



- (51) **International Patent Classification:**  
G06F 17/30 (2006.01)
- (21) **International Application Number:**  
PCT/GB20 15/053026
- (22) **International Filing Date:**  
14 October 2015 (14.10.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
1418261.2 15 October 2014 (15.10.2014) GB
- (71) **Applicant:** BRITISH BROADCASTING CORPORATION [GB/GB]; Broadcasting House, London Greater London W1A 1AA (GB).
- (72) **Inventors:** HUGHES, Chris; Broadcasting House, London Greater London W1A 1AA (GB). ARMSTRONG, Mike; Broadcasting House, London Greater London W1A 1AA (GB).

- (74) **Agent:** LOVELESS, Ian; 16 Theobalds Road, London Greater London WC1X 8PL (GB).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on nextpage]

(54) **Title:** SUBTITLING METHOD AND SYSTEM

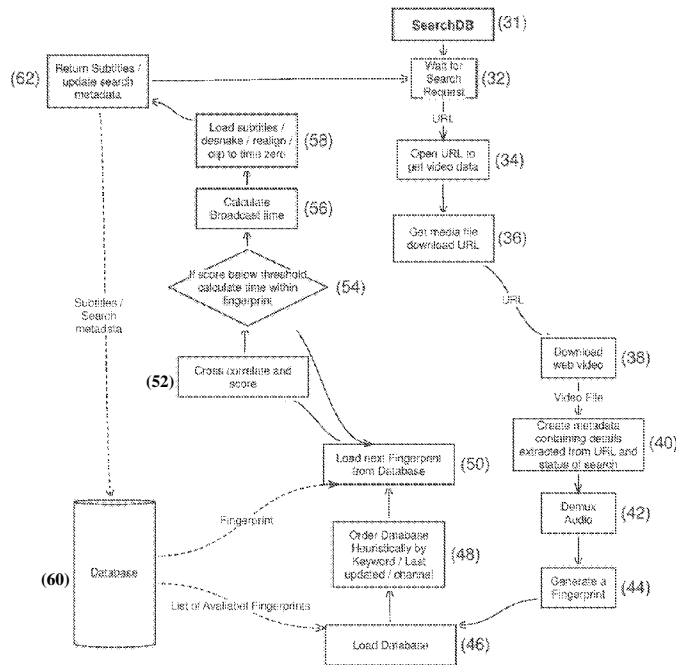


Figure 2

(57) **Abstract:** A method of retrieving supplementary data related to audio-video content with reference to a reference version of that content involves deriving an audio signature from audio-video content. The audio signature is searched against reference audio signatures of reference versions of the audio-video content. If a match is found, supplementary data related to the reference audio-video content is retrieved. The search is a directed search using data extracted from the transmission route by which the audio-video content is to be delivered to a consumer. The search process is thereby more efficient as the search may be started at an appropriate location or conducted in an appropriate direction, rather than searching blindly.



## Subtitling Method and System

### BACKGROUND OF THE INVENTION

5

This invention relates to obtaining data related to audio-video content in situations in which such data may already be available, but not linked to the audio-video content.

10

Audio-video content such as television programmes and the like may be distributed to consumers via a variety of media. Traditional linear television programmes are broadcast over a broadcast network, whether terrestrial, cable or satellite and may be consumed by users by a variety of receiver devices. Such audio-video content includes the audio and video components and may also include additional content such as subtitles, audio description and other additional data. Such audio-video content may also be distributed via additional routes, in particular via on demand services such as on the internet. In the process of repurposing audio-video content for other distribution channels, though, additional content such as subtitles, audio description or other data or metadata is not routinely copied with the result that such additional content is not available on the version distributed by the additional distribution route. Similar issues can occur when audio-video content is re-versioned for broadcast.

15

20

We have appreciated the need to allow for data related to audio-video content, in particular supplementary data such as subtitles, should be easily retrieved for use with a complete or partial copy of that audio-video content.

25

### SUMMARY OF THE INVENTION

30

The present invention provides a system and method for retrieving supplementary data related to audio-video content in relation to a complete or partial copy of that audio-video content using a signature of the audio component of the audio-video content and matching such a signature to a reference signature to extract supplementary data for the audio-video content.

35

The embodiment of the invention preferably uses a feature extraction technique that produces a signature of the audio component. The signature may be variable in length so as to allow matching of an entire television programme, portion of a programme, or edited version, thereby allowing matching of an edited  
5 version of a programme against an original programme.

The signature (which may also be referred to as a fingerprint) may be any process or function which reduces the amount of information and complexity in the representation of the data which retains a low likelihood of correlation with a  
10 signature from an unrelated set of data whilst ensuring repeatable correlation when comparing similar signals. In the embodiment, the signature preferably retains a temporal resolution of around a second so that the location of the matching audio can be located to a similar resolution.

15 The search strategy may be to match using audio fingerprints alone, but preferably additional data is used to direct the search, in particular a date time field that specifies a date related to the audio video content such as the date it was first broadcast, which can then be used to direct the order of the search against the database containing the audio features and the supplementary data.  
20 Such an approach is particularly applicable to topical audio-video content presented on websites shortly before or after a date of first broadcast over a traditional channel. Additional data is also preferably used to direct the search, such as a category, channel or keywords.

## 25 BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described in more detail by way of example with reference to the accompanying drawings, in which:

**Fig. 1** shows the process for creating a database of audio signatures for  
30 subsequent searching;

**Fig. 2** shows the process for searching the database to find matching audio signatures;

**Fig. 3** shows a sliding search illustrating the matching of fingerprints; and

**Fig. 4** shows the partial matching of signatures to allow identification of  
35 clips within a programme.

## DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT OF THE INVENTION

The main embodiment that will be described is a method and system for obtaining subtitles that have already been created for one copy of audio-video content, that may be referred to as a reference copy, and providing them for another copy of that audio-video content. In the example given, the reference copy of the audio-video content may be an originally broadcast version sent over a communication channel such as terrestrial, satellite or cable. The second copy of the same audio-video content may be the whole or a portion of that content made available via the internet such as a website, dedicated online player or similar. The approach is operated by an embodiment to search for the same video clip in an archive or library of broadcast content for which subtitles have already been authored. These subtitles are then reformatted to a form suitable for that clip including error corrections and retiming.

15

This approach could also be used to retrieve any other data which is stored alongside the audio and video content, such as timing triggers for an interactive app or a list of shot changes or script information describing the scene or the characters/actors in the scene. Such data may generally be referred to as supplementary data.

20

The search is done using available data about the clip, and its audio and/or video content. The data in a web page may include the programme title and date of broadcast, or other text that suggests its approximate time and date of broadcast that can narrow the search down to a single day or programme. It may even contain the start time of the clip. The method enables the location of corresponding supplementary data to be found, even without supporting data, provided it exists in the archive/library.

25

The preferred method is to perform a comparison between the video clip audio signal and the contents of the archive/library. The audio signal is preferred as the search method because it relates directly to the main example of supplementary data, namely words in the subtitles, whereas the video may have been reused with a different soundtrack, which would contain different words to the clip being searched. In the case of data that relates to the video signal, such as audio description, the video signal would be preferred.

30

35

Refinements to this method can be applied to speed up the process and make the implementation practicable. These include methods such as using fingerprints generated from the audio and/or video content of the clip and a database of fingerprints for the archive/library. It may also be possible to use  
5 speech to text technology to generate semantic data from the clips to directly interrogate an archive of subtitles files to find a match.

The embodiment focuses matching the audio extracted from a web clip to  
10 audio stored in an archive. The example implementation does this by generating a fingerprint for the audio of all broadcast content and storing it in a database, alongside the matching subtitles, making it possible to rapidly search the database for the matching clips and provide the matching subtitles. However, this search could also use video fingerprints and any other metadata to identify when  
15 and where the clip was (or will be) broadcast. The preferred embodiment also includes various search heuristics to identify the fingerprint match and novel approaches to tidying up and retiming the subtitles to improve their quality, particularly if the subtitles were generated live.

## 20 ***System Overview***

An overview of the main components of a system embodying the invention will first be described, followed by an explanation of the process for creating a fingerprint database and then a description of the process for  
25 searching the database.

A server is used to maintain and update a database of fingerprints. The system preferably maintains such a database for all content broadcast on traditional linear broadcast for at least the main channels. As new content is  
30 broadcast on the conventional channels the matching audio is cached and a fingerprint is generated providing a small representation of the audio file (-2% of the size of the equivalent MP3 file). This fingerprint is then written to the database along with the matching subtitles that were broadcast live. Any other metadata (such as channel, programme name etc.) are also stored in the database.

35

The Server provides a search mechanism, which can be invoked by providing a URL from a website with web clips identified by a unique URL address. Once the search starts it visits the URL, downloads the video clip from the page, extracts the audio and generates a fingerprint representation for the clip. It also extracts meta-data from the web page (such as the creation time and date of the page, if it is news etc.) This meta-data is used to create a heuristic search through the database, by weighting each of the database entries based on time and date, channel and keywords found. The search algorithm then does a brute force search through the database, doing a cross correlation between the clip fingerprint and each entry in the database. Two thresholds are specified in the parameters, if the correlation error is below the first threshold it is assumed to be a highly likely match and returns an immediate result. If this is not found but the lowest error is below the second threshold once the search has completed, it is believed to be a very likely match and that result returned. A match will not be found if the clip has been edited together from a number of separate excerpts as the fingerprint will not be continuous within the reference database.

In order to allow for edited clips to be found, the algorithm firstly searches for the beginning of the web clip. Once it finds a match, it then grows the fingerprint length until it no longer matches signifying that an edit has occurred. The search then starts again from the point of the known edit to identify the subtitles for several sub clips. The retrieved subtitles may then be cleaned up such as by removing the repeated words present in live subtitle streams. Natural language processing may be used to identify the most likely start and end of the required subtitles and to correct incorrect words where the subtitler provided a correction. Also colours are corrected and subtitle files are generated. Finally a web-based editor provides full editorial control, to correct any errors that occurred in the live subtitles. It also includes other tools, such as a one-click manual retiming tool, the ability to import text from other sources, such as transcripts and original scripts (in case no matches were found). It also includes the functionality to pass the edited subtitles and audio through our phonetic retiming service (running on the same server) to provide an autonomous retiming tool. The web editor also provides warnings to indicate whether the current subtitles meet guidelines, such as showing the reading rate in words per minute for each subtitle.

### ***Database Creation***

Figure 1 shows the process by which a database of fingerprints may be created. The embodiment may be implemented as dedicated hardware or as routines in software and so Figure 1 may also be considered a functional diagram of separate modules. A variety of fingerprinting techniques may be used by which the audio features are extracted. Common to all such techniques is a reduction in the complexity of the original audio signal to the audio feature signal. In addition, the feature extraction process preferably avoids problems such as differences in amplitude of signals by deriving inherent characteristics of the signals. As previously noted, the embodiment of the invention preferably uses a feature extraction technique that produces a signature of the audio component. The signature may be variable in length so as to allow matching of an entire television programme, portion of a programme, or edited version, thereby allowing matching of an edited version of a programme against an original programme.

The signature (which may also be referred to as a fingerprint) may be any process or function which reduces the amount of information and complexity in the representation of the data which retains a low likelihood of correlation with a signature from an unrelated set of data whilst ensuring repeatable correlation when comparing similar signals. In the embodiment, the signature preferably retains a temporal resolution of around a second so that the location of the matching audio can be located to a similar resolution. A variety of different fingerprint techniques will now be discussed.

The embodiment of the invention is arranged to create a fingerprint of audio-video content whenever an update to a database of audio-video content is detected. In this way, the system may be used to automatically create a fingerprint database from continually delivered audio-video content. In particular, the system can monitor content broadcast on a range of channels and automatically record the associated fingerprints and supplementary data such as subtitles to database. At step 10 of Figure 1 an update to a database of audio-video content is detected and at step 12 a token from a recording service is loaded. A connection to the media store is established at step 14. The media store could be built from recordings of off-air broadcasts and/or a library/archive



of the broadcast files. A login to the media store is provided at step 8. At step 16, a list of channels for which fingerprints are to be created is searched and at step 18 the programmes for each channel listed. Once the appropriate channels and programmes have been identified, the process proceeds to download a file (such as WAV or MP3) representing the audio component of each programme and generates a fingerprint at step 22.

The downloaded audio file may then be removed at step 24 and a file containing the subtitle data (in a format such as SRT or TTML) downloaded at step 26. At step 28, if all programmes have been completed, the fingerprint and subtitle data is stored in a database at step 30 along with any other data available such as a programme name, date of broadcast and so on. If at step 28 it is determined that there are further programmes to be fingerprinted, the process reverts to step 20 and the next audio file downloaded.

Using the above approach, the database 30 may be created which contains a fingerprint for each portion of audio-video content (programme), the accompanying subtitles including timestamps for the presentation of the subtitles, a channel on which the programme is broadcast, a programme name, programme ID and other data that may be useful in the directed search that will be described later.

A fingerprinting technique (such as 'FPCALC') is used in order to create a repeatable set of features that represent the audio in a compressed format. Using a fingerprinting technique reduces the size of the representation, which reduces the time and processing required to complete the search and the storage for the data. For the purpose of locating the subtitle content a granularity of 1 to 2 seconds is appropriate. The fingerprinting technique used may be any of a variety of commercially available algorithms for reducing an audio track to features.

The feature extraction preferably provides a time varying output that varies with the time period of the audio-video content thereby allowing alignment between the reference database and a sample audio-video content so as to find matching parts of the content and also the matching timing for subtitles.

### ***Data Retrieval***

A Database interrogation service is implemented and used to interrogate the database in order to identify where and when a video clip was broadcast by rapidly matching its audio track to the audio stored in the database, using the stored fingerprints. The Service is invoked by providing a URL (for example: <http://www.bbc.co.uk/news/uk-england-28685666>) - All such web clips can be identified by their unique URL address.

The service then retrieves the video that is hosted on the URL along with its relevant metadata, performs a search against the database and if it is found the matching subtitles is returned as a subtitle file (such as an STL or TTML file) to match the video clip. Caching is also used to improve performance at each stage of the process. Figure 2 provides an overview of the structure of the Database interrogation architecture.

The search method shown in Figure 2 uses the fact that the request for subtitles derives from web content on a website identified by a URL and this allows data associated with the URL to be extracted and used as part of a search strategy. At step 31 the search is initialised and step 32 awaits for a first search request. On receiving a URL related to a piece of audio-video content, the programme ID of that content is retrieved and the relevant media server opened to retrieve the highest quality audio-video content available related to the URL and programme ID at step 36. At step 38 the video is downloaded as an audio/video file and at step 40 a metadata containing details extracted from the URL is stored in an exchangeable data format such as JSON or XML. At step 42 the audio component is extracted and the fingerprint generated at step 44 from the audio and is loaded to a database at step 46. This is extracted from the main database containing previously generated fingerprints and the associated subtitles as well as ancillary information such time of broadcast, programme ID, channel and so on.

The database is then ordered by an appropriate heuristic depending upon the information available from the URL for searching. For example, if the URL provides keywords related to the audio-video content, the database may be ordered by keyword. If the URL contains a date that the audio-video content was first broadcast, the database may be ordered by date and/or the channel on which the content was broadcast. A combination of such heuristics may be used to increase the likelihood that the fingerprint would be found quickly. At step 50, each fingerprint is loaded in turn and cross-correlated at step 52 with the fingerprint generated from the content at step 44. If this generates a score below a threshold, the relative offset between the database fingerprint and the fingerprint of the content being analysed may be generated. If the score is above a certain threshold, the next fingerprint is selected from the database and cross-correlation performed again so that the fingerprints are retrieved in turn in the order defined by the heuristic until a match is found. When the match is found and the time within the fingerprint calculated, this is compared to a broadcast time at step 56 and the subtitles retrieved and realigned using the offset time at step 58 and returned to the metadata file at step 62 for returning to the user that requested the subtitles.

The directed search progress will now be described in more detail with reference to some examples.

The video clip is first demuxed (using a tool such as 'ffmpeg') to extract its audio track. A fingerprint matching the audio file is then created and stored. A pass is made through the database cross correlating the clip fingerprint to each entry in the database. A confidence value is calculated for every comparison and compared to two threshold values. If the confidence value is less than the first value we are certain to have received a match and the search is stopped. If the confidence value is higher than the first threshold but less than the second value we continue searching to the end in case we find a better match. If we find a match we can extract the matching subtitles from the database and return them to the user.

A database search can be run in one of several modes as determined by the user when starting the service. Examples of operation include:

1. A single exhaustive search is performed based on a supplied URL.
- 5 2. A single exhaustive search is performed based on supplied programme identifier, this however means that there is no meta-data available to perform a heuristic based search.
3. A batch search is done on a list of URLs such as that found in an RSS feed.
- 10 4. A batch search is performed by creating a list of URLs by following links on a webpage, to a specified depth.
5. A single search is performed with parameters that are all specified within a file.
- 15 6. A rapid search by limiting the search depth into the database. This makes it possible to use a search in near real-time whilst browsing video clips, however it does mean that difficult to find video clips may be missed. Any videos not found may be tagged for later exhaustive search.
7. A debugging mode that allow search operations to be repeated.

20 The search algorithm then does a brute force search through the database, doing a cross correlation between the clip fingerprint and each entry in the database. Our cross Correlation algorithm performs a sliding dot product comparison between the sub fingerprint  $f(t)$  within each database entry fingerprint  
25 as  $g(t)$ .

$$f * g = f * (-t) * g.$$

To enable the database to be searched in near real-time, a heuristics model is used to search the most likely areas of the database first. The meta-data  
30 stored when the file is downloaded is used to create a search through the database, by weighting each of the database entries based on the time the clip was uploaded, most likely channels and keywords on the page.

Figure 3 shows the sliding dot product between a fingerprint from the  
35 database and the fingerprint of content being analysed.

As the subtitles stored in the database come from the live broadcast it may be necessary to reconstruct them from the live format into a pseudo-prepared block format. A request is made to the database for all of the subtitles stored between the start and end times of the programme. An offset is provided  
5 to ensure we retrieve subtitles from earlier and later than the clip to therefore getting too many subtitled returned at both the beginning and end and we can intelligently trim the subtitles later to ensure we have not missed the start and end point.

10 The subtitles are initially converted to a single text string, where timing is maintained on a per word basis. Each word is extracted from the live subtitle file and stored in a subtitle object alongside its broadcast timing. Due to the nature of live subtitles, once each line is completed, it is resent as a block. The system simply ignores anything that is not a single word.

15 Once a text string made up from the live subtitles is retrieved, the system will try and find the most likely start and end point for the text, this generally relies upon looking to start with the beginning of a sentence and ending on a punctuation mark. Starting at the point that the Database interrogation identified  
20 as the start time we search forwards and backwards until we find the beginning of a sentence, denoted by both a capital letter and the previous sentence ending with a punctuation mark. Similarly at the end of the subtitle file we look for the closest point at which a sentence ends and then we drop the words that fall outside of this. A parameter in the reconstruction process allows the user to  
25 specify that they want to look further forwards or backwards that forces the service to include more or less sentences.

The text string also maintains the original colours that were broadcast live. As it is likely the video has been clipped from the middle of an article the subtitles  
30 are likely to start on the wrong colour. We therefore re index the colour back into the standard order of white, yellow, green then blue by creating a lookup table.

The subtitles are reconstructed into blocks, by adhering to a maximum character per line rule. By default this is set at 38 characters as recommended in  
35 some industry guidelines, although can be overwritten by the user. Each word is

appended to a subtitle block until the next word is too long and forces the letter count over the limit. At which point a break is inserted and we start to fill the next line. Additional rules also avoid leaving orphaned words, by forcing a newline if we receive a punctuation mark over half way through the subtitle line and breaks  
5 are always forced when there is a change of speaker. Each line is then assembled into a two linen pair to complete the block.

Finally the subtitles are retimed. As the subtitles have been clipped from a subtitle file representing the entire programme each of the timings relate to the  
10 position within the programme. We subtract the found start time from each subtitle in order to realign it with the clip. We also make an assumption that short clips are always cut relatively tightly to the dialog. Therefore we have found reasonable results with scaling our subtitles to fit the video clip. This is done by pulling the first subtitle to the beginning of the clip and pushing the last subtitle so  
15 that it ends at the end of the clip. Each subtitle in between is then scaled to fit, maintaining the time ratio from broadcast.

The resulting subtitles and additional metadata are then written to the database cache in a format (such as JSON or XML). This enables the export of  
20 the subtitles to be provided in any standard format (such as TTML or STL).

The approach we have described may operate for complete audio-video programmes or portions of audio-video programmes. To do this, instead of comparing the entire signature of a programme against the entire signature of a  
25 corresponding programme in the database, portions of signatures may be compared as described below. Figure 4 shows the approach used.

Instead of searching for the entire fingerprint; the system only searches to find the first 5 seconds. If a match is found the fingerprint is grown in second  
30 increments and compared to the database until the fingerprint no longer matches. This is then interpreted as the boundary to the point that the video was edited as shown in Figure 4. Subtitles are returned for the first section and the process is repeated starting from the boundary point, until each section of the clip had been identified. The user when starting the Database interrogation service can specify  
35 a number of parameters. These may include:

1. The URL to search
2. A filename to log the output of each search
- 5 3. A list of channels to limit the search
4. A flag to output more detailed debug information during the search
5. A flag which dictates whether to output graphs for checking during the search
6. A flag to specify if the video should be played after the search is complete
- 10 7. A flag as to whether to use heuristic search
8. A flag that if set ignores the cache and always overwrites the files
9. The desired bitrate for the downloaded video
10. The maximum length of the search fingerprint
11. An offset value to ignore the beginning of clips which may have a fade
- 15 12. The immediate threshold at which point we are confident we have a correct match
13. The secondary threshold where confidence is high enough after an exhaustive search has completed
14. The resulting language - if a language is supplied all subtitles are passed
- 20 15. The maximum character width for subtitle blocks
16. The number of CPU's to use during cross correlation

**CLAIMS**

1. A method of retrieving supplementary data related to audio-video content with reference to a reference version of that content, comprising:
  - 5 - processing an audio component of the audio-video content to derive an audio signature of the audio-video content;
  - searching reference audio signatures derived from audio components of reference audio-video content by comparing the audio signature of the audio-video content to the reference audio signatures; and
  - 10 - if a match is found, retrieving supplementary data related to the reference audio-video content for which the match was found;
  - wherein the search is a directed search using data extracted from the transmission route by which the audio-video content is to be delivered to a consumer.
- 15 2. A method according to claim 1, wherein the data extracted from the transmission route comprises a time and date stamp.
3. A method according to claim 2, wherein the directed search comprises  
20 searching the reference content by comparing a date for each piece of reference audio-video content starting with a date near the date indicated by the time and date stamp.
4. A method according to claim 3, wherein the directed search comprises  
25 searching from the date indicated by the date and time stamp in both increasing and decreasing date order.
5. A method according to claim 3, wherein the date of each piece of  
30 reference audio-video content is a date of broadcast of the reference audio-video content.
6. A method according to any preceding claim, wherein the data extracted  
35 from the transmission route comprises one or more of a programme name, programme ID or channel and the directed search initially limits the search based on the extracted data.



7. A method according to any preceding claim, wherein the audio video content is from a downloadable source and the retrieving supplementary data comprises making the supplementary data available on the downloadable source.
- 5 8. A method according to any preceding claim, wherein the transmission route comprises download from the Internet and the data extracted from the transmission route is from a URL of the audio-video content.
9. A method according any preceding claim, wherein the audio signature is a  
10 time varying function and wherein comparing the audio signature of the audio-video content to the reference audio signatures comprises comparing a portion of the time varying function to the reference audio signatures.
10. A method according to claim 9, wherein the comparing comprises a  
15 sliding comparison of the portion of the signature relative to each reference signature.
11. A method according to claim 10, wherein the sliding comparison is  
20 arranged to establish both a match and relative time offset of the portion of the signature and the reference audio signature.
12. A method according to claim 9, 10 or 11, further comprising varying the duration of the portion of the signature.
- 25 13. A method according to claim 12, wherein varying the duration comprises increasing the duration after a corresponding reference signature has been found until the portion of the signature and reference signature no longer match.
14. A method according to claim 13, further comprising selecting a second  
30 portion of the signature after the first portion and comparing the second portion to the reference signature for which a match has been found.

15. A method of any of claims 9 to 13, comprising selecting a portion of the signature, comparing the portion to the reference signatures to establish a match and the repeated steps of:

- 5 - increasing the selected portion of the signature until the point that the portion would no longer match the previously matched reference signature;
- selecting a new selected portion adjacent the previously selected portion;

10 whereby multiple portions of the signature are matched to portions of one of the reference signatures.

16 A method according to claim 15, comprising producing a shot change list from the start and end times indicated by the matching portions.

15 17. A method according to any preceding claim, wherein the searching comprises determining in turn whether a function of the audio signature and each of the reference audio signatures meets a threshold test.

20 18. A method according to claim 17, wherein the threshold test is met and the search stopped if the function output is below a first threshold.

25 19. A method according to claim 17 or 18, wherein the threshold test is met at the end of the search if the function output is below a second threshold for one of the reference signatures.

20. A method according to any preceding claim, wherein the supplementary data comprises subtitles.

30 21. A method according to claim 20, further comprising re-timing the subtitles using information indicating the timing offset between the signature and matching reference signature.

35 22. A method according to any preceding claim, further comprising monitoring the storing of content broadcast on a broadcast network and adding signatures of such content to a database as the content is broadcast or stored.

23. A method according to any preceding claim, wherein searching the reference audio signatures comprises searching a database containing the reference audio signatures.

5

24. A method according to any preceding claim, wherein processing the audio component of the audio-video content to derive an audio signature comprises using feature extraction or a mathematical transform.

10 25. A system for retrieving supplementary data related to audio-video content with reference to a reference version of that content, comprising:

- means for processing an audio component of the audio-video content to derive an audio signature of the audio-video content;

15 - means for searching reference audio signatures derived from audio components of reference audio-video content by comparing the audio signature of the audio-video content to the reference audio signatures; and

- means, if a match is found, for retrieving supplementary data related to the reference audio-video content for which the match was found;

20 - wherein the search is a directed search using data extracted from the transmission route by which the audio-video content is to be delivered to a consumer.

26. A system according to claim 25, wherein the data extracted from the transmission route comprises a time and date stamp.

25

27. A system according to claim 26, wherein the directed search comprises searching the reference content by comparing a date for each piece of reference audio-video content starting with a date near the date indicated by the time and date stamp.

30

28. A system according to claim 26, wherein the directed search comprises searching from the date indicated by the date and time stamp in both increasing and decreasing date order.

29. A system according to claim 27, wherein the date of each piece of reference audio-video content is a date of broadcast of the reference audio-video content.
- 5 30. A system according to any of claims 25 to 29, wherein the data extracted from the transmission route comprises one or more of a programme name, programme ID or channel and the directed search initially limits the search based on the extracted data.
- 10 31. A system according to any of claims 25 to 30, wherein the audio video content is from a downloadable source and the retrieving supplementary data comprises making the supplementary data available on the downloadable source.
- 15 32. A system according to any of claims 25 to 31, wherein the transmission route comprises download from the Internet and the data extracted from the transmission route is from a URL of the audio-video content.
- 20 33. A system according to any of claims 25 to 32, wherein the audio signature is a time varying function and wherein comparing the audio signature of the audio-video content to the reference audio signatures comprises comparing a portion of the time varying function to the reference audio signatures.
- 25 34. A system according to claim 33, wherein the comparing comprises a sliding comparison of the portion of the signature relative to each reference signature.
- 30 35. A system according to claim 34, wherein the sliding comparison is arranged to establish both a match and relative time offset of the portion of the signature and the reference audio signature.
36. A system according to claim 33, 34 or 35, further comprising varying the duration of the portion of the signature.

37. A system according to claim 36, wherein varying the duration comprises increasing the duration after a corresponding reference signature has been found until the portion of the signature and reference signature no longer match.
- 5 38. A system according to claim 37, further comprising selecting a second portion of the signature after the first portion and comparing the second portion to the reference signature for which a match has been found.
39. A system of any of claims 33 to 38, comprising selecting a portion of the  
10 signature, comparing the portion to the reference signatures to establish a match and the repeated steps of:
- increasing the selected portion of the signature until the point that the portion would no longer match the previously matched reference signature;
  - 15 - selecting a new selected portion adjacent the previously selected portion;
- whereby multiple portions of the signature are matched to portions of one of the reference signatures.
- 20 40 A system according to claim 39, comprising producing a shot change list from the start and end times indicated by the matching portions.
41. A system according to any of claims 25 to 40, wherein the searching comprises determining in turn whether a function of the audio signature and each  
25 of the reference audio signatures meets a threshold test.
42. A system according to claim 41, wherein the threshold test is met and the search stopped if the function output is below a first threshold.
- 30 43. A system according to claim 41 or 42, wherein the threshold test is met at the end of the search if the function output is below a second threshold for one of the reference signatures.
44. A system according to any of claims 25 to 43, wherein the supplementary  
35 data comprises subtitles.

45. A system according to claim 44, further comprising re-timing the subtitles using information indicating the timing offset between the signature and matching reference signature.

5

46. A system according to any of claims 25 to 45, further comprising monitoring the storing of content broadcast on a broadcast network and adding signatures of such content to a database as the content is broadcast or stored.

10 47. A system according to any of claims 25 to 46, wherein searching the reference audio signatures comprises searching a database containing the reference audio signatures.

15 48. A system according to any of claims 25 to 47, wherein processing the audio component of the audio-video content to derive an audio signature comprises using feature extraction or a mathematical transform.

49. A method according to any of claims 1 to 24, comprising providing the supplementary data to a client device.

20

50. A method according to claim 49, further comprising asserting the supplementary data on a client device by one or more of displaying data or providing audio output of the data.

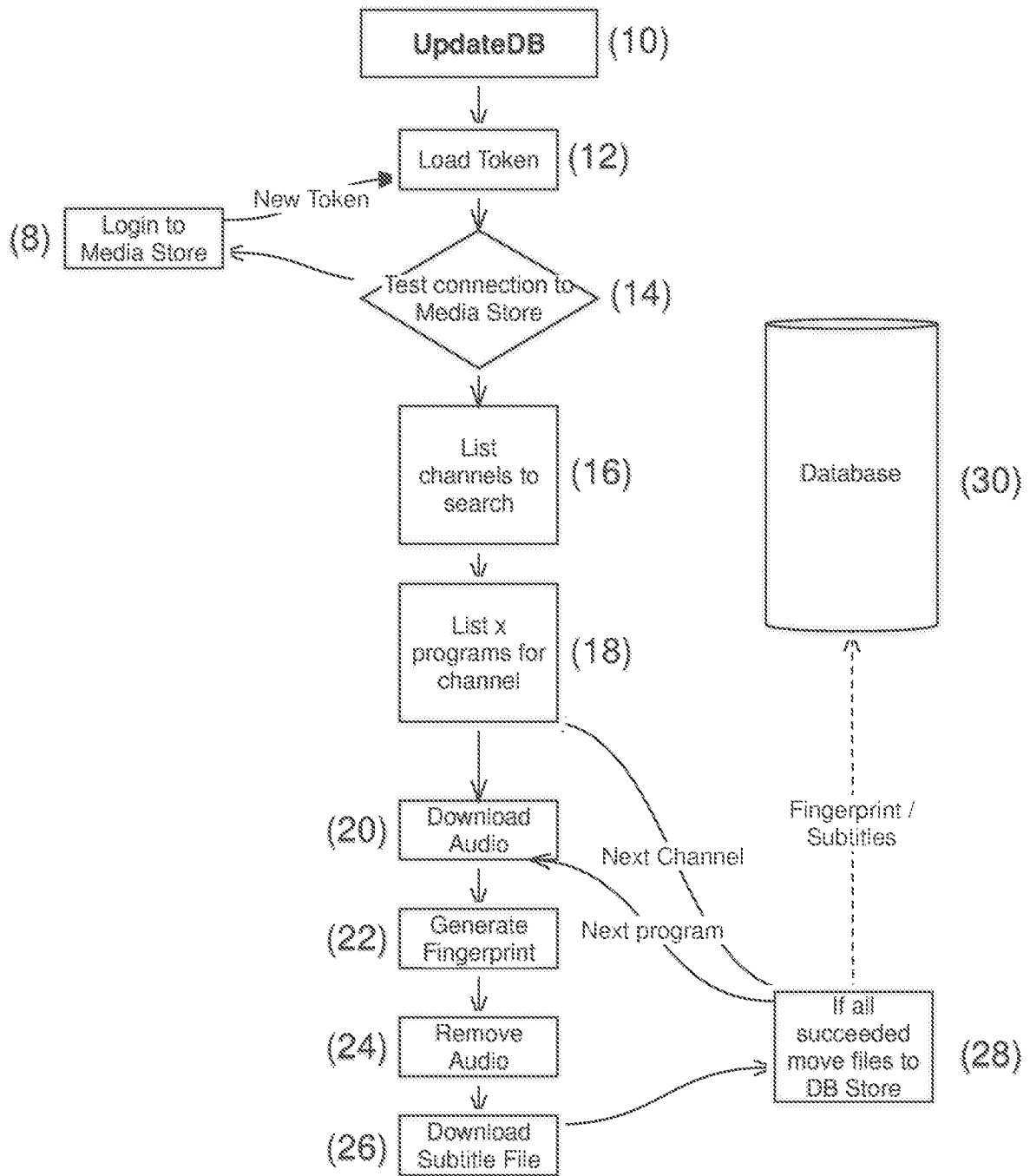


Figure 1

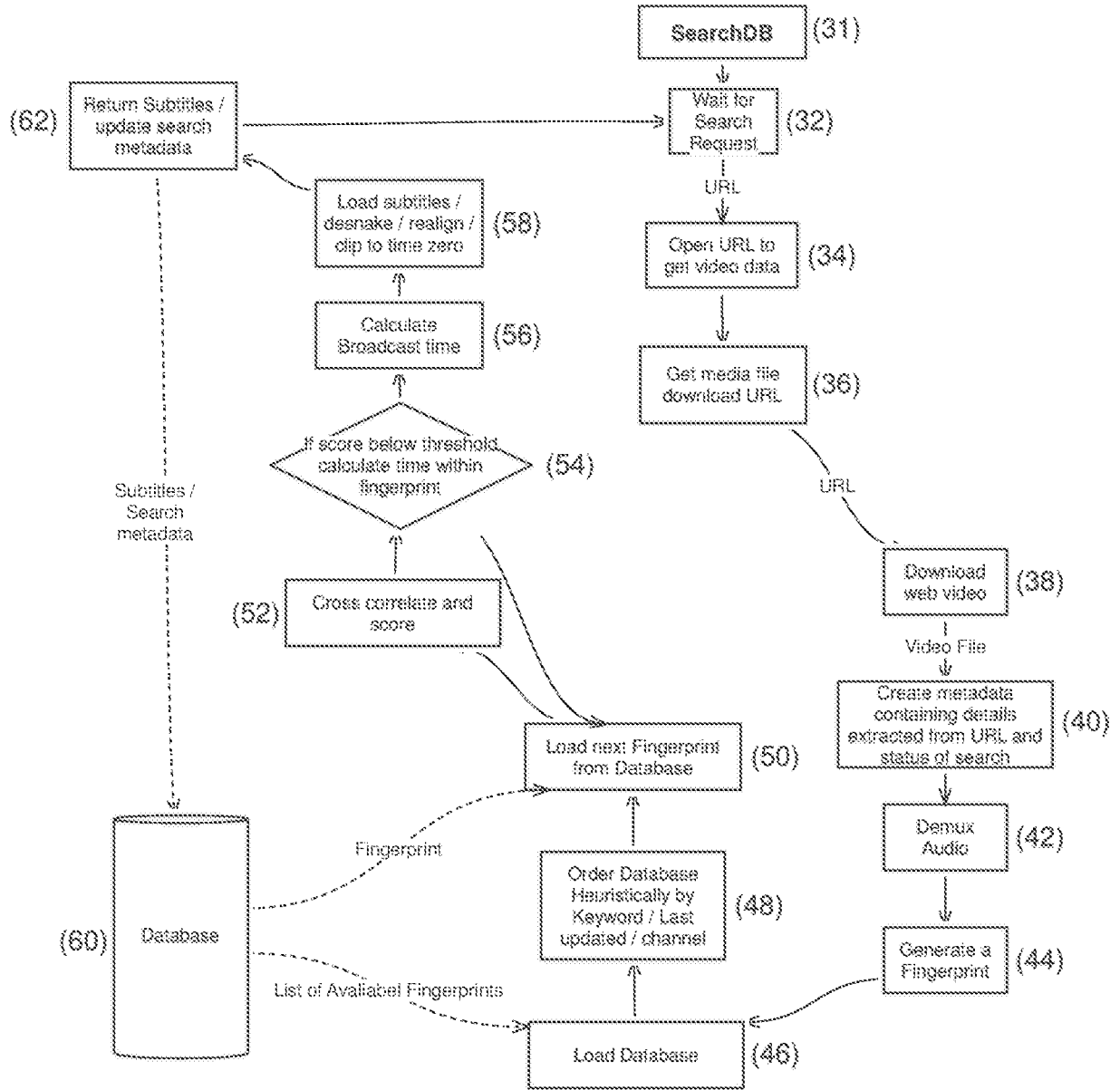


Figure 2



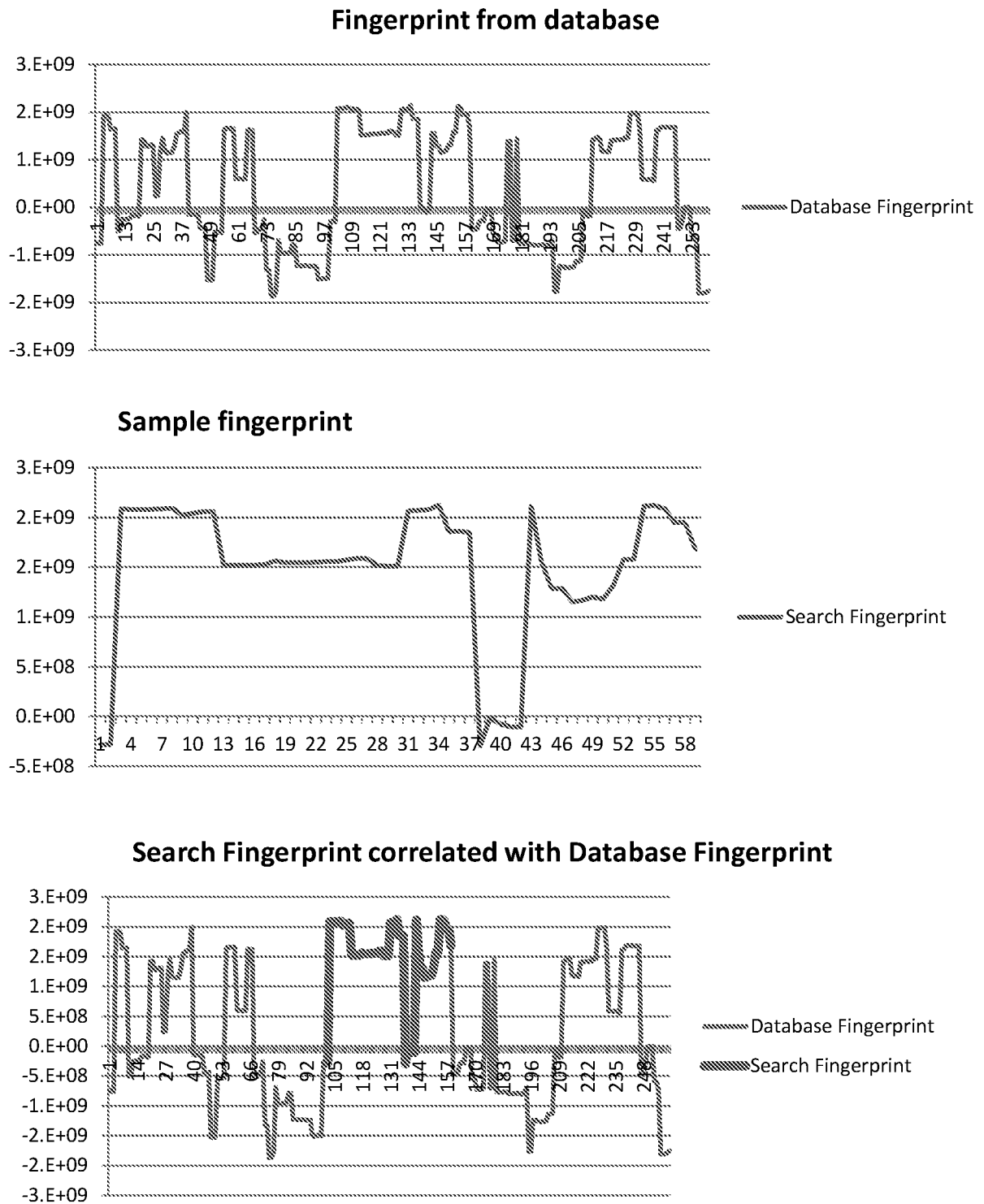


Figure 3

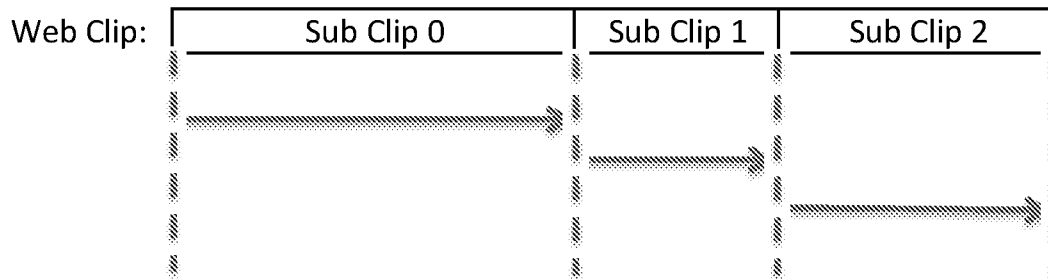


Figure 4

INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2015/053026

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G06F17/30  
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)  
G06F G06K

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	US 2011/247042 AI (MALLINSON DOMINIC S [US]) 6 October 2011 (2011-10-06) paragraph [0003] - paragraph [0034] -----	1-50
X	US 2014/067863 AI (BI LOBR0V SERGIY [US]) 6 March 2014 (2014-03-06) paragraph [0001] - paragraph [0042] -----	1-50
X	US 2006/013451 AI (HAITSMA JAAP A [NL]) 19 January 2006 (2006-01-19) page 1, line 2 - page 10, line 17 -----	1-50
X	EP 2 388 721 AI (GOOGLE INC [US]) 23 November 2011 (2011-11-23) paragraph [0001] - paragraph [0004] paragraph [0024] - paragraph [0047] ----- -/- .	1-50

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

15 December 2015

Date of mailing of the international search report

22/12/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

Chi nzer, Azzurra

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2015/053026

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012/124623 A1 (PERKINS ROBERT M [GB] ) 17 May 2012 (2012-05-17) paragraph [0002] paragraph [0014] - paragraph [0049] -----	1-50
A	US 2013/325888 A1 (ONEPPO MICHAEL [FR] ET AL) 5 December 2013 (2013-12-05) paragraph [0008] - paragraph [0036] -----	1-50
A	US 2011/276157 A1 (WANG AVERY LI-CHUN [US] ET AL) 10 November 2011 (2011-11-10) paragraph [0027] - paragraph [0169] -----	1-50
A	US 2014/059065 A1 (CORTES RICARDO D [US] ET AL) 27 February 2014 (2014-02-27) paragraph [0006] - paragraph [0046] -----	1-50

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No <b>PCT/GB2015/053026</b>
--

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011247042	AI	06-10 -2011	EP 2553652 A2 06-02-2013
			JP 5651231 B2 07-01-2015
			JP 2013529325 A 18-07-2013
			US 2011247042 AI 06-10-2011
			Wo 2011123718 A2 06-10-2011
us 2014067863	AI	06-03 -2014	US 2011116719 AI 19-05-2011
			US 2014067863 AI 06-03-2014
us 2006013451	AI	19-01 -2006	AU 2003264774 AI 25-05-2004
			CN 1708758 A 14-12-2005
			EP 1561176 A2 10-08-2005
			JP 2006506659 A 23-02-2006
			KR 20050061594 A 22-06-2005
			US 2006013451 AI 19-01-2006
			Wo 2004040475 A2 13-05-2004
EP 2388721	AI	23-11 -2011	EP 2388721 AI 23-11-2011
			JP 2011243204 A 01-12-2011
			US 2011289098 AI 24-11-2011
			US 2014195557 AI 10-07-2014
US 2012124623	AI	17-05 -2012	NONE
US 2013325888	AI	05-12 -2013	US 2013325609 AI 05-12-2013
			US 2013325807 AI 05-12-2013
			US 2013325888 AI 05-12-2013
			US 2013326041 AI 05-12-2013
us 2011276157	AI	10-11 -2011	CA 2798093 AI 10-11-2011
			CN 102959543 A 06-03-2013
			EP 2567332 AI 13-03-2013
			KR 20130029082 A 21-03-2013
			KR 20150095957 A 21-08-2015
			US 2011276157 AI 10-11-2011
			wo 2011140269 AI 10-11-2011
us 2014059065	AI	27-02 -2014	NONE