



## 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)                           |
|-----------------------------|-----------------|-------------------|-------------------|--------------------|---|--|
| <b>Whisper<br/>(OpenAI)</b> | 2022-09<br>①    | ○ ②               | ○ (フレーズ単位) ②      | ×                  | Tiny 39M～<br>Large 1.55B<br>③ (Largeで約10GB VRAM)  | 無料 (MIT)<br>③                          |
| <b>WhisperX<br/>(拡張版)</b>   | 2023-07<br>(推定) | ○ (Whisper 準拠)    | ○ (単語レベル)         | ○                  | Whisper Large相当 (~1.5B)                           | 無料 (BSD-2-Clause)<br>④                 |
| <b>Google Cloud STT</b>     | 2016-03<br>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)<br>⑨                        |
| <b>Meta (MMS)</b>           | 2023-05<br>⑩    | ○ (1100言語以上)<br>⑪ | ○ (内部モデルで対応)      | ×                  | 10億 ⑫   | 無料 (研究用、CC-BY-NC)<br>⑬                 |
| <b>NVIDIA Riva</b>          | 2021-11<br>⑭    | ○ ⑮               | ○ (Conformer ベース) | ○ (ストリーミング対応)<br>⑯ | 例: Parakeet-CTC 0.6B,<br>Parakeet-RNN-T 1.1B<br>⑰ | 無料 (NVIDIA ライセンス、企業向け有償サポート有)<br>⑯ ⑰ ⑱ |
| <b>AssemblyAI</b>           | 2018-01?<br>⑯   | ○ ⑯               | ○                 | ○ (日本語話者分離対応)<br>⑳ | 非公開 (クラウド)  | 有料 (API利用)                             |
| <b>Rev AI<br/>(Rev.com)</b> | 2016-03?<br>㉑   | ○ (58言語対応)<br>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/WhisperX ① ③ ④ 、 Google Cloud STT ⑤ ⑥ ⑦ ⑧ 、 Meta MMS ⑩ ㉒ 、 NVIDIA Riva ⑭ ⑮ ⑯ 、 AssemblyAI ⑯ ⑳ 、 Rev AI ㉑ (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/>