

(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
2022 年 1 月 20 日 (20.01.2022)



(10) 国际公布号
WO 2022/012164 A1

- (51) 国际专利分类号:
G10L 13/02 (2013.01) *G10H 1/00* (2006.01)
G10L 15/04 (2013.01)
- (21) 国际申请号: PCT/CN2021/095236
- (22) 国际申请日: 2021 年 5 月 21 日 (21.05.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010688502.3 2020 年 7 月 16 日 (16.07.2020) CN
- (71) 申请人: 百果园技术 (新加坡) 有限公司
(**BIGO TECHNOLOGY PTE. LTD.**) [SG/SG]; 新加坡巴西班让路枫树商业城 30 号楼 15 层 31A, Singapore 117440 (SG)。
- (72) 发明人: 及
- (71) 申请人 (仅对 US): 徐雯(**XU, Wen**) [CN/CN]; 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。
- (74) 代理人: 北京品源专利代理有限公司(**BEYOND ATTORNEYS AT LAW**); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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,

(54) **Title:** METHOD AND APPARATUS FOR CONVERTING VOICE INTO RAP MUSIC, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 将语音转换为说唱音乐的方法、装置、设备及存储介质

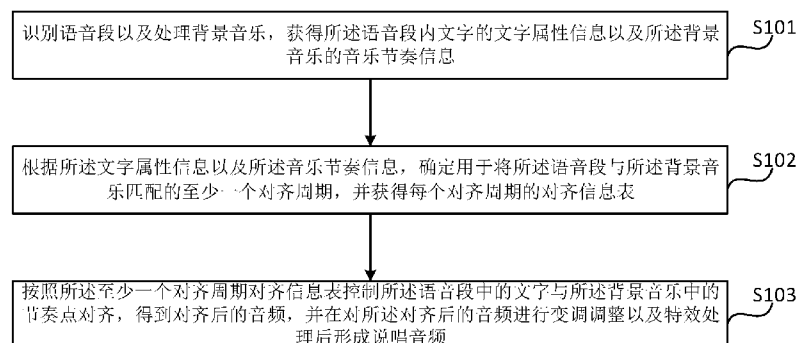


图 1

- S101 Identify a voice segment and process background music to obtain the character attribute information of characters in the voice segment and the music rhythm information of the background music
- S102 According to the character attribute information and the music rhythm information, determine at least one alignment period used for matching the voice segment with the background music, and obtain the alignment information table of each alignment period
- S103 According to the alignment information table of the at least one alignment period, control the characters in the voice segment to be aligned with rhythm points in the background music so as to obtain aligned audio, and form rap audio after performing tone change adjustment and special effect processing on the aligned audio

(57) **Abstract:** A method and apparatus for converting voice into rap music, a computer device, and a computer readable storage medium. The method comprises: identifying a voice segment and processing background music to obtain the character attribute information of characters in the voice segment and the music rhythm information of the background music (S101); according to the character attribute information and the music rhythm information, determining at least one alignment period used for aligning the voice segment with the background music, and obtaining the alignment information table of each alignment period (S102); and according to

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利(细则
4.17(ii))
- 关于申请人有权要求在先申请的优先权(细则
4.17(iii))
- 发明人资格(细则4.17(iv))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

the alignment information table of the at least one alignment period, controlling the characters in the voice segment to be aligned with rhythm points in the background music so as to obtain aligned audio, and forming rap audio after performing tone change adjustment and special effect processing on the aligned audio (S103).

(57) 摘要: 一种将语音转换为说唱音乐的方法、装置、计算机设备及计算机可读存储介质。该方法包括: 识别语音段以及处理背景音乐, 获得语音段内文字的文字属性信息以及背景音乐的节奏信息(S101); 根据文字属性信息以及节奏信息, 确定用于将语音段与背景音乐对齐的至少一个对齐周期, 并获得每个对齐周期的对齐信息表(S102); 按照至少一个对齐周期的对齐信息表控制语音段中的文字与背景音乐中的节奏点对齐, 得到对齐后的音频, 并在对对齐后的音频进行变调调整以及特效处理后形成说唱音频(S103)。

将语音转换为说唱音乐的方法、装置、设备及存储介质

本申请要求在2020年07月16日提交中国专利局、申请号为202010688502.3的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及音乐制作技术领域，例如涉及将语音转换为说唱音乐的方法、装置、设备及存储介质。

背景技术

随着K歌软件的普及，修音算法以及人声转音乐算法的研究逐渐受到广泛的关注，人们对于自动修音以及说话变唱歌的兴趣也越来越高涨。说唱文化逐渐进入大众的视野，说唱音乐的特点是创作者在背景音乐下快速有节奏地说出一连串押韵的文字，在说唱音乐制作过程中往往要经过复杂的过程，对于大多数非音频处理人员来说学习使用专业的音频处理软件，以及对音频处理软件进行复杂的手动操作均会耗费较长的时间。

针对上述问题，出现了一些适合非音频处理人员操作的语音转换软件，然而，不同语音转换软件在实现语音转换说唱的过程中，存在不同的缺陷，如，一种语音转换软件中的语音转说唱的技术方案，其限定了需要朗读特定歌词，由于歌词与背景音乐完全匹配，因此字与节奏点的对齐位置是固定的，该方案对于未知歌词内容、长度的情况，不能很好地处理，由此缩小了用户应用该软件时的创作空间，进而限制了该软件的应用前景。又如，另一种语音转换软件中的语音转说唱的技术方案，其在音频分割以及音频对齐上的算法设计都较为复杂，增加了转换难度，同时存在语音文字与音乐节奏点错位的问题，该种转换方式并不利于对用户自行上传的音乐进行有效地处理。

发明内容

本申请提供了将语音转换为说唱音乐的方法、装置、设备及存储介质，以解决语音转换为说唱音乐的过程中语音内容受限以及语音转换效果差的问题。

提供了一种将语音转换为说唱音乐的方法，包括：

识别语音段以及处理背景音乐，获得所述语音段中的文字的文字属性信息以及所述背景音乐的节奏信息；

根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与

所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；

按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

还提供一种将语音转换为说唱音乐的装置，包括：

信息确定模块，设置为识别语音段以及处理背景音乐，获得所述语音段中的文字的文字属性信息以及所述背景音乐的节奏信息；

对齐信息确定模块，设置为根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；

转换控制模块，设置为按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

还提供了一种计算机设备，包括：

一个或多个处理器；

存储装置，设置为存储一个或多个程序；

所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现上述的将语音转换为说唱音乐的方法。

还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述的将语音转换为说唱音乐的方法。

附图说明

图 1 为本申请实施例一提供的一种将语音转换为说唱音乐的方法的流程示意图；

图 2 为本申请实施例二提供的一种将语音转换为说唱音乐的方法的流程示意图；

图 3 为本申请实施例二提供的一种将语音转换为说唱音乐方法中确定对齐周期的实现流程图；

图 4 为本申请实施例二提供的一种将语音转换为说唱音乐方法中确定对齐周期中的对齐单元以及对齐单元信息的实现流程图；

图 5 为本申请实施例二提供的一种确定对齐周期中对齐单元以及对齐单元

信息的展开流程图；

图 6 为本申请实施例三提供的一种将语音转换为说唱音乐的装置的结构框图；

图 7 为本申请实施例四提供的一种计算机设备的硬件结构示意图。

具体实施方式

下面将结合附图对本申请实施例方式进行描述。

在本申请的描述中，术语“第一”、“第二”、“第三”等仅用于区别类似的对象，而不必用于描述特定的顺序或先后次序，也不能理解为指示或暗示相对重要性。可以根据情况理解上述术语在本申请中的含义。

实施例一

图 1 为本申请实施例一提供的一种将语音转换为说唱音乐的方法的流程示意图，该方法适用于将用户录制的语音段转换为说唱音乐的情况，该方法可以由将语音转换为说唱音乐的装置执行，其中，该装置可以由软件和/或硬件实现，并一般可集成在计算机设备上。

在该应用模式下，可以先向用户提供一个背景音乐的选择界面，由此获得用户所选定的背景音乐；之后，还可以向用户提供一个语音内容的选择界面，由此获得用户通过触发录制按钮实时录制的语音段，或者，获得用户通过触发上传按钮而上传的预先录制语音段；然后就可以执行本实施例提供的将语音转换为说唱音乐的方法，实现将上述所获得语音段转换为配合背景音乐的说唱片段。

如图 1 所示，本申请实施例一提供的一种将语音转换为说唱音乐的方法，包括如下操作：

S101、识别语音段以及处理背景音乐，获得所述语音段内文字的文字属性信息以及所述背景音乐的节奏信息。

在本实施例中，语音段可理解为在执行 S101 之前获得到的用户实时录制或预先录制的语音片段，背景音乐可理解为在执行 S101 之前接收到的用户从背景音乐集中选定的待使用音乐。

本实施例中，可以对语音段进行语音识别，由此可以获取到语音段所包括的文字的文字序号、文字发音时长（文字起止时间）以及文字中的首个元音的起始位置等相关的文字属性信息；也可以对背景音乐进行音乐节拍的检测处理，由此可以获取到背景音乐所包括的节奏点的节奏点序号、节奏点所在位置以及

划分所形成的每个节拍周期内所包括的节奏点数等相关的音乐节奏信息。

本实施例不对语音识别、文字检测以及节奏点检测的方式进行限定，只要可以获得到所需的文字属性信息以及音乐节奏信息即可。

S102、根据所述文字属性信息以及所述音乐节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表。

对于将用户语音转换为音乐片段的过程，除了 S101 中的语音识别以及节拍检测，最重要的环节在于将语音段中对应的文字与背景音乐中的节奏点进行对齐。所谓语音段和背景音乐的对齐，可认为是将语音分割为单个文字后，将每个文字与强节奏、规律性的重音点对齐，其中可能还伴有一些首尾字或中间字的重复用来加强节奏感。由此，本实施例在实现由语音到说唱音乐的转换时，需要先通过 S102 来确定用于将语音段和背景音乐对齐的对齐周期以及相应的对齐信息表。

所述对齐周期可理解为一个包括了能够与语音段内所有文字对齐的节奏点的最小重复单元，即从一个时间 t 开始，该背景音乐的节奏以一个能够对齐语音段内所有文字的固定周期 T 进行重复。所述对齐信息表可理解为包含了在一个对齐周期内进行节奏点与文字对齐时，所需的节奏点与待对齐文字的对应关系信息（如节奏点序号、文字序号）以及变速比等信息的信息声明表。

S102 的实现可表述为：

首先可以从文字属性信息中确定语音段所包括的文字总量，以及可以从音乐节奏信息中确定背景音乐所包括的节奏点的节奏点总量，以及对这些节奏点进行划分所形成的节拍周期的周期信息。其中，所述节拍周期可理解为根据节奏点找到的一个最小的节奏重复单元，即，从一个时间开始，该背景音乐的节奏以一个固定周期 Z 进行重复。

之后，根据文字总量以及一个节拍周期中所包括的节奏点的个数，可以确定节拍周期能否满足作为一个对齐周期的条件，如果节拍周期能满足作为一个对齐周期的条件，则直接将每个节拍周期看作对齐周期，如果节拍周期不能满足作为一个对齐周期的条件，则需要更新节拍周期的周期长度，获得可作为对齐周期的节拍周期。

然后，因为至少一个对齐周期的节奏是重复的，可以随机选取一个对齐周期，结合文字属性信息中的文字起止时间、文字的首个元音的起始位置，以及从音乐节奏信息中提取的一个对齐周期内节奏点的节奏点信息，来确定语音段中每个文字在该对齐周期内待对齐的节奏点，以及对该待对齐的节奏点进行对

齐时所要具备的变速比，由此形成包括节奏点序号与所关联文字的文字序号以及相应变速比的信息表，来作为该对齐周期的对齐信息表。

最终，可以将该对齐信息表看作每个完整的对齐周期的对齐信息表，而对于非完整的对齐周期，则可以从对齐信息表中提取部分对齐信息构成相对应的对齐信息表，由此通过 S102 获得了至少一个对齐周期，以及与每个对齐周期对应的对齐信息表。

S103、按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

在本实施例中，可以直接通过上述对背景音乐的节奏点划分形成的对齐周期以及包括语音段文字与节奏点对齐关系的对齐信息表，确定相匹配的文字和节奏点，并控制语音段中的文字与背景音乐中的节奏点对齐以及基于相应的变速比对相对齐后的音频进行变速，在此之后，还可以根据背景音乐的音高对变速后的音频进行变调调整以及加入混响等特效，形成转换后的说唱音频。

本申请实施例一提供的一种将语音转换为说唱音乐的方法，首先可以识别语音段以及对背景音乐进行处理，获得语音段内文字的文字属性信息以及背景音乐的节奏信息；之后可以根据文字属性信息以及节奏信息来确定用于语音段与背景音乐匹配的至少一个对齐周期，并获得每个对齐周期的对齐信息表；最终按照至少一个对齐周期的对齐信息表控制语音段中文字与背景音乐中的节奏点对齐，得到对齐后的音频，并在对对齐后的音频进行变调调整以及特效处理后形成说唱音频。上述技术方案，有效实现了将用户随意录制的语音内容片段转化为配合背景音乐的说唱片段，简化了手动进行音频剪辑制作的繁琐过程，为非专业音频处理人员提供了说唱音乐制作的可能；同时，上述技术方案无需限制待转换语音内容，保证了待转换语音内容的自由化录制，还简化了语音转换的实现过程，避免了语音文字与音乐节奏点错位的情况，提升了语音转换说唱音乐的应用范围。

作为本申请实施例一的一个可选实施例，在根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表之前，还包括：如果所述文字属性信息中的文字总量大于所述节奏信息中的节奏点总量，则结束将所述语音段转换为说唱音乐的处理，并输出重新获得语音段或背景音乐的提示。

本实施例所提供的将语音转换为说唱音乐的方法的实现中，上述 S102 以及 S103 的执行条件可默认为：通过 S101 所获得的文字属性信息中的文字总量小于或等于节奏信息中的节奏点总量，即，所获得语音段中的总字数小于或等

于背景音乐中的节奏点数。当上述条件不满足时，可认为不具备继续进行语音转换为说唱音乐的条件，此时，可以执行本可选实施例的操作，即，在确定文字总量大于节奏点总量时，可以结束后续将语音段转换为说唱音乐的步骤的执行，同时输出一个重新录制语音段的提示，以告知用户进行语音段的重新录制。或者，还存在其他可选的操作，如，本可选实施例也可以输出一个重新选择背景音乐的提示，告知用户重新选择背景音乐。

本可选实施例的操作，保证了待转换语音段与背景音乐的有效匹配，从而提升语音转换为说唱音乐的用户体验。

实施例二

图 2 为本申请实施例二提供的一种将语音转换为说唱音乐的方法的流程示意图，本实施例二以实施例一为基础进行说明，在本实施例中，识别语音段以及处理背景音乐，获得所述语音段内文字的文字属性信息以及所述背景音乐的节奏点信息，包括：对用户选定的语音段进行降噪处理以及端点检测处理，通过对处理后的语音段的语音识别获得所述语音段内每个文字的文字序号、起止时间、首个元音的起始位置以及文字总量，构成所述语音段的文字属性信息；对用户选定的背景音乐进行节奏点检测和节拍周期划分，确定所述背景音乐中包含的节奏点总量、节奏点序号、以及每个节拍周期的周期信息，构成所述背景音乐的节奏点信息；其中，所述周期信息包括：周期号、节拍周期内所包括的节奏点的节奏点个数以及每个节奏点的节奏点序号和节奏点起始时间。

本实施例中，根据所述文字属性信息以及所述节奏点信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表，包括：根据所述文字属性信息中的文字总量以及所述节奏点信息中每个节拍周期的周期信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期；选取一个完整的对齐周期作为待对齐节奏段，根据所述文字属性信息以及待对齐节奏段内待对齐节奏点的节奏点信息，确定至少一个对齐单元以及相应的对齐单元信息；汇总所述至少一个对齐单元信息形成所述待对齐节奏段的当前对齐信息表，并根据所述当前对齐信息表确定余下每个对齐周期的对齐信息表。

如图 2 所示，本实施例二提供的一种将语音转换为说唱音乐的方法，包括如下操作：

S201、对用户选定的语音段进行降噪处理以及端点检测处理，通过对处理后的语音段的语音识别获得所述语音段内每个文字的文字序号、起止时间、首

个元音的起始位置以及文字总量，构成所述语音段的文字属性信息。

在本实施例中，可以采用音频处理中的噪声处理策略对所录制的语音段进行降噪处理，以及采用端点检测策略对降噪后的语音段进行静音段的去除，之后可以采用语音识别策略对经过处理的语音段进行识别，从而获得构成语音段的每个文字的相关信息。

上述所获得的信息中可以包括整个语音段所包括文字的文字总量、每个文字的文字序号、该文字在语音段中对应的起止时间，该文字所对应发音的首个元音的起始位置，其中，文字的起止时间以及首个元音的起始位置均可看作相对时间点，即，可以按照整个语音的播放顺序，将首个文字的起始时间看作 0 秒。本实施例可以将上述信息记为语音段对应的文字属性信息。

示例性的，表 1 为一个文字属性信息的数据表，如表 1 所示，表 1 中的每列可以看做一个文字属性项，至少可以包括文字序号、文字的开始时间、文字中首个元音的开始时间以及文字的结束时间，表 1 的行数则可看作语音段内所包括的文字的总数。

表 1 语音段内文字的文字属性信息

文字序号	开始时间 (s)	元音开始时间 (s)	结束时间 (s)
1	0	0	0.1
2	0.1	0.15	0.25
...	...	...	...

S202、对用户选定的背景音乐进行节奏点检测和节拍周期划分，确定所述背景音乐中包含的节奏点总量、节奏点序号、以及每个节拍周期的周期信息，构成所述背景音乐的节奏点信息。

在本实施例中，可以采用音频处理中的节奏点检测策略从背景音乐中先检测出强节奏的重音点（即，节奏点），然后采用节拍划分策略来发现所检测出节奏点的发音规律，从而划分出具备最小节奏重复单元的节拍周期。对于一段背景音乐来说，检测出的节奏点本身具备一定的属性信息，如节奏点的序号，节奏点的总量、节奏点所处的位置（即节奏点出现的相对时间），同时，进行节拍检测后，也会对应每个节拍周期形成相应的周期信息，示例性的，所述周期信息可以包括：周期号、节拍周期内所包括的节奏点的节奏点个数以及每个节奏点的节奏点序号和节奏点起始时间。本实施例可以把这些信息汇总形成一

个音乐节奏信息。

示例性的，本实施例给出了一个数据表形式的音乐节奏信息，由此将音乐节奏信息以信息表的形式进行展示。如表 2 所示，表 2 为一个音乐节奏信息的数据表，可以看出表 2 为一个级联表，表 2 的第一列展示了以周期号为标识的节拍周期，第二列给出了节奏点序号，同时以级联的形式展现了哪些节奏点包含在周期号为 1 的节拍周期内，每个周期号下级联了该节拍周期所包括的节奏点序号以及节奏点位置（即起始时间），每个周期号下所级联的节奏点的行数可以作为该节拍周期的节奏点个数。

表 2 背景音乐对应的音乐节奏信息

周期号	节奏点序号	节奏点位置(s)
1	1	0.2
	2	0.6
	3	0.9
2	4	1.1
	5	1.5
	6	1.8
...	...	...
	...	...
	...	...

本实施例下述 S203 至 S205 给出了通过文字属性信息以及音乐节奏信息确定用于将语音段与背景音乐对齐的对齐周期及对齐信息表的实现过程。

S203、根据所述文字属性信息中的文字总量以及所述音乐节奏信息中每个节拍周期的周期信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期。

在本实施例中，可以先确定整段背景音乐中可以包括多少个用于对齐语音段内所有文字的对齐周期，当对齐周期的数量大于 1 时，确定出的最后一个对齐周期可能是一个非完整周期（即，并未包含全部文字）。S203 相当于先对背

景音乐进行一个大概的对齐周期划分。整个划分的过程需要借助语音段内所包括的文字的文字总量，以及背景音乐中一个完整的节拍周期内的节奏点个数，通过文字总量与该节拍周期中的节奏点个数的比对，来确定是否直接将节拍周期作为对齐周期，或者通过对节拍周期进行合并来获得对齐周期。

图 3 为本申请实施例二提供的一种将语音转换为说唱音乐方法中确定对齐周期的实现流程图，如图 3 所示，所述根据所述文字属性信息中的文字总量以及所述音乐节奏信息中每个节拍周期的周期信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，包括：

S2031、选定一个完整的节拍周期，并获取该完整的节拍周期所对应的周期信息内节奏点个数。

整段背景音乐中可以检测出至少一个节拍周期，当检测出一个节拍周期时认为该节拍周期为一个完整周期，当检测出大于 1 个节拍周期时，划分形成的最后一个周期可能是非完整周期，即并没有包含一个完整节拍周期内的全部节奏点。本实施例可以从完整的节拍周期中挑选一个节拍周期，获取该节拍周期所对应的周期信息中的节奏点个数。其中，不同完整的节拍周期中的节奏点个数的值是相同的。

S2032、判断该节奏点个数是否大于或等于所述文字总量，若该节奏点个数大于或等于所述文字总量，则执行 S2033；若该节奏点个数小于所述文字总量，则执行 S2034。

S2032 的判定目的主要在于确定当前检测获得的一个完整的节拍周期能否容纳语音段中的所有文字，如果该完整的节拍周期可以容纳语音段中的所有文字，则执行 S2033；如果该完整的节拍周期不可以容纳语音段中的所有文字，则需要执行 S2034。

S2033、将每个节拍周期看作一个对齐周期。

在节奏点个数大于或等于文字总量时，可以直接将该节拍周期看作一个对齐周期。当判定一个完整的节拍周期可以看作一个对齐周期时，检测出的其他完整的节拍周期均可看做完整的对齐周期，一个非完整的节拍周期可看做一个非完整的对齐周期。

S2034、判断所述背景音乐中所包括的节拍周期的个数是否大于 1，若背景音乐中所包括的节拍周期的个数大于 1，则执行 S2035；若背景音乐中所包括的节拍周期的个数不大于 1，则执行 S2036。

在节奏点个数小于文字总量时，相当于一个完整的节拍周期不能容纳语音段中的所有文字，此时，需要对节拍周期进行合并，而合并的前提条件是背景

音乐中所包括的节拍周期的个数至少为两个。通过 S2034 判定背景音乐中的节拍周期的个数是否大于 1，若背景音乐中的节拍周期的个数大于 1，则满足合并条件，可继续执行 S2035；若背景音乐中的节拍周期的个数不大于 1，相当于该段背景音乐与语音段不匹配，则需要执行 S2036。

S2035、按照周期号的排列顺序进行节拍周期的两两合并，形成至少一个新的节拍周期，返回执行 S2031。

在本实施例中，当节拍周期的个数大于 1 时，可以按照周期号顺序进行节拍周期的两两合并，由此形成新的节拍周期，新形成的节拍周期，其对应的周期信息也将发生相应的变化。以上述表 2 为例，假设周期号为 1 与周期号为 2 的两个节拍周期合并，形成的新节拍周期内所包含的节奏点个数之前两个节拍周期内所包含的节奏点个数之和。按照周期号顺序进行节拍周期的两两合并后，所形成的节拍周期个数为原有节拍周期个数的一半或一半加 1，之后可返回 S2031 依据新形成的节拍周期的周期信息进行对齐周期的确定，如此循环，直至找到适合的对齐周期，或者在查找失败时结束后续的语音到说唱音乐的转换操作。

S2036、结束将所述语音段转换为说唱音乐的处理，并输出重新获得语音段或背景音乐的提示。

在本实施例中，如果节拍周期的个数只有一个，且该节拍周期中的节奏点个数小于文字总量，可认为语音段与选定的背景音乐不匹配，需要通过本实施例一可选实施例的操作重新获得再次上传或再次录制语音段，或者，重选背景音乐。

S204、选取一个完整的对齐周期作为待对齐节奏段，根据所述文字属性信息以及待对齐节奏段内待对齐节奏点的节奏点信息，确定至少一个对齐单元以及相应的对齐单元信息。

上述进行了对齐周期划分后，可以以一个对齐周期为基准，来确定语音段所包括的每个文字相对该对齐周期内的节奏点的匹配情况。本实施例将一段时长内所包括的节奏点与语音段内文字的匹配看作一个对齐单元，每个对齐单元信息包括了对齐单元中所存在的节奏点的节奏点序号，以及与该节奏点所匹配的文字的文字序号，以及将所存在的节奏点与所匹配的文字进行对齐时所需的变速比。

每个对齐单元存在至少包括了节奏点序号、文字序号以及变速比的对齐单元信息。同时，由于多个对齐周期所包括的节奏点数量相同，所具备的音乐节奏相同，因此，可以仅对任一个完整的对齐周期进行有关对齐单元以及对齐单

元信息的确定。

S204 中确定对齐单元以及对齐单元信息的实现过程可描述为：首先将选定进行信息确定的对齐周期记为待对齐节奏段，该对齐周期所具备的节奏点信息可直接作为待对齐节奏段内待对齐节奏点的节奏点信息；之后，可以根据文字属性信息以及节奏点信息确定一个用于将文字和节奏点对齐的对齐匹配值；然后确定该对齐匹配值在预先设定的节奏点-文字对齐规则表中所属的对齐范围；最终基于该对齐范围对应的对齐规则来进行待对齐节奏段内对齐单元的确定，以及每个对齐单元所对应的对齐单元信息的确定，其中，节奏点-文字对齐规则表中的对齐范围以及相应的对齐规则均可通过历史经验预先设定。

图 4 为本申请实施例二提供的一种将语音转换为说唱音乐方法中确定对齐周期中的对齐单元以及对齐单元信息的实现流程图，如图 4 所示，根据所述文字属性信息以及待对齐节奏段内待对齐节奏点的节奏点信息，确定至少一个对齐单元以及相应的对齐单元信息可包括：

S2041、选取一个完整的对齐周期作为待对齐节奏段，基于待对齐节奏段内多个待对齐节奏点的节奏点信息，形成与所述多个待对齐节奏点一一对应的多个待对齐节奏块，并将所述多个待对齐节奏点的数量记为初始的剩余点个数。

在本实施例中，可以从上述确定的对齐周期中，选定一个完整的对齐周期作为进行对齐信息表确定的待对齐节奏段，所述待对齐节奏段中的待对齐节奏点为该完整的对齐周期中包括的所有节奏点，该完整的对齐周期所包括的节奏点的节奏点信息为待对齐节奏点的节奏点信息。

本实施例可以将相邻两个待对齐节奏点之间形成的间隔记为一个待对齐节奏块，由此，可首先按照待对齐节奏段所包括的待对齐节奏点的数量形成相同数量个待对齐节奏块，即可认为所形成的待对齐节奏块分别与待对齐节奏点一一对应，可以为每个待对齐节奏块设置相应的块序号，还可以将待对齐节奏点的数量记为初始的剩余点个数。

S2042、确定所述剩余点个数与所述文字属性信息中的文字总量的比值，将所述比值记为对齐匹配值。

本实施例为了实现待对齐节奏段内每个待对齐节奏点与语音段内文字的匹配，首先确定待对齐节奏段内未与文字匹配的待对齐节奏点的数量与文字总量的比值，并将该比值记为对齐匹配值。

当在待对齐节奏段内不存在已匹配的节奏点时，需要匹配的节奏点数为全部待对齐节奏点，因此，最初时将剩余点个数初始为待对齐节奏段所包括的待对齐节奏点的数量。

S2043、查找预设的节奏点-文字对齐规则表，确定所述对齐匹配值归属的长度比值范围。

本实施例预先设定了一个节奏点-文字对齐规则表，该规则表为一个二元关联表，两个关联的对象分别为长度比值范围以及对齐规则。所述长度比值范围可通过一个对齐周期内未匹配的节奏点个数与整个语音段所包括的文字总量的比值来设定。本实施例基于历史经验形成 6 个不同区间的长度比值范围，分别为： $(0,0.2]$ 、 $(0.2,0.8]$ 、 $(0.8,1]$ 、 $(1,1.1]$ 、 $(1.1,1.3]$ 和 $(1.3,\infty)$ 。

在本实施例中，可以确定上述所获得的对齐匹配值在节奏点-文字对齐规则表中所处的长度比值范围。

S2044、按照所述长度比值范围对应的对齐规则确定存在相匹配文字的待对齐节奏块，并记为候选对齐单元。

确定对齐匹配值归属的长度比值范围后，可以获得该长度比值范围所关联的对齐规则，通过该对齐规则为该待对齐节奏段进行候选对齐单元的划分。

本实施例中，可以将文字与节奏点的匹配，看作文字与一个待对齐节奏块的匹配，基于长度比值范围对应的对齐规则，可以为每个待对齐节奏块确定出与其相匹配的文字（文字个数不确定，但文字个数至少为 1 个），且匹配后的待对齐节奏块就可作为一个候选对齐单元。

对应于不同的长度比值范围，本实施例设定了相应的对齐规则，示例性的，表 3 给出了预先设定的节奏点-文字对齐规则表。通过表 3 中多个长度比值范围对应的对齐规则为剩余节奏点（剩余的待对齐节奏块）进行文字匹配。

表 3 节奏点-文字对齐规则表

长度比值范围	对齐规则
$(0,0.2]$	先进行重复最后一个文字操作，再对剩余节奏点添加静音段
$(0.2,0.8]$	当 L 小于或等于 0.5，则随机选定 $L \times$ 文字总量的待重复文字，调整已匹配的节奏点-文字的位置，在所选定文字后进行重复添加； 当 L 大于 0.5，随机选定 50% 文字总量的待重复文字，调整已匹配的节奏点-文字的位置，在所选定文字后进行重复添加，并对余下 $(L-0.5) \times$ 文字总量的剩余节奏点进行静音段添加，其中，L 为

	对齐匹配值
(0.8,1]	随机从已匹配节奏点-文字中选择相邻两个文字合并至与一个节奏点匹配，直到余下的剩余节奏点与文字总量相同
(1,1.1]	从首个文字开始选择 10%文字总量的文字进行从首个剩余节奏点开始的匹配，并将余下的剩余节奏点分别按照文字顺序与文字匹配
(1.1,1.3]	从首个剩余节奏点开始与所有文字进行匹配，之后随机选定 10%文字总量的文字在所选定文字的位置后进行重复添加，然后对余下小于或等于 20%文字总量的剩余节奏点，从最后一个文字开始最多选择 20%文字总量的文字进行重复匹配
(1.3,∞)	从首个文字开始选择 10%文字总量的文字进行从首个剩余节奏点开始的匹配，之后将 100%文字总量的剩余节奏点分别按照文字顺序与文字匹配，然后对之后 20%文字总量的剩余节奏点，从最后一个文字开始选择 20%文字总量的文字进行重复匹配

S2045、统计剩余待对齐节奏块的块数，作为新的剩余点个数。

采用上述 S2044 进行一次对齐匹配后，还可能存在未匹配的待对齐节奏块，统计出待对齐节奏段中剩余的待对齐节奏块的块数，并将该块数作为新的剩余点个数。

S2046、确定所述剩余点个数是否为 0，若所述剩余点个数为 0，则执行 S2047；若所述剩余点个数不为 0，返回执行 S2042。

本实施例中，可以判定剩余点个数是否为 0，如果剩余点个数为 0，则可认为待对齐节奏段中剩余的待对齐节奏块为 0，即所有的待对齐节奏块均已完成了匹配，此时可以执行 S2047 的操作；如果剩余点个数不为 0，则可认为待对齐节奏段中还存在未匹配的待对齐节奏块，此时可以返回重新执行 S2042 的对齐匹配值确定操作。

基于 S2046 的操作，当一个待对齐节奏段内所有的待对齐节奏块均完成匹

配后，其所形成的候选对齐单元的个数实际与所包括待对齐节奏点的个数相同。即，可认为一个待对齐节奏点（待对齐节奏块）对应存在一个候选对齐单元，多个候选对齐单元的单元序号可以按照其对齐顺序从 0 开始依次递增标记。

为便于理解候选对齐单元的确定过程，本实施例给出了一个示例性描述。示例性的，假设一个待对齐节奏段中的待对齐节奏点数量为 8，且当前确定的剩余点个数为 8；用户所获得的语音段包括的文字总量为 5，如“淡黄色长裙”，则将“淡黄色长裙”与 8 个剩余节奏点进行匹配确定每个候选对齐单元的过程描述为：

1) 对齐匹配值为： $8/5=1.6$ ，落入 $(1.3, \infty)$ 的长度比值范围，查找上述表 3，可以获得其对应的对齐规则。

2) 按照长度比值范围 $(1.3, \infty)$ 所关联对齐规则进行文字与节奏点的匹配。

该对齐规则为：“从首个文字开始选择 10%文字总量的文字进行从首个剩余节奏点开始的匹配，之后将 100%文字总量的剩余节奏点分别按照文字顺序与文字匹配，然后对之后 20%文字总量的剩余节奏点，从最后一个文字开始选择 20%文字总量的文字进行重复匹配”。基于该对齐规则，首先需要从“淡黄色长裙”的首个字开始，选择 10%文字总量，即 0.5 个字进行重复。当遇到待重复的字长小于 1 时，就进行向下取整的操作，因此，当前需要重复的字数为 0。之后，可以直接从首个剩余节奏点开始，选择 100%文字总量的节奏点，分别与 5 个文字顺序匹配，此时，节奏点 0-4 形成的待对齐节奏块 0-4 分别对应了“淡”“黄”“色”“长”“裙”5 个字；然后，需要从“淡黄色长裙”的最后一个字开始，选择 20%文字总量的文字，即最后一个字“裙”，此时，节奏点 5 形成的待对齐节奏块对应了“裙”这个字。至此，完成了按照长度比值范围 $(1.3, \infty)$ 所关联对齐规则进行文字与节奏点匹配的操作，当前确定出的候选对齐单元的单元序号分别为 0-5，且 6 个候选匹配单元所对应的文字分别为：“淡”“黄”“色”“长”“裙”“裙”。

3) 上述操作后，8 个待对齐节奏块中还余下 2 个未匹配的待对齐节奏块，认为剩余点个数大于 0，由此可重新进行对齐匹配值确定，新的对齐匹配值为 $2/5=0.4$ ，落入 $(0.2, 0.8]$ 的长度比值范围，查找上述表 3，可以获得其对应的对齐规则。

4) 按照长度比值范围 $(0.2, 0.8]$ 所关联对齐规则进行文字与节奏点的匹配。

该对齐规则为：“当 L 小于或等于 0.5，则随机选定 L *文字总量的待重复文字，调整已匹配的节奏点-文字的位置，在所选定文字后进行重复添加；当 L 大于 0.5，随机选定 50%文字总量的待重复文字，调整已匹配的节奏点-文字的位

置，在所选定文字后进行重复添加，并对余下 $(L-0.5)*$ 文字总量的剩余节奏点进行静音段添加，其中， L 为对齐匹配值。”

分析该对齐匹配值 0.4，可知按照该对齐规则，该对齐匹配值 0.4 小于 0.5，因此可以直接进行随机选定 40%文字总量（即 2 个字）的操作，假设字号 0-4 中随机选定的字号为 1 和 3，分别对应的字为“黄”和“长”，则之后需要对上述已匹配形成的“淡黄色长裙裙”进行调整，使得待重复的字能够位于所选定文字的位置之后，按照该对齐规则，余下两个待对齐节奏块匹配的文字分别为“黄”和“长”，由此形成了分别匹配“黄”和“长”两个文字的新的候选对齐单元，由于对上述已匹配形成的“淡黄色长裙裙”进行了调整，基于该操作后，8 个候选匹配单元所对应的文字分别为：“淡”“黄”“黄”“色”“长”“长”“裙”“裙”。

5) 上述操作后，余下未匹配的待对齐节奏块为 0，即剩余点个数为 0，符合结束候选对齐单元的匹配条件，由此可以结束上述操作。

步骤 5) 之后，就可以形成 8 个单元序号依次为 0-7 的候选对齐单元。由此完成语音段内文字到待对齐节奏段的对齐匹配。

S2047、根据每个候选对齐单元的单元时长以及每个候选对齐单元所匹配的文字的匹配文字属性信息，确定至少一个对齐单元并获得相应的变速比。

根据上述描述可知，从待对齐节奏段内确定出的候选对齐单元的个数与待对齐节奏段内所包括的待对齐节奏块的块数相同，而一个待对齐节奏块为相应节奏点到相邻下一节奏点或者节奏结束点（该种情况主要针对最后一个节奏点）所形成的间隔块，即，一个待对齐节奏块的持续时长为两节奏点（或节奏结束点）的间隔时长。本实施例中，由于一个候选对齐单元对应一个待对齐节奏块，所以可将待对齐节奏块的持续时长作为相应候选对齐单元的单元时长。

确定出候选对齐单元相匹配的文字后，需要做的就是将所匹配文字与候选对齐单元进行文字发音与单元时长的对齐。一般情况下，该种对齐可以直接是在播放该候选对齐单元音频信号的同时混入所匹配文字的发音。考虑到有些文字的发音时间较短，但与其匹配的候选对齐单元的单元时长又较长，又或者，有些文字的发音时间较长，但与其匹配的候选对齐单元的单元时长又较短，为了实现文字与待对齐单元的对齐，则需要调整文字的发音速率，比如将文字的发音时间拉伸（减小发音速度）或者压缩（加快发音速度）以使其等于单元时长。

本实施例将文字需要拉伸或者压缩的比例值记为变速比，可以根据候选对齐单元的单元时长，以及与候选对齐单元所匹配的文字的匹配文字属性信息（如

所匹配的文字的文字起止时间，文字中首个元音的起始位置等），来确定所匹配的文字与相应候选对齐单元进行对齐时所需的变速比。

通过拉伸或压缩文字发音，将文字与待对齐单元对齐的实现中，文字发音能够拉伸或者压缩的程度是有限定的，如果仅考虑对齐而无限的拉伸或压缩文字发音，在实际执行对齐操作后，所形成的音频就存在失真的风险，因此，本实施例需要为文字发音的压缩或者拉伸设定一个合适的范围，即，需要保证文字所对应的变速比为处于一个正常的比值范围，该比值范围就可看作适合拉伸或压缩的适合条件。

由此，可以将上述计算所得的变速比与所设定的适合条件比对，来确定相应的候选对齐单元是否适合作为对齐单元，如果候选对齐单元适合作为对齐单元，则可直接将候选对齐单元确定为对齐单元，并将其对应的变速比确定为该对齐单元的变速比；如果候选对齐单元不适合作为对齐单元，则需要对候选对齐单元进行静音填补或者两个或多个候选对齐单元的合并处理，从而获得满足上述适合条件的对齐单元，且将进行适合条件判定的变速比作为该对齐单元的变速比。

上述所确定的待对齐节奏点数量个的候选对齐单元最终能形成至少一个对齐单元，每个对齐单元中至少可以包括一个节奏点，同时至少可以包括一个相匹配的文字，每个对齐单元的变速比就可看作将所包括的文字与所包括的节奏点进行对齐时，对文字进行拉伸或者伸缩所需的比例值。

S2048、将每个对齐单元的单元序号、所包括节奏点中的起始节奏点序号、所匹配的文字的文字序号以及变速比确定为相应的对齐单元信息。

在进行上述对齐单元以及相应变速比的确定操作时，也相应获得了每个对齐单元的单元序号，以及该对齐单元中所包括的每个节奏点的节奏点序号，同时，也可以获得该对齐单元中所匹配的每个文字的文字序号。可以针对每个对齐单元进行上述信息的汇总，由此对应每个对齐单元形成相应的对齐单元信息。

S205、汇总所述至少一个对齐单元对应的对齐单元信息形成所述待对齐节奏段的当前对齐信息表，并根据所述当前对齐信息表确定余下每个对齐周期的对齐信息表。

本实施例通过上述 S204 可以确定出待对齐节奏段中包括的至少一个对齐单元，以及相应的对齐单元信息，可以对上述确定的对齐单元信息按照对齐单元的单元序号顺序排列汇总，由此形成一个当前对齐信息表。之后还可以根据该当前对齐信息表确定上述 S203 确定出的其余每个对齐周期的对齐信息表。

对于剩余的对齐周期，如果是一个完整的对齐周期，则可复制上述当前对

齐信息表直接作为相应的对齐信息表；如果是一个非完整的对齐周期，则可从当前对齐信息表中取出与该对齐周期所包括的节奏点个数相同行的对齐单元信息形成相应的对齐信息表。

表 4 一个对齐周期中基于对齐单元的信息形成的对齐信息表

单元序号	起始节奏点序号	文字序号	变速比
1	1	2	1.0
2	2	3,4	1.2
3	3	5	0.9
...	...	...	...

示例性的，表 4 为一个对齐周期中基于对齐单元的信息形成的对齐信息表，如表 4 所示，该对齐信息表中的每列相当于对齐单元的属性信息，可以包括：对齐单元的单元序号、该对齐单元内起始节奏点的节奏点序号、所匹配的文字的文字序号以及进行对齐所需的变速比，该对齐信息表的行数代表了该对齐周期所具备对齐单元的单元个数。

所述根据所述当前对齐信息表确定余下每个对齐周期的对齐信息表，包括：针对余下的每个对齐周期，如果所述对齐周期为一个完整周期，则将所述当前对齐信息表作为所述对齐周期的对齐信息表；如果所述对齐周期为一个非完整周期且所述当前对齐信息表中的一行对应一个对齐单元，则确定所述对齐周期所包括的节奏点的节奏点个数；从所述当前对齐信息表中逆序选定所述节奏点个数行的对齐单元信息，形成所述对齐周期的对齐信息表。

本实施例中的上述描述给出了背景音乐内其余对齐周期的对齐信息表的确定过程，对于非完整的对齐周期，假设非完整的对齐周期中包括 2 个节奏点，则可直接从当前对齐信息表中由下至上选择两行对齐单元信息构成相应的对齐信息表。

S206、按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

在本实施例中，每个对齐周期所形成的对齐信息表中至少包括了一个对齐单元以及相应的对齐单元信息，而每个对齐单元信息中包括了实际用于文字和

节奏点对齐的节奏点序号、所匹配的文字序号以及对齐所需的变速比等。本实施例通过上述步骤获得每个对齐周期的对齐信息表后，就可以依据每个对齐信息表中包括的对齐单元信息，控制相应的节奏点与所匹配的文字按照对应的变速比进行对齐，由此来实现语音段中的文字与背景音乐中的节奏点的对齐匹配。

在控制语音段内的文字与所匹配的节奏点对齐时，对于每个对齐周期内的匹配，实际相当于先根据该对齐周期内每个对齐单元内所包括的文字的发音占据时长（该对齐单元的首个元音起始点到下一个单元的首个元音起始点的间隔时长）来获取该对齐单元在语音段内实际对应的音频数据，然后根据该对齐单元的变速比来对实际对应的音频数据进行变速调整，最终，可以对变速调整后的音频数据再进行变调调整以及特效处理等操作，从而形成转换后的说唱音乐。

本申请实施例二提供的一种将语音转换为说唱音乐方法，给出了文字属性信息以及音乐节奏信息的确定操作，还给出了确定用于将语音段与背景音乐对齐的对齐周期以及相关对齐信息表的操作。通过本实施例提供的方法，能够使用户在选定背景音乐以及录制一段随意内容的说话语音后，通过获得到的节奏点位置，单字的起止时间以及元音开始时间，确定字与节奏点匹配对齐以及变速的对齐策略，由此通过对齐策略短时间内就能获得字与节奏点对齐后形成的说唱音乐。整个技术方案的实现，简化了手动进行音频剪辑制作的繁琐过程，为非专业音频处理人员提供了说唱音乐制作的可能；同时，上述技术方案无需限制待转换语音内容，保证了待转换语音内容的自由化录制，还简化了语音转换的实现过程，避免了语音文字与音乐节奏点错位的情况，提升了语音转换说唱音乐的应用范围。

作为本申请实施例二的一个可选实施例，本可选实施例在执行上述 S202 中确定所述背景音乐中包含的节奏点总量、节奏点序号、以及每个节拍周期的周期信息，构成所述背景音乐的节奏点信息之前，还包括：获取检测出的多个初始节奏点，并确定相邻两个初始节奏点之间形成的间隔时长；根据所述语音段所包括的文字的平均字长时间结合所述间隔时长，确定所述多个初始节奏点中的待删除节奏点并删除待删除节奏点，获得所述背景音乐中有效的节奏点。

在本可选实施例中，给出了对从背景音乐中所检测出的节奏点进行处理的操作，通过该操作可以从所检测出的节奏点（本可选实施例记为初始节奏点）中去除掉相邻两个节奏点的间隔时长小于平均字长时间一半的间隔较密集的节奏点。

文字的平均字长时间为全部文字所占用的时长与文字总量的比值，一般来说，如果相邻两个节奏点间的间隔时长小于平均字长时间的一半，不利于文字与节奏点的对齐，因此，需要对相邻两个节奏点中的任一个进行删除，从而使

未删除的那个节奏点与所删除掉节奏点的前一个节奏点或者后一个节奏点构成新的间隔时长，且可以再次通过本可选实施例的方式对新形成的间隔时长再次进行判定，由此循环更新去除无效的节奏点，保留下有效的节奏点。

作为本申请实施例二的另一个可选实施例，对上述 S2047 的执行进行了说明，图 5 为本申请实施例二提供的一种确定对齐周期中对齐单元以及对齐单元信息的展开流程图，如图 5 所示，对根据每个候选对齐单元的单元时长以及每个候选对齐单元所匹配的文字的匹配文字属性信息，确定至少一个对齐单元并获得相应的变速比，进行说明：

本可选实施例为上述 S2047 的执行过程。通过上述 S2046 的操作，在待对齐节奏段内可以获得一定数量的候选对齐单元，本可选实施例下述操作可以确定候选对齐单元中的对齐单元以及对齐单元对应的变速比。

S1、按照单元序号的顺序选取一个未选中的候选对齐单元作为当前处理单元。

在本实施例中，待对齐节奏段内的候选对齐单元存在相应的单元序号，可以按照单元序号的顺序先选取一个之前未被选中的候选对齐单元，作为当前处理单元，未选中可理解为未被选取作为过当前处理单元。

示例性的，选取首个候选处理单元作为当前处理单元。

S2、根据所述当前处理单元的单元时长，结合所述当前处理单元以及相邻下一候选对齐单元分别所匹配的文字的起止时间和首个元音的起始位置，确定所述当前处理单元的当前变速比。

根据本实施例的上述描述，可知文字与候选对齐单元的对齐主要表现在文字实际的发音时长与候选对齐单元的单元时长的对齐，二者的对齐可通过对文字发音时长的拉伸或者压缩来实现，而对文字发音时长的拉伸或者压缩可通过一个变速比决定。变速比相当于发音时长与文字实际发音时长的比值。

对于一个文字而言，其实际的发音时长是从首个元音的起始位置处开始的，而实际发音的结束时间可看做相邻下一文字的首个元音起始位置处。将文字与候选对齐单元相结合来考虑的话，一个候选对齐单元中，所匹配的全部文字实际发音所占用的时长应该是从该候选对齐单元中首个匹配文字的首个元音位置开始，到与其相邻的下一候选对齐单元中首个匹配文字的首个元音位置结束。因此，可以通过当前处理单元以及相邻下一候选对齐单元分别所匹配的文字的起止时间和首个元音的起始位置来确定当前处理单元所匹配的全部文字的实际发音时长，并由此根据已知的单元时长以及确定出的实际发音时长来获得当前处理单元的当前变速比。

本实施例中，根据所述当前处理单元的单元时长，结合所述当前处理单元以及相邻下一候选对齐单元分别所匹配的文字的起止时间和首个元音的起始位置，确定所述当前处理单元的当前变速比，包括：

S21、根据所述当前处理单元相匹配的全部文字中的每个文字的起止时间以及首个元音的起始位置，确定相匹配的全部文字在所述当前处理单元内的发音占据时长。

本步骤可以获取到当前处理单元中相匹配的全部文字的匹配文字属性信息，匹配文字属性信息可以是每个文字的起止时间以及该文字首个原因的起始位置，基于这些信息，可以确定出相匹配的全部文字在所述当前处理单元内的发音占据时长。

示例性的，假设当前处理单元中仅有一个文字，该文字的起止时间分别为 $t1$ 和 $t2$ ，首个元音的起始位置为 $t3$ ，其中， $t1 < t3 < t2$ ，则该文字在当前处理单元内的发音占据时长实际为 $t2 - t3$ 。

假设当前处理单元中有两个文字，第一个文字的起止时间分别为 $t1$ 和 $t2$ ，首个元音的起始位置为 $t3$ ，第二个文字的起止时间分别为 $t2$ 和 $t4$ ，首个元音的起始位置为 $t5$ ，其中， $t1 < t3 < t2 < t5 < t4$ ，则这两个文字在当前处理单元内共同具备的发音占据时长为： $t4 - t3$ 或者为 $(t2 - t3) + (t4 - t5)$ 。可以看出，当前处理单元中所匹配的全部文字的发音占据时长仅为全部文字起止时长之和与第一个文字起始到首个元音的间隔时长的差值。

这里的发音占据时长并不是当前处理单元中相匹配的全部文字的实际发音时长，实际发音时长中还包括了当前处理单元相邻下一候选对齐单元所匹配的首个文字的元音间隔时长，该元音间隔时长可通过下述 S22 获得。本实施例采用上述方式确定发音占据时长的目的在于使语音段中每个文字的首个元音位置能够对齐到所匹配的候选对齐处理单元的节奏点上，由此来保证采用该种方式对齐后的文字和节奏点的播放效果更优于直接把字首与节奏点对齐的播放效果。

S22、根据所述当前处理单元相邻下一候选对齐单元所匹配的首个文字的起止时间以及首个元音的起始位置，确定所述首个文字的元音间隔时长。

以当前处理单元包括两个文字为例，假设当前处理单元相邻下一候选对齐单元所匹配的文字中首个文字的起止时间为 $t4$ 和 $t6$ ，首个元音的起始位置为 $t7$ ，其中， $t1 < t3 < t2 < t5 < t4 < t7 < t6$ ，则首个文字的元音间隔时长为 $t7 - t4$ 。

通过 S21 和 S22，可以确定当前处理单元中相匹配的全部文字的实际发音时长为当前处理单元内相匹配的所有文字的发音占据时长与相邻下一候选对齐单元所匹配的首个文字的元音间隔时长。当前处理单元中相匹配的全部文字的实

际发音时长为 $(t_4-t_3)+(t_7-t_4)$ 。

S23、将所述当前处理单元的单元时长与所确定的实际发音时长的比值作为所述当前处理单元的当前变速比，其中，所述实际发音时长为发音占据时长与元音间隔时长的和。

假设当前处理单元的单元时长为 t ，当前处理单元的当前变速比可表示为： $t/[(t_4-t_3)+(t_7-t_4)]$ 。

S3、将所述当前变速比与设定的第一变速比值以及第二变速比值进行比对，其中，所述第二变速比值大于所述第一变速比值。

通过 S2 确定出当前处理单元的当前变速比之后，可以将当前变速比与设定的第一变速比值以及第二变速比值进行比较，以此来确定通过当前变速比对相匹配的文字进行拉伸或者压缩是否满足常规的拉伸/压缩条件。

本实施例设定处于第一变速比值和第二变速比值之间的变速比满足拉伸/压缩条件，小于第一变速比值的不满足压缩条件，大于第二变速比值的不满足拉伸条件。

S4、如果所述当前变速比大于或等于第一变速比值且小于或等于第二变速比值，则将所述当前处理单元确定为对齐单元，并记所述当前变速比为所述对齐单元的变速比，之后执行 S7。

当所述当前变速比大于或等于第一变速比值且小于或等于第二变速比值时，认为该当前处理单元的当前变速比满足常规的拉伸/压缩条件，此时可直接将当前处理单元看作一个对齐单元，以及将当前变速比看作该对齐单元的变速比，并跳转至 S7 继续执行操作。

S5、如果所述当前变速比大于所述第二变速比值，则确定用于填补所述当前处理单元的静音时长，并根据所述静音时长确定新的当前变速比，之后执行 S3。

当所述当前变速比大于所述第二变速比值时，认为该当前处理单元的当前变速比不满足常规的拉伸条件，此时相当于当前处理单元的单元时长过长于相匹配的全部文字的实际发音时长，需要在当前处理单元中补入一个静音时长，以此来增加文字的实际发音时长。

一实施例中，确定所补入的静音时长为一个文字的起止时长，由此，根据该静音时长与已确定的实际发音时长相结合，重新确定一个以单元时长为分子，静音时长与实际发音时长之和为分母的当前变速比。并在此之后重新返回 S3 进行变速比的比对操作。

S6、如果所述当前变速比小于所述第一变速比值，则将当前处理单元与相邻下一候选对齐单元合并形成新的当前处理单元，返回执行 S2。

当所述当前变速比小于所述第一变速比值时，认为该当前处理单元的当前变速比不满足常规的压缩条件，此时相当于当前处理单元的单元时长过小于相匹配的全部文字的实际发音时长，需要在当前处理单元的基础上，再并入一个候选对齐单元，形成一个新的当前处理单元，以此来增加当前处理单元的单元时长。

一实施例中，待并入的候选对齐单元为当前处理单元相邻的下一候选对齐单元，此时，新形成的当前处理单元的单元时长为原有单元时长与下一候选对齐单元所对应单元时长的和，之后，可返回 S2 重新计算新形成的当前处理单元中相匹配的全部文字的实际发音时长。

本实施例中进行了选择下一候选对齐单元并入当前处理单元的操作，同样认为所选取的下一候选对齐单元已经被选中，后续在需要执行 S1 时，可以跳过对该下一候选对齐单元的选取，不再单独选取该下一候选对齐单元作为当前处理单元。

S7、判断所有候选对齐单元是否均被选中参与处理，若所有候选对齐单元均被选中参与处理，则执行 S8；若所有候选对齐单元中存在未被选中参与处理的候选对齐单元，返回执行 S1；

通过上述步骤确定出一个对齐单元后，该待对齐节奏段中可能还存在未被选中的候选对齐单元，可以通过 S7 进行一下判定，如果所有候选对齐单元均被选中参与了上述处理，则可执行 S8，如果所有候选对齐单元中存在未被选中参与处理的候选对齐单元，需要返回 S1 重新选择一个未被选中的候选对齐单元循环进行上述操作。

S8、汇总所确定的每个对齐单元以及相应的变速比。

可以将上述确定出的每个对齐单元以及相应的变速比进行汇总，以获得待对齐节奏段中包括的至少一个对齐单元以及变速比。

本可选实施例给出了确定待对齐节奏段中的对齐单元以及相应变速比的实现过程，通过本可选实施例的执行，能够保证待对齐节奏段中的节奏点与语音段中的文字的有效对齐，避免了语音文字与音乐节奏点错位的情况发生，由此为本实施例中语音到说唱音乐的转换提供了有效的理论支撑。

实施例三

图 6 为本申请实施例三提供的一种将语音转换为说唱音乐的装置的结构框图，该装置适用于对用户录制的语音进行说唱音乐转换的情况，其中，该装置可以由软件或硬件实现，并一般可集成在计算机设备上。如图 6 所示，该装置包括：信息确定模块 31、对齐信息确定模块 32 以及转换控制模块 33。

信息确定模块 31，设置为识别语音段以及处理背景音乐，获得所述语音段内文字的文字属性信息以及所述背景音乐的节奏信息；对齐信息确定模块 32，设置为根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；转换控制模块 33，设置为按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

本申请实施例三提供的一种将语音转换为说唱音乐的装置，有效实现了将用户随意录制的语音内容片段转化为配合背景音乐的说唱片段，简化了手动进行音频剪辑制作的繁琐过程，为非专业音频处理人员提供了说唱音乐制作的可能；同时，上述技术方案无需限制待转换语音内容，保证了待转换语音内容的自由化录制，还简化了语音转换的实现过程，避免了语音文字与音乐节奏点错位的情况，提升了语音转换说唱音乐的应用范围。

实施例四

图 7 为本申请实施例四提供的一种计算机设备的硬件结构示意图，该计算机设备包括：处理器和存储装置。存储装置中存储有至少一条指令，且指令由所述处理器执行，使得所述计算机设备执行如上述方法实施例所述的将语音转换为说唱音乐的方法。

参照图 7，该计算机设备可以包括：处理器 40、存储装置 41、显示屏 42、输入装置 43、输出装置 44 以及通信装置 45。该计算机设备中处理器 40 的数量可以是一个或者多个，图 6 中以一个处理器 40 为例。该计算机设备中存储装置 41 的数量可以是一个或者多个，图 7 中以一个存储装置 41 为例。该计算机设备的处理器 40、存储装置 41、显示屏 42、输入装置 43、输出装置 44 以及通信装置 45 可以通过总线或者其他方式连接，图 7 中以通过总线连接为例。

实施例中，处理器 40 执行存储装置 41 中存储的一个或多个程序时，实现如下操作：识别语音段以及处理背景音乐，获得所述语音段内文字的文字属性信息以及所述背景音乐的节奏信息；根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，

并获得每个对齐周期的对齐信息表；按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

本申请实施例还提供一种计算机可读存储介质，所述存储介质中的程序由计算机设备的处理器执行时，使得计算机设备能够执行如上述实施例所述的将语音转换为说唱音乐的方法。示例性的，上述实施例所述的将语音转换为说唱音乐的方法包括：识别语音段以及处理背景音乐，获得所述语音段内文字的文字属性信息以及所述背景音乐的节奏信息；根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

对于装置、计算机设备、存储介质实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

权 利 要 求 书

1、一种将语音转换为说唱音乐的方法，包括：

识别语音段以及处理背景音乐，获得所述语音段中的文字的文字属性信息以及所述背景音乐的节奏信息；

根据所述文字属性信息以及所述节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；

按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

2、根据权利要求1所述的方法，其中，所述识别语音段以及处理背景音乐，获得所述语音段中的文字的文字属性信息以及所述背景音乐的节奏信息，包括：

对用户选定的所述语音段进行降噪处理以及端点检测处理，通过对处理后的语音段的语音识别获得所述处理后的语音段中的多个文字中每个文字的文字序号、每个文字的起止时间、每个文字的首个元音的起始位置以及所述多个文字的文字总量，构成所述语音段的文字属性信息；

对所述用户选定的所述背景音乐进行节奏点检测和节拍周期划分，确定所述背景音乐中包含的多个节奏点的节奏点总量、每个节奏点的节奏点序号、以及所述背景音乐中包含的多个节拍周期中每个节拍周期的周期信息，构成所述背景音乐的节奏信息；

其中，所述周期信息包括：所述每个节拍周期的周期号、所述每个节拍周期所包含的多个节奏点的节奏点个数以及所述每个节拍周期所包含的每个节奏点的节奏点序号和节奏点起始时间。

3、根据权利要求2所述的方法，在所述确定所述背景音乐中包含的多个节奏点的节奏点总量、每个节奏点的节奏点序号、以及所述背景音乐中包含的多个节拍周期中每个节拍周期的周期信息，构成所述背景音乐的节奏信息之前，还包括：

获取从所述背景音乐中检测出的多个初始节奏点，并确定相邻两个初始节奏点之间形成的间隔时长；

根据所述语音段中的所述多个文字的平均字长时间以及所述间隔时长，确定所述多个初始节奏点中的待删除节奏点并删除所述待删除节奏点，获得所述背景音乐中的有效的节奏点。

4、根据权利要求2所述的方法，其中，所述根据所述文字属性信息以及所

述音乐节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表，包括：

根据所述文字属性信息中的所述文字总量以及所述音乐节奏信息中每个节拍周期的周期信息，确定用于将所述语音段与所述背景音乐对齐的所述至少一个对齐周期；

选取一个完整的对齐周期作为待对齐节奏段，根据所述文字属性信息以及所述待对齐节奏段中的待对齐节奏点的节奏点信息，确定至少一个对齐单元以及每个对齐单元对应的对齐单元信息；

汇总所述至少一个对齐单元对应的对齐单元信息形成所述待对齐节奏段的对齐信息表，并根据所述对齐信息表确定所述至少一个对齐周期中除所述一个完整的对齐周期外的对齐周期的对齐信息表。

5、根据权利要求4所述的方法，其中，所述根据所述文字属性信息中的所述文字总量以及所述音乐节奏信息中每个节拍周期的周期信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，包括：

判断所述背景音乐中的一个完整的节拍周期所对应的周期信息中的节奏点个数是否大于或等于所述文字总量；

响应于所述一个完整的节拍周期所对应的周期信息中的节奏点个数大于或等于所述文字总量，将每个节拍周期看作一个对齐周期；

响应于所述一个完整的节拍周期所对应的周期信息中的节奏点个数小于所述文字总量，在所述背景音乐中包括的节拍周期的个数大于1的情况下，按照周期号的排列顺序进行节拍周期的两两合并，形成至少一个新的节拍周期，并返回判定所述音乐节奏信息中一个完整的新的节拍周期所对应的周期信息中的节奏点个数是否大于或等于所述文字总量。

6、根据权利要求4所述的方法，其中，所述根据所述对齐信息表确定所述至少一个对齐周期中除所述一个完整的对齐周期外的对齐周期的对齐信息表，包括：

针对所述至少一个对齐周期中除所述一个完整的对齐周期外的每个对齐周期，在所述每个对齐周期为一个完整周期的情况下，将所述待对齐节奏段的对齐信息表作为所述每个对齐周期的对齐信息表；

在所述每个对齐周期为一个非完整周期，且所述待对齐节奏段的对齐信息表中的一行对应一个对齐单元的情况下，确定所述每个对齐周期所包括的节奏点的节奏点个数；从所述待对齐节奏段的对齐信息表中逆序选定所述每个对齐周期对应的节奏点个数行的对齐单元信息，将选定的所述节奏点个数行的对齐

单元信息作为所述每个对齐周期的对齐信息表。

7、根据权利要求4所述的方法，其中，所述根据所述文字属性信息以及所述待对齐节奏段中的待对齐节奏点的节奏点信息，确定至少一个对齐单元以及每个对齐单元对应的对齐单元信息，包括：

基于所述待对齐节奏段中的多个待对齐节奏点的节奏点信息，形成与所述多个待对齐节奏点一一对应的多个待对齐节奏块，并将所述多个待对齐节奏点的数量记为初始的剩余点个数；

确定所述剩余点个数与所述文字属性信息中的所述文字总量的比值，将所述比值记为对齐匹配值；

查找预设的节奏点-文字对齐规则表，确定所述对齐匹配值在所述节奏点-文字对齐规则表中所处的长度比值范围；

按照所述长度比值范围对应的对齐规则确定存在相匹配文字的待对齐节奏块，将所述存在相匹配文字的待对齐节奏块记为候选对齐单元；

统计未记为候选对齐单元的待对齐节奏块的块数，将所述块数作为新的剩余点个数，返回执行所述确定所述剩余点个数与所述文字属性信息中的所述文字总量的比值，将所述比值记为对齐匹配值的操作，直至最终的剩余点个数为0；

根据每个候选对齐单元的单元时长以及所述每个候选对齐单元所匹配的文字的匹配文字属性信息，确定所述至少一个对齐单元并获得每个对齐单元对应的变速比；

将每个对齐单元的单元序号、所包括的节奏点中的起始节奏点序号、所匹配的文字的文字序号以及变速比确定为所述每个对齐单元对应的对齐单元信息。

8、根据权利要求7所述的方法，其中，所述根据每个候选对齐单元的单元时长以及所述每个候选对齐单元所匹配的文字的匹配文字属性信息，确定所述至少一个对齐单元并获得每个对齐单元对应的变速比，包括：

按照单元序号的顺序选取一个未选中的候选对齐单元作为当前处理单元；

根据所述当前处理单元的单元时长，结合所述当前处理单元以及相邻下一候选对齐单元分别所匹配的文字的起止时间和首个元音的起始位置，确定所述当前处理单元的当前变速比；

将所述当前变速比与设定的第一变速比值以及第二变速比值进行比对，其中，所述第二变速比值大于所述第一变速比值；

在所述当前变速比大于或等于所述第一变速比值且小于或等于所述第二变速比值的情况下，将所述当前处理单元确定为所述对齐单元，并记所述当前变

速比为确定的所述对齐单元的变速比；判断所有候选对齐单元是否均被选中参与处理，响应于所述所有候选对齐单元均被选中参与处理，汇总所确定的每个对齐单元以及所述每个对齐单元对应的变速比，响应于所述所有候选对齐单元中存在未被选中参与处理的候选对齐单元，返回执行所述按照单元序号的顺序选取一个未选中的候选对齐单元作为当前处理单元的操作；

在所述当前变速比大于所述第二变速比值的情况下，确定用于填补所述当前处理单元的静音时长，并根据所述静音时长确定新的当前变速比，返回执行所述将所述当前变速比与设定的第一变速比值以及第二变速比值进行比对的操作；

在所述当前变速比小于所述第一变速比值的情况下，将所述当前处理单元与相邻下一候选对齐单元合并形成新的当前处理单元，并返回执行所述根据所述当前处理单元的单元时长，结合所述当前处理单元以及相邻下一候选对齐单元分别所匹配的文字的起止时间和首个元音的起始位置，确定所述当前处理单元的当前变速比的操作。

9、根据权利要求8所述的方法，其中，所述根据所述当前处理单元的单元时长，结合所述当前处理单元以及相邻下一候选对齐单元中分别所匹配的文字的起止时间和首个元音的起始位置，确定所述当前处理单元的当前变速比，包括：

根据所述当前处理单元相匹配的全部文字中的每个文字的起止时间以及首个元音的起始位置，确定所述相匹配的全部文字在所述当前处理单元中的发音占据时长；

根据所述当前处理单元相邻下一候选对齐单元所匹配的首个文字的起止时间以及首个元音的起始位置，确定所述首个文字的元音间隔时长；

将所述当前处理单元的单元时长与所确定的实际发音时长的比值作为所述当前处理单元的当前变速比，其中，所述实际发音时长为所述发音占据时长与所述元音间隔时长的和。

10、根据权利要求1-9任一项所述的方法，在所述根据所述文字属性信息以及所述音乐节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表之前，还包括：

在所述文字属性信息中的文字总量大于所述音乐节奏信息中的节奏点总量的情况下，结束将所述语音段转换为说唱音乐的处理，并输出重新获得语音段或背景音乐的提示。

11、一种将语音转换为说唱音乐的装置，包括：

信息确定模块，设置为识别语音段以及处理背景音乐，获得所述语音段中的文字的文字属性信息以及所述背景音乐的节奏信息；

对齐信息确定模块，设置为根据所述文字属性信息以及所述音乐节奏信息，确定用于将所述语音段与所述背景音乐对齐的至少一个对齐周期，并获得每个对齐周期的对齐信息表；

转换控制模块，设置为按照所述至少一个对齐周期的对齐信息表控制所述语音段中的文字与所述背景音乐中的节奏点对齐，得到对齐后的音频，并在对所述对齐后的音频进行变调调整以及特效处理后形成说唱音频。

12、一种计算机设备，包括：

至少一个处理器；

存储装置，设置为存储至少一个程序；

所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 1-10 任一项所述的将语音转换为说唱音乐的方法。

13、一种计算机可读存储介质，存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求 1-10 任一项所述的将语音转换为说唱音乐的方法。

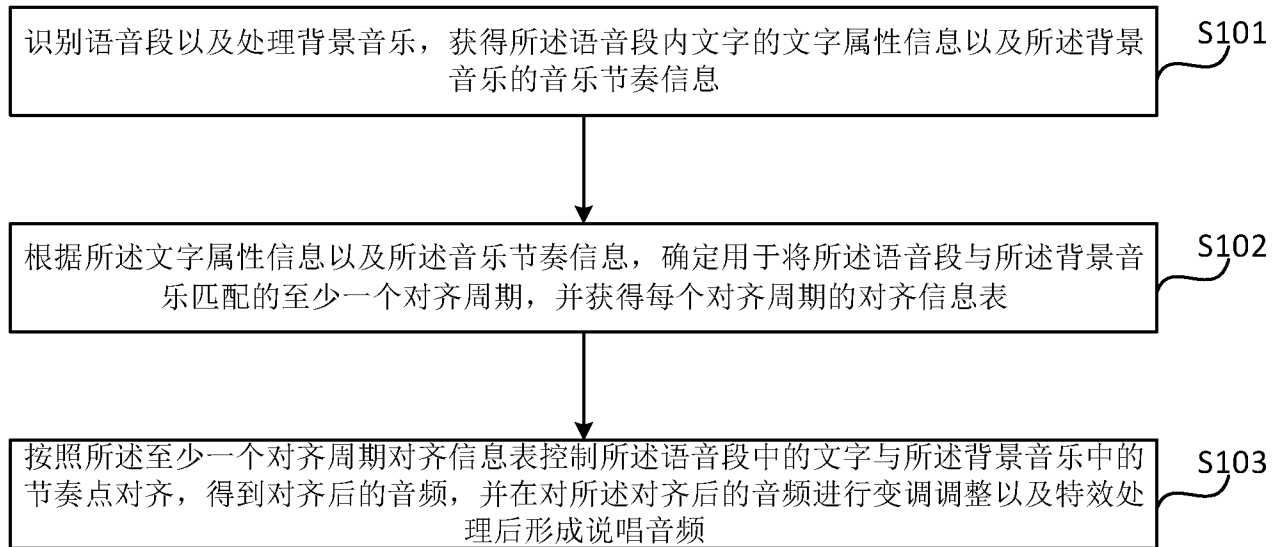


图 1

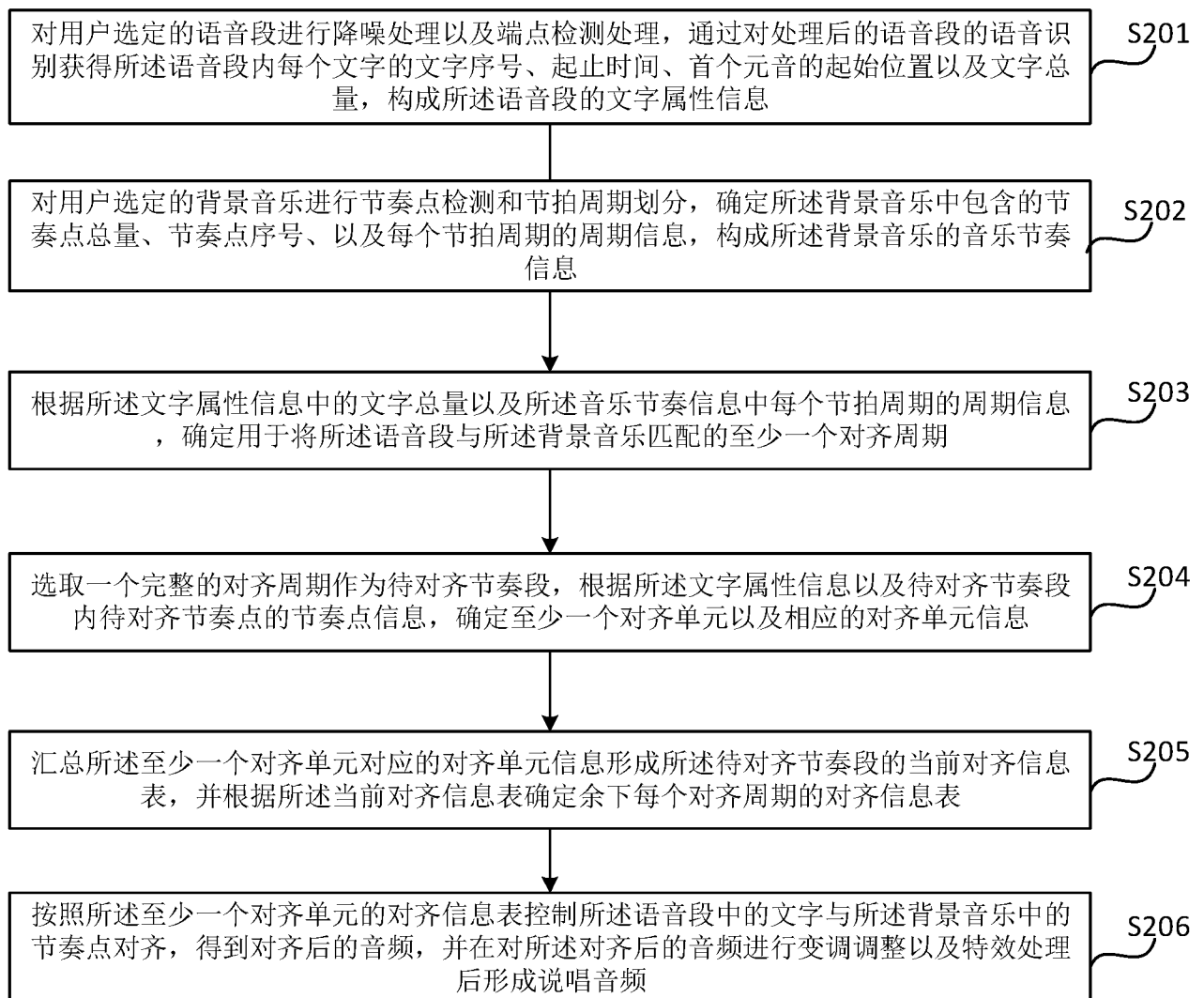


图 2

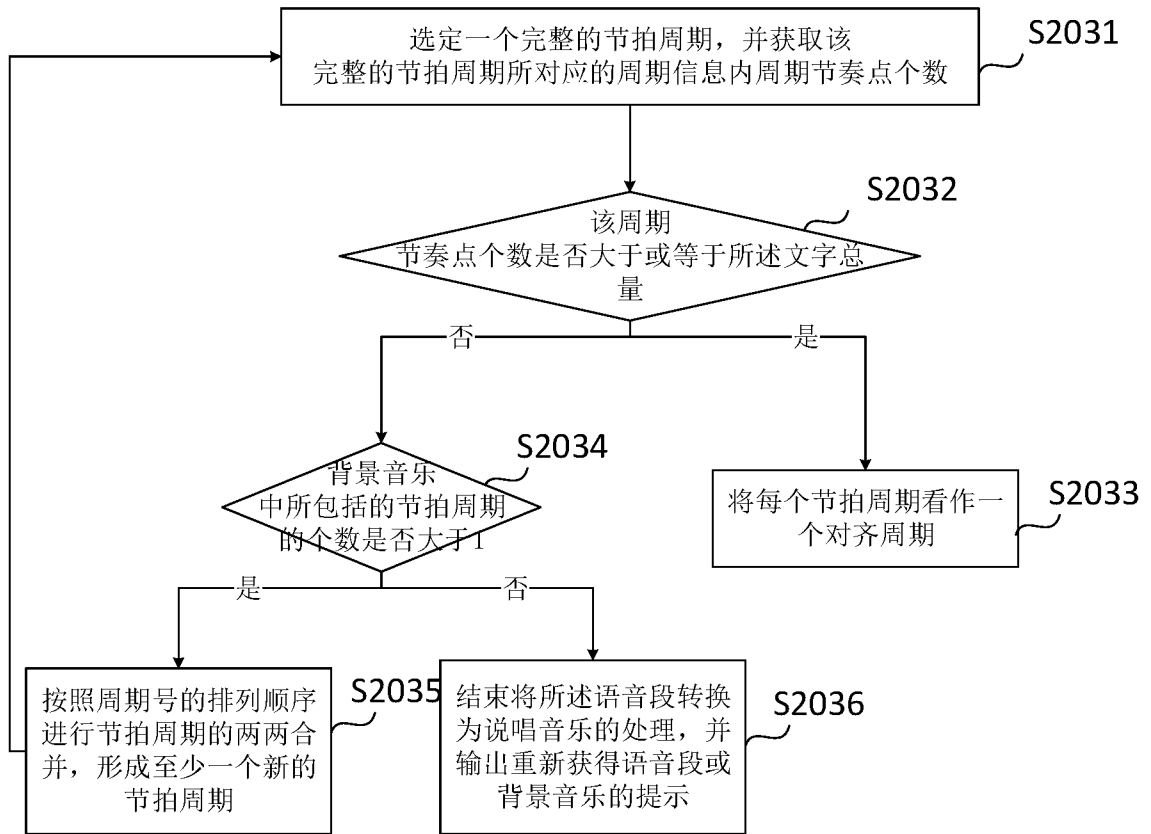


图 3

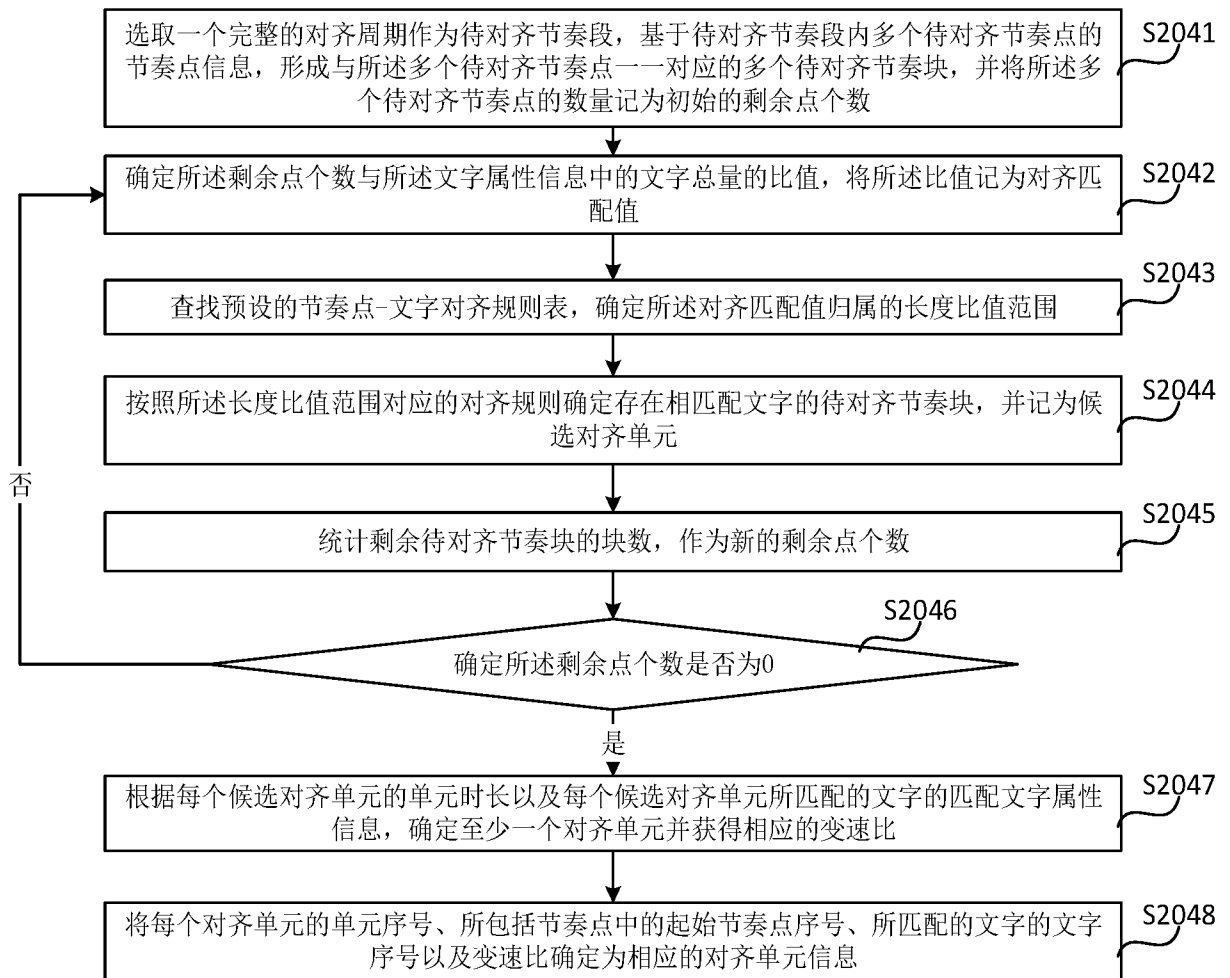


图 4

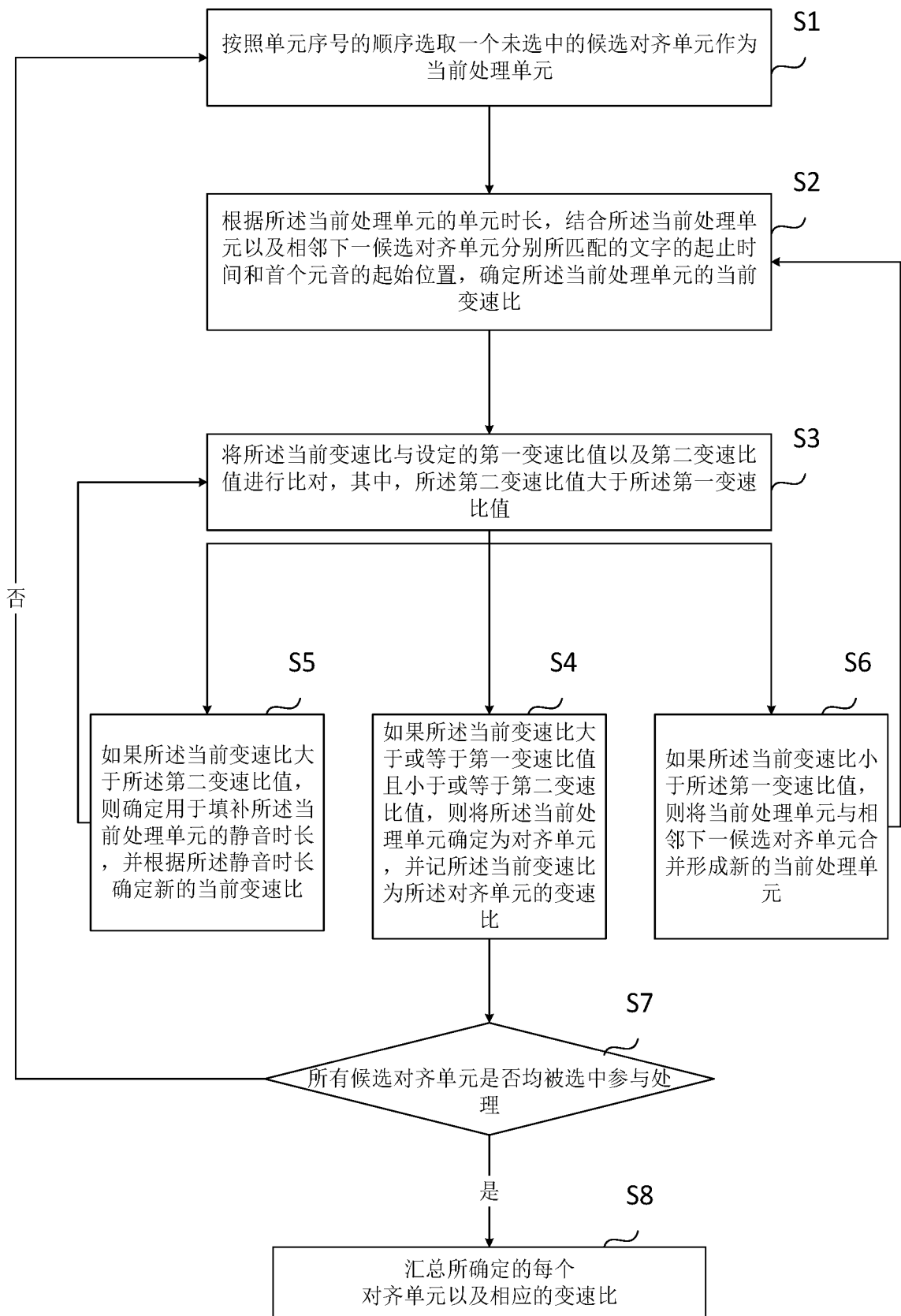


图 5



图 6

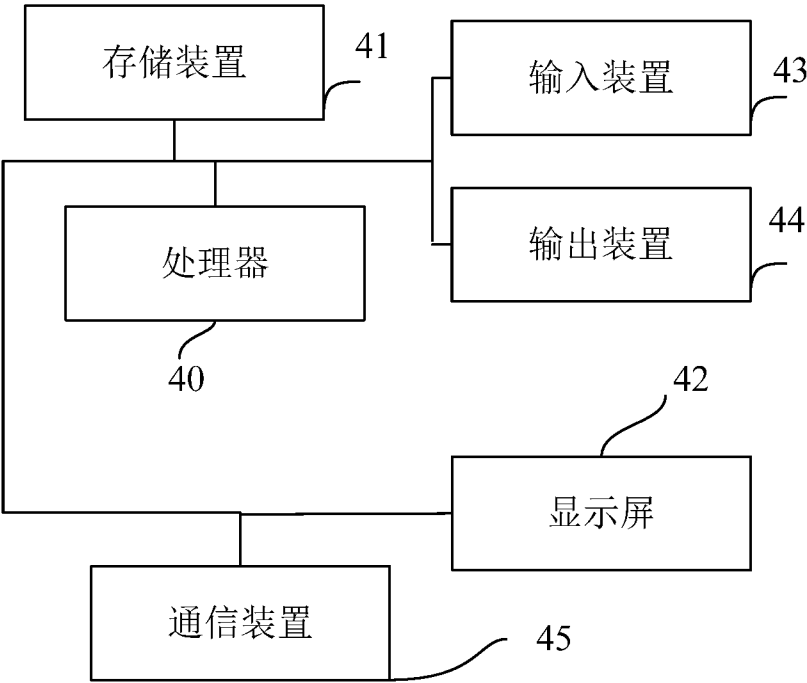


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/095236

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/02(2013.01)i; G10L 15/04(2013.01)i; G10H 1/00(2006.01)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, G10H

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)

CNABS, CNTXT, WPABSC; 语音, 音频, 转换, 合成, 说唱, 音乐, 背景, 伴奏, 节拍, 节奏, 韵律, 对应, 对齐, speech, voice, synthesis, convert+, mix+, music, background, accompan+, rhythm, correspond+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111862913 A (GUANGZHOU CITY PAGODA INFORMATION TECHNOLOGY CO., LTD.) 30 October 2020 (2020-10-30) claims 1-13, description paragraphs 0004-0192, figures 1-7	1-13
X	CN 101399036 A (SAMSUNG ELECTRONICS CO., LTD. et al.) 01 April 2009 (2009-04-01) description, page 4, paragraph 4 - page 13, paragraph 5, figures 1-12	1-13
A	CN 103440862 A (BEIJING QIYI CENTURY SCIENCE & TECHNOLOGY CO., LTD.) 11 December 2013 (2013-12-11) entire document	1-13
A	CN 103035235 A (SIEMENS AG.) 10 April 2013 (2013-04-10) entire document	1-13
A	CN 111402843 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 10 July 2020 (2020-07-10) entire document	1-13
A	CN 105931625 A (CHENGDU TUYA TECHNOLOGY CO., LTD.) 07 September 2016 (2016-09-07) entire document	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

19 August 2021

Date of mailing of the international search report

27 August 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/095236

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5811707 A (ROLAND K. K.) 22 September 1998 (1998-09-22) entire document	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/095236

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111862913	A	30 October 2020	None			
CN	101399036	A	01 April 2009	CN	101399036	B	29 May 2013
CN	103440862	A	11 December 2013	CN	103440862	B	09 March 2016
CN	103035235	A	10 April 2013	None			
CN	111402843	A	10 July 2020	None			
CN	105931625	A	07 September 2016	None			
US	5811707	A	22 September 1998	GB	2290685	A	03 January 1996
				CA	2151625	A1	25 December 1995
				CA	2151625	C	04 September 2001
				JP	H0816181	A	19 January 1996
				KR	319478	B1	22 April 2002

国际检索报告

国际申请号

PCT/CN2021/095236

A. 主题的分类

G10L 13/02(2013.01)i; G10L 15/04(2013.01)i; G10H 1/00(2006.01)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, G10H

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))

CNABS, CNTXT, WPABSC; 语音, 音频, 转换, 合成, 说唱, 音乐, 背景, 伴奏, 节拍, 节奏, 韵律, 对应, 对齐, speech, voice, synthesis, convert+, mix+, music, background, accompan+, rhythm, correspond+

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 111862913 A (广州市百果园信息技术有限公司) 2020年 10月 30日 (2020 - 10 - 30) 权利要求1-13, 说明书第0004-0192段, 附图1-7	1-13
X	CN 101399036 A (三星电子株式会社 等) 2009年 4月 1日 (2009 - 04 - 01) 说明书第4页第4段-第13页第5段, 附图1-12	1-13
A	CN 103440862 A (北京奇艺世纪科技有限公司) 2013年 12月 11日 (2013 - 12 - 11) 全文	1-13
A	CN 103035235 A (西门子子公司) 2013年 4月 10日 (2013 - 04 - 10) 全文	1-13
A	CN 111402843 A (北京字节跳动网络技术有限公司) 2020年 7月 10日 (2020 - 07 - 10) 全文	1-13
A	CN 105931625 A (成都涂鸦科技有限公司) 2016年 9月 7日 (2016 - 09 - 07) 全文	1-13
A	US 5811707 A (ROLAND KK.) 1998年 9月 22日 (1998 - 09 - 22) 全文	1-13

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“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

“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

国际检索实际完成的日期

2021年 8月 19日

国际检索报告邮寄日期

2021年 8月 27日

ISA/CN的名称和邮寄地址

中国国家知识产权局(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

房倩

电话号码 86-(010)-62089621

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/095236

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	111862913	A	2020年 10月 30日	无	
CN	101399036	A	2009年 4月 1日	CN 101399036	B 2013年 5月 29日
CN	103440862	A	2013年 12月 11日	CN 103440862	B 2016年 3月 9日
CN	103035235	A	2013年 4月 10日	无	
CN	111402843	A	2020年 7月 10日	无	
CN	105931625	A	2016年 9月 7日	无	
US	5811707	A	1998年 9月 22日	GB 2290685	A 1996年 1月 3日
				CA 2151625	A1 1995年 12月 25日
				CA 2151625	C 2001年 9月 4日
				JP H0816181	A 1996年 1月 19日
				KR 319478	B1 2002年 4月 22日