**STEP 1: CREATE A FOLDER**:

1. Open PuTTY and login:

Login as: **username**

Password: **password**

*Username: first letter of first name + family name. Lowercase letters.*

*e.g.: Piet van Dijk → pvandijk*

1. Create a folder under: /vol/tensusers2/**username.**

*Username: first letter of first name + family name. Lowercase letters.*

*e.g.: Piet van Dijk → pvandijk*

To create a folder, use command **mkdir**, e.g.

[pvandijk@applejack:/vol/tensusers2/] $ **mkdir** pvandijk

* pvandijk@applejack means you are logged in on Ponyland on applejack [computer cluster name] with account pvandijk.
* /vol/tensusers2/ means you are under this path.
* **mkdir** is the command to make a directory. To use this, follow it by a space and then the name you want to give to the folder [in this example *pvandijk*].

1. Go to the new folder and copy the data and the forced aligner to the new folder.

To go to the folder (this will change the path): use the command **cd**.

Example: [**username**@applejack:/vol/tensusers2/] $ **cd** **username**

   [**username**@applejack:/vol/tensusers2/**username**/] $

Current location: /vol/tensusers2/**username**

1. To copy the data: use command **cp**.

Example (copying forced aligner): [username@applejack:/vol/tensusers2/**username**/] $ **cp** –r /vol/tensusers/mganzeboom/clst-asr\_forced-aligner .

* **cp** = the command to copy things
* **-r** = the option for **cp** command. Here it means that you are copying a folder.
* /vol/tensusers/mganzeboom/clst-asr\_forced-aligner = the thing you want to copy.
* . (little dot) = you want to copy something to the current position/path which in this example is under /vol/tensusers2/**username**/.

Current situation: the data and the forced aligner are in your folder. You have two folders under your own folder username: clst-asr\_forced-aligner and data

**STEP 2: RUN THE FORCED ALIGNER**

General instructions, without errors:<https://www.cls.ru.nl/clst-asr/doku.php?id=forced-aligner>

Be sure: wave files + textgrid files have the same name without extensions.

*e.g. pp1\_reading.wav and pp1\_reading.Textgrid are compatible, but the combination pp1\_reading.wav and pp1\_readingstory.Textgrid is not.*

Steps to run the forced aligner (copied from website cls.ru.nl):

1. Login to one of the ponies (do not use (the old) applejack because of an older CUDA version).

2. Run from your home directory:

/vol/tensusers/mganzeboom/clst-asr\_forced-aligner/run.sh

<absolute-path-to-directory-with-recordings>.

The script will create a directory in your home directory and copy the default config and lexicon files.

3. Open ~/clst-asr-fa/align\_config.rc with your favourite editor and change the configuration settings to your liking (the defaults are fine on average).

4. Run step 2 once again and a job will be added to the Slurm queue manager starting the forced alignment of the provided directory. The logs of this job can be found in ~/clst-asr-fa/slurm-logs/slurm-<job-id>.out. Provide multiple input directories at the command line to queue multiple jobs at once.

5. The force alignment logs can be found at <absolute-path-to-directory-with-recordings>/logs when all Slurm jobs have completed.

     -aligner.tgscriptions are not in the lexicon provided with the acoustic models. Run the script /vol/tensusers/mganzeboom/clst-asr\_forced-aligner/list-missing-words.sh <absolute-path-to-directory-with-recordings> to print a list of these words and their corresponding transcription files to the command line. You could then add a phonemic transcription of these words to your custom lexicon file in ~/clst-asr-fa/lexicon.txt. It is recommended to base these new transcriptions on parts of already existing ones. Afterwards, rerun the script from step 4.

Current situation: stage 3 (of 7) is finished. Under audio folder you see a subfolder ‘log’. In this ‘log’ folder, another subfolder is created called ‘splits’.

To check in which stage you are: open the script called ‘run\_forced\_alignment.sh’. Use command **cat** (no editing in the script) or **nano** (editing and save, be careful using this) which is under path e.g. /vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/

E.g. cat: [**username**@applejack:/vol/tensusers2/**username**/] $ **cat** clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/run\_forced\_alignment.sh

E.g. **nano**:

* You want to open the script called “run\_forced\_alignment.sh” as shown in the command the light blue part run\_forced\_alignment.sh.
* But to open this, since it’s not in your current path, you need to specify the path to that script which is the purple part in the command clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/. Here you use the relative path since it’s in a subfolder in where you currently is. You could also use absolute path, it always works. In this example, the absolute path: /vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/
* You can see, the absolute path is combined with your current path (dark blue part) and the subfolder path (purple part).

**STEP 2: RUN FORCED ALIGNER MANUALLY**

1. When there is no final ali in the folder (check this with notepad), manually run the forced aligner:

/vol/tensusers2/**username/**clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/run\_forced\_alignment.sh --config ~/clst-asr-fa/align\_config.rc <absolute-path-to-directory-with-recordings>

/vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/run\_forced\_alignment.sh --config

/vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/align\_config.rc <absolute-path-to-directory-with-recordings>

Current situation: stage 4 is finished. You can see the output in the window.

8. Manually run the praat script: praat --run: /vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/align2praat/createtextgrid.praat <absolute-path-to-splits-directory-under-recordings-folder> < absolute-path-to-directory-with-recordings> <phone\_tier\_name>

Current situation: stage 5 is finished.

1. Manually run another praat script: praat --run:

/vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/align2praat/createWordTextGrids.praat <absolute-path-to-log-directory-with-recordings> <absolute-path-to-directory-with-recordings> <absolute-path-to-splits-directory-under-recordings-folder> <word\_tier\_name>

Current situation: stage 6 is finished.

1. Change stage to 7 (stage=7) in your config file in your home directory. Re-run: /vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/run\_forced\_alignment.sh –config ~/clst-asr-fa/align\_config.rc <absolute-path-to-directory-with-recordings>
2. Finally, manually run another praat script: praat --run /vol/tensusers2/**username**/clst-asr\_forced-aligner/kaldi/egs/clst-asr\_forced-aligner/s5/align2praat/stackTextGrids.praat <absolute-path-to-splits-directory-under-recordings-folder> <absolute-path-to-directory-with-recordings>

Output: text grid file named: -aligner.tg with word level & phoneme level alignments. You did it! Congratulations!

**Tips:**

**Map Ponyland on your own laptop**

If you have problems with running praat scripts in the command line, here is another option. First, map the Ponyland to your own laptop. This is easier than filezilla and looks like a windows thing.

1.       Open your file explorer.

2.       Right click “This PC”, then click “Map network drive…”

3.       In the new window, give the drive name, default is X:. Give Folder name, here is e.g. \\derpy.science.ru.nl\tensusers2\**username**

You probably need to login with your username and password. This could also work outside the campus once you connect the VPN.

Result: you see everything in that folder and editing things as you normally do in Windows.