

(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
2021 年 7 月 8 日 (08.07.2021)



(10) 国际公布号
WO 2021/134591 A1

(51) 国际专利分类号:

G10L 13/047 (2013.01) *G10L 13/10* (2013.01)
G10L 13/08 (2013.01) *G10L 25/30* (2013.01)

(21) 国际申请号: PCT/CN2019/130766

(22) 国际申请日: 2019 年 12 月 31 日 (31.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳市优必选科技股份有限公司
(UBTECH ROBOTICS CORP LTD) [CN/CN]; 中国
广东省深圳市南山区学苑大道1001号南山智园
C1栋16、22楼, Guangdong 518000 (CN)。

(72) 发明人: 李贤 (LI, Xian); 中国广东省深圳市南
山区学苑大道1001号南山智园C1栋16、22
楼, Guangdong 518000 (CN)。 黄东延 (HUANG,

Dongyan); 中国广东省深圳市南山区学苑大
道1001号南山智园C1栋16、22楼, Guangdong
518000 (CN)。 丁万 (DING, Wan); 中国广东省
深圳市南山区学苑大道1001号南山智园
C1栋16、22楼, Guangdong 518000 (CN)。 张皓
(ZHANG, Hao); 中国广东省深圳市南山区学苑
大道1001号南山智园C1栋16、22楼, Guangdong
518000 (CN)。 白洛玉 (BAI, Luoyu); 中国广东
省深圳市南山区学苑大道1001号南山智
园C1栋16、22楼, Guangdong 518000 (CN)。 熊
友军 (XIONG, Youjun); 中国广东省深圳市南
山区学苑大道1001号南山智园C1栋16、
22楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中细软知识产权代理有
限公司 (SHENZHEN CIPRUN INTELLECTUAL
PROPERTY AGENCY CO., LTD.); 中国广东省深

(54) Title: SPEECH SYNTHESIS METHOD, SPEECH SYNTHESIS APPARATUS, SMART TERMINAL AND STORAGE MEDIUM

(54) 发明名称: 语音合成方法、装置、终端及存储介质

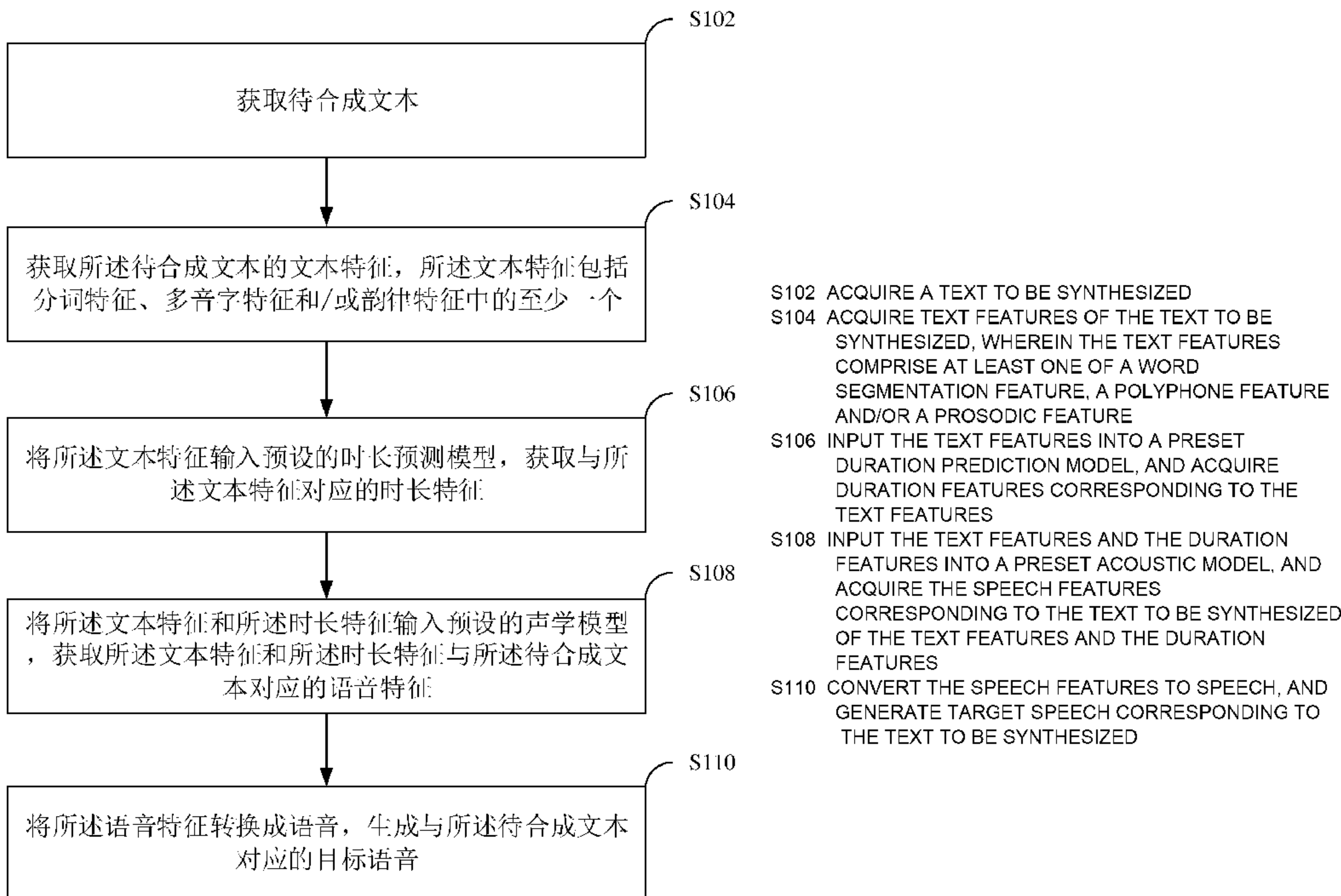


图 2

(57) Abstract: A speech synthesis method, a speech synthesis apparatus, a smart terminal and a storage medium. The method comprises: acquiring a text to be synthesized (S102); acquiring text features of the text to be synthesized, wherein the text features comprise at least one of a word segmentation feature, a polyphone feature and/or a prosodic feature (S104); inputting the text features into a preset duration prediction model, and acquiring duration features corresponding to the text features (S106); inputting the text features and the duration features into a preset acoustic model, and acquiring the speech features corresponding to the text to be synthesized (S108); and

圳市南山区粤海街道粤兴四道1号中山大学深圳
产学研大楼11楼, 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,
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))。

converting the speech features to speech, and generating target speech corresponding to the text to be synthesized (S110). According to the speech synthesis method, the speech features generated by various text features and duration features are considered, such that the synthesized speech is more accurate, thereby improving the speech synthesis accuracy, and improving the user experience.

(57) 摘要: 一种语音合成方法、语音合成装置、智能终端及存储介质, 该方法包括: 获取待合成文本 (S102); 获取待合成文本的文本特征, 文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个 (S104); 将文本特征输入预设的时长预测模型, 获取与文本特征对应的时长特征 (S106); 将文本特征和时长特征输入预设的声学模型, 获取与待合成文本对应的语音特征 (S108); 将语音特征转换成语音, 生成与待合成文本对应的目标语音 (S110)。该语音合成方法考虑了多种文本特征和时长特征生成的语音特征, 使得合成的语音更加准确, 提高了语音合成的准确性, 提高了用户体验。

语音合成方法、装置、终端及存储介质

技术领域

[0001] 本发明涉及人工智能技术领域，尤其涉及一种语音合成方法、装置、智能终端及计算机可读存储介质。

背景技术

[0002] 随着移动互联网和人工智能技术的快速发展，语音播报、听小说、听新闻、智能交互等一系列语音合成的场景越来越多。语音合成可以将文本等转换成自然语音输出。

发明概述

技术问题

[0003] 现有技术中语音合成多采用统计参数合成法，对于引得频谱特性参数进行建模，生成参数合成器，来构建文本序列映射到语音的映射关系，然后统计模型来产生每时每刻的语音参数（包括基频、共振峰频率等），然后把这些参数转化为语音对应的相关特征，最后生成输出的语音。但是上述语音合成方法中，每个步骤对应的单一子模块的计算结果不一定全部都是最优效果，从而导致了无法将文本准确转换为适应多语言、多音色场景的语音，影响了整体上的语音合成的质量，极大影响用户体验。

[0004] 也就是说，上述语音合成的方案中，因为单一子模块计算结果非最优的问题导致了最终合成的语音的质量不足。

问题的解决方案

技术解决方案

[0005] 基于此，有必要针对上述问题，提出了一种语音合成方法、装置、智能终端及计算机可读存储介质。

[0006] 在本发明的第一方面，提出了一种语音合成方法。

[0007] 一种语音合成方法，包括：

[0008] 获取待合成文本；

- [0009] 获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；
- [0010] 将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；
- [0011] 将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；
- [0012] 将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。
- [0013] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤之前，还包括：对所述待合成文本进行正则化处理。
- [0014] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤还包括：将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取所述待合成文本对应的韵律特征。
- [0015] 在一个实施例中，所述方法还包括：获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。
- [0016] 在一个实施例中，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征；所述方法包括：将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。

- [0017] 在一个实施例中，所述方法还包括：通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。
- [0018] 在本发明的第二方面，提出了一种语音合成装置。
- [0019] 一种语音合成装置，包括：
- [0020] 获取模块，用于获取待合成文本；
- [0021] 文本特征确定模块，用于获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；
- [0022] 时长特征确定模块，用于将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；
- [0023] 语音特征确定模块，用于将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；
- [0024] 转换模块，用于将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。
- [0025] 在一个实施例中，所述文本特征确定模块还包括：预处理单元，用于对所述待合成文本进行正则化处理。
- [0026] 在一个实施例中，所述文本特征确定模块还包括：分词特征确定单元，用于将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；多音字特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；韵律特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取所述待合成文本对应的韵律特征。
- [0027] 在一个实施例中，所述装置还包括：获取训练模块，用于获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；时长训练模块，用于将所述训练文本对应的文本参考特征作为

所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；语音训练模块，用于将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。

[0028] 在一个实施例中，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征，所述装置包括：分词训练模块，用于将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。多音字训练模块，用于将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。韵律训练模块，用于将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。

[0029] 在一个实施例中，所述装置还包括：文本获取模块，用于通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；文本预测模块，用于将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。

[0030] 在本发明的第三方面，提出了一种智能终端。

[0031] 一种智能终端，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

[0032] 获取待合成文本；

[0033] 获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；

[0034] 将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；

[0035] 将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；

[0036] 将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

[0037] 在本发明的第四方面，提出了一种计算机可读存储介质。

[0038] 一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行以下步骤：

[0039] 获取待合成文本；

[0040] 获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；

[0041] 将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；

[0042] 将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；

[0043] 将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

发明的有益效果

有益效果

[0044] 实施本发明实施例，将具有如下有益效果：

[0045] 采用本发明的语音合成方法、装置、终端及存储介质之后，在进行语音合成的过程中，首先获取待合成文本的待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征等文本特征；然后将文本特征输入预设的时长预测模型获取对应的时长特征；将文本特征和时长特征输入预设的声学模型获取对应的语音特征；最后将语音特征转换成语音，生成与待合成文本对应的目标语音。在进行语音合成的特征提取的过程中，考虑的文本特征包括了多音字特征和韵律特征等，并结合模型预测的时长特征，得到最终合成语音的过程中所需要的语音特征。也就是说，本发明提供的语音合成方法、装置、终端及存储介质考虑了多种文本特征和时长特征生成的语音特征，使得合成的语音更加准确，提高了语音合成的准确性，提高了用户体验。

对附图的简要说明

附图说明

[0046] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或

现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0047] 其中：

[0048] 图1为本申请中一个实施例中语音合成方法的应用环境图；

[0049] 图2为本申请中一个实施例中语音合成方法的流程示意图；

[0050] 图3为本申请中一个实施例中一种获取待合成文本的文本特征的过程的流程图示意图；

[0051] 图4为本申请中一个实施例中时长预测模型和声学模型训练的方法的流程示意图；

[0052] 图5为本申请中一个实施例中分词模型、多音字预测模型和/或韵律预测模型的流程示意图；

[0053] 图6为本申请中一个实施例中语音合成方法的流程图；

[0054] 图7为本申请中一个实施例中语音合成装置的结构框图；

[0055] 图8为本申请中一个实施例中文本特征确定模块的结构框图；

[0056] 图9为本申请中一个实施例中语音合成装置的结构框图；

[0057] 图10为本申请中一个实施例中语音合成装置的结构框图；

[0058] 图11为本申请中一个实施例中语音合成装置的结构框图；

[0059] 图12为本申请中一个实施例中执行前述语音合成方法的计算机设备的结构框图。

发明实施例

本发明的实施方式

[0060] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，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他实施例，都属于本发明保护的范围。

[0061] 图1为一个实施例中语音合成的方法的应用环境图。参照图1，该语音合成的方法应用于语音合成系统。该语音合成系统包括终端110和服务器120。终端110和

服务器120通过网络连接，终端110具体可以是台式终端或移动终端，移动终端具体可以是PC、手机、平板电脑、笔记本电脑等终端设备。服务器120可以用独立的服务器或者是多个服务器组成的服务器集群来实现。终端110用于获取待合成文本，服务器120用于对待合成文本进行分析和处理，合成待合成文本对应的目标语音。

[0062] 在另一个实施例中，上述基于语音合成的方法的执行还可以是基于终端设备，该终端可获取待合成文本、也可以对待合成文本进行分析，合成待合成文本对应的目标语音。

[0063] 考虑到该方法既可以应用于终端，也可以应用于服务器，且在具体的语音合成的过程是相同的，本实施例以应用于终端举例说明。

[0064] 如图2所示，在一个实施例中，提供了一种语音合成的方法。该语音合成的方法具体包括如下步骤S102-S110：

[0065] 步骤S102，获取待合成文本。

[0066] 具体的，待合成文本为需要进行语音合成的文本信息。例如，在语音聊天机器人、语音读报等场景下，需要转换成语音的文本信息。示例性的，待合成文本可以是“自从那一刻起，她便不再妄自菲薄。”。

[0067] 上述待合成文本可以是获取直接输入文本信息，也可以是通过摄像头等扫描识别得到文本信息。

[0068] 步骤S104，获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个。

[0069] 具体的，文本特征是待合成文本中文本信息对应的规律特征。

[0070] 在一个具体的实施例中，文本特征可以是分词特征、多音字特征和/或韵律特征中的一个。

[0071] 分词特征是将组成待合成文本的词语进行分类得到的词组特征，可以是名词、动词、介词和形容词等。

[0072] 多音字特征是待合成文本中包括的存在多种读音的字或词，由于读音有区别词性和词义的作用，因此使用情况或者环境不同，读音也不同。

[0073] 韵律特征是语言的一种韵律结构，与句法和语篇结构、信息结构等其他语言学

结构密切相关。韵律特征是自然语言的典型特征，是不同语言具有的共同特点，比如：音高下倾、重读、停顿等特征。韵律特征可以分为三个主要方面：语调、时域分布和重音，通过超音段特征实现。超音段特征包括音高，强度以及时间特性，由音位或音位群负载。韵律特征是语言和对应的情绪表达的重要形式。

[0074] 在获取待合成文本的文本特征前，还可以先对待合成文本进行预处理，避免一些微小的影响（例如格式问题）导致输出的文本特征存在偏差。

[0075] 在一个实施例中，在获取所述待合成文本的文本特征之前，对所述待合成文本进行正则化处理。

[0076] 其中，正则化处理是待合成文本进行规范化，将语言文字转换为预设形式的语言文字，例如，英文处理字母的大小写，可以根据需要去除标点符号，从而避免文本格式等问题导致输出的文本特征存在偏差。在另一个具体的实施例中，对待合成文本进行规范化处理还包括了将待合成文本中的数字、符号等文本转换成中文，以便于后续进行分词特征、多音字特征和/或韵律特征的提取，减少特征提取的误差。

[0077] 上述待合成文本的文本特征的获取，可以是将待合成文本输入预设的神经网络模型，预设的神经网络模型根据相应的算法计算得到相应的文本特征；或者，按照预设的特征提取算法，从待合成文本中提取对应的文本特征。

[0078] 在一个实施例中，对通过神经网络模型获取待合成文本的文本特征的过程进行描述。

[0079] 具体的，如图3所示，给出了一种获取待合成文本的文本特征的过程的流程示意图。

[0080] 如图3所述，上述获取待合成文本的文本特征的过程包括如图3所示的步骤S1041-S1043：

[0081] 步骤S1041：将待合成文本输入预设的分词模型，获取与待合成文本对应的分词特征，其中分词特征包含了待合成文本中应该从哪些地方进行断句、或者断开，从而确定待合成文本对应的分词结果对应的分词特征；

[0082] 步骤S1042：将待合成文本和/或分词特征输入预设的多音字预测模型，获取待

合成文本对应的多音字特征；

[0083] 步骤S1043：将待合成文本和/或分词特征输入预设的韵律预测模型，获取待合成文本对应的韵律特征。

[0084] 分词模型是将待合成文本进行分词处理得出分词特征的神经网络模型，通过分词模型可以对待合成文本的分词特征进行预测。其中，分词特征的确定是通过分词得到的字向量确定的，字向量是根据分词模型划分的单词或者词组对应的向量，用于确定待合成文本的分词特征。

[0085] 多音字预测模型可以预测待合成文本中或者分词特征中的多音字特征，可以为神经网络模型。

[0086] 韵律预测模型是预测待合成文本或者分词特征中的韵律特征的神经网络模型，可以对待合成文本的韵律特征进行预测，例如，韵律词特征、韵律短语特征以及语调短语特征。

[0087] 本实施例中的待合成文本中的文本特征不限于本实施例中的分词特征、多音字特征和韵律特征等文本特征。

[0088] 用户可以对本文中涉及的文本特征进行设置，涉及的文本特征不仅仅是分词特征、多音字特征和韵律特征，还可以是前后词关联特征等其他特征。用户也可以通过对计算图的建立对总的神经网络模型的结构进行建立，对输入数据进行选择。

[0089] 步骤S106，将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征。

[0090] 具体的，时长特征是待合成文本以及待合成文本对应的文本特征中包含的音素文本特征对应的时间长度。预设的时长预测模型是预测音素文本特征对应的时间长度的神经网络模型，确定待合成文本包含的每一个音素对应的时间长度，其中包含将拼音转化成音素的过程，是通过多音字预测模型得到字的读音（例如ou3），然后将读音转化为音素，然后用时长预测模型去预测音素的时长。示例性的，将读音转换为音素，“我在中国”中的“我”的读音ou3可以转换为ou的1个音素，“国”的读音guo2可以转换为g，uo的两个音素。

[0091] 步骤S108，将所述文本特征和所述时长特征输入预设的声学模型，获取与所述

待合成文本对应的语音特征。

[0092] 具体的，语音特征是根据文本特征和时长特征生成的特征，语音特征包括声强、响度、音高和/或基音周期等特征。其中，声强是单位时间内通过垂直于声波传播方向的单位面积的平均声能；响度反映了主观感觉到的声音强弱程度；音高反映了主观感觉到的声音频率高低；基音周期，是在发音时浊音波形呈现的准周期，反映了声门相邻两次开闭之间的时间间隔或开闭的频率。

[0093] 在本实施例中，将上述步骤S104中获取的文本特征以及步骤S106中获取的时长特征输入预设的声学模型，通过声学模型对待合成文本对应的语音特征。

[0094] 上述对语音特征进行预测的声学模型，为一神经网络模型，通过事先的训练声学模型具备根据文本特征和时长特征计算对应的语音特征的能力。

[0095] 步骤S110，将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

[0096] 目标语音是通过对待合成文本生成的语音。将语音特征转换为语音，可以将语音特征通过声码器进行合成，输出语音特征在声码器中对应的语音和语音时长等，得到目标语音，其中，声码器可以是parallel wavenet声码器。具体的，将语音特征作为输入，通过预设的声码器对待合成文本对应的语音特征进行语音合成，输出对应的目标语音。

[0097] 进一步的，上述时长预测模型和声学模型可以对待合成文本的相关特征进行很好的预测，并且，在使用相应的模型进行预测之前，还要根据训练数据对相应的模型进行训练。也就是说，对文本特征进行预测得到对应的时长特征和语音特征之前，还需要对时长预测模型和声学模型进行训练，使得相应的模型具备对文本特征对应的时长特征和语音特征进行准确的预测的能力。

[0098] 如图4所述，上述语音合成方法还包括如图4所示的步骤S1101-S1103：

[0099] 步骤S1101：获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；

[0100] 步骤S1102：将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；

- [0101] 步骤S1103: 将所述文本参考特征和所述时长参考特征作为所述声学模型的输入, 所述语音参考特征作为声学模型的输出, 对所述声学模型进行训练。
- [0102] 在进行模型训练之前, 首先需要对数据进行标识, 确定文本对应的时长参考特征和语音参考特征。其中, 时长参考特征是与待合成文本对应的时长特征, 语音参考特征是与待合成文本对应的语音特征。本实施例中通过预先训练样本集对时长预测模型和声学模型进行训练, 使得模型具备准确预测待合成文本对应的时长特征和语音特征的能力。
- [0103] 针对训练数据集包含的每一条训练文本, 将训练文本对应的文本参考特征作为输入, 将对应的时长参考特征作为输出, 对预设的时长预测模型进行训练, 以使时长训练模型具备时长特征预测的功能。
- [0104] 针对训练数据集包含的每一条训练文本, 将训练文本对应的文本参考特征和时长参考特征作为输入, 将对应的语音参考特征作为输出, 对预设的声学模型进行训练, 以使声学模型具备语音特征预测的功能。
- [0105] 进一步的, 在一个实施例中, 还需要对文本特征预测的各个模型进行模型的训练, 具体包含对分词模型、多音字预测模型、韵律预测模型的训练。
- [0106] 也就是说, 通过训练样本集对文本特征预测涉及的分词模型、多音字预测模型和韵律预测模型进行训练, 使得分词模型、多音字预测模型和韵律预测模型分别具备有根据待合成文本预测分词特征、多音字特征和韵律特征等文本特征的能力。
- [0107] 如图5所述, 上述语音合成方法还包括如图5所示的步骤S2101-S2103:
- [0108] 在一个实施例中, 所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征; 所述方法包括:
- [0109] 步骤S2101: 将所述训练文本作为所述分词模型的输入, 所述分词参考特征作为分词模型的输出, 对所述分词模型进行训练。
- [0110] 步骤S2102: 将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入, 所述多音字参考特征作为多音字预测模型的输出, 对所述多音字预测模型进行训练。
- [0111] 步骤S2103: 将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的

输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。

[0112] 其中，训练样本集还可以包括多个训练文本以及期望模型输出的分词特征、多音字特征和韵律特征。分词参考特征是期望分词模型根据训练文本输出的分词特征，多音字参考特征是期望多音字预测模型根据训练文本和对应的分词特征输出的多音字特征。韵律参考特征是期望韵律预测模型根据训练文本和对应的分词特征输出的韵律特征。

[0113] 针对训练数据集包含的每一条训练文本，将训练文本作为输入，将对应的分词参考特征作为输出，对预设的分词模型进行训练，以使分词模型具备分词特征预测的功能。

[0114] 针对训练数据集包含的每一条训练文本，将训练文本以及对应的分词参考特征作为输入，将对应的多音字参考特征作为输出，对预设的多音字预测模型进行训练，以使多音字预测模型具备多音字特征预测的功能。

[0115] 针对训练数据集包含的每一条训练文本，将训练文本以及对应的分词参考特征作为输入，将对应的韵律参考特征作为输出，对预设的韵律预测模型进行训练，以使韵律预测模型具备韵律特征预测的功能。

[0116] 本实施例中通过预先处理好的数据对分词模型、多音字预测模型和韵律预测模型进行训练，使得模型能够准确预测出训练文本对应的分词特征、多音字特征和韵律特征。

[0117] 在具体的预测过程中，可以同时获取多个待合成文本，获取每一个待合成文本对应的文本特征。将多个待合成文本对应的文本特征进行筛选和排序，输入预设的特征队列。获取特征队列中预设数量的文本特征，输入时长预测模型和声学模型进行预测，得到对应的特征。其中，生成每一个待合成文本对应的文本特征和对预设数量的文本特征进行预测的步骤是同步进行的。

[0118] 如图6所述，上述语音合成方法还包括如图6所示的步骤S3101-S3102：

[0119] 步骤S3101：通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；

[0120] 步骤S3102：将多个待合成文本对应的文本特征添加至预设的特征队列，当所

述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。

- [0121] 其中，文本迭代器用于获取多个待合成文本和对应的文本特征中连续的数据，可以在多个获取待合成文本中的文本特征的进程中连续、迭代出文本特征，特征队列是包含多个文本特征的有序的集合。预设条件是进行将文本特征输入时长预测模型的条件，可以是文本特征达到一定的数量，也可以是达到预设的文本特征的获取时间。预设数量是特征队列输出文本特征的数量，可以是固定数值，也可以是依据某种规律变换的数值。
- [0122] 本实施例中添加了队列特征和文本迭代器对多个待合成文本进行处理，使得更加有效和快速对待合成文本进行文本转换，提高了语音合成和模型训练的效率。
- [0123] 示例性的，在上述模型训练以及模型训练的过程中，时长预测模型、声学模型、分词模型、多音字预测模型和/或韵律预测模型为神经网络模型，在一个具体的实施例中，为双向长短期记忆神经网络模型（BiLSTM模型）。。
- [0124] 其中，BiLSTM(Bi-directional Long Short-Term Memory，双向长短期记忆网络)模型，使得数据具有时间依赖性，并且对数据进行全局化处理。能通过前后词等特征来更好的预测结果。
- [0125] 如图7所示，在一个实施例中，提出了一种语音合成装置，该装置包括：
- [0126] 获取模块702，用于获取待合成文本；
- [0127] 文本特征确定模块704，用于获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；
- [0128] 时长特征确定模块706，用于将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；
- [0129] 语音特征确定模块708，用于将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；
- [0130] 转换模块710，用于将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

- [0131] 如图8所示，在一个实施例中，所述文本特征确定模块704还包括：预处理单元，用于对所述待合成文本进行正则化处理。
- [0132] 如图8所示，在一个实施例中，所述文本特征确定模块704还包括：分词特征确定单元，用于将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；多音字特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；韵律特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取所述待合成文本对应的韵律特征。
- [0133] 如图9所示，在一个实施例中，所述装置还包括：获取训练模块703，用于获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；时长训练模块705，用于将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；语音训练模块707，用于将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。
- [0134] 如图10所示，在一个实施例中，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征，所述装置包括：分词训练模块7041，用于将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。多音字训练模块7043，用于将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。韵律训练模块7045，用于将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。
- [0135] 如图11所示，在一个实施例中，所述装置还包括：文本获取模块709，用于通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；文本预测模块711，用于将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，

获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。

[0136] 在一个实施例中，所述时长预测模型、声学模型、分词模型、多音字预测模型和/或韵律预测模型为BiLSTM模型。

[0137] 图12示出了一个实施例中智能终端的内部结构图。该智能终端具体可以是终端，也可以是服务器。如图12所示，该智能终端包括通过系统总线连接的处理器、存储器和网络接口。其中，存储器包括非易失性存储介质和内存储器。该智能终端的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现年龄识别方法。该内存储器中也可储存有计算机程序，该计算机程序被处理器执行时，可使得处理器执行年龄识别方法。本领域技术人员可以理解，图12中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0138] 在一个实施例中，提出了一种智能终端，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：获取待合成文本；获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

[0139] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤之前，还包括：对所述待合成文本进行正则化处理。

[0140] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤还包括：将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；将所述待合成文本和/或所述分词特征输入预设的韵

律预测模型，获取所述待合成文本对应的韵律特征。

- [0141] 在一个实施例中，所述方法还包括：获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练。将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。
- [0142] 在一个实施例中，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征；所述方法包括：将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。
- [0143] 在一个实施例中，所述方法还包括：通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。
- [0144] 在一个实施例中，提出了一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行以下步骤：获取待合成文本；获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

- [0145] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤之前，还包括：对所述待合成文本进行正则化处理。
- [0146] 在一个实施例中，所述获取所述待合成文本的文本特征的步骤还包括：将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取所述待合成文本对应的韵律特征。
- [0147] 在一个实施例中，所述方法还包括：获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练。将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。
- [0148] 在一个实施例中，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征；所述方法包括：将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。
- [0149] 在一个实施例中，所述方法还包括：通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。
- [0150] 采用本发明的语音合成方法、装置、终端及存储介质之后，在进行语音合成的

过程中，首先获取待合成文本的待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征等文本特征；然后将文本特征输入预设的时长预测模型获取对应的时长特征；将文本特征和时长特征输入预设的声学模型获取对应的语音特征；最后将语音特征转换成语音，生成与待合成文本对应的目标语音。在进行语音合成的特征提取的过程中，考虑的文本特征包括了多音字特征和韵律特征等，并结合模型预测的时长特征，得到最终合成语音的过程中所需要的语音特征。也就是说，本发明提供的语音合成方法、装置、终端及存储介质考虑了多种文本特征和时长特征生成的语音特征，使得合成的语音更加准确，提高了语音合成的准确性，提高了用户体验。

[0151] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一非易失性计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink) DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0152] 以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

[0153] 以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权

利要求为准。

权利要求书

- [权利要求 1] 一种语音合成方法，其特征在于，所述方法包括：
- 获取待合成文本；
- 获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；
- 将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；
- 将所述文本特征和所述时长特征输入预设的声学模型，获取与所述待合成文本对应的语音特征；
- 将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。
- [权利要求 2] 根据权利要求1所述的方法，其特征在于，所述获取所述待合成文本的文本特征的步骤之前，还包括：
- 对所述待合成文本进行正则化处理。
- [权利要求 3] 根据权利要求1所述的方法，其特征在于，所述获取所述待合成文本的文本特征的步骤还包括：
- 将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；
- 将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取与所述待合成文本对应的多音字特征；
- 将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取与所述待合成文本对应的韵律特征。
- [权利要求 4] 根据权利要求1所述的方法，其特征在于，所述方法还包括：
- 获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；
- 将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；

将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。

[权利要求 5]

根据权利要求3所述的方法，其特征在于，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征；

所述方法包括：

将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。

将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。

将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。

[权利要求 6]

根据权利要求1所述的方法，其特征在于，所述方法还包括：

通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；

将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。

[权利要求 7]

一种语音合成装置，其特征在于，所述装置包括：

获取模块，用于获取待合成文本；

文本特征确定模块，用于获取所述待合成文本的文本特征，所述文本特征包括分词特征、多音字特征和/或韵律特征中的至少一个；

时长特征确定模块，用于将所述文本特征输入预设的时长预测模型，获取与所述文本特征对应的时长特征；

语音特征确定模块，用于将所述文本特征和所述时长特征输入预设的

声学模型，获取与所述待合成文本对应的语音特征；

转换模块，用于将所述语音特征转换成语音，生成与所述待合成文本对应的目标语音。

[权利要求 8]

根据权利要求7所述的装置，其特征在于，所述文本特征确定模块还包括：

分词特征确定单元，用于将所述待合成文本输入预设的分词模型，获取与所述待合成文本对应的分词特征；

多音字特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的多音字预测模型，获取所述待合成文本对应的多音字特征；

韵律特征确定单元，用于将所述待合成文本和/或所述分词特征输入预设的韵律预测模型，获取所述待合成文本对应的韵律特征。

[权利要求 9]

根据权利要求7所述的装置，其特征在于，所述装置还包括：

获取训练模块，用于获取训练样本集，所述训练样本集包含多个训练文本以及对应的文本参考特征、时长参考特征和/或语音参考特征；

时长训练模块，用于将所述训练文本对应的文本参考特征作为所述时长预测模型的输入，所述时长参考特征作为时长预测模型的输出，对所述时长预测模型进行训练；

语音训练模块，用于将所述文本参考特征和所述时长参考特征作为所述声学模型的输入，所述语音参考特征作为声学模型的输出，对所述声学模型进行训练。

[权利要求 10]

根据权利要求8所述的装置，其特征在于，所述训练样本集还包含与所述多个训练文本对应的分词参考特征、多音字参考特征和/或韵律参考特征，所述装置包括：

分词训练模块，用于将所述训练文本作为所述分词模型的输入，所述分词参考特征作为分词模型的输出，对所述分词模型进行训练。

多音字训练模块，用于将所述训练文本和/或所述分词参考特征作为所述多音字预测模型的输入，所述多音字参考特征作为多音字预测模型的输出，对所述多音字预测模型进行训练。

韵律训练模块，用于将所述训练文本和/或所述分词参考特征作为所述韵律预测模型的输入，所述韵律参考特征作为韵律预测模型的输出，对所述韵律预测模型进行训练。

- [权利要求 11] 根据权利要求7所述的装置，其特征在于，所述装置还包括：
- 文本获取模块，用于通过文本迭代器获取多个待合成文本，针对每一个待合成文本，分别执行所述获取所述待合成文本的文本特征的步骤；
- 文本预测模块，用于将多个待合成文本对应的文本特征添加至预设的特征队列，当所述特征队列满足预设条件时，获取所述特征队列中的预设数量的文本特征，并将所述预设数量的文本特征分别输入所述时长预测模型，以使同步执行所述获取与所述文本特征对应的时长特征的步骤。
- [权利要求 12] 一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如权利要求1至6中任一项所述方法的步骤。
- [权利要求 13] 一种智能终端，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如权利要求1至6中任一项所述方法的步骤。

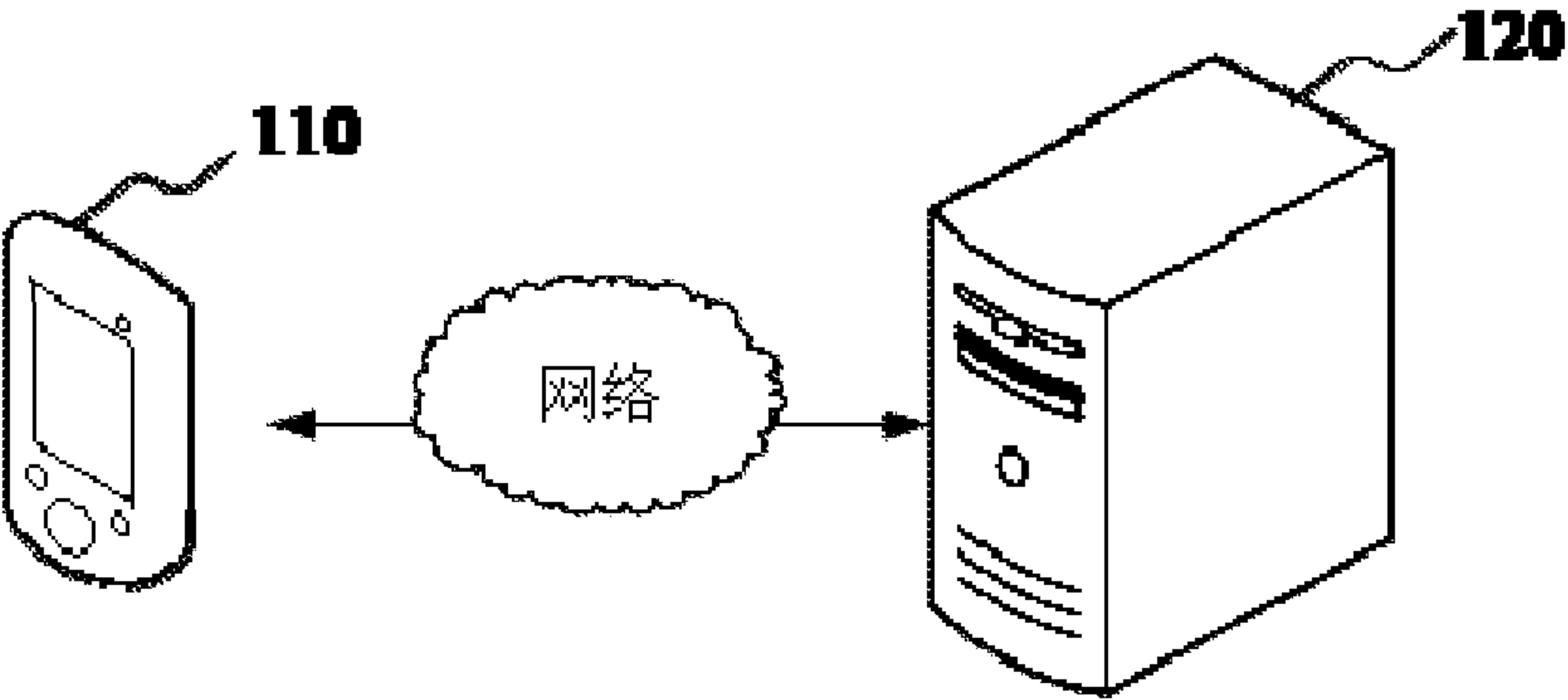


图 1

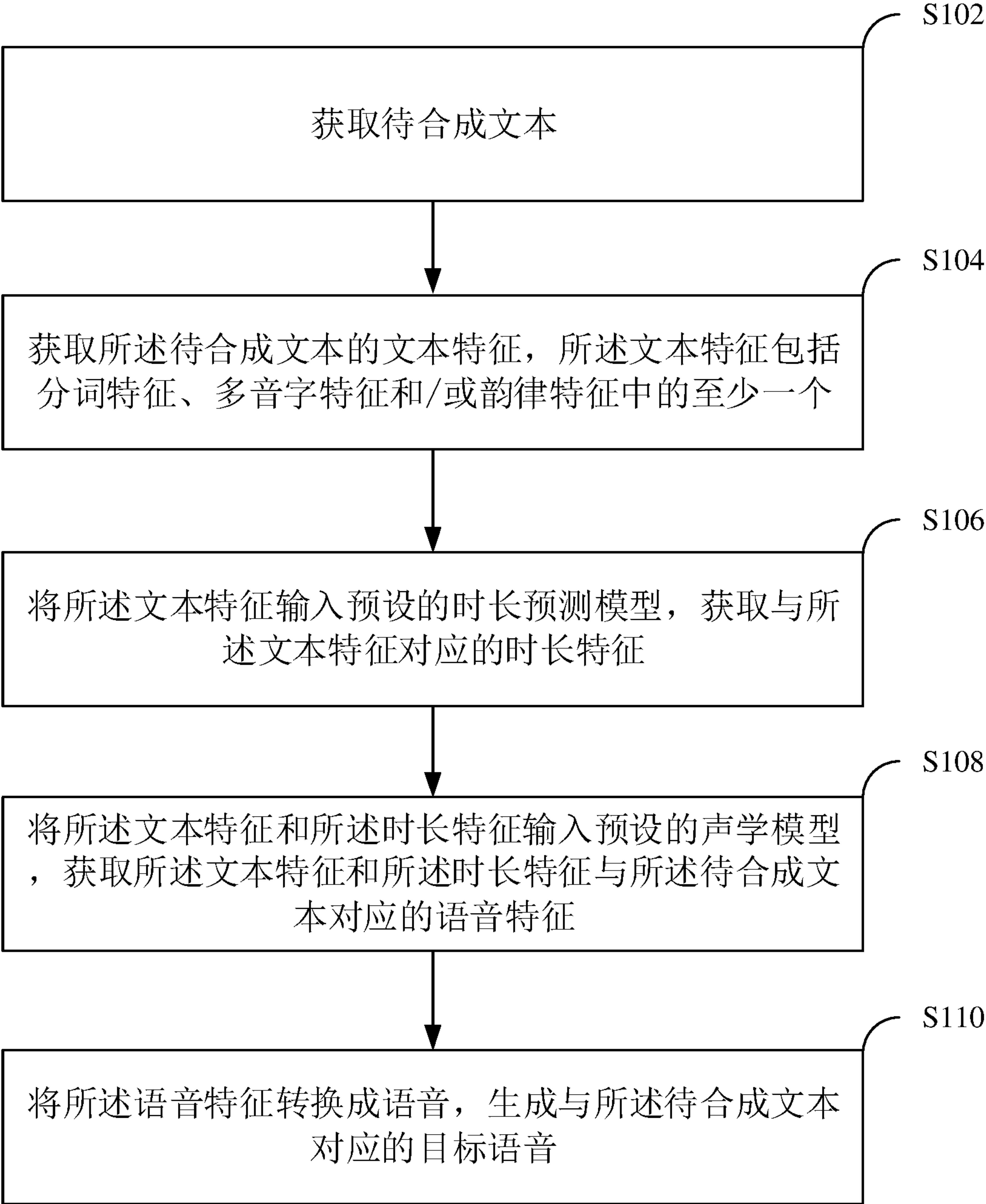


图 2

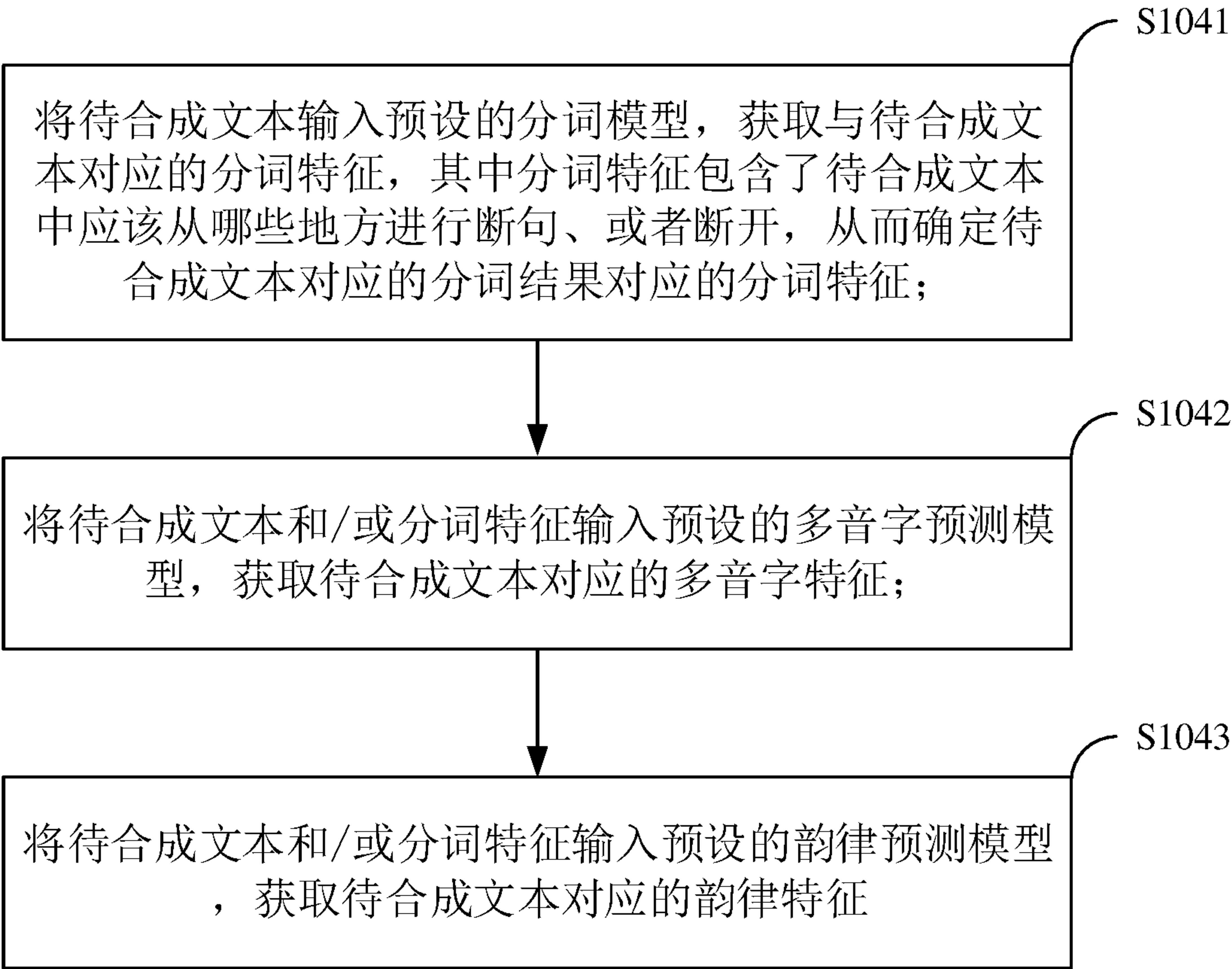


图 3

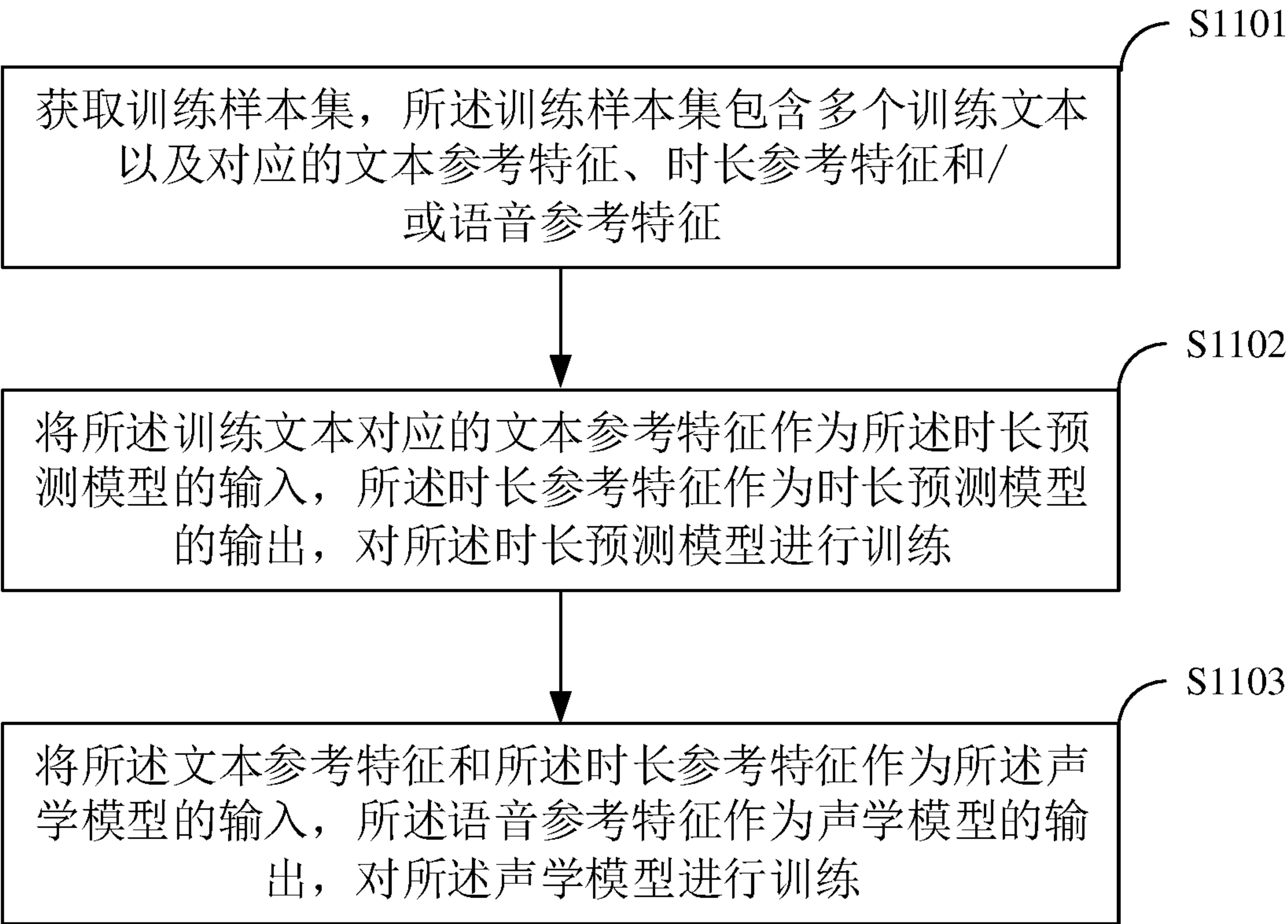


图 4

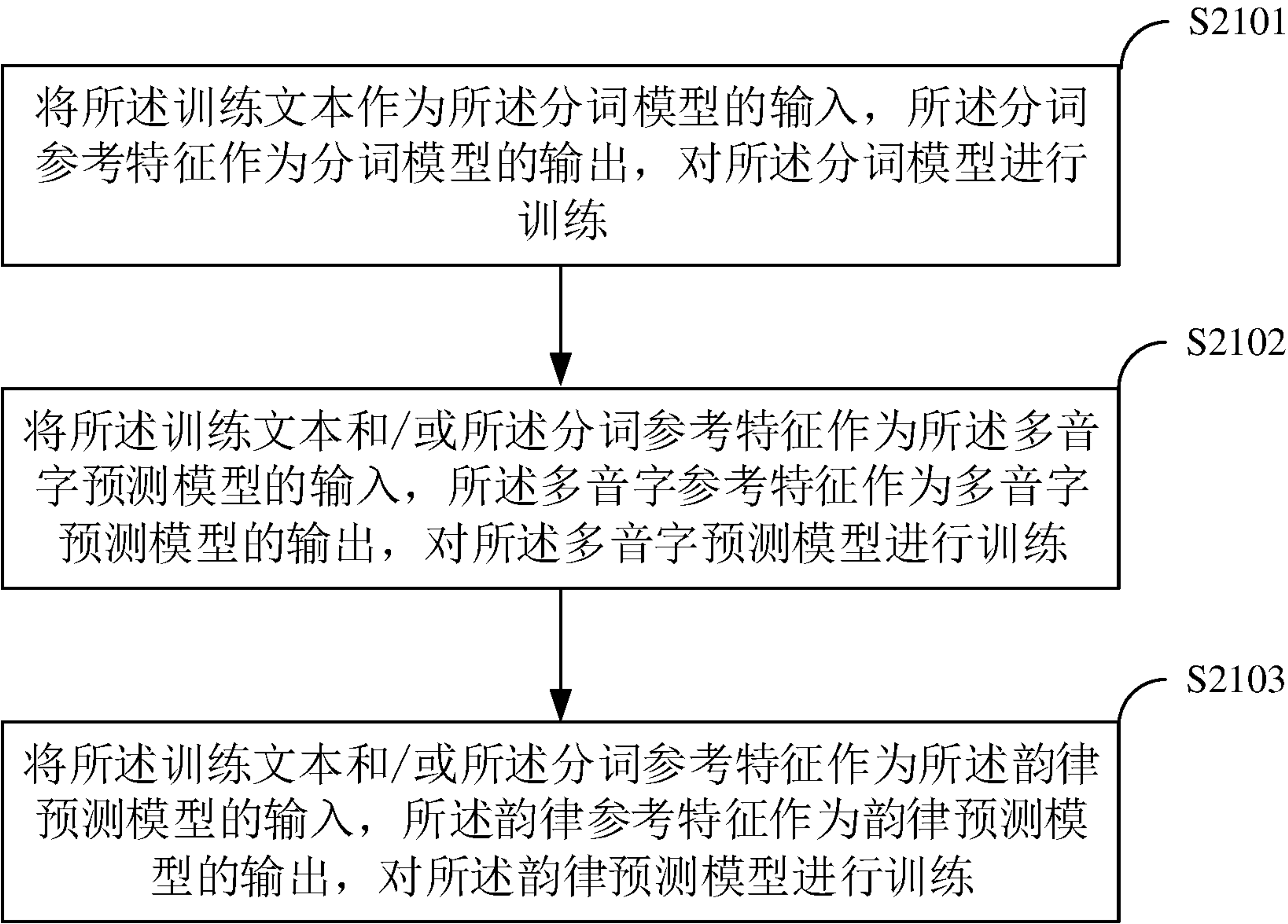


图 5

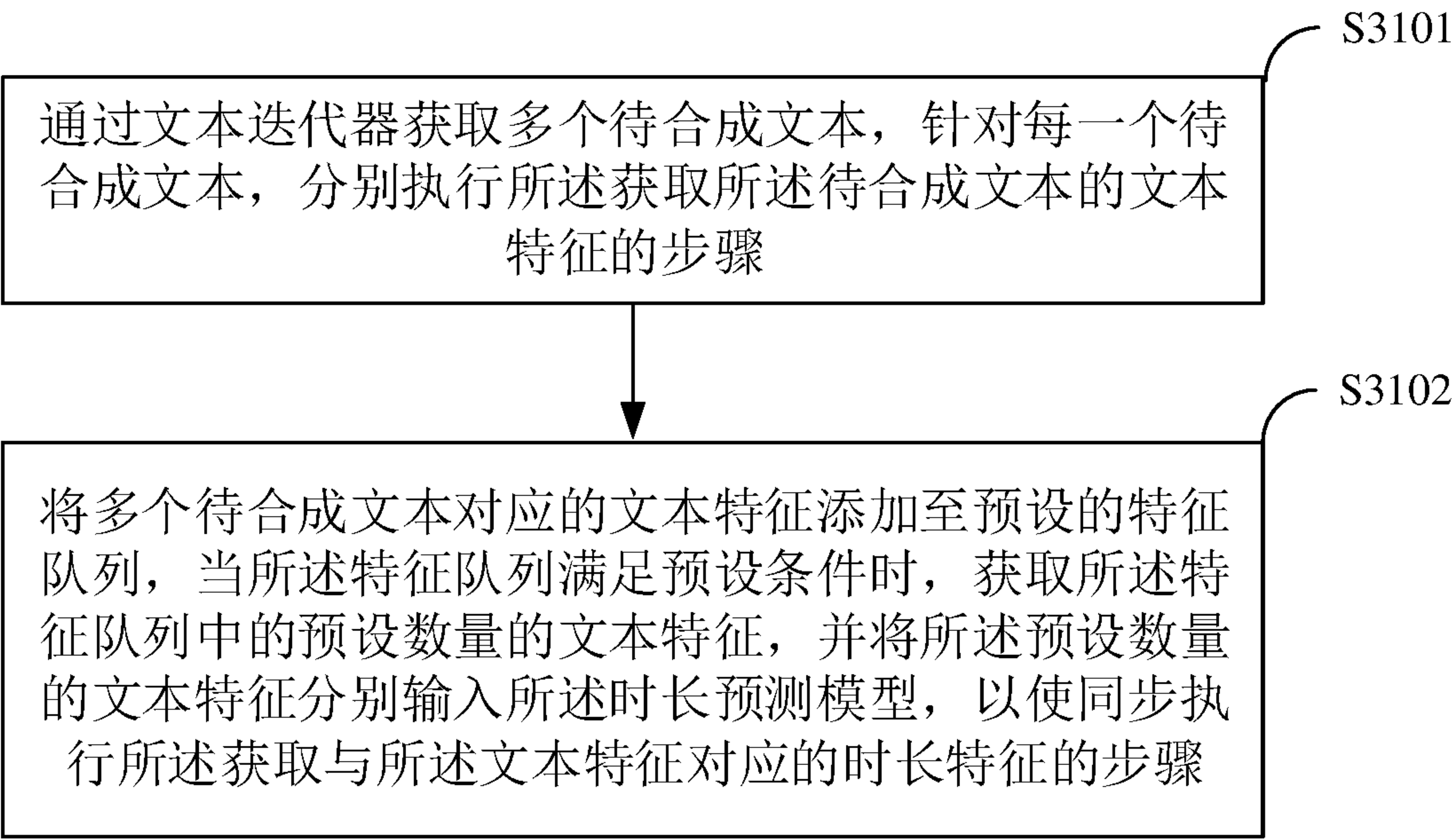


图 6

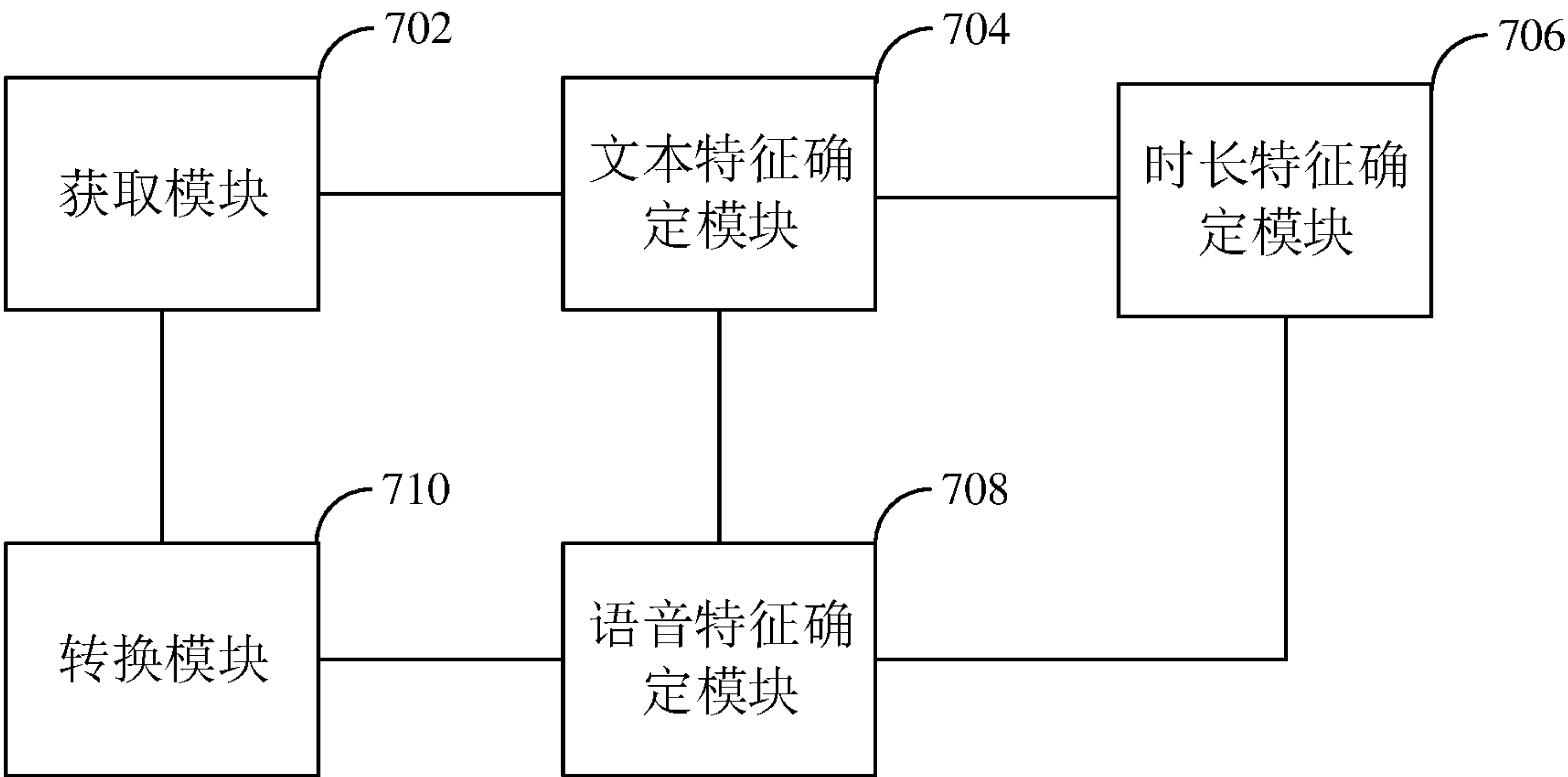


图 7

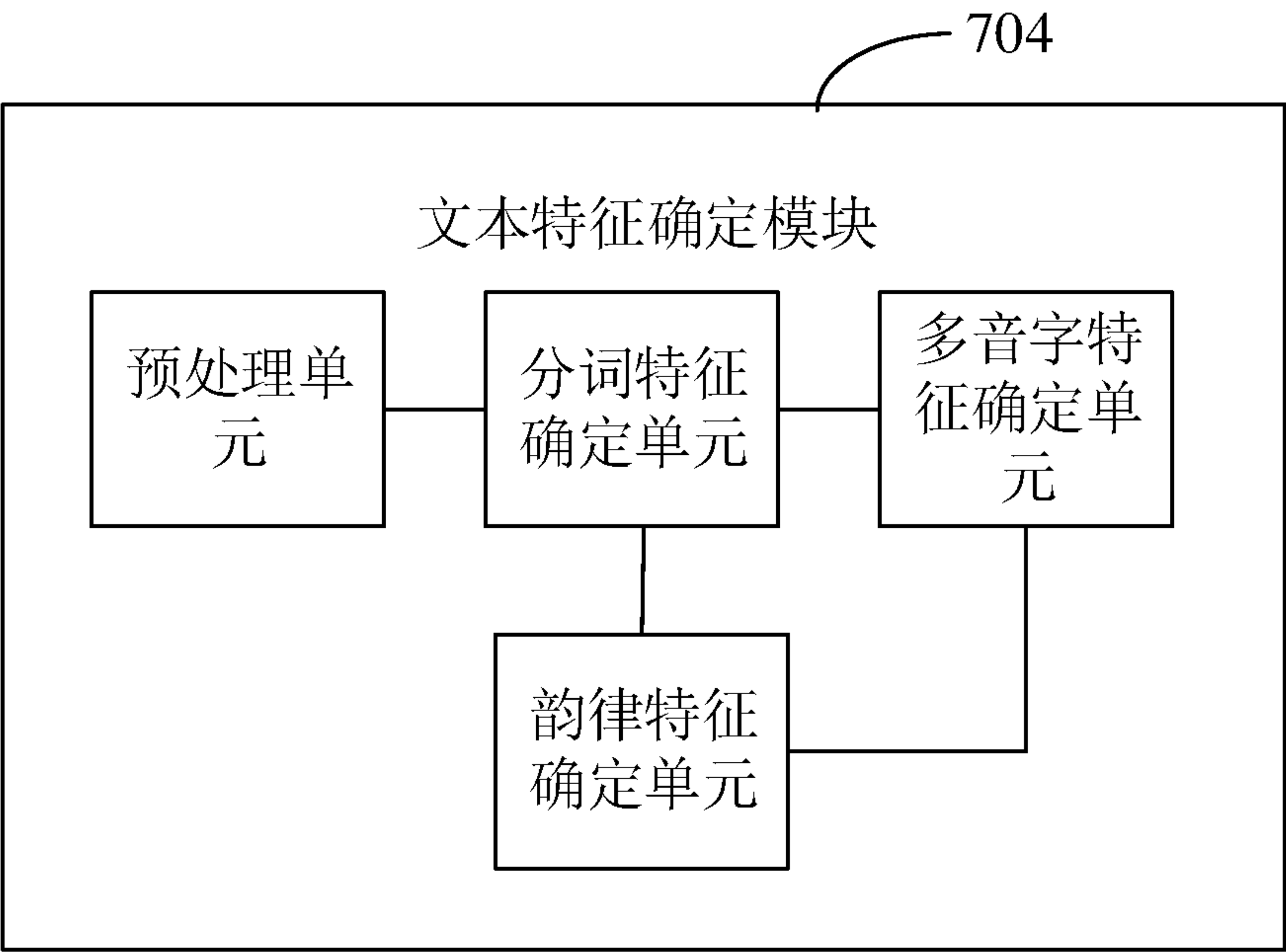


图 8

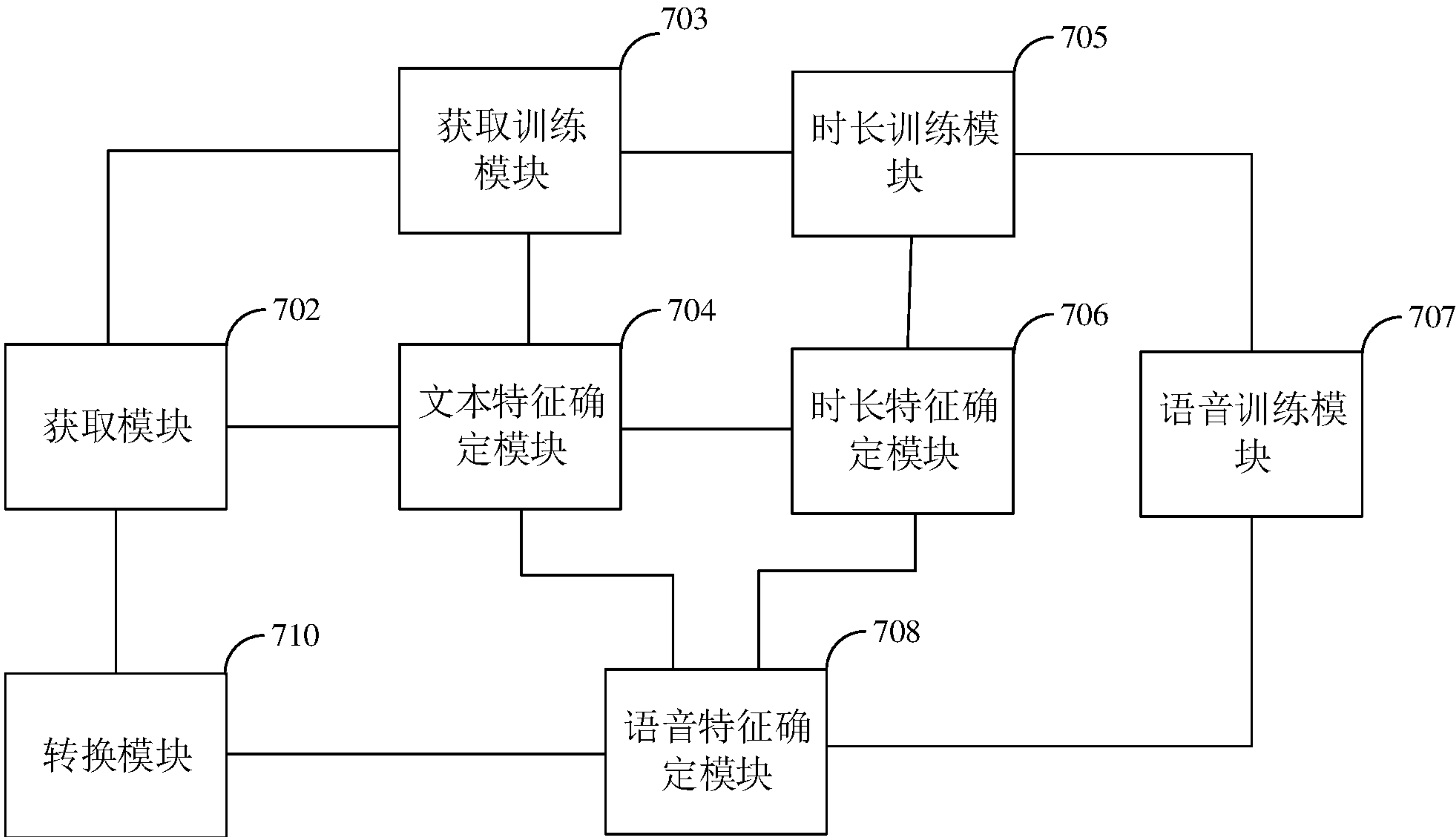


图 9

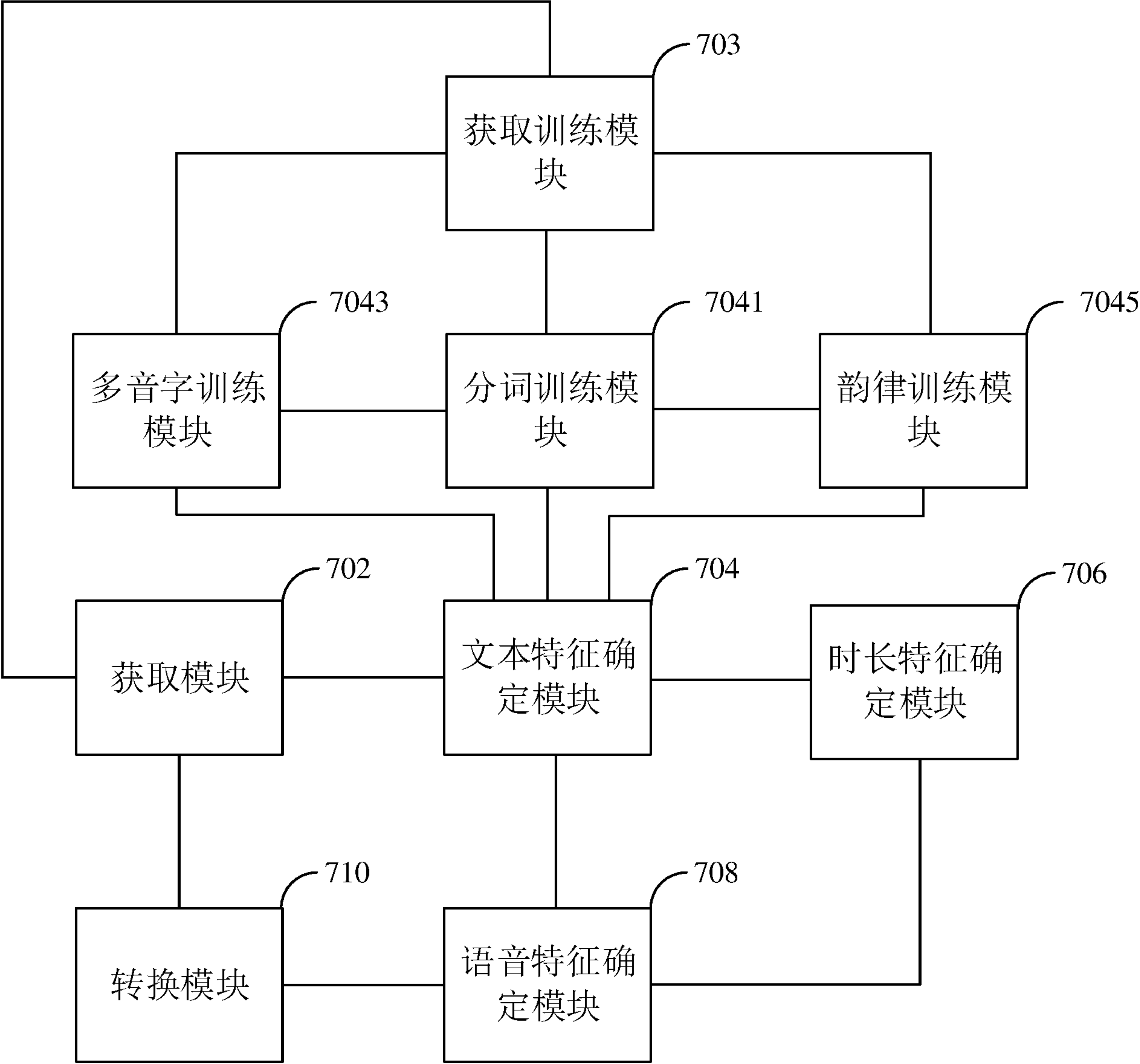


图 10

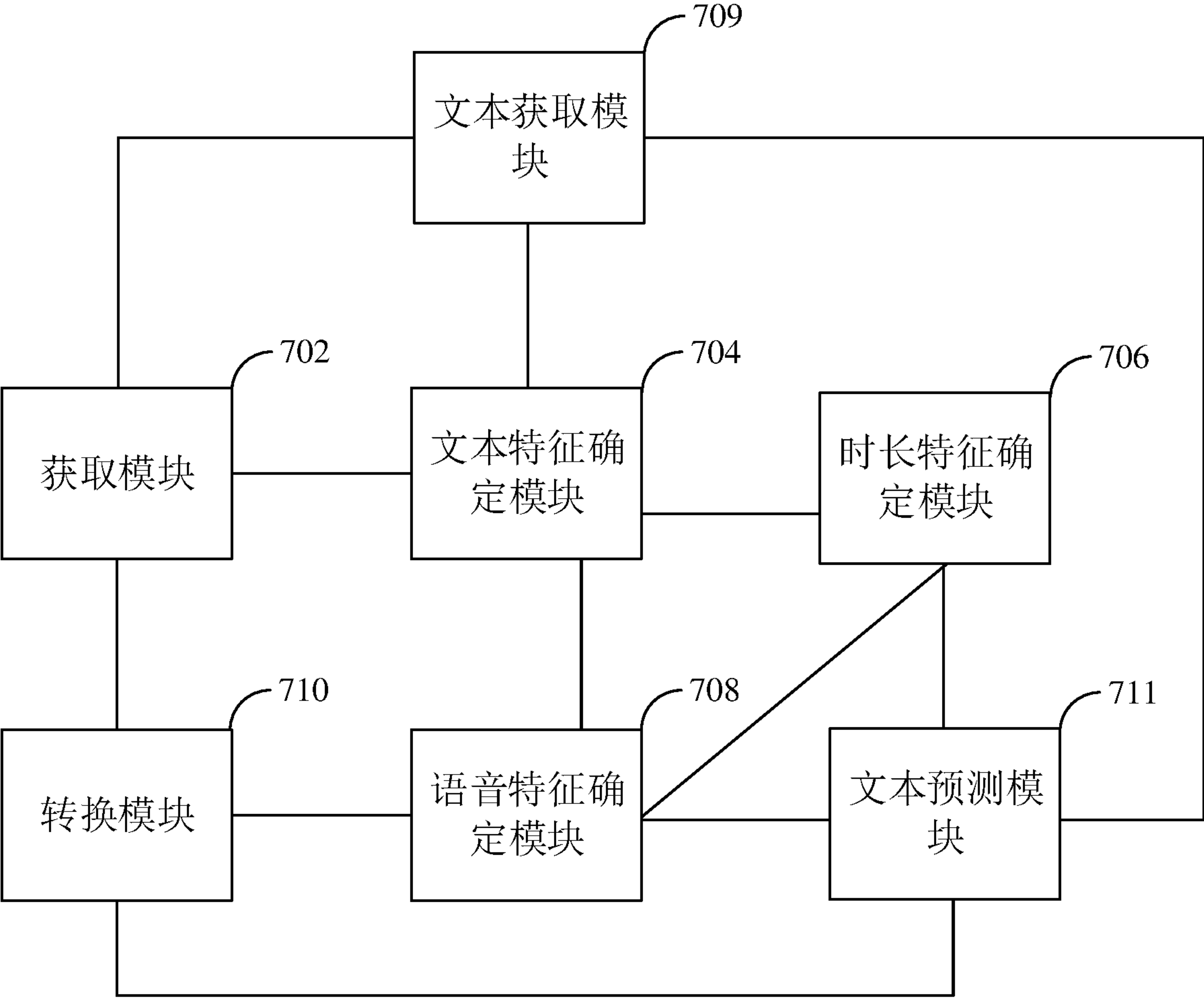


图 11

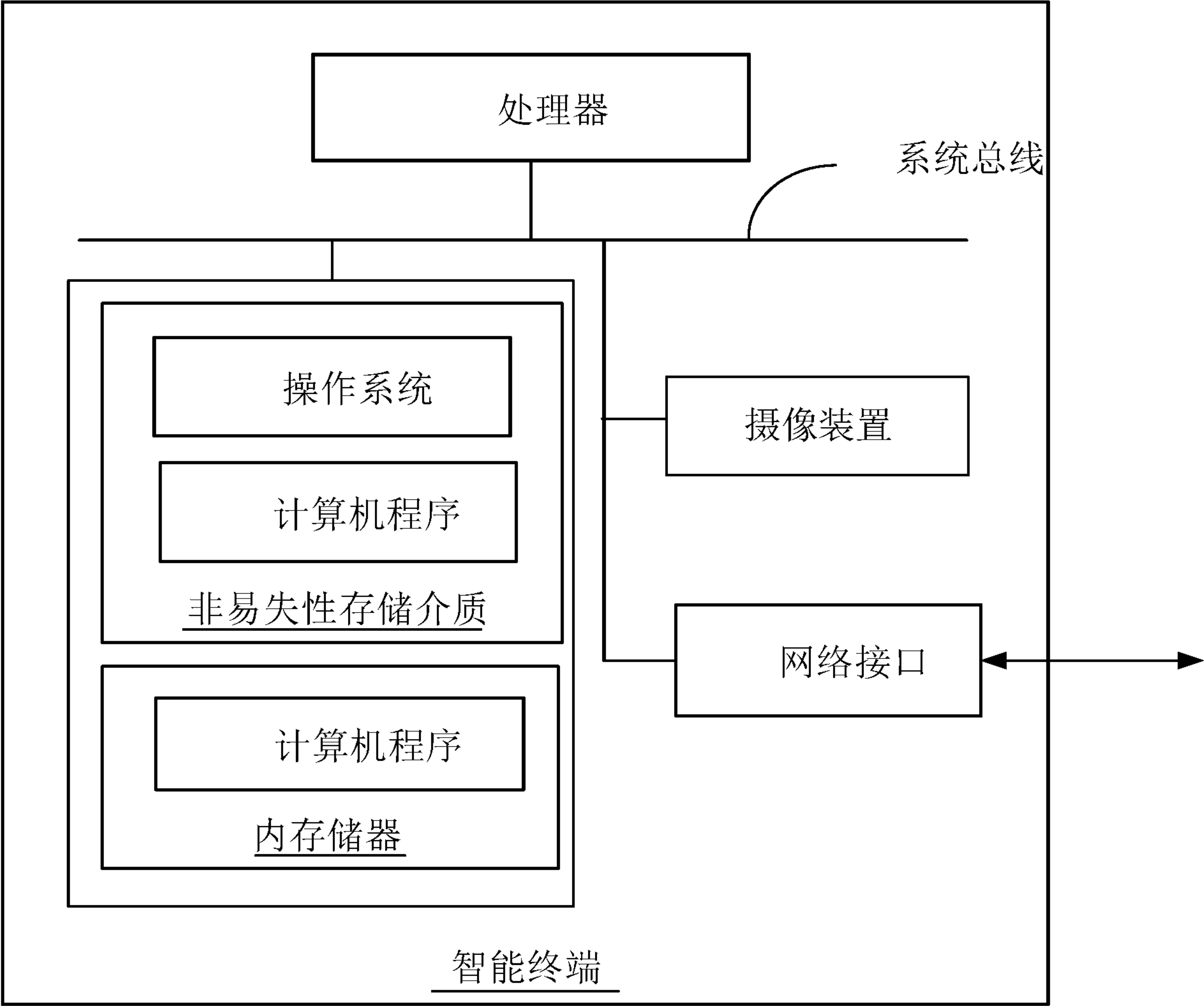


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/130766

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/047(2013.01)i; G10L 13/08(2013.01)i; G10L 13/10(2013.01)i; G10L 25/30(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

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)

CNKI, CNPAT, DWPI, EPODOC: 优必选, 李贤, 黄东延, 丁万, 张皓, 白洛玉, 熊友军, 语音, 合成, 音频, 韵律, 多音字, 训练, 神经网络, 模型, 分词, 时长, 正则化, speech, voice, synthesis+, acoustic+, model+, train+, criteri+, prosody, phonem+, neural network

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105336322 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 17 February 2016 (2016-02-17) description paragraphs [0037]-[0137] and figures 1-3, 7-10	1-13
A	CN 105185372 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 23 December 2015 (2015-12-23) entire document	1-13
A	CN 106507321 A (XINJIANG AGRICULTURAL UNIVERSITY) 15 March 2017 (2017-03-15) entire document	1-13
A	CN 106601228 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 26 April 2017 (2017-04-26) entire document	1-13
A	CN 110299131 A (SUZHOU QDREAMER NETWORK SCIENCE AND TECH CO., LTD.) 01 October 2019 (2019-10-01) entire document	1-13

☒ 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

01 September 2020

Date of mailing of the international search report

12 October 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
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/CN2019/130766

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107039034 A (IFLYTEK CO., LTD.) 11 August 2017 (2017-08-11) entire document	1-13
A	US 2016049144 A1 (AT&T INTELLECTUAL PROPERTY I, L.P.) 18 February 2016 (2016-02-18) entire document	1-13
A	US 2019172443 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 06 June 2019 (2019-06-06) entire document	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/130766

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105336322	A	17 February 2016	CN	105336322	B	10 May 2017
CN	105185372	A	23 December 2015	CN	105185372	B	22 March 2017
				US	10410621	B2	10 September 2019
				US	2018254034	A1	06 September 2018
				WO	2017067206	A1	27 April 2017
CN	106507321	A	15 March 2017	None			
CN	106601228	A	26 April 2017	CN	106601228	B	04 February 2020
CN	110299131	A	01 October 2019	None			
CN	107039034	A	11 August 2017	CN	107039034	B	01 May 2020
US	2016049144	A1	18 February 2016	US	10199034	B2	05 February 2019
US	2019172443	A1	06 June 2019	US	10418025	B2	17 September 2019

国际检索报告		国际申请号 PCT/CN2019/130766
A. 主题的分类 G10L 13/047(2013.01)i; G10L 13/08(2013.01)i; G10L 13/10(2013.01)i; G10L 25/30(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 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)) CNKI, CNPAT, DWPI, EPDOC:优必选, 李贤, 黄东延, 丁万, 张皓, 白洛玉, 熊友军, 语音, 合成, 音频, 韵律, 多音字, 训练, 神经网络, 模型, 分词, 时长, 正则化, speech, voice, synthesis+, acoustic+, model+, train+, criteri+, prosody, phonem+, neural network		
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 105336322 A (百度在线网络技术北京有限公司) 2016年 2月 17日 (2016 - 02 - 17) 说明书第[0037]-[0137]段及附图1-3, 7-10	1-13
A	CN 105185372 A (百度在线网络技术北京有限公司) 2015年 12月 23日 (2015 - 12 - 23) 全文	1-13
A	CN 106507321 A (新疆农业大学) 2017年 3月 15日 (2017 - 03 - 15) 全文	1-13
A	CN 106601228 A (百度在线网络技术北京有限公司) 2017年 4月 26日 (2017 - 04 - 26) 全文	1-13
A	CN 110299131 A (苏州奇梦者网络科技有限公司) 2019年 10月 1日 (2019 - 10 - 01) 全文	1-13
A	CN 107039034 A (科大讯飞股份有限公司) 2017年 8月 11日 (2017 - 08 - 11) 全文	1-13
A	US 2016049144 A1 (AT&T INTELLECTUAL PROPERTY I, L.P.) 2016年 2月 18日 (2016 - 02 - 18) 全文	1-13
☒ 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 2020年 9月 1日		国际检索报告邮寄日期 2020年 10月 12日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 刘畅 电话号码 86-(10)-53962364

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	US 2019172443 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2019年 6月 6日 (2019 - 06 - 06) 全文	1-13

国际申请号 PCT/CN2019/130766

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	105336322	A	2016年 2月 17日	CN	105336322	B	2017年 5月 10日
CN	105185372	A	2015年 12月 23日	CN	105185372	B	2017年 3月 22日
				US	10410621	B2	2019年 9月 10日
				US	2018254034	A1	2018年 9月 6日
				WO	2017067206	A1	2017年 4月 27日
CN	106507321	A	2017年 3月 15日	无			
CN	106601228	A	2017年 4月 26日	CN	106601228	B	2020年 2月 4日
CN	110299131	A	2019年 10月 1日	无			
CN	107039034	A	2017年 8月 11日	CN	107039034	B	2020年 5月 1日
US	2016049144	A1	2016年 2月 18日	US	10199034	B2	2019年 2月 5日
US	2019172443	A1	2019年 6月 6日	US	10418025	B2	2019年 9月 17日