

Technical Disclosure Commons

Defensive Publications Series

December 08, 2017

HYPER SUGGESTION SPANS FOR IMPROVED TEXT CORRECTION

Glen Shires

Jakob Foerster

Diego Melendo Casado

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

Recommended Citation

Shires, Glen; Foerster, Jakob; and Casado, Diego Melendo, "HYPER SUGGESTION SPANS FOR IMPROVED TEXT CORRECTION", Technical Disclosure Commons, (December 08, 2017)
http://www.tdcommons.org/dpubs_series/923



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

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.

Glen Shires, Jakob Foerster, Diego Melendo Casado

HYPER SUGGESTION SPANS FOR IMPROVED TEXT CORRECTION

ABSTRACT

Improved techniques for correcting text transcriptions of voice input is described herein. A computing device may perform speech recognition to transcribe voice input (e.g., speech) into text, enable a user to select any of one or more consecutive words in the text transcription, and, in response, determine and output a set of alternative hypotheses for the selected one or more consecutive words, so that the user may select one alternative hypothesis in the set of alternative hypotheses in order to replace the selected one or more consecutive words with the alternative hypothesis.

INTRODUCTION

A computing device may receive text input in a number of ways. Voice input is one common method for inputting text into a computing device. An audio input device of a computing device, such as the microphone connected to or included in the computing device, may receive audio signals that carry voice, or spoken, input. The computing device may perform techniques such as voice recognition to recognize and transcribe the spoken content of the voice input into text, such as characters, words, phrases, and/or sentences, for text input.

When transcribing voice input received by the audio input device into text, the computing device may determine a set of possible text transcriptions of the voice input, and may output for display the most likely text transcription out of the set of possible text transcriptions. For example, the computing device may assign a set of scores to the set of possible text transcriptions of the voice input, and may output the text transcription with the highest score as the text transcription for the voice input. However, currently, using voice input as a text input method has a word error rate of around 10%. As such, it may be useful for the computing device to enable a user to correct the text transcription of the voice input by editing, replacing, deleting, or otherwise correcting one or more words of the text transcription produced by the computing device.

The computing device may produce alternative hypotheses for one or more of the transcribed words or phrases of the text transcription of the voice input. Alternative hypotheses

for a transcribed word may be a set of alternative words or phrases for the transcribed word. The computing device may enable the user of the computing device to select a word of a text transcription to see a set of alternative hypotheses for the selected word, and may enable the user to select an alternative hypothesis from the set of alternative hypotheses to replace or otherwise substitute the selected word with the selected alternative hypothesis.

For example, the computing device may determine one or more words of the text transcription for which the computing device may determine an alternative hypothesis for each of the one or more words. The computing device may enable a user to select a particular word of a transcribed phrase and sentence. In response to receiving a selection of a word of the transcribed phrase or sentence, the computing device may display or output a set of alternative hypotheses for the selected word, and may enable the user to select one of the set of alternative hypotheses. In response to receiving a selection of one of the set of alternative hypotheses, the computing device may replace the word in the transcribed phrase or sentence with the selected alternative hypothesis.

The computing device may output all or a portion of the alternative hypotheses it has determined as the set of alternative hypotheses for a word. For example, the computing device may determine a score for each alternative hypothesis and may determine the N alternative hypotheses with the highest scores as the set of alternative hypotheses for a word. Alternatively, the computing device may determine that each alternative hypothesis having a score above a specified scoring threshold may be included in the set of alternative hypotheses for a word. The computing device may also use any other suitable technique for determine the set of alternative hypotheses for a word.

As shown below in FIG. 1, computing device 2 may perform voice recognition to transcribe voice input into the phrase “Let’s find a real example of lettuce is until they break,” and may display the text transcription of the voice input on a display. Computing device 2 may also determine one or more words in the text transcription for which computing device 2 may provide a set of alternative hypotheses, and may highlight, underline, change the color of, or otherwise visually emphasize the one or more words for which computing device 2 has determined a set of alternative hypotheses. As shown in FIG. 1, computing device 2 may highlight the word “lettuce” via underlining to indicate that computing device 2 has produced a set of alternative hypotheses 4A-4D (“set of alternative hypotheses 4”) for the word “lettuce,”

and that the user of computing device 2 may provide input, such as by selecting the word “lettuce” to see the set of alternative hypotheses 4.

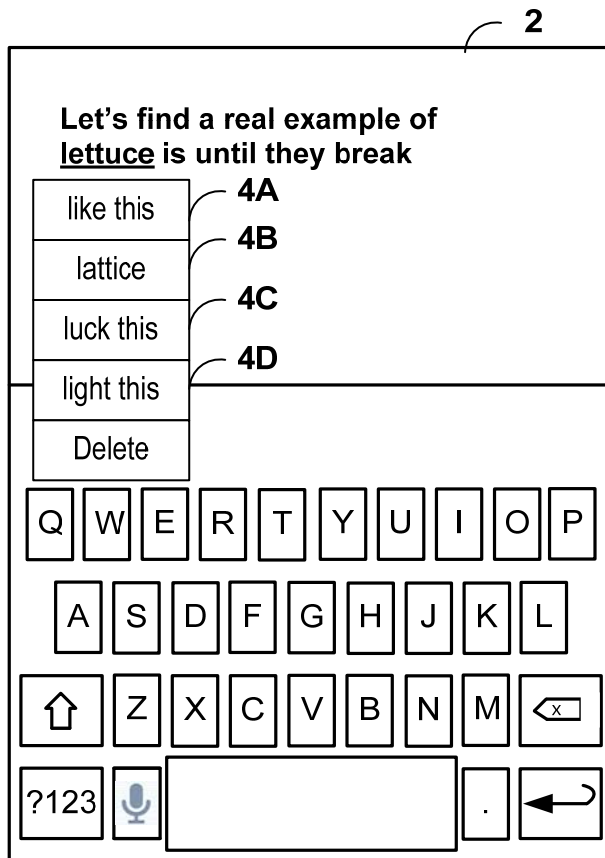


FIG. 1

If the display is a presence-sensitive display (e.g., a touch-sensitive display), the user may tap, touch, or otherwise provide input to select the highlighted word “lettuce” to provide an indication to computing device 2 to display the set of alternative hypotheses 4. Computing device 2 may enable a user to select (e.g., tap) one of the alternative hypotheses 4 and to replace the selected word “lettuce” with one of the alternative hypotheses 4 (e.g., “like this”). In this way, the user may correct the text transcription by selecting one of the visually highlighted words in the text transcription and selecting one of the alternative hypotheses 4 to replace and correct the selected word.

However, by pre-determining the one or more words of the text transcription that may be selected and replaced with an alternative hypothesis, computing device 2 may limit the ability of

the user to correct the text transcription because computing device 2 may only provide a set of alternative hypotheses for fewer than all of the words of the text transcription. Thus, to correct or replace a word in the text transcription that is not visually highlighted, the user may have to delete the word in the text transcription and manually enter a word to replace the deleted word, instead of being able to select the desired word of the text transcription and selecting one of the alternative hypothesis for the word to replace the word with the selected alternative hypothesis. Further, because computing device 2 only provides a set of alternative hypotheses for individual words, it may be cumbersome to correct two or more consecutive words.

DESCRIPTION

Improved techniques for correcting text transcriptions of voice input are described herein, which may enable a user to more efficiently and accurately correct text input, such as voice input that a computing device transcribes into text. Instead of pre-determining the words within a text transcription for which the computing device may provide a set of alternative hypotheses, the computing device may enable the user to select any of the words in the text transcription and may, in response, determine and output a set of alternative hypotheses for the selected word, so that the user may select one of the alternative hypothesis in the set of alternative hypotheses in order to replace the selected word with the alternative hypothesis. Further, instead of allowing the user to only select a single word at a time, and determining a set of alternative hypotheses for the single word, the techniques disclosed herein may enable the user to select any arbitrary combination of consecutive words (i.e., two or more consecutive words) of the text transcription and may enable the computing device to, in response to the selection of the two or more consecutive words, determine and output a set of alternative hypotheses for the selected two or more consecutive words, so that the user may select one of the alternative hypothesis in the set of alternative hypotheses in order to replace the selected two or more words with the alternative hypothesis.

As shown in FIG. 2, computing device 2 may perform voice recognition to transcribe voice input into the phrase “Let’s find a real example of lettuce is until they break”, and may display the text transcription of the voice input on a display. As discussed above, computing device 2 may enable any set of one or more consecutive words of the text transcription to be

selected, and may, in response determine and output a set of alternative hypotheses for the selected one or more consecutive words.

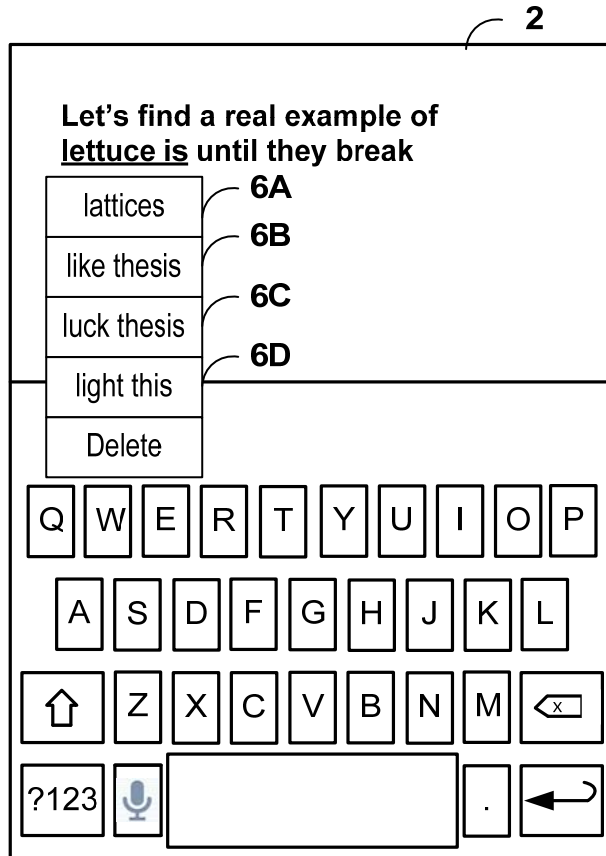


FIG. 2

As shown in FIG. 2, computing device 2 may enable the user to select arbitrary consecutive words “lettuce is” of the text transcription. For example, if the text transcription is displayed by a presence-sensitive display, the user may perform a drag gesture from the start of the word “lettuce” until the end of the letter “is” to select the phrase “lettuce is.” In response to determining that the phrase “lettuce is” has been selected, computing device 2 may determine and output a set of alternative hypotheses 6A-6D (“set of alternative hypotheses 6”) for the phrase “lettuce is”. Computing device 2 may enable a user to select (e.g., tap) one of the alternative hypotheses 6 and to replace the selected phrase “lettuce is” with one of the alternative hypotheses 6 (e.g., “lattices” or “like thesis”). In this way, the user may correct the text transcription by selecting one or more consecutive words in the text transcription and selecting

one of the alternative hypotheses 6 determined by computing device 2 to replace and correct the selected one or more consecutive words.

Enabling two or more consecutive words of the text transcription to be selected may enable computing device 2 to surface relatively more accurate alternate hypotheses than may be possible if only individual words of the text transcription may be selected. As can be seen in FIG. 2, computing device 2 may determine “lattices” to be an alternative hypothesis 6A for the phrase “lettuce is.” In contrast, as shown in FIG. 1, “lattices” is not included in the set of alternative hypotheses 4 when only the word “lettuce” is selected. Thus, the techniques disclosed herein may enable the computing device to determine and surface more accurate alternative hypotheses than is possible with prior techniques for correcting text transcriptions.

FIGS. 3A-3E illustrate the selection of arbitrary one or more consecutive words of a text transcription.

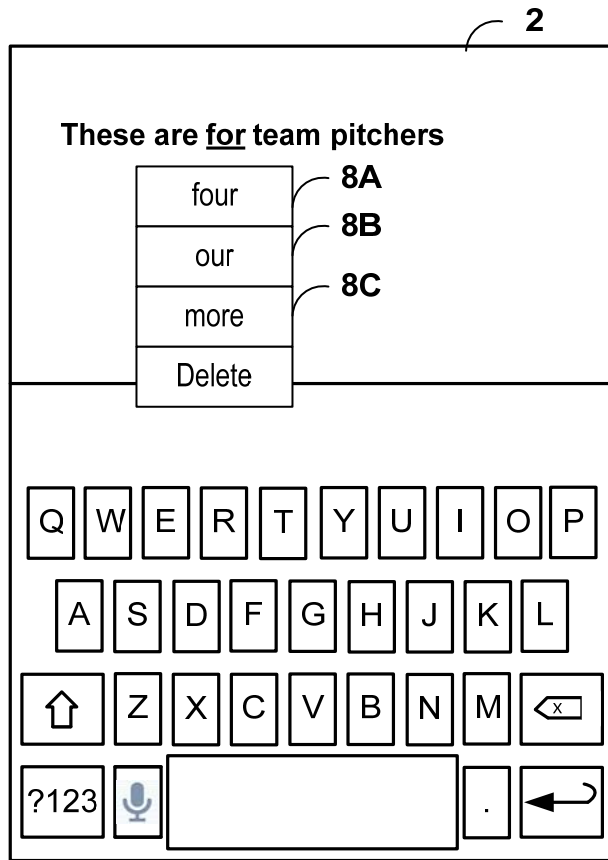


FIG. 3A

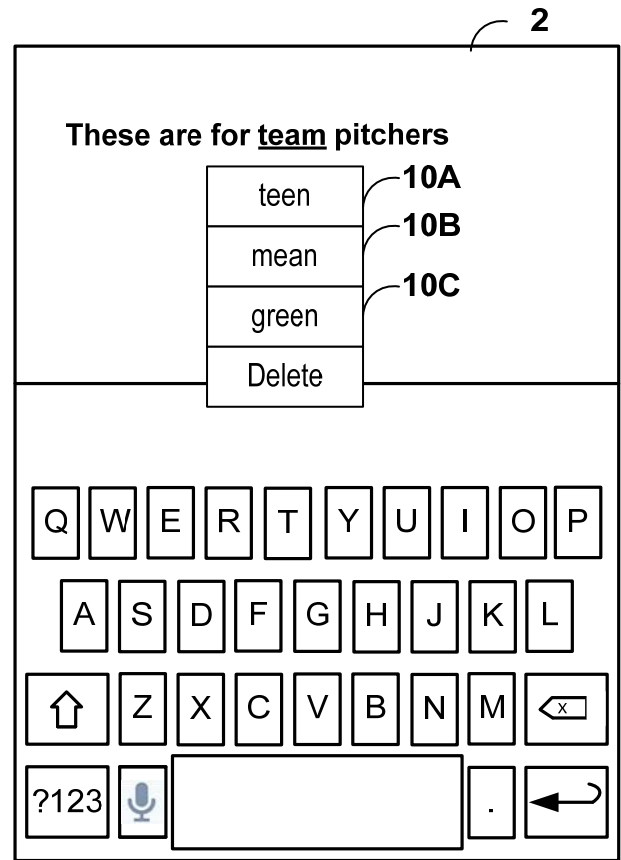


FIG. 3B

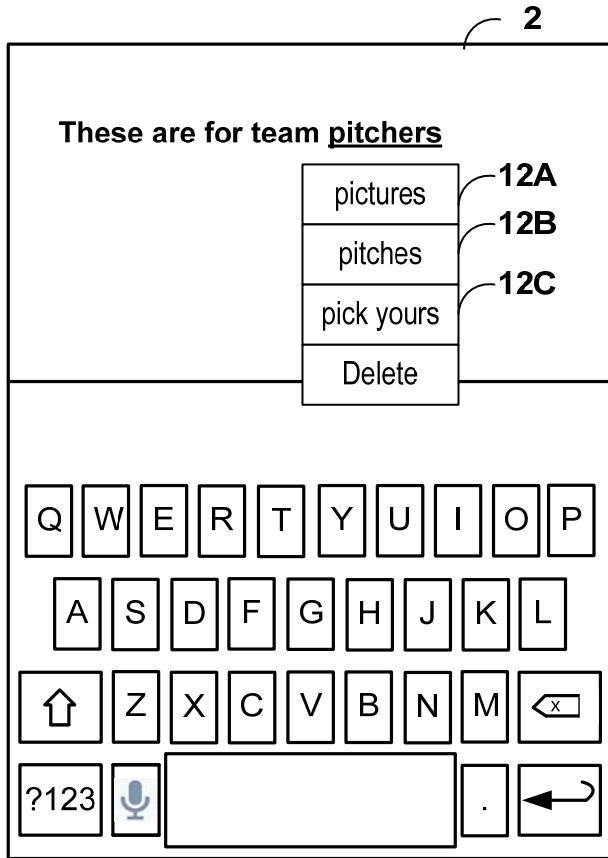


FIG. 3C

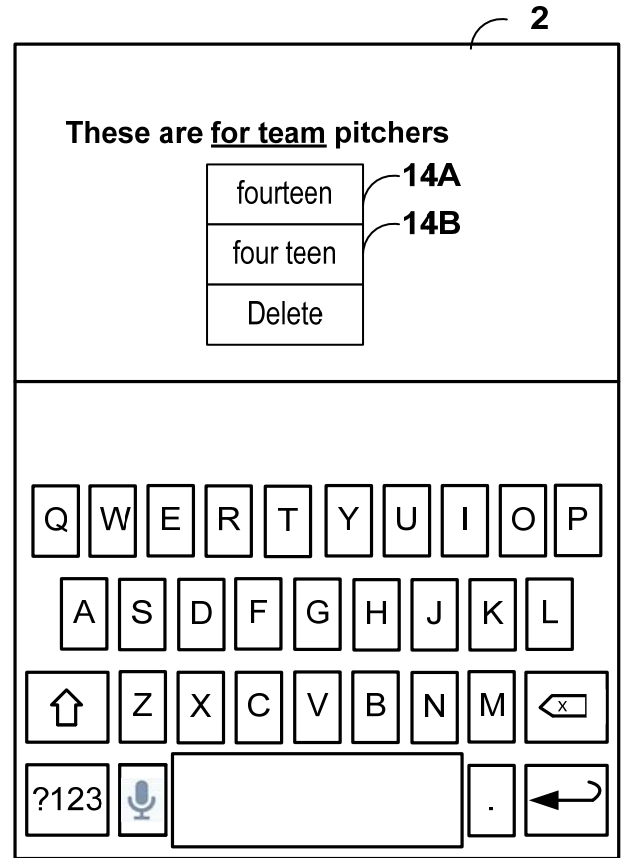


FIG. 3D

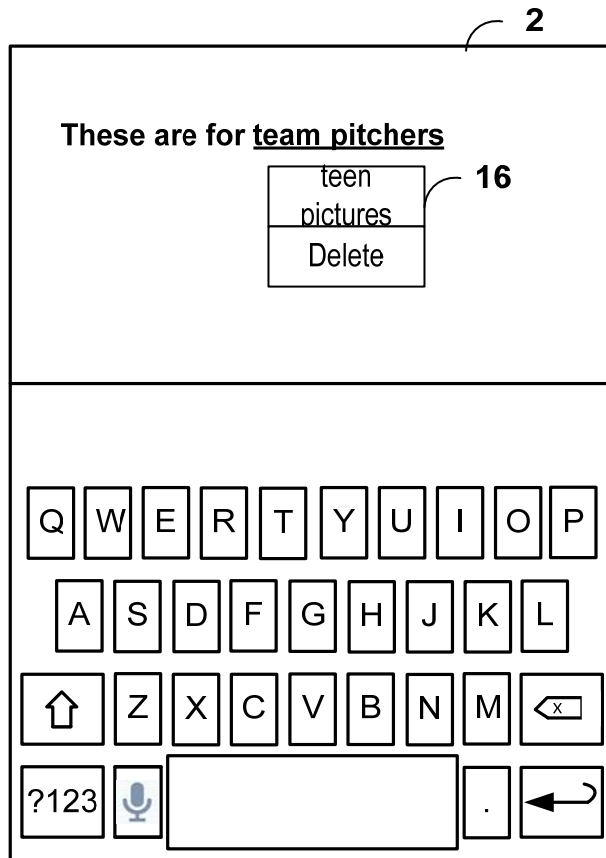


FIG. 3E

As shown in FIG. 3A, in response to the selection of the word “for” in the text transcription “These are for team pitchers,” computing device 2 may determine and output a set of alternative hypotheses 8A-8C (“alternative hypotheses 8”) for the selected word “for”. As shown in FIG. 3B, in response to the selection of the word “team”, computing device 2 may determine and output a set of alternative hypotheses 10A-10C (“alternative hypotheses 10”) for the selected word “team”. As shown in FIG. 3C, in response to the selection of the word “pitchers,” computing device 2 may determine and output a set of alternative hypotheses 12A-12C (“alternative hypotheses 12”) for the selected word “pitchers”.

As shown in FIG. 3D, in response to the selection of the two consecutive words “for team,” computing device 2 may determine and output a set of alternative hypotheses 14A and 14B (“alternative hypotheses 14”) for the selected two consecutive words “for team”. As shown in FIG. 3E, in response to the selection of the two consecutive words “team pitchers,” computing

device 2 may determine and output alternative hypothesis 16 for the selected two consecutive words “team pitchers.”

To determine the set of alternative hypotheses for the selected two or more consecutive words, the selection of the two or more consecutive words within the text transcription may indicate to the computing device that the adjacent unselected portions of the text transcription is correct. The user therefore may indicate, by selecting two or more consecutive words within the text transcription, that the selected two or more consecutive words may need to be corrected, while the adjacent unselected portions of the text transcription may not need to be corrected. Thus, the computing device may determine the set of alternative hypotheses on the basis that the adjacent unselected portions of the text transcription are correct, and may determine alternative hypotheses that fit into the unselected positions of the text transcription.

By enabling arbitrary one or more consecutive words of a text transcription to be selected, and by enabling the user to replace the selected one or more consecutive words with an alternative hypothesis, the techniques disclosed herein enables the user of a computing device to more quickly and easily correct mistakes made by the computing device when transcribing voice input into text. In this way, the techniques disclosed herein significantly reduces the need for the user to have to re-speak words into the computing device to have the user’s speech be transcribed again, and significantly reduces the need for the user to have to manually correct mistakes in the text transcription of the voice input.