Speechmatics

The Speechmatics Cookbook

Take your first step into speech-to-text using our handy how-to guide for Developers and Product Leaders.

Speechmatics.com

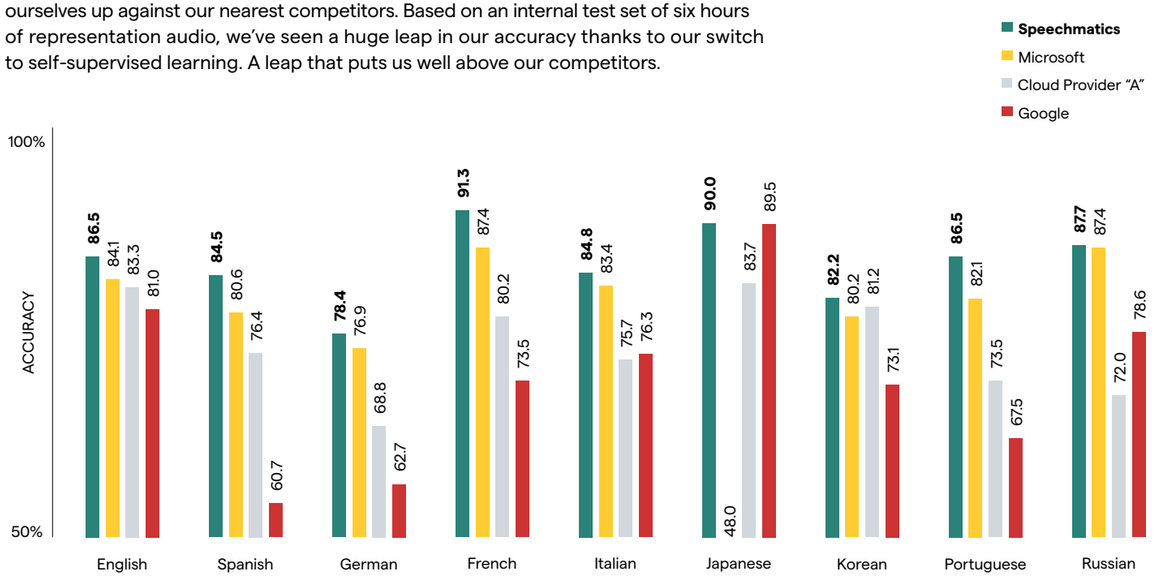
Contents

Intro

Welcome to the Speechmatics Cookbook. Over the following pages you'll find a wealth of information to get you started with speech-to-text. Whether you're a Developer, Product Leader, or someone looking to integrate speech-to-text into your product, this document offers a helpful overview of our product, its benefits, and how you can use it.

As experts in deep learning, we at Speechmatics have a mindset that voice technology should accurately represent the world around us. This has enabled us to build the most accurate and inclusive speech-to-text engine available to customers and businesses anywhere in the world.

To show how we offer the best voice-to-text transcription on the market, we put ourselves up against our nearest competitors. Based on an internal test set of six hours of representation audio, we've seen a huge leap in our accuracy thanks to our switch to self-supervised learning. A leap that puts us well above our competitors.



Step 1: Choose Your Deployment Options

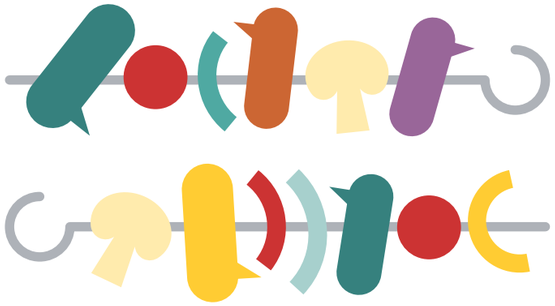
One of the first questions you'll need to answer is, how do you want to use Speechmatics? We have four deployment options to choose from, Virtual Appliance, Container, SaaS, and Hybrid.

Here's a brief breakdown of each, including how we use and store your data:

Virtual Appliance

The Speechmatics Virtual Appliance is a pre-configured virtual machine capable of doing Real-Time or Batch processing. Supported on multiple platforms, the Virtual Appliance can be deployed directly in your on-premises environment. No need to worry about maintenance or scaling as this happens automatically.

For Real-Time offerings no data is stored during the speech-to-text process. Batch contains a clean-up policy which automatically deletes the transcripts after a configurable period of time. The default is 24-hours.



Containers

The Speechmatics Docker Containers enable you to build scalable transcription services within your own infrastructure in Real-Time or Batch processing.

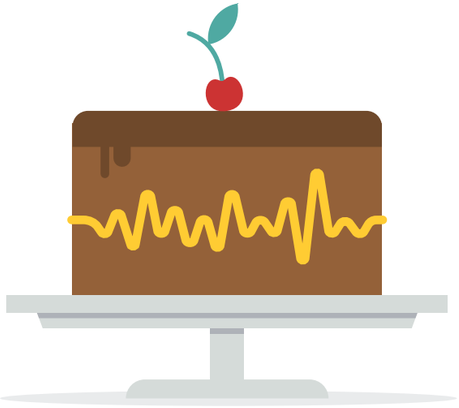
The Container does not store any audio or transcripts, making it easy to use within secure environments, and to maintain any audio and transcripts within the customer's own security boundaries.

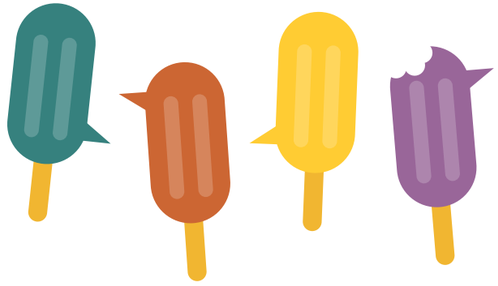
SaaS

Our SaaS delivers all the benefits of the Speechmatics ASR, without the complexities of deploying within your own team and environment. You can choose from the public cloud – hosted by Speechmatics – or your own cloud environment.

Speed up your time-to-market with a secure service and instant access to our new features, languages, and updates.

- All features available (see page 7 & 8).
- All languages available (see page 10).
- Available for pre-recorded (Batch) media files.
- Open and extensively documented APIs for integration and operational simplicity.





Hybrid

Hybrid deployment may also be right for you if you have a mixture of data requirements that use both cloud and on-premises processing. The close similarity between the APIs removes any operational overheads and complexities.

Step 2: Choose Your Offering

This section gives you a breakdown of both our offerings – Real-Time and Batch – as well as notes on their performance across the differing deployment options.

Batch

Transcribe pre-recorded media files at your convenience with Speechmatics' most powerful, accurate, and inclusive engine ever released. Schedule a transcription at a time that suits you to help optimize your available resource.

Docker Container:

- A transcript can be provided in 2x real-time for files >5 minutes.
- Parallelization has the potential to have faster turnaround times.
- 1vCPU per concurrent transcription job.
- Multiple containers can be executed on the same host at the same time or across many hosts to enable large scale operations.

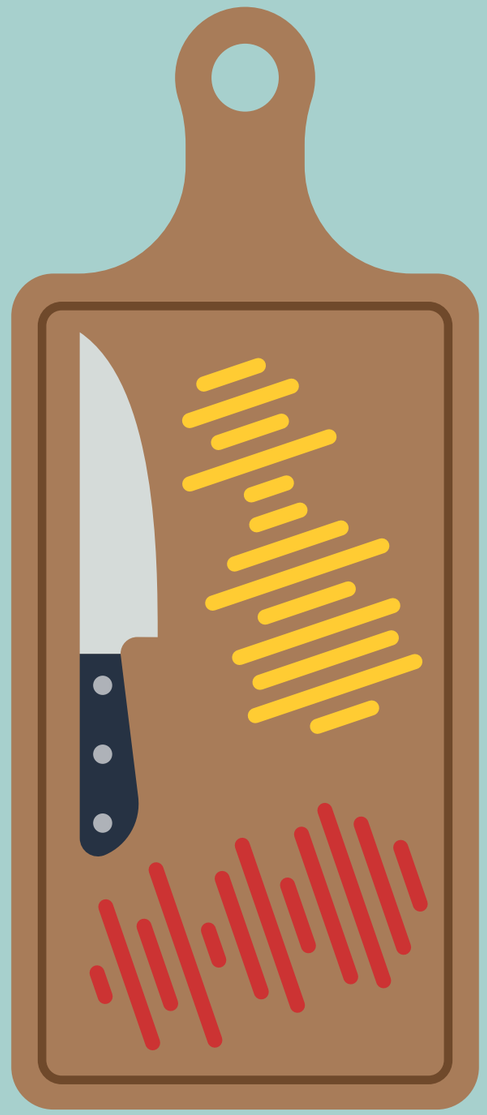
Virtual Appliance:

- A transcript can be provided in 2x real-time for files >5 minutes.
- 1vCPU per concurrent transcription job.
- Additional vCPUs can be added to enable multiple media files to be transcribed at the same time.
- The real-time factor is calculated from when the file starts being processed to its completion.

SaaS:

- A transcript can be provided in 2x real-time for files >5 minutes.
- The real-time factor is calculated from when the file starts being processed to its completion.

Find out more about our Batch offering by reading our Product Sheets at **speechmatics.com**



Real-Time

Transcribe in real-time and get results instantly with Speechmatics' most powerful, accurate, and inclusive engine ever released. Gather actionable data as soon as it's needed. Our proprietary technology delivers best-in-class accuracy even at low latencies.

Docker Container:

- 1vCPU per stream allows a transcript to be provided in real-time.
- Parallelization has the potential to have faster turnaround times.
- Multiple containers can be executed on the same host at the same time or across many hosts to enable large scale operations.

Virtual Appliance:

- 1vCPU per stream allows a transcript to be provided in real-time.
- Additional vCPUs can be added to the Virtual Appliance to enable multiple streams to be concurrently transcribed (in the languages you require).

Find out more about our Real-Time offering by reading our Product Sheets at **speechmatics.com**

Step 3: Choose Your Features

It's not just world-leading accuracy that sets us apart from the competition. We have a series of great features to make sure the transcripts you receive are as beneficial to you as possible.

Confidence Scores
Visualize the confidence of every word in the transcript.

Entity Formatting
Improve the professionalism of your transcripts with numeral recognition.

Speaker Diarization
Detect and label different speakers within the same channel.

Channel Diarization
Detect and label different speakers on up to six streams or channels.

All Major File Formats Supported
Support all major audio and video formats so you can reduce the time it takes to prepare files.

Advanced Punctuation
Use an extensive set of supported punctuation marks to optimize the speed and ease of transcription.

Custom Dictionary and Sounds Feature
Add a set of context-specific words to the dictionary to enhance your transcription accuracy.

Speaker Change
Easily identify a change of speaker within your transcript and improve its readability.

Notifications
Use callback to receive a notification when your job is complete. The notification can also include your transcription output.

Profanity Tagging
Words are automatically tagged as a profanity in the transcription JSON output for use in post-processing.

Disfluencies
Words are automatically tagged which are considered to be hesitation or indecision in transcription JSON output for use in post-processing.

Partials
Transcriptions are returned as soon as transcript data is available, without the need to wait for additional context.

Transcript Finalization
Provides highly accurate transcripts and can automatically correct words to match the given context.

Low Latency Finals
Define the context of transcriptions and use it to automatically correct words.

Flexible Endpointing
Ensure output formatting is kept consistent by flexibly overriding when a transcription's finals are returned.

Note: Some features are exclusively available in Batch, some in Real-Time. See website for more information: **speechmatics.com**

Step 4: Choose Your Languages

Our industry-leading language coverage ensures our technology can handle your business needs. Your ability to use any or all the languages will depend on what languages you're contracted to use.

Speechmatics supports the following languages:

Language	Language Code
Arabic	(ar)
Bulgarian	(bg)
Catalan	(ca)
Cantonese	(yue)
Croatian	(hr)
Czech	(cs)
Danish	(da)
Dutch	(nl)
English	(en)
Finnish	(fi)
French	(fr)
German	(de)
Greek	(el)
Hindi	(hi)
Hungarian	(hu)
Indonesian	(id)
Italian	(it)

Language	Language Code
Japanese	(ja)
Korean	(ko)
Latvian	(lv)
Lithuanian	(lt)
Malay	(ms)
Mandarin (simplified and traditional)	(cmn)
Norwegian	(no)
Polish	(pl)
Portuguese	(pt)
Romanian	(ro)
Russian	(ru)
Slovakian	(sk)
Slovenian	(sl)
Spanish	(es)
Swedish	(sv)
Turkish	(tr)

Languages outside this list are not yet supported. We can only accept one language within each request. Please provide the two-letter ISO639-1 code (above) with your transcription request (see page 11).

Step 5: Check Your Systems

Here's a brief breakdown of what you'll need to run the various offerings in differing deployment options.

Docker Container

Minimum specification: Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent).

Enhanced operating point - for optimal performance: Intel Cascade Lake with AVX512 VNNI support.

Linux Docker runtime host must be Advanced Vector Extension (AVX) compatible.

We strongly recommend AVX2 compatible hardware to take advantage of the latest performance improvements.

Compute requirement for Batch: An individual Docker image is required for each transcription language. Each running container requires: 1 vCPU, 2-5GB RAM, 100MB hard disk space.

Compute requirement for Real-Time: The container supports a single audio stream with the same or smaller footprint than our existing (appliance based) real-time containers, current footprint: 1 vCPU per container. 1.5GB RAM per container (default config). 3GB RAM per container (with Custom Dictionary).

Virtual Appliance

Minimum specification: Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent).

Enhanced operating point - for optimal performance: Intel Cascade Lake with AVX512 VNNI support.

At minimum we support AVX but we strongly recommend AVX2 compatible hardware to take advantage of the latest performance improvements.

Hypervisor: Oracle VirtualBox, VMWare ESXi 6.5 and onward, VMWare Workstation, Amazon Web Services EC2.

Compute requirement for Batch: Base config – 2 vCPU, 8GB RAM this will process approximately 2 hours of audio per hour. Additional resources: 1 vCPU, up to 5GB RAM for every additional worker.

Compute requirement for Real-Time: Each virtual machine: Base config: 2 vCPU, 8GB RAM required to process one continuous audio stream. Additional resources: 1 vCPU, up to 3GB RAM for ever additional worker able to process a continuous audio stream.

SaaS

Our production environment is hosted in both Western Europe and Western USA. Our trial environment is hosted in Western Europe.

Step 6: Start Your Project

Below is a very basic API overview of how to:

- Sign-up to Speechmatics SaaS Portal.
- Submit an audio file for batch transcription.
- Check the status of a job.
- Retrieve your final transcript in a supported format.

How to Submit

The simplest configuration for a transcription job is to specify:

- The `Type` of request you want. This is always `transcription` .
- The `language` you wish to use. This is a two-digit language code following ISO639-1 format. It must be one of the language codes supported by Speechmatics (see page 10)

Below is an example of a basic configuration:

```
config='{
    "type": "transcription",
    "transcription_config": { "language": "en" }
  }'
```

[Click here](#) to read our full API documentation.

In the examples below, job configuration is enclosed within a file called config.json.
This file can be extended or modified as you wish, when using Speechmatics additional features.

Submitting an Audio File

Unix/Ubuntu

```
curl -L -X POST https://asr.api.speechmatics.com/v2/jobs/ -H "Authorization: Bearer NDFj0TE3NGEt0WVm" -F data_file=@example.wav -F config="$(cat config.json)" | jq
```

Windows

```
curl.exe -L -X POST https://asr.api.speechmatics.com/v2/jobs/ -H "Authorization: Bearer NDFj0TE3NGEt0WVm" -F data_file=@example.wav -F config="<config.json" | jq
```

Check Job Status

To make a `GET` request to check the status of a job:

Unix/Ubuntu

```
curl -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

Windows

```
curl.exe -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

The supported default transcription output is JSON. Other formats supported are srt (SubRip subtitle format) and txt (plain text). The format is set using the format query string parameter.

Below are examples for retrieving a transcript in TXT format:

Unix/Ubuntu

```
curl -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i/transcript?format=txt -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

Windows

```
curl.exe -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i/transcript?format=txt -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

Receive a Transcript

Transcripts can be retrieved as follows:

Unix/Ubuntu

```
curl -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i/transcript  -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

Windows

```
curl.exe -L -X GET https://asr.api.speechmatics.com/v2/jobs/d1hsd8d69i/transcript  -H "Authorization: Bearer NDFj0TE3NGEt0WVm" | jq
```

Ready to Try Speechmatics?

Sign up for your free trial to start converting your audio in to text using our well-documented APIs. We pride ourselves on offering the best support for your business needs.

Try for Free

If you have any questions, just ask.

Contact Us

For any other questions or comments, call or send us an email.

Our office is open between 9am-5pm.

Speak to Sales

Phone

UK/Europe +44 (0)1223 907 818

USA/Canada +1 866 791 8546

UK office

Cambridge Science Park
Milton Road
Cambridge
CB4 0WD
United Kingdom