

ASRモデル比較表

モデル名	リリース 年月	日本語対 応	タイムスタン プ出力	話者分 離	モデルサイズ (パラメータ 数/推論リソー ス)	無料/有料 (ライセンス)
Whisper (OpenAI)	2022-09 ¹	○ ²	○ (フレーズ 単位) ²	×	Tiny 39M~ Large 1.55B ³ (Largeで 約10GB VRAM)	無料 (MIT) ³
WhisperX (拡張版)	2023-07 (推定)	○ (Whisper 準拠)	○ (単語レベ ル)	○	Whisper Large相当 (~1.5B)	無料 (BSD-2- Clause) ⁴
Google Cloud STT	2016-03 ⁵	○ (ja-JP 対応) ⁶	○ (単語レベ ル) ⁷	○ ⁸	非公開 (クラ ウドAPI)	有料 (商用 API) ⁹
Meta (MMS)	2023-05 ¹⁰	○ (1100言 語以上) ¹¹	○ (内部モデ ルで対応)	×	10億 ¹²	無料 (研究 用、CC-BY- NC) ¹³
NVIDIA Riva	2021-11 ¹⁴	○ ¹⁵	○ (Conformer ベース)	○ (ス トリー ミング 対応) ¹⁶	例: Parakeet- CTC 0.6B, Parakeet- RNN-T 1.1B ¹⁷	無料 (NVIDIA ライセン ス、企業向 け有償サ ポート有) ¹⁸
AssemblyAI	2018-01?	○ ¹⁹	○	○ (日 本語話 者分離 対応) ²⁰	非公開 (クラ ウド)	有料 (API利 用)
Rev AI (Rev.com)	2016-03?	○ (58言語 対応) ²¹	○ (単語レベ ル)	×	非公開 (クラ ウド)	有料 (API利 用)

各モデルの詳細は引用元を参照してください。例えばOpenAI発表のWhisperは2022年9月公開で多言語対応（日本語含む）²、フレーズ単位タイムスタンプ対応、話者分離非対応、最大1.55BパラメータでMITライセンス公開³しています。Google Cloud STTは2016年3月から提供（日本語対応⁶、単語レベルタイムスタンプ・話者分離対応⁷⁸）の有料サービスです。MetaのMMSは2023年5月発表（1100言語対応¹¹、1Bパラメータ、CC-BY-NCライセンス）など、引用資料をご覧ください。

出典: Whisper/WhisperX¹³⁴、Google Cloud STT⁵⁶⁷⁸、Meta MMS¹⁰²²、NVIDIA Riva¹⁴¹⁵¹⁶、AssemblyAI¹⁹²⁰、Rev AI²¹ (各社ドキュメントなど)。

1 2 **Introducing Whisper | OpenAI**

<https://openai.com/index/whisper/>

3 **GitHub - openai/whisper: Robust Speech Recognition via Large-Scale Weak Supervision**

<https://github.com/openai/whisper>

4 **GitHub - m-bain/whisperX: WhisperX: Automatic Speech Recognition with Word-level Timestamps (& Diarization)**

<https://github.com/m-bain/whisperX>

5 9 **Google opens access to its speech recognition API, going head to head with Nuance | TechCrunch**

<https://techcrunch.com/2016/03/23/google-opens-access-to-its-speech-recognition-api-going-head-to-head-with-nuance/>

6 **Speech-to-Text supported languages | Google Cloud**

<https://cloud.google.com/speech-to-text/docs/speech-to-text-supported-languages>

7 **Get word timestamps | Cloud Speech-to-Text | Google Cloud**

<https://cloud.google.com/speech-to-text/docs/async-time-offsets>

8 **Detect different speakers in an audio recording | Cloud Speech-to-Text | Google Cloud**

<https://cloud.google.com/speech-to-text/docs/multiple-voices>

10 **Introducing speech-to-text, text-to-speech, and more for 1,100+ languages - Argos Translate - LibreTranslate Community**

<https://community.libretranslate.com/t/introducing-speech-to-text-text-to-speech-and-more-for-1-100-languages/611>

11 12 13 22 **facebook/mms-1b-all · Hugging Face**

<https://huggingface.co/facebook/mms-1b-all>

14 18 **NVIDIA Announces Riva Speech AI and Large Language Modeling Software For Enterprise | NVIDIA Technical Blog**

<https://developer.nvidia.com/blog/nvidia-announces-riva-speech-ai-and-large-language-modeling-software-for-enterprise/>

15 **NVIDIA Riva: Speech AI Technology**

<https://www.datamonsters.com/technologies/riva>

16 17 **Release Notes — NVIDIA Riva**

<https://docs.nvidia.com/deeplearning/riva/user-guide/docs/release-notes.html>

19 **Supported languages | AssemblyAI | Documentation**

<https://www.assemblyai.com/docs/pre-recorded-audio/supported-languages>

20 **Newsletter #37: Speaker Diarization Now in 5 New Languages & Latest Speech AI tutorials**

<https://www.assemblyai.com/blog/assemblyai-newsletter-37>

21 **Speech to Text API | Speech Recognition Service - Rev AI**

<https://www.rev.ai/>