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

(51) **International Patent Classification:** Not classified

(21) **International Application Number:**
PCT/IB2005/053755

(22) **International Filing Date:**
15 November 2005 (15.11.2005)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
04105895.9 18 November 2004 (18.11.2004) EP
05102541.9 31 March 2005 (31.03.2005) EP

(71) **Applicant** *(for all designated States except US)*: **KONIN-KLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **VIGNOLI, Fabio** [IT/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **PAUWS, Steffen, C.** [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) **Agents: GROENENDAAL, Antonius, W., M.** et al.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— as to applicant's entitlement to apply for and be granted a patent *(Rule 4.17(ii))*

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD OF PROCESSING A SET OF CONTENT ITEMS, AND DATA- PROCESSING DEVICE
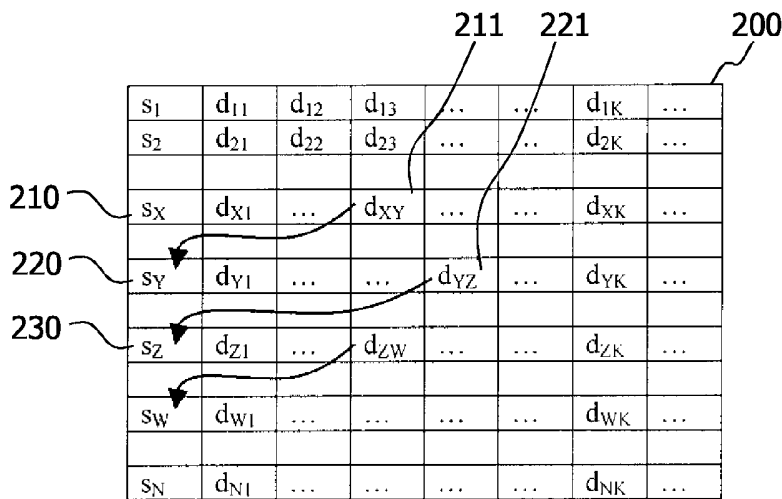
(57) **Abstract:** A method of processing a set of content items, the method comprising the steps of: (150) selecting a particular one (s_X, 210) of the set of content items, (160) selecting a first content item (s_Y, 220) on the basis of a first distance value (d_{XY}, 211) representative of the distance between the particular content item and the first content item, and (180) selecting a second content item (s_Z, 230) on the basis of a second distance value (d_{YZ}, 221) representative of the distance between the first content item and the second content item. The particular, first and second content items may be sequentially included in a sequence of content items, e.g. for creating a playlist of content items. The invention also relates to a data-processing device for processing a set of content items, as well as to a distance data structure and a computer program product.

Method of processing a set of content items, and data-processing device

The invention relates to a method of processing a set of content items, a data-processing device for processing a set of content items, as well as to a distance data structure and a computer program product, for instance, to create a sequence of content items.

US6545209B1 discloses a system for generating a playlist of songs. The
5   playlist is generated from a reference song and/or a user preference profile. The system uses music analysis algorithms to analyze intrinsic properties of the songs. A similarity value to the relationship between two respective songs is computed on the basis of the intrinsic properties of the songs. Songs which are closely related, i.e. songs with the highest similarity value, are included in the playlist.

10   It is a drawback of the system known from US6545209B1 that the playlist includes the same (or substantially the same) closely related (similar) songs whenever the same reference song or the same user preference profile is taken.

WO02/061750A1 discloses a method of sorting music files on the basis of a parameter. A random number is generated and added to a value of the parameter to calculate
15   sorting criteria for each music file. The music files are then sorted in accordance with their sorting criteria, generating a playlist of the music files.

It is a drawback of the method known from WO02/061750A1 that one music file is not similar to the preceding or subsequent music file of the paylist. In other words, the playlist is not coherent.

20   It is desirable to provide a method of processing a set of content items, which allows generating a sequence of content items which is not the same when the same reference (particular) song is taken whenever the sequence is generated. At the same time, every content item should be similar to the preceding content item in the sequence.

The method comprises the steps of:
25   -   selecting a particular one of the set of content items,
-    selecting a first content item on the basis of a first distance value representative of the distance between the particular content item and the first content item, and

-        selecting a second content item on the basis of a second distance value representative of the distance between the <u>first content item</u> and the second content item.

The first content item (fci) may be the closest content item to the particular content item. For instance,

$$fci = \frac{\arg\min(d(seed,s))}{C}$$

where d(seed,s) = distance between the seed and any item s belonging to the collection C.

Alternatively, the first content item may be selected randomly from a first sub-set of the first content items. The first sub-set is established in such a manner that each first content item of the first sub-set is associated with the respective first distance value which is smaller than a distance value representative of the distance between the particular content item and any content item excluded from the first sub-set. For example,

$$C^* = \{s : d(seed,s) < \varepsilon\}$$
$$fci = random(C^*)$$

wherein C* is the subset of the items s belonging to the collection C that have a distance with the seed item smaller than a certain threshold. Alternatively, the subset C* can be computed by selecting the first N items closer to the seed.

To select the second content item, the distance (or, alternatively, the similarity) between the first content item and the second content item is determined. Therefore, the second content item is similar (by at least one parameter, e.g. a content genre) to the first content item. The sequence comprises, in the order of the selection, the particular, first and second content items.

In contrast to US6545209B1, the distance between the particular content item and the second content item does not need to be decisive for the inclusion of the second content item in the sequence, but, preferably, there will be a certain similarity between the particular content item and the second content item caused by the calculation of the first and second distance values with respect to the first content item.

In WO02/061750A1, the distance (similarity) between the first and second content items is not considered at all. Therefore, for example, two content items that are quite

3

different from each other by their tempo, but are both related to the same genre as the particular content item are subsequently included in the playlist which thereby becomes inconsistent.

A data-processing device for processing a set of content items is provided, the device comprising a data processor configured to:

- select a particular one of the set of content items,
- select a first content item on the basis of a first distance value representative of the distance between the particular content item and the first content item, and
- select a second content item on the basis of a second distance value representative of the distance between the first content item and the second content item.

The device functions in accordance with the method of the present invention.

These and other aspects of the invention will be further explained and described, by way of example, with reference to the following drawings:

Figure 1 is an embodiment of the method of the present invention;

Figure 2 is an example of a distance data structure comprising distance values representative of distances between content items;

Figure 3 is an embodiment of a data-processing system comprising a data-processing device for processing a set of content items;

Figure 4A shows an example of an algorithm of calculating a distance matrix;

Figure 4B shows an example of a simple algorithm for selecting songs from a distance matrix;

Figure 4C shows an example of an algorithm for selecting songs from the distance matrix 200, wherein the selection is user-operably controllable;

Figure 4D shows an example of an algorithm for controlling the selection of content items with respect to the distance between the particular content item and further selected content items (e.g. the second content item).

For purposes of brevity and clarity, identical or corresponding elements such as distance values, will bear the same designations or numbering throughout the Figures.

Digital consumer electronics devices have made it possible to store a large amount of content on data carriers, e.g. hard disks. The amount of available content leads to

4

the necessity of new ways of access that are easy and less time-consuming. Automatic generation of playlists is considered as one of the most important ways for users of digital devices to maximize the experience of listening to music while reducing the amount of time and effort needed. The easiest way to automatically generate a playlist is the shuffle button

5   present on the majority of the devices. Unfortunately, random functions introduce too much variation in the playlist (e.g. a hard rock song right after a romantic jazz piece) and therefore do not provide the kind of experience the users are looking for.

A possible solution to the excessive randomness of the playlists is provided in WO02/31828. It enables the user to decide the amount of randomness by e.g. turning a knob

10   from 0% to 100%. A value of 0% means that all songs played come from the selection of favorites manually updated by the users and 50% means that half the number of songs are randomly selected from the entire collection, while 100% means that all songs are randomly selected from the user's collection. The problem with this solution is that, with a long list of favorites, it is still quite probable that very diverse songs will be played after each other, even

15   when the knob is set to 0%.

Another method of generating playlists is based on the user selection of a first song (seed song). The automatic playlist generator computes the distance d with all other songs in the collection, sorts the songs based on said distance (in ascending order) and proposes a playlist of the first songs. The distance between two songs can be based on very

20   simple algorithms based on metadata (e.g. if same artist d=0, if different artist d=1). Other methods of generating playlists based on the distance are described in US6545209B1 and WO02/061750A1, but these methods have also problems as discussed above.

According to one aspect of the present invention, a sequence of content items is generated from a set of available content items. The sequence of content items may be used

25   in various applications. For example, the content items are presented to a user in the order determined by the sequence. A playlist of audio (or video) content items may be generated. In another example, the content items are stored in the order determined by the sequence.

The content item may comprise at least one of, or any combination of, visual information (e.g. video images, photos, graphics), audio information, text information, and

30   other digital data such as e.g. metadata according to the MPEG-7 standard which may be used to describe and search digitized materials by means of sampling, as well as by using lexical search terms.

The expression "audio data", or "audio content", is hereinafter used as data pertaining to audio comprising audible tones, silence, speech, music, tranquility, external

noise or the like. The audio data may be in formats like the MPEG-1 layer III (mp3) standard (Moving Picture Experts Group), the AVI (Audio Video Interleave) format, or the WMA (Windows Media Audio) format, etc. The expression "video data", or "video content", is used as data which are visible such as a motion picture, "still pictures", video text, etc. The video data may be in formats like GIF (Graphic Interchange Format), JPEG (Joint Photographic Experts Group), or MPEG-4, etc. The text information may be, for example, in the ASCII (American Standard Code for Information Interchange) format, the PDF (Adobe Acrobat Format) format, or the HTML (HyperText Markup Language) format. The metadata may be in the XML (Extensible Markup Language) format, the MPEG7 format, stored in a SQL database or any other format.

The sequence of content items is created on the basis of the similarity/distance of content of the items. The terms "similarity" and "distance" may be used interchangeably (where similarity means a small distance with the seed item). To measure the similarity between the content items, various characteristics of the content items may be taken into account.

For example, songs may be compared by their tempo, sonic, melodic movement and musical consonance characterizations or even genre. Songs may be compared by their catalog metadata, e.g. performer, composer, album, period of time. It is known to compute the Fast Fourier Transform (FFT) of short frames of the audio signal (e.g. 20-40 ms) and derive some features from it such as Mel Frequency Cepstrum Coefficients as described in J.R. Deller, Jr., J.H.L. Hansen, and J.G. Proakis, Discrete Time Processing of Speech Signals (2nd ed.), New York: IEEE Press, 2000, pp. 380-390. These MFCC may be used to determine the distance and henceforth the similarity.

The song may have a particular rhythm time, rhythm type, pitch that designates the frequency of a voice, loudness, dynamics and timbre. For instance, the song having a dominant frequency spectrum at 3 kHz is interpreted as being of the "jazz" genre, e.g. in addition to other genres. A song with a frequency spectrum of more than 3kHz or less than 3 kHz will be of the "less of jazz genre". For instance, a "jazz" distance value representative of the distance between two "jazz" songs may be close to 0. A "jazz" distance value representative of the distance between the "jazz" song and a "jazz/pop" song may be around 0.5. When the genre of the song is determined on the basis of a plurality of characteristics, it is possible to calculate more precisely a ratio of values of the same parameter of two songs in order to express the distance value between songs.

It is possible that a content item, e.g. a song, is of more than one genre. For instance, rhythm is a phenomenon based on the relationships between events in time. Such events may be either changing or repeating notes or chords. The rhythm may be important to organize musical events throughout a song, while the concept of tempo may be a mere

5   quantitative parameter indicating the beats per minute. Tempo may be regarded to some extent independent of the rhythm of the music. Rhythm may not change if tempo is increased within reasonable limits. Therefore, if the song has a rhythm as in Latin music and tempo as in pop music, the song may be associated with both genres.

According to an embodiment of the method shown in Figure 1, distances

10   between available content items are calculated so that every content item is compared with other respective available content items. The expression "content item" is further referred to as "item" for the purpose of brevity.

In step 110, one of the N available items is selected. For example, an item s_1 is selected from all items (s_1, s_2, ... s_N). A distance value d_{12} representative of the distance

15   between the item s_1 and the item s_2 is calculated in step 120. The distance value d_{11} is equal to 0, of course, because it relates to the same item s_1. N distance values (d_{11}, d_{12}, ... d_{1N}) are calculated, which are representative of respective distances between the item s_1 and items (s_1, s_2, ... s_N). The distance values are in a range of, for example, 0-1.

Some distance values may be smaller than other distance values. The distance

20   values may be compared and, in step 130, only K smallest distance values (d_{11}, d_{12}, ... d_{1K}) are stored in a distance array (matrix) 200 shown in Figure 2. Items 1, 2, ...K constitute a sub-set from the set of N available items. Each K distance value is smaller than any distance value representative of the distance between the item s_1 and any item excluded from the sub-set.

25   In step 140, the calculations are repeated until N^2 distance values are computed for all N items. As a result, the distance matrix D 200 having N×K distance values and indexes is obtained. An example of an algorithm of calculating the distance matrix D is shown in Figure 4A.

Strictly speaking, it is not required that only K smallest distance values are

30   stored in the distance array 200. For example, the distance values (d_{11}, d_{12}, ... d_{1K}) may be computed on the basis of a first characteristic D that is common to the items (s_1, s_2, ... s_N), and similarity values (g_{11}, g_{12}, ... g_{1K}) on the basis of a second characteristic G. Characteristics D and G may be determined independently for any item. The total number of elements stored in the distance array may be K, including the smallest distance values and the

largest similarity values. In another example, all NxN distance values are stored in the distance matrix 200 if it is desirable to do so.

The distance matrix 200 obtained in steps 110-140 further serves as the basis for generating the sequence of items in steps 150-190.

In step 150, a particular item (seed s_x) 210 is selected from the set of N available items ($s_1, s_2, \ldots s_N$).

The K distance values ($d_{X1}, d_{X2}, \ldots d_{XK}$) corresponding to the particular item 210 are identified in the distance matrix 200. The distance values in the distance array may be identified on the basis of the indices (e.g. XK attached to the respective distance value, where X indicates one of N items, and K indicates the other item in respect of which the distance d_{XK} is computed). Items 1, 2, ...K related to the distance values ($d_{X1}, d_{X2}, \ldots d_{XK}$) constitute a first sub-set from the set of N available items. Each K distance value ($d_{X1}, d_{X2}, \ldots d_{XK}$) is smaller than any distance value representative of the distance between the item s_X and any item excluded from the first sub-set.

One of the K distance values ($d_{X1}, d_{X2}, \ldots d_{XK}$), e.g. d_{XY} 211, is selected in step 160. The distance value d_{XY} may be selected randomly from the K distance values. Alternatively, the distance value d_{XY} may be the smallest among the K distance values. In another example, M of the K distance values are extracted from the K distance values, where the M distance values are smaller than a predetermined threshold value. The distance value d_{XY} is selected (e.g. randomly) from the M items.

An example of a simple algorithm for selecting songs from the distance matrix 200 is shown in Figure 4B.

On the basis of the (first) distance value d_{XY} 211, a first item s_Y 220 is identified which is associated with the distance value d_{XY} 211 representative of the distance between the particular item s_X 210 and the first item s_Y 220. Thus, the first item s_Y 220 is selected from the set of N available items ($s_1, s_2, \ldots s_N$). In step 170, the particular item s_X 210 and the first item s_Y 220 are sequentially added to a sequence of content items in the respective order.

In step 180, K distance values ($d_{Y1}, d_{Y2}, \ldots d_{YK}$) corresponding to the first item s_Y 220 are identified in the distance array 200. Items 1, 2, ...K related to the distance values ($d_{Y1}, d_{Y2}, \ldots d_{YK}$) constitute a second sub-set from the set of N available items. Each K distance value ($d_{Y1}, d_{Y2}, \ldots d_{YK}$) is smaller than any distance value representative of the distance between the item s_Y and any item excluded from the second sub-set.

One of the K distance values (d_{Y1}, d_{Y2}, ... d_{YK}), e.g. d_{YZ} 221, is selected. The distance value d_{YZ} may be selected randomly from the K distance values. Alternatively, the distance value d_{YZ} may be the smallest among the K distance values. In another example, P of the K distance values are separated from the K distance values, where the P distance

5   values are smaller than a predetermined threshold value. The distance value d_{YZ} is selected (e.g. randomly) from the P items.

In a further example, the distance value d_{YZ} 221 is selected from the K distance values (d_{Y1}, d_{Y2}, ... d_{YK}) so that a distance value between the particular item s_X 210 and an item s_Z (not the item s_Y) which is associated with the distance value d_{YZ} is smaller

10   than a predetermined distance threshold. Advantageously, the item s_Z may not be too distant to the particular item s_X (see Figure 4D and the corresponding paragraph below).

On the basis of the (second) distance value d_{YZ} 221, a second item s_Z 230 is identified which is associated with the distance value d_{YZ} 221 representative of the distance between the first item s_Y 220 and the second item s_Z 230. Thus, the second item s_Z 230 is

15   selected from the set of N available items (s_1, s_2, ... s_N). In step 190, the second item 230 is included in the sequence of the particular item s_X 210 and the first item s_Y 220.

The distance values (d_{X1}, d_{X2}, ... d_{XK}) or (d_{Y1}, d_{Y2}, ... d_{YK}) may be calculated in step 120 or step 140, taking into account a rating of the items in accordance with a user profile expressing user preferences with respect to genres, styles, etc. of content items. In an

20   extreme case, the distance array comprises only one distance value for each N available content item.

It is possible to control the selection of the first item and/or the second item, e.g. by means of a user-operable slider presented on a display screen (not shown). For example, when the first item is randomly selected from the M items (1<M<K) of the first

25   sub-set, the increase of M will result in the increased randomness of the selection, and the decrease of M will lead to the selection of the first item from the smaller number of items of the first sub-set, and, therefore, the randomness of the selection will be smaller.

This control may enable a user to decide a degree of randomness the user wants in the generated playlist. The control is implemented as the user-operable slider (from

30   0% (no-random) to 100% (full-random)). A value of 50% can be interpreted as a selection of M=K/2. Figure 4C shows an example of an algorithm for selecting songs from the distance matrix 200, wherein the selection is user-operably controllable.

The described embodiment of the method solves the problem of having a sequence of items based on one seed song (particular song) which is coherent (two

consecutive items are close to each other) and varied (the sequence is different for two different runs with the same seed songs) and user-controlled (the user can specify how much randomness the sequence should have). However, the methods above might propose sequences that drift farther away from the original seed of the song. The problem of the drift may be solved as indicated in Figure 4D. Figure 4D shows an example of an algorithm for controlling the selection of content items with respect to the distance between the particular content item and further selected content items (e.g. the second content item). For example, a slider from 0% to 100% can be used to determine the amount of drift allowed from the selected seed. The value of ε may directly be related to the maximal distance found in the current collection. A drift of 100% corresponds to a complete liberty to explore every song in the collection and a drift of 0% is a more restricted assignment.

The embodiment of the method shown in Figure 1 relates to the generation of the sequence of the items from one seed. However, it is possible to generate the sequence on the basis of two or more seeds (two or more particular items).

The method may further comprise the following steps:

- selecting a further particular content item,

- selecting a further first content item on the basis of a further first distance value representative of the distance between the <u>further particular content item</u> and the further first content item,

- selecting a representative first content item from the first content item and the further first content item,

- selecting the second content item on the basis of a second distance value representative of the distance between the <u>representative first content item</u> and the second content item.

Figure 3 shows an embodiment of a data-processing system 300 comprising a data-processing device 350 (a sequence generator) for processing the set of content items.

The system 300 comprises a media content database 310 comprising the set of content items, a distance array generator 320, the sequence generator 350 and a playlist player 390. The system may be distributed, i.e. its elements are remote from each other and may be remotely communicated.

The media content database 310 may be arranged to store the content items such as audio data, video data, etc. on different data carriers such as audio or video tapes, optical storage discs, e.g. CD-ROM discs (Compact Disc Read Only Memory) or DVD discs (Digital Versatile Disc), floppy and hard-drive disks, etc. in any format, e.g. MPEG (Motion

Picture Experts Group), MIDI (Musical Instrument Digital Interface), Shockwave, QuickTime, WAV (Waveform Audio), etc. For example, the media content database may comprise a computer hard-disk drive, a versatile flash memory card, e.g. a "Memory Stick" device, etc.

5   The distance array generator 320 comprises a distance data processor 321 and distance memory means 322. The distance data processor 321 is configured to access the media content database 310 and compute the distance values as described with reference to steps 110-140 in Figure 1. After the distance values are determined and the NxN distance array having NxN distance values is obtained, the distance data processor 321 stores the
10   distance array in the distance memory means 322.

        The sequence generator 350 comprises a sequence data processor 351, selection memory means 352, and sequence memory means 353. The sequence data processor 351 is configured to communicate with the distance memory means 322 and receive only a part of the NxN distance array so that the NxK distance array is stored in the
15   selection memory means. In other words, it is not required that the entire NxN distance array is downloaded from the distance memory means 322 to the selection memory means 352 because the NxK distance array (K<N) comprises the smallest distance values, and the rest of the NxN distance array is preferably ignored for the further purpose of generating the sequence of content items.

20   The sequence data processor 351 coupled to the selection memory means further performs the steps 150-190 explained with reference to Figure 1, and obtains the sequence of content items comprising the particular content item, the first content item and the second content item. The sequence data processor 351 further stores the sequence in the sequence memory means 353.

25   The distance memory means 322, the selection memory means 352 or the sequence memory means 353 may be, for example, a known RAM (random access memory) memory module. The distance data processor 321 or the sequence data processor 351 may be a central processing unit (CPU) suitably arranged to allow operation of the distance array generator 320 or the sequence generator 350 as explained herein.

30   The playlist player 390 may be a consumer electronics device such as a television set (TV set) with a cable, satellite or other link, a videocassette or HDD-recorder, a home cinema system, a portable CD player, a remote control device such as an I-Pronto remote control, a cell phone, etc.

The playlist player communicates with the sequence generator 350 to retrieve the sequence of content items from the sequence memory means 353. Therefore, the playlist player is enabled to reproduce the content items indicated in the sequence in the respective order. The content items, e.g. the particular, first and second content items, may be retrieved by the playlist player 390 from the media content database 310.

The sequence generator 350 may be provided with user input means (not shown), e.g. a keyboard, such as the well-known QWERTY computer keyboard, a pointing device, a TV remote control unit, etc. The user input means is configured to enable a user to set, for example, the parameter M or P to control the randomness of generating the sequence of content items. Alternatively, the user input means may be comprised in the playlist player 390, and the playlist player is arranged to communicate the parameter M or P to the sequence generator 350.

When the sequence stored in the sequence memory means 353 is the video playlist of video content items, the playlist player 390 may be arranged to present the video content items to the user using display means (not shown), for example, a CRT (cathode ray tube), LCD (Liquid Crystal Diode), or LCOS (Liquid Crystal on Silicon) rear-projection technology, a DLP (Digital Light Processing) television/Projector, a Plasma Screen display device, etc.

Variations and modifications of the described embodiment are possible within the scope of the inventive concept.

For example, in one embodiment, the sequence generator 350 and the playlist player 390 are combined in a single device, and the distance array generator 320 functions as a service provider/server.

The distance data processor 321 or the sequence data processor 351 may execute a software program to allow execution of the respective steps of the method of the present invention. The software may enable the system of the present invention independently of where it is being run. The independent method claim and the computer program product claim may be used to protect the invention when the software is manufactured or exploited to run on the consumer electronics products. The elements of the system may communicate with each other, using existing technologies, such as Bluetooth, 802.11[a-g], UPnP (Universal Plug and Play), etc.

A "computer program" is to be understood to mean any software product stored on a computer-readable medium, such as a floppy disk, downloadable via a network, such as the Internet, or marketable in any other manner.

The various program products may implement the functions of the system and method of the present invention and may be combined in several ways with the hardware or located in different devices. The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer.

5   Use of the verb 'comprise' and its conjugations does not exclude the presence of elements or steps other than those defined in a claim. In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. All details may be replaced with other technically equivalent elements.

CLAIMS:

1.   A method of processing a set of content items, the method comprising the steps of:

-           (150) selecting a particular one (s_X, 210) of the set of content items,

-           (160) selecting a first content item (s_Y, 220) on the basis of a first distance value (d_{XY}, 211) representative of the distance between the <u>particular content item</u> and the first content item, and

-           (180) selecting a second content item (s_Z, 230) on the basis of a second distance value (d_{YZ}, 221) representative of the distance between the <u>first content item</u> and the second content item.

2.   The method of claim 1, further comprising a step of obtaining a first sub-set of the first content items (1, 2, ..., K) from the set of content items, each first content item of the first sub-set being associated with the respective first distance value (d_{X1}, d_{X2}, ..., d_{XK}) which is smaller than a distance value representative of the distance between the <u>particular content item</u> and any content item excluded from the first sub-set.

3.   The method of claim 2, wherein the first content item is selected randomly from the first sub-set.

4.   The method of claim 3, wherein the randomness is distributed inversely proportional to the distance.

5.   The method of claim 2, 3 or 4, wherein the first content item is selected randomly from only a specified first number (M) of the first content items of the first sub-set.

6.   The method of claim 5, enabling a user to control the first number (M).

7.   The method of claim 2, wherein the first content item is selected from the first sub-set if said first content item is associated with the smallest first distance value among all first content items of the first sub-set.

5   8.   The method of claim 1, further comprising a step of obtaining a second sub-set of the second content items (1, 2, ..., K) from the set of content items, each second content item of the second sub-set being associated with the respective second distance value (d_{Y1}, d_{Y2}, ..., d_{YK}) which is smaller than a distance value representative of the distance between the <u>first content item</u> and any content item excluded from the second sub-set.

10

9.   The method of claim 8, wherein the second content item is selected randomly from the second sub-set.

10.   The method of claim 8 or 9, wherein the second content item is selected
15   randomly from only a specified second number (P) of the second content items of the second sub-set.

11.   The method of claim 8, wherein the second content item is selected from the second sub-set if said second content item is associated with the smallest second distance
20   value among all second content items of the second sub-set.

12.   The method of claim 8 or 9, wherein a distance value representative of the distance between the particular content item and the selected second content item is smaller than a predetermined distance threshold.

25

13.   The method of claim 1, further comprising a step (170, 190) of sequentially including the particular, first and second content items in a sequence of content items.

14.   The method of claim 1, further comprising the steps of:
30   -   selecting a further particular content item,
     -           selecting a further first content item on the basis of a further first distance value representative of the distance between the <u>further particular content item</u> and the further first content item,
     -           selecting a representative first content item from the first content item and the

further first content item,

-        selecting the second content item on the basis of a second distance value representative of the distance between the <u>representative first content item</u> and the second content item.

15.   The method of claim 1, wherein the first and/or second distance value is determined on the basis of a parameter of the content item: a genre of the content item, or a rating of the content item in accordance with a user profile.

16.   A data-processing device (350) for processing a set of content items (310), the device comprising a data processor (351) configured to:

-        select a particular one (s_X, 210) of the set of content items,

-        select a first content item (s_Y, 220) on the basis of a first distance value (d_{XY}, 211) representative of the distance between the <u>particular content item</u> and the first content item, and

-        select a second content item (s_Z, 230) on the basis of a second distance value (d_{YZ}, 221) representative of the distance between the <u>first content item</u> and the second content item.

17.   A distance data structure (200) comprising,

-        at least one first distance value (d_{XY}, 211) representative of the distance between a particular content item (s_X, 210) from a set of content items (310) and a respective one of first content items (s_Y, 220) of the set, and

-        at least one second distance value (d_{YZ}, 221) representative of the distance between the respective first content item (s_Y, 220) and a respective one of second content items (s_Z, 230) of the set.

18.   A computer program product enabling a programmable device when executing said computer program product to function as the device as defined in claim 16.
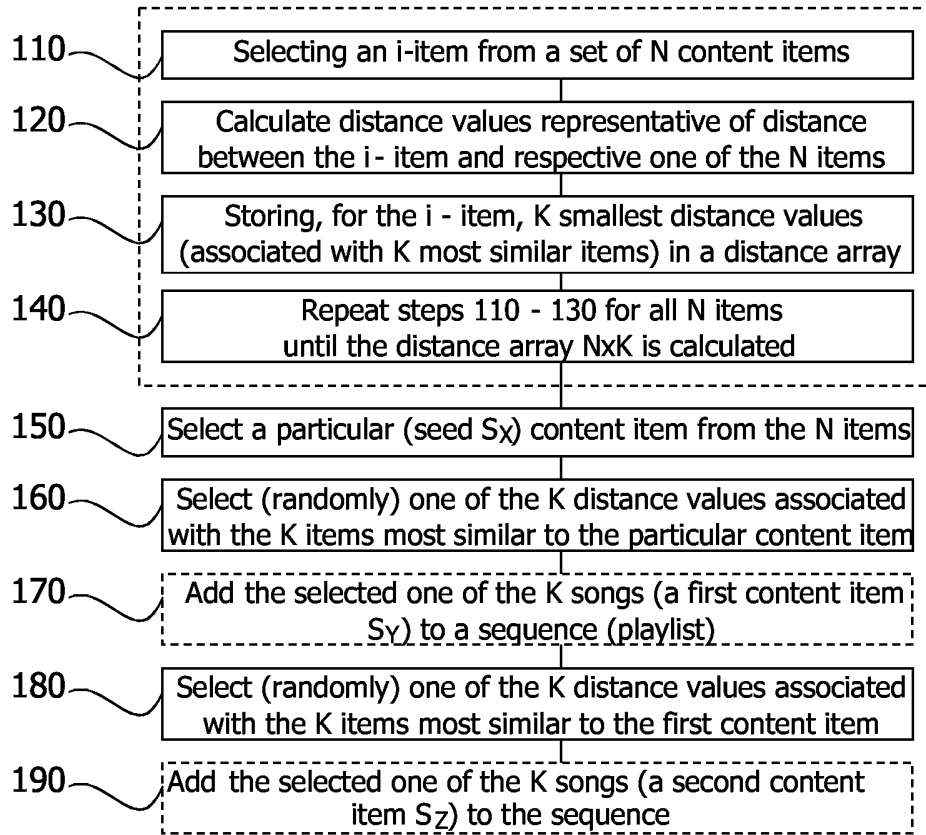
1/4

110 — Selecting an i-item from a set of N content items

120 — Calculate distance values representative of distance
between the i- item and respective one of the N items

130 — Storing, for the i - item, K smallest distance values
(associated with K most similar items) in a distance array

140 — Repeat steps 110 - 130 for all N items
until the distance array NxK is calculated

150 — Select a particular (seed S_X) content item from the N items

160 — Select (randomly) one of the K distance values associated
with the K items most similar to the particular content item

170 — Add the selected one of the K songs (a first content item
S_Y) to a sequence (playlist)

180 — Select (randomly) one of the K distance values associated
with the K items most similar to the first content item

190 — Add the selected one of the K songs (a second content
item S_Z) to the sequence

FIG.1

211 221   200

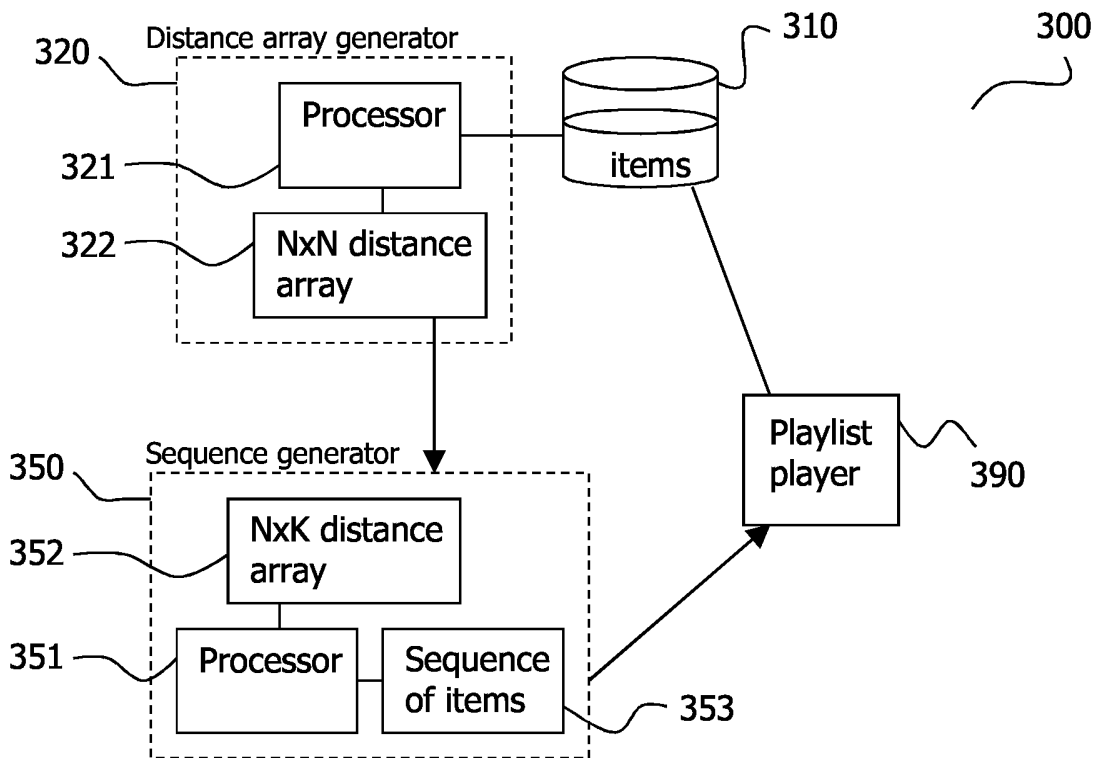| s_1	d_{11}	d_{12}	d_{13}	..	../	d_{1K}	...
s_2	d_{21}	d_{22}	d_{23}	...	..	d_{2K}	...
210 — s_X	d_{X1}	...	d_{XY}	...	...	d_{XK}	...
220 — s_Y	d_{Y1}	...	...	d_{YZ}	...	d_{YK}	...
230 — s_Z	d_{Z1}	...	d_{ZW}	...	...	d_{ZK}	...
s_W	d_{W1}	...	...	...	...	d_{WK}	...
s_N	d_{N1}	...	...	...	...	d_{NK}	...

FIG.2

FIG.3

COMPUTEDISTANCEMATRIX (NxK) ALGORITHM

For each song s_x in C
  $C' = C - s_x$
  For each song s_y in C'
  $C'' = C' - s_y$
  $D_{xy} = (\text{argmin}(d(s_x, C''), \min(d(s_x, C''))$
end

FIG.4A

SIMPLERANDOMSELECTNEXTSONG ALGORITHM

Init:
  Select seed s_x
  Initialize playlist P = ∅

Iteration:
  If $s_x \notin P$ then P = P + s_x
  $s^* = \text{random}(D_x)$
  $s_x = s^*$

End condition:
  card(P) < N

where N is the number of items desired by the user.
Random: D →N. The probability is distributed uniformly and
s* is
the random selected song.

DISTANCERANDOMSELECTNEXTSONG ALGORITHM

Same as SIMPLERANDOMSELECTNEXTSONG
But this time the probability is inversely distributed w.r.t.
the distance and s* is the random selected song.

FIG.4B

4/4

UserControlledRandomSelectNextSong ALGORITHEM

Init:

Select seed s_x

Initialize playlist $P = \varnothing$

Iteration:

If $s_x \notin P$ then $P = P + s_x$

$s^* = random(D'_x)$

$s_x = s^*$

End condition:

$card(P) < N$

where N is the number of items desired by the user. Random:$D \rightarrow N$. In this case the size of the set D is determined by the parameter M controlled by the user. The probability is distributed uniformly (or based on the distance) and s* is the random selected song.

FIG.4C

NonDriftingRandomSelectNextSong ALGORITHEM

Init:

Select seed s'

$s_x = s'$

Initialize playlist $P = \varnothing$

Iteration:

If $(s_x \notin P)$ and $(d(s_x, s') < \varepsilon)$ then $P = P + s_x$

$s^* = random(D'_x)$

$s_x = s^*$

End condition:

$card(P) < N$

where N is the number of items desired by the user. Random:$D \rightarrow N$. In this case the condition to check before adding a song to the playlist is that the distance with the first seed song s' is smaller than a certain threshold.

FIG.4D