DEEPGRAM

PRODUCT SHEET

Real-Time Streaming

Deepgram's real-time streaming transcriptions and captions are the fastest in the market. Our transcription latency is less than 300 milliseconds; i.e. the word text is output within 300 milliseconds of the end of the spoken word, for transcription processing. Coupled with an average accuracy rate of over 90%, Deepgram allows conversational AI companies to create near human response times for voicebots. No need for the sound of fake typing while the ASR is transcribing the text for analysis.

Now sales and support enablement and call analytics can be true real-time. Get immediate tips, recommendations, compliance alerts, or buyer intent data during the conversation. Close the deal while on the phone. Get to issue resolution faster. Interrupt a call if the agent is not complying with regulatory requirements.

And real-time streaming is available for on-premise deployment which will further reduce any network latency and keep your audio data more secure and private for financial, health, and government institutions.

USE CASES:

- Agent assist for contact centers
- IVR and Conversational AI voicebots
- Real-time captioning for events or classrooms
- Real-time meeting transcriptions for online or in-person meetings
- Sales and support enablement for all businesses

BENEFITS:

- Improve conversational AI UX with lower latency
- Increase accessibility and meet regulations with real-time accurate captioning
- Develop a true real-time sales or support enablement or analytics solution
- Build accurate, real-time products that scale

CAPABILITIES:

- Less than 300 ms transcription latency*
- 90%+ accuracy at real-time speeds
- Unlimited number of concurrent streams

**Represents average latency and accuracy on conversational audio. User connection speeds and audio quality may alter results.*

Speech-to-text features

FEATURE	DESCRIPTION
Base or enhanced architecture models	Base architecture is built for use with audio in common conversations and has been trained on the most common words used in each language and dialect. Our Enhanced architecture is built for audio applications that involve more specific terminology or jargon such as technical meetings, educational lectures, and financial transactions.
Callback	Callback allows you to supply a callback URL to which transcriptions can be returned. When passed, Deepgram will immediately respond with a request_id before processing your audio asynchronously.
Endpointing	Deepgram's Endpointing feature monitors incoming streaming audio and detects when a user has finished speaking or paused for a significant amount of time, indicating the completion of an idea. When Deepgram detects an endpoint, it assumes that no additional data will improve its prediction, so it immediately finalizes its results for the processed time range and returns the transcript with a speech_final parameter set to true.
Find and replace	Find and Replace searches for terms or phrases and replaces them in the response JSON object.
Interim results	Interim Results monitors streaming audio and provides interim transcripts, which are preliminary results provided during the real-time streaming process.
Keywords/custom vocabulary	Keywords are words you want the model to pay particular attention to help it understand context; this is known as keyword boosting. You can also suppress keywords.
Languages	Our supported languages are listed here.
Multi-channel	Multichannel recognizes multiple audio channels and transcribes each channel independently. When set to true, you will receive one transcript for each channel.
Numerals	Numerals feature numbers from written format to numerical format. For example, the number "nine hundred" would appear in your transcript as "900".
Punctuation and capitalization	Deepgram's Punctuation feature adds punctuation and capitalization to your transcript.
Profanity filtering	Deepgram's Profanity Filter feature looks for recognized profanity and converts it to the nearest recognized non-profane word or removes it from the transcript completely.

Speech-to-text features

FEATURE	DESCRIPTION
Search (audio)	Search searches for terms or phrases by matching acoustic patterns in audio (which we have found is more accurate than matching for text patterns in transcripts) and returns results in the response JSON object.
Tagging	Tagging allows you to label your API requests or API keys for the purpose of identification during usage reporting.
Tailored speech models	Tailored speech models allow you to customize a speech model for your difficult audio characteristics; i.e. terminology, jargon, noise, accents, dialects, and language mixtures.
Use case speech models	Use case speech models are trained for specific use cases and are more accurate than our general models for that use case.
Voice activity detection	Voice Activity Detection (VAD) feature monitors incoming audio and detects when a sufficiently long pause is detected.

Speech understanding features

FEATURE	DESCRIPTION
Diarization	Deepgram's Diarize feature recognizes speaker changes and assigns a speaker to each word in the transcript.
Named entity recognition (NER)	Named-Entity Recognition (NER) recognizes alphanumerics in audio and removes whitespace between the characters in the transcript.
Redaction	Redaction redacts sensitive information, replacing redacted content with asterisks (*).

Technical details

TECHNICAL ASPECT	DETAIL
Management interface	Console or REST API
Speech interface	TCP WebSocket
SDKs	Node.js, Python, .NET.Demp and Go
Input file formats	• Most containerized audio packets • Raw, headerless audio packets with the following encodings – Linear16, flac, mulaw, amr-nb, amr-wb, opus, and speex – Sample rates are required for amr encodings – Number of channels are also required
Output format	JSON

Is your speech recognition slowing down your voice solution? Contact us at deepgram.com/contact-us to explore what's possible for your business with real-time transcriptions.

Related information

- Getting Started with Live Streaming Audio
- How Deepgram Works
- The Definitive Guide to Speech Recognition

Is your speech recognition slowing down your voice solution?
Contact us at deepgram.com/contact-us to explore what's possible
for your business with real-time transcriptions.