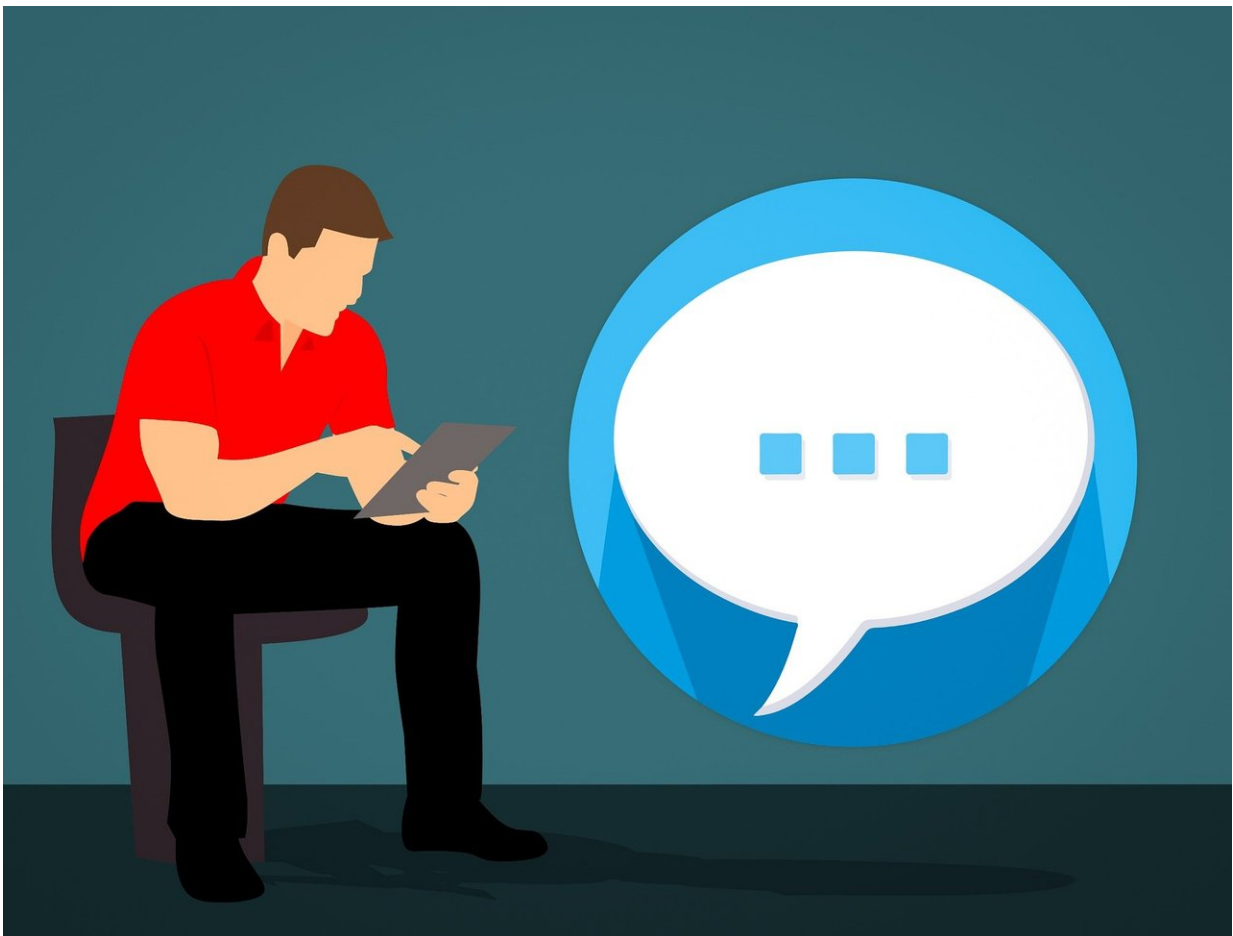


Introducing Cloud Text-to-Speech service for developers

March 29 2018, by Nancy Owano



Credit: CC0 Public Domain

The Google Cloud Platform Blog issued a Tuesday announcement,

introducing Cloud Text-to-Speech.

Dan Aharon, Product Manager, Cloud AI, said, "Developers have been telling us they'd like to add text-to-speech to their own applications, so today we're bringing this technology to Google Cloud [Platform](#) with Cloud Text-to-Speech."

Cloud Text-to-Speech is all about text to speech conversion powered by machine learning.

As an API, said the website for Cloud Text-to-Speech, you can create interactions with users, across applications and devices. Cloud Text-to-Speech supports applications or devices that can send a REST or gRPC request. That includes phones, PCs, tablets and IoT devices (e.g., cars, TVs, speakers).

What real-world applications would apply? Use-cases include call center automation and interactive responses from IoT devices.

He said that Cloud Text-to-Speech is already helping customers deliver a better experience to their end users.

(Robert Hof of *SiliconANGLE* said that "Several dozen alpha users have been [trying](#) it since November.")

Customers include Cisco and Dolphin ONE. The latter integrated Cloud Text-to-Speech into its products; their users can create "natural call center experiences."

What Is Google Cloud Platform? This is a suite of cloud computing services running on the same infrastructure that Google uses internally for products such as Google Search and YouTube. Now, said Frederic Lardinois in *TechCrunch*, "developers will get access to the same

DeepMind-developed text-to-speech engine that the company itself is current using for its Assistant and for its Google Maps direction."

Enter WaveNet neural network architecture—which directly generates a raw [audio waveform](#).

Aharon blogged, "Cloud Text-to-Speech also includes a selection of high-fidelity voices built using WaveNet, a generative model for raw audio created by DeepMind. WaveNet synthesizes more natural-sounding speech and, on average, produces speech audio that people prefer over other [text-to-speech](#) technologies."

The Cloud Text-to-Speech carries advanced speech technology; Deep Mind's research in machine learning models to generate speech that mimics human voices has succeeded. The speech sounds natural, and its team claimed it reduced the gap with human performance by over 50%.

Lardinois pointed to what makes WaveNet's contribution to speech special:

"Unlike previous efforts, WaveNet doesn't do speech synthesis based on a collection of short speech fragments, which tends to create the kind of robotic sounding voices you are surely familiar with. Instead, WaveNet models raw audio using a machine-learning [model](#) to create a far more natural-sounding speech."

Lardinois also provided a brief history of WaveNet and how it addressed all-important speed of response.

"Google first talked about WaveNet about a year ago. Since then, it moved these tools to a new infrastructure that sits on top of the company's own Tensor Processing Units. This allows it to generate these audio waveforms 1,000 times faster than before, so generating a second

of audio now only takes 50 [milliseconds](#)."

It lets developers synthesize natural-sounding speech with 30 voices. Moreover, it is available in multiple languages and variants. The site said it Supports 32 voices in 12 languages and variants.

(This writer [tried](#) it out in two languages. It seemed excellent in both attempts.)

Frederic Lardinois in *TechCrunch* pointed out that developers will be able to customize the pitch, speaking rate and volume gain of the MP3 or WAV files the service will generate.

Aharon in the blog provided a link for [pricing](#) information and for [documentation](#).

More information: cloudplatform.googleblog.com/2018/03/eNet-technology.html

© 2018 Tech Xplore

Citation: Introducing Cloud Text-to-Speech service for developers (2018, March 29) retrieved 11 September 2024 from <https://techxplore.com/news/2018-03-cloud-text-to-speech.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--