Office de la Propriété Intellectuelle du Canada

Un organisme d'Industrie Canada

Canadian Intellectual Property Office

An agency of Industry Canada

CA 2817340 A1 2012/05/18

(21) 2817 340

# (12) DEMANDE DE BREVET CANADIEN CANADIAN PATENT APPLICATION

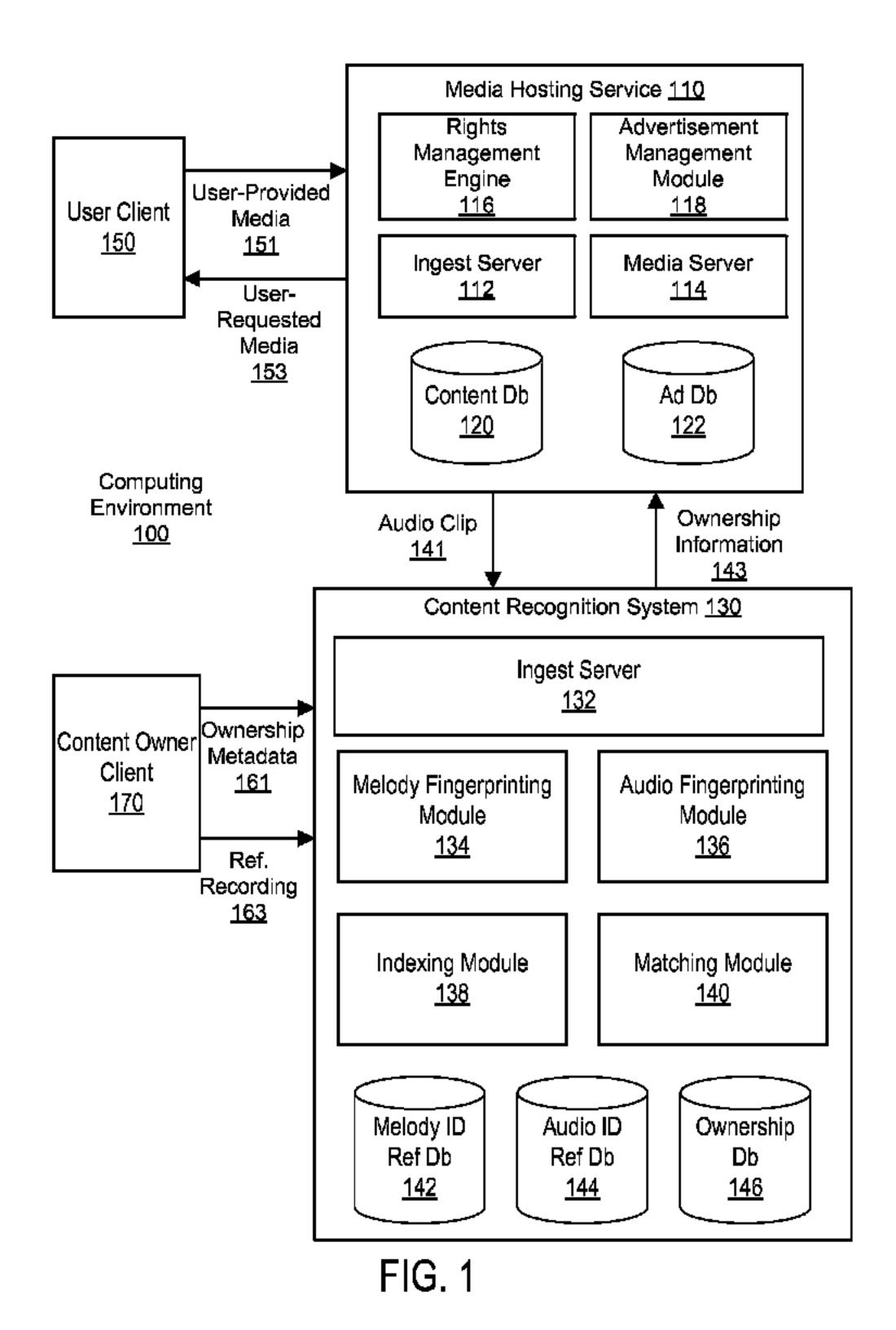
(13) **A1** 

- (86) Date de dépôt PCT/PCT Filing Date: 2011/11/10
- (87) Date publication PCT/PCT Publication Date: 2012/05/18
- (85) Entrée phase nationale/National Entry: 2013/05/08
- (86) N° demande PCT/PCT Application No.: US 2011/060169
- (87) N° publication PCT/PCT Publication No.: 2012/064945
- (30) Priorité/Priority: 2010/11/12 (US12/945,216)

- (51) Cl.Int./Int.Cl. *G06Q 50/00* (2012.01), *H04N 21/266* (2011.01), *H04N 21/4627* (2011.01), *H04N 21/6334* (2011.01), *H04N 21/835* (2011.01)
- (71) Demandeur/Applicant: GOOGLE INC., US
- (72) Inventeurs/Inventors:
  KING, DAVID G., US;
  SALEM, GEORGE, US;
  WANG, YILING SU, US;
  WISEMAN, MATTHEW, US
- (74) Agent: SMART & BIGGAR

(54) Titre: GESTION DE DROITS MULTIMEDIAS UTILISANT UNE IDENTIFICATION DE MELODIE

(54) Title: MEDIA RIGHTS MANAGEMENT USING MELODY IDENTIFICATION



#### (57) Abrégé/Abstract:

A content recognition system operates in conjunction with a media hosting service to identify hosted media content and ownership rights associated with the hosted content. By applying melody recognition, the content recognition system can identify compositions embodied in hosted media content even when these compositions do not precisely match any known sound recording. Thus, the content recognition system is beneficially able to detect, for example, recorded cover performances and recorded live performances embodied in hosted media content. Once identified, ownership information is determined and the media hosting service can carry out appropriate rights management policies associated with the content such as monetizing or blocking the protected content.

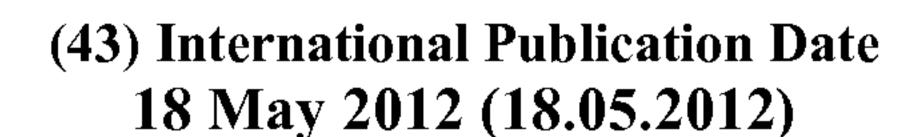




#### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

#### (19) World Intellectual Property Organization

International Bureau







## (10) International Publication Number WO 2012/064945 A3

(51) International Patent Classification:

*G06Q 50/00* (2006.01) *H04N 21/6334* (2011.01) *H04N 21/835* (2011.01) *H04N 21/266* (2011.01) *H04N 21/4627* (2011.01)

(21) International Application Number:

PCT/US2011/060169

(22) International Filing Date:

10 November 2011 (10.11.2011)

(25) Filing Language:

English

(26) Publication Language:

English

US

(30) Priority Data: 12/945,216 12 November 2010 (12.11.2010)

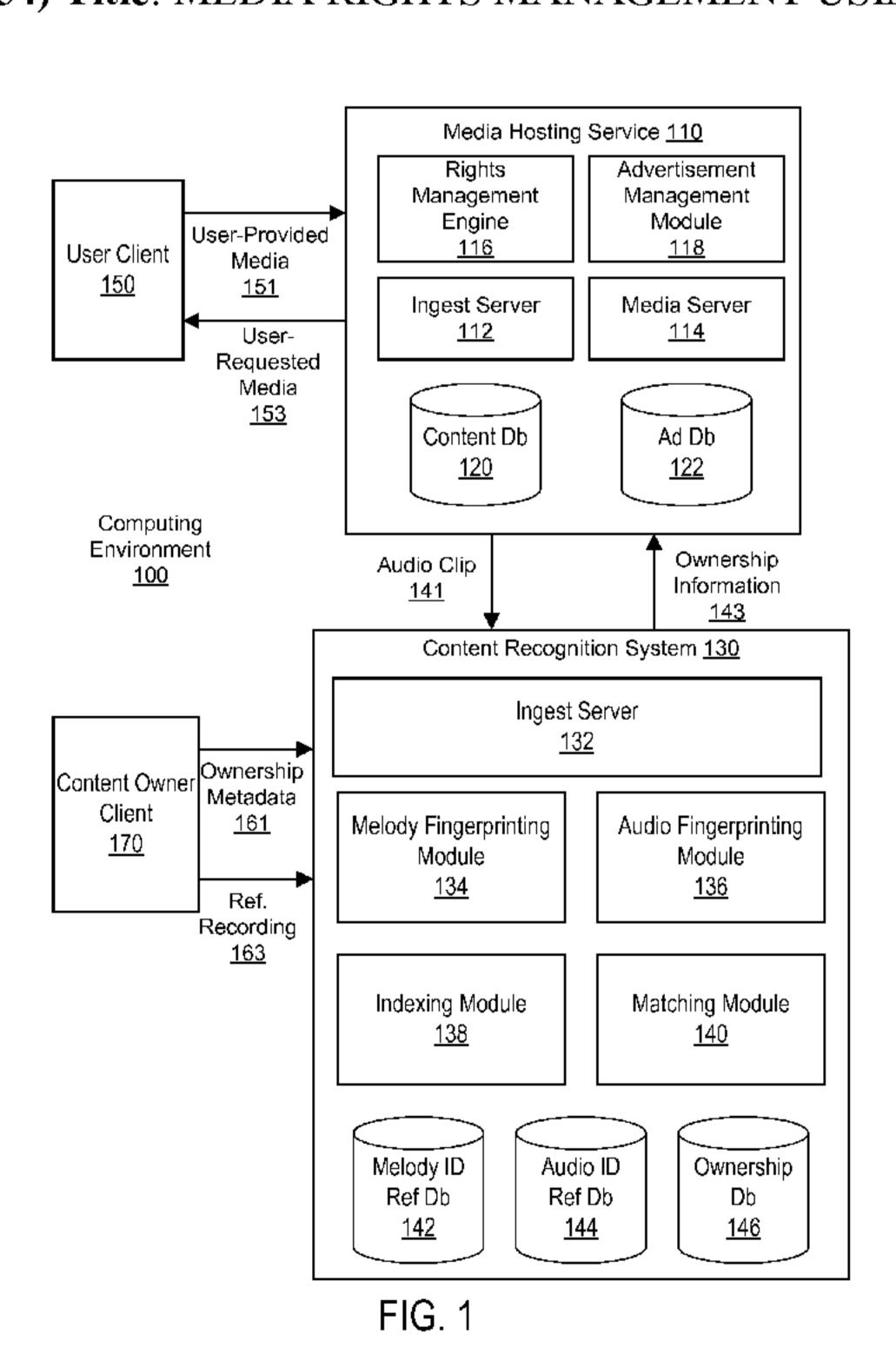
- (71) Applicant (for all designated States except US): GOOGLE INC. [US/US]; 1600 Amphitheatre Parkway, Mountain View, CA 94043 (US).
- (72) Inventors; and

(75) Inventors/Applicants (for US only): KING, David, G. [US/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Mountain, CA 94043 (US). SALEM, George [US/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043 (US). WANG, Yiling, Su [US/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Mountain View,

- CA 94043 (US). WISEMAN, Matthew [US/US]; C/o Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043 (US).
- (74) Agents: HOFFMAN, Brian, M. et al.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (US).
- kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

#### (54) Title: MEDIA RIGHTS MANAGEMENT USING MELODY IDENTIFICATION



(57) Abstract: A content recognition system operates in conjunction with a media hosting service to identify hosted media content and ownership rights associated with the hosted content. By applying melody recognition, the content recognition system can identify compositions embodied in hosted media content even when these compositions do not precisely match any known sound recording. Thus, the content recognition system is beneficially able to detect, for example, recorded cover performances and recorded live performances embodied in hosted media content. Once identified, ownership information is determined and the media hosting service can carry out appropriate rights management policies associated with the content such as monetizing or blocking the protected content.

## 

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,

### GW, ML, MR, NE, SN, TD, TG).

#### **Published**:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- (88) Date of publication of the international search report: 5 July 2012

# MEDIA RIGHTS MANAGEMENT USING MELODY IDENTIFICATION BACKGROUND

#### 1. FIELD OF ART

[0001] The invention generally relates to media sharing and more specifically to enforcing ownership rights to media content.

#### 2. Description of the Related Art

[0002] Generally, a "sound recording" refers to a particular musical performance stored on a tangible storage medium such as an optical disc (e.g., CD, DVD), magnetic disc or tape, solid state memory (e.g., memory card) or other persistent, tangible storage medium. In the music industry, sound recordings are typically produced and distributed by record labels, i.e., by companies that scout, develop, and manage recording artists, and coordinate the marketing, promotion, production, manufacture, and distribution of sound recordings. These record labels typically hold various rights under copyright law to the sound recordings they produce, although other entities may also hold these rights. In some instances, two or more recording labels or other entities may hold ownership rights to a single sound recording as the sound recording ownership may vary by country.

[0003] In contrast to a sound recording, a "composition" generally refers to an original piece of music (i.e., a song) that is not limited to one particular sound recording that memorializes a performance of the piece. For example, for a given composition by a song writer, there may be a studio recording by the song writer, a recorded live performance, and a recorded cover performance by another artist, each of which would be a distinct sound recording. Ownership rights to a composition are typically held by music publishers who collect royalties and distribute them to the songwriters, although other entities may also hold these rights. (In some cases, the music publisher is also the recording label).

[0004] Most recording labels directly collect royalties on the use of their sound recordings. By contrast, composers and music publishers typically collect royalties on the use of their compositions through the facilities of a copyright collecting agency (or a "performance rights organization"), such ASCAP, BMI, SESAC. For international performances, international collecting societies are typically responsible for collecting royalty payments on behalf of the rights holders. In some instances, two or more publishers or other entities hold ownership rights to a single composition. Furthermore, composition ownership may vary by country.

1

[0005] Media hosting services that allow users to upload multimedia content (e.g., music content and video content) for mass viewing have become increasingly popular in recent years. As the volume of hosted media content continues to grow, the management of ownership rights pertaining to the hosted media content has become an increasingly challenging problem for hosting services. For music content embedded in an audio or video file, for example, the songwriter, the publisher, and the recording label are just some of the different entities that may hold rights to the media content. For appropriate payments to be made to copyright holders, media content must be correctly identified. However, unlike television and radio environments where the content is typically identified prior to airing, media hosting services often handle user-provided media content that may initially be unidentified. Manual identification of such media content becomes onerous when media hosting sites receive thousands or millions of new media uploads every day, and traditional automated mechanisms lack the robustness and scalability required for modern media hosting services. The identification problem becomes even more complex when media uploads include live performances or cover performances that do not precisely match any sound recording known to the media hosting service, and their content is not identified in associated, uploaded, metadata. Thus, a method for identifying new sound recordings of known compositions is needed to facilitate accurate payment of royalties to copyright holders.

#### SUMMARY

[0006] A content recognition system determines ownership rights associated with media files uploaded to a media hosting service. In addition to identifying previously known sound recordings, the content recognition system also beneficially identifies compositions (e.g. songs) that are embodied in recorded live performances or cover performances that do not precisely match previously known sound recordings. Once the content recognition system identifies compositions and/or sound recordings, the content recognition system can determine ownership information pertaining to those compositions and/or sound recordings.

[0007] To identify ownership information pertaining to a composition, a fingerprinting module generates a melody fingerprint for an audio clip. The melody fingerprint represents a melody of the composition embodied in the audio clip by extracting features that are invariant to changes in the key, instrumentation, artistic interpretation or performance, or recording methods or artifacts. Thus, differences in the musical performance, recording, and processing do not substantially affect the melody fingerprint.

[0008] The content recognition system then queries a reference database for a reference melody fingerprint matching the input melody fingerprint. The reference database stores reference melody fingerprints of compositions embodied in a set of reference sound recordings. Each reference melody fingerprint in the reference database is associated with composition ownership information indicating at least one entity having ownership rights to the composition embodied in the reference sound recording from which the reference melody fingerprint was made. Responsive to finding a reference melody fingerprint that matches the input melody fingerprint in the reference database, the content recognition system determines the composition ownership information associated with the matching reference melody fingerprint.

[0009] To identify ownership pertaining to a sound recording, the content recognition system generates an audio fingerprint for the audio clip. Unlike the melody fingerprints discussed above, the audio fingerprints are generally unique to a specific recording, and typically vary with differences in performance, recording, and processing, and thus can be used to distinguish between different recordings of the same composition. The content recognition system then queries the reference database for a reference audio fingerprint that matches the audio fingerprint. Responsive to finding a matching reference audio fingerprint for the audio fingerprint in the reference database, the content recognition system determines the ownership information associated with the sound recording from which the matching reference audio fingerprint was made.

[0010] When a match is found for a melody fingerprint (corresponding to a composition) or an audio fingerprint (corresponding to a sound recording), the content recognition system provides ownership and usage policy information to the hosting service that allows the hosting service to manage the ownership rights. For example, the ownership policy may indicate that the media hosting service should block access to the media file containing the audio clip. Alternatively, the ownership policy may indicate that the media hosting service should monetize the media file containing the audio clip. Under this option, the media hosting service can place advertisements together with the monetized media file, and share the revenues generated from the advertisements with the content owners. In other instances, the ownership policy may indicate that the hosting service should statistically track usage of the media file containing the audio clip.

[0011] To generate the reference database of melody fingerprints, the content recognition system receives a reference sound recording embodying a composition and composition

ownership metadata indicating one or more entities having ownership rights to the composition. The fingerprinting module generates a melody fingerprint from the reference sound recording. The content recognition system then stores the melody fingerprint and the associated composition ownership metadata in the reference database.

**[0012]** Similarly, to generate the reference database of audio fingerprints, the content recognition system generates an audio fingerprint from the reference sound recording and stores the audio fingerprint and the associated composition ownership metadata in the reference database.

[0013] The features and advantages described in the specification are not all inclusive and, in particular, many additional features and advantages will be apparent to one of ordinary skill in the art in view of the drawings, specification, and claims. Moreover, it should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter.

#### BRIEF DESCRIPTION OF THE FIGURES

[0014] FIG. 1 illustrates an embodiment of a content recognition system operating in conjunction with a media hosting service.

[0015] FIG. 2 is a flowchart illustrating an embodiment of a process for generating an audio fingerprint reference database and a melody fingerprint reference database.

[0016] FIG. 3 is an embodiment of a graphical interface displaying various metadata associated with a known composition.

[0017] FIG. 4 is an embodiment of a graphical interface displaying various metadata associated with a known sound recording.

[0018] FIG. 5 is a flowchart illustrating an embodiment of a process for identifying ownership information pertaining to media content and generating claims on behalf of the owners.

[0019] FIG. 6 is a flowchart illustrating an embodiment of a process for implementing an ownership policy associated with hosted media content.

[0020] FIG. 7 is an embodiment of a graphical interface displaying various metadata associated with a claim generated on behalf of a content owner.

[0021] The figures depict various embodiments of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following discussion

that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles of the invention described herein.

#### DETAILED DESCRIPTION

#### **O**VERVIEW

[0022] A content recognition system automatically identifies sound recordings and compositions embedded in user-provided content (e.g., video and/or audio files) uploaded to a media hosting service. As used herein, a "sound recording" and a "composition" are each works that would be recognized as such under copyright law. By automatically identifying both sound recordings and compositions, the content recognition system is capable of detecting the use of both master recordings of a composition (e.g., a studio recording) released by a record label, and any other recordings of a composition, such as cover performances, newly released versions, alternative versions (e.g., acoustic versions) or live performance footage. Once media content is identified, a media hosting service can manage and monetize ownership rights on behalf of the content owners. Thus, for example, the media hosting service can automatically detect and block media content on behalf of the owners, or monetize the media content by placing targeted advertisements together with the media content and distributing royalties to the content owners.

[0023] Automated detection of media content is beneficial, if not necessary, for large scale media rights hosting and management solutions because manual review of all uploaded media content is at best impractical. Furthermore, it is difficult or impossible for humans to remember the ownership rights associated with all possible compositions or sound recordings that may be uploaded to a media hosting service. By automating the detection of sound recordings and compositions in an efficient and scalable manner, the media hosting service can minimize the amount of manual intervention required by rights holders. This automated detection is particularly beneficial for high traffic media hosting services which may receive thousands or millions of new user-provided media uploads every day. This results in increased efficiency in the overall usage of copyrighted works and the payment of royalties for the same, thereby benefiting the copyright holders of such recordings and compositions.

#### SYSTEM ARCHITECTURE

[0024] FIG. 1 illustrates an embodiment of a computing environment 100 for automatically identifying ownership rights pertaining to user-provided media content. The computing environment comprises a media hosting service 110, a content recognition system 130, a user

client 150 and a content owner client 170. In alternative configurations, the computing environment 100 may comprise different or additional components. The clients communicate with the service 110 over a network (not shown). Conventional elements are not necessarily shown in order to avoid obscuring the relevant aspects of this embodiment.

[0025] The media hosting service 110 is configured to enable users to upload, share, and view media content such as video and audio files. In one embodiment, users interact with the media hosting service 110 via a media hosting web site accessible by a web browser executing on a user client 150. Using the user client 150, users can upload user-provided media 151 to the media hosting service 110 and/or view user-requested media 153 hosted by the media hosting service 110 (e.g., via an online interface facilitated over a network). The media hosting service 110 utilizes the content recognition system 130 to identify ownership rights and policies pertaining to the media content hosted by the media hosting service 110. As illustrated, the content recognition system 130 receives an audio clip 141 from the media hosting service 110 and returns the corresponding ownership information 143.

[0026] In one embodiment, the media hosting service 110 comprises an ingest server 112, a media server 114, a rights management engine 116, an advertisement management module 118, a content database 120 and an advertisement database 122. In alternative configurations, the media hosting service 110 may comprise different or additional modules. [0027] The ingest server 112 receives the user-provided media 151 (e.g., an audio or video file) from the user client 150. The ingest server 112 optionally performs audio and/or video processing on the user-provided media 151, for example, to encode the user-provided media 151 in a standardized format. Once uploaded, the user-provided media content 151 is stored in the content database 120. Using the user client 150, a user can request to view hosted media content previously stored in the content database 120. Upon request, the media server 114 streams the user-requested media 153 from the content database 120 to the user client 150 for viewing by a user.

[0028] The advertisement database 122 stores advertising content to be presented along with the user-requested media 153. The advertising content may be in the form of images, videos, audio, text, hyperlinks, or a combination of formats. The advertisement management module 118 manages access to advertising content stored in the advertisement database 122 and determines advertising content to associate with certain user-requested media 153. In one embodiment, the advertisement management module 118 selects advertisements based on the identity of the sound recording embodied and/or the composition performed in the user-

requested media 153 and/or the ownership information associated with the user-requested media 153. For example, the advertisement management module 118 may select an advertisement with a hyperlink to a web site belonging to a record label that has ownership rights to a sound recording embodied in the user-requested media 153. In other embodiments, the advertising content may be selected based on other factors as well, such as user-specific information and preferences.

[0029] The rights management engine 116 manages and enforces ownership policies associated with media content stored in the content database 120. For example, in one embodiment, content owners can set an ownership policy associated with a media item to "track," "monetize," or "block." If the content owner chooses to block content, the rights management engine 116 removes the content from the content database 120 or otherwise prevents the user client 150 from accessing the content. If a content owner chooses to monetize the content, the advertising management module 118 is configured to provide advertisements together with the user-requested media 153, and the rights management engine 116 invokes steps to provide royalties generated from the advertisements to the content owners, typically based on a licensing agreement between the media hosting service and the content owner. If a content owner chooses to track content, statistics related to the content are tracked (e.g., number of views) and the rights management engine 116 provides the tracked statistics to the content owners.

[0030] The media hosting service 110 utilizes the content recognition system 130 to identify ownership rights pertaining to the media content hosted by the media hosting service 110. As illustrated, the content recognition system 130 receives an audio clip 141 from the media hosting service 110 and returns the corresponding ownership information 143. The content recognition system 130 also enables content owners (e.g., record labels and/or publishers) to provide ownership metadata 161 and reference recordings 163 via a content owner client 170. The reference recordings 163 and ownership metadata 161 correspond to media content (e.g., sound recordings or compositions) for which the content owners seek enforcement of their ownership rights. The content recognition system 130 seeks to match the audio clips 141 to one or more reference sound recordings 163 and returns the corresponding ownership information 143 when a match is found.

[0031] In one embodiment, the content recognition system 130 comprises an ingest server 132, a melody fingerprinting module 134, an audio fingerprinting module 136, an indexing module 138, a matching module 140, a melody ID reference database 142, an audio ID

reference database 144, and an ownership database 146. In alternative configurations, the content recognition system may comprise different or additional modules.

[0032] The ingest server 132 receives the reference recordings 163 and ownership metadata 161 from the content owner client 170. The reference recordings are sound recordings for which a record label or other entity has ownership rights. Typically a publisher or other entity will also have ownership rights to a composition embodied in the sound recording. The reference recordings 163 may comprise an audio file encoded in any type of audio codec (e.g., AAC, HE-AAC, MP3, FLAC, ALAC, OGG, WMA, and so forth), and may be an entire audio file (e.g., a recording of a complete musical performance) or a portion of an audio file. The ingest server 132 optionally performs audio processing on the reference recording 163, for example, to encode the reference recording 163 in a standardized format. The ownership metadata 161 typically comprises a text-based file that stores identifying information related to the reference recording 163 and the content owners. The ownership metadata 161 may be organized into various categories or fields such as, for example, artist, title, genre, label, publisher, etc.

[0033] The ingest server 132 is also configured to receive audio clips 141 from the media hosting service 110. Like the reference recordings 163, the audio clips 141 may comprise audio files encoded in any type of audio codec, and may be entire audio files or portions of audio files. Alternatively, the audio clips 141 may comprise the audio portions of video files (or portions of video files). The ingest server 132 optionally performs audio processing on the audio clips 141, for example, to encode the audio clips 141 in a standardized format or to extract the audio portions of video files.

[0034] The audio fingerprinting module 136 generates reference audio fingerprints (also referred to as "audio ID files") for the reference sound recordings 163 provided by content owners. The audio fingerprinting module 136 is configured to generate audio fingerprints that uniquely represent a particular sound recording owned by a record label or other entity. An audio fingerprint compactly represents the audio characteristics of a reference sound recording 163 in a format that can be efficiently compared and matched to other audio fingerprints. The audio fingerprinting module 136 similarly generates audio fingerprints for audio clips 141 received from the media hosting service 110 so that the audio fingerprints can be compared to the reference audio fingerprints.

[0035] The melody fingerprinting module 134 generates reference melody fingerprints (also referred to as "melody ID files") for reference sound recordings provided by content owners.

The melody fingerprints are designed to uniquely represent a composition (which may be embodied in various studio recordings, live performance recordings, or cover performances) based on the melody of the composition. A melody fingerprint compactly represents the melodic characteristics of a reference sound recording in a format that can be efficiently compared and matched to other melody fingerprints. In contrast to an audio fingerprint, which uniquely represents a particular recording of a performance, a melody fingerprint instead represents the melody of a composition that is embodied in the performance, and does so in such a way that variations in key, instrumentation, encoding formats, and other performing, recording, and processing variations do not substantially affect the features of the melody fingerprint. Thus, a melody fingerprint for a live performance of a particular composition will match a melody fingerprint for a studio recording of that composition, while the audio fingerprints for the live and studio performances will not match. The melody fingerprinting module 134 similarly generates melody fingerprints for audio clips 141 received from the media hosting service 110.

[0036] In one embodiment, the melody fingerprinting module 134 detects and compactly represents a sequence of pitch intervals occurring between different time points in the audio clip 141. Melody fingerprinting using a pitch interval representation is further described in U.S. Patent Application No. 12/826,623 entitled "Intervalgram Representation of Audio for Melody Recognition" to Richard Lyon, et al., the contents of which are incorporated by reference herein. In one such embodiment, the audio clip 141 is first processed to generate a Stabilized Auditory Image (SAI). The SAI represents the audio clip 141 using an auditory model designed to simulate how the human auditory system processes and represents sound. Using the SAI, representative features of the audio clip 141 can be extracted that are characteristic of the audio features perceived by the human ear. For example, the perceived dominant musical tones in the input audio clip 141 can be extracted at regular time intervals throughout the input audio clip 141. These extracted tones are largely independent of the particular instrumentation, recording parameters, encoding, or processing used to produce the input audio clip. Each extracted tone can correspond to, for example, one of the twelve notes in the musical scale. Alternatively, a finer scale may be used (e.g., 36 possible tones per octave instead of 12). Thus, the input audio clip 141 is reduced to a representation comprising a sequence of the perceivable tones occurring in the audio clip 141. In order to convert the representation to one invariant to key, the sequence of extracted tones is further processed to determine pitch intervals (e.g., number of whole and or half-steps) between

CA 02817340 2013-05-08

temporally consecutive tones. This sequence of pitch intervals forms a melody fingerprint that is invariant to the musical key. Furthermore, the melody fingerprint is substantially invariant to instrumentation, tempo changes, and other performing, recording, and processing differences. The melody fingerprint representation allows the content recognition system to find reference recordings of compositions that are similar enough that present copyright law may recognize them as embodying the same compositions. Thus, for example, melody fingerprints can be used to accurately match live performances and/or cover performances of a composition to a different reference recording of the composition.

[0037] The indexing module 108 indexes reference audio fingerprints and reference melody fingerprints stored in the audio ID database 144 and the melody ID database 142 respectively. A variety of different indexing schemes can be used, but generally, the indexing scheme is designed to improve the efficiency of comparing and matching an input fingerprint for an audio clip 141 against the reference fingerprints in the reference databases 142, 144. In one embodiment, the indexing module 138 applies a locality sensitive hashing (LSH) bands indexing scheme. In LSH bands indexing, reference fingerprints in the reference data bases 142, 144 are indexed by a set of unique fixed-length byte sequences (i.e., "index keys"), which in one embodiment, are 4 bytes wide. For each index key (i.e., a unique 4-byte sequence), the LSH index stores pointers to all reference fingerprints in the reference databases 142, 144 that contain that particular byte sequence. Thus, for example, if reference fingerprints A, D, and X each include the 4-byte sequence {A5 B1 43 67}, the LSH index stores pointers to the location of reference fingerprints A, D, and X in the reference databases 142, 144 in association with the index key {A5 B1 43 67}. The LSH index can be queried with an index key that is obtained from a fingerprint of an input recording, and can return pointers to the fingerprints of each reference audio clip that is stored in the reference databases 142, 144 that contains that particular index key. LSH bands indexing is just one example of an indexing scheme for indexing the reference fingerprints in the reference databases 142, 144. In alternative embodiments, the indexing module 138 can index reference fingerprints according to a different indexing scheme.

**[0038]** The matching module 140 compares audio and melody fingerprints (ID files) representing the audio clip 141 against reference audio and melody fingerprints in the reference databases 142, 144 to determine a reference sound recording and/or reference composition that best matches the audio clip 141. Based on the outcomes of the matches, different actions will be taken.

[0039] First, an audio ID match indicates that the audio clip 141 matches one of the reference sound recordings. An audio ID match also indicates that a composition embodied in the audio clip 141 matches a composition embodied in the reference sound recording. Thus, for an audio ID match, the matching module 140 typically identifies both sound recording and composition ownership.

[0040] Second, a melody ID match, in the absence of an audio ID match, indicates that a composition embodied in the audio clip 141 matches a composition embodied in at least one of the reference sound recordings, even though there is no sound recording match. An melody ID match may occur, for example, when the audio clip 141 embodies a cover performance or live performance of a composition, while the reference database includes a different recording (e.g., a studio recording) of the composition. Thus, for a melody ID match, in the absence of an audio ID match, the matching module typically identifies only the composition ownership, and does not identify any sound recording ownership.

[0041] The matching module 140 outputs ownership information 143 indicating the identified entities having ownership rights to the audio clip 141, based on the foregoing outcomes. This process is further described below with respect to FIG. 5.

[0042] As discussed above, the matching module 140 determines matches between an input fingerprint for an audio clip 141 and one or more reference fingerprints in the reference databases 142, 144. To improve the efficiency of finding matches, the matching module 140 typically operates in conjunction with the indexing module 138 to first locate candidate reference fingerprints that are most likely to match the fingerprint for the audio clip 141. For example, in one embodiment that utilizes LSH bands indexing, the indexing module 138 divides the input fingerprint for audio clip 141 into a plurality of bands (e.g., 4 byte wide bands) that serve as a set of index keys. The indexing module 138 uses these index keys to query an LSH bands index that returns a set of pointers to candidate reference fingerprints in reference databases 142, 144 that contain at least one of the index keys. Once a set of candidate reference fingerprints is identified, the matching module 140 calculates a match metric between the input fingerprint and each one of the candidate reference fingerprints. The match metric provides a figure of merit as to the quality of the match (e.g., a score, distance, probability, or other measure). For example, in one embodiment, the match metric is a Euclidian distance or a Mahalanobis distance between a fingerprint for the audio clip 141 and one or more candidate reference fingerprints in the reference databases 142, 144. A candidate reference fingerprint is considered to match the fingerprint for the input audio clip

141 when the calculated Euclidian or Mahalanobis distance between the candidate reference fingerprint and the fingerprint for the audio clip 141 is less than a threshold.

[0043] In alternative embodiments, the indexing module 138 or matching module 140 can receive a fingerprint representation of the audio clip 141 from a fingerprint source that is external to the content recognition system 130 rather than from one of the fingerprinting modules 134, 136. In these embodiments, the fingerprinting modules 134, 136 are omitted, and the ingest server 132 is configured to receive fingerprints representative of the audio clip 141 rather than the audio clip 141 itself.

**[0044]** The melody ID reference database 142 stores reference melody fingerprints for a plurality of reference recordings, each representative of a particular composition. Similarly, the audio ID reference database 144 stores reference audio fingerprints for a plurality of reference recordings, each representative of a particular sound recording.

[0045] The ownership database 146 stores ownership metadata identifying the ownership rights associated with the reference sound recordings and/or compositions embodied in the reference recordings 163. Examples of ownership metadata stored in the ownership database 146 will be described in further detail below with reference to FIG. 3-4. The ownership metadata also includes ownership policies indicating how, if at all, the content owner wants to enforce the rights associated with the sound recording and/or composition (e.g., block, track, or monetize). A process for handling different ownership policies will be described in further detail below with reference to FIG. 6.

[0046] In another embodiment, audio and/or melody fingerprints generated for audio clips 141 can be stored as additional reference fingerprints in the melody ID reference database 142 and the audio ID reference database 144. In this way, content owners can supplement the reference databases 142, 144 by providing additional recordings of a particular composition or additional instances of a particular sound recording.

[0047] Each of the various components (alternatively, modules) of the media hosting service 110 and the content recognition system 130, e.g., ingest server 112, media server 114, rights management engine 116, advertisement management system 118, ingest server 132, melody fingerprinting module 134, audio fingerprinting module 136, indexing module 138, and matching module 140 are implemented as part of a server-class computer system with one or more computers comprising a CPU, memory, network interface, peripheral interfaces, and other well known components. The computers themselves preferably run an operating system (e.g., LINUX), have generally high performance CPUs, 1G or more of memory, and

100G or more of disk storage. Of course, other types of computers can be used, including personal and handheld computers when the database is not too big for them, and it is expected that as more powerful computers are developed in the future, they can be configured in accordance with the teachings here. Generally, the modules comprise computerexecutable program instructions stored on a computer readable storage medium (e.g., a hard disk). In operation, the computer-executable program instructions are loaded into a memory and executed by one or more processors included as part of the system. When configured to execute the various operations described herein, a general purpose computer becomes a particular computer, as understood by those of skill in the art, as the particular functions and data being stored by such a computer configure it in a manner different from its native capabilities as may be provided by its underlying operating system and hardware logic. An example of a media hosting service 110 is, for example, the YOUTUBE<sup>TM</sup> website; other media hosting systems are known as well, and can be adapted to operate according to the teachings disclosed herein. It will be understood that the named components of the media hosting service 110 and content recognition system 130 described herein represent one embodiment of the present invention, and other embodiments may include other or differently named components. In addition, other embodiments may lack components described herein and/or distribute the described functionality among the modules in a different manner. Additionally, the functionalities attributed to more than one component can be incorporated into a single component.

[0048] Although only a single media hosting service 110 is illustrated for clarity of description, the content recognition system 130 may be adapted for use by any number of different media hosting services 110. In other alternative embodiments, the content recognition system 130 may be incorporated as a component of the media hosting service 110. Furthermore, the media hosting service 110 may interact with many different user clients 150. Similarly, the content recognition system 130 may interact with any number of content owner clients 170. Furthermore, a single client could be used as both a user client 150 and a content owner client 170.

[0049] In one embodiment, the media hosting service 110 provides the audio clips 141 to the content recognition system 130 as part of the upload flow of the media hosting service 110. Thus, in this embodiment, user-provided media content 151 is identified prior to, concurrently with, or shortly after being stored to the content database 120 and made accessible for download or viewing by other users, if permitted per the ownership metadata

CA 02817340 2013-05-08

found in the ownership rights database 146. In another embodiment, the content recognition system 130 is configured to perform legacy scanning of previously stored content in the content database 120. This embodiment allows, for example, the content recognition system 130 to identify ownership rights pertaining to hosted content that existed prior to the first use of the content recognition system 130 (e.g., before media hosting service 110 gained access to the content recognition system 130). Additionally, legacy scanning is useful for updating ownership information and usage policies associated with a content database 120 as new reference sound recordings 163 and the ever changing ownership metadata 161 become available to the content recognition system 130.

#### OPERATION AND USE

[0050] FIG. 2 illustrates an embodiment of a process performed by the content recognition system 130 for generating the reference databases 142, 144, 146. The content recognition system 130 receives 202 a reference sound recording 163 and/or the ownership metadata 161 corresponding to that sound recording (including desired ownership policies) from a content owner via the content owner client 170. In some instances, only a portion of the reference sound recording 163 and/or ownership metadata 161 is provided by a single content owner. For example, a publisher may provide only ownership metadata associated with a composition without providing a reference sound recording. In other instances, a record label may provide ownership information related to a sound recording without identifying the publisher(s) having ownership rights to the underlying composition. In yet other instances, a content owner may indicate that it has only partial ownership to a composition (e.g., 50%) ownership) without necessarily identifying the other entities having the remaining ownership rights. To assemble piecemeal information, the content recognition system 130 correlates 204 the received information and combines the information into a set of entries, each corresponding to a single sound recording or composition. Furthermore, composition entries may be linked to one or more sound recording entries that embody the composition. The correlation is typically performed by matching various metadata fields (e.g., song title, artist name, identification numbers, etc.) that are common to the partial information submissions. [0051] The audio fingerprinting module 136 generates 206 a reference audio fingerprint for the reference sound recording and stores 208 the reference audio fingerprint in association with the sound recording ownership metadata. The melody fingerprinting module 134 generates 210 a reference melody fingerprint representing the composition embodied in the reference sound recording and stores 212 the reference melody fingerprint in association with

corresponding composition ownership metadata. Thus, the content recognition system 130 produces both a reference audio fingerprint and a reference melody fingerprint for each reference recording provided.

[0052] FIG. 3 is a graphical interface illustrating an example of an ownership metadata entry associated with a composition. Such a graphical interface may be available, for example, to an administrator of the content recognition system 130, the media hosting service 110, and or a content owner. Alternatively, some or all of the metadata shown in FIG. 3 may be used only internally, and may therefore not be available for display in a graphical interface.

[0053] The ownership metadata is divided into a number of categories, each comprising different identifying fields. For example, in this embodiment, the ownership metadata is categorized into metadata 302, ownership information 304, rights 306, related assets 308, and reference content 310 categories. The metadata category 302 provides various fields identifying the composition including, for example, an identifier field (e.g., CMS asset ID), Type (e.g., composition or sound recording), Provider (i.e., the entity that submitted the reference data), Source, Custom ID, Added (i.e., date/time of submission), ISWC, Title, Category, and Writers. As illustrated, some of the fields may be empty indicating that the information is presently still unknown or incomplete.

[0054] The ownership information category 304 identifies the entities having ownership rights to the composition, the countries where the ownership applies (because ownership may be different between different countries), and a percent or fraction of ownership if applicable (because in some countries, ownership may be split between more than one entity). In the illustrated example, the ownership information indicates that "Publisher A" owns 66.66% of the composition in the United States and "Publisher B" owns 33.34% of the composition in the United States.

[0055] The rights category 306 indicates the ownership policies selected by the content owners ("Owner Policy"), if known, and the policy actually being applied by the hosting service ("Applied Policy"). As explained above, the policies can include, for example, monetize, track, or block. The rights category 306 includes a drop-down box 307 allowing a viewer to select "Match Claim" (as selected in the illustration), or "Embed Claim" (not shown). When "Match Claim" is selected (as illustrated) the ownership policies displayed are those selected and/or applied when a matching composition is detected. In the illustrated example, the owners have selected to "Monetize (and track) if Location of the viewer is the United States" and the hosting service is applying the same policy. If, alternatively, "Embed

Claim" is selected from the drop down box 307, the ownership policies are displayed for a sound recording that embed the composition. This would allow, for example, a publisher to block usage even if a label owning the sound recording chooses to track or monetize.

[0056] The related assets category 308 identifies other assets (e.g., sound recordings) that embed the composition. In the illustrated example, the related assets category identifies a sound recording ("Composition in A Major") that embodies the composition.

[0057] The reference content category 310 identifies reference recordings, if any, provided by the content owners of the composition. Here, none of the publishers have provided a reference recording representative of the composition. However, the composition may still be linked to a reference recording for the purpose of determining composition matches if the location of a reference recording for any of the related assets (e.g., the related sound recording titled "Composition in A Major") is known. The entry illustrated in FIG. 3 is just one example of a metadata entry for a composition. In other embodiments, the entries can have different categories, fields, data, and organizational structures.

[0058] FIG. 4 is graphical interface illustrating an example of an ownership metadata entry associated with a reference sound recording. Similar to the composition ownership metadata shown in FIG. 3, the sound recording ownership metadata may be used only internally, and may therefore not be available for display in a graphical interface. The sound recording ownership metadata is divided into a number of categories, each comprising different identifying fields. For example, in this embodiment, the ownership metadata is categorized into metadata 402, ownership information 404, related assets 408, and reference content 410 categories.

**[0059]** The metadata category 402 provides various information identifying the reference sound recording and includes many of the same fields as the composition metadata discussed above. Additionally, the metadata category 402 may include some fields specific to sound recordings such as, for example, Genre, Label, Audio ISRC, UPC, and GRid.

[0060] The ownership information category 404 indicates one or more entities having ownership rights to the sound recording. In this case, "Label A" owns the sound recording worldwide. The related assets category 408 identifies other assets (e.g., compositions) that the sound recording embodies. In the illustrated example, the sound recording embodies the composition, "Composition in A Major," discussed above with respect to FIG. 3.

[0061] The reference content category 410 identifies one or more reference recordings associated with the sound recording. In the illustrated embodiment, the owner (Label A) has

provided two different reference recordings that can be used by the content recognition system 130 to identify the sound recording. Various identifying fields are provided for each reference recording including, for example, Reference ID, Date (i.e., date/time of submission), Type (audio or video), Provider (i.e., the submitting entity), and Status (active or inactive). The entry illustrated in FIG. 4 is just one example of a metadata entry for a sound recording. In other embodiments, the entries can have different categories, fields, data, and organizational structures.

[0062] FIG. 5 is a flowchart illustrating a process performed by the content recognition system 130 for determining ownership information 143 associated with an audio clip 141. The content recognition system 130 receives 502 the audio clip 141 and generates 504 an audio fingerprint (i.e., audio ID file) representing the audio clip 141. The content recognition system 130 then determines 506 if the audio fingerprint for the audio clip 141 matches a reference audio fingerprint in the audio ID database 144. If an audio fingerprint match is found, a claim is generated 508 on behalf of the content owners of the sound recording. For an audio fingerprint match, a claim is typically generated on behalf of both the sound recording owner (typically a record label) and the composition owner (typically a publisher). As explained above, when an audio fingerprint from a clip 141 matches a reference audio fingerprint, the match allows both the ownership of the sound recording and the ownership of the composition that is embodied in the sound recording to be determined.

[0063] If no audio fingerprint match is found, the content recognition system 130 generates 510 a melody fingerprint representing the underlying melody in the audio clip 141. The content recognitions system 130 then determines 512 if the melody fingerprint for the input audio clip 141 matches a reference melody fingerprint in the melody ID database 142. If a match is found, a claim is generated 514 on behalf of the content owners of the composition that is embodied in the audio clip 141. However, since no audio ID match was found, no claim can be made on behalf of an owner of the sound recording embodied in the audio clip 141.

[0064] If neither an audio ID nor melody ID match is found, then no known match exists 516 for the audio clip 141in the content recognition system 130 and no claims are generated.

[0065] For efficiency, when an audio ID match is found in step 506, it is generally unnecessary to also generate and compare melody fingerprints in steps 510-514. Instead, once a sound recording match is detected, the underlying composition can generally be determined from the sound recording metadata, such as the related assets metadata 408 that

CA 02817340 2013-05-08

identifies the composition that is embodied in the sound recording. In other embodiments, the melody fingerprint can be generated in addition to the audio fingerprint, even if there is match.

[0066] In an alternative embodiment, audio and melody fingerprint matching is performed for every input audio clip 141. In this embodiment, the strengths of the best matching audio and melody fingerprints are considered in determining audio fingerprint and/or melody fingerprint matches. For example, the confidence of an otherwise weak (low confidence) audio fingerprint match may be boosted if a strong (high confidence) melody fingerprint match to the same reference sound recording exists. In this way, an audio fingerprint match may be detected even when the match would not have been apparent from comparing the audio fingerprints alone. In general, weights can be applied to the metrics found for the best matching audio and melody fingerprints, and different ways of combining these weighted metrics can be employed to determine whether the best matching audio and/or melody fingerprint is considered a matching audio and/or melody fingerprint.

[0067] The claims generated on behalf of the content owners invoke the ownership policies associated with the identified media content. FIG. 6 illustrates an example of a method for carrying out ownership policies based on a generated claim (performed, for example, by the rights management engine 116). The rights management engine 116 identifies 602 ownership policies for the media content identified by the content recognition system 130 by accessing the ownership information in the ownership database 146. If the rights management engine 116 determines 604 that 100% of the owners have requested to monetize the content, then the rights management engine 116 takes steps to monetize 606 the content and to proportionately distribute revenues between the content owners. Details of the monetization policy, including revenue distribution, are typically dictated by a licensing agreement between the media hosting service and the one or more content owners. Typically, monetizing content includes streaming targeted advertisements together with the userrequested content, and allocating at least a portion of the revenue generated from the advertisers for distribution to the content owners. If the rights management engine 116 instead determines 604 that less than 100% of the owners requested to monetize the content, the rights management engine 116 next determines 608 if at least one owner requests to block the content. If at least one owner requests to block the content, the content is blocked 610. Blocking may include removing the content from the content database 120, or otherwise preventing a user client 150 from accessing the content. If no owners request blocking the

content, but at least one owner fails to request monetizing the content, the rights management engine 116 will track 612 content usage and provide the tracking data to the owners.

Tracking typically includes collecting statistics related to user requests for the content and providing these statistics to the content owners.

[0068] FIG. 7 is a graphical interface illustrating examples of claim metadata generated by the rights management engine 116 in response to identifying uploaded media content. The metadata indicates that the user-uploaded media content comprises footage of a live performance of "Composition in A Major." No sound recording exactly matches the user-provided content (i.e., no audio ID match was found), but the content recognition system nevertheless determined that the melody in the user-provided content matched a melody fingerprint for the known composition "Composition in A Major." The metadata for the generated claim includes various information pertaining to the user-provided content and matched composition, as well as ownership information and associated claim policies. The metadata illustrated in FIG. 7 is just one example of a metadata entry for a generated claim. In other embodiments, different or additional metadata may be included.

**[0069]** Thus, the content recognition system 130 beneficially acts in conjunction with the media hosting service 110 to identify hosted media content, determine ownership rights, and apply claim policies to enforce the ownership rights. Additionally, the system benefits content owners by providing a platform to monetize their media content. Finally, the system benefits the users of the media hosting service because it allows them access to an expansive library of media content that is licensed for viewing.

[0070] Unlike conventional systems, the content recognition system beneficially utilizes melody recognition to efficiently identify compositions embodied in hosted media content. Thus, the content recognition system is able to detect, for example, known compositions that are embodied in previously unknown or uncatalogued performances, including cover recordings and live recordings. As a result, the content recognition system provides an efficient and scalable solution to the problem of enforcing ownership rights for hosted media content.

[0071] The present invention has been described in particular detail with respect to a limited number of embodiments. Those of skill in the art will appreciate that the invention may additionally be practiced in other embodiments. First, the particular naming of the components, capitalization of terms, the attributes, data structures, or any other programming or structural aspect is not mandatory or significant, and the mechanisms that implement the

invention or its features may have different names, formats, or protocols. Furthermore, the system may be implemented via a different combination of hardware and software from that described. Also, the particular division of functionality between the various system components described herein is merely exemplary, and not mandatory; functions performed by a single system component may instead be performed by multiple components, and functions performed by multiple components may instead be performed by a single component.

[0072] Some portions of the above description present the feature of the present invention in terms of algorithms and symbolic representations of operations on information. These algorithmic descriptions and representations are the means used by those skilled in the art to most effectively convey the substance of their work to others skilled in the art. These operations, while described functionally or logically, are understood to be implemented by computer programs stored in a memory and executed by one or more processors. Furthermore, it has also proven convenient at times, to refer to these arrangements of operations as modules or code devices, without loss of generality.

[0073] Unless specifically stated otherwise as apparent from the present discussion, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0074] Certain aspects of the present invention include process steps and instructions described herein in the form of an algorithm. It should be noted that the process steps and instructions of the present invention could be embodied in software, firmware or hardware, and when embodied in software, could be downloaded to reside on and be operated from different platforms used by real time network operating systems.

[0075] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, magnetic-optical disks, read-only memories (ROMs),

random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, application specific integrated circuits (ASICs), or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus. Furthermore, the computers referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0076] The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may also be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description above.

[0077] Finally, it should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention.

#### CLAIMS

- 1. A computer-implemented method for determining ownership rights to a composition, the method comprising:
  - receiving an input media file, the input media file comprising an audio clip; generating an input melody fingerprint for the audio clip, the input melody fingerprint representing a melody of a first composition that is embodied in the audio clip, the melody fingerprint being invariant to parameters used to perform or record the performance of the composition;
  - querying a reference database for a reference melody fingerprint that matches the input melody fingerprint, each reference melody fingerprint in the reference database representing a melody of a composition that is embodied in a reference sound recording; and
  - if a matching reference melody fingerprint for the input melody fingerprint is found, determining ownership information for the first composition from ownership information that is associated with the matching reference melody fingerprint.
- 2. The method of claim 1, wherein the parameters used to perform or record the performance of the composition include a key in which the composition was performed, an instrument used to perform the composition, or instrumentation used to make the audio clip.
- 3. The method of claim 1, further comprising:

  determining an ownership policy associated with the composition that is associated with the matching melody fingerprint.
- 4. The method of claim 3, further comprising providing the ownership policy to a media hosting service hosting the input media file.
- 5. The method of claim 3, further comprising blocking access to the input media file when the ownership policy indicates that at least one entity having an ownership interest in the composition wishes to block access to the composition.
- 6. The method of claim 3, further comprising monetizing the input media file when the ownership policy indicates that all entities having an ownership interest in the composition wish to monetize the composition.
- 7. The method of claim 6, wherein monetizing the input media file comprises: receiving a request for the input media file from a user of the media hosting service; selecting an advertisement for placement with the input media file;

- providing the input media file to the user together with the selected advertisement; and
- sharing the revenue generated from the advertisement with a content owner of the composition.
- 8. The method of claim 3, further comprising tracking the input media file when the ownership policy indicates that no entities having an ownership interest in the composition wish to block access to the composition and at least one entity having an ownership interest in the composition does not wish to monetize the composition.
- 9. The method of claim 1, wherein the input media file comprises a video clip, and wherein the audio clip comprises an audio track of the video clip.
- 10. A computer-implemented method for determining ownership rights to a sound recording, comprising:
  - receiving an input media file, the input media file comprising an audio clip; generating an input audio fingerprint for the audio clip, the audio fingerprint representing a first sound recording of a first composition that is embodied in the audio clip;
  - querying a reference database for a reference audio fingerprint that matches the input audio fingerprint, each reference audio fingerprint in the reference database representing a sound recording that embodies a composition; and
  - if a matching reference audio fingerprint for the input audio fingerprint in the reference database is found, determining the ownership information for the first sound recording from ownership information that is associated with the matching reference audio fingerprint;
  - if a matching reference audio fingerprint for the input audio fingerprint is not found: generating an input melody fingerprint for the audio clip, the input melody fingerprint representing a melody of the first composition, the melody fingerprint being invariant to parameters used to perform or record the performance of the first composition;
    - querying a reference database for a reference melody fingerprint that matches the input melody fingerprint, each reference melody fingerprint in the reference database representing a melody of a composition that is embodied in a reference sound recording; and

if a matching reference melody fingerprint for the input melody fingerprint is found, determining ownership information for the first composition from ownership information that is associated with the matching reference melody fingerprint.

- 11. The method of claim 10, wherein a matching reference audio fingerprint for the input audio fingerprint is found, further comprising determining ownership information for the first composition from ownership information that is associated with the matching reference audio fingerprint.
- 12. The method of claim 10, wherein the matching reference melody fingerprint is derived from a first recorded performance of the composition, and wherein the matching audio fingerprint is derived from a second recorded performance of the composition that is different from the first recorded performance.
- 13. A computer-implemented method for generating a searchable reference database of ownership rights for audio compositions, the method comprising: receiving a reference sound recording embodying a composition and composition ownership information indicating one or more entities having ownership rights to the composition;
  - generating a melody fingerprint from the reference sound recording, the melody
    fingerprint representing a melody of the composition embodied in the
    reference sound recording in a manner that is invariant to parameters used to
    perform or record the performance of the composition; and
    storing the melody fingerprint and the associated composition ownership information
    in the searchable reference database.
- 14. The method of claim 13, wherein the melody fingerprint is invariant to the key in which the composition is performed, the instrumentation used to make the reference sound recording, or the instrumentation used to record the reference sound recording.
- 15. The method of claim 13, further comprising:

  generating an audio fingerprint representing the reference sound recording;

  receiving reference sound recording ownership information indicating one or more

  entities having ownership rights to the reference sound recording; and

  storing the audio fingerprint and the associated reference sound recording ownership

  information in the searchable reference database.

16. The method of claim 13, wherein the composition ownership information comprises an identifying song title and songwriter of the composition embodied in the reference sound recording.

- 17. The method of claim 13, wherein the composition ownership information further comprises an ownership policy indicating a method for enforcing ownership rights pertaining to user-uploaded media having a melody fingerprint matching the reference melody fingerprint.
- 18. A system for determining ownership rights to a composition, the system comprising: a processor; and
  - a memory coupled to the processor, wherein the memory stores instructions operable to cause the processor to:

receive an input media file, the input media file comprising an audio clip;
generate an input melody fingerprint for the audio clip, the input melody fingerprint
representing a melody of a first composition that is embodied in the audio clip,
the melody fingerprint being invariant to parameters used to perform or record
the performance of the composition;

- query a reference database for a reference melody fingerprint that matches the input melody fingerprint, each reference melody fingerprint in the reference database representing a melody of a composition that is embodied in a reference sound recording; and
- if a matching reference melody fingerprint for the input melody fingerprint is found, determine ownership information for the first composition from ownership information that is associated with the matching reference melody fingerprint.
- 19. The system of claim 18, wherein the parameters used to perform or record the performance of the composition include a key in which the composition was performed, an instrument used to perform the composition, or instrumentation used to make the audio clip.
- 20. The system of claim 18, wherein the memory further stores instructions operable to cause the processor to determine an ownership policy associated with the composition that is associated with the matching melody fingerprint.
- 21. The system of claim 20, wherein the memory further stores instructions operable to cause the processor to provide the ownership policy to a media hosting service hosting the input media file.

22. The system of claim 20, wherein the memory further stores instructions operable to cause the processor to block access to the input media file when the ownership policy indicates that at least one entity having an ownership interest in the composition wishes to block access to the composition.

- 23. The system of claim 20, wherein the memory further stores instructions operable to cause the processor to monetize the input media file when the ownership policy indicates that all entities having an ownership interest in the composition wish to monetize the composition.
- 24. The system of claim 23, wherein the instructions stored in the memory operable to cause the processor to monetize the input media file further comprises instructions stored in the memory operable to cause the processor to:

  receive a request for the input media file from a user of the media hosting service; select an advertisement for placement with the input media file; provide the input media file to the user together with the selected advertisement; and share of revenue generated from the advertisement with a content owner of the composition.
- 25. The system of claim 20, wherein the memory further stores instructions operable to cause the processor to track the input media file when the ownership policy indicates that no entities having an ownership interest in the composition wish to block access to the composition and at least one entity having an ownership interest in the composition does not wish to monetize the composition.
- 26. The system of claim 18, wherein the input media file comprises a video clip, and wherein the audio clip comprises an audio track of the video clip.
- 27. A system for determining ownership rights to a sound recording, the system comprising: a processor; and
  - a memory coupled to the processor, wherein the memory stores instructions operable to cause the processor to:
    - receive an input media file, the input media file comprising an audio clip;
      generating an input audio fingerprint for the audio clip, the audio
      fingerprint representing a first sound recording of a first composition that
      is embodied in the audio clip;

CA 02817340 2013-05-08

query a reference database for a reference audio fingerprint that matches the input audio fingerprint, each reference audio fingerprint in the reference database representing a sound recording that embodies a composition; and if a matching reference audio fingerprint for the input audio fingerprint in the reference database is found, determine the ownership information for the first sound recording from ownership information that is associated with the matching reference audio fingerprint;

- if a matching reference audio fingerprint for the input audio fingerprint is not found:
  - generate an input melody fingerprint for the audio clip, the input melody fingerprint representing a melody of the first composition, the melody fingerprint being invariant to parameters used to perform or record the performance of the first composition;
  - query a reference database for a reference melody fingerprint that matches the input melody fingerprint, each reference melody fingerprint in the reference database representing a melody of a composition that is embodied in a reference sound recording; and
  - if a matching reference melody fingerprint for the input melody fingerprint is found, determine ownership information for the first composition from ownership information that is associated with the matching reference melody fingerprint.
- 28. The system of claim 27, wherein a matching reference audio fingerprint for the input audio fingerprint is found, further comprising determining ownership information for the first composition from ownership information that is associated with the matching reference audio fingerprint.
- 29. The system of claim 27, wherein the matching reference melody fingerprint is derived from a first recorded performance of the composition, and wherein the matching audio fingerprint is derived from a second recorded performance of the composition that is different from the first recorded performance.
- 30. A system for generating a searchable reference database of ownership rights for audio compositions, the system comprising:

  a processor; and

a memory coupled to the processor, wherein the memory stores instructions operable to cause the processor to:

- receive a reference sound recording embodying a composition and composition ownership information indicating one or more entities having ownership rights to the composition;
- generate a melody fingerprint from the reference sound recording, the melody
  fingerprint representing a melody of the composition embodied in the
  reference sound recording in a manner that is invariant to parameters used to
  perform or record the performance of the composition; and
  store the melody fingerprint and the associated composition ownership information in
  the searchable reference database.
- 31. The system of claim 30, wherein the melody fingerprint is invariant to the key in which the composition is performed, the instrumentation used to make the reference sound recording, or the instrumentation used to record the reference sound recording.
- 32. The system of claim 30, wherein the memory further stores instructions operable to cause the processor to:

  generate an audio fingerprint representing the reference sound recording;
  receive reference sound recording ownership information indicating one or more entities having ownership rights to the reference sound recording; and store the audio fingerprint and the associated reference sound recording ownership
- 33. The system of claim 30, wherein the composition ownership information comprises an identifying song title and songwriter of the composition embodied in the reference sound recording

information in the searchable reference database.

34. The system of claim 30, wherein the composition ownership information further comprises an ownership policy indicating a method for enforcing ownership rights pertaining to user-uploaded media having a melody fingerprint matching the reference melody fingerprint.

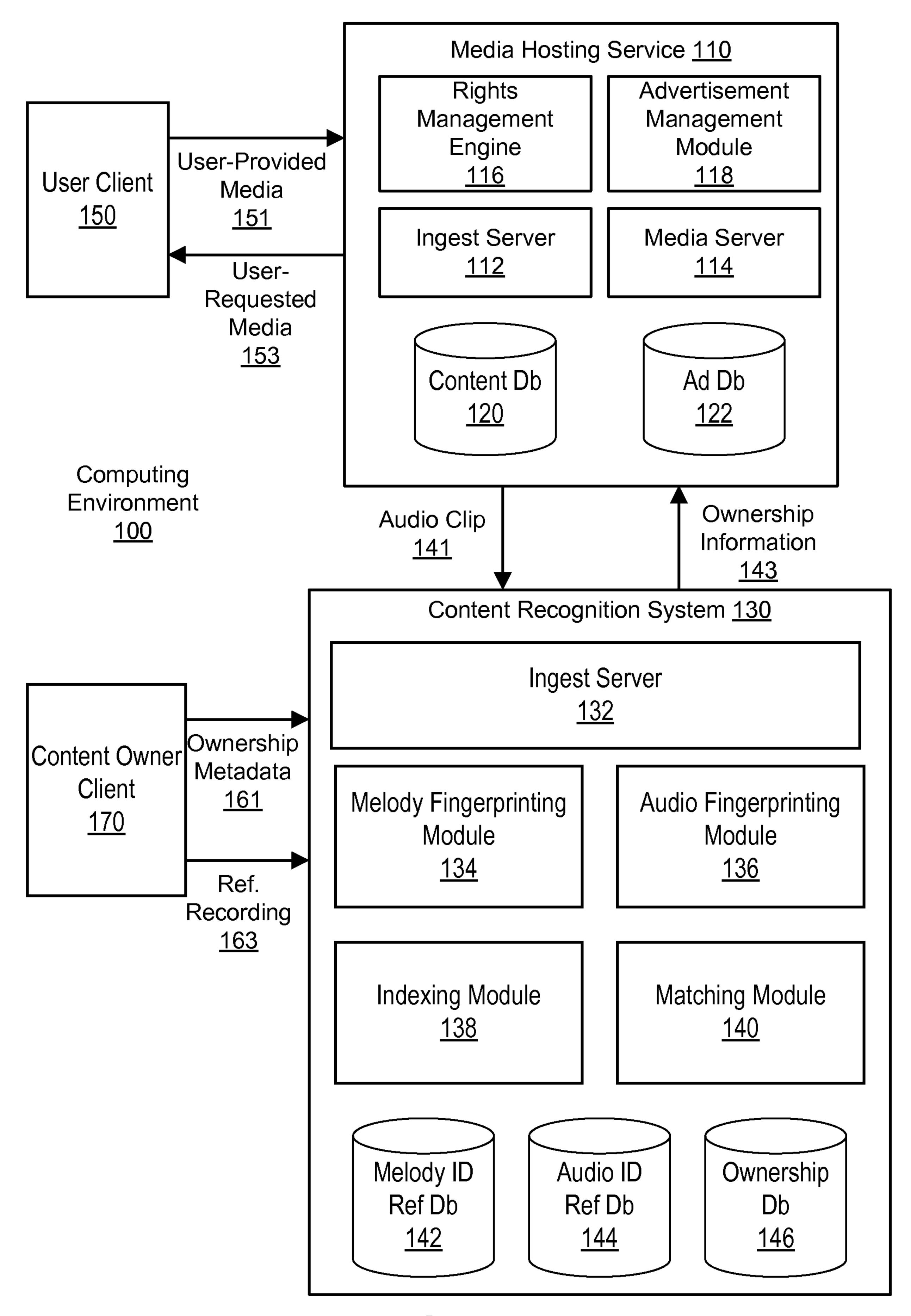
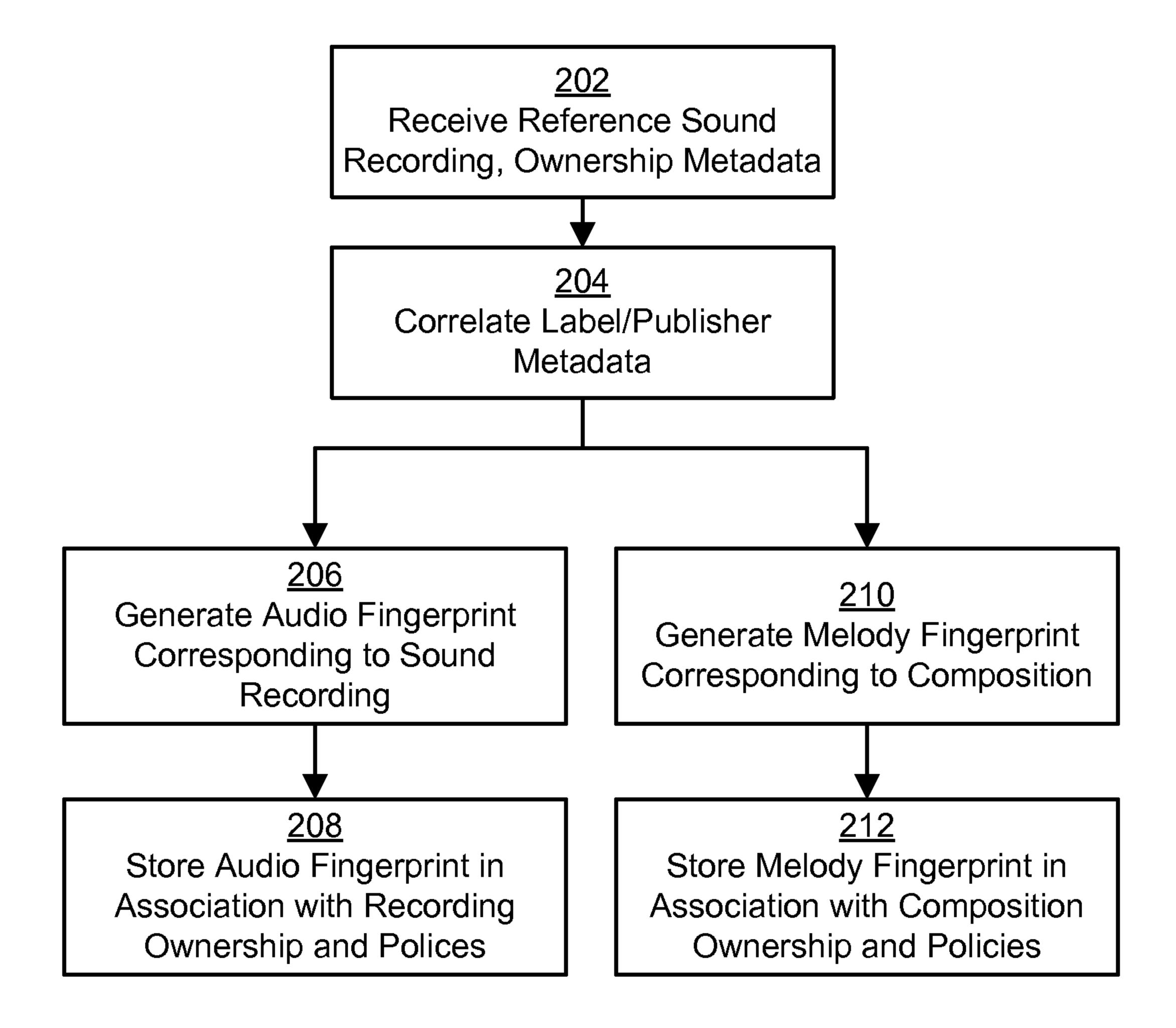


FIG. 1



CA 02817340 2013-05-08

4/7 relationships Edit 61wApziT5aw Mar 08, 2010, jFw7BzeS0u8 Mar 31, 2010, 408 sition in A Major (Composition) Content Label Active Label Active Audio Audio Results Mar Mar Assets embeds: Composition Search ä Reference Album asset Reference Reference Related Provider: Status: Provider: Status: Date: Type: Date: Type: This <<Back Reports Records 70999352 77259987 ⋖ owner Major Rock advanced View by: Claims Album: Artist: Song: Genre: -abel: Audio **GRid:** Viewing: 11:23 yacGydqDXU nd Recording Information Worldwide 2009, Search User Nov 02, de la composition de la compos AssetID: \_\_\_, Sound 2. Recording) Label CMS asset Claims Ownership Metadata this **Overview** View All **Provider:** (Sound Source: Custom Added: Label Type: owns CMS

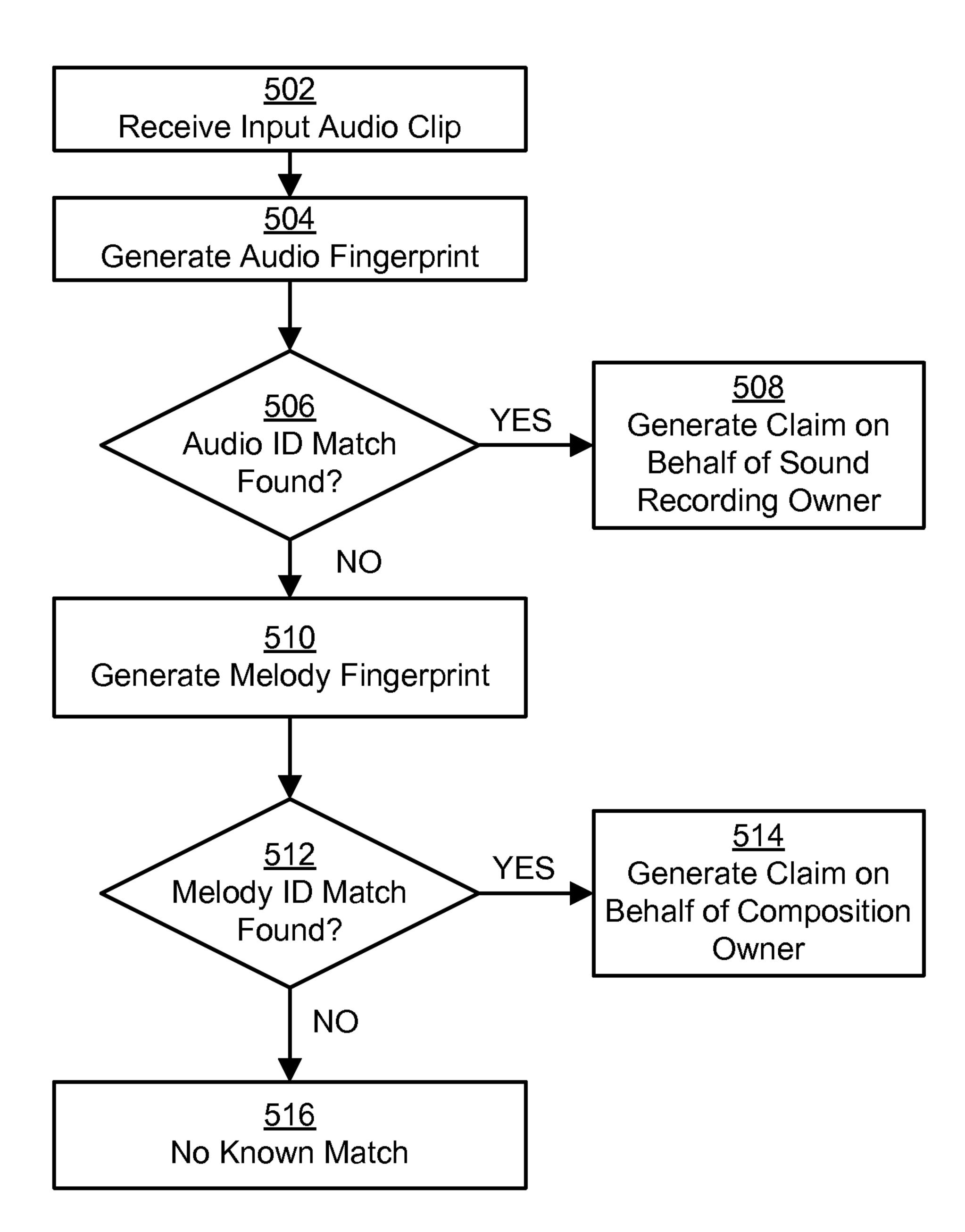


FIG. 5

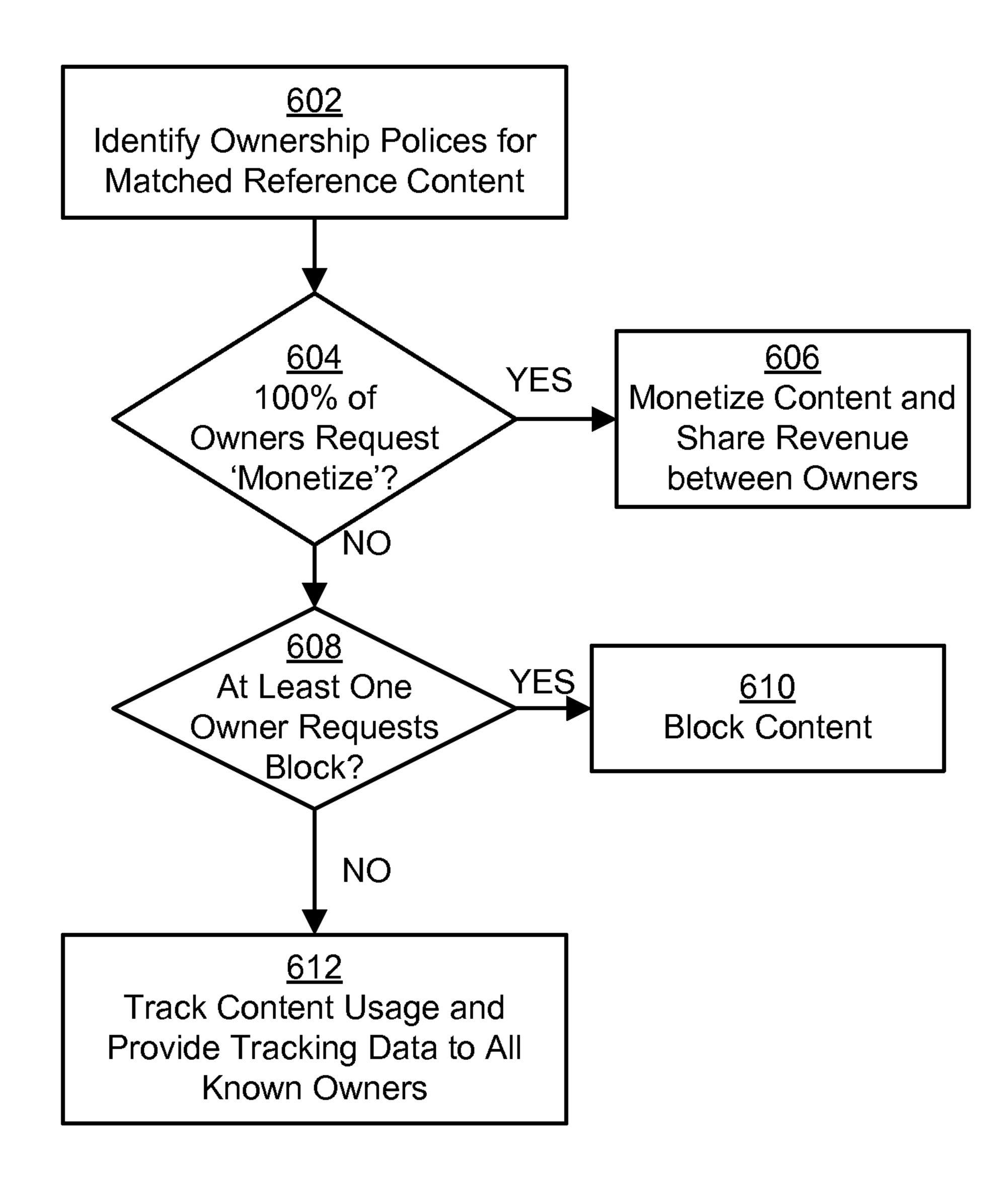


FIG. 6

Overview	Search	Uploads	Claims	Assets	orts		change permissions
Active claims	(61)	Match	against	reference	material		
Possible claims (	(4,024)	eference	j t	rtist A L	06.2010:	Active claim	
Released claims		is not available.		07.Composition in A User—unloaded file	Major	Play user	video at match offset
						Claim Th.	C2tuzQtB_Gi
						Claim	Jul 07, 2010
						air.	Melody Match
						origin: Claim	Audio only
						Type: Platform:	YouTube
				Matched offset	38	source:	
				Match Duration	n 3:42 (0:38–4:20)	Applied	Monetize the following
		Reference o	offset 0:55	Description	722	Policy:	:: Uni
		( )		Views	-	Claim	_
		Custom 1D Title	Composition in	Added	<b>2</b> 01	status:	
		S	Major-Album	ם ה	coma	Release	claim
				Tags	15. 15.		
				Video ID	(more) BrB0B8IobAY		

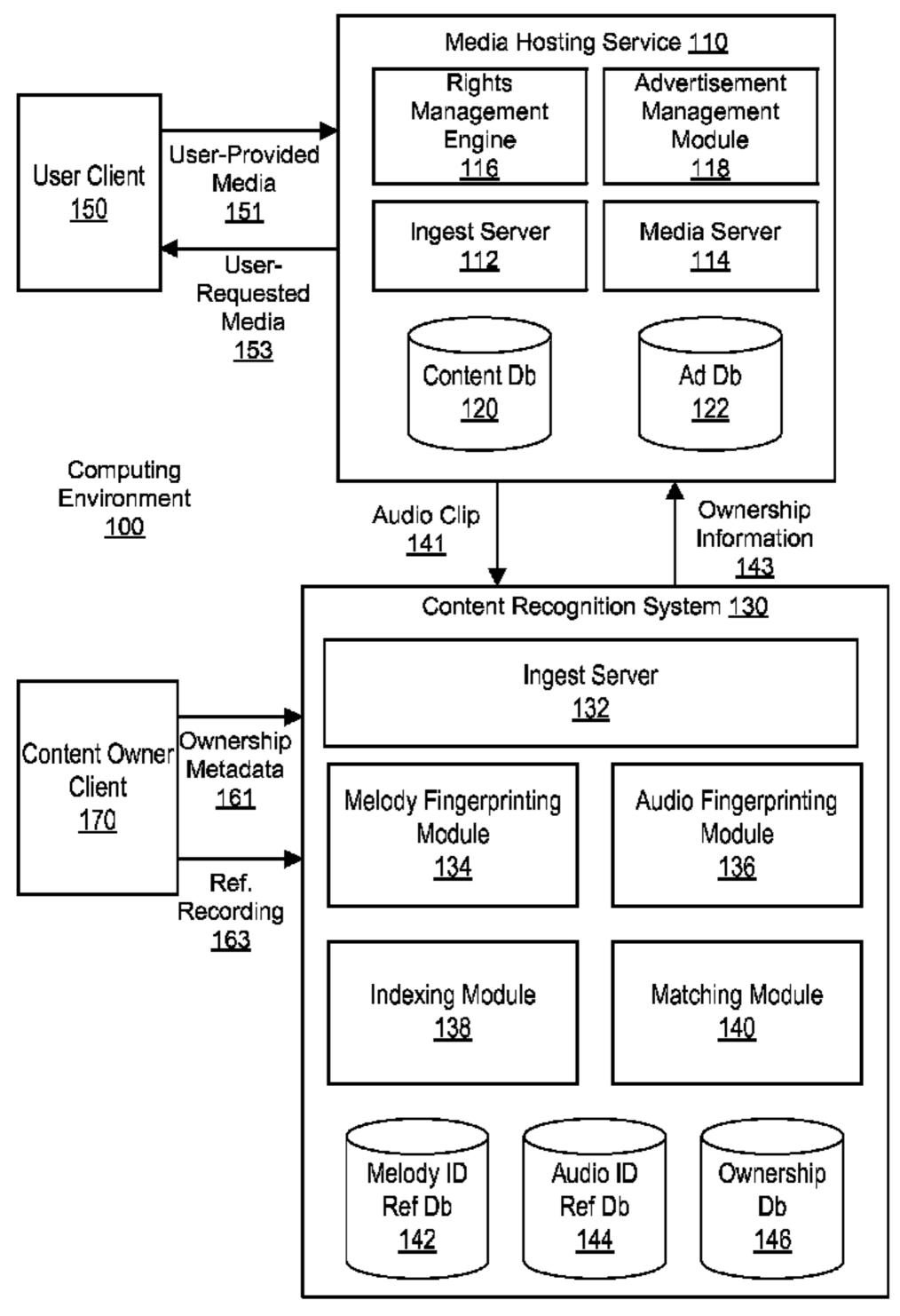


FIG. 1