US 20240038250A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2024/0038250 A1**

NESFIELD et al. (43) **Pub. Date: Feb. 1, 2024**

(54) **METHOD AND SYSTEM FOR TRIGGERING EVENTS**

(71) Applicant: **SONOS EXPERIENCE LIMITED**, Hayes (GB)

(72) Inventors: **James Andrew NESFIELD**, London (GB); **Daniel Jones**, London (GB)

(21) Appl. No.: **18/144,589**

(22) Filed: **May 8, 2023**

Related U.S. Application Data

(63) Continuation of application No. 16/623,160, filed on Dec. 16, 2019, now Pat. No. 11,682,405, filed as application No. PCT/GB2018/051645 on Jun. 14, 2018.

(30) **Foreign Application Priority Data**

May 16, 2017 (GB) 1709583.7

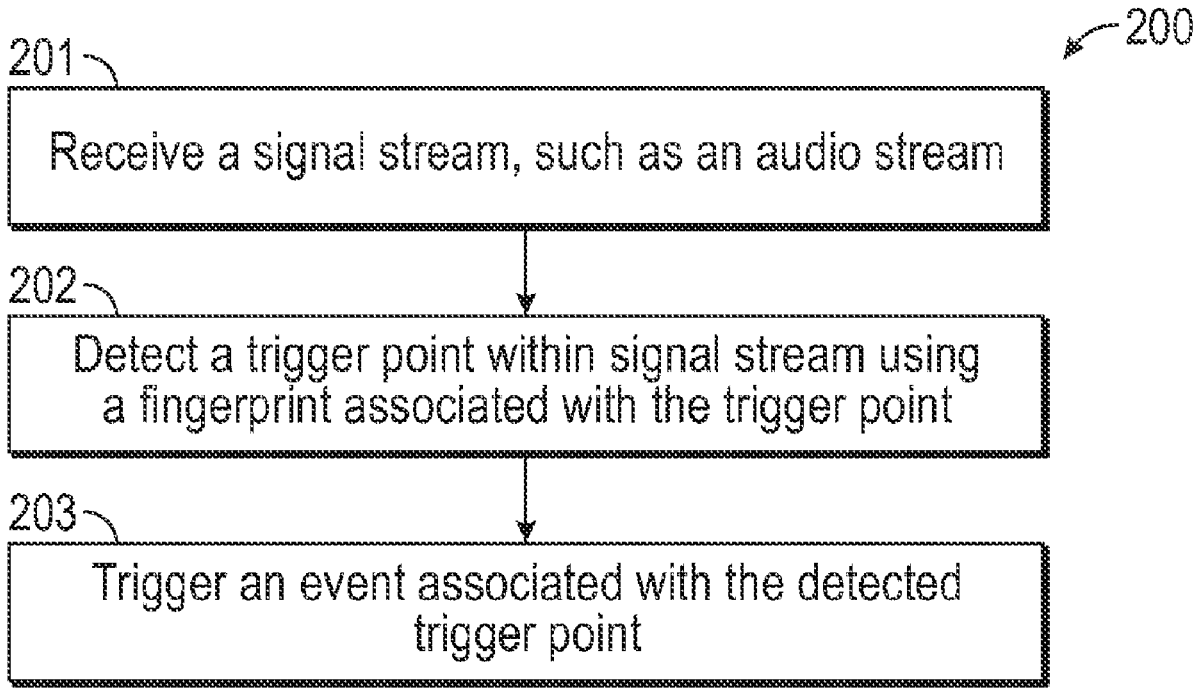
Publication Classification

(51) **Int. Cl.**
G10L 19/02 (2006.01)

(52) **U.S. Cl.**
CPC ***G10L 19/02*** (2013.01)

(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.



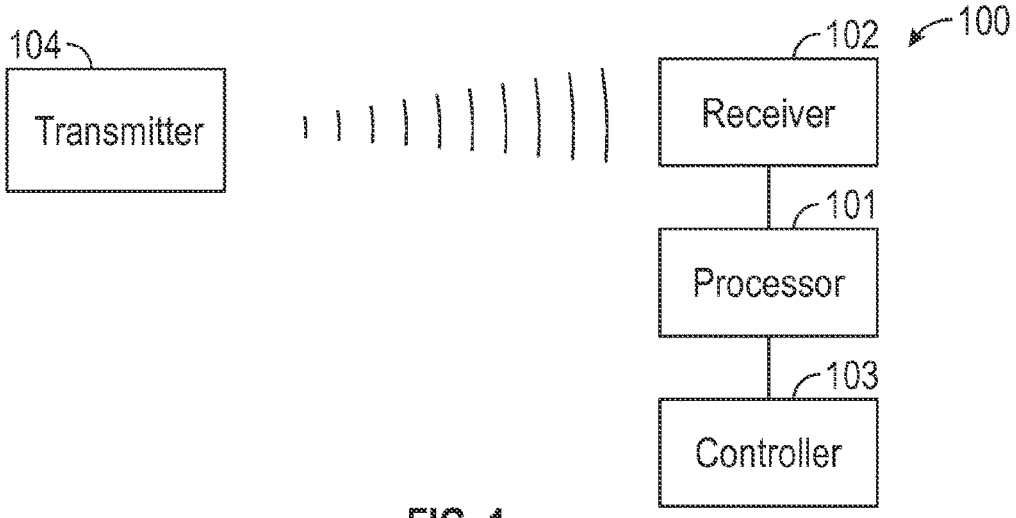


FIG. 1

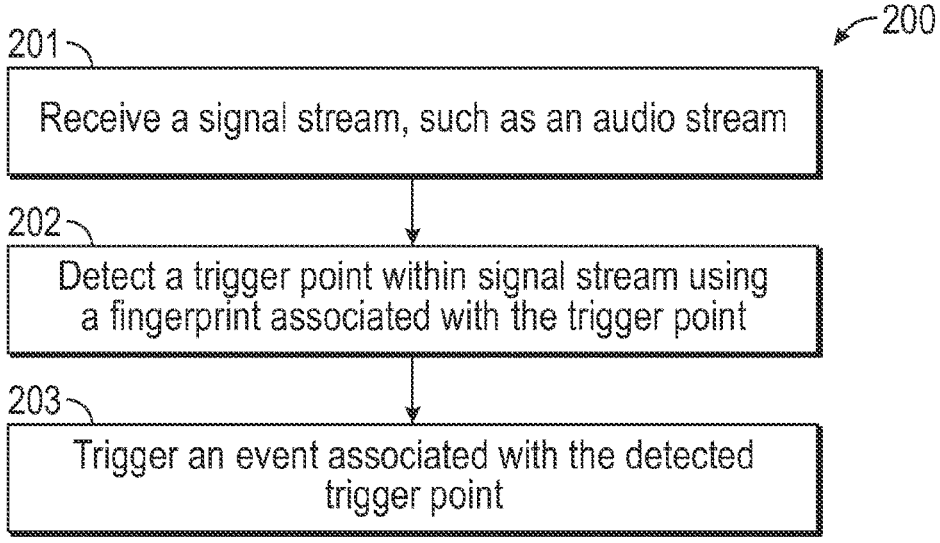


FIG. 2

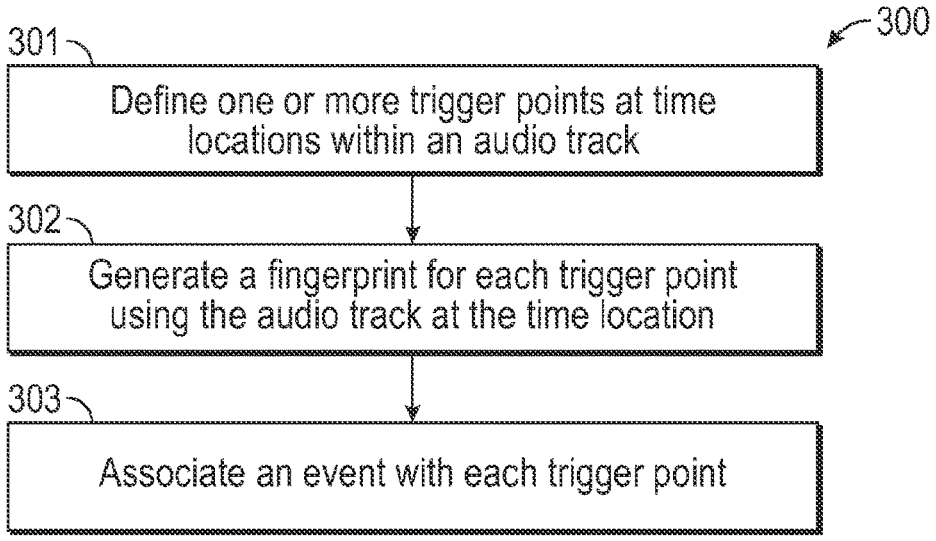


FIG. 3

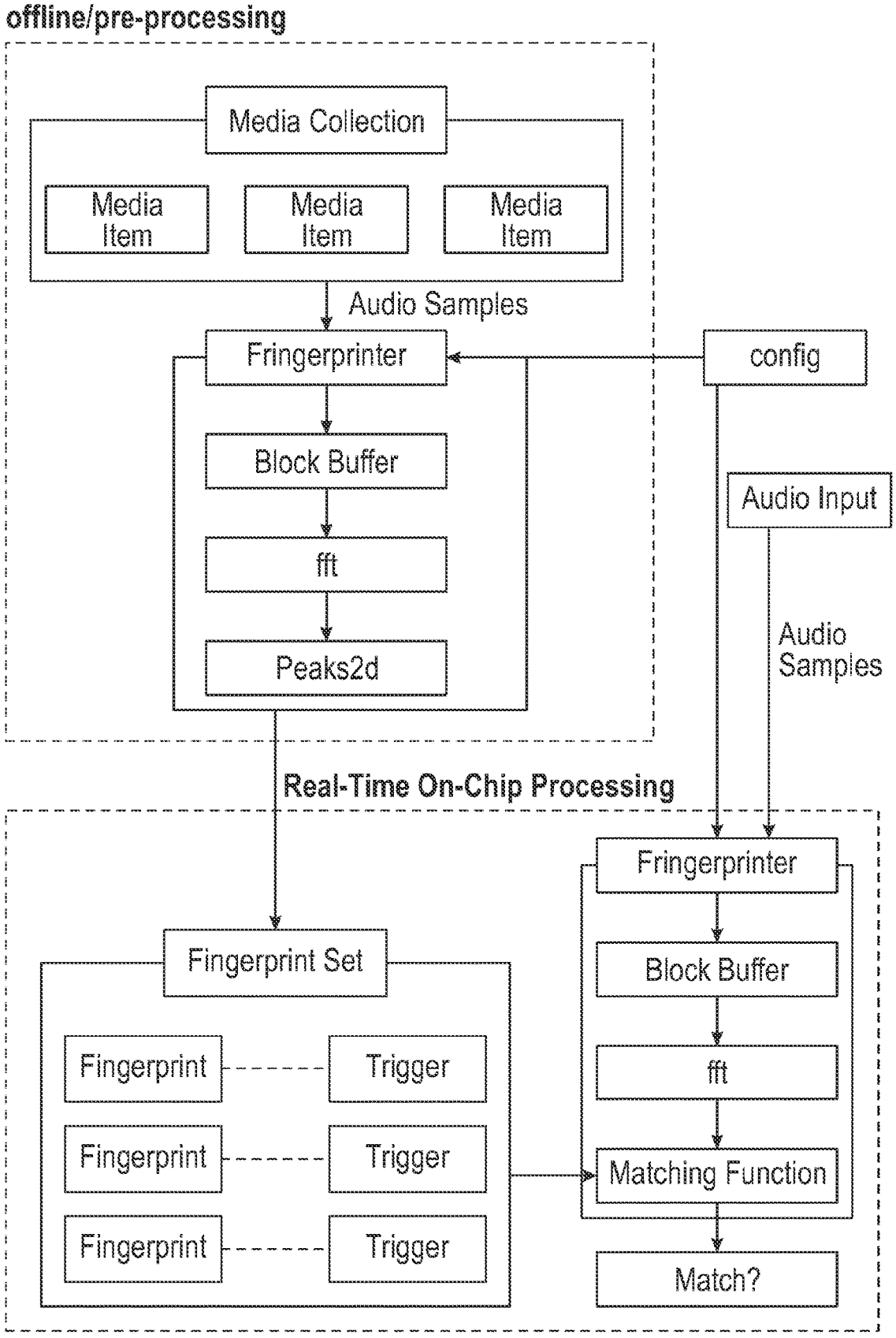

offline/pre-processing
Media Collection
Media Item
Media Item
Media Item
Audio Samples
Fringerprinter
config
Block Buffer
fft
Peaks2d
Audio Input
Audio Samples
Real-Time On-Chip Processing
Fingerprint Set
Fingerprint
Trigger
Fingerprint
Trigger
Fingerprint
Trigger
Fringerprinter
Block Buffer
fft
Matching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

FIG. 4

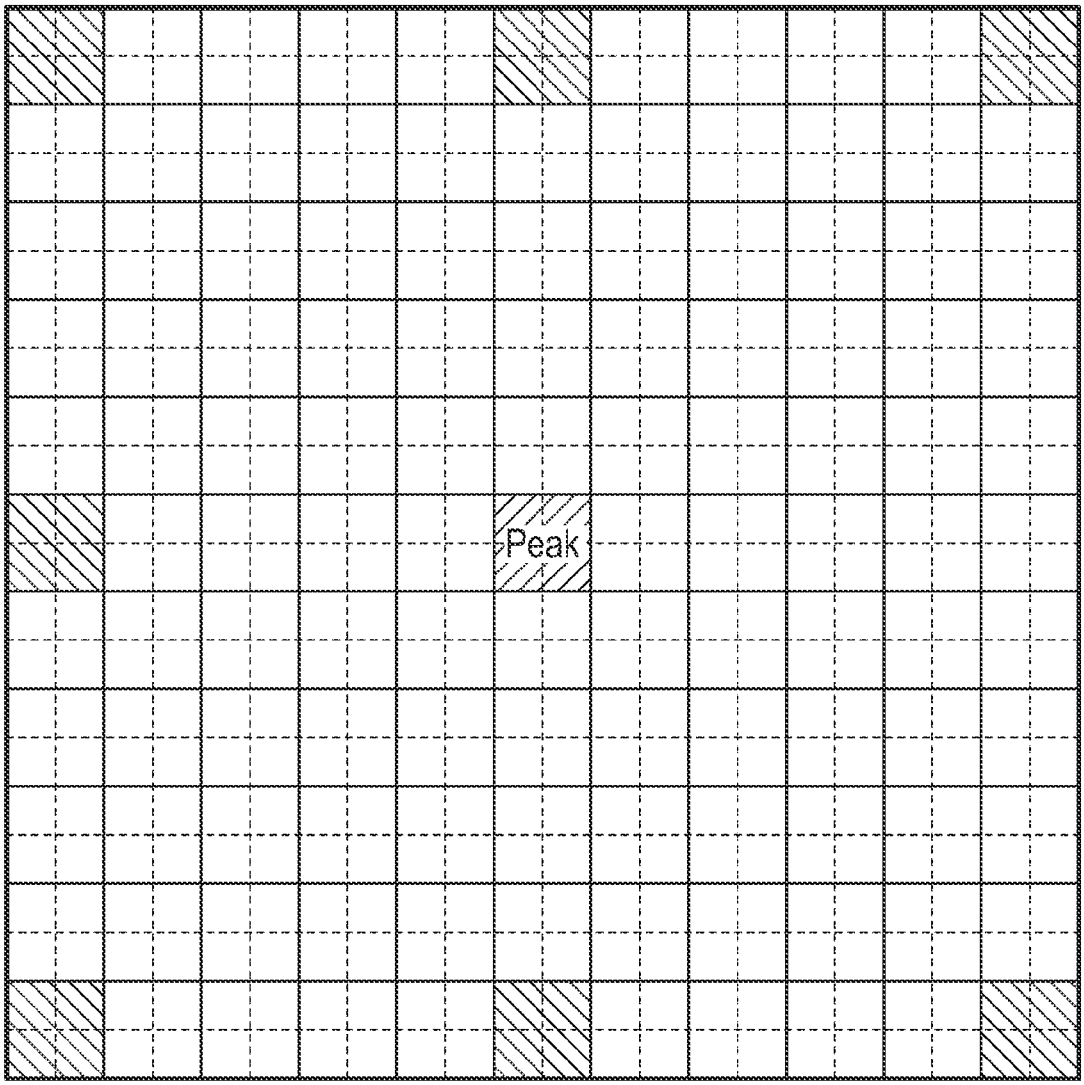


FIG. 5

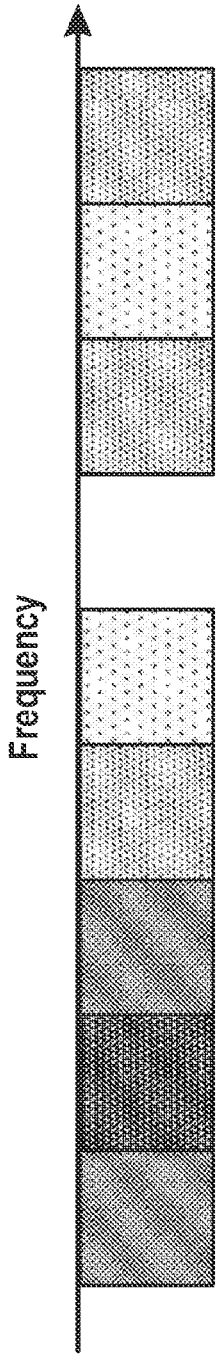


FIG. 6

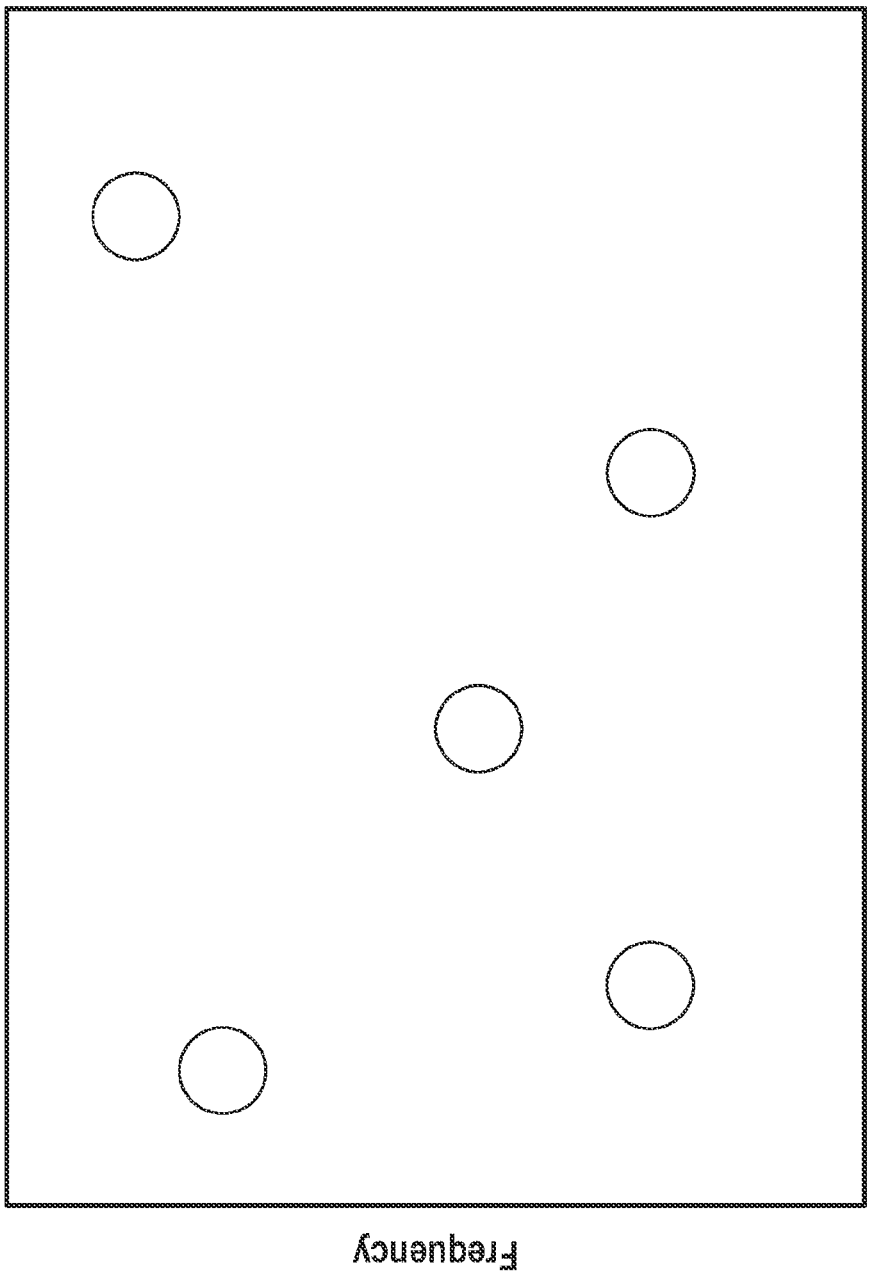


FIG. 7

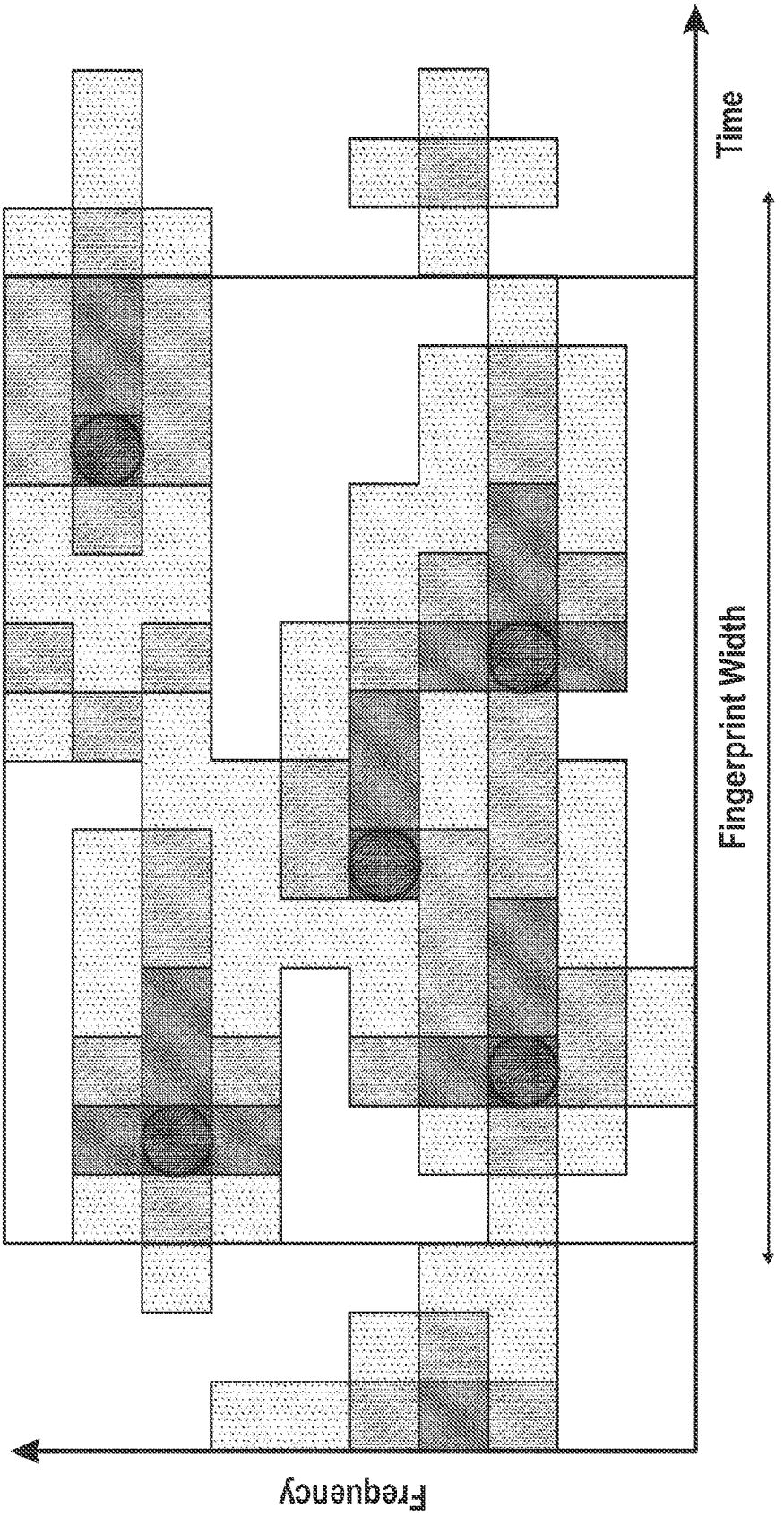


FIG. 8

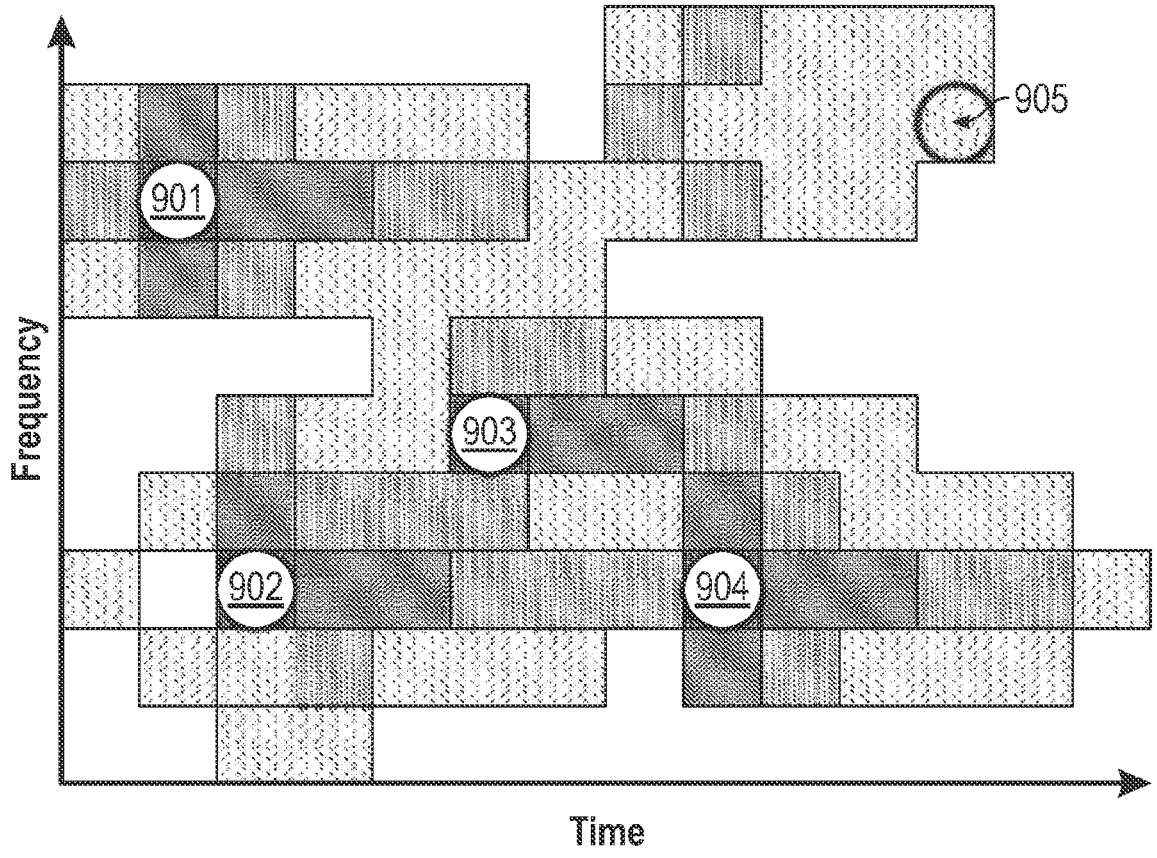


FIG. 9

METHOD AND SYSTEM FOR TRIGGERING EVENTS

FIELD OF INVENTION

[0001] 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

[0002] Signals, such as audio signals, can be processed to analyse various qualities of the signal.

[0003] 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.

[0004] 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.

[0005] 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.

[0006] 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

[0007] According to a first aspect of the invention there is provided a method of triggering an event, including:

[0008] a) receiving a signal stream;

[0009] b) detecting a trigger point within the signal stream using a fingerprint associated with the trigger point; and

[0010] c) triggering an event associated with the detected trigger point;

[0011] Other aspects of the invention are described within the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

[0013] FIG. **1**: shows a block diagram illustrating a system in accordance with an embodiment of the invention;

[0014] FIG. **2**: shows a flow diagram illustrating a method in accordance with an embodiment of the invention;

[0015] FIG. **3**: shows a flow diagram illustrating a method in accordance with an embodiment of the invention;

[0016] FIG. **4**: shows a block diagram illustrating a system in accordance with an embodiment of the invention;

[0017] FIG. **5**: shows a diagram illustrating a function for determining peak elevation using neighbour reference points in accordance with an embodiment of the invention;

[0018] FIG. **6**: shows a diagram illustrating a ID slice of a magnitude spectrum buffer in accordance with an embodiment of the invention;

[0019] FIG. **7**: shows a diagram illustrating peaks located using a system in accordance with an embodiment of the invention:

[0020] FIG. **8**: shows a diagram illustrating peaks located using a system in accordance with another embodiment of the invention; and

[0021] FIG. **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

[0022] The present invention provides a method and system for triggering events.

[0023] 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.

[0024] In FIG. **1**, a system **100** in accordance with an embodiment of the invention is shown.

[0025] 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**.

[0026] 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.

[0027] 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**.

[0028] The signal receiver **102** may be an input device such as a microphone, a camera, a radio frequency receiver or any analogue sensor.

[0029] 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.

[0030] 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.

[0031] In one embodiment, the audio signals may be in the audible range, inaudible range, or include components in both the audible and inaudible range.

[0032] Referring to FIG. **2**, a method **200** for triggering events in accordance with an embodiment of the invention will be described.

[0033] 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.

[0034] The signal stream may be received in real-time via the signal receiver.

[0035] The audio signal may be formed of audible, inaudible or a audible/inaudible components.

[0036] 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.

[0037] In one embodiment, the trigger points and associated fingerprints may be created as described in relation to FIG. 3.

[0038] 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.

[0039] The fingerprint may be formed of a set of peaks within 2D coordinate space of a magnitude spectrum.

[0040] 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).

[0041] 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.

[0042] 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.

[0043] 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.

[0044] In step 203, an event associated with the trigger point is triggered when the trigger point is detected.

[0045] 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.

[0046] Steps 201, 202 and 203 may be performed by one or more processors (e.g. processor(s) 101).

[0047] Referring to FIG. 3, a method 300 for creating a trigger points for an audio track in accordance with an embodiment of the invention will be described.

[0048] 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.

[0049] In step 302, an associated fingerprint is generated at the time location for each trigger point.

[0050] 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.

[0051] 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 FIG. 2.

[0052] Referring to FIGS. 4 to 9, a method of creating and detecting trigger points will be described in accordance with an embodiment of the invention.

[0053] 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:

[0054] minimal latency (less than 50 ms)

[0055] high reliability (99% recognition rate in typical acoustic environments over a distance of 1 m)

[0056] low false-positive rate (events should rarely be triggered at the wrong moment, if ever)

[0057] To perform audio triggering, two phases are involved.

[0058] 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.

[0059] 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.

[0060] Spectral Peaks

[0061] 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.

[0062] 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.

[0063] 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] An overview flow chart of the fingerprinter and recogniser ("scanner") is shown in FIG. **4**.

[0068] Peak Elevation Function

[0069] 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.

[0070] However, this may require a lot of calculations (up to '(NEIGHBOURHOOD_WIDTH*2+1)×(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.

[0071] 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 FIG. **5**.

[0072] Phase 1: Audio Fingerprinting

[0073] 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.

[0074] The fingerprinting phase proceeds as follows:

[0075] 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.

[0076] 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.

[0077] 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 FIG. **6**) running down the centre of the buffer, and checking if it is a local maxima in the Moore neighbourhood of size 'NEIGHBOURHOOD_WIDTH×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 12 dB), it is classified as a peak and added to another rolling buffer of width 'FINGERPRINT_WIDTH' as shown in FIG. **7**.

[0078] 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 FIG. **8**.

[0079] 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.

[0080] When the entire set of fingerprints have been collected, they are serialised to disk as a "fingerprint set". A fingerprint set is characterised by:

[0081] the number of fingerprints it contains;

[0082] 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).

[0083] Phase 2: Audio Recognition

[0084] The audio recogniser takes a set of fingerprints and a stream of audio (usually real-time), and attempts to recognise fingerprints within the stream.

[0085] It functions as follows:

[0086] 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.

[0087] 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. FIG. **9** shows examining of a fingerprint formed from the peaks of FIG. **7** against the spectral history buffer. Peak matches are shown at **901**, **902**, **903**, and **904**, and a peak non-match shown at **905**.

[0088] 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.

[0089] Matching then continues in case another fingerprint is matched at a higher confidence.

[0090] '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.

[0091] 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.

[0092] Audio Recognition: Additional Optimisations

[0093] Additional optimisations may be used to further reduce the CPU footprint of audio recognition, with minimal impact on recognition success rate:

[0094] Early Rejection

[0095] 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.

[0096] 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.

[0097] 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.

[0098] The values of 'CHIRP_FINGERPRINT_SCANNER_EARLY_REJECT_PROPORTION' and 'CHIRP_FINGERPRINT_SCANNER_EARLY_REJECT_LOWER_THRESHOL D' are selected to minimise the number of peaks that must be inspected on average, whilst minimising the number of actual matches that are missed.

[0099] Note that this technique can also function with an unordered set of peaks.

[0100] Disable Below Minimal Threshold

[0101] 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.

[0102] 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.

[0103] 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.

[0104] 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.

1. A system comprising:
memory; and
a processor operatively coupled to the memory and configured to:
define one or more trigger points at one or more time locations within an audio signal;
generate, based on the audio signal, a fingerprint for each trigger point; and
associate an event with each trigger point.

2. The system of claim **1**, wherein each trigger point is associated with a timing offset from a start of the audio signal.

3. The system of claim **1**, wherein the fingerprint is generated from a plurality of peaks identified within a Fast Fourier Transform (FFT) of the audio signal at the time location of the fingerprint.

4. The system of claim **3**, wherein each of the plurality of peaks is a local magnitude maxima of a 2D coordinate space created by each FFT block.

5. The system of claim **3**, wherein each of the plurality of peaks in the fingerprint is separated from other peaks in the fingerprint by a minimum spacing.

6. The system of claim **5**, wherein the minimum spacing includes a minimum width and a minimum height.

7. The system of claim **1**, wherein a plurality of trigger points are defined, a plurality of fingerprints corresponding to the plurality of trigger points are generated, and each of the fingerprints characterizes a different frame of the audio signal, each frame having a fixed duration of a same width.

8. The system of claim **1**, wherein the audio signal comprises a mono uncompressed audio.

9. The system of claim **1**, wherein each fingerprint comprises a plurality of peaks, and generating the fingerprint comprises:
reading the audio signal in Fast Fourier Transform (FFT) sized buffers for performing FFT on the audio signal;

deriving a magnitude spectrum from the FFT of the audio signal;

adding the magnitude spectrum to a rolling buffer;

identifying peak-candidates for each frame of the rolling buffer;

based on identifying a local maxima of the peak-candidate in a one-dimensional slice of the rolling buffer, applying an elevation function to determine the peak-candidate's elevation versus elevations of nearby cells; and

based on the elevation of the peak-candidate's elevation exceeding a threshold, classifying the peak-candidate as a peak of the fingerprint.

10. The system of claim **1**, wherein each fingerprint comprises a fingerprint width, fingerprint identification number, a number of peaks in the fingerprint, and coordinates of each peak in the fingerprint.

11. The system of claim **1**, wherein each fingerprint comprises a number of peaks in the fingerprint and coordinates of each peak in the fingerprint.

12. The system of claim **1**, further comparing:

a first device comprising the memory and the processor; and

a second device,

wherein the processor of the first device is further configured to transmit, to the second device, fingerprint information corresponding to the generated fingerprint,

wherein the second device is configured to:

identify a trigger point within an audio signal corresponding a fingerprint detected in the audio signal based on the fingerprint information; and

trigger an event associated with the fingerprint, a timing of the event synchronized to the identified trigger point within the audio signal.

13. The system of claim **1**, further comparing:

a first device comprising the memory and the processor; and

a second device,

wherein the processor of the first device is further configured to transmit, to the second device, fingerprint information corresponding to the generated fingerprint and audio signal,

wherein the second device is configured to:

receive the audio signal from the first device;

detect a trigger point within the audio signal, wherein detecting the trigger point within the audio signal includes:

processing the audio signal to provide a plurality of frames, and

identifying, in the plurality of frames, a frame including a fingerprint associated with the trigger point based on the fingerprint information; and

trigger an event associated with the fingerprint in the first frame, a timing of the event synchronized to the detected trigger point within the audio signal.

14. The system of claim **13**, wherein the second device comprises a microphone and the second device receives the audio signal via the microphone.

15. A method performed by a device comprising a processor and memory, the method comprising:

defining one or more trigger points at one or more time locations within an audio signal;

generating, based on the audio signal, a fingerprint for each trigger point; and

associating an event with each trigger point.

16. The method of claim **15**, wherein the fingerprint is generated from a plurality of peaks identified within a Fast Fourier Transform (FFT) of the audio track at the time location of the fingerprint.

17. The method of claim **16**, wherein each of the plurality of peaks is a local magnitude maxima of a 2D coordinate space created by each FFT block.

18. The method of claim **16**, wherein each of the plurality of peaks in the fingerprint is separated from other peaks in the fingerprint by a minimum spacing.

19. The method of claim **15**, wherein each fingerprint comprises a plurality of peaks and generating the fingerprint comprises:

reading the audio signal in Fast Fourier Transform (FFT) sized buffers for performing FFT on the audio signal;

deriving a magnitude spectrum from the FFT of the audio signal;

adding the magnitude spectrum to a rolling buffer;

identifying peak-candidates for each frame of the rolling buffer;

based on identifying a local maxima of the identified peak-candidate in a one-dimensional slice of the rolling buffer, applying an elevation function to determine the peak-candidate's elevation versus elevations of nearby cells; and

based on the elevation of the peak-candidate's elevation exceeding a threshold, classifying the peak-candidate as a peak of the fingerprint.

20. The method of claim **15**, wherein each fingerprint comprises a fingerprint width, fingerprint identification number, a number of peaks in the fingerprint, and coordinates of each peak in the fingerprint.

21. A non-transitory computer readable medium having stored therein computer-readable instructions that, when executed by one or more processors, cause an apparatus connected to the one or more processors to:

define one or more trigger points at one or more time locations within an audio signal;

generate, based on the audio signal, a fingerprint for each trigger point; and

associate an event with each trigger point.

22. An apparatus comprising:

memory; and

a processor operatively coupled to the memory and configured to control the apparatus to:

define a plurality of trigger points within an audio signal, each of the trigger points corresponding to a different time location within the audio signal;

generate, based on peaks identified within a Fast Fourier Transform (FFT) of the audio signal, a fingerprint for each trigger point, wherein each fingerprint comprises a plurality of peaks identified within the FFT of the audio signal at the time location corresponding to the trigger point; and

associate an event with each trigger point.

23. The apparatus of claim **22**, wherein each fingerprint comprises a fingerprint width, fingerprint identification number, a number of peaks in the fingerprint, and coordinates of each peak in the fingerprint.

* * * * *