

(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
2018 年 6 月 28 日 (28.06.2018)



(10) 国际公布号
WO 2018/113535 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/115194
- (22) 国际申请日: 2017 年 12 月 8 日 (08.12.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201611196447.6 2016 年 12 月 22 日 (22.12.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 阳鹤翔 (YANG, Hexiang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL

PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, 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, 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,

(54) **Title:** METHOD AND APPARATUS FOR AUTOMATICALLY GENERATING DUBBING CHARACTERS, AND ELECTRONIC DEVICE

(54) 发明名称: 一种自动生成配音文字的方法、装置以及电子设备

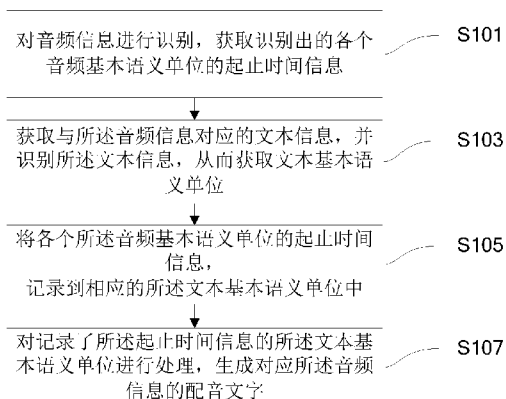


图 1

(57) **Abstract:** A method and apparatus for automatically generating dubbing characters, and an electronic device. The method for generating dubbing characters comprises: identifying audio information to acquire starting and ending time information about each identified audio basic semantic unit (S101); acquiring text information corresponding to the audio information, and identifying the text information, so as to acquire a text basic semantic unit (S103); recording the starting and ending time information about each of the audio basic semantic units in the corresponding text basic semantic unit (S105); and processing the text basic semantic unit into which the starting and ending time information is recorded, so as to generate dubbing characters corresponding to the audio information (S107). By means of the method, a dynamic lyrics file can be produced without using a manual method, thereby improving the production efficiency, reducing the production cost, and simplifying the production procedure.

- S101 Identifying audio information to acquire starting and ending time information about each identified audio basic semantic unit
- S103 Acquiring text information corresponding to the audio information, and identifying the text information, so as to acquire a text basic semantic unit
- S105 Recording the starting and ending time information about each of the audio basic semantic units in the corresponding text basic semantic unit
- S107 Processing the text basic semantic unit into which the starting and ending time information is recorded, so as to generate dubbing characters corresponding to the audio information

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)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种自动生成配音文字的方法、装置以及电子设备。其中，所述生成配音文字的方法，包括：对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息(S101)；获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位(S103)；将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中(S105)；对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字(S107)。采用上述方法，能够使动态歌词文件不再采用人工的方式进行制作，提高了制作的效率降低了制作成本，简化了制作的流程。

一种自动生成配音文字的方法、装置以及电子设备

本申请要求 2016 年 12 月 22 日递交的申请号为 201611196447.6、发明名称为“一种自动生成配音文字的方法、装置以及电子设备”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。

5

技术领域

本申请涉及计算机技术领域，具体涉及一种自动生成配音文字的方法；本申请同时涉及一种自动生成配音文字的装置以及一种电子设备。

10 背景技术

随着音频处理技术的发展，用户对试听体验有了更高的要求，不仅要求音频播放应用能够播放音频文件，还希望音频播放应用能够同步显示与音频文件相应的歌词文件。音频播放同步显示歌词功能使得人们在听到优美旋律的同时能够看到该音频文件的歌词，该功能已经成为了音频播放应用以及播放器的必备功能之一。

15 为了满足用户的需求，目前用于音频播放同步显示的歌词主要采用人工方式来进行的，人工边听音频边给歌词标注时间，为音频文件数据库中的每个音频文件生成相应的歌词文件，并将所生成的歌词文件导入到音频播放应用中，从而在播放音频文件时，同步显示相应地歌词文件。

由此可见，在现有的用于音频播放同步显示的歌词的制作方案下，采用人工方式生
20 成歌词文件制作过程比较繁琐，不仅效率低且成本高。随着音频曲库规模的不断扩大，人工方式所存在的弊端显得日益严重。

发明内容

本申请提供一种自动生成配音文字的方法，以解决现有技术中的上述问题。本申请
25 同时涉及一种自动生成配音文字的装置以及一种电子设备。

本申请实施例提供了一种自动生成配音文字的方法，所述自动生成配音文字的方法，包括：

对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息；

获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语
30 义单位；

将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中；

对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字。

- 5 可选的，所述对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字，包括：

针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位；

根据已获取的所述文本基本语义单位中记录的起止时间信息，确定所述单句的起止时间信息；

- 10 将确定了起止时间信息的所述单句进行整合，形成对应所述音频信息，且具有每一单句的起止时间信息的配音文字。

可选的，所述针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位时，若所述文本基本语义单位中记录了至少两组起止时间信息，则按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组。

- 15 可选的，在所述按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组的步骤之后，包括：

根据预定的计算方法，对每一所述文本基本语义单位组中，各个文本基本语义单位的所有起止时间信息进行筛选，确定组成所述单句的文本基本语义单位组。

可选的，所述预定的计算方法，包括：

- 20 计算各个所述文本基本语义单位组内，每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距，获取各个所述文本基本语义单位组中所述起始时间与所述终止时间的时间间距的和，将所述时间间距的和作为所述文本基本语义单位组的误差值。

- 25 可选的，所述对每一所述文本基本语义单位组中，各个文本基本语义单位的所有起止时间信息进行筛选，确定组成所述单句的文本基本语义单位组，包括：

对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组。

可选的，在所述保留误差值低于预设的阈值的文本基本语义单位组的步骤之后，包括：

- 30 计算保留的所述文本基本语义单位组内，每一文本基本语义单位中的起始时间大于

所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数，获取该次数最大的文本基本语义单位组。

可选的，所述识别所述文本信息获取文本基本语义单位，包括：

从所述文本信息中，按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

可选的，在将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中时，若所述音频基本语义单位的起止时间信息为空值，则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

可选的，在所述确定组成所述单句的文本基本语义单位组的步骤之后，包括：

按照预定的推算方式，对取值为空值的所述文本基本语义单位推算起止时间信息。

可选的，所述预定的推算方式，包括：

计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息；

将取值为空值的所述文本基本语义单位，上一个文本基本语义单位中的终止时间，放入取值为空值的所述文本基本语义单位的起始时间中；

将所述终止时间加上所述平均时间信息后，放入取值为空值的所述文本基本语义单位的终止时间中。

相应的，本申请实施例还提供了一种自动生成配音文字的装置，所述自动生成配音文字的装置，包括：

音频识别单元，用于对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息；

文本识别单元，用于获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位；

时间写入单元，用于将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中；

配音文字生成单元，用于对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字。

可选的，所述配音文字生成单元，包括：

文本语义获取子单元，用于针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位；

时间信息确定子单元，用于根据已获取的所述文本基本语义单位中记录的起止时间

信息确定所述单句的起止时间信息；

配音文字生成子单元，用于将确定了起止时间信息的所述单句进行整合，形成对应所述音频信息，且具有每一单句的起止时间信息的配音文字。

可选的，所述时文本语义获取子单元，具体用于针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位时，若所述文本基本语义单位中记录了至少两组起止时间信息，则按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组。

可选的，所述的自动生成配音文字的装置，还包括：

文本语义筛选子单元，用于在所述按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组之后，根据预定的计算方法，对每一所述文本基本语义单位组中，各个文本基本语义单位的所有起止时间信息进行筛选，确定组成所述单句的文本基本语义单位组。

可选的，所述文本语义筛选子单元，包括：

误差计算子单元，用于计算各个所述文本基本语义单位组内，每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距，获取各个所述文本基本语义单位组中所述起始时间与所述终止时间的时间间距的和，将所述时间间距的和作为所述文本基本语义单位组的误差值。

可选的，所述文本语义筛选子单元，还包括：

过滤子单元，用于对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组。

可选的，所述文本语义筛选子单元，还包括：

时间次数计算子单元，用于在所述保留误差值低于预设的阈值的文本基本语义单位组之后，计算保留的所述文本基本语义单位组内，每一文本基本语义单位中的起始时间大于所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数，获取该次数最大的文本基本语义单位组。

可选的，所述文本识别单元，具体用于从所述文本信息中，按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

可选的，所述时间写入单元，具体用于在将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中时，若所述音频基本语义单位的起止时间信息为空值，则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

可选的,所述的自动生成配音文字的装置,还包括:

时间推算单元,用于在所述确定组成所述单句的文本基本语义单位组之后,按照预定的推算方式,对取值为空值的所述文本基本语义单位推算起止时间信息。

可选的,所述时间推算单元,包括:

- 5 平均时间计算子单元,用于计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息;

起始时间写入子单元,用于将取值为空值的所述文本基本语义单位,上一个文本基本语义单位中的终止时间,放入取值为空值的所述文本基本语义单位的起始时间中;

- 10 终止时间写入子单元,用于将所述终止时间加上所述平均时间信息后,放入取值为空值的所述文本基本语义单位的终止时间中。

此外,本申请实施例还提供了一种电子设备,包括:

显示器;

处理器;

- 15 存储器,用于存储配音文字生成程序,所述程序在被所述处理器读取执行时,执行如下操作:对音频信息进行识别,获取识别出的各个音频基本语义单位的起止时间信息;获取与所述音频信息对应的文本信息,并识别所述文本信息,从而获取文本基本语义单位;将各个所述音频基本语义单位的起止时间信息,记录到相应的所述文本基本语义单位中;对记录了所述起止时间信息的所述文本基本语义单位进行处理,生成对应所述音频信息的配音文字。

- 20 与现有技术相比,本申请具有以下优点:

- 本申请提供了一种自动生成配音文字的方法、装置以及电子设备,通过对音频信息进行识别,获取识别出的各个音频基本语义单位的起止时间信息;获取与所述音频信息对应的文本信息,并识别所述文本信息,从而获取文本基本语义单位;将各个所述音频基本语义单位的起止时间信息,记录到相应的所述文本基本语义单位中;对记录了所述
25 起止时间信息的所述文本基本语义单位进行处理,生成对应所述音频信息的配音文字。
所述技术方案通过对音频信息进行语音识别,获取音频信息中每个音频基本语义单位起止时间信息,通过识别所述音频信息对应的文本信息,确定文本信息内每个单句内文本基本语义单位的数量与字形,使所述音频信息中识别出的音频基本语义单位与所述文本信息中识别出的文本基本语义单位相对应,在确立对应关系后,根据所述音频信息中每
30 个音频基本语义单位起止时间信息确定文本信息中对应单句的时间信息,使文本内的每

条单句带有时间信息，使动态歌词文件不再采用人工的方式进行制作，提高了制作的效率降低了制作成本，简化了制作的流程。

附图说明

5 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所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申请中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图 1 示出了根据本申请的实施例提供的自动生成配音文字的方法的流程图；

10 图 2 示出了根据本申请的实施例提供的对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字的流程图；

图 3 示出了根据本申请的实施例提供的自动生成配音文字的装置的示意图；

图 4 示出了根据本申请的实施例提供的电子设备的示意图。

15 具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此本申请不受下面公开的具体实施的限制。

20 为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施方式对本申请进行进一步的详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

本申请的实施例提供了一种自动生成配音文字的方法，本申请的实施例同时提供了一种自动生成配音文字的装置以及一种电子设备。在下面的实施例中逐一进行详细说明。

25 目前用于音频播放同步显示的歌词主要采用人工方式来进行的，人工边听音频边给歌词标注时间，为音频文件数据库中的每个音频文件生成相应的歌词文件，并将所生成的歌词文件导入到音频播放应用中，从而在播放音频文件时，同步显示相应地歌词文件。由此可见，在现有的用于音频播放同步显示的歌词的制作方案下，采用人工方式生成歌词文件制作过程比较繁琐，不仅效率低且成本高。随着音频曲库规模的不断扩大，人工方式所存在的弊端显得日益严重。针对这一问题，本申请的技术方案通过对音频信息
30 进行语音识别，获取音频信息中每个音频基本语义单位起止时间信息，通过识别所述音频

信息对应的文本信息，确定文本信息内每个单句内文本基本语义单位的数量与字形，使所述音频信息中识别出的音频基本语义单位与所述文本信息中识别出的文本基本语义单位相对应，在确立对应关系后，根据所述音频信息中每个音频基本语义单位起止时间信息确定文本信息中对应单句的时间信息，使文本内的歌词带有时间信息，从而实现了自动制作动态歌词文件的功能。

在详细描述本实施例的具体步骤之前，先对本技术方案涉及的动态歌词作简要说明。

动态歌词是通过编辑器把歌词按歌曲歌词出现的时间编辑到一起，然后在播放歌曲时同步依次将歌词显示出来。常用的动态歌词文件包括：lrc、qrc 等。

lrc 是英文 lyric（歌词）的缩写，被用做动态歌词文件的扩展名。以 lrc 为扩展名的歌词文件可以在各类数码播放器中同步显示。lrc 歌词是一种包含着“*:*:*”（其中，“*”是指通配符，用来代替一个或多个真正的字符。在实际的歌词文件中“*”是指歌词的时间(即时间内容)，例如：“01:01:00”是指 1 分 1 秒；“:”用来分割分、秒、毫秒的时间信息）形式的“标签(tag)”的、基于纯文本的歌词专用格式。这种歌词文件能以文字处理软件查看、编辑(用记事本按照上述格式写好后，将扩展名改为 lrc 即可做出“文件名.LRC”的歌词文件)。Lrc 动态歌词文件的标准格式为[分钟:秒:毫秒]歌词。

lrc 歌词文本中含有两类标签：

一是标识标签，其格式为“[标识名:值]”主要包含以下预定义的标签：

[ar:歌手名]、[ti:歌曲名]、[al:专辑名]、[by:编辑者(指 lrc 歌词的制作人)]。

二是时间标签，形式为“[mm:ss]”或“[mm:ss.ff]”，时间标签需位于某行歌词中的句首部分，一行歌词可以包含多个时间标签(比如歌词中的迭句部分)。当歌曲播放到达某一时间点时，就会寻找对应的时间标签并显示标签后面的歌词文本，这样就完成了“歌词同步”的功能。

lrc 动态歌词文件在使用时要求歌曲和 lrc 动态歌词文件的文件名相同（即除了扩展名.mp3、.wma、.lrc 等不同之外，点前面的文字、文字格式要一模一样）并且放在同一目录下（即同一文件夹中），用带显示歌词功能的播放器播放歌曲时歌词就可以同步显示。

本申请的实施例提供了一种生成配音文字的方法，所述生成配音文字的方法实施例如下：

请参考图 1，其示出了根据本申请的实施例提供的自动生成配音文字的方法的流程图。

所述自动生成配音文字的方法包括：

步骤 S101，对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息。

在本实施例中，所述对音频信息进行识别，主要是将所述音频信息的语音信号转换为可识别的文本信息，例如：以文本信息的形式获取将所述音频信息的语音信号转换为可以识别的音频基本语义单位。所述音频基本语义单位包括：中文文字、中文词语、拼音、数字、英文文字和/或英文词语等。具体的，语音识别过程可采用统计模式识别技术等语音识别方法。

在具体实施时，可以通过 CMU-Sphinx 语音识别系统对所述音频信息进行语音识别。
CMU-Sphinx 是大词汇量语音识别系统，采用连续隐含马尔可夫模型 CHMM 建模。支持多种模式操作，高精度模式扁平译码器以及快速搜索模式树译码器。

需要说明的是，所述文本信息中，包含从所述音频信息中识别出的音频基本语义单位以及所述音频基本语义单位在所述音频信息中起止时间信息。可以理解的，所述音频信息可以是 mp3 或其他音乐格式的歌曲文件，mp3 文件是具有一定时长直接记录了真实声音的音频文件，所以在对 mp3 文件进行识别，将识别出的音频基本语义单位采用文本信息的形式进行输出时会记录识别出的该音频基本语义单位在所述音频信息中播放时起止时间信息。

在本实施例中，在对所述音频信息进行识别后输出的所述文本信息中采用如下格式记录识别出的音频基本语义单位以及所述音频基本语义单位的时间信息：<word, TIMECLASS>。其中，word 是指识别出的音频基本语义单位，TIMECLASS 是指时间标注，该时间标注采用起始时间以及终止时间 {startTime,endTime} 的形式记录该音频基本语义单位在在所述音频信息中播放时出现时的时间信息，即：是相对于所述音频信息在开始播放 0 时刻时的偏移量，单位为毫秒。

下面通过一个具体的例子说明所述生成配音文字的方法，例如：所述音频信息为 mp3 文件，该 mp3 文件在播放时的时常为 10 秒，在该 mp3 文件播放到 1 秒时出现歌词：“我想了又想”，则通过识别所述音频信息获取的文本信息中记录的识别出的音频基本语义单位以及所述音频基本语义单位的时间信息为：

<word: “我”， {startTime:1000,endTime:1100} >;

<word: “想”， {startTime:1200,endTime:1300} >;

<word: “了”， {startTime:1400,endTime:1500} >;

<word: “又”， {startTime:1600,endTime:1700} >;

<word: “想”， {startTime:1800,endTime:1900} >。

需要说明的是，若所述音频信息为中文的音频信息，则在对所述音频信息进行识别后输出的所述文本信息中记录的识别出的音频基本语义单位为单个中文汉字；同样的道理，若所述音频信息为英文的音频信息，则在对所述音频信息进行识别后输出的所述文本信息中记录的识别出的音频基本语义单位为单个英文单词。

可以理解的，所述音频基本语义单位的起止时间信息是以毫秒为单位进行记录的，而歌词：“我想了又想”是在该 mp3 文件播放到 1 秒时出现，则音频基本语义单位“我”在该 mp3 文件播放到 1 秒至 1.1 秒时出现，所以记录的音频基本语义单位“我”的时间信息为 {startTime:1000,endTime:1100} 。

步骤 S103，获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位。

在本实施例中，所述获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位，可以采用如下方式实现：通过互联网搜索所述音频信息对应的文本信息，在获取所述文本信息后对所述文本信息中的每个基本语义单位进行识别，对识别出的每个基本语义单位形成时间信息为空值的文本基本语义单位，获取所述文本基本语义单位。

需要说明的是，所述基本语义单位是所述文本信息内的单字信息，包括：中文文字、中文词语、拼音、数字、英文文字和/或英文词语等。

沿用上述具体的例子进行说明：所述音频信息为 mp3 文件，通过互联网络搜索该 mp3 文件对应的歌词文本，所述歌词文本的具体内容为：“我想了又想”，在获取该 mp3 文件对应的歌词文本后，对所述文本信息中的每个基本语义单位进行识别，对识别出的每个基本语义单位形成时间信息为空值的文本基本语义单位：

<word: “我”， timeList{}>;

<word: “想”， timeList{}>;

<word: “了”， timeList{}>;

<word: “又”， timeList{}>;

<word: “想”， timeList{}>。

步骤 S105，将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中。

在本实施例中，所述将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中，可以采用如下方式实现：将在对所述音频信息进行识别后识别出的各个所述音频基本语义单位与从所述音频信息对应的文本信息中对每个基本语义单位进行识别形成的文本基本语义单位进行匹配，将所述音频基本语义单位的起止时间信息放入到与该音频基本语义单位相应的文本基本语义单位内。

例如：通过识别所述音频信息获取的文本信息中记录的识别出的音频基本语义单位以及所述音频基本语义单位的时间信息为：

<word: “我”， {startTime:1000,endTime:1100} >;

<word: “想”， {startTime:1200,endTime:1300} >;

对所述文本信息中的每个基本语义单位进行识别，对识别出的每个基本语义单位形成时间信息为空值的文本基本语义单位为：

<word: “我”， timeList{}>;

<word: “想”， timeList{}>;

进行识别形成的文本基本语义单位进行匹配

由于所述音频信息进行识别后识别出的音频基本语义单位“我”和“想”与对所述歌词文本中歌词的文本基本语义单位进行识别后形成的文本基本语义单位“我”和“想”的字形相同，则将音频基本语义单位“我”和“想”的起止时间信息放入到文本基本语义单位“我”和“想”中：

<word: “我”， timeList{startTime:1000,endTime:1100}>;

<word: “想”， timeList{startTime:1200,endTime:1300}>。

需要说明的是，由于音频信息中相同的音频基本语义单位出现的次数可能不唯一，例如：在一首歌曲中，某个相同的字可以多次出现，所以在执行步骤 S105 将各个所述音频基本语义单位的起止时间信息记录到相应的所述文本基本语义单位中时，当具有相同的音频基本语义单位时，可以采用如下方式实现：将从所述音频信息中获取的音频基本语义单位的起止时间信息放入每一个与所述音频基本语义单位相同的文本基本语义单位内。

沿用上述具体的例子进行说明：通过识别所述音频信息获取的文本信息中记录的识别出的音频基本语义单位以及所述音频基本语义单位的时间信息为：

<word: “我”， {startTime:1000,endTime:1100} >;

<word: “想”， {startTime:1200,endTime:1300} >;

<word: “了”, {startTime:1400,endTime:1500} >;
<word: “又”, {startTime:1600,endTime:1700} >;
<word: “想”, {startTime:1800,endTime:1900} >。

在获取所述文本信息后对所述文本信息中的每个基本语义单位进行识别, 对识别出的
5 的每个基本语义单位形成时间信息为空值的文本基本语义单位为:

<word: “我”, timeList{}>;
<word: “想”, timeList{}>;
<word: “了”, timeList{}>;
<word: “又”, timeList{}>;
10 <word: “想”, timeList{}>。

由于所述音频信息进行识别后识别出的音频基本语义单位“我”、“想”、“了”、“又”和“想”与对所述歌词文本中歌词的文本基本语义单位进行提取后形成的文本基本语义单位“我”、“想”、“了”、“又”和“想”的时间集中字形相同, 则将上述音频基本语义单位的起止时间信息放入到相应的文本基本语义单位中:

15 <word: “我”, timeList{startTime:1000,endTime:1100}>;
<word: “想”, timeList{startTime:1200,endTime:1300},
{startTime:1800,endTime:1900}>;
<word: “了”, timeList{startTime:1400,endTime:1500}>;
<word: “又”, timeList{startTime:1600,endTime:1700}>;
20 <word: “想”, timeList{startTime:1200,endTime:1300},
{startTime:1800,endTime:1900}>。

可以理解的, 在上述例子中, 由于在所述音频信息以及所述文本中“想”字出现了两次, 所以将从所述音频信息中获取的“想”的起止时间信息分别放入与“想”字对应的文本基本语义单位“想”中。

25 步骤 S107, 对记录了所述起止时间信息的所述文本基本语义单位进行处理, 生成对应所述音频信息的配音文字。

在本实施例中, 所述对记录了所述起止时间信息的所述文本基本语义单位进行处理, 生成对应所述音频信息的配音文字, 可以采用如下方式实现: 根据所述文本信息中的具体单句确定组成该单句的文本基本语义单位, 并根据组成该单句的所述文本基本语义单位中的起止时间信息确定该单句的起止时间信息, 整理所有的单句的起止时间信息, 生
30

成对应所述音频信息并确定了所有单句的起止时间信息的配音文字。

需要说明的是，在所述文本信息中确定单句时，可以通过单句与单句之间的换行符区分所述文本中的每一单句。

所述对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述
5 音频信息的配音文字，具体包括步骤 S107-1 至 S107-3，下面结合附图 2 作进一步说明。

请参考图 2，其示出了根据本申请的实施例提供的对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字的流程图。

所述对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字，包括：

10 步骤 S107-1，针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位。

在本实施例中，所述针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位，可以采用如下方式实现：根据换行符进行区分所述文本信息中的每一单句，并针对具体的某一单句获取组成所述单句的文本基本语义单位。

15 例如：所述文本信息中的具体单句为：“我想”和“你了”，则组成该单句的文本基本语义单位为“我”和“想”以及“你”和“了”，且文本基本语义单位“我”和“想”为：

<word: “我”，timeList{startTime:1000,endTime:1100}>;

<word: “想”，timeList{startTime:1200,endTime:1300}>;

20 文本基本语义单位“你”和“了”为：

<word: “你”，timeList{startTime:1400,endTime:1500}>;

<word: “了”，timeList{startTime:1600,endTime:1700}>; 。

步骤 S107-2，根据已获取的所述文本基本语义单位中记录的起止时间信息，确定所述单句的起止时间信息。

25 在本实施例中，所述根据已获取的所述文本基本语义单位中记录的起止时间信息，确定所述单句的起止时间信息，可以采用如下方式实现：以组成所述单句的文本基本语义单位中起始时间最早的时间信息作为所述单句的起始时间，以组成所述单句的文本基本语义单位的时间集中终止时间最晚的时间信息作为所述单句的终止时间，并将所述单句的起始时间以及终止时间作为所述单句的起止时间信息。

30 例如：根据上述两个文本基本语义单位的时间信息确定的单句“我想”的时间信息

为:

timeList{startTime:1000,endTime:1300},

根据上述两个文本基本语义单位的时间信息确定的单句“你了”的时间信息为:

timeList{startTime:1400,endTime:1700}。

- 5 步骤 S107-3, 将确定了起止时间信息的所述单句进行整合, 形成对应所述音频信息, 且具有每一单句的起止时间信息的配音文字。

例如: 在确定所述文本中所有的单句“我想”和“你了”的时间信息之后, 输出带有上述两句的时间信息的文本(即: 动态歌词 lrc):

[00:01:00]我想

- 10 [00:01:40]你了。

可以理解的, 在播放所述音频信息时, 在达到每一所述单句的显示时间时, 显示配音文字中相应的单句。

- 在本实施例中, 由于音频信息中相同的音频基本语义单位出现的次数可能不唯一, 例如: 在一首歌曲中, 某个相同的字可以多次出现, 所以在执行步骤 S107-1 针对所述文本信息中每一单句, 获取组成所述单句的文本基本语义单位时, 当具有相同的基本语义单位时, 可以采用如下方式实现: 若所述文本基本语义单位中记录了至少两组起止时间信息, 则按照起止时间信息的组数, 分别形成组成所述单句的文本基本语义单位组。

沿用上述具体的例子进行说明: 所述文本中的具体单句为: “我想了又想”, 则组成该单句的文本基本语义单位“我”、“想”、“了”、“又”和“想”为:

- 20 <word: “我”, timeList{startTime:1000,endTime:1100}>;

<word: “ 想 ” , timeList{startTime:1200,endTime:1300} , {startTime:1800,endTime:1900}>;

<word: “了”, timeList{startTime:1400,endTime:1500}>;

<word: “又”, timeList{startTime:1600,endTime:1700}>;

- 25 <word: “ 想 ” , timeList{startTime:1200,endTime:1300} , {startTime:1800,endTime:1900}>;

由于组成单句“我想了又想”的两个文本基本语义单位“想”中各具有两组时间信息, 则按照起止时间信息的组数, 分别形成组成所述单句的文本基本语义单位组包括如下四组: 第一组为:

- 30 <word: “我”, timeList{startTime:1000,endTime:1100}>;

<word: “想”, timeList{startTime:1200,endTime:1300}>;
<word: “了”, timeList{startTime:1400,endTime:1500}>;
<word: “又”, timeList{startTime:1600,endTime:1700}>;
<word: “想”, timeList{startTime:1200,endTime:1300}>;

5 第二组为:

<word: “我”, timeList{startTime:1000,endTime:1100}>;
<word: “想”, timeList{startTime:1200,endTime:1300}>;
<word: “了”, timeList{startTime:1400,endTime:1500}>;
<word: “又”, timeList{startTime:1600,endTime:1700}>;
10 <word: “想”, timeList{startTime:1800,endTime:1900}>;

第三组为:

<word: “我”, timeList{startTime:1000,endTime:1100}>;
<word: “想”, timeList{startTime:1800,endTime:1900}>;
<word: “了”, timeList{startTime:1400,endTime:1500}>;
15 <word: “又”, timeList{startTime:1600,endTime:1700}>;
<word: “想”, timeList{startTime:1200,endTime:1300}>;

第四组为:

<word: “我”, timeList{startTime:1000,endTime:1100}>;
<word: “想”, timeList{startTime:1800,endTime:1900}>;
20 <word: “了”, timeList{startTime:1400,endTime:1500}>;
<word: “又”, timeList{startTime:1600,endTime:1700}>;
<word: “想”, timeList{startTime:1800,endTime:1900}>。

由于真实的所述单句的文本基本语义单位中应该只具有一种时间信息,所以需要过滤掉时间信息不合理的文本基本语义单位组,所以在执行完按照起止时间信息的组数,
25 分别形成组成所述单句的文本基本语义单位组的步骤之后,还包括如下步骤:

根据预定的计算方法,对每一所述文本基本语义单位组中,各个文本基本语义单位的所有起止时间信息进行筛选,确定组成所述单句的文本基本语义单位组。

在本实施例中,所述预定的计算方法,采用如下方式进行计算:计算各个所述文本基本语义单位组内,每一文本基本语义单位中的起始时间与所述文本基本语义单位的上
30 一个文本基本语义单位的终止时间之间的时间间距,获取各个所述文本基本语义单位组

中所述起始时间与所述终止时间的时间间距的和，将所述时间间距的和作为所述文本基本语义单位组的误差值。

需要说明的是，所述时间间距是指：每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距，由于在形成的组成所述单句的文本基本语义单位组时，所述文本基本语义单位的起始时间可能会小于上一个文本基本语义单位的终止时间，为了防止在计算误差值时出现的负数时间间距影响误差值的计算，需要获取所述时间间距的正值。

获取所述时间间距的正值的方法包括：取绝对值、取平方等，下面以采用取平方的方式获取所述时间间距的正值来进行说明。可以理解的，由于要获取每一文本基本语义单位中的起始时间与上一个文本基本语义单位的终止时间之间的时间间距，所以通过差平方的计算方式获取所述时间间距的正值。

具体的，所述预定的计算方法的数学算法为：

$$\text{误差值} = (\text{startTime2} - \text{endTime1})^2 + (\text{startTime3} - \text{endTime2})^2 \dots + (\text{startTime n} - \text{endTime n-1})^2$$

下面分别对上述4组时间集进行计算进行详细说明。(为了方便计算进行举例说明，在计算时以秒为单位进行计算)

$$\text{第一组: } (1.2-1.1)^2 + (1.4-1.3)^2 + (1.6-1.5)^2 + (1.2-1.7)^2 = 0.28$$

$$\text{第二组: } (1.2-1.1)^2 + (1.4-1.3)^2 + (1.6-1.5)^2 + (1.8-1.7)^2 = 0.04$$

$$\text{第三组: } (1.8-1.1)^2 + (1.4-1.9)^2 + (1.6-1.5)^2 + (1.2-1.7)^2 = 1$$

$$\text{第四组: } (1.8-1.1)^2 + (1.4-1.9)^2 + (1.6-1.5)^2 + (1.8-1.7)^2 = 0.76$$

在本实施例中，所述预设的阈值可以通过本领域的技术人员根据经验进行配置的较为合理的数值，或者所述预设的阈值为数值最小的误差值，在计算完误差值之后，对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组。

当所述预设的阈值为数值最小的误差值时，所述对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组，可以采用如下方式实现：保留误差值最小的组成所述单句的文本基本语义单位组，将其他的组成所述单句的文本基本语义单位组过滤掉。

需要说明的是，在对组成所述单句的文本基本语义单位组进行过滤时，可能会出现具有相同误差值的组成所述单句的文本基本语义单位组，这时在根据误差值进行过滤后还是无法获取单一的只具有一种时间信息的文本基本语义单位组，为了解决上述问题，

本申请的实施例提供了一种优选的实施方式，在优选方式下，在执行所述对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组的步骤之后，还需要计算保留的所述文本基本语义单位组内，每一文本基本语义单位中的起始时间大于所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数，获取该

5 次数最大的文本基本语义单位组。

下面以一个具体的实例进行说明。

若组成组成所述单句的文本基本语义单位组还包括第五组：

<word: “我”，timeList{startTime:1000,endTime:1100}>;

<word: “想”，timeList{startTime:1200,endTime:1300}>;

10 <word: “了”，timeList{startTime:1400,endTime:1500}>;

<word: “又”，timeList{startTime:1600,endTime:1700}>;

<word: “想”，timeList{startTime:1600,endTime:1700}>;

则第五组的误差值为：

$$(1.2-1.1)^2+(1.4-1.3)^2+(1.6-1.5)^2+(1.6-1.7)^2=0.04$$

15 经过对误差值进行过滤后，保留误差值最小的组成所述单句的文本基本语义单位组为第二组以及第五组，则还需对第二组和第五组的按照单句中文本基本语义单位的时间顺序进行合理性判断，即：判断保留的组成所述单句的每一文本基本语义单位中的起始时间大于所述单句中的上一个文本基本语义单位的终止时间的次数。

例如：第二组“想”字的起始时间大于“想”字上一个文本基本语义单位“我”的

20 终止时间；“了”字的起始时间大于“了”字上一个文本基本语义单位“想”的终止时间；“又”字的起始时间大于“又”字上一个文本基本语义单位“了”的终止时间；“想”字的起始时间大于“想”字上一个文本基本语义单位“又”的终止时间，则第二组的合理次数为4次；同样的道理，第五组的合理次数为3次，则获取合理次数为4次的组成所述单句的文本基本语义单位的时间集组。

25 作为一个优选实施方式，本申请实施例提供的自动生成配音文字的方法中，在执行步骤S103获取与所述音频信息对应的文本信息，并识别所述文本信息获取文本基本语义单位时，是从所述文本信息中，按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

作为一个优选实施方式，本申请实施例提供的自动生成配音文字的方法中，由于语音识别存在识别率，即：不一定能使所述音频信息精确无误的被识别出来，所以在步骤

30

S101 中对音频信息进行识别时，可能会有未被识别出的音频基本语义单位，而在执行步骤 S103，获取与所述音频信息对应的文本信息，并识别所述文本信息获取文本基本语义单位时，由于文本信息内的信息是计算机可以识别的字符串，则能够将所述文本信息内的每个基本语义单位进行识别并形成文本基本语义单位，所以在执行步骤 S105 将各个所述音频基本语义单位的起止时间信息记录到相应的所述文本基本语义单位中时，若所述音频基本语义单位的起止时间信息为空值，则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

可以理解的，若所述音频信息在识别过程中，具有未识别出的音频基本语义单位，即：所述音频基本语义单位为空，且该音频基本语义单位中的起止时间信息的取值也为空值，则在执行步骤 S105 将各个所述音频基本语义单位的起止时间信息记录到相应的所述文本基本语义单位中时，形成的文本基本语义单位的数目会大于语音识别出的音频基本语义单位的数目，则使未匹配上的所述文本基本语义单位中的起止时间信息的取值为空值。

例如：通过识别所述音频信息识别出的音频基本语义单位以及所述音频基本语义单位的时间信息为：

<word: “我”， {startTime:1000,endTime:1100} >;

<word: “想”， {startTime:1200,endTime:1300} >;

<word: “又”， {startTime:1600,endTime:1700} >;

对所述歌词文本中歌词的每个文本基本语义单位形成时间信息为空值的文本基本语义单位为：

<word: “我”， timeList{}>;

<word: “想”， timeList{}>;

<word: “了”， timeList{}>;

<word: “又”， timeList{}>;

由于所述音频信息进行识别后只识别出了“我”、“想”和“又”，而对所述歌词文本中歌词的文本基本语义单位进行识别后形成的文本基本语义单位为：“我”、“想”、“了”、“又”，则将上述音频基本语义单位的时间信息放入到相应的文本基本语义单位中：

<word: “我”， timeList{startTime:1000,endTime:1100}>;

<word: “想”， timeList{startTime:1200,endTime:1300}>;

<word: “了”，timeList{}>;

<word: “又”，timeList{startTime:1600,endTime:1700}>。

作为一个优选实施方式，本申请实施例提供的自动生成配音文字的方法中，在执行步骤 S107-1 针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位时，

5 若具有取值为空值的文本基本语义单位时，在所述确定组成所述单句的文本基本语义单位组的步骤之后，为了使每一文本基本语义单位都具有起止时间信息，按照预定的推算方式，对取值为空值的所述文本基本语义单位推算起止时间信息。

所述预定的推算方式，包括：

计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息；

10 将取值为空值的所述文本基本语义单位的上一个基本语义单位中的终止时间，放入取值为空值的所述文本基本语义单位的起始时间中；

将所述终止时间加上所述平均时间信息后，放入取值为空值的所述文本基本语义单位的终止时间中。

在本实施例中，所述计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息，可以采用如下方式实现：将组成所述单句的每一文本基本语义单位中的终止时间减去起始时间，获得每一文本基本语义单位在音频信息中的播放时间，并根据该单句中

15 文本基本语义单位的播放时间的和除以该单句中文本基本语义单位的数量计算组成所述单句的文本基本语义单位的平均时间信息。

可以理解的，由于所述文本基本语义单位是按照文本信息的单句中每个基本语义单位

20 位的顺序形成的，所以能通过取值为空值的文本基本语义单位的上一文本基本语义单位的时间信息中的终止时间进行时间估算，将取值为空值的文本基本语义单位的上一个文本基本语义单位中的终止时间，放入取值为空值的文本基本语义单位的起始时间中，即：将与取值为空值的文本基本语义临近的文本基本语义的终止时间作为取值为空值的文本基本语义的起始时间。

25 在确定取值为空值的文本基本语义的起始时间后，根据该单句中每个文本基本语义单位在音频信息中的平均播放时间确定取值为空值的文本基本语义单位的终止时间，即：将取值为空值的文本基本语义单位已确定的起始时间加上所述平均时间信息后，放入取值为空值的文本基本语义的终止时间中。

需要说明的是，由于执行步骤 S103 获取与所述音频信息对应的文本信息，并识别所

30 述文本信息获取文本基本语义单位时，是从所述文本信息中，按照每句内的每个字的顺

序进行识别获取所述文本信息中的文本基本语义单位的，则对取值为空值的文本基本语义单位推算起止时间信息还可以采用另一种方式实现：直接以取值为空的文本基本语义单位的上一文本基本语义单位的时间信息中的终止时间以及以取值为空的文本基本语义单位的下一文本基本语义单位的时间信息中的开始时间，分别作为该取值为空的文本基本语义单位的时间信息中的开始时间以及终止时间。

可以理解的，由于所述文本基本语义单位是按照文本单句中每个文本基本语义单位的顺序形成的，所以取值为空值的文本基本语义单位的基本语义单位是出现在与其临近的前后文本基本语义单位之间的，所以能通过上一文本基本语义单位的时间信息中的结束时间以及下一文本基本语义单位的时间信息中的开始时间对取值为空值的文本基本语义单位进行时间估算。

在上述的实施例中，提供了一种自动生成配音文字的方法，与上述自动生成配音文字的方法相对应的，本申请还提供了一种自动生成配音文字的装置。由于装置的实施例基本相似于方法的实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。下述描述的装置实施例仅仅是示意性的。所述自动生成配音文字的装置实施例如下：

请参考图 3，其示出了根据本申请的实施例提供的自动生成配音文字的装置的示意图。

所述自动生成配音文字的装置，包括：音频识别单元 301、文本识别单元 303、时间写入单元 305 以及配音文字生成单元 307；

所述音频识别单元 301，用于对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息；

所述文本识别单元 303，用于获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位；

所述时间写入单元 305，用于将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中；

所述配音文字生成单元 307，用于对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字。

可选的，所述时间记录单元，包括：文本语义获取子单元、时间信息确定子单元以及配音文字生成子单元；

所述文本语义获取子单元，用于针对所述文本信息中每一单句，获取组成所述单句

的文本基本语义单位；

所述时间信息确定子单元，用于根据已获取的所述文本基本语义单位中记录的起止时间信息确定所述单句的起止时间信息；

所述配音文字生成子单元，用于将确定了起止时间信息的所述单句进行整合，形成
5 对应所述音频信息，且具有每一单句的起止时间信息的配音文字。

可选的，所述时文本语义获取子单元，具体用于针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位时，若所述文本基本语义单位中记录了至少两组起止时间信息，则按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组。

10 可选的，所述的自动生成配音文字的装置，还包括：文本语义筛选子单元；

所述文本语义筛选子单元，用于在所述按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组之后，根据预定的计算方法，对每一所述文本基本语义单位组中，各个文本基本语义单位的所有起止时间信息进行筛选，确定组成所述单句的文本基本语义单位组。

15 可选的，所述时间集组筛选子单元，包括：误差计算子单元；

所述误差计算子单元，用于计算各个所述文本基本语义单位组内，每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距，获取各个所述文本基本语义单位组中所述起始时间与所述终止时间的
时间间距的和，将所述时间间距的和作为所述文本基本语义单位组的误差值。

20 可选的，所述时间集组筛选子单元，还包括：过滤子单元；

所述过滤子单元，用于对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组。

可选的，所述时间集组筛选子单元，还包括：时间次数计算子单元；

所述时间次数计算子单元，用于在所述保留误差值低于预设的阈值的文本基本语义
25 单位组之后，计算保留的所述文本基本语义单位组内，每一文本基本语义单位中的起始时间大于所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数，获取该次数最大的文本基本语义单位组。

可选的，所述文本识别单元 303，具体用于从所述文本信息中，按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

30 可选的，所述时间写入单元 305，具体用于在将各个所述音频基本语义单位的起止

时间信息，记录到相应的所述文本基本语义单位中时，若所述音频基本语义单位的起止时间信息为空值，则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

可选的，所述的自动生成配音文字的装置，还包括：

- 5 时间推算单元，用于在所述确定组成所述单句的文本基本语义单位组之后，按照预定的推算方式，对取值为空值的所述文本基本语义单位推算起止时间信息

可选的，所述时间推算单元，包括：

平均时间计算子单元，用于计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息；

- 10 起始时间写入子单元，用于将取值为空值的所述文本基本语义单位，上一个文本基本语义单位中的终止时间，放入取值为空值的所述文本基本语义单位的起始时间中；

终止时间写入子单元，用于将所述终止时间加上所述平均时间信息后，放入取值为空值的所述文本基本语义单位的终止时间中。

- 15 在上述的实施例中，提供了一种自动生成配音文字的方法以及一种自动生成配音文字的装置，此外，本申请还提供了一种电子设备；所述电子设备实施例如下：

请参考图 4，其示出了根据本申请的实施例提供的电子设备的示意图。

所述电子设备，包括：显示器 401；处理器 403；存储器 405；

- 20 所述存储器 405，用于存储配音文字生成程序，所述程序在被所述处理器读取执行时，执行如下操作：对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息；获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位；将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中；对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字。

- 25 在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器 (RAM) 和/或非易失性内存等形式，如只读存储器 (ROM) 或闪存(flash RAM)。内存是计算机可读介质的示例。

- 30 1、计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数

据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁
5 带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非暂存电脑可读媒体 (transitory media)，如调制的数据信号和载波。

2、本领域技术人员应明白，本申请的实施例可提供为方法、系统或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例
10 的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本申请虽然以较佳实施例公开如上，但其并不是用来限定本申请，任何本领域技术人员在不脱离本申请的精神和范围内，都可以做出可能的变动和修改，因此本申请的保
15 护范围应当以本申请权利要求所界定的范围为准。

权 利 要 求 书

1.一种自动生成配音文字的方法，其特征在于，包括：

对音频信息进行识别，获取识别出的各个音频基本语义单位的起止时间信息；

获取与所述音频信息对应的文本信息，并识别所述文本信息，从而获取文本基本语义单位；

将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中；

对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字。

2.根据权利要求 1 所述的自动生成配音文字的方法，其特征在于，所述对记录了所述起止时间信息的所述文本基本语义单位进行处理，生成对应所述音频信息的配音文字，包括：

针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位；

根据已获取的所述文本基本语义单位中记录的起止时间信息，确定所述单句的起止时间信息；

将确定了起止时间信息的所述单句进行整合，形成对应所述音频信息，且具有每一单句的起止时间信息的配音文字。

3.根据权利要求 2 所述的自动生成配音文字的方法，其特征在于，所述针对所述文本信息中每一单句，获取组成所述单句的文本基本语义单位时，若所述文本基本语义单位中记录了至少两组起止时间信息，则按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组。

4.根据权利要求 3 所述的自动生成配音文字的方法，其特征在于，在所述按照起止时间信息的组数，分别形成组成所述单句的文本基本语义单位组的步骤之后，包括：

根据预定的计算方法，对每一所述文本基本语义单位组中，各个文本基本语义单位的所有起止时间信息进行筛选，确定组成所述单句的文本基本语义单位组。

5.根据权利要求 4 所述的自动生成配音文字的方法，其特征在于，所述预定的计算方法，包括：

计算各个所述文本基本语义单位组内，每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距，获取各个所述文本基本语义单位组中所述起始时间与所述终止时间的时间间距的和，将所述时间间

距的和作为所述文本基本语义单位组的误差值。

6.根据权利要求 5 所述的自动生成配音文字的方法,其特征在于,所述对每一所述文本基本语义单位组中,各个文本基本语义单位的所有起止时间信息进行筛选,确定组成所述单句的文本基本语义单位组,包括:

5 对各个所述文本基本语义单位组进行过滤,保留误差值低于预设的阈值的文本基本语义单位组。

7.根据权利要求 6 所述的自动生成配音文字的方法,其特征在于,在所述保留误差值低于预设的阈值的文本基本语义单位组的步骤之后,包括:

10 计算保留的所述文本基本语义单位组内,每一文本基本语义单位中的起始时间大于所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数,获取该次数最大的文本基本语义单位组。

8.根据权利要求 4-7 中任意一项所述的自动生成配音文字的方法,其特征在于,所述识别所述文本信息,从而获取文本基本语义单位包括:

15 从所述文本信息中,按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

9.根据权利要求 8 所述的自动生成配音文字的方法,其特征在于,在将各个所述音频基本语义单位的起止时间信息,记录到相应的所述文本基本语义单位中时,若所述音频基本语义单位的起止时间信息为空值,则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

20 10.根据权利要求 9 所述的自动生成配音文字的方法,其特征在于,在所述确定组成所述单句的文本基本语义单位组的步骤之后,包括:

按照预定的推算方式,对取值为空值的所述文本基本语义单位推算起止时间信息。

11.根据权利要求 10 所述的自动生成配音文字的方法,其特征在于,所述预定的推算方式,包括:

25 计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息;

将取值为空值的所述文本基本语义单位,上一个文本基本语义单位中的终止时间,放入取值为空值的所述文本基本语义单位的起始时间中;

将所述终止时间加上所述平均时间信息后,放入取值为空值的所述文本基本语义单位的终止时间中。

30 12.一种自动生成配音文字的装置,其特征在于,包括:

音频识别单元,用于对音频信息进行识别,获取识别出的各个音频基本语义单位的起止时间信息;

文本识别单元,用于获取与所述音频信息对应的文本信息,并识别所述文本信息,从而获取文本基本语义单位;

- 5 时间写入单元,用于将各个所述音频基本语义单位的起止时间信息,记录到相应的所述文本基本语义单位中;

配音文字生成单元,用于对记录了所述起止时间信息的所述文本基本语义单位进行处理,生成对应所述音频信息的配音文字。

- 10 13.根据权利要求 12 所述的自动生成配音文字的装置,其特征在于,所述配音文字生成单元,包括:

文本语义获取子单元,用于针对所述文本信息中每一单句,获取组成所述单句的文本基本语义单位;

时间信息确定子单元,用于根据已获取的所述文本基本语义单位中记录的起止时间信息确定所述单句的起止时间信息;

- 15 配音文字生成子单元,用于将确定了起止时间信息的所述单句进行整合,形成对应所述音频信息,且具有每一单句的起止时间信息的配音文字。

- 20 14.根据权利要求 13 所述的自动生成配音文字的装置,其特征在于,所述文本语义获取子单元,具体用于针对所述文本信息中每一单句,获取组成所述单句的文本基本语义单位时,若所述文本基本语义单位中记录了至少两组起止时间信息,则按照起止时间信息的组数,分别形成组成所述单句的文本基本语义单位组。

- 15.根据权利要求 14 所述的自动生成配音文字的装置,其特征在于,还包括:

- 25 文本语义筛选子单元,用于在所述按照起止时间信息的组数,分别形成组成所述单句的文本基本语义单位组之后,根据预定的计算方法,对每一所述文本基本语义单位组中,各个文本基本语义单位的所有起止时间信息进行筛选,确定组成所述单句的文本基本语义单位组。

- 16.根据权利要求 15 所述的自动生成配音文字的装置,其特征在于,所述文本语义筛选子单元,包括:

- 30 误差计算子单元,用于计算各个所述文本基本语义单位组内,每一文本基本语义单位中的起始时间与所述文本基本语义单位的上一个文本基本语义单位的终止时间之间的时间间距,获取各个所述文本基本语义单位组中所述起始时间与所述终止时间的时间

间距的和，将所述时间间距的和作为所述文本基本语义单位组的误差值。

17.根据权利要求 15 所述的自动生成配音文字的装置，其特征在于，所述文本语义筛选子单元，还包括：

5 过滤子单元，用于对各个所述文本基本语义单位组进行过滤，保留误差值低于预设的阈值的文本基本语义单位组。

18.根据权利要求 17 所述的自动生成配音文字的装置，其特征在于，所述文本语义筛选子单元，还包括：

10 时间次数计算子单元，用于在所述保留误差值低于预设的阈值的文本基本语义单位组的之后，计算保留的所述文本基本语义单位组内，每一文本基本语义单位中的起始时间大于所述文本基本语义单位的上一个文本基本语义单位的终止时间的次数，获取该次数最大的文本基本语义单位组。

19.根据权利要求 15- 18 中任意一项所述的自动生成配音文字的装置，其特征在于，所述文本识别单元，具体用于从所述文本信息中，按照每句内的每个字的顺序进行识别获取所述文本信息中的文本基本语义单位。

15 20.根据权利要求 19 所述的自动生成配音文字的装置，其特征在于，所述时间写入单元，具体用于在将各个所述音频基本语义单位的起止时间信息，记录到相应的所述文本基本语义单位中时，若所述音频基本语义单位的起止时间信息为空值，则使与所述音频基本语义单位相应的所述文本基本语义单位的取值为空值。

21.根据权利要求 20 所述的自动生成配音文字的装置，其特征在于，还包括：

20 时间推算单元，用于在所述确定组成所述单句的文本基本语义单位组之后，按照预定的推算方式，对取值为空值的所述文本基本语义单位推算起止时间信息。

22.根据权利要求 21 所述的自动生成配音文字的装置，其特征在于，所述时间推算单元，包括：

25 平均时间计算子单元，用于计算所述文本基本语义单位组中的文本基本语义单位的平均时间信息；

起始时间写入子单元，用于将取值为空值的所述文本基本语义单位，上一个文本基本语义单位中的终止时间，放入取值为空值的所述文本基本语义单位的起始时间中；

终止时间写入子单元，用于将所述终止时间加上所述平均时间信息后，放入取值为空值的所述文本基本语义单位的终止时间中。

30 23.一种电子设备，其特征在于，所述电子设备包括：

显示器;

处理器;

- 存储器, 用于存储配音文字生成程序, 所述程序在被所述处理器读取执行时, 执行如下操作: 对音频信息进行识别, 获取识别出的各个音频基本语义单位的起止时间信息; 获取与
5 所述音频信息对应的文本信息, 并识别所述文本信息, 从而获取文本基本语义单位; 将各个所述音频基本语义单位的起止时间信息, 记录到相应的所述文本基本语义单位中; 对记录了所述起止时间信息的所述文本基本语义单位进行处理, 生成对应所述音频信息的配音文字。

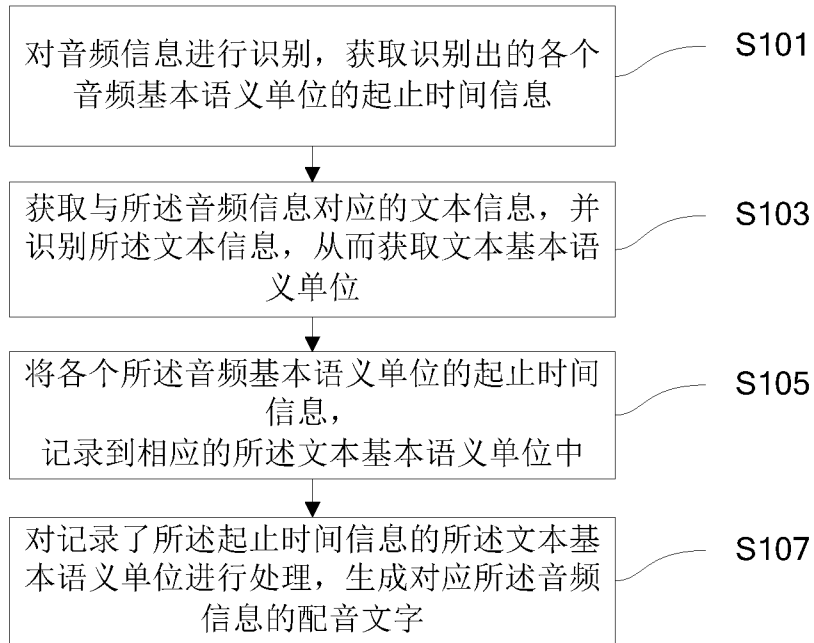


图 1

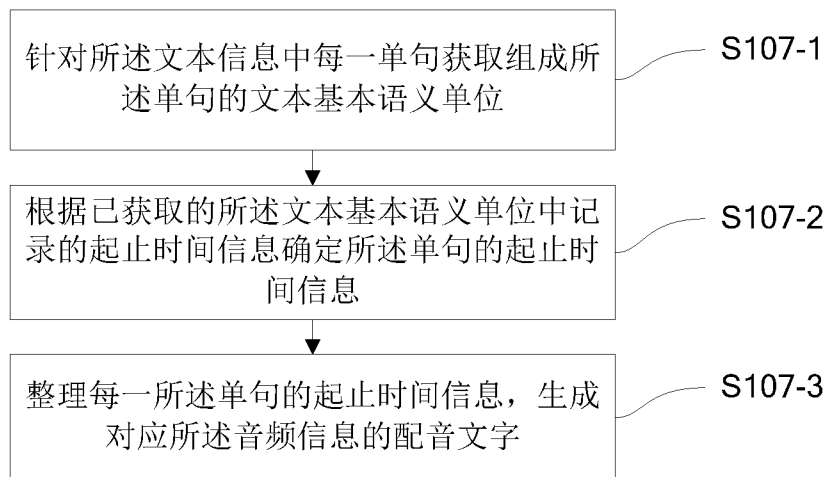


图 2

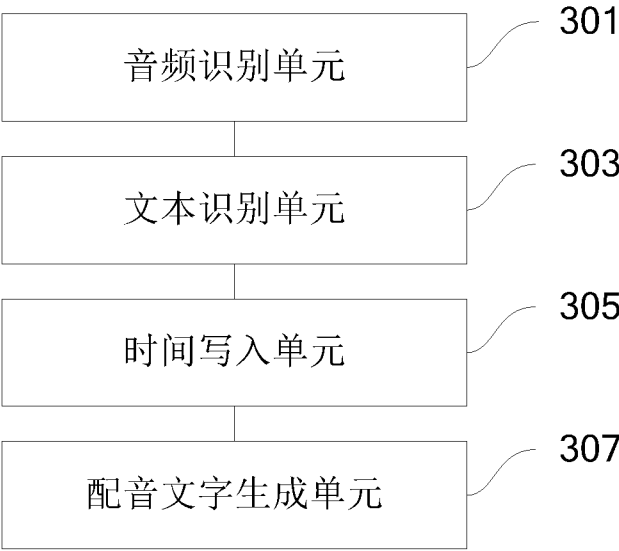


图 3

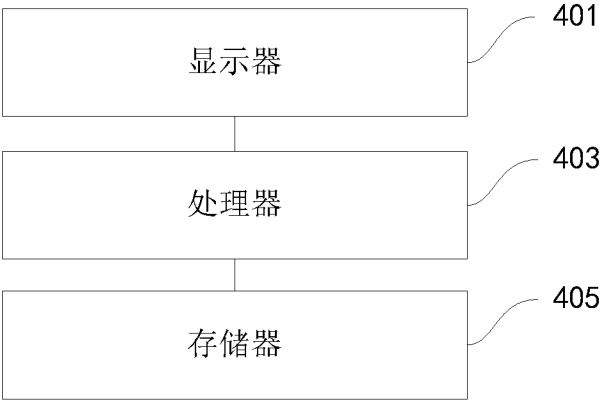


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/115194

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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)

G06F

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

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

WPI; EPODOC; CNPAT; CNKI; IEEE: 配音, 文字, 音频, 识别, 起止, 时间, 基本, 语义, 文本, dub, letter, audio, frequency, identify, time, text, start, end

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104599693 A (INTERNET OF LANGUAGES WUHAN INFORMATION TECHNOLOGY CO., LTD.) 06 May 2015 (06.05.2015), description, paragraphs [0035]-[0055], and figure 1	1-23
A	CN 1949227 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 18 April 2007 (18.04.2007), entire document	1-23
A	CN 101616264 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 30 December 2009 (30.12.2009), entire document	1-23
A	US 6345252 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 05 February 2002 (05.02.2002), entire document	1-23

☐ 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 05 February 2018	Date of mailing of the international search report 07 March 2018
Name and mailing address of the ISA 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 LI, Meng Telephone No. (86-10) 53961536

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/115194

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104599693 A	06 May 2015	None	
CN 1949227 A	18 April 2007	None	
CN 101616264 A	30 December 2009	None	
US 6345252 B1	05 February 2002	EP 1043665 A2	11 October 2000
		KR 20010006942 A	26 January 2001
		JP 2000348064 A	15 December 2000
		TW 469385 B	21 December 2001
		CN 1270361 A	18 October 2000

国际检索报告

国际申请号

PCT/CN2017/115194

A. 主题的分类 G06F 17/30 (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) G06F 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)) WPI; EPODOC; CNPAT; CNKI; IEEE: 配音, 文字, 音频, 识别, 起止, 时间, 基本, 语义, 文本, dub, letter, audio, frequency, identify, time, text, start, end																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 104599693 A (语联网武汉信息技术有限公司) 2015年 5月 6日 (2015 - 05 - 06) 说明书第 [0035] - [0055] 段, 附图 1</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 1949227 A (北京搜狗科技发展有限公司) 2007年 4月 18日 (2007 - 04 - 18) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 101616264 A (中国科学院自动化研究所) 2009年 12月 30日 (2009 - 12 - 30) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>US 6345252 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2002年 2月 5日 (2002 - 02 - 05) 全文</td> <td>1-23</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 104599693 A (语联网武汉信息技术有限公司) 2015年 5月 6日 (2015 - 05 - 06) 说明书第 [0035] - [0055] 段, 附图 1	1-23	A	CN 1949227 A (北京搜狗科技发展有限公司) 2007年 4月 18日 (2007 - 04 - 18) 全文	1-23	A	CN 101616264 A (中国科学院自动化研究所) 2009年 12月 30日 (2009 - 12 - 30) 全文	1-23	A	US 6345252 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2002年 2月 5日 (2002 - 02 - 05) 全文	1-23
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
A	CN 104599693 A (语联网武汉信息技术有限公司) 2015年 5月 6日 (2015 - 05 - 06) 说明书第 [0035] - [0055] 段, 附图 1	1-23															
A	CN 1949227 A (北京搜狗科技发展有限公司) 2007年 4月 18日 (2007 - 04 - 18) 全文	1-23															
A	CN 101616264 A (中国科学院自动化研究所) 2009年 12月 30日 (2009 - 12 - 30) 全文	1-23															
A	US 6345252 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2002年 2月 5日 (2002 - 02 - 05) 全文	1-23															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
<table border="0"> <tr> <td style="vertical-align: top;"> * 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 </td> <td style="vertical-align: top;"> “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 </td> </tr> </table>			* 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													
* 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 2018年 2月 5日		国际检索报告邮寄日期 2018年 3月 7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		授权官员 李萌 电话号码 (86-10) 53961536															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/115194

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	104599693	A	2015年 5月 6日	无			
CN	1949227	A	2007年 4月 18日	无			
CN	101616264	A	2009年 12月 30日	无			
US	6345252	B1	2002年 2月 5日	EP	1043665	A2	2000年 10月 11日
				KR	20010006942	A	2001年 1月 26日
				JP	2000348064	A	2000年 12月 15日
				TW	469385	B	2001年 12月 21日
				CN	1270361	A	2000年 10月 18日