



- (51) **International Patent Classification:**
G06F 17/30 (2006.01) *G06K 9/62* (2006.01)
- (21) **International Application Number:**
PCT/FI2014/051037
- (22) **International Filing Date:**
22 December 2014 (22.12.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** NOKIA TECHNOLOGIES OY [FI/FI];
Karaportti 3, FI-02610 Espoo (FI).
- (72) **Inventors:** ERONEN, Antti; Rintamäenkatu 13 A 6, FI-33820 Tampere (FI). LEPPÄNEN, Jussi; Kossinkatu 1 B, FI-33590 Tampere (FI). SAARI, Pasi; Mesikämmen 10 C 17, FI-40400 Jyväskylä (FI). LEHTINIEMI, Arto; Lomarantatie 17, FI-33880 Lempäälä (FI).
- (74) **Agents:** NOKIA TECHNOLOGIES OY et al.; Ari Aarnio, IPR Department, Karakaari 7, FI-02610 Espoo (FI).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

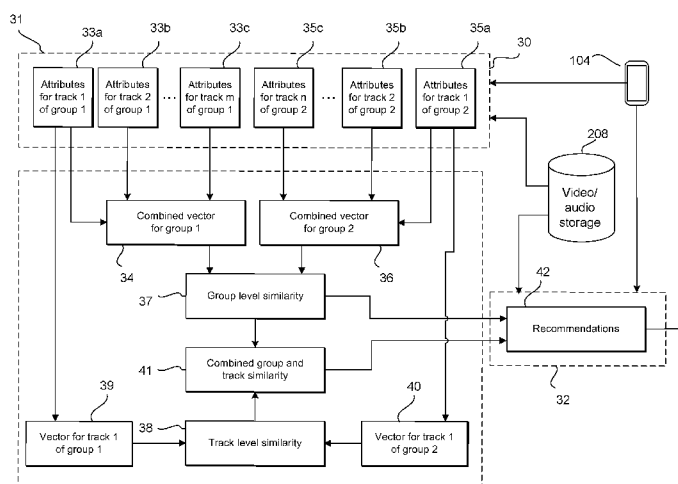
(54) **Title:** SIMILARITY DETERMINATION AND SELECTION OF MUSIC

Figure 3

(57) **Abstract:** A method comprises determining properties of first group of music tracks, e.g. by a first artist, and a second group of music tracks, e.g. by a second artist, based on track level attributes, determining a similarity between the first and second groups based at least in part on determined properties of the tracks, selecting one or more tracks from the second group based at least in part on said similarity, and outputting a list of said selected tracks. The similarity may include group level, track level, and combined group and track level similarities. The track level attributes may be acoustic features extracted from the tracks, tags, metadata or other data, such as keywords extracted from reviews of the tracks. The method may include ranking and/or revising the list based on one or more of user preferences, a user history and/or whether a user plays or skips the selected tracks.



Similarity determination and selection of music

Field

This disclosure relates to determining similarity and similarity-based selection of music tracks. In particular, this disclosure relates to assessing and selecting music tracks from a database based on acoustic similarities.

Background

Audio content databases, streaming services, online stores and media player software applications often include genre classifications, to allow a user to search for tracks to play stream and/or download.

Some databases, services, stores and applications also include a facility for recommending music tracks to a user based on a history of music that they have accessed in conjunction with other data, such as rankings of tracks or artists from the user, history data from other users who have accessed the same or similar tracks in the user's history or otherwise have similar user profiles, metadata assigned to the tracks by experts and/or users, and so on.

Summary

According to an aspect, an apparatus includes a controller and a memory in which is stored computer readable instructions that, when executed by the controller, cause the controller to receive input information regarding one or more music tracks, determine properties of a first plurality of music tracks belonging to a first group of music tracks defined based, at least in part, on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties of the second plurality of music tracks being based on track level attributes, determine a similarity between the first group of music tracks and the second group of music tracks based at least in part on the determined properties of the first plurality of music tracks and the second plurality of music tracks, select one or more music tracks from the second group of music tracks based at least in part on said similarity and output a list of said selected music tracks.

The input information may, for example, indicate a name of a music track, an album, an artist, a performer, a record label, a playlist, a producer, a musical genre, sub-genre or style. For example, where the input information indicates an artist,

- 2 -

the first group of music tracks may contain music tracks by that artist and, optionally, the second group of music tracks may contain music tracks by one or more second artists. If the second group of music tracks contains music tracks by one second artist, then the group level similarity would indicate similarity between
5 the first and second artists.

In some embodiments, the input information may indicate a particular music track, in which case the controller may obtain information regarding one or more of an album, an artist, a performer, a record label, a playlist, a producer, a musical genre,
10 sub-genre or style from metadata associated with the particular music track or from extracting information from a local or remote database, and define the first group based on the obtained information.

The computer readable instructions, when executed by the controller, may further
15 cause the controller to determine a similarity between a first one of said first plurality of music tracks and one of the second plurality of music tracks where said similarity between the first group of music tracks and the second group of music tracks is a combination of a group level similarity based on the properties of the first plurality of music tracks and the properties of the second plurality of music
20 tracks and a track level similarity based on the properties of said one of said first plurality of music tracks and the properties of said one of the second plurality of music tracks.

The track level attributes may include acoustic features extracted from said music
25 tracks and/or at least one of: tags associated with at least some of said music tracks, metadata associated with said music tracks and keywords extracted from text associated with said music tracks.

The properties may include at least one property based on a musical instrument
30 and at least one property based on a musical genre. For example, the properties may include probabilities that a tag for a musical instrument or genre applies to a respective one of the first and second pluralities of music tracks.

The computer readable instructions, when executed by the controller, may further
35 cause the controller to monitor a history of music tracks previously accessed by a user, revise said list of selected music tracks based on the properties of the selected

- 3 -

music tracks in said history and on whether the previously accessed music tracks in the history were played or skipped and output said revised list.

5 The computer readable instructions, when executed by the controller, may further cause the controller to rank the selected music tracks in the list based, at least in part, on user preferences for the properties included in the received input.

10 The computer readable instructions, when executed by the controller, may further cause the controller to rank the selected music tracks in the list based, at least in part, on properties of previously accessed music tracks indicated in a user history. Alternatively, or additionally, the computer readable instructions, when executed by the controller, may further cause the controller to rank the selected music tracks in the list based, at least in part, on a property of a previously accessed music track indicated in a user history, wherein said first plurality of music tracks does not
15 include said property. Where the selected music tracks based at least in part on the user history, the ranking may include adjusting similarities for the selected music tracks according to whether said previously accessed music track, or tracks, indicated in the user history was played or skipped by the user.

20 The computer readable instructions, when executed by the controller, may cause the controller to determine said properties by evaluating first level probabilities that a particular tag applies based on the track level attributes and evaluating a second level probability that the particular tag applies based on the first level probability. For example, the controller may be caused to evaluate the first level
25 probabilities using a first classifier and a second classifier and to evaluate the second level probabilities using a third classifier, wherein the first and third classifiers are non-probabilistic classifiers and the second classifier is a probabilistic classifier.

30 According to another aspect, a method includes receiving input information regarding at least one music track, determining properties of a first plurality of music tracks belonging to a first group of music tracks defined based, at least in part, on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of
35 music tracks and the properties of the second plurality of music tracks being based on track level attributes, determining a similarity between the first group of music tracks and the second group of music tracks based at least in part on the

- 4 -

determined properties of the first plurality of music tracks and the second plurality of music tracks, selecting one or more music tracks from the second group of music tracks based at least in part on said similarity and outputting a list of said selected music tracks.

5

Such a method may further include determining a similarity between one of said first plurality of music tracks and one of the second plurality of music tracks, wherein said similarity between the first group of music tracks and the second group of music tracks is a combination of a group level similarity based on the
10 properties of the first plurality of music tracks and the properties of the second plurality of music tracks and a track level similarity based on the properties of said one of said first plurality of music tracks and the properties of said one of the second plurality of music tracks.

15

The track level attributes may include acoustic features extracted from said music tracks and/or at least one of: tags associated with at least some of said music tracks, metadata associated with said music tracks and keywords extracted from text associated with said music tracks.

20

The method may also include monitoring a history of music tracks previously accessed by a user, revising the list of selected music tracks based on the properties of the previously accessed music tracks in said history and on whether the previously accessed music tracks in the history were played or skipped and outputting said revised list.

25

The method may include ranking the selected music tracks in the list based on one or more of user preferences for the properties included in the received input, properties of previously accessed music tracks indicated in a user history and a property of a previously music track indicated in a user history, wherein said first
30 plurality of music tracks does not include said property. For example, such ranking may include adjusting similarities for the selected music tracks according to whether said previously music track, or tracks, indicated in the user history was played or skipped by the user.

35

In an example embodiment, the method may include determining said properties comprises evaluating first level probabilities that a particular tag applies based on

- 5 -

the extracted acoustic features and evaluating a second level probability that the particular tag applies based on the first level probability.

A computer program comprising computer readable instructions which, when
5 executed by a computer, cause said computer to perform any of the above methods according to said aspect may also be provided.

According to yet another aspect, a non-transitory tangible computer program
product in which is stored computer readable instructions that, when executed by a
10 computer, cause the computer to receive input information regarding one or more music tracks, determine properties of a first plurality of music tracks belonging to a first group of music tracks defined based, at least in part, on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties
15 of the second plurality of music tracks being based on track level attributes, determine a similarity between the first group of music tracks and the second group of music tracks based at least in part on the determined properties of the first plurality of music tracks and the second plurality of music tracks, select one or more music tracks from the second group of music tracks based at least in part on
20 said similarity and output a list of said selected music tracks.

According to a further aspect, an apparatus is configured to receive input
information regarding one or more music tracks, determine properties of a first
plurality of music tracks belonging to a first group of music tracks defined based, at
25 least in part, on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties of the second plurality of music tracks being based on track level attributes, determine a similarity between the first group of music tracks and the second group of music tracks based at least in part on the
30 determined properties of the first plurality of music tracks and the second plurality of music tracks, select one or more music tracks from the second group of music tracks based at least in part on said similarity and output a list of said selected music tracks.

35 According to a yet further aspect, an apparatus includes an interface to receive input information regarding one or more music tracks and to output a list of selected music tracks, an extractor to extract track level attributes associated with a

- 6 -

first plurality of music tracks belonging to a first group of music tracks defined based, at least in part, on the input information and a second plurality of music tracks belonging to a second group of music tracks and to determine properties of the first plurality of music tracks and the second plurality of music tracks based on track level attributes of said music tracks, a similarity determination module to
5 determine a similarity between the first group of music tracks and the second group of music tracks based at least in part on the determined properties of the first plurality of music tracks and the second plurality of music tracks and a recommendation engine to select the selected music tracks from the second group
10 of music tracks based at least in part on said similarity.

Brief description of the drawings

Embodiments will now be described by way of non-limiting examples with reference to the accompanying drawings, of which:

15 Figure 1 is a schematic diagram of a system in which an embodiment may be included;

Figure 2 is a schematic diagram of components of an analysis server according to an embodiment, in the system of Figure 1;

Figure 3 is an overview of a method that may be performed by the analysis
20 server of Figure 2;

Figure 4 is a flowchart of the method shown in overview in Figure 3;

Figure 5 depicts a user interface for use in the method of Figure 3

Figure 6 is a flowchart of a method of extracting acoustic features from an input signal, for use in the method of Figure 4;

25 Figure 7 depicts an example of a blocked and windowed input signal;

Figure 8 depicts an example energy spectrum of a transformed input signal;

Figure 9 depicts a frequency response of an example filter bank for filtering the transformed input signal shown in Figure 8;

Figure 10 depicts an example mel-energy spectrum output from the filter
30 bank represented in Figure 8;

Figure 11 is an overview of a process for obtaining multiple types of acoustic features in the method of Figure 4;

Figure 12 is an overview of a method of tag obtaining probabilities for use in the method of Figure 4;

35 Figure 13 is a flowchart of the method of Figure 12;

Figure 14 shows example distributions for instrument-based tag probabilities;

- 7 -

Figure 15 shows the example probability distributions of Figure 14 after logarithmic transformation;

Figure 16 depicts an example track feature vector generated by the method of Figure 13;

5 Figure 17 is an overview of a recommendation procedure that may be performed as a part of the method of Figure 4;

Figure 18 is an overview of another recommendation procedure that may be performed as a part of the method of Figure 4;

10 Figure 19 is an overview of yet another recommendation procedure that may be performed as a part of the method of Figure 4.

Detailed description

Embodiments described herein concern assessing features of music tracks, determining similarities between music tracks and selecting music tracks based on such similarities, for example, for recommendation to a user.

Referring to Figure 1, an analysis server 100 is shown connected to a network 102, which can be any data network such as a Local Area Network (LAN), Wide Area Network (WAN) or the Internet. The analysis server 100 is configured to receive and process requests for audio content from one or more terminals 104, 105 via the network 102.

In the present example, two terminals 104, 105 are shown, each incorporating media playback hardware and software, such as a speaker (not shown) and/or audio output jack (not shown) and a processor (not shown) that executes a media player software application to stream and/or download audio content over the network 102 and to play audio content through the speaker. As well as audio content, the terminals 104, 105 may be capable of streaming or downloading video content over the network 102 and presenting the video content using the speaker and a display 106. Suitable terminals 104, 105 will be familiar to persons skilled in the art. For instance a smart phone could serve as a terminal 104, 105 in the context of this application although a laptop, tablet or desktop computer may be used instead. Such terminals 104, 105 include music and video playback and data storage functionality and can be connected to the music analysis sever 100 via a cellular network, Wi-fi, Bluetooth® or any other suitable connection such as a cable or wire. Optionally, the display 106 may be a touch screen display.

- 8 -

As shown in Figure 2, the analysis server 100 includes a controller 202, an input and output interface 204 configured to transmit and receive data via the network 102, a memory 206 and a mass storage device 208 for storing video and audio data.

5 The controller 202 is connected to each of the other components in order to control operation thereof. The controller 202 may take any suitable form. For instance, it may be a processing arrangement that includes a microcontroller, plural microcontrollers, a processor, or plural processors.

10 The memory 206 and mass storage device 208 may be in the form of a non-volatile memory such as read only memory (ROM) a hard disk drive (HDD) or a solid state drive (SSD). The memory 206 stores, amongst other things, an operating system 210 and at least one software application 212 to be executed by the controller 202. Random Access Memory (RAM) 214 is used by the controller 202 for the temporary
15 storage of data.

The operating system 210 may contain code which, when executed by the controller 202 in conjunction with the RAM 214, controls operation of analysis server 100 and provides an environment in which the or each software application 212 can run.

20

Software application 212 is configured to control and perform audio and video information processing by the controller 202 of the analysis server 100 to determine similarities between music tracks and, optionally, to generate music recommendations. The operation of this software application 212 according to a
25 first embodiment will now be described in detail, with reference to Figures 3 to 4. In the following, the accessed music tracks are referred to as input signals.

30

Figure 3 is an overview of a procedure for recommending music tracks to the user of the terminal 104, in which the controller 202 acts as an extractor 30 to extract track level attributes of a music track, similarity assessment module 31 and recommendation engine 32. The basis for the recommendation procedure may be information provided by a user of the terminal 104 via the network 102. For example, the user may indicate an artist and/or a track so that similar artists and/or tracks can be identified. Alternatively, or additionally, the user may provide
35 other information on which the recommendation procedure may be based. For example, the user may indicate one or more of an album, a performer, such as a

- 9 -

particular musician, a producer, a record label, a playlist, a musical genre, sub-genre or style, and so on.

Alternatively, or additionally, information regarding music tracks accessed by a user, such as the tracks that have been accessed the greatest number of times, the artist with the greatest number of tracks in a library in the terminal 104, recently accessed tracks or recently purchased tracks, may be used to identify an artist and, optionally, a track, referred to as track 1, to use as a basis for generating recommendations. If only a name for track 1 is indicated, an artist may be identified from metadata of track 1. If only an artist is indicated, track 1 may be a track selected automatically, such as the track by that artist accessed the most times by the user, or a most popular track by that artist as indicated by a remote database, such as a streaming database, rankings for tracks in a digital music store or information obtained from social media.

A first group of music tracks, group 1, is defined based on the information. Where the user has input artist or performer information, or an artist or performer has been identified from other information input by the user or obtained from the user history, the first group may contain multiple tracks by that artist or performer. In another example, if the user has input information identifying an album or record label, the first group may include music tracks from that album or record label.

As noted above, in other examples, information that may be used as that basis can include one or more of an artist, album, a performer, a producer, a record label, a playlist, a musical genre, sub-genre or style, and so on, by defining the first group of music tracks, group 1, according to the basis provided.

Attributes 33a to 33c for a first music track 1 of the first group, group 1, and one or more further tracks 2...m of group 1 are obtained from the data stored in video/audio storage 208 or remote database or obtained from social media information, other websites and so on. The similarity assessment module 31 defines a combined vector 34 for the first group, group 1, based on some or all of the attributes 33a to 33c obtained for tracks 1...m.

Similarly, attributes 35a to 35c are obtained for a plurality of tracks 1...n of the second group, group 2, from which the recommendations are to be drawn, and a combined vector 36 for group 2 is defined. For example, where the

- 10 -

recommendation is to be based on a first artist, album or performer, the second group may contain multiple tracks by a second artist or performer. The second artist may be selected automatically, based on an analysis of attributes 33a to 33c obtained for track 1 of group 1 and/or information from streaming databases, rankings in digital music stores, social media information and so on. For example, such databases may indicate that users who listened to the first artist often listen to certain other artists and one of those other artists may be selected as the second artist and the second group, group 2, defined to include multiple tracks by that second artist.

The similarity assessment module 31 then determines a group level similarity 37, based on the combined vectors 35, 36 for group 1 and group 2, based on the plurality of tracks 1...m of group 1 and the plurality of tracks 1...n of group 2.

The similarity assessment module 31 may also determine one or more track level similarities 38, each based on a vector 39, 40 combining attributes 33a of an individual track of group 1 and the attributes 35a of an individual track of group 2 respectively. A combined group and track similarity 41 may also be computed based on the group level similarity 37 and the track level similarity 38.

One or more of the group level similarity 37 and the combined group and track similarity 41 are input to the recommendation engine 32. The recommendation engine may then select music tracks from the video/audio storage 208 or another database, for example, a remote database accessed via the network 102 or other network, as recommendations 42 of music tracks that the user of the terminal 104 might enjoy, based on the input similarities 37, 41 and, optionally, further input from the user of the terminal 104. The recommendations 42 may be output via the I/O interface 204 and transmitted to the terminal 104 for presentation on the display 106.

Figure 4 is a flowchart showing further detail of the method described above in relation to Figure 3.

Beginning at step s4.0, a basis for generating the recommendations 42 is obtained (step s4.1).

- 11 -

In this example the basis may be provided by the user of the terminal 104. Figure 5 depicts a user interface 50, that may be presented by the display 106, through which the user can provide the basis. The user interface 50 includes fields 51, 52 in which a user can indicate a name for a first artist, and/or a name of a first music track by the first artist, track 1. As noted above, where a user indicates only one of the first artist and the first music track by the first artist, additional artist and track information may be obtained to supplement the user input as basis for the recommendations 42.

Optionally, one or more sliders 53, 54 may be provided to allow the user to indicate preferences for the type of music tracks to be recommended. In this example, sliders 53 are provided for indicating instrument-based preferences and sliders 54 are provided for indicating music genre-based preferences. While Figure 5 depicts sliders 53, 54, in other embodiments, alternative input techniques for obtaining user preferences may be used, such as numerical values indicating relative importance or rankings for the preferences or input arranging the preferences in order of importance to the user.

In steps s4.2 and s4.3, first and second groups of music tracks are defined according to the basis obtained in step s4.1. Examples of ways in which group 1 and group 2 may be defined are discussed above in relation to Figure 3. In one example, where the basis includes information identifying an artist, group 1 may contain one or more music tracks by that artist, while group 2 may contain one or more music tracks by a second artist.

Next, in step s4.4, the controller 202 obtains attributes of a plurality of tracks 1...m of group 1. Such attributes may be obtained from metadata associated with the plurality of tracks 1...m indicating, for example, genre of musical tracks or type of artist, obtained from the data storage 208, or a remote database, or information from a streaming service or digital music store. Additionally, or alternatively, attributes may be obtained by analysing text in social media pages or other webpages. For example, where group 1, is a collection of music tracks by a first artist, an analysis of text on a website for the first artist or a label on which the first artist's music is released and/or reviews of the first artist's music on websites and/or blog pages may be performed and keywords extracted from that text. Another option, which may be combined with one or both of such metadata and

- 12 -

such keywords, is to extract acoustic features from audio data of tracks 1...m of group 1.

Figure 4 shows attributes 33a to 33c being obtained for individual tracks 1...m of group 1 one by one and used to form a vector 34 for group 1, before attributes 35a to 35c for individual tracks 1...n of group are obtained in turn, in steps s4.4 to s4.8. This sequence may be used, for example, where group 2 is selected based on results from the analysis of the tracks of group 1. However, in other embodiments, attributes 33a to 33c, 35a to 35c for the tracks from groups 1 and 2 may be obtained in a different order than shown in Figure 4, or in parallel. In particular, it is not necessary to complete the obtaining of attributes 33a to 33c for the tracks 1...m from group 1 and/or create the vector 34 for group 1 before proceeding to obtain attributes 35a to 35c for tracks 1...n of group 2. Nor is it necessary to obtain the attributes of a particular track before beginning a process for obtaining attributes for another track.

In the following description, the attributes 33a to 33c, 35a to 35c are acoustic features extracted from audio data of tracks 1...m of group 1 and tracks 1...n of group 2, in the form of probabilities 116 that the tracks include a particular instrument or belong to a particular music genre. However, as noted above, the attributes may include one or more of metadata obtained from a database 208, streaming service or digital music store, keywords extracted from text relating to track 1 of group 1 and other audio features of the tracks, as well as, or instead of such tag probabilities 116.

An example procedure for extracting acoustic features and obtaining tag probabilities 116 at step s4.4 will now be described with reference to Figure 6.

Starting at step s6.0, if an input signal for track 1 of group 1 is in a compressed format, such as MPEG-1 Audio Layer 3 (MP3), Advanced Audio Coding (AAC) and so on, the input signal is decoded into pulse code modulation (PCM) data (step s6.1). In this particular example, the samples for decoding are taken at a rate of 44.1 kHz and have a resolution of 16 bits.

The controller 202 may, optionally, resample the decoded input signal at a lower rate, such as 22050 kHz (step s6.2).

- 13 -

An optional “pre-emphasis” process is shown as step s6.3. Since audio signals conveying music tend to have a large proportion of their energy at low frequencies, the pre-emphasis process filters the decoded input signal to flatten the spectrum of the decoded input signal. The relatively low sensitivity of the human ear to low frequency sounds may be modelled by such flattening. One example of a suitable filter for this purpose is a first-order Finite Impulse Response (FIR) filter with a transfer function of $1-0.98z^{-1}$.

At step s6.4, the controller 202 blocks the input signal into frames. The frames may include, for example, 1024 or 2048 samples of the input signal, and the subsequent frames may be overlapping or they may be adjacent to each other according to a hop-size of, for example, 50% and 0%, respectively. In other examples, the frames may be non-adjacent so that only part of the input signal is formed into frames. Figure 7 depicts an example in which an input signal 70 is divided into blocks to produce adjacent frames of about 30 ms in length which overlap one another by 25%. However, frames of other lengths and/or overlaps may be used.

A Hamming window, such as windows 72a to 72d, is applied to the frames at step s6.5, to reduce windowing artifacts. An enlarged portion in Figure 7 depicts a frame 74 following the application of a window to the input signal 70.

At step s7.6, a Fast Fourier Transform (FFT) is applied to the windowed signal 74 to produce a magnitude spectrum of the input signal. An example FFT spectrum is shown in Figure 8. Optionally, the FFT magnitudes may be squared to obtain a power spectrum of the signal for use in place of the magnitude spectrum in the following steps.

The spectrum produced by the FFT at step s6.6 may have a greater frequency resolution at high frequencies than is necessary, since the human auditory system is capable of better frequency resolution at lower frequencies but is capable of lower frequency resolution at higher frequencies. So, at step s6.7, the spectrum is filtered to simulate non-linear frequency resolution of the human ear.

In this example, the filtering at step s6.7 is performed using a filter bank having channels of equal bandwidths on the mel-frequency scale. The mel-frequency

- 14 -

scaling may be achieved by setting the channel centre frequencies equidistantly on a mel-frequency scale, given by the Equation (1),

$$mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

5

where f is the frequency in Hertz.

The output of each filtered channel is a sum of the FFT frequency bins belonging to that channel, weighted by a mel-scale frequency response. The weights for filters in an example filter bank are shown in Figure 9. In Figure 9, 40 triangular-shaped
10 bandpass filters are depicted whose center frequencies are evenly spaced on a perceptually motivated mel-frequency scale. The filters may span frequencies from 30 Hz to 11025 Hz, in the case of the input signal having a sampling rate of 22050 Hz. For the sake of example, the filter heights in Figure 9 have been scaled to unity.

15

Variations may be made in the filter bank in other embodiments, such as spanning the band center frequencies linearly below 1000 Hz, scaling the filters such that they have unit area instead of unity height, varying the number of frequency bands, or changing the range of frequencies spanned by the filters.

20

The weighted sum of the magnitudes from each of the filter bank channels may be referred to as mel-band energies $\tilde{m}_j$, where $j=1\dots N$, N being the number of filters.

In step s6.8, a logarithm, such as a logarithm of base 10, may be taken from the mel-band energies $\tilde{m}_j$, producing log mel-band energies m_j . An example of a log
25 mel-band energy spectrum is shown in Figure 10.

Next, at step s6.9, the MFCCs are obtained. In this particular example, a Discrete Cosine Transform is applied to a vector of the log mel-band energies m_j to obtain
30 the MFCCs according to Equation (2),

$$c_{mel}(i) = \sum_{j=1}^N m_j \cos \left(\frac{\pi \cdot i}{N} \left(j - \frac{1}{2} \right) \right) \quad (2)$$

- 15 -

where N is the number of filters, $i=0,\dots,I$ and I is the number of MFCCs. In an exemplary embodiment, $I=20$.

At step s6.10, further mathematical operations may be performed on the MFCCs produced at step s6.9, such as calculating a mean of the MFCCs and/or time derivatives of the MFCCs to produce the required acoustic features 33a on which the calculation of the tag probabilities 116 will be based.

In this particular embodiment, the acoustic features produced at step s6.10 include one or more of:

- a MFCC matrix for the music track;
- first and, optionally, second time derivatives of the MFCCs, also referred to as “delta MFCCs”;
- a mean of the MFCCs of the music track;
- a covariance matrix of the MFCCs of the music track;
- an average of mel-band energies over the music track, based on output from the channels of the filter bank obtained in step s5.6;
- a standard deviation of the mel-band energies over the music track;
- an average logarithmic energy over the frames of the music track, obtained as an average of $c_{mel}(o)$ over a period of time, for example, using Equation (2); and
- a standard deviation of the logarithmic energy.

The extracted features are then output (step s6.11).

As noted above, the features output at step s6.11 may also include a fluctuation pattern and danceability features for the track, such as:

- a median fluctuation pattern over the song;
- a fluctuation pattern bass feature;
- a fluctuation pattern gravity feature;
- a fluctuation pattern focus feature;
- a fluctuation pattern maximum feature;
- a fluctuation pattern sum feature;
- a fluctuation pattern aggressiveness feature;

- 16 -

- a fluctuation pattern low-frequency domination feature;
- a danceability feature (detrended fluctuation analysis exponent for at least one predetermined time scale);and
- a club-likeness value.

5

The mel-band energies calculated in step s6.8 may be used to calculate one or more of the fluctuation pattern features listed above in step s6.10. In an example method of fluctuation pattern analysis, a sequence of logarithmic domain mel-band magnitude frames are arranged into segments of a desired temporal duration and
10 the number of frequency bands is reduced. A FFT is applied over coefficients of each of the frequency bands across the frames of a segment to compute amplitude modulation frequencies of loudness in a described range, for example, in a range of 1 to 10 Hz. The amplitude modulation frequencies may be weighted and smoothing filters applied. The results of the fluctuation pattern analysis for each segment may
15 take the form of a matrix with rows corresponding to modulation frequencies and columns corresponding to the reduced frequency bands and/or a vector based on those parameters for the segment. The vectors for multiple segments may be averaged to generate a fluctuation pattern vector to describe the music track.

20 Danceability features and club-likeness values are related to beat strength, which may be loosely defined as a rhythmic characteristic that allows discrimination between pieces of music, or segments thereof, having the same tempo. Briefly, a piece of music characterised by a higher beat strength would be assumed to exhibit perceptually stronger and more pronounced beats than another piece of music
25 having a lower beat strength. As noted above, a danceability feature may be obtained at step s6.10 by detrended fluctuation analysis, which indicates correlations across different time scales, based on the mel-band energies obtained at step s6.8.

30 Examples of techniques of club-likeness analysis, fluctuation pattern analysis and detrended fluctuation analysis are disclosed in British patent application no. 1401626.5, as well as example methods for extracting MFCCs. The disclosure of GB 1401626.5 is incorporated herein by reference in its entirety.

35 The features obtained at step s6.10 may include features relating to tempo in beats per minute (BPM), such as:

- 17 -

- an average of an accent signal in a low, or lowest, frequency band;
- a standard deviation of said accent signal;
- a maximum value of a median or mean of periodicity vectors;
- a sum of values of the median or mean of the periodicity vectors;
- 5 - tempo indicator for indicating whether a tempo identified for the input signal is considered constant, or essentially constant, or is considered non-constant, or ambiguous;
- a first BPM estimate and its confidence;
- a second BPM estimate and its confidence;
- 10 - a tracked BPM estimate over the music track and its variation;
- a BPM estimate from a lightweight tempo estimator.

Example techniques for beat tracking, using accent information, are disclosed in US published patent application no. 2007/240558 A1, US patent application no.

- 15 14/302,057, International (PCT) published patent application nos. WO2013/164661 A1 and WO2014/001849 A1, the disclosures of which are hereby incorporated by reference in their entireties.

In one example beat tracking method, described in GB 1401626.5, one or more
20 accent signals are derived from the input signal 70, for detection of events and/or changes in the music track. A first one of the accent signals may be a chroma accent signal based on fundamental frequency F_0 salience estimation, while a second one of the accent signals may be based on a multi-rate filter bank decomposition of the input signal 70.

25 A BPM estimate may be obtained based on a periodicity analysis for extraction of a sequence of periodicity vectors on the basis of the accent signals, where each periodicity vector includes a plurality of periodicity values, each periodicity value describing the strength of periodicity for a respective period length, or "lag". A
30 point-wise mean or median of the periodicity vectors over time may be used to indicate a single representative periodicity vector over a time period of the music track. For example, the time period may be over the whole duration of the music track. Then, an analysis can be performed on the periodicity vector to determine a most likely tempo for the music track. One example approach comprises
35 performing k-nearest neighbours regression to determine the tempo. In this case,

- 18 -

the system is trained with representative music tracks with known tempo. The k-nearest neighbours regression is then used to predict the tempo value of the music track based on the tempi of k-nearest representative tracks. More details of such an approach have been described in Eronen, Klapuri, "Music Tempo Estimation
5 With k -NN Regression", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, Issue 1, pages 50-57, the disclosure of which is incorporated herein by reference.

Chorus related features that may be obtained at step s6.10 include:

- 10 - a chorus start time; and
 - a chorus end time.

Example systems and methods that can be used to detect chorus related features are disclosed in US 2008/236371 A1, the disclosure of which is hereby incorporated
15 by reference in its entirety.

Other features that may be obtained include:

- a duration of the music track in seconds,
 - an A-weighted sound pressure level (SPL);
20 - a standard deviation of the SPL;
 - an average brightness, or spectral centroid (SC), of the music track, calculated as a spectral balancing point of a windowed FFT signal magnitude in frames of, for example, 40 ms in length;
 - a standard deviation of the brightness;
25 - an average low frequency ratio (LFR), calculated as a ratio of energy of the input signal below 100Hz to total energy of the input signal, using a windowed FFT signal magnitude in 40 ms frames; and
 - a standard deviation of the low frequency ratio.

30 Figure 11 is an overview of a process of extracting multiple acoustic features, some or all of which may be obtained in step s6.9 and s6.10. Figure 11 shows how some input features are derived, at least in part, from computations of other input features. The features shown in Figure 11 include the MFCCs, delta MFCCs and mel-band energies discussed above in relation to Figure 6, indicated using bold text
35 and solid lines.

- 19 -

The dashed lines and standard text indicate other features that may be extracted and made available alongside, or instead of, the MFCCs, delta MFCCs and mel-band energies, for use in calculating the tag probabilities 116. For example, as discussed
5 above, the mel-band energies may be used to calculate fluctuation pattern features and/or danceability features through detrended fluctuation analysis. Results of fluctuation pattern analysis and detrended fluctuation analysis may then be used to obtain a club-likeness value. Also as noted above, beat tracking features, labeled as “beat tracking 2” in Figure 11, may be calculated based, in part, on a chroma accent
10 signal from a F_0 salience estimation.

Returning to Figure 6, at step s6.12, tag probabilities 116 and an overall track vector 39 for track 1 of group 1 are evaluated. An overview of an example method for obtaining tag probabilities 116 and creating a track vector 39 is shown in Figure 12.

15 The acoustic features 110 for track 1 of group 1 produced in steps 6.9 and s6.10 are input to first level classifiers 111 to generate first level probabilities for the music track. In this example, the first level classifiers 111 include first classifiers 112 and second classifiers 113 to generate first and second probabilities respectively, the
20 second classifiers 113 being different from the first classifiers 112. In the embodiments to be described below, the first classifiers 112 are non-probabilistic classifiers, while the second classifiers 113 are probabilistic classifiers.

In this embodiment, the first and second classifiers 112, 113 compute first level
25 probabilities that the music tracks include particular instruments and/or belong to particular musical genres. Optionally, probabilities based on other acoustic similarities may be included as will be noted hereinbelow.

The first level probabilities are input to at least one second level classifier 114. In
30 this embodiment, the second level classifier 114 includes a third classifier 115, which may be a non-probabilistic classifier. The third classifier 115 generates the tag probabilities 116 based, at least in part, on the first level probabilities output by the first level classifiers 111 and the second level probabilities output by the second classifiers 113.

35 Figure 13 is a flowchart depicting the method of Figure 12 in more detail. In this particular example, the first and third classifiers 112, 115 are support vector

- 20 -

machine (SVM) classifiers and the second classifiers 113 are based on Gaussian Mixture Models (GMM).

Starting at step s13.0, one, some or all of the extracted features 110 or descriptors
5 obtained in steps s6.9 and s6.10 are selected to be used as input to the first
classifiers 112 (step s13.1) and, optionally, normalised (step s13.2). For example, a
look up table 216 or database may be stored in the memory 206 of the for each of
the tag probabilities to be produced by the analysis server 100, that provides a list
of features to be used to generate each first classifier and statistics, such as mean
10 and variance of the selected features, that can be used in normalisation of the
extracted features 33a. In such an example, the controller 202 retrieves the list of
features from the look up table 216, and accordingly selects and normalises the
listed features for each of the first level probabilities to be generated. The
normalisation statistics for each first level probability in the database may be
15 determined during training of the first classifiers 112.

As noted above, in this example, the first classifiers 112 are SVM classifiers. The
SVM classifiers are trained using a database of music tracks for which information
regarding musical instruments and genre is already available. The database may
20 include tens of thousands of tracks that provide examples for each particular
musical instrument for which a tag probability 116 is to be evaluated.

Examples of musical instruments for which information may be provided in the
database include:

25

- Accordion;
- Acoustic guitar;
- Backing vocals;
- Banjo;
- 30 - Bass guitar;
- Bass synthesizer;
- Brass instruments;
- Glockenspiel;
- Drums;
- 35 - Eggs;

- 21 -

- Electric guitar;
- Electric piano;
- Guitar synthesizer;
- Keyboards;
- 5 - Lead vocals;
- Organ;
- Percussion;
- Piano;
- Saxophone;
- 10 - Stringed instruments;
- Synthesizer; and
- Woodwind instruments.

The training database includes indications of genres that the music tracks belong
15 to, as well as indications of genres that the music tracks do not belong to.
Examples of musical genres that may be indicated in the database include:

- Ambient and new age;
- Blues;
- 20 - Classical;
- Country and western;
- Dance;
- Easy listening;
- Electronica;
- 25 - Folk and roots;
- Indie and alternative;
- Jazz;
- Latin;
- Metal;
- 30 - Pop;
- Rap and hip hop;
- Reggae;

- 22 -

- Rock;
- Soul, R&B and funk; and
- World music.

5 By analysing acoustic features extracted from the music tracks in the training database, for which instruments and/or genre are known, a SVM classifier can be trained to determine whether or not a music track includes a particular instrument, for example, an electric guitar. Similarly, another SVM classifier can be trained to determine whether or not the music track belongs to a particular genre, such as
10 Metal.

In this embodiment, the training database provides a highly imbalanced selection of music tracks, in that a set of tracks for training a given SVM classifier includes many more positive examples than negative ones. In other words, for training a
15 SVM classifier to detect the presence of a particular instrument, a set of music tracks for training in which the number of tracks that include that instrument is significantly greater than the number of tracks that do not include that instrument will be used. Similarly, in an example where a SVM classifier is being trained to determine whether a music track belongs to a particular genre, the set of music
20 tracks for training might be selected so that the number of tracks that belong to that genre is significantly greater than the number of tracks that do not belong to that genre.

An error cost may be assigned to the different first level probabilities produced by
25 the first classifiers 112 to take account of the imbalances in the training sets. For example, if a minority class of the training set for a particular first classification includes 400 songs and an associated majority class contains 10,000 tracks, an error cost of 1 may be assigned to the minority set and an error cost of 400/10,000 may be assigned to the majority class. This allows all of the training data to be
30 retained, instead of downsampling data of the negative examples.

New SVM classifiers can be added by collecting new training data and training the new classifiers. Since the SVM classifiers are binary, new classifiers can be added alongside existing classifiers.

35 As mentioned above, the training process can include determining a selection of one or more acoustic features to be used input for particular first classifiers 112 and

- 23 -

statistics for normalising those features. The number of features available for selection, M , may be much greater than the number of features selected for determining a particular first classification, N ; that is, $M \gg N$. The selection of features to be used is determined iteratively, based on a development set of music tracks for which the relevant instrument or genre information is available, as follows.

Firstly, to reduce redundancy, a check is made as to whether two or more of the features are so highly correlated that the inclusion of more than one of those features would not be beneficial. For example, if two features have a correlation coefficient that is larger than 0.9, then only one of those features is considered available for selection.

The feature selection training starts using an initial selection of features, such as the average MFCCs for music tracks in the development set or a single “best” feature for a given first classification. For instance, a feature that yields the largest classification accuracy when used individually may be selected as the “best” feature and used as the sole feature in an initial feature selection.

An accuracy of the first classification based on the initial feature selection is determined. Further features are then added to the feature selection to determine whether or not the accuracy of the first classification is improved by their inclusion. Features to be tested for addition to the selection of features may be chosen using a method that combines forward feature selection and backward feature selection in a sequential floating feature selection. Such feature selection may be performed during the training stage, by evaluating the classification accuracy on a portion of the training set.

In each iteration, each of the features available for selection is added to the existing feature selection in turn, and the accuracy of the SVM classifier with each additional feature is determined. The feature selection is then updated to include the feature that, when added to the feature selection, provided the largest increase in accuracy for the development set.

After a feature is added to the feature selection, the accuracy of the SVM classifier is reassessed, by generating probabilities based on edited feature selections in which each of the features in the feature selection is omitted in turn. If it is found

- 24 -

that the omission of one or more features provides an improvement in the accuracy of a generated probability, then the feature that, when omitted, leads to the biggest improvement in accuracy is removed from the feature selection. If no improvements are found when any of the existing features are left out, but the accuracy does not change when a particular feature is omitted, that feature may also be removed from the feature selection in order to reduce redundancy.

The iterative process of adding and removing features to and from the feature selection continues until the addition of a further feature no longer provides a significant improvement in the accuracy of the SVM classifier. For example, if the improvement in accuracy falls below a given percentage, the iterative process may be considered complete, and the current selection of features is stored in the lookup table 216, for use in selecting features in step s13.1.

The normalisation of the selected features 110 at step s13.2 is optional. Where provided, the normalization of the selected features 110 at step s13.2 may potentially improve the accuracy of the first classifiers 112.

In another embodiment, at step s13.1, a linear feature transform may be applied to the available features 110 obtained in steps s6.9 and s6.10, instead of performing the feature selection procedure described above. For example, a Partial Least Squares Discriminant Analysis (PLS-DA) may be used to obtain a linear combination of features for calculating a corresponding first classification. Instead of using the above iterative process to select N features from the set of M features, a linear feature transform is applied to an initial high-dimensional set of features to arrive at a smaller set of features which provides a good discrimination between classes. The initial set of features may include some or all of the available features, such as those shown in Figure 11, from which a reduced set of features can be selected based on the result of the transform.

The PLS-DA transform parameters may be optimized and stored in a training stage. During the training stage, the transform parameters and its dimensionality may be optimized for each tag or output classification, such as an indication of an instrument or a genre. More specifically, the training of the system parameters can be done in a cross-validation manner, for example, as five-fold cross-validation, where all the available data is divided into five non-overlapping sets. At each fold, one of the sets is held out for evaluation and the four remaining sets are used for

- 25 -

training. Furthermore, the division of folds may be specific for each tag or classification.

For each fold and each tag or classification, the training set is split into 50%-50% inner training-test folds. Then, the PLS-DA transform may be trained on the inner training-test folds and the SVM classifier may be trained on the obtained dimensions. The accuracy of the SVM classifier using the transformed features transformed may be evaluated on the inner test fold. It is noted that, when a feature vector (track) is tested, it is subjected to the same PLS-DA transform, the parameters of which were obtained during training. This manner, an optimal dimensionality for the PLS-DA transform may be selected. For example, the dimensionality may be selected such that the area under the receiver operating characteristic (ROC) curve is maximized. In one example embodiment, an optimal dimensionality is selected among candidates between 5 to 40 dimensions. Hence, the PLS-DA transform is trained on the whole of the training set, using the optimal number of dimensions, and then used in training the SVM classifier. As an alternative to PLS-DA, other feature transforms such as Linear Discriminant Analysis (LDA), Principal Components Analysis (PCA), or Independent Component Analysis (ICA) could be used.

In the following, an example is discussed in which the selected features on which the first classifications are based are the mean of the MFCCs of track 1 of group 1 and the covariance matrix of the MFCCs of track 1 of group 1, although in other examples alternative and/or additional features, such as the other features shown in Figure 11, may be used.

At step s13.3, the controller 202 defines a feature vector based on each set of selected features 110 or selected combination of features 110 for track 1 of group 1. The feature vectors may then be normalized to have a zero mean and a variance of 1, based on statistics determined and stored during the training process.

At step s13.4, the controller 202 generates one or more first probabilities that track 1 of group 1 has a certain characteristic, based on the feature vector or vectors. The first classifiers 112 are used to calculate respective probabilities for each feature vector defined in step s13.3. In this manner, the number of first classifiers 112 corresponds to the number of tag probabilities 116 to be predicted for the music track.

- 26 -

In this particular example, a probability is generated by a respective first classifier 112 for each instrument tag probability and for each genre tag probability to be predicted for the music track, based on the mean MFCCs and the MFCC covariance matrix. In addition, a probability may be generated by the first classifiers 112 based on whether the music track is likely to be an instrumental track or a vocal track. Also, for vocal tracks, another probability may be generated by the first classifiers 112 based on whether the vocals are provided by a male or female vocalist. In other embodiments, the controller 202 may generate only one or some of these probabilities and/or calculate additional probabilities at step s13.4. The different classifications may be based on respective selections of features from the available features 110 selected in step s13.1.

The first classifiers 112 may use a radial basis function (RBF) kernel K , defined as:

$$K(\vec{u}, \vec{v}) = e^{-\gamma \|\vec{u} - \vec{v}\|^2} \quad (3)$$

where the default γ parameter is the reciprocal of the number of features in the feature vector, $\vec{u}$ is the input feature vector and $\vec{v}$ is a support vector.

The output from the first classifiers 112 may be in the form of first classifications based on an optimal predicted probability threshold that separates a positive prediction from a negative prediction for a particular tag probability, based on the probabilities computed by the first classifiers 112. The setting of an optimal predicted probability threshold may be learned in the training procedure to be described later below. Where there is no imbalance in data used to train the first classifiers 112, the optimal predicted probability threshold may be 0.5. However, where there is an imbalance between the number of tracks providing positive examples and the number of tracks provided negative examples in the training sets used to train the first classifiers 112, the threshold p_{thr} may be set to a prior probability of a minority class P_{min} in the first classification, using Equation (4) as follows:

$$p_{thr} = P_{min} = \frac{n_{min}}{n_{maj}} \quad (4)$$

- 27 -

where, in the set of n tracks used to train the SVM classifiers, n_{min} is the number of tracks in the minority class and n_{maj} is the number of tracks in a majority class. The prior probability P_{min} may be learned as part of the training of the SVM classifiers.

5

Probability distributions for examples of possible first classifications, based on an evaluation of a number n of tracks, are shown in Figure 14. The nine examples in Figure 14 suggest a correspondence between a prior probability for a given first classification and its probability distribution based on the n tracks. Such a
 10 correspondence is particularly marked where the SVM classifier was trained with an imbalanced training set of tracks. Consequently, the predicted probability threshold for the different examples vary over a considerable range.

Optionally, a logarithmic transformation may be applied to the probabilities
 15 produced by the first classifiers 112 at step s13.4, so that the probabilities are on the same scale and the optimal predicated probability threshold may correspond to a predetermined value, such as 0.5.

Equations (5) to (7) below provide an example normalization which adjusts the
 20 optimal predicted probability threshold to 0.5. Where the probability output by a SVM classifier is p and the prior probability P of a particular tag being applicable to a track is greater than 0.5, then the normalized probability p_{norm} is given by:

$$p_{norm} = 1 - (1 - p)^L \quad (5)$$

25

$$\text{where } L = \frac{\log(0.5)}{\log(1 - P)} \quad (6)$$

Meanwhile, where the prior probability P is less than or equal to 0.5, then the normalised probability p_{norm} is given by:

30

$$p_{norm} = p^{L'} \quad (7)$$

$$\text{where } L' = \frac{\log(0.5)}{\log(P)} \quad (8)$$

- 28 -

Figure 15 depicts the example probability distributions of Figure 14 after a logarithmic transformation has been applied, on which optimal predicated probability thresholds of 0.5 are marked.

5 The probabilities output by the first classifiers 112 correspond to a normalised probability p_{norm} that a respective one of the tags to be considered applies to track 1 of group 1. The first classifications may include probabilities p_{inst1} that a particular instrument is included in the music track and probabilities p_{gen1} that the music track belongs to a particular genre.

10

In steps s13.5 to s13.6, further first level probabilities are generated for the input signal by the second classifiers 113, based on the MFCCs and other parameters produced in step s4.4. Although Figure 13 shows steps s13.3 and s13.6 being performed in sequence, in another embodiment steps s13.5 and s13.6 may be performed before, or in parallel, with steps s13.4 and s13.5.

15

In this particular example, the acoustic features 110 of track 1 of group 1 on which the second classifications are based are the MFCC matrix for and the first time derivatives of the MFCCs, and probabilities are generated p_{inst2} , p_{gen2} for each instrument tag (step s13.5) and for each musical genre tag (step s13.6) to be predicted. Optionally, further probabilities may be generated based on whether the music track is likely to be an instrumental track or a vocal track and, for vocal tracks, another probability may be generated based on whether the vocals are provided by a male or female vocalist. In other embodiments, the controller 202 may generate only one or some of these second classifications and/or calculate additional second classifications at steps s13.5 and s13.6.

20

25

In this example, the second classifiers 113 compute probabilities p_{inst2} , p_{gen2} using probabilistic models that have been trained to represent the distribution of features extracted from audio signals captured from each instrument or genre. Such training can be performed using an expectation maximisation algorithm that iteratively adjusts the model parameters to maximise the likelihood of the model for a particular instrument or genre generating features matching one or more input features in the captured audio signals for that instrument or genre. The parameters of the trained probabilistic models may be stored in a database, for example, in the database 208 of the analysis server, or in remote storage that is accessible to the analysis server 100 via a network, such as the network 102.

30

35

- 29 -

For each instrument or genre, likelihoods L_{yes} , L_{no} are evaluated that the respective probabilistic model could have generated the selected or transformed features from the input signal. In this embodiment, in steps s13.5 and s13.6, the instrument-

5 based probabilities p_{inst2} are produced by the second classifiers 113 using first and second Gaussian Mixture Models (GMMs), based on the MFCCs and their first time derivatives calculated in step s13.5. Meanwhile, the probabilities p_{gen2} that the music track belongs to a particular musical genre are produced by the second classifiers 113 using third GMMs. However, the first and second GMMs used to

10 compute the instrument-based probabilities p_{inst2} may be trained and used slightly differently from third GMMs used to compute the genre-based probabilities p_{gen2} , as will now be explained.

The first and second GMMs used in step s13.5 may have been trained with an

15 Expectation Maximisation algorithm using a training set of examples which are known either to include the instrument and examples which are known to not include the instrument. For each track in the training set, MFCC feature vectors and their corresponding first time derivatives are computed. The MFCC feature vectors for the examples in the training set that contain the instrument are used to

20 train a first GMM for that instrument, while the MFCC feature vectors for the examples that do not contain the instrument are used to train a second GMM for that instrument. In this manner, for each instrument to be tagged, two GMMs are produced. The first GMM is for a track that includes the instrument and is used to evaluate the likelihood L_{yes} , while the second GMM is for a track that does not

25 include the instrument and is used to evaluate the likelihood L_{no} . In this example, the first and second GMMs each contain 64 component Gaussians.

The first and second GMMs may then be refined by discriminative training, for example using maximum mutual information (MMI) criterion on a balanced

30 training set where, for each instrument to be tagged, the number of example tracks that contain the instrument is equal to the number of example tracks that do not contain the instrument.

Returning to step s13.5, the two likelihoods L_{yes} , L_{no} are computed based on the

35 first and second GMMs and the MFCCs for the music track. The first is the likelihood L_{yes} that the corresponding instrument is included in the music track, while the second is the likelihood L_{no} that the instrument is not included in the

- 30 -

music track. The first and second likelihoods L_{yes} , L_{no} may be computed in a log-domain, and then converted to a linear domain.

The first and second likelihoods L_{yes} , L_{no} are then mapped to a tag probability p_{inst2} of the instrument being included in the track, as follows:

$$p_{inst2} = \frac{L_{yes}}{(L_{yes} + L_{no})} \quad (9)$$

As noted above, the third GMMs, used to compute genre-based probabilities p_{gen2} , are trained differently to the first and second GMMs. For each genre to be considered, a third GMM is trained based on MFCCs for a training set of tracks known to belong to that genre. One third GMM is produced for each genre to be considered. In this example, the third GMM includes 64 component Gaussians.

In step s13.6, for each of the genres to be considered, a likelihood L_{gen} is computed for the track 1 of group 1 belonging to that genre, based on the likelihood of each of the third GMMs being capable of outputting the MFCC feature vector of the music track. For example, to determine which of the eighteen genres in the list hereinabove might apply to the music track, eighteen likelihoods would be produced.

The genre likelihoods L_{gen} are then mapped to probabilities p_{gen2} , as follows:

$$p_{gen2}(i) = \frac{L(i)}{\sum_{j=1}^m L(j)} \quad (10)$$

where m is the number of genre tags to be considered.

In another embodiment, the first and second GMMs may be trained and used in the manner described above for the third GMMs. In yet further embodiments, the GMMs used for analysing genre may be trained and used in the same manner, using either of techniques described in relation to the first, second and third GMMs above.

- 31 -

The first classifications p_{inst1} and p_{gen1} and the second classifications p_{inst2} and p_{gen2} for track 1 of group 1 are then normalized to have a mean of zero and a variance of 1 (step s13.7) and collected to form a feature vector for input to the one or more second level classifiers 115 (step s13.8). In this particular example, the second level classifiers 115 include third classifiers 116, as noted above, and the third classifiers 116 are non-probabilistic classifiers, such as SVM classifiers trained in a similar manner to that described above in relation to the first classifiers 112. At the training stage, the first classifiers 112 and the second classifiers 113 may be used to output probabilities p_{inst1} , p_{gen1} , p_{inst2} , p_{gen2} for the training sets of example music tracks from the database. The outputs from the first and second classifiers 112, 113 are then used as input data to train the third classifiers 116.

The third classifiers 116 determine second level probabilities p_{inst3} for whether track 1 of group 1 contains a particular instrument and/or second level probabilities p_{gen3} for whether track 1 of group 1 belongs to a particular genre (step s13.9). In this example, where the third classifiers 116 are SVM classifiers, the second level probabilities p_{inst3} , p_{gen3} are generated in a similar manner to the first level probabilities p_{inst1} , p_{gen1} computed by the first classifiers 112.

The second level probabilities p_{inst3} , p_{gen3} are then log normalised (step s13.11), as described above in relation to the first level probabilities p_{inst1} , p_{gen1} from the first classifiers 112, and output as the tag probabilities 116 at step s13.11.

Optionally, tags based on the tag probabilities 116 may be associated with the music track at step s13.11. For example, the tag probabilities 116 exceed a probability threshold, such as 0.5 for normalised probabilities, tags corresponding to the instruments and/or genres may be stored in a database entry for the music track 208.

The track vector 39 is then generated at step s13.12 from the tag probabilities 116 output at step s13.11 and normalised. An example of a track vector 39 is shown in Figure 16. In this particular example, the track vector 39 reflects non-zero probabilities for the music track being a rock song including lead and backing vocals, bass, drums, electric guitar, keyboard and percussion.

- 32 -

Alternatively, or additionally, some or all of the first level and/or second level probabilities p_{inst1} , p_{gen1} , p_{inst2} , p_{gen2} , p_{inst3} , p_{gen3} themselves and/or the features 110 extracted at step s6.9 and s6.10 may be output for further analysis and/or storage.

5 The tag probability calculation process ends at step s13.13.

Returning to Figure 4, steps s4.4 and steps s4.5, including the processes of Figures 6 and 13, are repeated to obtain attributes 33b, 33c for tracks 2 to m of group 1, until no further tracks of group 1 remain to be analysed (step s4.5). However, the
 10 repetition of step s13.12 to create track vectors 39 for tracks 2...m of group 1 is optional. In this manner, attributes 33a to 33c are obtained for tracks 1...m of group 1, while a track vector 39 may be generated for one, some or all of the tracks 1...m of group 1.

15 The combined vector 34 for group 1 is then created (step s4.6), based on the tag probabilities 116 generated at step s4.4 for tracks 1...m of group 1. For example, the feature vector 34 may be created by summing the tag probabilities 116 for all of the analysed tracks 1 to m of group 1 and, optionally, normalising the sum.

20 Next, for each of the tracks 1...n of group 2 to be analysed, the attributes 35a to 35c are obtained in turn (steps s4.7, s4.8) and, for at least one of the tracks 1...n of group 2, a track vector 40 is created, as described above in relation to steps s4.4 to s4.6 and Figures 6 and 13, until no further tracks of group 2 remain to be analysed (step s4.8). As discussed above in relation to the tracks 1...m of group 1, the
 25 creation of a track vector 40 may be performed at step s13.12 for one, some or all of the tracks 1...n of group 2.

A combined vector 36 for group 2 is then created (step s4.9), for example by summing the tag probabilities 116 for the tracks 1...n of group 2, and, optionally,
 30 normalising the sum.

The group level similarity 37 for the tracks of groups 1 and 2 is calculated by evaluating the similarity between the combined vectors 36, 37 for groups 1 and 2 (step s4.10). For example, if the combined vectors 36, 37 for artists 1 and 2 are
 35 denoted by $\vec{a}$ and $\vec{b}$, their similarity $\text{sim}(\vec{a}, \vec{b})$ can be measured with a cosine similarity defined as shown by Equation (11):

- 33 -

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \bullet \vec{b}}{|\vec{a}| \times |\vec{b}|} \quad (11)$$

In other embodiments, a different technique may be used to evaluate $sim(\vec{a}, \vec{b})$ instead of the cosine similarity shown in Equation (11). For example, one
 5 alternative technique may include using the Euclidean distance and taking its inverse to obtain the similarity $sim(\vec{a}, \vec{b})$. Another example technique for assessing similarity may use the Kullback-Leibler divergence.

10 One or more track level similarities 38 are assessed at step s4.11, based on the similarity of the track vectors 39, 40. The similarity of the track vectors 39, 40 may be assessed using Equation (11) above.

A combined group and track similarity 41 may then be determined, for example, by
 15 summing the group level similarity 37 and the track level similarity 41 (step s4.12). In this particular example, the group level similarity 37, the track level similarity 38 and the combined group and track similarity 41 are normalised so that they have values in a range between 0 and 1.

20 Similarities between group 1 and one or more further groups of music tracks may be computed by repeating steps s4.7 to s4.12 for additional groups and generating respective group level similarities 37, track level similarities 38 and combined group and track similarities 41 for each additional group. For example, where
 25 group 1 contains tracks by a first artist and group 2 contains tracks by a second artist, further groups, groups 3 and 4 may be defined, containing tracks by a third artist, a fourth artist respectively, and so on.

In step s4.13, a list of recommendations 42 of tracks is compiled from the music tracks of group 2 and, where provided, any further groups of music tracks that have
 30 been analysed, based on one or both of the group level similarity 41 and, optionally, the combined group and track similarity 37. In one embodiment, a list of tracks exhibiting the highest combined group and track similarity 41 and/or other similarity 37, 38 may be compiled at step s4.11 and output to the user (step s4.14). Alternatively, the list of recommendations 42 may be ranked and/or revised as part
 35 of the compilation (step s4.15). Examples of compilation procedures that may be

- 34 -

performed at step s4.15 will now be described, with reference to Figures 17, 18 and 19.

Figures 17 to 19 show example procedures for generating the list of
5 recommendations at step s4.15, in which other ranking techniques are employed.

Beginning with the example of Figure 17, which starts at step s17.0, a preliminary list of candidate tracks from group 2 and any further groups of music tracks is compiled (s17.1), based on one or more of the similarities calculated in steps s4.10
10 to s4.12.

In the example method of Figure 17, the list of preliminary candidates is revised based on user preferences input by the user, for example, by using the sliders 53, 54 in the user interface 50 shown in Figure 5. For example, the user may have
15 indicated that they would like to receive recommendations of tracks that are jazzier and include more piano than a particular track indicated in field 52 of the user interface 50, corresponding to track 1 of group 1, but include less stringed instruments, using the sliders 53, 54.

20 Where the user has provided input indicating a preference has been received (step s17.2), the tag probability 116 corresponding to a first property indicated by the user is identified (step s17.3) and the relevant tag probabilities for the candidate tracks are retrieved or otherwise obtained (step s17.4) and adjusted as follows.

25 If the user input indicated a positive contribution for the property (step s17.5), such as “more jazz”, the tag probabilities 116 for a genre of “jazz” for the candidate tracks are added to the values calculated for one or more of the similarities between track 1 of group 1 and the candidate tracks.

30 If the user input indicated a positive contribution for the property (step s17.5), such as “less strings”, the tag probabilities 116 for stringed instruments is subtracted from the similarities for the candidate tracks are subtracted from the similarity values for the candidate tracks (step s17.7).

35 If another preference has been indicated by the user (step s17.8), then steps s17.3 to s17.7 are repeated for the next preference, until the similarity values have been adjusted for all the received user preferences (step s17.8).

- 35 -

The candidate tracks are then ranked based on their adjusted similarities (step s17.9), completing the procedure (step s17.10). The list of recommendations 42 output at step s4.14 may be based on a selected subset, or on all, of the candidate tracks in the ranked list of candidate tracks. For example, a predetermined number of the highest ranked candidate tracks may be selected for inclusion in the list of recommendations 42.

In another example method, shown in Figure 18, beginning at step s18.0, a preliminary list of candidate tracks from group 2 is obtained (s18.1), for example by compiling the list based on one or more of the similarities calculated in steps s4.10 to s4.12 as described above in relation to Figure 17. In this example, the candidate tracks in the preliminary list are ranked based on user preferences as indicated by the user's listening history, as will now be explained.

A user history is obtained in step s18.2. The user history may be based on the number of times the user has previously accessed music tracks stored on the terminal 104, other database or a streaming service, tracks ranked by a user on social media or online music database, and/or on tracks purchased by the user from a digital music store.

Next, the controller 202 obtains a tag indicating a user preference from the tracks in the history (step s18.3). For example, the controller 202 may determine the most common tag for the previously accessed tracks shown in the user history. The corresponding tag probabilities 116 for the candidate tracks from group 2 are then retrieved (step s18.4).

If the obtained tag seems to be viewed positively by the user (step s18.5), for example, if the tag occurs most often in previously accessed tracks that were played, downloaded or purchased by a user, then the tag probabilities 116 for the candidate tracks are added to one or more of the similarities calculated in steps s4.10 to s4.12 (step s18.6).

If the obtained tag seems to indicate a negative user preferences, for example if the tag occurs most often in previously accessed tracks that were skipped by the user (step s18.5), then the tag probabilities 116 for the candidate tracks are subtracted from one or more of the similarities calculated in steps s4.10 to s4.12 (step s18.7).

- 36 -

Steps s18.3 to s18.8 may be repeated for further tags, if required (step s18.8).

5 The candidate tracks are then ranked based on their adjusted similarities (step s18.9), completing the procedure (step s18.10). As noted above, in relation to Figure 17, the list of recommendations 42 output at step s4.14 may be based on a subset, or on all, of the candidate tracks in the ranked list of candidate tracks produced at step s18.9.

10 In yet another example method, shown in Figure 19, starting at step s19.0, a list of preliminary candidates is obtained (step s19.1), as discussed above in relation to Figure 17. In this example method, the preliminary candidates are ranked based on properties other than those on which the tag probabilities 116 are based.

15 In step s19.2, the controller 202 determines whether the user history includes a previously accessed music track that has a tag that was not included in track 1 of group 1 (step s19.2). For example, a previously accessed music track in the user history may include instruments that were not included in track 1 of group 1, or belong a different genre from track 1 of group 1. In the following, such a tag is
20 referred to as a “new tag”.

The tag probabilities 116 for the new tag are retrieved for the candidate tracks (step s19.3).

25 If the user history indicates that the user listened to the previously accessed music track with the new tag (step s19.4), then the tag probabilities 116 for the new tag in the candidate tracks is added to their respective similarities (step s19.5).

30 If the user history indicates that the user skipped the previously accessed music track with the new tag (step s19.4), then the tag probabilities 116 for the new tag in the candidate tracks is subtracted from their respective similarities (step s19.6).

If there are further new tags in the previously accessed track (steps s19.7, s19.8), then tag probabilities 116 for the candidate tracks for the further new tags are also
35 added to, or subtracted from, the similarities as appropriate (steps s19.5, s19.6).

- 37 -

If required, the controller 202 may then search for another previously accessed track with at least one new tag in the user history (steps s19.9, s19.2), to further adjust the similarities of the candidate tracks (steps s19.3 to s19.9).

- 5 The candidate tracks are then ranked based on their adjusted similarities (step s19.10), completing the procedure (step s19.11). The list of recommendations 42 output at step s4.14 may be based on subset, or on all, of the candidate tracks in the ranked list of candidate tracks.
- 10 In yet another embodiment, one or more of the methods described above with reference to Figures 17 to 19 may be used to compile the list of recommendations 42 at step s4.13.

The list of recommendations 42 is output at step s4.14 via the interface 202. In this
15 example, the list is transmitted to the user's terminal 104 via the network 102. The terminal 104 may present the list of recommendations 42 to the user as a list of music tracks, optionally with links to access the recommended tracks from a streaming service or database or to purchase the recommended tracks from a digital music store. Where at least some of the recommendations include music
20 tracks in a library accessible by the terminal 104, for example stored in storage 208, the list of recommendations 42 may include, or take the form of, a playlist to be followed by a media player software application stored in the terminal 104.

The procedure for recommending music tracks may end at this point (step s4.15)
25 Alternatively, the analysis server 100 may receive and monitor user history information (step s4.16) from the terminal 104 after the list of recommendations 42 has been output (step s4.14) and determine whether the list of recommendations 42 should be revised (step s14.17). For example, the controller 202 may determine that revision is needed to adjust the recommendations 42 based on whether the
30 user has listened to, or skipped, tracks in the existing list of recommendations 42.

If revision is required (step s4.17), then the controller 202 revises the list of recommendations 42 (step s4.18). For example, the controller 202 may update the list of recommendations 42 based on tags from the recommended tracks that the
35 user has listened to, or skipped, by performing the method of Figure 18, using the existing list of recommendations 42 as the preliminary list of candidate tracks in step s18.1 and received updated user history as the user history obtained in step

- 38 -

s18.2. Alternatively, or additionally, the controller 202 may update the list of recommendations 42 if a previously accessed music track appearing in the updated user history includes a new tag, using the method of Figure 19, using the existing list of recommendations 42 as the preliminary list of candidate tracks in step s19.1.

5

The revised list of recommendations 42 based on the new rankings is then output (step s4.14).

The monitoring of user history and revision of the list of recommendations 42 (steps s4.14 to s4.18) may continue until further revision is not needed (step s4.15), for example, if the user of the terminal 104 pauses or stops music playback, closes the media player application or switches of terminal 104.

The procedure then ends at step s4.19.

15

It will be appreciated that the above-described embodiments are not limiting on the scope of the invention, which is defined by the appended claims and their alternatives. Various alternative implementations will be envisaged by the skilled person, and all such alternatives are intended to be within the scope of the claims.

20

It is noted that the disclosure of the present application should be understood to include any novel features or any novel combination of features either explicitly or implicitly disclosed herein or any generalization thereof and during the prosecution of the present application or of any application derived therefrom, new claims may be formulated to cover any such features and/or combination of such features.

25

Embodiments of the present invention may be implemented in software, hardware, application logic or a combination of software, hardware and application logic. The software, application logic and/or hardware may reside on memory, or any computer media. In an example embodiment, the application logic, software or an instruction set is maintained on any one of various conventional computer-readable media. In the context of this document, a "computer-readable medium" may be any media or means that can contain, store, communicate, propagate or transport the instructions for use by or in connection with an instruction execution system, apparatus, or device, such as a computer.

35

- 39 -

A computer-readable medium may comprise a computer-readable storage medium that may be any tangible media or means that can contain or store the instructions for use by or in connection with an instruction execution system, apparatus, or device, such as a computer as defined previously. The computer-readable medium
5 may be a volatile medium or a non-volatile medium.

According to various embodiments of the previous aspect of the present invention, the computer program according to any of the above aspects, may be implemented in a computer program product comprising a tangible computer-readable medium
10 bearing computer program code embodied therein which can be used with the processor for the implementation of the functions described above.

Reference to "computer-readable storage medium", "computer program product", "tangibly embodied computer program" etc, or a "processor" or "processing circuit"
15 etc. should be understood to encompass not only computers having differing architectures such as single/multi processor architectures and sequencers/parallel architectures, but also specialised circuits such as field programmable gate arrays FPGA, application specify circuits ASIC, signal processing devices and other devices. References to computer program, instructions, code etc. should be
20 understood to express software for a programmable processor firmware such as the programmable content of a hardware device as instructions for a processor or configured or configuration settings for a fixed function device, gate array, programmable logic device, etc.

25 If desired, the different functions discussed herein may be performed in a different order and/or concurrently with each other. Furthermore, if desired, one or more of the above-described functions may be optional or may be combined.

Although various aspects of the invention are set out in the independent claims,
30 other aspects of the invention comprise other combinations of features from the described embodiments and/or the dependent claims with the features of the independent claims, and not solely the combinations explicitly set out in the claims.

Claims

1. An apparatus, comprising:
a controller; and
5 a memory in which is stored computer readable instructions that, when executed by the controller, cause the controller to:
receive input information regarding one or more music tracks;
determine properties of a first plurality of music tracks belonging to a first group of music tracks defined based, at least in part, on the input information and
10 properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties of the second plurality of music tracks being based on track level attributes;
determine a similarity between the first group of music tracks and the second group of music tracks based at least in part on the determined properties of
15 the first plurality of music tracks and the second plurality of music tracks;
select one or more music tracks from the second group of music tracks based at least in part on said similarity; and
output a list of said selected music tracks.
- 20 2. An apparatus according to claim 1, where said computer readable instructions, when executed by the controller, further cause the controller to:
determine a similarity between one of said first plurality of music tracks and one of the second plurality of music tracks;
wherein said similarity between the first group of music tracks and the
25 second group of music tracks is a combination of a group level similarity based on the properties of the first plurality of music tracks and the properties of the second plurality of music tracks and a track level similarity based on the properties of said one of said first plurality of music tracks and the properties of said one of the second plurality of music tracks.
- 30 3. An apparatus according to claim 1 or 2, wherein said track level attributes include acoustic features extracted from said first plurality of music tracks or said second plurality of music tracks.
- 35 4. An apparatus according to claim 1, 2 or 3, wherein said track level attributes include at least one of:

- 41 -

tags associated with at least some of said first plurality of music tracks and said second plurality of music tracks;

metadata associated with at least some of said first plurality of music tracks and said second plurality of music tracks; and

5 keywords extracted from text associated with at least some of said first plurality of music tracks and said second plurality of music tracks.

5. An apparatus according to any of the preceding claims, wherein said properties include at least one property based on a musical instrument and at least
10 one property based on a musical genre.

6. An apparatus according to claim 5, wherein said properties comprise probabilities that a tag for a musical instrument or genre applies to a respective one of the first and second pluralities of music tracks.

15

7. An apparatus according to any of the preceding claims, where said computer readable instructions, when executed by the controller, further cause the controller to:

monitor a history of music tracks previously accessed by a user;

20 revise said list of selected music tracks based on the properties of the previously accessed music tracks in said history and on whether the music tracks in the history were played or skipped; and

output said revised list.

25 8. An apparatus according to any of the preceding claims, where said computer readable instructions, when executed by the controller, further cause the controller to:

rank the selected music tracks in the list based, at least in part, on user preferences for the properties included in the received input.

30

9. An apparatus according to any of the preceding claims, wherein said computer readable instructions, when executed by the controller, further cause the controller to:

35 rank the selected music tracks in the list based, at least in part, on properties of previously accessed music tracks indicated in a user history.

- 42 -

10. An apparatus according to any of the preceding claims, wherein said computer readable instructions, when executed by the controller, further cause the controller to:

rank the selected music tracks in the list based, at least in part, on a further
5 property of a previously accessed music track indicated in a user history, wherein said first plurality of music tracks does not include said further property.

11. An apparatus according to claim 9 or 10, wherein said computer readable instructions, when executed by the controller, cause the controller to rank the
10 selected music tracks by adjusting similarities for the selected music tracks according to whether said previously accessed music track indicated in the user history was played or skipped by the user.

12. An apparatus according to any of the preceding claims, wherein said
15 computer readable instructions, when executed by the controller, cause the controller to determine said properties by evaluating first level probabilities that a particular tag applies based on the track level attributes and evaluating a second level probability that the particular tag applies based on the first level probability.

20 13. An apparatus according to claim 12, wherein said computer readable instructions, when executed by the controller, cause the controller to evaluate the first level probabilities using a first classifier and a second classifier and to evaluate the second level probabilities using a third classifier, wherein the first and third classifiers are non-probabilistic classifiers and the second classifier is a
25 probabilistic classifier.

14. A method comprising:

receiving input information regarding at least one music track;
determining properties of a first plurality of music tracks belonging to a first group
30 of music tracks defined based at least in part on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties of the second plurality of music tracks being based on track level attributes;

determining a similarity between the first group of music tracks and the
35 second group of music tracks based at least in part on the determined properties of the first plurality of music tracks and the second plurality of music tracks;

- 43 -

selecting one or more music tracks from the second group of music tracks based at least in part on said similarity; and
outputting a list of said selected music tracks.

- 5 15. A method according to claim 14, comprising:
determining a similarity between one of said first plurality of music tracks and one of the second plurality of music tracks;
wherein said similarity between the first group of music tracks and the second group of music tracks is a combination of a group level similarity based on
10 the properties of the first plurality of music tracks and the properties of the second plurality of music tracks and a track level similarity based on the properties of said one of said first plurality of music tracks and the properties of said one of the second plurality of music tracks.
- 15 16. A method according to claim 14 or 15, wherein said track level attributes include acoustic features extracted from said first and second pluralities of music tracks.
17. A method according to claim 14, 15 or 16, comprising:
20 monitoring a history of music tracks accessed by a user;
revising the list of selected music tracks based on the properties of the previously accessed music tracks in said user history and on whether the previously accessed music tracks were played or skipped; and
outputting said revised list.
- 25 18. A method according to any of claims 14 to 17, comprising:
ranking the selected music tracks in the list based on one or more of:
user preferences for the properties included in the received input;
properties of previously accessed music tracks indicated in a user
30 history; and
a property of a previously accessed music track indicated in a user history, wherein said first plurality of music tracks does not include said property.
- 35 19. A method according to any of claims 14 to 17, comprising:
ranking the selected music tracks in the list based on at least one property of at least one previously accessed music track indicated in a user history;

- 44 -

wherein said ranking the selected music tracks comprises adjusting similarities for the selected music tracks according to whether said previously accessed music track indicated in the user history was played or skipped by the user.

5

20. A computer program comprising computer readable instructions which, when executed by a computer, cause said computer to perform a method according to any of claims 14 to 19.

10 21. A non-transitory tangible computer program product in which is stored computer readable instructions that, when executed by a computer, cause the computer to:

receive input information regarding one or more music tracks;
determine properties of a first plurality of music tracks belonging to a first group of
15 music tracks defined based at least in part on the input information and properties of a second plurality of music tracks belonging to a second group of music tracks, the properties of the first plurality of music tracks and the properties of the second plurality of music tracks being based on track level attributes;

determine a similarity between the first group of music tracks and the
20 second group of music tracks based at least in part on the determined properties of the first plurality of music tracks and the second plurality of music tracks;

select one or more music tracks from the second group of music tracks based at least in part on said similarity; and

output a list of said selected music tracks.

25

22. An apparatus configured to:

receive input a first group of music tracks;

extract track level attributes associated with a first plurality of music tracks
belonging to the first group of music tracks and a second plurality of music tracks
30 belonging to a second group of music tracks;

determine properties of the first plurality of music tracks and the second plurality of music tracks, based on track level attributes;

determine a similarity between the first group of music tracks and the
second group of music tracks based at least in part on the determined properties of
35 the first plurality of music tracks and the second plurality of music tracks; select one or more music tracks from the second group of music tracks based at least in part on said similarity; and

- 45 -

output a list of said selected music tracks.

23. An apparatus according to claim 22, configured to:

determine a similarity between one of said first plurality of music tracks and
5 one of the second plurality of music tracks;

wherein said similarity between the first group of music tracks and the
second group of music tracks is a combination of a group level similarity based on
the properties of the first plurality of music tracks and the properties of the second
plurality of music tracks and a track level similarity based on the properties of said
10 one of said first plurality of music tracks and the properties of said one of the
second plurality of music tracks.

24. An apparatus according to claim 22 or 23, wherein said track level attributes
include acoustic features extracted from said first plurality of music tracks or said
15 second plurality of music tracks.

25. An apparatus according to claim 22, 23 or 24, wherein said track level
attributes include at least one of:

tags associated with at least some of said first plurality of music tracks and
20 said second plurality of music tracks;

metadata associated with at least some of said first plurality of music tracks
and said second plurality of music tracks; and

keywords extracted from text associated with at least some of said first
plurality of music tracks and said second plurality of music tracks.

26. An apparatus according to any of claims 22 to 25, wherein said properties
include at least one property based on a musical instrument and at least one
property based on a musical genre.

27. An apparatus according to claim 26, wherein said properties comprise
probabilities that a tag for a musical instrument or genre applies to a respective one
of the first and second pluralities of music tracks.

28. An apparatus according to any of claims 22 to 27, configured to:

35 monitor a history of music tracks previously accessed by a user;

- 46 -

revise said list of selected music tracks based on the properties of the previously accessed music tracks in said history and on whether the music tracks in the history were played or skipped; and
output said revised list.

5

29. An apparatus according to any of claims 22 to 28, configured to:
rank the selected music tracks in the list based, at least in part, on user preferences for the properties included in the received input.

10

30. An apparatus according to any of claims 22 to 29, configured to:
rank the selected music tracks in the list based, at least in part, on properties of previously accessed music tracks indicated in a user history.

15

31. An apparatus according to any of claims 22 to 30, configured to:
rank the selected music tracks in the list based, at least in part, on a further property of a previously accessed music track indicated in a user history, wherein said first plurality of music tracks does not include said further property.

20

32. An apparatus according to claim 30 or 31, configured to rank the selected music tracks by adjusting similarities for the selected music tracks according to whether said previously accessed music track indicated in the user history was played or skipped by the user.

25

33. An apparatus according to any of claims 22 to 32, configured to determine said properties by evaluating first level probabilities that a particular tag applies based on the track level attributes and evaluating a second level probability that the particular tag applies based on the first level probability.

30

34. An apparatus according to claim 33, configured to evaluate the first level probabilities using a first classifier and a second classifier and to evaluate the second level probabilities using a third classifier, wherein the first and third classifiers are non-probabilistic classifiers and the second classifier is a probabilistic classifier.

35

1/15

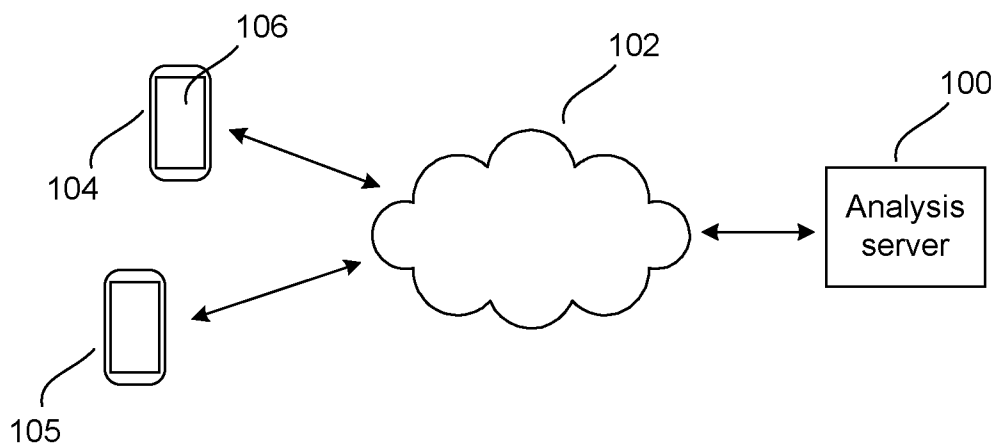


Figure 1

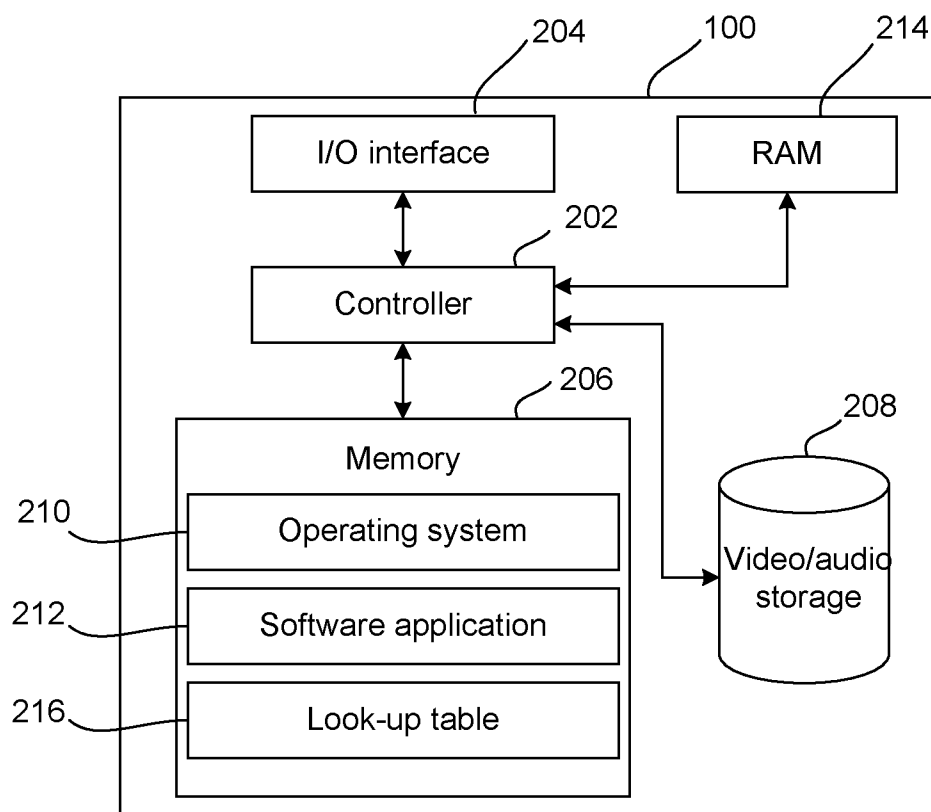


Figure 2

2/15

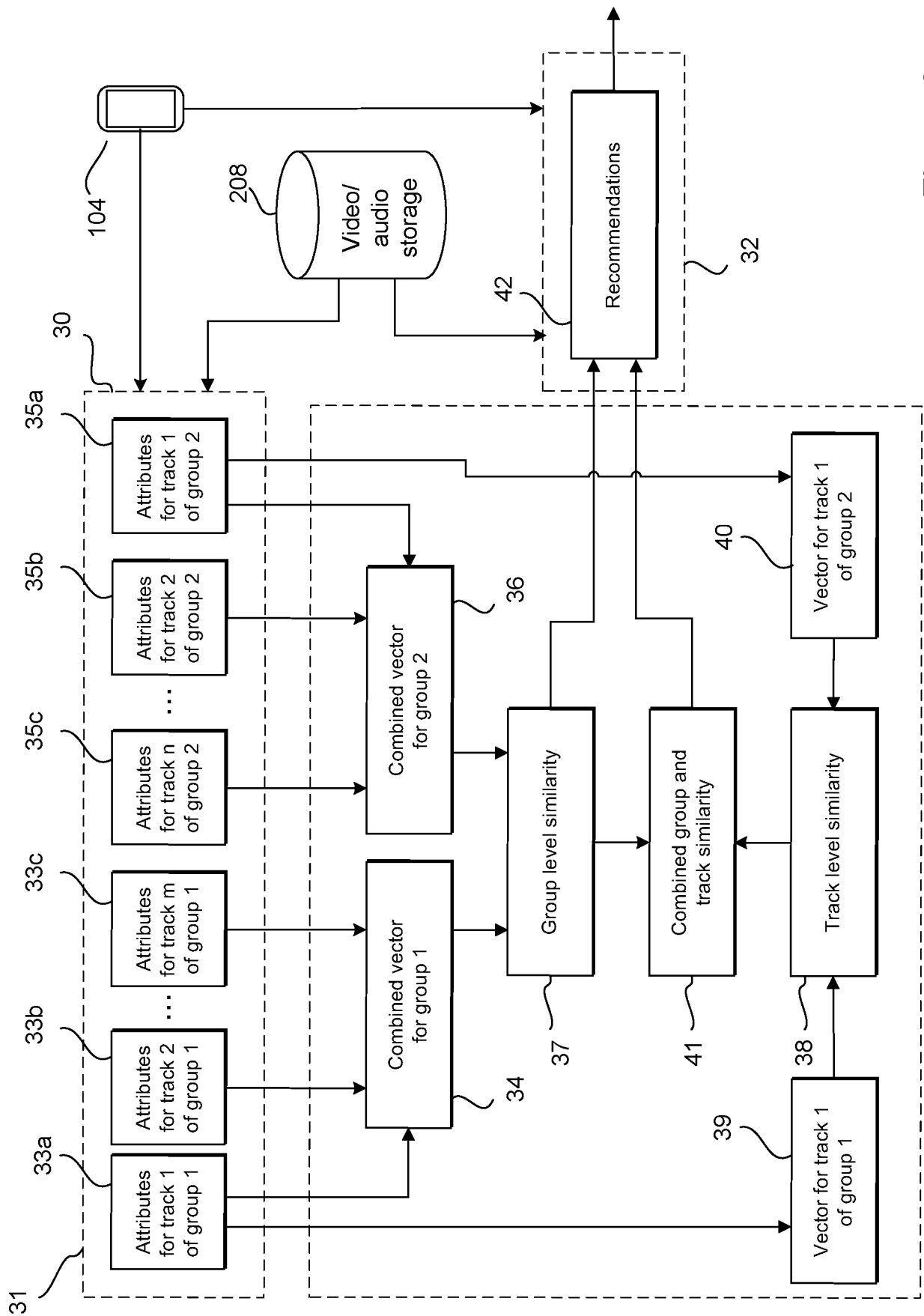


Figure 3

3/15

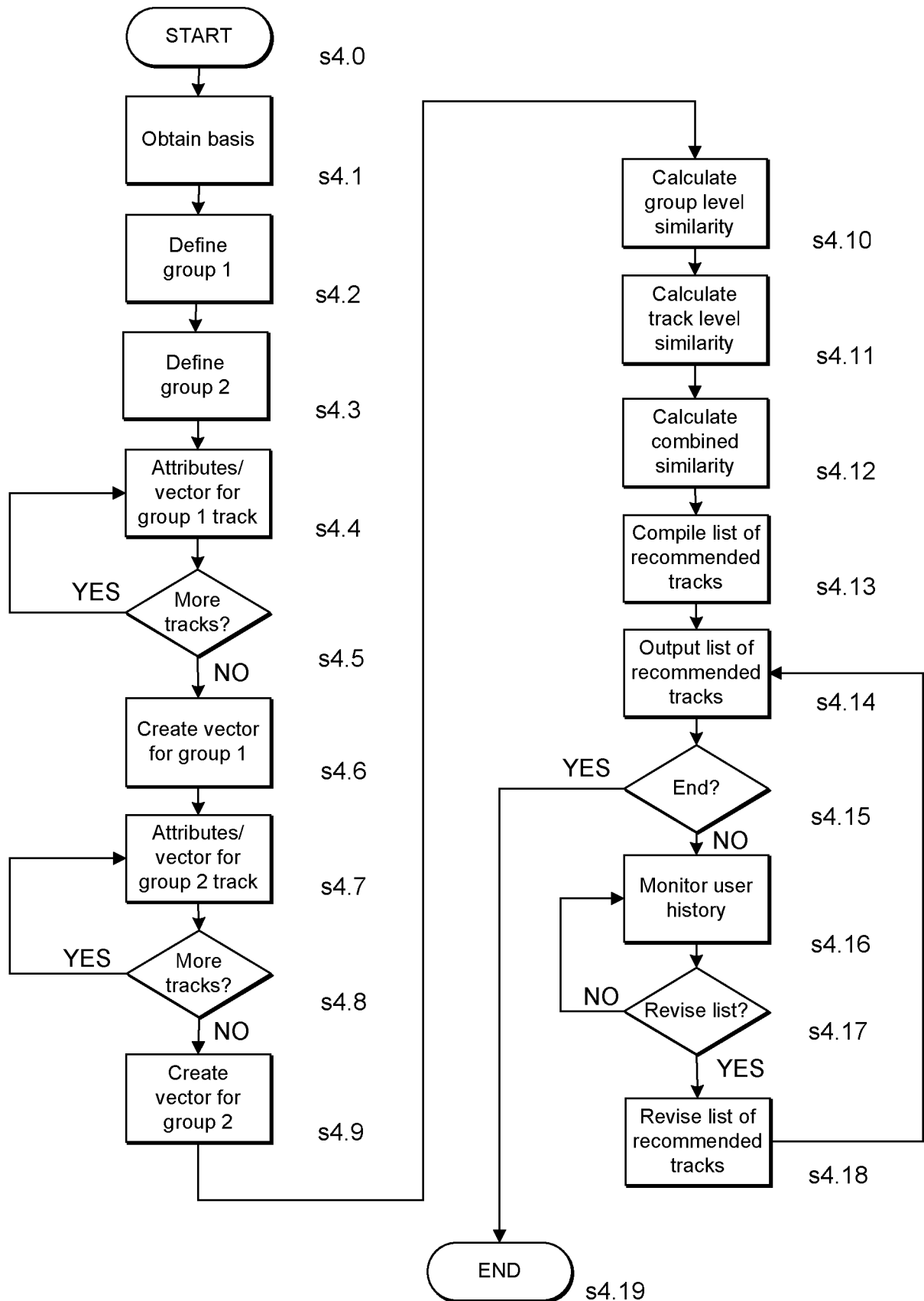


Figure 4

4/15

50

Search based on artist and/or song name

51

52

Artist name

Song name

SEARCH

Instruments

Less More

Accordion

Acoustic guitar

Banjo

Bass

Brass

Drums

Keyboards

Percussion

Piano

Saxophone

Strings

Synthesizer

Woodwind

53

Genres

Less More

Ambient and new age

Blues

Classical

Country and western

Dance

Easy listening

Electronica

Folk and roots

Indie and alternative

Jazz

Latin

Metal

Pop

Rap and hip hop

Reggae

Rock

Soul, R&B and funk

World music

54

Instruments	Less	More
Accordion		
Acoustic guitar		
Banjo		
Bass		
Brass		
Drums		
Keyboards		
Percussion		
Piano		
Saxophone		
Strings		
Synthesizer		
Woodwind		

Genres	Less	More
Ambient and new age		
Blues		
Classical		
Country and western		
Dance		
Easy listening		
Electronica		
Folk and roots		
Indie and alternative		
Jazz		
Latin		
Metal		
Pop		
Rap and hip hop		
Reggae		
Rock		
Soul, R&B and funk		
World music		

Figure 5

5/15

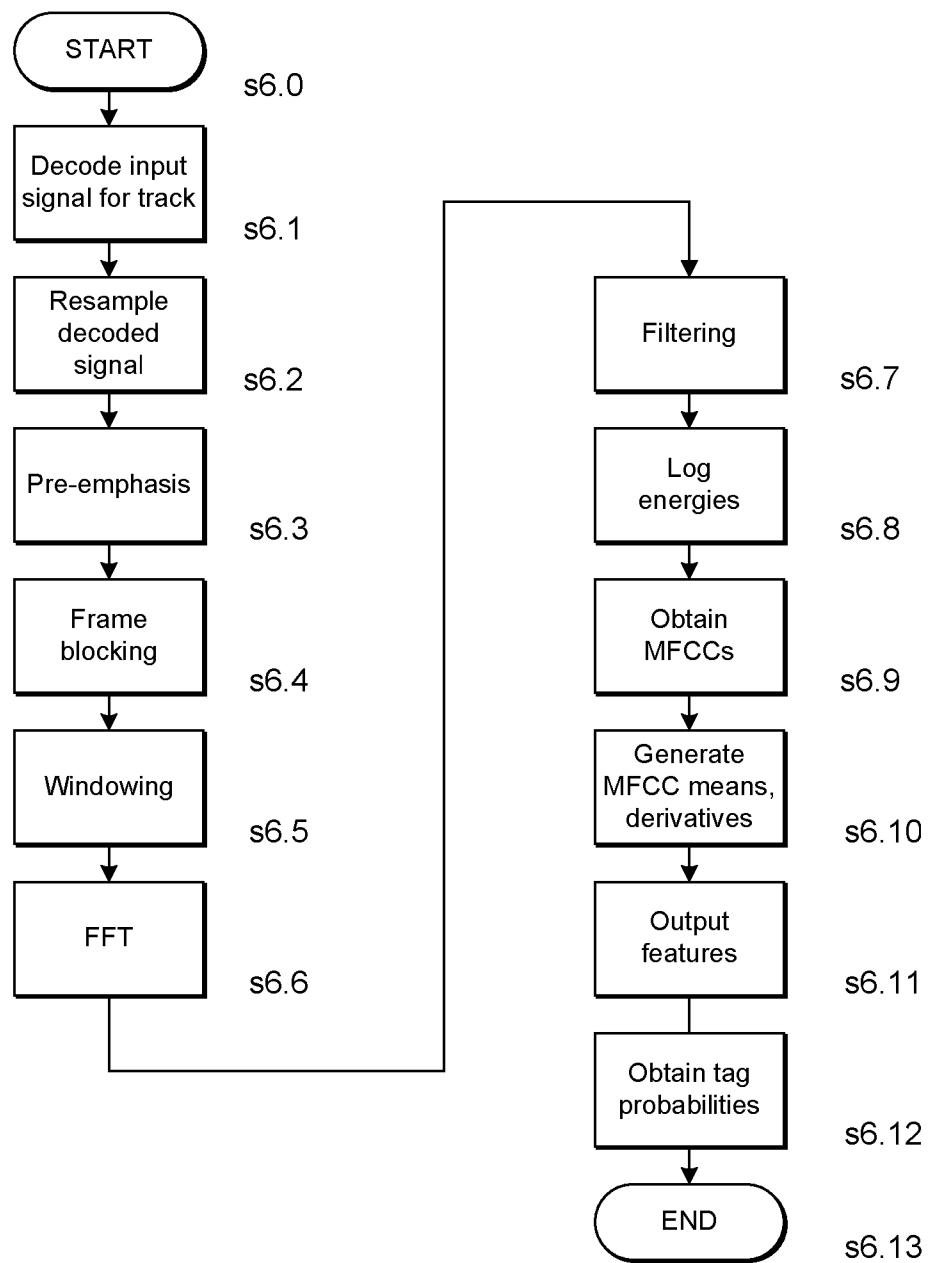


Figure 6

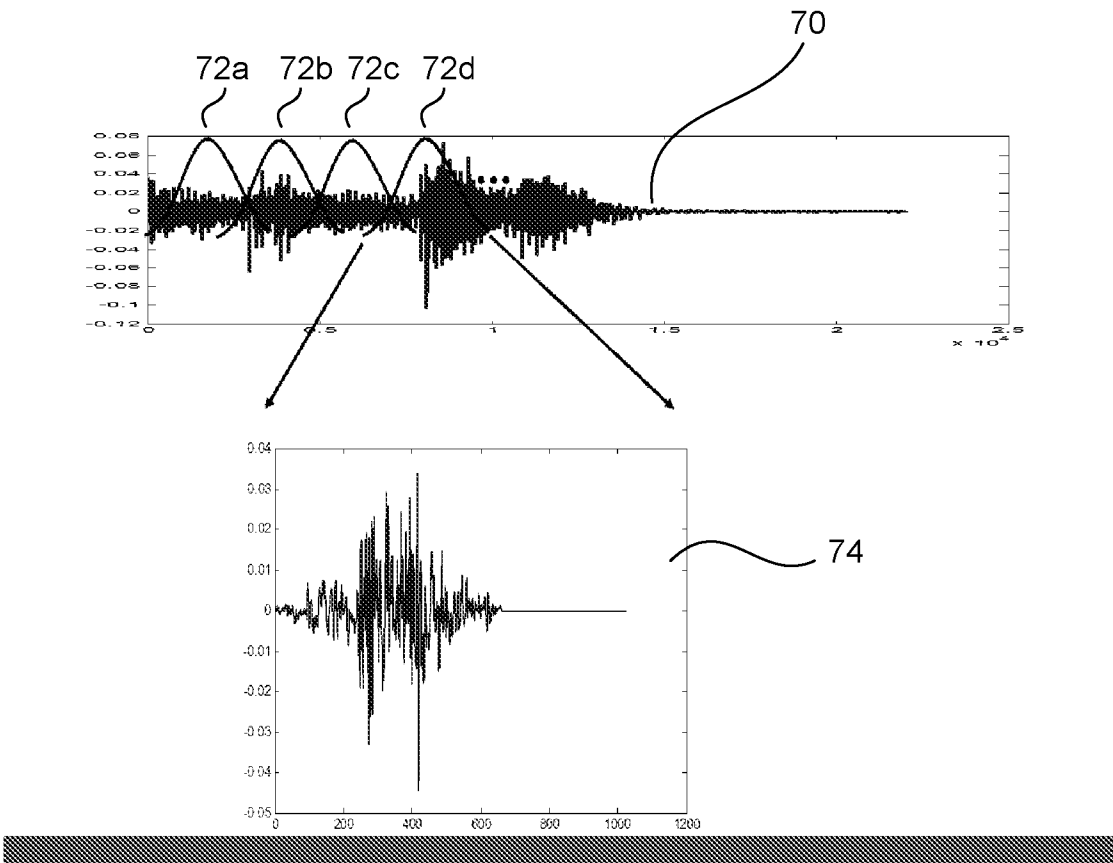


Figure 7

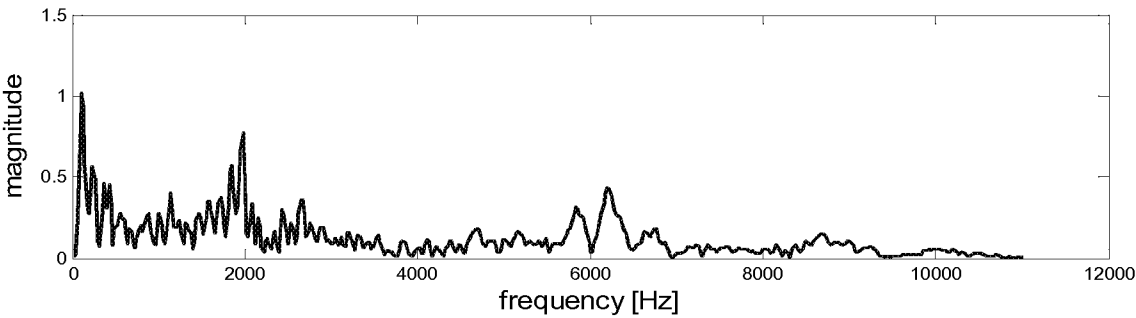


Figure 8

7/15

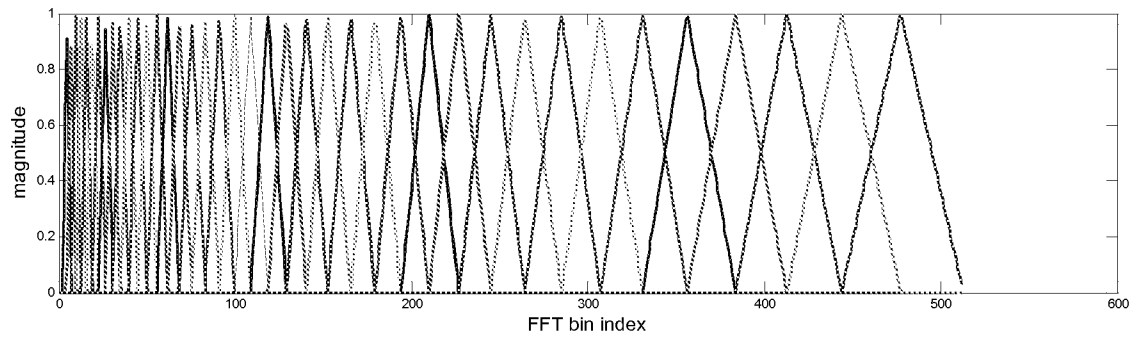


Figure 9

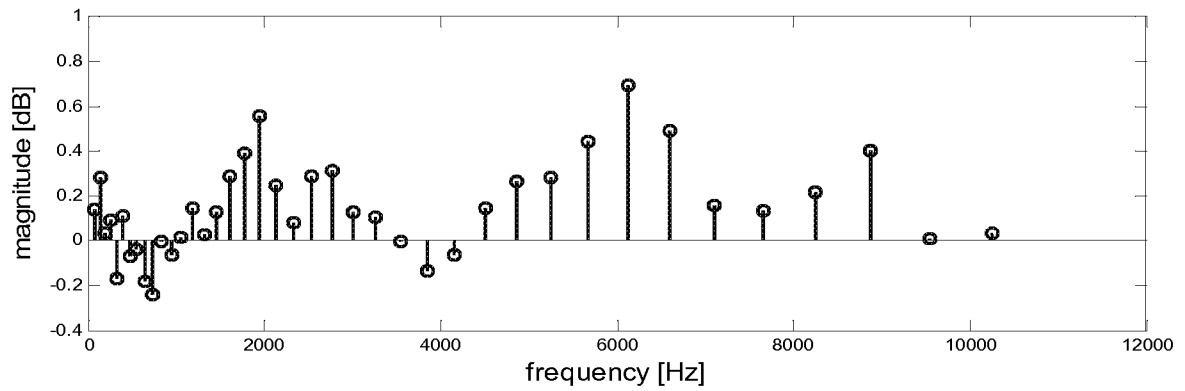


Figure 10

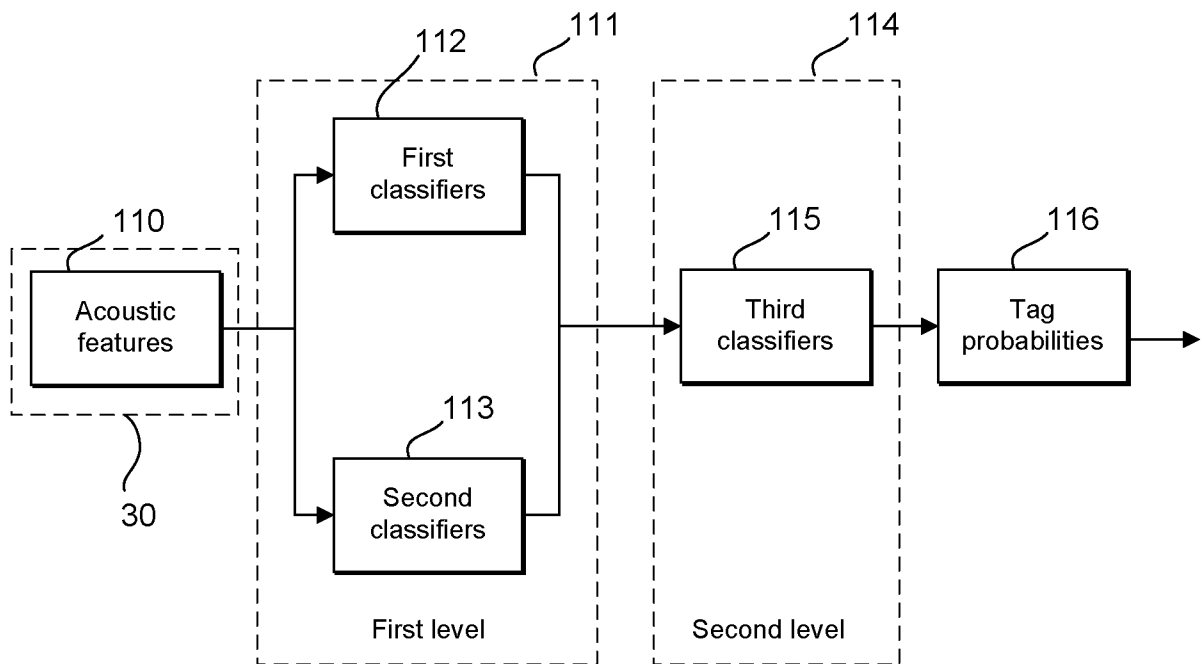


Figure 12

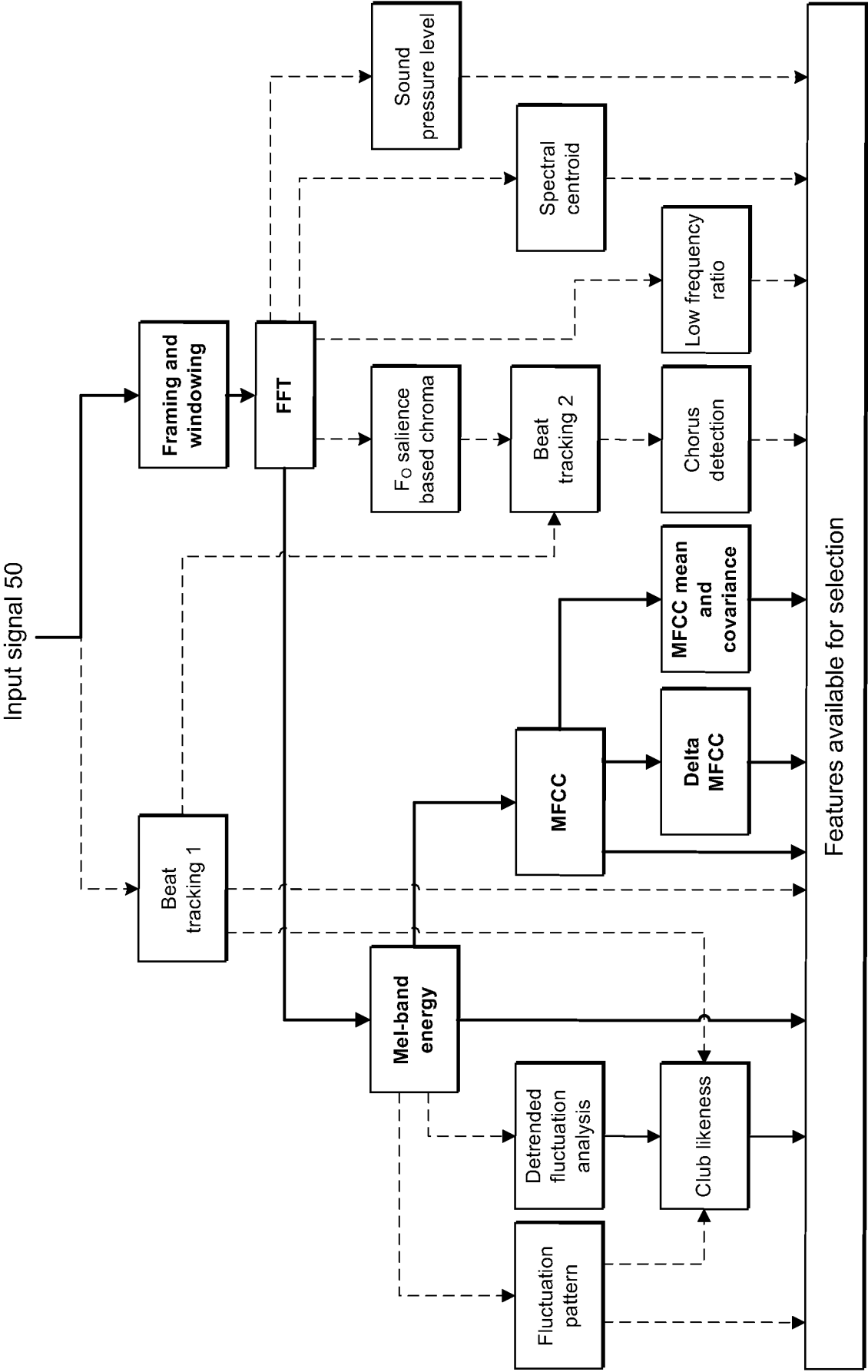


Figure 11

9/15

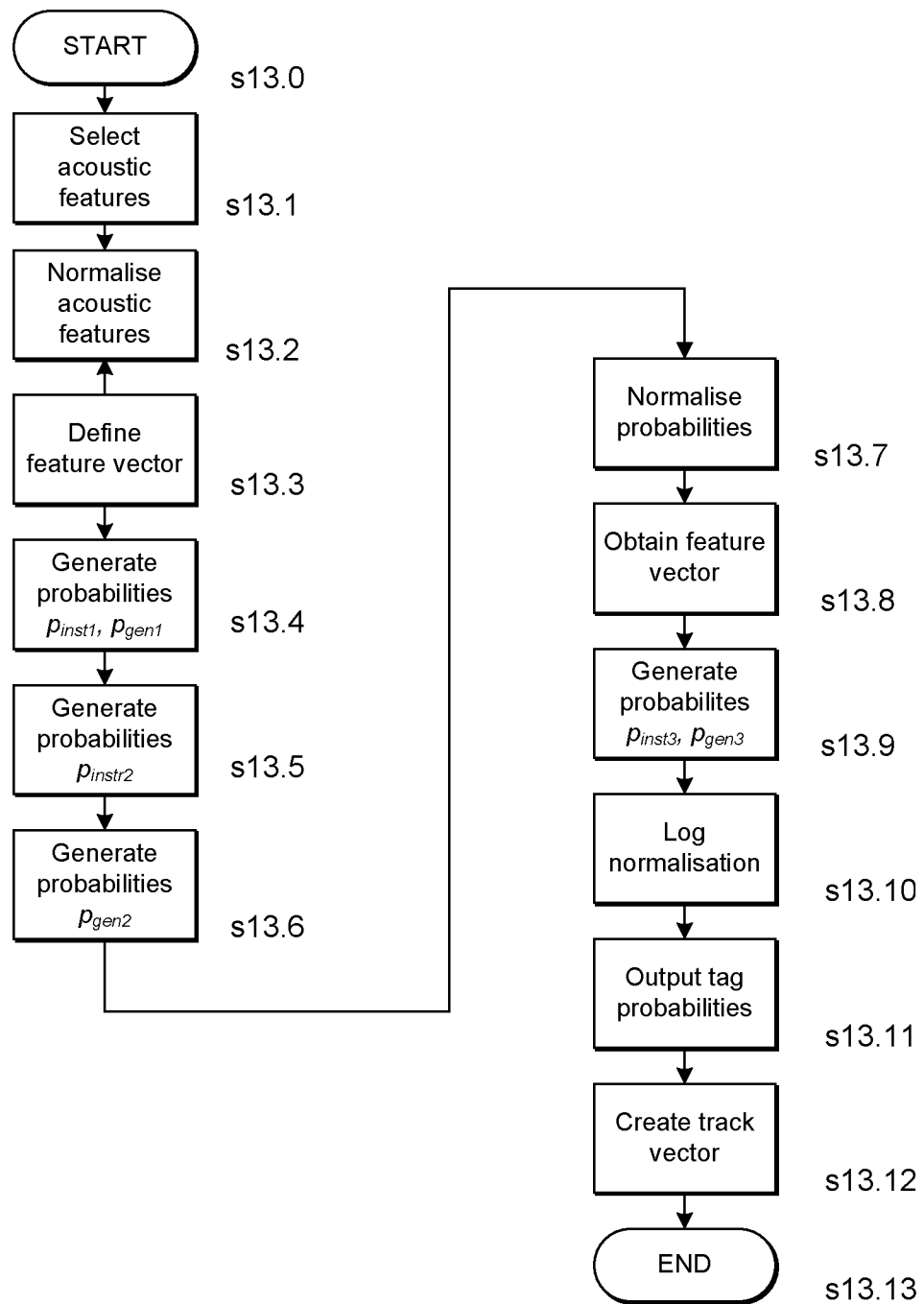


Figure 13

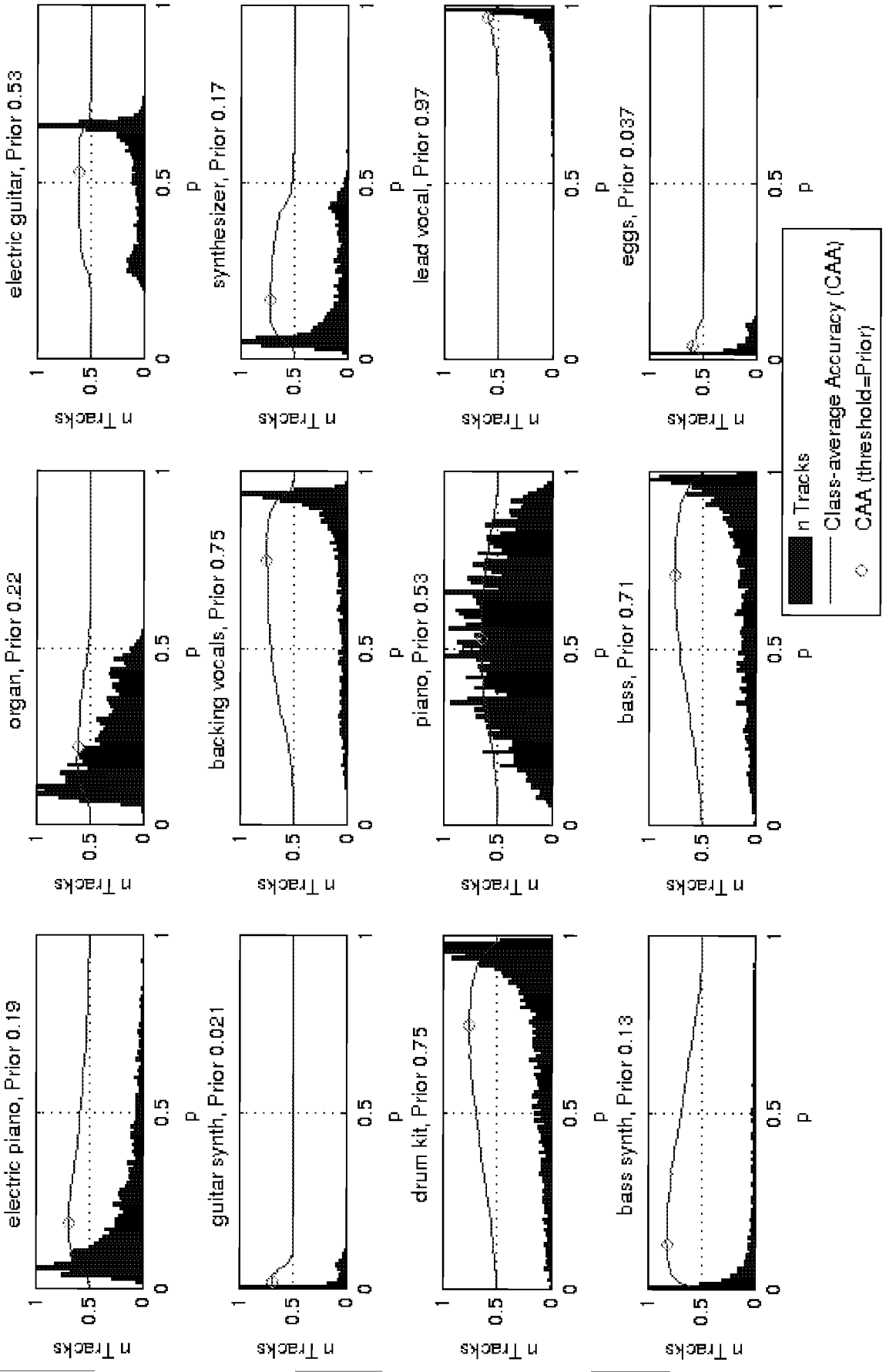


Figure 14

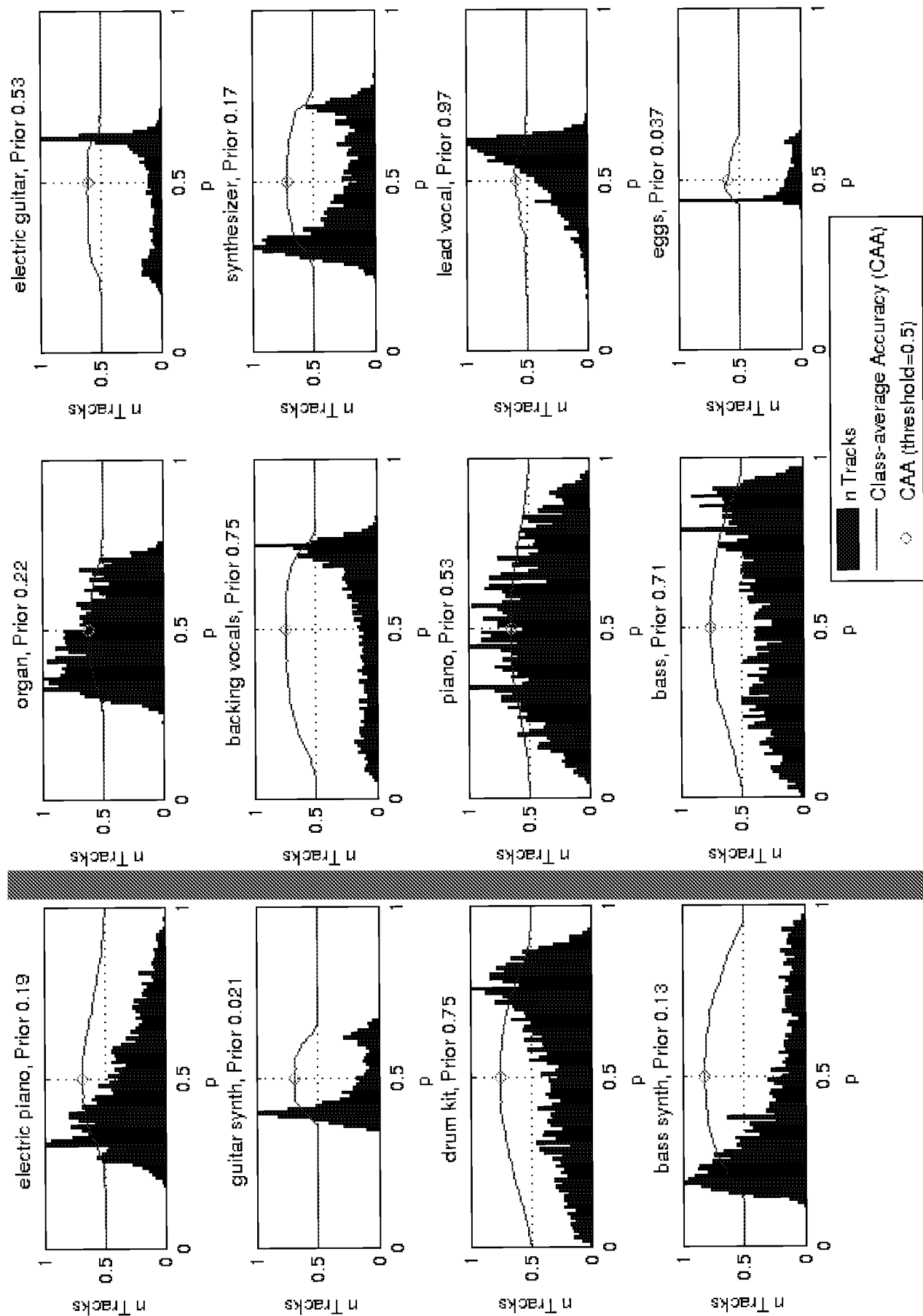


Figure 15

12/15

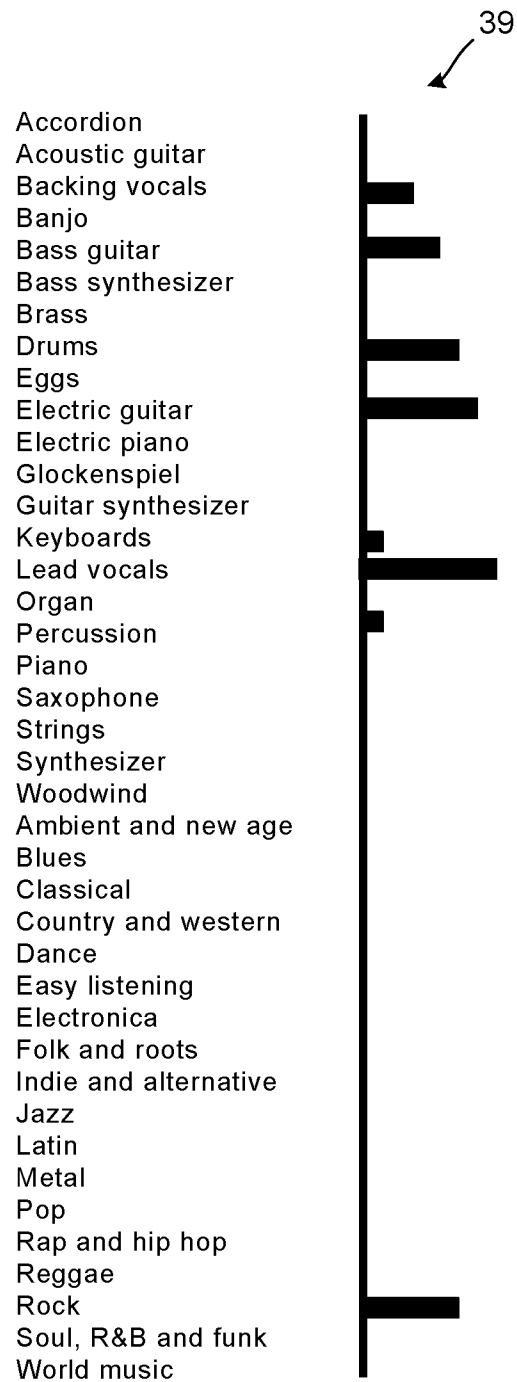


Figure 16

13/15

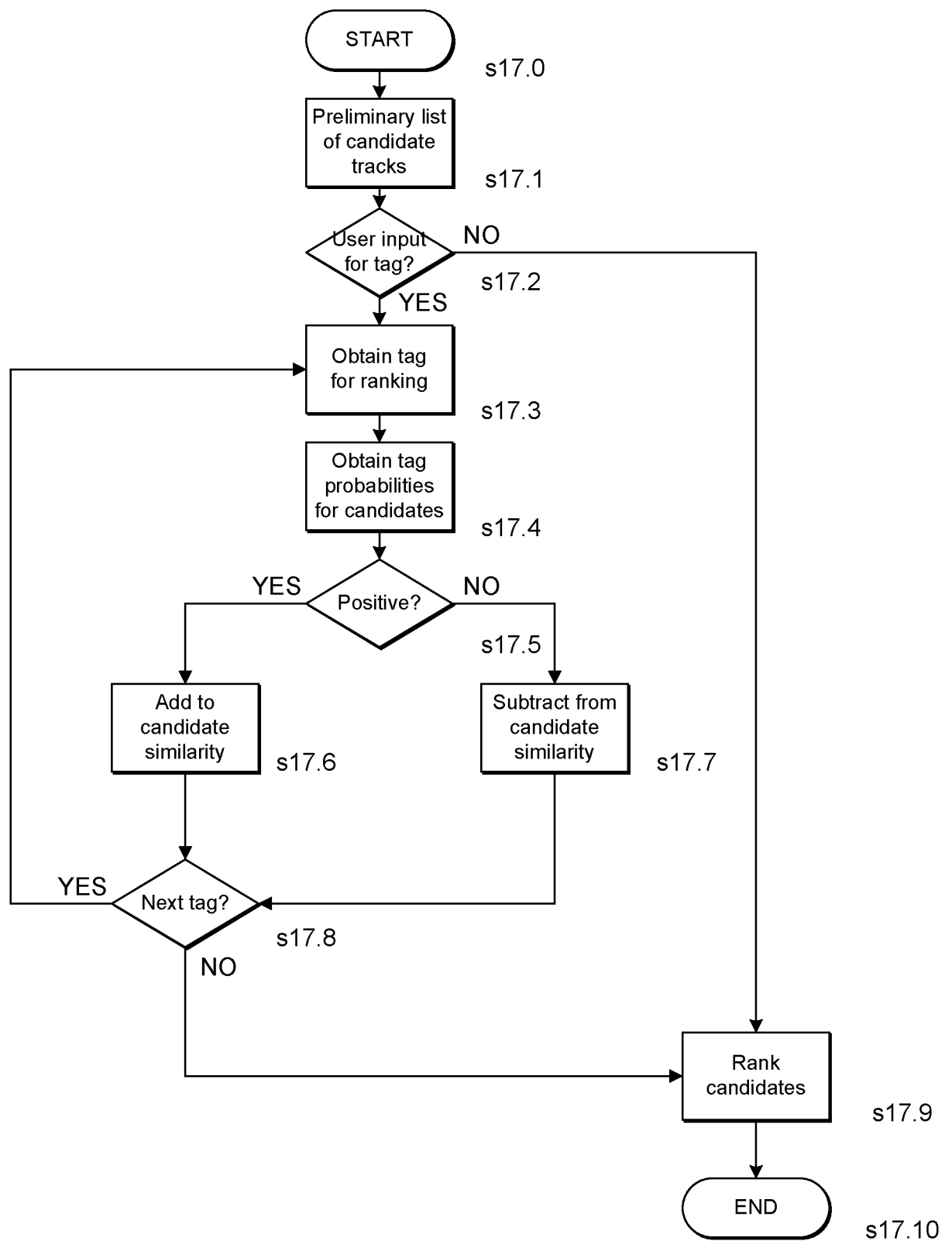


Figure 17

14/15

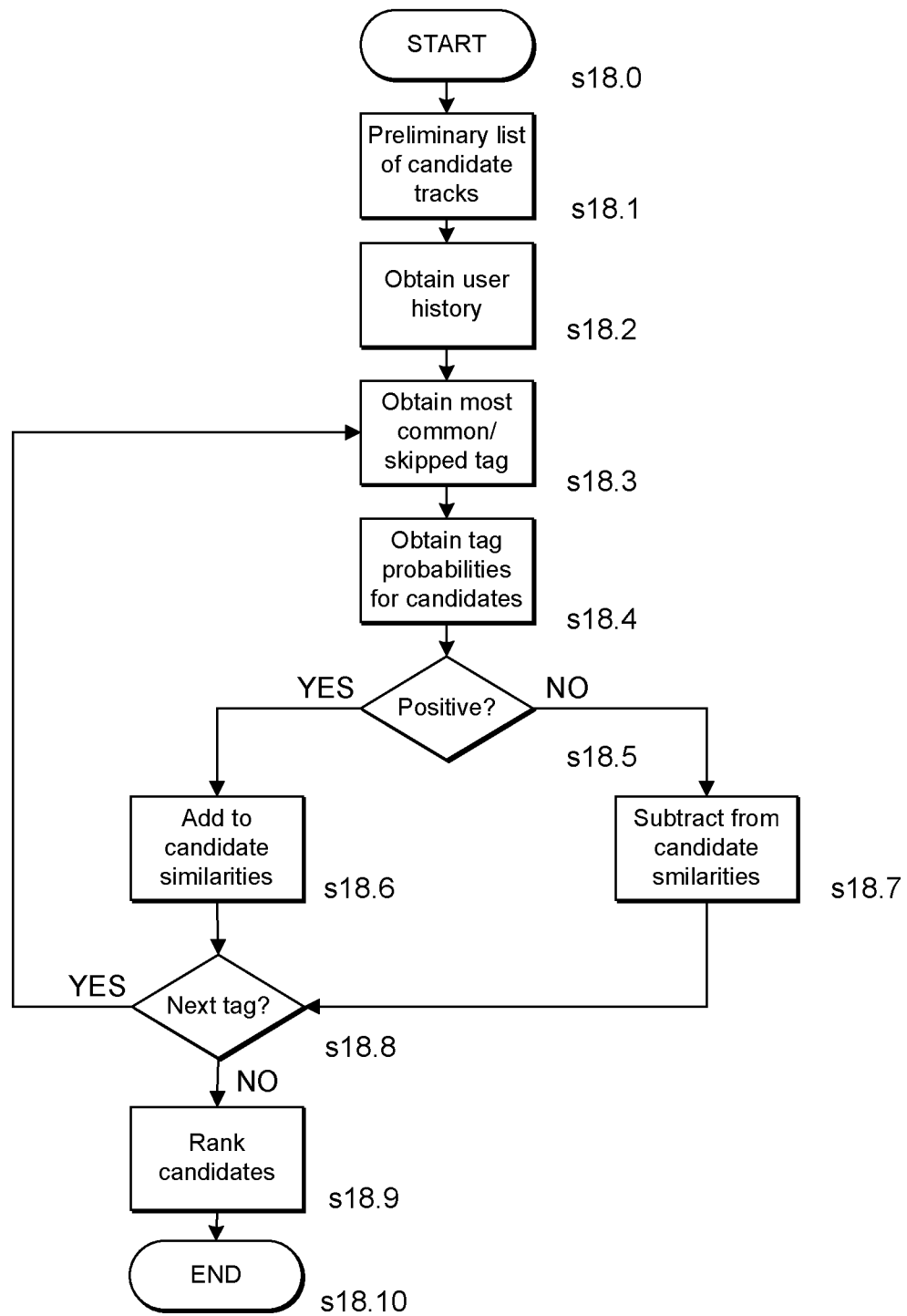


Figure 18

15/15

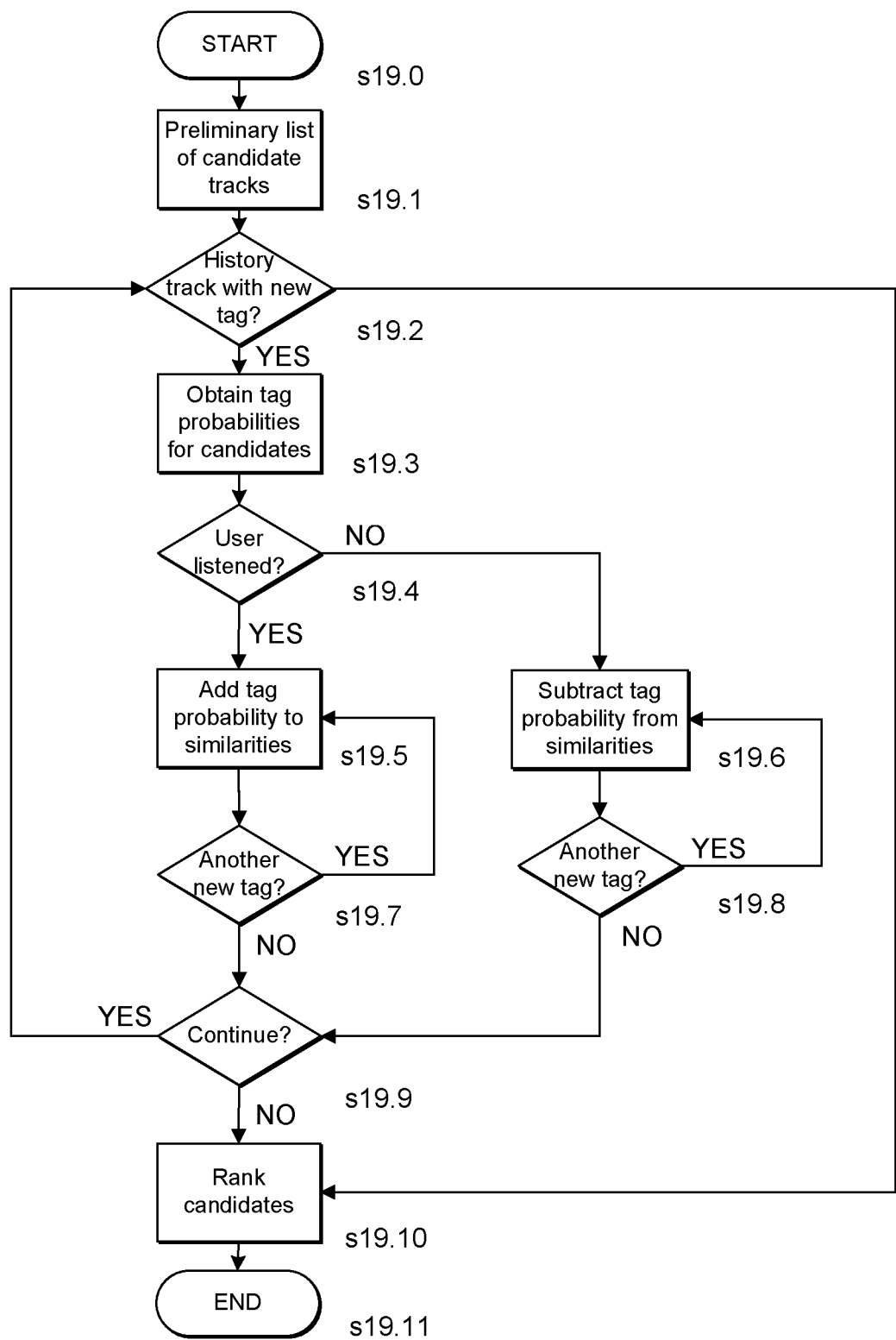


Figure 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2014/051037

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F, G06K, G10H, G10L

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

FI, SE, NO, DK

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

EPO-Internal, WPI, Google Scholar, IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2006107823 A1 (PLATT, J. et al.) 25 May 2006 (25.05.2006) abstract; sections 2.0 and 3.5; Figs. 2, 7	1-10, 14-18, 20-31
A	PAMPALK, E. et al.: "MusicRainbow: a new user interface to discover artists using audio-based similarity and web-based labeling", Int. Conf. on Music Information Retrieval (ISMIR2006), Victoria, Canada, 8-12 Oct. 2006, 4pp the whole document	1-34
A	KNEES, P. et al.: "A music search engine built upon audio-based and web-based similarity measures", Int. ACM SIGIR Conf. [on information retrieval], 23-27 July 2007, 8pp the whole document	1-34

☒ Further documents are listed in the continuation of Box C.
 ☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

21 May 2015 (21.05.2015)

Date of mailing of the international search report

27 May 2015 (27.05.2015)

Name and mailing address of the ISA/FI
Finnish Patent and Registration Office
P.O. Box 1160, FI-00101 HELSINKI, Finland
Facsimile No. +358 9 6939 5328

Authorized officer
Timo Laakso
Telephone No. +358 9 6939 500

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2014/051037

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CASEY, M. et al.: "Content-based information retrieval: current directions and future challenges", Proceedings of the IEEE, Vol. 96 No. 4, April 2008, pp.668-696 the whole document	1-34

International application No.
PCT/FI2014/051037

Form PCT/ISA/210 (patent family annex) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2014/051037

CLASSIFICATION OF SUBJECT MATTER

IPC

G06F 17/30 (2006.01)

G06K 9/62 (2006.01)