SPEECHMATICS

SPEECHMATICS
PRODUCT SHEET

Real-Time ASR

LIVE TRANSCRIPTS DELIVERED IN REAL-TIME.
TRANSCRIPTION PROVIDED IN PARALLEL TO INPUT AUDIO BEING STREAMED THROUGH THE SPEECHMATICS ENGINE

REAL-TIME ASR

Live transcripts delivered in real-time.

Transcription provided in parallel to input audio being streamed through the Speechmatics engine

The Speechmatics ASR means there is no need to complete the recording before accessing the transcript. Tunable transcript outputs allow integration into many use cases. The technology is available as a virtual appliance or a container, which can transcribe multiple language streams simultaneously.

Latency as low as 1 second - adaptive end pointing can be used to get the best accuracy within specified latency limits. Controls are provided to enable a continuous flow of output at a specified latency (the transcript accuracy will decrease as latency is reduced). This is known as Low latency Finals. Near real-time transcription with automatic word correction and the option of configurable latency setting. Words are updated after output as additional context becomes available. Latency between the initial transcription and the update is configurable to meet specific time Vs. accuracy KPIs.

	VIRTUAL APPLIANCE	DOCKER CONTAINER
OVERVIEW	Speechmatics' Real-time Virtual Appliance ASR provides a virtual machine that can be instantiated to interpret electronic audio streams containing speech into transcripts. The software is not intended to store any data persistently, audio and associated transcripts are not retained within the appliance.	The Speechmatics Docker container can be used by customers to build scalable, real-time transcription services in their own infrastructure.
	Preconfigured virtual machine easy to configure in a few clicks.	A lightweight, stand-alone software package that includes everything needed to run it: code, runtime, system tools, system libraries, settings.
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). 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.	**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.
COMPUTE REQUIREMENT	**Each virtual machine:** Base config: 2 vCPU, 8GB RAM required to process one continuous audio stream. **Additional resources:** 1 vCPU, up to 3 GB RAM for ever additional worker able to process a continuous audio stream.	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. 1.5 GB RAM per container (default config). 3 GB RAM per container (with Custom Dictionary).
MANAGEMENT INTERFACE	HTTPS REST API to manage the appliance including license config, log collection, scale config. For additional information see documentation.	N/A
SPEECH INTERFACE	TCP WebSocket based API using TLS (HTTPS) (wss://) is provided for the transmission of audio streams and retrieval of resulting transcripts.	TCP WebSocket based API for the transmission of audio streams and retrieval of resulting transcripts without TLS (ws://).
INPUT FILE FORMATS	**Raw audio:** PCM F32 LE raw audio stream (32-bit float), PCM S16 LE raw audio stream (16-bit signed int), mu-law **Files:** wav, mp3, aac, ogg, mpeg, amr, m4a, mp4, flac no additional formats are supported.	
OUTPUT FORMAT	**JSON format:** Operates on multiple streams at a time. Output for each stream is available continuously as each final phrase is transcribed. Can operate on continuous streams of audio data and includes metadata such as timing and confidence scores, speaker change tokens on a per word basis.	

REAL-TIME ASR

Live transcripts delivered in real-time.

Transcription provided in parallel to input audio being streamed through the Speechmatics engine

	VIRTUAL APPLIANCE			DOCKER CONTAINER
CONNECTIVITY REQUIREMENTS	Can operate within own security boundary allowing partners to keep control of their own data.			
PYTHON WRAPPER	**Supports:** Real time streaming protocol (RTSP). Direct microphone input. Support for plugging in custom inputs. Support for plugging in custom outputs.			New python client for container.
RESOURCE REQUIREMENTS	**Pack**	**Supported languages**	**Appliance size**	An individual Docker image is required for each transcription language. Each running container requires: 1 vCPU, 1.5 – 3 GB RAM.
	Nano	1	40GB	
	Mini	3	40GB	
	Midi	8	60GB	
	Maxi	15	80GB	
	Plus	17	80GB	
PERFORMANCE	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 (this is abstracted from the language models used).			1vCPU per stream allows a transcript to be provided in real-time.
ADMIN	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.			No ongoing maintenance needed for the containers. All administration is provided by direct use of Docker commands.
SUPPORT	Latest release (N) and previous release of the Speechmatics product (N-1).			

REAL-TIME ASR
Default Features

SPEECHMATICS

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 time frames.
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 (full stops), 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. These tags are identified against an internally curated list of values that are curated by Speechmatics.
DISFLUENCY TAGGING** ● ●	Speechmatics now 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 an internally curated list of values that are curated by Speechmatics.

*Only for certain languages

**Global English only

● DOCKER CONTAINER, ● VIRTUAL APPLIANCE

REAL-TIME ASR

Feature Matrix

FEATURES	DESCRIPTION
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. Not long Custom Dictionary Sounds is extension 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 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.
SPEAKER CHANGE (BETA) ● ●	Token passed in the output on sound wave analysis providing unique time and duration of each speaker turn.
OUTPUT LOCALE ● ●	Specify rules for transcription output for the Global English (en) language pack to specify American or British spellings.
PARTIALS ● ●	Transcription engine produces words instantly. Words can be updated after output as additional context becomes available.
LOW LATENCY FINALS ● ●	Near real-time transcription with automatic word correction with option of configurable latency setting. Words can be updated after output as additional context becomes available. Latency between the initial transcription and the update is configurable to meet specific time Vs. accuracy KPIs.

* The Virtual Appliance will manage this shared space, and automatically remove the oldest cached dictionaries when the cache is becomes full.

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

● DOCKER CONTAINER, ● VIRTUAL APPLIANCE

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 media & entertainment, contact centers, CRM, financial services, security, 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

QUEEN'S AWARD FOR INNOVATION (2019) WINNER