Groq Automatic Speech Recognition (ASR) API

The Automatic Speech Recognition (ASR) API from Groq provides developers access to powerful ASR models, including Whisper v3 Large and Distil-Whisper. Running on Groq® LPU™ AI inference technology, the Groq ASR API provides ultra-low latency audio transcription and translation. It is available at console.groq.com.

Whisper v3 Large and Distil Whisper v3 Large are available on-demand or can be deployed as private, dedicated instances in GroqCloud™. They can be fine-tuned for specific languages or industries.

Both models provide high-quality speech but they differ in terms of speed, accuracy, and cost. In this document, we'll guide you through the key features and performance of each model, helping you make an informed decision for your specific use case.

Model Overview

Whisper V3 Large

Whisper V3 Large is a multilingual ASR model that supports transcription and translation in multiple languages. It offers a high level of accuracy and speed, with a Word Error Rate (WER) of 10.3% and a real-time speed factor of up to 189x. It is ideal for applications that require fast and accurate transcription and translation in various languages.

Real-world examples:

- Global customer support: A multinational company uses Whisper V3 Large to provide customer support in English, Spanish, French, and Mandarin. The model transcribes and translates customer calls, enabling the company to respond promptly and effectively to customer inquiries.
- International business meetings: A company uses Whisper V3 Large to transcribe and translate business meetings, enabling participants to focus on the discussion rather than taking notes.
- Multilingual subtitling: A video streaming platform uses Whisper V3 Large to generate subtitles in multiple languages for TV shows and movies, making them more accessible to a global audience.

Distil-Whisper

Distil-Whisper is a compressed version of Whisper V3 Large, fine-tuned specifically for English speech recognition. It offers a remarkable balance of speed and accuracy, with a WER of 13% and a real-time speed factor up to 230x. This model is ideal for applications that require fast and accurate English transcription, such as real-time customer service chatbots, automated speech-to-text systems, and voice-controlled interfaces.

Real-world examples:

- Customer service chatbots: A company uses Distil-Whisper to power its English customer service chatbot, enabling customers to interact with the chatbot in real-time and receive accurate and helpful responses.
- Transcribing audio and video recordings: A media company uses Distil-Whisper to transcribe audio and video recordings in English, enabling journalists to focus on editing and analysis rather than transcription.
- Automated speech-to-text Systems: A healthcare organization uses Distil-Whisper to transcribe medical dictations in English, enabling doctors to focus on patient care rather than paperwork.

Model Specifications

Item	Whisper V3 Large	Distil-Whisper
Transcription Speed[1]	~189x real-time speed factor	~230x real-time speed factor
Accuracy[1]	~10.3% WER	~13% WER
Language Support	Multilingual	English only
Use Cases	Translation and transcription	Transcription only
Pricing[2]	$0.111 per hour[3]	$0.02 per hour
FIle Based	Support for file-based transcription only - no streaming support	
API Compatibility	Compatible with OpenAI API specification	
Supported Audio Formats	mp4, mpeg, m4a, wav, mp3, ogg, webm, opus	
Max Audio File Size	Up to 25 MB (~30 minute audio files) Note: Groq's Whisper implementation is fastest at audio files above 4 minutes	
SDKs Available	Python SDK, JavaScript SDK	
Audio Upload	Upload audio files directly, no audio URL required	
Rate Limits	Visit console.groq.com for the latest rate limits	

1. Speed and accuracy as of 9/8/2024 measured by ArtificialAnalysis.ai
2. Pricing as of 9/8/2024 see groq.com/pricing for the latest pricing. Minimum charge is 10 sec per request.
3. Price effective as of 10/1/2024

Whisper models require a minimum of 30 seconds of audio. Audio files shorter than 30 seconds are padded with silence. We recommend audio files longer than 30 seconds to optimize throughput.

Interested in Learning More?

For more information visit groq.com or contact us at sales@groq.com.

Legal Statements.

Groq Inc. HQ
301 Castro St. Suite 200
Mountain View, CA 94041

Mailing Address
PO Box 1778
Mountain View, CA 94041

www.groq.com
console.groq.com