SPEECHMATICS

SPEECHMATICS
PRODUCT SHEET

Batch ASR

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

BATCH ASR

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

SPEECHMATICS

Extensive deployment flexibility enables the integration of Speechmatics' ASR on-premises, 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	Transcription is provided after an audio file is passed to the Speechmatics engine within the container. 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.	Provides transcription of pre-recorded audio in a standalone environment provided directly by a virtual machine. The virtual appliance contains a clean-up policy which automatically deletes the transcripts after a configurable period of time. Default 24-hour period.	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. The Cloud offering accelerates time to market, while reducing operational complexities and cost.
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 Speechmatics language support webpage.		
SUPPORTED OPERATING ENVIRONMENT	**Minimum specification:** Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent). Linux Docker runtime host must be Advanced Vector Extension (AVX) compatible. We strongly recommend AVX2 compatible hardware to take advantage of latest performance improvements.	**Minimum specification:** Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent). At minimum we support AVX but we strongly recommend AVX2 compatible hardware to take advantage of latest performance improvements. **Hypervisor:** Oracle VirtualBox, VMWare ESXi 6.5 and onward, VMWare Workstation, Amazon Web Services EC2.	Production environment hosting: • West Europe • West US Trial environment hosting: • West Europe
COMPUTE REQUIREMENT	An individual Docker image is required for each transcription language. Each running container requires: 1 vCPU, 2-5GB RAM, 100MB hard disk space.	Base config – 2 vCPU, 8GB RAM this will process approximately 2 hours of audio per hour. Additional resources: 1 vCPU, up to 5 GB RAM for every additional worker.	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, aac, ogg, mpeg, amr, m4a, mp4, flac no additional formats are supported.		
TRANSCRIPTION FORMATS	16 kHz (Broadcast) and 8 kHz (Telephony) acoustic models built in, with automatic selection based on file sample rate.		
OUTPUTS FORMAT	**JSON** – Transcripts provided in JSON include metadata such as timing and confidence scores, speaker/channel labels plus more on a per word basis. **TXT –** TXT output does not include timing information and confidence scores, **SRT.**		

BATCH ASR

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

SPEECHMATICS

<table>
<tr><th></th><th>DOCKER CONTAINER</th><th>VIRTUAL APPLIANCE</th><th>CLOUD OFFERING</th></tr>
<tr><td>RESOURCE REQUIREMENTS</td><td>An individual Docker image is required for each transcription language.
Each running container requires:
1 vCPU, 2-5GB RAM, 100MB hard disk space.</td><td><table><tr><th>Pack</th><th>Supported languages</th><th>Appliance size</th></tr><tr><td>Nano</td><td>1</td><td>40GB</td></tr><tr><td>Mini</td><td>3</td><td>40GB</td></tr><tr><td>Midi</td><td>8</td><td>60GB</td></tr><tr><td>Maxi</td><td>15</td><td>80GB</td></tr><tr><td>Plus</td><td>17</td><td>80GB</td></tr></table></td><td>N/A</td></tr>
<tr><td rowspan="2">PERFORMANCE</td><td>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 Docker engine at the same time or across multiple Docker engines to enable large scale operations.</td><td>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.
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.</td><td>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.</td></tr>
<tr><td colspan="3">Performance figures are valid for ‘good’ audio. Very noisy audio may result in longer times. Enabling Speaker or Channel Diarization may slightly increase turnaround time.</td></tr>
<tr><td rowspan="2">DATA RETENTION PERIOD</td><td>0</td><td>Once the transcription is complete the transcript will be retained by the virtual appliance for 24 hours.
The media file is removed immediately after the transcription has completed.</td><td>7 days</td></tr>
<tr><td colspan="3">After this period, any media and transcription data will be purged unless already done so by the customer using the API.</td></tr>
<tr><td>LIMITATIONS</td><td colspan="2">Can operate on input file sizes of up to 2 hours recorded length or 4GB in size, whichever is reached first.</td><td>Can operate on input file sizes up to 2 hours recorded length or 1GB in size, whichever is reached first on POSTS within the API.</td></tr>
<tr><td>CONNECTIVITY REQUIREMENTS</td><td colspan="2">Can operate within own security boundary allowing you to keep control of your data.</td><td>HTTPS port 443 needed to enable access to the Cloud offering.
For connectivity requirements if call back is used, ingress port will need to be enabled.</td></tr>
<tr><td>ADMIN</td><td>No ongoing maintenance needed for the containers.
All administration is provided by direct use of Docker commands.</td><td>Full APIs for management and monitoring.
Monitoring of the appliance can be done via a web GUI or APIs.
Flexible licensing model.
Local administration available via APIs.</td><td>All administration of the Cloud offering is managed by Speechmatics.</td></tr>
<tr><td>SUPPORT</td><td colspan="3">Latest release (N) and previous release of the Speechmatics product (N-1).</td></tr>
</table>

BATCH ASR

Default features

FEATURES	DESCRIPTION
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 timeframes.
TIMINGS ● ● ●	Start and end time of each word (JSON).
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.
ADVANCED PUNCTUATION* ● ● ●	Support of periods, commas, exclamation marks and question marks.
PROFANITY TAGGING** ● ● ●	Speechmatics supports profanity tagging by providing a metadata tag within the JSON output informing users when words – matching those within a list of profanities curated by Speechmatics – have been transcribed.
DISFLUENCY TAGGING** ● ● ●	Speechmatics supports disfluency tagging by providing a tag within the JSON output informing users when grunts or non-lexical utterances such as "huh", "uh", "erm", "um", and "hmm" have been transcribed. These tags are identified against a list of values that are curated by Speechmatics.

*Only for certain languages

**Global English only

● DOCKER CONTAINER, ● VIRTUAL APPLIANCE, ● CLOUD OFFERING

BATCH ASR

Optional features

FEATURES	DESCRIPTION
SPEAKER DIARIZATION ● ● ●	Ability to detect and label speakers based on multiple speakers within the same channel.
CHANNEL DIARIZATION** ● ● ●	Supports multiple speakers on separate channels and streams. Up to 6 streams/channels. **Note:** Channel Diarization cannot be used with Speaker Diarization but it can be used with Speaker Change.
SPEAKER CHANGE (BETA) ● ● ●	Token passed in the output on sound wave analysis providing unique time and duration of each speaker turn.
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 an extension of Custom Dictionary which allows alternate spellings, pronunciations, acronyms and abbreviations to be used.
CUSTOM DICTIONARY CACHING* ● ●	Enable users to cache their Custom Dictionary word lists. Custom dictionaries are required to be loaded before each transcription session can start. Large word lists can slow down the start-up of the transcription engine. For users who repeatedly use the same word list, the ability for them to be ached optimizes session start-up speed.
ADVANCED PUNCTUATION OVERRIDES* ● ● ●	Customer configurable option to choose which punctuation characters are displayed in the output.
OUTPUT LOCALE ● ● ●	Specify rules for transcription output for the Global English (en) language pack to specify American or British spellings.
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.
NOTIFICATION CALLBACKS ● ● ●	A lightweight signal indicating that results are ready to be fetched through the normal API without needing to poll.
URL FETCHING ● ● ●	If digital media is stored in cloud storage (for example AWS S3 or Azure Blob Storage) jobs can be submitted by providing the URL of the audio file.
PARALLEL PROCESSING ●	Capability for customers to apply additional CPU resources to process multiple audio chunks in parallel for faster transcription turnaround time.

*Customers are responsible for the deployment, management, size and location of their own cache.

**Only for certain languages.

● DOCKER CONTAINER, ● VIRTUAL APPLIANCE, ● CLOUD OFFERING

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.