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, which can transcribe multiple language streams simultaneously by the same appliance, or via separate appliances.

Transcription

Latency as low as 1 second

- Adaptive end pointing can be used to get the best accuracy as fast as possible. Controls are provided to enable a continuous flow of output at a specified latency (the transcript accuracy will decrease as latency is reduced)
- 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 KPI's.
  - Adaptive updating of transcription provided by Speechmatics automatically gives you the best transcript possible.
- Transcript words and timing information
- 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 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.

Supported languages	**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)
Management Interface	REST Management APIs for admin and monitoring: • Stop and start appliance • Configure networking • Get status and logging information
Speech Interface	Websocket API: • Allows streaming of PCM_S16LE and F32LE directly • Ability to stream files • Asynchronously returns transcript • Low Latency Final transcripts available Python libraries: • Provide direct use of RTSP, microphones • Easy consumption of the transcripts
Supported Hosts	VM Ware - VMware vSphere® 6.0, VirtualBox - 5.2+, Amazon EC2
Connectivity requirements	Can operate within own security boundary allowing you to keep control of your own data
Resource requirements	**Nano:** 1 language: 1 vCPU, 11GB OVA 40GB Application size **Mini:** 3 languages: 1 vCPU, 16GB OVA 40GB Application size **Midi:** 8 Languages: 1 vCPU, 27GB OVA 40GB Application size **Maxi:** 15 Languages: 1 vCPU, 45GB OVA 60GB Application size Intel® Xeon® CPU E5-2630 v4 (Sandy Bridge) 2.20GHz (or equivalent) or better Each appliance required: • 1 vCPU per concurrent transcription stream • 4GB base + 2GB RAM per language stream
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 Stream input modes: RTSP
Outputs format	JSON
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)

Contact us for more information on features and pricing:

info@speechmatics.com | +44(0)1223 907818
www.speechmatics.com