Technical Disclosure Commons

Defensive Publications Series

October 02, 2017

SENDER VOICE MODELING FOR AUDIBLE PLAYBACK

Google Inc.

Follow this and additional works at: http://www.tdcommons.org/dpubs_series

Recommended Citation
Inc., Google, 'SENDER VOICE MODELING FOR AUDIBLE PLAYBACK', Technical Disclosure Commons, (October 02, 2017)
http://www.tdcommons.org/dpubs_series/732

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
SENDING VOICE MODELING FOR AUDIBLE PLAYBACK

ABSTRACT

A virtual, intelligent, or computational assistant (e.g., also referred to simply as an “assistant”) is described that is configured to use a voice modeled on a sender of text (e.g., text message, email, instant message, webpage, etc.) when synthesizing audio data to “read” the text aloud (e.g., narrate, generate text-to-speech (TTS), etc.). For instance, when a first user receives a text message from a second user, a computing device of the first user may perform TTS to read the received text message aloud using a voice that is modeled to sound like the second user. The assistant may model the voice of the sender based on voice samples obtained from the sender and/or from a bank of voice models at-which the sender is registered.

DESCRIPTION

Assistants execute on counter-top devices, mobile phones, automobiles, and many other type of computing devices. Assistants output useful information, responds to users’ needs, or otherwise performs certain operations to help users complete real-world and/or virtual tasks. Some assistants may perform TTS operations to read received text aloud by playing audio data synthesized based on the received text. In some cases, an assistant may always use a single voice. In other cases, an assistant may select from a small number of voices but may not vary from sender to sender then performing TTS operations. In many cases, it may be helpful for users to have their assistants perform TTS operations (e.g., disabled users, users who are driving or otherwise occupied, etc.).

The example system shown in FIG. 1 provides an assistant that tailors the voice used to perform TTS operations based on a sender of the text on-which the TTS operations are being
performed. For example, when an assistant is reading aloud a message sent to user A by user B, the assistant may use a voice that is modeled to sound like user B.

**FIG. 1**
The system of FIG. 1 includes one or more external systems and computing devices A–N communicating across a network with each of computing devices A–N executing an assistant that performs operations involving groups of people. The network of FIG. 1 represents a combination of any one or more public or private communication networks, for instance, television broadcast networks, cable or satellite networks, cellular networks, Wi-Fi networks, broadband networks, and/or other type of network for transmitting data (e.g., telecommunications and/or media data) between various computing devices, systems, and other communications and media equipment. Computing devices A–N represent any type of computing device, server, mainframe, cloud computing environment, or other system that is configured to execute an assistant and communicate on a network. The external systems represent any type of server or other computing system that is configured to support the assistants executing at computing devices A–N.

Computing devices A–N can be personal, corporate, or government owned computing devices. In some examples, computing devices A–N may be shared assets of multiple users. Examples of computing devices A–N include mobile phones, tablet computers, wearable computing devices, countertop computing devices, home automation computing devices, laptop computers, desktop computers, televisions, stereos, automobiles, and any and all other type of mobile and non-mobile computing device that is configured to execute an assistant. For example, computing device A may be a countertop assistant device and computing device N may be a mobile phone or automobile infotainment system.

An assistant executes across any combination of external systems one or more of computing devices A–N to provide assistant services to users of computing devices A–N. Examples of assistant services include: setting up reminders, creating calendar entries, booking travel, online ordering, sending messages or other communications, reading text aloud,
controlling televisions, lights, thermostats, appliances, or other computing devices, providing navigational instructions, or any other conceivable task or operation that may be performed by an assistant.

As a user interacts with the assistant, the assistant may obtain personal information about the user. Examples of personal information include: habits, voice samples, routines, preferences, notes, lists, contacts, communications, interests, location histories, and other types of user information. After receiving explicit permission from the user, the assistant may store, the personal information at user information data stores and in the course of providing assistant services, make use of the personal information stored at the user information data stores.

The external systems and computing devices A–N and the assistant treat the information stored at the information stores so that the information is protected, encrypted, or otherwise not susceptible to unauthorized use. The information stored at the information data stores may be stored locally at each of computing devices A–N and/or remotely (e.g., in a cloud computing environment provided by the external systems and which is accessible via the network of FIG. 1).

Further to the descriptions below, a user may be provided with controls allowing the user to make an election as to both if and when the assistant, the computing device, or the computing systems described herein can collect or make use of supplemental data (e.g., user information or contextual information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if and when the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user’s identity may be treated so that no personally identifiable information can be determined for the user, or a user’s geographic location may be generalized where location information is
obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what supplemental data is collected about the user, how that supplemental data is used, and what supplemental data is provided to the user.

The assistant can perform TTS using a voice of a sender in one of many ways. As one example, the assistant may model a voice of the sender based on voice samples of the sender (e.g., recordings of the sender’s voice). For instance, the assistant may use voicemail messages left by the sender, recordings of voice commands issued by the sender, or the like to generate a model of the sender’s voice.

As another example, the assistant may receive a model of the sender’s voice from a service that maintains a bank of voice models for various speakers. For instance, the sender may register with the service and provide the service with one or more voice samples. The service may model the sender’s voice based on the voice samples. The assistant may request the model of the sender’s voice from the service when performing TTS on text sent by the sender. As such, in some examples, the assistant may avoid having to develop and maintain voice models.

As another example, the assistant may obtain the model of the sender’s voice from the sender’s assistant. For instance, the sender’s assistant may append, attach, provide a resource locator, or otherwise include a model of the sender’s voice when sending text.

In any case, the assistant may perform TTS on the received text to sound like the sender. For instance, the assistant may use the model of the sender’s voice to synthesize audio data of received text.

By performing TTS to sound like a voice of the sender, the assistant may allow for text to be read aloud in a more lifelike manner than always using a single voice or selecting from a limited list of voices. The above examples are just some use cases for the assistant architecture shown in FIG. 1, the assistant architecture has many other applications and use cases.