(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 3 648 106 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
06.05.2020 Bulletin 2020/19

(21) Application number: **19206103.4**

(22) Date of filing: **30.10.2019**

(51) Int Cl.:
G11B 27/32 *(2006.01)* ***G11B 27/34*** *(2006.01)*
G06F 16/632 *(2019.01)* ***G10L 15/18*** *(2013.01)*

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(30) Priority: **02.11.2018 SE 1851369**

(71) Applicant: **Spotify AB**
111 53 Stockholm (SE)

(72) Inventors:
- **ROY, Brian**
  111 53 Stockholm (SE)
- **EDMONDS, Philip**
  111 53 Stockholm (SE)
- **KANE, Matthew**
  111 53 Stockholm (SE)
- **THOM, JENNIFER**
  111 53 Stockholm (SE)
- **KOTHARI, Neha**
  111 53 Stockholm (SE)
- **MENNICKEN, Sarah**
  111 53 Stockholm (SE)
- **HUMPHREYS, Karl**
  111 53 Stockholm (SE)
- **BRILLMAN, Ruth**
  111 53 Stockholm (SE)
- **REDDY, Sravana**
  111 53 Stockholm (SE)
- **CRAMER, Henriette**
  111 53 Stockholm (SE)
- **WILLIAMS, Robert**
  111 53 Stockholm (SE)
- **KUMAR, Rohit**
  111 53 Stockholm (SE)

(74) Representative: **Kransell & Wennborg KB**
P.O. Box 27834
115 93 Stockholm (SE)

(54) **MEDIA CONTENT STEERING**

(57) A media content steering solution is provided to identify a user query to steer playback of media content that is currently playing or has been played. The user steering query can include a voice request for playing media content that is relatively different from the media content being currently played or having been played. The media content steering solution analyzes the utterance of the user query and uses it to identify such different content that satisfies the user intent contained in the user query.

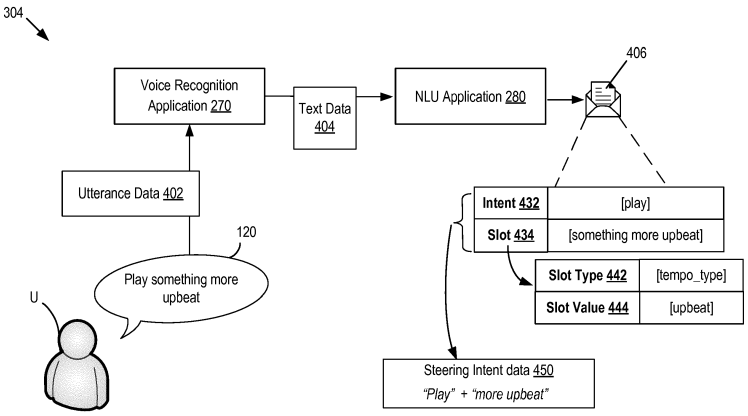

FIG. 4

EP 3 648 106 A1

Description

BACKGROUND

[0001] Media playback devices are used to receive streams of media content across networks from streaming services. The streaming services, in turn, provide media playback devices with access to relatively large collections of media content. The streaming services store the media content on servers remote from the media playback devices and then send the media content to the electronic devices when requested. Typically, a media playback device sends a server of the streaming service a playback request including an identification of media content to play.

[0002] Media content is information that is expressed as various forms, such as music, speech, writing, or any other arts. Media content is deliverable through a medium of various types, such as the Internet, mobile devices, CDs, DVDs, books, e-books, magazines, cinema, and live events. Examples of media content include songs, albums, videos, movies, radio stations, TV stations, TV shows, books, video games and the like.

[0003] Different pieces of media content are referred to as media content items that are typically identifiable through use of media content item identifiers. Each media content item is thus retrievable for playback via use of a media content item identifier thereof. A media content item can be also referred to as a media object, multimedia object, or the like.

[0004] One or more media content items can be grouped and identified as a whole as a media context to provide a particular context to the group of media content items. Examples of media contexts include playlists, albums, artists, stations, search result lists, and other things suitable to suggest context of media playback. In some examples, a media context includes one or more media content item identifiers for identifying the media content items associated with the media context. In some examples, a media context itself is identifiable through use of an identifier of the media content (also referred to herein as a media context identifier) and retrievable for playback via use of the media context identifier thereof.

[0005] When particular media content, such as a media content item or a media context, is being played via a media playback device, a user may want to listen to other media content that is relatively different from the currently-playing media content. For example, when listening to media content via a voice-enabled media playback device, a user can provide a voice command to change the playback to different media content. A voice command as used herein is a verbal instruction that is used to control a device. In some embodiments, a voice command is used to control a media playback device or a media delivery system via the media playback device.

[0006] Known media devices and streaming services typically receive user input (e.g., a user query) via a verbal or tactile input interface to change playback of media content. Existing media content recommendation techniques include mood-based recommendations systems as found in Mortensen, et al., Real-World Mood-Based Recommendation, AIRS 2008: Information Retrieval Technology, pp. 514-519.

[0007] While various media content recommendation techniques exist, such techniques do not provide for media content recommendation or search in conjunction with comparative-type instructions that include terms such as such as "more", "less", "faster", "slower" and the like. When a user utters "play something upbeat," typical services return a static playlist of music that is already defined as "upbeat" and is not relative to the user's currently listening media content. For example, if a user initially requests a service to play "some sad classical music," it will play a playlist such as "Solemn Classical Music" and play "Gymnopedie No.1", by Erik Satie. But if, during the playback of this piece, the service is requested to play "something more upbeat," it will not play more upbeat classical music but rather play media content that is very distant relative to, for example, genre, such as "Country Party Soundtrack" and play, for example, "This is How We Roll" by Florida Georgia Line.

[0008] Further, such a static playlist is not personalized to an account through which the request is made. For example, it is not personalized to the known or predicted preferences or activity of the account holder.

[0009] What is needed is a solution that identifies a voice command as a command containing comparative-type instructions, and determines media content in accordance with that voice command, particularly media content that is relatively different from the media content being currently played or queued to be played, where the relative difference is bounded by certain constraints.

[0010] The technical problem resides in the user interface. Existing interfaces of media playback devices and streaming services perform only simplistic interpretations of voice commands and lack the ability to identify utterances containing comparative-type instructions and process them to playback media content in any given direction that is constrained by the attributes of the media content items that are currently playing or queued to be played. Therefore, the existing solutions do not return media content items that accurately reflect the user's command relative to the media content items being currently played or queued to be played.

[0011] Further, if search results are not satisfactory, users typically retry searching with other voice commands, and accordingly the systems run searches multiple times, until the systems provide media content items that meet the users' expectations. Such multiple searches waste computing power, network and database resources, and cause inferior user experience.

[0012] Another technical problem resides in the recommendation component. Because known recommendation components of streaming devices and services are limited to such simplistic interpretations, they lack the ability to analyze, parse and interpret multiple parts of

the utterance to provide them with the additional information needed to change playback of media content to an appropriate direction and not stray too far from media content attributes of the media content item currently playing or queued to be played.

[0013] There exists, therefore, a need for improved technology that responds to user queries with comparative-type instructions associated with media content playback.

SUMMARY

[0014] In general terms, the present disclosure is directed to media content steering. In one possible configuration and by non-limiting example, the media content steering provides a solution to receive a user request for steering playback of media content and provide relatively different media content based on the user request. Various aspects are described in this disclosure, which include, but are not limited to, the following aspects.

[0015] According to an aspect of the present invention, there is provided a media playback system comprising a media playback device operable to play media content, and a media delivery system. The media delivery system is configured to receive an utterance from the media playback device. The media delivery system is also configured to identify at least one first media content item that is currently being played by the media playback device. The media delivery system is also configured to parse the utterance into parts including one or more slots. The media delivery system is also configured to identify a first facet of the at least one first media content item that matches the one or more slots, the first facet having a first facet type and a first facet value. The media delivery system is also configured to retrieve at least one second media content item from a catalog of media content items, the at least one second media content item having a second facet, the second facet having a second facet type and a second facet value, the second facet type being the same as the first facet type, and the second facet value having a relative difference from the first facet value, the relative difference corresponding to the one or more slots. The media delivery system is also configured to transmit the second media content item to the media playback device.

[0016] In some embodiments, the parsing of the utterance further includes parsing the utterance into parts including an intent and the one or more slots. In some embodiments, the identifying of the first facet comprises identifying the first facet that also matches the intent.

[0017] In some embodiments, each of the one or more slots has a slot type and a slot value. In some embodiments, the first facet matches at least one slot of the one or more slots by the facet type of the first facet being the same as the slot type of said at least one slot. Additionally or alternatively, in some embodiments, the relative difference between the first and second facet values corresponds to the one or more slots by corresponding to the slot value of at least one of said one or more slots.

[0018] According to another aspect of the present invention, there is provided a method performed by a media content steering engine of a media delivery system connected to a media playback device via a data communication network. The method comprises, from the media playback device, receiving a steering command from a user of said media playback device. The method also comprises, based on the received steering command, identifying a media content facet type for a first media content item which is currently playing on the media playback device. The method also comprises determining a first facet value of the identified facet type The method also comprises, based on the received steering command and in relation to the determined first facet value, selecting a second media content item having a second facet value of said identified facet type, different from the first facet value. The method also comprises, providing the second media content item to the media playback device in response to the received steering command.

[0019] In some embodiments of the present invention, the steering command is obtained from an utterance received from the media playback device in voice (i.e. audio) or text format. In some embodiments, the utterance is parsed before the steering command is obtained there from. According to another aspect of the present invention, there is provided a media delivery system provided with means for performing an embodiment of the method performed by a media content steering engine of the present disclosure.

[0020] According to another aspect of the present invention, there is provided a computer program product comprising computer-executable components for causing a media delivery system to perform an embodiment of a method of the present disclosure when the computer-executable components are run on processing circuitry comprised in the media delivery system.

[0021] According to another aspect of the present invention, there is provided a method performed by a media playback device connected to a media delivery system via a data communication network. The method comprises, from a user of the media playback device, receiving a steering command. The method also comprises, based on the received steering command, identifying a media content facet type for a first media content item which is currently playing on the media playback device. The method also comprises determining a first facet value of the identified facet type. The method also comprises transmitting the received steering command to the media delivery system. The method also comprises, in response to the transmitted steering command, obtaining a second media content item having a second facet value of said identified facet type, different from the first facet value. The method also comprises, playing said second media content item. Thus, in different embodiments of the present invention, some method steps may be performed in the media playback device instead of, or in addition to, in the media delivery system.

[0022] According to another aspect of the present invention, there is provided a media playback device comprising means for performing an embodiment of the method which is performed by a media playback device, in accordance with the present disclosure.

[0023] According to another aspect of the present invention, there is provided a computer program product comprising computer-executable components for causing a media playback device to perform an embodiment of a method of the present disclosure when the computer-executable components are run on processing circuitry comprised in the media playback device.

[0024] Some further aspects of the present invention are presented below.

[0025] According to an aspect of the present invention, there is provided a media playback system comprising a media playback device operable to play media content, and a media delivery system. The media delivery system is configured to receive an utterance from the media playback device. The media delivery system is also configured to identify at least one first media content item that is currently being played by the media playback device, the at least one first media content item having a first facet, the first facet having a first facet type and a first facet value. The media delivery system is also configured to parse the utterance into parts including information. The media delivery system is also configured to retrieve at least one second media content item from a catalog of media content items, the at least one second media content item having a second facet, the second facet having a second facet type and a second facet value, the second facet type being the same as the first facet type and the second facet value being different from the first facet value, the difference being representative of at least part of the information in the parsed utterance. The media delivery system is also configured to transmit the second media content item to the media playback device.

[0026] One other aspect is a method for playing media content. The method may include receiving an utterance, and identifying at least one first media content item that is currently being played. The first media content item may have a first facet that has a first facet type and a first facet value. The method may further include determining an intent and a slot from the utterance. The method may further include retrieving at least one second media content item from a catalog of media content items. The second media content item may have a second facet that has a second facet type and a second facet value. The second facet type may be the same as the first facet type. The second facet value may be different from the first facet value. The difference may represent at least one of the intent and the slot of the utterance.

[0027] In certain examples, the method may further include determining the first facet type and the second facet type based on at least one of the intent and the slot of the utterance.

[0028] In certain examples, the method may further include determining a slot type and a slot value from the slot, and determining the first facet type and the second facet type based on the slot type.

[0029] In certain examples, the first media content item has a plurality of facets including the first facet, and each facet has a facet type and a facet value. The second media content item has a plurality of facets including the second facet, and each facet has a facet type and a facet value. In certain examples, at least some of the facet types in the one second media content item are the same as at least some of the facet types in the first media content item. In certain examples, the facet values, other than the first facet value, of the at least some of the facet types in the second media content item are the same as the facet values, other than the second facet value, of the at least some of the facet types in the first media content item.

[0030] In certain examples, the method may further include identifying a media content steering intent from the utterance. The media content steering intent includes a command to retrieve at least one media content item having a facet that has a different facet value from a corresponding facet of at least one currently-playing media content item.

[0031] In certain examples, the method may further include retrieving user information associated with a user who provide the utterance. The second facet value of the second media content item may be determined based on the user profile.

[0032] In certain examples, the method may further include, prior to receiving the utterance, transmitting a media content signal to a voice-enabled device for playing the at least one first media content item. The method may further include receiving the utterance from the voice-enabled device, and converting the utterance to a text version. The intent and the slot of the utterance may be determined from the text version.

[0033] In certain examples, the first facet type and the second facet type include at least one of mood, genre, activity, tempo, acoustic attributes, artists, popularity, geographic information, release date, hits, and deep cuts.

[0034] In certain examples, the first media content item is a first media content playlist including a plurality of first tracks. In certain examples, the second media content item is a second media content playlist including a plurality of second tracks.

[0035] In certain examples, the first media content item is included in a first media content playlist. In certain examples, the second media content item is included in a second media content playlist.

[0036] Another aspect is a computer-readable medium having stored thereon instructions that, when executed by one or more processors cause execution of operations including receiving an utterance, and identifying at least one first media content item that is currently being played. The at least one first media content item has a first facet, and the first facet has a first facet type and a first facet value. The operations may further include determining an intent and a slot from the utterance. The operations

may further include retrieving at least one second media content item from a catalog of media content items. The at least one second media content item has a second facet, and the second facet has a second facet type and a second facet value. The second facet type may be the same as the first facet type. The second facet value may be different from the first facet value. The difference may be representative of at least one of the intent and the slot of the utterance.

[0037] In certain examples, the operations may further include determining a slot type and a slot value from the slot, and determining the first facet type and the second facet type based on the slot type.

[0038] In certain examples, wherein the at least one first media content item has a plurality of facets including the first facet, and each facet has a facet type and a facet value. The at least one second media content item has a plurality of facets including the second facet, and each facet having a facet type and a facet value. In certain examples, at least some of the facet types in the at least one second media content item may be the same as at least some of the facet types in the at least one first media content item. In certain examples, the facet values, other than the first facet value, of the at least some of the facet types in the at least one second media content item may be the same as the facet values, other than the second facet value, of the at least some of the facet types in the at least one first media content item.

[0039] In certain examples, the operations may further include identifying a media content steering intent from the utterance. The media content steering intent includes a command to retrieve at least one media content item having a facet that has a different facet value from a corresponding facet of at least one currently-playing media content item.

[0040] In certain examples, the operations may further include retrieving user information associated with a user who provide the utterance. The second facet value of the at least one second media content item may be determined based on the user profile.

[0041] In certain examples, the operations may further include, prior to receiving the utterance, transmitting a media content signal to a voice-enabled device for playing the at least one first media content item. The operations may further include receiving the utterance from the voice-enabled device, and converting the utterance to a text version. The intent and the slot of the utterance may be determined from the text version.

[0042] Still another aspect is a media playback system including a media playback device operable to play media content, and a media delivery system. The media delivery system may operate to receive an utterance from the media playback device; identify at least one first media content item that is currently being played, the at least one first media content item having a first facet, the first facet having a first facet type and a first facet value; determine an intent and a slot from the utterance; retrieve at least one second media content item from a catalog of media content items, the at least one second media content item having a second facet, the second facet having a second facet type and a second facet value, the second facet type being the same as the first facet type, and the second facet value being different from the first facet value, the difference representative of at least one of the intent and the slot of the utterance; and transmit the second media content item to the media playback device.

[0043] In certain examples, the media delivery system further operates to: prior to receiving the utterance, transmit a media content signal to the media playback device for playing the first media content item; receive the utterance from the media playback device; and convert the utterance to a text version, the intent and the slot of the utterance being determined from the text version.

[0044] In certain examples, the media delivery system further operates to: determine a slot type and a slot value from the slot; and determine the first facet type and the second facet type based on the slot type.

[0045] In certain examples, the at least one first media content item has a plurality of facets including the first facet that has a facet type and a facet value. The at least one second media content item has a plurality of facets including the second facet that has a facet type and a facet value. In certain examples, at least some of the facet types in the at least one second media content item are the same as at least some of the facet types in the at least one first media content item. In certain examples, the facet values, other than the first facet value, of the at least some of the facet types in the at least one second media content item are the same as the facet values, other than the second facet value, of the at least some of the facet types in the at least one first media content item.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] The features and advantages of the example embodiments of the invention presented herein will become more apparent from the detailed description set forth below when taken in conjunction with the following drawings.

FIG. 1 illustrates an example media playback system for providing media content to a user.

FIG. 2 is a block diagram of an example embodiment of a media playback device shown in FIG. 1.

FIG. 3 is a flowchart of an example method for steering media content in response to a user query.

FIG. 4 is a block diagram that illustrates an example utterance analysis operation in the method of FIG. 3.

FIG. 5 is a block diagram that illustrates an example subject facet determination operation in the method of FIG. 3.

FIG. 6 is a block diagram that illustrates an example candidate media content determination operation in the method of FIG. 3.

FIG. 7 illustrates the candidate media content determination operation of FIG. 6.

FIG. 8 is a block diagram that illustrates an example steered media content determination operation in the method of FIG. 3.

DETAILED DESCRIPTION

[0047] Various embodiments will be described in detail with reference to the drawings, wherein like reference numerals represent like parts and assemblies throughout the several views.

[0048] The example embodiments of the invention presented herein are directed to methods, systems and computer program products for media content steering. Aspects of the embodiments identify a steering command and determine media content in accordance with that steering command, particularly media content that is relatively different from the media content being currently played or queued to be played. In some examples, the relative difference is constrained by a distance between one or more media content attributes of the media content item currently playing or queued to be played and the steering command.

[0049] A steering command is a voice command from a user to steer a session being played (sometimes referred to as a "playback session" or "listening session") or queued to be played. In other embodiments, a steering command can be provided in other formats, such as a text-based input or other tactile input.

[0050] A steering command from an utterance is identified and processed to determine what media content to playback next. An utterance of the steering command (e.g., in the form of a user query) is parsed, and a search for media content items that meets the steering command is performed.

[0051] The terms utter, utterance and speak generally refer to a spoken word, statement or vocal sound that is used to convey information. Unless a specific meaning is specified for a given use of the terms "utter", "utterance", and "speak", they should be construed to mean any of these senses of the word, all of which are understood as a plain meaning of the terms "utter", "utterance", and "speak" by those having ordinary skill in the art.

[0052] In some embodiments, an utterance of a user query is converted into a text which is then parsed into an intent and one or more slots.

[0053] An intent in the utterance describes a general intent of the text version of the utterance.

[0054] A "slot" as used herein is in reference to an utterance. A slot describes a portion of the text version of the utterance having a specific meaning. In some embodiments, a slot includes a key-value pair. The key-value pair of a slot is also sometimes referred to herein as a pair of slot type and slot value, where the slot type is the key portion of the pair and the slot value is the value portion of the pair. By way of example, where an utterance is "play more upbeat," the slot can be "more upbeat" and expressed as a key-value pair of {mood:upbeat}.

[0055] A "facet" as used herein is in reference to a media content item. A facet of a media content item describes an attribute of that media content item. In some embodiments, a facet includes a key-value pair which can include a facet type (as a key) and a facet value (as a value). In certain examples, to determine recommended media content items, the existing solutions retrieve one or more media content items from a catalog of media content items and analyze facets of the retrieved media content items based on the parsed utterance of the user query, such as the intent and/or the slot(s) of the utterance.

[0056] Examples of a facet type include information associated with the acoustic attributes (e.g., temporal information such as tempo, rhythm, beats, downbeats, tatums, patterns, sections, or other structures; and acoustic information such as melody, pitch, harmony, timbre, chroma, loudness, vocalness, or other possible features), cultural information (e.g., styles, genres, moods, themes, similar artists and/or songs, rankings, hits, opinion, popularity, activities, social signaling, deep cuts, or any other information), explicit information (e.g., song titles, artist and composer names, other credits, release date, time period, era, epoch, recording label, organizational, album cover art, locale, geographical information, instrumentation, publisher name and product number, and any other information), and any other information available for media content.

[0057] The facet values include values for the corresponding facet types. Therefore, a facet is expressed as a key-value pair between a facet type and one or more facet values, such as: {tempo:60bpm} or {genre:rock}.

[0058] In some embodiments, a media content item has a plurality of facets that describe various attributes of the media content items. Each media content item can have a plurality of facets, each of which includes a pair of facet type and facet value. In some embodiments, facets of a media content item can be obtained from metadata associated with the media content item. Facets can be obtained in other ways in other embodiments.

[0059] The steering command can be a voice command that is used to control either a media playback device or a media delivery system via the media playback device. The steering command is particularly a voice command that steers the session being played or queued to be played. The steering command causes the media playback device (and/or media delivery system as the case may be) to play back media content that is relatively different from the media content being currently played. In certain examples, the steered media content may be personalized to the user.

[0060] In a non-limiting example, a user utters the steering command "play more upbeat" to a media playback device while the media playback device is playing a first media content item. A voice recognition system, for example trained with a natural language understanding (NLU) model, analyzes the utterance ("play more up-

beat"), transcribes the utterance into text and parses the utterance into its parts, the parts of which have syntactic roles. In some embodiments, the utterance is parsed into an intent and one or more slots. In this example, the intent is identified as a play-intent type intent (e.g., "play"). A play-intent indicates an intent of a user to play back media content. Alternatively, the intent in this example is identified as a steering intent (e.g., "play more ..."). A steering intent is an intent of a user to steer playback of media content to certain direction. For example, a steering intent represents a user request for retrieving a second media content item that is relatively different from the first media content item. In the same example, the slot of the utterance "play more upbeat" is expressed as a key-value pair of {mood: upbeat} .

[0061] It should be understood that the terms utter, utterance and speak generally refer to a spoken word, statement or vocal sound that is used to convey information. Unless a specific meaning is specified for a given use of the terms "utter", "utterance", and "speak", they should be construed to mean any of these senses of the word, all of which are understood as a plain meaning of the terms "utter", "utterance", and "speak" by those having ordinary skill in the art.

[0062] The parsed utterance, which include the intent and the slots, is used to identify at least one facet of the first media content item that matches, or is relevant to, at least part of the information in the parsed utterance. Once the facet of the first media content item is identified, the second media content item is determined which has a corresponding facet with a relatively different value thereof. As such, in the example embodiments described herein, facets are evaluated in view of the parsed utterance, and used to identify the second media content item that is relatively different from the first media content item and will be steered from the first media content item.

[0063] In some embodiments, the first media content item can have multiple facets, each of which has a pair of facet type and a facet value. The second media content item, which will be steered from the first media content item, is determined by comparing at least one facet of the first media content item and at least one facet of the second media content item.

[0064] Prior to the comparison, the facets being compared between the first media content item and the second media content item can be determined. The facets being compared between the first media content item and the second media content item can be determined based at least in part on the parsed utterance. In one example embodiment, the parsed utterance includes an intent and/or a slot. An intent, as described above, describes a general intent of the text version of the utterance. In this example, a slot of an utterance determines which facet of a first media content item should be considered.

[0065] In the case where the slot includes a slot type (i.e., key) of "mood" and a slot-value of "upbeat", the slot type (i.e., "mood") causes a tempo to be considered. For example, when the media playback device (and/or media delivery system as the case may be) determines that a slot contains a slot type of "mood", a determination is made as to which attribute to process. In this case, the attribute could be, for example, a tempo. If other facets are available, the slot type of the utterance causes the tempo to be considered among other facets as well. The second media content item can then be determined based on this identified facet type (i.e., tempo). For example, the second media content item is selected to have the same facet type (e.g., tempo) as the first media content item, and further have a facet value for that facet type that is relatively different from that of the first media content item. In some exemplary embodiments, this selection for the second media content item based on the identified facet type can be performed regardless of whether the other facets of the first media content item are at least partially identical to, or different from, the corresponding facets of the second media content item. In other embodiments, the second media content item is selected such that at least one of the other facets remain the same or similar between the first media content item and the second media content item. In this example, the second media content item is determined to have a higher tempo that the tempo value of the first media content item, in order to suit the user request for "more upbeat" songs.

[0066] In certain exemplary embodiments, a playback session may be steered substantially instantaneously upon receipt of a steering command. In other exemplary embodiments, steering of the playback session can be delayed to certain point in time (e.g., until after the media content item currently being played has finished). The playback session may be executed in any number of given manners. For example, the playback session is performed solely by the media playback device, by the media playback device in conjunction with a media delivery system, or solely by the media delivery system which is capable of receiving either the steering command itself or a derivative of the steering command.

[0067] The solution of the present disclosure is similarly applicable to a situation where a media context, or any other format of media content, is to be steered based on a steering command from a user.

[0068] As such, the solution of the present application allows to a user to modify upcoming recommendations that are based on the currently playing media content by simply uttering a voice command containing comparative terms. The solution can effectively identify a steering command from the utterance of the voice command. In response to the steering command, the solution can determine a media content item that will be steered, by considering the utterance and also evaluating and comparing both facets of a currently-playing media content item and facets of media content items from a catalog. This eliminates multiple searches in media content database via networks, and provides an effective and fast solution to search for a media content item having an attribute rel-

atively different from that of the currently-playing media content item, according to the comparative terms contained in the steering command.

[0069] The solution of the present application is described and illustrated in more detail with some exemplary embodiments below.

[0070] FIG. 1 illustrates an example media playback system 100 for providing media content to a user U. The system 100 includes a media playback device 102, a media delivery system 104, and a data communication network 106. The media delivery system 104 includes a media content steering engine 110. An example user query 120 and a media output 122 are also shown. As described herein, the media output 122 can change from a first media content item 124 to a second media content item 126 when the user query 120 includes a media content steering query.

[0071] The media playback device 102 operates to provide media content to a user U. As described herein, the media playback device 102 operates to receive the user query 120 and provide the media output 122 to the user U according to the user query 120. In some embodiments, the media playback device 102 operates to communicate with a system external to the media playback device 102, such as the media delivery system 104. The media playback device 102 can interact with the media delivery system 104 to process the user query 120 and identify media content in response to the user query 120. In some embodiments, the media playback device 102 operates to receive the media content that is identified and provided (e.g., streamed, transmitted, etc.) by the media delivery system 104. In some embodiments, the media playback device 102 operates to play the media content and generate the media output 122 using a media output device (e.g., a speaker) therein. In other embodiments, the media playback device 102 operates to transmit the media content to another device for playback, such as an external speaker or another media playback device (e.g., a vehicle entertainment system or a home entertainment system). An example of the media playback device 102 is illustrated and described in more detail herein, such as with reference to FIG. 2.

[0072] The media delivery system 104 operates to provide media content to one or more media playback devices, such as the media playback device 102, via the network 106. An example of the media delivery system 104 is illustrated and described in further detail herein, such as with reference to FIG. 2.

[0073] The network 106 is a data communication network that facilitates data communication between the media playback device 102 and the media delivery system 104. The network 106 typically includes a set of computing devices and communication links between the computing devices. The computing devices in the network 106 use the links to enable communication among the computing devices in the network. The network 106 can include one or more routers, switches, mobile access points, bridges, hubs, intrusion detection devices, storage devices, standalone server devices, blade server devices, sensors, desktop computers, firewall devices, laptop computers, handheld computers, mobile telephones, vehicular computing devices, and other types of computing devices.

[0074] In various embodiments, the network 106 includes various types of communication links. For example, the network 106 can include wired and/or wireless links, including cellular, Bluetooth®, Wi-Fi®, ultra-wideband (UWB), 802.11, ZigBee, near field communication (NFC), an ultrasonic data transmission, and other types of wireless links. Furthermore, in various embodiments, the network 106 is implemented at various scales. For example, the network 106 can be implemented as one or more vehicle area networks, local area networks (LANs), metropolitan area networks, subnets, wide area networks (WAN) (such as the Internet), or can be implemented at another scale. Further, in some embodiments, the network 106 includes multiple networks, which may be of the same type or of multiple different types.

[0075] Referring still to FIG. 1, in some embodiments, the media delivery system 104 includes the media content steering engine 110 that operates to change media content of the media output 122 in response to the user query 120.

[0076] As described herein, the user query 120 may include a media content steering query from the user U. In this document, therefore, the user query 120 can be referred to a media content steering query, a steering query, or the like. The media content steering query 120 is a request for playback of media content having one or more attributes different from those of other media content. In this example, such other media content can be represented as the first media content item 124, and the media content steered from the first media content item 124 can be represented as the second media content item 126.

[0077] The media content steering query 120 can be received either when media content is currently playing, or when no media content is currently playing. In some embodiments, when the first media content item 124 is being played via the media playback device 102, the media content steering query 120 is a request for determining a media content item (e.g., the second media content item 126) that is different from the first media content item 124 in one or more attributes. Such different attributes can be determined based at least in part on information contained in the media content steering query 120. In other embodiments, when no media content item is currently played, the media content steering query 120 can refer to media content that has been previously played. For example, the media content steering query 120 can be a request for determining a media content item (e.g., the second media content item 126) that is different from the first media content item 124, which has been previously played.

[0078] In this document, the media content that is currently playing, queued to be played, or has been previ-

ously played can be represented as the first media content item 124 as illustrated in FIG. 1. In addition, the media content that will be steered from the first media content item 124 in response to the user query 120 can be represented as the second media content item 126 as illustrated in FIG. 1. The first media content item 124 and the second media content item 126 can be of various types. In some embodiments, the first media content item 124 and/or the second media content item 126 can be media content items 230, 234 (FIG. 2). Alternatively, the first media content item 124 and/or the second media content item 126 can be playlists 239 (FIG. 2). In still other embodiments, the first media content item 124 and/or the second media content item 126 can be of other types of media contexts 238.

[0079] In some embodiments, as illustrated in FIG. 1, the user query 120 can be a voice request received through a sound detection device (e.g., a microphone). As described herein, the voice request can be processed into a text query suitable for media content search. In other embodiments, the user query 120 can be a text that is typed using the media playback device 102 or another computing device.

[0080] In some embodiments, the media content steering query 120 includes one or more relative words or phrases (also referred to herein as relative terms or descriptors, or comparative words, phrases, terms, or descriptors) that can be used as one or more keywords to identify media content (e.g., the second media content item 126) to be steered from the currently-playing or previously-played media content (e.g., the first media content item 124). In the illustrated example of FIG. 1, the media content steering query 120 can be a voice request that is converted into a text query of "play something more upbeat." In this example, the relative term in the steering query 120 can include "more" or "more upbeat." By way of other examples, a media content steering query 120, such as "play something new" and "only play the popular songs," can also be considered as a media content steering query 120 as they include relative terms, such as "new" (or "something new" or their variants) and "only" (or "only ... popular" or their variants) respectively. Other examples are also possible.

[0081] The media content steering engine 110 can operate to identify the user query 120 as a media content steering query and determine media content based at least in part on the information contained in the user query 120. In the illustrated example of the user query 120 of "play something more upbeat," the media content steering engine 110 can identify the comparative term "more" (or "more upbeat") and determine the second media content item 126 having an attribute that provides a more upbeat mood than the first media content item 124. An example of the media content steering engine 110 is illustrated and described in more detail herein, such as with reference to FIG. 3.

[0082] FIG. 2 is a block diagram of an example embodiment of the media playback device 102 of the system 100 shown in FIG. 1. In this example, the media playback device 102 includes a user input device 130, a display device 132, a wireless data communication device 134, a media content output device 140, a processing device 148, and a memory device 150.

[0083] The media playback device 102 operates to play media content. For example, the media playback device 102 is configured to play media content that is provided (e.g., streamed or transmitted) by a system external to the media playback device 102, such as the media delivery system 104, another system, or a peer device. In other examples, the media playback device 102 operates to play media content stored locally on the media playback device 102. In yet other examples, the media playback device 102 operates to play media content that is stored locally as well as media content provided by other systems.

[0084] In some embodiments, the media playback device 102 is a handheld or portable entertainment device, smartphone, tablet, watch, wearable device, or any other type of computing device capable of playing media content. In other embodiments, the media playback device 102 is a laptop computer, desktop computer, television, gaming console, set-top box, network appliance, blueray or DVD player, media player, stereo, or radio.

[0085] The user input device 130 operates to receive a user input 152 from a user U for controlling the media playback device 102. As illustrated, the user input 152 can include a manual input 154 and a voice input 156. In some embodiments, the user input device 130 includes a manual input device 160 and a sound detection device 162.

[0086] The manual input device 160 operates to receive the manual input 154 for controlling playback of media content via the media playback device 102. In some embodiments, the manual input device 160 includes one or more buttons, keys, touch levers, switches, and/or other mechanical input devices for receiving the manual input 154. For example, the manual input device 160 includes a text entry interface, such as a mechanical keyboard, a virtual keyboard, or a handwriting input device, which is configured to receive a text input, such as a text version of the user query 120. In addition, in some embodiments, the manual input 154 is received for managing various pieces of information transmitted via the media playback device 102 and/or controlling other functions or aspects associated with the media playback device 102.

[0087] The sound detection device 162 operates to detect and record sounds from proximate the media playback device 102. For example, the sound detection device 162 can detect sounds including the voice input 156. In some embodiments, the sound detection device 162 includes one or more acoustic sensors configured to detect sounds proximate the media playback device 102. For example, acoustic sensors of the sound detection device 162 include one or more microphones. Various types of microphones can be used for the sound detection

device 162 of the media playback device 102.

[0088] In some embodiments, the voice input 156 is a voice of a user (also referred to herein as an utterance) for controlling playback of media content via the media playback device 102. For example, the voice input 156 includes a voice version of the user query 120 received from the sound detection device 162 of the media playback device 102. In addition, the voice input 156 is a voice of a user for managing various data transmitted via the media playback device 102 and/or controlling other functions or aspects associated with the media playback device 102.

[0089] In some embodiments, the sounds detected by the sound detection device 162 can be processed by the sound processing engine 180 of the media playback device 102 as described below.

[0090] Referring still to FIG. 2, the display device 132 operates to display information to the user U. Examples of such information include media content playback information, notifications, and other information. In some embodiments, the display screen 132 is configured as a touch sensitive display and includes the manual input device 160 of the user input device 130 for receiving the manual input 154 from a selector (e.g., a finger, stylus etc.) controlled by the user U. In some embodiments, therefore, the display screen 132 operates as both a display device and a user input device. The touch sensitive display screen 132 operates to detect inputs based on one or both of touches and near-touches. In some embodiments, the display screen 132 displays a graphical user interface for interacting with the media playback device 102. Other embodiments of the display screen 132 do not include a touch sensitive display screen. Some embodiments include a display device and one or more separate user interface devices. Further, some embodiments do not include a display device.

[0091] The data communication device 134 operates to enable the media playback device 102 to communicate with one or more computing devices over one or more networks, such as the network 106. For example, the data communication device 134 is configured to communicate with the media delivery system 104 and receive media content from the media delivery system 104 at least partially via the network 106. The data communication device 134 can be a network interface of various types which connects the media playback device 102 to the network 106. Examples of the data communication device 134 include wired network interfaces and wireless network interfaces. Wireless network interfaces includes infrared, BLUETOOTH® wireless technology, 802.11a/b/g/n/ac, and cellular or other radio frequency interfaces in at least some possible embodiments. Examples of cellular network technologies include LTE, WiMAX, UMTS, CDMA2000, GSM, cellular digital packet data (CDPD), and Mobitex.

[0092] The media content output device 140 operates to output media content. In some embodiments, the media content output device 140 generates the media output 122 for the user U. In some embodiments, the media content output device 140 includes one or more embedded speakers 164 which are incorporated in the media playback device 102.

[0093] Alternatively or in addition, some embodiments of the media playback device 102 include an external speaker interface 166 as an alternative output of media content. The external speaker interface 166 is configured to connect the media playback device 102 to another system having one or more speakers, such as headphones, a portal speaker, and a vehicle entertainment system, so that the media output 122 is generated via the speakers of the other system external to the media playback device 102. Examples of the external speaker interface 166 include an audio output jack, a USB port, a Bluetooth transmitter, a display panel, and a video output jack. Other embodiments are possible as well. For example, the external speaker interface 166 is configured to transmit a signal that can be used to reproduce an audio signal by a connected or paired device such as headphones or a speaker.

[0094] The processing device 148, in some embodiments, comprises one or more central processing units (CPU). In other embodiments, the processing device 148 additionally or alternatively includes one or more digital signal processors, field-programmable gate arrays, or other electronic circuits.

[0095] The memory device 150 typically includes at least some form of computer-readable media. The memory device 150 can include at least one data storage device. Computer readable media includes any available media that can be accessed by the media playback device 102. By way of example, computer-readable media includes computer readable storage media and computer readable communication media.

[0096] Computer readable storage media includes volatile and nonvolatile, removable and non-removable media implemented in any device configured to store information such as computer readable instructions, data structures, program modules, or other data. Computer readable storage media includes, but is not limited to, random access memory, read only memory, electrically erasable programmable read only memory, flash memory and other memory technology, compact disc read only memory, blue ray discs, digital versatile discs or other optical storage, magnetic storage devices, or any other medium that can be used to store the desired information and that can be accessed by the media playback device 102. In some embodiments, computer readable storage media is non-transitory computer readable storage media.

[0097] Computer readable communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" refers to a signal that has one or more of its characteristics set or changed in

such a manner as to encode information in the signal. By way of example, computer readable communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency, infrared, and other wireless media. Combinations of any of the above are also included within the scope of computer readable media.

[0098] The memory device 150 operates to store data and instructions. In some embodiments, the memory device 150 stores instructions for a media content cache 172, a caching management engine 174, a media playback engine 176, a sound processing engine 180, and a voice interaction engine 182.

[0099] Some embodiments of the memory device 150 include the media content cache 172. The media content cache 172 stores media content items, such as media content items that have been received from the media delivery system 104. The media content items stored in the media content cache 172 may be stored in an encrypted or unencrypted format. In some embodiments, the media content cache 172 also stores metadata about media content items such as title, artist name, album name, length, genre, mood, era, etc. The media content cache 172 can further store playback information about the media content items and/or other information associated with the media content items.

[0100] The caching management engine 174 is configured to receive and cache media content in the media content cache 172 and manage the media content stored in the media content cache 172. In some embodiments, when media content is streamed from the media delivery system 104, the caching management engine 174 operates to cache at least a portion of the media content into the media content cache 172. In other embodiments, the caching management engine 174 operates to cache at least a portion of media content into the media content cache 172 while online so that the cached media content is retrieved for playback while the media playback device 102 is offline.

[0101] The media playback engine 176 operates to play media content to the user U. As described herein, the media playback engine 176 is configured to communicate with the media delivery system 104 to receive one or more media content items (e.g., through the media stream 232). In other embodiments, the media playback engine 176 is configured to play media content that is locally stored in the media playback device 102.

[0102] In some embodiments, the media playback engine 176 operates to retrieve one or more media content items that are either locally stored in the media playback device 102 or remotely stored in the media delivery system 104. In some embodiments, the media playback engine 176 is configured to send a request to the media delivery system 104 for media content items and receive information about such media content items for playback.

[0103] The sound processing engine 180 is configured to receive sound signals obtained from the sound detection device 162 and process the sound signals to identify different sources of the sounds received via the sound detection device 162. In some embodiments, the sound processing engine 180 operates to filter the voice input 156 (e.g., a voice request of the user query 120) from noises included in the detected sounds. Various noise cancellation technologies, such as active noise control or canceling technologies or passive noise control or cancelling technologies, can be used to filter the voice input from ambient noise. In examples, the sound processing engine 180 filters out omni-directional noise and preserves directional noise (e.g., an audio input difference between two microphones) in audio input. In examples, the sound processing engine 180 removes frequencies above or below human speaking voice frequencies. In examples, the sound processing engine 180 subtracts audio output of the device from the audio input to filter out the audio content being provided by the device, (e.g., to reduce the need of the user to shout over playing music). In examples, the sound processing engine 180 performs echo cancellation. By using one or more of these techniques, the sound processing engine 180 provides sound processing customized for use in a vehicle environment.

[0104] In other embodiments, the sound processing engine 180 operates to process the received sound signals to identify the sources of particular sounds of the sound signals, such as people's conversation in the vehicle, the vehicle engine sound, or other ambient sounds associated with the vehicle.

[0105] In some embodiments, the sound processing engine 180 at least partially operates to analyze a recording of sounds captured using the sound detection device 162, using speech recognition technology to identify words spoken by the user. In addition or alternatively, other computing devices, such as the media delivery system 104 (e.g., a voice interaction server 204 thereof) can cooperate with the media playback device 102 for such analysis. The words may be recognized as commands from the user that alter the playback of media content and/or other functions or aspects of the media playback device 102. In some embodiments, the words and/or the recordings may also be analyzed using natural language processing and/or intent recognition technology to determine appropriate actions to take based on the spoken words. Additionally or alternatively, the sound processing engine 180 may determine various sound properties about the sounds proximate the media playback device 102 such as volume, dominant frequency or frequencies, etc. These sound properties may be used to make inferences about the environment proximate to the media playback device 102.

[0106] The voice interaction engine 182 operates to cooperate with the media delivery system 104 (e.g., a voice interaction server 204 thereof) to identify a command (e.g., a user intent) that is conveyed by the voice input 156. In some embodiments, the voice interaction engine 182 transmits the voice input 156 (e.g., of a user) that is detected by the sound processing engine 180 to

the media delivery system 104 so that the media delivery system 104 operates to determine a command intended by the voice input 156. In other embodiments, at least some of the determination process of the command can be performed locally by the voice interaction engine 182.

[0107] In addition, some embodiments of the voice interaction engine 182 can operate to cooperate with the media delivery system 104 (e.g., the voice interaction server 204 thereof) to provide a voice assistant that performs various voice-based interactions with the user, such as voice feedbacks, voice notifications, voice recommendations, and other voice-related interactions and services.

[0108] Referring still to FIG. 2, the media delivery system 104 includes a media content server 200, a media content search server 202, a voice interaction server 204, and a user command interpretation server 206.

[0109] The media delivery system 104 comprises one or more computing devices and provides media content to the media playback device 102 and, in some embodiments, other media playback devices as well. In addition, the media delivery system 104 interacts with the media playback device 102 to provide the media playback device 102 with various functionalities.

[0110] In at least some embodiments, the media content server 200, the media content search server 202, the voice interaction server 204, and the user command interpretation server 206 are provided by separate computing devices. In other embodiments, the media content server 200, the media content search server 202, the voice interaction server 204, and the user command interpretation server 206 are provided by the same computing device(s). Further, in some embodiments, at least one of the media content server 200, the media content search server 202, the voice interaction server 204, and the user command interpretation server 206 is provided by multiple computing devices. For example, the media content server 200, the media content search server 202, the voice interaction server 204, and the user command interpretation server 206 may be provided by multiple redundant servers located in multiple geographic locations.

[0111] Although FIG. 2 shows a single media content server 200, a single media content search server 202, a single voice interaction server 204, and a single user command interpretation server 206, some embodiments include multiple media content servers, media content search servers, voice interaction servers, and user command interpretation servers. In these embodiments, each of the multiple media content servers, media content search servers, voice interaction servers, and user command interpretation servers may be identical or similar to the media content server 200, the media content search server 202, the voice interaction server 204, and the user command interpretation server 206, respectively, as described herein, and may provide similar functionality with, for example, greater capacity and redundancy and/or services from multiple geographic locations. Alternatively, in these embodiments, some of the multiple media content servers, the media content search servers, the voice interaction servers, and/or the user command interpretation servers may perform specialized functions to provide specialized services. Various combinations thereof are possible as well.

[0112] The media content server 200 transmits stream media to media playback devices such as the media playback device 102. In some embodiments, the media content server 200 includes a media server application 212, a processing device 214, a memory device 216, and a data communication device 218. The processing device 214 and the memory device 216 may be similar to the processing device 148 and the memory device 150, respectively, which have each been previously described. Therefore, the description of the processing device 214 and the memory device 216 are omitted for brevity purposes.

[0113] The data communication device 218 operates to communicate with other computing devices over one or more networks, such as the network 106. Examples of the data communication device include one or more wired network interfaces and wireless network interfaces. Examples of such wireless network interfaces of the data communication device 218 include wireless wide area network (WWAN) interfaces (including cellular networks) and wireless local area network (WLANs) interfaces. In other examples, other types of wireless interfaces can be used for the data communication device 218.

[0114] In some embodiments, the media server application 212 is configured to stream media content, such as music or other audio, video, or other suitable forms of media content. The media server application 212 includes a media stream service 222, a media application interface 224, and a media data store 226. The media stream service 222 operates to buffer media content, such as media content items 230A, 230B, and 230N (collectively 230), for streaming to one or more media streams 232A, 232B, and 232N (collectively 232).

[0115] The media application interface 224 can receive requests or other communication from media playback devices or other systems, such as the media playback device 102, to retrieve media content items from the media content server 200. For example, in FIG. 2, the media application interface 224 receives communication from the media playback device 102 to receive media content from the media content server 200.

[0116] In some embodiments, the media data store 226 stores media content items 234, media content metadata 236, media contexts 238, user accounts 240, and taste profiles 242. The media data store 226 may comprise one or more databases and file systems. Other embodiments are possible as well.

[0117] As described herein, the media content items 234 (including the media content items 230) may be audio, video, or any other type of media content, which may be stored in any format for storing media content.

[0118] The media content metadata 236 provides various information associated with the media content items 234. In addition or alternatively, the media content metadata 236 provides various information associated with the media contexts 238. In some embodiments, the media content metadata 236 includes one or more of title, artist name, album name, length, genre, mood, era, etc.

[0119] The media content metadata 236 operates to provide various pieces of information (also referred to herein as attribute(s)) associated with the media content items 234 and/or the media contexts 238. In some embodiments, the media content metadata 236 includes one or more of title, artist name, album name, length, genre, mood, era, etc.

[0120] In some embodiments, the media content metadata 236 includes acoustic metadata, cultural metadata, and explicit metadata. The acoustic metadata may be derived from analysis of the track and refers to a numerical or mathematical representation of the sound of a track. Acoustic metadata may include temporal information such as tempo, rhythm, beats, downbeats, tatums, patterns, sections, or other structures. Acoustic metadata may also include spectral information such as melody, pitch, harmony, timbre, chroma, loudness, vocalness, or other possible features. Acoustic metadata may take the form of one or more vectors, matrices, lists, tables, and other data structures. Acoustic metadata may be derived from analysis of the music signal. One form of acoustic metadata, commonly termed an acoustic fingerprint, may uniquely identify a specific track. Other forms of acoustic metadata may be formed by compressing the content of a track while retaining some or all of its musical characteristics.

[0121] The cultural metadata refers to text-based information describing listeners' reactions to a track or song, such as styles, genres, moods, themes, similar artists and/or songs, rankings, etc. Cultural metadata may be derived from expert opinion such as music reviews or classification of music into genres. Cultural metadata may be derived from listeners through websites, chatrooms, blogs, surveys, and the like. Cultural metadata may include sales data, shared collections, lists of favorite songs, and any text information that may be used to describe, rank, or interpret music. Cultural metadata may also be generated by a community of listeners and automatically retrieved from Internet sites, chat rooms, blogs, and the like. Cultural metadata may take the form of one or more vectors, matrices, lists, tables, and other data structures. A form of cultural metadata particularly useful for comparing music is a description vector. A description vector is a multi-dimensional vector associated with a track, album, or artist. Each term of the description vector indicates the probability that a corresponding word or phrase would be used to describe the associated track, album or artist.

[0122] The explicit metadata refers to factual or explicit information relating to music. Explicit metadata may include album and song titles, artist and composer names, other credits, album cover art, publisher name and product number, and other information. Explicit metadata is generally not derived from the music itself or from the reactions or opinions of listeners.

[0123] At least some of the metadata 236, such as explicit metadata (names, credits, product numbers, etc.) and cultural metadata (styles, genres, moods, themes, similar artists and/or songs, rankings, etc.), for a large library of songs or tracks can be evaluated and provided by one or more third party service providers. Acoustic and cultural metadata may take the form of parameters, lists, matrices, vectors, and other data structures. Acoustic and cultural metadata may be stored as XML files, for example, or any other appropriate file type. Explicit metadata may include numerical, text, pictorial, and other information. Explicit metadata may also be stored in an XML or other file. All or portions of the metadata may be stored in separate files associated with specific tracks. All or portions of the metadata, such as acoustic fingerprints and/or description vectors, may be stored in a searchable data structure, such as a k-D tree or other database format.

[0124] Referring still to FIG. 2, each of the media contexts 238 is used to identify one or more media content items 234. In some embodiments, the media contexts 238 are configured to group one or more media content items 234 and provide a particular context to the group of media content items 234. Some examples of the media contexts 238 include albums, artists, playlists, and individual media content items. By way of example, where a media context 238 is an album, the media context 238 can represent that the media content items 234 identified by the media context 238 are associated with that album.

[0125] As described above, the media contexts 238 can include playlists 239. The playlists 239 are used to identify one or more of the media content items 234. In some embodiments, the playlists 239 identify a group of the media content items 234 in a particular order. In other embodiments, the playlists 239 merely identify a group of the media content items 234 without specifying a particular order. Some, but not necessarily all, of the media content items 234 included in a particular one of the playlists 239 are associated with a common characteristic such as a common genre, mood, or era.

[0126] In some embodiments, a user can listen to media content items in a playlist 239 by selecting the playlist 239 via a media playback device, such as the media playback device 102. The media playback device then operates to communicate with the media delivery system 104 so that the media delivery system 104 retrieves the media content items identified by the playlist 239 and transmits data for the media content items to the media playback device for playback.

[0127] In some embodiments, the playlist 239 includes one or more playlist descriptions. The playlist descriptions include information associated with the playlist 239. The playlist descriptions can include a playlist title. In some embodiments, the playlist title can be provided by

a user using the media playback device 102. In other embodiments, the playlist title can be provided by a media content provider (or a media-streaming service provider). In yet other embodiments, the playlist title can be automatically generated.

[0128] Other examples of playlist descriptions include a descriptive text. The descriptive text can be provided by the user and/or the media content provider, which is to represent the corresponding playlist 239. In other embodiments, the descriptive text of the playlist description can be obtained from one or more other sources. Such other sources can include expert opinion (e.g., music reviews or classification of music into genres), user opinion (e.g., reviews through websites, chatrooms, blogs, surveys, and the like), statistics (e.g., sales data), shared collections, lists of favorite playlists, and any text information that may be used to describe, rank, or interpret the playlist or music associated with the playlist. In some embodiments, the playlist descriptions can also be generated by a community of listeners and automatically retrieved from Internet sites, chat rooms, blogs, and the like.

[0129] In some embodiments, the playlist descriptions can take the form of one or more vectors, matrices, lists, tables, and other data structures. A form of cultural metadata particularly useful for comparing music is a description vector. A description vector is a multi-dimensional vector associated with a track, album, or artist. Each term of the description vector indicates the probability that a corresponding word or phrase would be used to describe the associated track, album or artist. Each term of the description vector indicates the probability that a corresponding word or phrase would be used to describe the associated track, album or artist.

[0130] In some embodiments, the playlist 239 includes a list of media content item identifications (IDs). The list of media content item identifications includes one or more media content item identifications that refer to respective media content items 234. Each media content item is identified by a media content item ID and includes various pieces of information, such as a media content item title, artist identification (e.g., individual artist name or group name, or multiple artist names or group names), and media content item data. In some embodiments, the media content item title and the artist ID are part of the media content metadata 236, which can further include other attributes of the media content item, such as album name, length, genre, mood, era, etc. as described herein.

[0131] At least some of the playlists 239 may include user-created playlists. For example, a user of a media streaming service provided using the media delivery system 104 can create a playlist 239 and edit the playlist 239 by adding, removing, and rearranging media content items in the playlist 239. A playlist 239 can be created and/or edited by a group of users together to make it a collaborative playlist. In some embodiments, user-created playlists can be available to a particular user only, a group of users, or to the public based on a user-definable privacy setting.

[0132] In some embodiments, when a playlist is created by a user or a group of users, the media delivery system 104 operates to generate a list of media content items recommended for the particular user or the particular group of users. In some embodiments, such recommended media content items can be selected based at least on the taste profiles 242 as described herein. Other information or factors can be used to determine the recommended media content items. Examples of determining recommended media content items are described in U.S. Patent Application Serial No. 15/858,377, titled MEDIA CONTENT ITEM RECOMMENDATION SYSTEM, filed December 29, 2017, the disclosure of which is hereby incorporated by reference in its entirety.

[0133] In addition or alternatively, at least some of the playlists 239 are created by a media streaming service provider. For example, such provider-created playlists can be automatically created by the media delivery system 104. In some embodiments, a provider-created playlist can be customized to a particular user or a particular group of users. By way of example, a playlist for a particular user can be automatically created by the media delivery system 104 based on the user's listening history (e.g., the user's taste profile) and/or listening history of other users with similar tastes. In other embodiments, a provider-created playlist can be configured to be available for the public in general. Provider-created playlists can also be sharable with other users.

[0134] The user accounts 240 are used to identify users of a media streaming service provided by the media delivery system 104. In some embodiments, a user account 240 allows a user to authenticate to the media delivery system 104 and enable the user to access resources (e.g., media content items, playlists, etc.) provided by the media delivery system 104. In some embodiments, the user can use different devices to log into the user account and access data associated with the user account in the media delivery system 104. User authentication information, such as a username, an email account information, a password, and other credentials, can be used for the user to log into his or her user account. It is noted that, where user data is to be protected, the user data is handled according to robust privacy and data protection policies and technologies. For instance, whenever personally identifiable information and any other information associated with users is collected and stored, such information is managed and secured using security measures appropriate for the sensitivity of the data. Further, users can be provided with appropriate notice and control over how any such information is collected, shared, and used.

[0135] The taste profiles 242 contain records indicating media content tastes of users. A taste profile can be associated with a user and used to maintain an in-depth understanding of the music activity and preference of that user, enabling personalized recommendations, taste profiling and a wide range of social music applications. Libraries and wrappers can be accessed to create taste

profiles from a media library of the user, social website activity and other specialized databases to obtain music preferences.

[0136] In some embodiments, each taste profile 242 is a representation of musical activities, such as user preferences and historical information about the users' consumption of media content, and can include a wide range of information such as artist plays, song plays, skips, dates of listen by the user, songs per day, playlists, play counts, start/stop/skip data for portions of a song or album, contents of collections, user rankings, preferences, or other mentions received via a client device, or other media plays, such as websites visited, book titles, movies watched, playing activity during a movie or other presentations, ratings, or terms corresponding to the media, such as "comedy," etc.

[0137] In addition, the taste profiles 242 can include other information. For example, the taste profiles 242 can include libraries and/or playlists of media content items associated with the user. The taste profiles 242 can also include information about the user's relationships with other users (e.g., associations between users that are stored by the media delivery system 104 or on a separate social media site).

[0138] The taste profiles 242 can be used for a number of purposes. One use of taste profiles is for creating personalized playlists (e.g., personal playlisting). An API call associated with personal playlisting can be used to return a playlist customized to a particular user. For example, the media content items listed in the created playlist are constrained to the media content items in a taste profile associated with the particular user. Another example use case is for event recommendation. A taste profile can be created, for example, for a festival that contains all the artists in the festival. Music recommendations can be constrained to artists in the taste profile. Yet another use case is for personalized recommendation, where the contents of a taste profile are used to represent an individual's taste. This API call uses a taste profile as a seed for obtaining recommendations or playlists of similar artists. Yet another example of taste profile use case is referred to as bulk resolution. A bulk resolution API call is used to resolve taste profile items to pre-stored identifiers associated with a service, such as a service that provides metadata about items associated with the taste profile (e.g., song tempo for a large catalog of items). Yet another example use case for taste profiles is referred to as user-to-user recommendation. This API call is used to discover users with similar tastes by comparing the similarity of taste profile item(s) associated with users.

[0139] A taste profile 242 can represent a single user or multiple users. Conversely, a single user or entity can have multiple taste profiles 242. For example, one taste profile can be generated in connection with a user's media content play activity, whereas another separate taste profile can be generated for the same user based on the user's selection of media content items and/or artists for a playlist.

[0140] Referring still to FIG. 2, the media content search server 202 operates to perform media content search in response to a media content search request, such as the user query 120 (FIG. 1). In some embodiments, the media content search server 202 includes a media content search application 250, a processing device 252, a memory device 254, and a data communication device 256. The processing device 252, the memory device 254, and the data communication device 256 may be similar to the processing device 214, the memory device 216, and the data communication device 218, respectively, which have each been previously described.

[0141] In some embodiments, the media content search application 250 operates to interact with the media playback device 102 and provide selection of one or more media content items based on the user query 120. As described herein, the media content search application 250 can include the media content steering engine 110 (FIG. 1). The media content search application 250 can interact with other servers, such as the media content server 200, the voice interaction server 204, and the user command interpretation server 206, to perform media content search.

[0142] Referring still to FIG. 2, the voice interaction server 204 operates to provide various voice-related functionalities to the media playback device 102. In some embodiments, the voice interaction server 204 includes a voice recognition application 270, a speech synthesis application 272, a processing device 274, a memory device 276, and a data communication device 278. The processing device 274, the memory device 276, and the data communication device 278 may be similar to the processing device 214, the memory device 216, and the data communication device 218, respectively, which have each been previously described.

[0143] In some embodiments, the voice recognition application 270 and the speech synthesis application 272, either individually or in combination, operate to interact with the media playback device 102 and enable the media playback device 102 to perform various voice-related functions, such as voice media content search, voice feedback, voice notifications, etc.

[0144] In some embodiments, the voice recognition application 270 is configured to perform speech-to-text (STT) conversion, such as receiving a recording of voice command (e.g., an utterance) and converting the utterance to a text format.

[0145] In some embodiments, the speech synthesis application 272 is configured to perform text-to-speech (TTS) conversion, so that a language text is converted into speech. Then, the voice interaction server 204 can transmit an audio data or file for the speech to the media playback device 102 so that the media playback device 102 generates a voice assistance to the user using the transmitted audio data or file.

[0146] Referring still to FIG. 2, the user command interpretation server 206 operates to analyze the user command (e.g., the utterance) to determine appropriate ac-

tions to take according to the user command. In some embodiments, the user command interpretation server 206 analyzes a text version of a user command (e.g., a text version of the utterance). In other embodiments, a recording of the user command can be used for such analysis without converting into a text format.

[0147] In some embodiments, the user command interpretation server 206 includes a natural language understanding (NLU) application 280, a processing device 282, a memory device 284, and a data communication device 286. The processing device 282, the memory device 284, and the data communication device 286 may be similar to the processing device 214, the memory device 216, and the data communication device 218, respectively, which have each been previously described.

[0148] In some embodiments, the NLU application 280 operates to analyze the text format of the utterance to determine functions to perform based on the utterance. The NLU application 280 can use a natural language understanding algorithm that involves modeling human reading comprehension, such as parsing and translating an input according to natural language principles.

[0149] FIG. 3 is a flowchart of an example method 300 for steering media content in response to a user query 120. The method 300 is described with further reference to FIGS. 4-8. In at least some embodiment, the method 300 can be performed at least partially by the system 100 as described herein. For example, the method 300 is at least partially performed by the media delivery system 104 that interacts with the media playback device 102. In other embodiments, the method 300 can be at least partially performed by the media playback device 102 that interacts with the media delivery system 104. In yet other embodiments, the method 300 can be at least partially performed by the media playback device 102 and the media delivery system 104.

[0150] At operation 302, the system 100 operates to identify the first media content item 124. As described herein, the first media content item 124 includes media content that is currently playing via, for example, the media playback device 102. In alternative embodiments, where there is no media content that is currently playing, the first media content item 124 that the system 100 can operate to identify is the media content that has been previously played and is currently paused or stopped. Other media content can be the first media content item 124 in other embodiments.

[0151] The identification of the first media content item 124 playing media content can be performed in various manners. In some embodiments, the media delivery system 104 can receive a media content identifier (e.g., a media content identifier 198 in FIG. 2) from the media playback device 102, which identifies the media content item that is currently playing on the media playback device 102 or that has previously played on the media playback device 102. In some embodiments, such a media content identifier can be stored in the media content cache 172 (FIG. 2) of the media playback device 102, and transmitted to the media delivery system 104. In other embodiments, with or without the media content identifier, the media delivery system 104 can retrieve, from one or more databases thereof, playback information about the media content that is currently being streamed to the media playback device 102 or that has been streamed to the media playback device 102.

[0152] At operation 304, the system 100 receives the utterance of the user query 120. In some embodiments, the user query 120 is received through the media playback device 102. As illustrated in FIG. 4, the utterance of the user query 120 can be stored and received as utterance data 402 by the media delivery system 104. In the illustrated example, the utterance of the user query 120 includes "play something more upbeat."

[0153] The utterance data 402 is data describing the utterance of the user query 120. In at least some embodiments, the utterance data 402 is an audio recording that contains the utterance being spoken. In some examples, the utterance data 402 is received as an entire audio data file. For instance, the media playback device 102 buffers the utterance data 402 as it is obtained from the sound detection device 162. The buffered utterance data 402 is then sent to the media delivery system 104 for processing. In other instances, the media playback device 102 streams the utterance data 402 to the media delivery system 104 in real-time as the utterance data 402 is received from the sound detection device 162 of the media playback device 102. In some examples, the utterance data 402 is stored (e.g., by the media delivery system 104) in a data store after it is received.

[0154] In some embodiments, the media playback device 102 receives the user query 120 from user U while media content is being currently played. In other embodiments, the media playback device 102 receives the user query 120 when media content has been played and now paused or stopped. The user U may want to listen to other media content with different attribute than the currently-playing or previously-played media content. In the illustrated example, the user U wants to listen to media content that is more upbeat than the first media content item 124 (e.g., the currently-playing or previously-played media content), and thus provides the user query 120 of "play something more upbeat."

[0155] At operation 306, the system 100 operates to process the utterance of the user query 120. In some embodiments, the operation 306 can be performed by the media delivery system 104. Referring to FIG. 4, some embodiments of the operation 306 are described in which the media delivery system 104 in the system 100 processes the utterance data 402.

[0156] As illustrated in FIG. 4, the operation 306 can include performing automated speech recognition on the utterance data 402 to obtain text data 404. In some embodiments, the voice recognition application 270 in the voice interaction server 204 can receive and process the utterance data 402 to generate the text data 404.

[0157] In some embodiments, performing automated

speech recognition includes providing the utterance data 402 as input to an automated speech recognition system (e.g., the voice recognition application 270 of the voice interaction server 204) and receiving the text data 404 as output from the automated speech recognition system. Automated speech recognition can be performed using any of a variety of techniques (e.g., using hidden Markov models or neural networks). Examples of automated speech recognition systems include CMU SPHINX, maintained by CARNEGIE MELLON UNIVERSITY, and DEEPSPEECH, maintained by the MOZILLA FOUNDATION.

[0158] Once the text data 404 is obtained after processing the utterance data 402, the text data 404 is further processed to generate structured interpretation data 406. The structured interpretation data 406 include information that interprets the utterance of the user query 120.

[0159] In some embodiments, the structured interpretation data 406 is obtained by the natural language understanding (NLU) system (e.g., the NLU application 280). The structured interpretation data 406 is used to perform a fulfillment strategy based on a received utterance. As described herein, the media delivery system 104 can include the user command interpretation server 206 that operates the NLU application 280, and the media playback device 102 can receive an utterance of the user query 120 and provide the utterance to the media delivery system 104 for processing at the user command interpretation server 206. This is for example purposes only, and other configurations are possible. For instance, the NLU application 280 is locally saved and performed in the media playback device 102.

[0160] The structured interpretation data 406 include an intent 432 and a slot 434 that are determined from the text data 404. The intent 432 describes a general intent of the text data 404. The slot 434 is a key-value pair that describes a portion of the text data 404 having a specific meaning. In some embodiments, the slot 434 can be obtained from at least a portion of the text data 404 representing the utterance other than the intent 432. In other embodiments, the slot 434 can be obtained from at least a portion of the entire text data 404 (including the intent 432). Such a key-value pair of the slot 434 includes a slot type 442 (as a key) and a slot value 444 (as a value).

[0161] In the illustrated example herein, if the text data 404 represents the utterance of "play something more upbeat" as input, the intent 432 is "play" and the slot 434 can include "something more upbeat," "more upbeat," "upbeat," or any of their variants.

[0162] Further, the slot 434 can include at least one key-value pair (e.g., at least one pair of slot type 442 and slot value 444). In the illustrated example, the slot 434 can be a key-value pair of {tempo_type: upbeat}. Other variants are possible for the key-value pair of the slot 434, such as {mood: upbeat}, in other embodiments.

[0163] In some embodiments, the text data 404 can include a single slot 434 and a single intent 432. In other embodiments, the NLU application 280 can return a plurality of slots 434 and/or a plurality of intents 432 from the text data 404. In yet other embodiments, the text data 404 provides an intent 432 but no slot 434. For example, where the text data 404 is "play," the NLU application 280 can return the intent 432 being "play", but will not result in any slot 434 (e.g., the text data 404 does not include a description of what to play). In other examples, the text data 404 can include one or more slots 434 but no intent. For instance, where the text data 404 is "All Along the Watchtower by Jimi Hendrix," the operation 430 can result in two slots 434 (e.g., {Song: All Along the Watchtower} and {Artist: Jimi Hendrix}) but no intent 432 (e.g., the text data 404 does not include a description of what do to with the song and artist, such as search, play, or save).

[0164] Referring back to FIG. 3, at operation 308, the system 100 operates to identify a media content steering intent from the utterance of the user query 120. In some embodiments, as illustrated in FIG. 4, steering intent data 450 is generated to indicate the media content steering intent. The media content steering intent is a user's intent identified from the utterance of the user query 120 that requests for other media content with one or more attributes different from media content that is currently playing or has previously been played (e.g., the first media content item 124 in FIG. 1). In this regard, when the user query 120 includes such a media content steering intent, the user query 120 is determined as a media content steering query or request. As described herein, the media content steering intent is identified from the utterance of the user query that includes a command to retrieve media content having a facet different from a corresponding facet of the currently-playing or previously-played media content.

[0165] In some embodiments, the media content steering intent is identified when the user query 120 includes one or more words having a relative meaning or the like which can be used to compare between two or more media content items or contexts. Some examples of such relative words include any terms or expressions including comparative words (e.g., "new," "newer," "newest," "old," "older," "oldest," etc.), quantifying words (e.g., "all," "enough," "only," "more," "most," "less," "least," "no," "none," "not any," "some," "any," "a lot of," "lots of," "plenty of," "much," "a little/little/very little," "a great deal of," "a large amount/number/quantity of," "many," "a few/few/very few," "a number (of)," "several," "a majority of," etc.), and any other words or phrases which can convey relative meaning.

[0166] The relative words in the utterance can provide values of one or more of the slots 434 identified from the text data 404. In the illustrated example where the text data 404 includes "play something more upbeat" converted from the utterance of the user query 120, the values of the slots 434, such as "more" or "more upbeat" can be identified as relative words of the text data 404. In other embodiments, relative words of the text data 404

can be identified from the value of the intent 432 that is determined as "play... more" or "play ... more upbeat" in the above example. In yet other embodiments, relative words of the text data 404 can be identified in other methods.

[0167] As illustrated in FIG. 4, in some embodiments, the NLU application 280 in the media delivery system 104 can operate to identify the media content steering intent and generate the steering intent data 450 from the utterance of the user query 120. In some embodiments, the media content steering intent is identified from the structured interpretation data 406. For example, the media content steering intent is identified by determining whether at least one of the intent 432 and the slot 434 includes one or more relative words which can be used to compare between attributes of two or more media content items or contexts.

[0168] Referring back to FIG. 3, at operation 310, the system 100 operates to identify a facet 502A (FIG. 5) of the first media content item 124 based on the utterance of the user query 120. The operation 310 is described with further reference to FIG. 5. As illustrated in FIG. 5, the operation 310 can be performed by the media delivery system 104. In some embodiments, the operation 310 can be performed at least by the media content search application 250 of the media content search server 202 in the media delivery system 104, as shown in FIG. 2. For example, the media content steering engine 110 of the media content search application 250 can be configured to at least partially perform the operation 310. In the illustrated example of FIG. 5, a subject facet determination engine 500 operates to perform the operation 310. The subject facet determination engine 500 can be included in the media delivery system 104, such as in the media content steering engine 110 of the media content search application 250.

[0169] A facet 502 (including 502A and 502B) of media content includes information about an attribute of the media content. In some embodiments, a media content item has a plurality of facets 502 that describe various attributes of the media content items, such as attributes as identified in the media content metadata 236 (FIG. 2). For example, the facets 502 include information that is similar to various pieces of metadata as described with reference to FIG. 2. For example, the facets 502 can include information associated with the acoustic metadata (e.g., temporal information such as tempo, rhythm, beats, downbeats, tatums, patterns, sections, or other structures; and acoustic metadata such as melody, pitch, harmony, timbre, chroma, loudness, vocalness, or other possible features), cultural metadata (e.g., styles, genres, moods, themes, similar artists and/or songs, rankings, etc.), and/or explicit metadata (e.g., song titles, artist and composer names, other credits, album cover art, publisher name and product number, and other information). In some embodiments, the facets 502 of a media content item can be obtained from metadata 236 associated with the media content item. The facets 502 can be obtained in other ways in other embodiments.

[0170] In some embodiments, as illustrated in FIG. 5, each facet 502 includes a key-value pair which can include a facet type 504 (as a key) and a facet value 506 (as a value). Each media content item, such as the first media content item 124 and the second media content item 126, has a plurality of facets 502, each of which includes a pair of facet type 504 and facet value 506. Examples of the facet types 504 include a media content ID, title, artist, tempo, genre, activity, release date, popularity, mood, geographic information, hits, deep cut, and any other available attributes. Further examples of the facet types include any types of attributes as found in the media content metadata 236 (FIG. 2) and other available sources as described herein. The facet values 506 include values for the corresponding facet types 504 that are associated with the particular media content item.

[0171] In some embodiments, the subject facet determination engine 500 can receive facet data 510 associated with the first media content item 124. The facet data 510 include information about at least some of the facets 502A of the first media content item 124. Further, the subject facet determination engine 500 can receive information about the utterance of the user query 120 and use the information to identify at least one subject facet 520 from the facet data 510. The subject facet 520 can be selected from the facets 502A of the first media content item 124 and corresponds to the information of the utterance of the user query 120 (e.g., the slot 434, such as at least one of the slot type 442 and the slot value 444, as described below). The subject facet 520 is a facet 502A that needs to be referred to in order to determine the second media content item 126 according to the user query 120.

[0172] In some embodiments, the subject facet determination engine 500 can use the structured interpretation data 406 to identify at least one of the facets 502A (i.e., the subject facet 520) of the first media content item 124 that corresponds to the steering intent that is contained and identified in the structured interpretation data 406. For example, the slot 434 obtained from the structured interpretation data 406 can be used to determine a facet 502A of the first media content item 124 that will be used to determine the second media content item 126 that may appropriately reflects the steering intent in the utterance of the user query 120. Therefore, the subject facet 520 of the first media content item 124 is selected from the facets 502A of the first media content item 124 and corresponds to the information of the slot 434, such as at least one of the slot type 442 and the slot value 444.

[0173] In addition or alternatively, other information, such as the intent 432 and/or the steering intent data 450, can be used to determine a facet 502 that is need to determine the second media content item 126.

[0174] In the illustrated example where the utterance of the user query 120 includes "play something more upbeat," the structured interpretation data 406 includes the slot 434 of "something more upbeat." This slot 434 is

represented by a pair of slot type 442 and slot value 444. In the illustrated example of FIG. 6, the pair of slot type 442 and slot value 444 is {tempo_type:upbeat}. The slot type 442 can be used to identify the facet type 504 of the first media content item 124 that is needed for the media content steering process. In this example, as the slot type 442 is "tempo_type," the corresponding facet type 504 of "tempo" can be selected. Then, the facet value 506 for "tempo" in the first media content item 124 is found to be 130 bpm. Based on the identified facet type and value, the subject facet 520 can be determined as {tempo: 130 bpm} in this example.

[0175] Referring back to FIG. 3, the method 300 can continue at operation 312. The operation 312 is described with further reference to FIGS. 6-8.

[0176] At the operation 312, the system 100 operates to determine a second media content item 126 having a facet 502B that is relatively different from the corresponding facet 502A (i.e., the subject facet 520) of the first media content item 124. Such difference between the facet 502A (i.e., the subject facet 520) of the first media content item 124 and the corresponding facet 502B of the second media content item 126 can be determined based on the media content steering intent in the utterance of the user query 120. As described herein, the second media content item 126 is selected to be personalized to the particular user (e.g., the user U who provided the user query 120).

[0177] In some embodiments, the operation 312 includes a first operation 320 (FIG. 6). In addition, the operation 312 can further include a second operation 322 (FIG. 8). The first operation 320 is described with reference to FIGS. 6 and 7, and the second operation 322 is described with reference to FIG. 8. Although it is primarily described that the first operation 320 is performed prior to the second operation 322, it is understood that the second operation 322 can be performed before the first operation 320, or that the first operation 320 and the second operation 322 can be combined and performed as a single step.

[0178] The operation 312 (including the operations 320 and 322) can be performed by the media delivery system 104. In some embodiments, the operation 312 can be performed at least by the media content search application 250 of the media content search server 202 in the media delivery system 104, as shown in FIG. 2. For example, the media content steering engine 110 of the media content search application 250 can be configured to at least partially perform the operation 312. In the illustrated example of FIG. 6, a candidate media content determination engine 550 operates to at least part of the operation 312, such as the first operation 320. In the illustrated example of FIG. 8, a steered media content determination engine 590 operates to at least part of the operation 312, such as the second operation 322.

[0179] Referring to FIG. 6, at the first operation 320, the system 100 operates to determine one or more candidate media content items 552 (including 552A, 552B, 552C, etc.), from which the second media content item 126 can be selected. In some embodiments, the candidate media content determination engine 550 operates to receive data of the subject facet 520 and the information (e.g., the structured interpretation data 406) about the utterance of the user query 120. The candidate media content determination engine 550 further operates to retrieve a media content database 554 (e.g., the media data store 226 in FIG. 2), which stores a plurality of media content items 556 (e.g., the media content items 234 in FIG. 2), and select the candidate media content items 552 from the media content database 554 based on the subject facet 520 and the utterance of the user query 120.

[0180] In some embodiments, each of the candidate media content items 552 is selected to have a facet 502B that corresponds to the subject facet 520 and has a different value from the subject facet 520 according to the information (e.g., the structured interpretation data 406) about the utterance of the user query 120. An example method of selecting the candidate media content items 552 is illustrated in FIG. 7.

[0181] Referring to FIG. 7, one or more candidate media content items 552 are determined by comparing the facets 502A of the first media content item 124 with the facets 502B of the media content items 556 (including 556A, 556B, etc.) from the media content database 554. In some embodiments, the candidate media content items 552 are media content items 124 each having a facet 502B that matches a facet type 504 of the subject facet 520 and has a facet value 506 different from the subject facet 520. The difference in the facet values can be determined based on different facet value representing the information about the utterance of the user query 120, such as the structured interpretation data 406.

[0182] In the illustrated example of FIG. 7, the facets 502A of the first media content item 124 are compared with the facets 502B of each of the media content items 556. For example, among other facet types, the facet type 504 to be referred to for comparison is determined according to the subject facet 520. Because the subject facet 520 is associated with tempo, the facet type 504 to be referred to for comparison is selected to be tempo. For the determined facet type 504 (e.g., "tempo" in this example), the facet value 506 of the first media content item 124 is compared with the facet value of each of the media content items 556 so that media content items 556 having facet values 506 that satisfy the information (e.g., the structured interpretation data 406) about the utterance of the user query 120 are selected for the candidate media content items 552. In this example, because the structured interpretation data 406 (e.g., the slot value 444) indicates {more upbeat} or its variants, the media content items 556 that have facet values 506 of tempo higher than the facet value 506 of 130 bpm of the first media content item 124 are selected as the candidate media content items 552. By way of example, in FIG. 7, the media content item 556A has the tempo value greater than the tempo value of the first media content item 124

and thus can be a candidate media content item 522. However, because the media content item 556B has the tempo value less than the tempo value of the first media content item 124, the media content item 556B is not selected as a candidate media content item 552.

[0183] In some embodiments, other than the facet type determined based on the subject facet 520, at least one of the facet types of the candidate media content item 552 can match at least one of the facet types of the first media content item 124. In other embodiments, the facet type determined based on the subject facet 520 can be the only facet type that the first media content item 124 and the candidate media content item 552 have in common.

[0184] In some embodiments, where there are one or more common facet types between the candidate media content item 522 and the first media content item 124, the candidate media content item 552 can be selected from the pool of media content items 124 such that, for at least one of such common facet types other than the facet type determined based on the subject facet 520, the facet values are the same or similar between the candidate media content item 552 and the first media content item 124. In other embodiments, regardless of the facet values of the facet types other than the facet type determined based on the subject facet 520, the facet values of the facet type determined based on the subject facet 520 are only compared between the first media content item 124 and the media content items 556 to determine the candidate media content item 552.

[0185] In other embodiments, the second media content item 126 is determined such that, for the facet type of the subject facet 520, the facet value of the second media content item 126 is relatively different from the facet value of the first media content item 124. In some embodiments, the amount of the relative difference is determined based on a distance between one or more attributes of the first media content item 124 and the steering query 120. For example, the amount of relative difference is obtained using a Euclidean distance between a vector representing the attributes of the first media content item 124 and a vector representing the steering query 120. Other algorithms can be used to calculate the distance between the attributes of the first media content item 124 and the steering query 120.

[0186] Referring to FIG. 8, at the second operation 322 of the operation 310, the system 100 operates to determine the second media content item 126 from the candidate media content items 552. In some embodiments, the steered media content determination engine 590 operates to determine one from the candidate media content items 552 as the second media content item 126.

[0187] In some embodiments, the steered media content determination engine 590 operates to retrieve the user information (also referred to herein as a user profile), such as the user account 240 and/or the taste profile 242, to select the second media content item 126 from the candidate media content items 552.

[0188] In some embodiments, the user information is used to determine the facet value 506 of the second media content item 126 for the facet type 504 matching the subject facet 520. By way of example, if the user information indicates that the user likes music with higher tempo, the second media content item 126 can be selected to have the highest tempo among the candidate media content items 552. In other embodiments, various types of the user information can be used to select one of the candidate media content items 552 as the second media content item 126. Examples of such types of the user information include user preferences and historical information about the user's consumption of media content, and can include a wide range of information such as artist plays, song plays, skips, dates of listen by the user, songs per day, playlists, play counts, start/stop/skip data for portions of a song or album, contents of collections, user rankings, preferences, or other mentions received via a client device, or other media plays, such as websites visited, book titles, movies watched, playing activity during a movie or other presentations, ratings, or terms corresponding to the media.

[0189] In some embodiments, the steered media content determination engine 590 operates to generate a steered media content identifier 592 (FIG. 2) that identifies the second media content item 126. The steered media content identifier 592 can be used by one or more computing devices, such as the media playback device 102 and the media delivery system 104, to identify, retrieve, transmit, and/or receive the second media content item 126 for playback.

[0190] In the operations 320 and 322, it is primarily described that a plurality of candidate media content items 552 are determined and the second media content item 126 is selected from the candidate media content items 552. However, it is also possible in other embodiments that a single media content item can be determined in the operation 320, which then automatically becomes the second media content item 216 without the operation 322.

[0191] Referring back to FIG. 3, when the operation 312 is performed, the method 300 can move on to operation 314. At the operation 314, the media delivery system 104 operates to provide the second media content item 126 to the media playback device 102 for playback. In some embodiments, the media delivery system 104 operates to transmit the steered media content identifier 592 to the media playback device 102. In some embodiments, the media playback device 102 can use the steered media content identifier 592 to identify the second media content item 126. In other embodiments, the media playback device 102 operates to transmit the steered media content identifier 592 to the media delivery system 104 to enable the media delivery system 104 to retrieve the second media content item 126 associated with the identifier 592. When the second media content item 126 is determined in the media delivery system 104, the media delivery system 104 operates to stream the

second media content item 126 to the media playback device 102 via the network 106, and the media playback device 102 can store and/or play the second media content item 126.

[0192] Below follows an itemized list of some embodiments of the present invention.

1. A method (300) for playing media content, the method (300) comprising:

receiving an utterance of a user query (120);
identifying at least one first media content item that is currently being played, the at least one first media content item having a first facet (502), the first facet (502) having a first facet type (504) and a first facet value (506);
determining an intent (432) and a slot (434) from the utterance; and
retrieving at least one second media content item from a catalog of media content items (124), the at least one second media content item having a second facet (502), the second facet (502) having a second facet type (504) and a second facet value (506), the second facet type (504) being the same as the first facet type (504), and the second facet value (506) being different from the first facet value (506), the difference representative of at least one of the intent (432) and the slot (434) of the utterance.

2. The method (300) of item 1, further comprising:
determining the first facet type (504) and the second facet type (504) based on at least one of the intent (432) and the slot (434) of the utterance.

3. The method (300) of item 1, further comprising:

determining a slot type (442) and a slot value (444) from the slot (434); and
determining the first facet type (504) and the second facet type (504) based on the slot type (442).

4. The method (300) of any of the preceding items, wherein the at least one first media content item has a plurality of facets (502) including the first facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein the at least one second media content item has a plurality of facets (502) including the second facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein at least some of the facet types (504) in the at least one second media content item are the same as at least some of the facet types (504) in the at least one first media content item, and the facet values (506), other than the first facet value (506), of the at least some of the facet types (504) in the at least one second media content item are the same as the facet values (506), other than the second facet value (506), of the at least some of the facet types (504) in the at least one first media content item.

5. The method (300) of any of the preceding items, further comprising:
identifying a media content steering intent (450) from the utterance, the media content steering intent (450) including a command to retrieve at least one media content item having a facet (502) that has a different facet value (506) from a corresponding facet (502) of at least one currently-playing media content item.

6. The method (300) of any of the preceding items, further comprising:

retrieving user information associated with a user who provide the utterance;
wherein the second facet value (506) of the at least one second media content item is determined based on the user profile.

7. The method (300) of any of the preceding items, further comprising:

prior to receiving the utterance, transmitting a media content signal to a voice-enabled device for playing the at least one first media content item;
receiving the utterance from the voice-enabled device; and
converting the utterance to a text version, the intent (432) and the slot (434) of the utterance being determined from the text version.

8. The method (300) of any of the preceding items, wherein the first facet type (504) and the second facet type (504) include at least one of mood, genre, activity, tempo, acoustic attributes, artists, popularity, geographic information, release date, hits, and deep cuts.

9. The method (300) of any of the preceding items, wherein the first media content item is a first media content playlist (239) including a plurality of first tracks, and the second media content item is a second media content playlist (239) including a plurality of second tracks.

10. The method (300) of any of items 1-9, wherein the first media content item is included in a first media content playlist (239), and the second media content item is included in a second media content playlist (239).

11. A computer-readable medium having stored thereon instructions that, when executed by one or

more processors cause execution of operations including:

receiving an utterance;
identifying at least one first media content item that is currently being played, the at least one first media content item having a first facet (502), the first facet (502) having a first facet type (504) and a first facet value (506);
determining an intent (432) and a slot (434) from the utterance; and
retrieving at least one second media content item from a catalog of media content items (124), the at least one second media content item having a second facet (502), the second facet (502) having a second facet type (504) and a second facet value (506), the second facet type (504) being the same as the first facet type (504), and the second facet value (506) being different from the first facet value (506), the difference representative of at least one of the intent (432) and the slot (434) of the utterance.

12. The computer-readable medium of item 11, wherein the instructions further cause execution of operations including:

determining a slot type (442) and a slot value (444) from the slot (434); and
determining the first facet type (504) and the second facet type (504) based on the slot type (442).

13. The computer-readable medium of item 11 or 12, wherein the at least one first media content item has a plurality of facets (502) including the first facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein the at least one second media content item has a plurality of facets (502) including the second facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein at least some of the facet types (504) in the at least one second media content item are the same as at least some of the facet types (504) in the at least one first media content item, and the facet values (506), other than the first facet value (506), of the at least some of the facet types (504) in the at least one second media content item are the same as the facet values (506), other than the second facet value (506), of the at least some of the facet types (504) in the at least one first media content item.

14. The computer-readable medium of any of items 11-13, wherein the instructions further cause execution of operations including:
identifying a media content steering intent (450) from the utterance, the media content steering intent (450) including a command to retrieve at least one media content item having a facet (502) that has a different facet value (506) from a corresponding facet (502) of at least one currently-playing media content item.

15. The computer-readable medium of any of items 11-14, wherein the instructions further cause execution of operations including:

retrieving user information associated with a user who provide the utterance;
wherein the second facet value (506) of the at least one second media content item is determined based on the user profile.

16. The computer-readable medium of any of items 11-15, wherein the instructions further cause execution of operations including:

prior to receiving the utterance, transmitting a media content signal to a voice-enabled device for playing the at least one first media content item;
receiving the utterance from the voice-enabled device; and
converting the utterance to a text version, the intent (432) and the slot (434) of the utterance being determined from the text version.

17. A media playback system (100) comprising:

a media playback device (102) operable to play media content; and
a media delivery system (104) configured to:

receive an utterance from the media playback device (102);
identify at least one first media content item that is currently being played, the at least one first media content item having a first facet (502), the first facet (502) having a first facet type (504) and a first facet value (506);
determine an intent (432) and a slot (434) from the utterance;
retrieve at least one second media content item from a catalog of media content items (124), the at least one second media content item having a second facet (502), the second facet (502) having a second facet type (504) and a second facet value (506), the second facet type (504) being the same as the first facet type (504), and the second facet value (506) being different from the first facet value (506), the difference representative of at least one of the intent (432) and the slot (434) of the utterance; and
transmit the second media content item to the media playback device (102).

18. The media playback system (100) of item 17, wherein the media delivery system (104) is configured to:

prior to receiving the utterance, transmit a media content signal to the media playback device (102) for playing the first media content item;
receive the utterance from the media playback device (102); and
convert the utterance to a text version, the intent (432) and the slot (434) of the utterance being determined from the text version.

19. The media playback system (100) of item 17 or 18, wherein the media delivery system (104) is configured to:

determine a slot type (442) and a slot value (444) from the slot (434); and
determine the first facet type (504) and the second facet type (504) based on the slot type (442).

20. The media playback system (100) of any of items 17-19, wherein the at least one first media content item has a plurality of facets (502) including the first facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein the at least one second media content item has a plurality of facets (502) including the second facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein at least some of the facet types (504) in the at least one second media content item are the same as at least some of the facet types (504) in the at least one first media content item, and the facet values (506), other than the first facet value (506), of the at least some of the facet types (504) in the at least one second media content item are the same as the facet values (506), other than the second facet value (506), of the at least some of the facet types (504) in the at least one first media content item.

[0193] The various examples and teachings described above are provided by way of illustration only and should not be construed to limit the scope of the present disclosure. Those skilled in the art will readily recognize various modifications and changes that may be made without following the examples and applications illustrated and described herein, and without departing from the true spirit and scope of the present disclosure.

Claims

1. A media playback system (100) comprising:

a media playback device (102) operable to play media content; and

a media delivery system (104) configured to:

receive an utterance from the media playback device (102);
identify at least one first media content item (124) that is currently being played by the media playback device;
parse the utterance into parts including one or more slots (434);
identify a first facet (502) of the at least one first media content item (124) that matches the one or more slots, the first facet (502) having a first facet type (504) and a first facet value (506);
retrieve at least one second media content item (126) from a catalog of media content items (124), the at least one second media content item having a second facet (502), the second facet having a second facet type (504) and a second facet value (506), the second facet type (504) being the same as the first facet type (504), and the second facet value (506) having a relative difference from the first facet value (506), the relative difference corresponding to the one or more slots (434); and
transmit the second media content item to the media playback device (102).

2. The media playback system (100) of claim 1, wherein the media delivery system (104) is configured to:

prior to receiving the utterance, transmit a media content signal to the media playback device (102) for playing the first media content item;
receive the utterance from the media playback device (102); and
convert the utterance to a text version, the parsing of the utterance being determined from the text version.

3. The media playback system (100) of any preceding claim,
wherein the parsing of the utterance further includes parsing the utterance into parts including an intent (432) and the one or more slots (434), and
wherein the identifying of the first facet (502) comprises identifying the first facet (502) that also matches the intent (432).

4. The media playback system (100) of any preceding claim, wherein each of the one or more slots (434) has a slot type (442) and a slot value (444).

5. The media playback system (100) of claim 4, wherein the first facet (502) matches at least one slot (434) of the one or more slots by the facet type (504) of the first facet being the same as the slot type (442)

of said at least one slot.

6. The media playback system (100) of claim 4 or 5, wherein the relative difference between the first and second facet values (506) corresponds to the one or more slots (434) by corresponding to the slot value (444) of at least one of said one or more slots.

7. The media playback system (100) of any preceding claim, wherein the at least one first media content item (124) has a plurality of facets (502) including the first facet (502), each facet (502) having a facet type (504) and a facet value (506),
wherein the at least one second media content item (126) has a plurality of facets (502) including the second facet (502), each facet (502) having a facet type (504) and a facet value (506), and
wherein at least some of the facet types (504) in the at least one second media content item are the same as at least some of the facet types (504) in the at least one first media content item, and the facet values (506), other than the first facet value (506), of the at least some of the facet types (504) in the at least one second media content item are the same as the facet values (506), other than the second facet value (506), of the at least some of the facet types (504) in the at least one first media content item.

8. A method performed by a media content steering engine (110) of a media delivery system (104) connected to a media playback device (102) via a data communication network (106), the method comprising:

from the media playback device, receiving (304) a steering command from a user (U) of said media playback device;
based on the received (304) steering command, identifying (310) a media content facet type (504) of a first media item (124) which is currently playing on the media playback device;
determining a first facet value (506) of the identified facet type;
based on the received (304) steering command and in relation to the determined first facet value, selecting a second media content item (126) having a second facet value (506) of said identified facet type, different from the first facet value; and
providing (314) the second media content item to the media playback device in response to the received (304) steering command.

9. The method of claim 8, wherein the steering command is obtained from an utterance received (304) from the media playback device (102) in voice or text format.

10. The method of claim 9, wherein the utterance is parsed before the steering command is obtained there from.

11. A media delivery system (104) comprising:

means (218) for, from a media playback device (102), receiving (304) a steering command from a user (U) of said media playback device;
means for, based on the received (304) steering command, identifying (310) a media content facet type (504) for a first media content item (124) which is currently playing on the media playback device;
means for determining a first facet value (506) of the identified facet type;
means for, based on the received (304) steering command and in relation to the determined first facet value, selecting a second media content item (126) having a second facet value (506) of said identified facet type, different from the first facet value; and
means (218) for providing (314) the second media content item to the media playback device in response to the received (304) steering command.

12. A computer program product (216) comprising computer-executable components for causing a media delivery system (104) to perform the method of any one of claims 8-10 when the computer-executable components are run on processing circuitry (214) comprised in the media delivery system.

13. A method performed by a media playback device (102) connected to a media delivery system (104) via a data communication network (106), the method comprising:

from a user (U) of the media playback device, receiving (304) a steering command;
based on the received (304) steering command, identifying (310) a media content facet type (504) for a first media content item (124) which is currently playing on the media playback device;
determining a first facet value (506) of the identified facet type;
transmitting the received steering command to the media delivery system;
in response to the transmitted steering command, obtaining a second media content item (126) having a second facet value (506) of said identified facet type, different from the first facet value; and
playing said second media content item (126).

14. A media playback device (102) comprising:

means (130) for receiving (304) a steering command from a user (U) of the media playback device;
means for, based on the received (304) steering command, identifying (310) a media content facet type (504) for a first media content item (124) which is currently playing on the media playback device;
means for determining a first facet value (506) of the identified facet type;
means (134) for transmitting the received steering command to a media delivery system (104);
means (134) for, in response to the transmitted steering command, obtaining a second media content item (126) having a second facet value (506) of said identified facet type, different from the first facet value; and
means (140) for playing said second media content item (126).

15. A computer program product (150) comprising computer-executable components for causing a media playback device (102) to perform the method of claim 14 when the computer-executable components are run on processing circuitry (148) comprised in the media playback device.

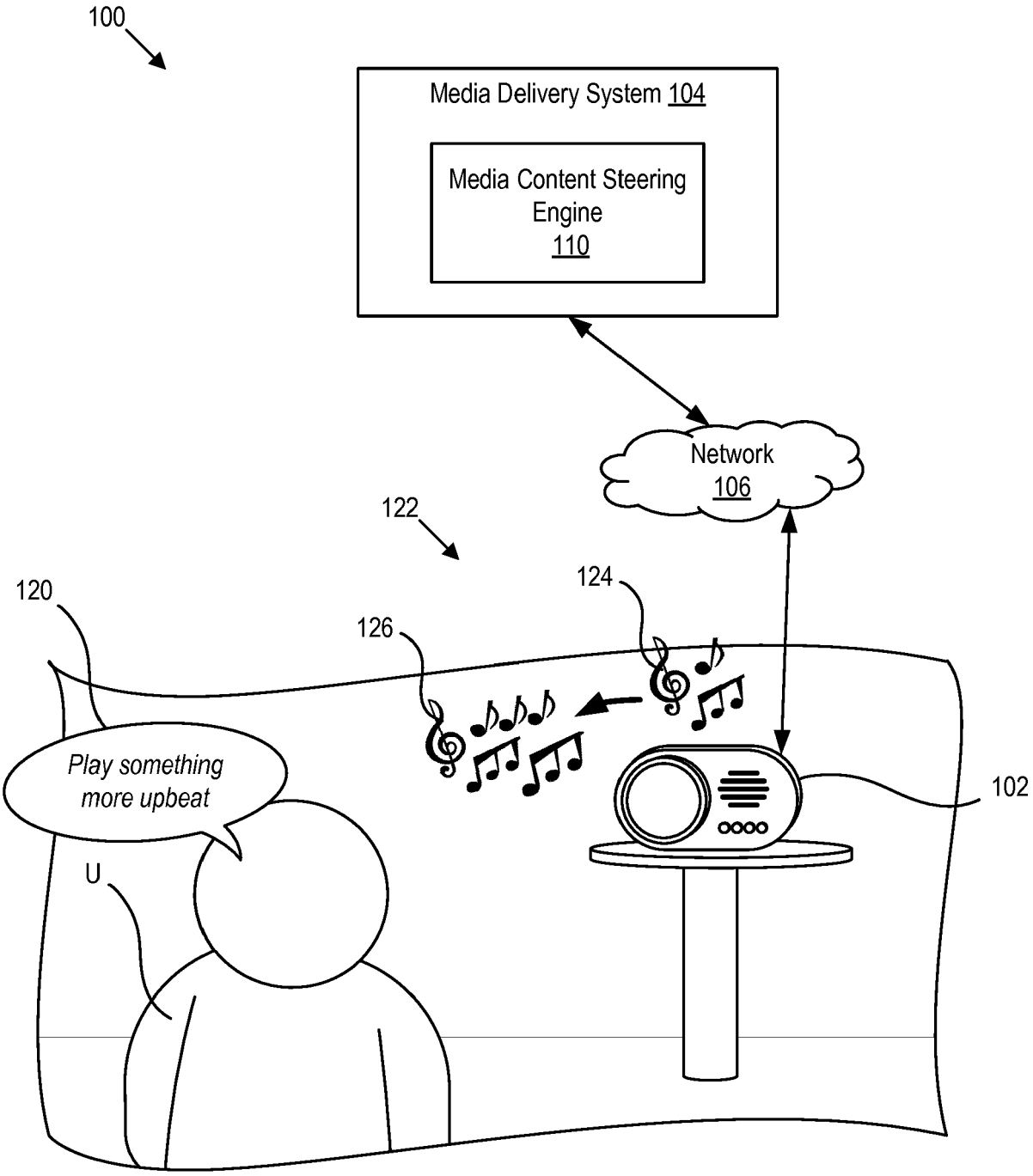
100
Media Delivery System 104
Media Content Steering Engine 110
Network 106
122
120
124
126
Play something more upbeat
U
102

FIG. 1

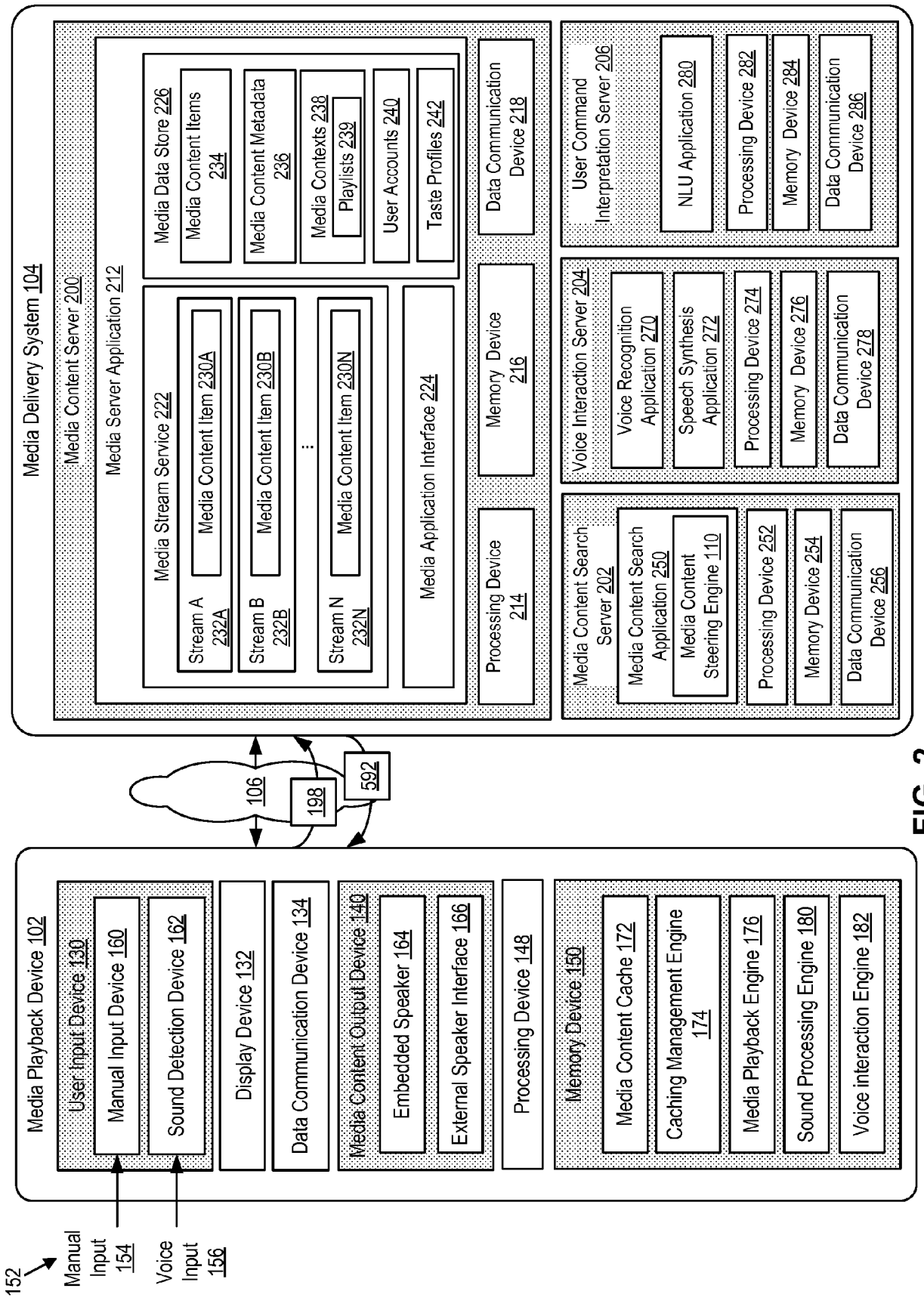
Media Delivery System 104
Media Content Server 200
Media Server Application 212
Media Stream Service 222
Stream A 232A
Media Content Item 230A
Stream B 232B
Media Content Item 230B
Stream N 232N
Media Content Item 230N
Media Application Interface 224
Media Data Store 226
Media Content Items 234
Media Content Metadata 236
Media Contexts 238
Playlists 239
User Accounts 240
Taste Profiles 242
Processing Device 214
Memory Device 216
Data Communication Device 218
Media Content Search Server 202
Media Content Search Application 250
Media Content Steering Engine 110
Processing Device 252
Memory Device 254
Data Communication Device 256
Voice Interaction Server 204
Voice Recognition Application 270
Speech Synthesis Application 272
Processing Device 274
Memory Device 276
Data Communication Device 278
User Command Interpretation Server 206
NLU Application 280
Processing Device 282
Memory Device 284
Data Communication Device 286
106
198
592
Media Playback Device 102
User Input Device 130
Manual Input Device 160
Sound Detection Device 162
Display Device 132
Data Communication Device 134
Media Content Output Device 140
Embedded Speaker 164
External Speaker Interface 166
Processing Device 148
Memory Device 150
Media Content Cache 172
Caching Management Engine 174
Media Playback Engine 176
Sound Processing Engine 180
Voice interaction Engine 182
152
Manual Input 154
Voice Input 156

FIG. 2

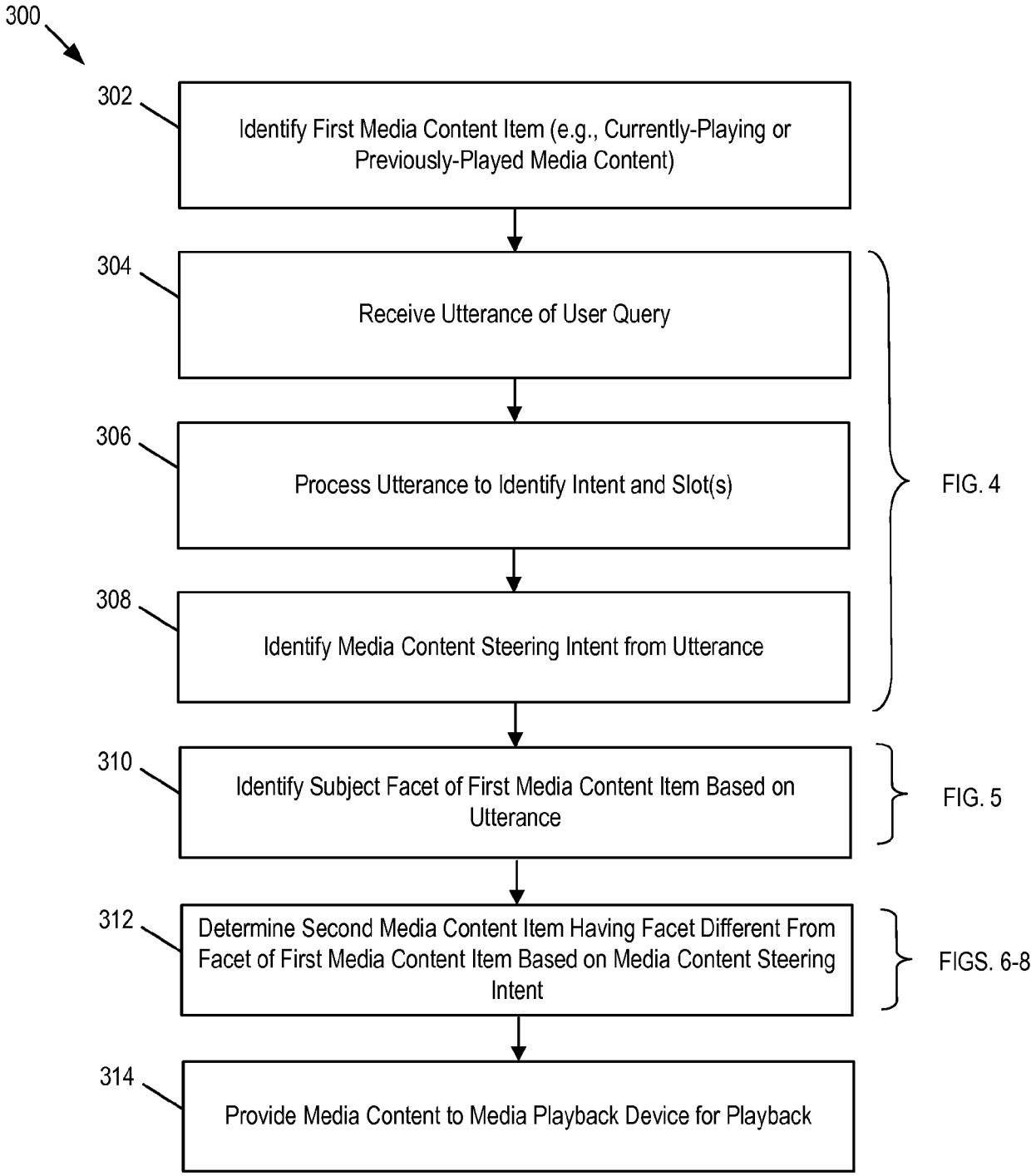
300
302
Identify First Media Content Item (e.g., Currently-Playing or Previously-Played Media Content)
304
Receive Utterance of User Query
306
Process Utterance to Identify Intent and Slot(s)
308
Identify Media Content Steering Intent from Utterance
FIG. 4
310
Identify Subject Facet of First Media Content Item Based on Utterance
FIG. 5
312
Determine Second Media Content Item Having Facet Different From Facet of First Media Content Item Based on Media Content Steering Intent
FIGS. 6-8
314
Provide Media Content to Media Playback Device for Playback

FIG. 3

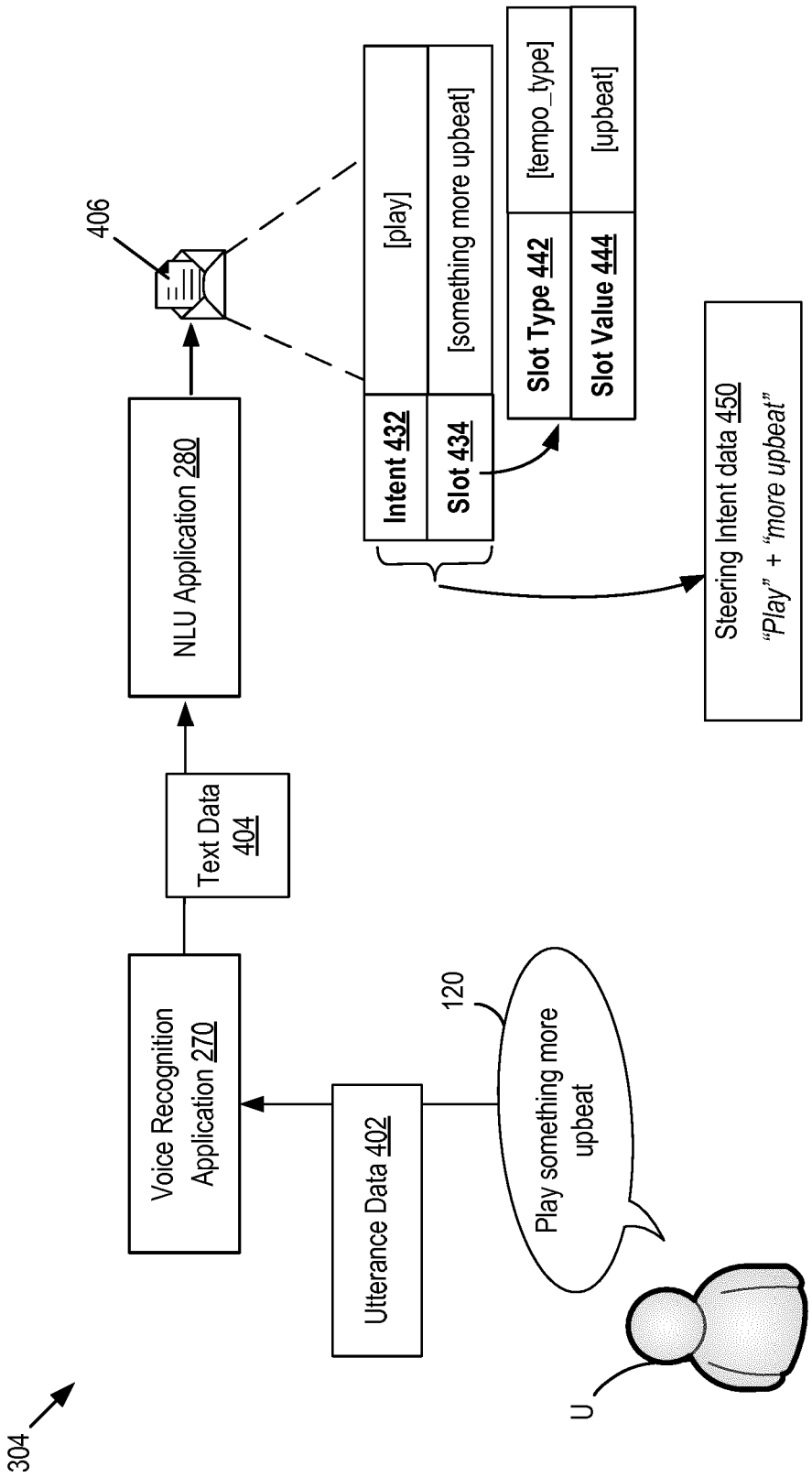
304
U
Play something more upbeat
120
Utterance Data 402
Voice Recognition Application 270
Text Data 404
NLU Application 280
406
Intent 432
[play]
Slot 434
[something more upbeat]
Slot Type 442
[tempo_type]
Slot Value 444
[upbeat]
Steering Intent data 450
"Play" + "more upbeat"

FIG. 4

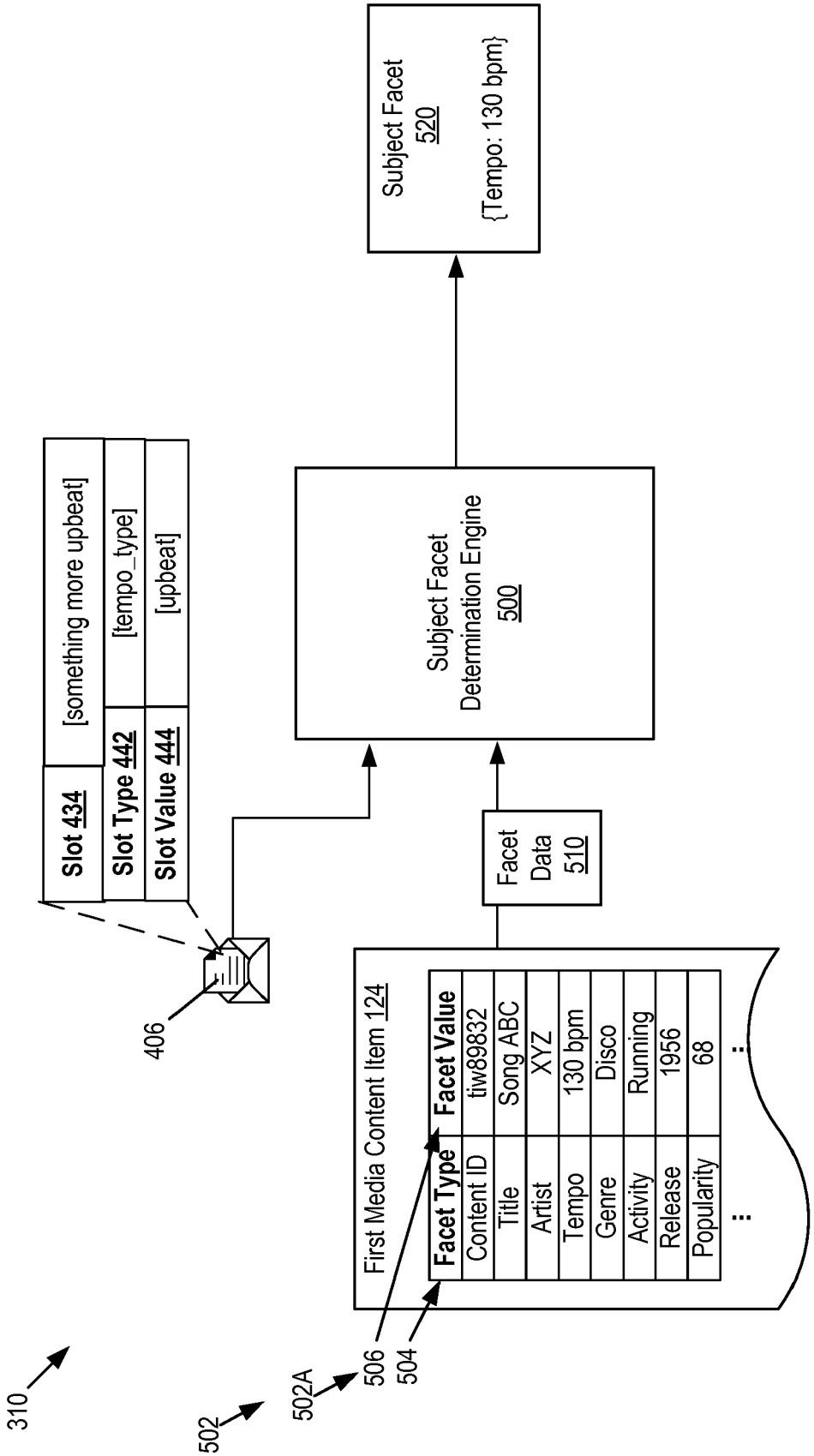
Subject Facet
520
{Tempo: 130 bpm}
Subject Facet
Determination Engine
500
Facet
Data
510
First Media Content Item 124
Facet Type
Facet Value
Content ID
tiw89832
Title
Song ABC
Artist
XYZ
Tempo
130 bpm
Genre
Disco
Activity
Running
Release
1956
Popularity
68
Slot 434
[something more upbeat]
Slot Type 442
[tempo_type]
Slot Value 444
[upbeat]
406
310
502
502A
506
504

FIG. 5

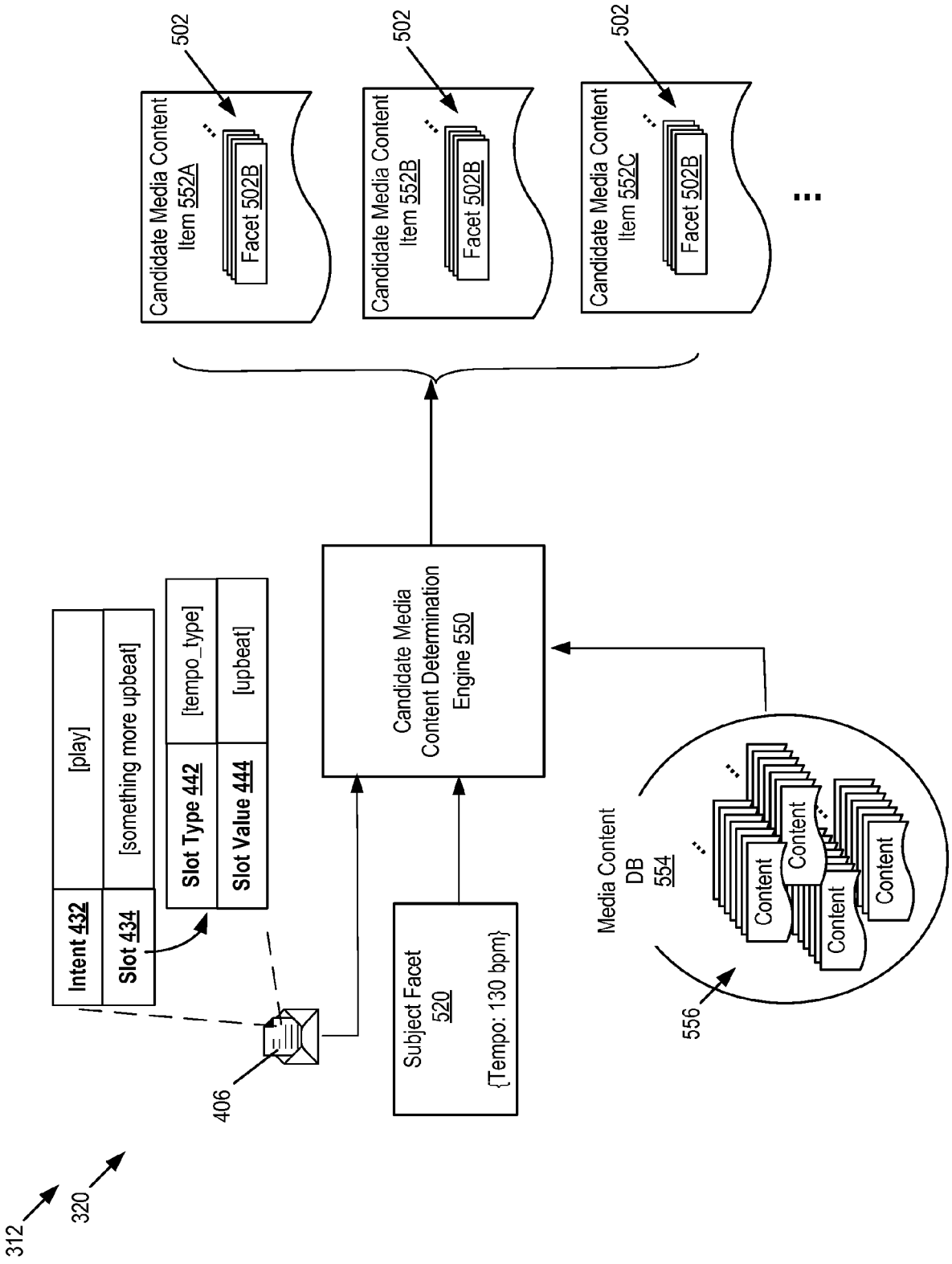
312
320
406
Intent 432
[play]
Slot 434
[something more upbeat]
Slot Type 442
[tempo_type]
Slot Value 444
[upbeat]
Candidate Media Content Determination Engine 550
Subject Facet 520
{Tempo: 130 bpm}
Media Content DB 554
Content
556
502
Candidate Media Content Item 552A
Facet 502B
Candidate Media Content Item 552B
Facet 502B
Candidate Media Content Item 552C
Facet 502B

FIG. 6

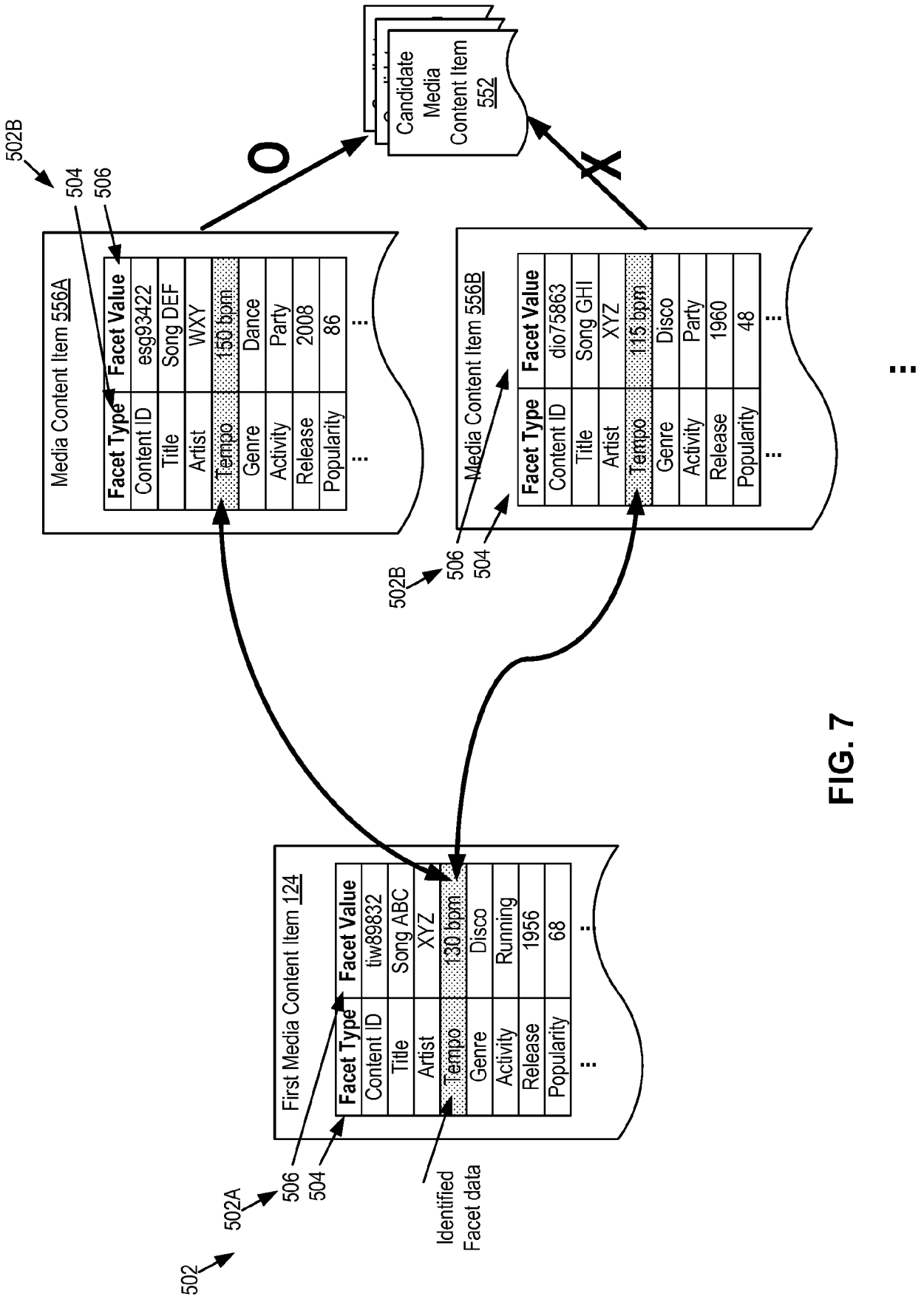
502
502A
506
504
First Media Content Item 124
Facet Type
Facet Value
Content ID
tiw89832
Title
Song ABC
Artist
XYZ
Tempo
130 bpm
Genre
Disco
Activity
Running
Release
1956
Popularity
68
Identified Facet data
502B
504
506
Media Content Item 556A
Facet Type
Facet Value
Content ID
esg93422
Title
Song DEF
Artist
WXY
Tempo
150 bpm
Genre
Dance
Activity
Party
Release
2008
Popularity
86
O
Candidate Media Content Item 552
X
Media Content Item 556B
Facet Type
Facet Value
Content ID
dio75863
Title
Song GHI
Artist
XYZ
Tempo
115 bpm
Genre
Disco
Activity
Party
Release
1960
Popularity
48

FIG. 7

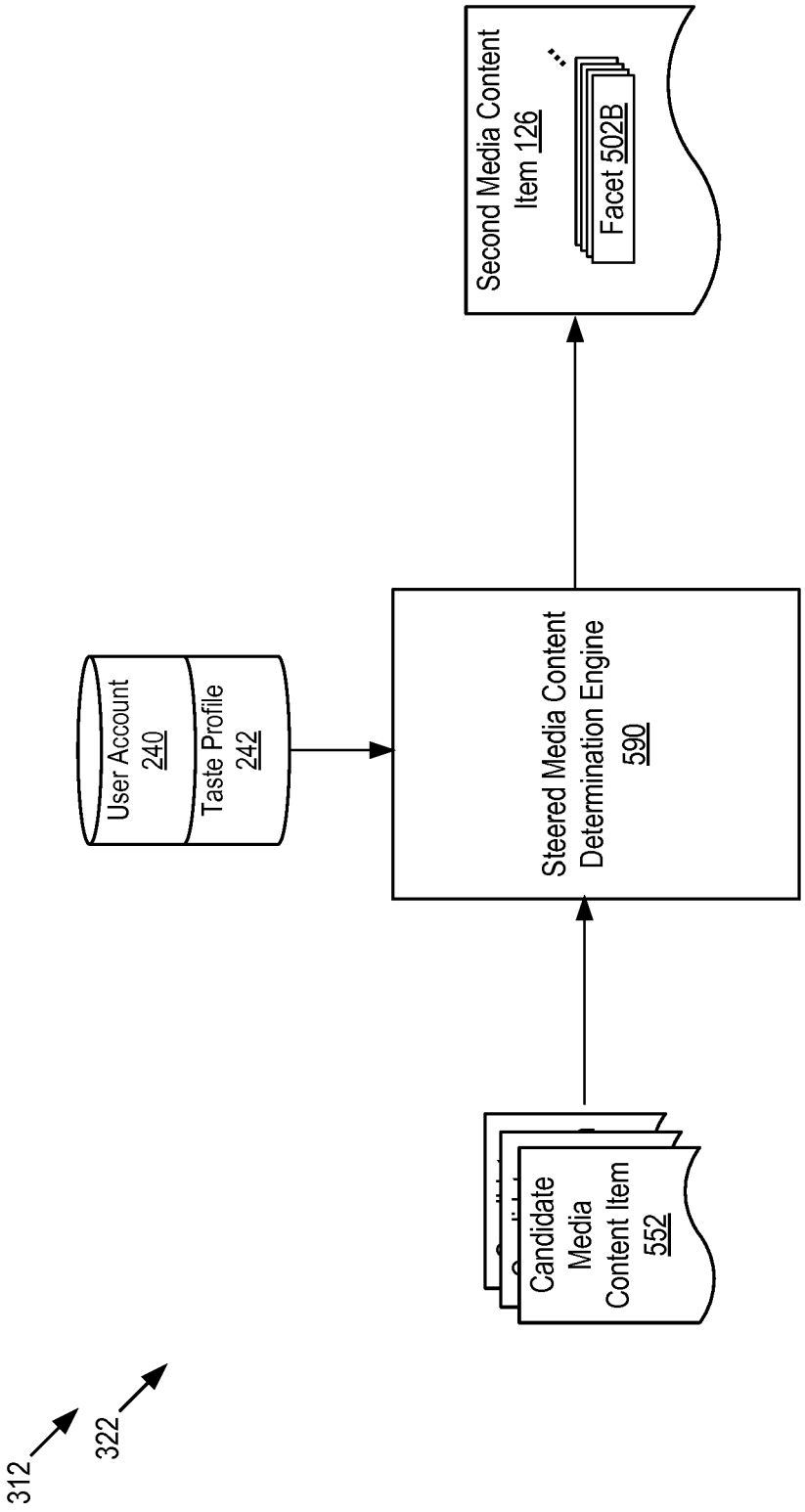
Second Media Content Item 126
Facet 502B
User Account 240
Taste Profile 242
Steered Media Content Determination Engine 590
Candidate Media Content Item 552
312
322

FIG. 8

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 19 20 6103

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	WO 2007/022533 A2 (GRACENOTE INC [US]; BRENNER VADIM [US] ET AL.) 22 February 2007 (2007-02-22) * paragraphs [0073], [0074] *	1-15	INV. G11B27/32 G11B27/34 G06F16/632 G10L15/18
A	US 2017/068423 A1 (NAPOLITANO LIA T [US] ET AL) 9 March 2017 (2017-03-09) * figures 6-all,7,8I-K *	1-15	
A	WO 2016/209924 A1 (AMAZON TECH INC [US]) 29 December 2016 (2016-12-29) * pages 7-11 *	1-15	
A	US 2007/089057 A1 (KINDIG BRADLEY D [US]) 19 April 2007 (2007-04-19) * the whole document *	1-15	
A	US 2017/243576 A1 (MILLINGTON NICHOLAS A J [US] ET AL) 24 August 2017 (2017-08-24) * the whole document *	1-15	
A	US 2017/025120 A1 (DAYAN YIGAL S [IL] ET AL) 26 January 2017 (2017-01-26) * the whole document *	1-15	TECHNICAL FIELDS SEARCHED (IPC) G11B G06F G10L
A,D	MAGNUS MORTENSEN ET AL: "Real-World Mood-Based Music Recommendation", 15 January 2008 (2008-01-15), INFORMATION RETRIEVAL TECHNOLOGY; [LECTURE NOTES IN COMPUTER SCIENCE], SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PAGE(S) 514 - 519, XP019076009, ISBN: 978-3-540-68633-0 * the whole document *	1-15	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	18 March 2020	Bruma, Cezar

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 19 20 6103

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-03-2020

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007022533 A2	22-02-2007	EP 1934828 A2	25-06-2008
		JP 2009505321 A	05-02-2009
		KR 20080043358 A	16-05-2008
		US 2009076821 A1	19-03-2009
		WO 2007022533 A2	22-02-2007
US 2017068423 A1	09-03-2017	AU 2016247040 A1	23-03-2017
		AU 2017204359 A1	20-07-2017
		AU 2018260958 A1	13-12-2018
		CN 107003797 A	01-08-2017
		CN 107577385 A	12-01-2018
		DK 201570826 A1	27-03-2017
		DK 201670575 A1	27-03-2017
		DK 201770287 A1	08-05-2017
		DK 201970273 A1	04-06-2019
		EP 3167618 A1	17-05-2017
		EP 3291564 A1	07-03-2018
		JP 6635982 B2	29-01-2020
		JP 2017215971 A	07-12-2017
		JP 2018190431 A	29-11-2018
		JP 2018502345 A	25-01-2018
		KR 20170036805 A	03-04-2017
		KR 20170100067 A	01-09-2017
		KR 20200011555 A	03-02-2020
		US 2017068423 A1	09-03-2017
		US 2017285915 A1	05-10-2017
		US 2019310765 A1	10-10-2019
		WO 2017044257 A1	16-03-2017
WO 2016209924 A1	29-12-2016	US 2016379638 A1	29-12-2016
		WO 2016209924 A1	29-12-2016
US 2007089057 A1	19-04-2007	NONE	
US 2017243576 A1	24-08-2017	US 2017243576 A1	24-08-2017
		US 2018226074 A1	09-08-2018
US 2017025120 A1	26-01-2017	US 2017024459 A1	26-01-2017
		US 2017025120 A1	26-01-2017

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 85837717 **[0132]**

Non-patent literature cited in the description

- **MORTENSEN et al.** Real-World Mood-Based Recommendation. *AIRS 2008: Information Retrieval Technology,* 2008, 514-519 **[0006]**