(19) Europäisches Patentamt / European Patent Office / Office européen des brevets

(11) EP 3 979 241 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
15.05.2024 Bulletin 2024/20

(21) Application number: **20815214.0**

(22) Date of filing: **22.05.2020**

(51) International Patent Classification (IPC):
G10L 25/27 *(2013.01)* ***G10L 25/54*** *(2013.01)*

(52) Cooperative Patent Classification (CPC):
G10L 25/54; G10L 25/27; Y02D 30/70

(86) International application number:
PCT/CN2020/091698

(87) International publication number:
WO 2020/238777 (03.12.2020 Gazette 2020/49)

(54) AUDIO CLIP MATCHING METHOD AND APPARATUS, COMPUTER-READABLE MEDIUM AND ELECTRONIC DEVICE

VERFAHREN UND VORRICHTUNG ZUR ABSTIMMUNG EINES AUDIOCLIPS, COMPUTERLESBARES MEDIUM UND ELEKTRONISCHE VORRICHTUNG

PROCÉDÉ ET APPAREIL DE MISE EN CORRESPONDANCE ENTRE CLIPS AUDIO, SUPPORT LISIBLE PAR ORDINATEUR ET DISPOSITIF ÉLECTRONIQUE

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **24.05.2019 CN 201910441366**

(43) Date of publication of application:
06.04.2022 Bulletin 2022/14

(73) Proprietor: **Tencent Technology (Shenzhen) Company Limited**
Shenzhen, Guangdong, 518057 (CN)

(72) Inventors:
- **LIN, Fangchao**
  Shenzhen, Guangdong 518057 (CN)
- **YUN, Weibiao**
  Shenzhen, Guangdong 518057 (CN)
- **ZENG, Peng**
  Shenzhen, Guangdong 518057 (CN)

(74) Representative: **Gunzelmann, Rainer**
Wuesthoff & Wuesthoff
Patentanwälte und Rechtsanwalt PartG mbB
Schweigerstraße 2
81541 München (DE)

(56) References cited:
EP-A1- 0 103 258 JP-A- S62 144 200

- **JYH-SHING ROGER JANG ET AL: "Hierarchical filtering method for content-based music retrieval via acoustic input", PROCEEDINGS / ACM MULTIMEDIA 2001, [9TH ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA] : OTTAWA, CANADA, SEPTEMBER 30 - OCTOBER 5, 2001, ASSOC. FOR COMPUTING MACHINERY, NEW YORK, NY, USA, 1 October 2001 (2001-10-01), pages 401-410, XP058310382, DOI: 10.1145/500141.500201 ISBN: 978-1-58113-394-3**
- **Myers Cory S: "A comparative study of several dynamic time warping algorithms for speech recognition", , 1 February 1980 (1980-02-01), pages 1-152, XP055931752, Retrieved from the Internet: URL:https://dspace.mit.edu/bitstream/handle/1721.1/27909/07888629-MIT.pdf?sequence=2&isAllowed=y [retrieved on 2022-06-15]**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 3 979 241 B1

Description

RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 201910441366.5, entitled "METHOD AND APPARATUS FOR MATCHING AUDIO CLIPS, COMPUTER-READABLE MEDIUM, AND ELECTRONIC DEVICE" and filed on May 24, 2019.

FIELD OF THE TECHNOLOGY

[0002] The invention relates to the field of computer and communication technologies, and in particular, to a method and an apparatus for matching audio clips, a computer-readable medium, and an electronic device.

BACKGROUND OF THE DISCLOSURE

[0003] In a scenario of matching audio clips, for example, query by humming or scoring by humming, an audio feature sequence of a hummed tune and a feature sequence of a to-be-retrieved audio are usually compared to obtain a degree of difference to measure a similarity between two audio clips. However, how to improve the accuracy of matching audio clips is a technical problem that urgently needs to be resolved.

[0004] D1 EP0103258A1 discloses a pattern matching apparatus.

[0005] D2 "Hierarchical filtering method forcontent-based music retrieval via acoustic input" discloses a content-based music retrieval system.

[0006] D3 JPS62144200A discloses a speech recognition apparatus.

[0007] D4 "A comparative study of several dynamic time warping algorithms for speech recognition" discloses a comaprative study of several dynamic time warping algorithms for speech recognition.

SUMMARY

[0008] Embodiments of the invention provide a method and an apparatus for matching audio clips, a computer-readable medium, and an electronic device, to improve the accuracy of matching audio clips.

[0009] Other features and advantages of the invention become obvious through the following detailed descriptions, or may be learned partially through the practice of the invention.

[0010] The present invention is defined in the independent claims, and the preferable features according to the present invention are defined in the dependent claims. Any embodiment in the present disclosure that does not fall within the scope of protection of the present invention shall be regarded as an example for understanding the present invention but not belonging to the present invention.

[0011] The technical solutions provided in the embodiments of the invention achieve at least the following beneficial effects.

[0012] A first accumulation distance between a start position in a distance matrix and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix are calculated, to obtain a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance. Therefore, a minimum distance between two feature sequences can be comprehensively calculated in two directions (that is, a direction from a start position to a target position in a distance matrix and a direction from an end position to the target position in the distance matrix), so that matching relationships between the feature sequences in the two directions are both taken into account, and it can be ensured that the calculated minimum distance between the two feature sequences is more accurate, which helps to improve the accuracy of matching audio clips.

[0013] It is to be understood that, the foregoing general descriptions and the following detailed descriptions are merely for illustration and explanation purposes and are not intended to limit the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] To describe the technical solutions of the embodiments of the invention more clearly, the following briefly introduces the accompanying drawings for describing the embodiments. The accompanying drawings in the following description show only some embodiments of the invention, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a schematic diagram of a system architecture according to an example embodiment of the invention.

FIG. 2 is a flowchart of a method for matching audio clips according to an example not falling within the scope of the invention.

FIG. 3 is a schematic diagram of a distance matrix according to an example embodiment of the invention.

FIG. 4 is a flowchart of calculating a first accumulation distance according to an example embodiment of the invention.

FIG. 5 is a schematic diagram of calculation directions of calculating accumulation distances from a start position in a distance matrix according to an example embodiment of the invention.

FIG. 6 is a flowchart of calculating a second accumulation distance according to an example embodiment of the invention.

FIG. 7 is a schematic diagram of calculation directions of calculating accumulation distances from an end position in a distance matrix according to an example embodiment of the invention.

FIG. 8 is a flowchart of calculating a minimum distance between a first feature sequence and a second feature sequence according to an example embodiment of the invention.

FIG. 9 is a flowchart of determining a degree of matching between a first audio clip and a second audio clip according to an example embodiment of the invention.

FIG. 10 is a flowchart of a method of scoring by humming according to an example embodiment of the invention.

FIG. 11 is a schematic diagram of a global optimal path that is obtained according to an example embodiment of the invention.

FIG. 12 is a block diagram of an apparatus for matching audio clips according to an example embodiment of the invention.

FIG. 13 is a schematic structural diagram of a computer system of an electronic device according to an example embodiment of the invention.

DESCRIPTION OF EMBODIMENTS

[0015] To make the objectives, technical solutions, and advantages of the invention clearer, the following further describes implementations of the invention in detail with reference to the accompanying drawings.

[0016] FIG. 1 is a schematic diagram of an example system architecture to which a technical solution according to an embodiment of the invention is applicable.

[0017] As shown in FIG. 1, the system architecture may include a terminal device (which may be any one or more of a smart phone 101, a tablet computer 102, and a portable computer 103 shown in FIG. 1, or may be a desktop computer, and the like), a network 104, and a server 105. The network 104 is configured to provide a medium of a communication link between the terminal device and the server 105. The network 104 may include various connection types, for example, a wired communication link, and a wireless communication link.

[0018] It is to be understood that the quantities of terminal devices, networks, and servers in FIG. 1 are only schematic. There may be any quantities of terminal devices, networks, and servers as required. For example, the server 105 may be a server cluster including a plurality of servers.

[0019] In an embodiment of the invention, a user may upload a first audio clip (for example, corresponding to a sound sung by the user) to the server 105 by using a terminal device. After obtaining the first audio clip uploaded by the terminal device, the server 105 may extract a first feature sequence corresponding to the first audio clip, obtain a second audio clip (for example, an audio clip pre-stored in the server 105) that needs to be matched against the first audio clip, and extract a second feature sequence corresponding to the second audio clip. The server 105 then constructs a distance matrix between the first feature sequence and the second feature sequence, each element in the distance matrix being used for representing a distance between a corresponding one of first positions and a corresponding one of second positions, the first positions being in the first feature sequence, the second positions being in the second feature sequence.

[0020] In an embodiment of the invention, after a distance matrix is constructed, the server 105 may calculate a first accumulation distance between a start position in the distance matrix and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix, then calculate a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance, and determine a degree of matching between the first audio clip and the second audio clip according to the minimum distance. Therefore, in the technical solution of the embodiments of the invention, a minimum distance between feature sequences of audio clips may be comprehensively calculated in two directions (that is, a direction from a start position to a target position in a distance matrix and a direction from an end position to the target position in the distance matrix), so that matching relationships between the feature sequences in the two directions are both taken into account, and it may be ensured that the calculated minimum distance between the two feature sequences is more accurate, which improves the accuracy of matching audio clips.

[0021] Example embodiments of the technical solutions in the embodiments of the invention are described below in detail.

[0022] FIG. 2 is a flowchart of a method for matching audio clips according to an example not falling within the scope of the invention. Embodiments of the invention require that multiple sequences of features must be extracted from each audio clip, as defined in par. 145 below. The method for matching audio clips may be performed by a device having a computing processing function, for example, the server 105 shown in FIG. 1. As shown in FIG. 2, the method for matching audio clips at least includes the following operations.

[0023] Operation S210: A server obtains a first feature sequence corresponding to a first audio clip and a second feature sequence corresponding to a second audio clip.

[0024] In an embodiment of the invention, the first audio clip and the second audio clip are two audio clips that need to be compared with each other to determine a degree of matching. For example, the first audio clip is an audio clip inputted by the user (such as an audio clip sung by the user, an audio clip recorded by the user), and the second audio clip is an audio clip stored in a database.

[0025] In an embodiment of the invention, the first feature sequence and the second feature sequence are obtained for the same type of an audio feature, and the audio feature includes at least one of a pitch feature, a musical tone energy, a Mel-frequency cepstrum coefficient, and a root mean square energy value of each frame.

[0026] In an embodiment of the invention, the feature sequences of the first audio clip and the second audio clip are extracted by using at least one of an autocorrelation function algorithm, a Yin algorithm, and a PYin algorithm.

[0027] Operation S220: The server constructs a distance matrix between the first feature sequence and the second feature sequence, each element in the distance matrix being used for representing a distance between a corresponding one of first positions and a corresponding one of second positions, the first positions being in the first feature sequence, the second positions being in the second feature sequence.

[0028] In some embodiments, the size of the distance matrix is related to lengths of the first feature sequence and the second feature sequence. For example, when the length of the first feature sequence is m and the length of the second feature sequence is n, the size of the distance matrix is $m \times n$. For example, as shown in FIG. 3, assuming that the first feature sequence is 301 and the second feature sequence is 302, and a distance matrix 303 with the size of $m \times n$ is constructed. Each position in the distance matrix represents a distance between a point on the first feature sequence 301 and a point on the second feature sequence 302. For example, (i, j) in the distance matrix represents a distance between an i^{th} point on the first feature sequence 301 and a j^{th} point on the second feature sequence 302, and the distance includes an Euclidean distance.

[0029] Operation S230: The server determines a first accumulation distance between a start position and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix.

[0030] In an embodiment of the invention, the start position in the distance matrix is a position corresponding to the first feature point on the first feature sequence and the first feature point on the second feature sequence in the distance matrix; the end position in the distance matrix is a position corresponding to the last feature point on the first feature sequence and the last feature point on the second feature sequence in the distance matrix; and the target position includes any position other than the start position and the end position in the distance matrix.

[0031] In an embodiment of the invention, as shown in FIG. 4, a calculation process of the first accumulation distance between the start position and the target position includes the following operations S410 to S430:

[0032] Operation S410: The server determines accumulation distances between the start position and first candidate positions, the first candidate positions being located between the start position and the target position.

[0033] In an embodiment of the invention, when the accumulation distances are calculated from the start position in the distance matrix, the accumulation distances may be calculated one by one in three directions on the matrix, that are, in a direction toward the top, in a direction to the right, and in a direction toward the upper right corner as shown in FIG. 5.

[0034] That is, associations exist between the first candidate positions and the target position, the associations being used for indicating that the first candidate positions are located within a preset distance range around the target position. For example, if coordinates of the target position are (i, j), coordinates of a plurality of first candidate positions include: $(i - 1, j - 1)$, $(i - 1, j - x)$, and $(i - x, j - 1)$, where x is a natural number less than i or j. For example, x is 0, 2, 3, or the like.

[0035] In an embodiment of the invention, in the foregoing embodiment, the value of x may be greater than 1. In this case, the accumulation is performed at intervals of $(x - 1)$ positions when the accumulation distances are calculated, thereby accelerating the calculation process of the accumulation distances, which increases the calculation rate. However, to ensure the accuracy of calculation results, the value of x may be set not to be excessively large, and may be set to an appropriate value according to an actual requirement. For example, the value of x may be set to 2.

[0036] In an embodiment of the invention, the process of calculating the accumulation distances between the start position and the first candidate positions may be similar to the process of calculating the accumulation distances between the start position and the target position.

[0037] Operation S420: The server determines first candidate accumulation distances between the start position and the target position according to the accumulation distances between the start position and the first candidate positions and distance values represented by values of the first candidate positions.

[0038] In an embodiment of the invention, the process of calculating a plurality of first candidate accumulation distances between the start position and the target position includes: adding the accumulation distances between the start position and the first candidate positions and the distance values represented by the first candidate positions to obtain distance sum values corresponding to the first candidate positions, and determining the distance sum values as the first candidate accumulation distances corresponding to the first candidate positions.

[0039] For example, if a first candidate position is $(i - 1, j - 1)$ and a distance value represented by the first candidate position is $d(i - 1, j - 1)$, a distance sum value corresponding to the first candidate position may be expressed as $D\_forward(i - 1, j - 1) + d(i - 1, j - 1)$, where $D\_forward(i - 1, j - 1)$ represents an accumulation distance between the start position and the first candidate position in the distance matrix.

[0040] In an embodiment of the invention, to calculate the first candidate accumulation distances between the start position and the target position, a weighted calculation is first performed on the distance values represented by the first candidate positions according to the distance values represented by the first candidate positions and weight values corresponding to the first candidate positions, to obtain weighted distance values corresponding to the first candidate positions; and the accumulation distances between the start position and the first candidate positions and the weighted distance values corresponding to the first candidate positions are added to obtain distance sum values corresponding to the first candidate positions, and the distance sum values are determined as the first candidate accumulation distances corresponding to the first candidate positions. For example, if a first candidate position is $(i - 1, j - 1)$, a distance value represented by the first candidate position is $d(i - 1, j - 1)$, and a weight value corresponding to the first candidate position is w, a distance sum value corresponding to the first candidate position may be expressed as $D\_forward(i - 1, j - 1) + d(i - 1, j - 1) \times w$, where $D\_forward(i - 1, j - 1)$ represents an accumulation distance between the start position and the first candidate position in the distance matrix.

[0041] In some embodiments, the weight value corresponding to the foregoing first candidate position is determined according to a distance between the first candidate position and a diagonal of the distance matrix. That is, distances between the first candidate positions and a diagonal of the distance matrix are determined, and the weight values corresponding to the first candidate positions are determined according to the distances between the first candidate positions and the diagonal. The diagonal is a straight line connecting the start position and the end position. That is, in the foregoing embodiments, since distances between the first candidate positions and the diagonal of the distance matrix may be different, the weight values corresponding to the first candidate positions are considered. To avoid positions selected from the first candidate positions from deviating too far from the diagonal of the distance matrix, the weight values corresponding to the first candidate positions are set according to the distances between the first candidate positions and the diagonal of the distance matrix. For example, if a position is closer to the diagonal, a weight corresponding to the position is closer to 1; and if a position is farther from the diagonal, a weight corresponding to the position is larger.

[0042] Operation S430: The server determines a minimum value among the first candidate accumulation distances as the first accumulation distance.

[0043] Based on the technical solution of an embodiment shown in FIG. 4, the first accumulation distance between the start position in the distance matrix and the target position in the distance matrix may be obtained.

[0044] The calculating a second accumulation distance between an end position and the target position in the distance matrix in an embodiment of the invention is described below in with reference to FIG. 6, and specifically includes the following operations S610 to S630:

[0045] Operation S610: The server determines accumulation distances between the end position and second candidate positions, the second candidate positions being located between the target position and the end position.

[0046] In an embodiment of the invention, when the accumulation distances are calculated from the end position in the distance matrix, the accumulation distances may be calculated one by one in three directions on the matrix, that are, in a direction toward the bottom, in a direction toward the left, and in a direction toward the lower left corner as shown in FIG. 7. That is, associations exist between the second candidate positions and the target position, the associations being used for indicating that the second candidate positions are located within a preset range around the target position. For example, if coordinates of the target position are (i, j), coordinates of a plurality of second candidate positions include: $(i + 1, j + 1)$, $(i + 1, j + y)$, and $(i + y, j + 1)$, where y is a natural number. For example, y is 0, 2, 3, or the like.

[0047] In an embodiment of the invention, in the foregoing embodiment, the value of y may be greater than 1. In this case, the accumulation is performed at intervals of $(y - 1)$ positions when the accumulation distances are calculated, thereby accelerating the calculation process of the accumulation distances, which increases the calculation rate. However, to ensure the accuracy of calculation results, the value of y may be set not to be excessively large, and may be set to an appropriate value according to an actual requirement, for example, set to 2.

[0048] In an embodiment of the invention, the process of calculating the accumulation distances between the end position and the second candidate positions may be similar to the process of calculating the accumulation distances between the end position and the target position.

[0049] In operation S620, the server determines second candidate accumulation distances between the end position and the target position according to the accumulation distances between the end position and the second candidate positions and distance values represented by values of the second candidate positions.

[0050] In an embodiment of the invention, the process of calculating the second candidate accumulation distances between the end position and the target position includes: adding the accumulation distances between the end position and the second candidate positions and the distance values represented by the second candidate positions to obtain distance sum values corresponding to the second candidate positions, and determining the distance sum values as the second candidate accumulation distances corresponding to the second candidate positions.

[0051] For example, if a second candidate position is $(i + 1, j + 1)$ and a distance value represented by the second candidate position is $d(i + 1, j + 1)$, a distance sum value corresponding to the second candidate position may be expressed as $D\_backward(i + 1, j + 1) + d(i + 1, j + 1)$, where $D\_backward(i + 1, j + 1)$ represents an accumulation distance between the end position and the second candidate position in the distance matrix.

[0052] In an embodiment of the invention, the process of calculating the second candidate accumulation distances between the end position and the target position includes: performing a weighted calculation on the distance values represented by the second candidate positions according to the distance values represented by the second candidate positions and weight values corresponding to the second candidate positions, to obtain weighted distance values corresponding to the second candidate positions; and adding the accumulation distances between the end position and the second candidate positions and the weighted distance values corresponding to the second candidate positions to obtain distance sum values corresponding to the second candidate positions, and determining the distance sum values as the second candidate accumulation distances corresponding to the second candidate positions. For example, if a second candidate position is $(i + 1, j + 1)$, a distance value represented by the second candidate position is $d(i + 1, j + 1)$, and a weight value corresponding to the second candidate position is w, a distance sum value corresponding to the second candidate position may be expressed as $D\_backward(i + 1, j + 1) + d(i + 1, j + 1) \times w$, where $D\_backward(i + 1, j + 1)$ represents an accumulation distance between the end position and the second candidate position in the distance matrix.

[0053] In some embodiments, the weight value corresponding to the foregoing second candidate position is determined according to a distance between the second candidate position and a diagonal of the distance matrix. That is, distances between the second candidate positions and the diagonal of the distance matrix are determined, and the weight values corresponding to the second candidate positions are determined according to the distances between the second candidate positions and the diagonal. The diagonal is a straight line connecting the start position and the end position. That is, in the foregoing embodiments, since distances between the second candidate positions and the diagonal of the distance matrix may be different, the weight values corresponding to the second candidate positions are considered. To avoid positions selected from the second candidate positions from deviating too far from the diagonal of the distance matrix, the weight values corresponding to the second candidate positions are set according to the distances between the second candidate positions and the diagonal of the distance matrix. For example, if a position is closer to the diagonal, a weight corresponding to the position is closer to 1; and if a position is farther from the diagonal, a weight corresponding to the position is larger.

[0054] Operation S630: The server determines a minimum value among the second candidate accumulation distances as the second accumulation distance.

[0055] Based on the technical solution of an embodiment shown in FIG. 6, the second accumulation distance between the end position in the distance matrix and the target position in the distance matrix may be obtained.

[0056] Operation S240: The server calculates a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance.

[0057] In an embodiment of the invention, as shown in FIG. 8, the process of calculating a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance in operation S240 includes the following operations S810 and S820:

[0058] Operation S810: The server determines minimum accumulation distances corresponding to the target position based on a distance value represented by a value of the target position, the first accumulation distance, and the second accumulation distance.

[0059] In an embodiment of the invention, the distance value represented by the target position, the first accumulation distance, and the second accumulation distance are added to obtain the minimum accumulation distances corresponding to the target position. For example, if a target position is (i, j), a distance value represented by the target position is $d(i, j)$, a first accumulation distance is $D\_forward(i, j)$, and a second accumulation distance is $D\_backward(i, j)$, the minimum accumulation distance corresponding to the target position is $D\_total(i, j) = D\_forward(i, j) + D\_backward(i, j) + d(i, j)$.

[0060] In an embodiment of the invention, a weighted calculation is performed on the distance value represented by the target position and a weight value corresponding to the target position to obtain a weighted distance value corresponding to the target position. The weighted distance value corresponding to the target position, the first accumulation distance, and the second accumulation distance are then added to obtain the minimum accumulation distances corresponding to the target position. For example, if a target position is (i, j), a distance value represented by the target position is $d(i, j)$, a weight value corresponding to the target position is w, a first accumulation distance is $D\_forward(i, j)$, and a second accumulation distance is $D\_backward (i, j)$, the minimum accumulation distance corresponding to the target position is $D\_total(i, j) = D\_forward(i, j) + D\_backward(i, j) + d(i, j) \times w$. The weight value corresponding to the target position may be determined according to the distance between the target position and the diagonal of the distance matrix.

[0061] Referring to FIG. 8, in operation S820, the server selects a minimum value from the minimum accumulation distances corresponding to the target position as the minimum distance between the first feature sequence and the second feature sequence.

[0062] In an embodiment of the invention, the minimum distance between the first feature sequence and the second feature sequence is $\min_{i,j}\{D\_total(i,j)\}$.

[0063] Operation S250: The server determines a degree of matching between the first audio clip and the second audio clip according to the minimum distance.

[0064] In an embodiment of the invention, as shown in FIG. 9, the process of determining a degree of matching between the first audio clip and the second audio clip according to the minimum distance between the first feature sequence and the second feature sequence in operation S240 includes the following operations S910 to S930:

[0065] Operation S910: The server obtains n minimum distances between n first feature sequences and n second feature sequences, n being a positive integer.

[0066] In some embodiments, the first audio clip corresponds to n first feature sequences, and the second audio clip corresponds to n second feature sequences, where an i^{th} first feature sequence and an i^{th} second feature sequence correspond to the same type of the feature, and when the minimum distances are calculated, a minimum distance between the i^{th} first feature sequence and the i^{th} second feature sequence is calculated, i being a positive integer, and $i \leq n$. The feature of the audio clip includes at least one of a pitch feature, a musical tone energy, a Mel-frequency cepstrum coefficient, and a root mean square energy value of each frame. That is, for each feature, the first feature sequence of the first audio clip and the second feature sequence of the second audio clip are respectively obtained, and the minimum distance between the two feature sequences is calculated accordingly. In this manner, the minimum distances between the first feature sequence and the second feature sequence respectively corresponding to various features are obtained.

[0067] Operation S920: The server performs a weighted summation on the n minimum distances to obtain a weighted distance value between the first audio clip and the second audio clip.

[0068] In an embodiment of the invention, the weight corresponding to the feature may be set according to the importance of the feature. For example, if a feature is important, the weight corresponding to the feature is set to a large value; and if a feature is not very important, the weight corresponding to the feature is set to a small value to highlight the impact of important features on the weighted distance value, and weaken the impact of non-important features on the weighted distance value.

[0069] Operation S930: The server determines the degree of matching between the first audio clip and the second audio clip according to the weighted distance value.

[0070] In an embodiment of the invention, after the weighted distance value between the first audio clip and the second audio clip is calculated, the weighted distance value may be divided by a reference value (such as the length of the first feature sequence or the length of the second feature sequence) to obtain a matching score, and the degree of matching between the first audio clip and the second audio clip is then determined according to the matching score. For example, if the matching score is high, it is determined that the degree of matching between the first audio clip and the second audio clip is strong; and if the matching score is low, it is determined that the degree of matching between the first audio clip and the second audio clip is weak.

[0071] In an embodiment of the invention, the minimum distance between the first feature sequence and the second feature sequence corresponding to a feature may be directly used as the distance between the first audio clip and the second audio clip to determine the degree of matching between the first audio clip and the second audio clip.

[0072] In the technical solution of the embodiments of the invention, a minimum distance between feature sequences of audio clips is comprehensively calculated in two directions (that is, a direction from a start position to a target position in a distance matrix and a direction from an end position to the target position in the distance matrix), so that matching relationships between the feature sequences in the two directions are both taken into account, and it may be ensured that the calculated minimum distance between the two feature sequences is more accurate, which improves the accuracy of matching audio clips.

[0073] Schematically, by using an example of a humming scoring scenario, the technical solutions of the embodiments of the invention are described in detail:

As shown in FIG. 10, a method of scoring by humming according to an embodiment of the invention includes the following operations.

[0074] Operation S1001: A server acquires an audio clip corresponding to a sound sung by a user.

[0075] In an embodiment of the invention, the user sings a part of a specific song, and the terminal acquires an audio clip of the user based thereon, and records the start and end time points of the audio clip to obtain an audio duration.

Optionally, if the audio duration of the audio clip acquired by the terminal is less than a preset duration, the audio clip is filtered out, and information about a scoring failure is returned.

[0076] Operation S1002: The server extracts a pitch sequence of the audio clip.

[0077] In an embodiment of the invention, an autocorrelation function method, a Yin algorithm or a PYin algorithm may be used to extract the pitch sequence of the audio clip according to a specified sampling rate.

[0078] Operation S1003: The server extracts a musical instrument digital interface (MIDI) pitch sequence of a target song clip.

[0079] In an embodiment of the invention, a bottom layer of the humming scoring depends on a MIDI library, which is the source of scoring criteria. The audio clip corresponding to a sound sung by the user has a start timestamp and an end timestamp correspondingly, which may accurately correspond to a note sequence in the MIDI library, and the pitch sequence is then obtained according to a conversion formula of MIDI notes and pitches. Optionally, the MIDI pitch sequence of the target song clip is generated in the MIDI library in advance.

[0080] Operation S1003, operation S1001, and operation S1002 are not arranged in a rigid order.

[0081] Operation S1004: The server calculates a minimum distance between the two pitch sequences by using an audio information matching algorithm in the embodiments of the invention.

[0082] Operation S1005: The server converts a result of the minimum distance into a standard score by using a conversion formula.

[0083] In an embodiment of the invention, since the minimum distance in operation S1004 is obtained by accumulation, when the pitch sequence is longer, the value of the calculated minimum distance is larger. To eliminate this impact, the minimum distance calculated in operation S1004 may be divided by the pitch sequence length of the audio clip of the user to obtain a standard score, and the standard score is then fed back to the user.

[0084] In operation S1004, it is assumed that two pitch sequences are a sequence p and a sequence q respectively, where the length of the sequence p is m and the length of the sequence q is n, that is, $p = (p_1, p_2, ..., p_i, ..., p_m)$; and $q = (q_1, q_2, ..., q_j, ..., q_n)$. The solution of calculating a minimum distance between the two pitch sequences by using the audio information matching algorithm in the embodiments of the invention mainly includes the following operations:

[0085] Operation (1): Calculate a distance matrix and a weight matrix of the sequence p and the sequence q.

[0086] In an embodiment of the invention, a position (i, j) in the distance matrix represents a distance $d(i, j)$ between p_i and q_j. If the distance is a Euclidean distance, $d(i, j) = (p_i - q_j)^2$.

[0087] In an embodiment of the invention, the weight matrix considers a distance between an element position (i, j) in the distance matrix and a diagonal (that is, a straight line formed by the point (1, 1) and the point (m, n)) of the distance matrix. If the sequence p and the sequence q are closer, a finally calculated optimal path from the start position (that is, the point (1, 1)) to the end position (that is, the point (m, n)) of the distance matrix is closer to the diagonal. Therefore, penalty weights may be set for element positions far away from the diagonal. That is, when the element position is closer to the diagonal, the corresponding weight is closer to 1, and when the element position is farther from the diagonal, the corresponding weight is larger.

[0088] In an embodiment of the invention, a distance $t(i, j)$ between the position (i, j) in the distance matrix and the diagonal of the distance matrix may be approximately:

$$t(i,j) = \frac{|i \times n - j \times m|}{\sqrt{n^2 + m^2}}.$$

[0089] In an embodiment of the invention, a calculation formula of the position (i, j) in the weight matrix is adaptive smoothing of $t(i, j)$. That is, a weight $w(i, j)$ corresponding to the position (i, j) in the distance matrix may be calculated by using the following Formula 1:

$$w(i,j) = \left[1 + t(i,j) \times 0.025\right] \times \left[1 + \log(1 + t(i,j) \times 0.025)\right] \quad \text{Formula 1}$$

[0090] Values in Formula 1 are only used as an example.

[0091] Operation (2): Calculate a forward accumulation distance matrix, a forward source node matrix, a backward accumulation distance matrix, and a backward source node matrix according to the distance matrix and the weight matrix obtained in operation (1).

[0092] In an embodiment of the invention, the shortest distance is found by backtracking respectively from the start position and the end position to a middle position in the distance matrix. That is, an improved dynamic time warping (DTW) algorithm is proposed in the embodiments of the invention. By this algorithm, a bidirectional calculation may be performed to take into account head matching and tail matching of a sequence, so that the matching is more comprehensive.

[0093] In an embodiment of the invention, in a process of forward calculation from the start position in the distance matrix, to accelerate the process of distance accumulation and take into account the degree of deviation of the element position in the distance matrix from the diagonal (that is, the weight corresponding to the element position), the accumulation distance of the position (i, j) starts from three positions (i - 1, j - 1), (i - 1, j - 2), and (i - 2, j - 1), and a forward local decision function *D_forward*(i, j) may be defined as shown in the following Formula 2, which is used for representing the accumulation distance from the start position in the distance matrix to the position (i, j) in the distance matrix, to obtain the forward accumulation distance matrix.

$$D\_forward(i,j)=\min\begin{Bmatrix}D\_forward(i-1,j-1)+d(i-1,j-1)\times w(i-1,j-1),\\ D\_forward(i-1,j-2)+d(i-1,j-2)\times w(i-1,j-2),\\ D\_forward(i-2,j-1)+d(i-2,j-1)\times w(i-2,j-1)\end{Bmatrix}$$

Formula 2

[0094] Formula 2 may be adjusted to obtain the following Formula 3:

$$D\_forward(i,j)=\min\begin{Bmatrix}D\_forward(i-1,j-1)+d(i-1,j-1)\times\dfrac{w(i,j)^2}{w(i-1,j-1)},\\ D\_forward(i-1,j-2)+d(i-1,j-2)\times\dfrac{1.5\times w(i,j)^2}{w(i-1,j-2)},\\ D\_forward(i-2,j-1)+d(i-2,j-1)\times\dfrac{1.5\times w(i,j)^2}{w(i-2,j-1)}\end{Bmatrix}$$

Formula 3

[0095] In the foregoing embodiment, the forward calculation from the start position in the distance matrix starts from the lower left corner (1,1) of the distance matrix, and a calculation is performed in each row from left to right. While a forward accumulation distance *D_forward*(i, j) is calculated, a subscript, that is, one of (i - 1, j - 1), (i - 1, j - 2), and (i - 2, j - 1), of a source node of *D_forward*(i, j) is stored at the position (i, j) in the forward source node matrix.

[0096] In an embodiment of the invention, a process of backward calculation from the end position in the distance matrix is similar to the solution of forward calculation in the previous embodiment, except that the calculation direction starts from the end position in the distance matrix, that is, starts from a position (m, n) at the upper right corner of the distance matrix, and the accumulation distance of the position (i, j) starts from three positions (i+1, j+1), (i+1, j+2), and (i+2, j+1). The backward local decision function *D_backward*(i, j) is defined as shown in the following Formula 4, which is used for representing the accumulation distance from the end position in the distance matrix to the position (i, j) in the distance matrix, to obtain the backward accumulation distance matrix.

$$D\_backward(i,j)=\min\begin{Bmatrix}D\_backward(i+1,j+1)+d(i+1,j+1)\times w(i+1,j+1),\\ D\_backward(i+1,j+2)+d(i+1,j+2)\times w(i+1,j+2),\\ D\_backward(i+2,j+1)+d(i+2,j+1)\times w(i+2,j+1)\end{Bmatrix}$$

Formula 4

[0097] Formula 4 may be adjusted to obtain the following Formula 5:

$$D\_backward(i,j)=\min\begin{Bmatrix} D\_backward(i+1,j+1)+d(i+1,j+1)\times\frac{w(i,j)^2}{w(i+1,j+1)}, \\ D\_backward(i+1,j+2)+d(i+1,j+2)\times\frac{1.5\times w(i,j)^2}{w(i+1,j+2)}, \\ D\_backward(i+2,j+1)+d(i+2,j+1)\times\frac{1.5\times w(i,j)^2}{w(i+2,j+1)} \end{Bmatrix}$$

Formula 5

where w represents a weight value, and d represents a distance value.

[0098] In the foregoing embodiment, the backward calculation from the end position in the distance matrix starts from the upper right corner (*m*, *n*) of the distance matrix, and a calculation is performed from right to left in each row. While a backward accumulation distance *D_backward*(*i*, *j*) is calculated, a subscript, that is, one of (*i*+1, *j*+1), (*i*+1, *j*+2), and (*i*+2, *j*+1) , of a source node of *D_backward*(*i*, *j*) is stored at the position (*i*, *j*) in the backward source node matrix.

[0099] Operation (3): Obtain the minimum distance and the shortest path from the forward accumulation distance matrix and the backward accumulation distance matrix.

[0100] In an embodiment of the invention, at any position (*i*, *j*) in the distance matrix, the shortest path connecting to (*i*, *j*) may be found from the lower left corner and the upper right corner, where a calculation formula of the shortest distance is shown in the following Formula 6:

$$D\_total(i,j)=d(i,j)\times w(i,j)+D\_forward(i,j)+D\_backward(i,j) \quad \text{Formula 6}$$

[0101] Based on Formula 6, the minimum distance min_*dist* is calculated by using the following Formula 7:

$$\min\_dist=\min_{i,j}\{D\_total(i,j)\} \quad \text{Formula 7}$$

[0102] The forward source node matrix and the backward source node matrix are found at the position corresponding to the minimum distance, and subscripts of previous nodes are obtained and traversed in turn to obtain a forward path (from (1,1) to (*i*, *j*)) and a backward path (from (*m*, *n*) to (*i*, *j*)). Based on the two paths, the minimum value of D_total (i, j) is the global optimal path corresponding to the minimum distance. Specifically, as shown in FIG. 11, the two pitch sequences are 1101 and 1102, and the global optimal path 1103 is finally obtained according to the technical solution of the embodiments of the invention.

[0103] The technical solutions in the embodiments of the invention take into consideration both the head priority matching and the tail priority matching of the pitch sequence, so that the matching is more comprehensive. Moreover, when the accumulation distance is calculated, an offset between the position and the diagonal of the distance matrix is taken into account, which prevents the final optimal path from greatly deviating from the diagonal of the distance matrix, thereby making the sequence matching more robust.

[0104] The following describes apparatus embodiments of the invention, which may be used for performing the method for matching audio clips in the foregoing embodiments of the invention. For details not disclosed in the apparatus embodiments of the invention, reference may be made to the method for matching audio clips in the foregoing embodiments of the invention.

[0105] FIG. 12 is a block diagram of an apparatus for matching audio clips according to an embodiment of the invention.

[0106] Referring to FIG. 12, an apparatus 1200 for matching audio clips according to an embodiment of the invention includes an obtaining unit 1202, a construction unit 1204, and a processing unit 1206.

[0107] The obtaining unit 1202 is configured to obtain a first feature sequence corresponding to a first audio clip and a second feature sequence corresponding to a second audio clip.

[0108] The construction unit 1204 is configured to: obtain the first feature sequence and the second feature sequence from the obtaining unit 1202, and construct a distance matrix between the first feature sequence and the second feature sequence, each element in the distance matrix being used for representing a distance between a corresponding one of first positions and a corresponding one of second positions, the first positions being in the first feature sequence, the second positions being in the second feature sequence.

[0109] The processing unit 1206 is configured to: obtain the distance matrix from the construction unit 1204, and determine a first accumulation distance between a start position and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix; determine a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance; and determine a degree of matching between the first audio clip and the second audio clip according to the minimum distance.

[0110] In an optional embodiment, the processing unit 1206 includes:

a determining subunit, configured to determine accumulation distances between the start position and first candidate positions, the first candidate positions being located between the start position and the target position,

the determining subunit being further configured to: determine first candidate accumulation distances between the start position and the target position according to the accumulation distances between the start position and the first candidate positions and distance values represented by values of the first candidate positions; and determine a minimum value among the first candidate accumulation distances as the first accumulation distance.

[0111] In an optional embodiment, the determining subunit is further configured to add the accumulation distances and the distance values represented by the first candidate positions to obtain distance sum values corresponding to the first candidate positions, the accumulation distances being distances between the start position and the first candidate positions; and

the determining subunit is further configured to determine the distance sum values as the first candidate accumulation distances corresponding to the first candidate positions.

[0112] In an optional embodiment, the determining subunit is further configured to perform a weighted calculation on the distance values represented by the first candidate positions according to the distance values represented by the first candidate positions and weight values corresponding to the first candidate positions, to obtain weighted distance values corresponding to the first candidate positions; and

the determining subunit is further configured to: add the accumulation distances between the start position and the first candidate positions and the weighted distance values corresponding to the first candidate positions to obtain distance sum values corresponding to the first candidate positions, and determine the distance sum values as the first candidate accumulation distances corresponding to the first candidate positions.

[0113] In an optional embodiment, the determining subunit is further configured to: determine distances between the first candidate positions and a diagonal of the distance matrix, the diagonal being a straight line connecting the start position and the end position; and determine the weight values corresponding to the first candidate positions according to the distances between the first candidate positions and the diagonal.

[0114] In an optional embodiment, associations exist between the first candidate positions and the target position, the associations being used for indicating that the first candidate positions are located within a preset distance range around the target position.

[0115] In an optional embodiment, the processing unit 1206 includes:

a determining subunit, configured to determine accumulation distances between the end position and second candidate positions, the second candidate positions being located between the target position and the end position,

the determining subunit being further configured to: determine second candidate accumulation distances between the end position and the target position according to the accumulation distances between the end position and the second candidate positions and distance values represented by values of the second candidate positions; and determine a minimum value among the second candidate accumulation distances as the second accumulation distance.

[0116] In an optional embodiment, associations exist between the second candidate positions and the target position, the associations being used for indicating that the second candidate positions are located within a preset distance range around the target position.

[0117] In an optional embodiment, the processing unit 1206 includes:

a determining subunit, configured to: determine minimum accumulation distances corresponding to the target position based on a distance value represented by a value of the target position, the first accumulation distance, and the second accumulation distance; and select a minimum value from the minimum accumulation distances corresponding to the target position, and determine the minimum value as the minimum distance between the first feature sequence and the second feature sequence.

[0118] In an optional embodiment, the determining subunit is further configured to add the distance value represented by the target position, the first accumulation distance, and the second accumulation distance to obtain the minimum accumulation distances corresponding to the target position; or the determining subunit is further configured to: perform a weighted calculation on the distance value represented by the target position and a weight value corresponding to the target position to obtain a weighted distance value corresponding to the target position; and add the weighted distance value, the first accumulation distance, and the second accumulation distance to obtain the minimum accumulation distances corresponding to the target position.

[0119] According to the invention, the first audio clip corresponds to n first feature sequences, and the second audio clip corresponds to n second feature sequences, n being a positive integer;

the obtaining unit 1202 is further configured to obtain n minimum distances between the n first feature sequences and the n second feature sequences; and

the processing unit 1206 is further configured to: perform a weighted summation on the n minimum distances to obtain a weighted distance value between the first audio clip and the second audio clip; and determine the degree of matching between the first audio clip and the second audio clip according to the weighted distance value.

[0120] In an optional embodiment, based on the foregoing solution, the features of an audio clip include a pitch feature, a musical tone energy, a Mel-frequency cepstrum coefficient, and a root mean square energy value of each frame.

[0121] The obtaining unit 1202 may be implemented by a memory in a computer device, or by a processor in a computer device, or by a memory and a processor together; and the construction unit 1204 and the processing unit 1206 are implemented by a processor in a computer device.

[0122] FIG. 13 is a schematic structural diagram of a computer system adapted to implement an electronic device according to an embodiment of the invention.

[0123] A computer system 1300 of the electronic device shown in FIG. 13 is merely an example, and does not constitute any limitation on functions and use ranges of the embodiments of the invention.

[0124] As shown in FIG. 13, the computer system 1300 includes a central processing unit (CPU) 1301, which may perform various suitable actions and processing based on a program stored in a read-only memory (ROM) 1302 or a program loaded from a storage part 1308 into a random access memory (RAM) 1303, for example, perform the method described in the foregoing embodiments. The RAM 1303 further stores various programs and data required for system operations. The CPU 1301, the ROM 1302, and the RAM 1303 are connected to each other through a bus 1304. An input/output (I/O) interface 1305 is also connected to the bus 1304.

[0125] The following components are connected to the I/O interface 1305: an input part 1306 including a keyboard, a mouse, or the like, an output part 1307 including a cathode ray tube (CRT), a liquid crystal display (LCD), a speaker, or the like, a storage part 1308 including a hard disk, or the like, and a communication part 1309 including a network interface card such as a local area network (LAN) card or a modem. The communication part 1309 performs communication processing by using a network such as the Internet. A driver 1310 is also connected to the I/O interface 1305 as required. A removable medium 1311, such as a magnetic disk, an optical disc, a magneto-optical disk, or a semiconductor memory, is installed on the drive 1310 as required, so that a computer program read from the removable medium is installed into the storage part 1308 as required.

[0126] Particularly, according to an embodiment of the invention, the processes described in the following by referring to the flowcharts may be implemented as computer software programs. For example, this embodiment of the invention includes a computer program product, the computer program product includes a computer program carried on a computer-readable medium, and the computer program includes program code used for performing the methods shown in the flowcharts. In such an embodiment, the computer program may be downloaded and installed from a network through the communication part 1309, and/or installed from the removable medium 1311. When the computer program is executed by the CPU 1301, the various functions defined in the system of the invention are executed.

[0127] The computer-readable medium shown in the embodiments of the invention may be a computer-readable signal medium or a computer-readable storage medium or any combination of the two. The computer-readable storage medium may be, for example, but is not limited to, an electric, magnetic, optical, electromagnetic, infrared, or semi-conductive system, apparatus, or component, or any combination of the above. A more specific example of the computer-readable storage medium may include but is not limited to: an electrical connection having one or more wires, a portable computer magnetic disk, a hard disk, a RAM, a ROM, an erasable programmable read-only memory (EPROM), a flash memory, an optical fiber, a compact disk read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any appropriate combination thereof. In the invention, the computer-readable storage medium may be any tangible medium containing or storing a program, and the program may be used by or used in combination with an instruction execution system, an apparatus, or a device. In the invention, a computer-readable signal medium may include a data signal being in a baseband or propagated as a part of a carrier wave, the data signal carrying computer-readable program code. A data signal propagated in such a way may assume a plurality of forms, including, but not limited to, an electromagnetic signal, an optical signal, or any appropriate combination thereof. The computer-readable signal medium may be further any computer readable medium in addition to a computer-readable storage medium. The computer readable medium may send, propagate, or transmit a program that is used by or used in conjunction with an instruction execution system, an apparatus, or a device. The program code included in the computer-readable medium may be transmitted by using any suitable medium, including but not limited to: a wireless medium, a wire, or the like, or any suitable combination thereof.

[0128] The flowcharts and block diagrams in the accompanying drawings illustrate possible system architectures, functions and operations that may be implemented by a system, a method, and a computer program product according to various embodiments of the invention. Each box in a flowchart or a block diagram may represent a module, a program segment, or a part of code. The module, the program segment, or the part of code includes one or more executable instructions used for implementing specified logic functions. In some implementations used as substitutes, functions annotated in boxes may alternatively occur in a sequence different from that annotated in an accompanying drawing. For example, two boxes shown in succession may be performed basically in parallel, and sometimes the two boxes may be performed in a reverse sequence. This is determined by a related function. Each box in a block diagram and/or a flowchart and a combination of boxes in the block diagram and/or the flowchart may be implemented by using a dedicated hardware-based system configured to perform a specified function or operation, or may be implemented by using a combination of dedicated hardware and a computer instruction.

[0129] A related unit described in the embodiments of the invention may be implemented in a software manner, or may be implemented in a hardware manner, and the unit described may also be set in a processor. Names of the units do not constitute a limitation on the units in a specific case.

[0130] According to another aspect, the invention further provides a computer-readable medium. The computer-readable medium may be included in the electronic device described in the foregoing embodiments, or may exist alone and is not disposed in the electronic device. The computer-readable medium carries one or more programs, the one or more programs, when executed by the electronic device, causing the electronic device to implement the method described in the foregoing embodiments.

[0131] Although a plurality of modules or units of a device configured to perform actions are discussed in the foregoing detailed description, such division is not mandatory. According to the implementations of the invention, the features and functions of two or more modules or units described above may be specifically implemented in one module or unit. On the contrary, the features and functions of one module or unit described above may be further divided to be embodied by a plurality of modules or units.

[0132] According to the foregoing descriptions of the implementations, a person skilled in the art may readily understand that the example implementations described herein may be implemented by using software, or may be implemented by combining software and necessary hardware. Therefore, the technical solutions of the embodiments of the invention may be implemented in a form of a software product. The software product may be stored in a non-volatile storage medium (which may be a CD-ROM, a USB flash drive, a removable hard disk, or the like) or on the network, including several instructions for instructing a computing device (which may be a personal computer, a server, a touch terminal, a network device, or the like) to perform the methods according to the embodiments of the invention.

[0133] It is to be understood that the invention is not limited to the precise structures described above and shown in the accompanying drawings, and various modifications and changes may be made without departing from the scope of the invention. The scope of the invention is subject only to the appended claims.

Claims

1. A method for matching audio clips, performed by a computer device, wherein the method for matching audio clips is performed based on an improved dynamic time warping algorithm, and the method comprises:

    obtaining (S210) a first feature sequence corresponding to a first audio clip and a second feature sequence corresponding to a second audio clip;
    constructing (S220) a distance matrix between the first feature sequence and the second feature sequence, elements in the distance matrix representing respective distances between first positions in the first feature sequence and second positions in the second feature sequence;
    determining (S230) a first accumulation distance between a start position and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix;
    determining (S240) a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance;
    characterized in that, the first audio clip corresponds to a plurality of n first feature sequences, and the second audio clip corresponds to a plurality of n second feature sequences, n being a positive integer;
    and **in that** method further comprises:

        obtaining (S910) n minimum distances between the n first feature sequences and the n second feature

sequences;
performing (S920) a weighted summation on the n minimum distances to obtain a weighted distance value between the first audio clip and the second audio clip; and
determining (S930) the degree of matching between the first audio clip and the second audio clip according to the weighted distance value.

2. The method according to claim 1, wherein the determining the first accumulation distance between the start position and the target position in the distance matrix comprises:

determining (S410) accumulation distances between the start position and first candidate positions, the first candidate positions being located between the start position and the target position in the distance matrix;
determining (S420) each of first candidate accumulation distances between the start position and the target position based on an accumulation distance between the start position and a corresponding first candidate position and a distance value represented by the corresponding first candidate position in the distance matrix; and
determining (S430), as the first accumulation distance, a minimum value among the first candidate accumulation distances.

3. The method according to claim 2, wherein the determining each of the first candidate accumulation distances between the start position and the target position based on the accumulation distance between the start position and the corresponding first candidate position and the distance value represented by the corresponding first candidate position comprises, with respect to each of the first candidate positions:

adding the accumulation distance between the start position and the corresponding first candidate position and the distance value represented by the corresponding first candidate position to obtain a distance sum value of the corresponding first candidate position, the accumulation distance of the corresponding first candidate position being based on distances between the start position and previous positions of the corresponding first candidate position, the previous positions being located between the start position and the corresponding first candidate position in the distance matrix; and
determining the distance sum value as a first candidate accumulation distance of the corresponding first candidate position.

4. The method according to claim 2, wherein the determining each of the first candidate accumulation distances between the start position and the target position based on the accumulation distance between the start position and the corresponding first candidate position and the distance value represented by the corresponding first candidate position comprises, with respect to each of the first candidate positions:

performing a weighted calculation on the distance value represented by the corresponding first candidate position by using a weight value of the corresponding first candidate position, to obtain a weighted distance value of the corresponding first candidate position;
adding the accumulation distance between the start position and the corresponding first candidate position and the weighted distance value of the corresponding first candidate position to obtain a distance sum value of the corresponding first candidate position, the accumulation distance of the corresponding first candidate position being based on distances between the start position and previous positions of the corresponding first candidate position, the previous positions being located between the start position and the corresponding first candidate position in the distance matrix; and
determining the distance sum value as a first candidate accumulation distance of the corresponding first candidate position.

5. The method according to claim 4, further comprising, prior to the determining the distance sum value as the first candidate accumulation distance of the corresponding first candidate position:

determining a distance between the corresponding first candidate position and a diagonal of the distance matrix, the diagonal being a straight line connecting the start position and the end position; and
determining the weight value of the corresponding first candidate position according to the distance between the corresponding first candidate position and the diagonal.

6. The method according to claim 2, wherein
the first candidate positions are located within a preset distance range around the target position.

7. The method according to claim 1, wherein the calculating the second accumulation distance between the end position and the target position in the distance matrix comprises:

determining (S610) accumulation distances between the end position and second candidate positions, the second candidate positions being located between the target position and the end position in the distance matrix;
determining (S620) each of second candidate accumulation distances between the end position and the target position based on an accumulation distance between the end position and a corresponding second candidate position and a distance value represented by the corresponding second candidate position in the distance matrix; and
determining (S630), as the second accumulation distance, a minimum value among the second candidate accumulation distances.

8. The method according to claim 7, wherein
the second candidate positions are located within a preset distance range around the target position.

9. The method according to any one of claims 1 to 8, wherein the determining the minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance comprises:

determining (S810) minimum accumulation distances corresponding to the target position based on a distance value represented by the target position, the first accumulation distance, and the second accumulation distance; and
selecting (S820) a minimum value from the minimum accumulation distances corresponding to the target position, and determining the minimum value as the minimum distance between the first feature sequence and the second feature sequence.

10. The method according to claim 9, wherein the determining the minimum accumulation distances corresponding to the target position based on the distance value represented by the target position, the first accumulation distance, and the second accumulation distance comprises:

adding the distance value represented by the target position, the first accumulation distance, and the second accumulation distance to obtain the minimum accumulation distances corresponding to the target position; or
performing a weighted calculation on the distance value represented by the target position by using a weight value corresponding to the target position to obtain a weighted distance value corresponding to the target position; and adding the weighted distance value, the first accumulation distance, and the second accumulation distance to obtain the minimum accumulation distances corresponding to the target position.

11. The method for matching audio clips according to claim 1, wherein the method is performed by a server.

12. A computer-readable medium, storing a computer program, the computer program, when executed by a processor, implementing the method for matching audio clips according to any one of claims 1 to 11.

13. An electronic device, comprising:

one or more processors; and
a storage apparatus, configured to store one or more programs, the one or more programs, when executed by the one or more processors, causing the one or more processors to implement the method for matching audio clips according to any one of claims 1 to 11.

Patentansprüche

1. Verfahren zum Abgleichen von Audioclips, das von einer Computervorrichtung durchgeführt wird, wobei das Verfahren zum Abgleichen von Audioclips auf der Grundlage eines verbesserten dynamischen Time-Warping-Algorithmus durchgeführt wird, und das Verfahren umfasst:

Erhalten (S210) einer ersten Merkmalssequenz, die einem ersten Audioclip entspricht, und einer zweiten Merkmalssequenz, die einem zweiten Audioclip entspricht;

Konstruieren (S220) einer Abstandsmatrix zwischen der ersten Merkmalssequenz und der zweiten Merkmalssequenz, wobei Elemente in der Abstandsmatrix jeweilige Abstände zwischen ersten Positionen in der ersten Merkmalssequenz und zweiten Positionen in der zweiten Merkmalssequenz darstellen;
Bestimmen (S230) eines ersten Summenabstandes zwischen einer Startposition und einer Zielposition in der Abstandsmatrix und eines zweiten Summenabstandes zwischen einer Endposition und der Zielposition in der Abstandsmatrix;
Bestimmen (S240) eines minimalen Abstands zwischen der ersten Merkmalssequenz und der zweiten Merkmalssequenz auf der Grundlage des ersten Summenabstands und des zweiten Summenabstands;
dadurch gekennzeichnet, dass der erste Audioclip einer Vielzahl von n ersten Merkmalssequenzen entspricht und der zweite Audioclip einer Vielzahl von n zweiten Merkmalssequenzen entspricht, wobei n eine positive ganze Zahl ist; und dadurch, dass das Verfahren ferner umfasst:

Erhalten (S910) von n minimalen Abständen zwischen den n ersten Merkmalssequenzen und den n zweiten Merkmalssequenzen;
Durchführen (S920) einer gewichteten Summierung über die n minimalen Abstände, um einen gewichteten Abstandswert zwischen dem ersten Audioclip und dem zweiten Audioclip zu erhalten; und
Bestimmen (S930) des Grades der Übereinstimmung zwischen dem ersten Audioclip und dem zweiten Audioclip gemäß dem gewichteten Abstandswert.

2. Verfahren gemäß Anspruch 1, wobei das Bestimmen der ersten Summenabstands zwischen der Startposition und der Zielposition in der Abstandsmatrix umfasst:

Bestimmen (S410) von Summenabständen zwischen der Startposition und ersten Kandidatenpositionen, wobei sich die ersten Kandidatenpositionen zwischen der Startposition und der Zielposition in der Abstandsmatrix befinden;
Bestimmen (S420) jeder von ersten Kandidatensummenabständen zwischen der Startposition und der Zielposition auf der Grundlage eines Summenabstands zwischen der Startposition und einer entsprechenden ersten Kandidatenposition und einem Abstandswert, der durch die entsprechende erste Kandidatenposition in der Abstandsmatrix dargestellt wird; und
Bestimmen (S430) eines Minimalwerts unter den ersten Kandidatensummenabständen als ersten Summenabstand.

3. Verfahren nach Anspruch 2, wobei das Bestimmen jeder der ersten Kandidatensummenabstände zwischen der Startposition und der Zielposition auf der Grundlage des Summenabstands zwischen der Startposition und der entsprechenden ersten Kandidatenposition und des Abstandswertes, der durch die entsprechende erste Kandidatenposition dargestellt wird, in Bezug auf jede der ersten Kandidatenpositionen umfasst:

Zugeben des Summenabstands zwischen der Startposition und der entsprechenden ersten Kandidatenposition und des Abstandswertes, der durch die entsprechende erste Kandidatenposition repräsentiert wird, um einen Summenabstandswert der entsprechenden ersten Kandidatenposition zu erhalten, wobei der Summenabstand der entsprechenden ersten Kandidatenposition auf Abständen zwischen der Startposition und vorherigen Positionen der entsprechenden ersten Kandidatenposition basiert, wobei die vorherigen Positionen zwischen der Startposition und der entsprechenden ersten Kandidatenposition in der Abstandsmatrix angeordnet sind; und
Bestimmen des Summenabstandswerts als erste Kandidatensummenabstand der entsprechenden ersten Kandidatenposition.

4. Verfahren nach Anspruch 2, wobei das Bestimmen jeder der ersten Kandidatensummenabstände zwischen der Startposition und der Zielposition auf der Grundlage des Summenabstands zwischen der Startposition und der entsprechenden ersten Kandidatenposition und des Abstandswertes, der durch die entsprechende erste Kandidatenposition dargestellt wird, in Bezug auf jede der ersten Kandidatenpositionen umfasst:

Durchführen einer gewichteten Berechnung des Abstandswertes, der durch die entsprechende erste Kandidatenposition repräsentiert wird, unter Anwendung eines Gewichts-% der entsprechenden ersten Kandidatenposition, um einen gewichteten Abstandswert der entsprechenden ersten Kandidatenposition zu erhalten;
Zugeben des Summenabstands zwischen der Startposition und der entsprechenden ersten Kandidatenposition und des gewichteten Abstandswertes der entsprechenden ersten Kandidatenposition, um einen Summenabstandswert der entsprechenden ersten Kandidatenposition zu erhalten, wobei der Summenabstand der entsprechenden ersten Kandidatenposition auf Abstände zwischen der Startposition und vorherigen Positionen

der entsprechenden ersten Kandidatenposition basiert, wobei die vorherigen Positionen zwischen der Startposition und der entsprechenden ersten Kandidatenposition in der Abstandsmatrix angeordnet sind; und
Bestimmen des Summenabstandswerts als ersten Kandidatensummenabstand der entsprechenden ersten Kandidatenposition.

5. Verfahren nach Anspruch 4, ferner umfassend, dass vor der Bestimmung des Summenabstandswerts als ersten Kandidatensummenabstand der entsprechenden ersten Kandidatenposition:

Bestimmen eines Abstands zwischen der entsprechenden ersten Kandidatenposition und einer Diagonale der Abstandsmatrix, wobei die Diagonale eine gerade Linie ist, die die Startposition und die Endposition verbindet; und
Bestimmen des Gewichts-% der entsprechenden ersten Kandidatenposition entsprechend dem Abstand zwischen der entsprechenden ersten Kandidatenposition und der Diagonale.

6. Verfahren nach Anspruch 2, wobei
sich die ersten Kandidatenpositionen innerhalb eines vorgegebenen Abstandsbereichs um die Zielposition befinden.

7. Verfahren gemäß Anspruch 1, wobei die Berechnung des zweiten Summenabstands zwischen der Endposition und der Zielposition in der Abstandsmatrix umfasst:

Bestimmen (S610) von Summenabständen zwischen der Endposition und zweiten Kandidatenpositionen, wobei sich die zweiten Kandidatenpositionen zwischen der Zielposition und der Endposition in der Abstandsmatrix befinden;
Bestimmen (S620) jeder von zweiten Kandidatensummenabstände zwischen der Endposition und der Zielposition auf der Grundlage eines Summenabstands zwischen der Endposition und einer entsprechenden zweiten Kandidatenposition und einem Abstandswert, der durch die entsprechende zweite Kandidatenposition in der Abstandsmatrix dargestellt wird; und
Bestimmen (S630) eines Minimalwerts unter den zweiten Kandidatensummenabständen als der zweite Summenabstand.

8. Verfahren nach Anspruch 7, wobei
sich die zweiten Kandidatenpositionen innerhalb eines vorgegebenen Abstandsbereichs um die Zielposition befinden.

9. Verfahren gemäß einem der Ansprüche 1 bis 8, wobei das Bestimmen des minimalen Abstands zwischen der ersten Merkmalssequenz und der zweiten Merkmalssequenz basierend auf dem ersten Summenabstand und dem zweiten Summenabstand umfasst:

Bestimmen (S810) minimaler Summenabstände, die der Zielposition entsprechen, basierend auf einem Abstandswert, der durch die Zielposition, den ersten Summenabstand und den zweiten Summenabstand dargestellt wird; und
Auswählen (S820) eines minimalen Wertes aus den minimalen Summenabständen, die der Zielposition entsprechen, und Bestimmen des minimalen Wertes als den minimalen Abstand zwischen der ersten Merkmalssequenz und der zweiten Merkmalssequenz.

10. Verfahren gemäß Anspruch 9, wobei das Bestimmen der minimalen Summenabstände, die der Zielposition entsprechen, basierend auf dem Abstandswert, der durch die Zielposition, den ersten Summenabstand und den zweiten Summenabstand dargestellt wird, umfasst:

Zugeben des Abstandswertes, der durch die Zielposition, den ersten Summenabstand und den zweiten Summenabstand dargestellt wird, um die minimalen Summenabstände zu erhalten, die der Zielposition entsprechen; oder
Durchführen eines gewichteten Berechnens des durch die Zielposition dargestellten Abstandswerts unter Anwendung eines der Zielposition entsprechenden Gewichts-%, um einen der Zielposition entsprechenden gewichteten Abstandswert zu erhalten; und Zugeben des gewichteten Abstandswerts, des ersten Summenabstands und des zweiten Summenabstands, um die der Zielposition entsprechenden minimalen Summenabstände zu erhalten.

11. Verfahren zum Abgleichen von Audioclips gemäß Anspruch 1, wobei das Verfahren von einem Server durchgeführt wird.

12. Computerlesbares Medium, das ein Computerprogramm speichert, wobei das Computerprogramm, wenn es von einem Prozessor ausgeführt wird, das Verfahren zum Abgleichen von Audioclips gemäß einem der Ansprüche 1 bis 11 implementiert.

13. Elektronische Vorrichtung, umfassend:

einen oder mehrere Prozessoren; und
eine Speichervorrichtung, die so konfiguriert ist, dass sie ein oder mehrere Programme speichert, wobei das eine oder die mehreren Programme, wenn sie von dem einen oder den mehreren Prozessoren ausgeführt werden, den einen oder die mehreren Prozessoren veranlassen, das Verfahren zum Abgleichen von Audioclips gemäß einem der Ansprüche 1 bis 11 zu implementieren.

Revendications

1. Procédé de mise en correspondance d'audioclips, exécuté par un dispositif informatique, le procédé de mise en correspondance d'audioclips étant exécuté sur la base d'un algorithme de déformation temporelle dynamique améliorée, et le procédé comprenant:

l'obtention (S210) d'une première séquence de caractéristiques correspondant à un premier audioclip et une seconde séquence de caractéristiques correspondant à un second audioclip;
la construction (S220) d'une matrice de distance entre la première séquence de caractéristiques et la seconde séquence de caractéristiques, les éléments dans la matrice de distance représentant des distances respectives entre des premières positions dans la première séquence de caractéristiques et des secondes positions dans la seconde séquence de caractéristiques;
la détermination (S230) d'une première distance d'accumulation entre une position de départ et une position cible dans la matrice de distance et d'une seconde distance d'accumulation entre une position finale et la position cible dans la matrice de distance;
la détermination (S240) d'une distance minimum entre la première séquence de caractéristiques et la seconde séquence de caractéristiques sur la base de la première distance d'accumulation et de la seconde distance d'accumulation;
caractérisé en ce que, le premier audioclip correspond à une pluralité de n premières séquences de caractéristiques, et le second audioclip correspondant à une pluralité de n secondes séquences de caractéristiques, n étant un nombre entier positif, et **en ce que** le procédé comprend en outre:

l'obtention (S910) de n distances minimum entre les n premières séquences de caractéristiques et les n secondes séquences de caractéristiques;
la réalisation (S920) d'une somme pondérée sur les n distances minimum pour obtenir une valeur de distance pondérée entre le premier audioclip et le second audioclip; et
la détermination (S930) du degré de mise en correspondance entre le premier audioclip et le second audioclip conformément à la valeur de distance pondérée.

2. Procédé selon la revendication 1, dans lequel la détermination de la première distance d'accumulation entre la position de départ et la position cible dans la matrice de distance comprend:

la détermination (S410) de distances d'accumulation entre la position de départ et des premières positions candidates, les premières positions candidates étant situées entre la position de départ et la position cible dans la matrice de distance;
la détermination (S410) de chacune des premières distances d'accumulation candidates entre la position de départ et la position cible sur la base d'une distance d'accumulation entre la position de départ et une première position candidate correspondante et une valeur de distance représentée par la première position candidate correspondante dans la matrice de distance; et
la détermination (S230), en tant que première distance d'accumulation, d'une valeur minimum parmi les premières distances d'accumulation candidates.

3. Procédé selon la revendication 2, dans lequel la détermination de chacune des premières distances d'accumulation candidates entre la position de départ et la position cible sur la base de la distance d'accumulation entre la position de départ et la première position candidate correspondante et de la valeur de distance représentée par la première position candidate correspondante comprend, par rapport à chacune des premières positions candidats:

l'ajout de la distance d'accumulation entre la position de départ et la première position candidate correspondante et la valeur de distance représentée par la première position candidate correspondante pour obtenir une valeur de somme de distance de la première position candidate correspondante, la distance d'accumulation de la première position candidate correspondante étant basée sur des distances entre la position de départ et des positions précédentes de la première position candidate correspondante, les premières positions précédentes étant situées entre la position de départ et la première position candidate correspondante dans la matrice de distance; et
la détermination de la valeur de somme de distance en tant que première distance d'accumulation candidate de la première position candidate correspondante.

4. Procédé selon la revendication 2, dans lequel la détermination de chacune des premières distances d'accumulation candidates entre la position de départ et la position cible sur la base de la distance d'accumulation entre la position de départ et la première position candidate correspondante et de la valeur de distance représentée par la première position candidate correspondante comprend, par rapport à chacune des premières positions candidats:

la réalisation d'un calcul pondéré sur la valeur de distance représentée par la première position candidate correspondante à l'aide d'une valeur de pondération de la première position candidate correspondante, pour obtenir une valeur de distance pondérée de la première position candidate correspondante;
l'ajout de la distance d'accumulation entre la position de départ et la première position candidate correspondante et la valeur de distance pondérée de la première position candidate correspondante pour obtenir une valeur de somme de distance de la première position candidate correspondante, la distance d'accumulation de la première position candidate correspondante étant basée sur des distances entre la position de départ et des positions précédentes de la première position candidate correspondante, les premières positions précédentes étant situées entre la position de départ et la première position candidate correspondante dans la matrice de distance; et
la détermination de la valeur de somme de distance en tant que première distance d'accumulation candidate de la première position candidate correspondante.

5. Procédé selon la revendication 4, comprenant en outre, avant la détermination de la valeur de somme de distance en tant que première distance d'accumulation candidate de la première position candidate correspondante:

la détermination d'une distance entre la première position candidate correspondante et une diagonale de la matrice de distance, la diagonale étant une ligne droite reliant la position de départ et la position finale; et
la détermination de la valeur de pondération de la première position candidate correspondante conformément à la distance entre la première position candidate correspondante et la diagonale.

6. Procédé selon la revendication 2, dans lequel
les premières positions candidates sont situées dans une plage de distance prédéfinie autour de la position cible.

7. Procédé selon la revendication 1, dans lequel le calcul de la première distance d'accumulation entre la position finale et la position cible dans la matrice de distance comprend:

la détermination (S610) de distances d'accumulation entre la position finale et des secondes positions candidates, les secondes positions candidates étant situées entre la position finale et la position finale dans la matrice de distance;
la détermination (S620) de chacune des secondes distances d'accumulation candidates entre la position finale et la position cible sur la base d'une distance d'accumulation entre la position finale et une seconde position candidate correspondante et d'une valeur de distance représentée par la seconde position candidate correspondante dans la matrice de distance; et
la détermination (S630), en tant que seconde distance d'accumulation, d'une valeur minimum parmi les secondes distances d'accumulation candidates.

8. Procédé selon la revendication 7, dans lequel les secondes positions candidates sont situées dans une plage de distance prédéfinie autour de la position cible.

9. Procédé selon l'une quelconque des revendications 1 à 8, dans lequel la détermination de la distance minimum entre la première séquence de caractéristiques et la seconde séquence de caractéristiques sur la base de la première distance d'accumulation et de la seconde distance d'accumulation comprend:

la détermination (S810) de distances d'accumulation minimum correspondant à la position cible sur la base d'une valeur de distance représentée par la position cible, de la première distance d'accumulation, et de la seconde distance d'accumulation; et
la sélection (S820) d'une valeur minimum dans les distances d'accumulation minimum correspondant à la position cible, et la détermination de la valeur minimum en tant que distance minimum entre la première séquence de caractéristiques et la seconde séquences de caractéristiques.

10. Procédé selon la revendication 9, dans lequel la détermination des distances d'accumulation minimum correspondant à la position cible sur la base de la valeur de distance représentée par la position cible, de la première distance d'accumulation, et de la seconde distance d'accumulation comprend:

l'ajout de la valeur de distance représentée par la position cible, de la première distance d'accumulation, et de la seconde distance d'accumulation pour obtenir les distances d'accumulation minimum correspondant à la position cible; ou
la réalisation d'un calcul pondéré sur la valeur de distance représentée par la position cible à l'aide d'une valeur de pondération correspondant à la position cible pour obtenir une valeur de distance pondérée correspondant à la position cible; et l'ajout de la valeur de la distance pondérée, de la première distance d'accumulation, et de la seconde distance d'accumulation pour obtenir les distances d'accumulation minimum correspondant à la position cible.

11. Procédé de mise en correspondance d'audioclips selon la revendication 1, le procédé étant exécuté par un serveur.

12. Support d'enregistrement lisible sur ordinateur, stockant un programme d'ordinateur, le programme d'ordinateur, lorsqu'il est exécuté par un processeur, mettant en oeuvre le procédé de mise en correspondance d'audioclips selon l'une quelconque des revendications 1 à 11.

13. Dispositif électronique, comprenant :

un ou plusieurs processeurs ; et
un appareil de stockage configuré pour stocker un ou plusieurs programmes, le ou les programmes, lorsqu'ils sont exécutés par le ou les processeurs, amenant le ou les processeurs à mettre en oeuvre le procédé de mise en correspondance d'audioclips selon l'une quelconque des revendications 1 à 11.

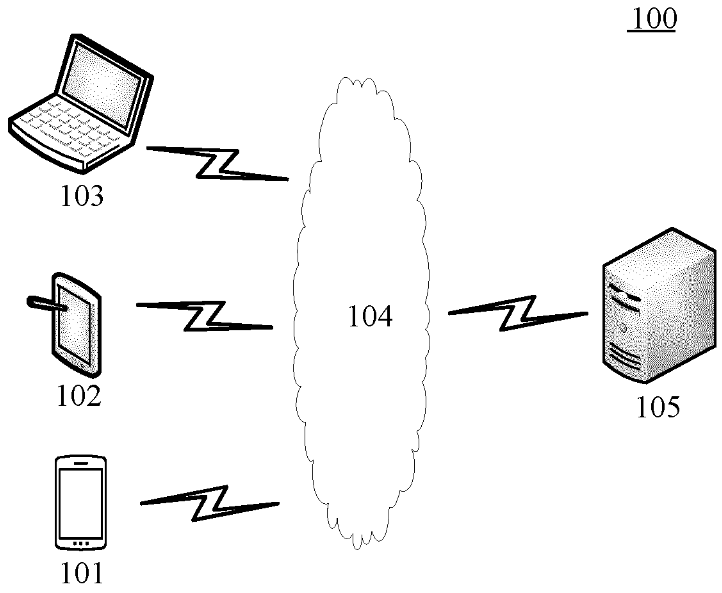
100
103
102
101
104
105

FIG. 1

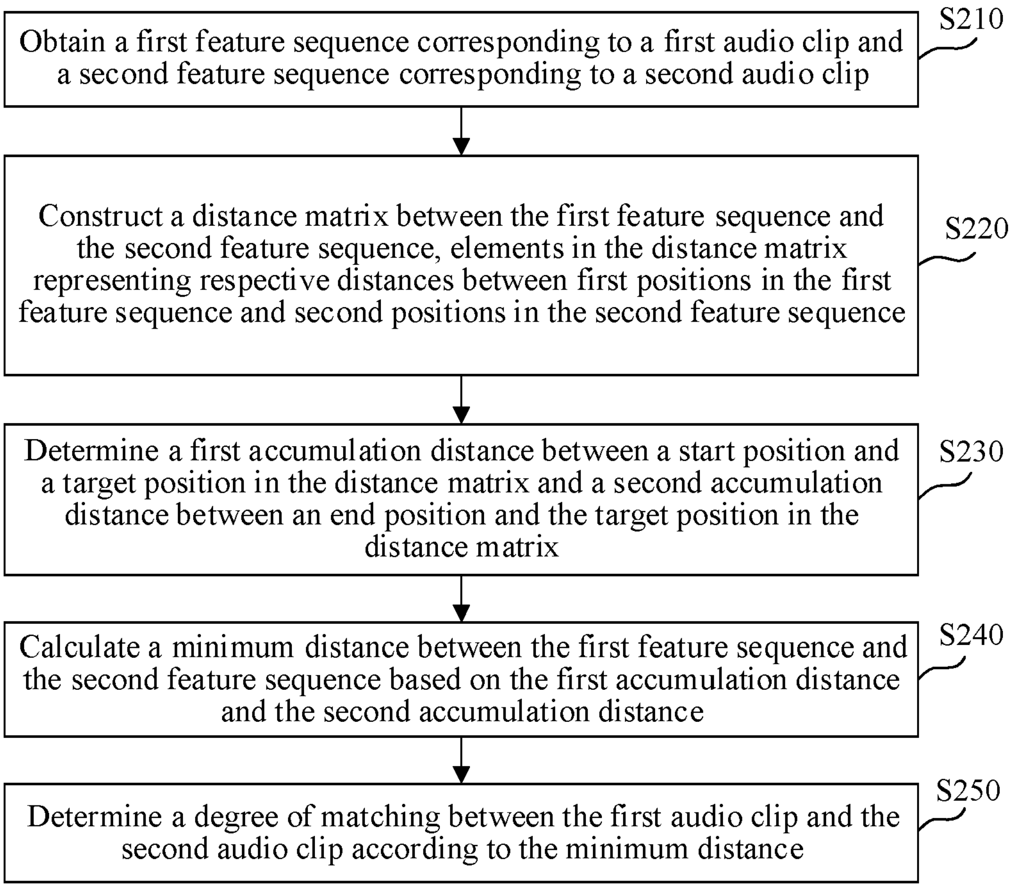
Obtain a first feature sequence corresponding to a first audio clip and a second feature sequence corresponding to a second audio clip
S210
Construct a distance matrix between the first feature sequence and the second feature sequence, elements in the distance matrix representing respective distances between first positions in the first feature sequence and second positions in the second feature sequence
S220
Determine a first accumulation distance between a start position and a target position in the distance matrix and a second accumulation distance between an end position and the target position in the distance matrix
S230
Calculate a minimum distance between the first feature sequence and the second feature sequence based on the first accumulation distance and the second accumulation distance
S240
Determine a degree of matching between the first audio clip and the second audio clip according to the minimum distance
S250

FIG. 2

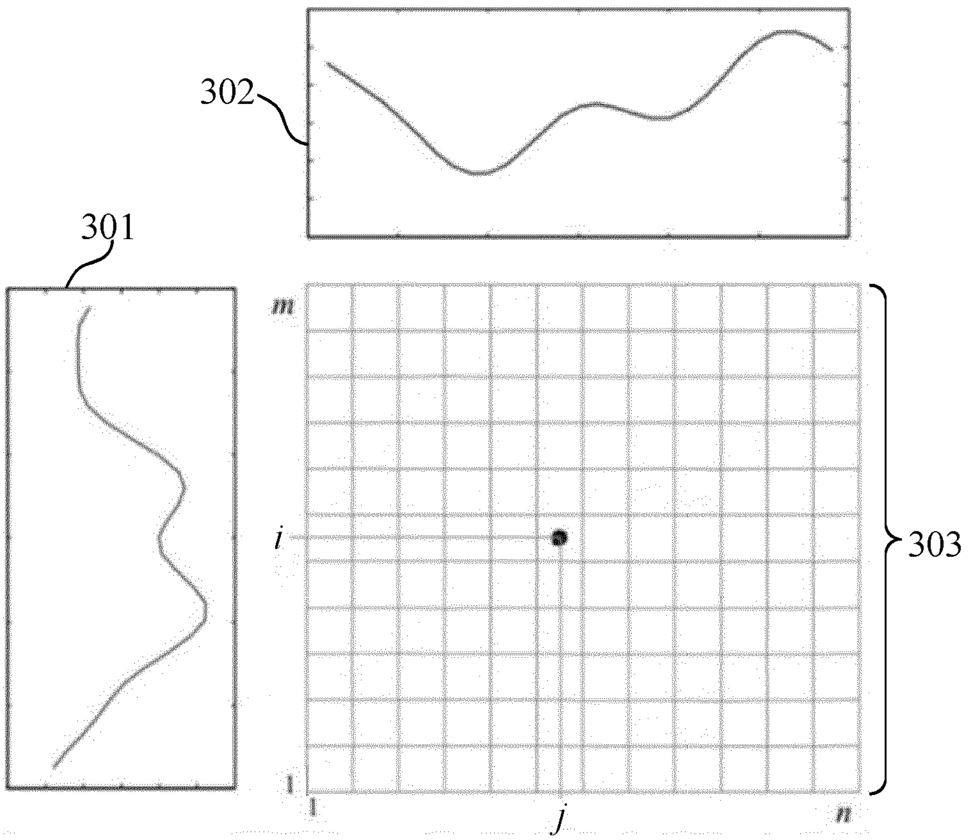
302
301
m
i
303
1
1
j
n

FIG. 3

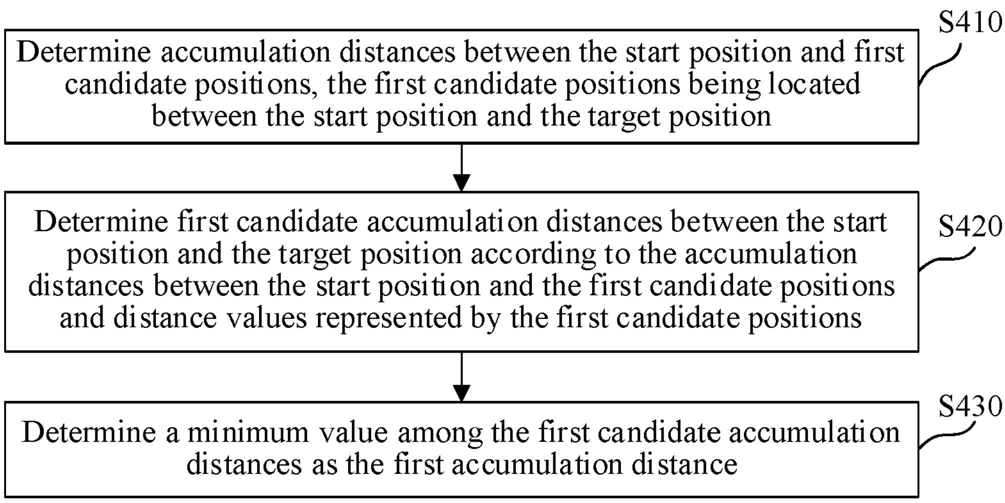
S410
Determine accumulation distances between the start position and first candidate positions, the first candidate positions being located between the start position and the target position
S420
Determine first candidate accumulation distances between the start position and the target position according to the accumulation distances between the start position and the first candidate positions and distance values represented by the first candidate positions
S430
Determine a minimum value among the first candidate accumulation distances as the first accumulation distance

FIG. 4

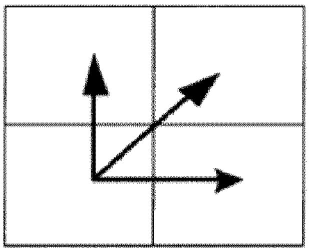

FIG. 5

Determine accumulation distances between the end position and second candidate positions, the second candidate positions being located between the target position and the end position — S610

↓

Determine second candidate accumulation distances between the end position and the target position according to the accumulation distances between the end position and the second candidate positions and distance values represented by the second candidate positions — S620

↓

Determine a minimum value among the second candidate accumulation distances as the second accumulation distance — S630

FIG. 6

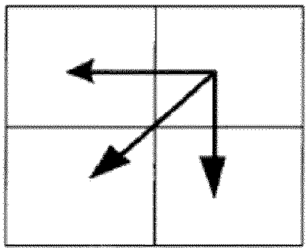

FIG. 7

Determine minimum accumulation distances corresponding to the target position based on a distance value represented by the target position, the first accumulation distance, and the second accumulation distance — S810

↓

Select a minimum value from the minimum accumulation distances corresponding to the target position as the minimum distance between the first feature sequence and the second feature sequence — S820

FIG. 8

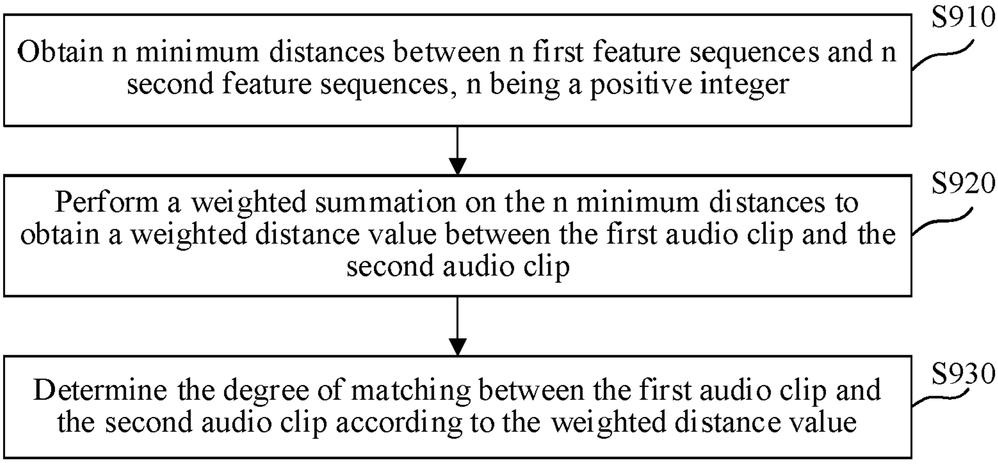
S910
Obtain n minimum distances between n first feature sequences and n second feature sequences, n being a positive integer
S920
Perform a weighted summation on the n minimum distances to obtain a weighted distance value between the first audio clip and the second audio clip
S930
Determine the degree of matching between the first audio clip and the second audio clip according to the weighted distance value

FIG. 9

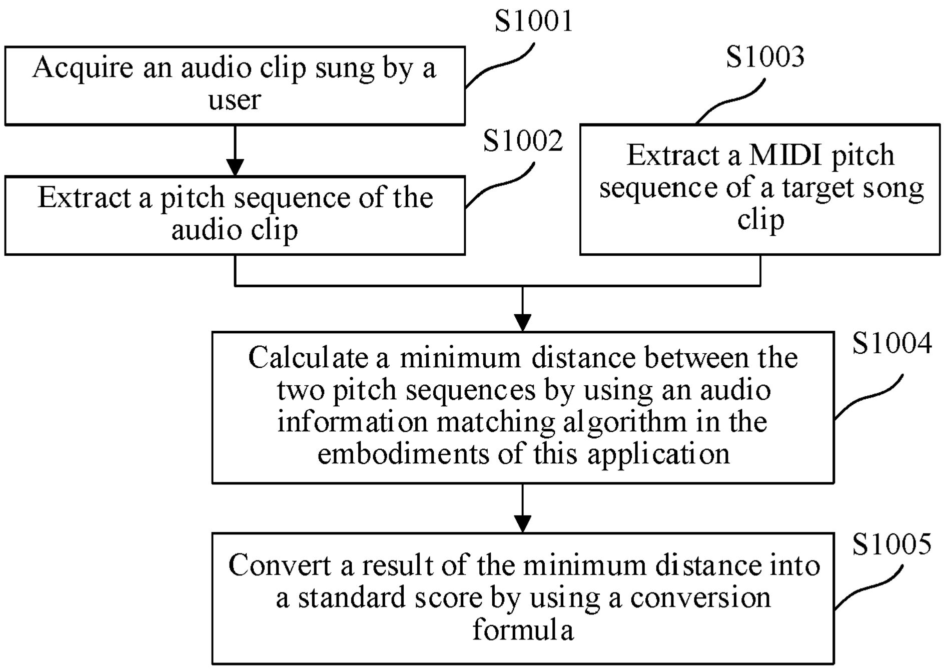
S1001
Acquire an audio clip sung by a user
S1002
Extract a pitch sequence of the audio clip
S1003
Extract a MIDI pitch sequence of a target song clip
S1004
Calculate a minimum distance between the two pitch sequences by using an audio information matching algorithm in the embodiments of this application
S1005
Convert a result of the minimum distance into a standard score by using a conversion formula

FIG. 10

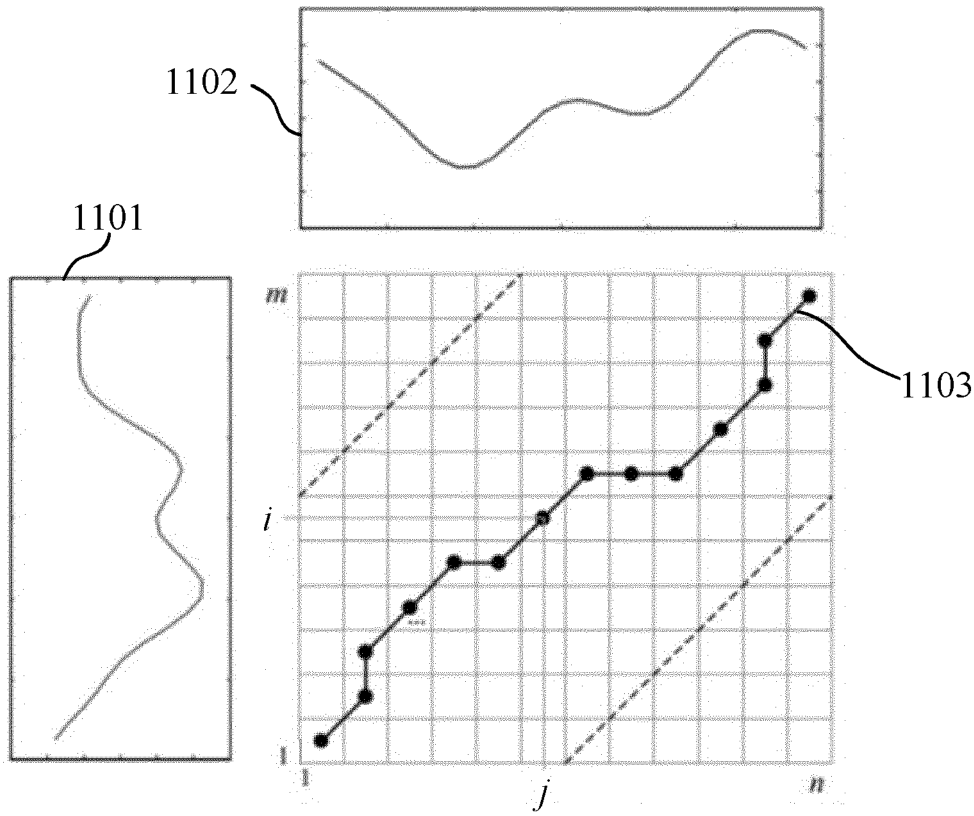
1102
1101
1103
m
i
1
1
j
n

FIG. 11

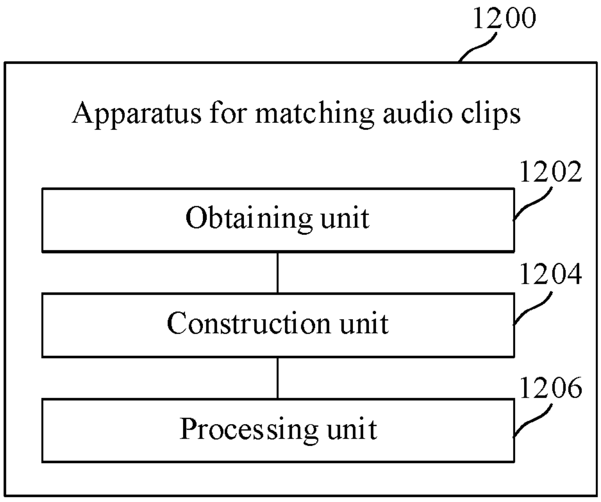
1200
Apparatus for matching audio clips
1202
Obtaining unit
1204
Construction unit
1206
Processing unit

FIG. 12

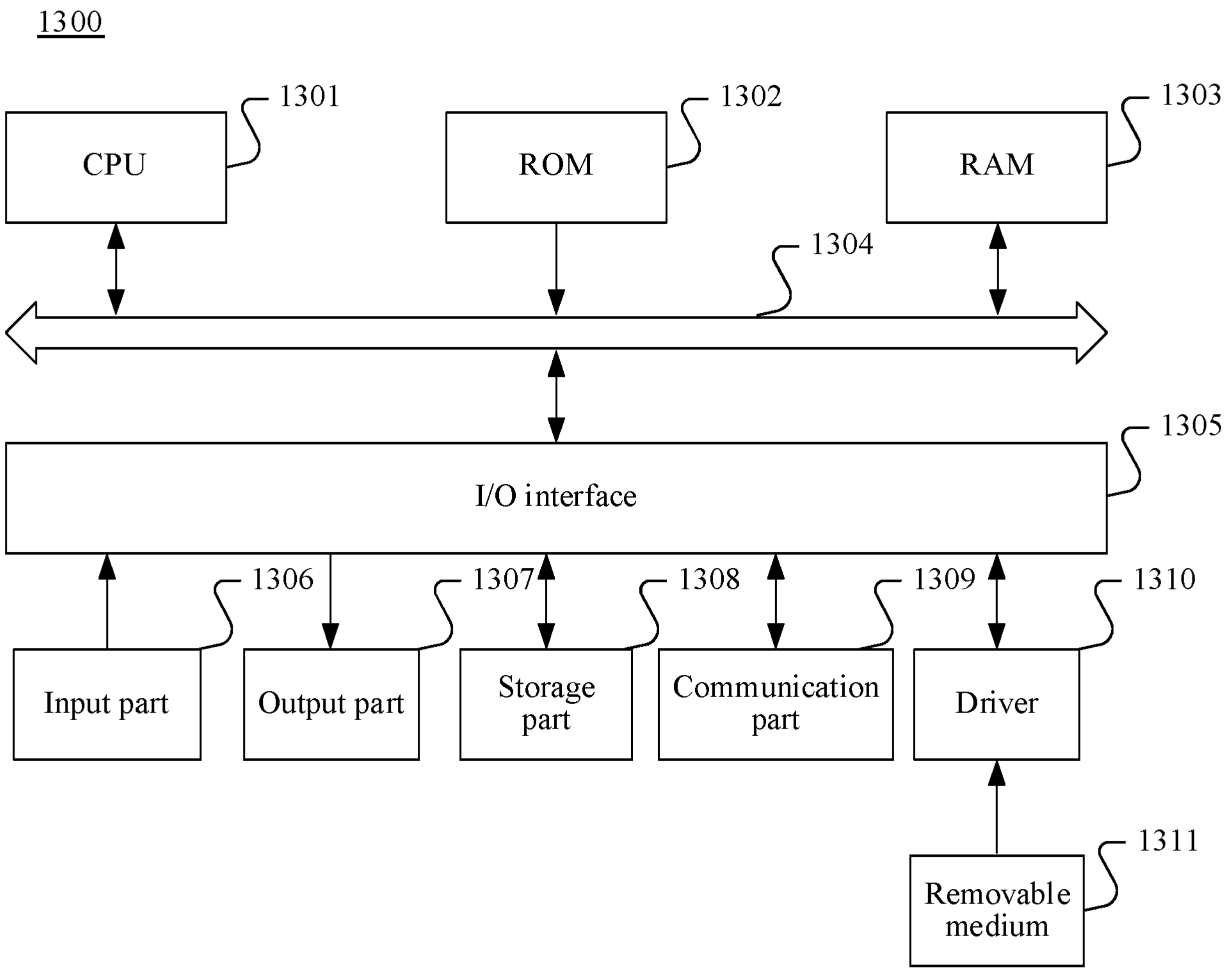
1300
CPU
1301
ROM
1302
RAM
1303
1304
I/O interface
1305
1306
1307
1308
1309
1310
Input part
Output part
Storage part
Communication part
Driver
Removable medium
1311

FIG. 13

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CN 201910441366 **[0001]**
- EP 0103258 A1 **[0004]**
- JP S62144200 A **[0006]**