Delivery of Content Across Multiple Interfaces in a Voice Activated Computing System

John D. Lanza
Foley & Lardner LLP
Voice activated computing systems provide a user with content or services in response to voice commands spoken by the user. Such systems can capture voice commands from a user, process the voice commands to determine requests and keywords in the voice commands, and provide the user with content or services related to the determined requests and keywords.

As discussed herein, a voice activated computing system processes the voice commands to identify non-sponsored content as well as sponsored content related to the voice commands. The system sends the non-sponsored content to a user device or interface over which the voice command was received. But the system sends the sponsored content to a different interface on the same user device, or a different user device. The system can poll the capabilities and availability of various interfaces or user devices, and select an interface or a device to which the sponsored content is sent. For example, a user may speak into a mic of a voice assistant device and state: “Is today a beach day?” or “I would like to go to the beach today.” The system can process the voice command received from the voice assistant device and generate a non-sponsored response that includes a weather forecast for an area beach, such as “it will be sunny and 80 degrees at the beach today, with a high tide at 3 pm.” The system can also generate sponsored content such as an advertisement for a sunscreen. But the system sends the advertisement for the sunscreen not to the speaker of the voice assistant, but to a different interface, such as a display of the voice assistant device, or to a different device, such as the user computer or cell-phone. The system can determine the capabilities and availabilities of the different interfaces and different user devices, and select one or more of these interfaces or user devices to which the sponsored content is sent.
Figure 1 shows an example voice activated computing system. The system includes a voice assistant device, a service provider, a data processing system and a content provider communicating over a network. The voice assistant device can be a device accepts voice commands, and provides audio or visual output. The voice assistant can include one or more mics and cameras, such that voice commands received by the user are converted into corresponding audio signals. The voice assistant can send the audio signals to the data processing system and the service provider. The voice assistant device also can receive data such as audio signals or video signals from the data processing system or the service provider. The voice assistant device also can include audio speakers that can convert the audio signals received from the data processing system or the service provider into sound.
The data processing system can process voice commands received from the voice assistant device. The data processing system includes a natural language processor, an audio signal generator, a task predictor, and a content selector. The natural language processor is capable of processing voice commands included in the audio signals received from the voice assistant device. The natural language processor can convert the audio signals into recognized text by comparing the audio signals against a stored, representative set of audio waveforms, and choosing the closest matches. The representative waveforms are generated across a large set of users, and can be augmented with speech samples. After the audio signals are converted into recognized text, the natural language processor can match the text to words that are associated, for example via training across users or through manual specification, with actions that the data processing system can serve. Basically, the natural language processor identifies requests and trigger words in the converted text, based on which the natural language processor can determine the content and actions to be carried out. The task predictor can predict tasks or actions based on the converted text, and in particular by identifying requests and trigger keywords in the converted text. The task predictor also can predict sponsored content related to the converted text. The content selector can select content, such as services or actions to be offered to the user based on the actions identified by the task predictor. The content selector also can communicate with the content server to obtain sponsored content identified by the task predictor. The audio signal generator can generate audio signals based on the actions selected by the content selector. The audio signals can be representative of voice responses or voice instructions provided to the user in response to the voice commands.

The service provider can provide one or more service to the user. For example, the service provider can provide weather forecast, traffic conditions, and the like. The service
provider can communicate with the data processing system to provide information related to the requested service. Alternatively, the service provider can communicate directly with the voice assistant device independently of the data processing system. The service provider can also include a natural language processor, similar to the one discussed above in relation to the data processing system, to convert user voice commands into text, and identify requests and keywords to determine the services requested by the user.

Referring again to the voice command example mentioned above, the user can speak the voice command “I would like to go to the beach,” to the voice assistant device. The mics at the voice assistant device can convert the voice commands into corresponding audio signals, which are be transmitted by the voice assistant device to the data processing system over the network.

At the data processing system, the natural language processor processes the audio signal received form the voice assistant device and identify requests for “beach” and “today.” The natural language processor also can identify a trigger keyword “go” or “to go to,” which can indicate travel to the beach, such as a need for items one may want to bring to the beach, or a need for transportation to the beach. The task predictor can generate or identify subject matter to presented to the user, such as ‘today’s weather at the beach’ and identify sponsored content related to the beach, such as sunscreen, swim-suit, shades, water, and the like.

The content selector can communicate with the service provider and request for information on the ‘today’s weather at the beach’ as well as communicate with the content provider to request for sponsored content related to items such as sunscreen, swim-suit, shades, water, and the like, identified by the task predictor.

The data processing system also includes an interface manager for determining an interface of rendering the non-sponsored content and an interface for rendering the sponsored
content. The interface manager can identify one or more candidate interfaces of the voice assistant device, which received the voice command from the user. Such interfaces can include speakers, display screen, apps, computer programs, web interfaces, and the like. The interfaces also can include social media accounts, text messaging applications, and email accounts associated with the user that provided the voice command. Interfaces also can include output of a smartphone, or an app based messaging device installed on the smartphone, or on a wearable computing device. To be able to determine whether an interface can render the non-sponsored or the sponsored content, the interface manager can determine the parameters or capabilities of various interfaces. For example, the interface manager can determine whether a user’s smartwatch includes an available visual interface of sufficient size or resolution to render the sponsored content. The interface manager also can determine utilization values for candidate interfaces, which can include processing power, battery status, memory utilization, power requirements, and the like. The interface manager also can determine the proximity of various interfaces to the voice assistant device, or to the user. For example, the interface manager may determine, with the user’s permission, the location of a smartphone or a smartwatch to the voice assistant device. Based on the above discussed factors, the interface manager can rank or prioritize various candidate interfaces.
Figure 2 depicts an example flow diagram for providing multiple content items to multiple user interfaces. The user speaks into the voice assistant device a voice command stating: “I would like to go to the beach today.” The voice assistant device converts the voice command into a corresponding audio file, and sends the audio file to the data processing system. The data processing system, processes the audio file and determines a non-sponsored action and sponsored content in relation to the voice command. For example, the data processing system determines that the non-sponsored action includes sending a voice response to the user stating “the weather today is 80° and sunny.” The data processing system sends an audio file with the voice response to the same voice assistant device from which the voice command was received. The data processing system also determines a sponsored content that can be sent to the user. For example, the sponsored content includes the message “hotel XYZ at the beach is offering a great deal for $ tonight, would you like to book a room?”
The data processing system also determines which of the candidate interfaces are available to send the sponsored content. For example, the candidate interfaces can include a smartphone display of the user, a smartwatch, a social network account of the user, and other interfaces discussed above. The data processing system sends the sponsored content to the selected candidate interface. In some instances, the data processing system may send an additional voice message to the voice assistant device indicating to the user that a sponsored message has been sent to an identified interface. The additional message may not include the sponsored content itself, but just a notification that the sponsored content was sent to the identified interface, if the user is interested. The user may access the sponsored content and may decide to respond. For example, the user may use the voice assistant device to take up on the offer by stating: “Sure, please book a room for tonight.” The user may also use the interface to which the sponsored content was sent to respond with the same message.
Abstract

This document describes a technique for processing voice commands received by a voice activated computing system to identify non-sponsored content as well as sponsored content related to the voice commands. The system sends the non-sponsored content to a user device or interface over which the voice command was received. But the system sends the sponsored content to a different interface on the same user device, or a different user device associated with the user. The system can poll the capabilities and availability of various interfaces or user devices, and select an interface or a device to which the sponsored content can be sent.