

# Manual ASRcorrector

## General

The ASRcorrector software is meant for correcting the results of the first transcription, done via Automatic Speech Recognition (ASR). To do so, the CTM-results of an ASR are read by the software and converted into a dedicated XML-file. This XML-file contains the automatic generated transcriptions, the by a human operator converted transcriptions, the recognised audio- or video-file and the progress status (the moment in the file where you were working just before saving the file).

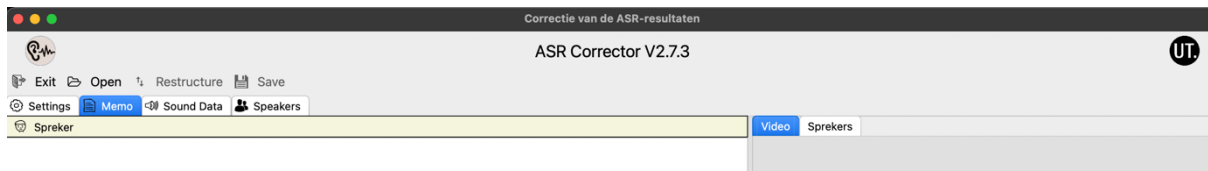


Figure 1: The view on the ASRcorrector if you start the software. The version showed (V 2.7.3) contains a mix of Dutch and English titles/headers. Once the software is “ready” we will change it into real English 😊

Once the corrections are done, the results can be exported to a special text file for the Forced Alignment: a different service meant to align the spoken and written text.

## Introduction

With the use of Deep Neural Networks (DNN) in modern ASR, speech recognition became good. I.e. most of the times our Dutch and English recognition engines reaches a better than 90% correct recognition of the words spoken. This means that in a common interview, it makes sense to use the ASR results as the basic input text that needs to be checked and eventually corrected.

In a classical transcription task, it takes between 6 and 8 hours to transcribe the spoken content in a verbatim way. However, with the combination of ASR and correction of the ASR-results, this drops to 2 to 4 hours (depending of course on the quality of the audio, the clearness and correctness of the spoken speech, etc. etc.).

Moreover, the software makes it possible to add the different speakers, to replace one or more words by another word/ words, to make phrases, add signs for unfinished phrases endings and more.

## Start of ASRcorrector

### SHOW EXPERT MODE

The first time you open ASRcorrector, you will see something like image 1 above: a rather empty screen. To read the CTM-file (the result of the ASR-engine), one goes to settings and clicks on “Toon expertmode” (=Show expert mode). As a result, you see a number of additions (see fig.2).

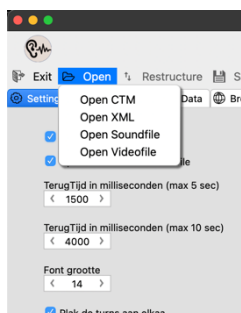


Figure 3: Open the CTM-file

Do not change the other fields the first time you open ASRcorrector!

### READ CTM-FILE

If you click now on the **Open** button, you can read the CTM- or CTM.SPK-file.

After you read the ctm-file, open or the audio or the video-file. Then, save the file as an XML-file. For the correction phase, you can now deselect the “Show Expertmode” checkbox.

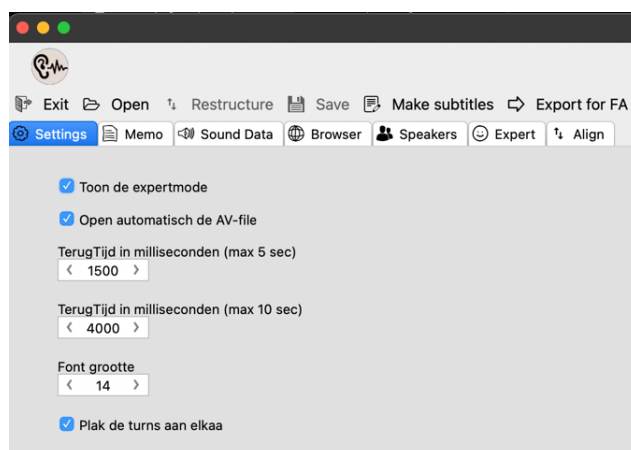


Figure 2: For the reading of the CTM-file, you need to click on “Toon expertmode” in the settings tab

## Working with the XML-file

After the initial phase where you read the CTM-file and store it, together with the audio/video file in the XML-file. For the correction phase, this XML-file is the working file.

### ADDING SPEAKERS

The next thing that needs to be done, is the addition of the different speakers. Click on the tab Speakers and you will see a screen as in figure 4.

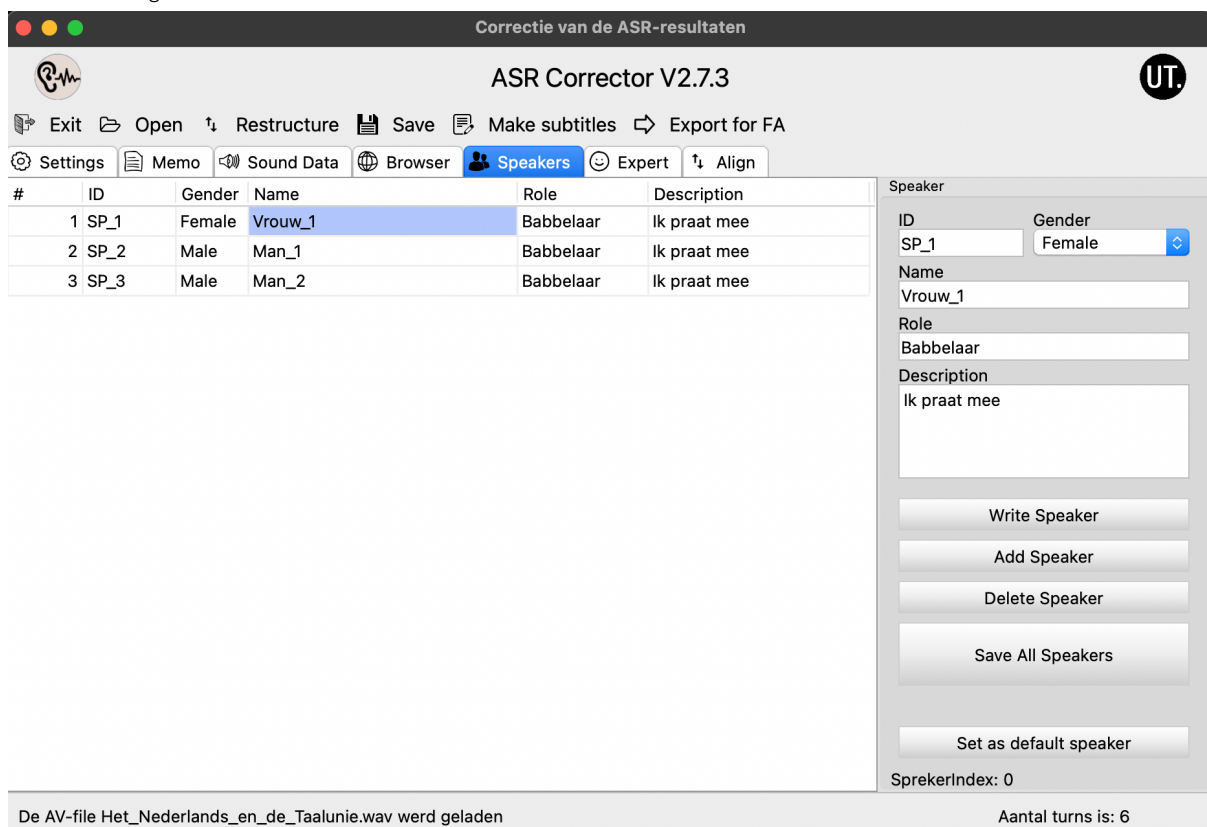


Figure 4: Addition of speakers.

For the addition of a new speaker, click on **Add Speaker**. As the ID you will get SP\_1, gender is set to other and name is set to male\_x or female\_x. Role and Description can be used for additional information. After filling in these parameters, click on **Write Speaker**. The results will be added on the left-side of the panel. Once you have added all the speakers, click on **Save All Speakers**.

### CORRECTION

Click on the **Memo** tab to start the correction of the ASR results. In the field on the left part of the panel, you see the recognised text.

Please correct this text.

Below this text-field, you see the audio buttons, bounded on both sides by the start and end time of this fragment. The 4 buttons in the centre are for playing/pausing the sound-file, limited by the start and end-time. You can use the bar as well (especially for longer fragments) to set the beginning of the audio-fragment.

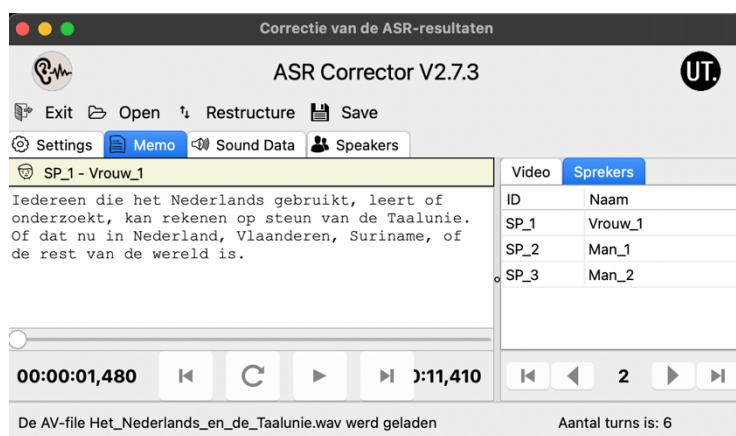


Figure 5: the correction window.

On the panel on the right, you see 2 tabs: **Video** and **Speakers**. If there is a video, you see that video in the video-tab. On the right-bottom you see the buttons for the chunks or turns (fragments of the audio/video). These chunks are set by the

input/CTM-file. If you select a \*.ctm.spk file, the turns of the speakers are used as input. If you use a \*.ctm file, each chunk will contain 100 words. The buttons are -10, -1, +1 and +10 to the right or the left.

## SPEAKER SELECTION

There are two ways of setting a speaker. The first one is to set the speaker for the chunk. You do this by right-clicking on the **head** above the text-field and then selecting the right speaker.

However, it happens often that you need to add another speaker somewhere in the text field when someone else starts to speak.

To do so, go to position in the text where you want to add a new speaker. Right-click with your mouse

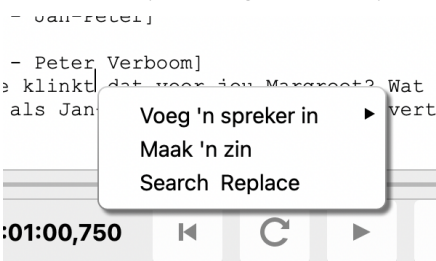


Figure 7: Adding another speaker in the middle of the text of the turn.

## COMBINING TURNS

It may happen that you want to combine the turn you are working on, with the previous or next turn. This occurs especially when the turns contain the same main speaker and are rather small.

To do so, right-click on the turn-number in the right-bottom window and select **prev** or **next**. The turns are then glued together, and the text is combined in one text. For an improved workability, do not make the turns too great. Turns with more or less a 100 words turn out to be ideal for the correction tasks.

## MAKING SENTENCES

As explained above, the ASR returns only words with small-caps and no punctuation marks. In order to make the reading of the ASR-results better/easier, we have made a phrase detector that starts a new phrase after each pause of 400 msec or longer. It turns out that this “400 sec boundary” works most of the time quite well. However, there are speakers and/or situations where people hesitate in the middle of a phrase, causing the phrase generator to add too many phrases and, on the other hand, speakers can talk fast so that the software sees no difference between two consecutive utterances and avoids turning them into two sentences.

Given a correct ASR-result, one can split the text into two sentences by right-clicking at the space between the two sentences and selecting “Making a sentence” (or Maak ‘n zin).

## SEARCH AND REPLACE WORDS

Due to the absence of all upper-case letters, geographical and personal names often have to be changed one by one. To simplify this process, we have created a simple search and replace routine that replaces all occurrences of a word or word combination with another word or word combination. To do so, select the word(s) to be replaced, right-click the selection, and select **Search & Replace**. The software copies the word in both the search and the replace field. Clicking Ok, results in a document wide replacement of the original word by the new word.

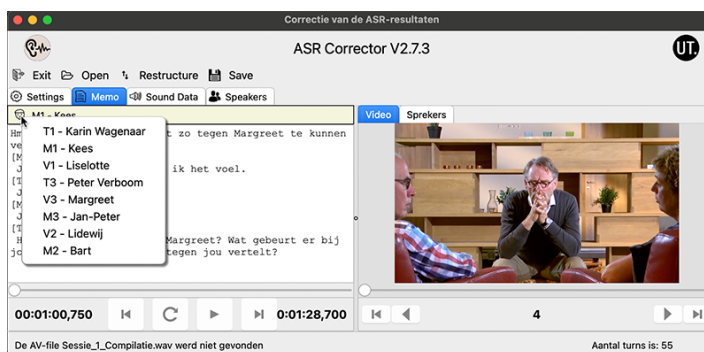


Figure 6: Adding the main speaker of a turn.

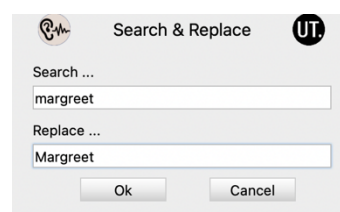
and select the upper item (**Add a speaker or Voeg ‘n spreker in**). Then, select again the right speaker.

The software adds the speaker with its ID and name (for example: [T3 - Peter Verboom]).

The hard returns do not have a meaning; they are just added to improve the readability of the text.



Figure 8: Glueing two adjacent turns together.



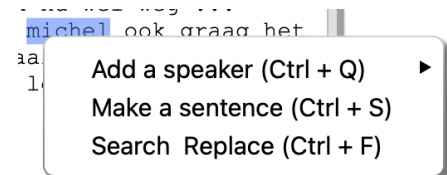
## Things-To-Do

It is difficult to give a complete set of instructions of things that must necessarily be done. However, we can make some suggestions to ensure that the transcriptions are done as well and as uniformly as possible.

### NAMES

Names are not recognised as such by the speech recogniser and therefore appear as "normal words" in the transcription. If for example michel is said once, then you can simply replace this michel by Michel. However, if you think that michel occurs more often, then it might be more convenient to select michel first and then click the right mouse button.

You are seeing the next image:



Then choose the "Search Replace" button, replace michel with Michel and click OK.

The programme will now replace all michel's with Michel.

Of course, you can also use this to replace other words and word combinations, such as "zuid holland" with "Zuid-Holland". It works well, but you have to be careful with words that are not only names but also normal words. Think for example of "former prime minister Kok" and "de kok in het restaurant". Here, unfortunately, you cannot simply replace "kok" with "Kok".

### PHRASES

Make sure you start every sentence with a capital letter. And remember to end an incomplete sentence with three dots at the end. For example, "and then I thought...". With those three dots you indicate that the sentence ends but that part of the text is actually missing.

You can also make sure that the sentences are not too long. So, even if you can make it one long sentence, it is preferable to split it into two shorter ones.

#### Example:

*"We're still working on it every **day and** that's what I want to give people as well that most people who are going to look at this now will also agree with me that it's never finished."*

Can also be written as:

*"We are still working on it every **day. And** that's what I want to give people, too, that most people who go and see it now will agree with me that it's never finished."*

### NUMBERS

Many numbers are very long (three hundred and sixty-six thousand two hundred and ninety-eight euros) and can often be written down as a number: €366298. But... this does not always have to be done. Especially in the case of smaller numbers, you may prefer just to write them out. However, try to be as consistent as possible with this.