(54) Title: AUDIO USER INTERFACE

(57) Abstract: An audio user interface that provides audio prompts that help a user interact with a user interface of an electronic device is disclosed. The audio prompts can provide audio indicators that allow a user to focus his or her visual attention upon other tasks such as driving an automobile, exercising, or crossing a street, yet still enable the user to interact with the user interface. An intelligent path can provide access to different types of audio prompts from a variety of different sources. The different types of audio prompts may be presented based on availability of a particular type of audio prompt. As examples, the audio prompts may include pre-recorded voice audio, such as celebrity voices or cartoon characters, obtained from a dedicated voice server. Absent availability of pre-recorded or synthesized audio data, non-vocal audio prompts may be provided.
Declarations under Rule 4.17:  
— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

Published:  
— with international search report (Art. 21(3))
AUDIO USER INTERFACE

CROSS-REFERENCES TO RELATED APPLICATIONS


FIELD OF THE INVENTION

[0002] This disclosure relates generally to audio user interfaces and, more specifically, to techniques for providing audio user interfaces for computing devices.

BACKGROUND OF THE INVENTION

[0003] Electronic devices, such as portable media players, cellular phones, personal digital assistants (PDAs), and the like, are prevalent in today’s marketplace, as are the peripheral electronic devices that support their use, such as docking stations and the like. As competition in the personal electronics marketplace becomes ever more heated, consumers have become more demanding in terms of both the functionality and use of such devices.

[0004] Users listen to, watch, or otherwise receive and consume content in a variety of contexts. For example, it is common to listen to music while driving, riding public transit, exercising, hiking, doing chores, or the like. Additionally, users now are more often using media players to receive radio, television, and satellite broadcasts, Global Positioning and other broadcast-based location services for navigation and recreation in addition to the playback of content stored on the media players.

[0005] Conventionally, a media player or portable media player can be capable of playing media, such as audio (e.g., songs) or video (e.g., movies) for the benefit of its user. When playing audio, if the media player includes a display, the display can present the name of the song, artist, and other information pertaining to the song. In the case of playing a video, the display can be used to present the video.
[0006] In order to achieve portability, many hand-held devices can use user interfaces that present various display screens to the user for interaction that is predominantly visual. Users can interact with the user interfaces to manipulate a scroll wheel and/or a set of buttons to navigate display screens to thereby access functions of the hand-held devices. However, these user interfaces can be difficult to use at times for various reasons. One reason is that the display screens may tend to be small in size and form factor and therefore difficult to see. Another reason is that a user may have poor reading vision or otherwise be visually impaired. Even if the display screens can be perceived, a user will have difficulty navigating the user interface in situations when a user cannot shift visual focus away from an important activity and towards the user interface. Such activities include, for example, driving an automobile, exercising, and crossing a street.

[0007] Accordingly, what is desired are improved methods and apparatus for solving some of the problems discussed above. Additionally, what is desired are improved methods and apparatus for reducing some of the drawbacks discussed above.

BRIEF SUMMARY OF THE INVENTION

[0008] In various embodiments, the experience of a user interactive with an electronic device, such as a media player or portable media device, can be enhanced through the incorporation of an audio user interface that provides an intelligent path for determining whether appropriate audio dialogs for the audio user interface are available. For example, based on whether the electronic device has a broadband connection to a communications network (e.g., the Internet), a determination may be made to request audio files of a first type or category (e.g., high quality voice recordings) from a voice server to be streamed to the electronic device for output with the audio user interface. In another example, a determination may be made to use only audio files of a second type or category (e.g., low quality voice recordings) which are available on a media storage device accessible to the electronic device. In yet another example, absent availability of pre-recorded voice audio data, a determination may be made to use one or more voice synthesis or text-to-speech techniques to create audio data of a third category for audio prompts for the audio user interface.

[0009] In some embodiments, a user of an electronic device, such as a media player or portable media device, can determine the quality of audio prompts that are to be presented
(e.g., played) for the audio user interface. The user may provide one or more user preferences indicating whether pre-recorded audio data should be used, whether audio prompts synthesized using one or more synthesis techniques should be used, or whether traditional beeps or other non-voice audio data should be used for the audio user interface. Accordingly, an electronic device, such as a media player or portable media device, with or without a display can be enhanced with an audio user interface to facilitate user interactions based on availability of service or other selection criteria.

[0010] In one embodiment, input indicative of an interaction of the user with a user interface associated with an electronic device, such as a media player or portable media device, may be received. The user may interact with the media player by pressing buttons, such as a play/pause button or selecting/highlighting a menu item of a graphical user interface. The electronic device may identify an audio prompt associated with audibilizing the interaction of the user with the user interface. The electronic device may determine whether one of a plurality of categories of audio data corresponding to the audio prompt is available to the media player. For example, the electronic device may determine whether pre-recorded celebrity-voice audio files are stored on an internal storage, whether a voice synthesis module or text-to-speech engine can synthesize numbers, or whether a voice server is able to stream audio data to the electronic device for the audio user interface.

[0011] A portion of audio data of a first category may then be output or otherwise presented at the electronic device. In some embodiments, playback of a media file can be paused or suspended in response to the output of the portion of audio data from the first source. A playback volume of a media file may be reduced or muted in response to the output of the portion of audio data from the first source.

[0012] A further understanding of the nature, advantages, and improvements offered by those inventions disclosed herein may be realized by reference to remaining portions of this disclosure and any accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] In order to better describe and illustrate embodiments and/or examples of any inventions presented within this disclosure, reference may be made to one or more accompanying drawings. The additional details or examples used to describe the accompanying drawings should not be considered as limitations to the scope of any of the
disclosed inventions, any of the presently described embodiments and/or examples, or the presently understood best mode of any invention presented within this disclosure.

[0014] FIG. 1 is a block diagram of a media player that may incorporate embodiments of the present invention;

[0015] FIG. 2 is a block diagram of a media player that may provide pre-recorded or synthesized audio prompts in one embodiment according to the present invention;

[0016] FIG. 3 is a block diagram of an audio user interface management system that may provide pre-recorded or synthesized audio prompts in one embodiment according to the present invention;

[0017] FIG. 4 is a block diagram of a streaming audio prompt system in one embodiment according to the present invention;

[0018] FIG. 5 illustrates a diagrammatic view of a media player and its associated user input controls in one embodiment according to the present invention;

[0019] FIG. 6 illustrates a diagrammatic view of a media player and its associated user input controls in an alternative embodiment according to the present invention;

[0020] FIG. 7 is a simplified flowchart of a method for providing an audio user interface to a user of an electronic device in one embodiment according to the present invention;

[0021] FIGS. 8A and 8B are a flowchart of a method for providing an electronic device with an audio user interface in an alternative embodiment according to the present invention;

[0022] FIG. 9 is a flowchart of a method for streaming audio prompts for an audio user interface in one embodiment according to the present invention;

[0023] FIG. 10 is a flowchart of a method for creating audio prompts at a host computer system using one or more voice or text to speech synthesis techniques in one embodiment according to the present invention;

[0024] FIG. 11 is a flowchart of a method for creating audio prompts using one or more voice or text to speech synthesis techniques in an alternative embodiment according to the present invention; and

[0025] FIG. 12 is a block diagram of an electronic device that may incorporate embodiments of the present invention.
DETAILED DESCRIPTION OF THE INVENTION

[0026] Various embodiments can be well suited for electronic devices having audio playback capabilities, such as media devices (e.g., digital media players or portable MP3 players) or other portable multi-function devices (e.g., mobile telephone or Personal Digital Assistant). For example, portable devices can often store and play digital media assets (media items), such as music (e.g., songs), videos (e.g., movies), audio books, podcasts, meeting recordings, and/or other multimedia recordings. Portable devices, such as portable media players or other portable multi-function devices, can also be small and highly portable. Often, portable devices are hand-held devices, such as hand-held media players or hand-held multi-function devices, which can be easily held by and within a single hand of a user. Portable devices can also be pocket-sized, miniaturized, or wearable.

[0027] In various embodiments, the experience of a user interactive with an electronic device, such as a media player or portable media device, can be enhanced through the incorporation of an audio user interface that provides an intelligent path for determining whether appropriate audio dialogs for the audio user interface are available. For example, based on whether the electronic device has a broadband connection to a communications network (e.g., the Internet), a determination may be made to request audio files of high quality voice recording from a voice server to be streamed to the electronic device for output with the audio user interface. In another example, a determination may be made to use only audio files of low quality voice recordings that are available on a media storage device accessible to the electronic device. In yet another example, absent availability of pre-recorded voice audio data, a determination may be made to use one or more voice synthesis or text-to-speech audio techniques to create audio prompts for the audio user interface.

[0028] In some embodiments, a user of an electronic device, such as a media player or portable media device, can determine the quality of audio prompts that are to be presented (e.g., played) for the audio user interface. The user may provide one or more user preferences indicating whether pre-recorded audio data should be used, whether audio prompts synthesized using one or more synthesis techniques should be used, or whether traditional beeps or other non-voice audio data should be used for the audio user interface. Accordingly, an electronic device, such as a media player or portable media device, with or without a display can be enhanced with an audio user interface to facilitate user interactions based on availability of service or other selection criteria.
[0029] Aspects of the environments within which various examples and/or embodiments of those invention found within the specification operate will first be described.

[0030] FIG. 1 is a block diagram of media player 100 that may incorporate embodiments of the present invention. In general, a media player stores content and/or media assets, such as audio tracks, movies, or photos that can be played or displayed on the media player. One example of media player 100 can be the iPod® media player, which is available from Apple, Inc. of Cupertino, CA. Another example of media player 100 can be a personal computer, such as a laptop or desktop.

[0031] In this example, media player 100 includes processor 110, storage device 120, user interface 130, and communications interface 140. Processor 110 can control various functionalities associated with media player 100. Media play 100 may output audio content, video content, image content, and the like. Media player 100 may also output metadata or other information associated with content, such as track information and album art.

[0032] Typically, a user may load or store content onto media player 100 using storage device 120. Storage device 120 can include read-only memory (ROM), random access memory (RAM), non-volatile memory, flash memory, floppy disk, hard disk, or the like. A user may interact with user interface 130 of media player 100 to view or consume content. Some examples of user interface 130 can include buttons, click wheels, touch pads, displays, touch screens, and other input/output devices.

[0033] Media player 100 can include one or more connectors or ports that can be used to load content, retrieve content, interact with applications running on media player 100, interface with external devices, and the like. In this example, media player 100 includes communications interface 140. Some examples of communications interface 140 can include universal serial bus (USB) interfaces, IEEE 1394 (or FireWire/iLink®) interfaces, universal asynchronous receiver/transmitters (UARTs), wired and wireless network interfaces, transceivers, and the like. Media player 100 may connect to devices, accessories, private and public communications networks (e.g., the Internet), or the like, using communications interface 140.

[0034] In one example, media player 100 can be coupled via a wired and/or wireless connector or port to output audio and/or other information to speakers 150. In another example, media player 100 may be coupled via a wired and/or wireless connector or port to output audio and/or other information to headphones 160. In yet another example, media
player 100 may be coupled via a wired and/or wireless connector or port to interface with an accessory 170 or a host computer 180. The same connector or port may enable different connections at different times.

[0035] Media player 100 can be physically inserted into docking system 190. Media player 100 may be coupled via a wired and/or wireless connector or port to interface with docking system 190. Docking system 190 may also enable one or more accessory devices 195 to couple with wires or wirelessly to interface with media player 100. Many different types and functionalities of accessory devices 170 and 195 can interconnect to or with media player 100. For example, an accessory may allow a remote control to wirelessly control media player 100. As another example, an automobile may include a connector into which media player 100 may be inserted such that an automobile media system can interact with media player 100, thereby allowing media content stored on media player 100 to be played within the automobile.

[0036] In various embodiments, media player 100 can receive content or other media assets from a computer system (e.g., host computer 160). The computer system may serve to enable a user to manage media assets stored on the computer system and/or stored on media player 100. As an example, communications interface 140 may allow media player 100 to interface with host computer 160. Host computer 160 may execute a media management application to manage media assets, such as loading songs, movies, photos, or the like, onto media player 100. The media management application may also create playlists, record or rip content, schedule content for playback or recording, or the like. One example of a media management application can be iTunes®, produced by Apple, Inc. of Cupertino, California.

[0037] In various embodiments, media player 100 may include an audio user interface. Embodiments of the audio user interface may present or otherwise output audio prompts selected from an audio dialog for playback as a user interacts with media player 100 (e.g., as the user presses a button, touches a touch screen, or selects items of a graphical user interface). The audio prompts may include audio indicators that allow a user to focus his or her visual attention upon other tasks such as driving an automobile, exercising, or crossing a street, yet still enable the user to interact with user interface 130. As examples, the audio prompts may audibilize the spoken name or description of a depressed hardware button, the spoken activation of a virtual button or control, or the spoken version of a user interface selection, such as a selected function or a selected (e.g., highlighted) menu item of a display.
menu. The audio prompts may include pre-recorded voice data or be produced by voice or speech generation techniques.

[0038] In one aspect, embodiments of media player 100 may include techniques for providing an audio user interface an electronic device that efficiently leveraging availability of audio prompt sources for the audio user interface. For example, media player 100 may selectively output audio prompts from different audio dialogs based on whether a connection to the source of an audio dialog is available, whether a higher quality source is available, or the like. In one example, a user of media player 100 may hear low quality voice audio prompts or audio prompts synthesized by media player 100 until connecting to the Internet, upon which higher quality pre-recorded voice audio prompts may be downloaded or streamed to the audio user interface. Thus, in various embodiments, media player 100 may determine the availability of a source of audio prompts for the audio user interface, and automatically switch form one source to another to selectively provide one the best available audio feedback to the user.

[0039] FIG. 2 is a block diagram of media player 200 that may provide pre-recorded or synthesized audio prompts in one embodiment according to the present invention. In this example, media player 200 can be embodied as media player 100 and may include a portable computing device dedicated to processing content or other media assets, such as audio, video, or images. For example, media player 200 can be a music player (e.g., MP3 player), a game player, a video player, a video recorder, a camera, an image viewer, a mobile phone (e.g., cell phones), a personal hand-held device, and the like. These devices are generally battery operated and highly portable so as to allow a user to listen to music, play games or video, record video or take pictures wherever the user travels.

[0040] In one implementation, media player 200 may include a hand-held device that is sized for placement into a pocket or hand of the user. By being hand-held, media player 200 may be relatively small and easily handled and utilized by its user. By being pocket-sized, the user does not have to directly carry media player 200 and therefore the device can be taken almost anywhere the user travels (e.g., the user is not limited by carrying a large, bulky and often heavy device, as in a portable computer). Furthermore, media player 200 may be operated by the user’s hands so that no reference surface such as a desktop is needed. In alternative embodiments, media player 200 may be computing devices that are not
specifically limited to playing media files. For example, media player 200 can also be a mobile telephone or a personal digital assistant.

[0041] In this example, media player 200 can include user interface control module 210, audio prompt database 220, and text-to-speech engine 230. User interface control module 210 can include hardware and/or software elements for managing a user interface that allows a user to interact (e.g., navigate, initiate content playback, etc.) with media player 200. The user interface can, for example, allow the user of media player 200 to browse, sort, search, play, etc. content or other media assets resident on or otherwise accessible to media player 200. The user interface can also allow the user of media player 200 to download (add) or delete (remove) media items from media player 200.

[0042] Interaction with the user interface of media player 200 can cause audio prompts for an audio user interface to be played back, such as through earphones or speakers. Audio prompt database 220 can include hardware and/or software elements for storing audio data and audio files for audio prompts. In some embodiments, the audio files may include audio prompts that have been pre-recorded and stored on media player 200. In further embodiments, the audio files may include audio prompts streamed from one or more computers and cached in audio prompt database 220 for subsequent use. In various embodiments, the audio files may include audio prompts generated by media player 200 or by another device using one or more voice synthesis techniques. Audio prompt database 220 may include other content or media assets.

[0043] Text-to-speech conversion engine 230 can include hardware and/or software elements for converting data, such as text, into audio files or audio data that can be played to generate a user interface audio prompt that audibilizes (e.g., verbalizes in a human-like voice or in spoken form) data, such as text string. Such text-to-speech (TTS) engines can use various techniques for creating the audio data or audio files. For example, some algorithms use a technique of breaking a word down into fragments or syllables for which a certain sound is then designated. Then, a word can be verbalized through combining individual sounds. In the case where the media content pertains to music, these text strings may, for example, correspond to song titles, album names, artist names, contact names, addresses, phone numbers, and playlist names.

[0044] In one example of operation, media player 200 may selectively provide audio prompts for an audio user interface based on availability of audio prompts for audio prompt
database 220 and TTS engine 230. For example, media player 200 may selectively output
audio prompts from audio prompts database 220 when pre-recorded audio prompts are
available or otherwise stored in audio prompt database 220. Media player 200 may further
selectively choose between audio prompts of various quality, such as presenting audio
prompts of a higher quality or bit rate that those of a lower quality or bit rate. In another
example, media player 100 may present voice prompts or audio prompts synthesized by TTS
engine 230 due to the lack of pre-recorded audio prompts stored in audio prompt database
220 or in response to a user's preference for a particular simulated voice profile. In various
embodiments, media player 100 may dynamically output audio prompts from audio prompts
database 220 or TTS engine 230, or in combination.

[0045] In further embodiments, an electronic device, such as a media player or portable
media device, may include an audio user interface provided by an audio user interface
management system. An audio user interface management system can include a media
playback device and one or more of a host computer or server computer system that
facilitates providing the audio user interface on the media playback device. As an example,
the host computer system may include a personal computer and the media playback device
can include an MP3 player. In some embodiments, the media playback device may provide
for multi-modal interaction with the user interface. For example, a user can interact with the
user interface through audio and visual prompts.

[0046] FIG. 3 is a block diagram of audio user interface management system 300 that may
provide pre-recorded or synthesized audio prompts in one embodiment according to the
present invention. In this example, management system 300 can include media player 310
and personal computer (host computer) 340. Media player 310 may be embodied as media
player 100, as discussed above, and can be linked or coupled to personal computer 340.

[0047] Media player 310 may be embodied as media player 100 of FIG. 1, and can include,
for example, a portable, battery-operated device. In one embodiment, media player 310
includes an MP3 player. Typically, media player 310 may store content or other media assets
to one or different data storage devices (e.g., disk drives). Media player 310 may store
content or other media assets in media files.

[0048] Media player 310 can include user interface control module 320 and audio prompt
database 330. User interface control module 320 can include hardware and/or software
elements for managing a user interface that allows a user to interact (e.g., navigate, initiate
content playback, etc.) with media player 310. Interaction with the user interface of media player 310 can cause audio prompts for an audio user interface to be played back, such as through earphones or speakers. Audio prompt database 330 can include hardware and/or software elements for storing audio data and audio files for audio prompts.

[0049] Personal computer 340 can include media manager 350, audio prompt database 360, and text-to-speech (TTS) engine 370. Personal computer 340 may serve as a host computer system to media player 310. Personal computer 340 can also be any type of computer that acts as a server to media player 310, which is the client.

[0050] Media manager 350 can include hardware and/or software elements for that enable a user of personal computer 350 to directly manage content or other media assets stored on personal computer 340. Media manager 350 may further be configured to directly or indirectly manage content or other media assets stored on media player 310. In one example, media player 310 and personal computer 340 can be coupled using a peripheral cable. Typically, a peripheral cable can couple together data ports provided on media player 310 and personal computer 340. In some embodiments, the data ports can be FIREWIRE ports and the peripheral cable can be a FIREWIRE cable. In another example, the data ports can be Universal Serial Bus (USB) ports and the peripheral cable can be a USB cable. More generally, the peripheral cable may act as a data link. Media items can be transferred between media player 310 and personal computer 340 over the peripheral cable, and vice versa.

[0051] In various embodiments, media manager 350 may also include a user interface that allows a user to browse, sort, search, play, make playlists from, burn Compact Discs (CDs) of, etc. the content or other media assets resident on personal computer 340. The user interface can also allow the user of personal computer 340 to download (add) or delete (remove) media items from personal computer 340. In one embodiment, media manager 350 and its associated user interface are provided by iTunes™, from Apple Inc. of Cupertino, California.

[0052] Audio prompt database 360 of personal computer 340 may include hardware and/or software elements for storing audio data or audio files for audio prompts for an audio user interface associated with media player 310 or personal computer 340. Audio prompt database 330 may include audio prompts for audio dialogs that have been downloaded from the Internet, ripped from a CD, recorded by the user, or generated by TTS engine 370. TTS
engine 370 may include hardware and/or software elements for converting information or data into audio files or voice data that can be played as audio prompts that audibilizes the information.

[0053] In one example, a synchronization operation between personal computer 340 and media player 310 may occur to upload audio prompts to or update audio prompts stored in audio prompt database 330 of media player 310 with audio prompts stored in audio prompt database 360 or generated by TTS engine 370. In one example, when comparison of the contents from respective databases indicate that there is a particular audio prompt resident on personal computer 340 that is not resident on media player 330, then the particular audio prompt may be transmitted (downloaded) to media player 310, e.g., using a wireless link or over a peripheral cable. Accordingly, a synchronization operation between personal computer 340 and media player 310 can ensure that media player 310 contains audio data or audio files appropriate for presenting a usable audio user interface.

[0054] The number of audio files to be downloaded onto media player 310 may depend upon user settings for the audio user interface. For example, a user may desire that audio files or other audio data stored in audio prompt database 360 be downloaded for association with all or a subset of options or features of the audio user interface on media player 310.

[0055] FIG. 4 is a block diagram of streaming audio prompt system 400 in one embodiment according to the present invention. In this example, media player 410 is linked to communications network 420. Media player 410 may be embodied as media player 200 of FIG. 2 or as media player 310 of FIG. 3. Voice server 430 is also linked to communications network 420, and is able to communicate with media player 410.

[0056] In various embodiments, media player 410 may detect that a connection exists to voice server 430 via communications network 420. In one example of operation, media player 410 may select to receive audio prompts from voice server 430 for presentation with an audio user interface of media player 410. Media player 410 may generate one or more requests for an audio prompt, and upon receiving a request, voice server 430 may stream a corresponding audio prompt to media player 410 for output to a user.

[0057] Voice server 430 can include audio prompt database 440 and TTS engine 450. Audio prompt database 440 of voice server 430 may include hardware and/or software elements for storing audio data or audio files for audio prompts for an audio user interface associated with media player 410. Audio prompt database 330 may include audio prompts
for audio dialogs that have been pre-recorded by one or more content producers, provided by content publishers, or generated by TTS engine 450. TTS engine 370 may include hardware and/or software elements for converting information or data into audio files or voice data that can be played as audio prompts that audibilizes the information.

Accordingly, media player 410 may selectively choose between the sources of audio prompts for an audio user interface to provide audio voice feedback to a user. Media player 410 may receive audio prompts (e.g., pre-recorded or synthesized) from voice server 430 until a connection is lost. At that point, media player 410 may automatically select audio prompts from a different source, such as an internal audio prompt database or speech synthesis module.

FIG. 5 illustrates a diagrammatic view of media player 500 and its associated user input controls in one embodiment according to the present invention. Media player 500 can include any computer device for playing media files, such as song files. Media player 500 may contain memory that stores a media database and a play module for presenting or playing content or other media assets stored in the media database. A set of nested menus 505 may represent at least part of a user interface that allows a user to navigate through, select, and thereby listen to desired song files. It may be possible to reach a certain media file through different paths using the set of nested menus 505. The user interface may also allow users to navigate and select a desired function provided by media player 500.

FIG. 5 also illustrates representative user interface controls 510 of media player 500. According to one embodiment, user interface controls 510 include menu button 515, next button 520, play/pause button 525, and previous button 530. User interface control 510 may include a scroll wheel implemented as a rotating wheel apparatus that can rotate or a touch pad apparatus that understands a rotation user gesture. A user may press, rubs, or otherwise interact with user interface controls 510 to navigate through nested menus 505.

FIG. 6 illustrates a diagrammatic view of media player 600 and its associated user input controls in an alternative embodiment according to the present invention. Media player 600 can include previous button 610, play/pause button 620, and next button 630. LEDs 640 and 650 may be used to convey information to the user, such as to indicate power status or media playback status. In this example, media player 600 may not include a display configured to a graphical user interface, such as nested menus 505 of FIG. 5. Accordingly, a
user interface that audibly conveys information about operation of media player 600 may
dramatically enhance user experience.

[0062] FIG. 7 is a simplified flowchart of a method for providing an audio user interface to
a user of an electronic device in one embodiment according to the present invention. The
processing of method 700 depicted in FIG. 7 may be performed by software (e.g., instructions
or code modules) when executed by a central processing unit (CPU or processor) of a logic
machine, such as a computer system or information processing device, by hardware
components of an electronic device or application-specific integrated circuits, or by
combinations of software and hardware elements. FIG. 7 begins in step 710.

[0063] In step 720, information indicative of an interaction of a user with a user interface is
received. The information may include signals, messages, interrupts, input, or the like. The
information may specify that a user has pressed or depressed a button, clicked a click-wheel,
touched a touch screen, performed a gesture, highlighted or selected an element of a graphical
user interface, or the like. The information may represent a single act of a user or a
combination of acts.

[0064] In step 730, an audio prompt corresponding to the interaction of the user is
identified. The audio prompt may include information identifying audio data that vocalizes,
audibilizes, or otherwise provides feedback to a user of a registered interaction. In step 740,
a type or category of audio data for the audio prompt is determined. In various embodiments,
the audio prompt may be represented by different types or categories of audio data. A type or
category of audio data may include, for example, audio data of different audible qualities,
voice vs. non-voice, bit rate, compression, encoding, source, delivery mechanism, or the like.
For example, synthesized audio data generated by a voice synthesis module may be used to
provide audio prompts for numbers, dates, or the like. In another example, pre-recorded
audio data that has been compressed may be used to provide audio prompts for button
interactions, such as play, pause, next, back, forward, reverse, or the like. In yet another
example, CD-quality pre-recorded audio data may be used to provide an entire set of audio
prompts for numbers, dates, button presses, menu selections, and any other interaction of the
user that may be included in a given audio user interface.

[0065] In step 750, availability of the determined type or category of audio data for the
audio prompt is determined. For example, a selection can be made to use pre-recorded audio
dialog (e.g., a set of pre-recorded audio files) for audio prompts of an audio user interface.
An electronic device may check its internal storage to determine whether an audio file for the audio prompt exists. Alternatively, the electronic device may requests the audio file for the audio prompt from a host computer or a streaming voice server. In another example, a selection may be made to use pre-recorded audio data for some audio prompts and synthesized audio data for other audio prompts if pre-recorded audio prompts are not stored locally at the electronic device.

[0066] In step 760, a portion of audio data of the determined type or category is output from an available source. Accordingly, various embodiments may provide dynamic selection of different types or categories of audio data for audio prompts of an audio user interface. Additionally, some embodiments may provide mechanisms for getting the selected or identified type or categories of audio data to an electronic device for use as part of an audio user interface. FIG. 7 ends in step 770.

[0067] FIGS. 8A and 8B are a flowchart of method 800 for providing an electronic device with an audio user interface in one embodiment according to the present invention. Method 800 generally involves an intelligent decision path that determines whether an appropriate audio dialog for an audio user interface is available and obtaining the best available audio dialog for output to a user. FIG. 8A begins in step 805.

[0068] In step 810, input indicative of a button press is received. For example, a user may interface with user interface controls 510 of media player 500 of FIG. 5. Media player 500 may generate one or more analog or digital signals representative of a button press, touch, pressure, gesture, motion, or the like.

[0069] In step 815, a determination is made whether to present an audio prompt for the button press. In some embodiments, control selections are accompanied by an indication to output an audio prompt that confirms the selection to the user. For example, “play” can be audibilized to the user to provide feedback that play/pause button 525 was actually depressed. These embodiments may involve a repeated user action to make a user interface control selection. For example, a user would make multiple “clicks” of a user interface control to make the selection. A first “click” can cause media player 500 to determine to audibilize the selected user interface control. For example, “play” may be audibilized when a user presses the play button. This first audio prompt can provide audio guidance as to which button has been depressed, which is helpful to a user when not directing visual attention upon the handheld device.
[0070] A subsequent “click” may then cause media player 500 to perform the action corresponding to the user interface control. For example, pressing the play button a second time may cause a media file to be played. On the other hand, the audio prompt may have informed the user that an unintended selection is about to be made. Therefore, the user can attempt to select a different user interface control. For example, the user may then attempt to press “next” button 520, rather than proceeding to press play button 525 for a second time.

[0071] If a determination is made to present an audio prompt for the button press in step 815, the processing follows an intelligent decision path that determines whether an appropriate dialog for the audio prompt is available and how to get the appropriate audio dialog onto an electronic device. The intelligent decision path may include, for example, discovery or identification of types or categories of audio data and whether the audio data is available.

[0072] In step 820, a determination is made whether a high quality sources is available. A high quality source may include digital audio files or audio data sampled above a pre-determine or recognized frequency, at a give bit rate, whose size exceeds a pre-determined threshold or limit, or the like, relative to a lower quality source. The determination may be made based on whether a wireless or wired connection to a communications network exists over which the high quality source is accessible. In one implementation, the determination may be made based on selection criteria or user preferences. For example in one mode of operation, a user may desire to hear an audio prompt for every action and menu item selected by the user. In another mode, a user may deactivate audio prompts for the control selections (e.g., the “play” button) and only hear audio prompts for the highlighted menu items. In another mode, audio prompts can be output for only top-level menu items.

[0073] If a determination is made that a high quality sources is available, in step 825, the audio prompt is retrieved from the high quality source corresponding to the button press. One example of a high quality source may include lossless or CD-quality pre-recorded audio data or audio files. The pre-recorded audio data or audio files may include professionally made recordings of celebrity voices, cartoon characters, or quotes from television shows or feature films.

[0074] If a determination is made in the alternative that a high quality source is not available, in step 830, a determination is made whether a low quality source is available. If a determination is made that a low quality sources is available, in step 835, the audio prompt is
retrieved from the low quality source corresponding to the button press. One example of a low quality source may include pre-recorded audio data or audio files compresses using one or more compression or encoding techniques, such as MP3, WMA, OGG, or the like. The pre-recorded audio data or audio files may include generic recordings of human voices or stored audio data or audio files generated using one or more voice or text synthesis techniques.

[0075] Referring now to FIG. 8B, if a determination is made in the alternative that a low quality source is not available, in step 840, a determination is made whether text-to-speech (TTS) or voice synthesis is available. If a determination is made that one or more synthesis sources are available, in step 845, the audio prompt is synthesized or generated using voice synthesis or TTS synthesis.

[0076] If no source of audio prompts for the audio user interface can be determined and selected, in step 850, one or more beeps or other generic sounds may be output corresponding to the button press. Preferably, in step 855, the audio prompt corresponding to the button press is output that was selectively obtained from a high quality source in step 825, obtained from a low quality source in step 835, or synthesized in step 845. The audio prompts, in some embodiments, can be played according to a selected audio interface mode. When a media player or portable media device is not playing an audio file, only audio files corresponding to the user interface may be played and made audible to the user.

[0077] In various embodiments, when a media file is being played back, the audio interface mode can be set to mix the media file and an audio prompt playback in different manners. According to one setting, the volume for playing back a media file may be dynamically reduced when an audio prompt is to be played. For example, the volume for playing back a song or a movie clip can be lowered during the playback of the audio prompt. According to another setting, playback of a media file is paused during the playback of an audio prompt and then restarted after the audio prompt has been played. If a user is making multiple user control selections in a certain time frame, the playback of the media file can be paused for a short period of time so that the playback of the media file need not be paused and restarted multiple times. This can avoid a repeated interruption of a song’s playback. For instance, playback of a media file can be paused for five seconds if a user makes at least three user control selections within 5 seconds. The time periods and number of user control selections
may vary depending upon a user’s preference. Some audio interface modes can designate
that the audio prompts be played through a left, right, or both speakers or earphone channels.

[0078] Accordingly, a determination may be made whether an appropriate audio dialog is
available, such as either on an electronic device or a host/server computer connected to the
device, and whether the best available audio dialog can be obtained for output to the user.
FIG. 8B ends in step 860.

[0079] FIG. 9 is a flowchart of method 900 for streaming audio prompts for an audio user
interface in one embodiment according to the present invention. Method 900 generally
involves streaming audio prompts to a media playback device based on a connection to a
voice server. FIG. 9 begins in step 910.

[0080] In step 920, a media playback device (e.g., media player 100) detects a broadband
connection. For example, the media playback device may successfully associate with a
wireless access point. In yet another example, the media playback device may recognize a
wired connection to the Internet.

[0081] In step 930, the media playback device determines to use a voice server to obtain
voice dialogs for the audio user interface. For example, a software program executed by the
media playback device may initiate and complete a handshake with one or more applications
hosted by the voice server. In another example, the media playback device may periodically
poll the voice server to determine whether a connection is available.

[0082] In step 940, the media playback device generates a request for the audio prompt.
The request may include information identifying the audio prompt, information identifying a
user’s interaction corresponding to the requested audio prompt, or the like. The request may
include one or more of headers, flags, fields, checks, hashes, or the like. In one embodiment,
the request may include hypertext transport protocol (HTTP) data or real-time transport
protocol (RTP) data.

[0083] In step 950, the voice server streams the audio prompt to the media playback device.
In step 960, the media playback device outputs the streamed audio prompt. The voice server
may utilize one or more streaming transport protocols (e.g., real-time or faster than real-time)
such that the media playback device buffers a portion of the audio prompt before playback.

[0084] In various embodiments, the voice server may be accessible on a pay per item or
subscription basis. The voice server may support the streaming of uncompressed and
compresses (e.g., lossless or lossy) audio data. The voice server may further support the
communication of information associated with content or other media assets with which a
user may interact (e.g., navigate), such as title information, album information, artist
information, genre information, metadata, or the like. FIG. 9 ends in step 970.

[0085] FIG. 10 is a flowchart of method 1000 for creating audio prompts at a host
computer system using one or more voice or text to speech synthesis techniques in one
embodiment according to the present invention. Method 1000 generally involves
synthesizing audio prompts for an audio user interface and transferring the synthesized audio
prompts to a media playback device. FIG. 10 begins in step 1010.

[0086] In step 1020, a media playback device (e.g., media player 100 of FIG. 1) detects a
connection to a host computer. For example, the media playback device may detect that the
media playback device is coupled to the host computer using a peripheral cable. In another
example, the media playback device may detect the proximity of the host computer and
establish a wireless connection, for example, using WiFi or Bluetooth modules.

[0087] In step 1030, the media playback device determines to use the host computer to
obtain voice dialogs for an audio user interface. For example, the media playback device
may determine to use the host computer when an internal storage of the media playback
device does not have sufficient space for storing audio prompts in addition to content or other
media assets. In another example, the media playback device may determine to use the host
computer when the media playback device does not include a TTS engine.

[0088] In step 1040, the host computer synthesizes an audio prompt. The host computer
may use one or more voice synthesis or text-to-speech synthesis techniques to generate the
audio prompt. For example, the host computer may determine a profile associated with the
media playback device. The profile may include textual descriptions of events registered
through button presses, menu selections, or other user interactions, specific to an electronic
device. The host computer may audibilize the textual descriptions of the profile by
generating and recording a synthesized voice reading. The host computer may generate a
single audio prompt for each textual description. The host computer may also generate a
single audio prompt that includes audio data for each textual description in addition to
information indicative of audio data for a given textual description within the single audio
prompt.
In step 1050, the host computer transfers the audio prompt to the media playback device. In one implementation, the host computer generates a plurality of audio prompts of an audio dialog for the audio user interface. The host computer then may transfer the entire audio dialog to the media playback device, for example, when managing content or other media assets on the device. In another example, the host computer may generate and transfer the audio prompt to the media playback device in substantially real-time. In step 1060, the media playback device outputs the audio prompt. FIG. 10 ends in step 1060.

FIG. 11 is a flowchart of method 1100 for creating audio prompts using one or more voice or text to speech synthesis techniques in an alternative embodiment according to the present invention. Method 1100 generally involves creating or synthesizing audio data representing textual descriptions of an event. FIG. 11 begins in step 1110.

In step 1120, an event is identified. An event may include any user interface possible with an electronic device. The event may be represented by a button press, a click, a scroll, a touch, a selection, a highlight, or the like by the user. In step 1130, a textual description of the identified event is determined. The textual description may include words, sentences, etc. that describe the event, a device, a user, a portion of content, or the like. The textual description may be generated by a user, a developer, or other third party.

In step 1140, voice audio is synthesized or otherwise generated based on the textual description of the event. In one example, a computer system may retrieve configuration settings for a text-to-speech conversion process. The configuration settings can control various aspects of the voice synthesis or text-to-speech conversion process. For example, the configuration settings can determine the certain text strings to be converted into audio files, quality of the TTS conversions, gender of the voice that verbalizes the text strings, the speed at which an audio prompt is audible (e.g., a speaking rate can be increased as the user gets more familiar with the audio prompts), and customizing voices to different subtasks (e.g., the controls and function can be audible with one voice while data such as songs and contact names can be audible with a different voice). Furthermore, a configuration setting can handle adept manipulation of user interface controls by playing only a part of an audio prompt as a user navigates. For example, while browsing through contact names lexicographically, only the letter (a, b, c . . .) is rendered until the user reaches the contact name that start with a desired letter. For example, j, as in Jones. Accordingly, it should be
understood that the TTS configuration settings can have various settings to correspond to a device, configuration, or user’s desires.

[0093] Various sound synthesizer rules and engines can be used to generate the audio file. A generalized example of a process for converting a word into an audio file can operate as follows. The process for converting the word “browse” begins by breaking the word into fragments that represent dipphone units or syllables, such as “b” “r” “ow” “s”. Then various techniques generate audio prompts for each component, which can then be combined to form an intelligible word or phrase. The audio file is typically given an extension that corresponds to the type of audio file created. For example, the audio file for “browse” can be identified by a browse.aiff filename, wherein the .aiff extension indicates an audio file.

[0094] In step 1150, the voice audio prompt is output. The voice audio prompt may be output in response to user interactions with a media playback device having an audio user interface. In one embodiment, the audio user interface may include includes pointers to the corresponding audio prompts or audio files. For example, a lookup table may be used for keeping track of correlating pointers to audio prompts. FIG. 11 ends in step 1160.

[0095] FIG. 12 is a simplified block diagram of a computer system 1200 that may incorporate embodiments of the present invention. FIG. 12 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[0096] In one embodiment, computer system 1200 includes processor(s) 1210, random access memory (RAM) 1220, disk drive 1230, input device(s) 1240, output device(s) 1250, display 1260, communications interface(s) 1270, and a system bus 1280 interconnecting the above components. Other components, such as file systems, storage disks, read only memory (ROM), cache memory, codecs, and the like may be present.

[0097] RAM 1220 and disk drive 1230 are examples of tangible media configured to store data such as audio, image, and movie files, operating system code, embodiments of the present invention, including executable computer code, human readable code, or the like. Other types of tangible media include floppy disks, removable hard disks, optical storage media such as CD-ROMS, DVDs and bar codes, semiconductor memories such as flash memories, read-only-memories (ROMS), battery-backed volatile memories, networked storage devices, and the like.
In various embodiments, input device 1240 is typically embodied as a computer mouse, a trackball, a track pad, a joystick, a wireless remote, a drawing tablet, a voice command system, an eye tracking system, a multi-touch interface, a scroll wheel, a click wheel, a touch screen, an FM/TV tuner, audio/video inputs, and the like. Input device 1240 may allow a user to select objects, icons, text, and the like, via a command such as a click of a button or the like. In various embodiments, output device 1250 is typically embodied as a display, a printer, a force-feedback mechanism, an audio output, a video component output, and the like. Display 1260 may include a CRT display, an LCD display, a Plasma display, and the like.

Embodiments of communications interface 1270 may include computer interfaces, such as include an Ethernet card, a modem (telephone, satellite, cable, ISDN), (asynchronous) digital subscriber line (DSL) unit, FireWire interface, USB interface, and the like. For example, these computer interfaces may be coupled to a computer network 1290, to a FireWire bus, or the like. In other embodiments, these computer interfaces may be physically integrated on the motherboard or system board of computer system 1200, and may be a software program, or the like.

In various embodiments, computer system 1200 may also include software that enables communications over a network such as the HTTP, TCP/IP, RTP/RTSP protocols, and the like. In alternative embodiments of the present invention, other communications software and transfer protocols may also be used, for example IPX, UDP or the like.

In various embodiments, computer system 1200 may also include an operating system, such as Microsoft Windows®, Linux®, Mac OS X®, real-time operating systems (RTOSs), open source and proprietary OSs, and the like.

FIG. 12 is representative of a media player and/or computer system capable of embodying the present invention. It will be readily apparent to one of ordinary skill in the art that many other hardware and software configurations are suitable for use with the present invention. For example, the media player may be a desktop, portable, rack-mounted or tablet configuration. Additionally, the media player may be a series of networked computers. Moreover, the media player may be a mobile device, an embedded device, a personal digital assistant, a smart phone, and the like. In still other embodiments, the techniques described above may be implemented upon a chip or an auxiliary processing board.
[0103] The present invention can be implemented in the form of control logic in software or hardware or a combination of both. The control logic may be stored in an information storage medium as a plurality of instructions adapted to direct an information-processing device to perform a set of steps disclosed in embodiments of the present invention. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will appreciate other ways and/or methods to implement the present invention.

[0104] The embodiments discussed herein are illustrative of one or more examples of the present invention. As these embodiments of the present invention are described with reference to illustrations, various modifications or adaptations of the methods and/or specific structures described may become apparent to those skilled in the art. All such modifications, adaptations, or variations that rely upon the teachings of the present invention, and through which these teachings have advanced the art, are considered to be within the scope of the present invention. Hence, the present descriptions and drawings should not be considered in a limiting sense, as it is understood that the present invention is in no way limited to only the embodiments illustrated.

[0105] The above description is illustrative but not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of the disclosure. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the pending claims along with their full scope or equivalents.
WHAT IS CLAIMED IS:

1. A method performed by a media player for providing an audio user interface to a user of the media player, the method comprising:
   receiving input indicative of an interaction of the user with a user interface associated with the media player;
   identifying an audio prompt associated with audibilizing the interaction of the user with the user interface;
   determining whether one of a plurality of categories of audio data corresponding to the audio prompt is available to the media player; and
   outputting a portion of audio data of a first category at the media player from a first source.

2. The method of claim 1 wherein determining whether one of a plurality of categories of audio data corresponding to the audio prompt is available comprises determining whether one of pre-recorded voice audio data, synthesized voice audio data, or non-voice audio data is available to the media player.

3. The method of claim 1 wherein determining whether one of a plurality of categories of audio data corresponding to the audio prompt is available to the media player comprises determining availability of the first source whose quality of audio data is greater than audio data associated with a second source.

4. The method of claim 1 wherein determining whether one of a plurality of categories of audio data corresponding to the audio prompt is available to the media player comprises determining availability of the first source based on unavailability of a second source.

5. The method of claim 1 further comprising:
   obtaining the audio data from the first source.

6. The method of claim 5 wherein obtaining the audio data from the first source comprises retrieving an audio file from a storage device local to the media player.

7. The method of claim 5 wherein obtaining the audio data from the first source comprises retrieving audio data from a host computer system.
8. The method of claim 5 wherein obtaining the audio data from the first source comprises receiving audio data from a voice synthesis module configured to generate audio data using one or more voice synthesis techniques.

9. The method of claim 8 wherein the voice synthesis module is local to the media player.

10. The method of claim 1 wherein the interaction of the user with the user interface associated with the media player comprises a play selection, a next selection, a previous selection, or a stop selection.

11. The method of claim 1 wherein the interaction of the user with the user interface associated with the media player comprises selection of a menu item from a menu displayed on the media player.

12. The method of claim 1 further comprising:
   pausing playback of a media file in response to the output of the portion of audio data from the first source.

13. The method of claim 1 further comprising:
   reducing a playback volume of a media file in response to the output of the portion of audio data from the first source.

14. The method of claim 1 wherein the media player comprises a handheld or wearable device.

15. A computer-readable medium configured to store a set of code modules which when executed by a processor of a media playback device become operational for providing an audio user interface at the media playback device, the computer-readable medium comprising:
   code for receiving input indicative of an interaction of the user with a user interface associated with the media playback device;
   code for determining a type of audio data to use for the audio user interface;
   code for determining whether a source for the type of audio data to be used is available to the media playback device; and
code for outputting a portion of audio data of the determined type at the media
playback device, the audio data of the determined type corresponding to an audio prompt
associated with audibilizing the interaction of the user with the user interface.

16. A portable media playback device comprising:
a media playback system;
a user interface; and
a processor configured to:
receive user input via the user interface;
identify an audio prompt of an audio user interface, the audio prompt
associated with audibilizing an interaction of a user with the user interface represented by the
user input;
determine whether one of a plurality of categories of audio data
corresponding to the audio prompt is available; and
initiate playback of a portion of audio data of a first category using the
media playback system.

17. The portable media playback device of claim 16 wherein the processor
is configured to determine whether one of the plurality of categories of audio data is available
by determining availability of the first source whose quality of audio data is greater than
audio data associated with a second source.

18. The portable media playback device of claim 16 wherein the processor
is configured to determine whether one of the plurality of categories of audio data is available
by determining availability of the first source based on unavailability of a second source.

19. The portable media playback device of claim 16 further comprising the
first source including a storage device configured to obtain voice audio data.

20. The portable media playback device of claim 16 wherein the processor
is further configured to obtain voice audio data from a host computer system.

21. The portable media playback device of claim 16 wherein the processor
is further configured to obtain synthesized voice audio data from a voice synthesis module.

22. A media player comprising:
a user interface;
a storage device;
a media playback subsystem;
a voice synthesis module; and
a processor configured to:
receive input indicative of an interaction of a user with the user interface;
identify an audio prompt associated with audibilizing the interaction of the user with the user interface;
determine whether one of a plurality of categories of audio data corresponding to the audio prompt is available;
instruct the media playback subsystem to output a portion of audio data of a first category from the storage device or a portion of audio data of a second category from the voice synthesis module corresponding to the audio prompt.

23. The media player of claim 22 wherein the processor is further configured to instruct the media playback subsystem to output a portion of audio data of a third category available from a voice server configured to stream audio data.

24. The media player of claim 22 wherein the processor is further configured to instruct the voice synthesis module to generate synthesized audio data of the second category corresponding to the audio prompt.

25. The media player of claim 22 wherein the processor is further configured to receive audio data of the first category or audio data of the second category from a host computer.
FIG. 1
FIG. 2
FIG. 3
FIG. 5

FIG. 6
BEGIN

RECEIVE INFORMATION INDICATIVE OF INTERACTION OF USER WITH USER INTERFACE

IDENTIFY AUDIO PROMPT CORRESPONDING TO INTERACTION OF USER

DETERMINE TYPE OR CATEGORY OF AUDIO DATA FOR AUDIO PROMPT

DETERMINE AVAILABILITY OF TYPE OR CATEGORY OF AUDIO DATA FOR AUDIO PROMPT

OUTPUT PORTION OF AUDIO DATA OF DETERMINED TYPE OR CATEGORY FROM AVAILABLE SOURCE

END

FIG. 7
BEGIN

RECEIVE INPUT INDICATIVE OF BUTTON PRESS

DETERMINE WHETHER TO PRESENT AUDIO PROMPT FOR BUTTON PRESS

HIGH QUALITY SOURCE AVAILABLE?

YES

RETRIEVE AUDIO PROMPT FROM HIGH QUALITY SOURCE CORRESPONDING TO BUTTON PRESS

NO

LOW QUALITY SOURCE AVAILABLE?

YES

RETRIEVE AUDIO PROMPT FROM LOW QUALITY SOURCE CORRESPONDING TO BUTTON PRESS

NO

A

B

FIG. 8A
SYNTHESIS AVAILABLE?

YES

GENERATE AUDIO PROMPT USING VOICE SYNTHESIS OR TTS SYNTHESIS

NO

OUTPUT ONE OR MORE BEEPS CORRESPONDING TO BUTTON PRESS

OUTPUT AUDIO PROMPT CORRESPONDING TO BUTTON PRESS

END

FIG. 8B
BEGIN

MEDIA PLAYBACK DEVICE DETECTS BROADBAND CONNECTION

MEDIA PLAYBACK DEVICE DETERMINES TO USE VOICE SERVER TO OBTAIN VOICE DIALOGS FOR AUDIO USER INTERFACE

MEDIA PLAYBACK DEVICE GENERATES REQUEST FOR AUDIO PROMPT

VOICE SERVER STREAMS AUDIO PROMPT TO MEDIA PLAYBACK DEVICE

MEDIA PLAYBACK DEVICE OUTPUTS STREAMED AUDIO PROMPT

END

FIG. 9
1000

BEGIN

1010

MEDIA PLAYBACK DEVICE DETECTS CONNECTION TO HOST COMPUTER

1020

MEDIA PLAYBACK DEVICE DETERMINES TO USE HOST COMPUTER TO OBTAIN VOICE DIALOGS FOR AUDIO USER INTERFACE

1030

HOST COMPUTER SYNTHESIZES AUDIO PROMPT

1040

HOST COMPUTER TRANSFERS AUDIO PROMPT TO MEDIA PLAYBACK DEVICE

1050

MEDIA PLAYBACK DEVICE OUTPUTS AUDIO PROMPT

1060

END

1070

FIG. 10
1100
BEGIN
1110
IDENTIFY EVENT
1120
DETERMINE TEXTUAL DESCRIPTION OF EVENT
1130
SYNTHESIZE VOICE AUDIO BASED ON TEXTUAL DESCRIPTION OF EVENT
1140
OUTPUT VOICE AUDIO
1150
END
1160

FIG. 11
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

**INV. G06F3/16**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06F G01C G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 2006/095848 A1 (NAIK DEVANG K [US])</td>
<td>1-25</td>
</tr>
<tr>
<td></td>
<td>4 May 2006 (2006-05-04) cited in the application</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0006]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0008]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0028]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0033] - paragraph [0034]; figure 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0043]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0051] - paragraph [0052]; figure 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0084] - paragraph [0085]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0092]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>claims 2,4,9,10</td>
<td></td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

- **A** document defining the general state of the art which is not considered to be of particular relevance
- **E** earlier document but published on or after the international filing date
- **L** document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- **O** document referring to an oral disclosure, use, exhibition or other means
- **P** document published prior to the international filing date but later than the priority date claimed

* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

* document member of the same patent family

**Date of the actual completion of the international search**

26 October 2009

**Name and mailing address of the ISA/IEA**

European Patent Office, P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk
Tel: (+31-70) 340-2040,
Fax: (+31-70) 340-3016

**Date of mailing of the international search report**

30/10/2009

**Authorized officer**

Limacher, Rolf
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>US 7 003 463 B1 (MAES STEPHANE H [US] ET AL) 21 February 2006 (2006-02-21) column 3, line 1 - line 23 column 10, line 54 - column 11, line 33; figure 2 column 12, line 56 - line 64</td>
<td>1-25</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>US 2006122836 A1</td>
<td>08-06-2006</td>
<td>CN 1786957 A</td>
</tr>
</tbody>
</table>