



SPEECHMATICS ALIGNMENT INSTRUCTIONS

This document gives user instructions on how to use the Speechmatics alignment API.

How it works

The new alignment products from Speechmatics allow a user to supply an audio file and a text file.

Speechmatics will then return the original text file with added mark-up giving timings (e.g., '<time=12.345>').

This allows you to find out exactly when each word in your text was spoken in your audio file.

Submitting an alignment job

Submitting an alignment job using curl is very similar to submitting a transcription job, except that you will specify a 'text_file' as well as a 'data_file'. For example:

```
curl -F data_file=@my_audio_file.mp3 -F text_file=@my_text_file.txt -F model=en-US  
"https://api.speechmatics.com/v1.0/user/$MY_API_USER_ID/jobs/?auth_token=$MY_API_A  
UTH_TOKEN"
```

Supported alignment language products are listed on this page:

<https://www.speechmatics.com/support>

Retrieving an alignment job

As with transcription, once an alignment job has finished you will be emailed (or if you have specified, a url callback made instead). You can then retrieve the result of your job in a similar way as you would retrieve a transcription job, substituting the word 'alignment' for 'transcript':

```
curl  
"https://api.speechmatics.com/v1.0/user/$MY_API_USER_ID/jobs/$MY_JOB_ID/alignment?a  
uth_token=$MY_API_AUTH_TOKEN"
```

The result will be your text file returned to you with time tags included. We will not change or remove anything in your original text file – we will just add the tags.

If you would like only one tag per line, you can specify this with the 'tags' argument:

```
curl "https://api.speechmatics.com/v1.0/user/$MY_API_USER_ID/jobs/$MY_JOB_ID/  
alignment?tags=one_per_line&auth_token=$MY_API_AUTH_TOKEN"
```

Failed jobs

Sometimes it will not be possible to align your text to your audio. This may be because the audio is unclear, or the text is not a true representation of what was said in the audio.

If this occurs your alignment will fail. You will be notified of this in your notification email, and if you attempt to retrieve the job you will be informed of the failure to align. Getting the job status will return a job_status=could_not_align.

Your credits will be reimbursed if your job cannot be aligned.

Supported file formats

As with our transcription products, we support most audio file types, including aac, aif, m4a, mov, mp3, mp4, wav and more! If your audio file is not supported, your api call will return with job_status 'rejected'.

At present we expect all text files to only contain UTF-8 (<https://en.wikipedia.org/wiki/UTF-8>) and we will reject any text files that we detect in another file format.

Text formatting

Input

The first step in alignment is extracting what we believe are words from your text file. We define a word as any string of characters separated by whitespace characters (space, tab, new line, etc.).

If there is any markup in your file we assume it corresponds to SGML with angle brackets as start- and end- tag delimiters. We therefore assume that nothing between comment delimiters ('<!--', '-->') or angle brackets ('<', '>') should be considered as a word.

So for a file containing this text:

```
'Hello <markup> world <!-- comment > comment --> how are you?'
```

We would attempt to align the following words to your audio:

```
'Hello' 'world' 'how' 'are' 'you?'
```

Output

As stated earlier, the only change we make to the input is to add timing tags.

From the earlier example the output then may look something like this:

```
Output text: '<time=0.12>Hello<time=0.23> <markup> <time=0.34>world<time=0.45> <!--  
comment > comment --> <time=0.56>how<time=0.67> <time=0.78>are<time=0.89>  
<time=0.90>you?<time=1.00>''
```

Each time tag should be considered self-closing and the number refers to the number of seconds from the start of the audio file.

If you have chosen to only have one tag per line the output would instead look like this:

```
[00:00:00.1] Hello <markup> world <!-- comment > comment --> how are you?
```

Cost

As with transcription you pay for alignment using credits. Your existing credits are useable for both transcription and alignment.

Currently, alignment will cost 1 credit per 15s, with a minimum of 4 credits (1 minute) per audio file. The number of words in the text file will not affect the cost. This is two thirds of the cost of transcription.

You can buy credits either through the Speechmatics website or by contacting us.

If your alignment does not succeed, we will reimburse your credits.



Support and feedback

If you have any feedback or questions, please don't hesitate to contact us at support@speechmatics.com. With your help, we can make this product even better.