SPEECHMATICS

Real-Time Virtual Appliance

Live transcripts provided in 'real time'. The transcription is provided whilst the input audio is being streamed to the Speechmatics engine.

There is no need to complete the recording before accessing the transcript. Partial, final and dynamic transcript outputs are available to allow integration into many use cases. The technology is available as a virtual appliance, enabling multiple streams to be transcribed in multiple languages at the same time in 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)
- 3 output modes:
  - Partial – almost instant transcription with automatic word correction at a later time once additional context is available*
  - Final – most accurate transcription. Output available a 'phrase at a time'**
  - Dynamic – customisable output timing based on time or character constraints. No corrections once transcription is provided***
- All modes provide the transcript words and timing information
- Custom Dictionary allows up to 1,000 additional words to be added to the dictionary on a per-transcription session basis
  - Allows users to quickly add context-specific words, for example company names, place names or unusual names
  - Custom Dictionary Sounds is an extension that allows alternate spellings or pronunciations to be used
  - 8 kHz and 16 kHz models support telephony and broadcast use cases

Languages supported:

- Global English (en), Dutch (nl) German (de), Japanese (ja), Spanish (es), French (fr), Korean (ko), Portuguese (pt)

Programatic use:

- Websocket API for direct use
  - Allows streaming of PCM_S16LE and F32LE directly (small set of samples at a time)
  - Files can be streamed
  - Asyncronously returns transcript
  - Partial, Dynamic and Final transcripts available
- Python libraries
  - Provide direct use of RTSP, microphones
  - Easy consumption of the transcripts

Administration:

- REST Management APIs for admin and monitoring:
  - Stop and start appliance
  - Configure networking
  - Get status and logging information

Hardware needs:

- Host hardware: Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, or better
- Each appliance required:
  - 1 vCPU per concurrent transcription stream
  - 4GB base + 2GB RAM per language stream

Contact us for more information on features and pricing:
info@speechmatics.com | +44 (0)1223 794 497 | **www.speechmatics.com**

SPEECHMATICS

Real-Time Virtual Appliance

Hypervisor support:

- VMWare
- VirtualBox

Storage:

- Appliance operates within a 34GB storage footprint
- No audio or transcripts are stored within the appliance

Performance:

- 1vCPU per stream allows a transcript to be provided in real-time
- Additional vCPU's can be added to enable multiple streams to be transcribed at the same time

Connectivity requirements:

- Can operate within own security boundary* allowing you to keep control of your own data

*The appliance needs connectivity to a licence service to be activated, and to enable licence changes/updates

*Partial transcripts:

- Produce words instantly
- Words can be updated after output as additional context becomes available
- Will be finalised once the final transcript is available

**Dynamic transcripts:

- Aimed at captioning where initial transcript needs to be available as soon as possible and cannot be updated after output. Three parameters can be set:
  - Max Character = max characters before output (forces output when a character limit is reached)
  - Min Context = minimum number of times that a word has been said before being output
  - Max Delay = maximum time a word is said before being output

Notes: The aim is to get the dynamic transcript as long as possible. Splitting transcripts into individual words reduces the accuracy, so the system prefers chunks to be as long as possible. It is essentially a transcript of any audio starting at max_delay (in the past) and ending min_context (in the past). So, for example, if you set min_context to 3s and max_delay to 5s then you get 2 seconds of audio starting 5 seconds in the past. If you want the words to appear ASAP, set max_delay to the same as min_context.

***Final transcripts:

- Adaptive end pointing provided by Speechmatics automatically gives you the best transcript possible as fast as possible. This achieves transcription a 'phrase' at a time. Any finer level of control will require use of partial or dynamic transcripts as defined above

Licensing:

- Can be licensed per hour or per stream

Contact us for more information on features and pricing:

info@speechmatics.com | +44 (0)1223 794 497 | **www.speechmatics.com**