[?????-??????????]

?????-??????????

?? TTS ????



The global market for text-to-speech (TTS) technology, while small, is growing at a compound annual rate of 15 percent. As a result, there is a spike in demand for services that can convey written text to spoken word with the right tone and voice to reflect a brand’s personality. Let’s take a closer look at how TTS is evolving.

TTS Defined

TTS is a type of technology that converts digital text to spoken word. As the name implies, TTS literally takes a text string and converts it into audio, which is why TTS is sometimes referred to as read-aloud technology. The most popular uses of TTS consist of voices for smart speakers, kiosks, chatbots, and accessibility services.

TTS Evolving

Until recently, businesses were satisfied with TTS applications in which the voice sounded largely robotic. So long as the TTS application did what it was supposed to do, nuances such as the tone of voice did not matter or required extensive editing work through standardized markup languages such as SSML to improve results. But thanks to advances in artificial intelligence, it's possible to make voices sound more human with all the nuances of speech that we associate with how real people talk. This branch of TTS is called neural text to speech. As a result, businesses are applying TTS in some categories where traditionally real human voices were used, such as in tutorial and advertisements.

For example, recently KFC celebrated National Fried Chicken Day *by reinventing the drive-through experience* with a simulation of KFC's international icon, Colonel Sanders. During the campaign, a voice-based Col. Sanders head gave drive-through customers a humorous experience of ordering from Col. Sanders himself. The experience used speech recognition, AI, and TTS and to make a KFC drive-through operator's voice sound as though Col. Sanders were speaking in a southern drawl evoking KFC's Kentucky roots. Here, TTS helped inject personality and humor into a global brand by enabling a playful experience.

Vyond, one of the largest platforms to create animated advertisements, is relying on TTS to create content that combines rich animation and voice. Vyond relies on the *Amazon Polly* text-to-speech technology to deliver spoken content from a machine that incorporates tones, accents and languages that people would expect from other people. For example, Kapitec Software uses Vyond to create whiteboard videos for Kapitec's eLearning software. On the Vyond website, Kapitec Software's CEO Sandrine Boarqueiro?Verdu testified to the power of effective TTS. According to Boarqueiro?Verdu, "The text-to-speech voices are very natural-sounding. Although we mainly use the French voices, having an extensive language selection provides us with the option to localize content for different regions. Our videos are well-received by our customers and viewed more than other content."

Why TTS is becoming more popular

A big reason why businesses are using TTS to emulate actual human voices is that neural TTS is getting better. Thanks to neural networks, TTS can understand the emotional inflection and rhythm of the voice based on the string's recognized intent – when to express sadness or a surprise. The ability to sense emotion has always been a drawback of TTS, but this issue is increasingly less of an obstacle. As a result, businesses can use TTS to replace voice actors for functions such as narrating corporate videos, ads, games, and other content.

As AI-enabled TTS improves, businesses can realize a number of benefits. For example, businesses can realize faster turnaround times and more cost-efficient production. That's because with the right parameters, a machine can translate text to voice perfectly, without requiring do-overs that inevitably happens with voice actors.

Also, in human-in-the-loop type of workflows, this technology allows linguists to make changes (to prosody, pitch, rate, and pronunciation), guaranteeing even better results thanks to the post-editing and the help of standards for speech synthesis editing such as SSML.

Where TTS is Headed

TTS is evolving in many ways. Voice cloning, for instance, can capture your brand essence and express it via a machine. With voice cloning, you can use TTS along with voice recordings data sets to incorporate the voices of recognizable people such as executives and celebrities, which can be useful for businesses in areas such as entertainment.

Cheetah Mobile is an example of a company that is moving to another branch of speech synthesis called speech-to-speech (basically speech in a source language translated in speech in a target language). The company recently rolled out on a larger scale a version of its CM Translator, a hand-held translation device. As Cheetah Mobile noted in a press release, the tool helps American travelers communicate abroad when performing tasks such as asking directions, and it can be useful for someone who has moved to the United States and needs assistance.

How to Think about TTS

We believe it is important for businesses to think about TTS intelligently. If you are considering the use of TTS, keep in mind a few considerations:

- TTS is not effective for videos where you need a person onscreen. When machine voices speak too quickly, they have a low quality. You lose the turnaround time and cost advantage.
- Not all text is ideal for conversion to speech. People comprehend information differently through with ears than with their eyes. Our brains process simpler, easier-to-digest content through our ears, which is why spoken ideas are more powerful when broken into smaller chunks of information.
- TTS does not replace people. People can still understand a given stimulus and respond to it better than a machine can. For instance, a human being can read emotions and respond with the proper voice inflection better than a machine can, and people can understand context better than a machine. In addition, human voices are still more ideal for live action narration. Machines cannot adjust quickly enough to changes in pace (such as narrating a live sporting event).

At Centific, we help companies incorporate TTS, employing our knowledge of AI, TTS, SSML and language translation expertise. Contact us for more insight on how we can help you.

-
-