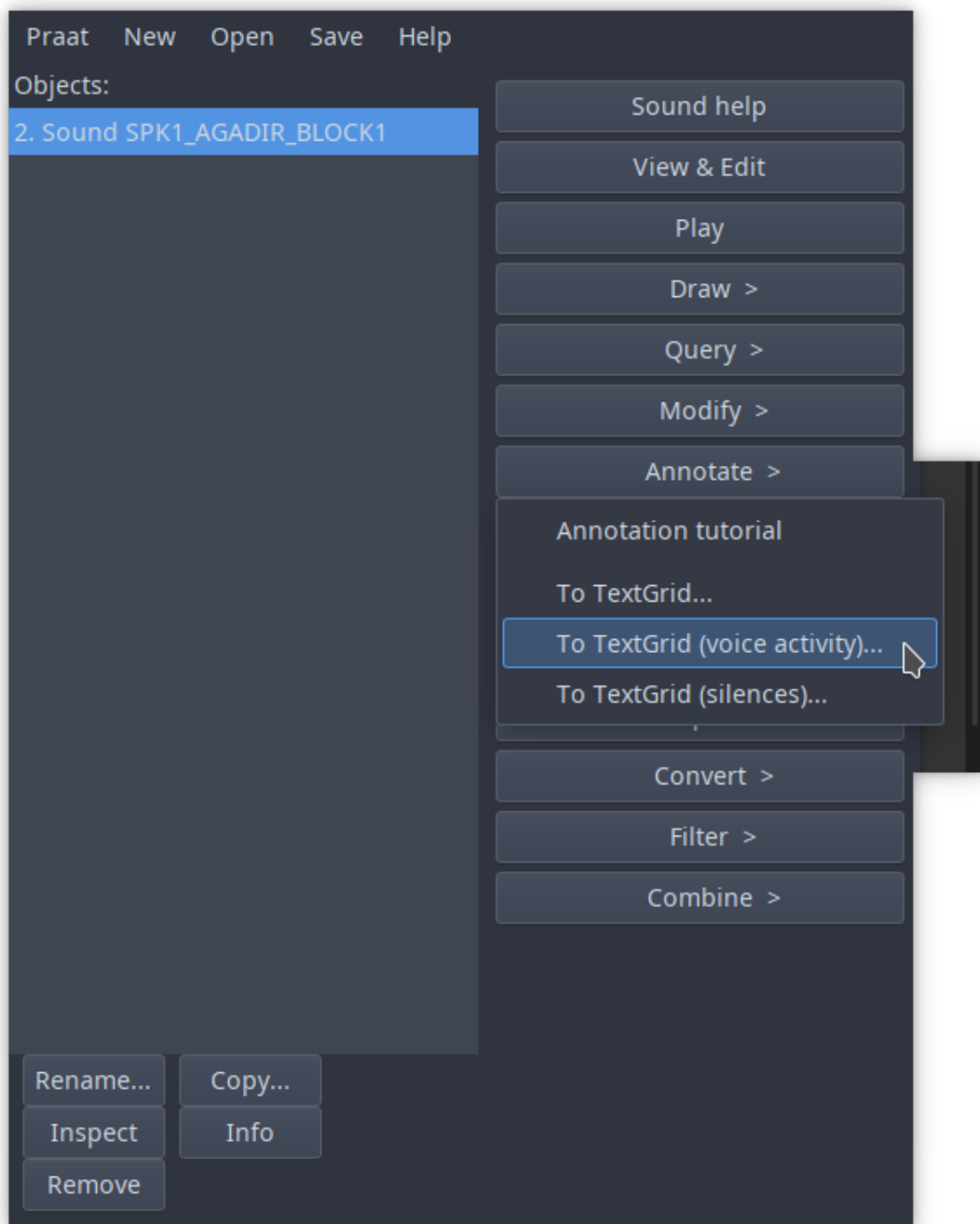


## 2 PRAAT

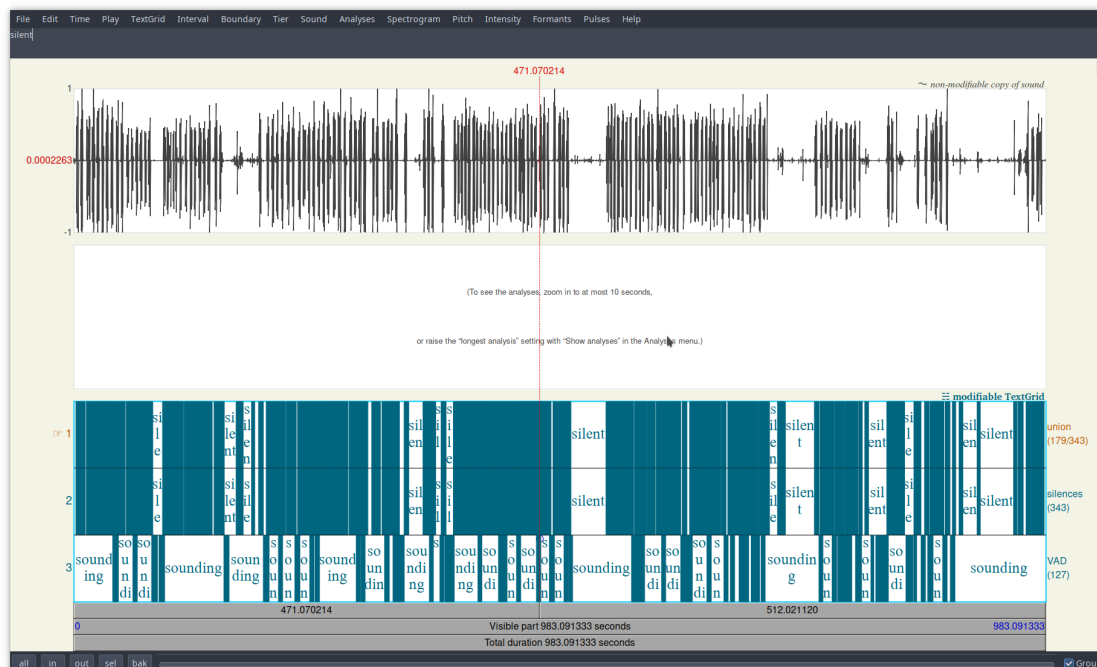
Pre-segmentation can also be done in PRAAT. In order to create a TextGrid with utterance intervals, open your audio in PRAAT. Then click on Annotate and To TextGrid (voice-activity)



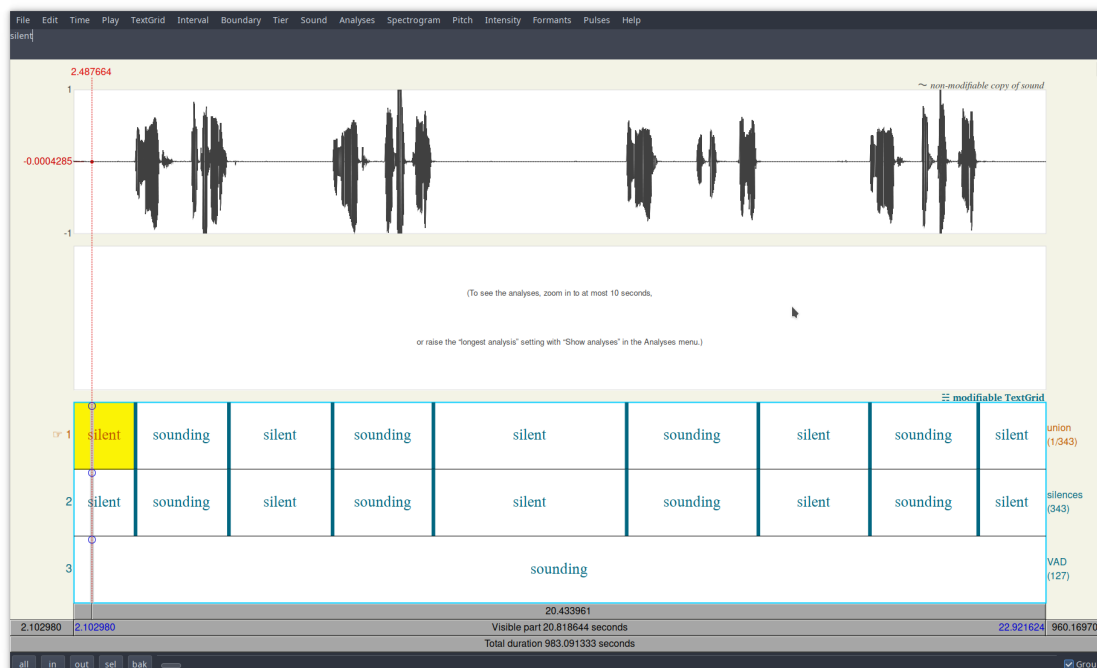
A new window will open where you can enter multiple parameters for the detection of voiced regions.



The standard settings may do a good job, however, they will also divide utterance that contain pauses longer than 0.1 s. In this case I changed the minimum silent interval to 1.0 s. Run the voice activity detection. The resulting TextGrid has three tiers, but only the first tier is important.



You will see that the intervals in the first two tiers contain silenced and voiced regions and in the voiced regions you can enter the transcription of these utterances.



After you did your annotation, open the MFA\_corpus\_preparation.praat script and modify the input and output paths. This script extracts the utterances of all audio and TextGrids files in a directory. If you have multiple speakers, run the script for each speaker. The corpus that was created this way can then be used for forced alignment.