Introduction: Text-to-speech

Speechify is an AI-powered text-to-speech (TTS) voice generator software that uses human-like voices to read text. The primary objective of Speechify is to improve one's comprehension, retention, and understanding of written material.

Speechify can be a valuable tool for students looking to improve their reading speed and listening skills. It provides audio versions of written texts in various languages, allowing students to listen to native and near-native speakers' voices. This aids them in learning while multitasking, managing their daily activities, and expanding their studying time.

Individuals with dyslexia and ADHD can benefit from Speechify, as it can reduce reading time and increase reading speed.

Speechify can also generate voiceovers in more than 40 languages, which caters to the diverse needs of not only English-speaking students but also Spanish, Mandarin, and any other language. Additionally, learners can customize the reading speed, choose their preferred voice, and make other settings according to their preferences. This level of personalization enables learners to tailor their education experience to their specific needs and learning styles, ultimately enhancing their study sessions' effectiveness.

Speechify's TTS technology increases accessibility for everyone, especially individuals with vision problems such as blindness, dyslexia, and ADHD.

For English TTS, Speechify offers AI-made voice selections of US celebrities such as Gwyneth Paltrow and Snoop Dogg; for other languages, narrators have native names.

Speechify has free and paid versions; most features are available for $139.99/year.

Background: From text to speech

The text-to-speech (TTS) concept dates back to the 18th and 19th centuries when inventors started experimenting with various speech synthesis techniques. Wolfgang von Kempelen made the earliest attempts at TTS in the 18th century with his mechanical speaking machine, which used bellows and reeds to produce sounds resembling human speech. However, as an electronic device, a Japanese researcher, Noriko Umeda, created the first English text-to-speech electronic device in 1968.

The primary function of a reading machine is to make written materials accessible to people who are blind through text-to-speech technology.

By the early 1980s, automatic speech recognition had advanced enough to allow for practical speech-driven data entry systems.

In the late 1970s, the first reading machine for people who are blind was made possible with optical character recognition combination and synthetic speech.

The Advanced Research Projects Agency and major corporations such as IBM, AT&T, and Texas Instruments have provided significant funding for speech technology research. Advancements in microprocessors and decoding algorithms have enabled the creation of voice command systems for computers and telephones since the 1990s.

Commercial PC-based applications grew significantly in the early 1990s with new platforms and tools.

Many early commercial systems were offered with little consideration of the difficulty of implementing these semi-human technologies. Recently, speech interface technology systems have attempted to overcome the natural limitations of speech understanding with more user-friendly interfaces.

Analysis: Advantages and disadvantages

- **Advantages of TTS:** Text-to-speech (TTS) technology is a beneficial tool for people with visual impairments, enabling them to access textual information easily. It works by converting written text into spoken words, which can be heard through a computer or mobile device. TTS technology uses algorithms to interpret written text, resulting in accurate pronunciation and intonation of words, making it an excellent alternative to traditional text-based communication.

  Not only does TTS technology allow visually impaired individuals to access information, but it also enables them to control the speed and the narrator's style of speaking, making it easier to understand. For example, users can slow down the reading speed if needed. Additionally, TTS technology provides an alternative means of accessing content without requiring the use of hands or eyes, allowing individuals to multitask, drive, or exercise while still accessing digital content.

  TTS systems can also enhance the user experience on digital platforms by providing an alternative mode of interaction that increases engagement and satisfaction. For instance, TTS can be used in e-books, websites, and mobile applications, allowing users to listen to content instead of reading it. Moreover, TTS technology enhances inclusivity in digital content by providing an alternative means of consuming information for individuals with dyslexia or other vision difficulties. This contributes to a more equitable and inclusive digital environment.

  TTS technology comes with several crucial features, such as clarity, intonation, punctuation, and pronunciation, making it an effective tool for promoting equality among

genders, accents, geographical locations, and age groups. With TTS, users can customize the voice to reflect their preferred gender, accent, or language. This quality is beneficial for people who may struggle with traditional text-based communication due to cultural or linguistic differences.

The effectiveness of the TTS system will be compromised if these features are not met.

- **Disadvantages of TTS:** Learning with the help of an inefficient text-to-speech (TTS) system can be a challenging and frustrating experience. For instance, the TTS system Speechify may read in a monotonous tone or may not follow the punctuation in the text. As a result, words may lose their intended meaning, regardless of their grammatical role in a sentence. The system only minimally mimics human emotions, such as pauses, breaths, and intonation.
  Speechify's reading settings can be customized, but the system occasionally fails to comply with rules or update itself.

  Text-to-speech (TTS) software has been developed to assist in various tasks such as reading aloud written text, generating synthesized speech, and providing audio feedback. While TTS software can be helpful, it may not be the most effective choice for some topics, such as language learning.

  One of the main issues with TTS software is that it lacks the ability to convey intonation nuances characteristic of native speakers. This is a significant drawback because intonation plays an essential role in language learning. TTS software cannot replicate the rise and fall of a speaker's voice, which can affect the meaning of a sentence. For example, a statement can become a question in Farsi and German just by changing the intonation.

  Furthermore, when it comes to learning a new language, having a teacher or instructor to guide you can be more beneficial than relying solely on TTS. A teacher can help you with pronunciation, grammar, and vocabulary and provide feedback on your progress. A TTS software, on the other hand, cannot provide you with the same level of personalized instruction.

  For example, Speechify supports multiple languages, including Persian Farsi, which is Iran's official language. However, despite offering Farsi as one of the supported languages, the female voice used in the text-to-speech (TTS) has a Dari accent that is specific to Afghanistan. This is an important distinction that a native Farsi speaker would immediately notice.

Although the TTS feature can be helpful for someone who already knows how to pronounce the words correctly, it may not be as beneficial to a beginner's learning experience. In fact, it could even be detrimental, as it may reinforce incorrect pronunciation or usage of certain words. This is particularly true for languages like Farsi, which have complex grammar structures and a wide range of vowel and consonant sounds that can be difficult for beginners to master.

Therefore, while Speechify's text-to-speech feature is a handy tool for certain users, it's important to note that it may not be the best option for beginners who are looking to improve their language skills.

Proposed Solutions:

One way to improve the user experience when dealing with text-to-speech software is to offer an on-demand service that allows users to order a person to read the document. This feature is useful when dealing with non-technical issues, such as intonation and monotoning. By allowing users to choose a human voice to read the text, the software makes it easier to engage with the content.

However, it is important to note that technical issues and glitches can still occur, even with the addition of this service. To minimize these issues, it is necessary to revise the software regularly to ensure it is working correctly and all bugs have been fixed.

Furthermore, when dealing with languages other than English, ensuring the quality of the text-to-speech output is crucial. This includes conducting quality assessments with the help of a native speaker to ensure that the pronunciation, intonation, and phrasing are correct. By taking these steps, you can ensure that your text-to-speech software is accurate, effective, and accessible for all users, regardless of their language or technical proficiency.

Sources:

https://speechify.com/blog/what-speechify-languages-are-available/#:~:text=Speechify%20supports%20English%2C%20Italian%2C%20French,%2C%20Finnish%2C%20and%20many%20others.

https://lovo.ai/post/7-essential-qualities-of-a-good-text-to-speech-platform

https://www.duolingo.com/learn

https://languagetech.lab.uiowa.edu/technology-tools-language-learning#heading526

https://www.pnas.org/doi/pdf/10.1073/pnas.92.22.10007