SESTEK

Supported Languages

This page provides a comprehensive list of languages supported by Knovvu Speech Recognition.

Single Models

We currently support the following languages through the endpoint:

Arabic, Azerbaijani, Bulgarian, Croatian, Czech, Danish, Dutch, English, Farsi, Finnish, Flemish, French, German, Greek, Hindi, Indonesian, Italian, Japanese, Kazakh, Kurmanji-Kurdish, Korean, Latvian, Malay, Mandarin, Mongolian, Norwegian, Pashto, Polish, Portuguese, Romanian, Russian, Spanish, Swahili, Swedish, Tagalog, Tamil, Turkish, Ukrainian, Urdu, Welsh.

Multilingual Models

We currently support the following multilingual models through the endpoint:

Model Name	Language Supported
EnglishTurkish	English, Turkish
DutchFrench	Dutch, French
ArabicEnglish	Arabic, English
ArabicEnglishTurkish	Arabic, English, Turkish
EnglishFrenchTurkish	English, French, Turkish
EnglishSpanish	English, Spanish
EnglishMalay	English, Malay
LatvianRussian	Latvian, Russian
Mena	Arabic, English, French, Portuguese, Spanish
North America	English, French, Portuguese, Spanish
Europe	English, French, Portuguese, Spanish, Dutch, German, Italian
Asia	Mandarin, Tamil, Malay, English
FourLanguagesMulti	English, Arabic, French Spanish
SixLanguagesMulti	English, Turkish, Arabic, Russian, French Spanish
Large	Arabic, Danish, Dutch, English, Finnish, French, German, Hindi, Italian, Latvian, Mandarin, Norwegian, Portuguese, Russian, Spanish, Swedish, Tagalog, Turkish, Urdu

When to Use Single vs. Multilingual Models

Single Models

Single models should be used when the language of the input is known before processing. These models provide the highest accuracy because they are specifically trained for a single language. Single models are ideal for:

- IVR-driven systems where the caller selects a language before speaking.
- Speech analytics for monolingual environments, such as customer service centers operating in a single language.
- Use cases where high accuracy is required, and the language does not need to be detected dynamically.

Multilingual Models

Multilingual models should be used when the language of the input is unknown at the time of recognition or when dynamic language switching is required. These models work best for:

- When the IVR system does not provide language information, making it impossible to route the call to a specific monolingual SR model.
- Applications that support multiple languages but do not allow dynamic switching between single models.
- Use cases where users may speak different languages interchangeably, but each utterance is typically monolingual.

In such cases, multilingual SR models provide high-accuracy transcriptions by detecting and transcribing the input in one of the languages they were trained on. These models work best with monolingual utterances—where the entire speech is in one of the supported languages.

Limitations with Code-Switching

Code-switching refers to scenarios where speakers mix multiple languages within the same utterance, such as a Turkish sentence containing French words. Multilingual SR models are not explicitly trained to handle code-switching, as their training data primarily consists of monolingual examples for each supported language.

As a result, multilingual models do not significantly improve the recognition of foreign words embedded in another language compared to a dedicated monolingual model. For example, an EnglishFrenchTurkish model is not necessarily better at recognizing a French word in a Turkish sentence than a standard Turkish model, since such occurrences are rare or absent in the training data.

Recommended Approach for Code-Switching Scenarios

For improved recognition of foreign words within a sentence, we recommend leveraging context-biasing (pronunciation support) available in our end-to-end (E2E) models. This approach enhances recognition accuracy for specific terms, providing a more effective solution for code-switching cases compared to
relying solely on a multilingual SR model.

> Info
>
> Multilingual models provide flexibility in handling diverse language scenarios but may not be as accurate as single models in cases where the input is already known. If the use case involves frequent code-switching within a single utterance, context-biasing techniques should be considered to improve recognition accuracy.

Request Sample with Curl:

JSON Copy

```
curl --location '{{Address}}/v1/speech/dictation/request' \
--header 'ModelName: FourLanguagesMulti' \
--header 'ModelVersion: 2' \
--header 'Content-Type: audio/wave' \
--header 'Authorization: Bearer Token
--data 'audio file.wav'
```

You can specify the language within a recognition request's ModelName parameter. For detailed instructions on sending a recognition request and specifying the language for transcription, please refer to the API Reference on performing speech recognition.

Third-Party Models

Knovvu Speech Recognition also integrates third-party models to expand language support and offer additional flexibility in various use cases.

Whisper Models

As part of Knovvu's commitment to providing comprehensive speech recognition solutions, we have strategically integrated Whisper models into our supported languages. This decision allows us to bring additional language support to Knovvu SR, particularly in regions or dialects we do not directly support.

Performance and Limitations

It is important to note that while Whisper models enable access to additional languages, Sestek does not assume responsibility for the performance or accuracy of these third-party models. Results may vary based on language and environmental factors, and Whisper models may not achieve the same performance standards as Knovvu's core SR technology. We recommend testing these models within specific environments to ensure they meet your project requirements.

Supported Languages for Whisper Models

We currently support the following languages through the endpoint:

Afrikaans, Amharic, Arabic, Assamese, Azerbaijani, Bashkir, Belarusian, Bulgarian, Bengali, Tibetan, Breton, Bosnian, Catalan, Czech, Welsh, Danish, German, Greek, English, Spanish, Estonian, Basque, Persian, Finnish, Faroese, French, Galician, Gujarati, Hawaiian, Hausa, Hebrew, Hindi, Croatian, Haitian Creole, Hungarian, Armenian, Indonesian, Icelandic, Italian, Japanese, Javanese, Georgian, Kazakh, Khmer, Kannada, Korean, Latin, Luxembourgish, Lingala, Lao, Lithuanian, Latvian, Malagasy, Māori, Macedonian, Malayalam, Mongolian, Marathi, Malay, Maltese, Burmese, Nepali, Dutch, Norwegian Nynorsk, Norwegian, Occitan, Punjabi, Polish, Pashto, Portuguese, Romanian, Russian, Sanskrit, Sindhi, Sinhala, Slovak, Slovenian, Shona, Somali, Albanian, Serbian, Sundanese, Swedish, Swahili, Tamil, Telugu, Tajik, Thai, Turkmen, Tagalog, Turkish, Tatar, Ukrainian, Urdu, Uzbek, Vietnamese, Yiddish, Yoruba, Yue Chinese, Chinese.

Model Options

To accommodate diverse performance needs, Knovvu offers two versions of Whisper models: **WhisperTiny** and **WhisperTurbo**. Each model provides a distinct balance of speed and recognition accuracy:

- **WhisperTiny:** This is a smaller, lightweight model designed for faster performance. While WhisperTiny provides quick processing speeds, its recognition accuracy may be limited in comparison to larger models. This option is ideal for applications where speed is prioritized over the highest accuracy levels, or where computing resources are limited.
- **WhisperTurbo:** Equivalent to the advanced "large v3 turbo" Whisper model, WhisperTurbo represents the latest in Whisper technology, offering improved recognition accuracy. However, this model requires more processing power and operates at a slower speed than WhisperTiny. WhisperTurbo is suited for applications that demand high accuracy, especially in complex or noisy environments, but can accommodate slower processing times.

Integrating Whisper Models

To utilize Whisper models in the Knovvu SR API, users simply need to specify the desired model in the ModelName parameter. Enter either WhisperTiny or WhisperTurbo as the value. There is no need to specify the language, as Whisper models automatically detect and process supported languages.

ABOUT SESTEK

SESTEK is a conversational automation company helping organizations with conversational solutions to be data-driven, increase efficiency and deliver better experiences for their customers. Sestek's AI-powered solutions are built on text-to-speech, speech recognition, natural language processing and voice biometrics technologies.

SESTEK is a part of UNIFONIC

www.sestek.com

Copyright © 2025 SESTEK. All rights reserved.