



**SPEECHMATICS**

SPEECHMATICS  
**PRODUCT SHEET**

# Batch ASR

TRANSCRIPTION OF PRE-RECORDED  
AUDIO AND VIDEO AVAILABLE IN  
FLEXIBLE DEPLOYMENT OPTIONS

[WWW.SPEECHMATICS.COM](http://WWW.SPEECHMATICS.COM)

# Batch ASR

Transcription of pre-recorded audio and video available in flexible deployment options

Extensive deployment flexibility enables the integration of Speechmatics ASR on premise, ensuring data remains within your private environment or within your choice of cloud provider. In addition, Speechmatics cloud offering offers a fully maintained, operated and managed solution.

	Docker Container	Virtual Appliance	Cloud Offering
Overview	The software comprises proprietary code, language models and open source software. The proprietary code is restricted to the core automatic speech recognition engine.		
Deployment method	<p>Transcription is provided after an audio file is passed to the Speechmatics engine within the container.</p> <p>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.</p>	<p>Provides transcription of pre-recorded audio in a standalone environment provided directly by a virtual machine.</p> <p>The virtual appliance contains a clean-up policy which automatically deletes the transcripts after a configurable period of time. Default 24 hour period.</p>	<p>Transcription is provided after the audio/video file is submitted to the Speechmatics engine which is deployed within a hosted environment, managed and operated by Speechmatics.</p> <p>The cloud offering accelerates time to market, while reducing operational complexities and cost.</p>
Description	A lightweight, stand-alone software package that includes everything needed to run it: code, runtime, system tools, system libraries, settings. Requires a Linux Docker runtime.	Preconfigured virtual machine easy to configure in a few clicks.	A low risk, high reward approach to highly accurate transcription. Deployed within a public cloud environment. Fully managed and operated by Speechmatics.
Supported languages	Supported languages for each deployment mode available from the <a href="#">Speechmatics language support webpage</a>		
Supported operating environment	<p><b>Minimum specification:</b> Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent). Linux Docker runtime host must be Advanced Vector Extension (AVX) compatible.</p>	<p><b>Minimum specification:</b> Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent) that is AVX compatible.</p> <p>Hypervisor: Oracle VirtualBox/VMWare ESXi / Workstation, Amazon Web Services EC2.</p>	<p>Production environment hosting:</p> <ul style="list-style-type: none"> <li>West Europe</li> <li>West US</li> </ul> <p>Trial environment hosting:</p> <ul style="list-style-type: none"> <li>West Europe</li> </ul>
Compute requirement	<p>An individual Docker image is required for each transcription language.</p> <p>Each running container requires: 1 vCPU, 2-5GB RAM, 100MB hard disk space.</p>	<p>Base config – 2 vCPU, 8GB RAM this will process approximately 2 hours of audio per hour.</p> <p>Additional resources: 1 vCPU, up to 5 GB RAM for every additional worker.</p>	N/A
Management interface	Standard Docker, Kubernetes or other orchestration tools.	HTTPS REST API to manage the appliance including license config, log collection, scale config. For additional information see documentation.	N/A
Speech interface	Console/STDIO	REST API	REST API
Input file formats	wav, mp3, m4a, aac, ogg, flac, wma, mpeg, amr, caf, mp4, mov, wmv, mpeg, mpg, m4v, avi, flv, mkv Other formats are not explicitly supported and should be validated as part of customer acceptance.		
Transcription formats	16 kHz (Broadcast) and 8 kHz (Telephony) acoustic model built in, with automatic selection based on file sample rate		
Outputs format	JSON*, TXT**		JSON*, TXT**, SRT

\*Transcripts provided in JSON which include metadata such as timing and confidence scores, speaker/channel labels plus more on a per word basis.

\*\*TXT output does not include timing information and confidence scores.



**SPEECHMATICS**

# Batch ASR

Transcription of pre-recorded audio and video available in flexible deployment options

Extensive deployment flexibility enables the integration of Speechmatics ASR on premise, ensuring data remains within your private environment or within your choice of cloud provider. In addition, Speechmatics cloud offering offers a fully maintained, operated and managed solution.

	Docker Container	Virtual Appliance			Cloud Offering
<b>Resource requirements</b>	An individual Docker image is required for each transcription language. Each running container requires: 1 vCPU, 2-5GB RAM, 100MB hard disk space.	Pack	Supported languages	Appliance size	N/A
		Nano	1	40GB	
		Mini	3	40GB	
		Midi	8	60GB	
		Maxi	15	80GB	
		Plus	17	80GB	
<b>Performance</b>	Transcript can be provided in better than 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 Docker engine at the same time or across multiple Docker engines to enable large scale operations.	Transcript can be provided in better than 2x real-time. 1vCPU per concurrent transcription job. Additional vCPUs can be added to enable multiple streams to be transcribed at the same time. Performance figures are valid for 'good' audio. Very noisy audio may result in longer times.			Transcript can be provided in 2x real-time. Example: transcript of a 60-minute file is transcribed in 30 minutes.
	Performance figures are valid for 'good' audio. Very noisy audio may result in longer times. Enabling diarization may slightly increase turnaround time.				
<b>Data retention period</b>	0	Once the transcription is complete the transcript will be retained by the virtual appliance for 24 hours. The media file removed immediately after the transcription has completed.			7 days
	After this period, any media and transcription data will be purged unless already done so by the customer using the API.				
<b>Limitations</b>	Can operate on input file sizes of up to 2 hours recorded length or 4GB in size, whichever is reached first.				Can operate on input file sizes up to 12 hours recorded length or 4GB in size, whichever is reached first.
<b>Connectivity requirements</b>	Can operate within own security boundary allowing you to keep control of your own data.				HTTPS port 443 needed to enable access to the cloud service. For connectivity requirements if call back is used ingress port will need to be enabled.
<b>Admin</b>	No ongoing maintenance needed for the containers. All administration is provided by direct use of Docker commands.	<b>Full APIs for management and monitoring.</b> Monitoring of the appliance can be done via a web GUI or APIs. Flexible licensing model. Local administration available via APIs.			All administration of the cloud is managed by Speechmatics.
<b>Support</b>	Latest release (N) and previous release of the Speechmatics product (N-1).				



**SPEECHMATICS**

# Batch ASR

## Feature Matrix

Features	Description	Optional Vs default	Docker Container	Virtual Appliance	Cloud Offering
Confidence Scores	The word confidence score is a measure of the cumulative score of the actual output expressed as a percentage of the cumulative score of all the words considered in those time frames.	Default	Yes	Yes	Yes
Timings	Start and end time of each word (JSON).	Default	Yes	Yes	Yes
Sentence boundary information	Defined by periods (full stops) available for all languages (JSON). Note: periods, question marks and exclamation marks (?! ) are used to dictate the end of sentence.	Default	Yes	Yes	Yes
Custom Dictionary	Custom Dictionary allows up to 1,000 additional words that can be added to the standard dictionary per input stream. Allows users to quickly add context-specific words, for example company names, place names or foreign words, proper nouns, acronyms, and abbreviations. Custom Dictionary Sounds is extension which allows alternate spellings, pronunciations, acronyms and abbreviations to be used.	Optional	Yes	Yes	Yes
Advanced Punctuation	Support of periods (full stops), commas, exclamation marks and question marks.	Default	Yes*	Yes*	Yes*
Advanced Punctuation overrides	Customer configurable option to choose which punctuation characters are displayed in the output.	Optional	Yes*	Yes*	Yes*
Speaker Change (BETA)	Token passed in the output on sound wave analysis providing unique time and duration of each speaker turn.	Optional	Yes	Yes	Yes
Speaker Diarization	Ability to detect and label speakers based on multiple speakers within the same channel.	Optional	Yes	Yes	Yes
Channel Diarization	Supports multiple speakers on separate channels and streams. Up to 6 streams/channels.	Optional	Yes	Yes	Yes
Output Locale	Specify rules for transcription output for the Global English (en) language pack to specify American or British spellings.	Optional	Yes*	Yes*	Yes*
Tracking Metadata (via JSON)	Attach richer metadata to a job using the tracking configuration. This allows users to track a job through their own management workflow using whatever information is relevant.	Optional	No	Yes	Yes
Notification Callbacks	A lightweight signal indicating that results are ready to be fetched through the normal API without needing to poll.	Optional	No	Yes	Yes
TLS for API access	All traffic sent to Speechmatics' ASR deployment of choice is secure during transit using HTTPS and encryption. It supports the TLS 1.2 protocol and 256bit encryption.	Default	N/A	Yes	Yes
Parallelization	Capability for customers to apply any additional CPU resources to processes multiple audio chunks in parallel for faster transcription turnaround time.	Optional	Yes	No	No
URL Fetching	If you store your digital media in cloud storage (for example AWS S3 or Azure Blob Storage) you can also submit a job by providing the URL of the audio file.	Optional	No	No	Yes

\*Only for certain languages.

Note: Speaker and Channel Diarization cannot be used together.





## Try Speechmatics

See how it could work for your business:

- Test against your use case
- Check accuracy and language coverage
- Deploy to suit your business needs and scale with growing demand



## Get in Touch

For more information

- Sales (UK): +44 (0)1223 907 818
- Sales (US/Canada): +1 866 791 8546

Want to know more about how Speechmatics can help you *innovate with voice*? [speak to an expert.](#)



Speechmatics® powers applications that require mission-critical, accurate speech recognition through its any-context speech recognition engine.

Speechmatics' speech recognition technology is used by enterprises in scenarios such as contact centers, CRM, consumer electronics, security, media & entertainment and software. Speechmatics processes millions of hours of transcription worldwide every month in 30+ languages.

Having pioneered machine learning voice engineering, Speechmatics is enabling companies to build applications that detect and transcribe voice in any context and in real-time. Its neural networks consider acoustics, languages, dialects, multiple speakers, punctuation, capitalization, context and implicit meanings.



**SPEECHMATICS**