# **Manual for the new Forced Aligner [October 2019] – Maarten Vos**

## **Step I. – Logging into Ponyland (To run commands)**

1. First you need to have an account. Please request an account from [admin@cls.ru.nl](mailto:admin@cls.ru.nl). Write your full name (preferably in capitals) and give your e-mail address. After your account has been created you will receive an e-mail with an invitation to visit the admin, so you can change the password.

Before the meeting, it is recommended to check the webpage and read the basic information about Ponyland so you can ask questions and the admin will help you (more information: <https://ponyland.science.ru.nl/wiki:ponyland:about>)

For more information about the availability of the admin: <https://ponyland.science.ru.nl/doku.php?id=wiki:ponyland:admin>

In case you would like to change your password, you can do it on: [http://diy.science.ru.nl](http://diy.science.ru.nl/)

1. After receiving your username and password you need to download PuTTY at <https://www.putty.org/>. Download the MSI installer for Windows. Choose the 32 bit installation if you are not sure whether your computer runs a 64 bits installation of Windows.
2. Install PuTTY and start it.
3. Set up your connection with the following settings (click *Save* to save it for subsequent sessions) and click *Open*:
   1. Host Name: *applejack.science.ru.nl*
   2. Port: *22*
   3. Connection type: *SSH*
4. For *login as,* fill in your username (which is either the first letter of your first name and your last name or the entirety of your first name and last name, so *jjansen* or *janjansen* or as in the example: *fhatvani).* You will get your username from the admin (see step 1). For password, fill in your password and press *enter* (note that you will not see your password or anything indicating that you are typing, but it does remember what you type).

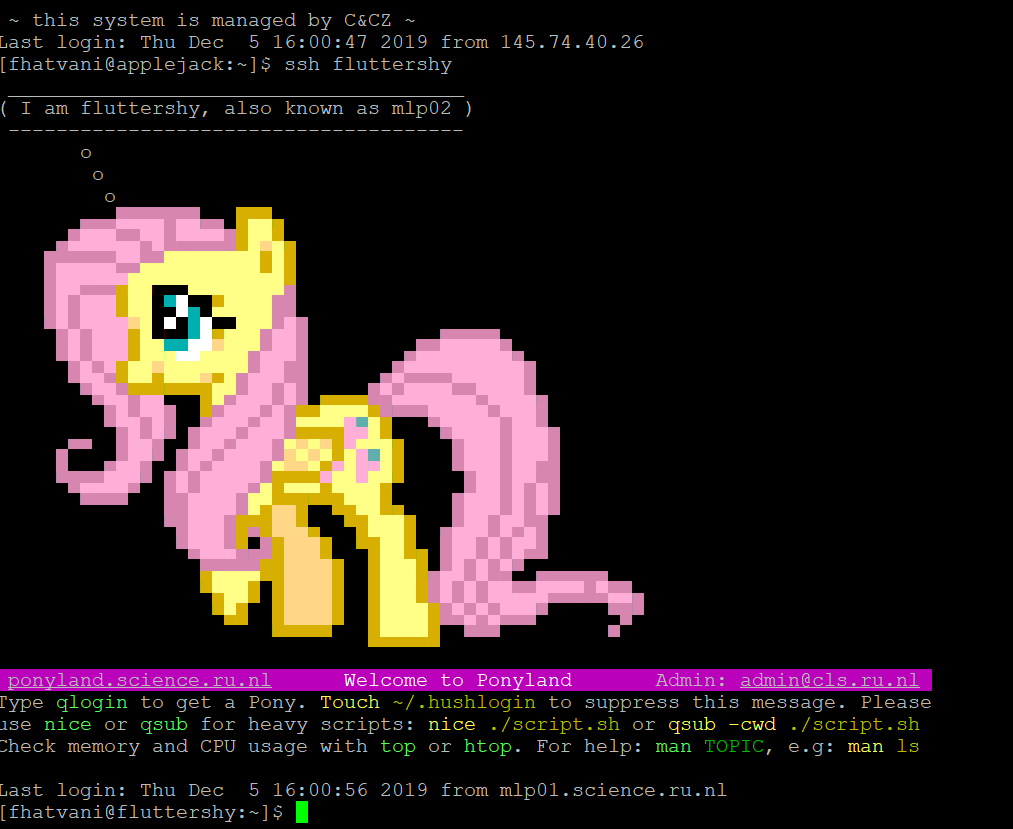
You will see this:



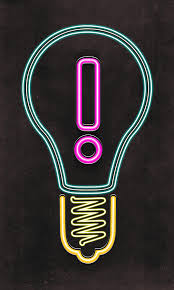
1. Now that you are logged in on Applejack, you have to determine which server you want to run your experiments on, as Applejack is not suitable for this.
2. Choose a server from the list of pony servers on <https://ponyland.science.ru.nl/wiki:ponyland:about>, for example *dinky, fluttershy* or *pipsqueak*. Check which one you need (depending on CPU). If you are not sure which one to use, *dinky* is a good option.
3. Type *ssh* and the name of the chosen pony and press Enter, for example: *ssh fluttershy*



1. You are now connected to the correct server and are ready to run the experiments. You will see this (your login name and chosen pony):



→ **Important**

You cannot access Ponyland directly from outside the RU network.

However, the Ponyland server Applejack is open to the entire internet and from there the ponies are reachable.

If you get stuck, first have a look at the website: <https://ponyland.science.ru.nl/doku.php?id=wiki:ponyland:about>

If you have problems to connect, maybe try to download the VPN for creating a connection to the university so you can access Ponyland: <https://www.ru.nl/ict-uk/staff/working-off-campus/vpn-virtual-private-network/>

→ **Are you new to the Unix command line?**

If you do not have any experience on the Unix command line, please take a look at the following links:

* [](https://ponyland.science.ru.nl/doku.php?id=wiki:unix:cli)https://ponyland.science.ru.nl/doku.php?id=wiki:unix:cli
* [https://ponyland.science.ru.nl/doku.php?id=wiki:unix:basics](https://ponyland.science.ru.nl/doku.php?id=wiki:unix:basics#_blank)
* This link gives good general information: [https://tutorials.ubuntu.com/tutorial/command-line-for-beginners#00](https://tutorials.ubuntu.com/tutorial/command-line-for-beginners#_blank)

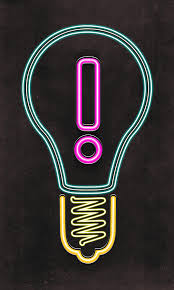
## **Step II. – Putting the files on the server**

1. Download FileZilla from <https://filezilla-project.org/download.php?type=client>. Choose the 32 bit installation when you are not sure if your version of Windows supports 64 bits. Note that the link will bring you to the 64 bit version. If you need 32 bit installation, click on “other platforms” and download Filezilla Client 32 bit.
2. Install FileZilla and start it. Note, that if the language of your computer is not English, the program will run in the same language (You can change it: Edit*→Settings→Language*).
3. Click on File in the upper left corner to open S*ite manager (File→Site Manager).* A new window will pop up. Click on *New site.* Fill in the following information:
4. Protocol: *SFTP*
5. Host: *applejack.science.ru.nl*
6. Login type: *Normal*
7. Put your credentials in the *User* and *Password* boxes (type the same username and password you used in Step I/1.).
8. Click on *Connect*.
9. On the left hand side of your screen, you will see the folders that are currently on your own computer. On the right hand side of your screen you will see the folders that are currently on the Ponyland servers. Note that these files are present on ALL Ponyland servers, not just Applejack, so it does not matter where you put your files.
10. You are currently in the /home/[user\_name] folder (for example: /home/jjansen).

This is also where the config, the lexicon and the log files are stored when running the forced aligner script. You can now navigate to where you want to store your files (you will go one folder back if you click on the folder with the ‘..’).

Just creating a new folder (e.g. *Data*) at your current location is a good idea, but you can create a folder at any location that you want, as long as you remember the path to that location. This is done by pressing the right mouse button and clicking on *Create new folder*.

If you already have one or more directories containing both wav-files and text grids, just remember their locations and skip the next steps.

1. Open the created folder and put both your files inside. This should be both the wav files of the audio that you want to force align, and the textgrids for these files, in the format [wav\_name].tg (so for example participant1.wav and participant1.tg).

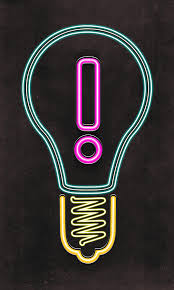
**Important**: Make sure that the wav and the text grid files have exactly the same name.

Putting these files onto the server can be done by:

- either dragging and dropping them from windows explorer onto the right-hand side of the screen

- or by navigating towards the folder on your own computer (left-hand side) and dragging them to the right side of the screen. Note: Copy-paste method is not working in this case.

1. Wait until all files are uploaded and remember the location where they are uploaded for the next step (it is recommended to save the route to your files in order to reach them easily), as you have to run the forced aligner script on these files.



→ **Note** that all of the wav-files should be in Dutch or all should be in English, no combination between the two is possible as only one lexicon is used to handle the files in all specified folders.

## **Step III. – Run the forced aligner**

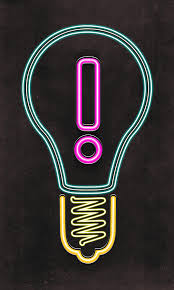
1. It is possible to run the forced aligner in multiple ways, but the result is the same.
2. If there is one folder to be used, you can type (from any location) into PuTTY

**/vol/tensusers2/mvos/clst-asr\_forced-aligner/run.sh [path/to/dir/]**,

so for example: */vol/tensusers2/mvos/clst-asr\_forced-aligner/run.sh /home/jjansen/data/*

**Important**: enter \**space*\* and / when you type the route to your folder. For example:



 **Remember:** to add a *space* between the command the and data folder.

**Remember**: to start the command with / (slash)

**Note**: If the command doesn’t work please double check what you wrote (every space, slash counts).



*b.* You can also go to the folder and then just run the command. In other words: first navigate to your folder then run the command: **/vol/tensusers2/mvos/clst-asr\_forced-aligner/run.sh**

How to navigate to your folder?

Navigating to a folder can be done by typing **cd** and then the name of a subdirectory of the current directory. So for example if you are in */home/jjansen/* you could type *cd data* to get to */home/jjansen/data.*

**Note**: don’t use / but use \**space*\*, for example:



Then type: **/vol/tensusers2/mvos/clst-asr\_forced-aligner/run.sh**



→ **Useful note**: If you first navigate to your folder, you can also check your subfolders.

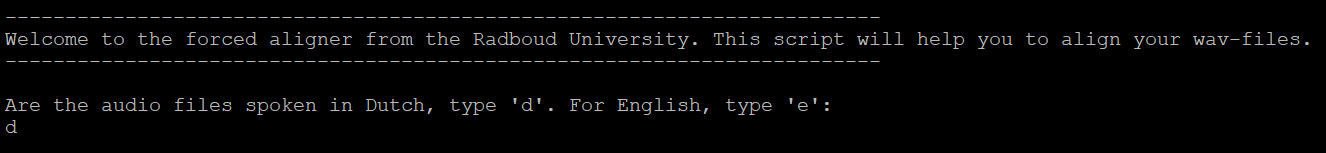
Type **ls** (stands for list) to view a list of the subdirectories that are present in the current directory. Type *cd ..* to go back to the superdirectory of the current directory.



You could also directly type *cd /home/jjansen/data/* to go to that folder.

**→** If you need help this tutorial can be useful: <https://www.youtube.com/watch?v=MKai-i4mtqA>.

1. After you completed step 1 you will see this:



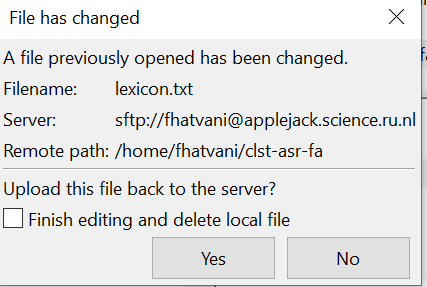
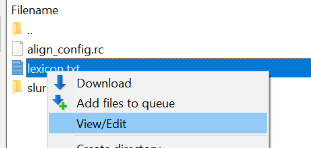
The script walks you through all the steps necessary to run the forced aligner. For this used input is required, which can be given by typing an answer (most often ‘y’ for Yes and ‘n’ for No) and then pressing enter. It will ask you, amongst other things, which lexicon to use and which configurations to use.

After answering the questions the job runs. If the job ran successfully you will see this:



**→ Missing words**:

After the jobs are run, the script will show the missing words from the lexicon. These words, along with their phoneme representation, should be added to */home/jjansen/clst-asr-fa/lexicon.txt,* which is easiest to do by navigating to this file with FileZilla, right-click on the file and clicking *view/edit*, editing it and uploading it back to the server by clicking on the pop-up in FileZilla.

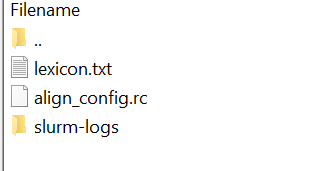
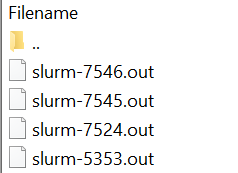


→ Edit the words in the txt file →save

These missing words can also be determined by running the script again and pressing ‘1’ when asked:

After these missing words are added, the script should be run again like in step 1, but now on the new lexicon (which will happen if you do not force it to download a new fresh lexicon).

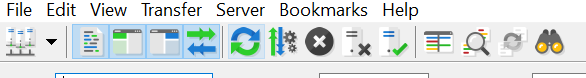
1. The script can be cancelled by pressing Ctrl+C. When the jobs are started you can see the progress by typing *squeue –u [user\_name] (squeue –u jjansen)*  into PuTTY and you can see the log files by opening the newest log file (*slurm-[digits].out*) in */home/jjansen/clst-asr-fa/slurm-logs* in FileZilla.



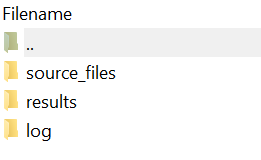
→



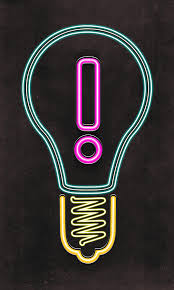
If you don’t see the log file try to refresh the file and folder lists by clicking on the icon

 This can be found in the header of Filezilla.

1. If everything has been run successfully, the specified data folder will contain three folders, namely *results* in which the resulting text grids are stored, *source-files* in which the original text grids and wav files are stored, and *log* in which log-files per audio file are stored.



1. If you want to exit just type *exit*.

→ **Important**: If something is wrong the system will inform you:

In this case go to the FileZilla and go to your folder */home/jjansen/clst-asr-fa/slurm-logs* and check what the problem is (right click on the file and choose *View/Edit*).

This should be all the information necessary for running the forced aligner (along with the information that is provided by the user when running the script, please **read first the output of the script** (in the log file, for example: */home/jjansen/clst-asr-fa/slurm-logs )*before asking questions!).

If you have any error messages, first try to Google it (type the error message into Google).

Should something be unclear, send an email to [maartenNVos@gmail.com](mailto:maartenNVos@gmail.com)