March 13, 2015

VOICE AND TOUCH BASED INPUT

Michael Cardosa

VOICE AND TOUCH BASED INPUT

ABSTRACT

A word prediction system predicts a word based on a combination of user touch input and user voice input. The system can receive the touch input via, for example, a keyboard or other input device. In some implementations, the system can receive voice input corresponding to the touch input at substantially the same time as the touch input.

PROBLEM STATEMENT

Swipe-based text input entered using a touchscreen keyboard generally involves sliding a finger or stylus from letter to subsequent letter displayed by the touchscreen keyboard to thereby input a desired word or string. Systems offering swipe-based text input functionality generally employ error-correcting algorithms and one or more language models to infer a word or string based on sliding user inputs. However, this process often produces imprecise results for short words or strings. More specifically, words containing two to five letters are often confused for other words, leading to erroneous text input and a frustrating input experience for the user. When this occurs, a user--upon realizing an incorrect word was interpreted by the keyboard--must delete the word and reenter the input gesture. Accordingly, a method and system that uses a hybrid model of voice-based input and touch input to achieve high accuracy word prediction is disclosed.

WORD PREDICTION SYSTEM

The system and techniques described in this disclosure relate to a word prediction system that analyzes a touch input in combination with a voice input to predict a word desired for input by a user. The word prediction system can be implemented for use by a client device connected to one or more other devices via the Internet, an intranet, or in any client-server environment. The word prediction system can include program instructions implemented locally on a client device or implemented across a client device and server environment. The client device can be any electronic device such as a mobile device, a smartphone, a tablet, a handheld electronic device, a wearable device, etc.

Fig. 1 illustrates an example method 100 for combining a touch input with a voice input to predict a word. Method 100 can be performed by a system that analyzes a touch input and a corresponding voice input to predict a word.

When a user is on an electronic device, the user inputs text using a keyboard in order to write messages, e-mails, documents, to-do lists, social posts, etc. The system can receive a touch input via the keyboard (110). The electronic device includes a touch interface for the user to provide the touch input on the keyboard. The electronic device’s keyboard may be one of various keyboard layouts, e.g., Qwerty, Dvorak, or Colemak. The user can simply tap on individual letters of the keyboard in order to enter a word or the user may use a swipe based input mechanism, i.e., drawing a path on the keyboard that covers all the letters of the word in sequence.

As the user types the word on the keyboard, the user may simultaneously speak the same word into a microphone of the electronic device. The system can then receive a voice input corresponding to the touch input (120). The user may speak only those words that he thinks can be easily confused by the system, e.g., words that have a high error rate with swiping. For example, the words “in” and “on” are easily confused with each other since the letters “i” and “o” on a Qwerty keyboard are immediately next to each other. The user can speak in any manner detectable by the microphone of the electronic device and by the system. The user might not want to speak words aloud in aloud fashion as he or she types on the keyboard. For example, the user may choose to softly speak, e.g., whisper, the words.

The system can next analyze the received voice input to identify the presence of consonants and vowels, and such hints can then be streamed in real-time to detect various letters of the word. Unlike speech recognition, this approach does not predict words based on the user’s voice input, but instead analyzes the user’s voice input to determine the probability that one or more consonants or vowels are present in the word. For example, when the user whispers the word “so”, the microphone detects the audio, which is then analyzed in real-time to determine a probability of the consonant “s” being present in the word. This approach can also determine the sequence of the consonants and vowels present in the word based on the voice input.

The system can further analyze the touch input and the voice input to predict the word (130). The system predicts the word by using the touch input, the voice input, or a combination of both the inputs. The system may analyze the touch input to determine the possible string of characters input by the user to form the word and the voice input to detect various consonants and vowels in the word. The system can rely on the detected consonants and vowels from the voice input to increase precision in predicting the swiped words.

For example, when the user attempts to enter the word “so” by drawing a straight line between the letters “s” and “o” of a Qwerty keyboard, the user may not perfectly place his or her finger precisely at a location of the touch-sensitive device associated with the letter “s.” If considering just the swipe input, the system may not be confident whether the user intended to swipe “do” or “so” because the letters “s” and “d” are immediately next to each other on the keyboard. While typing “so,” the user can whisper the word “so.” The system can then analyze the voice input to determine that the consonant “s” is present in the word. This analysis may indicate that swipe input with a “s” is of higher probability, and thus the system can then determine that the user intended to write “so” and consequently chooses the word “so” as the final input.

An alternative method to increase text input accuracy is to use traditional speech recognition algorithms to detect voice input in combination with received touch inputs. For example, the user can speak a word and then tap the first letter of the word on the keyboard. As a further example, the user may speak a word while making a low-precision swipe motion on the keyboard between the first and last letters of the word without attempting to connect the intermediate letters of the word as part of the motion. The system can use both speech recognition algorithms and the limited text input to predict the word.

The method 100 helps in reducing user frustration because the text input prediction will be more accurate, leading to a better user experience. The system additionally helps in energy and power savings as the method 100 uses simple sound processing algorithms to detect the presence of letters over using full-fledged speech recognition algorithms to detect exact entire words.

Fig. 2 illustrates an operation of the word prediction system. As shown in Fig. 2, a user 220 can swipe on a Qwerty keyboard of his electronic device using a swipe based input mechanism (210). The user 220 intends to input the word “quick” by starting the swipe from letter “q” and drawing a curvy line through the word’s letters ending at letter “k” (210). The user may not perfectly start swiping from the letter “q.” As the user’s swipe starts between “q” and “w,” the system may get confused whether the user intends to type “quick” or “wick.” Hence, the system detects an ambiguity in predicting the word based only on the swipe input.

While swiping on the touch screen, the user 220 may speak at least a portion of the word “quick” (230). A microphone of user’s electronic device can then pick up the voice input and transmits it to the system. The system may analyze the voice input in order to guess the presence of at least some of the consonants and vowels, e.g., “q,” “u,” “i,” “c,” “k” in the voice input. The hints received from the voice analysis of the system indicate that it is more likely that words with “q” are of higher probability. Hence, based on the hints and the swipe input, the system can determine that the user intends to type “quick.” The system can then display “quick” on the touch screen as a result of the user’s input (240).

Fig. 3 is a block diagram of an exemplary environment that shows components of a system for implementing the techniques described in this disclosure. The environment includes client devices 310, servers 330, and network 340. Network 340 connects client devices 310 to servers 330. Client device 310 is an electronic device. Client device 310 may be capable of requesting and receiving data/communications over network 340. Example client devices 310 are personal computers (e.g., laptops), mobile communication devices, (e.g. smartphones, tablet computing devices), set-top boxes, game-consoles, embedded systems, and other devices 310’ that can send and receive data/communications over network 340. Client device 310 may execute an application, such as a web browser 312 or 314 or a native application 316. Web applications 313 and 315 may be displayed via a web browser 312 or 314. Server 330 may be a web server capable of sending, receiving and storing web pages 332. Web page(s) 332 may be stored on or accessible via server 330. Web page(s) 332 may be associated with web application 313 or 315 and accessed using a web browser, e.g., 312. When accessed, webpage(s) 332 may be transmitted and displayed on a client device, e.g., 310 or 310'. Resources 318 and 318' are resources available to the client device 310 and/or applications thereon, or server(s) 330 and/or web pages(s) accessible therefrom, respectively. Resources 318' may be, for example, memory or storage resources; a text, image, video, audio, JavaScript, CSS, or other file or object; or other relevant resources. Network 340 may be any network or combination of networks that can carry data communication.

The subject matter described in this disclosure can be implemented in software and/or hardware (for example, computers, circuits, or processors). The subject matter can be implemented on a single device or across multiple devices (for example, a client device and a server device). Devices implementing the subject matter can be connected through a wired and/or wireless network. Such devices can receive inputs from a user (for example, from a mouse, keyboard, or touchscreen) and produce an output to a user (for example, through a display). Specific examples disclosed are provided for illustrative purposes and do not limit the scope of the disclosure.

DRAWINGS

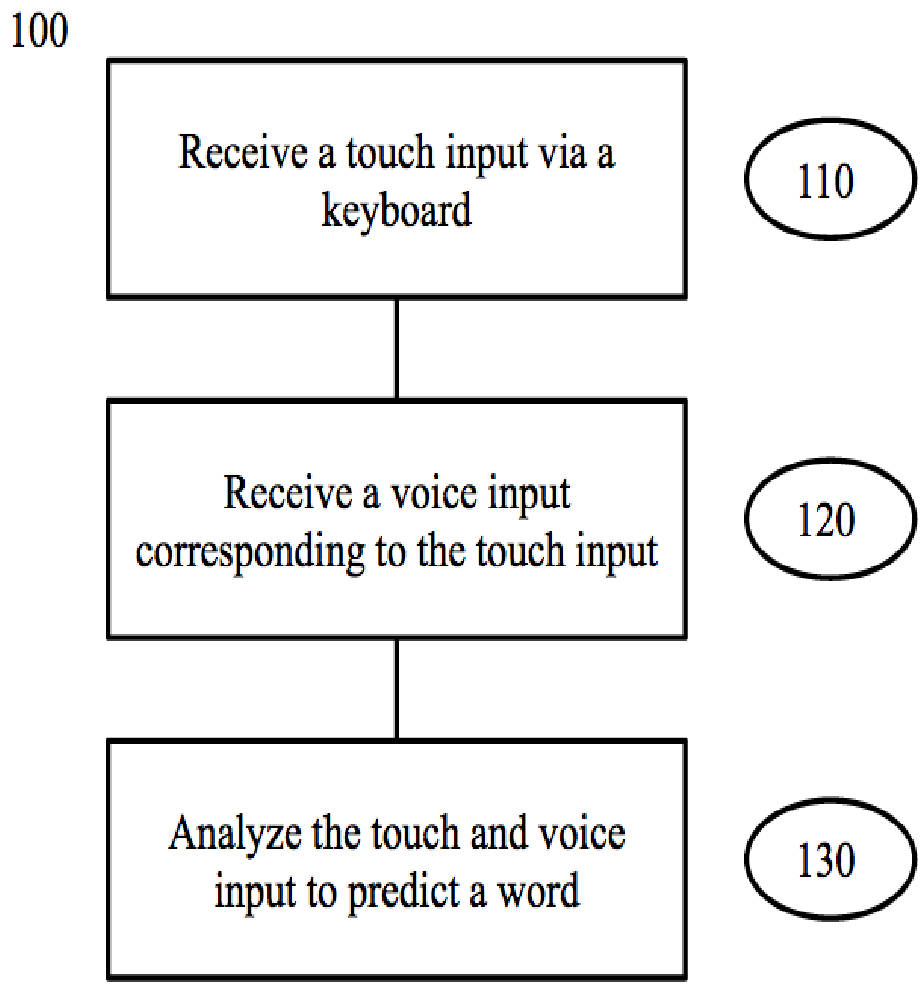

Fig. 1

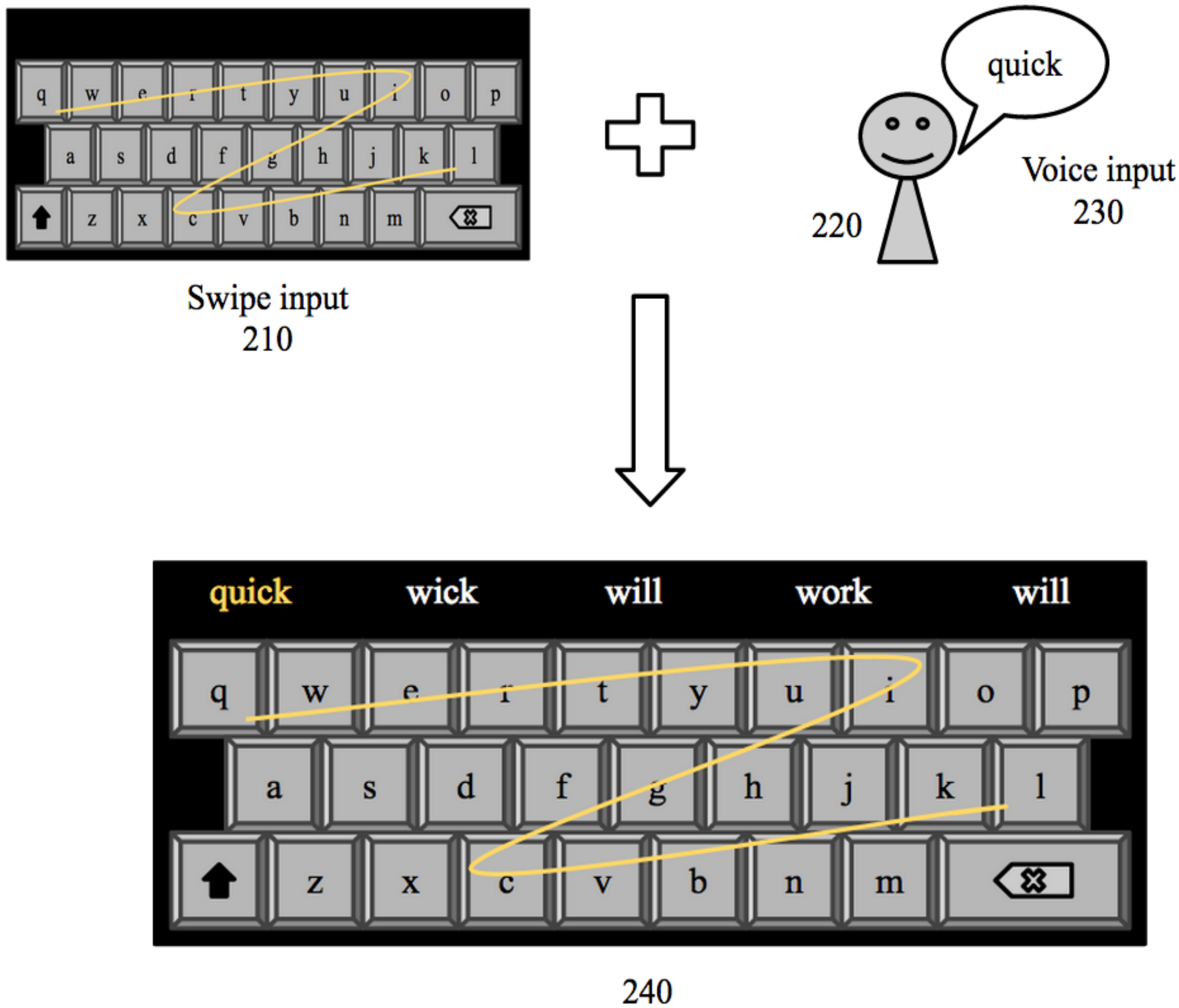
quick
Voice input
230
220
Swipe input
210
quick
wick
will
work
will
240

Fig. 2

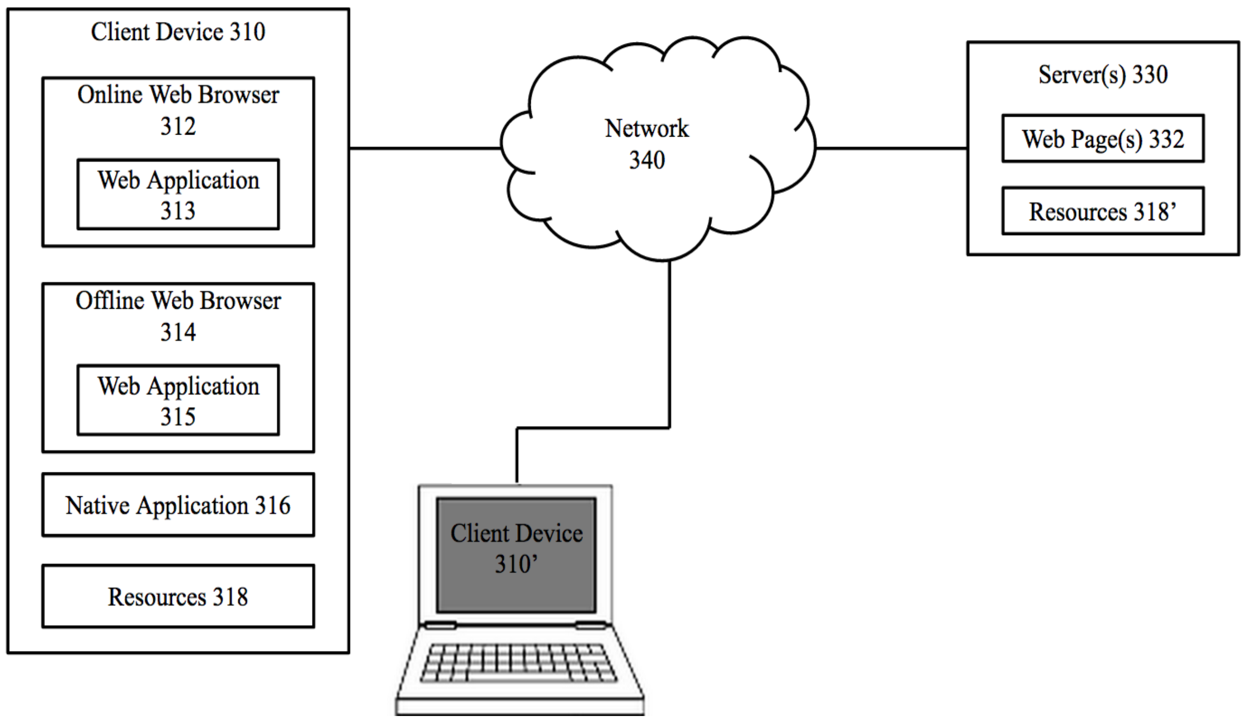
Client Device 310
Online Web Browser 312
Web Application 313
Offline Web Browser 314
Web Application 315
Native Application 316
Resources 318
Network 340
Server(s) 330
Web Page(s) 332
Resources 318'
Client Device 310'

Fig. 3