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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2018 (20.12.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/229497 A1

(51) International Patent Classification:
G10L 19/018 (2013.01)

(21) International Application Number:
PCT/GB2018/051645

(22) International Filing Date:
14 June 2018 (14.06.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1709583.7 15 June 2017 (15.06.2017) GB

(71) Applicant: **ASIO LTD** [GB/GB]; Wework The Cursitor, 38 Chancery Lane, London WC2A 1EN (GB).

(72) Inventors: **NESFIELD, James**; 40 Wilmot Street, London E2 0BS (GB). **JONES, Daniel**; 19 Firs Close, London SE23 1BB (GB).

(74) Agent: **RATIONAL IP LIMITED**; 81 Rivington Street, London Greater London EC2A 3AY (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, 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).

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

(54) Title: A METHOD AND SYSTEM FOR TRIGGERING EVENTS

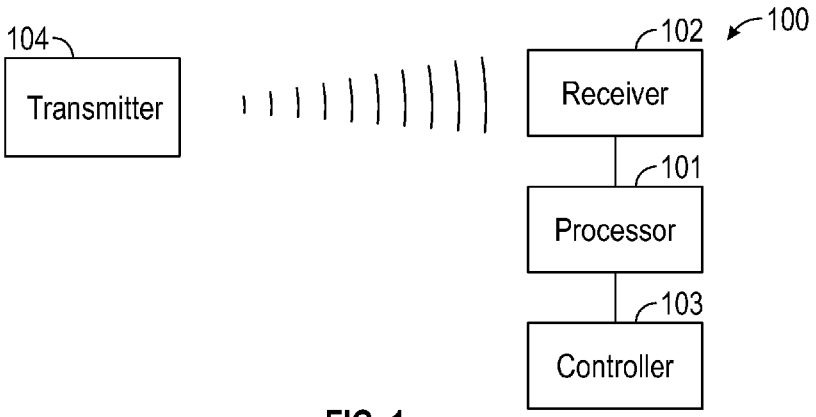

FIG. 1

(57) Abstract: The present invention relates to a method of triggering an event. The method includes receiving a signal stream, detecting a trigger point within the signal stream using a fingerprint associated with the trigger point and triggering an event associated with the detected trigger point.

WO 2018/229497 A1

A Method and System for Triggering Events

Field of Invention

The present invention is in the field of signal processing. More particularly, but not exclusively, the present invention relates to processing signals to trigger events.

Background

Signals, such as audio signals, can be processed to analyse various qualities of the signal.

For example, for streaming audio signals, Shazam technology analyses the audio signal to form a fingerprint of the audio signal. This fingerprint is then compared to a database of audio fingerprints to identify which music track the audio signal originates from.

The Shazam technology is optimised for hardware with sufficient compute to calculate fingerprints of the streaming audio signal and optimised for identifying one music track out of millions.

It would be desirable if a system could be developed which could be used on lower cost hardware to optimally analyse streaming signals to trigger events.

It is an object of the present invention to provide a method and system for triggering events which overcomes the disadvantages of the prior art, or at least provides a useful alternative.

Summary of Invention

According to a first aspect of the invention there is provided a method of triggering an event, including:

a) receiving a signal stream;

b) detecting a trigger point within the signal stream using a fingerprint associated with the trigger point; and

c) triggering an event associated with the detected trigger point;

Other aspects of the invention are described within the claims.

Brief Description of the Drawings

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1: shows a block diagram illustrating a system in accordance with an embodiment of the invention;

Figure 2: shows a flow diagram illustrating a method in accordance with an embodiment of the invention;

Figure 3: shows a flow diagram illustrating a method in accordance with an embodiment of the invention;

Figure 4: shows a block diagram illustrating a system in accordance with an embodiment of the invention;

Figure 5: shows a diagram illustrating a function for determining peak elevation using neighbour reference points in accordance with an embodiment of the invention;

Figure 6: shows a diagram illustrating a 1D slice of a magnitude spectrum buffer in accordance with an embodiment of the invention;

Figure 7: shows a diagram illustrating peaks located using a system in accordance with an embodiment of the invention;

Figure 8: shows a diagram illustrating peaks located using a system in accordance with another embodiment of the invention; and

Figure 9: shows a diagram illustrating examination of a magnitude spectral history buffer using a fingerprint in accordance with an embodiment of the invention.

Detailed Description of Preferred Embodiments

The present invention provides a method and system for triggering events.

The inventors have discovered that fingerprints can be used to analyse a streamed signal, such as an audio signal, directly. This can enable the triggering of events at specific time locations, or trigger points, within the streamed signal.

In Figure 1, a system 100 in accordance with an embodiment of the invention is shown.

One or more processors 101 are shown. The one or more processors 101 may be configured receive a signal stream from a signal receiver 102, to detect a trigger point within the signal stream using a fingerprint associated with the trigger point, and to trigger an event associated with the trigger point. Triggering the event may result in generation of one or more event instructions. The event instructions may control one or more apparatus via one or more controllers 103.

The one or more processors 101 may detect the trigger point from a set of trigger points. The one or more processors 101 may detect the trigger point may comparing one or more of the set of fingerprints associated with the trigger points to the signal stream. The signal stream may be processed for the comparison. In one embodiment, each frame of the signal stream is processed via a Fast Fourier Transform (FFT) and at least some of the frames are compared with at least one fingerprint associated with a trigger point of the set of trigger points.

The system 100 may include a signal receiver 102 configured to receive an analogue signal, such as an audio signal, from a signal transmitter 104 and to provide the signal as a signal stream to the one or more processors 101.

The signal receiver 102 may be an input device such as a microphone, a camera, a radio frequency receiver or any analogue sensor.

The system 100 may include one or more controllers 103. The controllers 103 may be configured for receiving events instructions from the one or more processors 101 to control one or more apparatus. For example, the one or more apparatus may include mechanical apparatus, transmitters, audio output, displays, or data storage.

The system 100 may include a signal transmitter 104. The signal transmitter 104 may correspond to the signal receiver 102 (e.g. a speaker for a microphone), and may be an audio speaker, a light transmission apparatus, a radio frequency transmitter, or any analogue transmitter.

In one embodiment, the audio signals may be in the audible range, inaudible range, or include components in both the audible and inaudible range.

Referring to Figure 2, a method 200 for triggering events in accordance with an embodiment of the invention will be described.

In step 201, a signal stream is received. The signal stream may be received via a signal receiver (e.g. receiver 102). The signal may be audio. The audio signal may be correspond to an audio track. The signal receiver may be a microphone.

The signal stream may be received in real-time via the signal receiver.

The audio signal may be formed of audible, inaudible or a audible/inaudible components.

In step 202, a trigger point is detected within the signal stream using a fingerprint associated with the trigger point. The trigger point may be one of a set of trigger points. A fingerprint for each of plurality of trigger points may be compared against the signal stream to detect a trigger point. A detected trigger point may be the fingerprint which matches the signal stream beyond a predefined or dynamic threshold, the closest match within the set of trigger points, or the closest match within a subset of the set of trigger points.

In one embodiment, the trigger points and associated fingerprints may be created as described in relation to Figure 3.

In one embodiment, the trigger points and associated fingerprints are predefined, and not generated from an existing signal (such as an audio track). The signal (e.g. the audio signal) may be generated for transmission based upon the fingerprint information. For example, a broadcast device may synthesis new audio with notes at the corresponding time/frequency offsets of the fingerprint.

The fingerprint may be formed of a set of peaks within 2D coordinate space of a magnitude spectrum.

Spectral magnitudes may be calculated from the signal stream to form a rolling buffer (with a size T forming a spectral frame within the buffer).

A trigger point may be detected within the signal by iterating over at least some of the peaks within the set of peaks for the fingerprint and examining the corresponding coordinate in the spectral frame in the buffer. A confidence level may be calculated for each peak examination by measuring properties such as the ration between the peak's intensity and the mean intensity of its neighbouring bins. An overall confidence interval may be calculated for the set of peaks by taking, for example, the mean of the individual peak confidences.

In on embodiment, fingerprints for multiple trigger points may be used to examine the spectral frame. In this case, the fingerprint with the highest confidence interval may be identified.

Where the confidence interval for an entire fingerprint exceeds a threshold (and is the highest confidence where multiple fingerprints are used), the trigger point associated with fingerprint is detected within the audio stream.

In step 203, an event associated with the trigger point is triggered when the trigger point is detected.

The event may result in a controller (e.g. controller 103) actuating, for example, play-back of audio corresponding to the trigger point, generation of mechanical movement in time with the audio signal, manifestation of electronic game-play coordinated to audio signal, display of related material synchronised in time to the signal (such as subtitles), generation of any time synchronisation action, or any other action.

Steps 201, 202 and 203 may be performed by one or more processors (e.g. processor(s) 101).

Referring to Figure 3, a method 300 for creating a trigger points for an audio track in accordance with an embodiment of the invention will be described.

In step 301, one or more trigger points are defined for the audio track at time locations within the audio track. Each trigger point may be associated with a timing offset from the start of the audio track.

In step 302, an associated fingerprint is generated at the time location for each trigger point.

The associated fingerprint may be generated from peaks identified within a FFT of the audio track at the time location. The peaks may be local magnitude maxima of the 2D coordinate space created by each FFT block and FFT bin.

In step 303, an event is associated with each trigger point.

The one or more trigger points with each associated fingerprint and event may then be used by the method described in relation to Figure 2.

Referring to Figures 4 to 9, a method of creating and detecting trigger points will be described in accordance with an embodiment of the invention.

This embodiment of the invention performs real-time recognition of pre-determined audio segments, triggering an event when a segment is recognised in the incoming audio stream. The objective is to be able to respond in real-time, with the following key properties:

- minimal latency (less than 50ms)

- high reliability (99% recognition rate in typical acoustic environments over a distance of 1m)
- low false-positive rate (events should rarely be triggered at the wrong moment, if ever)

To perform audio triggering, two phases are involved.

1. Audio fingerprinting (offline, non real-time): An input media file, plus an index of `FINGERPRINT_COUNT` unique trigger timestamps (all within the duration of the media file), are used to generate `FINGERPRINT_COUNT` audio "fingerprints". Each fingerprint characterises the `FINGERPRINT_WIDTH` frames of audio leading up to its corresponding timestamp in the trigger index, where `FINGERPRINT_WIDTH` is the fixed duration of a fingerprint.
2. **Audio recognition** (online, real-time): The set of fingerprints produced by Phase 1 are fed into a separate audio recognition system. This system listens to a live audio stream and attempts to recognise fingerprints from its database within the stream. When a fingerprint is recognised, the corresponding trigger is generated.

Spectral Peaks

Both phases utilise the concept of a "spectral peak". Given a 2D segment of an acoustic spectrum over some period of time, a spectral peak is a 2D point in space (where the X-axis is the time, delineated in FFT frames, and the Y-axis is frequency, delineated in FFT bins) which has some degree of "elevation" over the surrounding peaks.

The elevation of a peak is the difference between its decibel magnitude and the mean magnitude of some selection of peaks around it. A peak with a greater elevation is substantially louder than the spectrogram cells around it, meaning it is perceptually prominent. A short, sharp burst of a narrow

frequency band (for example, striking a glockenspiel) would result in a peak of high elevation.

Peaks may be used to characterise audio segments because they have fixed linear relationships in the time and frequency domains, and because they may be robust to background noise. When a given audio recording is played back over a reasonable-quality speaker, the resultant output will typically demonstrate peaks of roughly similar elevation. Even when background noise is present, peaks at the original locations will still mostly be evident.

This recognition system may require that peaks be identically distributed in the original fingerprint and the audio stream to recognise effectively. Therefore, it assumes that the audio will not be altered before playback. For example, if it played at a lower speed or pitch, the recognition may fail. However, this system should remain robust to processes such as distortion, filtering effects from acoustic transducers, and degradation from compression codecs, all of which do not heavily affect the linear relationships between peaks in time and frequency.

In this algorithm, `NEIGHBOURHOOD_WIDTH` and `NEIGHBOURHOOD_HEIGHT` are used define the minimum spacing between peaks. `NEIGHBOURHOOD_WIDTH` is measured in FFT frames, and `NEIGHBOURHOOD_HEIGHT` in FFT bins.

Different algorithms used to determine and match peaks for the fingerprinting and recognition phases. In the fingerprinting phase, the algorithm may be optimised for precision, to ensure that the best-possible peaks are selected. In the recognition phase, the algorithm may be optimised for speed and efficiency.

An overview flow chart of the fingerprinter and recogniser ("scanner") is shown in Figure 4.

Peak Elevation Function

The key characteristic of a peak is its elevation. In a time-unlimited system, this would typically be determined by taking the mean magnitude of all the cells in the peak's neighbourhood, and calculating the ratio of the peak's magnitude vs the mean surrounding magnitude.

However, this may require a lot of calculations (up to `(NEIGHBOURHOOD_WIDTH * 2 + 1) x (NEIGHBOURHOOD_HEIGHT * 2 + 1) - 1`: 120 calculations for width and height of 5). In one embodiment, to enhance efficiency, more economical neighbourhood functions to determine whether a peak's elevation can be determined with fewer reference points have been discovered to be effective.

One such current elevation function makes use of 7 reference points around the peak: in the 4 corners of its Moore neighbourhood, and the top, bottom and left edges. The right edge is omitted as this may be susceptible to be artificially amplified by note tails and reverberation from the peak's acoustic energy. This is shown in Figure 5.

Phase 1: Audio Fingerprinting

The audio fingerprinting phase is designed to identify and prioritise the peaks that uniquely and reliably characterise each segment of audio. It is typically performed offline prior to any recognition process, meaning it may not need to be designed for efficiency.

The fingerprinting phase proceeds as follows:

- The fingerprinter is given a media file (for example, a mono uncompressed audio file) and an index of times at which trigger events are to occur.
- The fingerprinter opens the media file and proceeds to read the contents in FFT-sized buffers (typically 256 frames). Each buffer, an FFT is performed and the magnitude spectrum derived. This magnitude spectrum is added to a rolling buffer.
- Each frame, the last `NEIGHBOURHOOD_WIDTH * 2 + 1` frames of the rolling buffer are inspected to find peaks. This is done by iterating over each cell in the 1D line (shown in Figure 6) running down the centre of the buffer, and checking if it is a local maxima in the Moore neighbourhood of size `NEIGHBOURHOOD_WIDTH x NEIGHBOURHOOD_HEIGHT`. If it is the maxima, the elevation function is applied to determine the peak-candidate's elevation versus nearby cells (see "Peak Elevation Function" above). If the peak-candidate's elevation exceeds a threshold in decibels (`FINGERPRINT_PEAK_THRESHOLD`, typically around 12dB), it is classified as a peak and added to another rolling buffer of width `FINGERPRINT_WIDTH` as shown in Figure 7.
- Note that the 1D-slice part of the above step may be performed for efficiency. An alternative approach would be to store a rolling spectrogram buffer of width `FINGERPRINT_WIDTH`, and do a brute-force search of the buffer for peaks as shown in Figure 8.
- When the fingerprinter reaches a frame whose timestamp corresponds to a trigger timestamp, it collects the peaks for the previous `FINGERPRINT_WIDTH` spectral frames. These peaks are ordered by their elevation (descending), and any peaks beyond the first `MAX_PEAK_COUNT` peaks are rejected. This ensures that each fingerprint has a maximum number of peaks. The prioritisation is imposed because peaks with a higher elevation are proportionately more likely survive acoustic playback and be detected by the decoder. This allows "weaker" peaks to be omitted, improving efficiency later.

- When the entire set of fingerprints have been collected, they are serialised to disk as a "fingerprint set". A fingerprint set is characterised by:
    - the number of fingerprints it contains;
    - for each fingerprint, its width, its ID number, the number of peaks it contains, and each peak's X/Y coordinate, with each peak ordered by magnitude (descending).

Phase 2: Audio Recognition

The audio recogniser takes a set of fingerprints and a stream of audio (usually real-time), and attempts to recognise fingerprints within the stream.

It functions as follows:

- Audio is read in FFT block-sized chunks. An FFT is performed to obtain the magnitude spectrum, which is appended to a rolling buffer of width `FINGERPRINT_WIDTH`. This acts as an acoustic spectrum history, containing the most recent frames of precisely the same width as a fingerprint.
- Each frame, the scanner iterates over every available fingerprint. Each peak's (x, y) coordinates are examined within the spectral history buffer, and its elevation calculated according to the same elevation function as the fingerprinter (e.g., for the EDGES elevation function, comparing its value to the mean of 7 points around it). If the decibel difference between the peak's magnitude and the surrounding background magnitude exceeds a fixed `CANDIDATE_PEAK_THRESHOLD`, the peak is classified as a match. The `CANDIDATE_PEAK_THRESHOLD` is measured in decibels, similar to the `FINGERPRINT_PEAK_THRESHOLD` used to select peaks for the fingerprint. However, the `CANDIDATE_PEAK_THRESHOLD` is substantially lower, as

background noise and filtering may reduce a peak's prominence in real-world playback. Figure 9 shows examining of a fingerprint formed from the peaks of Figure 7 against the spectral history buffer. Peak matches are shown at 901, 902, 903, and 904, and a peak non-match shown at 905.

- After all the peaks have been classified, a confidence level is determined for the fingerprint as a whole by calculating `PEAKS_MATCHED / TOTAL_NUM_PEAKS` (i.e., the proportion of peaks matched). If this exceeds a fixed `CANDIDATE_CONFIDENCE_THRESHOLD` (usually around 0.7), the fingerprint is classified as matching.
- Matching then continues in case another fingerprint is matched at a higher confidence.
- `WAIT_FOR_PEAK` operation: The fingerprint with the highest confidence is selected as a match. Matches can sometimes occur one or two frames earlier than expected (with confidence levels rising quickly to a peak and then dropping again). To ensure a match isn't triggered too soon, the match is recorded. The following frame, if the confidence level drops, the match is triggered. If the confidence level rises, the match is recorded again to be checked the following frame. Matches are thus always triggered with one frame of delay, _after_ the peak value.
- Once a fingerprint match has been triggered, a trigger can not then occur again for a fixed number of frames to ensure accidental re-triggers. Optionally, the specific trigger ID can be disabled automatically for some period.

Audio Recognition: Additional Optimisations

Additional optimisations may be used to further reduce the CPU footprint of audio recognition, with minimal impact on recognition success rate:

Early Rejection

As a given fingerprint's peaks are ordered by magnitude, descending, the first peaks are the most prominent and thus most likely to be successfully recognised in real-world playback.

For efficiency, an algorithm may be used that first inspects a small proportion (e.g. 20%) of a fingerprint's peaks. The mean confidence of these peaks is calculated, and compared to a threshold value that is lower than `CANDIDATE_CONFIDENCE_THRESHOLD` (e.g, `0.8` of the overall candidate threshold). If the mean confidence of this initial sample falls below the minimum threshold, it is unlikely that the fingerprint as a whole will be a match, and the rest of the peaks are not inspected.

If the mean confidence is above the threshold, the remainder of the peaks are inspected as normal, and the fingerprint as a whole either accepted or rejected.

The values of `CHIRP_FINGERPRINT_SCANNER_EARLY_REJECT_PROPORTION` and `CHIRP_FINGERPRINT_SCANNER_EARLY_REJECT_LOWER_THRESHOLD` are selected to minimise the number of peaks that must be inspected on average, whilst minimising the number of actual matches that are missed.

Note that this technique can also function with an unordered set of peaks.

Disable Below Minimal Threshold

As acoustic spectra tend to be correlated over time, is unlikely that a match (with confidence of 0.7 above, corresponding to 70% of the peaks matching up) will arise immediately after a spectral block with a match rate of a low value such as 0.01.

Therefore, a feature may be used to temporarily disable a fingerprint if its confidence is below some minimal threshold, removing it from the match pool for the next `CHIRP_FINGERPRINT_DISABLED_DURATION` frames.

A potential advantage of some embodiments of the present invention is that trigger points within streamed signals can be used to trigger events. This may provide various functionality such as synchronisation of events with the streamed signal. Furthermore, by detecting the trigger point using the fingerprint rather than calculating fingerprints from the streamed signal, computation is reduced enable deployment on lower cost hardware.

While the present invention has been illustrated by the description of the embodiments thereof, and while the embodiments have been described in considerable detail, it is not the intention of the applicant to restrict or in any way limit the scope of the appended claims to such detail. Additional advantages and modifications will readily appear to those skilled in the art. Therefore, the invention in its broader aspects is not limited to the specific details, representative apparatus and method, and illustrative examples shown and described. Accordingly, departures may be made from such details without departure from the spirit or scope of applicant's general inventive concept.

Claims

1. A method of triggering an event, including:
   a) receiving a signal stream;
   b) detecting a trigger point within the signal stream using a fingerprint associated with the trigger point; and
   c) triggering an event associated with the detected trigger point.

2. A method as claimed in claim 1, wherein the signal is an analogue signal.

3. A method as claimed in claim 2, wherein the analogue signal is an audio signal.

4. A method as claimed in claim 3, wherein the audio signal is from an audio track.

5. A method as claimed in any one of claims 3 to 4, wherein the audio signal stream is received via a microphone.

6. A method as claimed in any one of the preceding claims, wherein the trigger point is one from a set of trigger points.

7. A method as claimed in any one of the preceding claims, wherein each fingerprint comprises a plurality of spectral peaks.

8. A method as claimed in claim 7, wherein each fingerprint is constructed using the highest spectral peaks at a predefined time for the trigger point.

9. A method as claimed in any one of the preceding claims when dependent on claim 6, wherein the trigger point is detected by comparing fingerprints for at least some of the trigger points to the signal stream.

10. A method as claimed in claim 9, wherein the detected trigger point is the trigger point with the highest confidence match.

11. A method as claimed in any one of the preceding claims, wherein the signal stream is processed by FFT to produce a magnitude spectrum frame.

12. A method as claimed in claim 11, wherein the trigger point is detected within the magnitude spectrum frame.

13. A method as claimed in any one of the preceding claims, wherein a plurality of trigger points are detected over time within the signal stream.

14. A method as claimed in any one of the preceding claims when dependent on claim 7, wherein, during detecting of the trigger point, a fingerprint for a trigger point is rejected for a match if a defined proportion of the spectral peaks are not detected in the signal stream.

15. A method as claimed in claim 14, wherein the proportion of spectral peaks are the highest spectral peaks.

16. A method as claimed in any one of the preceding claims when dependent on claim 6, wherein, during detecting of the trigger point, a trigger point is not used for matching if the trigger point matched below a defined confidence level for the previous frame.

17. A method for creating trigger points for an audio track, including:

a) defining one or more trigger points at one or more time locations within an audio track;

b) generating a fingerprint for each trigger point based upon the audio track; and

c) associating an event with each trigger point.

18. A method as claimed in any one of claims 1 to 16, wherein the trigger point is created in accordance with the method of claim 17.

19. A system for triggering an event, including:

A signal receiver;

A processor configured for receiving a signal stream from the signal receiver, detecting a trigger point within the signal stream using a fingerprint associated with the trigger point and triggering an event associated with the detected trigger point.

20. A computer program configured for performing the method of any one of claims 1 to 18.

21. A computer readable medium configured for storing the computer program of claim 20.

22. A processor configured for detecting a trigger point within a signal stream using a fingerprint associated with the trigger point and triggering an event associated with the detected trigger point.

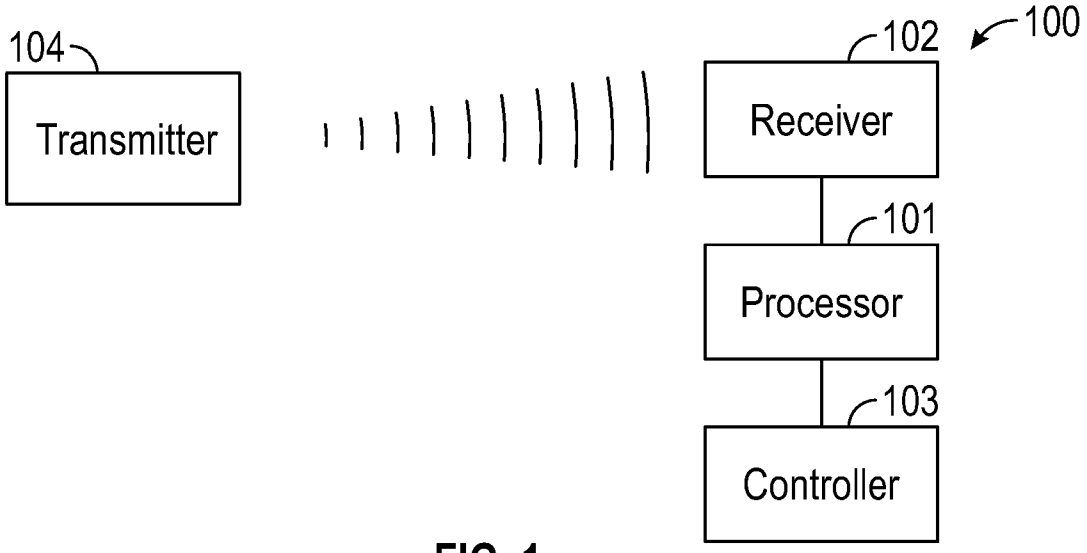

FIG. 1

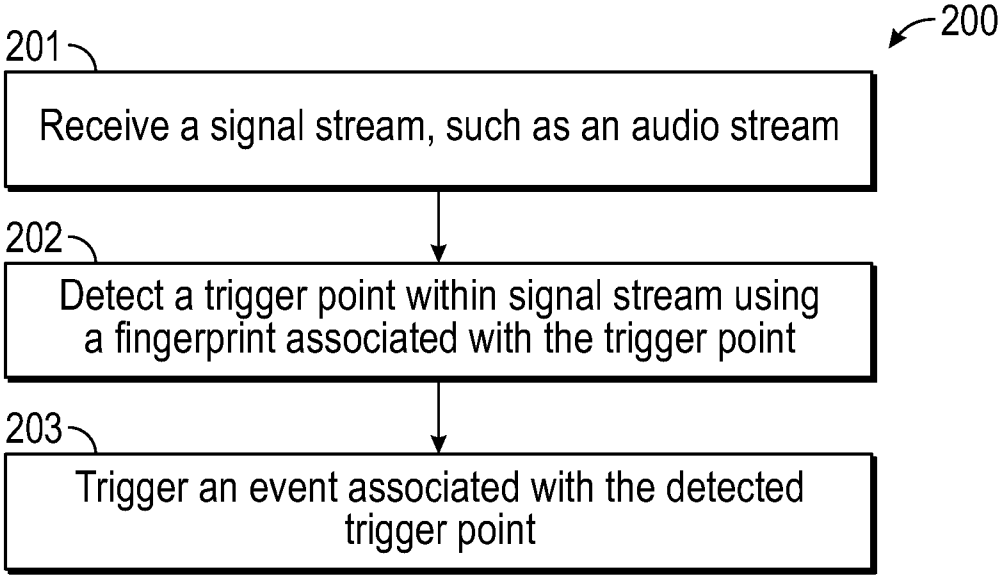

FIG. 2

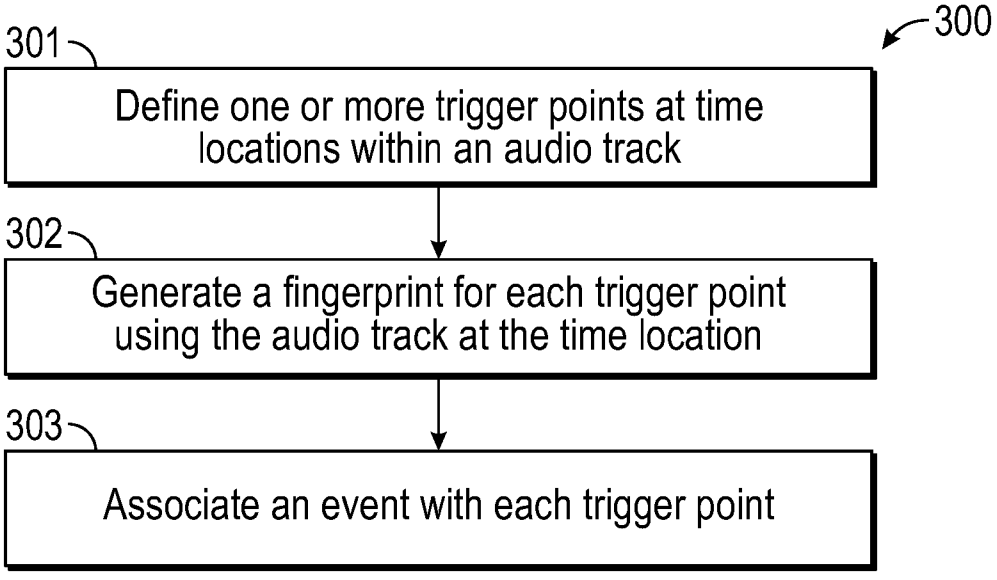

FIG. 3

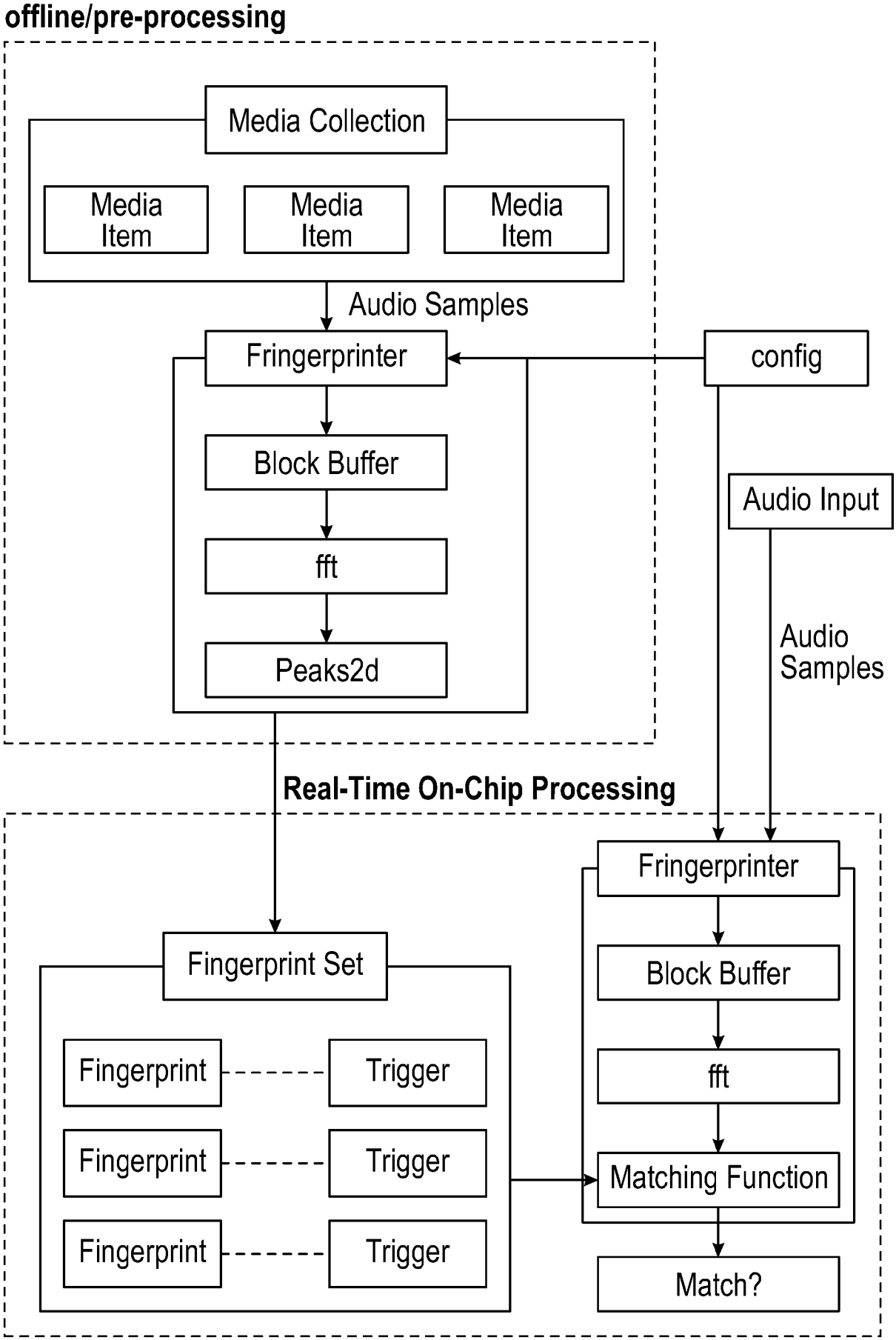
offline/pre-processing
Media Collection
Media Item
Media Item
Media Item
Audio Samples
Fringerprinter
config
Block Buffer
fft
Peaks2d
Audio Input
Audio Samples
Real-Time On-Chip Processing
Fingerprinter
Fingerprint Set
Block Buffer
Fingerprint
Trigger
fft
Fingerprint
Trigger
Matching Function
Fingerprint
Trigger
Match?

FIG. 4

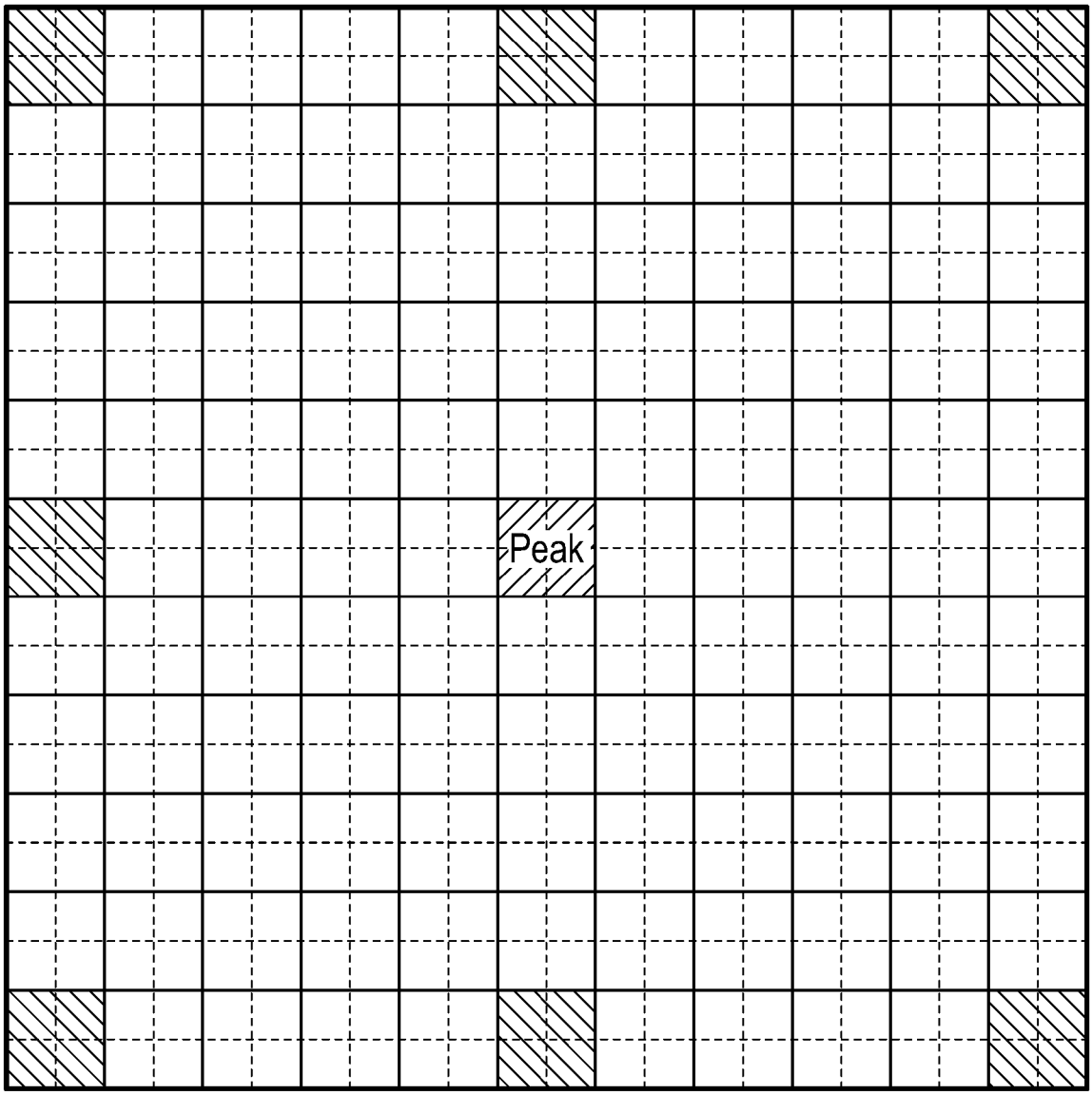

FIG. 5

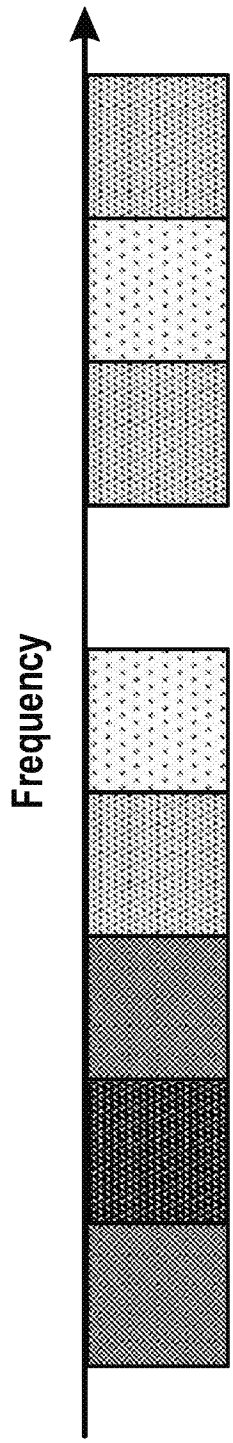

FIG. 6

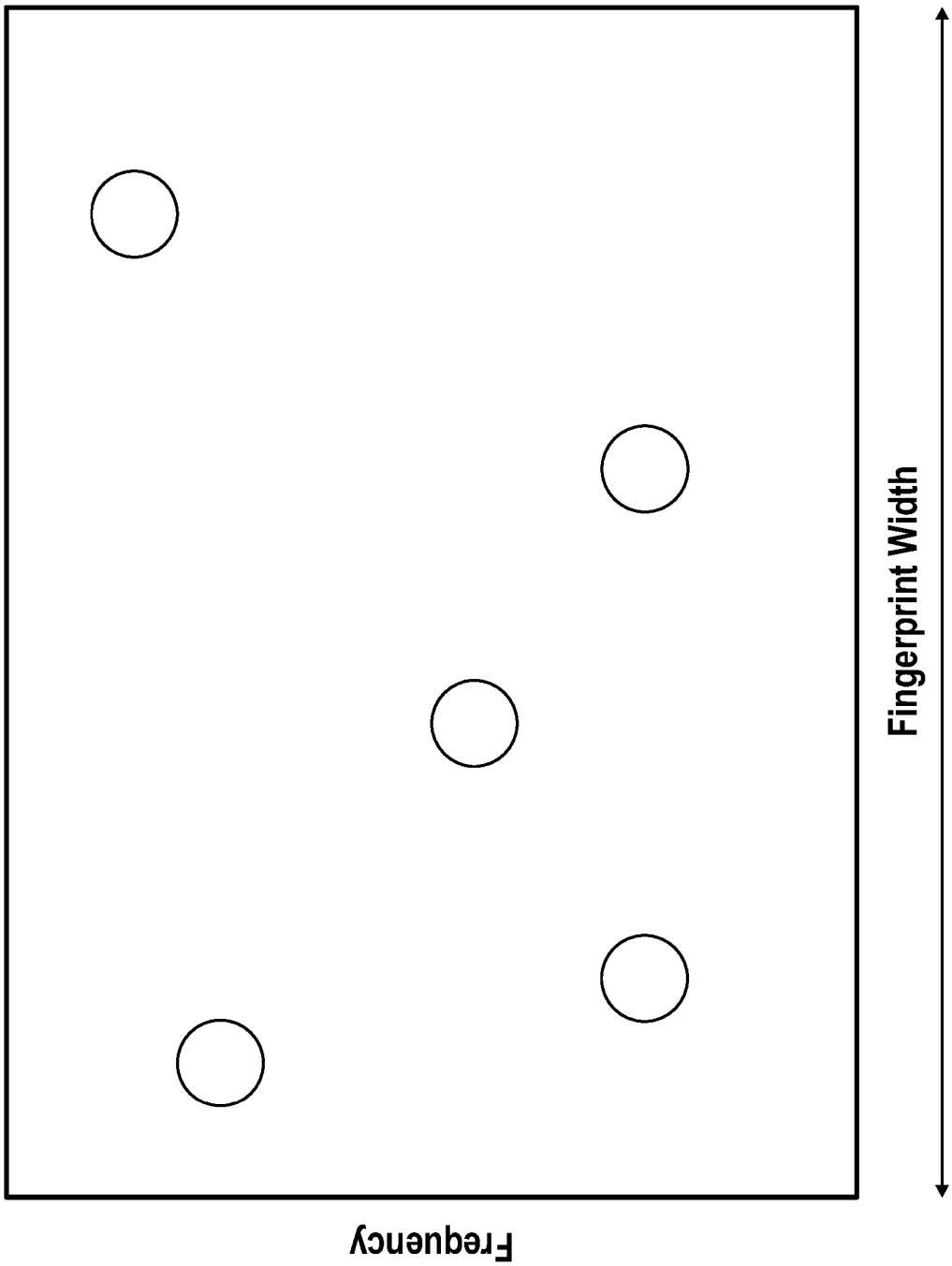

FIG. 7

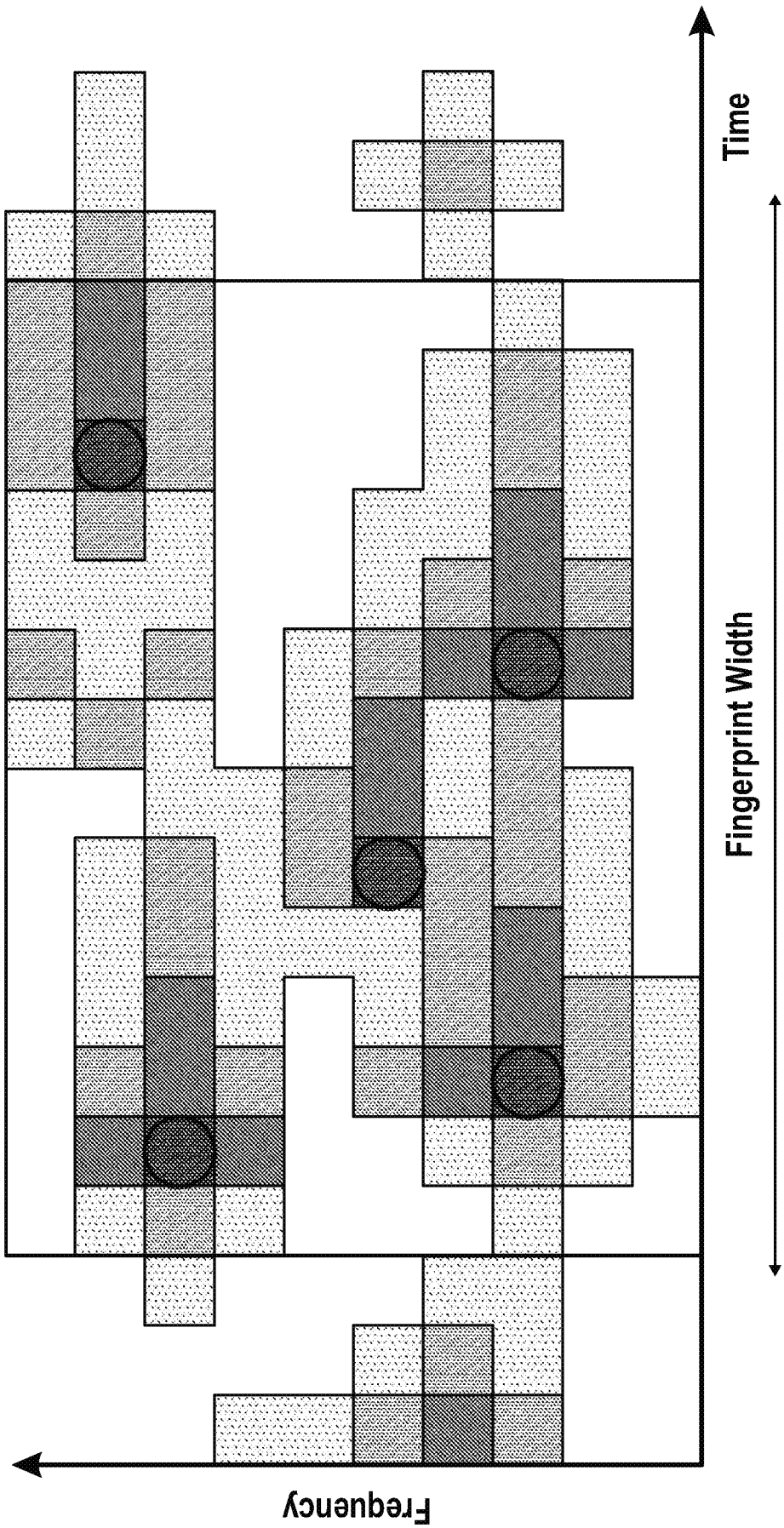

FIG. 8

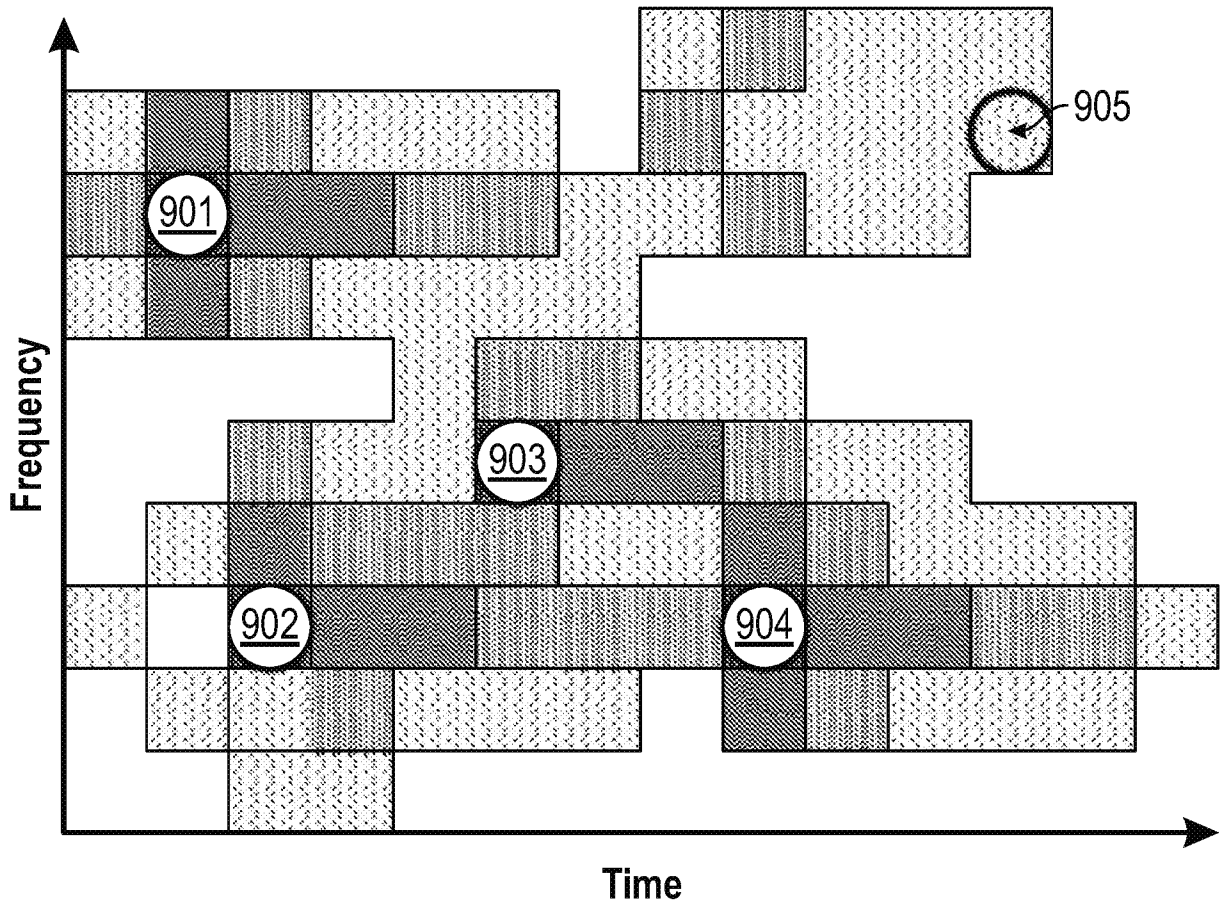

FIG. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2018/051645

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/018
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/61987 A2 (VERANCE CORP [US]; PETROVIC RADE [US]; WINOGRAD JOSEPH M [US]) 23 August 2001 (2001-08-23)	1-6,9, 10,13, 16-22
Y	abstract page 8, lines 5-27 page 9, lines 24-29 page 10, line 24 - page 11, line 16 figures 1,2	7,8,11, 12,14,15

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 2018	04/10/2018

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

2

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2018/051645

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WANG A: "An Industrial-Strength Audio Search Algorithm", PROCEEDINGS OF 4TH INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, BALTIMORE, MARYLAND, USA, 27 October 2003 (2003-10-27), XP002632246, abstract paragraph [[2.1RobustConst.]] - paragraph [[2.2FastCombin.]] paragraph [[2.3.1Significance]] figure 1 -----	7,8,11, 12,14,15
A	BOURGUET MARIE-LUCE ET AL: "A Robust Audio Feature Extraction Algorithm for Music Identification", AES CONVENTION 129; NOVEMBER 2010, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 4 November 2010 (2010-11-04), XP040567151, abstract paragraph [[2.2.ShazamMethod]] figures 2,3 -----	1,17, 19-22
A	EP 1 760 693 A1 (SEET INTERNET VENTURES INC [US]) 7 March 2007 (2007-03-07) abstract paragraphs [[0031]] - [[0034]] figures 1,2 -----	1,17, 19-22

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2018/051645

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 0161987	A2	23-08-2001	AU	4371601	A	27-08-2001
			US	6737957	B1	18-05-2004
			US	2004169581	A1	02-09-2004
			US	2007247278	A1	25-10-2007
			US	2011068898	A1	24-03-2011
			US	2012130719	A1	24-05-2012
			US	2013339029	A1	19-12-2013
			US	2015016228	A1	15-01-2015
			WO	0161987	A2	23-08-2001
EP 1760693	A1	07-03-2007	AT	403212	T	15-08-2008
			EP	1760693	A1	07-03-2007
			ES	2309924	T3	16-12-2008
			JP	5362178	B2	11-12-2013
			JP	2007065659	A	15-03-2007
			US	2007055500	A1	08-03-2007
			US	2009157391	A1	18-06-2009
			US	2013197913	A1	01-08-2013