(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
2019 年 6 月 27 日 (27.06.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/119552 A1

(51) 国际专利分类号:
G06F 17/28 (2006.01)

(21) 国际申请号: PCT/CN2018/072007

(22) 国际申请日: 2018 年 1 月 9 日 (09.01.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711388000.3 2017年12月20日 (20.12.2017) CN

(71) 申请人: 深圳市沃特沃德股份有限公司 (SHENZHEN WATER WORLD CO., LTD) [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1079号花园城数码大厦B座503, 602, Guangdong 518000 (CN)。

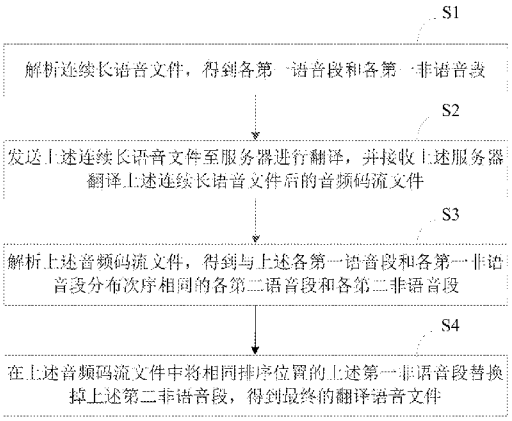
(72) 发明人: 郑勇 (ZHENG, Yong); 中国广东省深圳市南山区蛇口南海大道1079号花园城数码大厦B座503,602, Guangdong 518000 (CN)。 金志军 (JIN, Zhijun); 中国广东省深圳市南山区蛇口南海大道1079号花园城数码大厦B座503,602, Guangdong 518000 (CN)。 王文祺 (WANG, Wenqi); 中国广东省深圳市南山区蛇口南海大道1079号花园城数码大厦B座503,602, Guangdong 518000 (CN)。

(74) 代理人: 深圳市明日今典知识产权代理事务所（普通合伙） (SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL)); 中国广东省深圳市南山区南头街道智恒产业园E区01B栋405, 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, 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,

(54) **Title:** METHOD FOR TRANSLATING CONTINUOUS LONG SPEECH FILE, AND TRANSLATION MACHINE

(54) 发明名称: 连续长语音文件的翻译方法与翻译机



S1 Parse a continuous long speech file to obtain first speech segments and first non-speech segments

S2 Send the continuous long speech file to a server for translation, and receive an audio code stream file obtained after the server translates the continuous long speech file

S3 Parse the audio code stream file to obtain second speech segments and second non-speech segments having the same distribution sequence as that of the first speech segments and the first non-speech segments

S4 Replace the second non-speech segments with the first non-speech segments at the same ranking positions in the audio code stream file to obtain a final translated speech file

图 1

(57) **Abstract:** A method for translating a continuous long speech file, and a translation machine. The method for translating a continuous long speech file comprises: parsing a continuous long speech file to obtain first speech segments and first non-speech segments (S1), wherein the first speech segments and the first non-speech segments are distributed according to a time sequence generated in the continuous long speech; sending the continuous long speech file to a server for translation, and receiving an audio code stream file obtained after the server translates the continuous long speech file (S2); parsing the audio code stream file to obtain second speech segments and second non-speech segments having the same distribution sequence as that of the first speech segments and the first non-speech segments (S3); and replacing the second non-speech segments with the first non-speech segments at the same ranking positions in the audio code stream file to obtain a final translated speech file (S4). The rhythm, language background sound, and sentence natural interval of a continuous long speech file are reserved, and the use experience of users is improved.

WO 2019/119552 A1

[见续页]

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, 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：连续长语音文件的翻译方法与翻译机，其中，连续长语音文件的翻译方法，包括：解析连续长语音文件，得到各第一语音段和各第一非语音段（S1），其中，各第一语音段和各第一非语音段按照在连续长语音中生成的时序分布；发送连续长语音文件至服务器进行翻译，并接收服务器翻译连续长语音文件后的音频码流文件（S2）；解析音频码流文件，得到与各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段（S3）；在音频码流文件中将相同排序位置的第一非语音段替换掉第二非语音段，得到最终的翻译语音文件（S4）。保留了连续长语音文件的节奏、语言背景音和语句自然间隔，提高用户的使用体验。

连续长语音文件的翻译方法与翻译机

技术领域

本发明涉及到电子翻译技术，特别是涉及到连续长语音文件的翻译方法与翻译机。

背景技术

在电子翻译领域，教育、录音翻译等应用场景下的连续长语音文件通过语音识别引擎、翻译引擎、合成引擎协同作用，得到翻译后的语音文件，并通过翻译机等电子设备终端输出，方便了不同语言用户的相互交流沟通，为人们的生活提供了极大的便利。但现有翻译引擎对连续长语音文件中的每个语音段翻译输出的语音文件没有背景噪音信息，而且语音文件中的语句间隔是预先设定的固定输出间隔，使得翻译后的语音文件脱离了原始连续长语音文件的节奏和语句自然间隔，失去了原始连续长语音文件的语言环境和情趣，用户体验性能差。

因此，现有技术还有待改进。

技术问题

本发明的主要目的为提供一种连续长语音文件的翻译方法，旨在解决现有翻译技术中不能保留原始连续长语音文件的节奏和语句自然间隔的技术问题。

技术解决方案

本发明提出一种连续长语音文件的翻译方法，包括：

解析连续长语音文件，得到各第一语音段和各第一非语音段，其中，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布；

发送上述连续长语音文件至服务器进行翻译，并接收上述服务器

翻译上述连续长语音文件后的音频码流文件；

解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段；

在上述音频码流文件中将相同排序位置的上述第一非语音段替换掉上述第二非语音段，得到最终的翻译语音文件。

优选地，上述解析连续长语音文件，得到各第一语音段和各第一非语音段的步骤，包括：

通过语音活动检测分析技术处理上述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态；

根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

优选地，上述根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段的步骤，包括：

将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段。

优选地，上述将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段的步骤之后，包括：

提取各上述第一非语音段；

将各上述第一非语音段按照在上述连续长语音中生成的时序存储于非语音段缓存区；

优选地，上述发送上述连续长语音文件至服务器进行翻译，并接收上述服务器翻译上述连续长语音文件后的音频码流文件的步骤，包括：

将连续长语音文件发送至语音识别服务器；

接收上述语音识别服务器反馈的与上述连续长语音文件对应的第一文本文件；

将上述第一文本文件发送至翻译服务器；

接收上述翻译服务器反馈的翻译上述第一文本文件后的指定语种的第二文本文件；

将上述第二文本文件发送至语音合成服务器；

接收上述语音合成服务器转换上述第二文本文件后的音频码流文件。

优选地，上述解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段的步骤，包括：

将上述第一文本文件的第一字符串信息与上述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系；

通过语音活动检测分析技术处理上述音频码流文件，获得第二语音帧和第二非语音帧的排布状态；

根据上述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段；

根据上述第一类一一对应关系建立各上述第一语音段与各上述第二语音段的第二类一一对应关系；

根据上述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序，获得与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

本发明还提供了一种翻译机，包括：

第一解析模块，用于解析连续长语音文件，得到各第一语音段和各第一非语音段，其中，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布；

发送接收模块，用于发送上述连续长语音文件至服务器进行翻译，并接收上述服务器翻译上述连续长语音文件后的音频码流文件；

第二解析模块，用于解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段；

替换模块，用于在上述音频码流文件中将相同排序位置的上述第一非语音段替换掉上述第二非语音段，得到最终的翻译语音文件。

优选地，上述第一解析模块，包括：

第一处理单元，用于通过语音活动检测分析技术处理上述连续长

语音文件，获得第一语音帧和第一非语音帧的排布状态；

第一获得单元，用于根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

优选地，上述第一获得单元，包括：

合成子单元，用于根据上述排布状态将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段。

优选地，上述第一获得单元，还包括：

提取子单元，用于提取各上述第一非语音段；

存储子单元，用于将各上述第一非语音段按照在上述连续长语音中生成的时序存储于非语音段缓存区。

优选地，上述发送接收模块，包括：

第一发送单元，用于将连续长语音文件发送至语音识别服务器；

第一接收单元，用于接收上述语音识别服务器反馈的与上述连续长语音文件对应的第一文本文件；

第二发送单元，用于将上述第一文本文件发送至翻译服务器；

第二接收单元，用于接收上述翻译服务器反馈的翻译上述第一文本文件后的指定语种的第二文本文件；

第三发送单元，用于将上述第二文本文件发送至语音合成服务器；

第三接收单元，用于接收上述语音合成服务器转换上述第二文本文件后的音频码流文件。

优选地，上述第二解析模块，包括：

分析单元，用于将上述第一文本文件的第一字符串信息与上述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系；

第二处理单元，用于通过语音活动检测分析技术处理上述音频码流文件，获得第二语音帧和第二非语音帧的排布状态；

第二获得单元，用于根据上述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段；

建立单元，用于根据上述第一类一一对应关系建立各上述第一语音段与各上述第二语音段的第二类一一对应关系；

第三获得单元，用于根据上述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序，获得与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

有益效果

本发明通过将原始连续长语音文件区分为语音段和非语音段，并保留与原始连续长语音文件相同的非语音段，使得翻译后的音频流码文件与原始连续长语音文件相比，具有几乎相同的节奏、语言背景音和语句自然间隔，增加机器翻译的生动活力感，提高用户的使用体验。

附图说明

图1 本发明一实施例的连续长语音文件的翻译方法流程示意图；

图2 本发明一实施例的步骤S1的流程示意图；

图3 本发明一实施例的步骤S11的流程示意图；

图4 本发明一实施例的步骤S2的流程示意图；

图5 本发明一实施例的步骤S3的流程示意图；

图6 本发明一实施例的翻译机的结构示意图；

图7 本发明一实施例的第一解析模块的结构示意图；

图8 本发明一实施例的第一获得单元的结构示意图；

图9 本发明一实施例的发送接收模块的结构示意图；

图10 本发明一实施例的第二解析模块的结构示意图。

本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

本发明的最佳实施方式

应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

参照图 1，本发明一实施例的连续长语音文件的翻译方法，包括：

S1：解析上述连续长语音文件，得到各第一语音段和各第一非语

音段，其中，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布。

本实施例的终端设备以翻译机为例。本步骤中通过解析上述连续长语音文件，得到各第一语音段与各第一非语音段的交替间隔排布的数据文件，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布，比如表示为：第一语音段1、第一非语音段1、第一语音段2、第一非语音段2、第一语音段3、第一非语音段3、……、第一语音段N、第一非语音段N。

S2：发送上述连续长语音文件至服务器进行翻译，并接收上述服务器翻译上述连续长语音文件后的音频码流文件。

本步骤指连续长语音文件依次经翻译机发送到语音识别服务器、翻译服务器以及语音合成服务器进行翻译的过程。本实施例的音频码流文件指翻译连续长语音文件后得到的对应的音频数据，包括语音数据、非语音数据。

S3：解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

本实施例的音频码流文件是通过一一对应翻译连续长语音文件后得到的音频数据，所以音频码流文件中的各第二语音段和各第二非语音段与各第一语音段和各第一非语音段分布次序相同。

S4：在上述音频码流文件中将分布次序相同的相同排序位置的上述第一非语音段替换掉上述第二非语音段，得到最终的翻译语音文件。

本实施例在与连续长语音文件分布次序相同的音频码流文件中，将相同排序位置上的第二非语音文件替换为第一非语音段，并将第一非语音段与翻译后的音频码流文件合为一体，使得最终的翻译语音文件与原始连续长语音文件具有相同的节奏、语言背景音和语句自然间隔，增加机器翻译的生动活力感，提高用户的使用体验。

参照图2，进一步地，本发明一实施例中，步骤S1，包括：

S10：通过语音活动检测分析技术处理上述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态。

本实施例通过翻译机对连续长语音文件做VAD（Voice Activity

Detection，语音活动检测分析），将上述连续长语音文件中的第一语音段和第一非语音段区分开，以便于后续操作。举例地，按帧处理连续长语音文件，每帧时长根据语音信号特点来设定，比如GSM的20ms的时间为帧长度。首先通过VAD检测出连续长语音文件里的每个第一语音段的开始和结束，并通过算法处理得到连续长语音文件的各第一语音段的时间长度，如采用GSM通信系统中的ETSI VAD算法或者G.729 Annex B VAD算法，将VAD提取的连续长语音文件的参数特征值与门限值比较，以区分开上述第一语音段和上述第一非语音段。本步骤的排布状态是指，经语音活动检测分析处理后，连续长语音文件变为连续排布的第一语音帧与连续排布的第一非语音帧交替分布的数据文件时的排布信息。

S11：根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

本实施例中，将通过VAD区分开的各第一语音段和各第一非语音段，分别采用不同编码标记，以便识别。

参照图3，进一步地，本发明一实施例中，步骤S11，包括：

S112：将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段。

本实施例通过VAD的判决结果辨别第一语音帧和第一非语音帧，比如：判决结果是1，就是第一语音帧；判决结果是0，就是背景噪声帧（即第一非语音帧），并把连续长语音文件中的语句，变为经VAD处理后的连续排列的第一语音帧1到第一语音帧m合并成的第一语音段1，连续排列的第一非语音帧1到第一非语音帧k合并成的第一非语音段1，依次处理，将连续长语音文件变为连续排列的第一语音段1至第一语音段N，与连续排列的第一非语音段1至第一非语音段N一一交替分布的数据文件。

进一步地，本发明一实施例中，步骤S112之后，包括

S113：提取各上述第一非语音段。

根据各第一语音段和各第一非语音段的不同编码标记，把连续长语音文件里的第一非语音段提取出来，比如：将按照依次编码为T1、

T2…Tn 的第一非语音段 1、2...N 提取出来。

S114：将各上述第一非语音段按照在上述连续长语音中生成的时序存储于非语音段缓存区。

本步骤的非语音段缓存区设定于翻译机内存的指定区域，以便按照在上述连续长语音中生成时序中第一非语音段的相同排列次序，依次将翻译后的音频码流文件与第一非语音段合为一体后再输出。

参照图 4，本发明一实施例中，步骤 S2，包括：

S20：将连续长语音文件发送至语音识别服务器。

S21：接收上述语音识别服务器反馈的与上述连续长语音文件对应的第一文本文件。

本步骤通过语音识别服务器得到的与连续长语音文件相对应的第一文本文件。

S22：将上述第一文本文件发送至翻译服务器。

S23：接收上述翻译服务器反馈的翻译上述第一文本文件后的指定语种的第二文本文件。

本步骤通过翻译服务器对第一文本文件进行翻译，以形成指定语种的第二文本文件。比如：中文翻译成英文，中文的第一文本文件与英文的第二文本文件为一一对应关系，即中文一句话翻译成英文的一句话，每句话一一对应。

S24：将上述第二文本文件发送至语音合成服务器。

S25：接收上述语音合成服务器转换上述第二文本文件后的音频码流文件。

本步骤通过将第二文本文件顺序送入语音合成服务器，以便将第二文本文件顺序转换为指定语种的音频码流文件，比如，英文的音频码流文件。

参照图 5，进一步地，本发明一实施例中，步骤 S3，包括：

S30：将上述第一文本文件的第一字符串信息与上述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系。

本步骤通过对比分析文本文件的字符串信息，对第一文本文件和第二文本文件上的字符串组成的每句话分别做好标记，比如：第 1 句

话、第 2 句话、……、第 N 句话，以便通过对应比较获得第二文本文件与第一文本文件的一一对应关系。

S31：通过语音活动检测分析技术处理上述音频码流文件，获得第二语音帧和第二非语音帧的排布状态。

本步骤通过 VAD 处理音频码流文件，以区分开音频码流文件中的第二语音段和第二非语音段，每个第二语音段对应第二文本文件中的一句话，N 个第二非语音段的时间长度相同。

S32：根据上述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段。

本实施例中，将通过 VAD 区分开的各第二语音段和各第二非语音段，同样分别采用不同编码标记，以便识别。

S33：根据上述第一类一一对应关系建立各上述第一语音段与各上述第二语音段的第二类一一对应关系。

每个第一语音段对应第一文本文件中的一句话，每个第二语音段对应第二文本文件中的一句话，根据第二文本文件与第一文本文件的一一对应关系，找到各第一语音段与各第二语音段的一一对应关系，以便确定各第一非语音段与各第二非语音段的一一对应关系，以便精确替换。

S34：根据上述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序，获得与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

本实施例通过第二类一一对应关系，以及各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序，使翻译后的音频码流文件及连续长语音文件的语音段一一对应，使连续长语音文件的节奏（比如每句话与每句话之间的不同间隔）、语言背景音（比如背景音乐、掌声等）以及语句自然间隔（即非语音段的自然长度），能更好的与翻译后的音频码流文件合成为一体，使得最终的翻译语音文件更加贴近原语言环境，提高用户的使用体验。

参照图 6，本发明一实施例的翻译机，包括：

第一解析模块 1，用于解析上述连续长语音文件，得到各第一语音段和各第一非语音段，其中，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布。

本实施例的终端设备以翻译机为例。本实施例中通过第一解析模块 1 解析上述连续长语音文件，得到各第一语音段与各第一非语音段的交替间隔排布的数据文件，各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序分布，比如表示为：第一语音段 1、第一非语音段 1、第一语音段 2、第一非语音段 2、第一语音段 3、第一非语音段 3、……、第一语音段 N、第一非语音段 N。

发送接收模块 2，用于发送上述连续长语音文件至服务器进行翻译，并接收上述服务器翻译上述连续长语音文件后的音频码流文件。

本实施例通过发送接收模块 2，将连续长语音文件依次发送到语音识别服务器、翻译服务器以及语音合成服务器进行翻译。本实施例的音频码流文件指翻译连续长语音文件后得到的对应的音频数据，包括语音数据、非语音数据。

第二解析模块 3，用于解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

本实施例的音频码流文件是通过一一对应翻译连续长语音文件后得到的音频数据，所以通过第二解析模块 3 解析音频码流文件，得到的各第二语音段和各第二非语音段与各第一语音段和各第一非语音段分布次序相同。

替换模块 4，用于在上述音频码流文件中将相同排序位置的上述第一非语音段替换掉上述第二非语音段，得到最终的翻译语音文件。

本实施例在与连续长语音文件分布次序相同的音频码流文件中，将相同排序位置上的第二非语音文件通过替换模块 4 替换为第一非语音段，并将第一非语音段与翻译后的音频码流文件合为一体，使得最终的翻译语音文件与原始连续长语音文件具有相同的节奏、语言背景音和语句自然间隔，增加机器翻译的生动活力感，提高用户的使用体验。

参照图 7，进一步地，本发明一实施例中，上述第一解析模块 1，包括：

第一处理单元 10，用于通过语音活动检测分析技术处理上述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态。

本实施例通过第一处理单元 10 对连续长语音文件做 VAD，将上述连续长语音文件中的第一语音段和第一非语音段区分开，以便于后续操作。举例地，按帧处理连续长语音文件，每帧时长根据语音信号特点来设定，比如 GSM 的 20ms 的时间为帧长度。首先通过 VAD 检测出连续长语音文件里的每个第一语音段的开始和结束，并通过算法处理得到连续长语音文件的各第一语音段的时间长度，如采用 GSM 通信系统中的 ETSI VAD 算法或者 G.729 Annex B VAD 算法，将 VAD 提取的连续长语音文件的参数特征值与门限值比较，以区分开上述第一语音段和上述第一非语音段。本实施例的排布状态是指，经第一处理单元 10 语音活动检测分析处理后，连续长语音文件变为连续排布的第一语音帧与连续排布的第一非语音帧交替分布的数据文件时的排布信息。

第一获得单元 11，用于根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

本实施例中，将通过 VAD 区分开的各第一语音段和各第一非语音段，分别采用不同编码标记，以便识别。

参照图 8，进一步地，本发明一实施例中，上述第一获得单元 11，包括：

合成子单元 112，用于根据上述排布状态将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段。

本实施例通过 VAD 的判决结果辨别第一语音帧和第一非语音帧，比如：判决结果是 1，就是第一语音帧；判决结果是 0，就是背景噪声帧（即第一非语音帧），并通过合成子单元 112 把连续长语音文件中的语句，变为经 VAD 处理后的连续排列的第一语音帧 1 到第一语音

帧 m 合并成的第一语音段 1，连续排列的第一非语音帧 1 到第一非语音帧 k 合并成的第一非语音段 1，依次处理，将连续长语音文件变为连续排列的第一语音段 1 至第一语音段 N，与连续排列的第一非语音段 1 至第一非语音段 N 一一交替分布的数据文件。

进一步地，本发明一实施例中，上述第一获得单元 11，还包括：

提取子单元 113，用于提取各上述第一非语音段。

根据各第一语音段和各第一非语音段的不同编码标记，通过提取子单元 113 把连续长语音文件里的第一非语音段提取出来，比如：将按照依次编码为 T1、T2…Tn 的第一非语音段 1、2...N 提取出来。

存储子单元 114，用于将各上述第一非语音段按照在上述连续长语音中生成的时序存储于非语音段缓存区。

本实施例的非语音段缓存区设定于存储子单元 114 的指定区域，以便按照在上述连续长语音中生成时序中第一非语音段的相同排列次序，依次将翻译后的音频码流文件与第一非语音段合为一体后再输出。

参照图 9，本发明一实施例中，上述发送接收模块 2，包括：

第一发送单元 20，用于将连续长语音文件发送至语音识别服务器。

第一接收单元 21，用于接收上述语音识别服务器反馈的与上述连续长语音文件对应的第一文本文件。

本实施例通过第一发送单元 20 将连续长语音文件发送至语音识别服务器，经语音识别服务器转换为与连续长语音文件相对应的第一文本文件。

第二发送单元 22，用于将上述第一文本文件发送至翻译服务器。

第二接收单元 23，用于接收上述翻译服务器反馈的翻译上述第一文本文件后的指定语种的第二文本文件。

本实施例第二发送单元 22 将第一文本文件发送至翻译服务器，通过翻译服务器对第一文本文件进行翻译，以形成指定语种的第二文本文件。比如：中文翻译成英文，中文的第一文本文件与英文的第二文本文件为一一对应关系，即中文一句话翻译成英文的一句话，每句

话一一对应。

第三发送单元24，用于将上述第二文本文件发送至语音合成服务器。

第三接收单元25，用于接收上述语音合成服务器转换上述第二文本文件后的音频码流文件。

本实施例通过第三发送单元24将第二文本文件顺序送入语音合成服务器，以便将第二文本文件顺序转换为指定语种的音频码流文件，比如，英文的音频码流文件。

参照图10，进一步地，本发明一实施例中，上述第二解析模块3，包括：

分析单元30，用于将上述第一文本文件的第一字符串信息与上述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系。

本实施例通过分析单元30对比分析文本文件的字符串信息，对第一文本文件和第二文本文件上的字符串组成的每句话分别做好标记，比如：第1句话、第2句话、……、第N句话，以便通过对应比较获得第二文本文件与第一文本文件的一一对应关系。

第二处理单元31，用于通过语音活动检测分析技术处理上述音频码流文件。

本实施例通过第二处理单元31，对音频码流文件进行VAD处理，以区分开音频码流文件中的第二语音段和第二非语音段，每个第二语音段对应第二文本文件中的一句话，N个第二非语音段的时间长度相同。

第二获得单元32，用于根据上述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段。

本实施例中，通过第二获得单元32获得各第二语音段和各第二非语音段，同样分别采用不同编码标记，以便识别。

建立单元33，用于根据上述第一类一一对应关系建立各上述第一语音段与各上述第二语音段的第二类一一对应关系。

每个第一语音段对应第一文本文件中的一句话，每个第二语音段

对应第二文本文件中的一句话，建立单元 33 根据第二文本文件与第一文本文件的一一对应关系，找到各第一语音段与各第二语音段的一一对应关系，以便确定各第一非语音段与各第二非语音段的一一对应关系，以便精确替换。

第三获得单元 34，用于根据上述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序，获得与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

本实施例的第三获得单元34通过第二类一一对应关系，以及各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序，获得翻译后的音频码流文件及连续长语音文件的语音段一一对应关系，使连续长语音文件的节奏（比如每句话与每句话之间的不同间隔）、语言背景音（比如背景音乐、掌声等）以及语句自然间隔（即非语音段的自然长度），能更好的与翻译后的音频码流文件合成为一体，使得最终的翻译语音文件更加贴近原语言环境，提高用户的使用体验。

以上所述仅为本发明的优选实施例，并非因此限制本发明的专利范围，凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本发明的专利保护范围内。

权利要求书

1、一种连续长语音文件的翻译方法，其特征在于，包括：

解析连续长语音文件，得到各第一语音段和各第一非语音段，其中，各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序分布；

发送所述连续长语音文件至服务器进行翻译，并接收所述服务器翻译所述连续长语音文件后的音频码流文件；

解析所述音频码流文件，得到与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段；

在所述音频码流文件中将相同排序位置的所述第一非语音段替换掉所述第二非语音段，得到最终的翻译语音文件。

2、根据权利要求 1 所述的连续长语音文件的翻译方法，其特征在于，所述解析连续长语音文件，得到各第一语音段和各第一非语音段的步骤，包括：

通过语音活动检测分析技术处理所述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态；

根据所述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

3、根据权利要求 2 所述的连续长语音文件的翻译方法，其特征在于，所述根据所述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段的步骤，包括：

将连续排布的第一语音帧分别合成各所述第一语音段，将连续排布的第一非语音帧分别合成各所述第一非语音段。

4、根据权利要求 3 所述的连续长语音文件的翻译方法，其特征在于，所述将连续排布的第一语音帧分别合成各所述第一语音段，将连续排布的第一非语音帧分别合成各所述第一非语音段的步骤之后，

包括：

提取各所述第一非语音段；

将各所述第一非语音段按照在所述连续长语音中生成的时序存储于非语音段缓存区。

5、根据权利要求 1 所述的连续长语音文件的翻译方法，其特征在于，所述发送所述连续长语音文件至服务器进行翻译，并接收所述服务器翻译所述连续长语音文件后的音频码流文件的步骤，包括：

将连续长语音文件发送至语音识别服务器；

接收所述语音识别服务器反馈的与所述连续长语音文件对应的第一文本文件；

将所述第一文本文件发送至翻译服务器；

接收所述翻译服务器反馈的翻译所述第一文本文件后的指定语种的第二文本文件；

将所述第二文本文件发送至语音合成服务器；

接收所述语音合成服务器转换所述第二文本文件后的音频码流文件。

6、根据权利要求 5 所述的连续长语音文件的翻译方法，其特征在于，所述解析所述音频码流文件，得到与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段的步骤，包括：

将所述第一文本文件的第一字符串信息与所述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系；

通过语音活动检测分析技术处理所述音频码流文件，获得第二语音帧和第二非语音帧的排布状态；

根据所述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段；

根据所述第一类一一对应关系建立各所述第一语音段与各所述第二语音段的第二类一一对应关系；

根据所述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序，获得与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

7、一种翻译机，其特征在于，包括：

第一解析模块，用于解析连续长语音文件，得到各第一语音段和各第一非语音段，其中，各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序分布；

发送接收模块，用于发送所述连续长语音文件至服务器进行翻译，并接收所述服务器翻译所述连续长语音文件后的音频码流文件；

第二解析模块，用于解析所述音频码流文件，得到与所述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段；

替换模块，用于在所述音频码流文件中将相同排序位置的所述第一非语音段替换掉所述第二非语音段，得到最终的翻译语音文件。

8、根据权利要求 7 所述的翻译机，其特征在于，所述第一解析模块，包括：

第一处理单元，用于通过语音活动检测分析技术处理所述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态；

第一获得单元，用于根据所述，获得第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段。

9、根据权利要求 8 所述的翻译机，其特征在于，所述第一获得单元，包括：

合成子单元，用于根据所述排布状态将连续排布的第一语音帧分别合成各所述第一语音段，将连续排布的第一非语音帧分别合成各所述第一非语音段。

10、根据权利要求 9 所述的翻译机，其特征在于，所述第一获得

单元，还包括：

提取子单元，用于提取各所述第一非语音段；

存储子单元，用于将各所述第一非语音段按照在所述连续长语音中生成的时序存储于非语音段缓存区。

11、根据权利要求 7 所述的翻译机，其特征在于，所述发送接收模块，包括：

第一发送单元，用于将连续长语音文件发送至语音识别服务器；

第一接收单元，用于接收所述语音识别服务器反馈的与所述连续长语音文件对应的第一文本文件；

第二发送单元，用于将所述第一文本文件发送至翻译服务器；

第二接收单元，用于接收所述翻译服务器反馈的翻译所述第一文本文件后的指定语种的第二文本文件；

第三发送单元，用于将所述第二文本文件发送至语音合成服务器；

第三接收单元，用于接收所述语音合成服务器转换所述第二文本文件后的音频码流文件。

12、根据权利要求 11 所述的翻译机，其特征在于，所述第二解析模块，包括：

分析单元，用于将所述第一文本文件的第一字符串信息与所述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系；

第二处理单元，用于通过语音活动检测分析技术处理所述音频码流文件，获得第二语音帧和第二非语音帧的排布状态；

第二获得单元，用于根据所述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段；

建立单元，用于根据所述第一类一一对应关系建立各所述第一语音段与各所述第二语音段的第二类一一对应关系；

第三获得单元，用于根据所述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在所述连续长语音中生成的时序，获得与所

述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段。

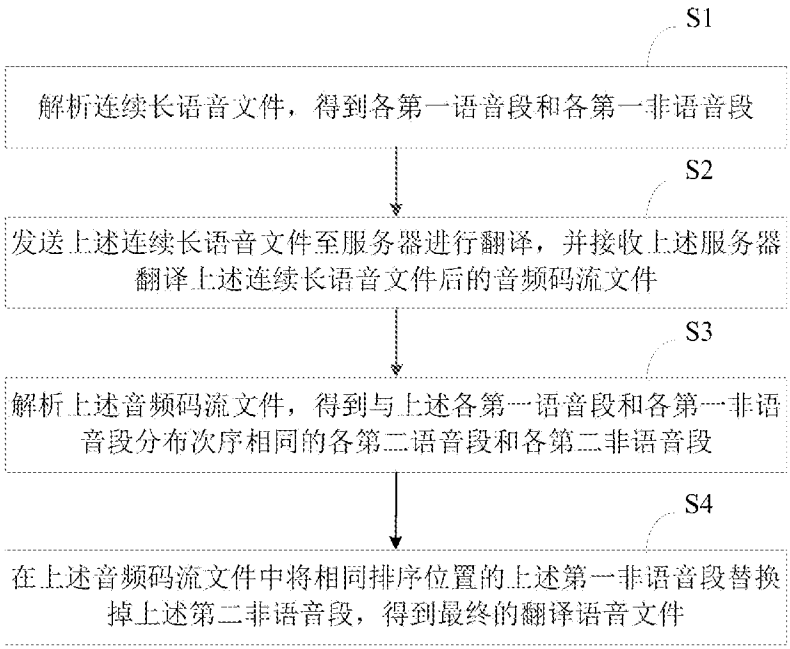

S1
解析连续长语音文件，得到各第一语音段和各第一非语音段
S2
发送上述连续长语音文件至服务器进行翻译，并接收上述服务器翻译上述连续长语音文件后的音频码流文件
S3
解析上述音频码流文件，得到与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段
S4
在上述音频码流文件中将相同排序位置的上述第一非语音段替换掉上述第二非语音段，得到最终的翻译语音文件

图 1

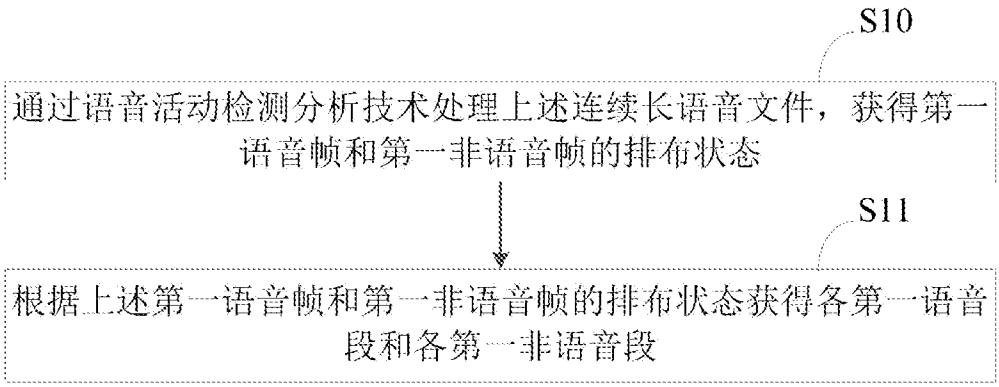

S10
通过语音活动检测分析技术处理上述连续长语音文件，获得第一语音帧和第一非语音帧的排布状态
S11
根据上述第一语音帧和第一非语音帧的排布状态获得各第一语音段和各第一非语音段

图 2

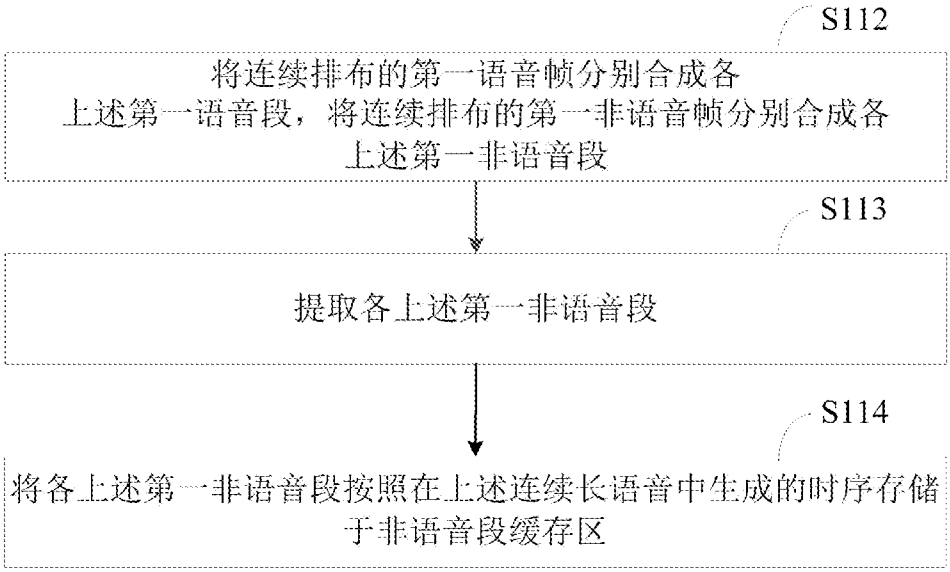

S112
将连续排布的第一语音帧分别合成各上述第一语音段，将连续排布的第一非语音帧分别合成各上述第一非语音段
S113
提取各上述第一非语音段
S114
将各上述第一非语音段按照在上述连续长语音中生成的时序存储于非语音段缓存区

图 3

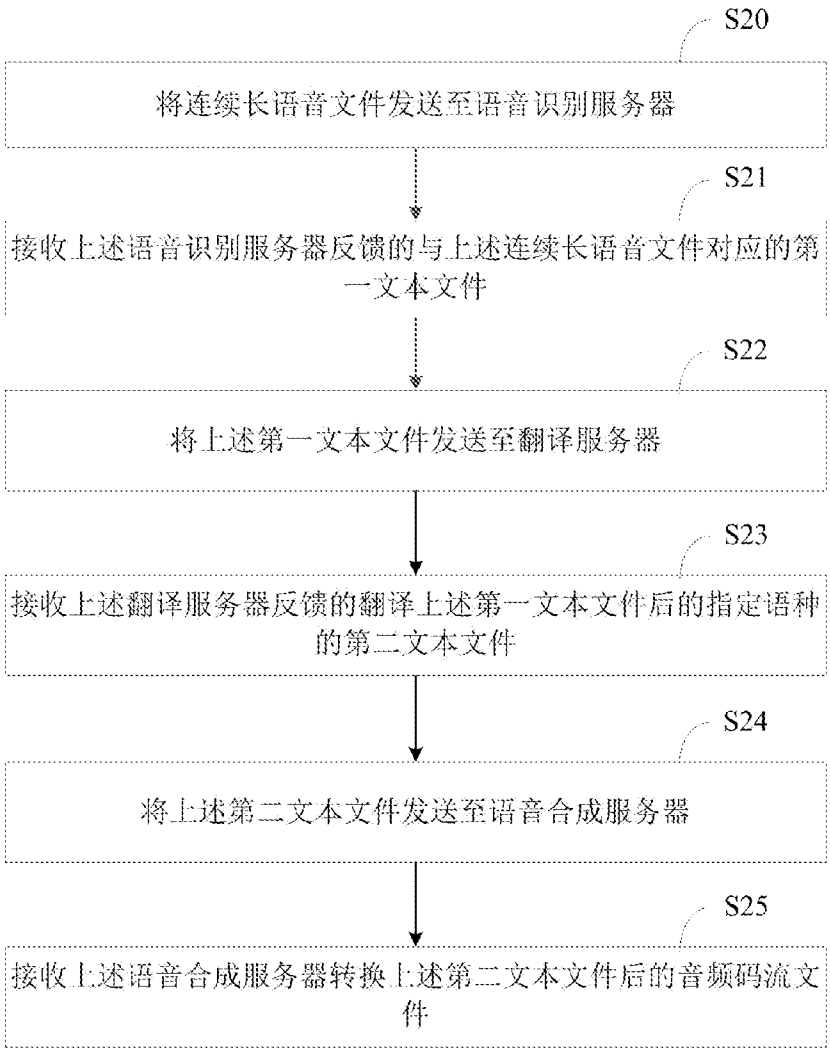

S20
将连续长语音文件发送至语音识别服务器
S21
接收上述语音识别服务器反馈的与上述连续长语音文件对应的第一文本文件
S22
将上述第一文本文件发送至翻译服务器
S23
接收上述翻译服务器反馈的翻译上述第一文本文件后的指定语种的第二文本文件
S24
将上述第二文本文件发送至语音合成服务器
S25
接收上述语音合成服务器转换上述第二文本文件后的音频码流文件

图 4

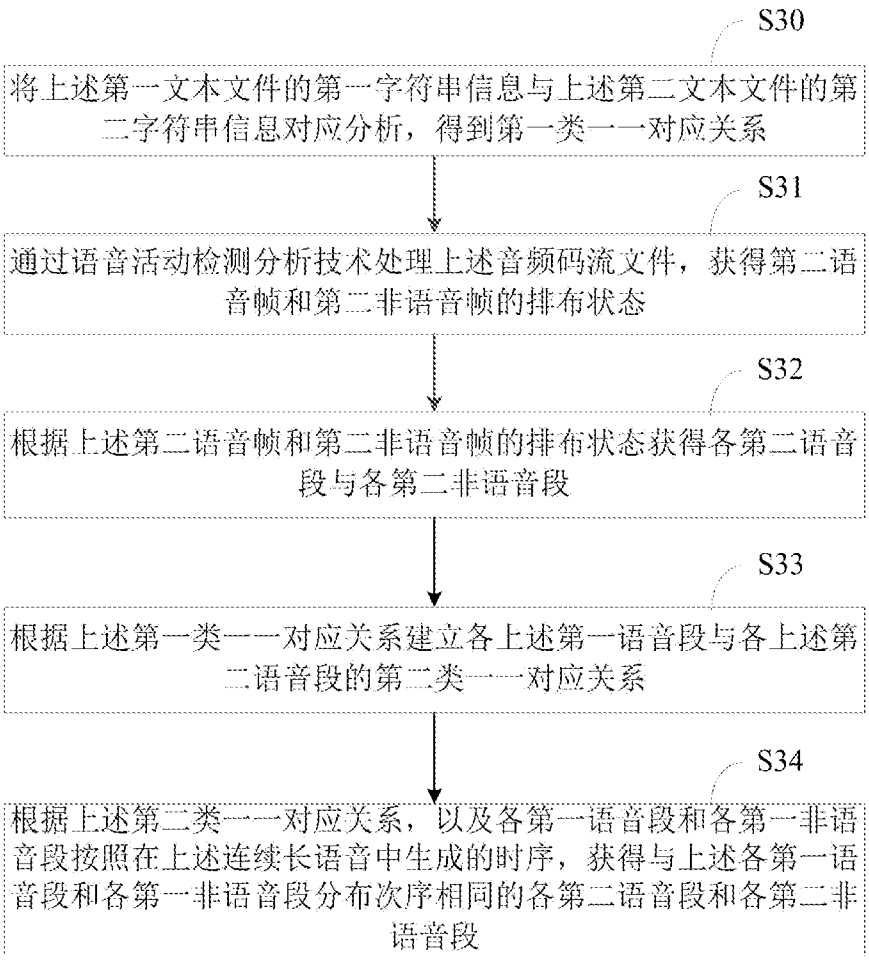

S30
将上述第一文本文件的第一字符串信息与上述第二文本文件的第二字符串信息对应分析，得到第一类一一对应关系
S31
通过语音活动检测分析技术处理上述音频码流文件，获得第二语音帧和第二非语音帧的排布状态
S32
根据上述第二语音帧和第二非语音帧的排布状态获得各第二语音段与各第二非语音段
S33
根据上述第一类一一对应关系建立各上述第一语音段与各上述第二语音段的第二类一一对应关系
S34
根据上述第二类一一对应关系，以及各第一语音段和各第一非语音段按照在上述连续长语音中生成的时序，获得与上述各第一语音段和各第一非语音段分布次序相同的各第二语音段和各第二非语音段

图 5

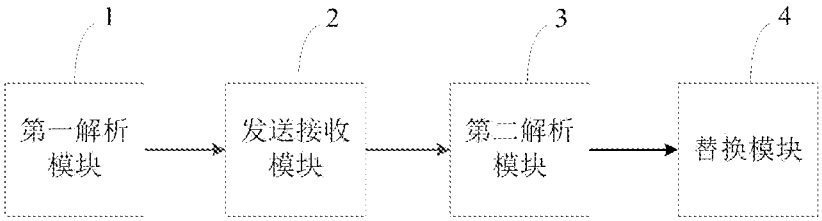

1
第一解析模块
2
发送接收模块
3
第二解析模块
4
替换模块

图 6

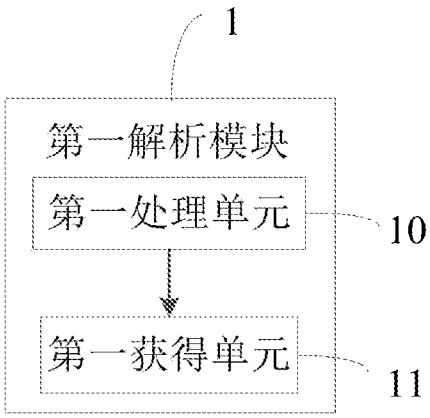

1
第一解析模块
第一处理单元
10
第一获得单元
11

图 7

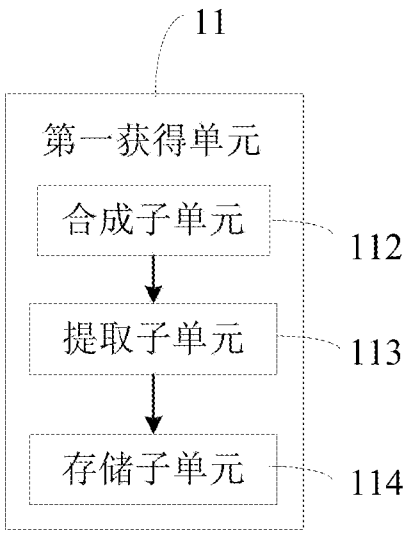

11
第一获得单元
合成子单元
112
提取子单元
113
存储子单元
114

图 8

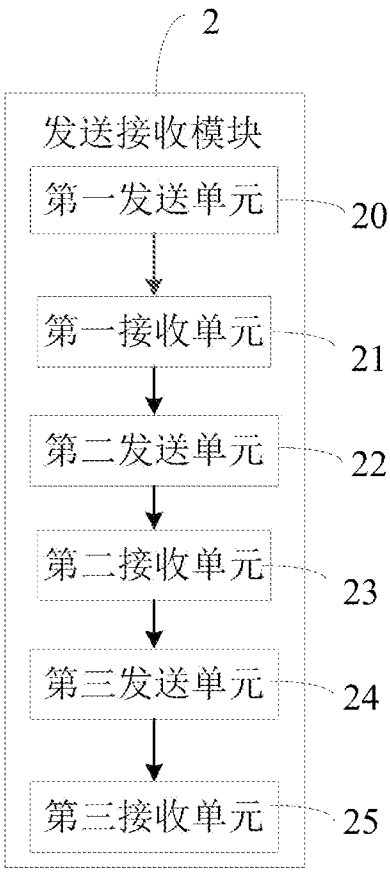

2
发送接收模块
第一发送单元
20
第一接收单元
21
第二发送单元
22
第二接收单元
23
第三发送单元
24
第三接收单元
25

图 9

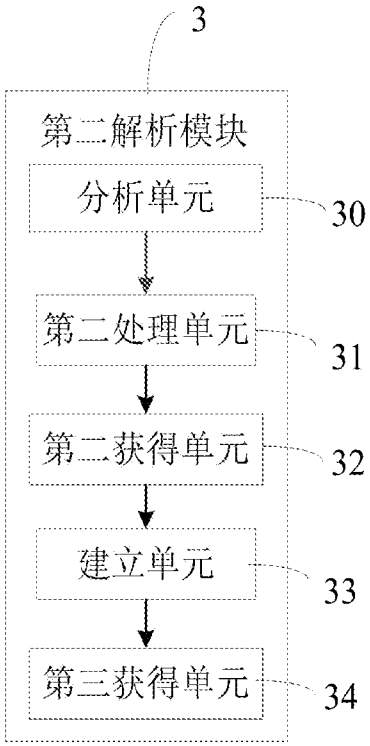


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/072007

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/28(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)

G06F17/-

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: 沃特沃德, 郑勇, 金志军, 王文祺, 语音, 翻译, 时序, 分段, 端点, 切分, 非语音, 替换, 替代, 取代, 置换; speech, voice, translat+, interpret+, time, section, subsection, segment, endpoint, terminal, separat+, replac+, displac+, substitut+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107391498 A (WATER WORLD TECHNOLOGY CO., LTD.) 24 November 2017 (2017-11-24) description, paragraphs [0089]-[0169], and figures 1-17	1-12
A	CN 101727904 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) 09 June 2010 (2010-06-09) entire document	1-12
A	CN 101458681 A (TOSHIBA CORPORATION) 17 June 2009 (2009-06-17) entire document	1-12
A	CN 105912533 A (SOOCHOW UNIVERSITY) 31 August 2016 (2016-08-31) entire document	1-12
A	CN 106303695 A (BEIJING GOCHINATV CULTURAL DEVELOPMENT CO., LTD.) 04 January 2017 (2017-01-04) entire document	1-12
A	CN 104252861 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 31 December 2014 (2014-12-31) entire document	1-12

☑ 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	Date of mailing of the international search report
23 August 2018	**20 September 2018**
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 Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/072007

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008243080 A (TOSHIBA CORPORATION) 09 October 2008 (2008-10-09) entire document	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/072007

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107391498	A	24 November 2017	None			
CN	101727904	A	09 June 2010	CN	101727904	B	24 April 2013
				US	2010114556	A1	06 May 2010
				US	9342509	B2	17 May 2016
CN	101458681	A	17 June 2009	US	2009150139	A1	11 June 2009
				JP	2009140503	A	25 June 2009
CN	105912533	A	31 August 2016	None			
CN	106303695	A	04 January 2017	None			
CN	104252861	A	31 December 2014	CN	104252861	B	13 April 2018
				WO	2016037440	A1	17 March 2016
JP	2008243080	A	09 October 2008	None			

国际检索报告

国际申请号
PCT/CN2018/072007

A. 主题的分类

G06F 17/28(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)

G06F17/-

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: 沃特沃德,郑勇,金志军,王文祺,语音,翻译,时序,分段,端点,切分,非语音,替换,替代,取代,置换; speech, voice, translat+, interpret+, time, section, subsection, segment, endpoint, terminal, separat+, replac+, displac+, substitut+

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
X	CN 107391498 A (深圳市沃特沃德股份有限公司) 2017年 11月 24日 (2017-11-24) 说明书第[0089]-[0169]段、附图1-17	1-12
A	CN 101727904 A (国际商业机器公司) 2010年 6月 9日 (2010-06-09) 全文	1-12
A	CN 101458681 A (株式会社东芝) 2009年 6月 17日 (2009-06-17) 全文	1-12
A	CN 105912533 A (苏州大学) 2016年 8月 31日 (2016-08-31) 全文	1-12
A	CN 106303695 A (北京东方嘉禾文化发展股份有限公司) 2017年 1月 4日 (2017-01-04) 全文	1-12
A	CN 104252861 A (百度在线网络技术北京有限公司) 2014年 12月 31日 (2014-12-31) 全文	1-12
A	JP 2008243080 A (株式会社东芝) 2008年 10月 9日 (2008-10-09) 全文	1-12

☐ 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	国际检索报告邮寄日期
2018年 8月 23日	2018年 9月 20日

ISA/CN的名称和邮寄地址	受权官员
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	陈红红
传真号 (86-10)62019451	电话号码 86-(10)-53962642

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/072007

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	107391498	A	2017年 11月 24日	无			
CN	101727904	A	2010年 6月 9日	CN	101727904	B	2013年 4月 24日
				US	2010114556	A1	2010年 5月 6日
				US	9342509	B2	2016年 5月 17日
CN	101458681	A	2009年 6月 17日	US	2009150139	A1	2009年 6月 11日
				JP	2009140503	A	2009年 6月 25日
CN	105912533	A	2016年 8月 31日	无			
CN	106303695	A	2017年 1月 4日	无			
CN	104252861	A	2014年 12月 31日	CN	104252861	B	2018年 4月 13日
				WO	2016037440	A1	2016年 3月 17日
JP	2008243080	A	2008年 10月 9日	无			