



| **Platform** | **Total Languages** | **Points** |
| --- | --- | --- |
| ElevenLabs  <https://elevenlabs.io/speech-to-text> | 99+ languages | * Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Punjabi, Sindhi, Tamil, Telugu. * Supports real-time streaming for English only. * Not clearly stated whether it supports Automatic Language Detection. |
| AssemblyAI  [https://www.assemblyai.com](https://www.assemblyai.com/) | 99+ languages | * Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Punjabi, Tamil. * Supports real-time streaming, but Indic support limited. * Supports Automatic Language Detection followed by Transcription. |
| GoogleSpeech-to-Text  <https://cloud.google.com/speech-to-text> | 125+ languages | * Hindi,Gujarati,Kannada,Malayalam,Marathi,Punjabi,Tamil,Telugu * Supports real-time streaming with these Indic languages * Supports Automatic Language Detection followed by Transcription |

| **Feature** | **Bhashini**  https://bhashini.gov.in/ulca/model/explore-models | **Sarvam**  <https://www.sarvam.ai/apis/speech-to-text> | Our Model |
| --- | --- | --- | --- |
| Speech Recognition | 13 languages | 11 languages (Saarika) | 23 languages |
| Input Language Parameter | Requires language input | Requires language input | Doesn’t require language input |
| Core Languages | Hindi, English, Bengali, Tamil, Telugu, Gujarati, Kannada, Malayalam, Marathi, Punjabi, Odia, Urdu, Assamese | Hindi, Bengali, Tamil, Telugu, Gujarati, Kannada, Malayalam, Marathi, Punjabi, Odia, English | Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Odia, Punjabi, Sanskrit, Santali, Sindhi, Tamil, Telugu, Urdu, English |
| Audio Duration and Processing Time | Lengthy audios are taking more than 4 minutes processing time. (tested with audio of size 3 minutes) | Audio files longer than 30 seconds are not supported; the batch API should be used for longer audios. Lengthy audios cannot be processed using sarvam.ai. | Can process lengthy audios and the processing time is 10-15 seconds. |
| Interface  Functionality | Transcription happens **after recording stops** | Transcription happens **after recording stops** | Supports **real-time streaming transcription** + additional features (Semantic Q&A search). |