

(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
2017 年 5 月 26 日 (26.05.2017)



(10) 国际公布号
WO 2017/084334 A1

- (51) 国际专利分类号:
G10L 15/00 (2013.01) G10L 15/02 (2006.01)
- (21) 国际申请号: PCT/CN2016/086357
- (22) 国际申请日: 2016 年 6 月 20 日 (20.06.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510792176.X 2015 年 11 月 17 日 (17.11.2015) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司
(BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 李骁 (LI, Xiao); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 李超 (LI, Chao); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。 关勇 (GUAN, Yong); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。
- (74) 代理人: 北京鸿德海业知识产权代理事务所(普通合伙) (BEIJING WISPRO INTELLECTUAL PROPERTY LLP.); 中国北京市海淀区知春路 6 号锦秋国际大厦 A508, Beijing 100088 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,

[见续页]

(54) Title: LANGUAGE RECOGNITION METHOD, APPARATUS AND DEVICE AND COMPUTER STORAGE MEDIUM

(54) 发明名称: 一种语种识别方法、装置、设备及计算机存储介质

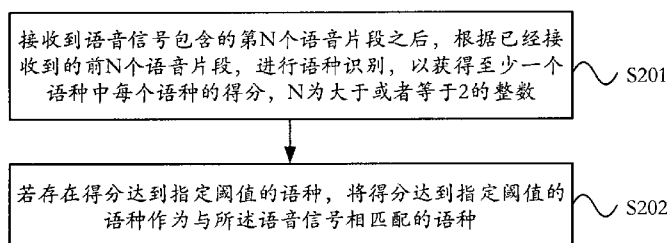


图 2

S201 After receiving the Nth speech segment contained in a speech signal, performing language recognition according to previous N speech segments which are already received, so as to obtain a score of each language in at least one language, where the value of N is an integer greater than or equal to 2
S202 If there is a language the score of which reaches a designated threshold value, taking the language the score of which reaches the designated threshold value as a language that matches the speech signal

(57) Abstract: Disclosed are a language recognition method and apparatus. The method comprises: after receiving the Nth speech segment contained in a speech signal, performing language recognition according to previous N speech segments which are already received, so as to obtain a score of each language in at least one language, where the value of N is 2, 3, 4, ... (S201); and if there is a language the score of which reaches a designated threshold value, taking the language the score of which reaches the designated threshold value as a language that matches the speech signal (S202). The present invention solves the problem in the prior art that the language recognition efficiency is relatively low, so that the language recognition cannot be applied to an application scenario where a recognition result needs to be obtained quickly.

(57) 摘要: 一种语种识别方法及装置。该方法包括: 接收到语音信号包含的第 N 个语音片段之后, 根据已经接收到的前 N 个语音片段, 进行语种识别, 以获得至少一个语种中每个语种的得分, N 的取值为 2、3、4、…… (S201); 若存在得分达到指定阈值的语种, 将得分达到指定阈值的语种作为与所述语音信号相匹配的语种 (S202)。解决了现有技术中语种识别的效率比较低, 使语种识别无法应用于需要快速获得识别结果的应用场景中的问题。



WO 2017/084334 A1

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

一种语种识别方法、装置、设备及计算机存储介质

本申请要求了申请日为 2015 年 11 月 17 日，申请号为 201510792176.X 发明名称为“一种语种识别方法及装置”的中国专利申请优先权。

技术领域

本发明涉及语音识别技术领域，尤其涉及一种语种识别方法、装置、设备及计算机存储介质。

背景技术

随着互联网应用与信息技术的不断发展，出现了语音识别技术，目前语音识别技术已经广泛应用于家电、工业、医疗等许多领域，给用户的生活带来极大便利。

现有技术中，在进行语音识别之前需要先进行语种识别，用以识别出输入的语音信号是采用的哪个语种，进而能够使用该语种对应的解码器对语音信号进行进一步的语音识别。

然而，现有技术中，客户端将用户输入的语音信号分成若干语音片段后，依次将语音片段发送到服务器进行语种识别的。服务器在接收到属于同一语音信号的所有语音片段之后，才开始进行语种识别，导致现有技术中语种识别的效率比较低，使语种识别无法应用于需要快速获得识别结果的应用场景中。

发明内容

有鉴于此，本发明实施例提供了一种语种识别方法、装置、设备及计算机存储介质，解决了现有技术中语种识别的效率比较低，使语种识别无法应用于需要快速获得识别结果的应用场景中的问题。

本发明实施例的一方面，提供一种语种识别方法，包括：

接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分， N 的取值为 2、3、4、... ..；

若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述方法还包括：

若不存在得分达到指定阈值的语种，在接收到第 $N+1$ 个语音信号之后，继续根据已经接收到第 $N+1$ 个语音信号，进行语种识别，直到获得与所述语音信号相匹配的语种为止。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分，包括：

从所述第 N 个语音片段提取语音特征；

根据所述语音特征，获得所述第 N 个语音片段的统计量；

根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量，获得第二累计统计量；

利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，从所述第 N 个语音片段提取语音特征之前，所述方法还包括：对所述第 N 个语音片段进行静音检测，以去除所述第 N 个语音片段中的静音信号。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述第二累计统计量包括:前 N 个语音片段的语音特征的权重、前 N 个语音片段的语音特征的均值和前 N 个语音片段的语音特征的方差。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述语种识别模型包括判决模型和背景模型;所述利用语种识别模型对所述第二累计统计量进行语种识别,以获得至少一个语种中每个语种的得分,包括:

利用背景模型将所述第二累计统计量转换成对应的高斯超向量;

计算判决模型中的向量与所述高斯超向量之间的空间距离,以作为所述判决模型对应的语种的得分。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述方法还包括:

若所述语音信号包含的所有语音片段都接收到之后,不存在得分达到指定阈值的语种,将得分最高的语种作为与所述语音信号相匹配的语种。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述方法还包括:将与所述语音信号相匹配的语种作为语种识别结果,并输出所述语种识别结果。

本发明实施例的一方面,提供一种语种识别装置,包括:

识别模块,用于接收到语音信号包含的第 N 个语音片段之后,根据已经接收到的前 N 个语音片段,进行语种识别,以获得至少一个语种中每个语种的得分, N 的取值为 2、3、4、... ...;

判断模块,用于若存在得分达到指定阈值的语种,将得分达到指定

阈值的语种作为与所述语音信号相匹配的语种。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述识别模块,还用于:

若不存在得分达到指定阈值的语种,在接收到第 $N+1$ 个语音信号之后,继续根据已经接收到第 $N+1$ 个语音信号,进行语种识别,直到获得与所述语音信号相匹配的语种为止。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述识别模块,具体用于:

从所述第 N 个语音片段提取语音特征;

根据所述语音特征,获得所述第 N 个语音片段的统计量;

根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量,获得第二累计统计量;

利用语种识别模型对所述第二累计统计量进行语种识别,以获得至少一个语种中每个语种的得分。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述识别模块,还用于:

对所述第 N 个语音片段进行静音检测,以去除所述第 N 个语音片段中的静音信号。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述第二累计统计量包括:前 N 个语音片段的语音特征的权重、前 N 个语音片段的语音特征的均值和前 N 个语音片段的语音特征的方差。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述语种识别模型包括判决模型和背景模型;所述识别模块用于利用语

种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分时，具体用于：

利用背景模型将所述第二累计统计量转换成对应的高斯超向量；

计算判决模型中的向量与所述高斯超向量之间的空间距离，以作为所述判决模型对应的语种的得分。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述判断模块，还用于：

若所述语音信号包含的所有语音片段都接收到之后，不存在得分达到指定阈值的语种，将得分最高的语种作为与所述语音信号相匹配的语种。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述装置还包括：

输出模块，用于将与所述语音信号相匹配的语种作为语种识别结果，并输出所述语种识别结果。

由以上技术方案可以看出，本发明实施例具有以下有益效果：

根据本发明实施例提供的技术方案，能够在每次接收到一个语音片段之后，就进行语种识别，从而可以在接收完所有语音片段之前就可以完成语种识别，与现有技术中，必须在接收到所有语音片段之后才开始进行语种识别的方法相比，本发明实施例解决了现有技术中语种识别的效率比较低，使语种识别无法应用于需要快速获得识别结果的应用场景中的问题，大大提高了语种识别的效率，使得语种识别能够应用于需要快速获得识别结果的应用场景中。

附图说明

图 1 是本发明实施例所提供的技术方案使用的系统示例图；

图 2 是本发明实施例所提供的语种识别方法的流程示意图；

图 3 是本发明实施例所提供的语种识别方法的优选实施例的流程示意图；

图 4 是本发明实施例所提供的语种识别装置的功能方块图。

具体实施方式

为了使本发明的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本发明进行详细描述。

应当明确，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。

在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

应当理解，本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

应当理解，尽管在本发明实施例中可能采用术语第一、第二等来描述累计统计量，但这些累计统计量不应限于这些术语。这些术语仅用来将累计统计量彼此区分开。例如，在不脱离本发明实施例范围的情况下，第一累计统计量也可以被称为第二累计统计量，类似地，第二累计统计

量也可以被称为第一累计统计量。

取决于语境，如在此所使用的词语“如果”可以被解释成为“在... ..时”或“当... ..时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测（陈述的条件或事件）”可以被解释成为“当确定时”或“响应于确定”或“当检测（陈述的条件或事件）时”或“响应于检测（陈述的条件或事件）”。

本发明实施例所提供的技术方案使用的系统如图 1 所示，主要由客户端和服务端组成，本发明实施例所提供的方法和装置在服务器侧实现，主要用于接收客户端发送的语音信号，然后进行语种识别，最后输出识别结果。本发明实施例对客户端的功能没有进行变更。

本发明实施例给出一种语种识别方法，请参考图 2，其为本发明实施例所提供的语种识别方法的流程示意图，如图所示，该方法包括以下步骤：

S201，接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分，N 为大于或者等于 2 的整数。

S202，若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

请参考图 3，其为本发明实施例所提供的语种识别方法的优选实施例的流程示意图，如图所示，在线上的实际场景中，客户端可以接收用户输入的语音信号，然后将语音信号分成若干语音片段，进而将语音片段依次发送给服务器。这意味着服务器并不是一次性的接收到用户的语音信号，而是一部分一部分的接收，依次接收语音片段。用户在一句话

还没说完时，实际上一部分的语音片段已经被发送到了服务器，这样，服务器就可以接收到客户端发送的语音片段，其中，客户端是按照顺序向服务器发送若干语音片段的，服务器一次也只接收一个语音片段。

可以理解的是，一个语音信号包含的各语音片段的长度可以相等，或者，也可以不等。例如，第一个语音片段的长度是 500ms，此后的每个语音片段的长度是 250ms。

本发明实施例中，服务器在接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分。

进一步的，若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。这里，由于服务器获得了与语音信号相匹配的语种，则服务器停止进行语种识别，即服务器即使再接收到第 N+1 个语音片段，服务器也不会继续根据前 N+1 个语音信号继续进行语种识别。

反之，若不存在得分达到指定阈值的语种，则在服务器接收到第 N+1 个语音信号之后，服务器继续根据已经接收到前 N+1 个语音信号，进行语种识别，直到获得与所述语音信号相匹配的语种为止。

若所述语音信号包含的所有语音片段都被服务器接收到之后，仍然不存在得分达到指定阈值的语种，则服务器将得分最高的语种作为与所述语音信号相匹配的语种。

本发明实施例中，N 为大于或者等于 2 的整数，N 为 2、3、4、...。相当于，服务器从接收的第 2 个语音片段开始，每收到一个语音片段之后，就根据已经收到所有语音片段进行一次语种识别，直到获得得分达

到指定阈值的语种时为止。

举例说明，本发明实施例中，服务器根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分的方法可以包括但不限于：

如图 3 所示，首先，对所述第 N 个语音片段进行静音检测，以去除所述第 N 个语音片段中的静音信号。然后，从所述第 N 个语音片段提取语音特征。接着，根据所述语音特征，获得所述第 N 个语音片段的统计量。然后，根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量，获得第二累计统计量。最后，利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分。

实际应用过程中，用户在向客户端输入语音信号时，在用户说话之前，客户端往往会采集到一定时长的静音信号。因此，服务器在收到一个语音片段之后，如第 N 个语音片段，需要先对该第 N 个语音片段进行静音检测，以确定静音信号在第 N 个语音片段中的位置，进而根据该位置，在该语音片段中去除该第 N 个语音片段的静音信号，从而避免静音信号的存在给语种识别带来的干扰。另外，还可以在静音检测的同时将语音片段中的噪声信号同时去除，进一步减少了给语种识别带来的干扰。

在一个具体的实现过程中，在对第 N 个语音片段进行静音检测后，进一步从静音检测后获得的第 N 个语音片段提取第 N 个语音片段的语音特征。其中，第 N 个语音片段的语音特征可以包括但不限于：感知线性预测系数(Perceptual Linear Predictive, PLP)和梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)等。

如图 3 所示，本发明实施例中，服务器在每接收到一个语音片段后，

都获得该语音片段的统计量，然后将该语音片段的统计量与之前收到的所有语音片段的累计统计量进行累计，从而获得新的累计统计量，相当于将该语音片段的统计量是在原有累计统计量上进行累计，并不是重新进行接收到所有语音片段的统计量的累计。

如图 3 所示，基于上述描述，服务器在接收到第 N 个语音片段后，就将第 N 个语音片段的统计量与前 N-1 个语音片段的第一累计统计量相加，以获得第二累计统计量，以实现将第 N 个语音片段的统计量累计到该第二累计统计量之中。可以理解的是，第二累计统计量代表了语音信号到目前为止的语种特征。

本发明实施例中，所述第 N 个语音片段的统计量包括：第 N 个语音片段的权重。

所述第一累计统计量包括：前 N-1 个语音片段的语音特征的权重、前 N-1 个语音片段的语音特征的均值和前 N-1 个语音片段的语音特征的方差。

相应的，所述第二累计统计量包括：前 N 个语音片段的语音特征的权重、前 N 个语音片段的语音特征的均值和前 N 个语音片段的语音特征的方差。

可以理解的是，为了能够实现语音片段的统计量的累计，可以在每次累计统计量之后，将获得的新的累计统计量保存在服务器的内存中，且在接收到重置信号之前，不会进行累计统计量的清零。因此，可以从服务器的内存中获得前 N-1 个语音片段的第一累计统计量。

举例说明，利用如下公式获得第 N 个语音片段的权重：

$$n_t = \Pr(i|x_t)$$

其中, n_i 表示第 N 个语音片段的权重。 x_t 表示在 t 时刻的语音片段, $\Pr(i|x_t)$ 表示 x_t 落在背景模型中第 i 个簇上的概率。

同理, 举例说明, 利用如下公式获得前 N 个语音片段的语音特征的权重:

$$n_i = \sum_{t=1}^T \Pr(i|x_t)$$

利用如下公式获得前 N 个语音片段的语音特征的均值:

$$E_i(x) = \frac{1}{n_i} \sum_{t=1}^T \Pr(i|x_t) x_t$$

利用如下公式获得前 N 个语音片段的语音特征的方差:

$$E_i(x^2) = \frac{1}{n_i} \sum_{t=1}^T \Pr(i|x_t) x_t^2$$

上述公式中, n_i 表示前 N 个语音片段的语音特征的权重; T 表示语音信号的总时长; $E_i(x)$ 表示前 N 个语音片段的语音特征的均值; $E_i(x^2)$ 表示前 N 个语音片段的语音特征的方差。

举例说明, 利用语种识别模型对所述第二累计统计量进行语种识别, 以获得至少一个语种中每个语种的得分的方法可以包括但不限于:

如图 3 所示, 本发明实施例中, 所述语种识别模型可以包括但不限于判决模型和背景模型。其中, 判决模型可以利用高斯超向量模型、深度神经网络模型或者空间向量模型实现。所述判决模型的数目可以为至少两个, 如英语对应的判决模型、汉语对应的判决模型、法语对应的判决模型等, 每个判决模型可以包括对应的语种的向量。背景模型是一个是利用大量不同语种、不同信道和不同性别的用户的语音信号进行训练

获得的高阶的高斯混合模型。

首先，可以利用背景模型将所述第二累计统计量转换成对应的高斯超向量。然后，计算判决模型中的向量与所述高斯超向量之间的空间距离，以作为所述判决模型对应的语种的得分。

可以理解的是，可以利用至少两个判断模型中每个判决模型分别对所述第二累计统计量进行打分，从而获得每个判决模型对应的语种的得分。例如，利用英语对应的判决模型进行打分，获得前 N 个语音片段对应的语种是英语的得分，利用汉语对应的判决模型进行打分，获得前 N 个语音片段对应的语种是汉语的得分，利用法语对应的判决模型进行打分，获得前 N 个语音片段对应的语种是法语的得分。

进一步，将至少一个语种中每个语种的得分，分别与该语种的指定阈值进行比较，若存在得分达到相应语种的指定阈值的语种，将该得分达到相应语种的指定阈值的语种，作为与所述前 N 个语音片段相匹配的语种，并停止语种识别，即使接收到第 N+1 个语音片段，服务器也不会对其执行上述的语种识别过程，因此已经识别出了与语音信号相匹配的语种。

反之，如果不存在得分达到相应语种的指定阈值的语种，即所有得分都没有达到相应语种的指定阈值，则继续对接收到的下一个语音片段，即第 N+1 个语音片段执行上述操作，用以根据前 N+1 个语音片段，再次获得每个语种的得分，直到存在得分达到相应语种的指定阈值的语种时停止上述语种识别过程。

例如，若前 N 个语音片段对应的语种是英语的得分达到英语的指定阈值，则认为英语是与前 N 个语音片段相匹配的语种。或者，又如，若

前 N 个语音片段对应的语种是英语的得分达到汉语的指定阈值，则认为汉语是与前 N 个语音片段相匹配的语种。或者，又如，若前 N 个语音片段对应的语种是法语的得分达到法语的指定阈值，则认为法语是与前 N 个语音片段相匹配的语种。

可以理解的是，一般的，要么没有得分达到指定阈值的语种，要么只有一个得分达到指定阈值的语种，不会存在两个以上得分达到指定阈值的语种。

可以理解的是，得分达到指定阈值指的是得分大于或者等于该指定阈值。

进一步的，在获得与所述语音信号相匹配的语种之后，将与所述语音信号相匹配的语种作为语种识别结果，并输出所述语种识别结果给控制模块。控制模块判断当前所使用的解码器是否为该语种对应的解码器，如果是，则可以利用当前所使用的解码器对已经收到的所有语音片段进行语音识别；如果不是，则切换到该语种对应的解码器，利用该语种对应的解码器对已经收到的所有语音片段进行语音识别。这样，即使服务器还没有获得完整的语音信号，服务器也可以开始进行语种识别和语音识别，提高了语种识别效率和语音识别效率。

现有技术中，服务器在接收到属于同一语音信号的所有语音片段之后，将完整的语音信号作为整体进行静音检测、语音特征提取以及利用语种识别模型进行识别。该技术方案的缺点是需要接收到所有语音片段之后，才开始进行语种识别，从而导致语种识别的效率比较低，使得语种识别技术无法应用于需要及时得到识别结果的应用场景中。例如，当所在国家未知的某搜索客户端在接收到用户输入的语音信号之后，服务

器需要先从客户端接收到完整的语音信号之后，再等待语种识别完成，然后再根据语种识别的结果选择相应的解码器来进行语音识别，最后根据语音识别结果进行搜索，并返回搜索结果给客户端。这一系列的处理过程效率比较低，让用户感受到了极大的延迟感，目前整体实时率在 0.5 左右，也就是说用户输入完一个 3s 长的语音信号之后，需要等待 1.5s 才能得到返回的结果。

为了解决现有技术中的问题，本发明实施例所提供的技术方案中，在每接收到一个语音片段后，就进行一次语种识别，若根据语种识别中的得分能够确定相匹配的语种，就立即获得语种识别结果，并不再对此后收到的语音片段进行语种识别，而是立即开始语音识别，从而能够提高语种识别效率和语音识别效率，使用户能够及时获得所需要的结果。

传统的语种识别技术由于受限于语音的完整性，难以与语音识别等其他语音技术相结合，不能应用于语音搜索、语音输入等与语音相关的常规应用场景。本发明实施例所提供的技术方案能够扩大语种识别的应用场景，使语种识别技术能够适用于更多的应用场景，如需要及时得到识别结果的应用场景。

另外，现有技术中，支持语音识别的应用往往只能接收针对某一特定语言的语音信号，如果用户需要使用其他语言来输入语音信号，只能选择使用其他应用，操作成本较高，用户体验较差。然而，如果结合使用本发明实施例所提供的语种识别方法，可以实现支持不同语言输入的语音信号，减少了操作成本，给用户带来良好体验。

本发明实施例进一步给出实现上述方法实施例中各步骤及方法的装置实施例。

请参考图 4，其为本发明实施例所提供的语种识别装置的功能方块图。如图所示，该装置包括：

识别模块 41，用于接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分， N 的取值为 2、3、4、... ...；

判断模块 42，用于若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

可选的，所述识别模块 41，还用于：

若不存在得分达到指定阈值的语种，在接收到第 $N+1$ 个语音信号之后，继续根据已经接收到第 $N+1$ 个语音信号，进行语种识别，直到获得与所述语音信号相匹配的语种为止。

在一个具体的实现过程中，所述识别模块 41，具体用于：

从所述第 N 个语音片段提取语音特征；

根据所述语音特征，获得所述第 N 个语音片段的统计量；

根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量，获得第二累计统计量；

利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分。

可选的，所述识别模块 41，还用于：

对所述第 N 个语音片段进行静音检测，以去除所述第 N 个语音片段中的静音信号。

在一个具体的实现过程中，所述第二累计统计量包括：前 N 个语音片段的语音特征的权重、前 N 个语音片段的语音特征的均值和前 N 个语

音片段的语音特征的方差。

在一个具体的实现过程中，所述语种识别模型包括判决模型和背景模型；所述识别模块用于利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分时，具体用于：

利用背景模型将所述第二累计统计量转换成对应的高斯超向量；

计算判决模型中的向量与所述高斯超向量之间的空间距离，以作为所述判决模型对应的语种的得分。

可选的，所述判断模块 42，还用于：

若所述语音信号包含的所有语音片段都接收到之后，不存在得分达到指定阈值的语种，将得分最高的语种作为与所述语音信号相匹配的语种。

可选的，所述装置还包括：

输出模块 43，用于将与所述语音信号相匹配的语种作为语种识别结果，并输出所述语种识别结果。

由于本实施例中的各单元能够执行图 2 所示的方法，本实施例未详细描述的部分，可参考对图 2 的相关说明。

本发明实施例的技术方案具有以下有益效果：

本发明实施例中，通过接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分，N 的取值为 2、3、4、... ...；从而，若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

根据本发明实施例提供的技术方案，能够在每次接收到一个语音片

段之后，就进行语种识别，从而可以在接收完所有语音片段之前就可以完成语种识别，与现有技术中，必须在接收到所有语音片段之后才开始进行语种识别的方法相比，本发明实施例解决了现有技术中语种识别的效率比较低，使语种识别无法应用于需要快速获得识别结果的应用场景中的问题，大大提高了语种识别的效率，使得语种识别能够应用于需要快速获得识别结果的应用场景中。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本发明所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如，多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集

成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机装置（可以是个人计算机，服务器，或者网络装置等）或处理器（Processor）执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

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

1、一种语种识别方法，其特征在于，所述方法包括：

接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分， N 的取值为 2、3、4、... ...；

若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

2、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

若不存在得分达到指定阈值的语种，在接收到第 $N+1$ 个语音信号之后，继续根据已经接收到第 $N+1$ 个语音信号，进行语种识别，直到获得与所述语音信号相匹配的语种为止。

3、根据权利要求 1 所述的方法，其特征在于，所述根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分，包括：

从所述第 N 个语音片段提取语音特征；

根据所述语音特征，获得所述第 N 个语音片段的统计量；

根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量，获得第二累计统计量；

利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分。

4、根据权利要求 3 所述的方法，其特征在于，从所述第 N 个语音片段提取语音特征之前，所述方法还包括：对所述第 N 个语音片段进行静音检测，以去除所述第 N 个语音片段中的静音信号。

5、根据权利要求3所述的方法，其特征在于，所述第二累计统计量包括：前N个语音片段的语音特征的权重、前N个语音片段的语音特征的均值和前N个语音片段的语音特征的方差。

6、根据权利要求3所述的方法，其特征在于，所述语种识别模型包括判决模型和背景模型；所述利用语种识别模型对所述第二累计统计量进行语种识别，以获得至少一个语种中每个语种的得分，包括：

利用背景模型将所述第二累计统计量转换成对应的高斯超向量；

计算判决模型中的向量与所述高斯超向量之间的空间距离，以作为所述判决模型对应的语种的得分。

7、根据权利要求1所述的方法，其特征在于，所述方法还包括：

若所述语音信号包含的所有语音片段都接收到之后，不存在得分达到指定阈值的语种，将得分最高的语种作为与所述语音信号相匹配的语种。

8、根据权利要求1至7中任一项所述的方法，其特征在于，所述方法还包括：将与所述语音信号相匹配的语种作为语种识别结果，并输出所述语种识别结果。

9、一种语种识别装置，其特征在于，所述装置包括：

识别模块，用于接收到语音信号包含的第N个语音片段之后，根据已经接收到的前N个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分，N的取值为2、3、4、... ..；

判断模块，用于若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

10、根据权利要求9所述的装置，其特征在于，所述识别模块，还

用于:

若不存在得分达到指定阈值的语种,在接收到第 $N+1$ 个语音信号之后,继续根据已经接收到第 $N+1$ 个语音信号,进行语种识别,直到获得与所述语音信号相匹配的语种为止。

11、根据权利要求 9 所述的装置,其特征在于,所述识别模块,具体用于:

从所述第 N 个语音片段提取语音特征;

根据所述语音特征,获得所述第 N 个语音片段的统计量;

根据所述第 N 个语音片段的统计量和前 $N-1$ 个语音片段的第一累计统计量,获得第二累计统计量;

利用语种识别模型对所述第二累计统计量进行语种识别,以获得至少一个语种中每个语种的得分。

12、根据权利要求 11 所述的装置,其特征在于,所述识别模块,还用于:

对所述第 N 个语音片段进行静音检测,以去除所述第 N 个语音片段中的静音信号。

13、根据权利要求 11 所述的装置,其特征在于,所述第二累计统计量包括:前 N 个语音片段的语音特征的权重、前 N 个语音片段的语音特征的均值和前 N 个语音片段的语音特征的方差。

14、根据权利要求 11 所述的装置,其特征在于,所述语种识别模型包括判决模型和背景模型;所述识别模块用于利用语种识别模型对所述第二累计统计量进行语种识别,以获得至少一个语种中每个语种的得分时,具体用于:

利用背景模型将所述第二累计统计量转换成对应的高斯超向量;

计算判决模型中的向量与所述高斯超向量之间的空间距离,以作为所述判决模型对应的语种的得分。

15、根据权利要求 9 所述的装置,其特征在于,所述判断模块,还用于:

若所述语音信号包含的所有语音片段都接收到之后,不存在得分达到指定阈值的语种,将得分最高的语种作为与所述语音信号相匹配的语种。

16、根据权利要求 9 至 15 中任一项所述的装置,其特征在于,所述装置还包括:

输出模块,用于将与所述语音信号相匹配的语种作为语种识别结果,并输出所述语种识别结果。

17、一种设备,包括

一个或者多个处理器;

存储器;

一个或者多个程序,所述一个或者多个程序存储在所述存储器中,当被所述一个或者多个处理器执行时:

接收到语音信号包含的第 N 个语音片段之后,根据已经接收到的前 N 个语音片段,进行语种识别,以获得至少一个语种中每个语种的得分, N 的取值为 2、3、4、... ...;

若存在得分达到指定阈值的语种,将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

18、一种计算机存储介质,所述计算机存储介质被编码有计算机程

序，所述程序在被一个或多个计算机执行时，使得所述一个或多个计算机执行如下操作：

接收到语音信号包含的第 N 个语音片段之后，根据已经接收到的前 N 个语音片段，进行语种识别，以获得至少一个语种中每个语种的得分， N 的取值为 2、3、4、... ..；

若存在得分达到指定阈值的语种，将得分达到指定阈值的语种作为与所述语音信号相匹配的语种。

1/3

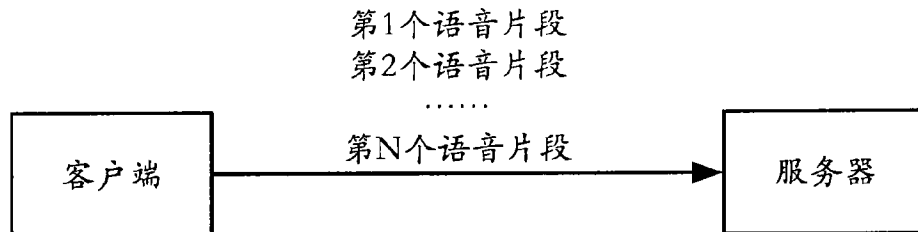
说明书附图

图 1

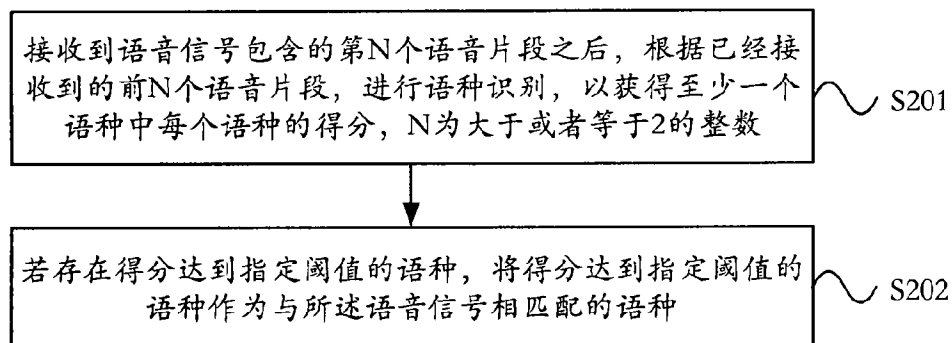


图 2

1/3

2/3

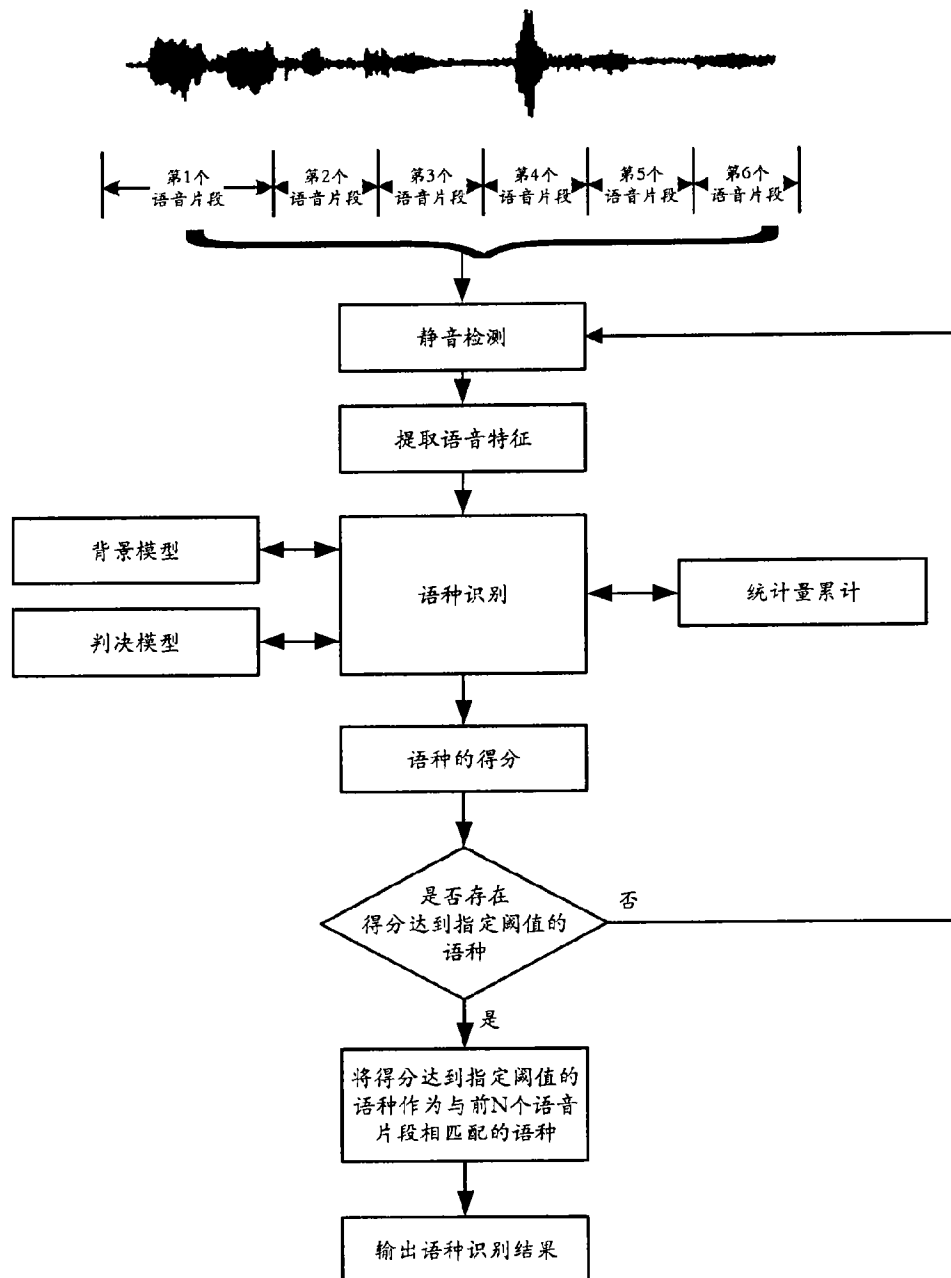


图 3

2/3

3/3

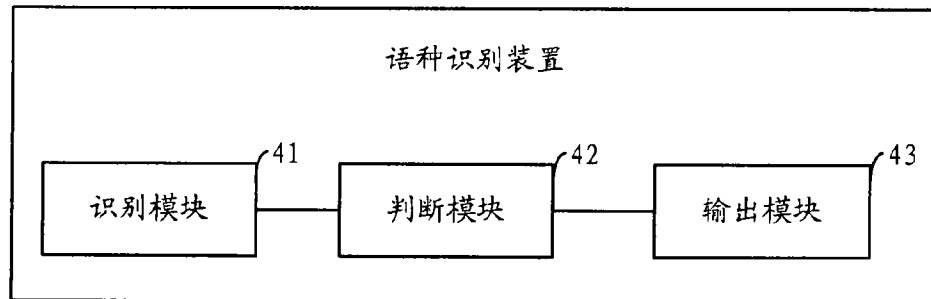


图 4

3/3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/086357

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/00 (2013.01) i; G10L 15/02 (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 15/-; G10L 21/-; G10L 17/-;

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
CNPAT, WPI, EPDOC, CNKI: fragment, beginning, language?, audio?, speech+, voice?, sound?, recogni+, partial, threshold

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105336324 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 17 February 2016 (17.02.2016), claims 1-13, and description, paragraphs [0058]-[0131]	1-18
Y	CN 103853703 A (LENOVO (BEIJING) CO., LTD.), 11 June 2014 (11.06.2014), description, paragraphs [0065]-[0091]	1-18
Y	US 5995928 A (SPEECHWORKS INT. INC.), 30 November 1999 (30.11.1999), abstract, description, column 7, line 33 to column 13, line 39, claims 1 and 2, and figure 2	1-18
Y	LIANG, Chunyan et al., "GSV-SVM based language recognition system", COMPUTER ENGINEERING AND APPLICATIONS, vol. 49, no. 2, 31 December 2013 (31.12.2013), ISSN: 1002-8331, section 2.1, and figure 1	3-6, 8, 11-14, 16
A	CN 103871401 A (LENOVO (BEIJING) CO., LTD.), 18 June 2014 (18.06.2014), the whole document	1-18
A	WO 0118793 A1 (SIEMENS A. G. et al.), 15 March 2001 (15.03.2001), the whole document	1-18

☒ 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	
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"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
"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	"&" document member of the same patent family

Date of the actual completion of the international search 30 August 2016 (30.08.2016)	Date of mailing of the international search report 20 September 2016 (20.09.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer KANG, Dandan Telephone No.: (86-10) 62413593

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/086357**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1892818 A (SIEMENS AG), 10 January 2007 (10.01.2007), the whole document	1-18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/086357

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105336324 A	17 February 2016	None	
CN 103853703 A	11 June 2014	None	
US 5995928 A	30 November 1999	None	
CN 103871401 A	18 June 2014	US 2014163984 A1	12 June 2014
WO 0118793 A1	15 March 2001	EP 1214704 A1	19 June 2002
		EP 1214704 B1	10 March 2004
		DE 50005615 D1	15 March 2001
CN 1892818 A	10 January 2007	EP 1739655 A2	03 January 2007
		DE 102005030380 B4	11 September 2014
		US 2007005358 A1	04 January 2007
		DE 102005030380 A1	04 January 2007

国际检索报告

国际申请号

PCT/CN2016/086357

A. 主题的分类

G10L 15/00(2013.01)i; G10L 15/02(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 15/-; G10L 21/-; G10L 17/-;

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, WPI, EPDOC, CNKI: 语种, 语言, 音频, 声音, 语音, 识别, 辨识, 部分, 片段, 开头, 阈值, 门限, 阈值, language?, audio?, speech+, voice?, sound?, recogni+, partial, threshold

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 105336324 A (百度在线网络技术北京有限公司) 2016年 2月 17日 (2016 - 02 - 17) 权利要求1-13, 说明书第[0058] - [0131]段	1-18
Y	CN 103853703 A (联想北京有限公司) 2014年 6月 11日 (2014 - 06 - 11) 说明书第[0065] - [0091]段	1-18
Y	US 5995928 A (SPEECHWORKS INT. INC.) 1999年 11月 30日 (1999 - 11 - 30) 摘要, 说明书第7栏第33行-第13栏第39行, 权利要求1-2, 图2	1-18
Y	梁春燕等. “高斯超向量-支持向量机鉴别性语种识别系统” 计算机工程与应用, 第49卷, 第2期, 2013年 12月 31日 (2013 - 12 - 31), ISSN: 1002-8331, 第2.1节, 图1	3-6, 8, 11-14, 16
A	CN 103871401 A (联想北京有限公司) 2014年 6月 18日 (2014 - 06 - 18) 全文	1-18
A	WO 0118793 A1 (SIEMENS A.G. 等) 2001年 3月 15日 (2001 - 03 - 15) 全文	1-18

☒ 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

2016年 8月 30日

国际检索报告邮寄日期

2016年 9月 20日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

康丹丹

电话号码 (86-10)62413593

国际检索报告

国际申请号

PCT/CN2016/086357

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 1892818 A（西门子公司）2007年 1月 10日（2007 - 01 - 10） 全文	1-18

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/086357

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	105336324	A	2016年 2月 17日	无			
CN	103853703	A	2014年 6月 11日	无			
US	5995928	A	1999年 11月 30日	无			
CN	103871401	A	2014年 6月 18日	US	2014163984	A1	2014年 6月 12日
WO	0118793	A1	2001年 3月 15日	EP	1214704	A1	2002年 6月 19日
				EP	1214704	B1	2004年 3月 10日
				DE	50005615	D1	2001年 3月 15日
CN	1892818	A	2007年 1月 10日	EP	1739655	A2	2007年 1月 3日
				DE	102005030380	B4	2014年 9月 11日
				US	2007005358	A1	2007年 1月 4日
				DE	102005030380	A1	2007年 1月 4日

表 PCT/ISA/210 (同族专利附件) (2009年7月)