

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局



(43) 国际公布日
2016 年 12 月 1 日 (01.12.2016)

WIPO | PCT



(10) 国际公布号

WO 2016/188329 A1

(51) 国际专利分类号:

G10L 15/04 (2013.01) G10L 25/48 (2013.01)
G10L 15/06 (2013.01) G10L 25/03 (2013.01)

(74) 代理人: 北京三高永信知识产权代理有限责任公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (CN).

(21) 国际申请号:

PCT/CN2016/081999

(22) 国际申请日:

2016 年 5 月 13 日 (13.05.2016)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201510270567.5 2015 年 5 月 25 日 (25.05.2015) CN
201510271769.1 2015 年 5 月 25 日 (25.05.2015) CN
201510271014.1 2015 年 5 月 25 日 (25.05.2015) CN

(71) 申请人: 广州酷狗计算机科技有限公司 (GUANG-ZHOU KUGOU COMPUTER TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市天河区科韵路 16 号自编 2 栋 1301 房, Guangdong 510665 (CN).

(72) 发明人: 赵伟峰 (ZHAO, Weifeng); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518000 (CN)。

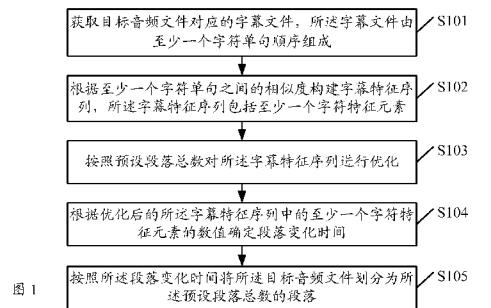
(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,

[见续页]

(54) Title: AUDIO PROCESSING METHOD AND APPARATUS, AND TERMINAL

(54) 发明名称: 一种音频处理方法、装置及终端



S101 ACQUIRE A SUBTITLE FILE CORRESPONDING TO A TARGET AUDIO FILE, THE SUBTITLE FILE BEING SEQUENTIALLY FORMED BY AT LEAST ONE CHARACTER SIMPLE SENTENCE
S102 CONSTRUCT A SUBTITLE CHARACTERISTIC SEQUENCE ACCORDING TO THE SIMILARITY
BETWEEN THE AT LEAST ONE CHARACTER SIMPLE SENTENCE, THE SUBTITLE
CHARACTERISTIC SEQUENCE COMPRISING AT LEAST ONE CHARACTER CHARACTERISTIC
ELEMENT
S103 OPTIMIZE THE SUBTITLE CHARACTERISTIC SEQUENCE ACCORDING TO A PRESET TOTAL
QUANTITY OF PARAGRAPHS
S104 DETERMINE THE PARAGRAPH CHANGE TIME ACCORDING TO A VALUE OF AT LEAST ONE
CHARACTERISTIC ELEMENT IN THE OPTIMIZED SUBTITLE CHARACTERISTIC SEQUENCE
S105 DIVIDE, ACCORDING TO THE PARAGRAPH CHANGE TIME, THE TARGET AUDIO FILE INTO
PARAGRAPHS THE PARAGRAPH QUANTITY OF WHICH IS THE PRESET TOTAL QUANTITY
OF PARAGRAPHS

(57) Abstract: An audio processing method and apparatus, and a terminal. The method comprises: acquiring file data of a target audio file; constructing a correlation characteristic sequence according to correlation characteristic data among component elements of the file data; optimizing the correlation characteristic sequence according to a preset total quantity of paragraphs; determining the paragraph change time according to a value of at least one characteristic element in the optimized correlation characteristic sequence; and dividing, according to the paragraph change time, the target audio file into paragraphs the paragraph quantity of which is the preset total quantity of paragraphs. In the method, a target audio file is divided into paragraphs according to the similarity between and time interval characteristics of character simple sentences in a subtitle file corresponding to an audio file as well as correlation characteristics of audio frames between audio paragraphs, thereby improving the efficiency of the paragraph division processing and the intelligence of audio processing.

(57) 摘要:

[见续页]

WO 2016/188329 A1



RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, 本国际公布:
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。
TG)。

一种音频处理方法、装置及终端。其中的方法包括：获取目标音频文件的文件数据；根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列；按照预设段落总数对所述相关性特征序列进行优化；根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间；按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。该方法基于音频文件对应的字幕文件中的字符单句的相似性和时间间隔特点，及音频段落之间的音频帧的相关性特点实现对目标音频文件的段落划分，提升分段处理效率，提升音频处理的智能性。

说 明 书

一种音频处理方法、装置及终端

5 本申请要求于 2015 年 5 月 25 日提交中国专利局、申请号为 201510270567.5、发明名称为“一种音频处理方法、装置及终端”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

10 本申请要求于 2015 年 5 月 25 日提交中国专利局、申请号为 201510271769.1、发明名称为“一种音频处理方法、装置及终端”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

本申请要求于 2015 年 5 月 25 日提交中国专利局、申请号为 201510271014.1、发明名称为“一种音频处理方法、装置及终端”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

15 技术领域

本发明互联网技术领域，具体涉及音频处理技术领域，尤其涉及一种音频处理方法、装置及终端。

背景技术

20 随着互联网技术的发展，互联网音频库中收录了大量的诸如歌曲、歌曲片段等音频文件，关于互联网音频的应用也日渐增多，例如：K 歌系统、听歌系统等等。许多音频文件的应用场景需要对音频文件进行段落划分，例如：K 歌系统中要实现歌曲分段合唱时，通常需要对歌曲进行段落划分；再如：听歌系统中需要重点听取歌曲片段时，通常需要对歌曲进行段落划分；等等。目前，
25 通常采用人工对音频文件进行段落划分，分段处理效率较低，无法满足用户对音频文件的使用需求，从而降低了音频处理的智能性。

发明内容

为了提高音频处理的智能性，本发明实施例提供了一种音频处理方法、装
30 置及终端。所述技术方案如下：

本发明实施例提供了一种音频处理方法，包括：

5 获取目标音频文件的文件数据；根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列；按照预设段落总数对所述相关性特征序列进行优化；根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间；按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。

其中，可根据目标音频文件的文件数据内的组成元素之间的相关性，如字符单句之间的相似度、字符单句之间的时间间隔或音频帧之间的相关性等，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

10 本发明的一个实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

20 本发明的另一个实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频25 处理的智能性。

30 本发明的又一个实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段

落划分，可提升分段处理效率，提升音频处理的智能性。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所
5 需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明
的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，
还可以根据这些附图获得其他的附图。

图 1 为本发明实施例提供的一种音频处理方法的流程图；

图 2 为本发明实施例提供的另一种音频处理方法的流程图；

10 图 3 为本发明实施例提供的一种音频处理装置的结构示意图；

图 4 为图 3 所示的构建单元的实施例的结构示意图；

图 5 为图 3 所示的优化单元的实施例的结构示意图；

图 6 为图 5 所示的优化处理单元的实施例的结构示意图；

图 7 为图 3 所示的确定单元的实施例的结构示意图；

15 图 8 为本发明实施例提供的一种音频处理方法的流程图；

图 9 为本发明实施例提供的另一种音频处理方法的流程图；

图 10 为本发明实施例提供的一种音频处理装置的结构示意图；

图 11 为图 10 所示的构建单元的实施例的结构示意图；

图 12 为图 10 所示的调整单元的实施例的结构示意图；

20 图 13 为图 10 所示的确定单元的实施例的结构示意图；

图 14 为本发明实施例提供的一种音频处理方法的流程图；

图 15 为本发明实施例提供的另一种音频处理方法的流程图；

图 16 为本发明实施例提供的一种音频处理装置的结构示意图；

图 17 为图 16 所示的获取单元的实施例的结构示意图；

25 图 18 为图 16 所示的构建单元的实施例的结构示意图；

图 19 为图 16 所示的规整处理单元的实施例的结构示意图；

图 20 为图 16 所示的确定单元的实施例的结构示意图。

具体实施方式

30 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明

实施方式作进一步地详细描述。

本发明实施例中，音频文件可以包括但不限于：歌曲、歌曲片段等文件。字幕文件可以包括但不限于：歌词、歌词片段等文件。一个音频文件可对应一个字幕文件。一个字幕文件可由至少一个字符单句顺序排列而成，以歌曲 A 为例，歌曲 A 对应的字幕文件可表示如下：

[641, 770], [641, 20] a_1 [661, 60] a_2 [721, 170] a_3 [891, 200] a_4 [1091, 70] a_5 [1161, 180] a_6 [1341, 20] a_7 [1361, 50] a_8
[1541, 180], [1541, 20] b_1 [1561, 50] b_2 [1611, 20] b_3 [1631, 30] b_4
[1661, 0] b_5 [1661, 10] b_6 [1671, 20] b_7 [1701, 30] b_8
[1871, 730], [1871, 60] c_1 [1931, 100] c_2 [2031, 110] c_3 [2141, 200]
 c_4 [2341, 70] c_5 [2411, 60] c_6 [2471, 50] c_7 [2421, 80] c_8

.....

上述歌曲 A 对应的字幕文件中，诸如“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”、“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”、“ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”可分别用于表示一个字符单句，各字符单句之前的“[]”用于描述对应的字符单句的时间属性，其单位时间通常为 ms，例如：上述[641, 770]用于描述字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”的时间属性，其中的“641”表示字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”的开始时间，“770”表示字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”的持续时间，假设歌曲 A 共 5 分钟，字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”则从第 641ms 开始演唱，持续 770ms 结束演唱。各字符单句中，每个字符之前的“[]”用于描述对应的字符的时间属性，其单位时间通常为 ms，例如：上述[641, 20]用于描述字符“ a_1 ”的时间属性，其中的“641”表示字符“ a_1 ”的开始时间，“20”表示字符“ a_1 ”的持续时间。按照开始时间的先后顺序，可确定字幕文件包含的各字符单句的顺序，例如：根据上述歌曲 A 对应的字幕文件的描述，字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”为第一个字符单句；字符单句“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”为第二个字符单句；字符单句“ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”为第三个字符单句，以此类推。其中，字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”和字符单句“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”为字符单句“ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”的在先字符单句，字符单句“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”和字符单句“ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”为字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”的在后字符单句，以此类推。进一步，字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”为字符单句“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”的相邻在先字符单句；字符单句“ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”为字符单句“ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”的相邻在后字符单句，以此类推。

一个音频文件可以被划分为多个音频段落，音频段落之间通常具备一定的

重复性；那么，一个字幕文件可对应分为多个字幕段落，字幕段落之间具备一定的相似性，也就是说，字幕段落之间所包含的字符单句存在一定的相似性。本发明实施例可利用上述的字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分。

5 一个音频文件可以被划分为多个音频段落，音频段落之间通常会有较长的停顿，即音频段落之间通常会有较长的时间间隔；那么，一个字幕文件可对应分为多个字幕段落，字幕段落之间存在较长的时间间隔，也就是说，字幕段落之间所包含的字符单句之间存在较长的时间间隔。本发明实施例可利用上述的字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分。
10

15 本发明的又一个实施例中，一个音频文件包括音频数据，对音频文件进行解码（例如：PCM 解码）可以获得音频数据（例如：PCM 数据）。一个音频文件的音频数据可包括至少一个音频帧，也就是说，一个音频文件的音频数据可以表示为多个音频帧顺序组成的一个帧序列。一个音频文件可以被划分为多个音频段落，音频段落之间通常具备一定的重复性，也就是说，音频段落之间所包含的音频帧存在一定的相关性。本发明实施例可利用上述的音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分。
20

25 基于上述描述，本发明实施例提供了一种音频处理方法，具体包括：获取目标音频文件的文件数据；根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列；按照预设段落总数对所述相关性特征序列进行优化；根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间；按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。其中，可根据目标音频文件的文件数据内的组成元素之间的相关性，如字符单句之间的相似度、字符单句之间的时间间隔或音频帧之间的相关性等，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

为了便于对本发明的理解，下面将结合附图 1-附图 2，对本发明实施例提供的音频处理方法进行详细介绍。

30 请参见图 1，为本发明实施例提供的一种音频处理方法的流程图；该方法可包括以下步骤 S101-步骤 S105。

S101，获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符单句顺序组成。

一个音频文件对应一个字幕文件。通常，互联网音频库中会存储多个音频文件、每个音频文件的属性以及每个音频文件对应的字幕文件，其中，音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识等等。本步骤中，可以从互联网音频库中获取目标音频文件对应的字幕文件；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件对应的字幕文件，并获取所查找到的字幕文件；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在互联网音频库中定位目标音频文件，并获取对应的字幕文件。

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推， $p(N-1)$ 用于表示第 N 个字符单句。

S102，根据所述至少一个字符单句之间的相似度构建字幕特征序列，所述字幕特征序列包括至少一个字符特征元素。

所述字幕特征序列可用于反映所述至少一个字符单句之间的相似度。本步骤中，首先可以采用相似度算法计算所述至少一个字符单句之间的相似度，此处需要计算每个字符单句与其在后字符单句之间的相似度，也就是说，需要计算 $p(0)$ 与 $p(1)$ 之间的相似度、 $p(0)$ 与 $p(2)$ 之间的相似度…… $p(0)$ 与 $p(N-1)$ 之间的相似度；计算 $p(1)$ 与 $p(2)$ 之间的相似度、 $p(1)$ 与 $p(3)$ 之间的相似度…… $p(1)$ 与 $p(N-1)$ 之间的相似度；以此类推。其中，相似度算法可包括但不限于：编辑距离算法（Levenshtein Distance）、最长公共子串算法（Longest Common Subsequences, LCS）、Heckel 算法、贪心字符串匹配算法（Greedy String Tiling, GST）等等。其次可以根据所述至少一个字符单句的数量、顺序以及计算获得的相似度构建所述字幕特征序列。

按照本实施例所示例子，假设采用 $s(n)$ 来表示所述字幕特征序列，则所构建的字幕特征序列 $s(n)$ 共包括 N 个字符特征元素，分别为 $s(0)$ 、 $s(1) \dots s(N-1)$ 。其中， $s(0)$ 的数值可用于描述 $p(0)$ 与其在后字符单句之间的相似情况； $s(1)$ 的数

值可用于描述 $p(1)$ 与其在后字符单句之间的相似情况；以此类推。

S103，按照预设段落总数对所述字幕特征序列进行优化。

所述预设段落总数可以根据用户对目标音频文件的实际分段需求设定。假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，则按照预设段落总数 M 对所述字幕特征序列 $s(n)$ 进行优化的目的在于，使优化后的所述字幕特征序列 $s(n)$ 刚好能够被划分为预设段落总数 M 个字幕段落，以满足对目标音频文件的实际分段需求。
5

S104，根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间。

10 其中，优化后的所述字幕特征序列 $s(n)$ 刚好能够被划分为预设段落总数 M 个字幕段落，同时，所述字幕特征序列 $s(n)$ 中的字符特征元素的数值可用于描述字符单句之间的相似情况，那么，根据优化后的所述字幕特征序列 $s(n)$ 中的字符特征元素的数值可以确定 M 个字幕段落的转折点，进一步可从字幕文件中获得 M 个字幕段落的起止时间。

15 S105，按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。由于音频文件与字幕文件相互对应，那么，按照所获得的 M 个字幕段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符
20 单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分
25 分，可提升分段处理效率，提升音频处理的智能性。

请参见图 2，为本发明实施例提供的另一种音频处理方法的流程图；该方法可包括以下步骤 S201-步骤 S213。

S201，获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符
30 单句顺序组成。

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的

结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推，
5 $p(N-1)$ 用于表示第 N 个字符单句。

本实施例的步骤 S201 可参见图 1 所示实施例的步骤 S101，在此不赘述。

S202，根据所述至少一个字符单句的数量确定构建字幕特征序列的字符特征元素的数量。

所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，即所述至少一个字符单句的数量为 N ，那么，本步骤可确定所述字幕特征序列的字符特征元素的数量也为 N ，即所述字幕特征序列的长度为 N 。假设采用 $s(n)$ 来表示所述字幕特征序列，则所构建的字幕特征序列 $s(n)$ 共包括 N 个字符特征元素，分别为
10 $s(0)$ 、 $s(1) \dots s(N-1)$ 。

S203，根据所述至少一个字符单句中各字符单句的顺序，确定构建所述字幕特征序列的各字符特征元素的索引。
15

所述字幕文件 N 个字符单句的顺序排列为 $p(0)$ 、 $p(1) \dots p(N-1)$ ，假设所述字幕特征序列 $s(n)$ 中： $s(0)$ 对应 $p(0)$ ， $s(1)$ 对应 $p(1)$ ，以此类推、 $s(N-1)$ 对应
20 $p(N-1)$ 。那么，所述字幕特征序列 $s(n)$ 中 $s(0)$ 的索引为 1，即第一个字符特征元素； $s(1)$ 的索引为 2，即第二个字符特征元素；以此类推， $s(N-1)$ 的索引为 N ，即第 N 个字符特征元素。

S204，将构建所述字幕特征序列的各字符特征元素的数值均设置为初始值。

所述初始值可以根据实际需要进行设定，本实施例中可假设所述初始值为 0，那么，本步骤可设置所述字幕特征序列 $s(n)$ 中所有字符特征元素的数值均为
25 0，也就是说， $s(0)=0$ ， $s(1)=0 \dots s(N-1)=0$ 。

S205，针对所述至少一个字符单句中的任一个目标字符单句，若所述目标字符单句与所述目标字符单句的后字符单句之间的最大相似度大于预设相似阀值，将所述目标字符单句对应的字符特征元素的数值从初始值变更为目标值。

30 本步骤 S205 的具体处理过程可包括如下 s11-s13：

s11，采用相似度算法计算所述至少一个字符单句之间的相似度，此处需

要计算每个字符单句与其在后字符单句之间的相似度，也就是说，需要计算 $p(0)$ 与 $p(1)$ 之间的相似度、 $p(0)$ 与 $p(2)$ 之间的相似度…… $p(0)$ 与 $p(N-1)$ 之间的相似度；计算 $p(1)$ 与 $p(2)$ 之间的相似度、 $p(1)$ 与 $p(3)$ 之间的相似度…… $p(1)$ 与 $p(N-1)$ 之间的相似度；以此类推。其中，相似度算法可包括但不限于：编辑距离算法、最长公共子串算法、Heckel 算法、贪心字符串匹配算法等等。需要说明的是，为了方便计算，计算获得的相似度均归一化至 $[0, 1]$ 的区间内，若某两个字符单句之间的相似度等于 0，表明该两个字符单句完全不同；若某两个字符单句之间的相似度等于 1，表明该两个字符单句完全相同。

s12，提取每个字符单句与其在后字符单句之间的最大相似度，例如：假设经过计算 $p(0)$ 与其在后字符单句 $p(1)$ 至 $p(N-1)$ 之间， $p(0)$ 与 $p(2)$ 之间的相似度最大为 Q_{02} ，则提取 Q_{02} ；再如：假设经过计算 $p(1)$ 与其在后字符单句 $p(2)$ 至 $p(N-1)$ 之间， $p(1)$ 与 $p(5)$ 之间的相似度最大为 Q_{15} ，则提取 Q_{15} ，等等。

s13，分别判断所提取的最大相似度是否大于预设相似阀值，根据判断结果变更设置相应字符特征元素的数值。其中，所述预设相似阀值可以根据实际需要进行设定，该预设相似阀值可采用 Th 表示，且 $0 \leq Th \leq 1$ 。所述目标值可以根据实际需要进行设定，且所述目标值大于所述初始值，本实施例可设定所述目标值为 1。按照步骤 s12 所示例子，例如：判断 Q_{02} 是否大于预设相似阀值 Th ，如果 $Q_{02} > Th$ ，那么将 $p(0)$ 对应的 $s(0)$ 的数值从 0 变更为 1，即 $s(0) = 1$ ；再如：判断 Q_{15} 是否大于预设相似阀值 Th ，如果 $Q_{15} > Th$ ，那么将 $p(1)$ 对应的 $s(1)$ 的数值从 0 变更为 1，即 $s(1) = 1$ ，等等。

S206，按照构建所述字幕特征序列的字符特征元素的数量、索引及数值，构建所述字幕特征序列。

所构建的所述字幕特征序列为 $s(n)$ ， $s(n)$ 是由 N 个字符特征元素 $s(0)$ 、 $s(1) \dots s(N-1)$ 顺序组成，且所述字幕特征序列 $s(n)$ 中各字符特征元素的数值形成由 0 和 1 组成的序列。

本实施例的步骤 S202-步骤 S206 可以为图 1 所示实施例的步骤 S102 的具体细化步骤。

S207，统计所述字幕特征序列中数值为目标值的字符特征元素的数量。按照本实施例所示例子，本步骤需要统计所述字幕特征序列 $s(n)$ 中数值为 1 的字符特征元素的数量。

S208，判断所述数量是否位于所述预设段落总数对应的容错区间内；若判

断结果为是，转入步骤 S210；若判断结果为否，转入步骤 S209。

假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，所述预设段落总数 M 对应的容错区间可表示为 $[M - u, M + u]$ (u 为整数)，其中 u 表示一个整数范围区间，可以根据实际需要进行设定。本步骤中，需要判断所统计的所述字幕特征序列 $s(n)$ 中数值为 1 的字符特征元素的数量是否位于 $[M - u, M + u]$ 的区间内，如果判断结果为是，表明所述字幕特征序列 $s(n)$ 能够被划分为预设段落总数 M 个字幕段落，以满足对目标音频文件的实际分段需求。如果判断结果为否，表明所述字幕特征序列 $s(n)$ 无法很好地被划分为预设段落总数 M 个字幕段落，无法满足对目标音频文件的实际分段需求，需要进行一些调整。

S209，调整所述预设相似阀值的大小以调整所述字幕特征序列中的各字符特征元素的数值。

本步骤的调整过程可包括以下步骤 s21-s22：

s21，若所述数量大于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长增大所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

如果所述数量大于 $M + u$ ，则需要按照预设步长增大所述预设相似阀值 Th 的值，并重新执行上述步骤 s13 以调整所述字幕特征序列中的各字符特征元素的数值。

s22，若所述数量小于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长减小所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

如果所述数量小于 $M - u$ ，则需要按照预设步长减小所述预设相似阀值 Th 的值，并重新执行上述步骤 s13 以调整所述字幕特征序列中的各字符特征元素的数值。

步骤 s21-s22 中，所述预设步长可以根据实际需要进行设定，所述预设步长可以为固定步长，即每次采用固定步长增大或减小所述预设相似阀值 Th 的值；所述预设步长也可以为随机步长，即每次采用不同步长增大或减小所述预设相似阀值 Th 的值。

本实施例的步骤 S207-步骤 S209 可以为图 1 所示实施例的步骤 S103 的具体细化步骤。

S210，从优化后的所述字幕特征序列中获取数值为目标值的字符特征元素

对应的目标索引。假设优化后的所述字幕特征序列 $s(n)$ 中 $s(0)=0$ ， $s(1)=0 \dots s(4)=1 \dots s(10)=1 \dots s(N-1)=0$ ，由于 $s(4)=1$ 和 $s(10)=1$ ， $s(4)$ 对应的索引为 5， $s(10)$ 对应的索引为 11，本步骤可获得目标索引为 5 和 11。

S211，根据所述目标索引在所述字幕文件中定位段落转折的字符单句。

5 所述目标索引为 5 和 11，则可在所述字幕文件中定位段落转折的字符单句为第 5 个字符单句和第 11 个字符单句，也就是说，第 5 个字符单句为一个字幕段落的起始位置，即所述字幕文件中第 1-4 个字符单句构成一个字幕段落；第 11 个字符单句为另一个字幕段落的起始位置，即所述字幕文件中第 5-10 个字符单句构成一个字幕段落。

10 S212，根据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

由于所述字幕文件中记录了每个字符单句的时间属性，包括每个字符单句的开始时间、持续时间、结束时间；本步骤可以从所述字幕文件中读取段落变化时间，按照本实施例所示例子，所述字幕文件中第 1-4 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 4 个字符单句的结束时间和第 5 个字符单句的开始时间；所述字幕文件中第 5-10 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 10 个字符单句的结束时间和第 11 个字符单句的开始时间。

20 本实施例的步骤 S210-步骤 S212 可以为图 1 所示实施例的步骤 S104 的具体细化步骤。依据步骤 S210-步骤 S212 可获得 M 个字幕段落的起止时间。

S213，按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。由于音频文件与字幕文件相互对应，那么，按照所获得的 M 个字幕段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

25 本实施例的步骤 S213 可参见图 1 所示实施例的步骤 S105，在此不赘述。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划

分，可提升分段处理效率，提升音频处理的智能性。

下述将结合附图 3-附图 7，对本发明实施例提供的音频处理装置的结构和功能进行详细介绍。需要说明的是，下述附图 3-附图 7 的所示的装置可以运行于终端中，以被应用于执行上述附图 1-附图 2 所示的方法。
5

请参见图 3，为本发明实施例提供的一种音频处理装置的结构示意图；该装置可包括：获取单元 301、构建单元 302、优化单元 303、确定单元 304 和分段单元 305。

10 获取单元 301，用于获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符单句顺序组成。

一个音频文件对应一个字幕文件。通常，互联网音频库中会存储多个音频文件、每个音频文件的属性以及每个音频文件对应的字幕文件，其中，音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识等等。所述获取单元 301 可以从互联网音频库中获取目标音频文件对应的字幕文件；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件对应的字幕文件，并获取所查找到的字幕文件；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在互联网音频库中定位目标音频文件，并获取对应的字幕文件。
15

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推， $p(N-1)$ 用于表示第 N 个字符单句。
20

25 构建单元 302，用于根据所述至少一个字符单句之间的相似度构建字幕特征序列，所述字幕特征序列包括至少一个字符特征元素。

所述字幕特征序列可用于反映所述至少一个字符单句之间的相似度。首先所述构建单元 302 可以采用相似度算法计算所述至少一个字符单句之间的相似度，此处需要计算每个字符单句与其在后字符单句之间的相似度，也就是说，
30 需要计算 $p(0)$ 与 $p(1)$ 之间的相似度、 $p(0)$ 与 $p(2)$ 之间的相似度…… $p(0)$ 与 $p(N-1)$ 之间的相似度；计算 $p(1)$ 与 $p(2)$ 之间的相似度、 $p(1)$ 与 $p(3)$ 之间的相似度…… $p(1)$

与 $p(N-1)$ 之间的相似度；以此类推。其中，相似度算法可包括但不限于：编辑距离算法、最长公共子串算法、Heckel 算法、贪心字符串匹配算法等等。其次所述构建单元 302 可以根据所述至少一个字符串单句的数量、顺序以及计算获得的相似度构建所述字幕特征序列。

5 按照本实施例所示例子，假设采用 $s(n)$ 来表示所述字幕特征序列，则所构建的字幕特征序列 $s(n)$ 共包括 N 个字符特征元素，分别为 $s(0)、s(1) \dots s(N-1)$ 。其中， $s(0)$ 的数值可用于描述 $p(0)$ 与其在后字符串单句之间的相似情况； $s(1)$ 的数值可用于描述 $p(1)$ 与其在后字符串单句之间的相似情况；以此类推。

优化单元 303，用于按照预设段落总数对所述字幕特征序列进行优化。

10 所述预设段落总数可以根据用户对目标音频文件的实际分段需求设定。假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，则所述优化单元 303 按照预设段落总数 M 对所述字幕特征序列 $s(n)$ 进行优化的目的在于，使优化后的所述字幕特征序列 $s(n)$ 刚好能够被划分为预设段落总数 M 个字幕段落，以满足对目标音频文件的实际分段需求。

15 确定单元 304，用于根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间。

其中，优化后的所述字幕特征序列 $s(n)$ 刚好能够被划分为预设段落总数 M 个字幕段落，同时，所述字幕特征序列 $s(n)$ 中的字符特征元素的数值可用于描述字符串单句之间的相似情况，那么，所述确定单元 304 根据优化后的所述字幕特征序列 $s(n)$ 中的字符特征元素的数值可以确定 M 个字幕段落的转折点，进一步可从字幕文件中获得 M 个字幕段落的起止时间。

分段单元 305，用于按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。

由于音频文件与字幕文件相互对应，那么，所述分段单元 305 按照所获得的 M 个字幕段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符串单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符串单句的相似

性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 4，为图 3 所示的构建单元的实施例的结构示意图；该构建单元 5 302 可包括：数量确定单元 401、索引确定单元 402、数值设置单元 403、数值变更单元 404 和序列构建单元 405。

数量确定单元 401，用于根据所述至少一个字符单句的数量确定构建字幕特征序列的字符特征元素的数量。

所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，即所述至少一个字符单句的数量为 N ，那么，所述数量确定单元 401 可确定所述字幕特征序列的字符特征元素的数量也为 N ，即所述字幕特征序列的长度为 N 。假设采用 $s(n)$ 来表示所述字幕特征序列，则所构建的字幕特征序列 $s(n)$ 共包括 N 个字符特征元素，分别为 $s(0)$ 、 $s(1) \dots s(N-1)$ 。
10
15

索引确定单元 402，用于根据所述至少一个字符单句中各字符单句的顺序，确定构建所述字幕特征序列的各字符特征元素的索引。

所述字幕文件 N 个字符单句的顺序排列为 $p(0)$ 、 $p(1) \dots p(N-1)$ ，假设所述字幕特征序列 $s(n)$ 中： $s(0)$ 对应 $p(0)$ ， $s(1)$ 对应 $p(1)$ ，以此类推、 $s(N-1)$ 对应 $p(N-1)$ 。那么，所述字幕特征序列 $s(n)$ 中 $s(0)$ 的索引为 1，即第一个字符特征元素； $s(1)$ 的索引为 2，即第二个字符特征元素；以此类推， $s(N-1)$ 的索引为 N ，即第 N 个字符特征元素。
20

数值设置单元 403，用于将构建所述字幕特征序列的各字符特征元素的数值均设置为初始值。

所述初始值可以根据实际需要进行设定，本实施例中可假设所述初始值为 0，那么，所述数值设置单元 403 可设置所述字幕特征序列 $s(n)$ 中所有字符特征元素的数值均为 0，也就是说， $s(0)=0$ ， $s(1)=0 \dots s(N-1)=0$ 。
25

数值变更单元 404，用于针对所述至少一个字符单句中的任一个目标字符单句，若所述目标字符单句与所述目标字符单句的后字符单句之间的最大相似度大于预设相似阀值，将所述目标字符单句对应的字符特征元素的数值从初始值变更为目标值。

所述数据变更单元 404 的具体处理过程可包括以下 A-C：

A、采用相似度算法计算所述至少一个字符单句之间的相似度，此处需要

计算每个字符单句与其在后字符单句之间的相似度，也就是说，需要计算 $p(0)$ 与 $p(1)$ 之间的相似度、 $p(0)$ 与 $p(2)$ 之间的相似度…… $p(0)$ 与 $p(N-1)$ 之间的相似度；计算 $p(1)$ 与 $p(2)$ 之间的相似度、 $p(1)$ 与 $p(3)$ 之间的相似度…… $p(1)$ 与 $p(N-1)$ 之间的相似度；以此类推。其中，相似度算法可包括但不限于：编辑距离算法、
5 最长公共子串算法、Heckel 算法、贪心字符串匹配算法等等。需要说明的是，为了方便计算，计算获得的相似度均归一化至 [0, 1] 的区间内，若某两个字符单句之间的相似度等于 0，表明该两个字符单句完全不同；若某两个字符单句之间的相似度等于 1，表明该两个字符单句完全相同。

B、提取每个字符单句与其在后字符单句之间的最大相似度，例如：假设
10 经过计算 $p(0)$ 与其在后字符单句 $p(1)$ 至 $p(N-1)$ 之间， $p(0)$ 与 $p(2)$ 之间的相似度最大为 Q_{02} ，则提取 Q_{02} ；再如：假设经过计算 $p(1)$ 与其在后字符单句 $p(2)$ 至
15 $p(N-1)$ 之间， $p(1)$ 与 $p(5)$ 之间的相似度最大为 Q_{15} ，则提取 Q_{15} ，等等。

C、分别判断所提取的最大相似度是否大于预设相似阀值，根据判断结果
20 变更设置相应字符特征元素的数值。其中，所述预设相似阀值可以根据实际需要进行设定，该预设相似阀值可采用 Th 表示，且 $0 \leq Th \leq 1$ 。所述目标值可以根
据实际需要进行设定，且所述目标值大于所述初始值，本实施例可设定所述目
标值为 1。按照本实施例所示例子，例如：判断 Q_{02} 是否大于预设相似阀值 Th ，
如果 $Q_{02} > Th$ ，那么将 $p(0)$ 对应的 $s(0)$ 的数值从 0 变更为 1，即 $s(0) = 1$ ；再如：判
断 Q_{15} 是否大于预设相似阀值 Th ，如果 $Q_{15} > Th$ ，那么将 $p(1)$ 对应的 $s(1)$ 的数值从
25 0 变更为 1，即 $s(1) = 1$ ，等等。

序列构建单元 405，用于按照构建所述字幕特征序列的字符特征元素的数量、索引及数值，构建所述字幕特征序列。

所构建的所述字幕特征序列为 $s(n)$ ， $s(n)$ 是由 N 个字符特征元素 $s(0)$ 、
25 $s(1) \dots s(N-1)$ 顺序组成，且所述字幕特征序列 $s(n)$ 中各字符特征元素的数值形成
由 0 和 1 组成的序列。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符
单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列
进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数
值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为
30 所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似
性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划

分，可提升分段处理效率，提升音频处理的智能性。

请参见图 5，为图 3 所示的优化单元的实施例的结构示意图；该优化单元 303 可包括：数量统计单元 501、判断单元 502 和优化处理单元 503。

5 数量统计单元 501，用于统计所述字幕特征序列中数值为目标值的字符特征元素的数量。按照图 4 所示实施例的例子，所述数量统计单元 501 需要统计所述字幕特征序列 $s(n)$ 中数值为 1 的字符特征元素的数量。

判断单元 502，用于判断所述数量是否位于所述预设段落总数对应的容错区间内。

10 假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，所述预设段落总数 M 对应的容错区间可表示为 $[M - u, M + u]$ (u 为整数)，其中 u 表示一个整数范围区间，可以根据实际需要进行设定。所述判断单元 502 需要判断所统计的所述字幕特征序列 $s(n)$ 中数值为 1 的字符特征元素的数量是否位于 $[M - u, M + u]$ 的区间内，如果判断结果为是，表明所述字幕特征序列 $s(n)$ 能够被划分为预设段落总数 M 个字幕段落，以满足对目标音频文件的实际分段需求。如果判断结果为否，表明所述字幕特征序列 $s(n)$ 无法很好地被划分为预设段落总数 M 个字幕段落，无法满足对目标音频文件的实际分段需求，需要进行一些调整。

20 优化处理单元 503，用于若判断结果为否，调整所述预设相似阀值的大小以调整所述字幕特征序列中的各字符特征元素的数值。

请一并参见图 6，为图 5 所示的优化处理单元的实施例的结构示意图；该优化处理单元 503 包括：第一调整单元 601 和第二调整单元 602。

25 第一调整单元 601，用于若所述数量大于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长增大所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

如果所述数量大于 $M + u$ ，则所述第一调整单元 601 需要按照预设步长增大所述预设相似阀值 Th 的值，并重新调整所述字幕特征序列中的各字符特征元素的数值。

30 第二调整单元 602，用于若所述数量小于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长减小所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

如果所述数量小于 $M-u$ ，则所述第二调整单元 602 需要按照预设步长减小所述预设相似阀值 Th 的值，并重新调整所述字幕特征序列中的各字符特征元素的数值。其中，所述预设步长可以根据实际需要进行设定，所述预设步长可以为固定步长，即每次采用固定步长增大或减小所述预设相似阀值 Th 的值；所述预设步长也可以为随机步长，即每次采用不同步长增大或减小所述预设相似阀值 Th 的值。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 7，为图 3 所示的确定单元 304 的实施例的结构示意图；该确定单元 304 可包括：目标索引获取单元 701、定位单元 702 和时间读取单元 703。

目标索引获取单元 701，用于从优化后的所述字幕特征序列中获取数值为目标值的字符特征元素对应的目标索引。

假设优化后的所述字幕特征序列 $s(n)$ 中 $s(0)=0$ ，
20 $s(1)=0 \dots s(4)=1 \dots s(10)=1 \dots s(N-1)=0$ ，由于 $s(4)=1$ 和 $s(10)=1$ ， $s(4)$ 对应的索引为 5， $s(10)$ 对应的索引为 11，所述目标索引获取单元 701 可获得目标索引为 5 和 11。

定位单元 702，用于根据所述目标索引在所述字幕文件中定位段落转折的字符单句。

所述目标索引为 5 和 11，所述定位单元 702 则可在所述字幕文件中定位段落转折的字符单句为第 5 个字符单句和第 11 个字符单句，也就是说，第 5 个字符单句为一个字幕段落的起始位置，即所述字幕文件中第 1-4 个字符单句构成一个字幕段落；第 11 个字符单句为另一个字幕段落的起始位置，即所述字幕文件中第 5-10 个字符单句构成一个字幕段落。

时间读取单元 703，用于根据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

由于所述字幕文件中记录了每个字符单句的时间属性，包括每个字符单句的开始时间、持续时间、结束时间；所述时间读取单元 703 可以从所述字幕文件中读取段落变化时间，按照本实施例所示例子，所述字幕文件中第 1-4 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 4 个字符单句的结束时间和第 5 个字符单句的开始时间；所述字幕文件中第 5-10 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 10 个字符单句的结束时间和第 11 个字符单句的开始时间。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

15

本发明实施例还公开了一种终端，该终端可以为 PC (Personal Computer, 个人计算机)、笔记本电脑、手机、PAD (平板电脑)、车载终端、智能可穿戴设备等设备。该终端中可包括一音频处理装置，该装置的结构和功能可参见上述图 3-图 7 所示实施例的相关描述，在此不赘述。

20

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的相似度构建字幕特征序列，按照预设段落总数对所述字幕特征序列进行优化，并根据优化后的所述字幕特征序列中的至少一个字符特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，音频处理过程利用字幕段落之间的字符单句的相似性特点，基于字幕文件中的字符单句的相似性实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

25

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

基于上述描述，下面将结合附图 8-附图 9，对本发明实施例提供的音频处理方法进行详细介绍。

请参见图 8，为本发明实施例提供的一种音频处理方法的流程图；该方法可包括以下步骤 S801-步骤 S805。

5 S801，获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符单句顺序组成。

一个音频文件对应一个字幕文件。所述字幕文件包括至少一个字符单句及各字符单句的关键信息；一个字符单句的关键信息包括：标识 (ID)、开始时间 (start_time) 和结束时间 (end_time)。通常，互联网音频库中会存储多个音频文件、每个音频文件的属性以及每个音频文件对应的字幕文件，其中，音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识等等。本步骤中，可以从互联网音频库中获取目标音频文件对应的字幕文件；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件对应的字幕文件，并获取所查找到的字幕文件；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在互联网音频库中定位目标音频文件，并获取对应的字幕文件。

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推， $p(N-1)$ 用于表示第 N 个字符单句。

S802，根据所述至少一个字符单句之间的时间间隔构建时间特征序列，所述时间特征序列包括至少一个时间特征元素。

25 所述时间特征序列可用于反映所述至少一个字符单句之间的时间间隔程度。本步骤中，首先计算所述至少一个字符单句之间的时间间隔，此处需要计算 $p(1)$ 与 $p(0)$ 之间的时间间隔 $p(1).start_time - p(0).end_time$ ；计算 $p(2)$ 与 $p(1)$ 之间的时间间隔 $p(2).start_time - p(1).end_time$ ；以此类推，计算 $p(N-1)$ 与 $p(N-2)$ 之间的时间间隔 $p(N-1).start_time - p(N-2).end_time$ 。其次可以根据所述至少一个字符单句的数量、顺序以及计算获得的时间间隔构建所述时间特征序列。

按照本实施例所示例子，假设采用 $t(n)$ 来表示所述时间特征序列，则所构

建的时间特征序列 $t(n)$ 共包括 N 个时间特征元素，分别为 $t(0)、t(1) \dots t(N-1)$ 。其中， $t(0)$ 的数值可设置为 0， $t(1)$ 的数值用于表示 $p(1)$ 与 $p(0)$ 之间的时间间隔； $t(2)$ 的数值用于表示 $p(2)$ 与 $p(1)$ 之间的时间间隔；以此类推， $t(N-1)$ 的数值用于表示 $p(N-1)$ 与 $p(N-2)$ 之间的时间间隔。

5 S803，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值。

所述预设段落总数可以根据用户对目标音频文件的实际分段需求设定。假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，则按照预设段落总数 M 调整所述时间特征序列 $t(n)$ 中的各时间特征元素的数值目的在于，使调整 10 后的所述时间特征序列 $t(n)$ 刚好能够提取到 M 个字幕段落对应的转折点，从而实现对目标音频文件的实际分段需求。

S804，根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间。

所述调整后的时间特征序列 $t(n)$ 中的各时间特征元素的数值能够反映 M 个 15 字幕段落对应的转折点，那么，本步骤可以根据调整后的所述时间特征序列中的至少一个时间特征元素的数值，从字幕文件中获得 M 个字幕段落的起止时间。

S805，按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。由于音频文件与字幕文件相互对应，那么，按照所获得的 M 个字幕 20 段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

30

请参见图 9，为本发明实施例提供的另一种音频处理方法的流程图；该方

法可包括以下步骤 S901-步骤 S905。

S901，获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符单句顺序组成。

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的 5 结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推， $p(N-1)$ 用于表示第 N 个字符单句。

10 本实施例的步骤 S901 可参见图 1 所示实施例的步骤 S801，在此不赘述。

S902，根据所述至少一个字符单句的数量确定构建时间特征序列的时间特征元素的数量。

所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，即所述至少一个字符单句的数量为 N ，那么，本步骤可确定所述时间特征序列的时间特征元素的数量也为 N ，即所述时间特征序列的长度为 N 。假设采用 $t(n)$ 来表示所述时间特征序列，则所构建的时间特征序列 $t(n)$ 共包括 N 个时间特征元素，分别为 15 $t(0)$ 、 $t(1)$ … $t(N-1)$ 。

S903，根据所述至少一个字符单句中各字符单句的顺序，确定构建所述时间特征序列的各时间特征元素的索引。

20 所述字幕文件 N 个字符单句的顺序排列为 $p(0)$ 、 $p(1)$ … $p(N-1)$ ，假设所述时间特征序列 $t(n)$ 中： $t(0)$ 对应 $p(0)$ ， $t(1)$ 对应 $p(1)$ ，以此类推、 $t(N-1)$ 对应 $p(N-1)$ 。那么，所述时间特征序列 $t(n)$ 中 $t(0)$ 的索引为 1，即第一个时间特征元素； $t(1)$ 的索引为 2，即第二个时间特征元素；以此类推， $t(N-1)$ 的索引为 N ，即第 N 个时间特征元素。

25 S904，针对所述至少一个字符单句中的任一个目标字符单句，将所述目标字符单句与所述目标字符单句的相邻在先字符单句之间的时间间隔设置为所述目标字符单句对应的时间特征元素的数值。

本步骤 S904 的具体处理过程可包括以下步骤 s11-s12：

s11，计算每一个字符单句与其相邻在先字符单句之间的时间间隔，此处 30 需要计算 $p(1)$ 与 $p(0)$ 之间的时间间隔 $p(1).start_time - p(0).end_time$ ；计算 $p(2)$ 与 $p(1)$ 之间的时间间隔 $p(2).start_time - p(1).end_time$ ；以此类推，计算 $p(N-1)$ 与

$p(N-2)$ 之间的时间间隔 $p(N-1).start_time - p(N-2).end_time$ 。

s12，将计算获得的时间间隔设置为相应时间特征元素的数值；那么，可设
 置 $t(0) = 0$, $t(1) = p(1).start_time - p(0).end_time$,
 $t(2) = p(2).start_time - p(1).end_time$, 以此类推 ,
 5 $t(N-1) = p(N-1).start_time - p(N-2).end_time$ 。

S905，按照构建所述时间特征序列的时间特征元素的数量、索引及数值，构建所述时间特征序列。

所构建的所述时间特征序列为 $t(n)$ ， $t(n)$ 是由 N 个时间特征元素 $t(0)$ 、
 $t(1) \dots t(N-1)$ 顺序组成，且所述时间特征序列 $t(n)$ 中各时间特征元素的数值为
 10 $t(0) = 0$, $t(1) = p(1).start_time - p(0).end_time$, $t(2) = p(2).start_time - p(1).end_time$,
 以此类推， $t(N-1) = p(N-1).start_time - p(N-2).end_time$ 。

本实施例的步骤 S902-步骤 S905 可以为图 8 所示实施例的步骤 S802 的具体细化步骤。

S906，从所述时间特征序列中查找前预设段落数量减 1 个最大数值的时间
 15 特征元素。假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，本步
 骤需要从所述时间特征序列 $t(n)$ 中查找前 $M-1$ 个最大数值的时间特征元素。

S907，将查找到的时间特征元素的数值调整为目标值，将所述时间特征序
 列中除查找到的时间特征元素之外的其他时间特征元素的数值调整为参考值。
 所述目标值和所述特征值可以根据实际需要进行设定，本发明实施例可设置所
 20 述目标值为 1，所述参考值为 0。

步骤 S906-S907 的具体处理过程可以为：首先遍历所述时间特征序列 $t(n)$ 中
 各时间特征元素的数值，从中找到最大数值对应的时间特征元素；排除查找到
 的时间特征元素之后，再次遍历所述时间特征序列 $t(n)$ 中各时间特征元素的数
 值，从中找到最大数值对应的时间特征元素；循环上述遍历过程，直到查找到
 25 $M-1$ 个最大数值为止。最后将所述时间特征序列 $t(n)$ 中查找到的 $M-1$ 个最大数
 值均调整为 1，将其他数值调整为 0。

本实施例的步骤 S906-步骤 S907 可以为图 8 所示实施例的步骤 S803 的具
 体细化步骤。由于 M 个字幕段落刚好对应 $M-1$ 个段落转折点，经过步骤 S906-
 步骤 S907 可使调整后的所述时间特征序列 $t(n)$ 刚好能够提取到 M 个字幕段落
 30 对应的 $M-1$ 段落转折点，从而实现对目标音频文件的实际分段需求。

S908，从调整后的所述时间特征序列中获取数值为目标值的时间特征元素

对应的目标索引。本步骤需要获取数值为 1 的时间特征元素对应的目标索引，即需要获取所查找到的 $M - 1$ 个时间特征元素的索引。

S909，根据所述目标索引在所述字幕文件中定位段落转折的字符单句。

假设其中一个目标索引为 5，则可在所述字幕文件中定位段落转折的字符单句为第 5 个字符单句，也就是说，第 5 个字符单句为一个字幕段落的起始位置，即所述字幕文件中第 1-4 个字符单句构成一个字幕段落。同理，可以定位 $M - 1$ 个段落转折的字符单句。

S910，根据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

由于所述字幕文件中记录了每个字符单句的关键信息，包括每个字符单句的开始时间和结束时间；本步骤可以从所述字幕文件中读取段落变化时间，按照本实施例所示例子，所述字幕文件中第 1-4 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 4 个字符单句的结束时间和第 5 个字符单句的开始时间。

本实施例的步骤 S908-步骤 S910 可以为图 8 所示实施例的步骤 S804 的具体细化步骤。依据步骤 S908-步骤 S910 可获得 M 个字幕段落的起止时间。

S911，按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。由于音频文件与字幕文件相互对应，那么，按照所获得的 M 个字幕段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

本实施例的步骤 S911 可参见图 8 所示实施例的步骤 S805，在此不赘述。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

30

下述将结合附图 10-附图 13，对本发明实施例提供的音频处理装置的结构

和功能进行详细介绍。需要说明的是，下述附图 10-附图 13 的所示的装置可以运行于终端中，以被应用于执行上述附图 8-附图 9 所示的方法。

请参见图 10，为本发明实施例提供的一种音频处理装置的结构示意图；该装置可包括：获取单元 1001、构建单元 1002、调整单元 1003、确定单元 1004 和分段单元 1005。
5

获取单元 1001，用于获取目标音频文件对应的字幕文件，所述字幕文件由至少一个字符单句顺序组成。

一个音频文件对应一个字幕文件。所述字幕文件包括至少一个字符单句及各字符单句的关键信息；一个字符单句的关键信息包括：标识（ID）、开始时间（start_time）和结束时间（end_time）。通常，互联网音频库中会存储多个音频文件、每个音频文件的属性以及每个音频文件对应的字幕文件，其中，音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识等等。所述获取单元 1001 可以从互联网音频库中获取目标音频文件对应的字幕文件；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件对应的字幕文件，并获取所查找到的字幕文件；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在互联网音频库中定位目标音频文件，并获取对应的字幕文件。
10
15

本发明实施例中，假设目标音频文件为歌曲 A，歌曲 A 对应的字幕文件的结构可参见本实施例所示例子，假设所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，假设该 N 个字符单句采用 $p(0)$ 至 $p(N-1)$ 进行表示，那么， $p(0)$ 可用于表示第一个字符单句 “ $a_1a_2a_3a_4a_5a_6a_7a_8$ ”， $p(1)$ 可用于表示第二个字符单句 “ $b_1b_2b_3b_4b_5b_6b_7b_8$ ”， $p(2)$ 可用于表示第三个字符单句 “ $c_1c_2c_3c_4c_5c_6c_7c_8$ ”，以此类推， $p(N-1)$ 用于表示第 N 个字符单句。
20
25

构建单元 1002，用于根据所述至少一个字符单句之间的时间间隔构建时间特征序列，所述时间特征序列包括至少一个时间特征元素。

所述时间特征序列可用于反映所述至少一个字符单句之间的时间间隔程度。首先所述构建单元 1002 计算所述至少一个字符单句之间的时间间隔，此处需要计算 $p(1)$ 与 $p(0)$ 之间的时间间隔 $p(1).start_time - p(0).end_time$ ；计算 $p(2)$ 与 $p(1)$ 之间的时间间隔 $p(2).start_time - p(1).end_time$ ；以此类推，计算 $p(N-1)$ 与 $p(N-2)$ 之间的时间间隔 $p(N-1).start_time - p(N-2).end_time$ 。其次所述构建单元
30

1002 可以根据所述至少一个字符单句的数量、顺序以及计算获得的时间间隔构建所述时间特征序列。

按照本实施例所示例子，假设采用 $t(n)$ 来表示所述时间特征序列，则所构建的时间特征序列 $t(n)$ 共包括 N 个时间特征元素，分别为 $t(0)、t(1) \dots t(N-1)$ 。其中， $t(0)$ 的数值可设置为 0， $t(1)$ 的数值用于表示 $p(1)$ 与 $p(0)$ 之间的时间间隔； $t(2)$ 的数值用于表示 $p(2)$ 与 $p(1)$ 之间的时间间隔；以此类推， $t(N-1)$ 的数值用于表示 $p(N-1)$ 与 $p(N-2)$ 之间的时间间隔。

调整单元 1003，用于按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值。

所述预设段落总数可以根据用户对目标音频文件的实际分段需求设定。假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，则所述调整单元 1003 按照预设段落总数 M 调整所述时间特征序列 $t(n)$ 中的各时间特征元素的数值目的在于，使调整后的所述时间特征序列 $t(n)$ 刚好能够提取到 M 个字幕段落对应的转折点，从而实现对目标音频文件的实际分段需求。

确定单元 1004，用于根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间。

所述调整后的时间特征序列 $t(n)$ 中的各时间特征元素的数值能够反映 M 个字幕段落对应的转折点，那么，所述确定单元 1004 可以根据调整后的所述时间特征序列中的至少一个时间特征元素的数值，从字幕文件中获得 M 个字幕段落的起止时间。

分段单元 1005，用于按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。

由于音频文件与字幕文件相互对应，那么，所述分段单元 1005 按照所获得的 M 个字幕段落的起止时间，对应地可以对所述目标音频文件进行段落划分，获得 M 个音频段落。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔。

隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 11，为图 10 所示的构建单元的实施例的结构示意图；该构建单元 1002 可包括：数量确定单元 1101、索引确定单元 1102、数值设置单元 1103 和序列构建单元 1104。

数量确定单元 1101，用于根据所述至少一个字符单句的数量确定构建时间特征序列的时间特征元素的数量。

所述字幕文件由 N (N 为正整数) 个字符单句顺序组成，即所述至少一个字符单句的数量为 N ，那么，所述数量确定单元 1101 可确定所述时间特征序列的时间特征元素的数量也为 N ，即所述时间特征序列的长度为 N 。假设采用 $t(n)$ 来表示所述时间特征序列，则所构建的时间特征序列 $t(n)$ 共包括 N 个时间特征元素，分别为 $t(0)$ 、 $t(1) \dots t(N-1)$ 。

索引确定单元 1102，用于根据所述至少一个字符单句中各字符单句的顺序，确定构建所述时间特征序列的各时间特征元素的索引。

所述字幕文件 N 个字符单句的顺序排列为 $p(0)$ 、 $p(1) \dots p(N-1)$ ，假设所述时间特征序列 $t(n)$ 中： $t(0)$ 对应 $p(0)$ ， $t(1)$ 对应 $p(1)$ ，以此类推、 $t(N-1)$ 对应 $p(N-1)$ 。那么，所述时间特征序列 $t(n)$ 中 $t(0)$ 的索引为 1，即第一个时间特征元素； $t(1)$ 的索引为 2，即第二个时间特征元素；以此类推， $t(N-1)$ 的索引为 N ，即第 N 个时间特征元素。

数值设置单元 1103，用于针对所述至少一个字符单句中的任一个目标字符单句，将所述目标字符单句与所述目标字符单句的相邻在先字符单句之间的时间间隔设置为所述目标字符单句对应的时间特征元素的数值。

所述数值设置单元 1103 的具体处理过程可包括以下 A-B：

A、计算每一个字符单句与其相邻在先字符单句之间的时间间隔，此处需要计算 $p(1)$ 与 $p(0)$ 之间的时间间隔 $p(1).start_time - p(0).end_time$ ；计算 $p(2)$ 与 $p(1)$ 之间的时间间隔 $p(2).start_time - p(1).end_time$ ；以此类推，计算 $p(N-1)$ 与 $p(N-2)$ 之间的时间间隔 $p(N-1).start_time - p(N-2).end_time$ 。

B、将计算获得的时间间隔设置为相应时间特征元素的数值；那么，可设置 $t(0) = 0$ ， $t(1) = p(1).start_time - p(0).end_time$ ， $t(2) = p(2).start_time - p(1).end_time$ ，以此类推， $t(N-1) = p(N-1).start_time - p(N-2).end_time$ 。

序列构建单元 1104，用于按照构建所述时间特征序列的时间特征元素的数量、索引及数值，构建所述时间特征序列。

所构建的所述时间特征序列为 $t(n)$ ， $t(n)$ 是由 N 个时间特征元素 $t(0)$ 、 $t(1) \dots t(N-1)$ 顺序组成，且所述时间特征序列 $t(n)$ 中各时间特征元素的数值为
5 $t(0)=0$ ， $t(1)=p(1).start_time-p(0).end_time$ ， $t(2)=p(2).start_time-p(1).end_time$ ，以此类推， $t(N-1)=p(N-1).start_time-p(N-2).end_time$ 。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。
10
15

请参见图 12，为图 10 所示的调整单元的实施例的结构示意图；该调整单元 1003 可包括：元素查找单元 1201 和数值调整单元 1202。

元素查找单元 1201，用于从所述时间特征序列中查找前预设段落数量减 1 个最大数值的时间特征元素。

假设采用 M (M 为正整数且 $M > 1$) 表示所述预设段落总数，所述元素查找单元 1201 需要从所述时间特征序列 $t(n)$ 中查找前 $M-1$ 个最大数值的时间特征元素。
20

数值调整单元 1202，用于将查找到的时间特征元素的数值调整为目标值，将所述时间特征序列中除查找到的时间特征元素之外的其他时间特征元素的数值调整为参考值。所述目标值和所述特征值可以根据实际需要进行设定，本发明实施例可设置所述目标值为 1，所述参考值为 0。
25

所述元素查找单元 1201 和所述数值调整单元 1202 的具体处理过程可以为：首先所述元素查找单元 1201 遍历所述时间特征序列 $t(n)$ 中各时间特征元素的数值，从中找到最大数值对应的时间特征元素；排除查找到的时间特征元素之后，再次遍历所述时间特征序列 $t(n)$ 中各时间特征元素的数值，从中找到最大数值对应的时间特征元素；循环上述遍历过程，直到查找到 $M-1$ 个最大数值
30

为止。最后所述数值调整单元 1202 将所述时间特征序列 $t(n)$ 中查找到的 $M-1$ 个最大数值均调整为 1，将其他数值调整为 0。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。
5
10

请参见图 13，为图 10 所示的确定单元的实施例的结构示意图；该确定单元 1004 可包括：目标索引获取单元 1301、定位单元 1302 和时间读取单元 1303。

目标索引获取单元 1301，用于从调整后的所述时间特征序列中获取数值为目标值的时间特征元素对应的目标索引。
15

按照图 5 所示实施例的例子，所述目标索引获取单元 1301 需要获取数值为 1 的时间特征元素对应的目标索引，即需要获取所查找到的 $M-1$ 个时间特征元素的索引。

定位单元 1302，用于根据所述目标索引在所述字幕文件中定位段落转折的字符单句。
20

假设其中一个目标索引为 5，所述定位单元 1302 则可在所述字幕文件中定位段落转折的字符单句为第 5 个字符单句，也就是说，第 5 个字符单句为一个字幕段落的起始位置，即所述字幕文件中第 1-4 个字符单句构成一个字幕段落。同理，可以定位 $M-1$ 个段落转折的字符单句。
25

时间读取单元 1303，用于根据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

由于所述字幕文件中记录了每个字符单句的关键信息，包括每个字符单句的开始时间和结束时间；所述时间读取单元 1303 以从所述字幕文件中读取段落变化时间，按照本实施例所示例子，所述字幕文件中第 1-4 个字符单句构成一个字幕段落，那么所读取的段落变化时间为：第 4 个字符单句的结束时间和第 5 个字符单句的开始时间。
30

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

本发明实施例还公开了一种终端，该终端可以为 PC (Personal Computer, 个人计算机)、笔记本电脑、手机、PAD (平板电脑)、车载终端、智能可穿戴设备等设备。该终端中可包括一音频处理装置，该装置的结构和功能可参见上述图 10-图 13 所示实施例的相关描述，在此不赘述。

本发明实施例中，可根据目标音频文件对应的字幕文件中的至少一个字符单句之间的时间间隔构建时间特征序列，按照预设段落总数调整所述时间特征序列中的各时间特征元素的数值，并根据调整后的所述时间特征序列中的至少一个时间特征元素的数值确定段落变化时间，然后按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落，该音频处理过程利用字幕段落之间的字符单句的时间间隔特点，基于字幕文件中的字符单句之间的时间间隔实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

基于上述描述，下面将结合附图 14-附图 15，对本发明实施例提供的音频处理方法进行详细介绍。

请参见图 14，为本发明实施例提供的一种音频处理方法的流程图；该方法可包括以下步骤 S1401-步骤 S1405。

S1401，获取目标音频文件的音频数据，所述音频数据包括至少一个音频帧。

一个音频文件包括音频数据，对音频文件进行解码（例如：PCM 解码）可以获得音频数据（例如：PCM 数据）。本步骤可以对目标音频文件进行解码，获得所述目标音频文件的音频数据。所述音频数据可包括至少一个音频帧，所

述音频数据可以表示为所述至少一个音频帧顺序组成的一个帧序列。

本发明实施例中，设定所述音频数据包含 N 个音频帧， N 为正整数，即 N 为所述音频数据的采样点数，所述音频数据可表示为 $x(n)$ ，其中， n 为正整数且 $n=0,1,2,N-1$ 。

5 S1402，根据所述至少一个音频帧的相关性构建峰值特征序列，所述峰值特征序列包括至少一个峰值特征元素。

所述峰值特征序列可用于反映所述至少一个音频帧的相似性。本步骤中，首先可采用相关计算公式计算所述至少一个音频帧的相关性，此处通过计算可获得所述至少一个音频帧的相关函数序列，假设采用 $r()$ 表示相关函数，那么，
10 相关计算可获得 $r(n)$ 、 $r(n+1)$ 、 $r(n+2) \dots r(N-2)$ 、 $r(N-1)$ 。其次可通过对所述至少一个音频帧的相关函数序列进行最大值、峰值等分析，构建峰值特征序列。

本发明实施例中，假设采用 $v(n)$ 表示所述峰值特征序列。则所构建的峰值特征序列 $v(n)$ 共包括 N 个波峰特征元素，分别为 $v(0)$ 、 $v(1) \dots v(N-1)$ 。其中， $v(0)$ 的数值可用于描述音频帧 $x(0)$ 与其在后音频帧之间的相关性； $v(1)$ 的数值可用于
15 描述 $x(1)$ 与其在后音频帧之间的相关性；以此类推。

S1403，对所述峰值特征序列进行规整处理。

本步骤中，可以采用预设的间隔系数对应的扫描区间对所述峰值特征序列 $v(n)$ 进行规整处理。规整处理的目的在于：使得所述峰值特征序列 $v(n)$ 在所述预设的间隔系数对应的扫描区间内仅存在一个最大峰值，以保证后续段落划分的准确性。
20

S1404，根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间。

所述规整后的所述峰值特征序列 $v(n)$ 中的各峰值特征元素的数值可用于描述音频帧之间的相关性，那么，本步骤可以根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定出现音频段落变化的时间。
25

S1405，按照所述段落变化时间对所述目标音频文件进行段落划分。按照所获得的音频段落变化的时间，可以对所述目标音频文件进行段落划分。

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，
30 按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利

用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 15，为本发明实施例提供的另一种音频处理方法的流程图；该方法可包括以下步骤 S1501-步骤 S1510。

S1501，获取所述目标音频文件的类型，所述类型包括：双声道类型或单声道类型。

通常，互联网音频库中会存储多个音频文件及每个音频文件的属性，其中，音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识、音频文件的类型等等。本步骤中，可以从互联网音频库中获取目标音频文件的类型；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件的类型；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在互联网音频库中定位目标音频文件，并获取所述目标音频文件的类型。

S1502，若所述目标音频文件的类型为单声道类型，对所述目标音频文件从所述单声道输出的内容进行解码获得音频数据；或者，若所述目标音频文件的类型为双声道类型，从所述双声道中选取一个声道，对所述目标音频文件从所选取声道输出的内容进行解码获得音频数据；或者将所述双声道处理为混合声道，对所述目标音频文件从所述混合声道输出的内容进行解码获得音频数据。

其中，若所述目标音频文件的类型为单声道类型，则所述目标音频文件通过一个声道输出音频内容，本步骤需要对该单声道输出的音频内容进行解码获得音频数据。若所述目标音频文件的类型为双声道类型，则所述目标音频文件通过两个声道输出音频内容，本步骤可以选取一个声道输出的音频内容进行解码获得音频数据，另外，本步骤也可以首先可以采用 Downmix 等处理方式将两个声道处理为混合声道，再对混合声道输出的音频内容进行解码获得音频数据。

本发明实施例中，设定所述音频数据包含 N 个音频帧， N 为正整数，即 N 为所述音频数据的采样点数，所述音频数据可表示为 $x(n)$ ，其中， n 为正整数且 $n = 0, 1, 2, N-1$ 。

本实施例的步骤 S1501-步骤 S1502 可以为图 14 所示实施例的步骤 S1401

的具体细化步骤。

S1503，对所述至少一个音频帧中各音频帧进行相关计算，获得所述至少一个音频帧对应的相关函数序列。

可采用相关计算公式计算所述至少一个音频帧的相关性，该相关计算公式

5 可表示如下：

$$r(n+i) = \frac{\sum_{m=0}^L x(n+m)x(n+i+m)}{L * M * M} \quad (1)$$

上述公式(1)中， i 为整数且 $0 \leq i \leq N-1$ ； m 为整数且 $0 \leq m \leq L$ ； L 为所述音频数据长度，假设所述音频数据的采样时间为 T 及采样率为 f ，则 $L=f*T$ ； M 为采样值的最大值，例如：若采样值为 16bit， $M=32767$ ；若采样值为 8bit，
10 $M=255$ ，等等。

通过上述公式(1)可计算获得所述至少一个音频帧的相关函数序列为 $r(n)$ 、
 $r(n+1)$ 、 $r(n+2) \dots r(N-2)$ 、 $r(N-1)$ 。

S1504，对所述至少一个音频帧对应的相关函数序列进行最大值计算，生成基准序列。

15 所述基准序列可表示为 $D(n)$ ，本步骤可采用最大值计算公式求取所述基准序列，该最大值计算公式可表示如下：

$$D(n) = \max(r(n), r(n+1), r(n+2), \dots, r(N-2), r(N-1)) \quad (2)$$

上述公式(2)中， $\max()$ 为最大值求取函数。

经过上述公式(2)所获得的基准序列 $D(n)$ ，其共包括 N 个元素，分别为 $d(0)$ 、
20 $d(1) \dots d(N-1)$ 。

S1505，对所述基准序列进行峰值求取计算，获得所述峰值特征序列。

假设采用 $v(n)$ 表示所述峰值特征序列。则所构建的峰值特征序列 $v(n)$ 共包括 N 个波峰特征元素，分别为 $v(0)$ 、 $v(1) \dots v(N-1)$ 。其中， $v(0)$ 的数值可用于描述音频帧 $x(0)$ 与其在后音频帧之间的相关性； $v(1)$ 的数值可用于描述 $x(1)$ 与其在后音频帧之间的相关性；以此类推。本步骤中，对所述基准序列 $D(n)$ 进行峰值求取计算，其计算原则为，如果元素 $d(i)$ （其中， i 为整数且 $0 \leq i \leq N-1$ ）的数值大于或等于与 $d(i)$ 前后相邻的元素的数值，则使得 $v(i)=d(i)$ ；如果元素 $d(i)$ 的数值小于与 $d(i)$ 前后相邻的任一元素的数值，则使得 $v(i)=0$ 。通过此计算原则，可以获得所述峰值特征序列 $v(n)$ 的各峰值特征元素的数值。

30 本实施例的步骤 S1503-步骤 S1505 可以为图 14 所示实施例的步骤 S1402 的具体细化步骤。

S1506，获取预设的间隔系数对应的扫描区间。

所述预设的间隔系数可以根据实际需要进行设定，假设所述预设的间隔系数为 Q ，那么，所述预设的间隔系数对应的扫描区间可以为 $[i-Q/2, i+Q/2]$ （其中， i 为整数且 $0 \leq i \leq N-1$ ）。

5 S1507，采用所述预设的间隔系数对应的扫描区间对所述峰值特征序列进行规整，将所述预设的间隔系数对应的扫描区间内的最大峰值对应的峰值特征元素的数值设置为目标值，将所述预设的间隔系数对应的扫描区间内除所述最大峰值对应的峰值特征元素之外的其他峰值特征元素的数值设置为初始值。

10 所述目标值和所述特征值可以根据实际需要进行设定，本发明实施例可设置所述目标值为1，所述参考值为0。

步骤S1506-步骤S1507对所述峰值特征序列 $v(n)$ 规整处理的目的在于：使得所述峰值特征序列 $v(n)$ 在所述预设的间隔系数对应的扫描区间内仅存在一个最大峰值，以保证后续段落划分的准确性。本实施例的步骤S1506-步骤S1507可以为图14所示实施例的步骤S1403的具体细化步骤。

15 S1508，从规整后的所述峰值特征序列中获取数值为目标值的峰值特征元素对应的目标索引。本步骤需要获取数值为1的峰值特征元素对应的目标索引，例如：假设 $v(i)=1$ ，本步骤可获得的目标索引即为 i 。

S1509，根据所述目标索引及所述目标音频文件的采样率，计算段落变化时间。

20 本步骤可以采用所述目标索引除以所述目标音频文件的采样率来获得段落变化时间，按照本实施例所示例子，所获得的目标索引为 i ，采样率为 f ，则某个段落变化时间为 i/f ，例如：若目标索引 $i=441000$ ，采样率 $f=44100$ ，那么 $i/f=100$ ，也就是说，所述目标音频文件中在100s的地方出现音频段落的变化。

25 S1510，按照所述段落变化时间对所述目标音频文件进行段落划分。按照所获得的音频段落变化的时间，可以对所述目标音频文件进行段落划分。

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

下述将结合附图 16-附图 20，对本发明实施例提供的音频处理装置的结构和功能进行详细介绍。需要说明的是，下述附图 16-附图 20 的所示的装置可以运行于终端中，以被应用于执行上述附图 14-附图 15 所示的方法。

5 请参见图 16，为本发明实施例提供的一种音频处理装置的结构示意图；该装置可包括：获取单元 1601、构建单元 1602、规整处理单元 1603、确定单元 1604 和分段单元 1605。

10 获取单元 1601，用于获取目标音频文件的音频数据，所述音频数据包括至少一个音频帧。

15 一个音频文件包括音频数据，对音频文件进行解码（例如：PCM 解码）可以获得音频数据（例如：PCM 数据）。所述获取单元 1601 可以对目标音频文件进行解码，获得所述目标音频文件的音频数据。所述音频数据可包括至少一个音频帧，所述音频数据可以表示为所述至少一个音频帧顺序组成的一个帧序列。

20 本发明实施例中，设定所述音频数据包含 N 个音频帧， N 为正整数，即 N 为所述音频数据的采样点数，所述音频数据可表示为 $x(n)$ ，其中， n 为正整数且 $n = 0, 1, 2, N-1$ 。

25 构建单元 1602，用于根据所述至少一个音频帧的相关性构建峰值特征序列，所述峰值特征序列包括至少一个峰值特征元素。

20 所述峰值特征序列可用于反映所述至少一个音频帧的相似性。首先所述构建单元 1602 可采用相关计算公式计算所述至少一个音频帧的相关性，此处通过计算可获得所述至少一个音频帧的相关函数序列，假设采用 $r()$ 表示相关函数，那么，相关计算可获得 $r(n)、r(n+1)、r(n+2) \dots r(N-2)、r(N-1)$ 。其次所述构建单元 1602 可通过对所述至少一个音频帧的相关函数序列进行最大值、峰值等分析，构建峰值特征序列。

30 本发明实施例中，假设采用 $v(n)$ 表示所述峰值特征序列。则所构建的峰值特征序列 $v(n)$ 共包括 N 个波峰特征元素，分别为 $v(0)、v(1) \dots v(N-1)$ 。其中， $v(0)$ 的数值可用于描述音频帧 $x(0)$ 与其在后音频帧之间的相关性； $v(1)$ 的数值可用于描述 $x(1)$ 与其在后音频帧之间的相关性；以此类推。

35 规整处理单元 1603，用于对所述峰值特征序列进行规整处理。

所述规整处理单元 1603 可以采用预设的间隔系数对应的扫描区间对所述

峰值特征序列 $v(n)$ 进行规整处理。规整处理的目的在于：使得所述峰值特征序列 $v(n)$ 在所述预设的间隔系数对应的扫描区间内仅存在一个最大峰值，以保证后续段落划分的准确性。

确定单元 1604，用于根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间。
5

所述规整后的所述峰值特征序列 $v(n)$ 中的各峰值特征元素的数值可用于描述音频帧之间的相关性，那么，所述确定单元 1604 可以根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定出现音频段落变化的时间。

分段单元 1605，用于按照所述段落变化时间对所述目标音频文件进行段落
10 划分。

按照所获得的音频段落变化的时间，所述分段单元 1605 可以对所述目标音频文件进行段落划分。

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整
15 后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

20 请参见图 17，为图 16 所示的获取单元的实施例的结构示意图；该获取单元 1601 可包括：类型获取单元 1701 和解码单元 1702。

类型获取单元 1701，用于获取所述目标音频文件的类型，所述类型包括：双声道类型或单声道类型。

通常，互联网音频库中会存储多个音频文件及每个音频文件的属性，其中，
25 音频文件的属性可包括但不限于：音频文件的音频特征、音频文件的标识、音频文件的类型等等。所述类型获取单元 1701 可以从互联网音频库中获取目标音频文件的类型；具体的获取方式可包括但不限于：可以根据目标音频文件的标识，在互联网音频库中查找该目标音频文件的类型；或者，可以提取目标音频文件的音频特征与互联网音频库中的音频文件的音频特征进行匹配，由此在
30 互联网音频库中定位目标音频文件，并获取所述目标音频文件的类型。

解码单元 1702，用于若所述目标音频文件的类型为单声道类型，对所述目

标音频文件从所述单声道输出的内容进行解码获得音频数据；或者，用于若所述目标音频文件的类型为双声道类型，从所述双声道中选取一个声道，对所述目标音频文件从所选取声道输出的内容进行解码获得音频数据；或者将所述双声道处理为混合声道，对所述目标音频文件从所述混合声道输出的内容进行解码获得音频数据。

其中，若所述目标音频文件的类型为单声道类型，则所述目标音频文件通过一个声道输出音频内容，所述解码单元 1702 需要对该单声道输出的音频内容进行解码获得音频数据。若所述目标音频文件的类型为双声道类型，则所述目标音频文件通过两个声道输出音频内容，所述解码单元 1702 可以选取一个声道输出的音频内容进行解码获得音频数据，另外，所述解码单元 1702 也可以首先可以采用 Downmix 等处理方式将两个声道处理为混合声道，再对混合声道输出的音频内容进行解码获得音频数据。

本发明实施例中，设定所述音频数据包含 N 个音频帧， N 为正整数，即 N 为所述音频数据的采样点数，所述音频数据可表示为 $x(n)$ ，其中， n 为正整数且 $n = 0, 1, 2, N-1$ 。

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 18，为图 16 所示的构建单元的实施例的结构示意图；该构建单元 1602 可包括：相关计算单元 1801、生成单元 1802 和序列求取单元 1803。

相关计算单元 1801，用于对所述至少一个音频帧中各音频帧进行相关计算，获得所述至少一个音频帧对应的相关函数序列。

所述相关计算单元 1801 可采用相关计算公式计算所述至少一个音频帧的相关性，该相关计算公式可表示如图 2 所示实施例中的公式（1）。通过上述公式（1）可计算获得所述至少一个音频帧的相关函数序列为 $r(n)、r(n+1)、r(n+2) \dots r(N-2)、r(N-1)$ 。

生成单元 1802，用于对所述至少一个音频帧对应的相关函数序列进行最大

值计算，生成基准序列。

所述基准序列可表示为 $D(n)$ ，所述生成单元 1802 可采用最大值计算公式求取所述基准序列，该最大值计算公式可表示为图 2 所示实施例中的公式(2)。经过上述公式(2)所获得的基准序列 $D(n)$ ，其共包括 N 个元素，分别为 $d(0)$ 、
5 $d(1) \dots d(N-1)$ 。

序列求取单元 1803，用于对所述基准序列进行峰值求取计算，获得所述峰值特征序列。

假设采用 $v(n)$ 表示所述峰值特征序列。则所构建的峰值特征序列 $v(n)$ 共包括 N 个波峰特征元素，分别为 $v(0)$ 、 $v(1) \dots v(N-1)$ 。其中， $v(0)$ 的数值可用于描述音频帧 $x(0)$ 与其在后音频帧之间的相关性； $v(1)$ 的数值可用于描述 $x(1)$ 与其在后音频帧之间的相关性；以此类推。所述序列求取单元 1803 对所述基准序列 $D(n)$ 进行峰值求取计算，其计算原则为，如果元素 $d(i)$ （其中， i 为整数且 $0 \leq i \leq N-1$ ）的数值大于或等于与 $d(i)$ 前后相邻的元素的数值，则使得 $v(i) = d(i)$ ；如果元素 $d(i)$ 的数值小于与 $d(i)$ 前后相邻的任一元素的数值，则使得 $v(i) = 0$ 。
10 通过此计算原则，可以获得所述峰值特征序列 $v(n)$ 的各峰值特征元素的数值。
15

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。
20

请参见图 19，为图 16 所示的规整处理单元的实施例的结构示意图；该规整处理单元 1603 可包括：区间获取单元 1901 和规整单元 1902。

25 区间获取单元 1901，用于获取预设的间隔系数对应的扫描区间。

所述预设的间隔系数可以根据实际需要进行设定，假设所述预设的间隔系数为 Q ，那么，所述预设的间隔系数对应的扫描区间可以为 $[i-Q/2, i+Q/2]$ （其中， i 为整数且 $0 \leq i \leq N-1$ ）。

30 规整单元 1902，用于采用所述预设的间隔系数对应的扫描区间对所述峰值特征序列进行规整，将所述预设的间隔系数对应的扫描区间内的最大峰值对应的峰值特征元素的数值设置为目标值，将所述预设的间隔系数对应的扫描区间

内除所述最大峰值对应的峰值特征元素之外的其他峰值特征元素的数值设置为初始值。所述目标值和所述特征值可以根据实际需要进行设定，本发明实施例可设置所述目标值为 1，所述参考值为 0。

对所述峰值特征序列 $v(n)$ 规整处理的目的在于：使得所述峰值特征序列 $v(n)$ 在所述预设的间隔系数对应的扫描区间内仅存在一个最大峰值，以保证后续段落划分的准确性。
5

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，
10 按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

请参见图 20，为图 16 所示的确定单元的实施例的结构示意图；该确定单元 1604 可包括：目标索引获取单元 2001 和时间计算单元 2002。
15

目标索引获取单元 2001，用于从规整后的所述峰值特征序列中获取数值为目标值的峰值特征元素对应的目标索引。

按照图 19 所示实施例所示的例子，所述目标索引获取单元 2001 需要获取数值为 1 的峰值特征元素对应的目标索引，例如：假设 $v(i)=1$ ，所述目标索引获取单元 2001 可获得的目标索引即为 i 。
20

时间计算单元 2002，用于根据所述目标索引及所述目标音频文件的采样率，计算段落变化时间。

所述时间计算单元 2002，可以采用所述目标索引除以所述目标音频文件的采样率来获得段落变化时间，按照本实施例所示例子，所获得的目标索引为 i ，采样率为 f ，则某个段落变化时间为 i/f ，例如：若目标索引 $i=441000$ ，采样率 $f=44100$ ，那么 $i/f=100$ ，也就是说，所述目标音频文件中在 100s 的地方出现音频段落的变化。
25

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，
30 按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利

用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

本发明实施例还公开了一种终端，该终端可以为 PC (Personal Computer, 5 个人计算机)、笔记本电脑、手机、PAD (平板电脑)、车载终端、智能可穿戴设备等设备。该终端中可包括一音频处理装置，该装置的结构和功能可参见上述图 16-图 20 所示实施例的相关描述，在此不赘述。

本发明实施例中，可根据目标音频文件的音频数据包含的至少一个音频帧的相关性构建峰值特征序列，对所述峰值特征序列进行规整处理，并根据规整 10 后的所述峰值特征序列中的至少一个峰值特征元素的数值确定段落变化时间，按照所述段落变化时间对所述目标音频文件进行段落划分，该音频处理过程利用音频段落之间的音频帧的相关性特点，实现对目标音频文件的段落划分，可提升分段处理效率，提升音频处理的智能性。

15 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种音频处理方法，其特征在于，包括：

5 获取目标音频文件的文件数据；

根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列；

10 按照预设段落总数对所述相关性特征序列进行优化；

根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间；

15 按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。

2、根据权利要求 1 所述的方法，其特征在于，根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列包括：

15 所述文件数据是指字幕文件，所述字幕文件由至少一个字符单句顺序组成，根据所述至少一个字符单句之间的相似度构建字幕特征序列，所述字幕特征序列包括至少一个字符特征元素。

20 3、如权利要求 2 所述的方法，其特征在于，所述根据所述至少一个字符单句之间的相似度构建字幕特征序列，包括：

根据所述至少一个字符单句的数量确定构建字幕特征序列的字符特征元素的数量；根据所述至少一个字符单句中各字符单句的顺序，确定构建所述字幕特征序列的各字符特征元素的索引；将构建所述字幕特征序列的各字符特征元素的数值均设置为初始值；针对所述至少一个字符单句中的任一个目标字符单句，若所述目标字符单句与所述目标字符单句的在后字符单句之间的最大相似度大于预设相似阈值，将所述目标字符单句对应的字符特征元素的数值从初始值变更为目标值；按照构建所述字幕特征序列的字符特征元素的数量、索引及数值，构建所述字幕特征序列。

30 4、如权利要求 3 所述的方法，其特征在于，所述按照预设段落总数对所述相关性特征序列进行优化，包括：

统计所述字幕特征序列中数值为目标值的字符特征元素的数量；判断所述数量是否位于所述预设段落总数对应的容错区间内；若否，调整所述预设相似阀值的大小以调整所述字幕特征序列中的各字符特征元素的数值。

5 5、如权利要求4所述的方法，其特征在于，所述若否，调整所述预设相似阀值的大小以调整所述字幕特征序列中的各字符特征元素的数值，包括：

若所述数量大于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长增大所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值；若所述数量小于所述预设段落总数对应的容错区间内的最大容错值，
10 按照预设步长减小所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

6、如权利要求5所述的方法，其特征在于，所述根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间，包括：

15 从优化后的所述字幕特征序列中获取数值为目标值的字符特征元素对应的目标索引；根据所述目标索引在所述字幕文件中定位段落转折的字符单句；根据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

7、根据权利要求1所述的方法，其特征在于，根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列包括：

所述文件数据是指字幕文件，所述字幕文件由至少一个字符单句顺序组成，根据所述至少一个字符单句之间的时间间隔构建时间特征序列，所述时间特征序列包括至少一个时间特征元素。

25 8、如权利要求7所述的方法，其特征在于，所述根据所述至少一个字符单句之间的时间间隔构建时间特征序列，包括：

根据所述至少一个字符单句的数量确定构建时间特征序列的时间特征元素的数量；根据所述至少一个字符单句中各字符单句的顺序，确定构建所述时间特征序列的各时间特征元素的索引；针对所述至少一个字符单句中的任一个目标字符单句，将所述目标字符单句与所述目标字符单句的相邻在先字符单句之间的时间间隔设置为所述目标字符单句对应的时间特征元素的数值；按照构建

所述时间特征序列的时间特征元素的数量、索引及数值，构建所述时间特征序列。

9、如权利要求 8 所述的方法，其特征在于，所述按照预设段落总数对所述
5 相关性特征序列进行优化，包括：

从所述时间特征序列中查找前预设段落数量减 1 个最大数值的时间特征元
素； 将查找到的时间特征元素的数值调整为目标值，将所述时间特征序列中除
查找到的时间特征元素之外的其他时间特征元素的数值调整为参考值。

10 10、如权利要求 9 所述的方法，其特征在于，所述根据优化后的所述相关
性特征序列中的至少一个特征元素的数值确定段落变化时间，包括：

从调整后的所述时间特征序列中获取数值为目标值的时间特征元素对应
的目标索引； 根据所述目标索引在所述字幕文件中定位段落转折的字符单句； 根
据所述段落转折的字符单句从所述字幕文件中读取段落变化时间。

15

11、根据权利要求 1 所述的方法，其特征在于，根据所述文件数据的组成
元素之间的相关性特征数据，构建相关性特征序列包括：

所述文件数据是指音频数据，所述音频数据包括至少一个音频帧，根据所
述至少一个音频帧的相关性构建峰值特征序列，所述峰值特征序列包括至少一
20 个峰值特征元素。

12、如权利要求 11 所述的方法，其特征在于，所述根据所述至少一个音频
帧的相关性构建峰值特征序列，包括：

对所述至少一个音频帧中各音频帧进行相关计算，获得所述至少一个音频
帧对应的相关函数序列； 对所述至少一个音频帧对应的相关函数序列进行最大
25 值计算，生成基准序列； 对所述基准序列进行峰值求取计算，获得所述峰值特
征序列。

13、如权利要求 12 所述的方法，其特征在于，所述按照预设段落总数对所
30 述相关性特征序列进行优化，包括：

获取预设的间隔系数对应的扫描区间； 采用所述预设的间隔系数对应的扫

描区间对所述峰值特征序列进行规整，将所述预设的间隔系数对应的扫描区间内的最大峰值对应的峰值特征元素的数值设置为目标值，将所述预设的间隔系数对应的扫描区间内除所述最大峰值对应的峰值特征元素之外的其他峰值特征元素的数值设置为初始值。

5

14、如权利要求 13 所述的方法，其特征在于，所述根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间，包括：

从规整后的所述峰值特征序列中获取数值为目标值的峰值特征元素对应的目标索引；根据所述目标索引及所述目标音频文件的采样率，计算段落变化时间。
10

15、如权利要求 11 所述的方法，其特征在于，所述获取目标音频文件的文件数据，包括：

获取所述目标音频文件的类型，所述类型包括：双声道类型或单声道类型；

15 若所述目标音频文件的类型为单声道类型，对所述目标音频文件从所述单声道输出的内容进行解码获得音频数据；

若所述目标音频文件的类型为双声道类型，从所述双声道中选取一个声道，对所述目标音频文件从所选取声道输出的内容进行解码获得音频数据；或者将所述双声道处理为混合声道，对所述目标音频文件从所述混合声道输出的内容
20 进行解码获得音频数据。

16、如权利要求 1-10 任一项所述的方法，其特征在于，所述字幕文件包括至少一个字符单句及各字符单句的关键信息；一个字符单句的关键信息包括：标识、开始时间和结束时间。

25

17、一种音频处理装置，其特征在于，包括：

获取单元，用于获取目标音频文件的文件数据；

构建单元，用于根据所述文件数据的组成元素之间的相关性特征数据，构建相关性特征序列；

30 优化单元，用于按照预设段落总数对所述相关性特征序列进行优化；

确定单元，用于根据优化后的所述相关性特征序列中的至少一个特征元素的数值确定段落变化时间；

分段单元，用于按照所述段落变化时间将所述目标音频文件划分为所述预设段落总数的段落。

5

18、如权利要求 17 所述的装置，其特征在于，所述构建单元用于所述文件数据是指字幕文件，所述字幕文件由至少一个字符单句顺序组成，根据所述至少一个字符单句之间的相似度构建字幕特征序列，所述字幕特征序列包括至少一个字符特征元素；或，

10 所述构建单元用于所述文件数据是指字幕文件，所述字幕文件由至少一个字符单句顺序组成，根据所述至少一个字符单句之间的时间间隔构建时间特征序列，所述时间特征序列包括至少一个时间特征元素；或，

15 所述构建单元用于所述文件数据是指音频数据，所述音频数据包括至少一个音频帧，根据所述至少一个音频帧的相关性构建峰值特征序列，所述峰值特征序列包括至少一个峰值特征元素。

19、如权利要求 18 所述的装置，其特征在于，所述构建单元包括：

数量确定单元，用于根据所述至少一个字符单句的数量确定构建字幕特征序列的字符特征元素的数量；

20 索引确定单元，用于根据所述至少一个字符单句中各字符单句的顺序，确定构建所述字幕特征序列的各字符特征元素的索引；

数值设置单元，用于将构建所述字幕特征序列的各字符特征元素的数值均设置为初始值；

25 数值变更单元，用于针对所述至少一个字符单句中的任一个目标字符单句，若所述目标字符单句与所述目标字符单句的在后字符单句之间的最大相似度大于预设相似阈值，将所述目标字符单句对应的字符特征元素的数值从初始值变更为目标值；

序列构建单元，用于按照构建所述字幕特征序列的字符特征元素的数量、索引及数值，构建所述字幕特征序列。

30

20、如权利要求 18 所述的装置，其特征在于，所述构建单元包括：

数量确定单元，用于根据所述至少一个字符单句的数量确定构建时间特征序列的时间特征元素的数量；

索引确定单元，用于根据所述至少一个字符单句中各字符单句的顺序，确定构建所述时间特征序列的各时间特征元素的索引；

5 数值设置单元，用于针对所述至少一个字符单句中的任一个目标字符单句，将所述目标字符单句与所述目标字符单句的相邻在先字符单句之间的时间间隔设置为所述目标字符单句对应的时间特征元素的数值；

序列构建单元，用于按照构建所述时间特征序列的时间特征元素的数量、索引及数值，构建所述时间特征序列。

10

21、如权利要求 18 所述的装置，其特征在于，所述构建单元包括：

相关计算单元，用于对所述至少一个音频帧中各音频帧进行相关计算，获得所述至少一个音频帧对应的相关函数序列；

15 生成单元，用于对所述至少一个音频帧对应的相关函数序列进行最大值计算，生成基准序列；

序列求取单元，用于对所述基准序列进行峰值求取计算，获得所述峰值特征序列。

22、如权利要求 19 所述的装置，其特征在于，所述优化单元包括：

20 数量统计单元，用于统计所述字幕特征序列中数值为目标值的字符特征元素的数量；

判断单元，用于判断所述数量是否位于所述预设段落总数对应的容错区间内；

25 优化处理单元，用于若判断结果为否，调整所述预设相似阀值的大小以调整所述字幕特征序列中的各字符特征元素的数值。

23、如权利要求 22 所述的装置，其特征在于，所述优化处理单元包括：

第一调整单元，用于若所述数量大于所述预设段落总数对应的容错区间内的最大容错值，按照预设步长增大所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值；

第二调整单元，用于若所述数量小于所述预设段落总数对应的容错区间内

的最大容错值，按照预设步长减小所述预设相似阀值以调整所述字幕特征序列中的各字符特征元素的数值。

24、如权利要求 19 所述的装置，其特征在于，所述确定单元包括：

5 目标索引获取单元，用于从优化后的所述字幕特征序列中获取数值为目标值的字符特征元素对应的目标索引；

定位单元，用于根据所述目标索引在所述字幕文件中定位段落转折的字符串单句；

时间读取单元，用于根据所述段落转折的字符串单句从所述字幕文件中读取
10 段落变化时间。

25、如权利要求 19 所述的装置，其特征在于，所述优化单元包括：

元素查找单元，用于从所述时间特征序列中查找前预设段落数量减 1 个最大数值的时间特征元素；

15 数值调整单元，用于将查找到的时间特征元素的数值调整为目标值，将所述时间特征序列中除查找到的时间特征元素之外的其他时间特征元素的数值调整为参考值。

26、如权利要求 19 所述的装置，其特征在于，所述确定单元包括：

20 目标索引获取单元，用于从调整后的所述时间特征序列中获取数值为目标值的时间特征元素对应的目标索引；

定位单元，用于根据所述目标索引在所述字幕文件中定位段落转折的字符串单句；

时间读取单元，用于根据所述段落转折的字符串单句从所述字幕文件中读取
25 段落变化时间。

27、如权利要求 19 至 26 任一项所述的装置，其特征在于，所述字幕文件包括至少一个字符串单句及各字符串单句的关键信息；

一个字符串单句的关键信息包括：标识、开始时间和结束时间。

30

28、如权利要求 19 所述的装置，其特征在于，所述优化单元包括：

区间获取单元，用于获取预设的间隔系数对应的扫描区间；
规整单元，用于采用所述预设的间隔系数对应的扫描区间对所述峰值特征序列进行规整，将所述预设的间隔系数对应的扫描区间内的最大峰值对应的峰值特征元素的数值设置为目标值，将所述预设的间隔系数对应的扫描区间内除
5 所述最大峰值对应的峰值特征元素之外的其他峰值特征元素的数值设置为初始值。

29、如权利要求 19 所述的装置，其特征在于，所述确定单元包括：

目标索引获取单元，用于从规整后的所述峰值特征序列中获取数值为目标
10 值的峰值特征元素对应的目标索引；

时间计算单元，用于根据所述目标索引及所述目标音频文件的采样率，计算段落变化时间。

30、如权利要求 19 所述的装置，其特征在于，所述获取单元包括：

15 类型获取单元，用于获取所述目标音频文件的类型，所述类型包括：双声道类型或单声道类型；

解码单元，用于若所述目标音频文件的类型为单声道类型，对所述目标音频文件从所述单声道输出的内容进行解码获得音频数据；或者，用于若所述目标音频文件的类型为双声道类型，从所述双声道中选取一个声道，对所述目标
20 音频文件从所选取声道输出的内容进行解码获得音频数据；或者将所述双声道处理为混合声道，对所述目标音频文件从所述混合声道输出的内容进行解码获得音频数据。

31、一种终端，其特征在于，包括如权利要求 17-30 任一项所述的音频处理
25 装置。

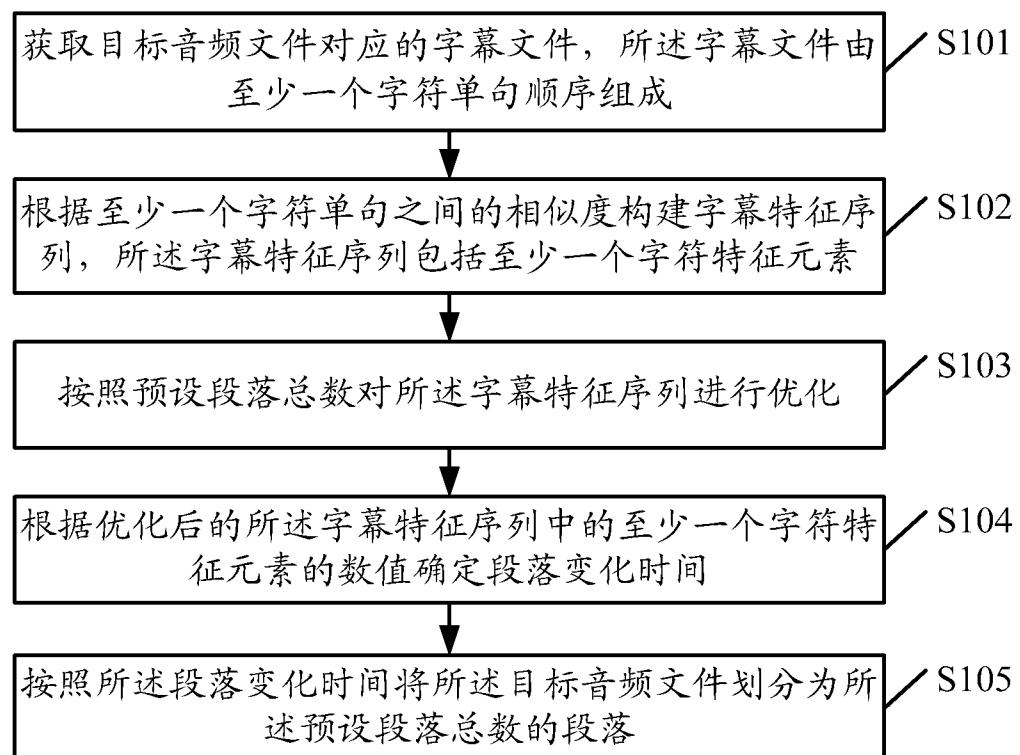


图 1

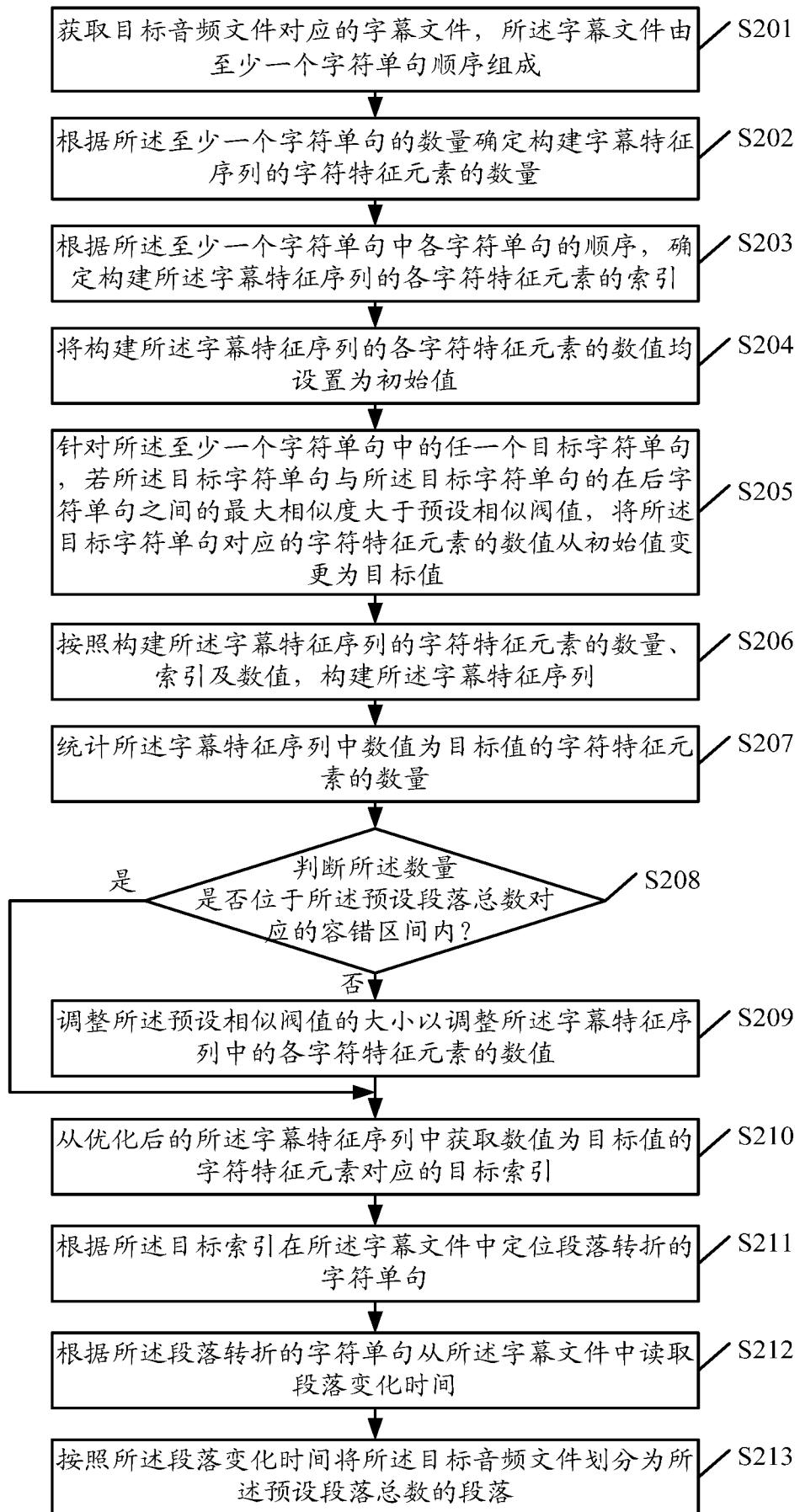


图 2

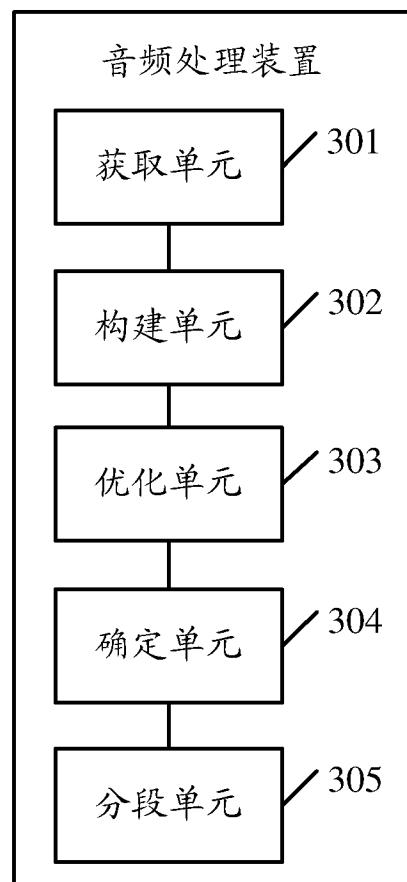


图 3

4/14

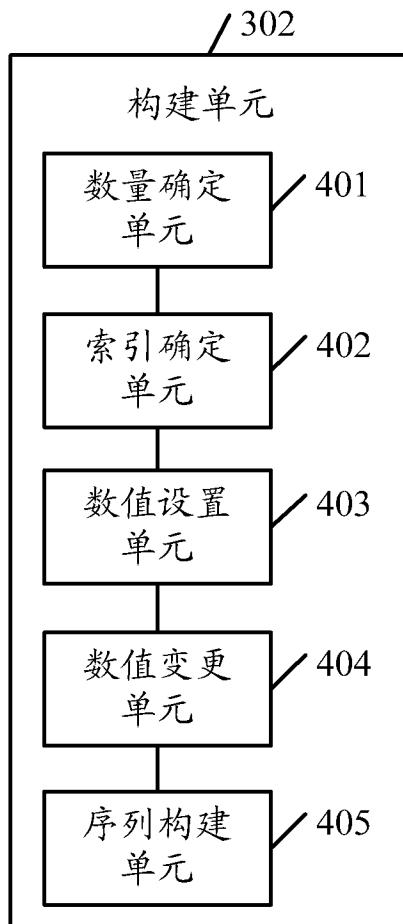


图 4

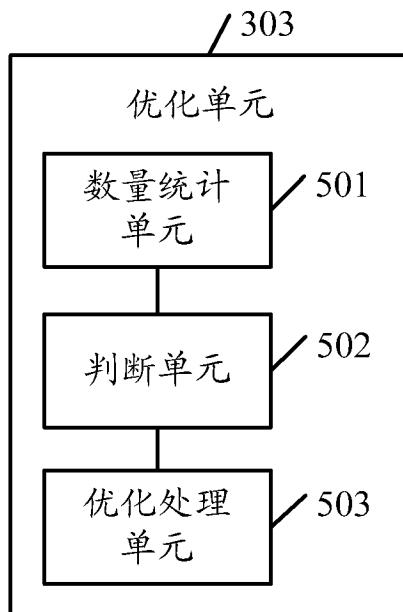


图 5

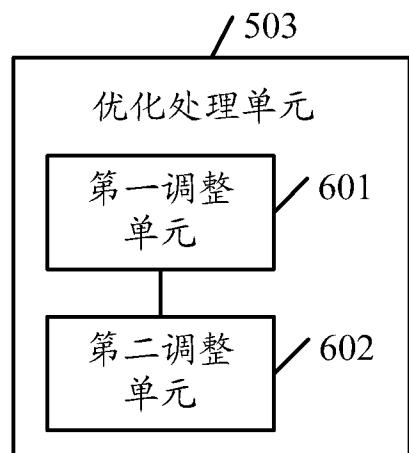


图 6

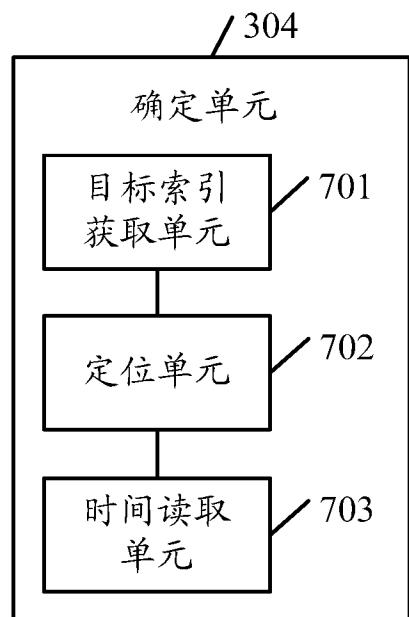


图 7

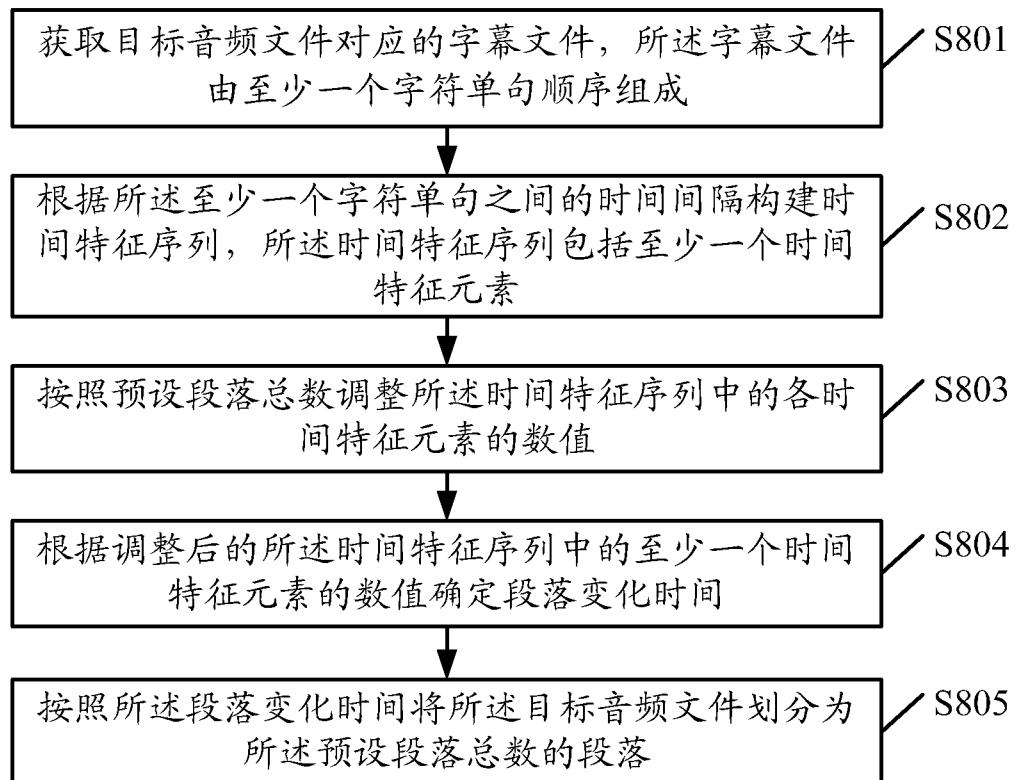


图 8

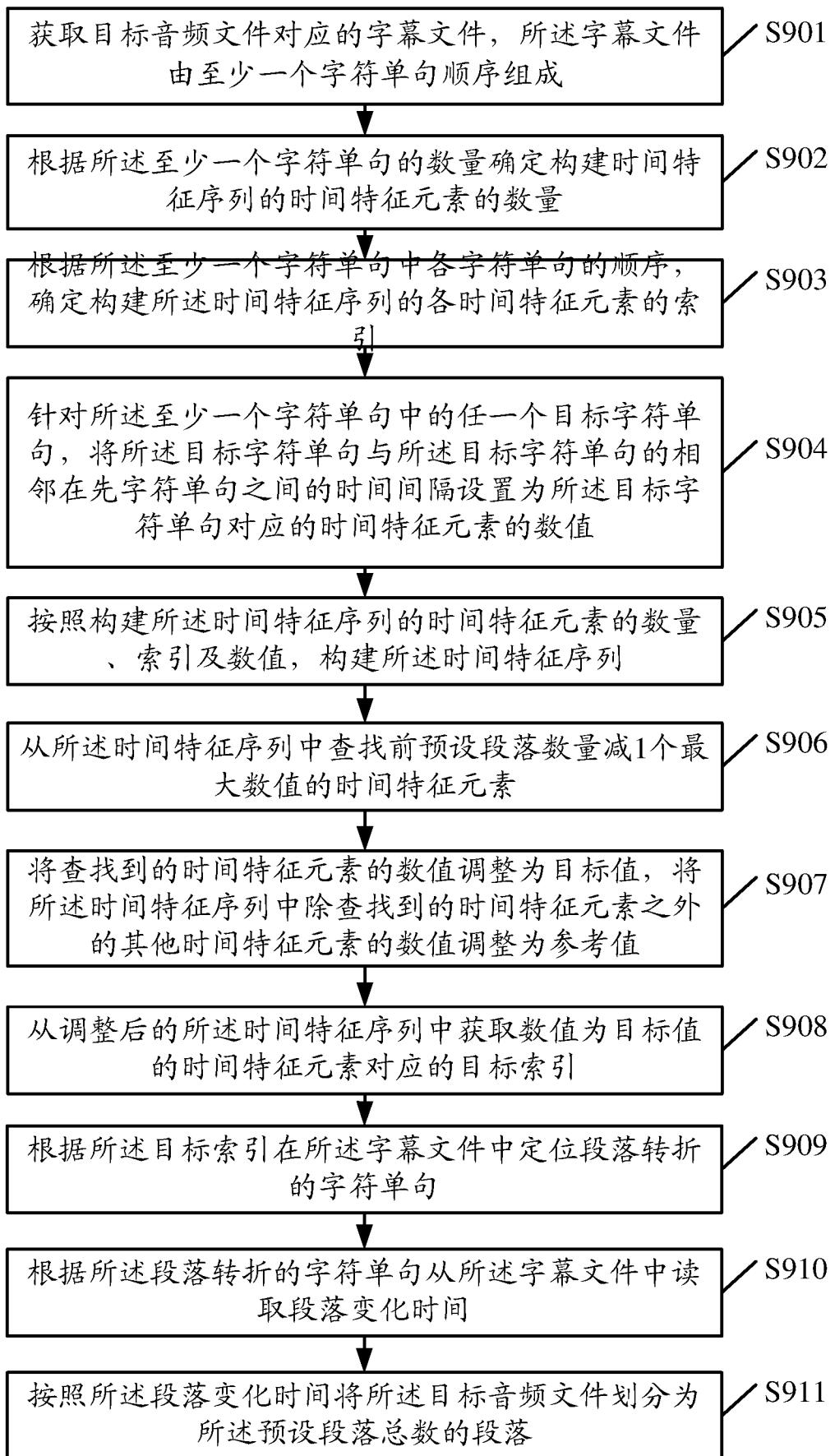


图 9

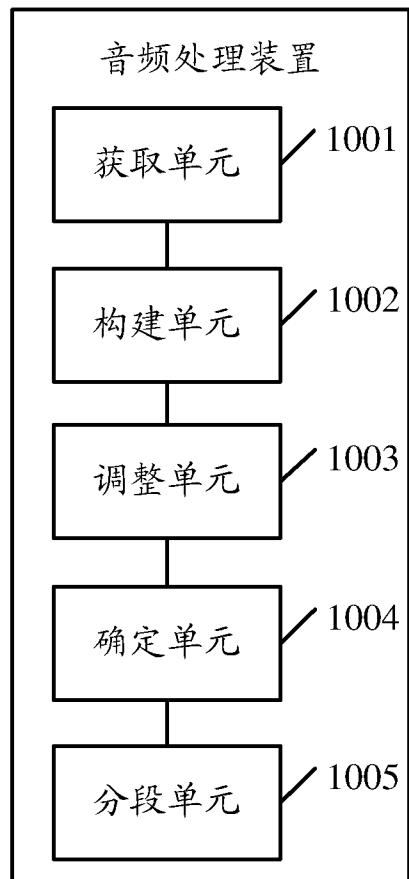


图 10

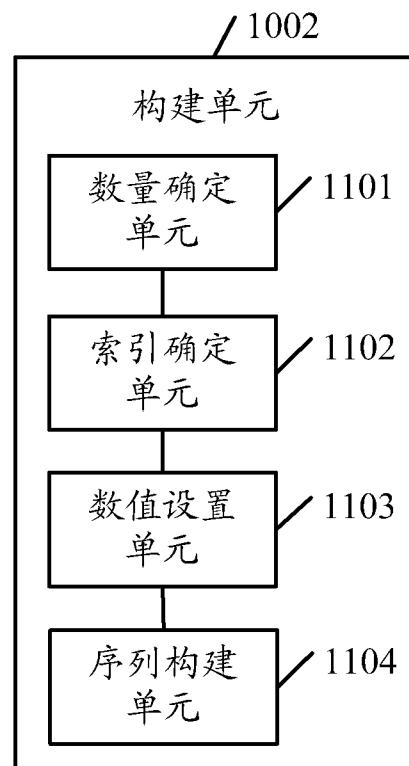


图 11

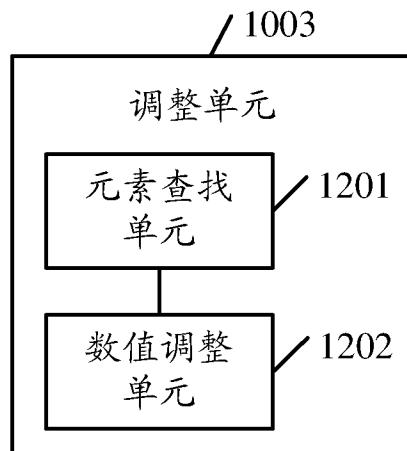


图 12

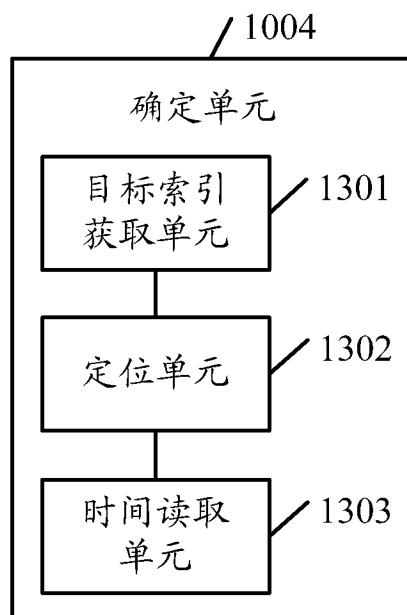


图 13

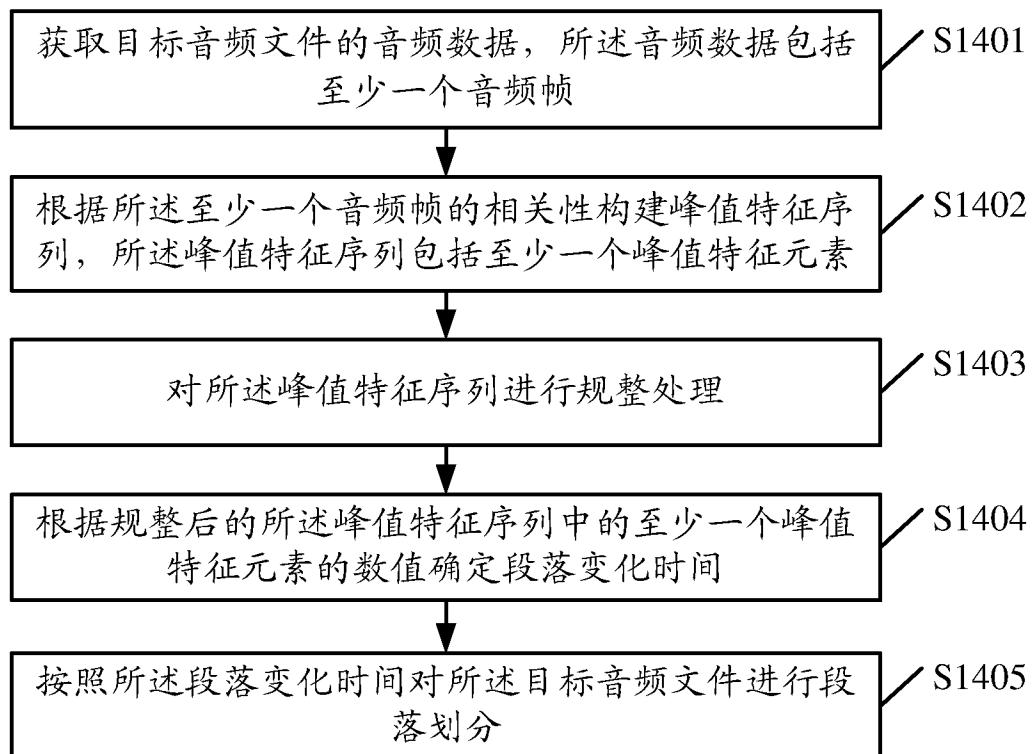


图 14

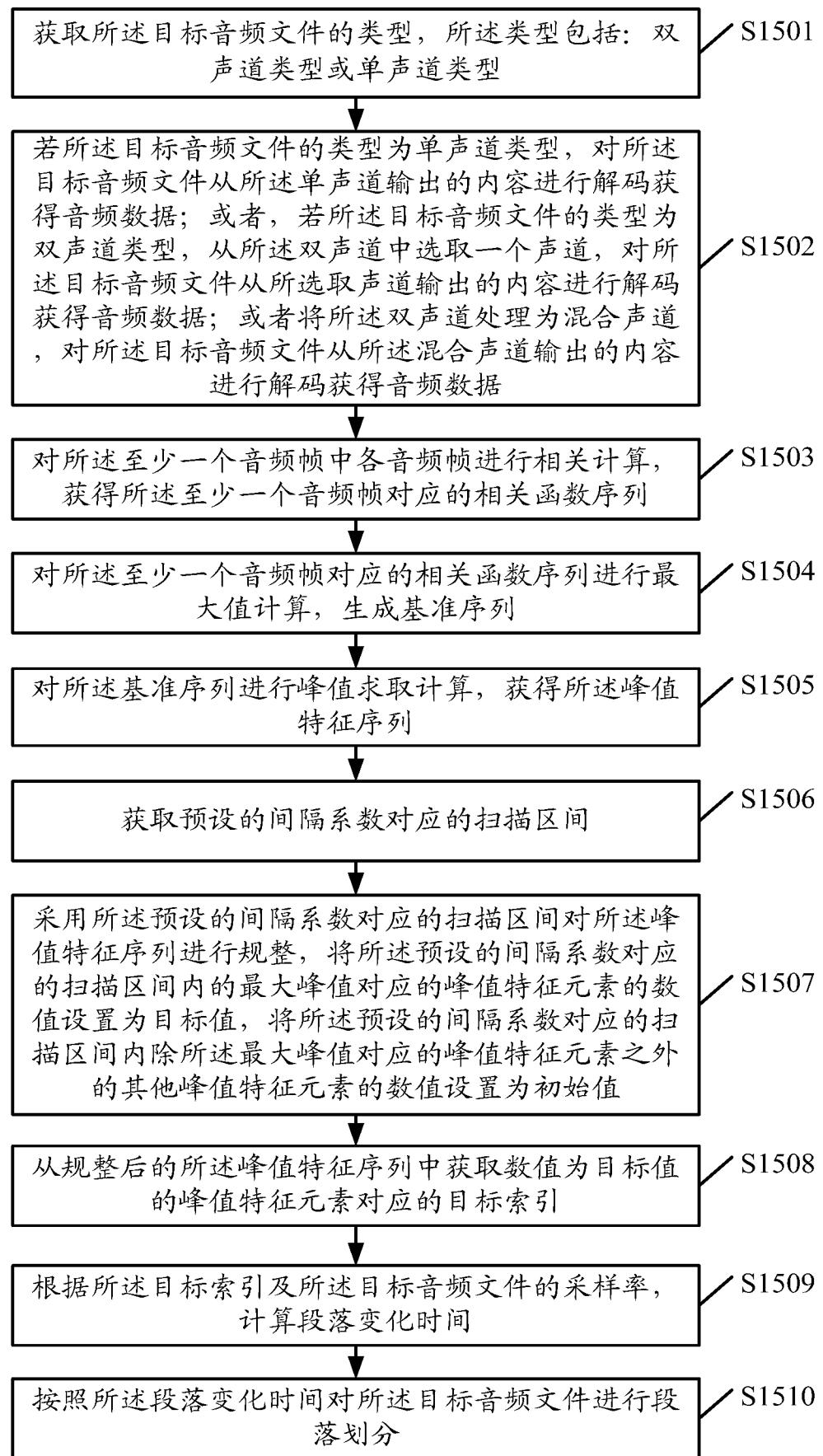


图 15

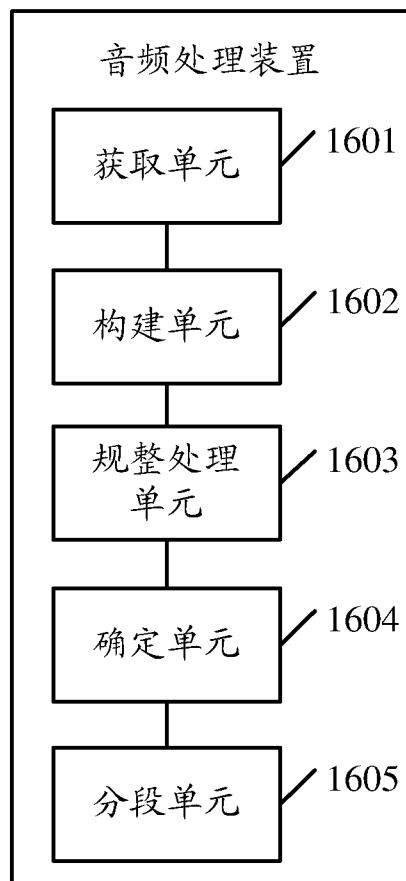


图 16

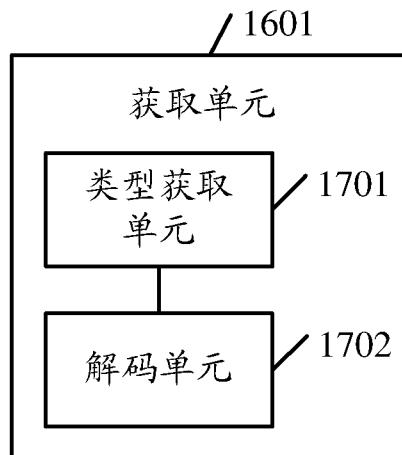


图 17

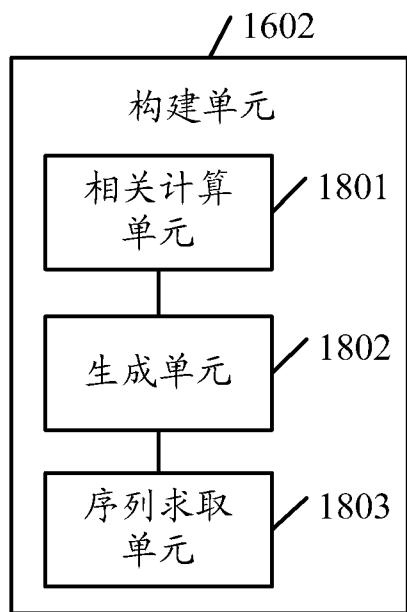


图 18

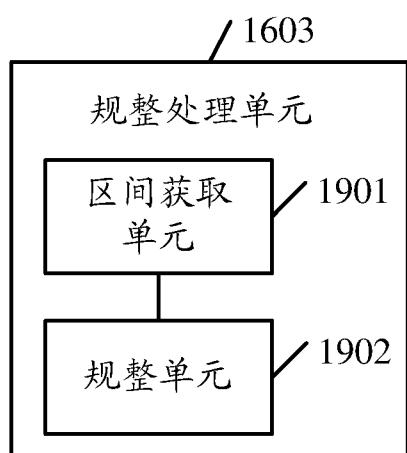


图 19

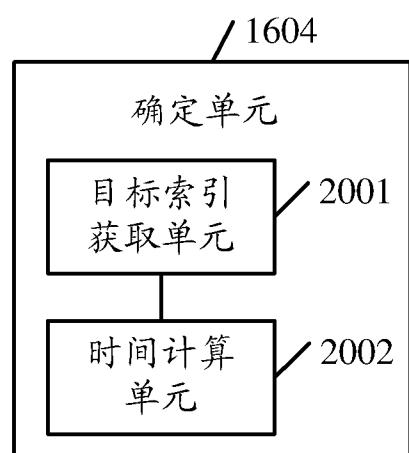


图 20

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/081999

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/04 (2013.01) i; G10L 15/06 (2013.01) i; G10L 25/48 (2013.01) i; G10L 25/03 (2013.01) i
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/-; G10L 25/-; G06F 17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: paragraph, word, similar, speech+, voice?, sound?, audio?, divid+, division, segment+, partition+, subtitle?, caption?, lyric?, text?, correlat+, time, peak?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104978961 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.), 14 October 2015 (14.10.2015), claims 1-11, and description, paragraphs [0003]-[0137]	1-6, 16-19, 22-24, 27, 31
PX	CN 105047203 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.), 11 November 2015 (11.11.2015), claims 1-11, and description, paragraphs [0003]-[0118]	1, 7-10, 16-18, 20, 25-27, 31
PX	CN 105047202 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.), 11 November 2015 (11.11.2015), claims 1-11, and description, paragraphs [0003]-[0116]	1, 11-15, 17-18, 21, 28-31
A	CN 1595397 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY), 16 March 2005 (16.03.2005), abstract, and description, page 2, paragraph 2 to page 3, paragraph 4	1-31
A	JP 20011175294 A (CASIO COMPUTER CO., LTD.), 29 June 2001 (29.06.2001), abstract, and description, paragraphs [0018]-[0041]	1-31
A	CN 1685345 A (MITSUBISHI ELECTRIC CORP.), 19 October 2005 (19.10.2005), the whole document	1-31

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 01 August 2016 (01.08.2016)	Date of mailing of the international search report 16 August 2016 (16.08.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer KANG, Dandan Telephone No.: (86-10) 62413593

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/081999**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1983276 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 20 June 2007 (20.06.2007), the whole document	1-31
A	JP 2007206183 A (YAMAHA CORP.), 16 August 2007 (16.08.2007), the whole document	1-31
A	US 6243676 B1 (OPENWAVE SYS. INC.), 05 June 2001 (05.06.2001), the whole document	1-31

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2016/081999

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104978961 A	14 October 2015	None	
CN 105047203 A	11 November 2015	None	
CN 105047202 A	11 November 2015	None	
CN 1595397 A	16 March 2005	CN 100559368 C	11 November 2009
JP 2001175294 A	29 June 2001	None	
CN 1685345 A	19 October 2005	EP 1461731 A1 WO 2004040479 A1 JP 2006505040 A US 7375731 B2 US 2004085323 A1	29 September 2004 13 May 2004 09 February 2006 20 May 2008 06 May 2004
CN 1983276 A	20 June 2007	US 7865501 B2 US 2007112855 A1 US 2011060743 A1 US 8103657 B2 CN 1983276 B	04 January 2011 17 May 2007 10 March 2011 24 January 2012 23 March 2011
JP 2007206183 A	16 August 2007	JP 4862413 B2 WO 2007088820 A1	25 January 2012 09 August 2007
US 6243676 B1	05 June 2001	US 2002055950 A1	09 May 2002

国际检索报告

国际申请号

PCT/CN2016/081999

A. 主题的分类 G10L 15/04(2013.01)i; G10L 15/06(2013.01)i; G10L 25/48(2013.01)i; G10L 25/03(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G10L 15/-; G10L 25/-; G06F 17/- 包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNPAT, WPI, EPODOC, CNKI: 语音, 声音, 话音, 音频, 分段, 分割, 划分, 段落, 字幕, 歌词, 文本, 文字, 相关, 相似, 时间, 峰值, speech+, voice?, sound?, audio?, divid+, division, segment+, partition+, subtitle?, caption?, lyric?, text?, correlat+, time, peak?.		
C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104978961 A (腾讯科技深圳有限公司) 2015年 10月 14日 (2015 - 10 - 14) 权利要求1-11, 说明书第[0003]-[0137]段	1-6, 16-19, 22-24, 27, 31
PX	CN 105047203 A (腾讯科技深圳有限公司) 2015年 11月 11日 (2015 - 11 - 11) 权利要求1-11, 说明书第[0003]-[0118]段	1, 7-10, 16-18, 20, 25-27, 31
PX	CN 105047202 A (腾讯科技深圳有限公司) 2015年 11月 11日 (2015 - 11 - 11) 权利要求1-11, 说明书第[0003]-[0116]段	1, 11-15, 17- 18, 21, 28-31
A	CN 1595397 A (华南理工大学) 2005年 3月 16日 (2005 - 03 - 16) 摘要, 说明书第2页第2段-第3页第4段	1-31
A	JP 2001175294 A (CASIO COMPUTER CO., LTD.) 2001年 6月 29日 (2001 - 06 - 29) 摘要, 说明书第[0018]-[0041]段	1-31
A	CN 1685345 A (三菱电机株式会社) 2005年 10月 19日 (2005 - 10 - 19) 全文	1-31
<input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。		<input checked="" type="checkbox"/> 见同族专利附件。
<p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件</p>		
国际检索实际完成的日期 2016年 8月 1日		国际检索报告邮寄日期 2016年 8月 16日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 康丹丹 电话号码 (86-10)62413593

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告

国际申请号

PCT/CN2016/081999

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 1983276 A (国际商业机器公司) 2007年 6月 20日 (2007 - 06 - 20) 全文	1-31
A	JP 2007206183 A (YAMAHA CORP.) 2007年 8月 16日 (2007 - 08 - 16) 全文	1-31
A	US 6243676 B1 (OPENWAVE SYS. INC.) 2001年 6月 5日 (2001 - 06 - 05) 全文	1-31

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/081999

检索报告引用的专利文件			公布日 (年/月/日)		同族专利		公布日 (年/月/日)			
CN	104978961	A	2015年 10月 14日					无		
CN	105047203	A	2015年 11月 11日					无		
CN	105047202	A	2015年 11月 11日					无		
CN	1595397	A	2005年 3月 16日	CN	100559368	C	2009年 11月 11日			
JP	2001175294	A	2001年 6月 29日		无					
CN	1685345	A	2005年 10月 19日	EP	1461731	A1	2004年 9月 29日			
				WO	2004040479	A1	2004年 5月 13日			
				JP	2006505040	A	2006年 2月 9日			
				US	7375731	B2	2008年 5月 20日			
				US	2004085323	A1	2004年 5月 6日			
CN	1983276	A	2007年 6月 20日	US	7865501	B2	2011年 1月 4日			
				US	2007112855	A1	2007年 5月 17日			
				US	2011060743	A1	2011年 3月 10日			
				US	8103657	B2	2012年 1月 24日			
				CN	1983276	B	2011年 3月 23日			
JP	2007206183	A	2007年 8月 16日	JP	4862413	B2	2012年 1月 25日			
				WO	2007088820	A1	2007年 8月 9日			
US	6243676	B1	2001年 6月 5日	US	2002055950	A1	2002年 5月 9日			