



(51) International Patent Classification:

G06F 17/30 (2006.01) G10L 21/0272 (2013.01)
G10L 19/008 (2013.01) G10L 25/51 (2013.01)
G10L 19/20 (2013.01) H04R 3/00 (2006.01)
G10L 21/0208 (2013.01)

(21) International Application Number:

PCT/IB2013/054243

(22) International Filing Date:

22 May 2013 (22.05.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: NOKIA CORPORATION [FI/FI]; Karakaari
7, FI-02610 Espoo (FI).(72) Inventor: OJANPERA, Juha, Petteri; Luuppikatu 3, FI-
327120 Nokia (FI).(74) Agent: SMITH, John, Gary; Page White & Farrer, Bed-
ford House, John Street, London Greater London WC1N
2BF (GB).(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, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU,
RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA,
ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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,
ML, MR, NE, SN, TD, TG).

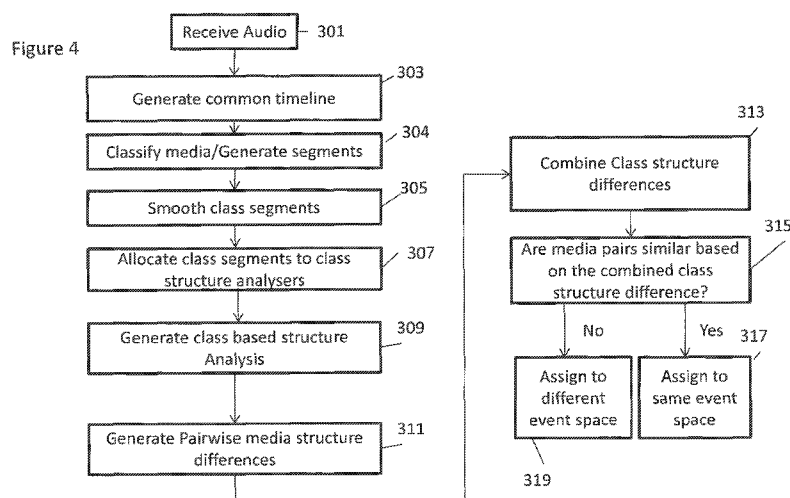
Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

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

(54) Title: A SHARED AUDIO SCENE APPARATUS



(57) Abstract: An apparatus comprising: an input configured to select at least two audio signals; a classifier configured to segment the at least two audio signals based on at least two defined class definitions; a class segment analyser configured to determine a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and a difference analyser configured to determine the at least two audio signal common class segments are within a common event space based on the difference measure.

A SHARED AUDIO SCENE APPARATUS

Field

5 The present application relates to apparatus for the processing of audio and additionally audio-video signals to enable sharing of audio scene captured audio signals. The invention further relates to, but is not limited to, apparatus for processing audio and additionally audio-video signals to enable sharing of audio scene captured audio signals from mobile devices.

10

Background

Viewing recorded or streamed audio-video or audio content is well known. Commercial broadcasters covering an event often have more than one recording device (video-camera/microphone) and a programme director will select a 'mix' where an output from a recording device or combination of recording devices is selected for transmission.

Multiple 'feeds' may be found in sharing services for video and audio signals (such as those employed by YouTube). Such systems, which are known and are widely used to share user generated content recorded and uploaded or up-streamed to a server and then downloaded or down-streamed to a viewing/listening user. Such systems rely on users recording and uploading or up-streaming a recording of an event using the recording facilities at hand to the user. This may typically be in the form of the camera and microphone arrangement of a mobile device such as a mobile phone.

Often the event is attended and recorded from more than one position by different recording users at the same time. The viewing/listening end user may then select one of the up-streamed or uploaded data to view or listen.

Summary

Aspects of this application thus provide a shared audio capture for audio signals from the same audio scene whereby multiple devices or apparatus can record
5 and combine the audio signals to permit a better audio listening experience.

There is provided according to a first aspect an apparatus comprising at least one processor and at least one memory including computer code for one or more programs, the at least one memory and the computer code configured to with the
10 at least one processor cause the apparatus to at least: select at least two audio signals; segment the at least two audio signals based on at least two defined class definitions; determine a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and determine the at least two audio signal common class segments
15 are within a common event space based on the difference measure.

The apparatus may be further caused to generate a common time line incorporating the at least two audio signals.

20 The apparatus may be further caused to align the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least two audio signals for a pair of common class segments from the at least two audio signals.

25

Segmenting the at least two audio signals based on the at least two classes may cause the apparatus to: analyse the at least two audio signals to determine at least one parameter; and segment the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range
30 of values associated with the at least one parameter.

Analysing the at least two audio signals to determine at least one parameter may cause the apparatus to: divide at least one of the audio signals into a number of

frames; analyse for at least one frame of the number of frames of the at least one audio signal to determine the at least one parameter value; and determine a class for the at least one frame based on at least one defined range of parameter values.

5

The at least two classes may comprise at least two of: music; speech; and noise.

Determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser may cause the apparatus to: allocate a pair of common class segments which overlap to an associated class based analyser; and determine a distance value using the associated class based analyser for the pair of common class segments.

10

Determining a distance value using the associated class based analyser for the pair of common class segments may further cause the apparatus to determine a binary distance value.

15

The apparatus may be further caused to determine whether the at least two audio signals are within a common event space based on the determination of the at least two audio signal common class segments are within a common event space.

20

According to a second aspect there is provided an apparatus comprising: means for selecting at least two audio signals; means for segmenting the at least two audio signals based on at least two defined class definitions; means for determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and means for determining the at least two audio signal common class segments are within a common event space based on the difference measure.

25

30

The apparatus may further comprise means for generating a common time line incorporating the at least two audio signals.

The apparatus may further comprise means for aligning the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least two audio signals for a pair of
5 common class segments from the at least two audio signals.

The means for segmenting the at least two audio signals based on the at least two classes may comprise: means for analysing the at least two audio signals to determine at least one parameter; and means for segmenting the at least two
10 audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

The means for analysing the at least two audio signals to determine at least one
15 parameter may comprise: means for dividing at least one of the audio signals into a number of frames; means for analysing for at least one frame of the number of frames of the at least one audio signal to determine the at least one parameter value; and means for determining a class for the at least one frame based on at least one defined range of parameter values.

20 The at least two classes may comprise at least two of: music; speech; and noise.

The means for determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based
25 analyser may comprise: means for allocating a pair of common class segments which overlap to an associated class based analyser; and means for determining a distance value using the associated class based analyser for the pair of common class segments.

30 The means for determining a distance value using the associated class based analyser for the pair of common class segments may comprise means for determining a binary distance value.

The apparatus may further comprise means for determining whether the at least two audio signals are within a common event space based on the determination of the at least two audio signal common class segments are within a common event space.

5

According to a third aspect there is provided a method comprising: selecting at least two audio signals; segmenting the at least two audio signals based on at least two defined class definitions; determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and determining the at least two audio signal common class segments are within a common event space based on the difference measure.

The method may further comprise generating a common time line incorporating the at least two audio signals.

The method may further comprise aligning the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least two audio signals for a pair of common class segments from the at least two audio signals.

Segmenting the at least two audio signals based on the at least two classes may comprise: analysing the at least two audio signals to determine at least one parameter; and segmenting the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

Analysing the at least two audio signals to determine at least one parameter may comprise: dividing at least one of the audio signals into a number of frames; analysing for at least one frame of the number of frames of the at least one audio signal to determine the at least one parameter value; and determining a class for the at least one frame based on at least one defined range of parameter values.

The at least two classes may comprise at least two of: music; speech; and noise.

- 5 Determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser may comprise: allocating a pair of common class segments which overlap to an associated class based analyser; and determining a distance value using the associated class based analyser for the pair of common class segments.
- 10 Determining a distance value using the associated class based analyser for the pair of common class segments may comprise determining a binary distance value.

- 15 The method may further comprise determining whether the at least two audio signals are within a common event space based on the determination of the at least two audio signal common class segments are within a common event space.

- 20 According to a fourth aspect there is provided an apparatus comprising: an input configured to select at least two audio signals; a classifier configured to segment the at least two audio signals based on at least two defined class definitions; a class segment analyser configured to determine a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and a difference analyser configured to determine the at
- 25 least two audio signal common class segments are within a common event space based on the difference measure.

- The apparatus may further comprise a segment smoother configured to generate a common time line incorporating the at least two audio signals.

30

The segment smoother may be configured to align the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common

time line incorporating the at least two audio signals for a pair of common class segments from the at least two audio signals.

5 The classifier may be configured to: analyse the at least two audio signals to determine at least one parameter; and segment the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

10 The classifier may comprise: a framer configured to divide at least one of the audio signals into a number of frames; an analyser configured to analyse for at least one frame of the number of frames of the at least one audio signal to determine the at least one parameter value; and a frame classifier configured to determine a class for the at least one frame based on at least one defined range of parameter values.

15 The at least two classes may comprise at least two of: music; speech; and noise.

20 The class segment analyser may be configured to: allocate a pair of common class segments which overlap to an associated class based analyser; and determine a distance value using the associated class based analyser for the pair of common class segments.

25 The class segment analyser may be configured to determine a binary distance value.

30 The apparatus may further comprise an event space assigner configured to determine whether the at least two audio signals are within a common event space based on the determination of the at least two audio signal common class segments are within a common event space.

A computer program product stored on a medium may cause an apparatus to perform the method as described herein.

An electronic device may comprise apparatus as described herein.

A chipset may comprise apparatus as described herein.

- 5 Embodiments of the present application aim to address problems associated with the state of the art.

Summary of the Figures

- 10 For better understanding of the present application, reference will now be made by way of example to the accompanying drawings in which:

Figure 1 shows schematically a multi-user free-viewpoint service sharing system which may encompass embodiments of the application;

- 15 Figure 2 shows schematically an apparatus suitable for being employed in embodiments of the application;

Figure 3 shows schematically an example content co-ordinating apparatus according to some embodiments;

- 20 Figure 4 shows a flow diagram of the operation of the example content co-ordinating apparatus shown in Figure 3 according to some embodiments;

Figure 5 shows an example audio segment; and

Figures 6 to 9 show audio alignment examples according to some embodiments.

Embodiments of the Application

The following describes in further detail suitable apparatus and possible mechanism for the provision of effective audio signal capture sharing. In the following examples, audio signals and audio capture signals are described.

- 30 However it would be appreciated that in some embodiments the audio signal/audio capture is a part of an audio-video system.

The concept of this application is related to assisting in the production of immersive person-to-person communication and can include video. It would be understood that the space within which the devices record the audio signal can be arbitrarily positioned within an event space. The captured signals as described
5 herein are transmitted or alternatively stored for later consumption where the end user can select the listening point based on their preference from the reconstructed audio space. The rendering part then can provide one or more down mixed signals from which the multiple recordings that correspond to the selective listening point. It would be understood that each recording device can
10 record the event seen and upload or upstream the recorded content. The uploaded or upstream process can include implicitly positioning information about where the content is being recorded.

Furthermore an audio scene can be defined as a region or area within which a
15 device or recording apparatus effectively captures the same audio signal. Recording apparatus operating within an audio scene and forwarding the captured or recorded audio signals or content to a co-ordinating or management apparatus effectively transmit many copies of the same or very similar audio signal. The redundancy of many devices capturing the same audio signal permits
20 the effective sharing of the audio recording or capture operation.

Content or audio signal discontinuities can occur, especially when the recorded content is uploaded to the content server after some time the recording has taken place that the uploaded content represents an edited version rather than the
25 actual recorded content. For example the user can edit any recorded content before uploading the content to the content server. The editing can for example involve removing unwanted segments from the original recording. The signal discontinuity can create significant challenges to the content server as typically an implicit assumption is made that the uploaded content represents the audio signal
30 or clip from a continuous timeline. Where segments are removed (or added) after recording has ended then the continuity assumption or condition no longer holds for the particular content.

- Furthermore to be able to jointly utilize the multi-user recorded content for various media rendering methods, such as audio mixing from multiple users and video view switching from one user to the other, the content between different users must employ a 'common' time or timeline. Furthermore, the common timeline
- 5 should be constructed such that the content from different devices or apparatus shares the same event space. For example users and their apparatus or devices may move in and out of a defined audio event space during recording or capturing resulting in a situation where there may be time periods when some apparatus do not share the same event space even though they share the same timeline.
- 10 Furthermore depending on the event venue there may exist multiple event spaces that are not correlated. For example an event venue with different rooms and/or floors can result in multiple event spaces from the content capturing and rendering point of view.
- 15 The concept as described herein in embodiments is to analyse and segment the recorded or captured content from an event venue into different event spaces. This invention outlines method for creating event spaces from multi-user captured content. The concept can further be summarized according to the following steps:
- 20 • Classifying recorded or captured media content to generate media segments associated with a defined class
- Applying analysis to media segments based the associated class
 - Determining similarities between segments of different user/apparatus media
- 25 • Creating event spaces based on similarity status for different user/apparatus media

In some embodiments the classification comprises at least 2 classes, for example a music class and a non-music class. Furthermore in some embodiments the

30 classes can furthermore be sub-divided into sub-classes of which the sub-classes are grouped, for example the music class can be divided into a music-classical, and music-rock sub-classes.

In some embodiments the media analysis is applied to each class present in the segment from different user media.

5 In some embodiments the audio domain properties are used to provide event space separation resulting fast and computationally efficient operation.

With respect to Figure 1 an overview of a suitable system within which embodiments of the application can be located is shown. The audio space 1 can have located within it at least one recording or capturing device or apparatus 19
10 which are arbitrarily positioned within the audio space to record suitable audio scenes. The apparatus 19 shown in Figure 1 are represented as microphones with a polar gain pattern 101 showing the directional audio capture gain associated with each apparatus. The apparatus 19 in Figure 1 are shown such that some of the apparatus are capable of attempting to capture the audio scene
15 or activity 103 within the audio space. The activity 103 can be any event the user of the apparatus wishes to capture. For example the event could be a music event or audio of a "news worthy" event. The apparatus 19 although being shown having a directional microphone gain pattern 101 would be appreciated that in some embodiments the microphone or microphone array of the recording
20 apparatus 19 has a omnidirectional gain or different gain profile to that shown in Figure 1.

Each recording apparatus 19 can in some embodiments transmit or alternatively store for later consumption the captured audio signals via a transmission channel
25 107 to an audio scene server 109. The recording apparatus 19 in some embodiments can encode the audio signal to compress the audio signal in a known way in order to reduce the bandwidth required in "uploading" the audio signal to the audio scene server 109.

30 The recording apparatus 19 in some embodiments can be configured to estimate and upload via the transmission channel 107 to the audio scene server 109 an estimation of the location and/or the orientation or direction of the apparatus. The position information can be obtained, for example, using GPS coordinates, cell-ID

or a-GPS or any other suitable location estimation methods and the orientation/direction can be obtained, for example using a digital compass, accelerometer, or gyroscope information.

5 In some embodiments the recording apparatus 19 can be configured to capture or record one or more audio signals for example the apparatus in some embodiments have multiple microphones each configured to capture the audio signal from different directions. In such embodiments the recording device or apparatus 19 can record and provide more than one signal from different the
10 direction/orientations and further supply position/direction information for each signal. With respect to the application described herein an audio or sound source can be defined as each of the captured or audio recorded signal. In some embodiments each audio source can be defined as having a position or location which can be an absolute or relative value. For example in some embodiments
15 the audio source can be defined as having a position relative to a desired listening location or position. Furthermore in some embodiments the audio source can be defined as having an orientation, for example where the audio source is a beamformed processed combination of multiple microphones in the recording apparatus, or a directional microphone. In some embodiments the orientation may
20 have both a directionality and a range, for example defining the 3dB gain range of a directional microphone.

The capturing and encoding of the audio signal and the estimation of the position/direction of the apparatus is shown in Figure 1 by step 1001.

25

The uploading of the audio and position/direction estimate to the audio scene server 109 is shown in Figure 1 by step 1003.

The audio scene server 109 furthermore can in some embodiments communicate
30 via a further transmission channel 111 to a listening device 113.

In some embodiments the listening device 113, which is represented in Figure 1 by a set of headphones, can prior to or during downloading via the further

transmission channel 111 select a listening point, in other words select a position such as indicated in Figure 1 by the selected listening point 105. In such embodiments the listening device 113 can communicate via the further transmission channel 111 to the audio scene server 109 the request.

5

The selection of a listening position by the listening device 113 is shown in Figure 1 by step 1005.

The audio scene server 109 can as discussed above in some embodiments receive from each of the recording apparatus 19 an approximation or estimation of the location and/or direction of the recording apparatus 19. The audio scene server 109 can in some embodiments from the various captured audio signals from recording apparatus 19 produce a composite audio signal representing the desired listening position and the composite audio signal can be passed via the further transmission channel 111 to the listening device 113.

The generation or supply of a suitable audio signal based on the selected listening position indicator is shown in Figure 1 by step 1007.

20 In some embodiments the listening device 113 can request a multiple channel audio signal or a mono-channel audio signal. This request can in some embodiments be received by the audio scene server 109 which can generate the requested multiple channel data.

25 The audio scene server 109 in some embodiments can receive each uploaded audio signal and can keep track of the positions and the associated direction/orientation associated with each audio source. In some embodiments the audio scene server 109 can provide a high level coordinate system which corresponds to locations where the uploaded/upstreamed content source is available to the listening device 113. The "high level" coordinates can be provided for example as a map to the listening device 113 for selection of the listening position. The listening device (end user or an application used by the end user) can in such embodiments be responsible for determining or selecting the listening

30

position and sending this information to the audio scene server 109. The audio scene server 109 can in some embodiments receive the selection/determination and transmit the downmixed signal corresponding to the specified location to the listening device. In some embodiments the listening device/end user can be
5 configured to select or determine other aspects of the desired audio signal, for example signal quality, number of channels of audio desired, etc. In some embodiments the audio scene server 109 can provide in some embodiments a selected set of downmixed signals which correspond to listening points neighbouring the desired location/direction and the listening device 113 selects
10 the audio signal desired.

In this regard reference is first made to Figure 2 which shows a schematic block diagram of an exemplary apparatus or electronic device 10, which may be used to record (or operate as a recording or capturing apparatus 19) or listen (or operate
15 as a listening apparatus 113) to the audio signals (and similarly to record or view the audio-visual images and data). Furthermore in some embodiments the apparatus or electronic device can function as the audio scene server 109.

The electronic device 10 may for example be a mobile terminal or user equipment
20 of a wireless communication system when functioning as the recording device or listening device 113. In some embodiments the apparatus can be an audio player or audio recorder, such as an MP3 player, a media recorder/player (also known as an MP4 player), or any suitable portable device suitable for recording audio or audio/video camcorder/memory audio or video recorder.

25

The apparatus 10 can in some embodiments comprise an audio subsystem. The audio subsystem for example can comprise in some embodiments a microphone or array of microphones 11 for audio signal capture. In some embodiments the microphone or array of microphones can be a solid state microphone, in other
30 words capable of capturing audio signals and outputting a suitable digital format signal. In some other embodiments the microphone or array of microphones 11 can comprise any suitable microphone or audio capture means, for example a condenser microphone, capacitor microphone, electrostatic microphone, electret

condenser microphone, dynamic microphone, ribbon microphone, carbon microphone, piezoelectric microphone, or microelectrical-mechanical system (MEMS) microphone. The microphone 11 or array of microphones can in some embodiments output the audio captured signal to an analogue-to-digital converter (ADC) 14.

In some embodiments the apparatus can further comprise an analogue-to-digital converter (ADC) 14 configured to receive the analogue captured audio signal from the microphones and outputting the audio captured signal in a suitable digital form. The analogue-to-digital converter 14 can be any suitable analogue-to-digital conversion or processing means.

In some embodiments the apparatus 10 audio subsystem further comprises a digital-to-analogue converter 32 for converting digital audio signals from a processor 21 to a suitable analogue format. The digital-to-analogue converter (DAC) or signal processing means 32 can in some embodiments be any suitable DAC technology.

Furthermore the audio subsystem can comprise in some embodiments a speaker 33. The speaker 33 can in some embodiments receive the output from the digital-to-analogue converter 32 and present the analogue audio signal to the user. In some embodiments the speaker 33 can be representative of a headset, for example a set of headphones, or cordless headphones.

Although the apparatus 10 is shown having both audio capture and audio presentation components, it would be understood that in some embodiments the apparatus 10 can comprise one or the other of the audio capture and audio presentation parts of the audio subsystem such that in some embodiments of the apparatus the microphone (for audio capture) or the speaker (for audio presentation) are present.

In some embodiments the apparatus 10 comprises a processor 21. The processor 21 is coupled to the audio subsystem and specifically in some

examples the analogue-to-digital converter 14 for receiving digital signals representing audio signals from the microphone 11, and the digital-to-analogue converter (DAC) 12 configured to output processed digital audio signals. The processor 21 can be configured to execute various program codes. The
5 implemented program codes can comprise for example audio signal or content shot detection routines.

In some embodiments the apparatus further comprises a memory 22. In some embodiments the processor is coupled to memory 22. The memory can be any
10 suitable storage means. In some embodiments the memory 22 comprises a program code section 23 for storing program codes implementable upon the processor 21. Furthermore in some embodiments the memory 22 can further comprise a stored data section 24 for storing data, for example data that has
15 been analysed and classified in accordance with the application or data to be analysed or classified via the application embodiments as described later. The implemented program code stored within the program code section 23, and the data stored within the stored data section 24 can be retrieved by the processor 21 whenever needed via the memory-processor coupling.

20 In some further embodiments the apparatus 10 can comprise a user interface 15. The user interface 15 can be coupled in some embodiments to the processor 21. In some embodiments the processor can control the operation of the user interface and receive inputs from the user interface 15. In some embodiments the user interface 15 can enable a user to input commands to the electronic device or
25 apparatus 10, for example via a keypad, and/or to obtain information from the apparatus 10, for example via a display which is part of the user interface 15. The user interface 15 can in some embodiments comprise a touch screen or touch interface capable of both enabling information to be entered to the apparatus 10 and further displaying information to the user of the apparatus 10.

30

In some embodiments the apparatus further comprises a transceiver 13, the transceiver in such embodiments can be coupled to the processor and configured to enable a communication with other apparatus or electronic devices, for

example via a wireless communications network. The transceiver 13 or any suitable transceiver or transmitter and/or receiver means can in some embodiments be configured to communicate with other electronic devices or apparatus via a wire or wired coupling.

5

The coupling can, as shown in Figure 1, be the transmission channel 107 (where the apparatus is functioning as the recording device 19 or audio scene server 109) or further transmission channel 111 (where the device is functioning as the listening device 113 or audio scene server 109). The transceiver 13 can
10 communicate with further devices by any suitable known communications protocol, for example in some embodiments the transceiver 13 or transceiver means can use a suitable universal mobile telecommunications system (UMTS) protocol, a wireless local area network (WLAN) protocol such as for example IEEE 802.X, a suitable short-range radio frequency communication protocol such
15 as Bluetooth, or infrared data communication pathway (IRDA).

In some embodiments the apparatus comprises a position sensor 16 configured to estimate the position of the apparatus 10. The position sensor 16 can in some embodiments be a satellite positioning sensor such as a GPS (Global Positioning
20 System), GLONASS or Galileo receiver.

In some embodiments the positioning sensor can be a cellular ID system or an assisted GPS system.

25 In some embodiments the apparatus 10 further comprises a direction or orientation sensor. The orientation/direction sensor can in some embodiments be an electronic compass, accelerometer, a gyroscope or be determined by the motion of the apparatus using the positioning estimate.

30 It is to be understood again that the structure of the electronic device 10 could be supplemented and varied in many ways.

Furthermore it could be understood that the above apparatus 10 in some embodiments can be operated as an audio scene server 109. In some further embodiments the audio scene server 109 can comprise a processor, memory and transceiver combination.

5

In the following examples there are described an audio scene/content recording or capturing apparatus which correspond to the recording device 19 and an audio scene/content co-ordinating or management apparatus which corresponds to the audio scene server 109. However it would be understood that in some
10 embodiments the audio scene management apparatus can be located within the recording or capture apparatus as described herein and similarly the audio scene recording or content capture apparatus can be a part of an audio scene server 109 capturing audio signals either locally or via a wireless microphone coupling.

15 With respect to Figure 3 an example content co-ordinating apparatus according to some embodiments is shown which can be implemented within the recording device 19, the audio scene server, or the listening device (when acting as a content aggregator). Furthermore Figure 4 shows a flow diagram of the operation of the example content co-ordinating apparatus shown in Figure 3 according to
20 some embodiments.

In some embodiments the content coordinating apparatus comprises an audio input 201. The audio input 201 can in some embodiments be the microphone input, or a received input via the transceiver or other wire or wireless coupling to
25 the apparatus. In some embodiments the audio input 201 is the memory 22 and in particular the stored data memory 24 where any edited or unedited audio signal is stored.

The operation of receiving the audio input is shown in Figure 4 by step 301.

30

In some embodiments the content coordinating apparatus comprises a content classifier 203. The content classifier 203 can in some embodiments receive the audio input signal and be configured (where the input signal is not originally) to

align the input audio signal according to its initial time stamp value. In the following example the input audio signal has a start timestamp $T=x$ and length or end time stamp $T=y$, in other words the input audio signal is defined by the pair wise value of (x, y) .

5

In some embodiments the initial time stamp based alignment can be performed with respect to one or more reference audio content parts. In some embodiments the input audio signal is aligned against a reference audio content time stamp where both the input audio signal and reference audio signal are known to use a common clock time stamp. For example in some embodiments the recording of the audio signal can be performed with an initial time stamp provided the apparatus internal clock or a received clock signal, such as a cellular clock time stamp, a positioning or GPS clock time stamp or any other received clock signal.

15 The operation of initially aligning the input audio signal against a reference signal or generating a common timeline is shown in Figure 4 by step 303.

With respect to Figure 6 an example audio signal or media set is shown. In this example there are three audio signals or media signals. A first audio or media signal A 501, a second audio or media signal B 503 and a third audio or media signal C 505. In this example the three example audio signals A 501, B 503 and C 505 are received and aligned relative to each other. In the example shown in Figure 6 the audio signals are ordered such that the audio signal A 501 is the first audio signal 'received' or the first to start, the audio signal B 503 is the second audio signal 'received' or the second to start, and the audio signal C 505 is the third audio signal received or the third to start. Furthermore in the audio signals are ordered such that the audio signal A 501 is the second audio signal to finish, the audio signal B 503 is the third to finish, and the audio signal C 505 is the first to finish with the overall length of the audio signal being audio signal B 503 being longest, the audio signal C 505 the shortest and audio signal A 501 being slightly shorter than audio signal B 503.

Furthermore in some embodiments the content classifier 203 is configured to analyse the audio signal and determine a classification of the audio signal segment.

- 5 In some embodiments the content classifier 203 is configured to analyse the input audio signals and segment the audio signal according to a determined or associated class. For example in some embodiments the content classifier 203 can be configured analyse the received audio (or media) signal and assign parts or segments to classes such as a 'music' or 'non-music' class.

10

It would be understood that in some embodiments there can be more than two classes or sub-classes. For example in some embodiments there can be sub-classes within each class. For example in some embodiments the content classifier 203 can be configured to determine a 'classical music' segment or
15 assign or associate a 'classical music' sub-class to an audio segment and a 'rock music' segment or assign or associate a 'rock music' sub-class to a different segment. For example the audio signal captured or recorded by the same apparatus can change class as the apparatus moves from a first room playing classical music to a second room playing rock music.

20

For example Figure 5 shows a representation of a captured or recorded audio signal or media which is analysed by the content classifier 203, and segmented into three parts based on the determined audio signal class. In this example the captured or recorded audio signal comprises a first part or segment 401 which is
25 determined as being or associated with a non-music class, a second part or segment 403 which is determined as being or associated with a music class, and a third part or segment 405 which is determined as being or associated with a non-music class.

30

Furthermore with respect to Figure 7 the segmentation of the example audio signal or media set shown in Figure 5 is shown. In this example the first audio or media signal A 501 comprises a first non-music segment 601 followed by a music segment 603. The second audio or media signal B 503 comprises a first non-

music segment 611 followed by a music segment 613. The third audio or media signal C 505 comprises a music segment 623.

In the example shown in Figure 7 the segmentation of the example audio signal or media set is such that the boundary between the non-music segments 601, 611 from the first audio signal A 501 and the second audio signal B 503 and the music segments 603, 613, 623 from the first audio signal A 501, the second audio signal B 503 and the third audio signal C 505 respectively are not aligned.

- 10 The operation of classifying media or generating segments with associated classes is shown in Figure 4 by step 304.

In some embodiments the audio input can be pre-classified, in other words the audio signal is received with metadata with associated classification values associated with the audio signal and defining audio signal segments.

In some embodiments the content classifier 203 can be configured to output the classified content or captured audio to a content segment smoother 205.

- 20 In some embodiments the content classifier 203 can be configured to receive the audio signal and generate frames or sub-frames time divided parts of the audio signal. For example in some embodiments the content classifier 203 can be configured to generate a frame of 20ms where each frame comprises a sub-frame which overlaps by 10ms with the preceding frame and a second sub-frame which overlaps by 10ms with the succeeding frame.

Furthermore in some embodiments the content classifier 203 is configured to analyse the audio signal on a frame by frame (or sub-frame by sub-frame) basis, and for each frame (or sub-frame) determine at least one possible feature or parameter value. In such embodiments each classification or class or sub-class can have an assigned or associated feature value range against which the determined feature or parameter value or feature (or parameter) values can then be compared to determine a classification or class for the frame (or sub-frame).

For example the feature values for a frame can in some embodiments be located within a space or vector map within which are determined classification boundaries defining audio classifications and from which can be determined a
5 classification for each frame.

For example a classifier which can be used in some embodiments is the one described in "Features for Audio and Music Classification" by McKinney and Breebaart, Proc. 4th Int. Conf. on Music Information Retrieval, which is configured
10 to determine classifications such as Classical Music, Jazz, Folk, Electronica, R&B, Rock, Reggae, Vocal, Speech, Noise, and Crowd Noise.

The analysis features can in some embodiments be any suitable features such as spectral features such as cepstral coefficients, frequency warping, magnitude
15 warping, Mel-frequency cepstral coefficients, spectral centroid, bandwidth, temporal features such as rise time, onset asynchrony at different frequencies, frequency modulation (amplitude and rate), amplitude modulation (amplitude and rate), zero crossing rate, short-time energy values, etc.

20 In some embodiments the features are selected for analysis according to any suitable manner, for example data normalisation, Sequential backward selection (SBS), principal component analysis, Eigenanalysis (determining the eigenvectors and eigenvalues of the data set), or feature transformation (linear or otherwise) can be used.

25

The classifier can in some embodiments generate classification from the feature values according to any suitable manner, such as for example by a supervised (or taught) classifier or unsupervised classifier. The classifier can for example in some embodiments be configured to use a minimum distance classification
30 method. In some embodiments the classifier be configured to use a k-nearest neighbour (k-NN) classifier where the k nearest neighbours are picked to the feature value x and then choose the class which was most often picked. In some embodiments the classifier employs statistical classification techniques where the

feature vector value is interpreted as a random variable whose distribution depends on the class (for example by applying Bayesian, Gaussian mixture models, or maximum a posteriori MAP, Hidden Markov model HMM, methods).

- 5 The exact set of classes or classifications can in some embodiments vary depending on the audio signals being analysed and the environment within which the audio signals were recorded or are being analysed. For example in some embodiments there can be a user interface input selecting the set of classes, or the set of classes can be chosen by an automatic or semi-automatic means.

10

In some embodiments the content coordinating apparatus comprises a content segment smoother 205. The content segment smoother 205 can be configured to receive the audio signal or media content which has been analysed and segmented by the content classifier 203 and filter audio signals. The purpose of
15 this filtering is to adjust the class segment boundaries such that small differences in the start and end boundaries between different audio signals media are removed.

For example with respect to the example segmentation as shown in Figure 6 the
20 content segment smoother 205 filtering is shown as action 651, the results of which are shown in Figure 8.

With respect to Figure 8 the segmentation of the example audio signal or media set shown in Figure 5 and 6 is shown having been filtered to attempt to align any
25 small difference between audio signal segments with the same class. For example using the first audio or media signal A 501 as a reference signal with a segment boundary between the non-music segment 701 and the music segment 703 at time instant t_2 733 and with the music segment ends at time instant t_4 737. The content segment smoother 205 can then in some embodiments be
30 configured to filter or shift the second audio or media signal B 503 such that the non-music segment 711 starts at time t_1 731 so that the non-music segment 711 ends at time t_2 733 (in other words aligns the end of the second audio signal B 503 non-music segment 711 to the first audio signal A 501 non-music segment).

This shift furthermore aligns the second audio signal B 503 music segment 713 to the first audio signal A 501 music segment 703.

5 The content segment smoother 205 can then in some embodiments be configured to filter or shift the third audio or media signal C 505 such that the music segment 723 starts at time t_2 733 in other words aligns the start of the third audio signal C 505 music segment 723 to the first audio signal A 501 music segment 703. This shift furthermore aligns the second audio signal B 503 music segment 713 to the third audio signal C 505 music segment 723. The shift of the
10 third audio signal means that the third audio signal C 505 music segment ends at time t_3 735, which occurs before t_4 737.

The content segment smoother 205 can then, in some embodiments, output the 'filtered' or aligned signals to a class segment analyser 207.
15

The smoothing or filtering of the class segments is shown in Figure 4 by step 305.

In some embodiments the content coordinating apparatus comprises a class segment analyser 207. The class segment analyser 207 can be configured to
20 receive the segmented and smoothed audio signals or media content. The class segment analyser 207 can in some embodiments comprise a class based signal structure analyser for the determined classes. Thus for example as shown in Figure 3 the class segment analyser 207 comprises a non-music signal structure analyser 221 and a music signal structure analyser 223. Furthermore the generic
25 class signal structure analyser is represented within Figure 3 by the <class> structure analyser 225.

In some embodiments the class segment analyser 207 is configured to allocate class segments to their associated class signal structure analysers. Thus for
30 example with respect to the audio signals shown in Figure 8 the class segment analyser 207 is configured to allocate the audio signal A 501 non-music segment 701 to the non-music signal structure analyser 221, and the music segment 703 to the music segment analyser 223. Furthermore the class segment analyser 207

is configured to allocate the audio signal B 503 non-music segment 711 to the non-music signal structure analyser 221, and the music segment 713 to the music segment analyser 223. With respect to the audio signal C 505 the music segment 723 is allocated by the class segment analyser 207 to the music segment
5 analyser 223.

The operation of allocating class segment to class structure analysers is shown in Figure 4 by step 307.

10 In some embodiments the class signal structure analysers are then configured to analyse the allocated audio or media segments for any overlapping audio segments. In other words the class signal structure analysers are configured to analyse pairs of audio signals where there are at least two audio signals with the same segment class at the same time. The number of analyses to be applied for
15 a media segment depends on the amount of different classes within the overlapping class segment. For example, if only music segments are present within the overlapping class segment, then music based analysis is applied for each media segment. However, if the overlapping class segment contains 2 or more different classes (that is, one media segment may get assigned to 'music'
20 whereas the some other media segment may get assigned to 'non-music'), then the same number of analyses are applied to each media segment regardless whether a particular class was assigned initially to the media segment or not. The class signal structure analysis results can then be passed to the pairwise difference analyser 209.

25

Thus for example with respect to the audio signals shown in Figure 8 the timeline comprises 3 overlapping class segments. These overlapping class segments are the time period from $t_1 - t_2$, $t_2 - t_3$, and $t_3 - t_4$. The first overlapping class segment from $t_1 - t_2$ comprises part of audio signal A 501 non-music segment 701 and
30 audio signal B 503 non-music segment 711 and is analysed by the non-music signal structure analyser 221. The non-music signal structure analyser 221 can be configured to analyse these of audio signals and pass the results to the pairwise difference analyser 209.

The second overlapping class segment from $t_2 - t_3$ comprises part of audio signal A 501 music segment 703, part of audio signal B 503 music segment 713 and audio signal C 505 music segment 723 and is analysed by the music signal structure analyser 223. The music signal structure analyser 223 can be configured to analyse these of audio signals and pass the results to the pairwise difference analyser 209.

The third overlapping class segment from $t_3 - t_4$ comprises a latter part of audio signal A 501 music segment 703, and a latter part of audio signal B 503 music segment 713 and is analysed by the music signal structure analyser 223. The music signal structure analyser 223 can be configured to analyse these of audio signals and pass the results to the pairwise difference analyser 209.

In some embodiments the signal structure analysers, such as the non-music signal structure analyser 221 and the music structure analyser 223 are is configured to analyse the audio signal segments on a frame by frame (or sub-frame by sub-frame) basis, and for each frame (or sub-frame) determine at least one possible class based feature or parameter value. In such embodiments the class based at least one feature or parameter has values which can then be compared to determine differences within the pairwise difference analyser 209.

In some embodiments the class based at least one feature or parameter value for a frame can be the same values which were used by the content classifier to define the classes. For example a classifier which can be used in some embodiments is the one described in "Features for Audio and Music Classification" by McKinney and Breebaart, Proc. 4th Int. Conf. on Music Information Retrieval, which is configured to determine classifications such as Classical Music, Jazz, Folk, Electronica, R&B, Rock, Reggae, Vocal, Speech, Noise, and Crowd Noise.

The analysis features can in some embodiments be any suitable features such as spectral features such as cepstral coefficients, frequency warping, magnitude

warping, Mel-frequency cepstral coefficients, spectral centroid, bandwidth, temporal features such as rise time, onset asynchrony at different frequencies, frequency modulation (amplitude and rate), amplitude modulation (amplitude and rate), zero crossing rate, short-time energy values, etc.

5

It would be understood that in some embodiments a first class signal structure analyser can be configured to generate or determine a first set of features or parameters while a second class signal structure analyser can be configured to generate or determine a second set of features or parameters. In some
10 embodiments the first set of features overlaps at least partially the second set of features.

For example for the 'music' class the music class dependent analysis can comprise any suitable music structure analysis techniques. For example, in some
15 embodiments the bars (or beats) of a music segment are determined and compared.

In some embodiments the signal structure analysers are configured to filter the feature or parameter values determined by the content classifier 203 and pass the
20 filtered feature or parameter values to the pairwise difference analyser 209.

The operation of generating class based structure analysis is shown in Figure 4 by step 309.

25 In some embodiments the content coordinating apparatus comprises a pairwise difference analyser 209.

The pairwise difference analyser 209 can be configured in some embodiments to receive the signal structure analysis results and pairwise analyse these to
30 determine differences which are passed to an event space assigner 211. In some embodiments the pairwise difference analyser 209 is configured to perform a decision based on the difference to determine whether the pairwise selection is similar or not. In other words the pairwise difference analyser can be configured

to compare on an audio signal or media segment pairwise manner to determine whether the signal structure analysis results are similar enough (indicating same event space) or not (indicating different event space).

- 5 The operation of generating pairwise media structure difference is shown in Figure 4 by step 311.

In some embodiments the comparison is applied with respect to the other audio signals or media segments within the same overlapping class segment.

10

In other words the class structure differences in some embodiments can be combined.

15

The operation of combining class structure differences is shown in Figure 4 by step 313.

The analysis comparison can in some embodiments be configured to return a binary decision value 0 or 1 which can be then summed across all applied analyses classes.

20

For example with respect to the music segments and where the feature value is bar or beat times where the difference in bar or beat times is too great (for example in the order of a second or more), the media in the pair are not similar and a binary decision of similarity of 0 is generated, and where the difference is less than the determined value (for example one second) then a binary decision of similarity of 1 is generated.

25

30

The content coordinating apparatus can in some embodiments comprise an event space assigner 211. The event space assigner 211 can be configured to receive the output of the pairwise difference analyser (for example the similarity binary decision or the difference values combined) and then determine whether the media pairs or audio signals are similar enough to be assigned to the same event space or not.

In some embodiments the event space assigner can therefore assign the same event space to the pair of audio signals by analysing the binary decision. In some embodiments the same event space determination can be made from the difference values output by the pairwise difference analyser.

Thus for example the event space assigner 211 can be configured to determine whether the media pairs are similar based on the combined class structure difference.

The operation of determining whether the media pairs are similar based on the combined class structure difference is shown in Figure 4 by step 315.

Where the media pairs are similar based on the combined class structure difference then the event space assigner 211 can be configured to assign both to the same event space.

The operation of assigning both of the audio signals (media pair) to the same event space is shown in Figure 4 by step 317.

Where the media pairs are not similar based on the combined class structure difference then the event space assigner 211 can be configured to assign the audio signals or media to different event spaces.

The operation of assigning the audio signals (media) to difference event spaces is shown in Figure 4 by step 319.

With respect to Figure 9 the example audio signals as shown in Figures 6 to 8 are shown having been assigned. The event space assignment for each overlapping class segment operates such that one of the media in the pair has already been assigned to at least one event space and the other media is assigned at this stage. For example the audio signals A 501, B 503, and C 505 for some arbitrary overlapping class segment, the event space assignment can be as follows:

The audio signal pairs are: A-B (for the first overlap period from $t_1 - t_2$), A-C (for the second overlap period from $t_2 - t_3$), and B-C (for the second overlap period from $t_1 - t_2$), and A-B (for the third overlap period from $t_3 - t_4$)

5

In some embodiments the event space assigner can be configured to assign for the audio signal A 501 non-music segment 701 event space 1 801.

The first audio signal pairing is A-B (for the first overlap period from $t_1 - t_2$) where
10 A non-music segment 701 is part of event space 1

Audio signal B non-music segment 711 therefore is assigned to event space 1 if the audio signals are similar or otherwise assigned to event space 2 (which would be a new event space). In the example as shown in Figure 9 the Audio signal B
15 non-music segment 711 is similar and therefore is assigned to event space 1.

In some embodiments the event space assigner can be configured to assign for the audio signal A 501 music segment 703 event space 2 803.

20 The second audio signal or media pair is A-C (for the second overlap period from $t_2 - t_3$) where audio signal A 501 music segment 703 is part of event space 2 803. Audio signal or Media C therefore assigned to event space 2 where they are similar or otherwise to event space 3 (a new event space). In the example as shown in Figure 9 the Audio signal C music segment 723 is similar and therefore
25 is assigned to event space 2.

The third audio signal or media pair is B-C (for the second overlap period from $t_2 - t_3$) where audio signal C 505 music segment 723 is part of event space 2 803 (it would be understood that a similar pairing can be A-B which would lead to a
30 similar result).

Audio signal or Media B is therefore assigned to event space 2 where they are similar or otherwise to event space 3 (a new event space). In the example as

shown in Figure 9 the Audio signal B music segment 713 is similar and therefore is assigned to event space 2.

A fourth audio signal or media pair is A-B (for the third overlap period from $t_3 - t_4$)
5 where audio signal A 501 music segment 703 is part of event space 2 803.

Audio signal or Media B is therefore assigned to event space 2 for the third overlap period where they are similar or otherwise to event space 3 (a new event space). In the example as shown in Figure 9 the Audio signal B music segment
10 713 for the third overlap period is similar and therefore is assigned to event space 2.

After all overlapping class segments have been processed, each media has been assigned to at least one event space.

15

In some embodiments the audio signal may get over-segmented, that is, it is assigned to too many event spaces. This may occur especially for example where the classification is not able to detect classes correctly (for example, it is not able to decide whether media segment belongs to 'music' or 'non-music' and class
20 segment alternates as a function of time). In some embodiments in order to reduce the risk of assigning too many event spaces to one media then event spaces for an audio signal or media are filtered such that a higher priority is given to event spaces with longer duration.

25 For example, an event space with a short duration (less than 10sec) is between longer duration event spaces (longer than 20sec), then the short duration event space is discarded and assigned to the longer duration event space.

The foregoing description has provided by way of exemplary and non-limiting
30 examples a full and informative description of the exemplary embodiment of this invention. However, various modifications and adaptations may become apparent to those skilled in the relevant arts in view of the foregoing description, when read in conjunction with the accompanying drawings.

Although the above has been described with regards to audio signals, or audio-visual signals it would be appreciated that embodiments may also be applied to audio-video signals where the audio signal components of the recorded data are processed in terms of the determining of the base signal and the determination of the time alignment factors for the remaining signals and the video signal components may be synchronised using the above embodiments of the invention. In other words the video parts may be synchronised using the audio synchronisation information.

It shall be appreciated that the term user equipment is intended to cover any suitable type of wireless user equipment, such as mobile telephones, portable data processing devices or portable web browsers.

Furthermore elements of a public land mobile network (PLMN) may also comprise apparatus as described above.

In general, the various embodiments of the invention may be implemented in hardware or special purpose circuits, software, logic or any combination thereof.

For example, some aspects may be implemented in hardware, while other aspects may be implemented in firmware or software which may be executed by a controller, microprocessor or other computing device, although the invention is not limited thereto. While various aspects of the invention may be illustrated and described as block diagrams, flow charts, or using some other pictorial representation, it is well understood that these blocks, apparatus, systems, techniques or methods described herein may be implemented in, as non-limiting examples, hardware, software, firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination thereof.

The embodiments of this invention may be implemented by computer software executable by a data processor of the mobile device, such as in the processor entity, or by hardware, or by a combination of software and hardware. Further in

this regard it should be noted that any blocks of the logic flow as in the Figures may represent program steps, or interconnected logic circuits, blocks and functions, or a combination of program steps and logic circuits, blocks and functions. The software may be stored on such physical media as memory chips, or memory blocks implemented within the processor, magnetic media such as hard disk or floppy disks, and optical media such as for example DVD and the data variants thereof, CD.

The memory may be of any type suitable to the local technical environment and may be implemented using any suitable data storage technology, such as semiconductor-based memory devices, magnetic memory devices and systems, optical memory devices and systems, fixed memory and removable memory. The data processors may be of any type suitable to the local technical environment, and may include one or more of general purpose computers, special purpose computers, microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASIC), gate level circuits and processors based on multi-core processor architecture, as non-limiting examples.

Embodiments of the inventions may be practiced in various components such as integrated circuit modules. The design of integrated circuits is by and large a highly automated process. Complex and powerful software tools are available for converting a logic level design into a semiconductor circuit design ready to be etched and formed on a semiconductor substrate.

Programs, such as those provided by Synopsys, Inc. of Mountain View, California and Cadence Design, of San Jose, California automatically route conductors and locate components on a semiconductor chip using well established rules of design as well as libraries of pre-stored design modules. Once the design for a semiconductor circuit has been completed, the resultant design, in a standardized electronic format (e.g., Opus, GDSII, or the like) may be transmitted to a semiconductor fabrication facility or "fab" for fabrication.

The foregoing description has provided by way of exemplary and non-limiting examples a full and informative description of the exemplary embodiment of this invention. However, various modifications and adaptations may become apparent to those skilled in the relevant arts in view of the foregoing description, when read
5 in conjunction with the accompanying drawings and the appended claims. However, all such and similar modifications of the teachings of this invention will still fall within the scope of this invention as defined in the appended claims.

CLAIMS:

1. Apparatus comprising at least one processor and at least one memory
5 including computer code for one or more programs, the at least one memory and
the computer code configured to with the at least one processor cause the
apparatus to at least:
 select at least two audio signals;
 segment the at least two audio signals based on at least two defined class
10 definitions;
 determine a difference measure between at least a pair of common class
segments from the at least two audio signals using a class based analyser; and
 determine the at least two audio signal common class segments are within
a common event space based on the difference measure.
15
2. The apparatus as claimed in claim 1, further caused to generate a
common time line incorporating the at least two audio signals.
3. The apparatus as claimed in claim 2, further caused to align the at least
20 two audio signals when a time difference between two of the at least two audio
signals is less than a defined threshold, wherein the time difference is the
difference on a common time line incorporating the at least two audio signals for a
pair of common class segments from the at least two audio signals.
- 25 4. The apparatus as claimed in claims 1 to 3, wherein segmenting the at least
two audio signals based on the at least two classes causes the apparatus to:
 analyse the at least two audio signals to determine at least one parameter;
 segment the at least two audio signals into parts where the parts of the at
least two audio signals are associated with at least one range of values
30 associated with the at least one parameter.
5. The apparatus as claimed in claims 4, wherein analysing the at least two
audio signals to determine at least one parameter causes the apparatus to:

divide at least one of the audio signals into a number of frames;
analyse for at least one frame of the number of frames of the at least one
audio signal to determine the at least one parameter value; and
determine a class for the at least one frame based on at least one defined
5 range of parameter values.

6. The apparatus as claimed in claims 1 to 5, wherein the at least two classes
comprise at least two of:

10 music;
speech; and
noise.

7. The apparatus as claimed in claims 1 to 6, wherein determining a
difference measure between at least a pair of common class segments from the
15 at least two audio signals using a class based analyser causes the apparatus to:
allocate a pair of common class segments which overlap to an associated
class based analyser; and
determine a distance value using the associated class based analyser for
the pair of common class segments.

20

8. The apparatus as claimed in claim 7, wherein determining a distance value
using the associated class based analyser for the pair of common class segments
further causes the apparatus to determine a binary distance value.

25 9. The apparatus as claimed in claims 1 to 8, further caused to determine
whether the at least two audio signals are within a common event space based on
the determination of the at least two audio signal common class segments are
within a common event space.

30 10. Apparatus comprising:
means for selecting at least two audio signals;
means for segmenting the at least two audio signals based on at least two
defined class definitions;

means for determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and

5 means for determining the at least two audio signal common class segments are within a common event space based on the difference measure.

11. The apparatus as claimed in claim 10, further comprising means for generating a common time line incorporating the at least two audio signals.

10 12. The apparatus as claimed in claim 11, further comprising means for aligning the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least two audio signals for a pair of common class segments from the at least two audio
15 signals.

13. The apparatus as claimed in claims 10 to 12, wherein the means for segmenting the at least two audio signals based on the at least two classes comprises:

20 means for analysing the at least two audio signals to determine at least one parameter; and

means for segmenting the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

25

14. A method comprising:

selecting at least two audio signals;

segmenting the at least two audio signals based on at least two defined class definitions;

30 determining a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and

determining the at least two audio signal common class segments are within a common event space based on the difference measure.

15. The method as claimed in claim 14, further comprising generating a common time line incorporating the at least two audio signals.

5 16. The method as claimed in claim 15, further comprising aligning the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least two audio signals for a pair of common class segments from the at least two audio signals.

10

17. The method as claimed in claims 14 to 16, wherein segmenting the at least two audio signals based on the at least two classes comprises:

analysing the at least two audio signals to determine at least one parameter; and

15 segmenting the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

18. An apparatus comprising:

20 an input configured to select at least two audio signals;

a classifier configured to segment the at least two audio signals based on at least two defined class definitions;

25 a class segment analyser configured to determine a difference measure between at least a pair of common class segments from the at least two audio signals using a class based analyser; and

a difference analyser configured to determine the at least two audio signal common class segments are within a common event space based on the difference measure.

30 19. The apparatus as claimed in claim 18, further comprising a segment smoother configured to generate a common time line incorporating the at least two audio signals.

20. The apparatus as claimed in claim 19, wherein the segment smoother is configured to align the at least two audio signals when a time difference between two of the at least two audio signals is less than a defined threshold, wherein the time difference is the difference on a common time line incorporating the at least
5 two audio signals for a pair of common class segments from the at least two audio signals.

21. The apparatus as claimed in claims 18 to 20, wherein the classifier is configured to:
10 analyse the at least two audio signals to determine at least one parameter;
and
segment the at least two audio signals into parts where the parts of the at least two audio signals are associated with at least one range of values associated with the at least one parameter.

15

Figure 1

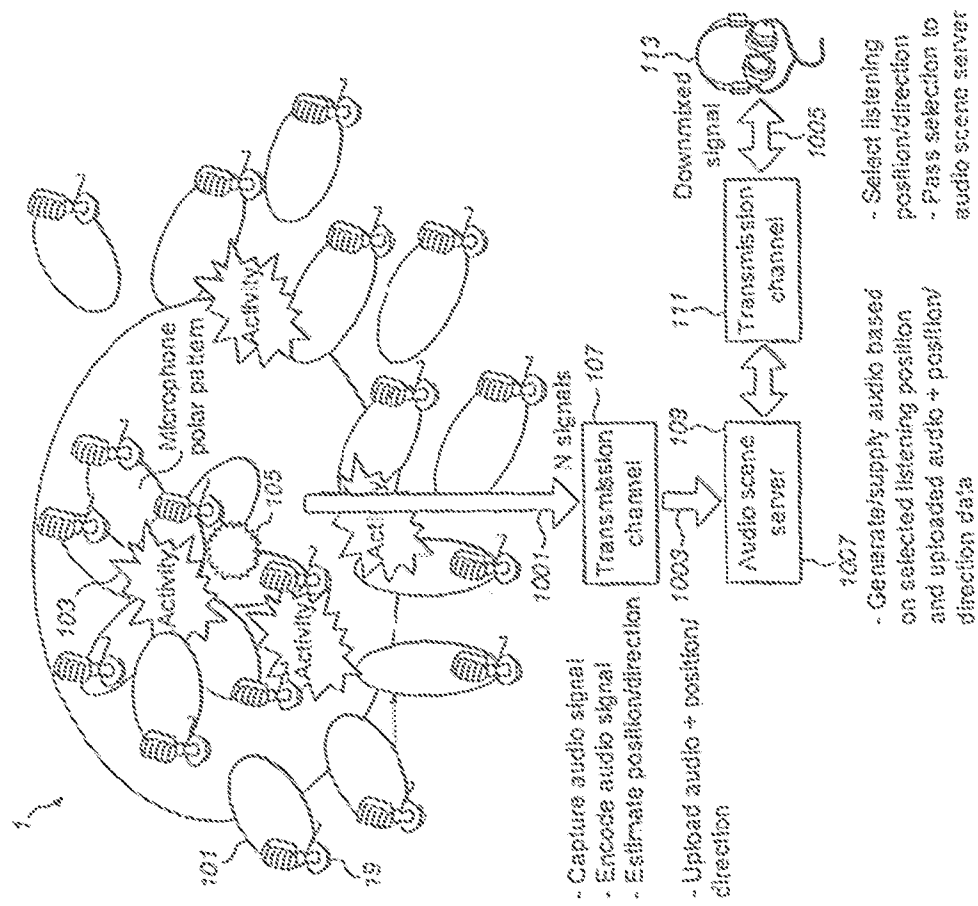


Figure 2

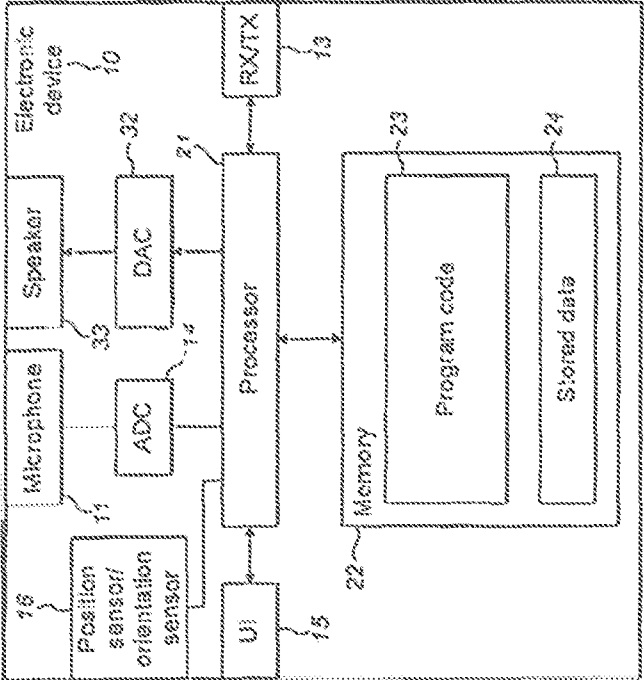
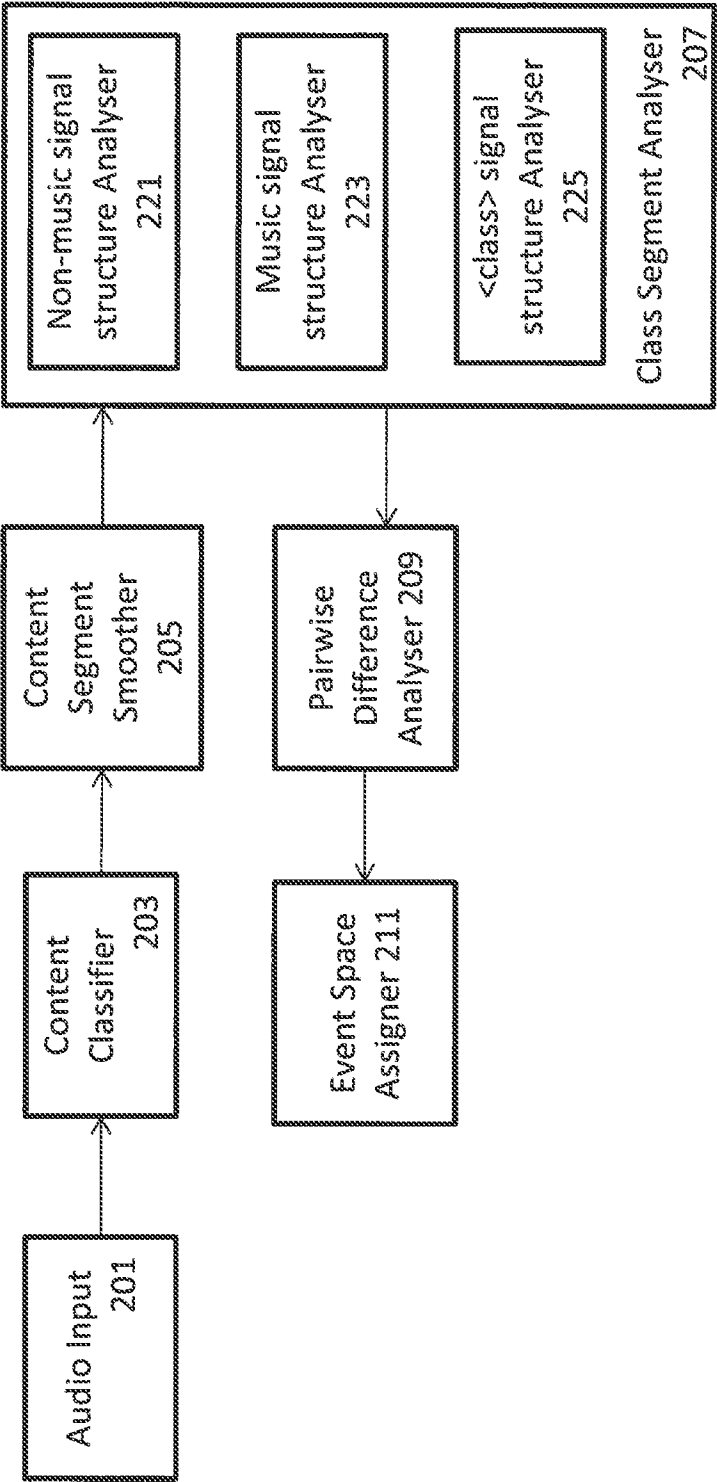


Figure 3



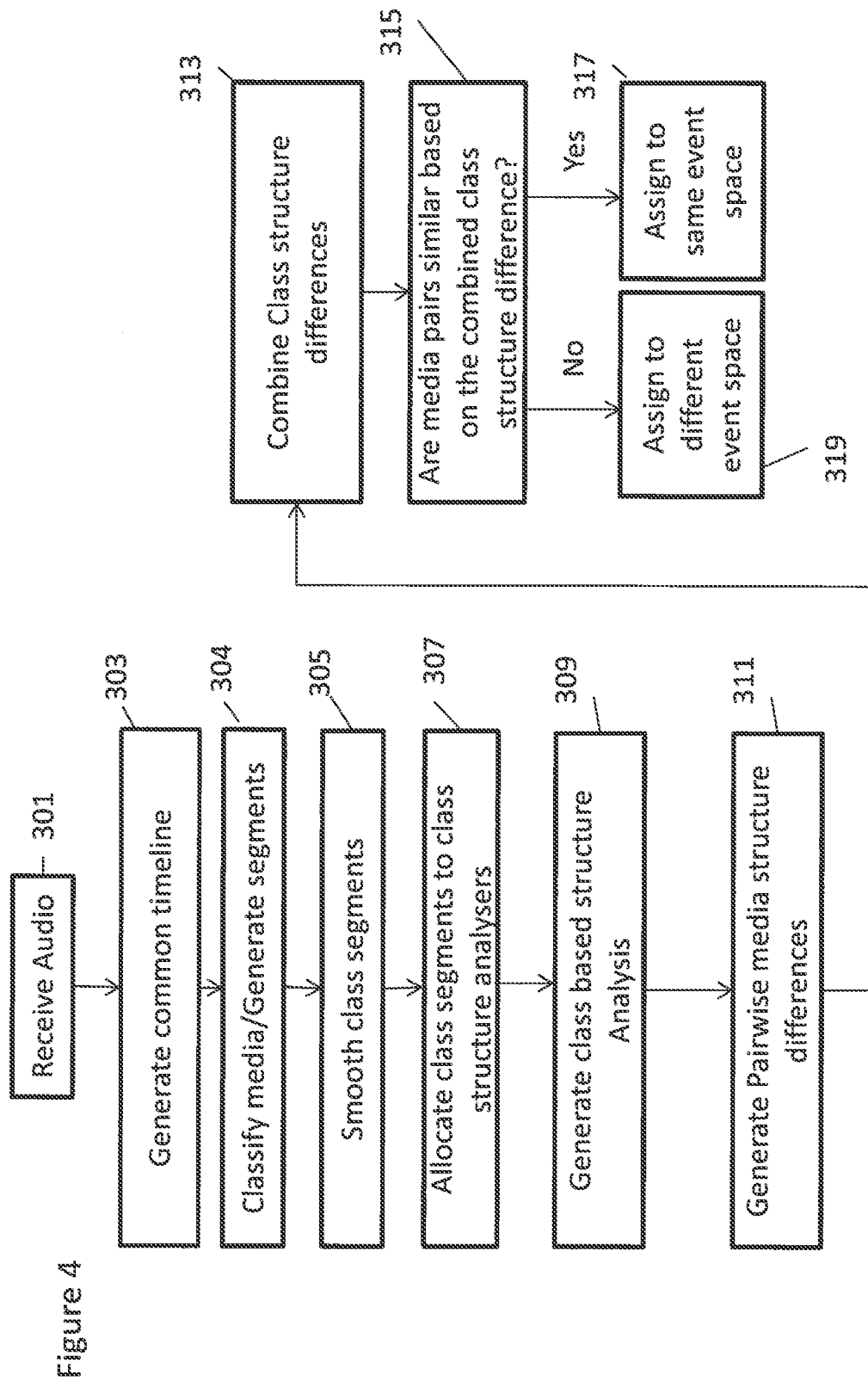


Figure 5

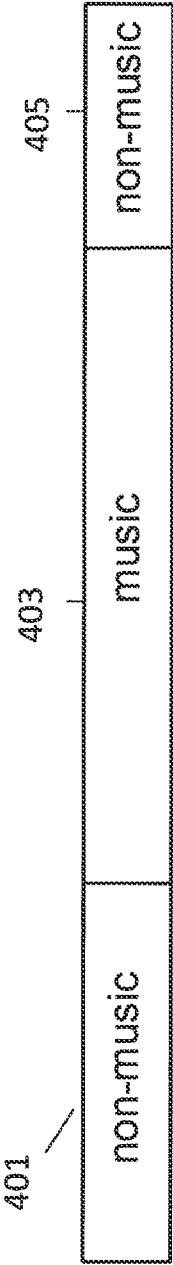


Figure 6

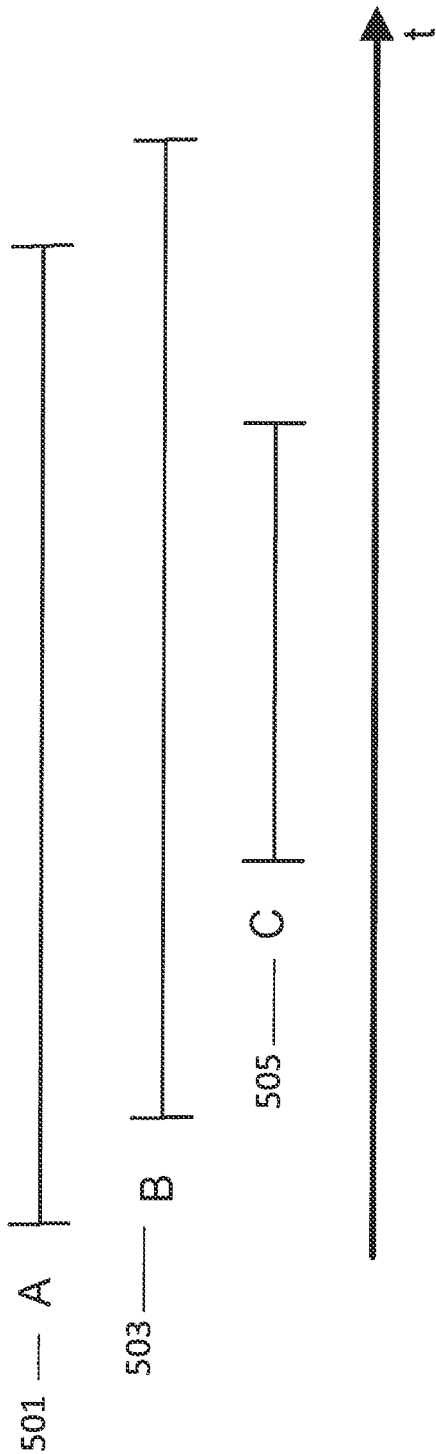


Figure 7

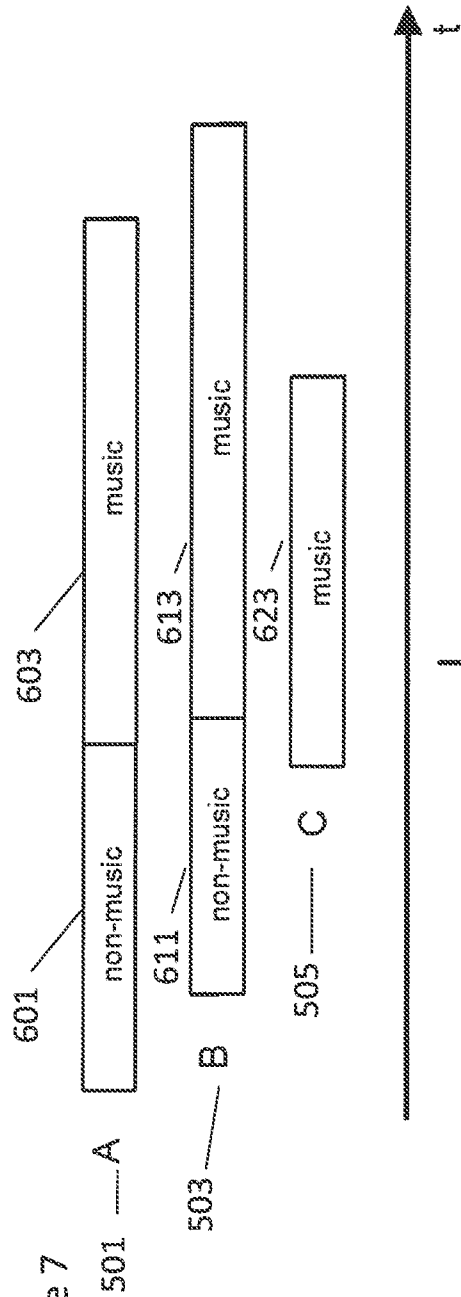


Figure 8

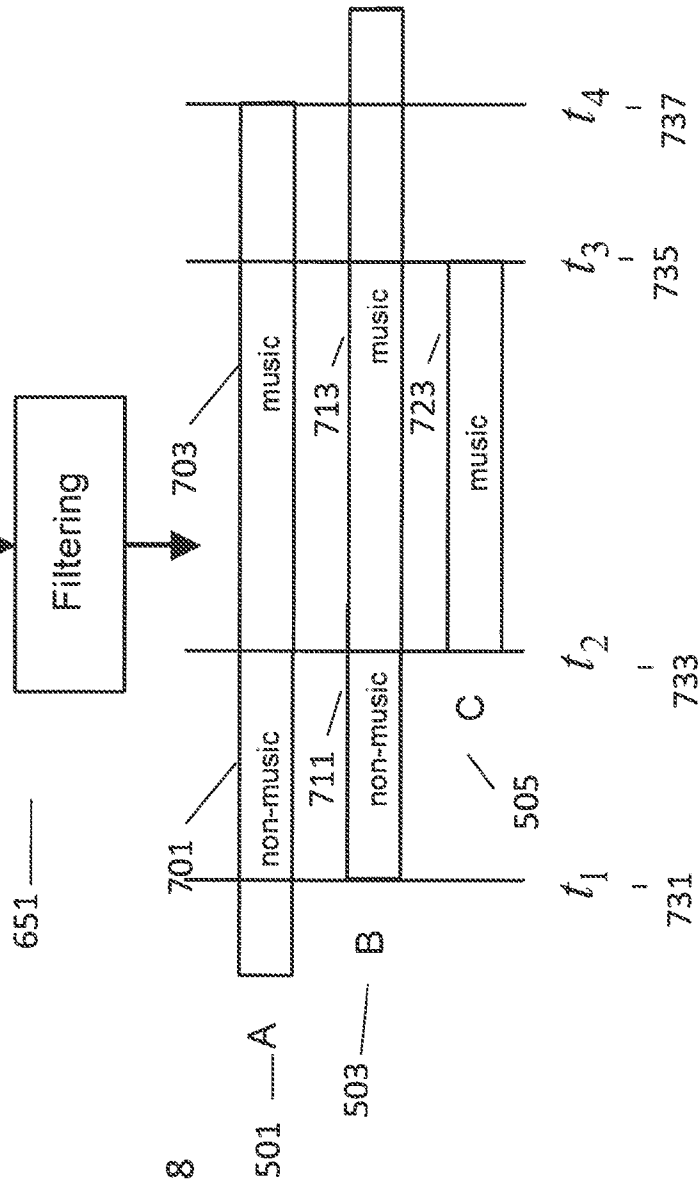
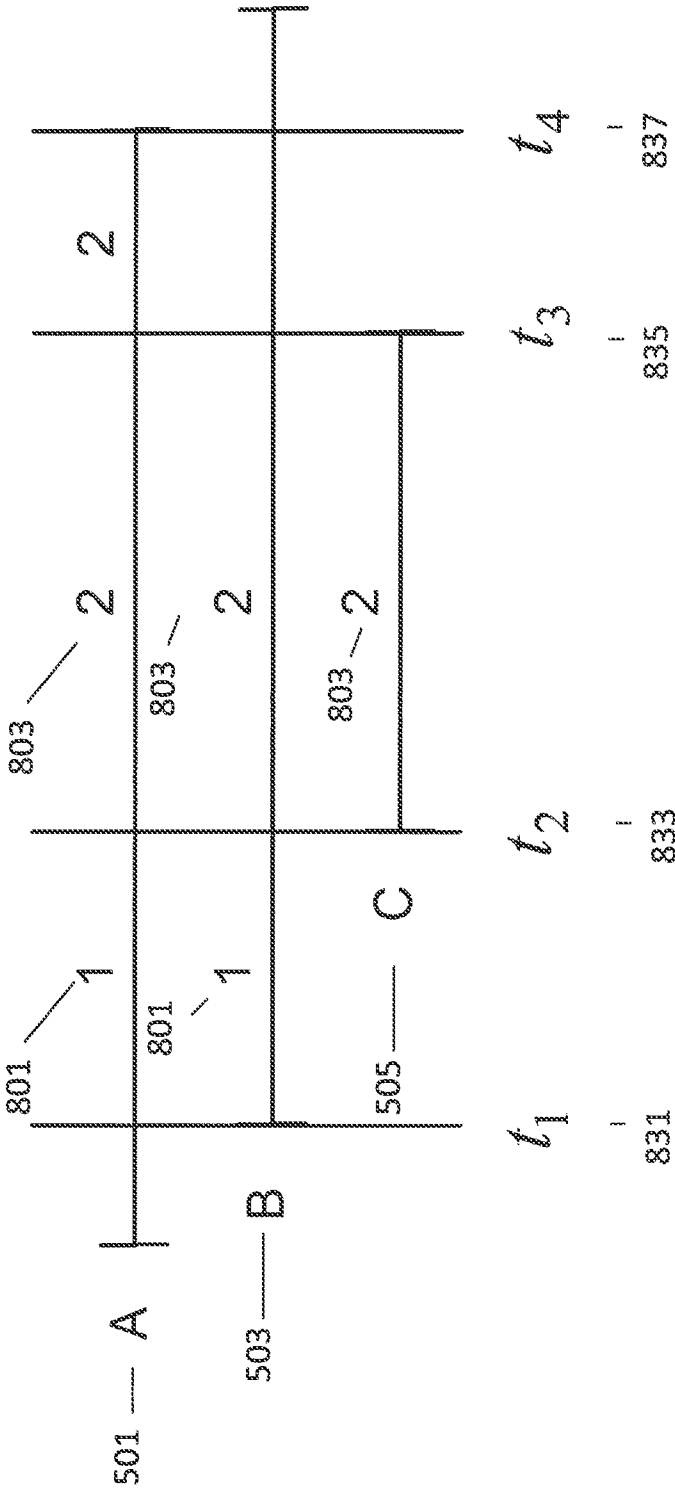


Figure 9



INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2013/054243

A. CLASSIFICATION OF SUBJECT MATTER

IPC: 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, G10L, H04R

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

SE, DK, FI, NO classes as above

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

EPO-Internal, PAJ, WPI data, COMPENDEX, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013043393 A1 (DIGIMARC CORP ET AL), 28 March 2013 (2013-03-28); whole document --	1-21
A	US 20110182432 A1 (ISHIKAWA TOMOKAZU ET AL), 28 July 2011 (2011-07-28); whole document --	1-21
A	US 20050097312 A1 (MIHCAK M K ET AL), 5 May 2005 (2005-05-05); whole document --	1-21
A	US 20130058488 A1 (CHENG BIN ET AL), 7 March 2013 (2013-03-07); whole document --	1-21

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

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

22-04-2014

Date of mailing of the international search report

22-04-2014

Name and mailing address of the ISA/SE

Patent- och registreringsverket

Box 5055

S-102 42 STOCKHOLM

Facsimile No. + 46 8 666 02 86

Authorized officer

Henrik Andersson

Telephone No. + 46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2013/054243

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 20110161074 A1 (PANCE ALEKSANDAR ET AL), 30 June 2011 (2011-06-30); whole document --	1-21
A	WO 2012098425 A1 (NOKIA CORP ET AL), 26 July 2012 (2012-07-26); whole document --	1-21
A	WO 2010037427 A1 (NOKIA CORP ET AL), 8 April 2010 (2010-04-08); whole document --	1-21
A	US 20090190780 A1 (NAGARAJA NAGENDRA ET AL), 30 July 2009 (2009-07-30); whole document --	1-21
A	US 20040210436 A1 (JIANG HAO ET AL), 21 October 2004 (2004-10-21); whole document --	1-21
A	US 20100004926 A1 (NEORAN ITAI ET AL), 7 January 2010 (2010-01-07); whole document -- -----	1-21

Continuation of: second sheet**International Patent Classification (IPC)****G06F 17/30** (2006.01)**G10L 19/008** (2013.01)**G10L 19/20** (2013.01)**G10L 21/0208** (2013.01)**G10L 21/0272** (2013.01)**G10L 25/51** (2013.01)**H04R 3/00** (2006.01)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/IB2013/054243

WO	2013043393 A1	28/03/2013	US	20130150117 A1	13/06/2013
US	20110182432 A1	28/07/2011	CN	102171754 A	31/08/2011
			EP	2461321 A1	06/06/2012
			WO	2011013381 A1	03/02/2011
US	20050097312 A1	05/05/2005	EP	1253525 A3	29/03/2006
			JP	2003005771 A	08/01/2003
			JP	4902565 B2	21/03/2012
			JP	2008191675 A	21/08/2008
			US	6971013 B2	29/11/2005
			US	7152163 B2	19/12/2006
			US	6973574 B2	06/12/2005
			US	7707425 B2	27/04/2010
			US	7657752 B2	02/02/2010
			US	7240210 B2	03/07/2007
			US	7188065 B2	06/03/2007
			US	20050084103 A1	21/04/2005
			US	20050076229 A1	07/04/2005
			US	20050071377 A1	31/03/2005
			US	20050066177 A1	24/03/2005
			US	20050066176 A1	24/03/2005
			US	20050065974 A1	24/03/2005
			US	20020184505 A1	05/12/2002
US	20130058488 A1	07/03/2013	CN	102982804 A	20/03/2013
			EP	2579256 A1	10/04/2013
US	20110161074 A1	30/06/2011	US	8560309 B2	15/10/2013
WO	2012098425 A1	26/07/2012	EP	2666160 A1	27/11/2013
			US	20130297053 A1	07/11/2013
WO	2010037427 A1	08/04/2010	EP	2345026 A1	20/07/2011
			US	20110206223 A1	25/08/2011

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/IB2013/054243

US	20090190780 A1	30/07/2009	CN	101903947 A	01/12/2010
			CN	101896971 A	24/11/2010
			CN	101896970 A	24/11/2010
			CN	101896969 A	24/11/2010
			CN	101896964 A	24/11/2010
			EP	2245619 A1	03/11/2010
			EP	2245624 A1	03/11/2010
			EP	2245623 A1	03/11/2010
			EP	2245626 A1	03/11/2010
			EP	2245625 A1	03/11/2010
			JP	2011511962 A	14/04/2011
			JP	2011511961 A	14/04/2011
			JP	2011516901 A	26/05/2011
			JP	2011512550 A	21/04/2011
			JP	2011512549 A	21/04/2011
			KR	20100113144 A	20/10/2010
			KR	20100129283 A	08/12/2010
			KR	20100125272 A	30/11/2010
			KR	20100125271 A	30/11/2010
			KR	20100113145 A	20/10/2010
			TW	200947422 A	16/11/2009
			TW	200933610 A	01/08/2009
			TW	200933609 A	01/08/2009
			TW	200933608 A	01/08/2009
			TW	200947423 A	16/11/2009
			US	20090192791 A1	30/07/2009
			US	20090192790 A1	30/07/2009
			US	20090192803 A1	30/07/2009
			US	20090192802 A1	30/07/2009
			US	8600740 B2	03/12/2013
			US	8560307 B2	15/10/2013
			US	8554551 B2	08/10/2013
			US	8554550 B2	08/10/2013
			US	8483854 B2	09/07/2013
			WO	2009097023 A1	06/08/2009
			WO	2009097022 A1	06/08/2009
			WO	2009097021 A1	06/08/2009
			WO	2009097020 A1	06/08/2009
			WO	2009097019 A1	06/08/2009

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/IB2013/054243

US	20040210436 A1	21/10/2004	US	20060178877 A1	10/08/2006
			US	20060136211 A1	22/06/2006
			US	20050075863 A1	07/04/2005
			US	20050060152 A1	17/03/2005
			US	7035793 B2	25/04/2006
			US	6901362 B1	31/05/2005
			US	7328149 B2	05/02/2008
			US	7249015 B2	24/07/2007
			US	7080008 B2	18/07/2006
US	20100004926 A1	07/01/2010	US	8428949 B2	23/04/2013
			WO	2010001393 A1	07/01/2010