(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
05 November 2020 (05.11.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/222957 A1

(51) International Patent Classification:
G06F 16/332 (2019.01) *G06F 16/683* (2019.01)

(21) International Application Number:
PCT/US2020/025729

(22) International Filing Date:
30 March 2020 (30.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/397,004 29 April 2019 (29.04.2019) US

(71) Applicant: **ROVI GUIDES, INC.** [US/US]; 2160 Gold Street, San Jose, CA 95002 (US).

(72) Inventors: **AHER, Ankur**; 11, Giriaj Krupa C.H.S., Rambaug Lane No. 2, Kalyan, Maharashtra 421301 (IN). **CHONAT SRI, Sindhuja**; D 311, Ramani Srishti, Opp Kannan Dept. Stores, Trichy Road, Singanallur, Coimbatore, Tamil Nadu 641005 (IN). **PUNIYANI, Aman**; A 302, Sobha Garnet, Iblur Village, Bellandur, near Sarjapur Junction, Bangalore, Karnataka 560103 (IN). **MAHAJAN, Nishchit**; 46 New Tehsil Pura, Amritsar, Punjab 143001 (IN).

(74) Agent: **GUILIANO, Joseph M.** et al.; Haley Guiliano LLP, 75 Broad Street, Suite 1000, New York, NY 10004 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,

(54) Title: SYSTEMS AND METHODS FOR DISAMBIGUATING A VOICE SEARCH QUERY

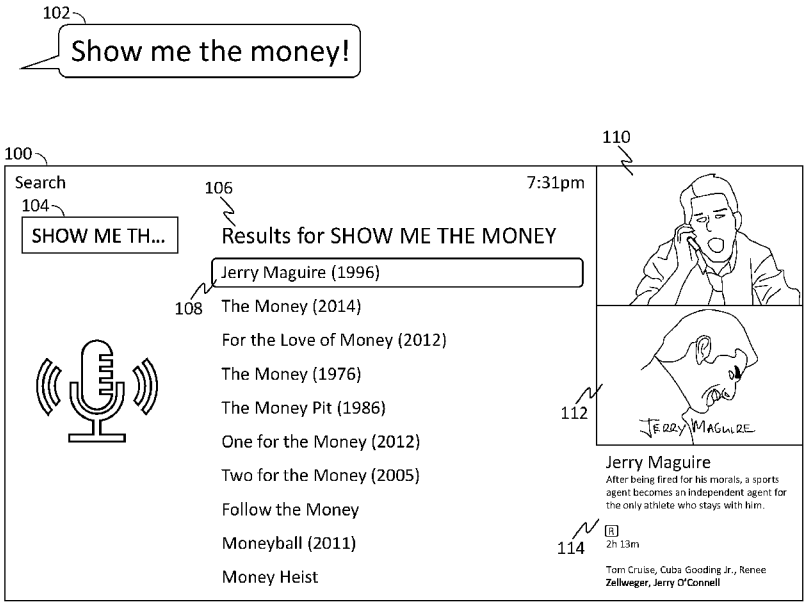

FIG. 1

(57) Abstract: Systems and methods are described herein for disambiguating a voice search query that contains a command keyword by determining whether the user spoke a quotation from a content item and whether the user mimicked or approximated the way the quotation is spoken in the content item. The voice search query is transcribed into a string, and an audio signature of the voice search query is identified. Metadata of a quotation matching the string is retrieved from a database that includes audio signature information for the string as spoken within the content item. The audio signature of the voice search query is compared with the audio signature information in the metadata to determine whether the audio signature matches the audio signature information in the quotation metadata. If a match is detected, then a search result comprising an identifier of the content item from which the quotation comes is generated.

WO 2020/222957 A1

[Continued on next page]

MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

SYSTEMS AND METHODS FOR DISAMBIGUATING A VOICE SEARCH QUERY

Background

[0001] The present disclosure relates to providing search results and, more particularly, disambiguation of a voice search query based on audio properties of the voice input.

Summary

[0002] Voice search applications on content platforms allow the user to search for content using voice commands. Using command keywords in conjunction with search parameters, the user can instruct the application to perform a search query for particular content items. Users can also use a famous quote from a particular content item as a search query for that content item. When quotes contain command keywords, however, the application may not recognize that the user is attempting to search for the particular content item from which the quote comes and instead performs a search using the words of the quote that follow the recognized command keyword.

[0003] Systems and methods are described herein for disambiguating a voice search query that contains a command keyword by determining whether the user spoke a quotation from a content item and whether the user mimicked or approximated the way the quotation is spoken in the content item. If so, a search result comprising an identifier of the content item is generated. The voice search query may also be processed based on the command keyword, which may return a number of search results. The search result representing the content item from which the quotation comes may be ranked highest among the search results returned and therefore presented first in a list of search results. If the user did not mimic or approximate the way the quotation is spoken in the content time, then a search result may not be generated for the content item or may be ranked lowest among other search results.

[0004] Upon receiving the voice search query, the system transcribes the voice search query into a string of text. An audio signature of the voice search query is also identified. A query is made to a database of quotations using the string. Metadata of a quotation matching the string is received in response to the query. The metadata includes audio signature information for the string as spoken within the content item and an identifier of the content item from which the quotation comes. The audio signature of the voice search query is compared with the audio signature information in the metadata of the content item, and the

system determines whether the audio signature of the voice search query matches the audio signature information in the quotation metadata. If a match is detected, then a search result comprising an identifier of the content item from which the quotation comes is generated. The system may establish upper and lower thresholds of similarity by transposing the audio signature information in a positive direction and negative direction, respectively, by a predetermined amount. The system determines that the audio signature matches the audio signature information if the audio signature falls between the upper and lower thresholds.

[0005] As part of determining whether the audio signature of the voice search query matches the audio signature information in the quotation metadata, a cadence of each word in the voice search query may be determined. The system compares the cadence of each word with cadence information in the metadata to determine whether the voice search query was spoken with identical or similar cadence to that with which the words are spoken in the content item. If so, the system determines that the audio signature matches the audio signature information in the quotation metadata. The system may establish upper and lower thresholds of similarity by modulating the cadence information in a positive direction and a negative direction, respectively, by a predetermined amount. The system determines that the cadence of each word matches the cadence information for the corresponding word if the cadence falls between the upper and lower thresholds.

[0006] The system may, alternatively or additionally, analyze emphasis placed on each word in the voice search query and compare the relative emphasis of each word with relative emphasis information in the quotation metadata. If the relative emphasis placed on each word of the voice search query is identical or similar to that with which the words are spoken in the content item, the system determines that the audio signature matches the audio signature information in the quotation metadata.

[0007] The system may also analyze the duration of each word in the voice search query and compare the duration of each word with duration information of each word in the quotation metadata. If the duration of each word is identical or similar to that of each word as spoken in the content item, the system determines that the audio signature matches the audio signature information in the quotation metadata. The system may establish upper and lower thresholds of similarity increasing and decreasing the duration information, respectively, by a predetermined amount. The system determines that the duration matches the duration information if the duration falls between the upper and lower thresholds.

[0008] The overall rhythm of the words spoken in the voice search query may also be analyzed and compared with rhythm information in the metadata of a content item. If the

rhythm is identical or similar to the rhythm information, the system determines that the audio signature matches the audio signature information in the quotation metadata. The system may establish upper and lower thresholds of similarity by modulating the rhythm information in a positive direction and a negative direction, respectively, by a predetermined amount. The system determines that the rhythm matches the rhythm information in the quotation metadata if the rhythm falls between the upper and lower thresholds.

Brief Description of the Drawings

[0009] The above and other objects and advantages of the disclosure will be apparent upon consideration of the following detailed description, taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

[0010] FIG. 1 shows an exemplary search interface, in accordance with some embodiments of the disclosure;

[0011] FIG. 2 shows another exemplary search interface, in accordance with some embodiments of the disclosure;

[0012] FIG. 3 shows exemplary voice search data and quotation metadata to be compared, in accordance with some embodiments of the disclosure;

[0013] FIG. 4 is a block diagram representing control circuitry, components, and data flow therebetween for disambiguating a voice search query, in accordance with some embodiments of the disclosure;

[0014] FIG. 5 is a flowchart representing a process for disambiguating a voice search query, in accordance with some embodiments of the disclosure;

[0015] FIG. 6 is a flowchart representing a process for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure;

[0016] FIG. 7 is a flowchart representing a second process for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure;

[0017] FIG. 8 is a flowchart representing a third process for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure;

[0018] FIG. 9 is a flowchart representing a fourth process for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure;

[0019] FIG. 10 is a flowchart representing a fifth process for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure;

[0020] FIG. 11 is a flowchart representing a process for determining whether a cadence of each word of a plurality of words matches cadence information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure;

[0021] FIG. 12 is a flowchart representing a process for determining whether a relative emphasis of each word of a plurality of words matches relative emphasis information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure;

[0022] FIG. 13 is a flowchart representing a process for determining whether a duration of each word of a plurality of words matches duration information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure; and

[0023] FIG. 14 is a flowchart representing a process for determining whether the overall rhythm of a plurality of words matches rhythm information for a second plurality of words, in accordance with some embodiments of the disclosure.

Detailed Description

[0024] Voice-based search applications are optimized for natural language input. Certain words or phrases are designated as command keywords, which inform the application of what function the user wants to activate. For example, in natural language, the user may say "Show me" or "I want to see," followed by the title of a particular content item or a description of the type of content the user wishes to search for. The application therefore designates "Show me" and "I want to see" as command keywords indicating that the application should perform a search on the words that immediately follow the command keyword. Thus, if the user says "Show me the money," the application will search for content entitled "the money." However, the user may say "Show me the money" in a way that mimics how actors Tom Cruise and Cuba Gooding Jr. say the phrase in the movie "Jerry Maguire," which indicates that the user wants to see that movie. The application compares an audio signature of the voice search with audio signature information of content items with

known quotations matching the words of the voice search. If the audio signature of the voice search is the same as or similar to the audio signature information of a content item, the application generates a search result for that particular content item. In some embodiments, the application may assign a confidence score to the content item and perform a regular content search based on the voice input as well, assigning a confidence score to each resulting content item. The application then generates search results for the content items having the highest confidence scores. If the audio signature of the voice search is different from the audio signature information or is extremely consistent throughout, the application may assign a lowest confidence score to the content item from which the quotation comes.

[0025] FIG. 1 shows an exemplary search interface 100, according to some embodiments of the disclosure. The search application receives the voice search query 102 comprising the words "Show me the money!" The application transcribes the voice search query 102 into a string of text 104 (displayed in a truncated fashion in FIG. 1). The application queries a database for content items matching text 104, and content items with known quotations matching or similar to text 104. The application receives, in response to the query, metadata of at least one quotation. Included in the metadata of each quotation is audio signature information for the quotation. The application compares an audio signature of the voice search query with the audio signature information in the metadata of each quotation. To account for variations in user vocal patterns, and accuracy of user memory and mimicry of the quotation, the application may establish upper and lower thresholds of similarity. The upper threshold is established by scaling, modulating, and/or transposing parameters of the audio signature information in a positive direction. For example, increasing the pitch, tone, or speed of the audio signature. Similarly, the lower threshold is established by scaling, modulating, and/or transposing the parameters of the audio signature in a negative direction. The application then determines whether each corresponding parameter of the audio signature of the voice search query is between the upper and lower thresholds. If so, the application determines that the voice search query comprises the quotation and generates, as the first result of a plurality of search results 106, a search result 108 comprising an identifier of the content item from which the quotation comes. For example, if the audio signature of the voice search query indicates that the user said "Show me the money!" in a way that is similar to how actors Tom Cruise and Cuba Gooding Jr. say the phrase in the movie "Jerry Maguire," the application generates a search result 108 for the movie "Jerry Maguire." The application also generates for display a still image 110 from the movie of a scene in which the quotation is said, as well as a thumbnail image 112 representing the movie and summary

information 114 describing the movie. Search results may be ordered based on rank, where higher ranks are associated with matches closer to the search string. In some embodiments, search result 108 may be ranked highest among all search results. If the audio signature of the voice search query indicates that the user did not say the quotation in a way that is similar to how the actors say the phrase, a search result for "Jerry Maguire" may not be generated, or may be ranked lower among the search results.

[0026] FIG. 2 shows another exemplary search interface presented on a mobile device 200, in accordance with some embodiments of the disclosure. Mobile device 200 receives voice search query 102 and displays the transcribed text of the voice search query in search box 202. A thumbnail image 204 of the movie "Jerry Maguire" is displayed as the first search result in response to voice search query 102.

[0027] FIG. 3 shows exemplary voice search data and quotation metadata to be compared, in accordance with some embodiments of the disclosure. Voice search data 300 represents the text 302 and audio signature parameters 304, 306, 308, 310, and 312 of the voice search query 102. Volume parameter 304 represents a scaled volume level at which the user spoke the voice search query. The volume level may be scaled based on an absolute decibel level or by comparing the volume level with an average volume level at which the user speaks based on a voice profile of the user. Tone parameter 306 indicates an emotion corresponding to the way in which the user spoke the voice query. The tone may be determined by analyzing the volume, speed, emphasis and other factors related to how the user spoke the voice search query.

[0028] Word emphasis parameter 308 indicates a relative emphasis of each word of the transcribed string 302. Continuing the above example, the string "show me the money" contains four words, and the emphasis placed on each word is represented by a corresponding entry 308a, 308b, 308c, and 308d in word emphasis parameter 308. Similarly, word duration parameter 310 indicates the duration of each word, for example in milliseconds, with each word of the transcribed string 302 having a corresponding entry 310a, 310b, 310c, and 310d in word duration parameter 310.

[0029] The application may compare the audio signature to a vocal profile of the user to determine whether the user has spoken the voice search query using an altered voice. For example, the user may attempt to mimic the sound of an Tom Cruise's voice when saying "Show me the money." The altered voice parameter 312 indicates whether the user has altered his or her voice when speaking the voice search query.

[0030] Quotation metadata 314 represents information about the quotation, including string information 316, volume parameter 318, tone parameter 320, word emphasis parameter 322 including entries 322a, 322b, 322c, and 322d, and word duration parameter 324 including entries 324a, 324b, 324c, and 324d. The application compares these parameters with the corresponding parameters of the voice search data 300 to determine whether the user said the quotation. Quotation metadata 314 also includes source identifier 326 identifying the content item from which the quotation comes.

[0031] FIG. 4 is a block diagram representing control circuitry, components, and data flow therebetween for disambiguating a voice search query, in accordance with some embodiments of the disclosure. Voice input 400 (*e.g.*, voice search query 102) is received using input circuitry 402. Input circuitry 402 may be a data interface such as a Bluetooth module, WiFi module, or other suitable data interface through which audio data captured by another device can be received. Alternatively, input circuitry 402 may be a microphone through which audio information is captured directly. Input circuitry 402 may convert the audio to a digital format such as WAV. Input circuitry 402 communicates voice input 400 to control circuitry 404. Control circuitry 404 may be based on one or more microprocessors, microcontrollers, digital signal processors, programmable logic devices, field-programmable gate arrays (FPGAs), application specific integrated circuits (ASICs), etc., and may include a multi-core processor (*e.g.*, dual-core, quad-core, hexa-core, or any suitable number of cores) or supercomputer. In some embodiments, processing circuitry may be distributed across multiple separate processors or processing units, for example, multiple of the same type of processing units (*e.g.*, two Intel Core i7 processors) or multiple different processors (*e.g.*, an Intel Core i5 processor and an Intel Core i7 processor). Input circuitry 402 transfers 406 voice input 400 to transcription circuitry 408 of control circuitry 404. Transcription circuitry 408 comprises speech-to-text circuitry and/or programming which transcribes voice input 400 into a string of text (*e.g.*, text 104, string 302). Input circuitry 402 also communicates 410 voice input 400 to comparison circuitry 412 of control circuitry 404. Comparison circuitry 412 compares the audio signature of voice input 400 with audio signature information in metadata of at least one content item.

[0032] Transcription circuitry 408 transfers 414 the string to transceiver circuitry 416. Transceiver circuitry 416 may be a network connection such as an Ethernet port, WiFi module, or any other data connection suitable for communicating with a remote server. Transceiver circuitry transmits 418 a query to quotation database 420 for quotations which match the string. The query may be an SQL “SELECT” command, or any other suitable

query format. Transceiver circuitry 416 receives 422, in response to the query, quotation metadata from quotation database 420. Transceiver circuitry 416 transfers 424 the quotation metadata to comparison circuitry 412. Comparison circuitry 412 compares the audio signature of the voice input 400 with audio signature information in the quotation metadata. Control circuitry 404 may establish upper and lower thresholds of similarity for the audio signature as described above. Comparison circuitry 412 may determine whether the audio signature of the voice input 400 falls between the upper and lower thresholds. If comparison circuitry 412 determines that the audio signature of voice input 400 matches audio signature information corresponding to a particular content item, comparison circuitry 412 transmits 426 a signal to output circuitry 428 to generate for display a content recommendation comprising an identifier of the particular content item. Output circuitry 428, which may be a GPU, VGA port, HDMI port, or any other suitable graphical output component, then generates for display 430 a search result comprising an identifier of the particular content item.

[0033] FIG. 5 is a flowchart representing an illustrative process 500 for disambiguating a voice search query, in accordance with some embodiments of the disclosure. Process 500 may be implemented on control circuitry 404. In addition, one or more actions of process 500 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0034] At 502, control circuitry 404 receives, from input circuitry 402, a voice search query. At 504, control circuitry 404, using transcription circuitry 408, transcribes the voice search query into a string comprising a plurality of words. Transcription circuitry 408 may use any suitable text-to-speech technique to transcribe the voice search query.

[0035] At 506, control circuitry 404 determines an audio signature of the voice search query. Control circuitry 404 analyzes the voice search query to identify audio parameters of the voice search query including tone, pitch, word emphasis, word duration, and volume. For example, control circuitry 404 may analyze the wave amplitude of the voice search query to determine the volume at which the user spoke the voice search query. Control circuitry 404 may analyze frequency data of the voice search query to determine the pitch and tone of the voice search query. Control circuitry 404 may isolate a portion of the voice search query corresponding to each of the plurality of words and determine the duration of each portion. Control circuitry 404 may compare each isolated portion of the voice search query to determine the relative level of emphasis placed on each word by the user.

[0036] At 508, control circuitry 404 queries the content database with the string. For example, control circuitry 404 may construct and transmit an SQL "SELECT" command to the content database to retrieve quotation metadata of all quotations matching the string, or significant portions thereof. At 510, control circuitry 404 receives, in response to the query, metadata of a plurality of quotations. The metadata includes as audio signature information of the quotation and an identifier of the content item from which the quotation comes.

[0037] At 512, control circuitry 404 initializes a counter variable N and sets its value to 0. At 514, control circuitry 404 determines whether the audio signature of the voice search query matches the audio signature information in the metadata of the Nth quotation. Methods for determining whether the audio signature is a match are described below in connection with FIGS. 6-10. If the audio signature of the voice search query matches the audio signature information in the metadata of the Nth quotation then, at 516, control circuitry 404, using output circuitry 428, generates for display a search result comprising an identifier of the content item from which the Nth quotation comes.

[0038] If the audio signature of the voice search query does not match the audio signature information in the metadata of the Nth quotation, or after generating a search result for the Nth quotation, then, at 518, control circuitry 404 determines whether there is metadata of additional quotations available. If so, then, at 520, control circuitry increments the value of counter variable N by one, and processing returns to step 514.

[0039] The actions and descriptions of FIG. 5 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 5 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0040] FIG. 6 is a flowchart representing an illustrative process 600 for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure. Process 600 may be implemented on control circuitry 404. In addition, one or more actions of process 600 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0041] A cadence of a word can be identified by tracking changes in pitch over time as the word spoken. At 602, control circuitry 404 determines a cadence of each word of the plurality of words. For each word, control circuitry 404 identifies a cadence of the word by analyzing changes in pitch during pronunciation of the word. Control circuitry 404 generates a data structure that tracks the changes in pitch over time for each word.

[0042] At 604, control circuitry 404 retrieves metadata of at least one quotation having a string that is similar to the words of the voice search query. The metadata also includes cadence information for each word of the string.

[0043] At 606, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 608, control circuitry determines whether the cadence of the Nth word matches the cadence information of the corresponding word in the metadata. Methods for determining whether the cadence matches the cadence information are described below in connection with FIG. 11. If the cadence of the Nth word matches the cadence information of the corresponding word in the metadata then, at 610, control circuitry increments the value of counter variable C by one. After incrementing the value of C, or if the cadence of the Nth word does not match the cadence information, at 612, control circuitry compares N with T to determine whether there are additional words for which cadence information is to be compared. If N is not equal to T, then, at 614, control circuitry 404 increments the value of N by one and processing returns to step 608. If N is equal to T, meaning that the cadence of each word of the voice search query has been compared with cadence information of the corresponding word in the metadata of a quotation, then, at 616, control circuitry determines whether the value of counter variable C, representing the number of words having matching cadences, is equal to T, meaning that the cadences of all words match the cadence information. If so, then, at 618, control circuitry 404 determines that the audio signature of the voice search matches the audio signature information in the metadata of the quotation.

[0044] The actions and descriptions of FIG. 6 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 6 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0045] FIG. 7 is a flowchart representing a second illustrative process 700 for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure. Process 700 may be implemented on control circuitry 404. In addition, one or more actions of process 700 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0046] At 702, control circuitry 404 determines an emphasis of each word of the plurality of words. For each word, control circuitry 404 analyzes the volume of the word relative to other words of the plurality of words. Control circuitry 404 generates a data structure that tracks the changes in volume over time for the plurality of words.

[0047] At 704, control circuitry 404 retrieves metadata of at least one quotation having a string that is similar to the words of the voice search query. The metadata also includes emphasis information for each word of the string.

[0048] At 706, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 708, control circuitry determines whether the emphasis of the Nth word matches the emphasis information of the corresponding word in the metadata. Methods for determining whether the emphasis matches the emphasis information are described below in connection with FIG. 12. If the emphasis of the Nth word matches the emphasis information of the corresponding word in the metadata then, at 710, control circuitry 404 increments the value of counter variable C by one. After incrementing the value of C, or if the emphasis of the Nth word does not match the emphasis information, at 712, control circuitry 404 compares N with T to determine whether there are additional words for which emphasis information is to be compared. If N is not equal to T then, at 714, control circuitry 404 increments the value of N by one and processing returns to step 708. If N is equal to T, meaning that the emphasis of each word of the voice search query has been compared with emphasis information of the corresponding word in the metadata of a quotation, then, at 716, control circuitry 404 determines whether the value of counter variable C, representing the number of words having matching emphases, is equal to T, meaning that the emphases of all words match the emphasis information. If so then, at 718, control circuitry 404 determines that the audio signature of the voice search matches the audio signature information in the metadata of the quotation.

[0049] The actions and descriptions of FIG. 7 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 7 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0050] FIG. 8 is a flowchart representing a third illustrative process 800 for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure. Process 800 may be implemented on control circuitry 404. In addition, one or more actions of process 800 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0051] At 802, control circuitry 404 determines a duration of each word of the plurality of word. For each word, control circuitry 404 identifies the amount of time taken for the word. Control circuitry 404 generates a data structure that tracks the duration of each word.

[0052] At 804, control circuitry 404 retrieves metadata of at least one content item having a string that is similar to the words of the voice search query. The metadata also includes duration information for each word of the string.

[0053] At 806, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 808, control circuitry determines whether the duration of the Nth word matches the duration information of the corresponding word in the metadata. Methods for determining whether the duration matches the duration information are described below in connection with FIG. 13. If the duration of the Nth word matches the duration information of the corresponding word in the metadata, then, at 810, control circuitry 404 increments the value of counter variable C by one. After incrementing the value of C, or if the duration of the Nth word does not match the duration information, at 812, control circuitry 404 compares N with T to determine whether there are additional words for which duration information is to be compared. If N is not equal to T, then, at 814, control circuitry 404 increments the value of N by one and processing returns to step 808. If N is equal to T, meaning that the duration of each word of the voice search query has been compared with duration information of the corresponding word in the metadata of a quotation, then, at 816, control circuitry 404 determines whether the value of counter variable C, representing the number of words having matching durations, is equal to T, meaning that the durations of all words match the duration information. If so, then, at 818, control circuitry 404 determines that the audio signature of the voice search matches the audio signature information in the metadata of the quotation.

[0054] The actions and descriptions of FIG. 8 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 8 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0055] FIG. 9 is a flowchart representing a fourth illustrative process 900 for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure. Process 900 may be implemented on control circuitry 404. In addition, one or more actions of process 900 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0056] At 902, control circuitry 404 determines an overall rhythm of the plurality of words. Control circuitry 404 analyzes word duration and word emphasis together to identify a rhythm of the plurality of words. Control circuitry 404 generates a data structure that tracks the rhythm of the plurality of words. At 904, control circuitry 404 retrieves metadata of at

least one quotation having a string that is similar to the words of the voice search query. The metadata also includes rhythm information for the string.

[0057] At 906, control circuitry determines whether the overall rhythm of the plurality of words matches the rhythm information. Methods for determining whether the rhythm matches the rhythm information are described below in connection with FIG. 14. If control circuitry 404 determines that the overall rhythm of the voice search query matches the rhythm information, then, at 908, control circuitry 404 determines that the audio signature of the voice search query matches the audio signature information in the metadata of the quotation.

[0058] The actions and descriptions of FIG. 9 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 9 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0059] FIG. 10 is a flowchart representing a fifth illustrative process 1000 for determining whether an audio signature matches audio signature information, in accordance with some embodiments of the disclosure. Process 1000 may be implemented on control circuitry 404. In addition, one or more actions of process 1000 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0060] At 1002, control circuitry 404 establishes a lower audio signature similarity threshold by negatively transposing (*i.e.*, decrease the value of each parameter) the audio signature information in the metadata of a quotation by a predetermined amount. For example, control circuitry may reduce the volume, word duration, emphasis, pitch, tone, and other parameters comprising the audio signature by 25%. At 1004, control circuitry 404 similarly establishes an upper audio signature similarity threshold by positively transposing (*i.e.*, increasing the value of each parameter) the audio signature information in the metadata of a quotation by the same predetermined amount.

[0061] At 1006, control circuitry 404 determines whether the audio signature of the voice search query is between the lower and upper thresholds. Control circuitry 404 may analyze the audio signature as a whole or may determine whether each individual parameter is between the corresponding parameter value of the lower and upper thresholds. If the audio signature of the voice search query is between the lower and upper thresholds, then, at 1008, control circuitry 404 determines that the audio signature matches the audio signature information in the metadata of the quotation.

[0062] The actions and descriptions of FIG. 10 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 10

may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0063] FIG. 11 is a flowchart representing an illustrative process 1100 for determining whether a cadence of each word of a plurality of words matches cadence information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure. Process 1100 may be implemented on control circuitry 404. In addition, one or more actions of process 1100 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0064] At 1102, control circuitry 404 establishes a lower cadence similarity threshold by negatively modulating the cadence information for each word in the metadata of a quotation by a predetermined amount. As described above, cadence information is represented by a change in pitch over time for each word. Control circuitry 404 modulates the change in pitch to a smaller change in pitch by a predetermined amount, such as 25%. Similarly, at 1104, control circuitry 404 establishes an upper cadence similarity threshold by positively modulating the cadence information for each word by the predetermined amount.

[0065] At 1106, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 1108, control circuitry determines whether the cadence of the Nth word is between the upper and lower cadence similarity thresholds of the corresponding word in the metadata. If the cadence of the Nth word is between the upper and lower thresholds of the corresponding word in the metadata then, at 1110, control circuitry 404 increments the value of counter variable C by one. After incrementing the value of C, or if the cadence of the Nth word is not between the upper and lower thresholds then, at 1112, control circuitry 404 compares N with T to determine whether there are additional words for which cadence information is to be compared. If N is not equal to T then, at 1114, control circuitry 404 increments the value of N by one and processing returns to step 1108. If N is equal to T, meaning that the cadence of each word of the voice search query has been compared with the upper and lower thresholds of the corresponding words in the metadata of a quotation, then, at 1116, control circuitry 404 determines whether the value of counter variable C, representing the number of words having cadences between their respective upper and lower thresholds, is equal to T, meaning that the cadences of all words are between their respective upper and lower thresholds. If so, then, at 1118, control circuitry 404 determines that the cadence of each word matches the cadence information for each corresponding word in the metadata of the quotation.

[0066] The actions and descriptions of FIG. 11 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 11 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0067] FIG. 12 is a flowchart representing an illustrative process 1200 for determining whether a relative emphasis of each word of a plurality of words matches relative emphasis information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure. Process 1200 may be implemented on control circuitry 404. In addition, one or more actions of process 1200 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0068] At 1202, control circuitry 404 determines a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words. As discussed above, control circuitry 404 analyzes a volume of each word of the voice search query and determines a relative volume level for each word, with louder words having greater emphasis. At 1204, control circuitry 404 determines a second plurality of relative emphasis levels between each word in the metadata of a quotation. Control circuitry 404 may retrieve emphasis information from the metadata or determine relative emphasis levels for each word based on other audio information such as volume information in the metadata of the quotation.

[0069] At 1206, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 1208, control circuitry determines whether the Nth relative emphasis level of the first plurality of emphasis levels is within a threshold amount of the Nth relative emphasis level of the second plurality of emphasis levels. For example, control circuitry 404 determines whether the change in relative emphasis level between two words in the voice search query is within 25% of the change in relative emphasis level between the same two words in the metadata of the quotation. If the Nth relative emphasis level of the first plurality of emphasis levels is within the threshold amount of the Nth relative emphasis level of the second plurality of emphasis levels, then, at 1210, control circuitry 404 increments the value of counter variable C by one. After incrementing the value of C, or if the cadence of the Nth word is not between the upper and lower thresholds then, at 1212, control circuitry 404 compares N with T to determine whether there are additional words for which cadence information is to be compared. If N is not equal to T, then, at 1214, control circuitry 404 increments the value of N by one and processing returns to step 1208. If N is

equal to T, meaning that the each relative emphasis level of the first plurality of emphasis levels is within the threshold amount of the corresponding emphasis levels of the second plurality of emphasis levels, then, at 1216, control circuitry 404 determines whether the value of counter variable C, representing the number of relative emphasis levels of the first plurality of emphasis levels within the threshold amount of the corresponding emphasis levels of the second plurality of emphasis levels, is equal to T. If so, then, at 1218, control circuitry 404 determines that the emphasis of each word matches the emphasis information for each corresponding word in the metadata of the quotation.

[0070] The actions and descriptions of FIG. 12 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 12 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0071] FIG. 13 is a flowchart representing an illustrative process 1300 for determining whether a duration of each word of a plurality of words matches duration information for each corresponding word of a second plurality of words, in accordance with some embodiments of the disclosure. Process 1300 may be implemented on control circuitry 404. In addition, one or more actions of process 1300 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0072] At 1302, control circuitry 404 establishes a lower duration similarity threshold by reducing the duration information for each word in the metadata of a quotation by a predetermined amount, such as 25%. Similarly, at 1304, control circuitry 404 establishes an upper duration similarity threshold by increasing the duration information for each word by the predetermined amount.

[0073] At 1306, control circuitry 404 initializes a counter variable N with a value of 0, a second counter variable C with a value of 0, and a variable T with a value equal to the number of words in the string. At 1308, control circuitry determines whether the duration of the Nth word is between the upper and lower duration similarity thresholds of the corresponding word in the metadata. If the duration of the Nth word is between the upper and lower thresholds of the corresponding word in the metadata then, at 1310, control circuitry 404 increments the value of counter variable C by one. After incrementing the value of C, or if the duration of the Nth word is not between the upper and lower thresholds, then, at 1312, control circuitry 404 compares N with T to determine whether there are additional words for which duration information is to be compared. If N is not equal to T, then, at 1114, control circuitry 404 increments the value of N by one and processing returns to step 1108. If

N is equal to T, meaning that the duration of each word of the voice search query has been compared with the upper and lower thresholds of the corresponding words in the metadata of a quotation, then, at 1316, control circuitry 404 determines whether the value of counter variable C, representing the number of words having durations between their respective upper and lower thresholds, is equal to T, meaning that the durations of all words are between their respective upper and lower thresholds. If so, then, at 1318, control circuitry 404 determines that the duration of each word matches the duration information for each corresponding word in the metadata of the quotation.

[0074] The actions and descriptions of FIG. 13 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 13 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0075] FIG. 14 is a flowchart representing an illustrative process 1400 for determining whether the overall rhythm of a plurality of words matches rhythm information for a second plurality of words, in accordance with some embodiments of the disclosure. Process 1400 may be implemented on control circuitry 404. In addition, one or more actions of process 1400 may be incorporated into or combined with one or more actions of any other process or embodiment described herein.

[0076] At 1402, control circuitry 404 establishes a lower rhythm similarity threshold by negatively modulating the rhythm information by a predetermined amount, such as 25%. For example, control circuitry 404 may reduce the duration and relative emphasis parameters of each word by the predetermined amount. Similarly, at 1404, control circuitry 404 establishes an upper rhythm similarity threshold by positively modulating the rhythm information by the predetermined amount.

[0077] At 1406, control circuitry 404 determines whether the overall rhythm of the voice search query is between the upper and lower rhythm similarity thresholds. Control circuitry 404 may compare the rhythm as a whole, or each individual rhythm parameter, with the upper and lower thresholds. If the overall rhythm is between the upper and lower thresholds, then, at 1408, control circuitry 404 determines that the overall rhythm of the plurality of words matches the rhythm information in the metadata of the quotation.

[0078] The actions and descriptions of FIG. 14 may be used with any other embodiment of this disclosure. In addition, the actions and descriptions described in relation to FIG. 14 may be done in suitable alternative orders or in parallel to further the purposes of this disclosure.

[0079] The processes described above are intended to be illustrative and not limiting. One skilled in the art would appreciate that the steps of the processes discussed herein may be omitted, modified, combined, and/or rearranged, and any additional steps may be performed without departing from the scope of the invention. More generally, the above disclosure is meant to be exemplary and not limiting. Only the claims that follow are meant to set bounds as to what the present invention includes. Furthermore, it should be noted that the features and limitations described in any one embodiment may be applied to any other embodiment herein, and flowcharts or examples relating to one embodiment may be combined with any other embodiment in a suitable manner, done in different orders, or done in parallel. In addition, the systems and methods described herein may be performed in real time. It should also be noted that the systems and/or methods described above may be applied to, or used in accordance with, other systems and/or methods.

This specification discloses embodiments including, but not limited to, the following items.

1. A method for disambiguating a voice search query, the method comprising:

receiving a voice search query;

transcribing the voice search query into a string comprising a plurality of words;

determining an audio signature of the voice search query;

querying a database with the string;

receiving, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

comparing the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determining, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generating for display a search result comprising an identifier of the quotation.

2. The method of item 1, further comprising:

determining a cadence of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

comparing a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

3. The method of item 1, further comprising:

determining an emphasis of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

comparing an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

4. The method of item 1, further comprising:

determining a duration of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

comparing a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

5. The method of item 1, further comprising:

determining an overall rhythm of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

comparing the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

6. The method of item 1, wherein determining whether the audio signature matches audio signature information on the metadata of a quotation comprises:

establishing a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

establishing an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

determining whether the audio signature is between the lower threshold and the upper threshold.

7. The method of item 2, wherein determining whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words comprises:

establishing a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

establishing an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

determining whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

8. The method of item 3, wherein determining whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words comprises:

determining a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

determining a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

determining, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

9. The method of item 4, wherein determining whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words comprises:

establishing, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

establishing, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

determining, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

10. The method of item 5, wherein determining whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words comprises:

establishing a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

establishing an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

determining whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

11. A system for disambiguating a voice search query, the system comprising:

input circuitry configured to receive a voice search query; and

control circuitry configured to:

transcribe the voice search query into a string comprising a plurality of words;

determine an audio signature of the voice search query;

query a database with the string;

receive, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

compare the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determine, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generate for display a search result comprising an identifier of the quotation.

12. The system of item 11, wherein the control circuitry is further configured to:

determine a cadence of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

compare a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

13. The system of item 11, wherein the control circuitry is further configured to:

determine an emphasis of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

compare an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

14. The system of item 11, wherein the control circuitry is further configured to:

determine a duration of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

compare a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the duration of each word of the plurality of

words matches the duration information for each corresponding word of the second plurality of words.

15. The system of item 11, wherein the control circuitry is further configured to:

determine an overall rhythm of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

compare the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

16. The system of item 11, wherein the control circuitry configured to determine whether the audio signature matches audio signature information on the metadata of a quotation is further configured to:

establish a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

establish an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

determine whether the audio signature is between the lower threshold and the upper threshold.

17. The system of item 12, wherein the control circuitry configured to determine whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words is further configured to:

establish a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

establish an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

determine whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

18. The system of item 13, wherein the control circuitry configured to determine whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words is further configured to:

determine a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

determine a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

determine, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

19. The system of item 14, wherein the control circuitry configured to determine whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words is further configured to:

establish, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

establish, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

determine, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

20. The system of item 15, wherein the control circuitry configured to determine whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words is further configured to:

establish a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

establish an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

determine whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

21. A system for disambiguating a voice search query, the system comprising:

means for receiving a voice search query;

means for transcribing the voice search query into a string comprising a plurality of words;

means for determining an audio signature of the voice search query;

means for querying a database with the string;

means for receiving, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

means for comparing the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

means for determining, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

means for, in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generating for display a search result comprising an identifier of the quotation.

22. The system of item 21, further comprising:

means for determining a cadence of each word of the plurality of words;

means for retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

means for comparing a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein the means for determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises means for determining, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

23. The system of item 21, further comprising:

means for determining an emphasis of each word of the plurality of words;

means for retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

means for comparing an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein the means for determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises means for determining, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

24. The system of item 21, further comprising:

means for determining a duration of each word of the plurality of words;

means for retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

means for comparing a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein the means for determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises means for determining, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

25. The system of item 21, further comprising:

means for determining an overall rhythm of the plurality of words;

means for retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

means for comparing the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein the means for determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises means for determining, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

26. The system of item 21, wherein the means for determining whether the audio signature matches audio signature information on the metadata of a quotation comprises:

means for establishing a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

means for establishing an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

means for determining whether the audio signature is between the lower threshold and the upper threshold.

27. The system of item 22, wherein the means for determining whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words comprises:

means for establishing a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

means for establishing an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

means for determining whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

28. The system of item 23, wherein the means for determining whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words comprises:

means for determining a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

means for determining a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

means for determining, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

29. The system of item 24, wherein the means for determining whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words comprises:

means for establishing, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

means for establishing, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

means for determining, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

30. The system of item 25, wherein the means for determining whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words comprises:

means for establishing a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

means for establishing an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

means for determining whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

31. A non-transitory computer-readable medium having non-transitory computer-readable instructions encoded thereon for disambiguating a voice search query that, when executed by control circuitry, cause the control circuitry to:

receive a voice search query;

transcribe the voice search query into a string comprising a plurality of words;

determine an audio signature of the voice search query;

query a database with the string;

receive, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

compare the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determine, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generate for display a search result comprising an identifier of the quotation.

32. The non-transitory computer-readable medium of item 31, wherein execution of the instructions further causes the control circuitry to:

determine a cadence of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

compare a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein execution of the instruction to determine whether the audio signature matches the audio signature information in the metadata of a quotation causes the control circuitry to determine, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

33. The non-transitory computer-readable medium of item 31, wherein execution of the instructions further causes the control circuitry to:

determine an emphasis of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

compare an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein execution of the instruction to determine whether the audio signature matches the audio signature information in the metadata of a quotation causes the control circuitry to determine, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

34. The non-transitory computer-readable medium of item 31, wherein execution of the instructions further causes the control circuitry to:

determine a duration of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

compare a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein execution of the instruction to determine whether the audio signature matches the audio signature information in the metadata of a quotation causes the control circuitry to determine, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

35. The non-transitory computer-readable medium of item 31, wherein execution of the instructions further causes the control circuitry to:

determine an overall rhythm of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

compare the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein execution of the instruction to determine whether the audio signature matches the audio signature information in the metadata of a quotation causes the control circuitry to determine, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

36. The non-transitory computer-readable medium of item 31, wherein execution of the instruction to determine whether the audio signature matches audio signature information on the metadata of a quotation causes the control circuitry to:

establish a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

establish an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

determine whether the audio signature is between the lower threshold and the upper threshold.

37. The non-transitory computer-readable medium of item 32, wherein execution of the instruction to determine whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words causes the control circuitry to:

establish a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

establish an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

determine whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

38. The non-transitory computer-readable medium of item 33, wherein execution of the instruction to determine whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words causes the control circuitry to:

determine a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

determine a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

determine, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

39. The non-transitory computer-readable medium of item 34, wherein execution of the instruction to determine whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words causes the control circuitry to:

establish, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

establish, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

determine, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

40. The non-transitory computer-readable medium of item 35, wherein execution of the instruction to determine whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words causes the control circuitry to:

establish a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

establish an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

determine whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

41. A method for disambiguating a voice search query, the method comprising:

receiving a voice search query;

transcribing the voice search query into a string comprising a plurality of words;

determining whether the plurality of words includes a command word; and

determining whether to act on the command word by:

determining an audio signature of the voice search query;

querying a database with the string;

receiving, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

comparing the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determining, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generating for display a search result comprising an identifier of the quotation.

42. The method of item 41, further comprising:

determining a cadence of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

comparing a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

43. The method of any of items 41-42, further comprising:

determining an emphasis of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

comparing an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

44. The method of any of items 41-43, further comprising:

determining a duration of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

comparing a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

45. The method of any of items 41-44, further comprising:

determining an overall rhythm of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

comparing the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

46. The method of any of items 41-45, wherein determining whether the audio signature matches audio signature information on the metadata of a quotation comprises:

establishing a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

establishing an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

determining whether the audio signature is between the lower threshold and the upper threshold.

47. The method of item 42, wherein determining whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words comprises:

establishing a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

establishing an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

determining whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

48. The method of item 43, wherein determining whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words comprises:

determining a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

determining a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

determining, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

49. The method of item 44, wherein determining whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words comprises:

establishing, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

establishing, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

determining, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

50. The method of item 45, wherein determining whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words comprises:

establishing a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

establishing an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

determining whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

What is claimed is:

1. A method for disambiguating a voice search query, the method comprising:

receiving a voice search query;

transcribing the voice search query into a string comprising a plurality of words;

determining an audio signature of the voice search query;

querying a database with the string;

receiving, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

comparing the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determining, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generating for display a search result comprising an identifier of the quotation.

2. The method of claim 1, further comprising:

determining a cadence of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

comparing a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

3. The method of claim 1, further comprising:

determining an emphasis of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

comparing an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

4. The method of claim 1, further comprising:

determining a duration of each word of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

comparing a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

5. The method of claim 1, further comprising:

determining an overall rhythm of the plurality of words;

retrieving metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words; and

comparing the overall rhythm of the plurality of words with rhythm information in the metadata for the second plurality of words;

wherein determining whether the audio signature matches the audio signature information in the metadata of a quotation comprises determining, based on the comparing, whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words.

6. The method of claim 1, wherein determining whether the audio signature matches audio signature information on the metadata of a quotation comprises:

establishing a lower threshold by negatively transposing the audio signature information in the metadata by a predetermined amount;

establishing an upper threshold by positively transposing the audio signature information in the metadata by the predetermined amount; and

determining whether the audio signature is between the lower threshold and the upper threshold.

7. The method of claim 2, wherein determining whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words comprises:

establishing a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount;

establishing an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount; and

determining whether the cadence of each word of the plurality of words is between the lower threshold and the upper threshold for the corresponding word of the second plurality of words.

8. The method of claim 3, wherein determining whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words comprises:

determining a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words;

determining a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words; and

determining, for each relative emphasis level of the first plurality of relative emphasis levels, whether the respective relative emphasis level is within a threshold amount of the corresponding relative emphasis level of the second plurality of emphasis levels.

9. The method of claim 4, wherein determining whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words comprises:

establishing, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount;

establishing, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount; and

determining, for each word of the plurality of words, whether the duration of each respective word is between the lower threshold duration and the upper threshold duration for the corresponding word of the second plurality of words.

10. The method of claim 5, wherein determining whether the overall rhythm of the plurality of words matches the rhythm information for the second plurality of words comprises:

establishing a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount;

establishing an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount; and

determining whether the overall rhythm of the plurality of words is between the lower threshold rhythm and the upper threshold rhythm.

11. A system for disambiguating a voice search query, the system comprising:

input circuitry configured to receive a voice search query; and

control circuitry configured to execute the method of any of claims 1-10.

12. A system for disambiguating a voice search query, the system comprising:

input circuitry configured to receive a voice search query; and

control circuitry configured to:

transcribe the voice search query into a string comprising a plurality of words;

determine an audio signature of the voice search query;

query a database with the string;

receive, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information;

compare the audio signature with audio signature information in the metadata of each quotation of the plurality of quotations;

determine, based on the comparing, whether the audio signature matches audio signature information in the metadata of a quotation; and

in response to determining that the audio signature matches audio signature information in the metadata of a quotation, generate for display a search result comprising an identifier of the quotation.

13. The system of claim 12, wherein the control circuitry is further configured to:

determine a cadence of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words; and

compare a cadence of each word of the plurality of words with cadence information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the cadence of each word of the plurality of words matches the cadence information for each corresponding word of the second plurality of words.

14. The system of claim 12 or 13, wherein the control circuitry is further configured to:

determine an emphasis of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words; and

compare an emphasis of each word of the plurality of words with emphasis information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words.

15. The system of any of claims 12-14, wherein the control circuitry is further configured to:

determine a duration of each word of the plurality of words;

retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words; and

compare a duration of each word of the plurality of words with duration information in the metadata for each corresponding word of the second plurality of words;

wherein the control circuitry configured to determine whether the audio signature matches the audio signature information in the metadata of a quotation is further configured to determine, based on the comparing, whether the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words.

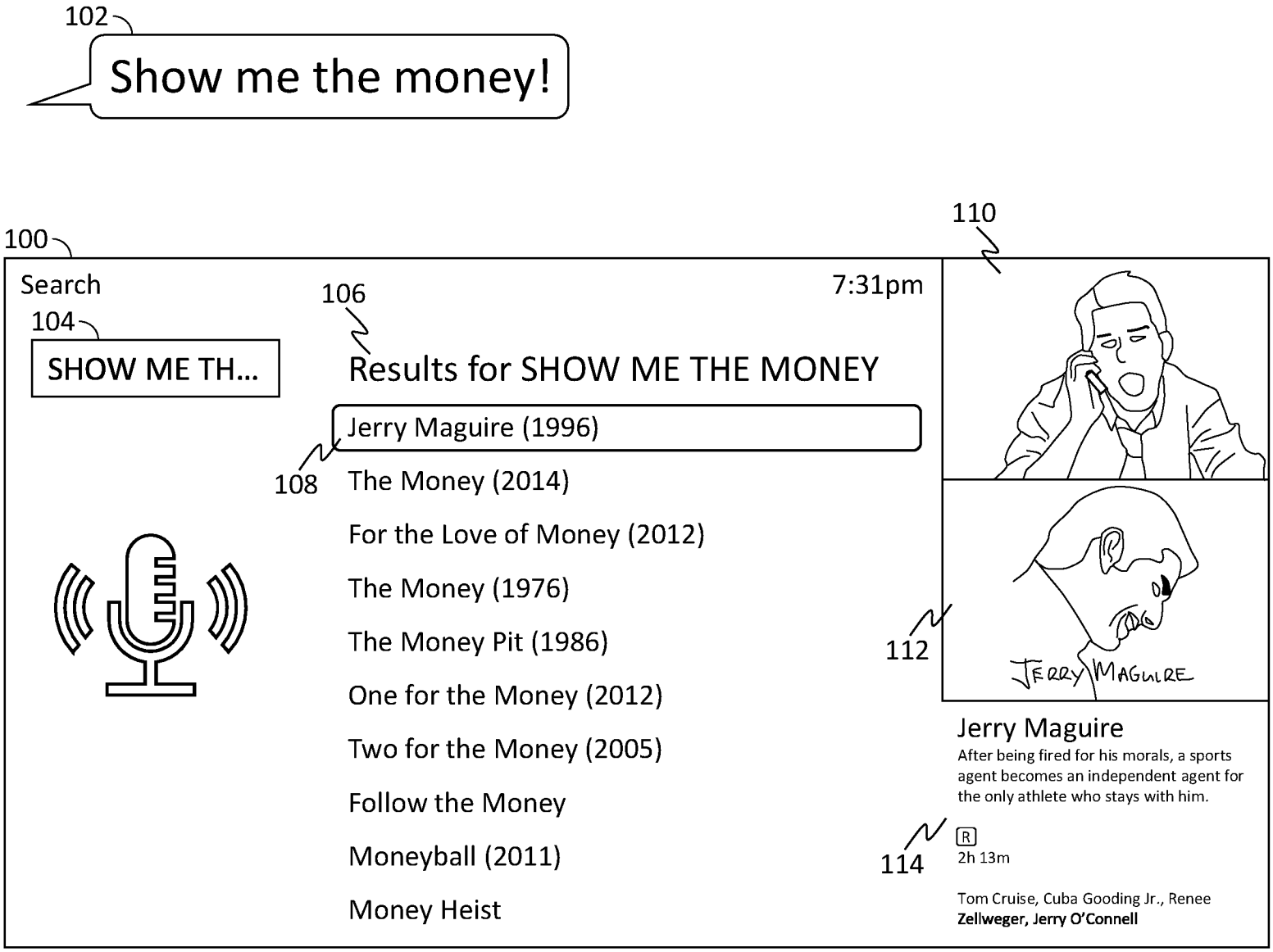
102
Show me the money!
100
Search
104
SHOW ME TH...
106
7:31pm
Results for SHOW ME THE MONEY
Jerry Maguire (1996)
108
The Money (2014)
For the Love of Money (2012)
The Money (1976)
The Money Pit (1986)
One for the Money (2012)
Two for the Money (2005)
Follow the Money
Moneyball (2011)
Money Heist
110
112
JERRY MAGUIRE
Jerry Maguire
After being fired for his morals, a sports agent becomes an independent agent for the only athlete who stays with him.
114
R
2h 13m
Tom Cruise, Cuba Gooding Jr., Renee Zellweger, Jerry O'Connell

FIG. 1

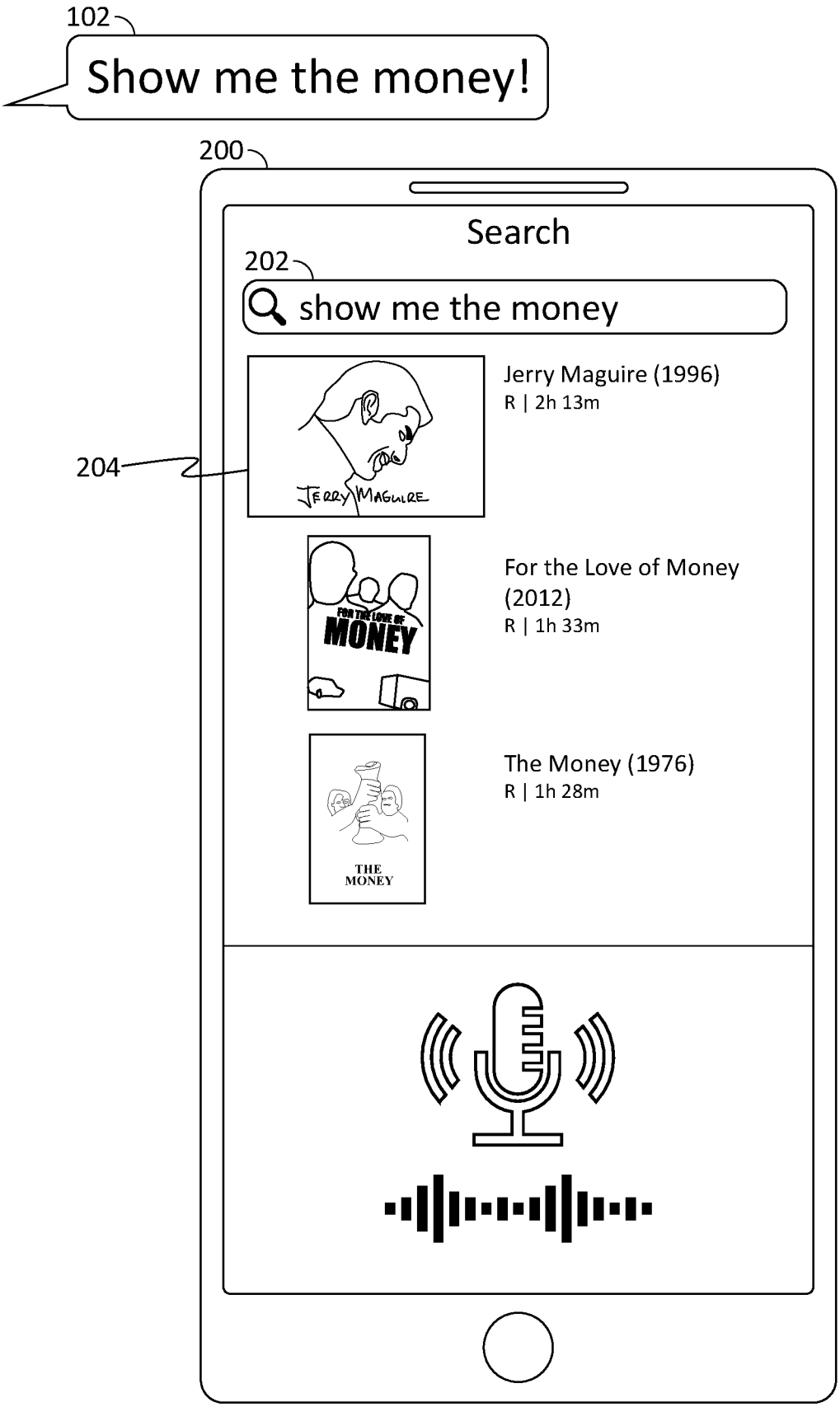
102
Show me the money!
200
Search
202
show me the money
204
Jerry Maguire (1996)
R | 2h 13m
For the Love of Money (2012)
R | 1h 33m
The Money (1976)
R | 1h 28m
THE MONEY

FIG. 2

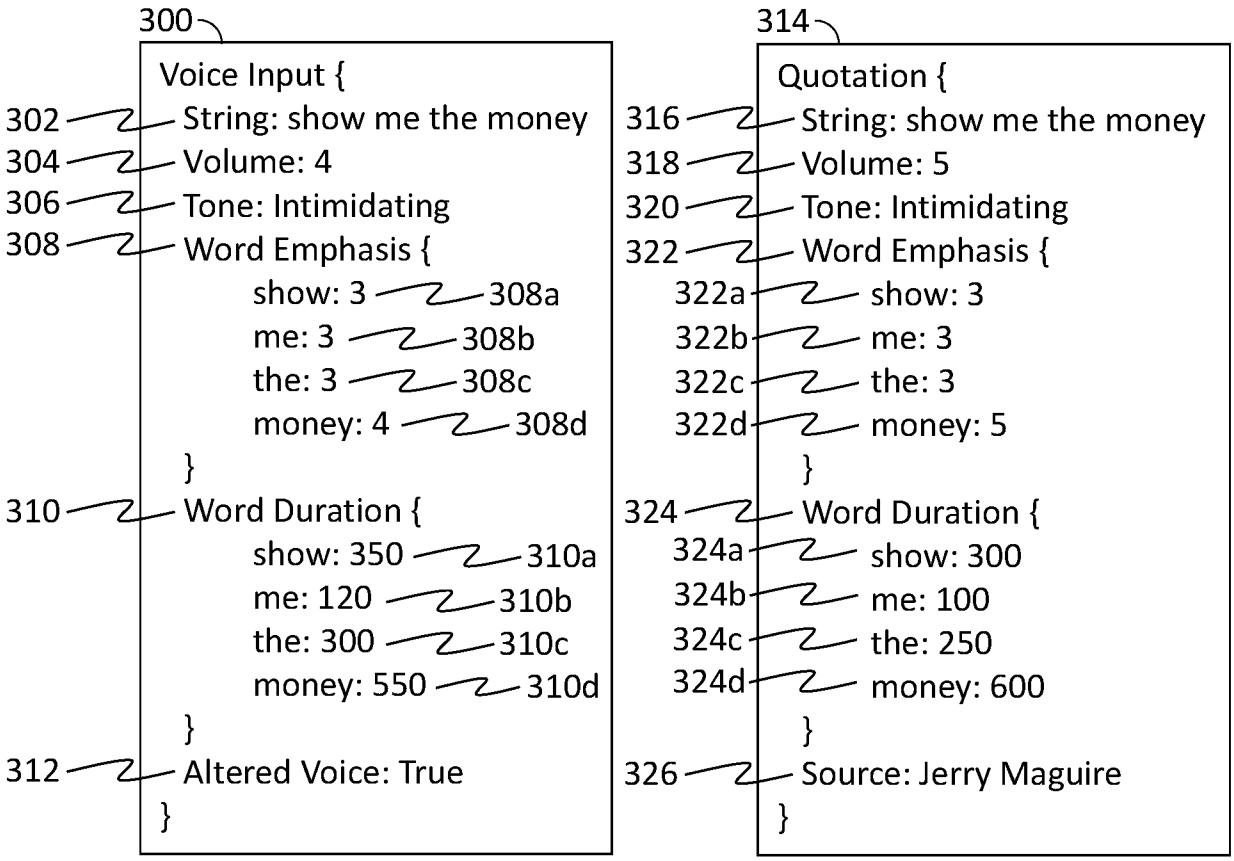
300
Voice Input {
302 String: show me the money
304 Volume: 4
306 Tone: Intimidating
308 Word Emphasis {
show: 3 308a
me: 3 308b
the: 3 308c
money: 4 308d
}
310 Word Duration {
show: 350 310a
me: 120 310b
the: 300 310c
money: 550 310d
}
312 Altered Voice: True
}
314
Quotation {
316 String: show me the money
318 Volume: 5
320 Tone: Intimidating
322 Word Emphasis {
322a show: 3
322b me: 3
322c the: 3
322d money: 5
}
324 Word Duration {
324a show: 300
324b me: 100
324c the: 250
324d money: 600
}
326 Source: Jerry Maguire
}

FIG. 3

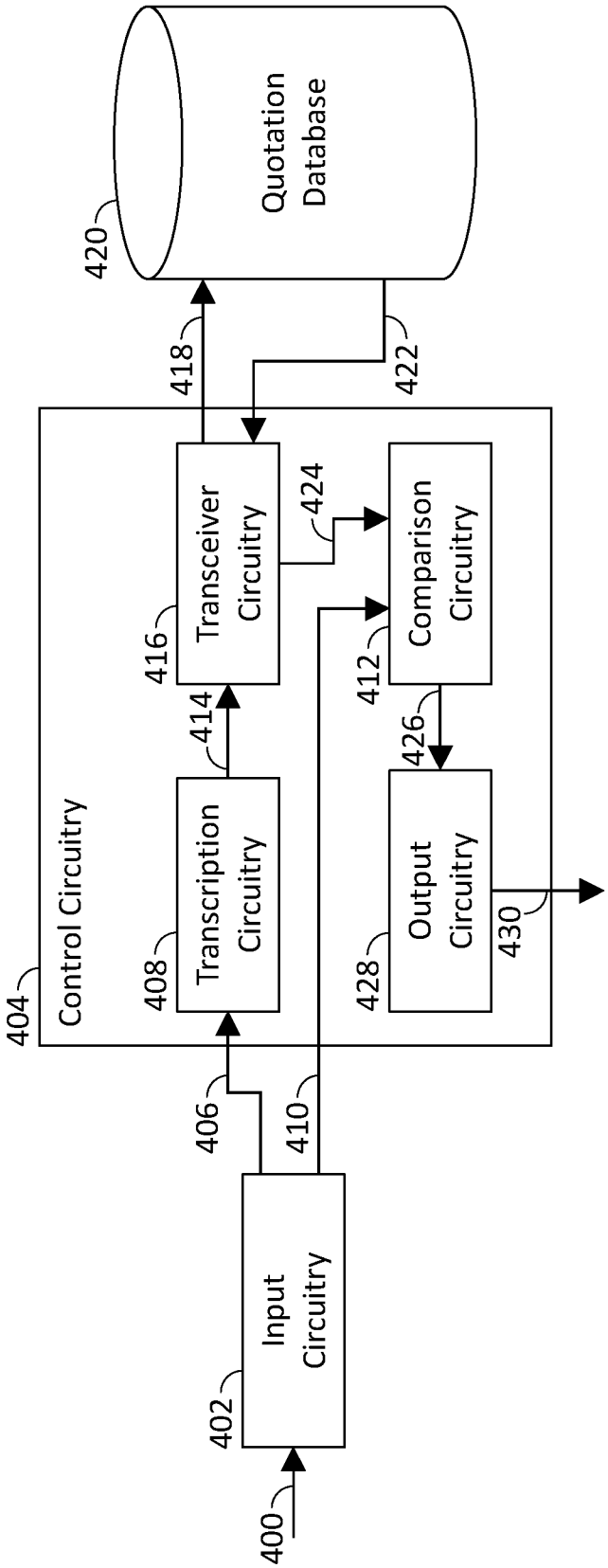
Quotation Database
420
418
422
Control Circuitry
404
Transceiver Circuitry
416
424
Comparison Circuitry
412
414
426
Transcription Circuitry
408
Output Circuitry
428
430
406
410
Input Circuitry
402
400

FIG. 4

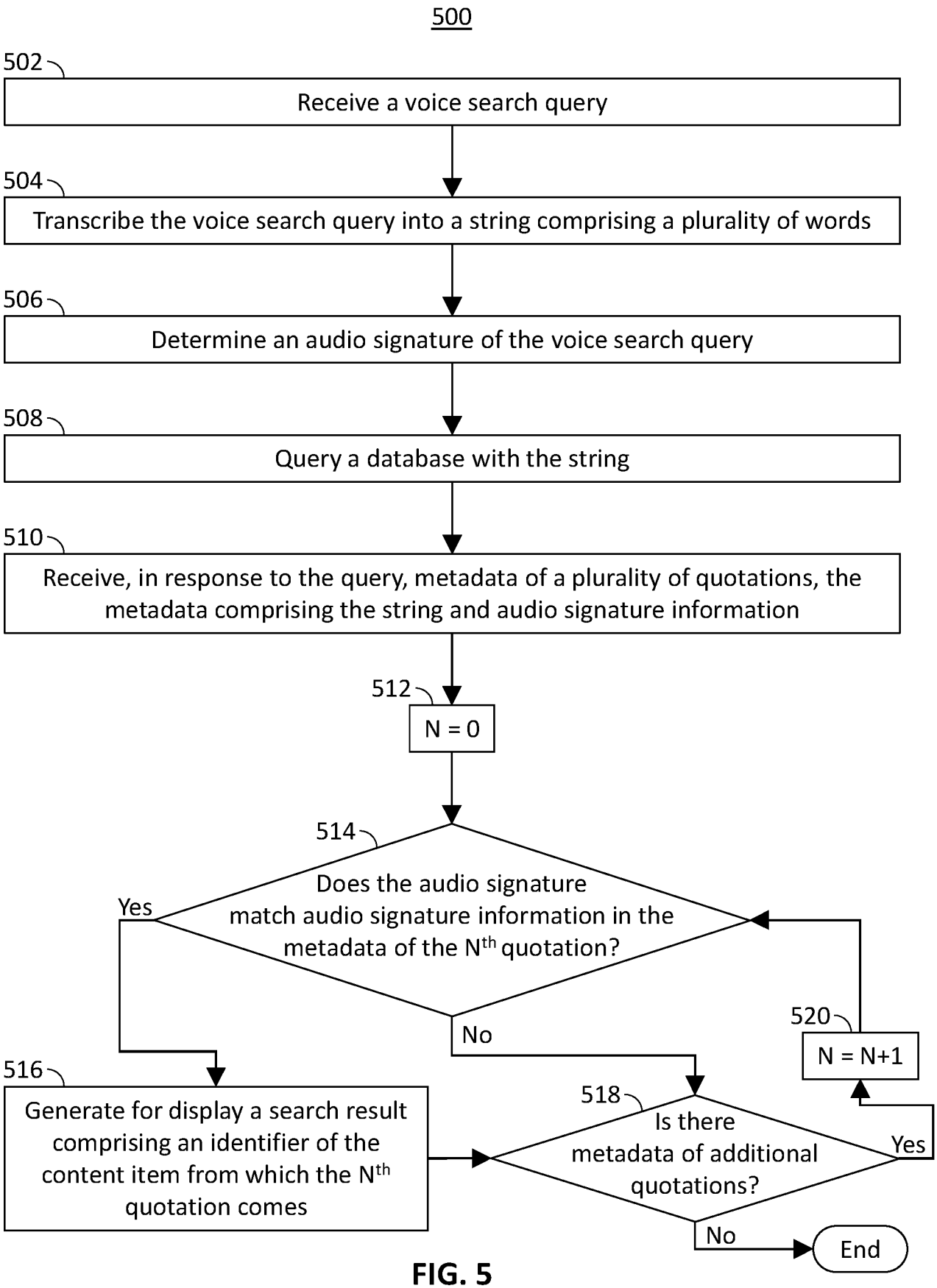
500
502
Receive a voice search query
504
Transcribe the voice search query into a string comprising a plurality of words
506
Determine an audio signature of the voice search query
508
Query a database with the string
510
Receive, in response to the query, metadata of a plurality of quotations, the metadata comprising the string and audio signature information
512
N = 0
514
Does the audio signature match audio signature information in the metadata of the Nth quotation?
Yes
No
516
Generate for display a search result comprising an identifier of the content item from which the Nth quotation comes
518
Is there metadata of additional quotations?
Yes
No
520
N = N+1
End

FIG. 5

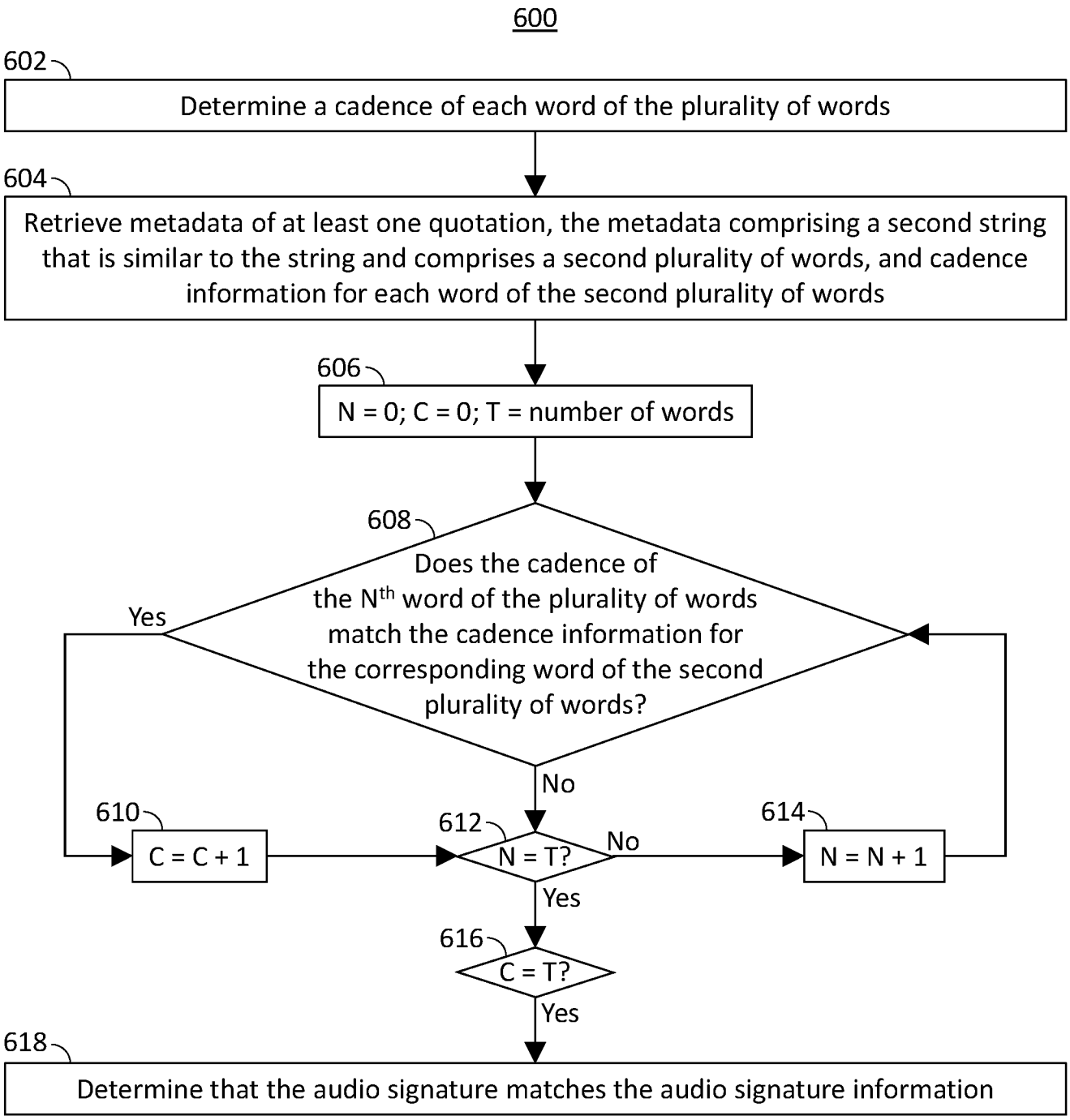
600
602
Determine a cadence of each word of the plurality of words
604
Retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and cadence information for each word of the second plurality of words
606
N = 0; C = 0; T = number of words
608
Does the cadence of the Nth word of the plurality of words match the cadence information for the corresponding word of the second plurality of words?
Yes
No
610
C = C + 1
612
N = T?
No
Yes
614
N = N + 1
616
C = T?
Yes
618
Determine that the audio signature matches the audio signature information

FIG. 6

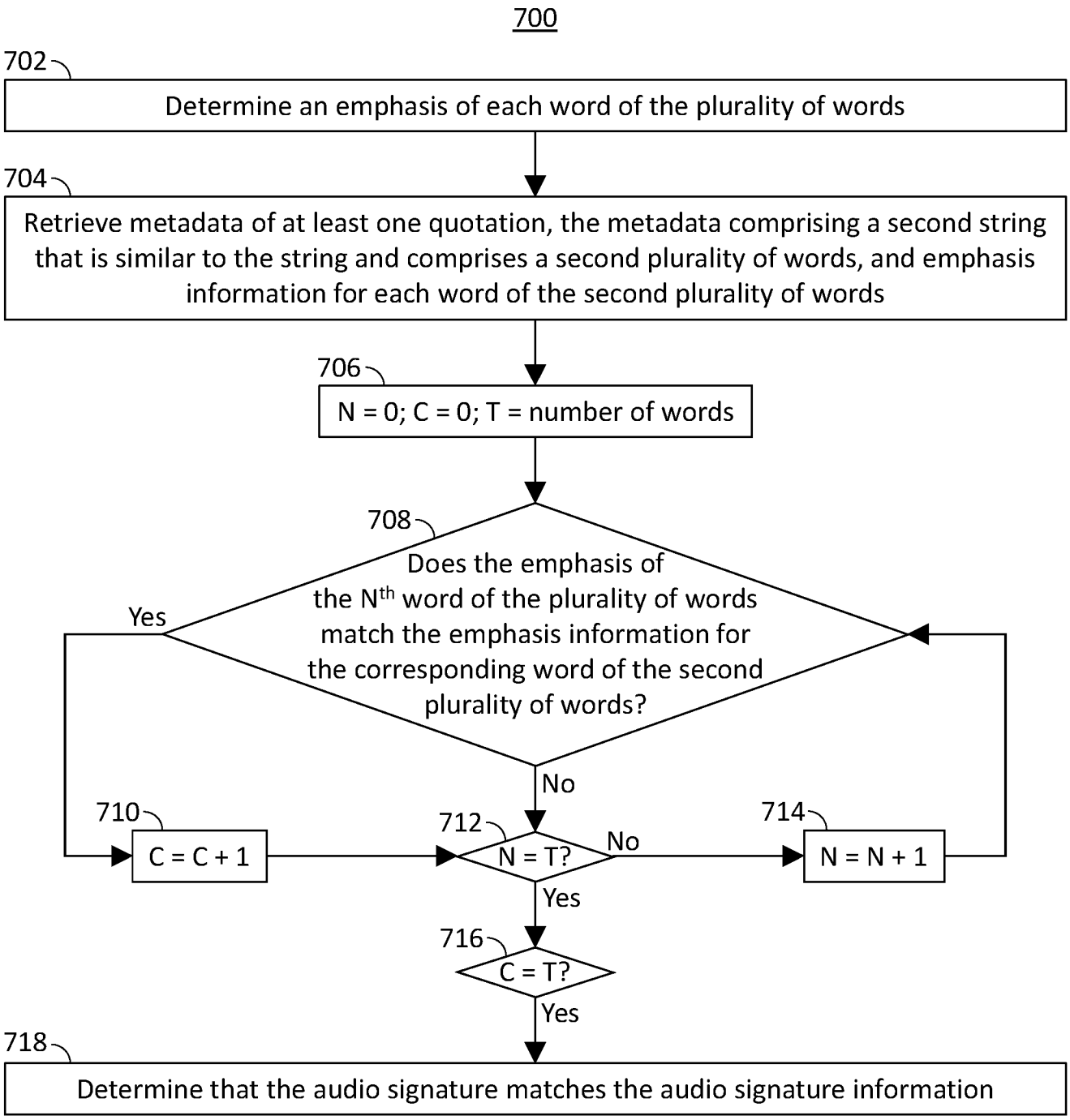
700
702
Determine an emphasis of each word of the plurality of words
704
Retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and emphasis information for each word of the second plurality of words
706
N = 0; C = 0; T = number of words
708
Does the emphasis of the Nth word of the plurality of words match the emphasis information for the corresponding word of the second plurality of words?
Yes
No
710
C = C + 1
712
N = T?
No
Yes
714
N = N + 1
716
C = T?
Yes
718
Determine that the audio signature matches the audio signature information

FIG. 7

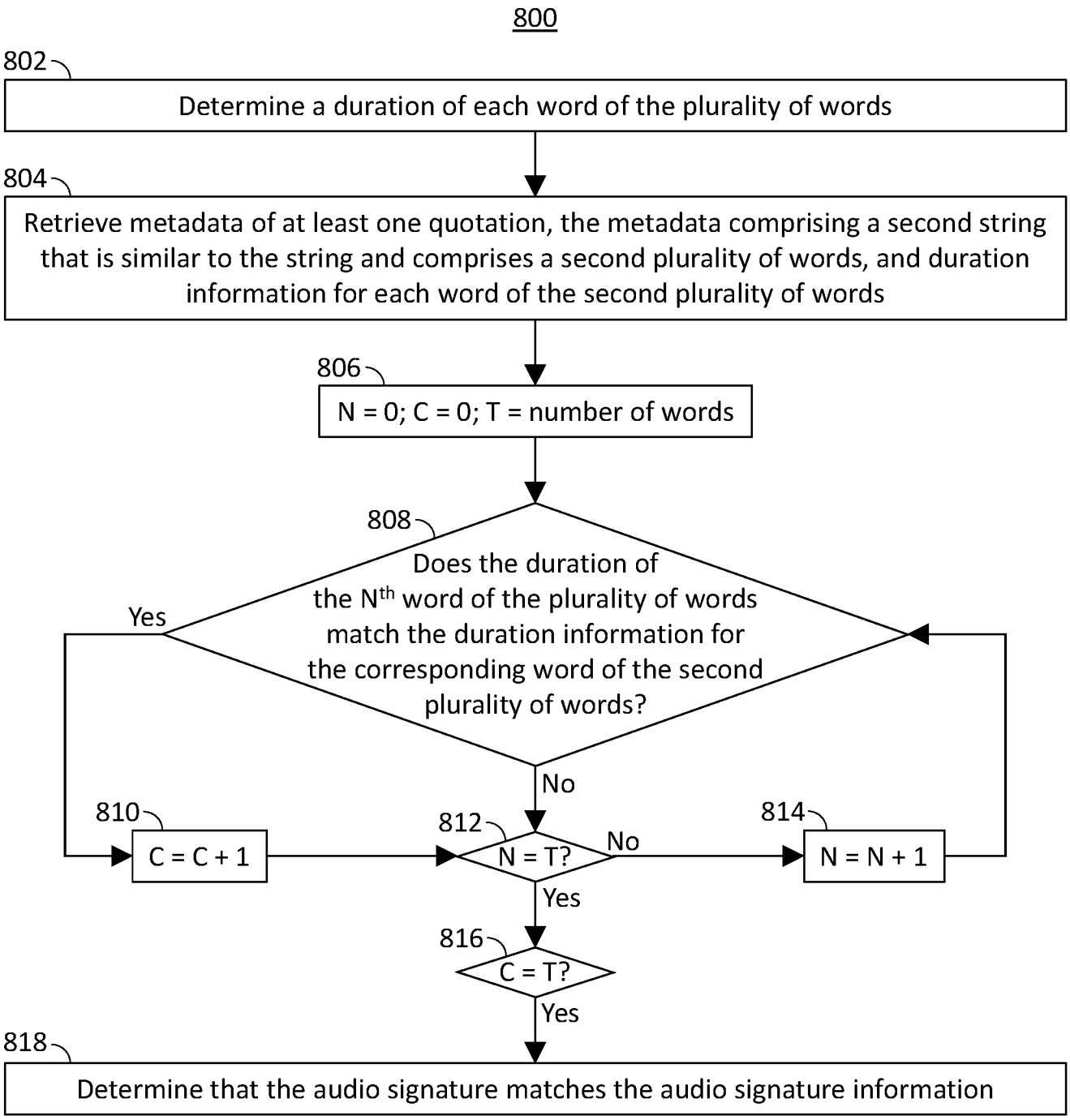
800
802
Determine a duration of each word of the plurality of words
804
Retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and duration information for each word of the second plurality of words
806
N = 0; C = 0; T = number of words
808
Does the duration of the Nth word of the plurality of words match the duration information for the corresponding word of the second plurality of words?
Yes
No
810
C = C + 1
812
N = T?
No
Yes
814
N = N + 1
816
C = T?
Yes
818
Determine that the audio signature matches the audio signature information

FIG. 8

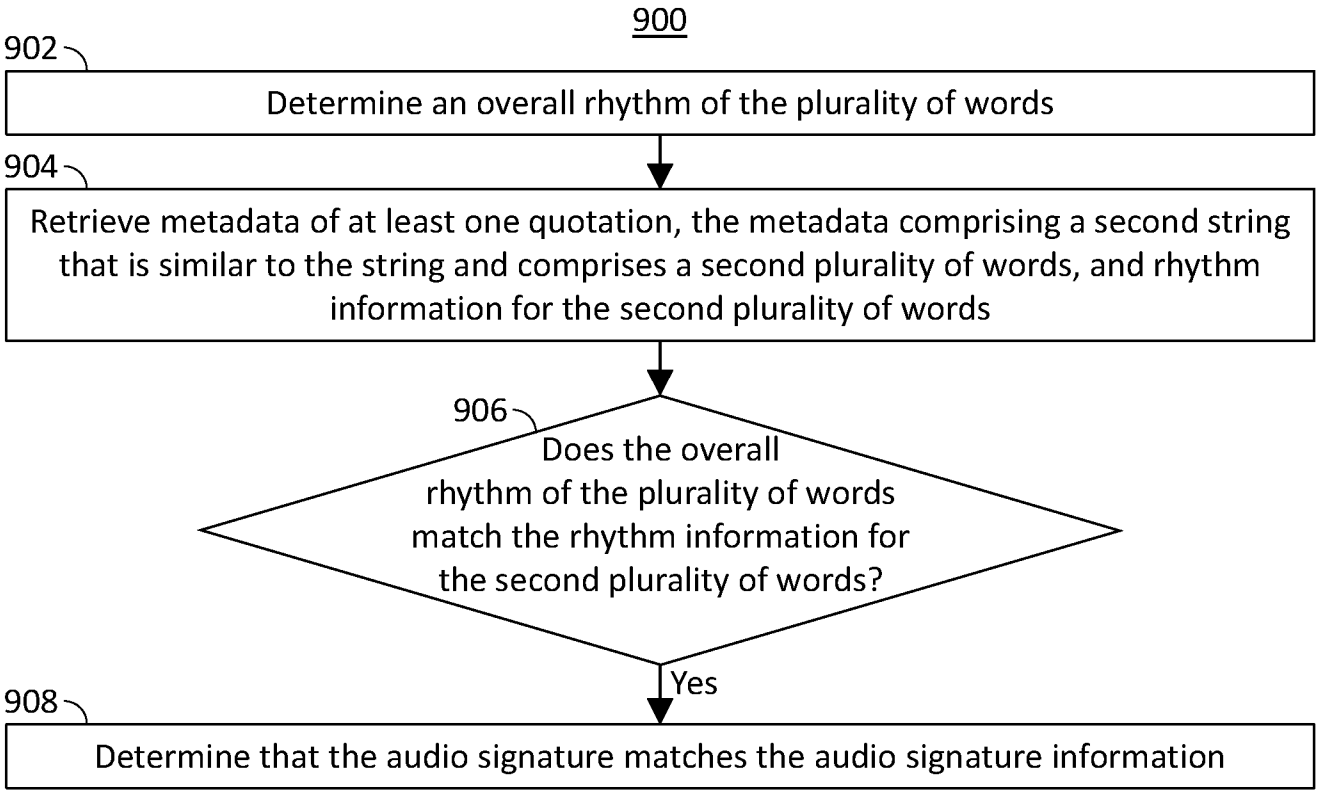
900
902
Determine an overall rhythm of the plurality of words
904
Retrieve metadata of at least one quotation, the metadata comprising a second string that is similar to the string and comprises a second plurality of words, and rhythm information for the second plurality of words
906
Does the overall rhythm of the plurality of words match the rhythm information for the second plurality of words?
Yes
908
Determine that the audio signature matches the audio signature information

FIG. 9

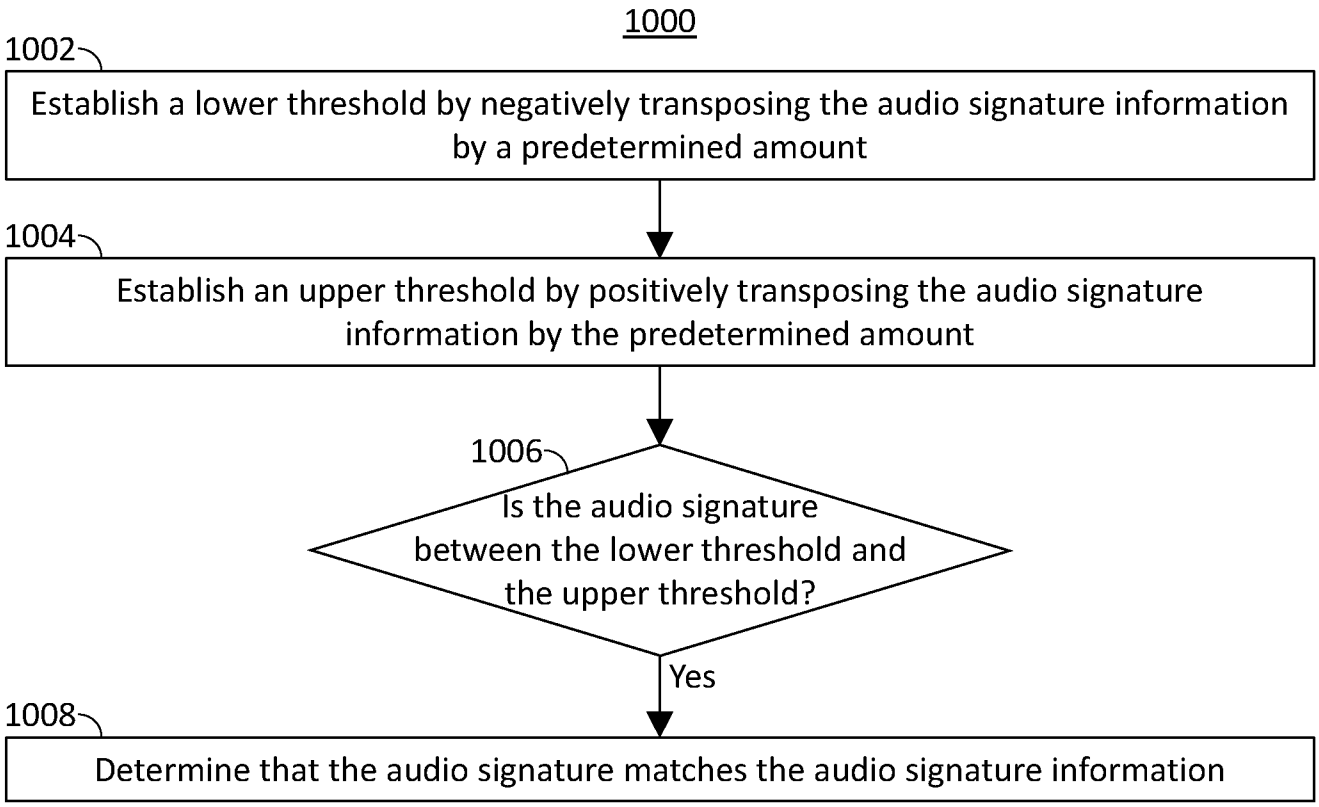
1000
1002
Establish a lower threshold by negatively transposing the audio signature information by a predetermined amount
1004
Establish an upper threshold by positively transposing the audio signature information by the predetermined amount
1006
Is the audio signature between the lower threshold and the upper threshold?
Yes
1008
Determine that the audio signature matches the audio signature information

FIG. 10

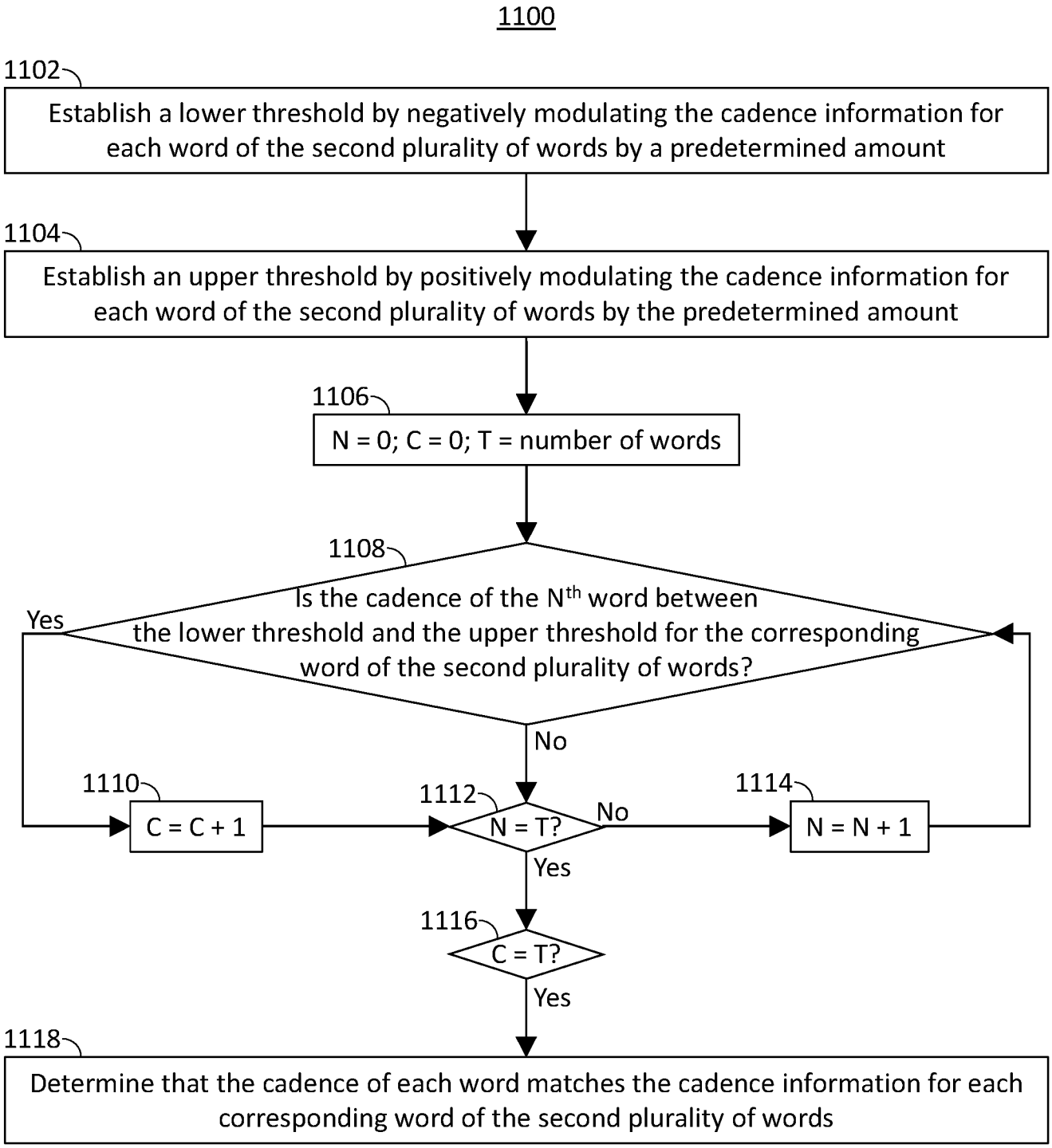
1100
1102
Establish a lower threshold by negatively modulating the cadence information for each word of the second plurality of words by a predetermined amount
1104
Establish an upper threshold by positively modulating the cadence information for each word of the second plurality of words by the predetermined amount
1106
N = 0; C = 0; T = number of words
1108
Is the cadence of the Nth word between the lower threshold and the upper threshold for the corresponding word of the second plurality of words?
Yes
No
1110
C = C + 1
1112
N = T?
No
Yes
1114
N = N + 1
1116
C = T?
Yes
1118
Determine that the cadence of each word matches the cadence information for each corresponding word of the second plurality of words

FIG. 11

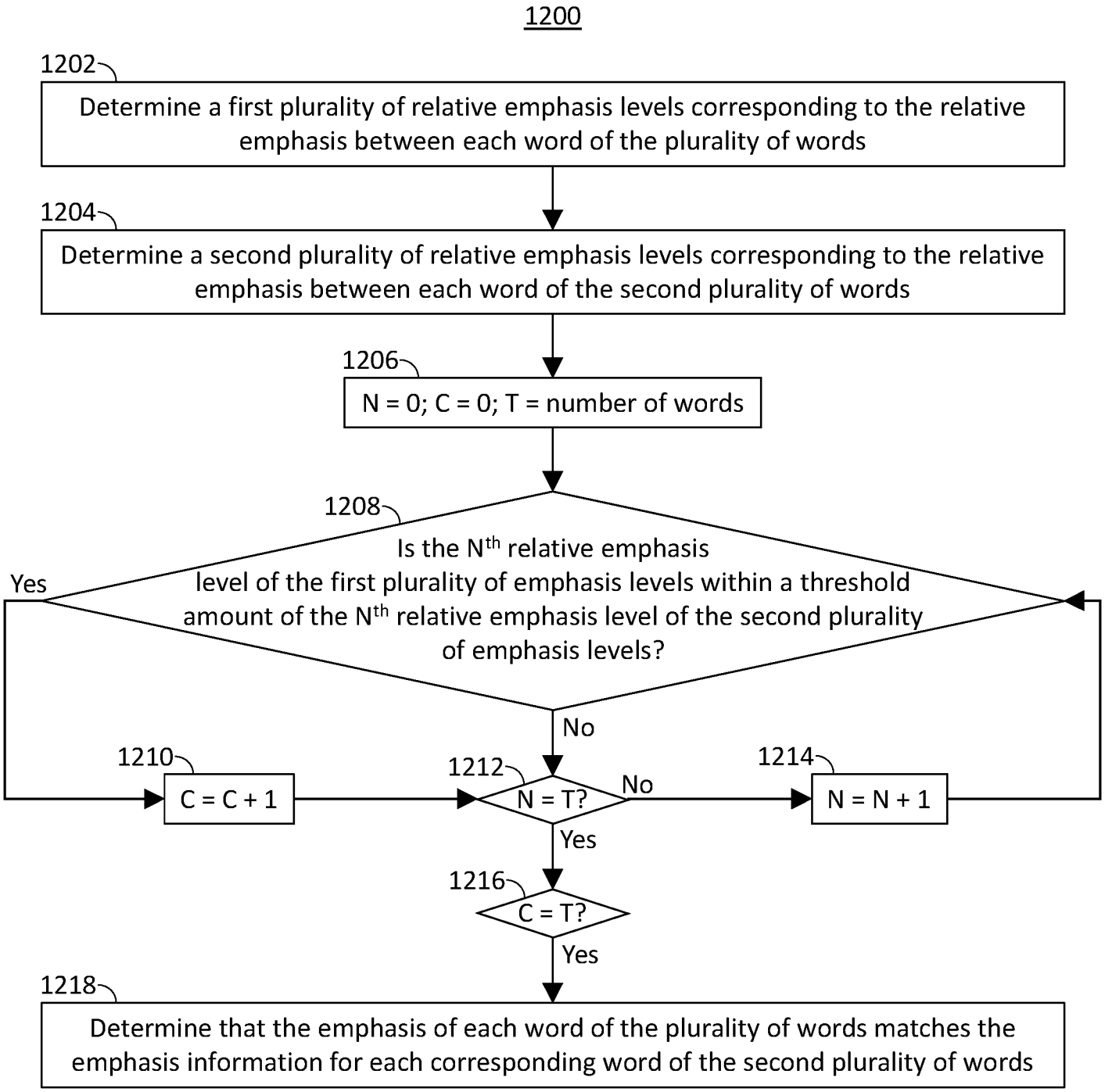
1200
1202
Determine a first plurality of relative emphasis levels corresponding to the relative emphasis between each word of the plurality of words
1204
Determine a second plurality of relative emphasis levels corresponding to the relative emphasis between each word of the second plurality of words
1206
N = 0; C = 0; T = number of words
1208
Is the Nth relative emphasis level of the first plurality of emphasis levels within a threshold amount of the Nth relative emphasis level of the second plurality of emphasis levels?
Yes
No
1210
C = C + 1
1212
N = T?
No
Yes
1214
N = N + 1
1216
C = T?
Yes
1218
Determine that the emphasis of each word of the plurality of words matches the emphasis information for each corresponding word of the second plurality of words

FIG. 12

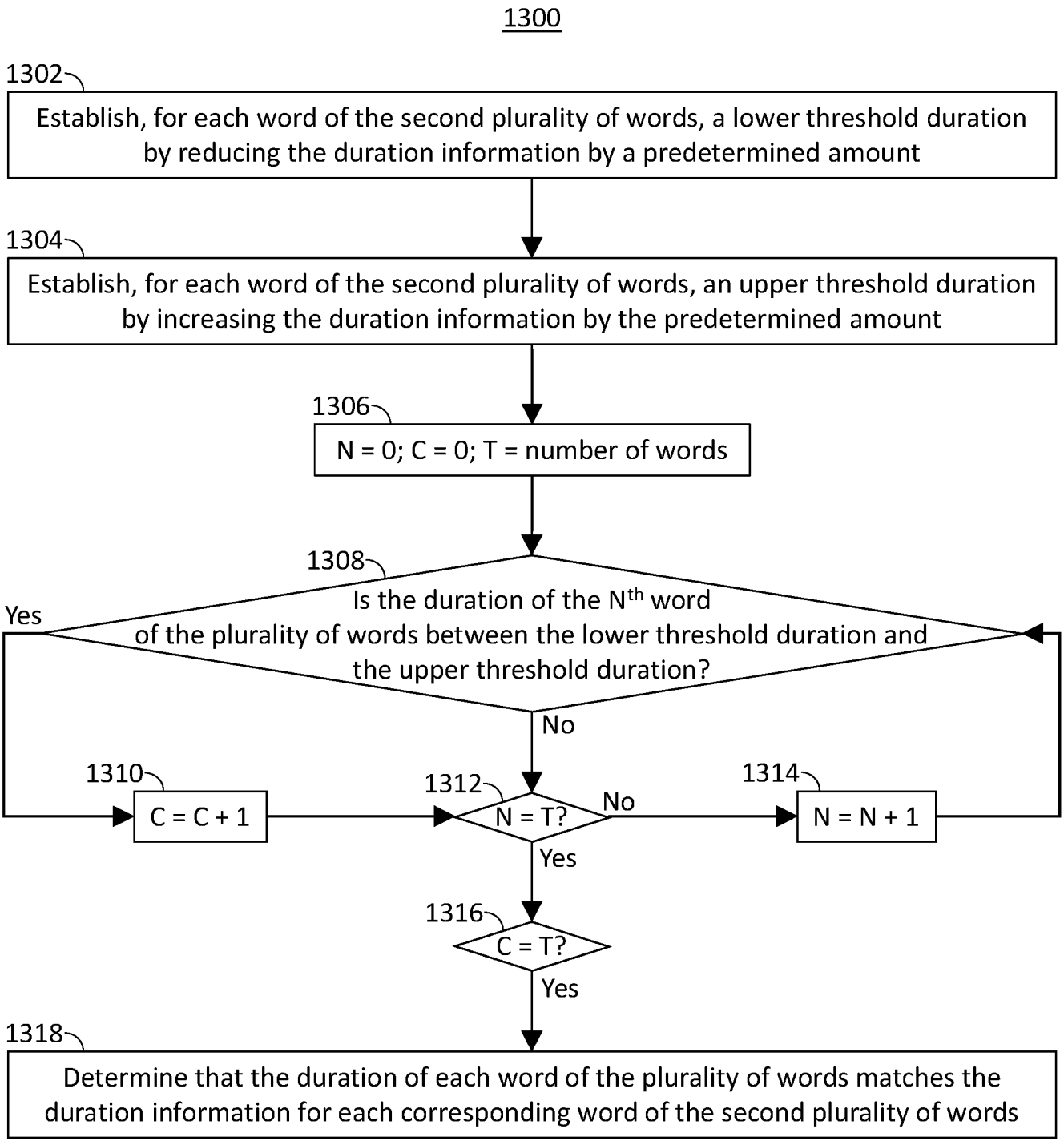
1300
1302
Establish, for each word of the second plurality of words, a lower threshold duration by reducing the duration information by a predetermined amount
1304
Establish, for each word of the second plurality of words, an upper threshold duration by increasing the duration information by the predetermined amount
1306
N = 0; C = 0; T = number of words
1308
Is the duration of the Nth word of the plurality of words between the lower threshold duration and the upper threshold duration?
Yes
No
1310
C = C + 1
1312
N = T?
No
Yes
1314
N = N + 1
1316
C = T?
Yes
1318
Determine that the duration of each word of the plurality of words matches the duration information for each corresponding word of the second plurality of words

FIG. 13

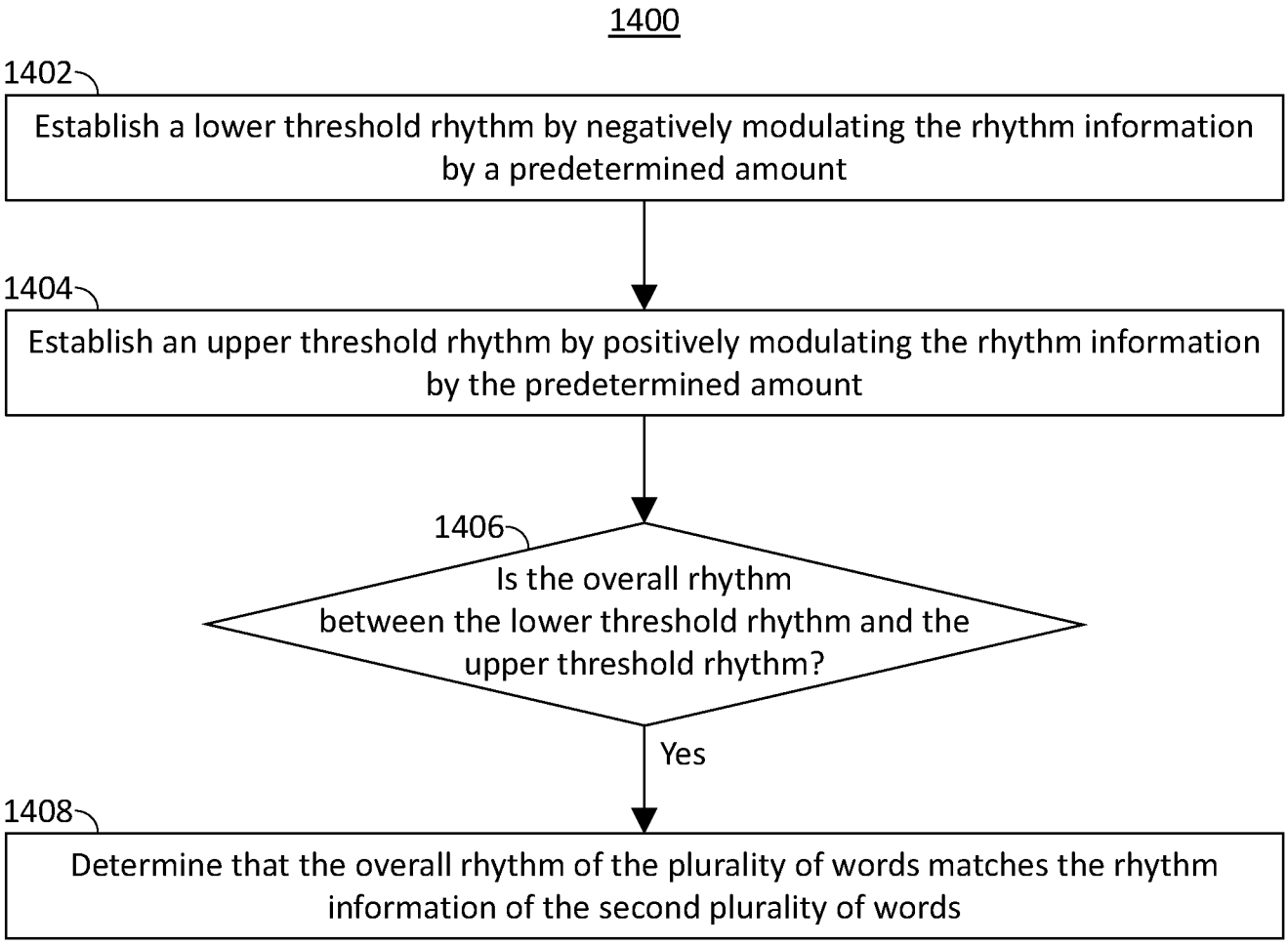
1400
1402
Establish a lower threshold rhythm by negatively modulating the rhythm information by a predetermined amount
1404
Establish an upper threshold rhythm by positively modulating the rhythm information by the predetermined amount
1406
Is the overall rhythm between the lower threshold rhythm and the upper threshold rhythm?
Yes
1408
Determine that the overall rhythm of the plurality of words matches the rhythm information of the second plurality of words

FIG. 14

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/025729

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F16/332 G06F16/683
ADD.

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

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 practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/145214 A1 (ZHANG FAN [CA] ET AL) 16 June 2011 (2011-06-16) figure 3	1-15
A	US 2018/097836 A1 (STOLARZ DAMIEN PHELAN [US] ET AL) 5 April 2018 (2018-04-05) paragraphs [0186] - [0188]	1-15

☐ 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 application or patent 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

"T" 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

"X" 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

"Y" 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	Date of mailing of the international search report
10 June 2020	24/06/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hackelbusch, Richard

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/025729

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011145214 A1	16-06-2011	NONE	
US 2018097836 A1	05-04-2018	EP 3520361 A1	07-08-2019
		US 2018097810 A1	05-04-2018
		US 2018097827 A1	05-04-2018
		US 2018097836 A1	05-04-2018
		US 2018097837 A1	05-04-2018
		US 2018097838 A1	05-04-2018
		US 2018097841 A1	05-04-2018
		WO 2018067393 A1	12-04-2018