

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 'Speech as a Service' (SaaS) offers a fully maintained, operated and managed solution.

<p>Transcription</p>	<ul style="list-style-type: none"> • Supports pre-recorded audio • Transcripts provided in JSON or TXT format, containing: <ul style="list-style-type: none"> – Timing and confidence scores for each transcribed word – Sentence boundary information • Diarization: <ul style="list-style-type: none"> – Speaker diarization - Provide the 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 	<ul style="list-style-type: none"> • Custom Dictionary allows up to 1,000 additional words that can be added to the standard dictionary per input stream <ul style="list-style-type: none"> – 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 • 16 kHz (Broadcast) and 8 kHz (Telephony) acoustic model built in, with automatic selection based on file sample rate
-----------------------------	---	--



Docker container



Virtual appliance



Cloud (SaaS)

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. SaaS accelerates time to market, while reducing operational complexities and cost. The SaaS has a configurable clean-up policy enabling operators to define the period transcripts will be stored.

Supported languages

Global English (en), Spanish (es), German (de), French (fr), Dutch (nl), Portuguese (pt), Japanese (ja), Korean (ko), Italian (it), Swedish (sv), Danish (da), Polish (pl), Catalan (ca), Hindi (hi), Russian (ru), Mandarin (cmn), Norwegian (no), Arabic (ar), Bulgarian (bg), Czech (cs), Greek (el), Finnish (fi), Hungarian (hu), Croatian (hr), Lithuanian (lt), Latvian (lv), Romanian (ro), Slovakian (sk), Slovenian (sl)

Nano: Global English (en)
Mini: Global English (en), German (de), Spanish (es)
Midi: Global English (en), German (de), Spanish (es), French (fr), Dutch (nl), Portuguese (pt), Japanese (ja), Korean (ko)
Maxi: Global English (en), German (de), Spanish (es), French (fr), Dutch (nl), Portuguese (pt), Japanese (ja), Korean (ko), Italian (it), Swedish (sv), Danish (da), Polish (pl), Catalan (ca), Hindi (hi), Russian (ru)

Global English (en), Spanish (es), German (de), French (fr), Dutch (nl), Portuguese (pt), Japanese (ja), Korean (ko), Italian (it), Swedish (sv), Danish (da), Polish (pl), Catalan (ca), Hindi (hi), Russian (ru), Mandarin (cmn), Norwegian (no), Arabic (ar), Bulgarian (bg), Czech (cs), Greek (el), Finnish (fi), Hungarian (hu), Croatian (hr), Lithuanian (lt), Latvian (lv), Romanian (ro), Slovakian (sk), Slovenian (sl)

Management Interface

Standard Docker or Kubernetes management tools

REST API

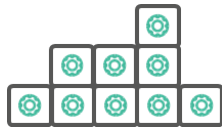
N/A

Speech Interface

Console/STUDIO

REST API

REST API



Docker container



Virtual appliance



Cloud (SaaS)

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	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	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 outputs format	JSON only	Transcripts provided in JSON or TXT format, containing: <ul style="list-style-type: none"> • Timing and confidence scores for each transcribed word • Sentence boundary information <i>.txt output does not include timing info and confidence scores</i>	Transcripts provided in JSON or TXT format, containing: <ul style="list-style-type: none"> • Timing and confidence scores for each transcribed word • Sentence boundary information
Resource requirements	Container execution: The container needs to run in a supported Docker environment: 2-5GB RAM for each executing container* *At least 100 MB free disk space must be available for operating overheads and, if streaming in audio file space, for a copy of the audio file being transcribed. Memory requirements will depend on audio quality 1 vCPU per concurrent transcription	Nano: 1 language: 1 vCPU, 13GB OVA, 40GB Application size Mini: 3 languages: 1 vCPU, 21GB OVA, 40GB Application size Midi: 8 Languages: 1 vCPU, 41GB OVA, 60GB Application size Maxi: 15 Languages: 1 vCPU, 65GB OVA, 80GB Application size Host hardware: Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, or better	N/A
Deployment Platform	Docker, Kubernetes	VMWare, VirtualBox, Amazon EC2	N/A
Performance	<ul style="list-style-type: none"> • Transcript can be provided in better than 2x real time • 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 	<ul style="list-style-type: none"> • Transcript can be provided in better than 2x real time • 1vCPU per concurrent transcription job • Additional vCPU's can be added to enable multiple streams to be transcribed at the same time 	Transcript can be provided in 2x real time Example: transcript of A 60 minute file provided in 30 minutes.
Connectivity requirements	Can operate within your own security boundary allowing you to keep control of your own data	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 *for connectivity requirements if call back is used ingress port will need to be enabled
Admin	No ongoing maintenance needed for the containers All administration is provided by direct use of Docker commands Flexible licensing model	Full management API for management tasks Full management API for Monitoring – monitoring of the appliance can be done via a console interface or API's Flexible licensing model	All administration of the cloud is managed by Speechmatics