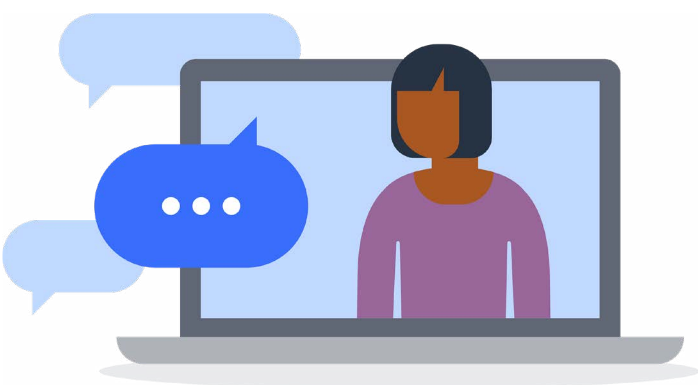
Real-Time ASR

Product Sheet

Transcribe in real-time and get data results instantly with Speechmatics' most powerful, inclusive, and accurate engine ever.



Speechmatics.com

Speechmatics

Real-Time ASR

With transcription provided in real-time, you can gather actionable data instantly. Our proprietary Low Latency Finals deliver best-in-class accuracy through automatic word correction re-scoring.

With our real-time ASR, there's no need to wait until the recording is finished before you can access the information within the transcript.

With latency as low as a second, it can be adapted to get the best accuracy within your specific limits. Low Latency Finals enable the most accurate real-time transcription by leveraging Speechmatics' proprietary rescoring outputs. Here, Speechmatics can define the context of a transcript and automatically correct words to match the context. You can also configure the latency between the initial transcription and the update to balance time and accuracy indicators.

	Virtual Appliance	Docker Container
Overview	Inside our Virtual Appliance, you can turn speech-to-text without storing any data. Audio and the associated transcripts are not retained within the appliance.	Using our Docker Container ASR, you can build scalable, real-time transcription services within your own infrastructure.
Description	A preconfigured virtual machine that's simple and easy to configure.	A fully-inclusive, lightweight, stand-alone software package.
Languages	You can find all supported languages for both Virtual Appliance and Container at [Speechmatics.com]().	
Operating Points	**Standard Operating Point:** • Requires Intel Broadwell class architecture minimum. **Enhanced Operating Point:** • Requires Intel Cascade Lake class architecture minimum. • Recommended using hardware that supports AVX512_VNNI flag to improve transcription processing speed.	
Supported Operating Environment	**Minimum specification:** Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent). We support 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.	**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 support AVX2 compatible hardware to take advantage of the 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 3GB RAM for every 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. 1.5GB RAM per container (default config). 3GB 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.	

Speechmatics

<table>
<tr><th></th><th colspan="3">Virtual Appliance</th><th>Docker Container</th></tr>
<tr><td>Connectivity Requirements</td><td colspan="4">Regardless of your deployment method, you can operate your ASR within your own security boundary. This allows you to keep control of your own data.</td></tr>
<tr><td>Python Wrapper</td><td colspan="3">Speechmatics provides reference Python libraries that can be used to wrap the WebSocket interface. They provide the ability to connect directly to a microphone or RTSP feed. It offers support for plugging in custom inputs and outputs.</td><td>New Python client for container.</td></tr>
<tr><td rowspan="6">Resource Requirements</td><td>Pack</td><td>Supported Languages</td><td>Appliance Size</td><td rowspan="6">An individual Docker image is required for each transcription language. Each running container requires: 1 vCPU, 1.5 – 3GB RAM.</td></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>19</td><td>80GB</td></tr>
<tr><td>Performance</td><td colspan="3">1 vCPU 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).</td><td>1 vCPU per stream allows a transcript to be provided in real-time.</td></tr>
<tr><td>Admin</td><td colspan="3">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>No ongoing maintenance needed for the containers.
All administration is provided by direct use of Docker commands.
We provide health and readiness checks to confirm the status of the Container.</td></tr>
<tr><td>Support</td><td colspan="4">Latest release (N) and previous release of the Speechmatics product (N-1).</td></tr>
</table>

Speechmatics

Additional Features

Features	Description
Confidence Scores	Visualize the confidence of every word in the transcript.
Low Latency Finals	Define the context of transcriptions and use it to automatically correct words.
All Major Files Formats Supported	Support for major audio and video formats, with automatic sample rate detection to get you started quickly.
Advanced Punctuation	Use an extensive set of supported punctuation marks to optimize the speed and ease of transcription.
Custom Dictionary and Sounds Feature	Add unique words to the dictionary to enhance your transcription accuracy.
Speaker Change	Easily identify a change of speaker within your transcript and improve its readability.
Flexible Endpointing	Ensure output formatting is kept consistent by flexibly overriding when transcriptions finals are returned.

Speechmatics

Ready to try Speechmatics?

Sign up for your free trial and we'll guide you through the implementation of our API. We pride ourselves on offering the best support for your business needs. 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.

Email

hello@speechmatics.com

Phone

+44 (0)1223 794 497

UK office

Cambridge Science Park
Milton Road
Cambridge CB4 0WD
United Kingdom