



- (51) International Patent Classification:
G06F 17/30 (2006.01)
- (21) International Application Number:
PCT/GB2015/051380
- (22) International Filing Date:
11 May 2015 (11.05.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1408257.2 9 May 2014 (09.05.2014) GB
- (71) Applicant: OMNIFONE LTD [GB/GB]; The Old School,
50 Brook Green, London W6 7BJ (GB).
- (72) Inventors: WHITE, Matthew; c/o Omnifone Ltd, The
Old School, 50 Brook Green, London W6 7BJ (GB).
SANT, Philip; c/o Omnifone Ltd, The Old School, 50
Brook Green, London W6 7BJ (GB).
- (74) Agent: ORIGIN LIMITED; Twisden Works, Twisden
Road, London NW5 1DN (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, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

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

(54) Title: METHODS, SYSTEMS AND COMPUTER PROGRAM PRODUCTS FOR IDENTIFYING COMMONALITIES OF RHYTHM BETWEEN DISPARATE MUSICAL TRACKS AND USING THAT INFORMATION TO MAKE MUSIC RECOMMENDATIONS

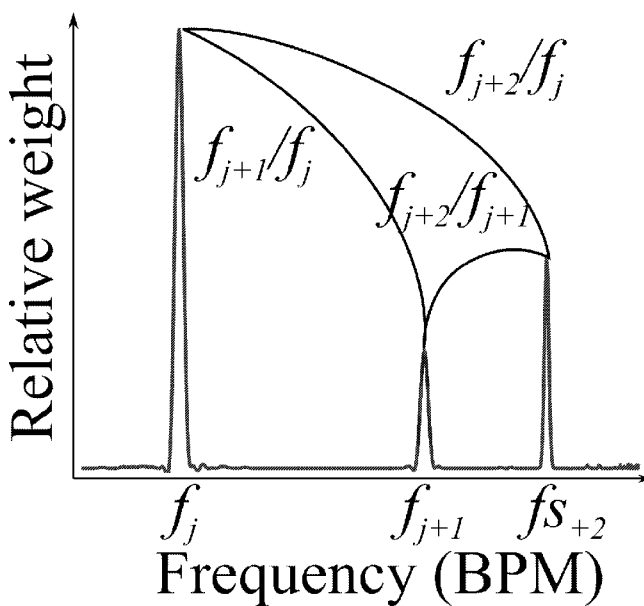


FIGURE 2

(57) Abstract: There is provided a computer program product embodied on a non-transitory storage medium or on a cellular mobile telephone device or on another hardware device, the computer program product operable to automatically generate recommendations for digital media content, in which that system is adapted to automatically generate recommendations of digital media content for an end user by (a) analysing a first item or set of digital media content and its associated metadata to calculate metric-based metadata for that digital media content and (b) using that analysis to identify other digital media content which have a specifically-defined relationship with that first digital media content; and (c) provide, to the end user, recommendations of one or more items of that digital media content which were identified in the previous step.

WO 2015/170126 A1

METHODS, SYSTEMS AND COMPUTER PROGRAM PRODUCTS FOR IDENTIFYING COMMONALITIES OF RHYTHM BETWEEN DISPARATE MUSICAL TRACKS AND USING THAT INFORMATION TO MAKE MUSIC RECOMMENDATIONS

5

BACKGROUND OF THE INVENTION

1. Field of the Invention

10 The field of the invention relates to methods, systems and computer program products for identifying commonalities of rhythm between disparate musical tracks and using that information to make music recommendations.

2. Technical Background

15

A common problem in the field of music provision is the problem of which music tracks to recommend, both in terms of making explicit recommendations to individual consumers based on their existing musical taste and in terms of selecting a musical track to be played after any given musical track has concluded.

20

One approach to that problem which has been tried at various times is to analyse the current musical track (or the corpus of music which a given end user is known to enjoy) and identify common features with other tracks which fit well with the first track.

25 However, the ways in which similarities between tracks are able to be automatically identified have historically been limited. The present invention teaches a method for more efficiently and reliably identifying commonalities of rhythm between disparate musical tracks and using that information to make recommendations accordingly.

30

SUMMARY OF THE INVENTION

The present invention provides a method for analysing audio media to extract metadata describing the rhythm, tempo and other descriptive metadata about the audio track and
5 utilising that metadata to assist in cross-matching tracks, providing recommendations, perform additional Digital Signal Processing of tracks and for sundry other purposes as disclosed in this document.

The present invention teaches a method for efficiently and reliably identifying
10 commonalities of rhythm between disparate musical tracks and using that information to make music recommendations accordingly.

Key Aspects of the Present Invention

15 1. A computer program product embodied on a non-transitory storage medium or on a cellular mobile telephone device or on another hardware device, the computer program product operable to automatically generate recommendations for digital media content, in which that system is adapted to automatically generate recommendations of
20 digital media content for an end user by (a) analysing a first item or set of digital media content and its associated metadata to calculate metric-based metadata for that digital media content and (b) using that analysis to identify other digital media content which have a specifically-defined relationship with that first digital media content; and (c)
provide, to the end user, recommendations of one or more items of that digital media content which were identified in the previous step.

25

2. The method of point 1 where a first digital media content is a track which is currently being listened to by an end user.

30 3. The method of point 1 where a first digital media content consists of a set of multiple tracks which are (i) analysed individually to calculate metric-based metadata and (ii) that calculated metric-based metadata is aggregated and subsequently treated, for the purposes of subsequent steps in point 1, as if it derives from a single item of digital media content.

4. The method of point 3 where the set of multiple tracks comprises the set of digital media content files which a user enjoys, whether determined by observing that user's listening preferences or by the user specifically indicating his digital media content preferences or by any other means.

5

5. The method of any preceding point where the metric-based metadata comprises one or more of: the rhythmic value or values of a track, the time signature or time signatures of a track; the rhythmic simplicity or complexity of a track; the tempo or cadence drift of a track; the rhythm salience and rhythmicity; confidence values for any
10 of the listed derived metrics; any other calculated metrics for a track.

6. The method of any preceding Claim where the specifically defined relationship with the first digital media content is that the identified other track or tracks share similar or identical calculated metric-based metadata with that first digital media content.

15

7. The method of point 6 where the calculated metric-based metadata for two or more tracks is defined as being similar if the two sets of metadata have values for a given metric which lie within a defined range from the metric-based metadata values calculated from a first item or set of digital media content.

20

8. The method of point 7 where the defined range may be defined separately for each metric-based metadata value which is utilised.

9. The method of any preceding point where the metric-based metadata is
25 calculated in advance and stored in a database for comparison, search and retrieval purposes.

10. The method of any preceding point where the metric-based metadata is calculated as and when needed and discarded after use.

30

11. The method of any preceding point where the metric-based metadata is first calculated as and when needed and is then stored in a database for use subsequently.

12. The method of any preceding point where the one or more items of digital content recommended to the end user form a subset of those digital media content items identified as having similar or identical metric-based metadata values to the first item(s) of digital media content.

5

13. The method of point 12 where the subset of digital media content items is determined by filtering the list using additional criteria, including but not limited to the artist, album, genre, musical taste, current date, time of playback, date/time of most recent playback, user demographic data or by using any other additional criteria.

10

14. A computer based system adapted to perform the method of any preceding point.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other aspects of the invention will now be described, by way of example only, with reference to the following Figures, in which:

5

Figure 1 shows a Tree representation for metrical hierarchy. (a) A simple duple hierarchy dividing the lower level into two groups of two (as with a ⁴ time signature). (b) A triple duple hierarchy dividing the lower level into three groups of two (e.g. ³ ⁴ time signature). (c) A compound-duple hierarchy dividing the lower level into two groups of three (e.g. ⁶ ⁸ time signature).

10

Figure 2 shows an example of a special case with two metrical levels that can't coexist in the same hierarchy, hence requiring the creation of two parallel hierarchies: usage of triplets and eight notes at the same time (a.k.a 3:2 polyrhythm) represented by f_{j+2} and

15 f_{j+1} respectively with f_j the quarter note level. Then $\frac{f_{j+1}}{f_j} = 2$ and $\frac{f_{j+2}}{f_j} = 3$ whereas $\frac{f_{j+2}}{f_{j+1}} = \frac{3}{2} \notin \mathbf{N}$.

Figure 3 shows elementary patterns and the sequence of amplitudes associated to both their plain and accented interpretations.

20

Figure 4 shows Spectrum $S(f)$ with extracted metrical hierarchy H , metrical levels rates F and relative weights vector W for (a) Billie Jean by Michael Jackson, (b) Home at last by Steely Dan and (c) Vivaldi's Violin Concerto in G minor (Summer), RV 315, movement III (Presto), as performed on the 1989 recording by Nigel Kennedy.

25

DETAILED DESCRIPTION

Definitions

5 For convenience, and to avoid needless repetition, the terms “music” and “media content” in this document are to be taken to encompass all “media content” which is in digital form or which it is possible to convert to digital form - including but not limited to books, magazines, newspapers and other periodicals, video in the form of digital video, motion pictures, television shows (as series, as seasons and as individual episodes),
10 computer games and other interactive media, images (photographic or otherwise) and music.

Similarly, the term “track” indicates a specific item of media content, whether that be a song, a television show, an eBook or portion thereof, a computer game or any other
15 discreet item of media content.

The terms “playlist”, “timeline” and “album” are used interchangeably to indicate collections of “tracks” and/or interstitials which have been conjoined together such that they may be treated as a single entity for the purposes of analysis or recommendation.
20

The terms “digital media catalogue”, “digital music catalogue”, “media catalogue” and “catalogue” are used interchangeably to indicate a collection of tracks and/or albums to which a user may be allowed access for listening purposes. The digital media catalogue may aggregate both digital media files and their associated metadata or, in another
25 example embodiment, the digital media and metadata may be delivered from multiple such catalogues. There is no implication that only one such catalogue exists, and the term encompasses access to multiple separate catalogues simultaneously, whether consecutively, concurrently or by aggregation. The actual catalogue utilised by any given operation may be fixed or may vary over time and/or according to the location or access
30 rights of a particular device or end-user.

The abbreviation “DRM” is used to refer to a “Digital Rights Management” system or mechanism used to grant access rights to a digital media file.

The verbs “to listen”, “to view” and “to play” are to be taken as encompassing any interaction between a human and media content, whether that be listening to audio content, watching video or image content, reading books or other textual content, playing a computer game, interacting with interactive media content or some
5 combination of such activities. In addition, the playing of digital media content is to be further construed as occurring where some or all of one digital media content file is incorporated within another digital media content file such that playing the latter would necessitate the former also being played.

10 The terms “user”, “consumer”, “end user” and “individual” are used interchangeably to refer to the person, or group of people making use of the facilities provided by the interface. In all cases, the masculine includes the feminine and vice versa.

The terms “device” and “media player” are used interchangeably to refer to any
15 computational device which is capable of playing digital media content, including but not limited to MP3 players, television sets, home entertainment system, home computer systems, mobile computing devices, games consoles, handheld games consoles, IVEs or other vehicular-based media players or any other applicable device or software media player on such a device. Something essentially capable of playback of media.

20 The term “DSP” (“Digital Signal Processing”) refers to any computational processing of digital media content in order to extract additional metadata from that content. Such calculated metadata may take a variety of forms, including deriving the tempo of a musical track or identifying one or more spots within the digital media file which are
25 gauged to be representative of that content as a whole.

The term “hook” is used to refer to one or more portions of a digital media file which have been identified, whether via DSP or manually or by some other method, as being representative of the content as a whole. For example, a movie trailer consists of a series
30 of one or more “hooks” from the movie while particularly apposite riffs or lines from a musical track serve a similar identifying purpose.

The terms “partial playback” and “extract” are used interchangeably to refer to the playback of an item of digital media content where that item is not played back in its entirety or is incorporated into another item of digital media content.

- 5 The abbreviation “BPM” is used to denote the count of the number of Beats Per Minute in an audio track.

The abbreviation “FFT” is used for the mathematical term “Fast Fourier Transform”.

- 10 The abbreviation “MIR” is used to denote “Music Information Retrieval”, a field of study whereby audio signals are analysed to extract metadata such as, and most relevantly to the present invention, the extraction of rhythm metadata from audio tracks.

- 15 The terms “DJ” and “Disk Jockey” are used interchangeably to refer to individuals who compile and/or play collections, or “mix tapes”, of musical tracks.

Tempo

- Tempo estimation has been an important aspect of Music Information Retrieval (MIR) research since its inception. A large body of research has been devoted to the task. Tempo is commonly considered as a fundamental property of the music, which certainly explains why it has been one of the first MIR tasks to be investigated.
- 20

- Tempo is often associated to our perception of the speed or pace of the music or, in other words, how fast we feel the music is. Its relevance for music databases navigation, recommendation systems or automatic curation is then obvious.
- 25

- However, the definition of the term “tempo” itself is not consistent in the related MIR literature thus leaving room for ambiguity and fuzziness in the discussions.
- 30

Metronomic indications vs. tempo

Published academic papers almost exclusively represent tempo as a BPM (Beats per Minute) value, which is effectively a frequency measurement. The implicit assumption

made there is that tempo can be represented by a single rate (frequency). Given that tempo is commonly associated with speed or pace of music, that is to say our sense of how fast the music is is governed by one rate.

5 A nuance between metronomic indication and tempo need to be made here as are generally represented in the form of a BPM value. A metronomic indication is written on a music score and associates a note value with an absolute frequency measurement, thereby providing an absolute point of reference for speed of execution - usually using a metronome. This has several implications. First, a metronomic indication makes sense
10 only in regard to the score it is attached to. Secondly, the “correct” metronomic indication for a given score is not necessarily unique. Numerous examples of this can be found in music history, Chopin being an example of a composer who used different notations on different editions of his scores.

15 A metronomic indication is relevant for performers, not necessarily for listeners. The score is a cue to the performers so that they can deliver the musical feel intended by the composer (including the sense of pace) but it does not mean a single element of the score (the metronomic indication) bears this information. In that sense, Italian tempo indications such as *presto*, *allegro*, *lento* provide information which would be a lot more
20 meaningful to listeners than just a metronomic indication since they describe what the music should sound like rather than the prescription of a technical performance element. Of course music notation is subject to conventions that influence the performance of musicians. For example, a cut time score usually bears a connotation for fast music. But the point being made here is that although metronomic indication and sense of speed or
25 pace are not uncorrelated, there isn't a strict equivalence between them.

Tactus vs. tempo

With the design of tempo estimation algorithms from audio comes the need to evaluate
30 them. This is commonly achieved by comparing the output of the algorithm with some ground truth data. There are two common possibility for ground truth tempo known as “notated tempo” and “perceived tempo”.

The “notated tempo” is obtained from the score and effectively refers to the metronomic indication. The access to, or even the existence of, a music score with a metronomic indication is not easy, and is furthermore subject to copyright law. For that reason, unless working on symbolic data, this source of ground truth tends not to be commonly used.

- 5 The “perceived tempo” ground truth data is typically collected by asking people to tap along a music piece, record the hits and measure the tapping rate. This is what the ground truth data used for the MIREX audio tempo estimation task consists of.

In more rigorous terms, the rate at which listeners tap along to music is called ‘tactus’.

- 10 That type of evaluation has become a standard in tempo estimation evaluation and is based on the implicit assumption that tactus equals tempo.

However, that assumption is challengeable.

15 **Going beyond tempo**

- The present invention provides a descriptor that is more closely related to the temporal feel of music than a single rate (whether it is called tempo or tactus). Such a descriptor has direct applications in the context of today’s music distribution and consumption such as content-based recommendation for streaming and download services, automatic radio, curation assistance etc.
- 20

- The metrical structure appears to play an important role in perception of speed: differences in surface rhythm and metrical structure do interfere with judgments of tempo, even if two passages have the same beat rate. Consequently, an important focus of the present invention is on metrical structure extraction from audio. It is also safe to assume that the metrical structure influences more than just the sense of speed. For instance, music in a compound meter feels different from a simple meter.
- 25

30 **Detailed Description**

Rhythm patterns in music are complex signals. In music which can be considered to have a sense of meter, the rhythm will have multiple related periodicities forming a

hierarchical structure; the common musical terms for which include bars, beats and beat subdivisions.

5 The present invention introduces a mid-level feature which captures the weighting of salient pulse rates present in the music, which correspond to the most likely metrical hierarchy. Feature extraction is achieved by computing a novelty curve from the audio then processing this in time and frequency to produce a rhythmic spectrogram or rhythmogram. From this, we derive multiple possible candidates for a metrical hierarchy that fits the rhythmic content the best. That feature provides a meaningful description of
10 the rhythm content without applying specific labels, such as tempo or tactus, to any particular level.

Overview

15 The temporal occurrences of musical events (e.g. notes) in a musical piece can be, and often are, organised such that periodic patterns emerge. These can be decomposed into different levels of periodicity or rates that together form a metrical hierarchy. Rhythm-related research has been an important area in Music Information Retrieval (MIR) since the discipline's earliest days. Audio tempo induction in particular has seen a lot of interest
20 in the academic community. The aim of such systems is to extract a single-value descriptor that tells us something about the speed or pace of the music.

Tempo is a minimal descriptor of rhythm content that represents one of the levels in the metrical hierarchy and, alone, does not provide information about broader structure.

25 Induction of more than one metrical level is sometimes used as an intermediate step in beat tracking or onset detection methods. With the exception of Degara's method, the metrical levels extracted in the aforementioned algorithms are labelled with respect to their musical or perceptive role. Gouyon *et. al* proposed a system differentiating between
30 duple and triple meter implying the estimation of two metrical levels. Further back, Temperley and Sleator designed an algorithm to analyse meter and harmony in symbolic music data, which allowed the presence of more than three metrical levels.

The present invention introduces a method to extract a complete metrical hierarchy from audio, representing it in a compact feature without applying semantic labels to any particular level. The extraction is performed using standard time-frequency analysis and a peak picking strategy that incorporates hierarchical structure constraints. The energy associated with each metrical level in our feature corresponds to the density of events of that rhythmic subdivision in the music signal.

Appendix A of this document presents, in some detail, the mathematical algorithms employed by an example of the present invention to identify a consistent mathematical signature for the rhythmic structure of a given piece of audio content.

Track Similarity

The present invention requires that the pool of audio content from which recommendations are to be drawn first be analysed using the algorithmic techniques detailed for example in Appendix A.

Once that step is completed, whether carried out on the fly or, in the preferred embodiment, carried out in advance, one or more of the following items of metadata will be available for each musical track which is so analysed:

- A metric which represents the rhythmic value or values of that piece of audio
- A metric which represents the time signature or signatures of that piece of audio
- A metric which represents the rhythmic simplicity or complexity of that piece of audio
- Values to establish the tempo or cadence drift of that piece of audio
- A metric representing the rhythm salience and rhythmicity of that piece of audio
- A confidence value for the extracted rhythm metric.

In the preferred embodiment of the present invention, all such values are calculated in advance for all musical tracks from which recommendations are to be drawn. The results of such calculations are, in the preferred embodiment, stored in a database via which those values may be searched.

When recommending additional music based on a given musical track (the “seed track”), the present invention takes the following steps:

1. Calculate (or retrieve pre-calculated values from a database) the desired metadata for the given track;
- 5 2. Locate other musical tracks which share similar or identical metrics within the database of such calculated metadata;
3. Locate other musical tracks whose metrics have a specific relationship to the seed. This relationship is not necessarily similarity and can be specified in the embodiment of the present invention. In a further
10 example embodiment of the present invention, multiple types of inter-track compatibility (defined by the given relationship, and not necessarily equal to similarity) may be employed.

In the preferred embodiment of the present invention, such calculated metadata is
15 calculated in advance and stored to assist in retrieving matching tracks which show similarities of rhythm, time signature(s), rhythmic complexity and/or cadence drift. In another embodiment of the present invention, tracks are deemed to be similar if one such calculated metadata metric matches exactly. In other example embodiment, tracks are deemed to be similar if one or more of such metrics lie within a defined range from
20 that of the seed track. In the preferred embodiment, the extent of such a range is configurable. In the preferred embodiment, tracks are deemed to be similar if a defined number of such metadata metrics match exactly or lie within a defined range. In the preferred embodiment, the defined range is separately configurable for each such calculated metadata metric.

25

Uses in Recommendations

When one or more tracks are identified, using the method disclosed above, as being
“similar” to the seed track, those tracks may, in the preferred embodiment, be
30 recommended to the end user whose seed track formed the basis for those recommendations.

In the preferred embodiment, the list of recommended tracks is further narrowed down by reference to other criteria defined within the computer program or service within which the end user is listening to music.

- 5 In another example embodiment, those tracks known (by inference or by direct selection by the end user or, in the preferred embodiment, by both methods) to be liked by a given end user have their calculated metadata metrics compared to attempt to divine a commonality of preferred music for that user. The aggregated metrics of those tracks is then, in the preferred embodiment, directly used to locate similar tracks to be
10 recommended for that user, the search for similar tracks being performed in the same manner as disclosed for the “seed track” above, though utilising the aggregated metrics rather than those derived from the seed track to locate similar tracks.

- In a further example embodiment, the calculated metadata metrics for a seed track is
15 used to locate similar tracks for playback as a playlist of music which is related in style, such as by having similar rhythmic structures.

Additional Use Cases

- 20 In addition to the preferred embodiment of the present invention disclosed above, the present invention may be further embodied to assist with both the provision of music recommendations and with other scenarios, including the following:
- Segmentation of an audio track based on the varying rhythmic values detected in that piece of audio
 - 25 • Similarity-based computation based on rhythmic content and/or segmentation, to identify audio tracks which are similar to one another, possibly for the purpose of recommending music based on an individual’s existing taste in music.
 - Calculate difference values for pitch and tempo changes for beat matching of different audio tracks
 - 30 • Identification of machine-generated tempo clocks vs live clocks
 - Identification of hook points in a given audio track
 - Identification of tempo changes in a given audio track
 - Use of the start and end tempos for sequencing of media

- A combination of all elements disclosed herein alongside existing metadata to improve or expand on recommendations, sequencers or AI media suggestions i.e. using time of day, genre, to identify the rhythmical style of music via speed, frequency or complexity in order to recommend music at different times of day.
- 5 • Use the present invention to inform a beat tracking system, so that the periodicities of the beats that should be found are already known.
- Measurement of rhythm similarity allows the present invention to be utilised to:
 - Recommend music which shares a similar rhythm to a given track
 - Assisted curation of music library (meaning that propositions are made by
10 a computer to a human user). For example, this usage would enable the pre-selection of those tracks that would merge nicely for a given DJ.
 - Estimate the time stretching factor necessary to match two songs that don't exactly match according to the previous point.
 - Evaluation of listeners tastes: do they always listen to something that has
15 the same signature in terms of feature values (in other word is the statistical inter track distance low?)?
 - Similarly, are there patterns in mix tapes, set lists or playlists made by musicians, DJs, etc... This would possibly give the ability to spot the "signature" of DJs.
- 20 • Reverse-engineer user or DJ signature: generate a playlist that follows the pattern detected for a DJ or user.
- An automatic DJ software can be designed by determining the periodicities of beats of various musical tracks and then generating playlists comprising tracks which have similar rhythms.
- 25 • Segmentation of a piece
- Intra-track similarity: using result of segmentation algorithm, use the similarity measure disclosed above to assess similarity between parts of a track.
- Extraction of a speed feeling value. This would answer the question: how fast does that music piece feels?
- 30 • Detection of rhythmic modulations using segmentation and similarity measures
- Generate "creative" recommendation. This would recommend a track that makes use of a rhythm structure that is a rhythmic modulation of the seed track. (this is a trick used by musicians, composers and some DJs)

- Music classification. Based on the idea that different genres potentially have different rhythmic "signature".
- Estimation of the presence/absence of drums/percussions in a music piece.
- Entrainment for biomusicology – rhythm may be matched to the users activity or bio feedback monitors
- 5 • For use in biomusicology to affect a user or set of users in a desired way
- In car/vehicle matching of various aspects of things like speed, gear changes, etc
- For journeys where rhythm can be applied to an ETA
- For musical sequencing where tempo, pulse or rhythm is part of the track to track journey or tempo curve over time and multiple songs/playlists/radio stations
- 10 • For use in Radio sequencers or schedulers
- For use in track identification or matching
- For use in searching for musical pieces that could fit into a music production in a studio or other similar creative environment
- 15 • For use as part of a sub set filing system that locates music for production by matching rhythm
- For use in machine learning
- For use in synchronisation of musical or audio pieces
- 20 • Identification of sub set rhythmic patterns or harmonically diverse (or simple) pieces of music
- Identification of non audible rhythmic patterns for use in data analyses of high resolution audio or media
- Scientific data analyses of any rhythmic pattern in data sets (beyond audio)
- 25 • Identification of 'harding' or pulsing of images and light in video/visual mediums
- Use in live audio situations for teaching or performance
- Use for multiplexing and de-multiplexing periodic signals
- navigation of large data sets, libraries and catalogues

30 **Note**

It is to be understood that the above-referenced arrangements are only illustrative of the application for the principles of the present invention. Numerous modifications and

alternative arrangements can be devised without departing from the spirit and scope of the present invention. While the present invention has been shown in the drawings and fully described above with particularity and detail in connection with what is presently deemed to be the most practical and preferred example(s) of the invention, it will be
5 apparent to those of ordinary skill in the art that numerous modifications can be made without departing from the principles and concepts of the invention as set forth herein.

APPENDIX A

A mid-level feature for metrical hierarchy

5 Rhythm patterns in music are complex signals. In music that can be considered to have a sense of meter, the rhythm will have multiple related periodicities forming a hierarchical structure; the common musical terms for which include bars, beats and beat subdivisions. In this Appendix, we introduce a mid-level feature that captures the weighting of salient pulse rates present in the music, which correspond to the most likely metrical hierarchy.

10 Feature extraction is achieved by computing a novelty curve from the audio then processing this in time and frequency to produce a rhythmic spectrogram or rhythmogram. From this, we derive multiple possible candidates for a metrical hierarchy that fits the rhythmic content the best. We believe that this feature provides a meaningful description of the rhythm content without applying specific labels, such as tempo or

15 tactus, to any particular level.

1. Introduction

The temporal occurrences of musical events (e.g. notes) in a musical piece can be, and often are, organised such that periodic patterns emerge. These can be decomposed into

20 different levels of periodicity or rates that together form a metrical hierarchy. Rhythm-related research has been an important area in Music Information Retrieval (MIR) since the discipline's earliest days. Audio tempo induction in particular has seen a lot of interest in the community and is a long running track in the annual MIREX competition. The

25 aim of such systems is to extract a single-value descriptor that tells us something about the speed or pace of the music.

Tempo is a minimal descriptor of rhythm content that represents one of the levels in the metrical hierarchy and, alone, does not provide information about broader structure.

30 Methods for extracting more than one metrical level from musical audio have also been proposed in the literature. Seppanen introduce a method to extract the 'beat and tatum levels' jointly and Klapuri describes a method to extract the 'tatum, tactus and measure level'.

Induction of more than one metrical level is sometimes used as an intermediate step in beat tracking or onset detection methods. With the exception of Degara's method, the metrical levels extracted in the aforementioned algorithms are labelled with respect to their musical or perceptive role. Gouyon *et. al* proposed a system differentiating between
 5 duple and triple meter implying the estimation of two metrical levels. Further back, Temperley and Sleator designed an algorithm to analyse meter and harmony in symbolic music data, which allowed the presence of more than three metrical levels.

In this Appendix we introduce a method to extract a complete metrical hierarchy from
 10 audio, representing it in a compact feature without applying semantic labels to any particular level. The extraction is performed using standard time-frequency analysis and a peak picking strategy that incorporates hierarchical structure constraints. The energy associated with each metrical level in our feature corresponds to the density of events of that rhythmic subdivision in the music signal.

15

2. Formalising metrical hierarchy

The metrical hierarchy that describes the rhythm structure of a piece of music may be represented as a tree as shown in **Figure 1**. Each horizontal level of nodes on the tree
 20 accounts for one metrical level (index $i \in [0, N]$), which is associated with a frequency, or pulse rate f_i measured in BPM (Beats Per Minute). The number of metrical levels necessary to represent the rhythm hierarchy of a piece of music is therefore $N + 1$. These rates can be grouped in ascending order in a vector $F = (f_0, f_1, \dots, f_N)$. Hierarchical relationships are defined by the number $\lambda_i \in \mathbf{N}$ of child nodes each level generates. This

implies that $\lambda_i = \frac{f_i}{f_{i-1}}$. As a consequence, the full hierarchical structure can be described
 25 as a sequence of frequency ratios $\Lambda = \langle \lambda_1, \dots, \lambda_N \rangle$.

Metrical hierarchy is related in musical stave notation terms to the time signature and the
 30 3 note values used in a composition. As an example, a musical piece using eighth notes in a

30

4 time signature can be represented by **Figure 1** (b), with metrical level $i-1$ being the bar

level, level i the quarter note level (three quarter notes in one

3

4 bar) and level $i+1$ being the eighth note level (quarter note divides into two eighth notes).

5

With the ratios defined in Λ , specification of frequency f_i for a single metrical level allows the frequencies of all other levels to be calculated. A simple expression of the metrical hierarchy can therefore be given as $H = f_0 \mathring{a} \Lambda$ containing equivalent information to vector F . (The symbol \mathring{a} is used to represent the fact that the frequency f_0 can be recursively multiplied by the elements λ_i of Λ , to regenerate vector F . For example $f_2 = f_0 \lambda_1 \lambda_2$.)

10

The feature we introduce in this Appendix can be fully represented by F or H coupled with a vector of weights W eg. as described below. This provides a rich and compact description of the rhythmic content by associating a hierarchical structure with a relative weighting of the metrical levels.

15

3 Time-frequency analysis

20

Rhythmograms are a standard tool for rhythm analysis; a typical rhythmogram processing chain comprises two parts: estimation of a novelty curve $\Delta(t)$ capturing the sudden variations (e.g. transients) in the input audio signal $x(t)$ and a time-frequency analysis that shows the periodicities present in $\Delta(t)$.

25

To extract our rhythm feature, we exploit rhythmograms obtained from standard computation of Fourier transform and autocorrelation function. Our input signal $x(t)$ is mono audio sampled at 44.1kHz. We generate a simple novelty curve by full-wave rectifying $x(t)$, taking the first derivative and half-wave rectifying the result:

$$\Delta(t) = \left| \frac{d}{dt} (|x(t)|) \right|_{>0} \quad (1)$$

30

A framed FFT of $\Delta(t)$ is computed (using a 12 second Hamming window with 97% overlap) resulting in the rhythmogram $R_F(t, f)$.

The autocorrelation rhythmogram is obtained with Grosche's toolbox also using a 12 second window length. For each window, the unbiased autocorrelation of $\Delta(t)$ is computed. The lag scale is then mapped to BPM scale to generate rhythmogram $R_A(t, f)$.

For this Appendix we calculate our metrical hierarchy feature across a whole song so we average the rhythmograms $R_F(t, f)$ and $R_A(t, f)$ over time. To obtain the average rhythmic spectrum for each rhythmogram we calculate

$$\Omega_X(f) = \sum_t R_X(t, f), \quad (2)$$

where X denotes which rhythmogram (F or A) is being averaged.

FFT rhythmograms reveal harmonics of the periodicities present in the novelty curve whereas ACF rhythmograms reveal sub-harmonics of these same periodicities. Multiplying the spectra $\Omega_F(f)$ and $\Omega_A(f)$ therefore allows us to keep only the common periodicities, the most salient of which correlate with the metrical rates used in the music. Following that rationale, to produce a final rhythmic spectrum vector $S(f)$ we calculate the Hadamard product (an element by element multiplication denoted as \circ) of the spectra $\Omega_F(f)$ and $\Omega_A(f)$ and normalise the result:

$$S(f) = \frac{(\Omega_A(f) \circ \Omega_F(f))}{\max_f (\Omega_A(f) \circ \Omega_F(f))} \quad (3)$$

4 Hierarchy feature extraction

25

Extraction of the feature from $S(f)$ is achieved using a peak-picking strategy constrained by the meter hierarchy model. A simple peak detection algorithm that detects a local maxima if an element is larger than both of its neighbours incorporating a

threshold filter is employed (this was implemented using the standard Matlab function `findpeaks`). A threshold value of 0.085 was found to work well.

The abscissa of biggest peak in $S(f)$, frequency f^{max} , is used as a point of reference to
 5 extract the metrical hierarchy. It represents the rhythmic rate containing the most energy
 in the audio signal and is therefore assumed to represent an important metrical level. If
 the peak at f^{max} is part of the hierarchical structure, this directly implies that the rates
 of all other metrical levels f_j are related to f^{max} by integer ratios. In order to
 implement this constraint, the abscissa f_j of the j^{th} peak in $S(f)$ is compared to
 10 f^{max} and is considered as a candidate level if, and only if, it satisfies one of the
 following conditions

$$\exists q \in \mathbb{N} \setminus \left\{ \begin{array}{l} \frac{f_j}{f^{max}} = q \quad \text{if } f_j > f^{max} \\ \frac{f^{max}}{f_j} = q \quad \text{if } f_j < f^{max} \end{array} \right. \quad (4)$$

otherwise it is rejected.

15 By finding all the peaks that are integer ratios of f^{max} we cannot guarantee that they
 form a hierarchy consistent with the model. Moreover a single hierarchy might not be
 enough to describe a rhythmic structure (polyrhythmic music may imply multiple
 competing hierarchies for example). A further constraint therefore needs to be added to
 build a consistent hierarchy: the rate of each metrical level and its immediate neighbour
 20 must be related by an integer ratio.

As before, the biggest peak (at f^{max}) is taken as a reference as it must be part of the
 hierarchical structure. Starting with f^{max} , iterative comparison of metrical levels is
 performed upwards (comparison with metrical levels with higher rates) and downwards
 25 (comparison with metrical levels with lower rates). Moving upwards, a metrical level f_{j+1}

is accepted as part of the hierarchy if $\frac{f_{j+1}}{f_j}$ is an integer number, otherwise it is rejected.

In this case f_j will then be compared to the next rate f_{j+2} and so on until an integer ratio is found or the peaks are exhausted.

A special case occurs when $\frac{f_{j+1}}{f_j}$ and $\frac{f_{j+2}}{f_j}$ are integer ratios but $\frac{f_{j+2}}{f_{j+1}}$ is not. This means
 5 that the metrical level associated to f_j could equally be subdivided in levels f_{j+1} or f_{j+2} whereas these two levels can't coexist in the same hierarchy (as shown in **Figure 2**). In such a situation, two parallel hierarchy candidates are generated; one using f_{j+1} (the
 10 duple meter in the example) and the other using f_{j+2} (the triple meter in the example). At this stage, hierarchy candidates have been generated, and are equally represented by their vector F or H .

Finally, for each hierarchy candidate, each one of the metrical levels f_i is associated with a weight $w_i = S(f_i)$ stored in $W = (w_0, w_1, \dots, w_N)$. Each hierarchy candidate is
 15 weighted by the sum of the weights of its metrical levels $\Theta = \sum_i w_i$. The hierarchy with the biggest weight Θ is the one receiving the more energy from the musical signal, and is therefore considered as the hierarchy that best fits the data.

5 Examples

20 5.1 Synthesised test tracks

A set of test rhythms were synthesised comprising all the permutations of groups of 3 or 4 events (24 patterns in total, which we labelled A-X), a selection of which is shown in **Figure 3** tracks schema. An individual pattern would be synthesised as a 0.5 second
 25 audio sample using a cross-stick sound and each audio track is made up of that pattern repeating for 30 seconds.

[plainE] [plainK] [plainQ]

[accentsE] [accentsK] [accentsQ]

Spectrums $S(f)$ with extracted hierarchy H , metrical levels rates F and relative weights vector W , for both plain and accented interpretation of patterns E, K and Q.

Two interpretations for each pattern were produced: Firstly, a plain interpretation for which black circles represent sounded notes (amplitude=1), and dashes represent rests (amplitude=0). Secondly, an accented interpretation for which circles are interpreted as accented notes (amplitude=1) whereas dashes are interpreted as unaccented notes (in this case amplitude= $\frac{1}{2}$).

10 [Pattern E

Plain $\langle 1,1,0,0 \rangle$

Accented $\langle 1,1,\frac{1}{2},\frac{1}{2} \rangle$] (0.2,0) circle (0.2cm); (1,0) circle (0.2cm); [thick] (1.6,0) – (2,0); [thick] (2.4,0) – (2.8,0); [Pattern Q

Plain $\langle 1,0,0 \rangle$

15 Accented $\langle 1,\frac{1}{2},\frac{1}{2} \rangle$] (0.2,0) circle (0.2cm); [thick] (0.8,0) – (1.2,0); [thick] (1.6,0) – (2,0); [Pattern K

Plain $\langle 1,1,1,0 \rangle$

Accented $\langle 1,1,1,\frac{1}{2} \rangle$] (0.2,0) circle (0.2cm); (1,0) circle (0.2cm); (1.8,0) circle (0.2cm); [thick] (2.4,0) – (2.8,0); [Pattern T

20 Plain $\langle 1,1,0 \rangle$

Accented $\langle 1,1,\frac{1}{2} \rangle$] (0.2,0) circle (0.2cm); (1,0) circle (0.2cm); [thick] (1.6,0) – (2,0);

The plain interpretation of pattern Q is effectively an isochronous event train at 120 BPM. As expected, the corresponding spectrum in **Figure 3** features only one peak at 120 BPM. Changing pattern Q to accented interpretation implies that the sequence of

amplitudes of sonic events is modified from $\langle 1,0,0 \rangle$ to $\langle 1,\frac{1}{2},\frac{1}{2} \rangle$. In other words a rhythmic carrier frequency is excited, and therefore a peak appears at 360 BPM. Plain

interpretation of pattern E excites two metrical levels: the carrier at 480 BPM because of the spacing between the two sounded notes and a 120 BPM level because of the repetition of the pattern. Plain interpretation of pattern K excites three metrical levels: the 480 BPM carrier level because of the spacing between first and second and second and third note, the 120 BPM level due to the pattern repetition and finally 240 BPM, the presence of which is explained by notes in first and third position being sounded.

Using accented interpretation effectively means that there are more notes played at the carrier metrical level, and therefore more energy present at that frequency. This fact is captured in W where the relative weight of carrier level is bigger. This demonstrates that the feature not only captures which metrical levels are used but also how much they are used.

5.2 Commercial tracks

15

We now apply the same analysis to commercial music recordings with results shown in **Figure 4**. The extracted hierarchy for Michael Jackson's 'Billie Jean' comprises only duple ratios. It is consistent with the fact that this track has a straight feel, easily scored in

⁴. In contrast, the high end of the extracted hierarchy for Steely Dan's 'Home at last' includes a triple ratio, consistent with the half-time 'Purdie shuffle' feel (easily scored in ¹²⁸). Finally, Vivaldi's 'Summer Presto' hierarchy also includes a triple ratio, but in the

low-end of the hierarchy this time. It is consistent with the scored meter of ³⁴. In fact, assuming the metrical level at 149 BPM is the quarter note, the other extracted levels

would then correspond to: two bars, one ³⁴ bar, quarter note, eighth notes and sixteenth notes. Here again the presence of the 607BPM metrical level and its relatively high weighting correspond to the heavy use of sixteenth notes made throughout the piece. Similarly, in Billie Jean, considering the 117 BPM metrical level as the quarter note, the composition is heavily based on eighth notes, in accordance with the predominant importance of the 234 BPM metrical level.

30

6 Conclusion

In this Appendix we have introduced a mid-level descriptor that captures rhythm structure information from audio. The examples shown here have allowed us to demonstrate that a feature of low dimensionality can jointly capture metrical hierarchy and the relative weighting of each metrical level along with a quantitative speed reference. It therefore provides a compact representation of musically meaningful information about the rhythmic content of a music piece.

Note

10

It is to be understood that the above-referenced arrangements are only illustrative of the application for the principles of the present invention. Numerous modifications and alternative arrangements can be devised without departing from the spirit and scope of the present invention. While the present invention has been shown in the drawings and fully described above with particularity and detail in connection with what is presently deemed to be the most practical and preferred example(s) of the invention, it will be apparent to those of ordinary skill in the art that numerous modifications can be made without departing from the principles and concepts of the invention as set forth herein.

CLAIMS

1. A computer program product embodied on a non-transitory storage medium or on a cellular mobile telephone device or on another hardware device, the computer
5 program product operable to automatically generate recommendations for digital media content, in which that system is adapted to automatically generate recommendations of digital media content for an end user by (a) analysing a first item or set of digital media content and its associated metadata to calculate metric-based metadata for that digital media content and (b) using that analysis to identify other digital media content which
10 have a specifically-defined relationship with that first digital media content; and (c) provide, to the end user, recommendations of one or more items of that digital media content which were identified in the previous step.
2. The method of preceding Claim 1 where a first digital media content is a track
15 which is currently being listened to by an end user.
3. The method of preceding Claim 1 where a first digital media content consists of a set of multiple tracks which are (i) analysed individually to calculate metric-based metadata and (ii) that calculated metric-based metadata is aggregated and subsequently
20 treated, for the purposes of subsequent steps in claim 1, as if it derives from a single item of digital media content.
4. The method of preceding Claim 3 where the set of multiple tracks comprises the set of digital media content files which a user enjoys, whether determined by observing
25 that user's listening preferences or by the user specifically indicating his digital media content preferences or by any other means.
5. The method of any preceding Claim where the metric-based metadata comprises one or more of: the rhythmic value or values of a track, the time signature or time
30 signatures of a track; the rhythmic simplicity or complexity of a track; the tempo or cadence drift of a track; the rhythm salience and rhythmicity; confidence values for any of the listed derived metrics; any other calculated metrics for a track.

6. The method of any preceding Claim where the specifically defined relationship with the first digital media content is that the identified other track or tracks share similar or identical calculated metric-based metadata with that first digital media content.
- 5 7. The method of preceding Claim 6 where the calculated metric-based metadata for two or more tracks is defined as being similar if the two sets of metadata have values for a given metric which lie within a defined range from the metric-based metadata values calculated from a first item or set of digital media content.
- 10 8. The method of preceding Claim 7 where the defined range may be defined separately for each metric-based metadata value which is utilised.
9. The method of any preceding Claim where the metric-based metadata is calculated in advance and stored in a database for comparison, search and retrieval
15 purposes.
10. The method of any preceding Claim where the metric-based metadata is calculated as and when needed and discarded after use.
- 20 11. The method of any preceding Claim where the metric-based metadata is first calculated as and when needed and is then stored in a database for use subsequently.
12. The method of any preceding Claim where the one or more items of digital content recommended to the end user form a subset of those digital media content items
25 identified as having similar or identical metric-based metadata values to the first item(s) of digital media content.
13. The method of preceding Claim 12 where the subset of digital media content items is determined by filtering the list using additional criteria, including but not limited
30 to the artist, album, genre, musical taste, current date, time of playback, date/time of most recent playback, user demographic data or by using any other additional criteria.
14. A computer based system adapted to perform the method of any preceding Claim.

1/4

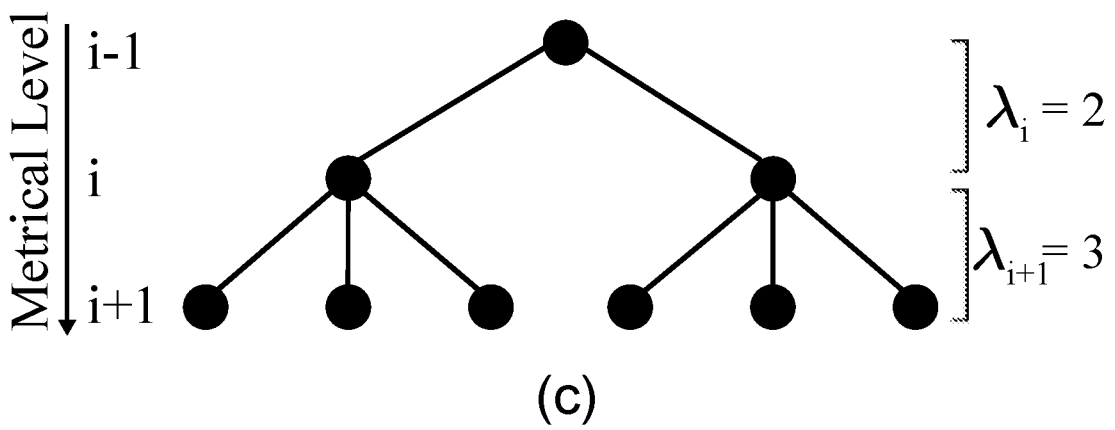
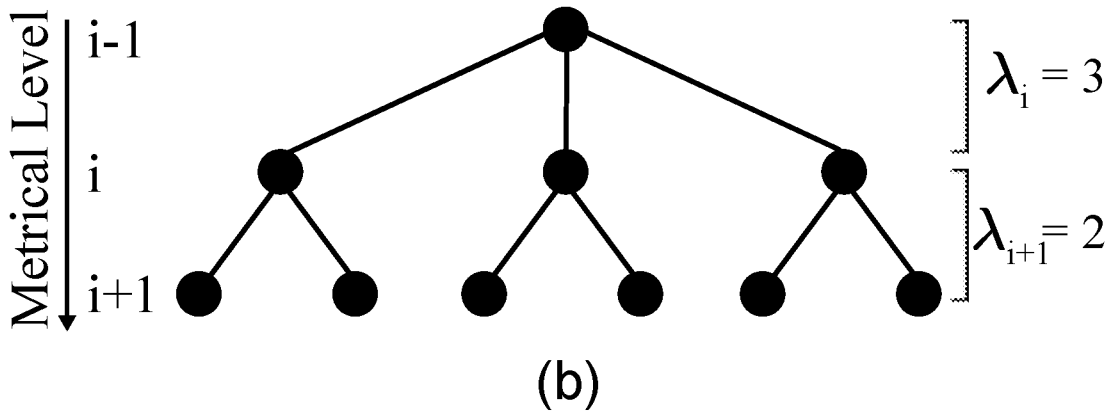
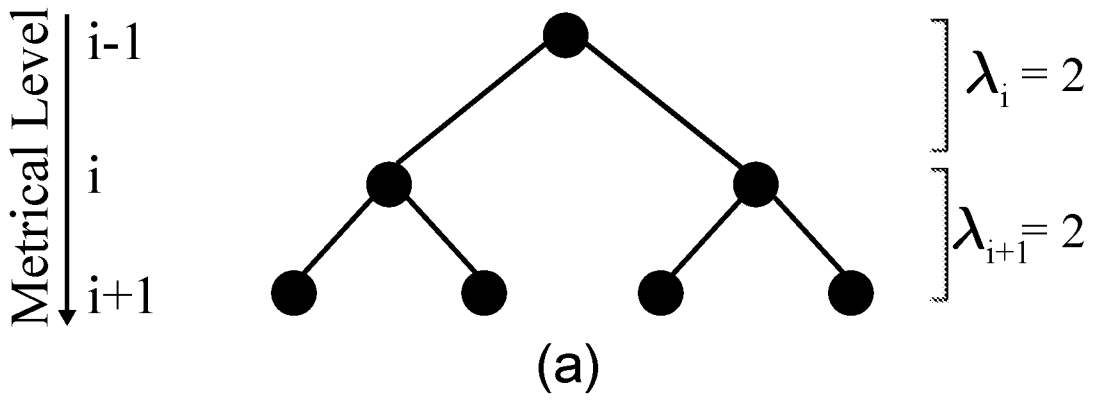


FIGURE 1

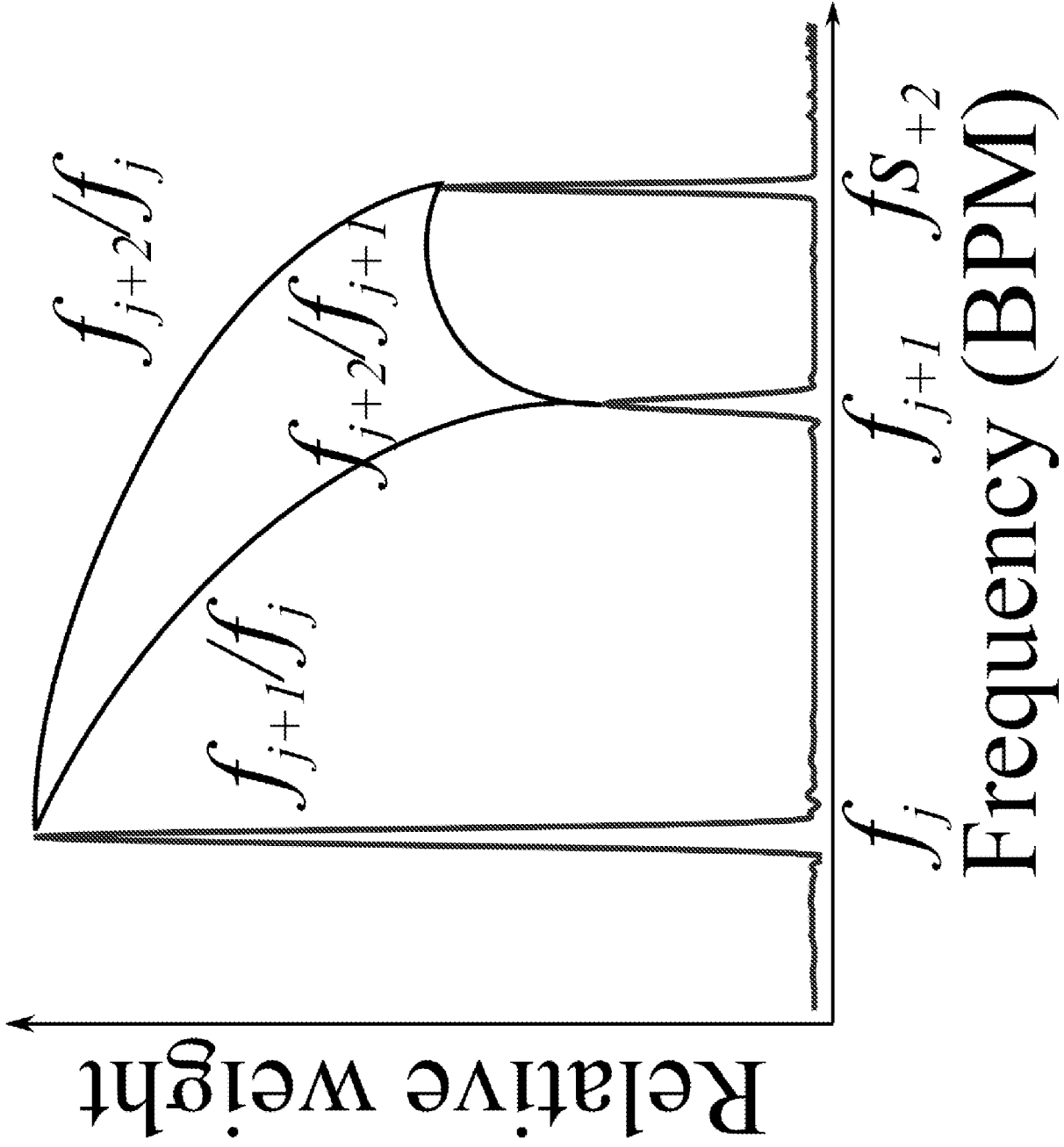


FIGURE 2

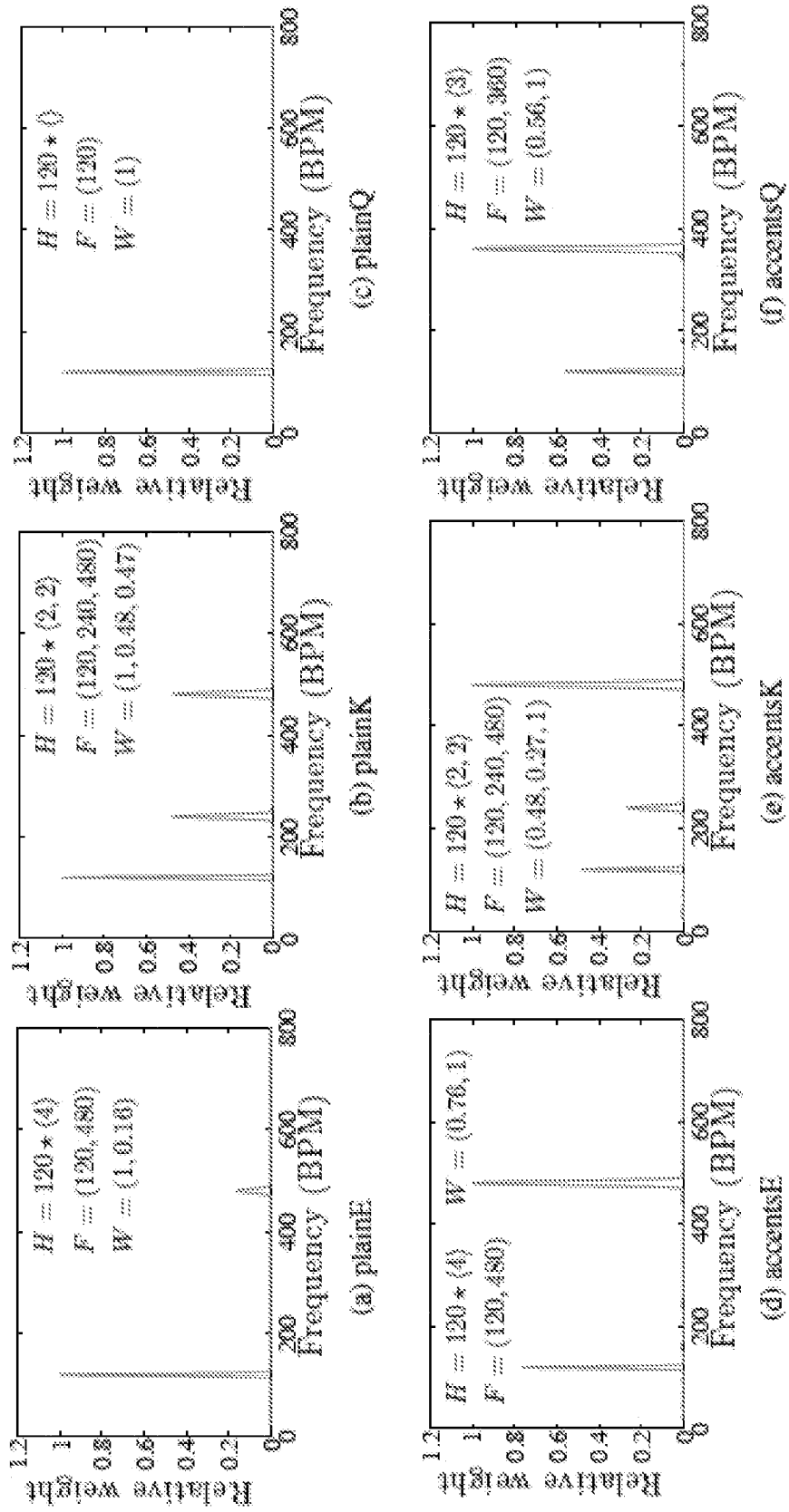


FIGURE 3

4/4

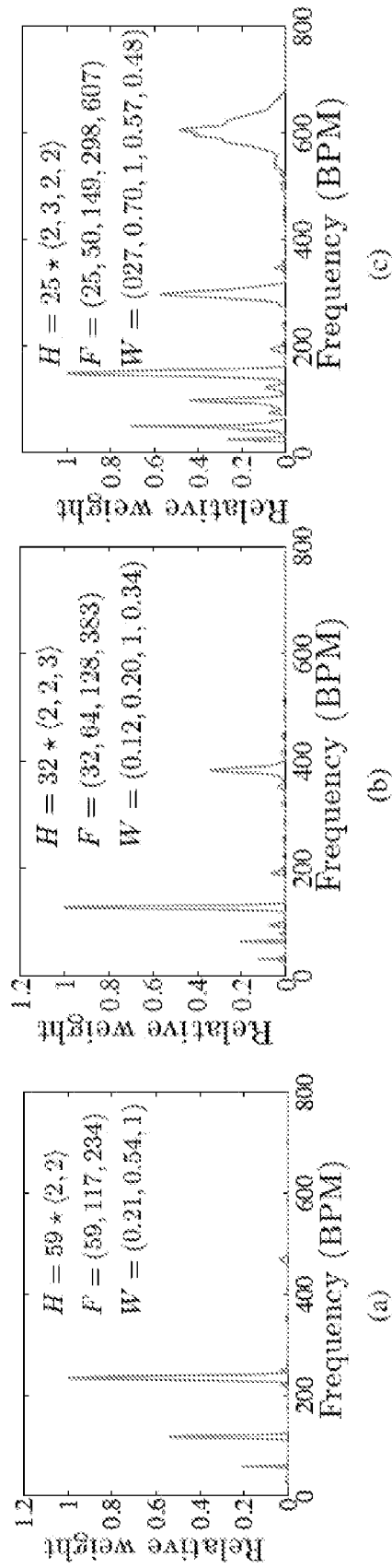


FIGURE 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/051380

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

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 Y	US 2013/031162 A1 (WILLIS MYK [US] ET AL) 31 January 2013 (2013-01-31) abstract; figures 2A,3A,4,5A,11 paragraph [0142] - paragraph [0143] paragraph [0155] - paragraph [0158] paragraph [0164]	1,2,5,7, 9,14 3,4,6,8, 10-13
X Y	WO 2014/002064 A1 (ECOLE POLYTECH [CH]) 3 January 2014 (2014-01-03) abstract; figures 3,4,6,8 page 8, line 6 - page 9, line 10 page 10, line 23 - page 11, line 5 page 13, line 19 - line 23	1,2,5,7, 9,14 3,4,6,8, 10-13
X A	US 2011/295843 A1 (INGRASSIA JR MICHAEL I [US] ET AL) 1 December 2011 (2011-12-01) abstract; figures 2,4,7 paragraphs [0037], [0048]	1,2,5,14 3,4,7-13

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 July 2015	Date of mailing of the international search report 22/07/2015
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 Nazzaro, Antonio

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2015/051380

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
US 2013031162	A1	31-01-2013	US 2013031162 A1	31-01-2013
			US 2013031177 A1	31-01-2013
			US 2013031216 A1	31-01-2013
			US 2013218942 A1	22-08-2013

WO 2014002064	A1	03-01-2014	NONE	

US 2011295843	A1	01-12-2011	NONE	
