Benefits of Using Audio-To-Text Technology in Business

TechRounder PDF Edition

Live article: https://www.techrounder.com/technology/benefits-of-using-audio-to-text-technology-in-business/

By Vipin PG | Published January 25, 2023 | Updated March 8, 2026 | Format: Article | 7 min read

In brief

Audio-to-text technology powered by AI is used for hands-free taking notes, live labeling, offering better customer support, and much more. Audio-to-text is used to:.

Audio-to-text technology powered by AI is used for hands-free taking notes, live labeling, offering better customer support, and much more. Audio-to-text is used to:

- produce emails
- give valuable comments in the form of meeting and event transcripts
- enable accessibility swiftly and effectively

Audio-to-text technologies promote workplace inclusivity and assist everyone in completing jobs more effectively. It's intended to become more competent with each usage, eventually taking over duties that people have previously performed. In addition, audio-to-text technology can make or break both content and workplaces for those with impairments, such as those who are deaf or have hearing loss.

What exactly is audio-to-text?

Audio-to-text software is voice recognition software that is frequently based on AI. It uses computational linguistics to recognize and translate spoken conversation into text. Audio-to-text creates transcripts, captions, and other written information that businesses require today. It operates by translating speech into printed word-for-word representations. When you use Siri or watch films with subtitles, you most likely witness voice-to-text in action.

The Automatic Speech Recognition (ASR) technique converts audio-to-text. The technology that converts speech or an audio source into text is known as ASR. The text is created using the expertise of linguistics, computer science, and electrical engineering. It's frequently used as the foundation for captioning and transcribing solutions.

What is the best method for converting audio-to-text?

Converting voice-to-text is done manually or automatically using built-in solutions for your devices and platforms. However, this isn't advised. Manually converting text sometimes takes time and effort. In addition, many automatic solutions leave you with mistakes, which won't create a professional feel or access for persons with disabilities.

In this situation, accuracy refers to the number of correct predictions provided by a different speech model or human assistance. Greater precision correlates to an excellent performance by the audio-to-text supplier, which is especially significant for people with impairments who depend on audio-to-text technologies at work.

Why is precision necessary while converting audio-to-text?

Because of their lack of accuracy, automatic speech-to-text techniques (without human intelligence) are insufficient to ensure fairness. According to Google, 27% of the worldwide internet population uses voice search on smartphones, but how many automatic audio-to-text tools are genuinely accurate? Moreover, while Siri and Google Assistant are helpful and entertaining, they don't always translate audio-to-text precisely as intended.

Cell phone speech-to-text issues

Inaccuracies in telephone numbers are a classic illustration of this. When pronouncing numerals out loud, one has to use 'oh' instead of 'zero' or double/triple digits such as 'triple three.' Context is also essential since language has many subtleties and ambiguities to consider. For example, "Pounds" either refer to weight or cash.

For organizations wishing to develop professional transcripts, audio-to-text conversion must be as precise as possible to speed rather than impede the workflow. Working with a service like an audio-to-text converter, which employs human editors in addition to automated technologies, provides the most significant degree of accuracy.

The benefits of using speech-to-text

Using voice recognition to translate audio and video into correct text allows corporate operations to run more smoothly and effectively while increasing accessibility. Some of the most frequent business applications for voice-to-text include:

1. Customer calls: Transcribing customer calls using audio-to-text helps you have a record and document to draw out meaningful info from customer talks easily. These transcripts give valuable input that is useful for improving both customer engagement and employee performance.
2. Searchable corporate content: It's helpful in indexing audio and video data. Searchable transcripts are especially useful for HR departments, marketing departments, and event producers who reference chats or extract quotations from conferences, podcasts, or other content they're streaming or recording. Furthermore, including transcripts with video material helps the information to be SEO-friendly since browsers such as Google scans the transcripts and put them higher in search results. This feature thus aids in the discovery of businesses and their content.
3. Accessibility for live meetings and events: Audio-to-text technology assists businesses in creating live video subtitles for regular conferences and huge events. Providing subtitles helps create knowledge support for all participants and serves as a handy tool when guests must listen in without a sound, and it's also necessary for accessibility. In addition, speech-to-text and manual editing support the requirement to make audio-visual material accessible to people who are deaf or hard of hearing.
4. Reporting and note-taking: Several companies and sectors use audio-to-text technology to take notes in real-time or to have notes to refer to following calls. Audio-to-text eliminates the need for professionals to scribble down notes manually, allowing them to focus more on the discussions, interviews, or events they attend.

More and more organizations are turning to Artificial Intelligence (AI) and voice-to-text tools, often recognizing that these technologies are enabling them to run more effective operations. However, while the advantages of converting audio-to-text appear limitless, it's critical to acquire the most accurate results possible for accessibility and professionalism.

Best practices for data submission to the Speech-to-Text API

These are some suggestions for supplying voice data to the Speech-to-Text API. These recommendations are intended to improve efficiency and accuracy and provide appropriate service response times. The Speech-to-Text API works best when the data given to the service falls inside the following parameters:

For best results,

- Use a lossless codec to capture and transmit audio with a sampling rate of more than 16,000 Hz. It's best to use FLAC or LINEAR16.
- Without further noise-canceling, the recognizer is designed to disregard background speech and noise. However, for best performance, place the microphone as near to the user as feasible, especially if there is background noise.
- If you're collecting audio from more than one individual, send each channel separately to achieve the most significant recognition results. Send the recording as is if all speakers are blended in a single-channel recording.
- Use word and phrase clues to expand the vocabulary and improve the accuracy of certain words and phrases.
- Use StreamingRecognize with a single utterance set to true for brief searches or commands, which improves recognition for short utterances while simultaneously reducing delay.

Avoid performing the following if possible:

- Lower sample rates result in decreased accuracy. Re-sampling, on the other hand, should be avoided. In telephony, for example, the native rate is typically 8000 Hz, which is the rate you supply to the service.
- Using lossy codecs such as mp3, mp4, m4a, mu-law, a-law, or others during recording or transmission diminishes accuracy. If your audio is already in an encoding that the API doesn't allow, convert it to lossless FLAC or LINEAR16. If you have to employ a lossy codec to save bandwidth, using AMR WB, OGG OPUS, or SPEEX WITH HEADER BYTE codecs is recommended.
- Excessive background noise and echoes degrade accuracy, mainly if a lossy codec is used.
- Many people speaking simultaneously or at varying volumes are considered background noise and disregarded.
- The recognizer has a broad vocabulary; nevertheless, phrases and proper names that aren't in the recognizer's language won't be recognized.
- For brief queries or command usages, use Recognize or LongRunningRecognize.
- Rate of sampling: Set the audio source's sample rate to 16000 Hz if feasible. Otherwise, adjust the sample rate hertz to match the audio source's native sample rate (instead of re-sampling).
- Size of the frame: Streaming recognition detects live audio taken from a microphone or other audio source. The audio stream is divided into frames and sent in a series of StreamingRecognizeRequest messages. Any size frame is allowed. Larger frames are more economical, but they introduce delay. A frame size of 100 milliseconds is a suitable compromise between delay and efficiency.
- Pre-processing of audio: Delivering as pure an audio as possible is essential by employing a high-quality, well-positioned microphone. On the other hand, noise-reduction signal processing on the audio before transmitting it to the service usually affects recognition accuracy. The service is intended to handle loud audio.

For optimal results, follow these steps:

Place the microphone as near to the person speaking as feasible, mainly if background noise exists.

- Prevent audio clipping
- Avoid using automatic gain control (AGC)
- Disable all noise reduction processes

- Play some audio samples. It has to be free of distortion and unexpected noise.
- Configuration request

Ensure that the audio data submitted with your request to the Speech-to-Text API is described appropriately. Help ensure that your request's RecognitionConfig represents the suitable sample RateHertz, encoding, and languageCode ensures that your request is transcribed and billed as accurately as possible.

With a lengthy history of research and invention, the advent of these AI voice-controlled assistants, or digital assistants, into the speech recognition industry in the twenty-first century revolutionized the landscape of this technology.

Audio-to-text technology applications

Audio-to-text technology and digital assistants have swiftly expanded from our cell phones to our homes. They are used in business, banking, marketing, healthcare, IoT, and language learning.

Applications in the future

Audio-to-text technology is still evolving. Still, based on the principle of hyper adoption, which states that customers prefer to acquire new technologies faster than they did previously, it's expected that this technology will expand and improve swiftly.

At this stage in the technology's life cycle, it is critical to know its potential and possibility of becoming common in our everyday lives. As a result, businesses are proactive in adopting audio-to-text technology into their digital marketing plan and budget. At the same time, people continue investigating audio-to-text benefits in their daily activities.

As accuracy rates rise and customer buy-in grows, companies anticipate being forced to adjust to becoming more speech-centric and, hence, more human than previously thought conceivable.

References

1. usabilitygeek.com - automatic-speech-recognition-asr-software-an-introduction - https://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/
2. happyscribe.com - audio-to-text - https://www.happyscribe.com/audio-to-text
3. xda-developers.com - lossless-audio - https://www.xda-developers.com/lossless-audio/