

# (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11月23日 (23.11.2017)



(10) 国际公布号  
WO 2017/197809 A1

- (51) 国际专利分类号:  
*G10L 13/047* (2013.01) *G06F 17/28* (2006.01)  
*G10L 13/06* (2013.01)
- (21) 国际申请号: PCT/CN2016/098126
- (22) 国际申请日: 2016年9月5日 (05.09.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201610329738.1 2016年5月18日 (18.05.2016) 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: 李昊(LI, Hao); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。康永国(KANG, Yongguo); 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: 北京清亦华知识产权代理事务所(普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区清华园清华大学照澜院商业楼301室, Beijing 100084 (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,

(54) Title: SPEECH SYNTHESIS METHOD AND SPEECH SYNTHESIS DEVICE

(54) 发明名称: 语音合成方法和语音合成装置

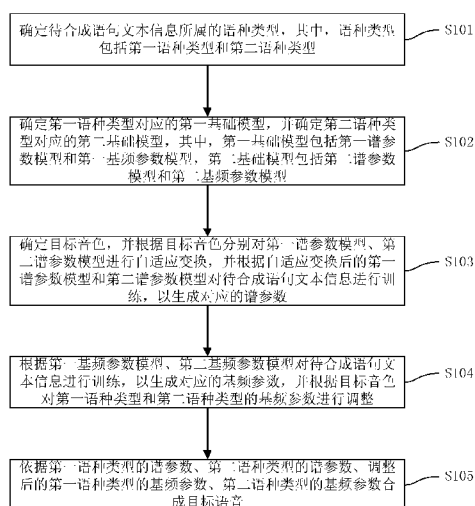


图 1

- S101 Determine a language type of statement text information to be synthesized, the language type comprising a first language type and a second language type
- S102 Determine a first base model corresponding to the first language type and a second base model corresponding to the second language type, the first base model comprising a first spectrum parameter model and a first fundamental frequency parameter model, and the second base model comprising a second spectrum parameter model and a second fundamental frequency parameter model
- S103 Determine a target timbre, separately perform adaptive transformation on the first spectrum parameter model and the second spectrum parameter model according to the target timbre, and train the statement text information to be synthesized according to the first and second spectrum parameter models subjected to adaptive transformation to generate corresponding spectrum parameters
- S104 Train the statement text information to be synthesized according to the first and second fundamental frequency parameter models to generate corresponding fundamental frequency parameters, and adjust the fundamental frequency parameters of the first language type and the second language type according to the target timbre
- S105 Synthesize a target speech according to the spectrum parameters and adjusted fundamental frequency parameters of the first and second language types

(57) Abstract: A speech synthesis method and device. The method comprises: determining a language type of statement text information to be synthesized, the language type comprising a first language type and a second language type (S101); determining a first base model corresponding to the first language type and a second base model corresponding to the second language type (S102); determining a target timbre, separately performing adaptive transformation on the first base model and the second base model according to the target timbre, and training the statement text information to be synthesized according to the first and second base models subjected



WO 2017/197809 A1

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, 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))。

---

to adaptive transformation to generate corresponding spectrum parameters and fundamental frequency parameters (S103); adjusting the fundamental frequency parameters of the first language type and the second language type according to the target timbre (S104); and synthesizing a target speech according to the spectrum parameters and adjusted fundamental frequency parameters of the first and second language types (S105).

**(57) 摘要:** 一种语音合成方法以及装置, 该方法包括: 确定待合成语句文本信息所属的语种类型, 其中, 语种类型包括第一语种类型和第二语种类型 (S101); 确定第一语种类型对应的第一基础模型, 并确定第二语种类型对应的第二基础模型 (S102); 确定目标音色, 并根据目标音色分别对第一基础模型、第二基础模型进行自适应变换, 并根据自适应变换后的第一基础模型、第二基础模型对待合成语句文本信息进行训练, 以生成对应的谱参数和基频参数 (S103); 根据目标音色对第一语种类型和第二语种类型的基频参数进行调整 (S104); 依据第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数合成目标语音 (S105)。

## 语音合成方法和语音合成装置

### 相关申请的交叉引用

本申请要求百度在线网络技术（北京）有限公司于 2016 年 5 月 18 日提交的、发明名称为“语音合成方法和语音合成装置”的、中国专利申请号“201610329738.1”的优先权。

### 技术领域

本发明涉及语音合成技术领域，尤其涉及一种语音合成方法和语音合成装置。

### 10 背景技术

随着语音合成技术的发展和应用的普及，语音合成业务正越来越多的被用户接受和使用。在语音合成业务的用户中，有很大一部分是双语言或多语言用户，而语音合成也越来越多的应用到多语种内容的场合。因此产生了多语种语音合成的需求，其中，尤其以中英混读最为普遍。用户对多语种语音合成通常的要求首先是易懂，其次是发音准确、自然且音色统一。在当前语音合成技术已经基本解决易懂度的情况下，如何合成自然、准确的、音色统一多语种语音，成了语音合成的一个技术难题。

相关技术中，通常在涉及到多语种合成的场合会采用分别对不同语言使用不同说话人数据建模的方式，或寻求各种语言均发音比较标准的发音人数据进行建模。

但是目前存在的问题是：（1）针对不同语言使用不同母语说话人的数据的方法，会造成合成音色不统一的问题，以至于影响语音合成的自然度和用户体验；（2）采用多语种说话人数据的方法，大多数发音人除母语外的语言并不地道，带有口音，与母语说话人有较大差距，降低用户体验，而采用这样的数据合成的语音除说话人母语外，读音均不够标准，而多种语言均标准的发音人通常是专业人员，数据采集成本又较高。

因此，如何低成本、高效率地将多种语言文本合成自然、准确的、音色统一的多语种语音已经成为亟待解决的问题。

### 发明内容

本发明的目的旨在至少在一定程度上解决上述的技术问题之一。

为此，本发明的第一个目的在于提出一种语音合成方法。该方法可以降低语言基础模型建立的数据成本和实现难度，减少了多语种合成对专业的多语发音人数据的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，提升了用户体验。

本发明的第二个目的在于提出一种语音合成装置。

本发明的第三个目的在于提出一种终端。

本发明的第四个目的在于提出一种存储介质。

为达到上述目的，本发明第一方面实施例的语音合成方法，包括：确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和第二语种类型；确定所述第一语种类型对应的第一基础模型，并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；确定目标音色，并根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换，并根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数，并根据所述目标音色对所述第一语种类型和第二语种类型的基频参数进行调整；依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

本发明实施例的语音合成方法，确定待合成语句文本中包含哪些语种类型，之后根据目标音色对各语种类型的谱参数模型进行自适应训练，并采用自适应训练后的谱参数模型生成对应的谱参数，并根据目标音色将生成的各语种类型的基频参数进行调整，以得到音色统一的多语种语音，可以理解，上述语言基础模型利用的大多是单语数据建立的模型，降低了模型建立的数据成本和实现难度，减少了多语种合成对专业的多语发音人数据的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，提升了用户体验。

为达到上述目的，本发明第二方面实施例的语音合成装置，包括：第一确定模块，用于确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和第二语种类型；第二确定模块，用于确定所述第一语种类型对应的第一基础模型，并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；第三确定模块，用于确定目标音色；自适应变换模块，用于根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换；谱参数生成模块，用于根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；基频参数生成模块，用于根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数；基频参数调整模块，用于根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，

以生成对应的基频参数；语音合成模块，用于依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

5 本发明实施例的语音合成装置，确定待合成语句文本中包含哪些语种类型，之后根据目标音色对各语种类型的谱参数模型进行自适应训练，并采用自适应训练后的谱参数模型生成对应的谱参数，并根据目标音色将生成的各语种类型的基频参数进行调整，以得到音色统一的多语种语音，可以理解，上述语言基础模型利用的大多是单语数据建立的模型，降低了模型建立的数据成本和实现难度，减少了多语种合成对专业的多语发音人数据的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，  
10 提升了用户体验。

为达到上述目的，本发明第三方面实施例的终端，包括：一个或者多个处理器；存储器；一个或多个程序，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进行如下操作：确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和第二语种类型；确定所述第一语种类型对应的第一基础模型，  
15 并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；确定目标音色，并根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换，并根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；根据所述第一基频参数模块、第二  
20 基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数，并根据所述目标音色对所述第一语种类型和第二语种类型的基频参数进行调整；依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

为达到上述目的，本发明第四方面实施例的存储介质，用于存储应用程序，所述应用程序用于执行本发明第一方面实施例所述的语音合成方法。  
25

本发明附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

## 附图说明

30 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 是根据本发明一个实施例的语音合成方法的流程图；

图 2 是根据本发明一个具体实施例的语音合成方法的流程图；

图 3 是根据本发明一个实施例的语音合成方法的示例图；

图 4 是根据本发明一个实施例的语音合成装置的结构框图；

图 5 是根据本发明一个具体实施例的语音合成装置的结构框图；

5 图 6 是根据本发明另一个具体实施例的语音合成装置的结构框图。

## 具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施

10 所述的实施例是示例性的，旨在用于解释本发明，而不能理解为对本发明的限制。

可以理解，在日常生活中，多语种语音合成应用已经逐渐被人们所需要，例如，以移动终端中的新闻应用程序为例，当用户使用新闻应用程序通过语音合成的功能听取新闻时，新闻内容，尤其是科技新闻，除中文外还夹杂有大量的英文，因此这一应用是典型的多语种语音合成，然而合成语音的自然度、准确度以及音色是否统一都会对用户体验造成影响。

15 为此，本发明提出了一种语音合成方法和装置，以有效地解决发音准确性和统一音色的问题。具体地，下面参考附图描述本发明实施例的语音合成方法和语音合成装置。

图 1 是根据本发明一个实施例的语音合成方法的流程图。需要说明的是，本发明实施例的语音合成方法可应用于移动终端（如手机、平板电脑、个人数字助理等）、终端（如 PC 机）等具有语音合成功能的电子设备中。此外，本发明实施例的语音合成方法适用于多种语言文本合成多种语种语音的场景。

20 多种语言文本合成多种语种语音的场景。

如图 1 所示，该语音合成方法可以包括：

S101，确定待合成语句文本信息所属的语种类型，其中，语种类型包括第一语种类型和第二语种类型。

具体地，可先获取待合成语句文本信息，该待合成语句文本信息可理解为待合成语句

25 文本的文本内容，之后，可对该待合成语句文本信息进行语言判别，以确定待合成语句文本信息中语句文本所属的语种类型。

作为一种示例，可根据语种的字符和上下文内容信息来对待合成语句文本信息中的文字内容进行断句，并判断出每个断句片段的语种类型，该语种类型可包括第一语种类型和第二语种类型，其中，该第二语种类型可为一个或多个，也就是说，该待合成语句文本中

30 语句文本所属的语种类型可以是两种语种，也可以是三种或三种以上语种类型。

S102，确定第一语种类型对应的第一基础模型，并确定第二语种类型对应的第二基础模型，其中，第一基础模型包括第一谱参数模型和第一基频参数模块，第二基础模型包括

第二谱参数模型和第二基频参数模块。

具体地，在确定待合成语句文本信息中所属哪些语种类型之后，可确定这些语种类型所对应的语言基础模型。例如，以待合成语句文本包括中英文混合语句文本为例，则可确定该中英文混合语句文本所属的语音类型包括中文语种类型和英文语种类型，之后，可确定中文语种类型对应的中文基础模型和英文语种类型对应的英文基础模型。

可以理解，每种语言基础模型可包括上下文相关的 HMM 模型（Hidden Markov Model，隐马尔可夫模型）及该 HMM 模型对应的状态聚类决策树。其中，HMM 模型的每一个状态表示为一个高斯模型，决策树的作用为对训练数据进行聚类，以使得每一个状态都获得足够的训练数据。

需要说明的是，上述第一基础模型可理解是使用以第一语种类型为母语的训练说话人的语音数据而建立的模型，而该训练说话人可以会说第二语种，但对于该训练说话人的第二语种发音的标准性不做要求。上述第二基础模型可理解是使用以第二语种类型为母语的训练说话人的语音数据而建立的模型，该训练说话人可以会说第一语种，但对于该训练说话人的第一语种发音的标准性不做要求。

也就是说，在进行多语种语音合模型训练时，不必要求某一个发音人具有非常标准的双语发音，只要其中一个语言标准即可，其他语言的基础模型可以用其他发音标准的发音人数据进行训练。由此，可以减少多语种合成中对专业的多语发音人数据的依赖，而利用更多的单语数据，降低数据成本和实现难度。

S103，确定目标音色，并根据目标音色分别对第一谱参数模型、第二谱参数模型进行自适应变换，并根据自适应变换后的第一谱参数模型和第二谱参数模型对待合成语句文本信息进行训练，以生成对应的谱参数。

可以理解，在本发明的实施例中，目标音色的确定方式可以有很多种，例如，可以通过确定用户的母语所属种类来确定目标音色，还可以通过确定用户所使用的电子设备（如移动终端、终端等）的语言设置中选择的是哪种语种类型来确定目标音色，还可以通过其他方式来确定目标音色，在此不再一一举例示出。

作为一种示例，确定目标音色的具体实现过程可如下：获取用户的用户信息（如用户名或账户名等），并根据用户信息确定用户的母语所属种类，其中，母语所属种类包含于语种类型中；将用户的母语所属种类的基础模型所对应的训练说话人的音色作为目标音色。例如，获取用户 A 的用户信息，并根据该用户信息确定该用户 A 的母语所属种类为中文，此时可将该用户 A 的母语中文的基础模型（即中文基础模型）所对应的训练说话人的音色作为目标音色。

在确定目标音色之后，可根据该目标音色分别对第一谱参数模型、第二谱参数模型进

行自适应变换，使得应用该变换后的第一谱参数模型、第二谱参数模型可以生成具有相同或相近音色的谱参数。也就是说，在确定目标音色之后，可根据该目标音色对第一基础模型和第二基础模型进行自适应训练，以使得利用自适应训练后的第一基础模型和第二基础模型所生成的谱参数相同或相近。具体的自适应变换的实现方式可参照后续实施例的描述。

5 S104，根据第一基频参数模块、第二基频参数模块对待合成语句文本信息进行训练，以生成对应的基频参数，并根据目标音色对第一语种类型和第二语种类型的基频参数进行调整。

具体地，可根据第一基频参数模块、第二基频参数模块对待合成语句文本信息中各语种类型所对应的待合成语句文本进行相应的训练，以生成与第一语种类型的待合成语句文本对应的基频参数以及与第二语种类型的待合成语句文本对应的基频参数。在生成基频参数之后，可根据上述目标音色对第一语种类型和第二语种类型的基频参数进行调整，例如，可将第一语种类型以及第二语种类型的基频参数中的基频曲线的全局均值和方差，统一调整与目标音色所对应的基频参数中的基频曲线的全局均值和方差相同，以使得将第一基础模型得到的语音音色与第二基础模型得到的语音音色统一成该目标音色，保证将多语言文  
10 本合成音色统一的多语种语音。

S105，依据第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

作为一种示例，可将第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数经由声码器合成目标语音。可以理解，该  
20 目标语音为多语种语音。

本发明实施例的语音合成方法，首先，确定待合成语句文本信息所属的语种类型，其中，语种类型包括第一语种类型和第二语种类型，之后，确定第一语种类型对应的第一基础模型，并确定第二语种类型对应的第二基础模型，然后，确定目标音色，并根据目标音色分别对第一基础模型、第二基础模型进行自适应变换，并根据自适应变换后的第一基础模型、第二基础模型对待合成语句文本信息进行训练，以生成对应的谱参数和基频参数，  
25 之后，根据目标音色对第一语种类型和第二语种类型的基频参数进行调整，最后，依据第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。即先确定待合成语句文本中包含哪些语种类型，之后根据目标音色对各语种类型的谱参数模型进行自适应训练，并采用自适应训练后的谱参数  
30 模型生成对应的谱参数，并根据目标音色将生成的各语种类型的基频参数进行调整，以得到音色统一的多语种语音，可以理解，上述语言基础模型利用的大多是单语数据建立的模型，降低了模型建立的数据成本和实现难度，减少了多语种合成对专业的多语发音人数据

的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，提升了用户体验。

图 2 是根据本发明一个具体实施例的语音合成方法的流程图。

可以理解，目标音色可以是用户更想该合成语音的音色更倾向于哪种说话人的音色，例如，该目标音色可以是第一基础模型所对应的训练说话人的音色，也可以是第二基础模型所对应的训练说话人的音色。

在本发明的实施例中，以目标音色为第一基础模型所对应的训练说话人的音色为例，如图 2 所示，当目标音色为第一基础模型所对应的训练说话人的音色时，该语音合成方法可以包括：

10 S201，确定待合成语句文本信息所属的语种类型，其中，语种类型包括第一语种类型和第二语种类型。

S202，确定第一语种类型对应的第一基础模型，并确定第二语种类型对应的第二基础模型，其中，第一基础模型包括第一谱参数模型和第一基频参数模块，第二基础模型包括第二谱参数模型和第二基频参数模块。

15 S203，确定目标音色，并根据目标音色对第二谱参数模型进行自适应变换。

可以理解，当确定目标音色为第一基础模型所对应的训练说话人的音色时，也就是，将第二基础模型所生成的参数调整至与第一基础模型所对应的训练说话人的音色相同，而第一基础模型可不进行自适应训练而可以直接用来参数生成。

作为一种示例，在确定目标音色为第一基础模型所对应的训练说话人的音色时，可获取第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据，并根据第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据对第二谱参数模型进行自适应变换。可以理解，谱参数模型的自适应变换是在参数生成之前完成的。

具体地，在获取到第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据（如包括第二语种类型的训练语句及其标注等）时，可以该第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据作为输入，经由第二谱参数模型的决策树进行聚类，得到每一个状态的训练数据，并以每一个状态的训练数据对每个谱参数的 HMM 状态估计变换矩阵，使得后续可以应用变换矩阵后的状态高斯模型可以生成与第一语种类型训练说话人相近的谱参数。

25 S204，根据第一谱参数模型对待合成语句文本信息中第一语种类型对应的待合成语句进行训练，以生成第一语种类型的谱参数，并根据自适应变换后的第二谱参数模型对待合成语句文本信息中第二语种类型对应的待合成语句进行训练，以生成第二语种类型的谱参数。

具体地，当确定目标音色为第一基础模型所对应的训练说话人的音色时，可直接根据第一谱参数模型对待合成语句文本信息中第一语种类型对应的待合成语句进行训练，以生成第一语种类型的谱参数。而对于第二语种类型对应的待合成语句，可根据自适应变换后的第二谱参数模型对该第二语种类型对应的待合成语句进行训练，以生成第二语种类型的谱参数。

S205，根据第一基频参数模块、第二基频参数模块对待合成语句文本信息进行训练，以生成对应的基频参数，并根据目标音色对第二语种类型的基频参数进行调整。

具体地，可根据第一基频参数模块、第二基频参数模块对待合成语句文本信息中各语种类型所对应的待合成语句进行相应的训练，以生成各语种类型对应的基频参数，即第一语种类型的基频参数和第二语种类型的基频参数。

可以理解，在确定目标音色为第一基础模型所对应的训练说话人的音色时，可对第一语种类型的基频参数不进行调整，而需对第二语种类型的基频参数进行调整。

作为一种示例，根据目标音色对第二语种类型的基频参数进行调整的具体实现过程可包括：可先获取第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据，然后，根据第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据对第二基频参数模型进行训练，以生成目标音色对应的目标说话人基频参数，最后，根据目标说话人基频参数对第二语种类型的基频参数进行调整。

可以理解，基频参数的调整是在参数生成之后完成的。在确定目标音色为第一基础模型所对应的训练说话人的音色时，可先获取第一基础模型所对应的训练说话人针对第二语种类型的训练语音数据（如包括第二语种类型的训练语句及其标注等），并以该训练语音数据为输入，经由第二基频参数模型的决策树进行聚类，得到每一个状态的训练数据，并以此每一个状态的训练数据对每个基频的 HMM 状态进行训练，得到 HMM 状态的高斯参数，称为目标说话人基频模型。合成时，先以目标说话人基频模型进行参数生成，计算生成的基频曲线的全局均值和方差，将其保存。然后，以第二基础模型进行基频参数生成，并将生成的基频曲线经过线性变换，使得其均值和方差转换为与目标说话人基频模型生成的基频全局均值和方差相同，完成基频曲线的调整。

S206，依据第一语种类型的谱参数、第二语种类型的谱参数、第一语种类型的基频参数以及调整后的第二语种类型的基频参数合成目标语音。

可以理解，第二语种语音的音色通过自适应和基频参数调整的方式转换为第一基础模型所对应的训练说话人的音色，保留了第二语种语音原有的时长、语调信息，使原本第一基础模型所对应的训练说话人讲的并不地道的第二语种语音变得与第二基础模型所对应的训练说话人发音接近。

为了使得本领域技术人员能够更加清楚地了解本发明，下面以中文和英文两种语言混合合成，并统一为中文说话人音色为例介绍本发明的方法。

举例而言，假设第一基础模型为中文基础模型，第二基础模型为英文基础模型，假设中文基础模型是利用一名以中文为母语的中英双语说话人的双语语音数据建立的模型，英文基础模型是利用一名以英语母语说话人的英语语音数据建立的模型，其中，对于中文母语说话人的英文发音的标准性不做要求。

如图 3 所示，在获取到待合成语句文本信息之后，可根据语种的字符和上下文来对待合成语句文本中的文字内容进行断句，并判断出每一个句子片段的语种（S301）。由于追中音色需要调整至中文发音人发音，因此中文基础模型被直接用来进行参数生成，英文基础模型需要进行转换处理。即，在参数生成之前，可以英文模型转换为中文说话人音色为例，可获取中文说话人的英文训练语句数据（如英文语句及其标注）为输入，经由英文基础模型中的英文谱参数模型的决策树进行聚类，得到每一个状态的训练数据，并以此数据对每个谱参数的 HMM 状态估计变换矩阵，使得应用变换矩阵后的状态高斯模型可以生成与中文说话人相近的谱参数，用以进行参数生成（S302）。在参数生成之后，可以中文说话人的英文训练语句数据（如英文语句及其标注）为输入，经由英文基础模型中的英文基频参数模型的决策树进行聚类，得到每一个状态的训练数据，并以此数据对每个基频的 HMM 状态进行训练，得到 HMM 状态的高斯参数，称为目标说话人基频模型。合成时，先以目标说话人基频模型进行参数生成，计算生成的基频曲线的全局均值和方差，将其保存。然后，以英文基础模型进行基频参数生成，并将生成的基频曲线经过线性变换，使得其均值和方差转换为与目标说话人基频模型生成的基频全局均值和方差相同，完成基频曲线的转换（S303）。最后，将生成的中文语句文本对应的谱参数、基频参数、适应后得到的英文语句文本对应的谱参数、调整后的英文语句文本对应的基频参数经由声码器进行语音合成，以得到中英文混合语音（S304）。

综上，本发明实施例的语音合成方法通过不依赖某一个标准的多语发音人数据的情况下，合成音色统一且发音标准的语音，可以减少多语种合成中对专业的多语发音人数据的依赖，而利用更多的单语数据，降低数据成本和实现难度。

本发明实施例的语音合成方法，确定待合成语句文本中包含哪些语种类型，并在确定目标音色为第一基础模型所对应的训练说话人的音色时，根据该目标音色对第二谱参数模型进行自适应变换，而第一基础模型可不进行自适应训练而可以直接用来参数生成，并根据自适应训练后的第二谱参数模型生成对应的第二语种类型谱参数，并根据目标音色将生成的第二语种类型的基频参数进行调整，以将第二语种说话人的音色调整为与第一语种说话人的音色相同或相近，可以降低模型建立的数据成本和实现难度，减少了多语种合成对

专业的多语发音人数据的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，提升了用户体验。

与上述几种实施例提供的语音合成方法相对应，本发明的一种实施例还提供一种语音合成装置，由于本发明实施例提供的语音合成装置与上述几种实施例提供的语音合成方法相对应，因此在前述语音合成方法的实施方式也适用于本实施例提供的语音合成装置，在本实施例中不再详细描述。图 4 是根据本发明一个实施例的语音合成装置的结构框图。如图 4 所示，该语音合成装置可以包括：第一确定模块 10、第二确定模块 20、第三确定模块 30、自适应变换模块 40、谱参数生成模块 50、基频参数生成模块 60、基频参数调整模块 70 和语音合成模块 80。

具体地，第一确定模块 10 可用于确定待合成语句文本信息所属的语种类型，其中，语种类型包括第一语种类型和第二语种类型。

第二确定模块 20 可用于确定第一语种类型对应的第一基础模型，并确定第二语种类型对应的第二基础模型，其中，第一基础模型包括第一谱参数模型和第一基频参数模块，第二基础模型包括第二谱参数模型和第二基频参数模块。

第三确定模块 30 可用于确定目标音色。具体而言，在本发明的一个实施例中，如图 5 所示，该第三确定模块 30 可包括：第一确定单元 31 和第二确定单元 32。其中，第一确定单元 31 用于获取用户的用户信息，并根据用户信息确定用户的母语所属种类，其中，母语所属种类包含于语种类型中。第二确定单元 32 用于将用户的母语所属种类的基础模型所对应的训练说话人的音色作为目标音色。

自适应变换模块 40 可用于根据目标音色分别对第一谱参数模型、第二谱参数模型进行自适应变换。

谱参数生成模块 50 可用于根据自适应变换后的第一谱参数模型和第二谱参数模型对待合成语句文本信息进行训练，以生成对应的谱参数。

基频参数生成模块 60 可用于根据第一基频参数模块、第二基频参数模块对待合成语句文本信息进行训练，以生成对应的基频参数。

基频参数调整模块 70 可用于根据第一基频参数模块、第二基频参数模块对待合成语句文本信息进行训练，以生成对应的基频参数。

语音合成模块 80 可用于依据第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

作为一种示例，在目标音色为第一基础模型所对应的训练说话人的音色时，自适应变换模块 40 还用于根据目标音色对第二谱参数模型进行自适应变换。谱参数生成模块 50 还

用于根据第一谱参数模型对待合成语句文本信息中第一语种类型对应的待合成语句进行训练，以生成第一语种类型的谱参数，并根据自适应变换后的第二谱参数模型对待合成语句文本信息中第二语种类型对应的待合成语句进行训练，以生成第二语种类型的谱参数。基频参数调整模块 70 还用于根据目标音色对第二语种类型的基频参数进行调整。语音合成模块 80 还用于依据第一语种类型的谱参数、第二语种类型的谱参数、第一语种类型的基频参数以及调整后的第二语种类型的基频参数合成目标语音。

在本实施例中，如图 6 所示，该自适应变换模块 40 可包括：获取单元 41 和自适应变换单元 42。其中，获取单元 41 用于获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据。自适应变换单元 42 用于根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二谱参数模型进行自适应变换。

在本实施例中，如图 6 所示，该基频参数调整模块 70 可包括：获取单元 71、目标说话人基频参数生成单元 72 和基频参数调整单元 73。其中，获取单元 71 用于获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据。目标说话人基频参数生成单元 72 用于根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二基频参数模型进行训练，以生成所述目标音色对应的目标说话人基频参数。基频参数调整单元 73 用于根据所述目标说话人基频参数对所述第二语种类型的基频参数进行调整。

本发明实施例的语音合成装置，确定待合成语句文本中包含哪些语种类型，之后根据目标音色对各语种类型的谱参数模型进行自适应训练，并采用自适应训练后的谱参数模型生成对应的谱参数，并根据目标音色将生成的各语种类型的基频参数进行调整，以得到音色统一的多语种语音，可以理解，上述语言基础模型利用的大多是单语数据建立的模型，降低了模型建立的数据成本和实现难度，减少了多语种合成对专业的多语发音人数据的依赖，可以有效地将多语言待合成语句文本合成自然的、准确的、音色统一的多语种语音，提升了用户体验。

为了实现上述实施例，本发明还提出了一种终端，包括：一个或者多个处理器；存储器；一个或多个程序，一个或者多个程序存储在存储器中，当被一个或者多个处理器执行时进行如下操作：

S101'，确定待合成语句文本信息所属的语种类型，其中，语种类型包括第一语种类型和第二语种类型。

S102'，确定第一语种类型对应的第一基础模型，并确定第二语种类型对应的第二基础模型，其中，第一基础模型包括第一谱参数模型和第一基频参数模块，第二基础模型包括第二谱参数模型和第二基频参数模块。

S103'，确定目标音色，并根据目标音色分别对第一谱参数模型、第二谱参数模型进行自适应变换，并根据自适应变换后的第一谱参数模型和第二谱参数模型对待合成语句文本信息进行训练，以生成对应的谱参数。

5 S104'，根据第一基频参数模块、第二基频参数模块对待合成语句文本信息进行训练，以生成对应的基频参数，并根据目标音色对第一语种类型和第二语种类型的基频参数进行调整。

S105'，依据第一语种类型的谱参数、第二语种类型的谱参数、调整后的第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

10 为了实现上述实施例，本发明还提出了一种存储介质，用于存储应用程序，应用程序用于执行本发明上述任一个实施例所述的语音合成方法。

在本发明的描述中，需要理解的是，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中，“多种”的含义是至少两种，例如两种，三种等，除非另有明确具体的限定。

15 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的  
20 技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本发明的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，  
25 包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本发明的实施例所属技术领域的技术人员所理解。

在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是用于实现逻辑功能的可执行指令的定序列列表，可以具体实现在任何计算机可读介质中，以供指令  
30 执行系统、装置或设备（如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统）使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用

的装置。计算机可读介质的更具体的示例（非穷尽性列表）包括以下：具有一个或多个布线的电连接部（电子装置），便携式计算机盘盒（磁装置），随机存取存储器（RAM），只读存储器（ROM），可擦除可编辑只读存储器（EPROM 或闪速存储器），光纤装置，以及便携式光盘只读存储器（CDROM）。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

应当理解，本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列（PGA），现场可编程门阵列（FPGA）等。

本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

此外，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模块中，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模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用，也可以存储在一个计算机可读存储介质中。

上述提到的存储介质可以是只读存储器，磁盘或光盘等。尽管上面已经示出和描述了本发明的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

## 权 利 要 求 书

1、一种语音合成方法，其特征在于，包括以下步骤：

5 确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和  
第二语种类型；

确定所述第一语种类型对应的第一基础模型，并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；

10 确定目标音色，并根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换，并根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；

根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数，并根据所述目标音色对所述第一语种类型和第二语种类型的基频参数进行调整；

15 依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

2、如权利要求 1 所述的语音合成方法，其特征在于，所述确定目标音色，包括：

获取用户的用户信息，并根据所述用户信息确定所述用户的母语所属种类，其中，所述母语所属种类包含于所述语种类型中；

20 将所述用户的母语所属种类的基础模型所对应的训练说话人的音色作为所述目标音色。

3、如权利要求 1 或 2 所述的语音合成方法，其特征在于，当所述目标音色为所述第一基础模型所对应的训练说话人的音色时，

25 所述根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换，包括：

根据所述目标音色对所述第二谱参数模型进行自适应变换；

所述根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数，包括：

30 根据所述第一谱参数模型对所述待合成语句文本信息中所述第一语种类型对应的待合成语句进行训练，以生成所述第一语种类型的谱参数，并根据自适应变换后的所述第二谱参数模型对所述待合成语句文本信息中所述第二语种类型对应的待合成语句进行训练，以生成所述第二语种类型的谱参数；

所述根据所述目标音色对所述第一语种类型和第二语种类型的基频参数进行调整，包括：

根据所述目标音色对所述第二语种类型的基频参数进行调整；

5 所述依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音，包括：

依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、所述第一语种类型的基频参数以及调整后的所述第二语种类型的基频参数合成所述目标语音。

4、如权利要求3所述的语音合成方法，其特征在于，所述根据所述目标音色对所述第二谱参数模型进行自适应变换，包括：

10 获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据；

根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二谱参数模型进行自适应变换。

5、如权利要求3所述的语音合成方法，其特征在于，所述根据所述目标音色对所述第二语种类型的基频参数进行调整，包括：

15 获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据；

根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二基频参数模型进行训练，以生成所述目标音色对应的目标说话人基频参数；

根据所述目标说话人基频参数对所述第二语种类型的基频参数进行调整。

20 6、一种语音合成装置，其特征在于，包括：

第一确定模块，用于确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和第二语种类型；

25 第二确定模块，用于确定所述第一语种类型对应的第一基础模型，并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；

第三确定模块，用于确定目标音色；

自适应变换模块，用于根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换；

30 谱参数生成模块，用于根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；

基频参数生成模块，用于根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数；

基频参数调整模块，用于根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数；

语音合成模块，用于依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

5 7、如权利要求6所述的语音合成装置，其特征在于，所述第三确定模块包括：

第一确定单元，用于获取用户的用户信息，并根据所述用户信息确定所述用户的母语所属种类，其中，所述母语所属种类包含于所述语种类型中；

第二确定单元，用于将所述用户的母语所属种类的基础模型所对应的训练说话人的音色作为所述目标音色。

10 8、如权利要求6或7所述的语音合成装置，其特征在于，在所述目标音色为所述第一基础模型所对应的训练说话人的音色时，

所述自适应变换模块还用于根据所述目标音色对所述第二谱参数模型进行自适应变换；

15 所述谱参数生成模块还用于根据所述第一谱参数模型对所述待合成语句文本信息中所述第一语种类型对应的待合成语句进行训练，以生成所述第一语种类型的谱参数，并根据自适应变换后的所述第二谱参数模型对所述待合成语句文本信息中所述第二语种类型对应的待合成语句进行训练，以生成所述第二语种类型的谱参数；

所述基频参数调整模块还用于根据所述目标音色对所述第二语种类型的基频参数进行调整；

20 所述语音合成模块还用于依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、所述第一语种类型的基频参数以及调整后的所述第二语种类型的基频参数合成所述目标语音。

9、如权利要求8所述的语音合成装置，其特征在于，所述自适应变换模块包括：

25 获取单元，用于获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据；

自适应变换单元，用于根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二谱参数模型进行自适应变换。

10、如权利要求8所述的语音合成装置，其特征在于，所述基频参数调整模块包括：

30 获取单元，用于获取所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据；

目标说话人基频参数生成单元，用于根据所述第一基础模型所对应的训练说话人针对所述第二语种类型的训练语音数据对所述第二基频参数模型进行训练，以生成所述目标音

色对应的目标说话人基频参数；

基频参数调整单元，用于根据所述目标说话人基频参数对所述第二语种类型的基频参数进行调整。

5 11、一种终端，其特征在于，包括：

一个或者多个处理器；

存储器；

一个或多个程序，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进行如下操作：

10 确定待合成语句文本信息所属的语种类型，其中，所述语种类型包括第一语种类型和第二语种类型；

确定所述第一语种类型对应的第一基础模型，并确定所述第二语种类型对应的第二基础模型，其中，所述第一基础模型包括第一谱参数模型和第一基频参数模块，所述第二基础模型包括第二谱参数模型和第二基频参数模块；

15 确定目标音色，并根据所述目标音色分别对所述第一谱参数模型、所述第二谱参数模型进行自适应变换，并根据自适应变换后的所述第一谱参数模型和所述第二谱参数模型对所述待合成语句文本信息进行训练，以生成对应的谱参数；

根据所述第一基频参数模块、第二基频参数模块对所述待合成语句文本信息进行训练，以生成对应的基频参数，并根据所述目标音色对所述第一语种类型和第二语种类型的基频参数进行调整；

20 依据所述第一语种类型的谱参数、所述第二语种类型的谱参数、调整后的所述第一语种类型的基频参数、第二语种类型的基频参数合成目标语音。

12、一种存储介质，其特征在于，用于存储应用程序，所述应用程序用于执行权利要求 1 至 5 中任一项所述的语音合成方法。

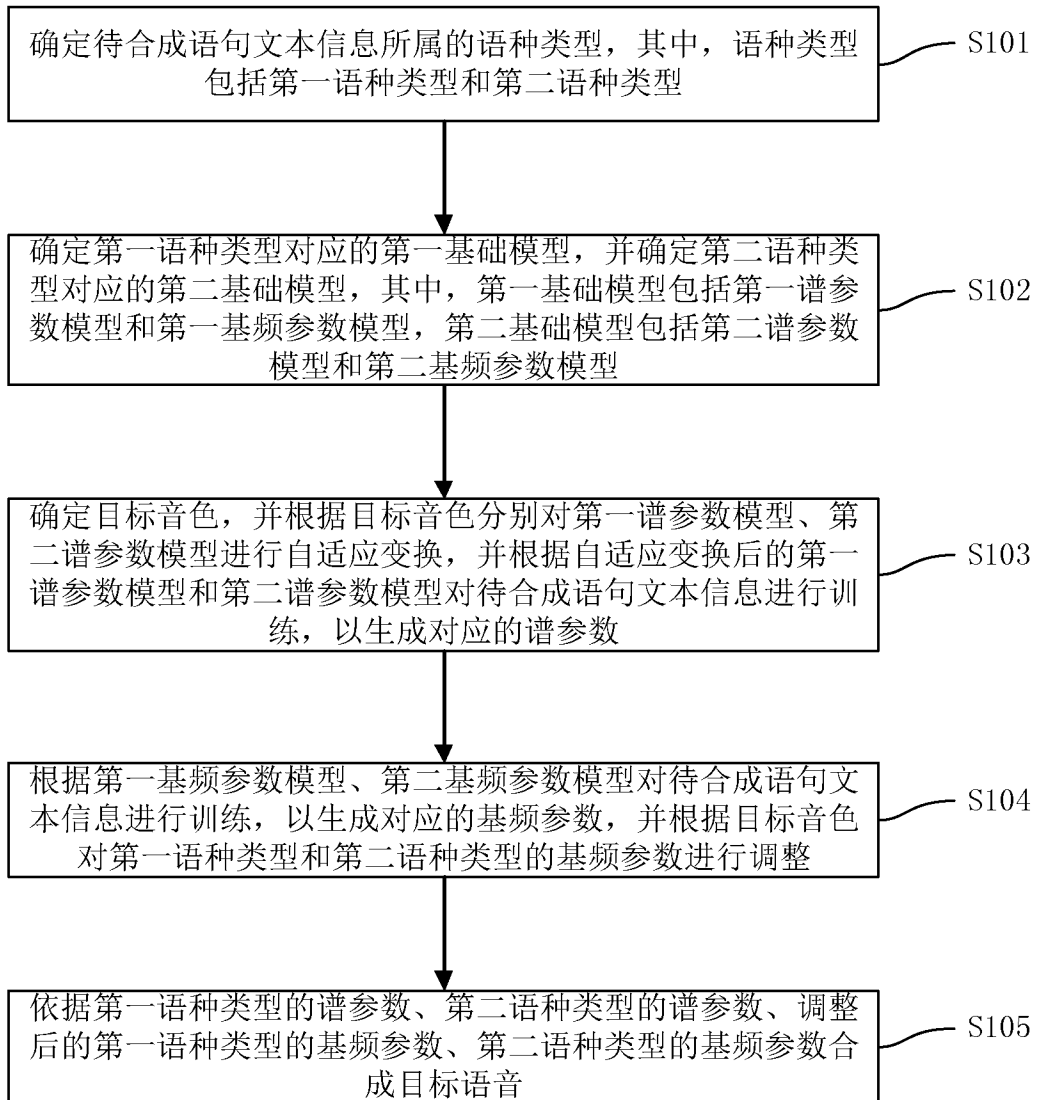


图 1

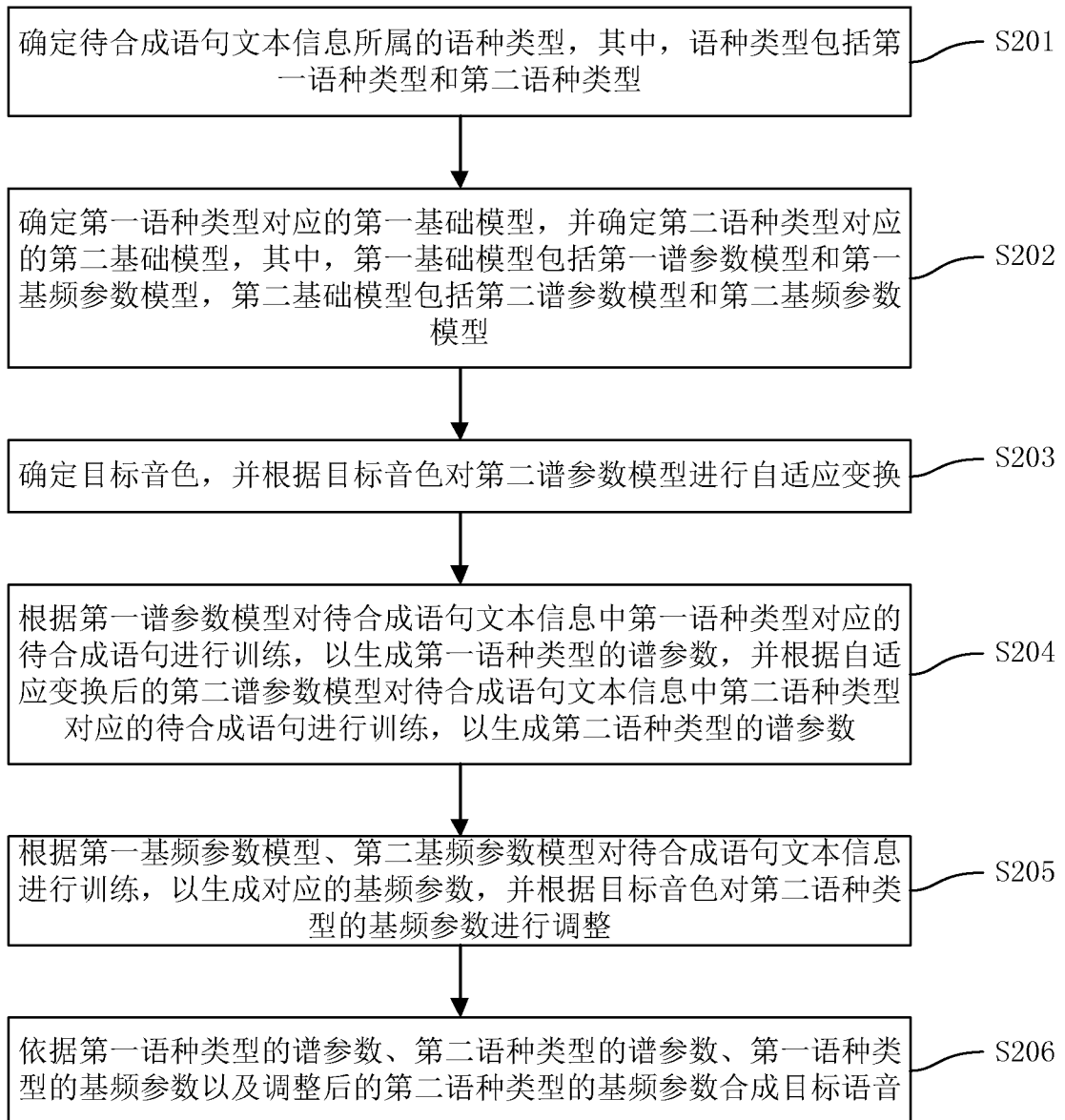


图 2

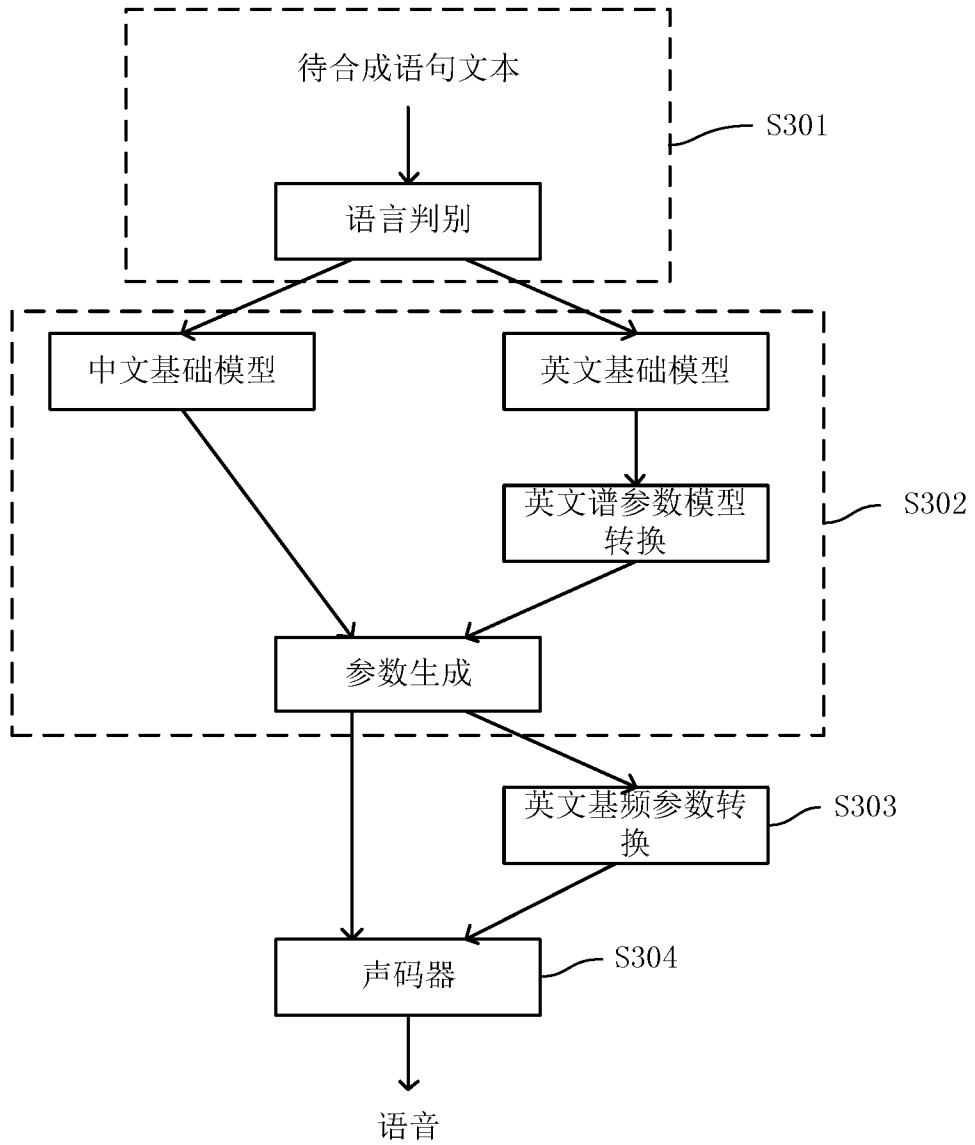


图 3

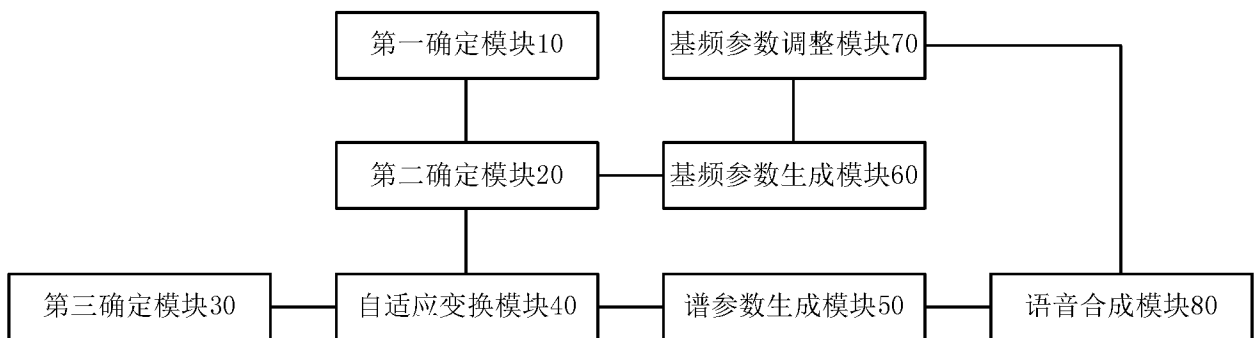


图 4

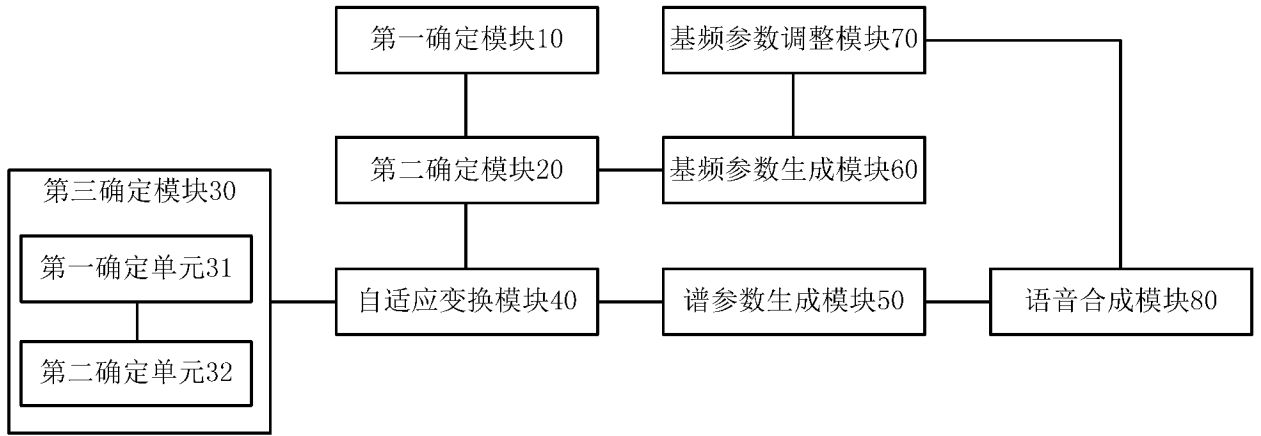


图 5

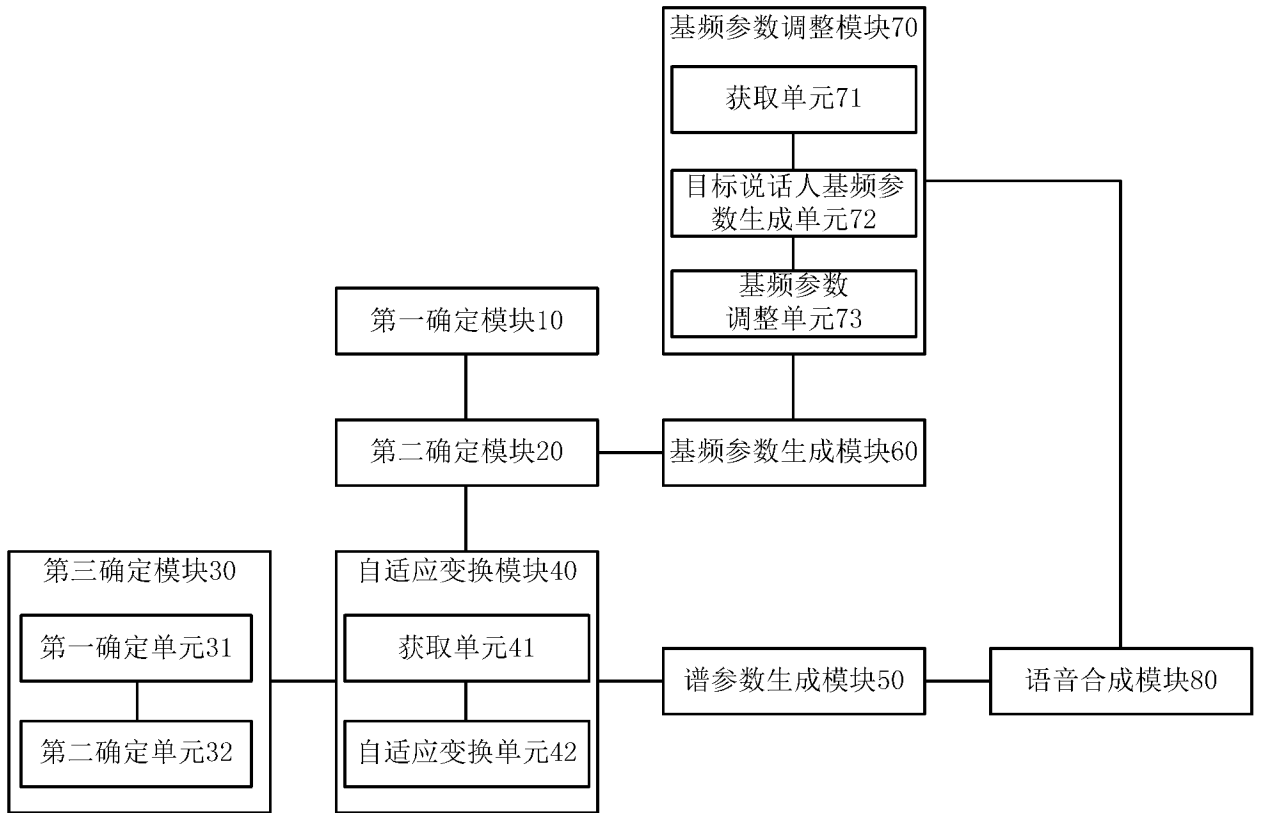


图 6

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2016/098126**

## A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/047 (2013.01) i; G10L 13/06 (2013.01) i; 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)

G10L

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)

VEN, CNABS: multi-language, hidden markov, HMM, speech, audio, voice, language?, type, kind, parameter, base, fundament, target, synthesis

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105845125 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 10 August 2016 (10.08.2016), claims 1-10, and description, paragraphs [0081]-[0084]	1-12
A	CN 102543069 A (INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE), 04 July 2012 (04.07.2012), description, paragraphs [0055]-[0092], and figures 1-6	1-12
A	CN 102360543 A (MICROSOFT CORPORATION), 22 February 2012 (22.02.2012), the whole document	1-12
A	JP 2011237886 A (CANON KK), 24 November 2011 (24.11.2011), the whole document	1-12
A	JP 2008139631 A (NIPPON TELEGRAPH & TELEPHONE CORP.), 19 June 2008 (19.06.2008), the whole document	1-12
A	WO 2009014465 A3 (JOVICIC, S. et al.), 07 May 2009 (07.05.2009), the whole document	1-12

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>
---	---

Date of the actual completion of the international search  
13 February 2017 (13.02.2017)

Date of mailing of the international search report  
**24 February 2017 (24.02.2017)**

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  
**ZHANG, Yaling**  
Telephone No.: (86-10) **62085722**

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
**PCT/CN2016/098126**

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105845125 A	10 August 2016	None	
CN 102543069 A	04 July 2012	TW I413105 B	21 October 2013
		US 8898066 B2	25 November 2014
		TW 201227715 A	01 July 2012
		CN 102543069 B	16 October 2013
		US 2012173241 A1	05 July 2012
CN 102360543 A	22 February 2012	WO 2009026270 A3	30 April 2009
		WO 2009026270 A2	26 February 2009
		CN 101785048 B	10 October 2012
		US 2009055162 A1	26 February 2009
		US 8244534 B2	14 August 2012
		CN 101785048 A	21 July 2010
		CN 102360543 B	27 March 2013
JP 2011237886 A	24 November 2011	None	
JP 2008139631 A	19 June 2008	None	
WO 2009014465 A3	07 May 2009	WO 2009014465 A2	29 January 2009
		RS 20070316 A	04 April 2008
		RS 50004 B	29 September 2008

国际检索报告

国际申请号

PCT/CN2016/098126

<p>A. 主题的分类</p> <p>G10L 13/047(2013.01)i; G10L 13/06(2013.01)i; G06F 17/28(2006.01)i</p> <p>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</p>																							
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G10L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>VEN, CNABS: 多语言, 语音, 语种, 合成, 种类, 类型, 马尔可夫隐形, HMM, speech, audio, voice, language?, type, kind, parameter, base, fundament, target, synthesis</p>																							
<p>C. 相关文件</p> <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>PX</td> <td>CN 105845125 A (百度在线网络技术北京有限公司) 2016年 8月 10日 (2016 - 08 - 10) 权利要求1-10, 说明书第[0081]-[0084]段</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 102543069 A (财团法人工业技术研究院) 2012年 7月 4日 (2012 - 07 - 04) 说明书第[0055]-[0092]段, 图1-6</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 102360543 A (微软公司) 2012年 2月 22日 (2012 - 02 - 22) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>JP 2011237886 A (CANON KK) 2011年 11月 24日 (2011 - 11 - 24) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>JP 2008139631 A (NIPPON TELEGRAPH &amp; TELEPHONE CORP.) 2008年 6月 19日 (2008 - 06 - 19) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>WO 2009014465 A3 (JOVICIC SLOBODAN等) 2009年 5月 7日 (2009 - 05 - 07) 全文</td> <td>1-12</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	PX	CN 105845125 A (百度在线网络技术北京有限公司) 2016年 8月 10日 (2016 - 08 - 10) 权利要求1-10, 说明书第[0081]-[0084]段	1-12	A	CN 102543069 A (财团法人工业技术研究院) 2012年 7月 4日 (2012 - 07 - 04) 说明书第[0055]-[0092]段, 图1-6	1-12	A	CN 102360543 A (微软公司) 2012年 2月 22日 (2012 - 02 - 22) 全文	1-12	A	JP 2011237886 A (CANON KK) 2011年 11月 24日 (2011 - 11 - 24) 全文	1-12	A	JP 2008139631 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2008年 6月 19日 (2008 - 06 - 19) 全文	1-12	A	WO 2009014465 A3 (JOVICIC SLOBODAN等) 2009年 5月 7日 (2009 - 05 - 07) 全文	1-12
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 105845125 A (百度在线网络技术北京有限公司) 2016年 8月 10日 (2016 - 08 - 10) 权利要求1-10, 说明书第[0081]-[0084]段	1-12																					
A	CN 102543069 A (财团法人工业技术研究院) 2012年 7月 4日 (2012 - 07 - 04) 说明书第[0055]-[0092]段, 图1-6	1-12																					
A	CN 102360543 A (微软公司) 2012年 2月 22日 (2012 - 02 - 22) 全文	1-12																					
A	JP 2011237886 A (CANON KK) 2011年 11月 24日 (2011 - 11 - 24) 全文	1-12																					
A	JP 2008139631 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2008年 6月 19日 (2008 - 06 - 19) 全文	1-12																					
A	WO 2009014465 A3 (JOVICIC SLOBODAN等) 2009年 5月 7日 (2009 - 05 - 07) 全文	1-12																					
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																							
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” 同族专利的文件</p>																							
<p>国际检索实际完成的日期</p> <p>2017年 2月 13日</p>	<p>国际检索报告邮寄日期</p> <p>2017年 2月 24日</p>																						
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10)62019451</p>	<p>受权官员</p> <p>张亚玲</p> <p>电话号码 (86-10)62085722</p>																						

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2016/098126

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	105845125	A	2016年 8月 10日	无			
CN	102543069	A	2012年 7月 4日	TW	I413105	B	2013年 10月 21日
				US	8898066	B2	2014年 11月 25日
				TW	201227715	A	2012年 7月 1日
				CN	102543069	B	2013年 10月 16日
				US	2012173241	A1	2012年 7月 5日
CN	102360543	A	2012年 2月 22日	WO	2009026270	A3	2009年 4月 30日
				WO	2009026270	A2	2009年 2月 26日
				CN	101785048	B	2012年 10月 10日
				US	2009055162	A1	2009年 2月 26日
				US	8244534	B2	2012年 8月 14日
				CN	101785048	A	2010年 7月 21日
				CN	102360543	B	2013年 3月 27日
JP	2011237886	A	2011年 11月 24日	无			
JP	2008139631	A	2008年 6月 19日	无			
WO	2009014465	A3	2009年 5月 7日	WO	2009014465	A2	2009年 1月 29日
				RS	20070316	A	2008年 4月 4日
				RS	50004	B	2008年 9月 29日

表 PCT/ISA/210 (同族专利附件) (2009年7月)