February 2023

Suggested Edits and Next Words for Voice Input via Virtual Keyboard

Cristen Anderson Ho

Yohan Launay

Muhammed Rajeef M K

Melissa McCoy

Garen Checkley

Suggested Edits and Next Words for Voice Input via Virtual Keyboard

ABSTRACT

Voice input is a convenient mode that enables users to overcome literacy and/or accessibility barriers because it is typically easier than typing and enables hands-free use of the device. On devices such as smartphones or tablets, a virtual keyboard with voice input capability is convenient and enables voice input to be provided to any application. However, virtual keyboard voice interfaces lack editing capabilities, often requiring users to start from scratch multiple times because of input mistakes, transcription errors, or automatic turning off of the microphone during pauses. This disclosure describes techniques that automatically provide users with suggestions for edits, deletions, and next words during voice interaction with a virtual keyboard. With user permission, the suggestions can be generated using suitable speech and language models to analyze the user's speech, keeping the microphone on until the user ends. The proposed suggestions can be presented in a visual interface with any suitable interactive mechanisms that enable the user to incorporate or dismiss the suggestions.

KEYWORDS

- Voice input
- Voice typing
- Speech transcription
- Virtual keyboard
- Typing correction
- Word suggestions
- Input prediction

BACKGROUND

Many users use voice input to interact with and control their devices such as smartphones, tablets, etc. Some applications include in-app voice interfaces that users can utilize to issue voice commands to control the applications and access various features. For example, a user can issue a voice command to a video playback application to pause the video content being played. Apart from convenience, voice input can also enable users to overcome literacy and/or accessibility barriers since voice input is easier than typing and enables hands-free use of the device.

However, not all applications provide custom voice interfaces. In such cases, users who desire voice-based interaction can use the voice input option included with on-screen keyboards. However, the user experience (UX) of using voice to provide input via the on-screen keyboard is cumbersome and limited. For instance, users cannot make edits to voice input that is misinterpreted by the device. Moreover, the voice input mode lacks dynamic editing functionality that is common for typed input, e.g., automatic correction of typing errors, receiving suggestions for next words or phrases based on past input, etc. As a result, if the voice input is not correctly recognized, the user needs to clear the entire input and start from scratch. Since voice recognition errors are not uncommon, the user may sometimes need to try multiple times before their voice input is accurately transcribed, which can lead to frustration and can discourage the use of voice as an input mode.

Device microphones are usually set to turn off automatically if no voice input is captured within a specified time interval. If a user pauses for too long to compose their thoughts when providing voice input, the microphone may turn off before the user is finished providing voice input. This results in a loss in the voice capture. Such loss creates input gaps that require the user

to start the input process all over again because of the inability to edit the voice input stream or append to previously provided voice input. As a result, the automatic turning off of the microphone during voice interaction can lead to inefficiencies and make users anxious about the microphone turning off when they pause.

DESCRIPTION

This disclosure describes techniques that enable users to edit their previous voice input as detected by the voice input mode of an on-screen keyboard (software keyboard) that includes voice-based input capabilities. Users can additionally be provided with suggestions for upcoming words or phrases in the input that are generated on-the-fly as users speak the desired input as well as after users finish speaking.

With user permission, a suitably trained voice interaction model (running on the user device) is employed to analyze the user's speech while the user is speaking. The analysis can generate suggestions of applicable corrections to the detected voice input and propose alternative words or phrases as likely replacements for those spoken by the user. The model can additionally be utilized to suggest the next word or phrase to be added before the user has spoken it. Further, the user can be provided the ability to delete specific words from the voice transcription.

The proposed corrections, alternate words, and suggested words can be presented visually in a user interface in elements that are displayed in any suitable location, e.g., just above or below the keyboard on the screen. As appropriate, the suggestions can include individual words or phrases that can be corrected or appended, and/or to present the transcription for the user's entire input. The user can tap on the displayed suggestions to edit the transcribed input, append suggested words, or replace the entire text.

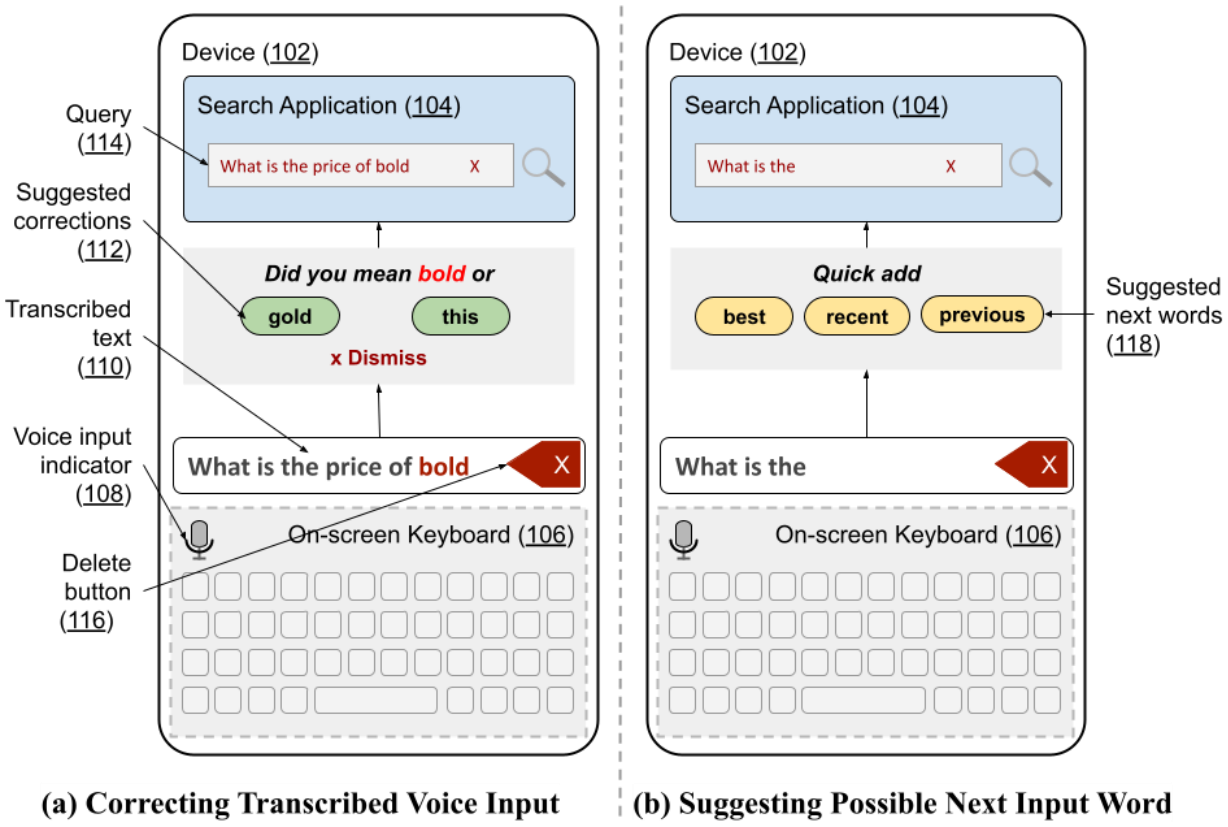

Fig. 1: Suggesting edits and predicting likely next words based on user's voice input

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A user provides a text query (114) to a search application (104) using the voice input feature (108) of the on-screen keyboard (106) of a device (102). As seen in Fig. 1(a), the text is transcribed from the speech (110) using a text-to-speech (TTS) model and is displayed to the user. Suggested corrections (112) - "gold" and "this" - are provided to correct the likely erroneous transcription "bold." The user can choose to accept either of the suggested corrections by tapping it or dismiss the suggestions and continue with the transcribed voice input. The user can also use the delete button (116) to delete the erroneous word in the input and begin the voice input from that word instead of starting the entire input from scratch. When the user explicitly

indicates being done with providing voice input, the microphone is turned off and the captured edited input is relayed as the query to be executed via the search application.

Fig 1(b) illustrates an alternate situation within the same scenario depicting suggestions for possible next words (118) based on the voice input received until that moment. The user can append any of the suggested words to the existing transcript of the voice input by tapping the corresponding button or can ignore the suggestions and continue providing voice input. The suggestions can contain multiple words, phrases, sentences, or entire input text. The user can choose to append the suggested text to the existing speech transcript or replace the existing transcript entirely with the suggested input.

The device microphone can stay activated without timing out after a specified period of silence until the user explicitly indicates that they have finished providing voice input. Not turning the microphone off automatically when interacting with the keyboard can help users compose their voice input similar to input composition when typing. Moreover, the user can take their time to provide the voice input without having to worry about the microphone turning off during pauses.

The techniques described in this disclosure can leverage any suitable existing underlying models for speech and language processing. Therefore, improvements in the voice input experience as described above can therefore be achieved with little to no change in the underlying models. Moreover, the techniques can be applied for voice input in any language by simply employing suitably trained models for that language.

With user permission, the techniques can be implemented on any device that includes a virtual keyboard that provides voice interaction capabilities. Since the virtual keyboard is shared across applications on such devices, implementation of the techniques can provide a voice input

experience that is consistent across applications. Implementation of the techniques can make the process of providing text input via voice more efficient and less frustrating. The enhanced user experience can in turn make applications accessible and facilitate better user engagement. The techniques can be implemented in a device operating system (e.g., as part of the virtual keyboard provided by the operating system), in a virtual keyboard application, a virtual assistant, or any other application that includes voice input features.

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 spoken input, language and keyboard settings, 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

This disclosure describes techniques that automatically provide users with suggestions for edits, deletions, and next words during voice interaction with a virtual keyboard. With user permission, the suggestions can be generated using suitable speech and language models to analyze the user's speech, keeping the microphone on until the user ends. The proposed

suggestions can be presented in a visual interface with any suitable interactive mechanisms that enable the user to incorporate or dismiss the suggestions.

REFERENCES

1. Cherepanov, Evgeny A., Jakob Nicolaus Foerster, Vikram Sridar, Ishai Rabinovitz, and Omer Tabach. “Biasing voice correction suggestions.” U.S. Patent 10,049,655, issued August 14, 2018.
2. Portugal, Beatriz. “EDITE: Voice-based mobile text editing” available online at https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997260163/79175-beatriz-portugal-resumo.pdf accessed 22 February 2023.
3. Weinberg, David. “Voice Typing on an Android Tablet” available online at https://smallbusiness.chron.com/voice-typing-android-tablet-39030.html accessed 22 February 2023.