February 2021

Automatically Switching Query Response Modality Based On Physical Gestures

Alexander James Faaborg

Shengzhi Wu

Automatically Switching Query Response Modality Based On Physical Gestures

ABSTRACT

When a user is using earbuds to receive device audio, responses to spoken queries are delivered to the earbuds in audio format. Audio delivery can be slower and inefficient compared to the user looking at the same response on a screen. Moreover, long responses delivered as audio can be cognitively difficult to process. Further, it is difficult to skim the content of a long response or skip portions, when the response is delivered as audio. This disclosure describes techniques that enable a user to perform a physical gesture in response to which the query response is provided on the screen of an available device. For example, the user can raise their wrist to view the response on the screen of a smartwatch.

KEYWORDS

- Virtual assistant
- Gesture interaction
- Earbuds
- Audio response
- Spoken query
- Text-to-speech (TTS)
- Ambient computing

BACKGROUND

People often use voice queries to interact with virtual assistants provided via various devices, such as smartphones. Responses to the user queries can be delivered via audio using text-to-speech (TTS) mechanisms. When the user is using an audio playback device, e.g., earbuds, the query response can be provided conveniently as audio played back via the earbuds.

For example, a user who asks for information on “current weather in New York city” is provided a readout of the detailed weather parameters via the earbuds. In another example, if the user asks “who are the top 10 richest people in the world?” a readout of the names may be provided via the earbuds.

Receiving such information via audio is slower and inefficient compared to the user looking at the same information on a screen that allows for richer presentation. Moreover, long responses delivered in the audio format can be cognitively difficult to process. In such cases, users may benefit from viewing the response on the display of an available device instead of receiving it via audio playback. Since earbuds lack a display, the only mechanism for delivering the audio response via the earbuds is to play it from start to finish, with no easy mechanisms to skim the content and limited options to skip to specific portions of interest. For example, the user may be interested in whether rain is likely in New York city, but has to listen to the entire report (that may include other information such as a temperature forecast for different parts of the day, cloudy/sunny conditions, etc.) to obtain that information.

DESCRIPTION

This disclosure describes techniques that enable a user to utilize gestures to automatically switch from an audio response to a visual response displayed on the screen of a user device such as a smartwatch or a smartphone. The user can indicate the desire to switch response delivery from audio mode to visual mode via interaction triggered seamlessly by a suitable physical gesture, such as raising the wrist on which the user wears the smartwatch or picking up and raising the smartphone to a position suitable for reading the screen.

Upon recognition of the triggering gesture, the query response is automatically shown on the screen of the corresponding device, thus permitting the user to derive the benefits of efficient interaction made possible by presentation in the richer, visual format. For example, the user can seamlessly switch from audio to visual format during the playback of a query response, or choose to receive both audio and visual information simultaneously.

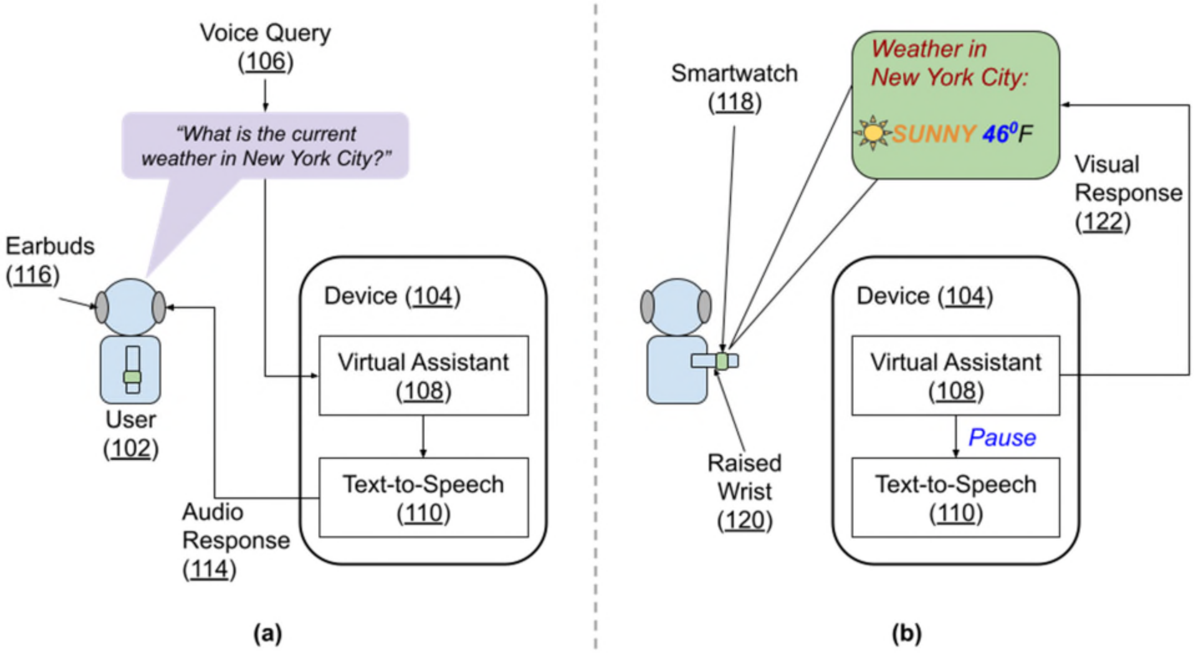

Fig. 1: Raising wrist to pause audio response and view the information on the smartwatch

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A user (102) issues a voice query (106) to a virtual assistant (108) provided via a user device (104), asking for current weather information for New York City. As shown in Fig. 1(a), the response to the user's query is converted from text to speech (110) (or otherwise obtained in audio format) and delivered as audio (114) via the user's wireless earbuds (116).

The user raises the wrist (120) on which a smartwatch (118) is worn. As shown in Fig. 1(b), the user's gesture is detected (with user permission) by use of appropriate sensors, e.g., inertial measurement unit (IMU) on the smartwatch. In response to the detection, the response is delivered in visual format (122) via the smartwatch display.

When switching to a visual mode of receiving query responses from a virtual assistant, the user can choose to pause the ongoing audio response as depicted in Fig 1(b). Alternatively, the user can let the audio response continue playing while looking at the same information on the device screen in the visual format. Delivery of the response via the display allows the user to grasp the information quickly.

With user permission, gesture detection can be performed using a suitably trained machine learning model using appropriate threshold values for detecting the triggering gesture with sufficient confidence. The threshold values can be set by the developers and/or specified by the user and/or determined dynamically at runtime.

While the foregoing example illustrates switching from audio to visual mode of delivery, the reverse can also be achieved via suitable gestures. For example, the user can switch from reading information on the screen to having it delivered as audio to their earbuds by performing a gesture that indicates that they're no longer looking at the screen, e.g., orienting the wrist away from the eyes.

The techniques described in this disclosure can be implemented within any device, service, or platform that provides audio-based virtual assistant capabilities. The operation described above can work with any suitable mobile and wearable devices, such as smartwatches, smartphones, tablets, fitness trackers, as well as any wired or wireless earbuds or earphones. Implementation of the techniques with user permission facilitates a seamless user experience (UX) that permits users to use seamless physical gestures that switch query response delivery between audio delivered to the earbuds and visual presentation on a screen.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if 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 information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

When a user is using earbuds to receive device audio, responses to spoken queries are delivered to the earbuds in audio format. Audio delivery can be slower and inefficient compared to the user looking at the same response on a screen. Moreover, long responses delivered as audio can be cognitively difficult to process. Further, it is difficult to skim the content of a long response or skip portions, when the response is delivered as audio. This disclosure describes techniques that enable a user to perform a physical gesture in response to which the query response is provided on the screen of an available device. For example, the user can raise their wrist to view the response on the screen of a smartwatch.

REFERENCES

1. "Announce Messages with Siri on AirPods" available online at https://support.apple.com/en-in/HT210406, accessed January 28, 2021