December 2019

SUGGESTING TITLES FOR AUDIO RECORDINGS

Itay Inbar

Amit Pitaru

Isaac Blankensmith

Dror Ayalon

Tiago Camolesi

See next page for additional authors

Inventor(s)
Itay Inbar, Amit Pitaru, Isaac Blankensmith, Dror Ayalon, Tiago Camolesi, Guilherme Santos, James Lemieux, Sherry Lin, and Jason Cho

SUGGESTING TITLES FOR AUDIO RECORDINGS

ABSTRACT

Techniques of this disclosure may enable a computing device to suggest one or more titles based on the content of audio being recorded or audio that was previously recorded, and other data such as time and location. Rather than applying a general default title or audio file name, the computing device may request authorization from a user to analyze the contents of a recorded audio file and, after receiving explicit authorization from the user, analyze the audio, including speech, and automatically suggest titles that are indicative of the content of the audio and/or other data. The computing device may convert speech included in the audio into text and extract a plurality of terms from the text based on various factors, such as word classes (e.g., convert audio that includes "this meatball recipe adds parmesan cheese" into text and extract a plurality of nouns such as "meatball," "recipe," "parmesan," and "cheese" from the text). Based on various factors, such as term frequency in the text and the relative uniqueness of the terms in the spoken language, the computing device may identify a plurality of words from the plurality of terms to represent the overall content of the audio (e.g., identify "meatball" and "recipe" from "meatball," "recipe," "parmesan," and "cheese" based on term frequency in the text). The computing device may also classify non-speech audio (e.g. applause, dog barking, music) and use the classification, including metadata associated with the classified audio object, such as song titles, to identify a plurality of words to represent the overall content of the audio. The speech terms, non-speech audio classification, classified audio object metadata, and other data may be combined to identify a plurality of words to represent the overall content of the audio. The computing device may display the identified words as suggested words to be included in the title of the audio file. The user may select one or more of the identified words as the title or

combine one or more of the identified words with one or more other words entered by the user. The computing device may use the selected and/or entered words as the title for the audio and/or for the name of the audio file.

DESCRIPTION

FIG. 1 is a conceptual diagram illustrating an example computing device 102 that is configured to suggest titles that are indicative of the content of the audio, in accordance with one or more aspects of the present disclosure. Examples of computing device 102 include a mobile phone, a tablet computer, a laptop computer, a desktop computer, a server, a mainframe, a set-top box, a television, a wearable device (e.g., a computerized watch, a computerized eyewear, a computerized glove, etc.), a home automation device or system (e.g., an intelligent thermostat or security system), a voice-interface or countertop home assistant device, a personal digital assistant (PDA), a gaming system, a media player, an e-book reader, a mobile television platform, an automobile navigation or infotainment system, or any other type of mobile, non-mobile, wearable, and non-wearable computing device.

In the example of FIG. 1, computing device 102 may include user interface component (UIC) 104, UI module 106, audio input component 108, audio recording application module 110, and title suggestion module 122. While title suggestion module 122 is shown as included in audio recording application module 110, in various instances, title suggestion module 122 may be a standalone module. UI module 106, audio input component 108, audio recording application module 110 may perform operations described using hardware, or a combination of hardware and software and/or firmware residing in and/or executing at computing device 102. Computing device 102 may execute UI module 106, audio input component 108, audio recording application module 110 with multiple processors, or multiple devices. In some cases, computing

device 102 may execute UI module 106 as virtual machines executing on the underlying hardware. UI module 106 may also execute as one or more services of an operating system or computing platform, or as one or more executable programs at an application layer of a computing platform.

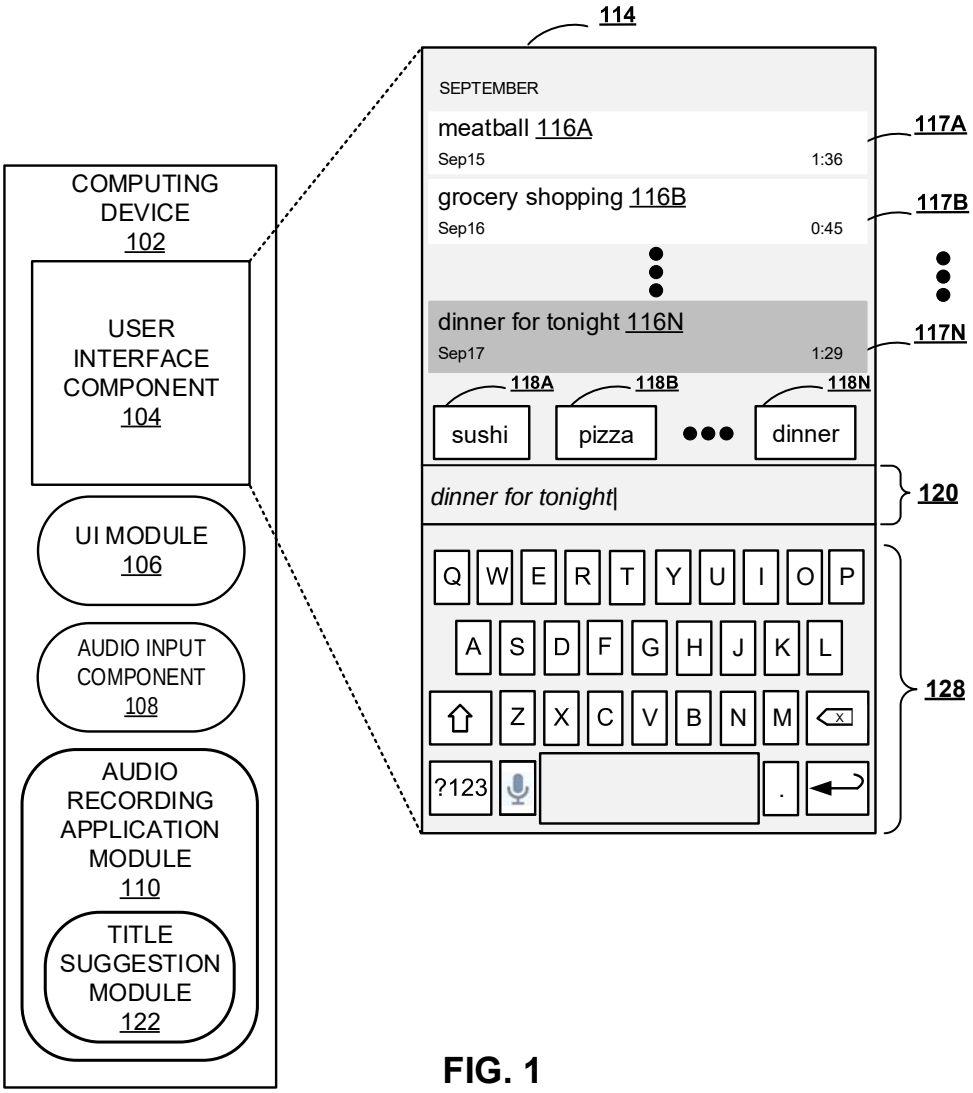

FIG. 1

UIC 104 of computing device 102 may function as an input and/or output device for computing device 102. UIC 104 may be implemented using various technologies. For instance, UIC 104 may function as an input device using presence-sensitive input screens, such as resistive touchscreens, surface acoustic wave touchscreens, capacitive touchscreens, projective

capacitance touchscreens, pressure-sensitive screens, acoustic pulse recognition touchscreens, radar, or another presence-sensitive display technology.

UIC 104 may function as input devices using a microphone or other transducer technologies, infrared sensor technologies, or other input device technologies for use in receiving user inputs. For example, UIC 104 may detect, using built-in microphone technology, audio inputs. As another example, UIC 104 may include a presence-sensitive display that may receive tactile inputs from a user of computing device 102. As such, UIC 104 may receive indications of tactile inputs by detecting one or more gestures from the user (e.g., the user touching or pointing to one or more locations of UIC 104 with a finger or a stylus pen).

UIC 104 may function as output (e.g., display) devices and present output to a user. For example, UIC 104 may function as an output device using any one or more display devices, such as liquid crystal displays (LCD), dot matrix displays, light-emitting diode (LED) displays, organic light-emitting diode (OLED) displays, e-ink, or similar monochrome or color displays capable of outputting visible information to a user of computing device 102. As another example, UIC 104 may function as an output device using speaker technologies, haptic feedback technologies, or other output device technologies for use in outputting information to the user. Furthermore, UIC 104 may present a user interface related to other features of computing platforms, operating systems, applications, and/or services executing at and/or accessible from computing device 102 (e.g., e-mail, chat, online services, telephone, gaming, etc.).

UI module 106 may manage user interactions with UIC 104 and other components of computing device 102. UI module 106 and UIC 104 may receive one or more indications of inputs (e.g., voice inputs, gesture inputs, etc.) from a user as the user interacts with UIC 104 at different times and when the user and computing device 102 are at different locations. UI

module 106 and UIC 104 may interpret inputs detected at UIC 104 and may relay information about the inputs detected at UIC 104 to audio recording application module 110, and/or one or more other associated platforms, operating systems, applications, and/or services executing at computing device 102 to, for example, cause computing device 102 to perform functions.

UI module 106 may cause UIC 104 to output, display, or otherwise present a user interface while a user of computing device 102 views output and/or provides input at UIC 104. For example, as shown in FIG. 1, UI module 106 may send instructions to UIC 104 that cause UIC 104 to display a graphical user interface (GUI) at a display screen of UIC 104. In other examples, UI module 106 may also cause UIC 104 to output a user interface in non-visual form, such as audio output. For example, if computing device 102 is an audio player device, UI module 106 may send instructions to UIC 104 that cause UIC 104 to output audio.

UI module 106 and UIC 104 may receive one or more indications of inputs (e.g., touch inputs, non-touch or presence-sensitive inputs, video inputs, audio inputs, etc.) from a user as the user interacts with the user interface output by UIC 104, at different times and when the user and computing device 102 are at different locations. UI module 106 and UIC 104 may interpret inputs detected at UIC 104 and may relay information about the inputs detected at UIC 104 to audio recording application module 110, and/or one or more other associated platforms, operating systems, applications, and/or services executing at computing device 102, for example, to cause computing device 102 to perform various functions.

Audio input component 108 may represent a microphone that records audio. Audio recording application module 110 may represent an application, service, or component executing at or accessible to audio input component 108 that records audio.

Title suggestion module 122 may represent an application, service, or component executing at or accessible to computing device 102 that suggest titles that are indicative of the content of audio currently being recorded or audio that was previously recorded. In some examples, title suggestion module 122 may be a sub-component of an operating system controlling the operation of computing device 102. For example, title suggestion module 122 may be integrated into audio recording application module 110 executing at computing device 102. In other examples, title suggestion module 122 may be a stand-alone application or subroutine that is invoked by an application or operating platform of computing device 102 any time an application or operating platform requires a title of an audio file. In some examples, computing device 102 may download and install title suggestion module 122 from an application repository of a service provider (e.g., via the Internet). In other examples, title suggestion module 122 may be pre-loaded as part of the operating system of computing device 102.

In accordance with the techniques of this disclosure, computing device 102 may record audio using audio input component 108 and automatically suggest titles for the recorded audio based on the contents of the audio recording. For example, a user may initiate the execution of audio recording application module 110, which causes UIC 104 to output a graphical user interface for audio recording application module 110. In some instances, prior to beginning to record audio, audio recording application module 110 requests approval from a user to analyze audio recorded and stored by audio recording application module 110. Audio recording application module 110 may request authorization upon initial execution of audio recording application module 110, periodically request reauthorization or request authorization each time audio recording application module 110 is executed. In various instances, rather than or in addition to requesting authorization to analyze audio recordings when audio recording

application module 110 is executed, audio recording application module 110 may request authorization in response to a user initiating a recording, terminating a recording, or any other time prior to performing any analysis of the content of the audio recording. That is, computing device 102 and audio recording application module 110 are configured to only analyze the content of audio recordings if a user of computing device 102 explicitly authorizes computing device 102 and audio recording application module 110 to analyze the content of the audio recordings. Absent the user's explicit authorization, computing device 102, and audio recording application module 110 will not analyze the content of the audio recordings.

After audio recording application module 110 begins recording audio, title suggestion module 122 may begin converting the audio into text using speech-to-text conversion techniques. For example, the computing device 102 may record audio that includes "this meatball recipe adds parmesan cheese." Using speech-to-text conversion techniques, title suggestion module 122 may generate the text "this meatball recipe adds parmesan cheese" from the audio recording. In various instances, title suggestion module 122 may generate a text version of speech included in the audio recording as the audio is being recorded or after the audio recording has been completed. In examples where the text is generated while the audio is being recorded, title suggestion module 122 may continue generating additional text until the audio recording is stopped. Title suggestion module 110 may continually update a list of suggested titles as the additional text is being generated or generate the list of suggested titles after computing device 102 has ceased recording the audio.

In the example of FIG. 1, audio recording application module 110 causes UIC 104 to output graphical user interface (GUI) 114 which includes audio file icons 117A–117N (collectively "audio file icons 117"), title suggestion regions 118A–118N (collectively "title

suggestion regions 118"), title edit region 120, and graphical keyboard 128. Each file of audio file icons 117-117N may have a respective file title region 116A–116N (collectively "file title regions 116") that includes the title of the file. The title of the file shown in file title regions 116 may be a default title, an automatically generated suggested title, a user-provided title, or a combination thereof. Title suggestion regions 118 may be generated by title suggestion module 122 in accordance with the techniques of this disclosure.

After generating at least a portion of the text from the recorded audio, title suggestion module 122 may, based on various factors, such as grammar and word classes, extract a plurality of unigrams and/or a plurality of bigrams from the text. For example, title suggestion module 122 may extract a plurality of unigrams "meatball," "recipe," "parmesan," and "cheese" and a plurality of bigrams "meatball recipe" and "parmesan cheese" from nouns included in the text.

Title suggestion module 122 may then identify a plurality of suggested titles reflecting the text from the plurality of unigrams and/or the plurality of bigrams based on various factors, such as term frequency in the text. For example, title suggestion module 122 may identify "meatball" and "recipe" as the terms with the highest frequency for non-generic terms in the text and may identify "meatball" and "recipe" as suggested titles reflecting the overall content of the audio.

In some examples, title suggestion module 122 may select a suggested title from the plurality of suggested titles and include the selected title within the file title region. For example, title suggestion module 122 may select "meatball" as the title for audio file icon 117A and may include "meatball" within file title region 116A as the title of audio file icon 117A.

In some examples, title suggestion module 122 may output one or more suggested titles of the plurality of suggested titles to user interface, and a user may select or more words to use as

the title for the audio. For example, title suggestion module 122 may output suggested titles "sushi," "pizza," and "dinner" as suggested titles for audio file icon 117N and may provide suggested titles to UI module 106 for display by UIC 104 within GUI 114. UIC 104 may include "sushi," "pizza," and "dinner" within title suggestion regions 118A -118N.

UIC 104 may detect a user input selecting one or more suggested titles and, in response to receiving the user input, including the one or more selected suggested titles as text within title edit region 120. For example, the user may select the suggested title "dinner" at title suggestion region 118N, and UCI 104 may present the selected title "dinner" within title edit region 128. The user may further edit the text within title edit region 120 by providing input to graphical keyboard 128. For example, after the user selects the suggested title "dinner," the user may further edit the title by providing user input "for tonight" to graphical keyboard 128. When the user is satisfied with the title, the user may store the desired title "dinner for tonight" within file title region 116N for audio file icon 117N.

By automatically generating a suggested title based on the content of audio, title suggestion module 122 may improve a title generation process for the audio content of audio recording and may summarize the content of the audio, thereby providing a more satisfying experience to an end-user. The automatically generated title is more likely to be descriptive of the contents of the audio, which may enable the user to be able to more quickly identify the desired audio recording without having to listen to all or parts of several of the audio recordings until the user finds the desired audio recording. As such, the techniques of this disclosure may save additional battery power by reducing the amount of time a screen may need to be powered on to find the desired audio recording and may save network bandwidth in instances where the audio recordings are not stored locally on the computing device.

FIG. 2 is a block diagram illustrating an example computing device that is configured to suggest titles that are indicative of the content of the audio, in accordance with one or more aspects of the present disclosure. FIG. 2 illustrates only one particular example of computing device 202, and many other examples of computing device 202 may be used in other instances and may include a subset of the components not shown in FIG. 2.

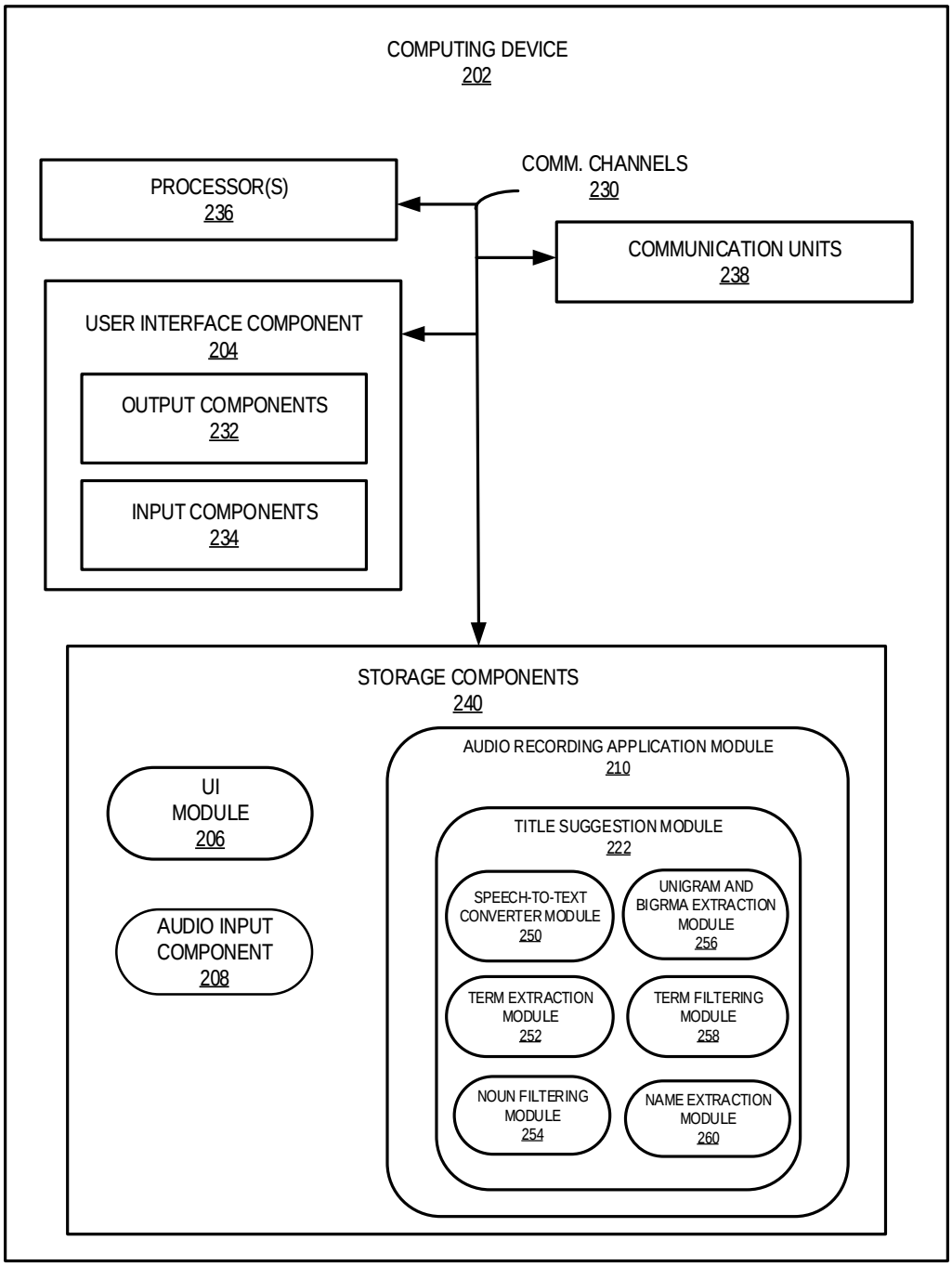

FIG. 2

As shown in the example of FIG. 2, computing device 202 includes user interface component (UIC) 204, one or more processors 236, one or more communication units 238, and one or more storage components 240. UIC 204 includes output component 232 and input component 234. Storage components 240 of computing device 202 includes UI module 206,

audio input component 208, and audio recording application module 210. Audio recording application module 210 includes title suggestion module 222, which includes speech-to-text converter module 250, term extraction module 252, noun filtering module 254, unigram and bigram extraction module 256, term filtering module 258, and name extraction module 260.

Communication channels 230 may interconnect each of processors 236, user interface component 204, communication unit 238, and storage components 240 for inter-component communications (physically, communicatively, and/or operatively). In some examples, communication channels 230 may include a system bus, a network connection, an inter-process communication data structure, or any other method for communicating data.

One or more communication units 238 of computing device 202 may communicate with external devices (e.g., an application provider server) via one or more wired and/or wireless networks by transmitting and/or receiving network signals on the one or more networks. Examples of communication units 238 include a network interface card (e.g., such as an Ethernet card), an optical transceiver, a radio frequency transceiver, a Global Navigation Satellite System (GNSS) (e.g., global positioning satellite (GPS)) receiver, or any other type of device that can send and/or receive information. Other examples of communication units 238 may include short wave radios, cellular data radios, wireless network radios, as well as universal serial bus (USB) controllers.

One or more input components 234 of computing device 202 may receive input. Examples of input are tactile, audio, and video input. Input components 234 of computing device 202, in one example, includes a presence-sensitive input device (e.g., a touch-sensitive screen, a presence-sensitive display), mouse, keyboard, voice responsive system, video camera, microphone or any other type of device for detecting input from a human or machine. In some

examples, input components 234 may include one or more sensor components such as one or more location sensors (GNSS components, Wi-Fi components, cellular components), one or more temperature sensors, one or more movement sensors (e.g., accelerometers, gyros, radar), one or more pressure sensors (e.g., barometer), one or more ambient light sensors, and one or more other sensors (e.g., microphone, camera, infrared proximity sensor, hygrometer, and the like). Other sensors may include a heart rate sensor, magnetometer, glucose sensor, hygrometer sensor, olfactory sensor, compass sensor, step counter sensor, to name a few other non-limiting examples.

One or more output components 232 of computing device 202 may generate output. Examples of output are tactile, audio, and video output. Output components 232 of computing device 202, in one example, includes a PSD, sound card, video graphics adapter card, speaker, cathode ray tube (CRT) monitor, liquid crystal display (LCD), organic light-emitting diode display (OLED), or any other type of device for generating output to a human or machine.

UIC 204 may include output component 232 and input component 234. Output component 232 may be a display component, such as a screen at which information is displayed by UIC 204, and input component 234 may be a presence-sensitive input component that detects an object at and/or near output component 232. Output component 232 and input component 234 may be a speaker and microphone pair or any other combination of one or more input and output components, such as input components 234 and output components 232. In the example of FIG. 2, UIC 204 may present a user interface.

While illustrated as an internal component of computing device 202, UIC 204 may also represent an external component that shares a data path with computing device 202 for transmitting and/or receiving input and output. For instance, in one example, UIC 204 represents

a built-in component of computing device 202 located within and physically connected to the external packaging of computing device 202 (e.g., a screen on a mobile phone). In another example, UIC 204 represents an external component of computing device 202 located outside and physically separated from the packaging or housing of computing device 202 (e.g., a monitor, a projector, etc. that shares a wired and/or wireless data path with computing device 202).

One or more storage components 240 within computing device 202 may store information for processing during operation of computing device 202 (e.g., computing device 102 may store data accessed by UI module 206, audio input component 208, and audio recording application module 210 during execution at computing device 202). In some examples, storage component 240 is a temporary memory, meaning that a primary purpose of storage component 28 is not long-term storage. Storage components 240 on computing device 202 may be configured for short-term storage of information as volatile memory and, therefore, not retain stored contents if powered off. Examples of volatile memories include random access memories (RAM), dynamic random-access memories (DRAM), static random-access memories (SRAM), and other forms of volatile memories known in the art.

Storage components 240, in some examples, also include one or more computer-readable storage media. Storage components 240 in some examples include one or more non-transitory computer-readable storage mediums. Storage components 240 may be configured to store larger amounts of information than typically stored by volatile memory. Storage components 240 may further be configured for long-term storage of information as non-volatile memory space and retain information after power on/off cycles. Examples of non-volatile memories include magnetic hard discs, optical discs, floppy discs, flash memories, or forms of electrically

programmable memories (EPROM) or electrically erasable and programmable (EEPROM) memories. Storage components 240 may store program instructions and/or information (e.g., data) associated with UI module 206, audio input component 208, and audio recording application module 210. Storage components 240 may include a memory configured to store data or other information associated with UI module 206, audio input component 208, and audio recording application module 210.

One or more processors 236 may implement functionality and/or execute instructions associated with computing device 202. Examples of processors 236 include application processors, display controllers, auxiliary processors, one or more sensor hubs, and any other hardware configured to function as a processor, a processing unit, or a processing device. UI module 206, audio input component 208, and audio recording application module 210 may be operable by processors 236 to perform various actions, operations, or functions of computing device 202. For example, processors 36 of computing device 202 may retrieve and execute instructions stored by storage components 240 that cause processors 236 to perform the operations of UI module 206, audio input component 208, and audio recording application module 210. The instructions, when executed by processors 236, may cause computing device 202 to store information within storage components 240.

UI module 206 may manage user interactions with UIC 204 and other components of computing device 202. UI module 206 may cause UIC 204 to output a user interface as a user of computing device 202 views output and/or provides input at UIC 204. For example, the user may provide a user input to select a suggested title. In response to the user input, UIC 204 may output the selected title on a user interface.

Audio input component 208 may record audio based on receiving explicit authorization from the user. For example, audio input component 208 may record audio that includes "Sushi or Pizza for dinner" after receiving explicit authorization from the user.

Audio recording application module 210 may represent an application, service, or component executing at or accessible to audio input component 208 that records audio. Audio recording application module 210 may be an application that is already installed at computing device 202, and may be executed by processors 236 to perform various functions. For example, audio recording application module 210 may initiate an audio recording, terminate the audio recording, and perform analysis of the content of the audio recording.

Title suggestion module 222 may suggest titles that are indicative of the content of audio, in accordance with one or more aspects of the present disclosure. Title suggestion module 222 may include various submodules, such as speech-to-text converter module 250, term extraction module 252, noun filtering module 254, unigram and bigram extraction module 256, term filtering module 258, and name extraction module 260.

Speech-to-text converter module 250 may convert audio into text. For example, for audio contains speech such as "Sushi or Pizza for dinner," speech-to-text converter module 250 may generate the text "Sushi or Pizza for dinner" from the audio. In some examples, speech-to-text converter module 228 may further include a language translator, which may convert the text from a first language into one or more other languages. For example, speech-to-text converter module 228 may convert the text from English to Spanish.

Term extraction module 252 may extract a plurality of terms from the text. Title suggestion module 232 may first divide the text into a sequence of terms based on prefixes, suffixes, punctuation marks, etc. For example, term extraction module 252 may divide text

"Sushi or Pizza for dinner" into terms "Sushi," "or," "Pizza," "for," and "dinner." Term extraction module 252 may then normalize the terms by turning the terms to lower case. For example, term extraction module 232 may turn terms "Sushi," "or," "Pizza," for," and "dinner" to lower case terms "sushi," "or," "pizza," for," and "dinner."

Noun filtering module 254 may identify a plurality of nouns included in the normalized terms. Noun filtering module 254 may first iterate over all the terms of the normalized terms to identify a plurality of common nouns and a plurality of proper nouns included in the plurality of terms. A proper noun is a specific name for a particular person, thing, or place, such as "Walt Disney," "Empire State Building," or "United States." A common noun is a generic name for a person, thing, or place, such as "men," "building," or "country." For terms "sushi," "or," "pizza," for," and "dinner," noun filtering module 254 may extract common nouns "sushi," "pizza," and "dinner." Noun filtering module 254 may then transform each identified noun into a uniform form, such as a singular form or a plural form. For example, noun filtering module 254 may transfer both terms "apple" and "apples" into a uniform form, such as a singular form "apple." By transforming each identified noun into a uniform form, title suggestion module 254 may be able to count each noun in both singular and plural forms in the text.

Unigram and bigram extraction module 256 may extract a plurality of unigrams and bigrams from the identified nouns and may assign a respective score to each unigram and bigram. Unigrams are one-word phrases, such as "apple," "pear," "peach," etc. Bigrams are two-word phrases, such as "ice cream," "apple pie," "chocolate cake," etc. For example, unigram and bigram extraction module 256 may extract unigrams "sushi," "pizza," and "dinner" from the identified nouns "sushi," "pizza," and "dinner." After extracting a plurality of unigrams and bigrams, unigram and bigram extraction module 256 may assign a respective score to each

unigram and bigram. The respective scores may be assigned based on term frequency, inverse document frequency, and specificity.

In one of the examples, the respective scores may be assigned based in part on term frequency (e.g., the frequency of the term in the text). For example, a unigram with higher term frequency may be assigned a higher score than a unigram with a lower term frequency. In one of the examples, the respective scores may be assigned based in part on inverse document frequency (e.g., the frequency of the term in other training documents, such as conversation data). For example, for unigrams that commonly appear in training documents, such as the term "pizza," term frequency will tend to incorrectly emphasize the commonly used terms without giving enough weight to the less frequently used terms, such as the term "sushi." To correctly evaluate each term, a commonly used term may be assigned a lower inverse document frequency than less frequently used terms. Unigram and bigram extraction module 256 may apply inverse document frequency as a weighting factor to term frequency. In one of the examples, the respective scores may be assigned based in part on specificity (e.g., the frequency of other terms in other training documents with a given term). Similar to inverse document frequency, unigram and bigram extraction module 256 may apply specificity as a weighting factor to term frequency.

Term filtering module 258 may filter the scored unigrams and bigrams to remove stop words and swear words. In one of the examples, term filtering module 258 may filter the scored unigrams and bigrams by removing unigrams that are included in the scored bigrams from the scored unigrams. For example, term filtering module 258 may remove unigram "computer," which is part of bigram "portable computer" from the scored unigrams.

In one of the examples, term filtering module 258 may remove stop words from the scored unigrams and bigrams based on stop words data. Stop words are commonly used words

that can be ignored. Term filtering module 258 may identify stop words based on stop words data and remove stop words from the scored unigrams and bigrams. In one of the examples, term filtering module 258 may remove swear words from the scored unigrams and bigrams based on swear words data. Swear words are offensive words. Similar to stop words, term filtering module 258 may identify swear words based on swear words data and remove swear words from the scored unigrams and bigrams.

Name extraction module 260 may capitalize names in the filtered unigrams and bigrams. Name extraction module 260 may first determine the number of capitalized appearances in the text for each unigram and bigram from the filtered unigrams and bigrams. Name extraction module 260 may then determine the number of capitalized appearances in the middle of a sentence for each unigram and bigram from the filtered unigrams and bigrams. If a term appears all the time capitalized in the text, name extraction module 236 may identify the term as a name and may capitalize the term.

Title suggestion module 222 may then suggest titles based on the respective score of each unigram and bigram from the name capitalized unigrams and bigrams. Title suggestion module 222 may first add a pre-determined boost to the respective score of each name from the name capitalized unigrams and bigrams. Title suggestion module 222 may then rank unigrams and bigrams from the name capitalized unigrams and bigrams based on the respective score and may suggest titles for the audio based on the rank.

In some examples, title suggestion module 222 may select the highest-ranked unigram or bigram as the default title for the audio and may include the selected title within the file title region. For example, title suggestion module 222 may select “grocery shopping” as the highest-

ranked bigram and include “grocery shopping” within file title region 116B as the title of audio file icon 117B.

In some examples, title suggestion module 222 may select one or more unigram and bigram based on the rank and output the selected unigram and bigram as suggested titles to UI module 206 and UI module 206 may instruct UIC 204 to present the selected unigram and bigram. For example, title suggestion module 222 may select unigrams “sushi,” “pizza,” and “dinner” as the three highest-ranked unigrams and provide the three unigrams to UI module 206. UI module 206 may instruct UIC 204 to present the selected terms “sushi,” “pizza,” and “dinner,” as suggested titles. A user may select one or more suggested titles and may further edit the selected suggested titles by providing input to UIC 204.

FIG. 3 is a flowchart illustrating example operations of a computing device that is configured to suggest titles that are indicative of the content of the audio, in accordance with one or more aspects of the present disclosure. The operations of FIG. 3 may be performed by one or more processors of a computing device, such as computing devices 102 and 202 in FIG.1 and FIG. 2. For purposes of illustration only, FIG. 3 are described below within the context of computing device 102 in FIG. 1.

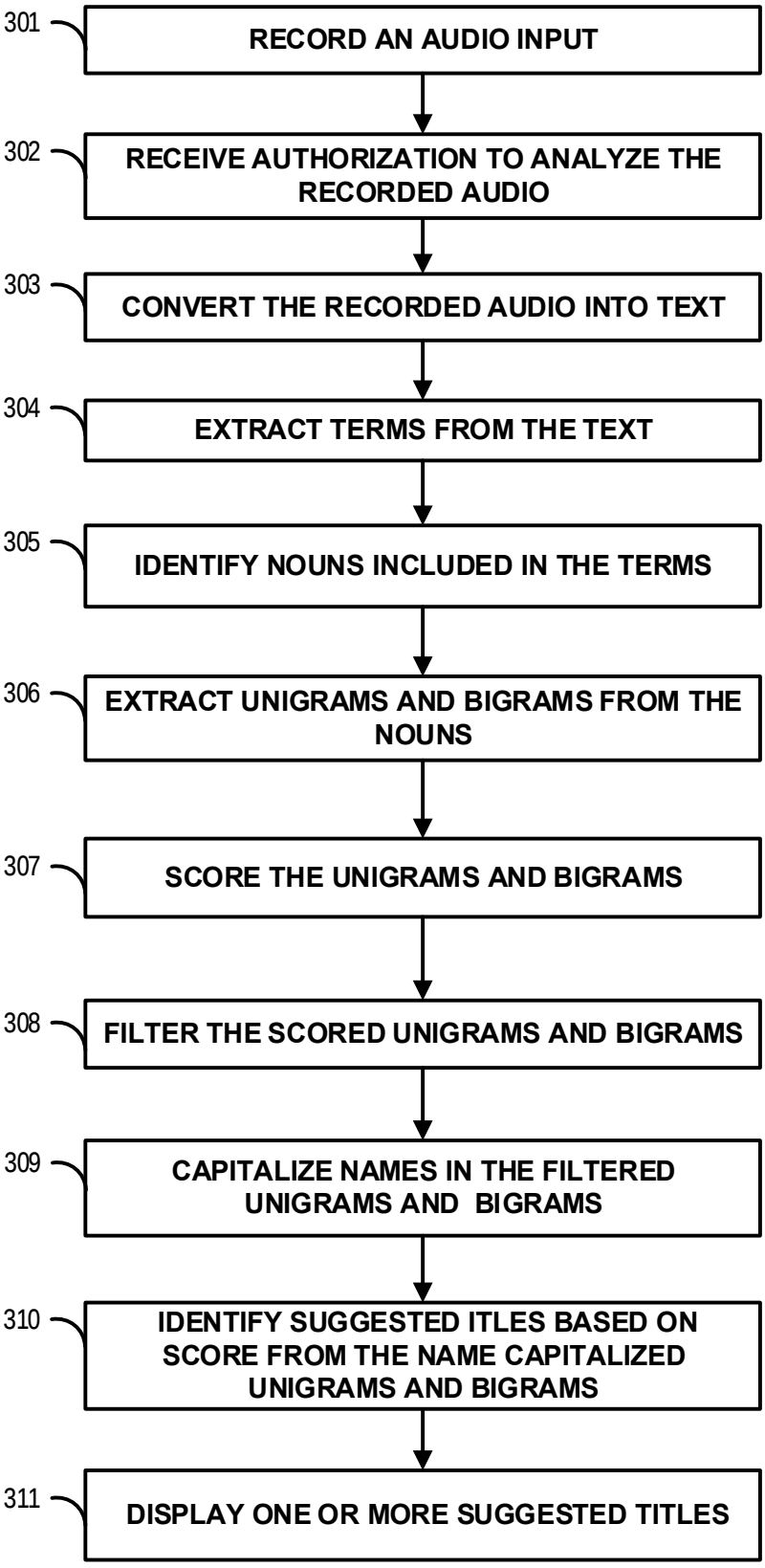
301
RECORD AN AUDIO INPUT
302
RECEIVE AUTHORIZATION TO ANALYZE THE RECORDED AUDIO
303
CONVERT THE RECORDED AUDIO INTO TEXT
304
EXTRACT TERMS FROM THE TEXT
305
IDENTIFY NOUNS INCLUDED IN THE TERMS
306
EXTRACT UNIGRAMS AND BIGRAMS FROM THE NOUNS
307
SCORE THE UNIGRAMS AND BIGRAMS
308
FILTER THE SCORED UNIGRAMS AND BIGRAMS
309
CAPITALIZE NAMES IN THE FILTERED UNIGRAMS AND BIGRAMS
310
IDENTIFY SUGGESTED ITLES BASED ON SCORE FROM THE NAME CAPITALIZED UNIGRAMS AND BIGRAMS
311
DISPLAY ONE OR MORE SUGGESTED TITLES

FIG. 3

In operation, computing device 102 may record an audio input (301). Upon receiving authorization to analyze the recorded audio (302), computing device 102 may convert the recorded audio into text (303). For example, the computing device 102 may record audio that includes "this meatball recipe adds parmesan cheese," and may generate the text "this meatball recipe adds parmesan cheese" from the audio recording.

Computing device 102 may extract a plurality of terms from the text (304) based on prefixes, suffixes, punctuation marks, and normalize the terms by turning the terms to lower case. For example, computing device 102 may extract terms "this," "meatball," "recipe," "adds parmesan," and "cheese" from the text. Computing device 102 may then identify nouns included in extracted terms (305). For example, computing device 102 may identify nouns "meatball," "recipe," "parmesan," and "cheese" from the extracted terms.

Computing device 102 may then extract unigrams and bigrams from the identified nouns (306) and may score the extracted unigrams and bigrams (307) based on various factors such as term frequency in the text.

Computing device 102 may further filter the scored unigrams and bigrams (308) by removing unigrams that are included in the scored bigrams, removing stop words based on stop words data, and removing swear words based on swear word data. Computing device 102 may also capitalize names in the filtered unigrams and bigrams (309).

Computing device 102 may then identify suggested titles that are indicative of the content of the audio based on the respective score of each unigram and bigram (310). For example, computing device 102 may rank unigrams and bigrams based on the respective score and may suggest titles based on the rank. After identifying suggested titles, computing device 102 may output, for display, one or more suggested titles (311).

In one or more examples, the functions described may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the functions may be stored on or transmitted over, as one or more instructions or code, a computer-readable medium and executed by a hardware-based processing unit. Computer-readable media may include computer-readable storage media, which correspond to tangible media such as data storage media, or communication media including any medium that facilitates transfer of a computer program from one place to another, e.g., according to a communication protocol. In this manner, computer-readable media generally may correspond to (1) tangible computer-readable storage media, which are non-transitory or (2) communication media such as signals or carrier waves. Data storage media may be any available media that can be accessed by one or more computers or one or more processors to retrieve instructions, code and/or data structures for implementation of the techniques described in this disclosure. A computer program product may include a computer-readable medium.

By way of example, and not limitation, such computer-readable storage media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage, or other magnetic storage devices, flash memory, or any other medium that can be used to store desired program code in the form of instructions or data structures and that can be accessed by a computer. Also, any connection is properly termed a computer-readable medium. For example, if instructions are transmitted from a website, server, or other remote source using a coaxial cable, fiber optic cable, twisted pair, digital subscriber line (DSL), or wireless technologies such as infrared, radio, and microwave, then the coaxial cable, fiber optic cable, twisted pair, DSL, or wireless technologies such as infrared, radio, and microwave are included in the definition of medium. It should be understood, however, that computer-readable storage media and data

storage media do not include connections, carrier waves, signals, or other transient media, but are instead directed to non-transient, tangible storage media. Disk and disc, as used, includes compact disc (CD), laser disc, optical disc, digital versatile disc (DVD), floppy disk and Blu-ray disc, where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media.

Instructions may be executed by one or more processors, such as one or more digital signal processors (DSPs), general purpose microprocessors, application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. Accordingly, the term "processor," as used, may refer to any of the foregoing structure or any other structure suitable for implementation of the techniques described. In addition, in some aspects, the functionality described may be provided within dedicated hardware and/or software modules. Also, the techniques could be fully implemented in one or more circuits or logic elements.

The techniques of this disclosure may be implemented in a wide variety of devices or apparatuses, including a wireless handset, an integrated circuit (IC) or a set of ICs (e.g., a chip set). Various components, modules, or units are described in this disclosure to emphasize functional aspects of devices configured to perform the disclosed techniques, but do not necessarily require realization by different hardware units. Rather, as described above, various units may be combined in a hardware unit or provided by a collection of interoperative hardware units, including one or more processors as described above, in conjunction with suitable software and/or firmware.