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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 May 2019 (23.05.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/097227 A1

(51) International Patent Classification:
*G10L 19/00* (2013.01) *G10L 25/00* (2013.01)

(21) International Application Number:
PCT/GB2018/053299

(22) International Filing Date:
14 November 2018 (14.11.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1718800.4 14 November 2017 (14.11.2017) GB

(71) Applicant: **QUEEN MARY UNIVERSITY OF LONDON** [GB/GB]; Mile End Road, London Greater London E1 4NS (GB).

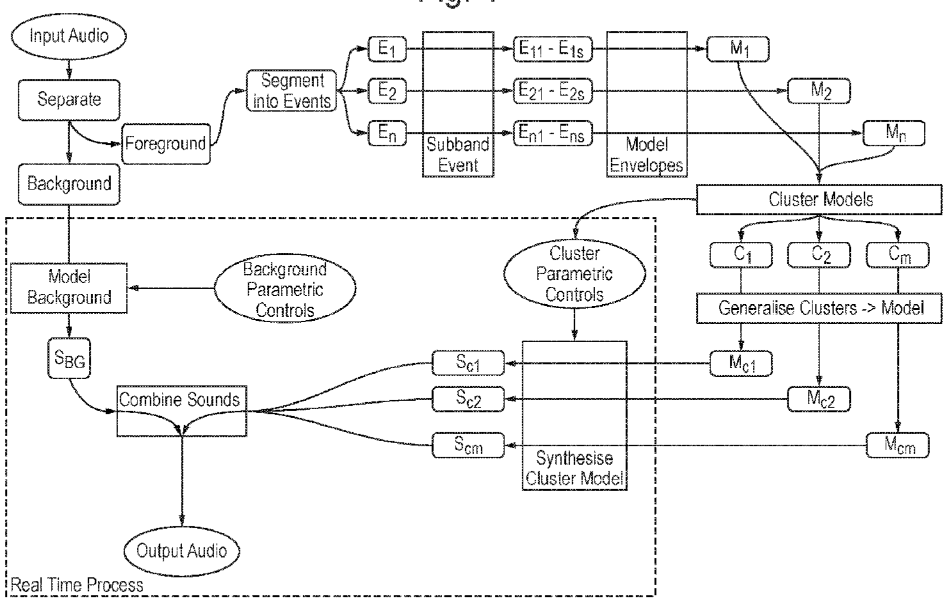
(72) Inventors: **MOFFAT, David**; Electronic Enginerring & Computer Science, Queen Mary University Of London, Mile End Road, London Greater London E1 4NS (GB). **REISS, Joshua**; Electronic Engineering & Computer Science, Mile End Road, Queen Mary University of London, London Greater London E1 4NS (GB).

(74) Agent: **HUNT-GRUBBE, Henry**; J A KEMP, 14 South Square, Gray's Inn, London Greater London WC1R 5JJ (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.

**(54) Title:** GENERATION OF SOUND SYNTHESIS MODELS

Fig. 1




**(57) Abstract:** Disclosed herein is a method of generating a sound synthesis model of an audio signal, the method comprising: receiving an audio signal; identifying audio events in the received an audio signal; separating each of the identified audio events into audio sub-bands; generating a model of each of the audio sub-bands; applying a machine learning technique to each generated model of each of the audio sub-bands so as to determine a plurality of different types of audio event in the received audio signal; generating, for each of the determined plurality of different types of audio event, a model of the type of audio event; and generating a sound synthesis model of the received audio signal in dependence on each of the generated models for types of audio event.

WO 2019/097227 A1

*[Continued on next page]*

**(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))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

# GENERATION OF SOUND SYNTHESIS MODELS

## Field

The field of the invention is the synthesis of sounds. Embodiments use machine learning to automatically generate sound effect models for synthesising sounds. Advantageously, sound effect models are easily generated.

## Background

Sound effects have a number of applications in creative industries. There are libraries of sound effects that store pre-recorded sound samples. A problem experienced by libraries of sound effects is that it is necessary for each library to store a very large number of pre-recorded sound samples in order for a broad range of sound effects to be generated. In addition, adapting pre-recorded sound effects for use in different scenarios is difficult.

There is therefore a need to improve known techniques for generating sound effects.

## Summary

According to a first aspect of the invention, there is provided a method of generating a sound synthesis model of an audio signal, the method comprising: receiving an audio signal; identifying audio events in the received an audio signal; separating each of the identified audio events into audio sub-bands; generating a model of each of the audio sub-bands; applying a machine learning technique to each generated model of each of the audio sub-bands so as to determine a plurality of different types of audio event in the received audio signal; generating, for each of the determined plurality of different types of audio event, a model of the type of audio event; and generating a sound synthesis model of the received audio signal in dependence on each of the generated models for types of audio event.

Preferably, the machine learning technique is an unsupervised clustering technique.

Preferably, each model for a type of audio event is a probability distribution.

Preferably, the method further comprises: separating the received audio signal into a foreground component and a background component; wherein said identification of audio events in the received an audio signal is only performed on the foreground component of the received audio signal; and the sound synthesis model of the received audio signal is generated in dependence on the background component.

Preferably, separating the received audio signal into a foreground component and a background component comprises performing a Short Time Fourier Transform.

Preferably, the model of each of the audio sub-bands is generated by using a gamma distribution model and/or a polynomial regression model.

Preferably, the method further comprises determining the principle components of each of the audio sub-bands; wherein the machine learning technique is applied in dependence on the determined principle components.

Preferably, the received audio signal is a plurality of audio signals.

Preferably, the method is computer-implemented.

Preferably, the method is implemented by software modules of a computing device.

According to a second aspect of the invention, there is provided a computer program comprising instructions that, when executed by a computing device, cause the computing device to perform the method according to the first aspect.

According to a third aspect of the invention, there is provided a computing device arranged to perform the method according to the first aspect.

According to a fourth aspect of the invention, there is provided a method of synthesising an audio signal, the method comprising: receiving a model for generating a synthesised background component of an audio signal; receiving a plurality of models for generating a synthesised foreground component of the audio signal, wherein each of the plurality of models is for a different type of audio event; receiving control parameters of the models; generating a synthesised background component of the audio signal in dependence on the model for generating a synthesised background component of the audio signal and one or more of the control parameters; and generating a synthesised foreground component of the audio signal in dependence on the received plurality of models for generating a synthesised foreground component of the audio signal and one or more of the control parameters; wherein generating the synthesised foreground component of the audio signal comprises: generating vectors by sampling each of the models for generating a synthesised foreground component of an audio signal; generating sub-band envelopes in dependence on the generated vectors; and applying the envelopes to one or more sub-band filtered probability density distributions.

Preferably, the received model for generating a synthesised background component and/or plurality of models for generating a synthesised foreground component are generated according to the method of the first aspect.

Preferably, the received model for generating a synthesised background component is a filtered probability density distribution.

Preferably, the method further comprises expanding the generated vectors using an inverse principal components analysis; wherein the sub-band envelopes are generated in dependence on the expanded vectors.

Preferably, the control parameters for the models for generating a synthesised foreground component comprise one or more of density, density distribution, gain, gain distribution, timbral and tonal control parameters.

Preferably, the audio signal is synthesised substantially in real-time.

Preferably, one or more of the model for generating a synthesised background component of an audio signal, plurality of models for generating a synthesised foreground component of the audio signal and control parameters of the models are received via a communications network; and the synthesised background and foreground components of the audio signal are transmitted over the communications network.

Preferably, the received audio signal is a plurality of audio signals.

Preferably, the method is computer-implemented.

Preferably, the method is implemented by software modules of a computing device.

According to a fifth aspect of the invention, there is provided a computer program comprising instructions that, when executed by a computing device, cause the computing device to perform the method according to the fourth aspect.

According to a sixth aspect of the invention, there is provided a computing device arranged to perform the method according to the fourth aspect.

**List of Figures**

Figure 1 shows processing steps in sound synthesis model generation and audio signal generation techniques according to embodiments.

**Description**

Embodiments provide a sound synthesis system that uses sound effect models to synthesise sounds. This improves on the use of sound effect libraries since large libraries of pre-recorded sounds are not required. The control of each generated sound effect by a user is also improved.

Embodiments also provide techniques for automatically generating sound effect models for use by sound synthesis systems. Advantageously, the generation of sound effect models according to embodiments is quick and accurate.

According to embodiments, machine learning techniques are used to process audio signals in order to generate a framework of sound effect models. The audio signals that the sound effect models are generated in dependence on may be pre-recorded sounds from a sound effect library.

According to embodiments, an audio signal, which may be any type of audio source/sample and/or a combination of audio signals, is first separated into background and foreground components. The background component is modelled as comprising a constant filtered noise signal. The foreground component is modelled as comprising a number of regular sound events. The sound events are then identified, separated and analysed independently. Each sound event then has a model representation of it created based on analysing the time and frequency properties of the sound event. These individual models are then grouped together into clusters by an unsupervised machine learning clustering system that identifies a number of sound categories, i.e. sound types. That is to say, each cluster corresponds to a different sound category/type. Each of the clusters is then modelled, with each parameter for recreating a sound being obtained from a probability distribution, such as a Gaussian distribution.

The modelled clusters and background component can be used as sound synthesis models for synthesising the modelled sound. In a sound synthesis system, the overall controls for each cluster are presented to the user who can change, for example, the volume, rate and synchronisation of each of the clusters as well as the background component. A user can also control the timbral and tonal properties of the synthesised sound. A controllable synthesised audio signal is therefore generated that is similar to the original audio signal.

Embodiments are described in more detail below.

Embodiments comprise the separate techniques of sound synthesis model generation and audio signal generation.

Sound synthesis model generation comprises the generation sound synthesis models for modelling one or more audio signals. Sound synthesis model generation is a processing operation of one or more audio signals and is a pre-processing that is performed prior to sound synthesis.

Audio signal generation, also referred to as sound signal generation or audio/sound synthesis, comprises the generation of one or more audio signals. The audio signal generation is a real-time synthesis approach. The audio signal generation may be a recreation of an audio signal that has been modelled by the sound synthesis model generation.

The sound synthesis model generation techniques according to embodiments can be performed offline whereas the audio signal generation according to embodiments is preferably performed online and substantially in real-time.

Figure 1 shows the processing steps in the sound synthesis model generation and audio signal generation techniques according to embodiments. The sound synthesis model generation processes are shown outside of the dashed box and the audio signal generation techniques are shown inside the dashed box.

The sound synthesis model generation is described in detail below.

The processes for sound synthesis model generation are performed by a model generation system. The model generation system may be any type of computing system. The components of the model generation system may be implemented as software modules.

An audio signal is input into the model generation system. The audio signal may be from any type of sound source and it may be a combination of a plurality of audio signals from a variety of sound sources.

The model generation system then analyses the input audio signal. In the analysis process, the background component of the audio signal is separated from the input audio signal. The separation of the background component may be performed by, for example, median filtering of the Short Time Fourier Transform (STFT). However, embodiments also include other techniques for separating the background component. The median spectrum of the background component is calculated. The modelling technique uses the median spectrum of the background component as a fixed, i.e. constant, background spectrum.

The foreground component of the input audio signal is the rest of the audio signal after the background component has been removed. The foreground component can be generated by the same process that separates the background component from the input audio signal. A difference between the foreground component and the background component is that the foreground component is not constant and the foreground component can change.

The foreground component of the input audio signal is modelled as comprising individual audio events that are identifiable as peaks in the amplitude spectrum of the foreground component. For example, the start and the end of an audio event can be detected through local minima identification of the amplitude spectrum using the techniques disclosed in 'Sadjad Siddiq. Morphing of granular sounds. In Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15), pages 4–11, 2015' and 'Sadjad Siddiq. Data-driven granular synthesis. In Audio Engineering Society Convention 142, May 2017', the entire contents of both of which are incorporated herein by reference. Each of the identifiable audio events can be separately processed. As shown in Figure 1, the number of identified audio events is referred to herein as 'n'.

The model generation system then processes each audio event separately. Each audio event is decomposed into a number of individual audio sub-bands. The decomposition of the audio events into sub-bands may be based on the Equivalent Rectangular Band (ERB) scale, however embodiments include using other techniques for the decomposition of the audio events into sub-bands. The decomposition may, for example, be performed using the techniques disclosed in 'Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. Hearing research, 47(1):103–138, 1990', the entire contents of which are incorporated herein by reference. As shown in Figure 1, the number of sub-bands for each of the 'n' events is referred to herein as 's'.

Each of the sub-bands has an envelope. The next processing step generates an approximate model of each individual envelope. For example, each envelope may be modelled by a gamma distribution model and/or each envelope may be modelled as a polynomial using a polynomial regression model. The model may provide a vector representation of the envelope of each sub-band for each audio event. This can be considered to be a vector, referred to as a sound event vector for each of the sound events. The principal components of each of the sound event vectors may then be taken, so as to perform dimensionality reduction.

A machine learning technique is then performed. The machine learning technique is preferably an unsupervised clustering technique. The machine learning technique determines a number of different types of sound event. The machine learning technique can determine the optimal number of different types of sound event that were identified in the separated foreground component of the input audio signal. The machine learning technique can be based on any of a number of techniques in machine learning, such as neural network techniques. The output from the machine learning technique is one or more models for each of a plurality of different types of audio event.

As shown in Figure 1, the number of outputs from the unsupervised clustering technique, i.e. clusters of models for a type of sound event, is 'm'. A generalisable model is then generated, in dependence on these clusters of models, where each parameter can be considered to be a value sampled from any of a number of probability distributions. For example, a Gaussian distribution may be used to represent all of the cluster parameters.

Accordingly, a model is generated for each of the 'm' clusters. The models may be considered to be sound synthesis models since, as explained below, the generated model for each of the 'm' clusters, in addition to other data generated in the sound synthesis model generation techniques by the model generation system, can be used as inputs to a real-time sound synthesis system.

Embodiments also include the generation of a synthesised audio signal by a sound synthesis system. The audio signal may be generated in real-time.

The generated audio signal may be one or more audio signals that have been modelled by the above-described model generation system according to embodiments. However, the sound synthesis system according to embodiments may also be used to generate audio signals that have been obtained by a different sound synthesis model generation technique. The sound synthesis system according to embodiments may also be used to generate audio signals that are not a recreation of a modelled specific audio signal but instead constructed from different modelled components from a plurality of modelled audio signals. For example, the sound synthesis system may generate a broad range of audio signals from modelled components that each correspond to a different class of sound.

The processes performed by the sound synthesis system according to embodiments are shown within the dashed line in Figure 1.

The sound synthesis system according to embodiments generates a synthesised audio signal from the following inputs to the sound synthesis system:

- A spectrum model of the background component of the audio signal being synthesised
- The number of clusters produced by the unsupervised clustering technique, i.e. the ‘m’ clusters of models, and the model of each of the ‘m’ clusters
- User-defined inputs from which control parameters of the components of the sound synthesis process can be generated.

If any of the user-defined inputs are not received then the sound synthesis system is configured to use default values of the user inputs that are not provided when generating the audio signal being synthesised.

The synthesis process uses the received spectrum model of the background component to generate a synthesised background sound. This can be performed by filtering Gaussian White Noise (GWN) so that it has the same spectrum properties as the spectrum model of the background component. A gain control is provided for the synthesised background component.

For each model of the 'm' clusters, the parameter controls preferably include all of density, density distribution, gain and gain distribution. The parameter controls allow the properties of each of the clusters in the sound synthesis process to be changed.

After each parameter has been set, models of each cluster, which are probability distributions, are sampled using a probabilistically triggered Monte Carlo sampling method and cluster parameter vectors are obtained.

The cluster parameter vectors are expanded using an inverse Principal Component Analysis (PCA) and, in dependence on this, approximations of each of the envelopes of sub-bands of events are constructed. The sub-band envelopes are applied to a sub-band filtered GWN, and combined to generate an entire sound event.

All of the sound events in the foreground of a synthesised signal are constructed and combined with the synthesised background component so as to generate a synthesised audio signal.

The probability, density and gain control parameters for each sound event allow for tonal and timbral control of the sonic texture.

Advantageously, embodiments provide controllable and interactive sound effect models, so that a sound designer may produce and control the type of sound that they desire. Embodiments allow a designer to take a sound recording, or range of sound recordings, and feed them into a machine learning system in order to generate sound synthesis models.

The designer can then use the sound synthesis models to generate a controllable sound synthesis system.

A particularly advantageous aspect of embodiments is that the features used to classify and describe the sounds correspond to perceptually meaningful descriptions, e.g., roughness, strength of an impact, noisiness, boomy, etc. This allows for the automatic generation of user controls that can be refined as appropriate.

Embodiments include a number of modifications and variations of the techniques as described above.

Embodiments include the use of any probability distribution models in the sound synthesis model generation and audio signal generation processes. For example, the synthesised background sound may alternative be generated using a different model than GWN.

The audio signal generation processes preferably comprise an API so that the inputs to the audio signal generation processes and synthesised audio signal generated by the audio signal generation processes can be received from and provided to a remote user. For example, the synthesised audio signal may be provided over the Internet to a user who is remote from the sound synthesis system.

The flow charts and descriptions thereof herein should not be understood to prescribe a fixed order of performing the method steps described therein. Rather, the method steps may be performed in any order that is practicable. Although the present invention has been described in connection with specific exemplary embodiments, it should be understood that various changes, substitutions, and alterations apparent to those skilled in the art can be made to the disclosed embodiments without departing from the spirit and scope of the invention as set forth in the appended claims.

Methods and processes described herein can be embodied as code (e.g., software code) and/or data. Such code and data can be stored on one or more computer-readable media, which may include any device or medium that can store code and/or data for use by a computer system. When a computer system reads and executes the code and/or data stored on a computer-readable medium, the computer system performs the methods and processes embodied as data structures and code stored within the computer-readable storage medium. In certain embodiments, one or more of the steps of the methods and processes described herein can be performed by a processor (e.g., a processor of a computer system or data storage system). It should be appreciated by those skilled in the art that computer-readable media include removable and non-removable structures/devices that can be used for storage of information, such as computer-readable instructions, data structures, program modules, and other data used by a computing system/environment. A computer-readable medium includes, but is not limited to, volatile memory such as random access memories (RAM, DRAM, SRAM); and non-volatile memory such as flash memory, various read-only-memories (ROM, PROM, EPROM, EEPROM), magnetic and ferromagnetic/ferroelectric memories (MRAM, FeRAM), phase-change memory and magnetic and optical storage devices (hard drives, magnetic tape, CDs, DVDs); network devices; or other media now known or later developed that is capable of storing computer-readable information/data. Computer-readable media should not be construed or interpreted to include any propagating signals.

## Claims

1. A method of generating a sound synthesis model of an audio signal, the method comprising:
receiving an audio signal;
identifying audio events in the received an audio signal;
separating each of the identified audio events into audio sub-bands;
generating a model of each of the audio sub-bands;
applying a machine learning technique to each generated model of each of the audio sub-bands so as to determine a plurality of different types of audio event in the received audio signal;
generating, for each of the determined plurality of different types of audio event, a model of the type of audio event; and
generating a sound synthesis model of the received audio signal in dependence on each of the generated models for types of audio event.

2. The method according to claim 1, wherein the machine learning technique is an unsupervised clustering technique.

3. The method according to claim 1 or 2, wherein each model for a type of audio event is a probability distribution.

4. The method according to any preceding claim, further comprising:
separating the received audio signal into a foreground component and a background component;
wherein said identification of audio events in the received an audio signal is only performed on the foreground component of the received audio signal; and
the sound synthesis model of the received audio signal is generated in dependence on the background component.

5. The method according to claim 4, wherein separating the received audio signal into a foreground component and a background component comprises performing a Short Time Fourier Transform.

6. The method according to any preceding claim, wherein the model of each of the audio sub-bands is generated by using a gamma distribution model and/or a polynomial regression model.

7. The method according to any preceding claim, further comprising determining the principle components of each of the audio sub-bands;

   wherein the machine learning technique is applied in dependence on the determined principle components.

8. The method according to any preceding claim, wherein the received audio signal is a plurality of audio signals.

9. The method according to any preceding claim, wherein the method is computer-implemented.

10. The method according to any preceding claim, wherein the method is implemented by software modules of a computing device.

11. A computer program comprising instructions that, when executed by a computing device, cause the computing device to perform the method according to any of claims 1 to 10.

12. A computing device arranged to perform the method according to any of claims 1 to 10.

13. A method of synthesising an audio signal, the method comprising:
    receiving a model for generating a synthesised background component of an audio signal;

receiving a plurality of models for generating a synthesised foreground component of the audio signal, wherein each of the plurality of models is for a different type of audio event;

receiving control parameters of the models;

generating a synthesised background component of the audio signal in dependence on the model for generating a synthesised background component of the audio signal and one or more of the control parameters; and

generating a synthesised foreground component of the audio signal in dependence on the received plurality of models for generating a synthesised foreground component of the audio signal and one or more of the control parameters;

wherein generating the synthesised foreground component of the audio signal comprises:

generating vectors by sampling each of the models for generating a synthesised foreground component of an audio signal;

generating sub-band envelopes in dependence on the generated vectors; and

applying the envelopes to one or more sub-band filtered probability density distributions.

14. The method according to claim 13, wherein the received model for generating a synthesised background component and/or plurality of models for generating a synthesised foreground component are generated according to the method of any of claims 1 to 10.

15. The method according to claim 13 or 14, wherein the received model for generating a synthesised background component is a filtered probability density distribution.

16. The method according to any of claims 13 to 15, further comprising expanding the generated vectors using an inverse principal components analysis;

wherein the sub-band envelopes are generated in dependence on the expanded vectors.

17. The method according to any of claims 13 to 16, wherein the control parameters for the models for generating a synthesised foreground component comprise one or more of density, density distribution, gain, gain distribution, timbral and tonal control parameters.

18. The method according to any of claims 13 to 17, wherein the audio signal is synthesised substantially in real-time.

19. The method according to any of claims 13 to 18, wherein one or more of the model for generating a synthesised background component of an audio signal, plurality of models for generating a synthesised foreground component of the audio signal and control parameters of the models are received via a communications network; and

the synthesised background and foreground components of the audio signal are transmitted over the communications network.

20. The method according to any of claims 13 to 19, wherein the received audio signal is a plurality of audio signals.

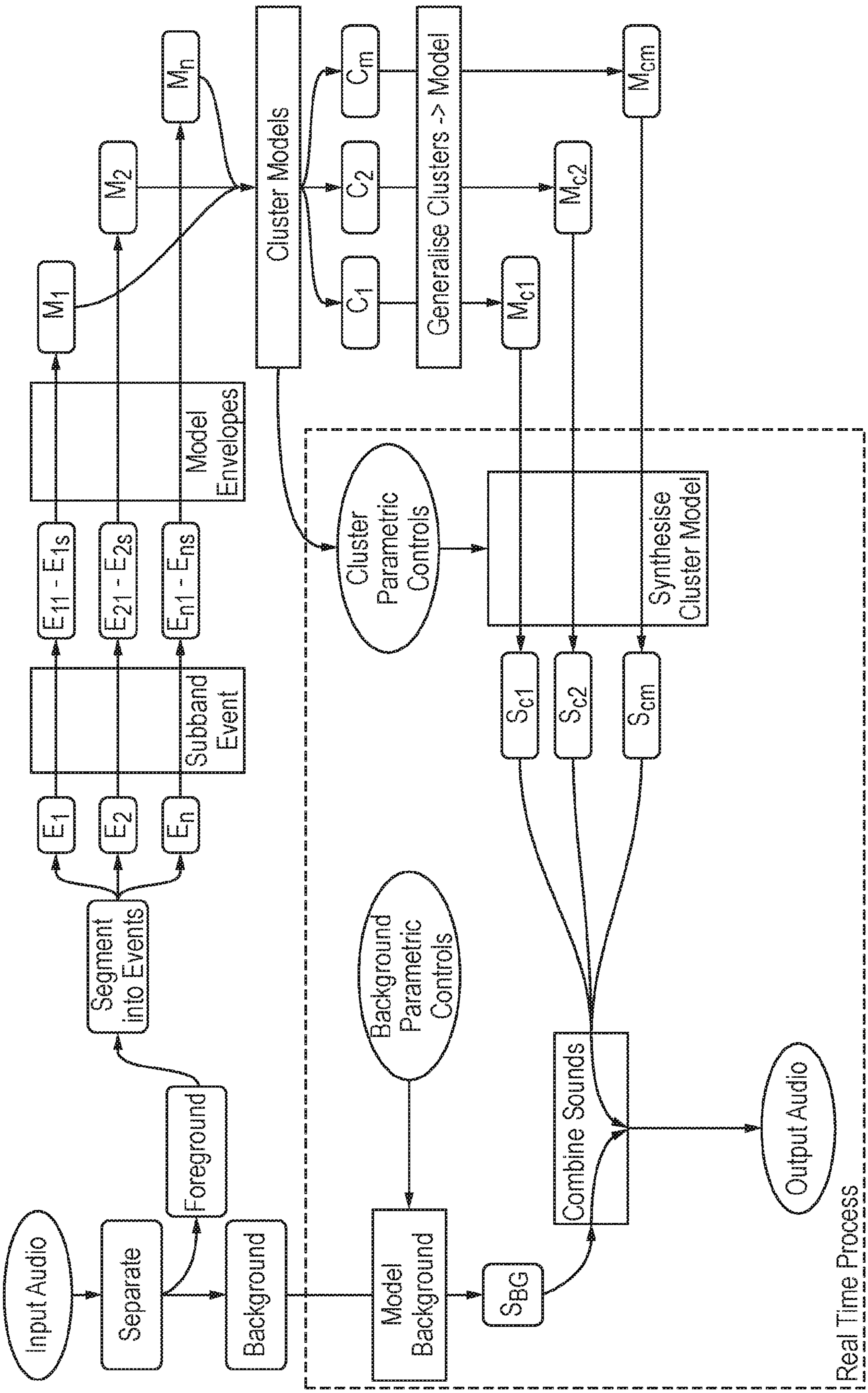
21. The method according to any of claims 13 to 20, wherein the method is computer-implemented.

22. The method according to any of claims 13 to 21, wherein the method is implemented by software modules of a computing device.

23. A computer program comprising instructions that, when executed by a computing device, cause the computing device to perform the method according to any of claims 13 to 22.

24. A computing device arranged to perform the method according to any of claims 13 to 22.

Fig. 1


Input Audio
Separate
Foreground
Background
Segment into Events
E1
E2
En
Subband Event
E11 - E1s
E21 - E2s
En1 - Ens
Model Envelopes
M1
M2
Mn
Cluster Models
C1
C2
Cm
Generalise Clusters -> Model
Mc1
Mc2
Mcm
Cluster Parametric Controls
Synthesise Cluster Model
Sc1
Sc2
Scm
Background Parametric Controls
Model Background
SBG
Combine Sounds
Output Audio
Real Time Process

# INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2018/053299

**A. CLASSIFICATION OF SUBJECT MATTER**
INV. G10L19/00 G10L25/00
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 | LEVY ET AL: "Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation", ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2006. ICASSP 2006 PROCEEDINGS . 2006 IEEE INTERNATIONAL CONFERENCE ON TOULOUSE, FRANCE 14-19 MAY 2006, PISCATAWAY, NJ, USA,IEEE, PISCATAWAY, NJ, USA, 14 May 2006 (2006-05-14), - 19 May 2006 (2006-05-19), pages V-V, XP031015952, DOI: 10.1109/ICASSP.2006.1661200 ISBN: 978-1-4244-0469-8 | 1-3,7-12 |
| A | page 13, left-hand column, line 1 - page 14, left-hand column, line 31 | 4-6 |
| | ----- | |
| | -/-- | |

[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 |
|---|---|
| 28 February 2019 | 27/03/2019 |

| 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 | Dobler, Ervin |

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X<br><br>A | US 2001/044719 A1 (CASEY MICHAEL A [US]) 22 November 2001 (2001-11-22)<br>paragraph [0038] - paragraph [0050]; figure 1<br>paragraph [0112] - paragraph [0132]; figure 13<br>----- | 1-3,8-12<br><br>4-7 |
| A | JONATHAN DOHERTY ET AL: "Pattern Matching Techniques for Replacing Missing Sections of Audio Streamed across Wireless Networks", ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY (TIST), ASSOCIATION FOR COMPUTING MACHINERY CORPORATION, 2 PENN PLAZA, SUITE 701 NEW YORK NY 10121-0701 USA,<br>vol. 6, no. 2, 31 March 2015 (2015-03-31), pages 1-38, XP058067860,<br>ISSN: 2157-6904, DOI: 10.1145/2663358<br>Sections 2.1, 5,6 and 7 on pages 25:5-25:7 and 25:15-25:26;<br>figures 1,4,5,7-12<br>----- | 13-24 |
| A | RADHAKRISHNAN R ET AL: "Modelling sports highlights using a time series clustering framework & model interpretation", VISUAL COMMUNICATIONS AND IMAGE PROCESSING; 20-1-2004 - 20-1-2004; SAN JOSE,,<br>vol. 5682, 1 January 2005 (2005-01-01), pages 269-276, XP009120135,<br>DOI: 10.1117/12.588059<br>ISBN: 978-1-62841-730-2<br>page 270, line 1 - page 275, line 2;<br>figures 1-7<br>----- | 13-24 |

2

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2018/053299

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [ ] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [X] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

[ ] The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

[ ] The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

[X] No protest accompanied the payment of additional search fees.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-12

   Generating a sound synthesis model

   ---

2. claims: 13-24

   Synthesizing an audio signal with foreground and background components

   ---

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2018/053299

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| US 2001044719 A1 | 22-11-2001 | DE 60203436 T2 | 09-02-2006 |
| | | EP 1260968 A1 | 27-11-2002 |
| | | JP 2003015684 A | 17-01-2003 |
| | | US 2001044719 A1 | 22-11-2001 |