

(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 7 月 6 日 (06.07.2017)



(10) 国际公布号
WO 2017/114172 A1

- (51) 国际专利分类号:
G10L 15/06 (2013.01)
- (21) 国际申请号: PCT/CN2016/110125
- (22) 国际申请日: 2016 年 12 月 15 日 (15.12.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201511016459.1 2015 年 12 月 29 日 (29.12.2015) 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: 及
- (71) 申请人 (仅对美国): 王志铭 (WANG, Zhiming) [CN/CN]; 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)。 李晓辉 (LI, Xiaohui) [CN/CN]; 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)。 李宏言 (LI, Hongyan) [CN/CN]; 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: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, 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, RO,

[见续页]

(54) Title: METHOD AND DEVICE FOR CONSTRUCTING PRONUNCIATION DICTIONARY

(54) 发明名称: 一种发音词典的构建方法及装置

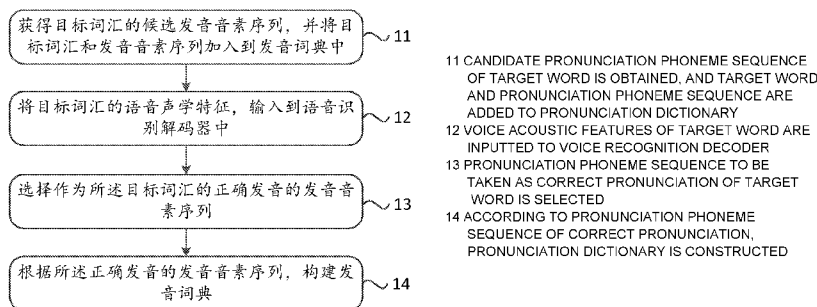


图 1

(57) Abstract: Provided are a method and device for constructing a pronunciation dictionary, used for solving the problem of the poor quality of pronunciation dictionaries constructed according to the prior art. The method comprises: inputting to a voice recognition decoder the voice acoustic features of a target word (12); the pronunciation dictionary in the voice recognition decoder comprising: a target word and a candidate pronunciation phoneme sequence of said target word; according to the candidate pronunciation phoneme sequence outputted by the voice recognition decoder, determining a probability distribution corresponding to the outputted candidate pronunciation phoneme sequence of the target word; according to the probability distribution, selecting from the outputted candidate pronunciation phoneme sequence a pronunciation phoneme sequence to be taken as the correct pronunciation of the target word (13); according to the pronunciation phoneme sequence of the correct pronunciation, constructing a pronunciation dictionary (14).

(57) 摘要: 一种发音词典的构建方法及装置, 用以解决按照现有技术构建的发音词典的质量较差问题。其中, 该方法包括: 将目标词汇的语音声学特征, 输入语音识别解码器 (12); 其中, 语音识别解码器中的发音词典包括: 目标词汇和目标词汇的候选发音音素序列; 根据语音识别解码器输出的候选发音音素序列, 确定目标词汇对应于输出的候选发音音素序列的概率分布; 根据该概率分布, 从输出的候选发音音素序列中, 选择作为目标词汇的正确发音的发音音素序列 (13); 根据正确发音的发音音素序列, 构建发音词典 (14)。

WO 2017/114172 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)。

一种发音词典的构建方法及装置

技术领域

本申请涉及计算机技术领域，尤其涉及一种发音词典的构建方法及装置。

5

背景技术

语音交互技术早在二十世纪中期就已经开始出现，近几年随着智能手机的普及，大量的语音交互产品相继出现，语音交互产品走进了普通用户的日常生活之中。例如，语音输入法就是通过接收并识别用户发出的语音，
10 然后将用户的语音转换成文字，省去了打字的繁琐输入；来电报号功能可以将文字以语音的形式输出，在用户不看屏幕的情况下，即可获知来电方身份。

在语音交互技术中，发音词典是语音交互系统中重要的组成部分，是联接声学模型和语言模型之间的桥梁，其覆盖面和发音质量对系统的整体
15 性能具有重大的影响。

发音词典中包含词和发音音素序列之间的映射关系，通常可以采用词转换为音素（Grapheme-to-Phoneme, G2P）方法建立该映射关系。一般情况下，发音词典经过语言学相关方面的专家审核校正，规模大小相对固定，因此其不可能覆盖所有的词汇，从而在实际应用中，有可能会根据需要，
20 利用 G2P 方法确定新增词汇所匹配的发音音素序列，即确定新增词汇的正确发音，进而根据新增词汇和与其匹配的发音音素序列，对现有的发音词典进行扩充。

目前，采用 G2P 方法，基本能够准确确定常规词汇的正确发音。但是，对于一些特别的词汇，比如包含多音字的词汇，采用该方法确定出的词汇
25 的正确发音的准确度往往较低，从而影响发音词典的质量。

发明内容

本申请实施例提供一种发音词典的构建方法，用以解决按照现有技术

构建的发音词典的质量较差的问题。

本申请实施例还提供一种发音词典的构建装置，用以解决按照现有技术构建的发音词典的质量较差的问题。

本申请实施例采用下述技术方案：

5 一种发音词典的构建方法，包括：

将目标词汇的语音声学特征，输入语音识别解码器；其中，所述语音识别解码器中的发音词典包括：目标词汇和目标词汇的候选发音音素序列；

10 根据所述语音识别解码器以所述语音声学特征作为输入而输出的候选发音音素序列，确定所述目标词汇对应于输出的候选发音音素序列的概率分布；

根据所述概率分布，从所述输出的候选发音音素序列中，选择作为所述目标词汇的正确发音的发音音素序列；

根据所述正确发音的发音音素序列，构建发音词典。

15 一种发音词典的构建装置，包括：

解码单元：用于将目标词汇的语音声学特征，输入语音识别解码器中；其中，所述语音识别解码器中的发音词典包括：目标词汇和目标词汇的候选发音音素序列；

20 发音确定单元：用于根据所述语音识别解码器以所述语音声学特征作为输入而输出的候选发音音素序列，确定所述目标词汇对应于输出的候选发音音素序列的概率分布；根据所述概率分布，从所述输出的候选发音音素序列中，选择作为所述目标词汇的正确发音的发音音素序列；

词典构建单元：用于根据所述正确发音的发音音素序列，构建发音词典。

25 本申请实施例采用的上述至少一个技术方案能够达到以下有益效果：

由于引入了待预测发音的目标词的语音声学特征，作为预测词汇正确发音的依据之一，从而相对于仅依靠词汇和发音音素序列的映射关系来作

为预测词汇正确发音依据的现有技术而言，可以更为准确地预测目标词汇正确发音，提升了基于确定出的正确发音构建的发音词典的质量。

5 附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

图 1 为本申请实施例提供的一种发音词典的构建方法的实现流程示意图；

图 2 为本实施例提供的一种发音词典的构建装置的具体结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

以下结合附图，详细说明本申请各实施例提供的技术方案。

实施例 1

现有的发音预测方法通常是基于 G2P 转换的方法，G2P 方法通过建立词汇和发音音素之间的映射关系，将词汇转换为发音音素序列。采用 G2P 方法，基本能够准确得到与常规词汇匹配的发音音素序列，但是由于该方法只利用了词汇（字序列）和发音音素的映射关系，因此对于一些特别的词汇，比如包含多音字的词汇，采用该方法确定出的与词汇匹配的发音音素序列的准确度往往较低，从而影响发音词典的质量。

为解决由于现有技术不能准确预测词汇的正确发音从而影响发音词典的质量的问题，本申请实施例 1 提供了一种发音词典的构建方法。

本申请实施例提供的发音词典的构建方法的执行主体可以是服务器也可以是不同于服务器的其他设备，等等。所述的执行主体并不构成对本
5 申请的限定，为了便于描述，本申请实施例均以执行主体是服务器为例进行说明。

为便于描述，在本实施例中，存在对应关系的词汇和语音声学特征可以用词汇-语音声学特征来表示。

类似的，存在对应关系的词汇（字序列）和音素序列，以及存在对应
10 关系的语音声学特征和语音音素序列，也可用上述表示方式表示。例如，存在对应关系的词汇和音素序列，可以用词汇-语音音素序列来表示。

以下对本申请实施例提供该方法进行详细介绍。

该方法的实现流程示意图如图 1 所示，包括下述步骤：

步骤 11：服务器将目标词汇的语音声学特征，输入到嵌入有发音词典、
15 声学模型和语言模型的语音识别解码器；

本申请实施例中，所述的目标词汇，可以是任何词汇，比如中文词汇、英文词汇或者其他语言的词汇。若针对语音识别解码器中已有的发音词典而言，所述的目标词汇，可以是指该发音词典当前不包含的词汇，即相对于该发音词典的新增词汇。

本申请实施例中所述的目标词汇的语音声学特征，可以但不限于包括
20 从说出该目标词汇所产生的语音信号中提取出的 Filter Bank 特征、MFCC（Mel Frequency Cepstrum Coefficient）特征以及 PLP（Perceptual Linear Predictive）特征等等中的至少一种。

本申请实施例中，所述的语音信号，比如可以是根据目标词汇对应的
25 音频样本。

目标词汇对应的音频样本，可以但不限于采用下述方式中的至少一种获得的：

一、委托专业的语音数据供应商进行人工录音，从而获得目标词汇对

应的音频样本;

二、采用众包的形式,以用户的真实使用感受和切身体验为出发点,将录音任务以自由自愿的形式委托给非特定的(而且通常是大型的)网络大众,从而获得目标词汇对应的音频样本;

5 三、分析用户反馈的记录日志,从而获得目标词汇对应的音频样本。例如,在语音搜索任务中,用户先通过语音输入目标词汇,如果语音识别系统识别错误,用户继续通过键盘输入正确的目标词汇,这一系列的行为可以通过日志的形式记录下来。

10 本申请实施例中,可以从目标词汇对应的音频样本中分别获得语音声学特征,进而将获得的各语音声学特征作为所述目标词汇的语音声学特征,分别输入所述语音识别解码器。

以下进一步介绍步骤 11 中提及的语音识别解码器的工作原理。

15 一般地,语音识别解码器,是用于针对输入的语音信号(或语音声学特征),根据声学模型、语言模型及发音词典,寻找能够以最大概率发出该语音信号(或与该语音声学特征相匹配的语音信号)的词的虚拟或者实体设备。

20 在语音识别领域,对语音信号进行解码的目标,就是寻找字序列 W^* (相当于上文所述的“词”),使得对应的语音声学特征 X 似然概率最大化,实质上就是一个基于贝叶斯准则的机器学习问题,即利用贝叶斯公式来计算最佳字序列 W^* ,如公式[1.1]所示:

$$W^* = \arg \max_i P(W_i | X) = \arg \max_i P(X | W_i) P(W_i)$$

[1.1]

其中 $P(X | W_i)$ 为声学模型, $P(W_i)$ 为语言模型。

25 声学模型,是字序列 W_i 的语音声学特征为 X 的概率。一般可以利用大量的数据(包括语音声学特征以及对应的标签序列)训练得到声学模型。

语言模型,是词汇对应的字序列 W_i 的出现概率。该出现概率的含义一般为:构成词汇的各个字依照所述各个字在该词汇中的排列顺序依次出现的概率。

考虑到字序列一般会对应的不同的发音音素序列，比如用带不同地方口音发出某个词汇（可由字序列表示）的发音可能对应不同的发音音素，又或者包含多音字的词汇也有可能对应不同的发音音素，因此，若假设 Q_i^j 是字序列 W_i 对应的各发音音素序列，那么公式[1.1]可变为：

$$5 \quad W^* = \arg \max_i P(X | W_i) P(W_i) = \arg \max_{i,j} P(X | Q_i^j) P(Q_i^j | W_i) P(W_i) \quad [1.2]$$

其中， W_i 为字序列； $P(X | Q_i^j)$ 为声学模型； $P(W_i)$ 为语言模型； $P(Q_i^j | W_i)$ 为发音词典中的词汇（由字序列 W_i 表示）的发音音素序列为 Q_i^j 的概率。

对于发音学习的问题，进一步假定字序列 W_i 和对应的语音声学特征 X 是已知的，则公式[1.2]的计算目标，可以转换是为了寻找字序列 W_i 对应的最佳发音音素序列 Q^* 。这样，公式[1.2]进一步可变为：

$$10 \quad Q^* = \arg \max_j P(X | Q_i^j) P(Q_i^j | W_i) P(W_i) = \arg \max_j P(X | Q_i^j) P(Q_i^j | W_i) \quad [1.3]$$

公式[1.3]中：

Q^* 为使得公式[1.3]中等号右侧的值最大的发音音素序列，也即字序列 W_i 对应的候选发音音素序列的概率分布的最大值；

W_i 为字序列， i 为词汇的编号；

X 表示 W_i 对应的语音声学特征；

Q 表示发音音素序列；

j 为发音音素序列的编号；

20 Q_i^j 表示编号为 i 的词汇对应的语音音素序列中的、编号为 j 的发音音素序列。

$P(X | Q_i^j)$ 为声学模型，即发音音素序列 Q_i^j 对应的语音声学特征为 X 的概率。

目前，相关的语音识别技术中用到的声学模型一般是对隐马尔科夫-深度神经网络（Hidden Markov Model- Deep Neural Network, HMM-DNN）

的混合模型训练得到的，或者也可以是对 DNN 模型训练得到的。本申请实施例中，可以预先通过海量语音声学特征，对 HMM-DNN 的混合模型或 DNN 模型进行训练得到声学模型，并设置在本申请实施例所述的该语音识别解码器中。

5 $P(W_i)$ 为语言模型——本实施例中的语言模型可以是 N-Gram 模型，该模型基于这样一种假设，第 N 个词的出现只与前面 N-1 个词相关，而与其它任何词都不相关，整句的概率就是各个词出现概率的乘积，各个词出现的概率可以通过直接从语料中统计 N 个词同时出现的次数得到。本实施例中的语言模型也可以是基于条件随机场或者基于深度神经网络策略的语言模型。该语言模型可以预先生成并设置在本申请实施例所述的该语音识别解码器中。

$P(Q_i^j | W_i)$ 为基于给定的发音词典中的词汇（由字序列 W_i 表示）的发音音素序列为 Q_i^j 的概率。

15 这里所说的发音词典，比如可以为加入了目标词汇对应的各个候选发音音素序列的发音词典。

目标词汇的候选发音音素序列，是指可能作为目标词汇正确发音的发音音素序列。本申请实施例中，可以但不限于采用 G2P 方法，为目标词汇生成发音音素序列（本申请实施例中称“候选发音音素序列”），并将所述目标词汇和生成的各候选发音音素序列，加入到发音词典中。

20 其中，将所述目标词汇和生成的各候选发音音素序列，加入到发音词典中，可以是指，将包含目标词汇-候选发音音素序列的词条，添加到发音词典中。

需要说明的是，当当前不存在发音词典时，将所述词条添加到发音词典中，可以是指根据所述词条构建发音词典；当当前已存在发音词典时，
25 将所述词条添加到发音词典中，可以是指根据所述词条对该已有的发音词典进行更新，得到更新后的发音词典。

以便于描述，本申请实施例中假设当前已存在发音词典。在这样的场

景下，所述目标词汇为相对于当前已存在的发音词典而言的新增词汇。

本实施例中，为目标词汇生成的对应的候选发音音素序列的个数视实际情况而定。

如，采用 G2P 方法，可以为目标词汇“阿里巴巴”生成十个以上候选的
5 发音音素序列。以该些发音音素序列中的某一个发音音素序列为例，其可以表示为“a1/li3/ba1/ba1/”。该发音音素序列中，符号“/”用于区分不同发音音素，即“/”前后的符号表示不同的音素。比如，a1 和 li3 为不同音素。音素中的数字代表声调，即 1 代表声调一声，2 代表声调二声，3 代表声调三声，4 代表声调四声。

10 基于嵌入有上述发音词典、公式[1.3]中所示的声学模型 $P(X|Q_i')$ 和语言模型 $P(W_i)$ 的语音识别解码器，本申请实施例中，将目标词汇的语音声学特征输入到该语音识别解码器中，可以触发该语音识别解码器通过对语音样本声学特征的解码，输出该语音样本声学特征对应的发音音素序列。

以下进一步介绍本申请实施例提供的该方法包含的后续步骤。

15 步骤 12：确定语音识别解码器以步骤 11 中所述的语音声学特征作为输入而输出的候选发音音素序列；并根据目标词汇对应于所述输出的候选发音音素序列的统计规律，确定目标词汇对应于输出的候选发音音素序列的概率分布；根据所述概率分布，从所述输出的候选发音音素序列中，选择作为目标词汇的正确发音的发音音素序列；

20 比如，若假定目标词汇 T 对应的候选发音音素序列有 2 个，分别为 A1 A2 和 B1 B2，且它们被添加到语音识别解码器包含的发音词典中。进一步地，若假设采集到的 T 的音频样本有 100 个，从而可以获得这 100 个音频样本各自的语音声学特征（共 100 个语音声学特征），通过执行步骤 11，将这 100 个语音声学特征分别输入到嵌入发音词典、声学模型和语言
25 模型的语音识别解码器中。

那么，语音识别解码器对这 100 个语音声学特征进行识别解码，可以

输出候选发音音素序列，如输出 A1、A2、B1、B2 的组合。

进一步地，假设根据设置于该语音识别解码器中的发音词典，确定目标词汇对应于所述输出的候选发音音素序列的统计规律为：

这 100 个语音声学特征中：有 75 个语音声学特征是通过发音词典的
5 词条“T-A1 A2”映射到 T，有 25 个语音声学特征是通过发音词典的词条
“T-B1 B2”映射到 T。

那么，根据该统计规律，可以得到如下概率分布：

T 对应于 A1 A2 的概率为 $75/100=0.75$

T 对应于 B1 B2 的概率为 $25/100=0.25$

10 一般地，服务器可以将所述概率分布中的最大概率值对应的候选发音音素序列，确定为所述目标词汇正确的发音的发音音素序列。

沿用上例，则服务器可以将所述概率分布中的最大概率值 0.75 对应的候选发音音素序列 A1 A2，确定为 T 正确发音的发音音素序列。

步骤 13: 根据作为目标词汇正确发音的发音音素序列，构建发音词典。

15 具体地，服务器比如可以从加入了目标词汇对应的各个候选发音音素序列的发音词典中，删除除作为目标词汇正确发音的发音音素序列外的、与该目标词汇对应的其他候选发音音素序列。或者，服务器也可以根据作为目标词汇正确发音的发音音素序列，重新构建新的发音词典。

20 采用本申请实施例 1 提供的上述方法，由于引入了待预测发音的目标词的语音声学特征，作为预测词汇正确发音的依据之一，从而相对于仅依靠词汇和发音音素序列的映射关系来做为预测词汇正确发音依据的现有技术而言，可以更为准确地预测目标词汇正确发音，从而提升了语音词典的质量。

25 实施例 2

为解决采用现有技术会导致与词汇匹配的发音音素序列的准确性较低的问题，本申请实施例提供一种发音词典的构建装置。该词汇发音预测装置的结构示意图如图 3 所示，主要包括下述功能单元：

解码单元 21, 用于将目标词汇的语音声学特征, 输入语音识别解码器中; 其中, 语音识别解码器中的发音词典包括: 目标词汇和目标词汇的候选发音音素序列;

5 发音确定单元 22, 用于根据所述语音识别解码器以所述语音声学特征作为输入而输出的候选发音音素序列, 确定所述目标词汇对应于输出的候选发音音素序列的概率分布; 根据所述概率分布, 从所述输出的候选发音音素序列中, 选择作为所述目标词汇的正确发音的发音音素序列;

词典构建单元 23, 用于根据所述正确发音的发音音素序列, 构建发音词典。

10 在一种实施方式中, 本申请实施例提供的该装置还可以包括音素序列处理单元。该单元用于在于将目标词汇的语音声学特征, 输入语音识别解码器中前, 获得目标词汇的候选发音音素序列; 并将目标词汇和获得的候选发音音素序列, 加入到所述语音识别解码器中的发音词典中。

15 在一种实施方式中, 音素序列处理单元, 具体可以用于利用 G2P 方法, 获得目标词汇的候选发音音素序列。

在一种实施方式中, 所述解码单元 21, 具体可以用于采集目标词汇对应的音频样本; 根据所述音频样本, 获得所述语音声学特征; 将获得的所述语音声学特征, 输入所述语音识别解码器中。

20 在一种实施方式中, 所述发音确定单元 22, 具体可以用于确定所述概率分布中的最大概率值; 从所述输出的候选发音音素序列中, 选择所述最大概率值对应的候选发音音素序列, 作为所述目标词汇的正确发音的发音音素序列。

25 在一种实施方式中, 所述词典构建单元 23, 具体可以用于根据作为所述目标词汇正确发音的发音音素序列, 从加入了目标词汇和获得的候选发音音素序列的发音词典中, 删除目标词汇对应的、除所述正确发音的发音音素序列外的其他候选发音音素序列。

采用本申请实施例 2 提供的上述装置, 由于引入了待预测发音的目标词的语音声学特征, 作为预测词汇正确发音的依据之一, 从而相对于仅依

靠词汇和发音音素序列的映射关系来做为预测词汇正确发音依据的现有技术而言，可以更为准确地预测目标词汇正确发音。

本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施
5 例、或结合软件和硬件方面的实施例的形式。而且，本发明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。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程
10 序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令
15 产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一
20 个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定
25 的功能的步骤。

以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权 利 要 求 书

1、一种发音词典的构建方法，其特征在于，所述方法包括：

将目标词汇的语音声学特征，输入语音识别解码器；其中，所述语音识别解码器中的发音词典包括：目标词汇和目标词汇的候选发音音素序列；

根据所述语音识别解码器以所述语音声学特征作为输入而输出的候选发音音素序列，确定所述目标词汇对应于输出的候选发音音素序列的概率分布；

根据所述概率分布，从所述输出的候选发音音素序列中，选择作为所述目标词汇的正确发音的发音音素序列；

根据所述正确发音的发音音素序列，构建发音词典。

2、如权利要求 1 所述的方法，其特征在于，将所述语音声学特征，输入所述语音识别解码器前，所述方法还包括：

获得目标词汇的候选发音音素序列；

将目标词汇和获得的候选发音音素序列，加入到所述语音识别解码器中的发音词典中。

3、如权利要求 2 所述的方法，其特征在于，获得目标词汇的候选发音音素序列，包括：

利用词转换为音素 G2P 方法，获得目标词汇的候选发音音素序列。

4、如权利要求 1 所述的方法，其特征在于，所述语音识别解码器中嵌入的声学模型，是对深度神经网络进行训练得到的。

5、如权利要求 1 所述的方法，其特征在于，将目标词汇的语音声学特征，输入所述语音识别解码器中，包括：

采集目标词汇对应的音频样本；

根据所述音频样本，获得所述语音声学特征；

将获得的所述语音声学特征，输入所述语音识别解码器中。

6、如权利要求 1 所述的方法，其特征在于，根据所述概率分布，从

所述输出的候选发音音素序列中，选择作为所述目标词汇的正确发音的发音音素序列，包括：

确定所述概率分布中的最大概率值；

5 从所述输出的候选发音音素序列中，选择所述最大概率值对应的候选发音音素序列，作为所述目标词汇的正确发音的发音音素序列。

7、如权利要求 1~6 任一权项所述的方法，其特征在于，根据所述正确发音的发音音素序列，构建发音词典，包括：

10 根据作为所述目标词汇正确发音的发音音素序列，从加入了目标词汇和获得的候选发音音素序列的发音词典中，删除目标词汇对应的、除所述正确发音的发音音素序列外的其他候选发音音素序列。

8、一种发音词典的构建装置，其特征在于，所述装置包括：

解码单元：用于将目标词汇的语音声学特征，输入语音识别解码器中；其中，所述语音识别解码器中的发音词典包括：目标词汇和目标词汇的的候选发音音素序列；

15 发音确定单元：用于根据所述语音识别解码器以所述语音声学特征作为输入而输出的候选发音音素序列，确定所述目标词汇对应于输出的候选发音音素序列的概率分布；根据所述概率分布，从所述输出的候选发音音素序列中，选择作为所述目标词汇的正确发音的发音音素序列；

20 词典构建单元：用于根据所述正确发音的发音音素序列，构建发音词典。

9、如权利要求 8 所述的装置，其特征在于，所述装置还包括：

音素序列处理单元，用于在于将目标词汇的语音声学特征，输入语音识别解码器中前，获得目标词汇的候选发音音素序列；并将目标词汇和获得的候选发音音素序列，加入到所述语音识别解码器中的发音词典中。

25 10、如权利要求 9 所述的装置，其特征在于，所述音素序列处理单元，具体可以用于：

利用词转换为音素 G2P 方法，获得目标词汇的候选发音音素序列。

11、如权利要求 8 所述的装置，其特征在于，所述语音识别解码器中嵌入的声学模型，是对深度神经网络进行训练得到的。

12、如权利要求 8 所述的装置，其特征在于：

5 所述解码单元，具体用于采集目标词汇对应的音频样本；根据所述音频样本，获得所述语音声学特征；将获得的所述语音声学特征，输入所述语音识别解码器中。

13、如权利要求 8 所述的装置，其特征在于，所述发音确定单元，具体用于：

确定所述概率分布中的最大概率值；

10 从所述输出的候选发音音素序列中，选择所述最大概率值对应的候选发音音素序列，作为所述目标词汇的正确发音的发音音素序列。

14、如权利要求 8~13 任一权项所述的装置，其特征在于：

15 所述词典构建单元，具体用于根据作为所述目标词汇正确发音的发音音素序列，从加入了目标词汇和获得的候选发音音素序列的发音词典中，删除目标词汇对应的、除所述正确发音的发音音素序列外的其他候选发音音素序列。

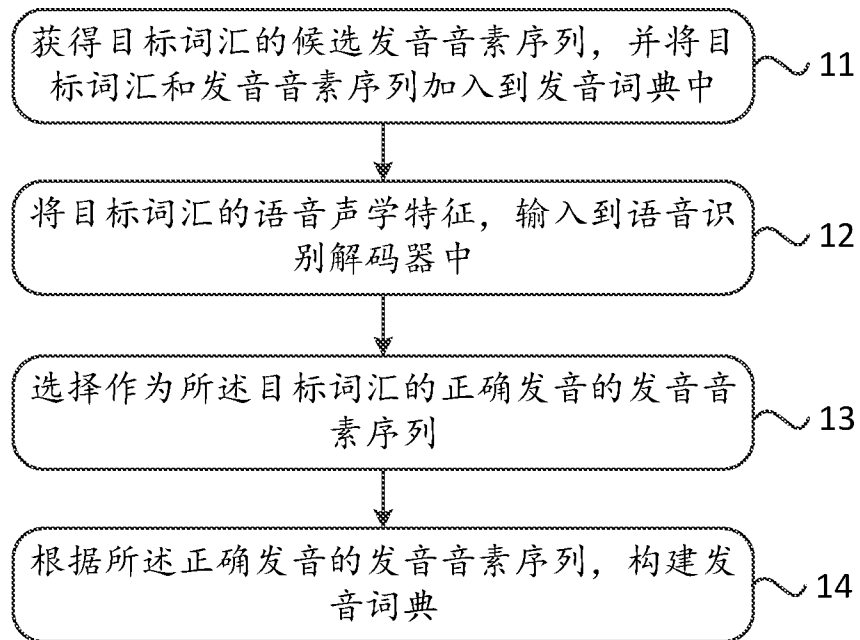


图 1

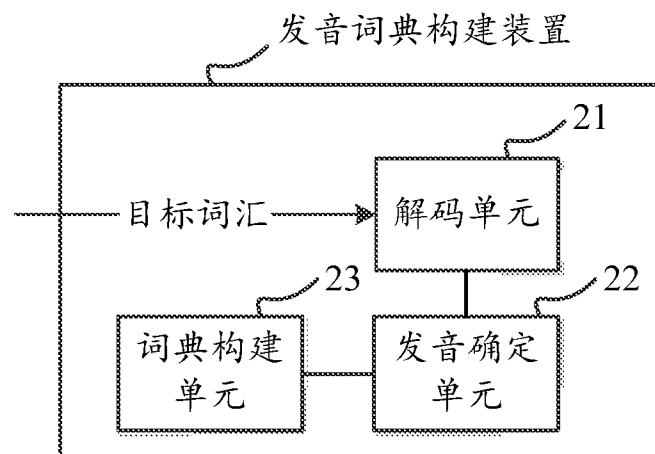


图 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/110125

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/06 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L; 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)

CN^S, C^TXT, CNKI^VEN: dictionary, lexicon+, construct+, build+, mak+ up, updat+, renew+, expan+, exten +, optimi+, add+, increas+, modif+, correct+, enganc+, precision, accuracy, quality, veracity, delete, acoustic feature, pronounce, voice, acoustics, phoneme

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	JP 2016011995 A (IBM CORP.) 21 January 2016 (21.01.2016) description, paragraphs [0041]-[0073], and figures 2-9	1-14
Y	US 2010312550 A1 (LEE GIL HO) 09 December 2010 (09.12.2010) description, paragraphs [0039]-[0076], and figures 1-6	1-14
Y	CN 1667700 A (MICROSOFT CORPORATION) 14 September 2005 (14.09.2005) description, page 6, line 30 to page 8, line 10, and figure 3	1-14
A	WO 2014209449 A1 (GOOGLE INC.) 31 December 2014 (31.12.2014) the whole document	1-14
A	US 7280963 B1 (NUANCE COMMUNICATIONS INC.) 09 October 2007 (09.10.2007) the whole document	1-14

☒ 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 14 March 2017	Date of mailing of the international search report 24 March 2017
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 YANG, Shilin Telephone No. (86-10) 62085717

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/110125

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 102201235 A (MITSUBISHI ELECTRIC CORP.) 28 September 2011 (28.09.2011) the whole document	1-14
A	CN 101432801 A (NEC CORP.) 13 May 2009 (13.05.2009) the whole document	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/110125

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2016011995 A	21 January 2016	None	
US 2010312550 A1	09 December 2010	KR 20100130263 A	13 December 2010
		KR 101587866 B1	25 January 2016
		US 8645139 B2	04 February 2014
CN 1667700 A	14 September 2005	EP 1575030 A1	14 September 2005
		DE 602005001125 T2	13 September 2007
		US 2005203738 A1	15 September 2005
		CN 1667700 B	06 October 2010
		US 7590533 B2	15 September 2009
		EP 1575030 B1	16 May 2007
		JP 2005258443 A	22 September 2005
		DE 602005001125 D1	28 June 2007
		AT 362633 T	15 June 2007
		KR 20060043845 A	15 May 2006
		IN 200500451 I1	08 December 2006
WO 2014209449 A1	31 December 2014	CN 105340004 A	17 February 2016
		US 2015006178 A1	01 January 2015
		EP 3014608 A1	04 May 2016
US 7280963 B1	09 October 2007	None	
CN 102201235 A	28 September 2011	JP 2011209704 A	20 October 2011
		EP 2378514 A1	19 October 2011
		US 2011238412 A1	29 September 2011
CN 101432801 A	13 May 2009	WO 2007097176 A1	30 August 2007
		CN 101432801 B	18 April 2012
		US 2009024392 A1	22 January 2009
		JP 5040909 B2	03 October 2012
		US 8719021 B2	06 May 2014

国际检索报告

国际申请号

PCT/CN2016/110125

A. 主题的分类

G10L 15/06(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L; 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))

CNABS, CNTXT, CNKI, VEN: 词典, 字典, 建, 更新, 增, 扩充, 优化, 增补, 修补, 添加, 校正, 删, 声学特征, 发音, 语音, 声学, 音素: dictionary, lexicon+, construct+, build+, mak+ up, updat+, renew+, expan+, exten+, optimi+, add+, increas+, modif+, correct+, enganc+, precision, accuracy, quality, veracity

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	JP 2016011995 A (IBM CORP.) 2016年 1月 21日 (2016 - 01 - 21) 说明书第[0041]-[0073]段, 附图2-9	1-14
Y	US 2010312550 A1 (LEE GIL HO) 2010年 12月 9日 (2010 - 12 - 09) 说明书第[0039]-[0076]段, 附图1-6	1-14
Y	CN 1667700 A (微软公司) 2005年 9月 14日 (2005 - 09 - 14) 说明书第6页第30行-第8页第10行, 附图3	1-14
A	WO 2014209449 A1 (GOOGLE INC.) 2014年 12月 31日 (2014 - 12 - 31) 全文	1-14
A	US 7280963 B1 (NUANCE COMMUNICATIONS INC.) 2007年 10月 9日 (2007 - 10 - 09) 全文	1-14
A	CN 102201235 A (三菱电机株式会社) 2011年 9月 28日 (2011 - 09 - 28) 全文	1-14
A	CN 101432801 A (日本电气株式会社) 2009年 5月 13日 (2009 - 05 - 13) 全文	1-14

☐ 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

2017年 3月 14日

国际检索报告邮寄日期

2017年 3月 24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

杨士林

电话号码 (86-10)62085717

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/110125

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)
JP	2016011995	A	2016年 1月 21日	无			
US	2010312550	A1	2010年 12月 9日	KR	20100130263	A	2010年 12月 13日
				KR	101587866	B1	2016年 1月 25日
				US	8645139	B2	2014年 2月 4日
CN	1667700	A	2005年 9月 14日	EP	1575030	A1	2005年 9月 14日
				DE	602005001125	T2	2007年 9月 13日
				US	2005203738	A1	2005年 9月 15日
				CN	1667700	B	2010年 10月 6日
				US	7590533	B2	2009年 9月 15日
				EP	1575030	B1	2007年 5月 16日
				JP	2005258443	A	2005年 9月 22日
				DE	602005001125	D1	2007年 6月 28日
				AT	362633	T	2007年 6月 15日
				KR	20060043845	A	2006年 5月 15日
				IN	200500451	I1	2006年 12月 8日
WO	2014209449	A1	2014年 12月 31日	CN	105340004	A	2016年 2月 17日
				US	2015006178	A1	2015年 1月 1日
				EP	3014608	A1	2016年 5月 4日
US	7280963	B1	2007年 10月 9日	无			
CN	102201235	A	2011年 9月 28日	JP	2011209704	A	2011年 10月 20日
				EP	2378514	A1	2011年 10月 19日
				US	2011238412	A1	2011年 9月 29日
CN	101432801	A	2009年 5月 13日	WO	2007097176	A1	2007年 8月 30日
				CN	101432801	B	2012年 4月 18日
				US	2009024392	A1	2009年 1月 22日
				JP	5040909	B2	2012年 10月 3日
				US	8719021	B2	2014年 5月 6日

表 PCT/ISA/210 (同族专利附件) (2009年7月)