November 2023

Personalized Phrase Dictionary for Voice Dictation

Pu-sen Chao

Personalized Phrase Dictionary for Voice Dictation

ABSTRACT

Users can use voice dictation capabilities of their devices to provide text input (or commands) and/or to edit text by using their voice. Speech recognition mechanisms are imperfect, leading to personalized names and phrases not being correctly transcribed, and requiring cumbersome manual corrections. This disclosure describes personalized automatic corrections to the transcribed text generated from voice dictation using a user-editable personal dictionary of correction pairs. With user permission, the personal dictionary can be generated automatically based on manual corrections to the transcribed text and shared across devices and applications associated with the user. Entries are added to the dictionary only after the user performs the corresponding corrections at least a threshold number of times. With user permission, user-specific and/or contextually relevant terms can be inferred from data sources such as the user's contacts, calendar, interaction data, locations, etc. The techniques reduce the effort to provide manual corrections and enhance the user experience of voice dictation

KEYWORDS

- Voice dictation
- Speech input
- Hands-free text input
- Speech-to-Text (STT)
- Automated speech recognition (ASR)
- Personal dictionary
- Speech biasing
- Automatic correction
- Personalized correction
- Context-specific correction
- Virtual assistant

BACKGROUND

Users can use voice dictation capabilities of their devices to provide text input (or commands) and/or to edit text by using their voice. When converting the user's speech to text, speech biasing mechanisms are employed to recognize context-specific and (if the user permits), user-specific words or phrases such as names, locations, slang, jargon, acronyms, etc. However, such mechanisms are imperfect, leading to personalized names and phrases not being correctly transcribed in many instances.

Accurate transcription of user-specific or context-specific words and phrases in spoken input is challenging because logs of manual user corrections of transcribed text do not capture the reasons behind the correction. As a result, analytics related to voice dictation cannot distinguish personalized or context-specific corrections from other types of corrections. Even if such corrections were detected and flagged separately, the proportion of personalized utterances in a user's aggregated voice input may be small. The small proportion would result in the weight assigned to such corrections being insignificant in general metrics, such as word revert rate. The small weight may not accurately reflect the large contribution of such recognition errors on the overall user satisfaction with voice dictation.

Incorrect transcription of such utterances requires users to interrupt the voice dictation and manually correct the transcription by editing it with the device keyboard. Unlike the availability of a personal dictionary for saving common corrections to errors in typed text, corrections to the transcription of voice dictation are typically not remembered for subsequent occurrences of the same term. As a result, users must correct the same transcription error multiple times, once for each utterance of the incorrectly transcribed term. The need to input such manual corrections with the keyboard prevents or deters users from using the hands-free

mode to input the entirety of the desired text with proper personalized and contextual terms. This interrupts the flow of the voice dictation and degrades the user experience (UX).

DESCRIPTION

This disclosure describes techniques, implemented with user permission, that enable the provision of a personal dictionary for correcting transcribed text generated via voice dictation capabilities. The dictionary can include multiple entries, with each entry being a correction pair (X, Y) such that dictation transcription X is automatically replaced with Y for the user.

With user permission, the dictionary can be generated and curated automatically based on manual corrections made by the user to the transcribed text generated from the user's spoken input. Entries can be added to the dictionary only after the user performs the corresponding corrections at least a threshold number of times, thus providing high confidence in the corrections compared to that of existing speech biasing correction logic. Users can choose to share the dictionary across multiple devices and services with any appropriate mechanism, such as a common user account across the devices/services. When sharing the dictionary across multiple devices and services, users can additionally permit logging and analyzing the correction data in aggregate across the devices/services. Such a configuration can enable users to avail of the personalized and contextually appropriate dictation corrections on all their devices and services.

With user permission, the dictionary can additionally be employed for making personalized and context-specific corrections to voice dictation, such as replacing "my wife" with the name of the user's wife, replacing "the restaurant" with the name of the restaurant where the user is currently dining, etc. The appropriate text replacements for such corrections can be obtained from relevant contextual information, such as the user's contacts, location,

conversation history, etc. For instance, names appropriate for the current context can be obtained from the user's contact list, the names of the individuals the user is currently interacting with, the names of people mentioned in the current conversation, etc.

For example, the transcribed spelling of the name "Catherine" can be replaced with "Katherine" if the user is currently messaging a person named Katherine. Similarly, other user-specific and/or uncommon terms can be inferred with permission from one or more relevant pieces of information, such as locations visited by the user, events in the user's calendar, names of known entities such as brands, organizations, works of art, consumer goods, etc. In addition, users can employ the dictionary for various other kinds of corrections, such as acronyms (e.g., “looks good to me” replaced with “LGTM”), emojis (e.g., “basketball” replaced with the basketball emoji 🏀) etc.

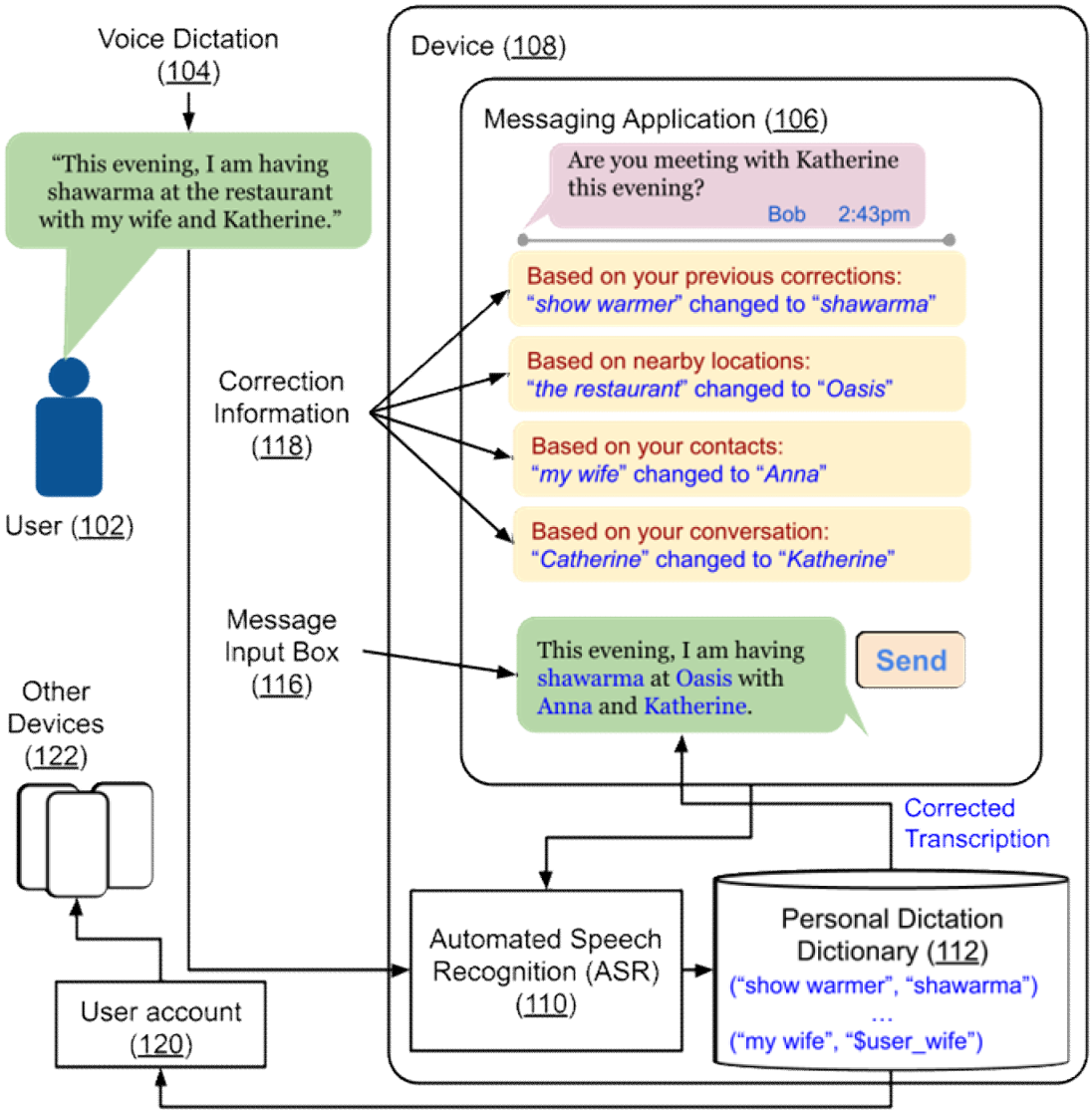

Fig. 1: Applying personalized corrections to transcribed voice dictation

Fig. 1 shows an example operational implementation of the techniques described in this disclosure. A user (102) is responding via voice dictation (104) to a message from Bob received in a messaging application (106) on a device (108). The user's speech is transcribed to text via automated speech recognition (ASR) (110). The transcribed text is then transformed by applying corrections according to the paired entries in a personal dictation dictionary (112), entries in the dictionary being generated based on the user's past corrections as described above.

In the example of Fig. 1, the application of the dictionary results in correcting a transcription error ("show warmer" to "shawarma"), applying the correct spelling of a contact's name ("Katherine" in place of "Catherine"), including the name of the user's wife ("Anna") from the user's contacts, and providing the name of the restaurant ("Oasis") based on the user's context (e.g., location). The corrected transcription is inserted in the message input box (116) of the messaging application. The user receives information about the personalized corrections (118), e.g., as transient pop ups in the user interface (UI) of the application. With user permission, the personal dictation dictionary can be shared across the user's other devices (122), e.g., via the user's online account (120).

The techniques described above can function reliably given that the transcription generated by the ASR applied to voice dictation is consistent. Users can be provided the ability to edit or remove entries in the personal dictionary for voice dictation via any suitable editing user interface (UI). In addition, users can choose to disable or reset the functionality via corresponding settings. The threshold number of user corrections to transcribed dictation that results in a corresponding entry being added to the personal dictionary can be set by the developers and/or specified by the users and/or determined dynamically at runtime. The user can turn off the personalized transcription at any time and can request deletion of the personalized dictionary. Further, the user can limit the dictionary to a particular device or set of devices. The dictionary is stored and updated in accordance with user settings.

As shown in Fig. 1, when an automatic correction from the personal dictionary is applied and stored, users can be informed via any suitable user UI mechanisms, such as a popup, or a UI element such as a chip or bubble, etc. The correction information can be displayed the first few times a specific correction is applied and subsequently, applied automatically.

With user permission, the functionality described in this disclosure can be implemented within any application and device that supports the use of voice dictation for text input. The functionality can be provided via one or more of any suitable mechanisms, such as a voice-based virtual assistant, application programming interface (API), etc.

Implementation of the techniques can enable users to employ voice dictation for text input without needing to interrupt the flow for correcting transcription errors resulting from user- and context-specific text, such as names, locations, jargon, slang, acronyms, etc. The ability to provide error-free and contextually appropriate text input can enable users to use voice dictation to produce high-quality, accurate text input. Moreover, the ability to dictate text input without requiring periodic manual correction can boost convenience when completely hand-free device interactions are necessary, such as when carrying objects, driving, etc. The personalized and context-appropriate transcription of voice dictation enabled by the techniques described in this disclosure can enhance the user experience of applications and devices that include voice dictation and promote greater use of the feature.

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 contacts and 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

This disclosure describes personalized automatic corrections to the transcribed text generated from voice dictation using a user-editable personal dictionary of correction pairs. With user permission, the personal dictionary can be generated automatically based on manual corrections to the transcribed text and shared across devices and applications associated with the user. Entries are added to the dictionary only after the user performs the corresponding corrections at least a threshold number of times. With user permission, user-specific and/or contextually relevant terms can be inferred from data sources such as the user's contacts, calendar, interaction data, locations, etc. The techniques reduce the effort to provide manual corrections and enhance the user experience of voice dictation

REFERENCES

1. "Dragon Anywhere User Guide" available online at https://www.nuance.com/asset/en_us/collateral/dragon/guide/gd-dragon-anywhere-user-manual-en-us.pdf accessed November 6, 2023.
2. Voicebase. "Custom Vocabulary" available online at https://docs.voicebase.com/docs/custom-vocabulary accessed November 6, 2023.
3. Bruguier, Antoine Jean, Fuchun Peng, and Francoise Beaufays. "Learning personalized entity pronunciations." U.S. Patent 10,152,965, issued December 11, 2018. Assigned to Google LLC.

4. “How to use Auto-Correction and predictive text on your iPhone, iPad, or iPod touch” https://support.apple.com/en-us/HT207525 available online at accessed November 6, 2023.