A method, apparatus, and electronic device for voice navigation are disclosed. A voice input mechanism 310 may receive a verbal input from a user to a voice user interface program invisible to the user. A processor 104 may identify in a graphical user interface (GUI) a set of GUI items. The processor 104 may convert the set of GUI items to a set of voice searchable indices 400. The processor 104 may correlate a matching GUI item of the set of GUI items to a phonemic representation of the verbal input.
Figure 1

Network Interface 102

Volatile Memory 106

User Interface 110

Component Interface 114

Processor 104

Non-Volatile Memory 108

User Output 112

Power Supply 116
Start

502 Display GUI

504 Identify GUI

506 Non-Textual? Yes

508 Develop Alternate Labels

506 Non-Textual? No

510 Convert GUI to LD

512 Organize into Index

End

Figure 5
Receive Verbal Input
Convert to LST
Identify Set of Possible Matching GUIs
Designate Primary Match and Alternates
Present Matching GUI CO
Present Alternates
Alternate?
Present Selected Alternate
End

Figure 6
SEARCH-BASED DYNAMIC VOICE ACTIVATION

FIELD OF THE INVENTION

[0001] The present invention relates to a method and system for voice navigation. The present invention further relates to voice navigation as relating to graphical user interface items.

INTRODUCTION

[0002] Voice recognition software has historically performed less than ideally. Most software programs that perform voice recognition based navigation have previously done so by constructing a voice dialogue application statically for each view of a graphical user interface (GUI). To do this, for each view of a GUI, a dialogue application has to anticipate every grammar and vocabulary choice of the user. This process may significantly impede browsing and navigation.

[0003] Web content providers may currently use VoiceXML® for voice navigation or browsing by voice enabling web pages. VoiceXML® is a static voice navigation system, which does not allow for much flexibility. VoiceXML® coverage may not extend to the entire webpage.

SUMMARY OF THE INVENTION

[0004] A method, apparatus, and electronic device for voice navigation are disclosed. A voice input mechanism may receive a verbal input from a user to a voice user interface program invisible to the user. A processor may identify in a graphical user interface (GUI) a set of GUI items. The processor may convert the set of GUI items to a set of voice searchable indices. The processor may correlate a matching GUI item of the set of GUI items to a phonemic representation of the verbal input.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered to be limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0006] FIG. 1 illustrates a block diagram one embodiment of a computing device that may be used to implement the communication protocol management method.

[0007] FIG. 2 illustrates in a block diagram one embodiment of a graphical user interface.

[0008] FIG. 3 illustrates in a block diagram one embodiment of a voicemail user interface software application.

[0009] FIG. 4 illustrates in a block diagram one embodiment of voice searchable indices.

[0010] FIG. 5 illustrates in a flowchart one embodiment of a method for developing voice searchable indices.

[0011] FIG. 6 illustrates in a flowchart one embodiment of a method for invisible voice navigation.

[0012] FIG. 7 may illustrate in a block diagram one embodiment of a graphical voice navigation response.

DETAILED DESCRIPTION OF THE INVENTION

[0013] Additional features and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The features and advantages of the invention may be realized and obtained by means of the instruments and combinations particularly pointed out in the appended claims. These and other features of the present invention will become more fully apparent from the following description and appended claims, or may be learned by the practice of the invention as set forth herein.

[0014] Various embodiments of the invention are discussed in detail below. While specific implementations are discussed, it should be understood that this is done for illustrative purposes only. A person skilled in the relevant art will recognize that other components and configurations may be used without parting from the spirit and scope of the invention.

[0015] The present invention comprises a variety of embodiments, such as a method, an apparatus, and an electronic device, and other embodiments that relate to the basic concepts of the invention. The electronic device may be any manner of computer, mobile device, or wireless communication device.

[0016] A method, apparatus, and electronic device for voice navigation are disclosed. A voice input mechanism may receive a verbal input from a user to a voice user interface program invisible to the user. A processor may identify in a graphical user interface (GUI) a set of GUI items. The processor may convert the set of GUI items to a set of voice searchable indices. The processor may correlate a matching GUI item of the set of GUI items to a phonemic representation of the verbal input.

[0017] FIG. 1 illustrates in a block diagram one embodiment of a computing device 100 that may be used to implement a voice navigation method. Any computing device, such as a desktop computer, handheld device, or a server, may implement the voice navigation method. The computing device 100 may access the information or data stored in a network. The computing device 100 may support one or more applications for performing various communications with the network. The computing device 100 may implement any operating system, such as Windows or UNIX, for example. Client and server software may be written in any programming language, such as C, C++, Java or Visual Basic, for example. The computing device 100 may be a mobile phone, a laptop, a personal digital assistant (PDA), or other portable device. For some embodiments of the present invention, the computing device 100 may be a WiFi capable device, which may be used to access the network for data or by voice using voice over internet protocol (VoIP). The computing device 100 may include a network interface 102, such as a transceiver, to send and receive data over the network.

[0018] The computing device 100 may include a controller or processor 104 that executes stored programs. The controller or processor 104 may be any programmed processor known to one of skill in the art. However, the decision support method may also be implemented on a general-purpose or a special purpose computer, a programmed microprocessor or microcontroller, peripheral integrated circuit elements, an application-specific integrated circuit or other circuits, hardware/electronic logic circuits, such as a discrete element circuit, a programmable logic device, such as a programmable logic array, field programmable gate-array, or the like. In general, any device or devices capable of implementing the decision support method as described herein can be used to implement the decision support system functions of this invention.
The computing device 100 may also include a volatile memory 106 and a non-volatile memory 108 to be used by the processor 104. The volatile 106 and nonvolatile data storage 108 may include one or more electrical, magnetic or optical memories such as a random access memory (RAM), cache, hard drive, or other memory device. The memory may have a cache to speed access to specific data. The memory may also be connected to a compact disc-read only memory (CD-ROM), digital video disc-read only memory (DVD-ROM), DVD read-write input, tape drive or other removable memory device that allows media content to be directly uploaded into the system.

The computing device 100 may include a user input interface 110 that may comprise elements such as a keypad, display, touch screen, or any other device that accepts input. The computing device 100 may also include a user output device that may comprise a display screen and an audio interface 112 that may comprise elements such as a microphone, earphone, and speaker. The computing device 100 may also include a component interface 114 to which additional elements may be attached, for example, a universal serial bus (USB) interface or an audio-video capture mechanism. Finally, the computing device 100 may include a power supply 116.

Client software and databases may be accessed by the controller or processor 104 from the memory, and may include, for example, database applications, word processing applications, video processing applications as well as components that embody the decision support functionality of the present invention. The user access data may be stored in the database accessible through a database interface or in the memory. The computing device 100 may implement any operating system, such as Windows or UNIX, for example. Client and server software may be written in any programming language, such as C, C++, Java or Visual Basic, for example.

A graphical user interface (GUI) may allow the user to interact with a series of data objects stored in a computer or on the internet. A data object may be a file, webpage, an application, or other coherent set of computer data. The term “computer data” may refer to data found on the internet. The GUI may represent each data object with a GUI item, such as a hyperlink, soft-button, image, icon, or other representation of the data object. The GUI need not distinguish between GUI-viewed data objects from a computer or the internet.

FIG. 2 illustrates in a block diagram one embodiment of a GUI. The user interface 110 of the computing device 100 may be a display 202. The computing device 100 may interact with the user using a graphical user interface 204. A standard GUI 204 may present to a user one or more GUI items, such as icons 206 representing one or more data file objects on the display 202. A GUI item may be any representation shown in a GUI that acts as an input signal to open some type of data object. For a computing device 100 connected to a network, such as the internet, the GUI may be a browser 208 to present a webpage to a user. The webpage may have images 210 that link to other web pages. The web pages may have an icon or button 212 to activate a web application. Further, the webpage may have hyperlinks 214 linking to other web pages buried within the set of text 216 presented on the webpage. For items such as these browsers 208, where a number of new hyperlinks 214 are presented each time the browser is reloaded, voice recognition software that requires the construction of grammars to reflect the various ways user pronounces the hyperlinks may be impractical. GUIs with a large number of GUI items may be impractical for prompted voice navigation.

The GUI items from a view of a GUI may be harvested and dynamically translated into voice search indices. A voice user interface (VUI) may use the search indices to form a view-specific searchable database. The view of the display 202 may be voice-enabled just in time. FIG. 3 illustrates in a block diagram one embodiment of an invisible verbal user interface program 300. The display 202 may show a GUI 302 to the user. A GUI items harvester module 304 may search the GUI 302 for GUI items. GUI items may include hyperlinks 214, images 210, application icons 206, and other graphic images that lead to a data object. A data object may be a file, webpage, an application, or other coherent set of computer data. The GUI items harvester module 304 may collect all the GUI items in the GUI 302, as well as any contextual data associated with the GUI items. A parser 306, such as a text normalization module or a grapheme to phoneme module, may convert each GUI item in the GUI 302 into a searchable index in the form of a linguistic document. The parser 306 may take into account linguistic surface form, surrounding texts, hyperlinked webpage titles, metadata, and other data associated with the GUI item. A database of GUI item indices 308 may organize the linguistic documents into a searchable database to facilitate searching.

The VUI may convert a verbal input into a phoneme lattice to match against the searchable indices from the view-specific searchable database. A voice input mechanism 310 may receive a verbal input from a user. A phoneme decoder 312, or other voice recognition technology, may convert the verbal input into a phoneme lattice. A search term generator 314 may extract linguistic search terms from the phoneme lattice, such as a phoneme, syllable, or word string. A GUI items search engine 316 may take the linguistic search term and search the GUI items index 308. The GUI items search engine 316 may select a GUI item and may perform a navigation action associated with the matching GUI item to the GUI 302.

FIG. 4 illustrates in a block diagram one embodiment of voice searchable indices 400. The voice searchable indices 400 may be initially sorted by number of words (WRD) 402. The voice searchable indices 400 may be further sorted by phonemes (PHO) 404, the phonemes arranged in spoken order. The voice searchable indices 400 may include a GUI item type 406, such as image, hyperlink, application icon, or other GUI item type. The voice searchable indices 400 may also include an associated grapheme or commonly used name of the GUI item (GRPH) 408, such as picture, button, arrow, or other names. The voice searchable indices 400 may have a set of alternate linguistic labels (ALT) 410 to identify the GUI item, especially if the GUI item is an image or other GUI item that may be thought to have multiple label names by the user. The voice searchable indices 400 may include a link to the computer object (OBJ) 412 represented by the GUI item.

The VUI 300 may create a just-in-time, voice-enabled searchable database from a view of the GUI. FIG. 5 illustrates in a flowchart one embodiment of a method 500 for developing a voice searchable indices 400. The computing device 100 may display a GUI to the user (Block 502). The computing device 100 may identify a GUI item (GUI) of the GUI (Block 504). If the GUI is a non-textual GUI (Block 506), such as an image or unlabeled icon, the computing device 100 may develop alternate linguistic labels for the
GUII (Block 508). The computing device 100 may create textual description based on metadata for a GUII, commonly depicted names, surrounding text, labels, grapheme, and other data. The computing device 100 may convert the GUII to a linguistic document (LD) (Block 510). The computing device 100 may organize the LDs into a searchable database of GUII indices (Block 512).

Upon receiving a verbal input from the user, the VUII 300 may use the GUI item index 400 to select the GUI item best matched with the verbal input to the user. The VUII 300 may also select and present to the user a set of all alternate GUI items that are next best matched to the verbal input. FIG. 6 illustrates a flowchart one embodiment of a method 600 for invisible voice navigation. The VUII 300 may receive a verbal input from the user (Block 602). The VUII 300 may identify a set of possible matching GUIIs (Block 604). The VUII 300 may designate a primary matching GUII, or closest verbal match, and a set of one or more alternate GUIIs from the set of possible matching GUIIs (Block 606). The VUII 300 may identify a primary matching GUII and one or more alternate GUIIs (Block 608). The VUII 300 may present a computer object (CO) associated with the primary matching GUII (Block 610). The VUII 300 may present the alternate GUIIs to the user (Block 612). Alternatively, the VUII 300 may present an approximation of the computer objects associated with the alternate GUIIs. If the user selects one of the alternate GUIIs (Block 614), the VUII 300 may present the computer object associated with the selected alternate GUII (Block 616).

To account for the verbal ties of a user, the VUII 300 may keep a history of various users in order to determine which GUII to present as the primary matching GUII and which GUIIs to present as the alternates during repeated uses of the VUII 300. The VUII 300 may track if a specific verbal input is repeatedly used when referring to a specific GUII of a specific GUI. The VUII 300 may then present that GUII as the primary matching GUII. Further, for an initial use of a GUI by a user, the VUII 300 may use the histories of other users to determine a primary matching GUII when multiple GUIIs have a similar linguistic document.

The VUII 300 may briefly present the alternate GUI items to the user in a pop-up window. The pop-up window may be removed if no item is selected after a set period of time. If one of the alternate GUI items is selected, the VUII 300 may execute the action associated with the selected GUI item and override the initially presented view. FIG. 7 may illustrate in a block diagram one embodiment of a graphical voice navigation response 702 with alternate computer objects. The browser 208 may present a computer object 702 associated with the matching GUII. The browser 208 may also present approximate representations of the computer objects 704 associated with the next closest matches to the LST. If the user does not select one of the alternates after a set period of time, the alternate computer objects may be removed from the browser 208.

Embodiments may also be practiced in distributed computing environments where tasks are performed by local and remote processing devices that are linked (either by hard-wired links, wireless links, or by a combination thereof) through a communications network.

Embodiments within the scope of the present invention may also include computer-readable media for carrying or having computer-executable instructions or data structures stored thereon. Such computer-readable media can be any available media that can be accessed by a general purpose or special purpose computer. By way of example, and not limitation, such computer-readable media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to carry or store desired program code means in the form of computer-executable instructions or data structures. When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or combination thereof to a computer, the computer properly views the connection as a computer-readable medium. Thus, any such connection is properly termed a computer-readable medium. Combinations of the above should also be included within the scope of the computer-readable media.

Computer-executable instructions include, for example, instructions and data which cause a general purpose computer, special purpose computer, or special purpose processing device to perform a certain function or group of functions. Computer-executable instructions also include computer programs that are executed by computers in stand-alone or network environments. Generally, computer programs include routines, programs, objects, components, and data structures, etc. that perform particular tasks or implement particular abstract data types. Computer-executable instructions, associated computer programs, and computer modules represent examples of the computer code means for executing steps of the methods disclosed herein. The particular sequence of such executable instructions or associated data structures represents examples of corresponding acts for implementing the functions described in such steps.

Although the above description may contain specific details, they should not be construed as limiting the claims in any way. Other configurations of the described embodiments of the invention are part of the scope of this invention. For example, the principles of the invention may be applied to each individual user where each user may individually deploy such a system. This enables each user to utilize the benefits of the invention even if any one of the large number of possible applications do not need the functionality described herein. In other words, there may be multiple instances of the electronic device each processing the content in various possible ways. It does not necessarily need to be one system used by all end users. Accordingly, the appended claims and their legal equivalents should only define the invention, rather than any specific examples given.

We claim:

1. A method for voice navigation, comprising:
   identifying in a view of a graphical user interface (GUI) a set of GUI items;
   converting the set of GUI items to a set of voice searchable indices;
   creating at least one phonemic representation of a verbal input via a search-based voice user interface program invisible to the user; and
   identifying a matching GUI item of the set of GUI items to the phonemic representation.

2. The method of claim 1, further comprising:
   presenting a matching computer object associated with the matching GUI item to the user via the graphical user interface.

3. The method of claim 2, further comprising:
   identifying an alternate matching GUI item of the set of GUI items;
   presenting the alternate matching GUI item to the user; and
   receiving a user input.
4. The method of claim 1, further comprising:
identifying a non-textual GUI item in the set of GUI items;
and
developing an alternate linguistic label for the non-textual icon.

5. The method of claim 1, wherein the set of GUI items includes at least one of a hyperlink, application icon, file name, or image.

6. The method of claim 1, wherein the phonemic representation is a linguistic search term.

7. The method of claim 1, further comprising
converting a GUI item of the set of GUI items to a corresponding linguistic document; and
organizing each corresponding linguistic document into the set of voice searchable indices.

8. A telecommunications apparatus for voice navigation, comprising:
voice input mechanism that receives a verbal input from a user to a voice user interface program invisible to the user; and
a processor that identifies in a graphical user interface (GUI) a set of GUI items, converts the set of GUI items to a set of voice searchable indices, and correlates a matching GUI item of the set of GUI items to at least one phonemic representation of the verbal input.

9. The telecommunications apparatus of claim 8, further comprising:
(a) a display that presents a matching computer object associated with the matching GUI item to the user via the graphical user interface.

10. The telecommunications apparatus of claim 9, wherein:
the processor identifies an alternate matching GUI item of the set of GUI items;
the display presents the alternate matching GUI item to the user; and
the voice input mechanism receives a user input.

11. The telecommunications apparatus of claim 8, wherein the processor identifies a non-textual GUI item in the set of GUI items and develops an alternate linguistic label for the non-textual icon.

12. The telecommunications apparatus of claim 8, wherein the set of GUI items includes at least one of a hyperlink, application icon, file name, or image.

13. The telecommunications apparatus of claim 8, wherein the phonemic representation is a linguistic search term.

14. The telecommunications apparatus of claim 8, wherein the processor converts a GUI item of the set of GUI items to a corresponding linguistic document and organizes each corresponding linguistic document into the set of voice searchable indices.

15. An electronic device for voice navigation, comprising:
(a) a voice input mechanism that receives a verbal input from a user to a voice user interface program invisible to the user; and
(b) a processor that identifies in a graphical user interface (GUI) a set of GUI items, converts the set of GUI items to a set of voice searchable indices, and correlates a matching GUI item of the set of GUI items to at least one phonemic representation of the verbal input.

16. The electronic device of claim 15, further comprising:
(a) a display that presents a matching computer object associated with the matching GUI item to the user via the graphical user interface.

17. The electronic device of claim 16, wherein:
the processor identifies an alternate matching GUI item of the set of GUI items;
the display presents the alternate matching GUI item to the user; and
the voice input mechanism receives a user input.

18. The electronic device of claim 15, wherein the processor identifies a non-textual GUI item in the set of GUI items and develops an alternate linguistic label for the non-textual icon.

19. The electronic device of claim 15, wherein the set of GUI items includes at least one of a hyperlink, application icon, file name, or image.

20. The electronic device of claim 15, wherein the processor converts a GUI item of the set of GUI items to a corresponding linguistic document and organizes each corresponding linguistic document into the set of voice searchable indices.