



## Voice-powered solutions with Cloud AI

*Dan Aharon<sup>1</sup>*

<sup>1</sup>Google, Inc.  
daharon@google.com

### Abstract

Google has been a leader in research into speech recognition, speech synthesis and natural language and much of that innovation has powered Google products like Search, Google Assistant, Google Maps, Android and others.

Google Cloud AI aims to help enterprises, startups, students and any other developer use AI to build great end user experiences and benefiting from investments Google has made in this space.

Google Cloud Speech-to-Text (formerly known as Cloud Speech API) was first released in Beta in 2016, was followed by Dialogflow Enterprise Edition in 2017 and Cloud Text-to-Speech in 2018. Each of these products is backed by cutting-edge research that Google has conducted in these spaces.

Google Cloud Speech-to-Text uses the same speech recognition that powers Google products. Google Cloud Speech-to-Text enables developers to convert audio to text by applying powerful neural network models in an easy to use API. The API recognizes 120 languages and variants. You can enable voice command-and-control, transcribe audio from call centers, and more. It can process real-time streaming or pre-recorded audio, using Google's machine learning technology.

In April 2018 we introduced the biggest update to the service in 2 years and announced new models that improve the transcription quality for phone calls by 55%+ and videos by 65%+. Cloud Speech-to-Text is considered by many to be the leading speech recognition product.

Dialogflow is an end-to-end development suite for building conversational interfaces for websites, mobile applications, popular messaging platforms, and IoT devices. You can use it to build interfaces (e.g., chatbots) that are capable of natural and rich interactions between your users and your business. It is powered by machine learning to recognize the intent and context of what a user says, allowing your conversational interface to provide highly efficient and accurate responses. Dialogflow Enterprise Edition is the enterprise tier of Dialogflow and is part of Google Cloud Platform.

Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 30+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's powerful neural networks to deliver the highest fidelity possible. You can create lifelike interactions with your users, across many applications and devices.

DeepMind's WaveNet is a brand way of generating speech which far exceeds concatenative and parametric speech synthesis in quality. However, Google had to make WaveNet 1000x faster and better before it was ready to be made available at scale. This involved many hardware, software and machine learning components coming together to solve the problem.

This set of products is expected to be generally available before this conference starts and we're excited to see how participants would use them to build new speech-powered experiences including mobile apps, connected devices (cars, TVs, speakers), robots, etc.

We will showcase multiple demos at the show, that show how you could build your own personal shopper or personal assistant using Cloud AI technology.

We will also show how easy it is to connect this to a phone line and turn this into a conversational IVR that can assist your users the way they'd want.