(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
2022 年 9 月 1 日 (01.09.2022)

WIPO | PCT

(10) 国际公布号
WO 2022/178996 A1

(51) 国际专利分类号:
G10L 13/047 (2013.01) *G10L 25/30* (2013.01)
G10L 21/02 (2013.01)

(21) 国际申请号: PCT/CN2021/096668

(22) 国际申请日: 2021 年 5 月 28 日 (28.05.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110219479.8 2021年2月26日 (26.02.2021) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

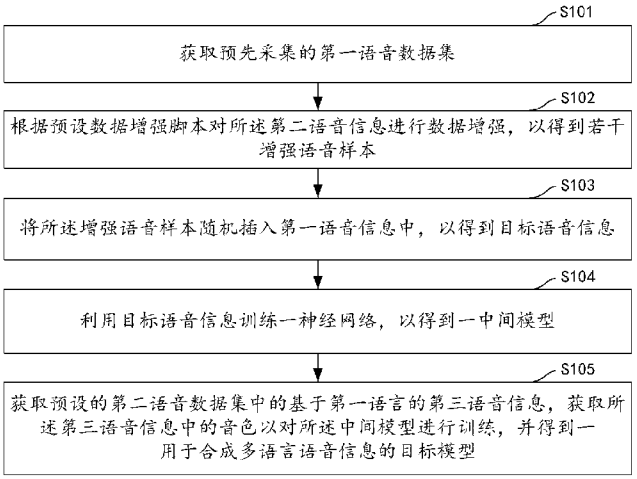
(72) 发明人: 陈闽川 (CHEN, Minchuan); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 马骏 (MA, Jun); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 王少军 (WANG, Shaojun); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所 (SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路6009号绿景广场B栋20层B, 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,

(54) **Title:** MULTI-LANGUAGE SPEECH MODEL GENERATION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) **发明名称:** 多语言语音模型生成方法、装置、计算机设备及存储介质



S101 Obtain a pre-collected first speech data set
S102 According to a preset data enhancement script, perform data enhancement on second speech information to obtain a plurality of enhanced speech samples
S103 Randomly insert the enhanced speech samples into first speech information to obtain target speech information
S104 Train a neural network by using the target speech information, so as to obtain an intermediate model
S105 Obtain third speech information based on a first language in a preset second speech data set, obtain timbre in the third speech information to train the intermediate model, and obtain a target model used for synthesizing multi-language speech information

图 1

(57) **Abstract:** A multi-language speech model generation method comprises: obtaining a pre-collected first speech data set (S101); according to a preset data enhancement script, performing data enhancement on second speech information to obtain enhanced speech samples (S102); randomly inserting the enhanced speech samples into first speech information to obtain target speech information (S103); training a neural network by using the target speech information, so as to obtain an intermediate model (S104); and obtaining third speech information based on a first language in a preset second speech data set, obtaining timbre in the third speech information to train the intermediate model, and obtaining a target model used for synthesizing multi-language speech information (S105). According to the multi-language speech model generation method, a multi-language speech model generation apparatus (100), a computer device (600), and a storage medium, data collection is convenient, the target model for synthesizing multi-language speech so as to satisfy a single timbre requirement can be obtained, and the present application can further be applied to scenarios, such as smart government affairs, thereby promoting the construction of a smart city, and improving the user experience of a user.

WO 2022/178996 A1

[见续页]

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：多语言语音模型生成方法包括获取预先采集的第一语音数据集（S101）；根据预设数据增强脚本对第二语音信息进行数据增强，以得到增强语音样本（S102）；将增强语音样本随机插入第一语音信息以得到目标语音信息（S103）；利用目标语音信息训练一神经网络，以得到一中间模型（S104）；获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取第三语音信息中的音色以对中间模型进行训练，并得到一用于合成多语言语音信息的目标模型（S105）。多语言语音模型生成方法、多语言语音模型生成装置（100）、计算机设备（600）及存储介质不仅数据收集便捷，可得到对多语言语音合成满足单音色需求的目标模型，还能应用于智慧政务等场景中，从而推动智慧城市的建设，提高用户使用体验度。

多语言语音模型生成方法、装置、计算机设备及存储介质

本申请要求于2021年2月26日提交中国专利局、申请号为202110219479.8，发明名称为“多语言语音模型生成方法、装置、计算机设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能领域，尤其涉及一种多语言语音模型生成方法、装置、计算机设备及存储介质。

背景技术

目前，多语言的语音合成是学术界的热门话题，完整的高质量的多语音合成体系和模型仍然是大家讨论的焦点所在。一般来说，在说话过程中，若存在两个语言体系，即会存在一个主体系及一个从体系，通常以自己的母语为主体系，而另一个语言为从体系，作为从体系的语言在语句中不会大段地连续出现，只是以单词的形式嵌入到句子中起辅助作用。例如人们在说汉语的时候，有时会夹杂着英语的字母或者部分英文的单词。

现今多语言语音合成体系主要包括作为主体系的主要语言和以及作为从体系的辅助语言，若只使用主要语言训练神经网络模型，即难以获得针对辅助语言的准确语音。在神经网络模型训练过程中获取包含多种语言的数据集是很困难的，尤其是较难找到精通多种语言的说话人进行数据采集。而使用多说话人的混合数据集训练神经网络模型时，需要收集大量的分别以主体语言为母语的说话人的预料以及以辅助语言为母语的说话人的预料等数据，但是发明人意识到在训练过程种由于不同说话人的音色不同，会导致神经网络模型生成的语句种存在明显的两个音色现象。故可知，在多语言语音合成模型的训练上，存在单说话人训练数据难收集以及多说话人训练得到的模型质量欠佳等问题。

发明内容

本申请实施例提供一种多语言语音模型生成方法、装置、计算机设备及存储介质，其不仅数据收集便捷，可得到用于生成对多语言语音合成单音色需求的多语言语音信息的目标模型，还能应用于智慧政务等场景中，从而推动智慧城市的建设，提高用户使用体验度。

第一方面，本申请实施例提供了一种多语言语音模型生成方法，该方法包括：

获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

第二方面，本申请实施例还提供了一种多语言语音模型生成装置，该装置包括：

数据获取单元，用于获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

数据增强单元，用于根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

语音处理单元，用于将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

第一训练单元，用于利用目标语音信息训练一神经网络，以得到一中间模型；

第二训练单元，用于调用预设的第二语音数据集中的基于第一语言的第三语音信息，以获取所述第三语音信息中的音色对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

第三方面，本申请实施例还提供了一种计算机设备，其包括存储器及处理器，所述存储器上存储有计算机程序，所述处理器执行所述计算机程序时实现如下步骤：

获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

第四方面，本申请实施例还提供了一种计算机可读存储介质，所述存储介质存储有计算机程序，所述计算机程序当被处理器执行时可实现如下步骤：

获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

本申请实施例提供了一种多语言语音模型生成方法、装置、计算机设备及存储介质，该申请由于数据收集便捷，可得到用于生成对多语言语音合成单音色需求的多语言语音信息的目标模型，还能应用于智慧政务、智慧安防等场景中，从而推动智慧城市的建设，提高用户使用体验度。

附图说明

为了更清楚地说明本申请实施例技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，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。

图 1 是本申请实施例提供的一种多语言语音模型生成方法的流程示意图；

图 1a 是本申请实施例提供的一种多语言语音模型生成方法的应用场景示意图；

图 2 是本申请实施例提供的一种多语言语音模型生成方法的子流程示意图；

图 3 是本申请实施例提供的一种多语言语音模型生成方法的子流程示意图；

图 4 是本申请实施例提供的一种多语言语音模型生成方法的子流程示意图；

图 5 是本申请实施例提供的一种多语言语音模型生成方法的子流程示意图；

图 6 是本申请实施例提供的一种多语言语音模型生成装置的示意性框图；

图 7 是本申请实施例提供的一种多语言语音模型生成装置的数据增强单元的示意性框图；

图 8 是本申请实施例提供的一种多语言语音模型生成装置的数据拼接单元的示意性框图；

图 9 是本申请实施例提供的一种多语言语音模型生成装置的标志插入单元的示意性框图；

图 10 是本申请实施例提供的一种多语言语音模型生成装置的语音处理单元的示意性框图；

图 11 是本申请实施例提供的一种计算机设备结构组成示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请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。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和 “包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解，在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使
用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

请参阅图 1 和图 1a，图 1 是本申请实施例提供的一种多语言语音模型生成方法的示意流程图，图 1a 是本申请实施例中多语言语音模型生成方法的场景示意图。该多语言语音模型生成方法应用于管理服务器 10 中。该管理服务器 10 能够通过对基于第一语言的第一主体 20 的第一语音信息的采集以及基于第二语言的第一主体 20 的第二语音信息的采集，实现对第一语音信息的数据增强，即得到混合有第二语音信息的目标语音信息，其中所有的目标语音信息构成一目标数据集；利用目标数据集训练一神经网络，以得到一中间模型；再次获取基于第一语言的第二主体 30 的第三语音信息，利用第三语音信息对中间模型再次进行训练，以获取第三语音信息的音色，从而得到一目标模型。以下将以处理服务器 10 的角度详细地介绍该多语言语音模型生成方法的各个步骤。

请参阅图 1，图 1 是本申请实施例提供的一种多语言语音模型生成方法的示意流程图。如图 1 所示，该方法的步骤包括步骤 S101~S105。

步骤S101，获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息。

在本实施例中，预先采集的基于第一语言的第一语音信息为同一个主体提供，其可以是一个完整的数据集，该数据集可以涵盖第一语言中的大部分语音数据的类型。预先采集的基于第二语言的第二语音信息也由上述主体提供，其可以包括第二语言中的基础语音数据。

例如，所述第一语言可以是汉语，所述第二语言可以是英语。作为优选地，在第一语音数据集中，汉语语音信息所占的比重远大于英语语信息据所占的比重。通常，第一语音信息作为一个较为完整的汉语语音数据集，可以涵盖人们生活中的大部分日常语言，从而保证后续训练得到的模型的质量。第二语音信息作为英语语音数据集，可以是上述主体提供的26个英文字母的读音，因为对于大量的英语语音信息，上述主体可能没有能力提供，而我们在日常生活中离不开英语字母，即车牌、楼门号等等计数或者号码牌都是需要字母的；通过26个英文字母的读音的获取，即可通过相应地处理得到更多的英语语音信息。

步骤S102，根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本。

在本实施例中，由于第一语音信息和第二语音信息是分开的，为了得到多语言的混合语音信息，此时需要预先设置一个数据增强脚本，该数据增强脚本用于随机组合所述第二语音信息，从而得到一个较长的语音信息，即得到若干增强语音样本。例如，当第二语音信息为英语字母时，可以随机抽取若干英语字母以组成多个英语单词，进而构成一个或多个简短的日常英语语句。

在另一实施例中，如图2所示，所述第二语音信息包括若干条第二语音数据，所述步骤102包括步骤S201~S202。

步骤S201，运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据；其中，每组中的第二语音数据的数量至少为两个。

在本实施例中，管理服务器可以调取预设数据增强脚本，并进行运行，从而可以从第二语音信息中随机抽取多组第二语音数据，为了确保拼接的相关语音信息更为合理，此时每组中的第二语音数据至少为两个，如可以是两个或者三个及以上的数量。

步骤S202，将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本。

在本实施例中，管理服务器需要将所抽取的每组第二语音数据进行拼接，从而得到多个相应的增强语音样本，例如，可以通过随机组合少量的英语语音数据，构成日常生活用语和句子等。

在进一步的实施例中，如图3所示，所述步骤S202包括步骤S301~S303。

步骤S301，将所抽取的每组第二语音数据进行拼接，以得到多个中间词组。

其中，管理服务器可以将所抽取的每组第二语音数据进行拼接，由于通常第二语音数据为第二语言中的基础发音和组成部分，故此时通过拼接，可以得到多个中间词组。

步骤S302，将所得到的中间词组进行组合，以得到多个中间语句。

其中，管理服务器还能够再次对所得到的中间词组进行组合，从而得到多个中间语句，

通过中间语句的组合，可以得到多个相关日常用语。

步骤 S303，将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本，其中，不同的预设静音标志关联有不同的静音时间段。

其中，管理服务器将所得到的中间语句进行拼接时，可以在拼接位置插入静音标志，通过设置静音标志，可以保证语句的整体韵律不出现问题，不同的静音标志关联有不同的静音时间段，静音时间段中通常没有语音内容，代表录音中的停顿。

作为可选地，所述预设静音标志为标点符号，所述标点符号包括逗号、顿号、句号、问号以及感叹号中的一种或多种。例如，拼接两段中间语句，通过设置标点符号来表示相应的静音，一般逗号和顿号 200-300 毫秒，句号、问号、感叹号停顿 400-500 毫秒等。

在一实施例中，如图 4 所示，若所述标点符号包括逗号和句号，所述逗号关联的静音时间段为 200~300 毫秒，所述句号关联的静音时间段为 400~500 毫秒，所述步骤 S303 包括步骤 S401~S403。

步骤 S401，将所得到的中间语句进行拼接以得到一拼接语句。步骤 S402，判断所述拼接语句的字符数是否超过第一预设值。步骤 S403，若所述拼接语句的字符数超过第一预设值，在拼接位置插入句号，以得到增强语音样本。所述方法还包括以下步骤，步骤 S404，若所述拼接语句的字符数没有超过第一预设值，在拼接位置插入逗号，以得到增强语音样本。

其中，管理服务器将所得到的中间语句进行拼接后，可以计算该拼接语句的字符数是否超过第一预设值，通常当字符数过多时，不便进行停顿，会导致语句的整体韵律出现问题，故可以再凭借语句中加入预设静音标志如句号或逗号，此时逗号所关联的停顿时间可以是 200~300 毫秒，句号所关联的停顿时间可以是 400~500 毫秒，即当字符数超过第一预设值时，此时需要使用句号进行停顿，若字符数没有超过第一预设值，此时则是使用逗号进行停顿。

步骤 S103，将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息。

在本实施例中，管理服务器能够将所述增强语音样本随机地插入第一语音信息中，从而得到多个包括有增强语音样本地目标语音信息。

例如，当第一语音信息为汉语语音信息时，若其中一条第一语音信息为“那儿有一辆车”，此时，其中一条增强语音样本为“red”时，可以将两者合并为目标语音信息“那儿有一辆 red 车”。

在另一实施例中，如图 5 所示，所述步骤 S103 可以包括步骤 S501~S502。

步骤 S501，获取所述增强语音样本。其中，管理服务器能够获取所述增强语音样本，以便于将增强语音样本插入到相应的第一语音信息中。

步骤 S502，将所获取的增强语音样本随机插入不同的第一语音信息中，以得到多个不同的目标语音信息。其中，管理服务器能够将所获取的增强语音样本随机插入到不同的第一语音信息中，从而得到多个不同的目标语音信息。通常插入增强语音样本的第一语音信息可以时包含多语言的语音信息。

步骤 S104，利用目标语音信息训练一神经网络，以得到一中间模型。

在本实施例中，管理服务器能够获取目标语音信息，从而训练一神经网络，通过训练后

的神经网络可以得到一中间模型，该神经网络可以采用循环神经网络（如：LSTM 等），训练的损失函数为均方差损失函数。

在一实施例中，所述目标语音信息为上述主体的读音信息，与该主体的说话的音色无关，以便于后续对相关模型的进一步的提升训练，从而得到更为完善的模型。

步骤 S105，获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

在本申请中，预设的第二语音数据集包括另一主体所提供的基于第一语言的第三语音信息，并通过获取所述第三语音信息中的音色来实现对中间模型的进一步的训练。管理服务器获取第三语音信息的音色，能够使得中间模型在再次训练之后得到一个可生成符合用户的对多语言语音合成需求的多语言语音信息的目标模型。

综上，本申请可以不仅数据收集更为便捷，还能生成符合用户的对多语言语音合成单音色需求的多语言语音信息的目标模型，提高了用户的使用体验度。

本领域普通技术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，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory，ROM）等。

请参阅图 6，对应上述一种多语言语音模型生成方法，本申请实施例还提出一种多语言语音模型生成装置，该装置 100 包括：数据获取单元 101、数据增强单元 102、语音处理单元 103、第一训练单元 104 以及第二训练单元 105。

所述数据获取单元 101，用于获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息。

在本实施例中，预先采集的基于第一语言的第一语音信息为同一个主体提供，其可以是一个完整的数据集，该数据集可以涵盖第一语言中的大部分语音数据的类型。预先采集的基于第二语言的第二语音信息也由上述主体提供，其可以包括第二语言中的基础语音数据。

例如，所述第一语言可以是汉语，所述第二语言可以是英语。作为优选地，在第一语音数据集中，汉语语音信息所占的比重远大于英语语信息据所占的比重。通常，第一语音信息作为一个较为完整的汉语语音数据集，可以涵盖人们生活中的大部分日常语言，从而保证后续训练得到的模型的质量。第二语音信息作为英语语音数据集，可以是上述主体提供的 26 个英文字母的读音，因为对于大量的英语语音信息，上述主体可能没有能力提供，而我们在日常生活中离不开英语字母，即车牌、楼门号等等计数或者号码牌都是需要字母的；通过 26 个英文字母的读音的获取，即可通过相应地处理得到更多的英语语音信息。

所述数据增强单元 102，用于根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本。

在本实施例中，由于第一语音信息和第二语音信息是分开的，为了得到多语言的混合语音信息，此时需要预先设置一个数据增强脚本，该数据增强脚本用于随机组合所述第二语音

信息，从而得到一个较长的语音信息，即得到若干增强语音样本。例如，当第二语音信息为英语字母时，可以随机抽取若干英语字母以组成多个英语单词，进而构成一个或多个简短的日常英语语句。

在另一实施例中，如图 7 所示，所述第二语音信息包括若干条第二语音数据，所述数据增强单元 102 包括脚本运行单元 201 以及数据拼接单元 202。

所述脚本运行单元 201，用于运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据；其中，每组中的第二语音数据的数量至少为两个。

在本实施例中，管理服务器可以调取预设数据增强脚本，并进行运行，从而可以从第二语音信息中随机抽取多组第二语音数据，为了确保拼接的相关语音信息更为合理，此时每组中的第二语音数据至少为两个，如可以是两个或者三个及以上的数量。

所述数据拼接单元 202，用于将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本。

在本实施例中，管理服务器需要将所抽取的每组第二语音数据进行拼接，从而得到多个相应的增强语音样本，例如，可以通过随机组合少量的英语语音数据，构成日常生活用语和句子等。

在进一步的实施例中，如图 8 所示，所述所述数据拼接单元 202 包括数据抽取单元 301、词组组合单元 302 以及标志插入单元 303。

所述数据抽取单元 301，用于将所抽取的每组第二语音数据进行拼接，以得到多个中间词组。

其中，管理服务器可以将所抽取的每组第二语音数据进行拼接，由于通常第二语音数据为第二语言中的基础发音和组成部分，故此时通过拼接，可以得到多个中间词组。

所述词组组合单元 302，用于将所得到的中间词组进行组合，以得到多个中间语句。

其中，管理服务器还能够再次对所得到的中间词组进行组合，从而得到多个中间语句，通过中间语句的组合，可以得到多个相关日常用语。

所述标志插入单元 303，用于将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本，其中，不同的预设静音标志关联有不同的静音时间段。

其中，管理服务器将所得到的中间语句进行拼接时，可以在拼接位置插入静音标志，通过设置静音标志，可以保证语句的整体韵律不出现问题，不同的静音标志关联有不同的静音时间段，静音时间段中通常没有语音内容，代表录音中的停顿。

作为可选地，所述预设静音标志为标点符号，所述标点符号包括逗号、顿号、句号、问号以及感叹号中的一种或多种。例如，拼接两段中间语句，通过设置标点符号来表示相应的静音，一般逗号和顿号 200-300 毫秒，句号、问号、感叹号停顿 400-500 毫秒等。

在一实施例中，如图 9 所示，若所述标点符号包括逗号和句号，所述逗号关联的静音时间段为 200~300 毫秒，所述句号关联的静音时间段为 400~500 毫秒，所述标志插入单元 303 包括语句拼接单元 401、数值判断单元 402 以及第一插入单元 403。

所述语句拼接单元 401，用于将所得到的中间语句进行拼接以得到一拼接语句。所述数

值判断单元 402，用于判断所述拼接语句的字符数是否超过第一预设值。所述第一插入单元 403，用于若所述拼接语句的字符数超过第一预设值，在拼接位置插入句号，以得到增强语音样本。所述装置 100 还包括第二插入单元 404，用于若所述拼接语句的字符数没有超过第一预设值，在拼接位置插入逗号，以得到增强语音样本。

其中，管理服务器将所得到的中间语句进行拼接后，可以计算该拼接语句的字符数是否超过第一预设值，通常当字符数过多时，不便进行停顿，会导致语句的整体韵律出现问题，故可以再凭借语句中加入预设静音标志如句号或逗号，此时逗号所关联的停顿时间可以是 200~300 毫秒，句号所关联的停顿时间可以是 400~500 毫秒，即当字符数超过第一预设值时，此时需要使用句号进行停顿，若字符数没有超过第一预设值，此时则是使用逗号进行停顿。

所述语音处理单元 103，用于将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息。

在本实施例中，管理服务器能够将所述增强语音样本随机地插入第一语音信息中，从而得到多个包括有增强语音样本地目标语音信息。

例如，当第一语音信息为汉语语音信息时，若其中一条第一语音信息为“那儿有一辆车”，此时，其中一条增强语音样本为“red”时，可以将两者合并为目标语音信息“那儿有一辆 red 车”。

在另一实施例中，如图 10 所示，所述语音处理单元 103 可以包括样本获取单元 501 以及样本插入单元 502。

所述样本获取单元 501，用于获取所述增强语音样本。其中，管理服务器能够获取所述增强语音样本，以便于将增强语音样本插入到相应的第一语音信息中。

所述样本插入单元 502，用于将所获取的增强语音样本随机插入不同的第一语音信息中，以得到多个不同的目标语音信息。其中，管理服务器能够将所获取的增强语音样本随机插入到不同的第一语音信息中，从而得到多个不同的目标语音信息。通常插入增强语音样本的第一语音信息可以时包含多语言的语音信息。

所述第一训练单元 104，用于利用目标语音信息训练一神经网络，以得到一中间模型。

在本实施例中，管理服务器能够获取目标语音信息，从而训练一神经网络，通过训练后的神经网络可以得到一中间模型，该神经网络可以采用循环神经网络（如：LSTM 等），训练的损失函数为均方差损失函数。

在一实施例中，所述目标语音信息为上述主体的读音信息，与该主体的说话的音色无关，以便于后续对相关模型的进一步的提升训练，从而得到更为完善的模型。

所述第二训练单元 105，用于调用预设的第二语音数据集中的基于第一语言的第三语音信息，以获取所述第三语音信息中的音色对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

在本申请中，预设的第二语音数据集包括另一主体所提供的基于第一语言的第三语音信息，并通过获取所述第三语音信息中的音色来实现对中间模型的进一步的训练。管理服务器获取第三语音信息的音色，能够使得中间模型在再次训练之后得到一个可生成符合用户的对

多语言语音合成需求的多语言语音信息的目标模型。

需要说明的是，所属领域的技术人员可以清楚地了解到，上述多语言语音模型生成装置100和各单元的具体实现过程，可以参考前述方法实施例中的相应描述，为了描述的方便和简洁，在此不再赘述。

由以上可见，在硬件实现上，以上数据获取单元101、数据增强单元102、语音处理单元103、第一训练单元104以及第二训练单元105等可以以硬件形式内嵌于或独立于多语言语音模型生成装置中，也可以以软件形式存储于多语言语音模型生成装置的存储器中，以便处理器调用执行以上各个单元对应的操作。该处理器可以为中央处理单元（CPU）、微处理器、单片机等。

上述多语言语音模型生成装置可以实现为一种计算机程序的形式，计算机程序可以在如图11所示的计算机设备上运行。

图11为本申请一种计算机设备的结构组成示意图。该设备可以是服务器，其中，服务器可以是独立的服务器，也可以是多个服务器组成的服务器集群。

参照图11，该计算机设备600包括通过系统总线601连接的处理器602、存储器、内存储器604和网络接口605，其中，存储器可以包括非易失性存储介质603和内存储器604。

该非易失性存储介质603可存储操作系统6031和计算机程序6032，该计算机程序6032被执行时，可使得处理器602执行一种多语言语音模型生成方法。

该处理器602用于提供计算和控制能力，支撑整个计算机设备600的运行。

该内存储器604为非易失性存储介质603中的计算机程序6032的运行提供环境，该计算机程序6032被处理器602执行时，可使得处理器602执行一种多语言语音模型生成方法。

该网络接口605用于与其它设备进行网络通信。本领域技术人员可以理解，图11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600的限定，具体的计算机设备600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。

其中，所述处理器602用于运行存储在存储器中的计算机程序6032，以实现如上述实施例中的多语言语音模型生成方法中的步骤。应当理解，在本申请实施例中，处理器602可以是中央处理单元（Central Processing Unit，CPU），该处理器602还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中，通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

本领域普通技术人员可以理解的是实现上述实施例的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。该计算机程序可存储于一存储介质中，该存储介质为计算机可读存储介质。该计算机程序被该计算机系统中的至少一个处理器执行，以实现上述方法的实施例的流程步骤。

因此，本申请还提供一种存储介质。该存储介质可以为计算机可读存储介质，所述计算

机可读存储介质可以是非易失性，也可以是易失性。该存储介质存储有计算机程序，该计算机程序被处理器执行时使处理器执行如上述实施例中的多语言语音模型生成方法中的步骤。

所述存储介质为实体的、非瞬时性的存储介质，例如可以是U盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、磁碟或者光盘等各种可以存储程序代码的实体存储介质。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，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，各个单元的划分，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。

本申请实施例方法中的步骤可以根据实际需要进行顺序调整、合并和删减。本申请实施例装置中的单元可以根据实际需要进行合并、划分和删减。另外，在本申请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，也可以是两个或两个以上单元集成在一个单元中。

该集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分，或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品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设备（可以是个人计算机，终端，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1.一种多语言语音模型生成方法，包括：

获取预先采集的第一语音数据集，该语音数据集包括均为同一主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

2.如权利要求 1 所述的方法，其中，所述第二语音信息包括若干条第二语音数据，所述根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本的步骤，包括：

运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据；其中，每组中的第二语音数据的数量至少为两个；

将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本。

3.如权利要求 2 所述的方法，其中，所述将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本的步骤，包括：

将所抽取的每组第二语音数据进行拼接，以得到多个中间词组；

将所得到的中间词组进行组合，以得到多个中间语句；

将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本，其中，不同的预设静音标志关联有不同的静音时间段。

4.如权利要求 3 所述的方法，其中，所述预设静音标志为标点符号，所述标点符号包括逗号、顿号、句号、问号以及感叹号中的一种或多种。

5.如权利要求 4 所述的方法，其中，若所述标点符号包括逗号和句号，所述逗号关联的静音时间段为 200~300 毫秒，所述句号关联的静音时间段为 400~500 毫秒，所述将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本的步骤，包括：

将所得到的中间语句进行拼接以得到一拼接语句；

判断所述拼接语句的字符数是否超过第一预设值；

若所述拼接语句的字符数超过第一预设值，在拼接位置插入句号，以得到增强语音样本。

6.如权利要求 5 所述的方法，其中，所述方法还包括：

若所述拼接语句的字符数没有超过第一预设值，在拼接位置插入逗号，以得到增强语音样本。

7.如权利要求 1 所述的方法，其中，所述将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息的步骤，包括：

获取所述增强语音样本；

将所获取的增强语音样本随机插入不同的第一语音信息中，以得到多个不同的目标语音

信息。

8.一种多语言语音模型生成装置，包括：

数据获取单元，用于获取预先采集的第一语音数据集，该语音数据集包括均为同意主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

数据增强单元，用于根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

语音处理单元，用于将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

第一训练单元，用于利用目标语音信息训练一神经网络，以得到一中间模型；

第二训练单元，用于调用预设的第二语音数据集中的基于第一语言的第三语音信息，以获取所述第三语音信息中的音色对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

9.一种计算机设备，其中，所述计算机设备包括存储器及处理器，所述存储器上存储有计算机程序，所述处理器执行所述计算机程序时实现如下步骤：

获取预先采集的第一语音数据集，该语音数据集包括均为同一主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

10.如权利要求 9 所述的计算机设备，其中，所述第二语音信息包括若干条第二语音数据，所述根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本的步骤，包括：

运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据；其中，每组中的第二语音数据的数量至少为两个；

将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本。

11.如权利要求 10 所述的计算机设备，其中，所述将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本的步骤，包括：

将所抽取的每组第二语音数据进行拼接，以得到多个中间词组；

将所得到的中间词组进行组合，以得到多个中间语句；

将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本，其中，不同的预设静音标志关联有不同的静音时间段。

12.如权利要求 11 所述的计算机设备，其中，所述预设静音标志为标点符号，所述标点符号包括逗号、顿号、句号、问号以及感叹号中的一种或多种。

13.如权利要求 12 所述的计算机设备，其中，若所述标点符号包括逗号和句号，所述逗

号关联的静音时间段为 200~300 毫秒，所述句号关联的静音时间段为 400~500 毫秒，所述将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本的步骤，包括：

将所得到的中间语句进行拼接以得到一拼接语句；

判断所述拼接语句的字符数是否超过第一预设值；

若所述拼接语句的字符数超过第一预设值，在拼接位置插入句号，以得到增强语音样本。

14.如权利要求 13 所述的计算机设备，其中，所述处理器执行所述计算机程序时还实现如下步骤：

若所述拼接语句的字符数没有超过第一预设值，在拼接位置插入逗号，以得到增强语音样本。

15.如权利要求 9 所述的计算机设备，其中，所述将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息的步骤，包括：

获取所述增强语音样本；

将所获取的增强语音样本随机插入不同的第一语音信息中，以得到多个不同的目标语音信息。

16. 一种计算机可读存储介质，其中，所述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所述处理器执行如下步骤：

获取预先采集的第一语音数据集，该语音数据集包括均为同一主体提供的基于第一语言的第一语音信息以及基于第二语言的第二语音信息；

根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本；

将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息；

利用目标语音信息训练一神经网络，以得到一中间模型；

获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型。

17.如权利要求 16 所述的计算机可读存储介质，其中，所述第二语音信息包括若干条第二语音数据，所述根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本的步骤，包括：

运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据；其中，每组中的第二语音数据的数量至少为两个；

将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本。

18.如权利要求 17 所述的计算机可读存储介质，其中，所述将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本的步骤，包括：

将所抽取的每组第二语音数据进行拼接，以得到多个中间词组；

将所得到的中间词组进行组合，以得到多个中间语句；

将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本，其中，不同的预设静音标志关联有不同的静音时间段。

19.如权利要求 18 所述的计算机可读存储介质，其中，所述预设静音标志为标点符号，所述标点符号包括逗号、顿号、句号、问号以及感叹号中的一种或多种。

20.如权利要求 19 所述的计算机可读存储介质，其中，若所述标点符号包括逗号和句号，所述逗号关联的静音时间段为 200~300 毫秒，所述句号关联的静音时间段为 400~500 毫秒，所述将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本的步骤，包括：

将所得到的中间语句进行拼接以得到一拼接语句；

判断所述拼接语句的字符数是否超过第一预设值；

若所述拼接语句的字符数超过第一预设值，在拼接位置插入句号，以得到增强语音样本。

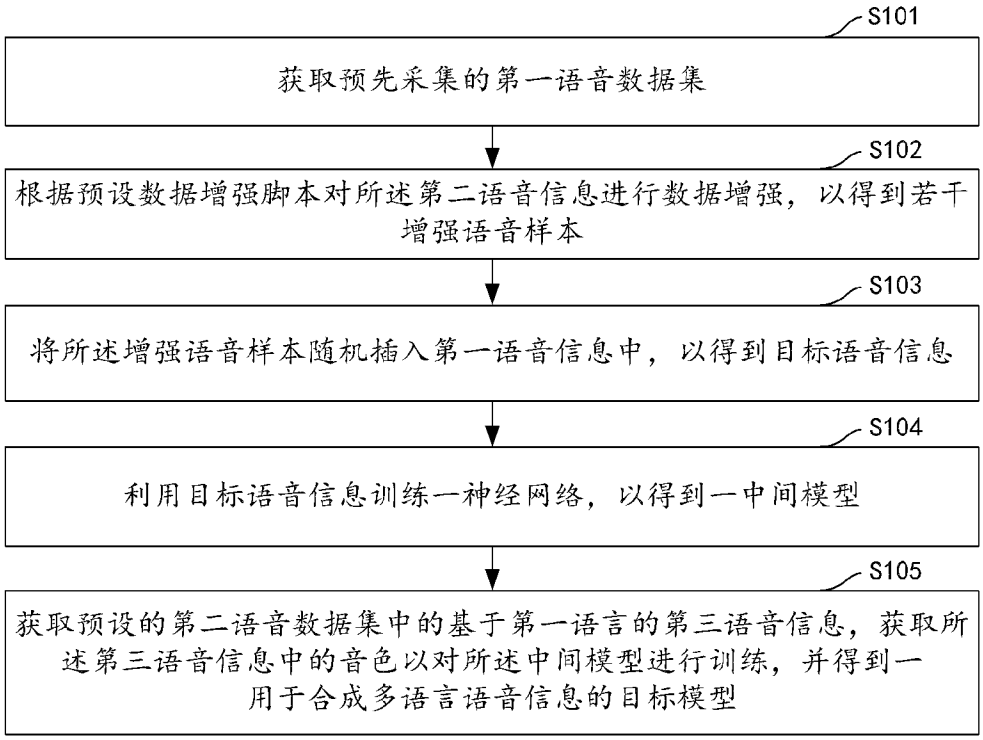

S101
获取预先采集的第一语音数据集
S102
根据预设数据增强脚本对所述第二语音信息进行数据增强，以得到若干增强语音样本
S103
将所述增强语音样本随机插入第一语音信息中，以得到目标语音信息
S104
利用目标语音信息训练一神经网络，以得到一中间模型
S105
获取预设的第二语音数据集中的基于第一语言的第三语音信息，获取所述第三语音信息中的音色以对所述中间模型进行训练，并得到一用于合成多语言语音信息的目标模型

图 1

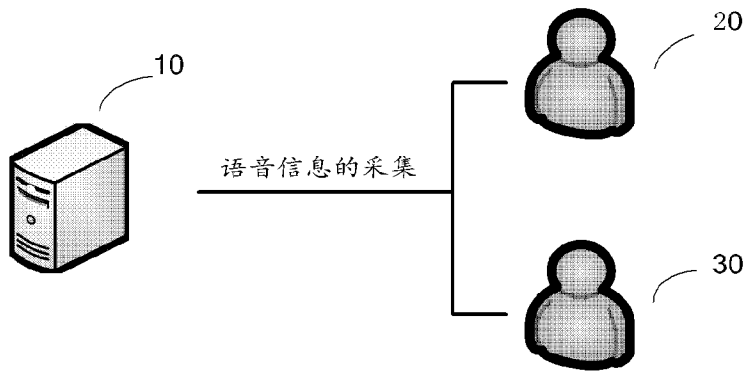

10
20
语音信息的采集
30

图 1a

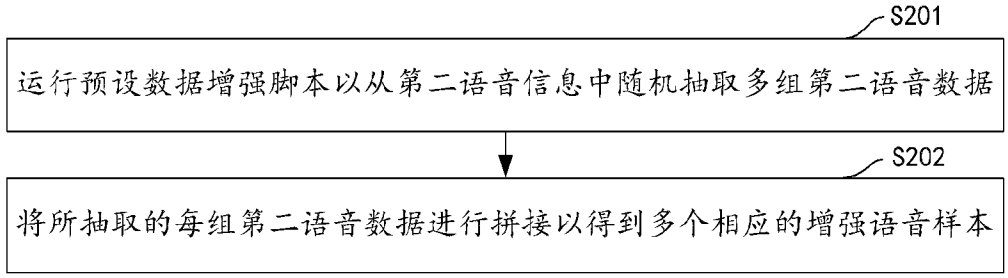

S201
运行预设数据增强脚本以从第二语音信息中随机抽取多组第二语音数据
S202
将所抽取的每组第二语音数据进行拼接以得到多个相应的增强语音样本

图 2

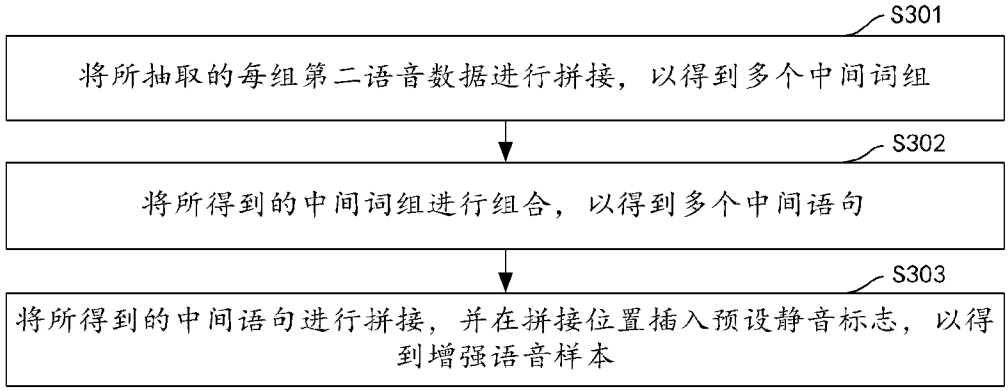

S301
将所抽取的每组第二语音数据进行拼接，以得到多个中间词组
S302
将所得到的中间词组进行组合，以得到多个中间语句
S303
将所得到的中间语句进行拼接，并在拼接位置插入预设静音标志，以得到增强语音样本

图 3

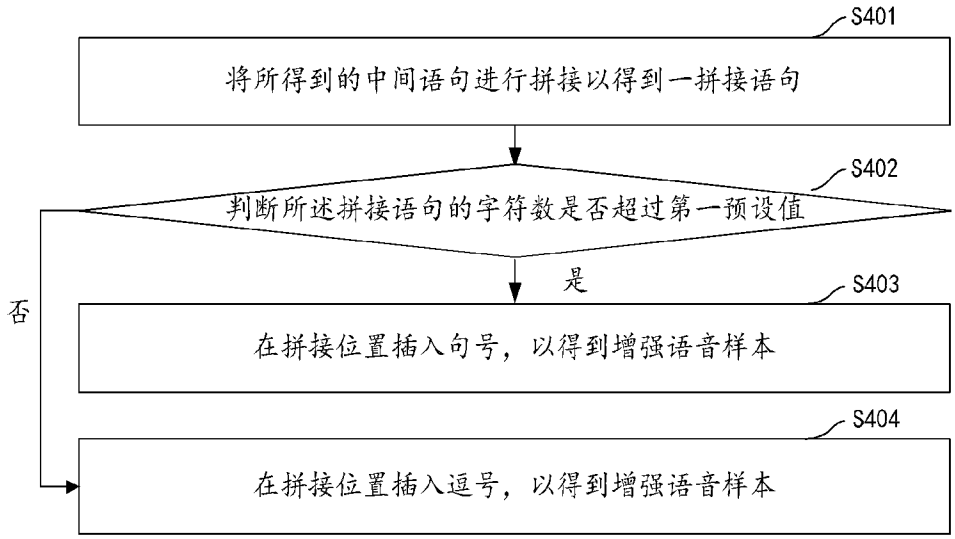

S401
将所得到的中间语句进行拼接以得到一拼接语句
S402
判断所述拼接语句的字符数是否超过第一预设值
是
否
S403
在拼接位置插入句号，以得到增强语音样本
S404
在拼接位置插入逗号，以得到增强语音样本

图 4

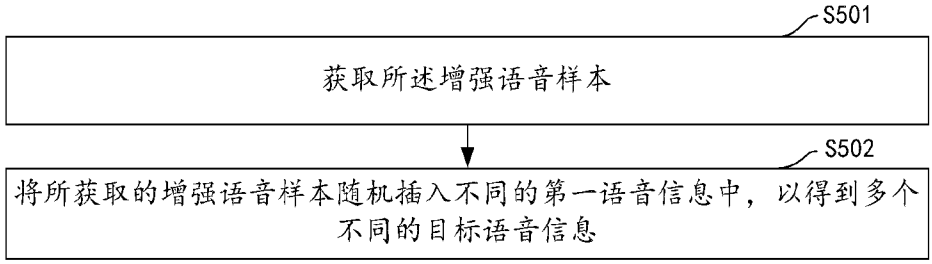

S501
获取所述增强语音样本
S502
将所获取的增强语音样本随机插入不同的第一语音信息中，以得到多个不同的目标语音信息

图 5

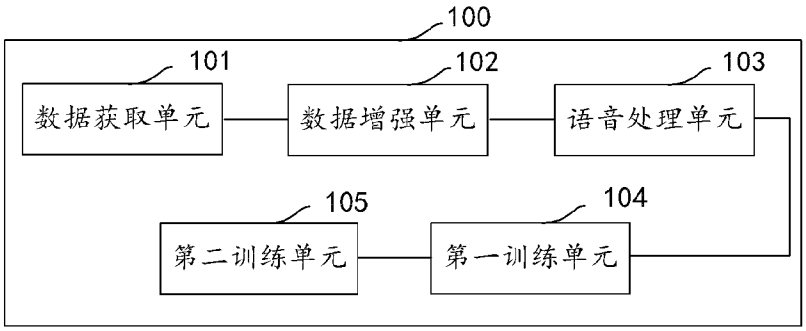

100
101
数据获取单元
102
数据增强单元
103
语音处理单元
105
第二训练单元
104
第一训练单元

图 6

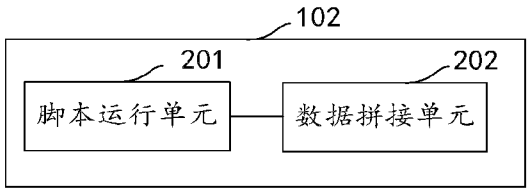

102
201
脚本运行单元
202
数据拼接单元

图 7

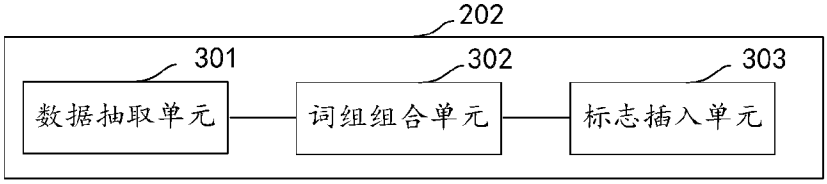

202
301
数据抽取单元
302
词组组合单元
303
标志插入单元

图 8

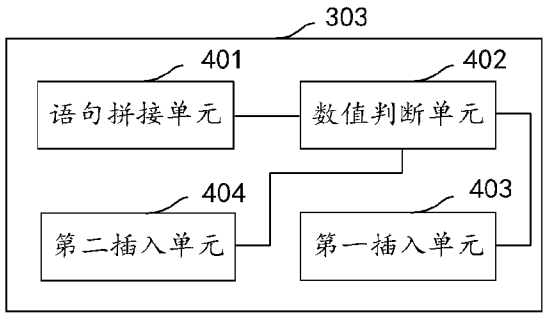

303
401
语句拼接单元
402
数值判断单元
404
第二插入单元
403
第一插入单元

图 9

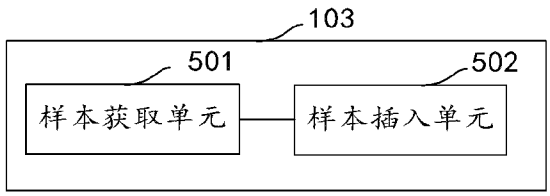

103
501
样本获取单元
502
样本插入单元

图 10

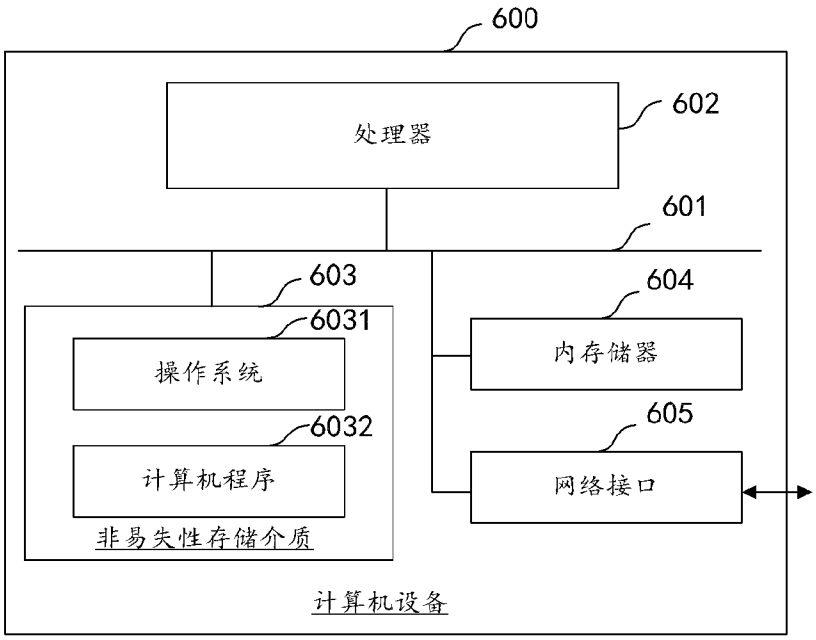
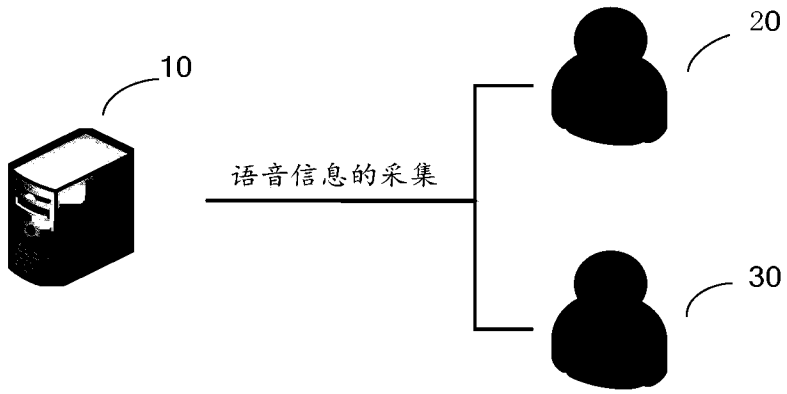

600
处理器
602
601
603
6031
操作系统
6032
计算机程序
非易失性存储介质
604
内存储器
605
网络接口
计算机设备

图 11


10
20
语音信息的采集
30

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/096668

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/047(2013.01)i; G10L 21/02(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)

CNPAT, CNKI, EPODOC, WPI, USTXT, EPTXT, WOTXT: 多语种, 多语言, 混合语种, 混合语言, 多方言, 混合方言, 多种语种, 多种语言, 两种语言, 两种语种, 双语, 增强, 音色, 语音, 模型, 建模, multilingual, multi+, two, mix+, hybrid+, language?, model+, enhanc+, sound, speech, voice, wave, quality, tone

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110827805 A (AI SPEECH LTD.) 21 February 2020 (2020-02-21) description, paragraphs [0053]-[0097], and figures 1-6	1-20
A	CN 108711420 A (BEIJING ORION STAR TECHNOLOGY CO., LTD.) 26 October 2018 (2018-10-26) entire document	1-20
A	CN 111816169 A (AI SPEECH LTD.) 23 October 2020 (2020-10-23) entire document	1-20
A	CN 112397051 A (WUHAN TCL GROUP INDUSTRIAL RESEARCH INSTITUTE CO., LTD.) 23 February 2021 (2021-02-23) entire document	1-20
A	CN 109616096 A (INTELLIGENT STEWARD CO., LTD.) 12 April 2019 (2019-04-12) entire document	1-20
A	CN 107481713 A (TSINGHUA UNIVERSITY et al.) 15 December 2017 (2017-12-15) entire document	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 November 2021	**19 November 2021**
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.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105845125 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 10 August 2016 (2016-08-10) entire document	1-20
A	CN 108831481 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 16 November 2018 (2018-11-16) entire document	1-20
A	US 2020160836 A1 (GOOGLE L.L.C.) 21 May 2020 (2020-05-21) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2021/096668

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110827805	A	21 February 2020	None			
CN	108711420	A	26 October 2018	CN	108711420	B	09 July 2021
CN	111816169	A	23 October 2020	None			
CN	112397051	A	23 February 2021	None			
CN	109616096	A	12 April 2019	None			
CN	107481713	A	15 December 2017	CN	107481713	B	02 June 2020
CN	105845125	A	10 August 2016	WO	2017197809	A1	23 November 2017
				US	2019213995	A1	11 July 2019
				US	10789938	B2	29 September 2020
				CN	105845125	B	03 May 2019
CN	108831481	A	16 November 2018	WO	2020024352	A1	06 February 2020
US	2020160836	A1	21 May 2020	None			

国际检索报告

国际申请号
PCT/CN2021/096668

A. 主题的分类

G10L 13/047(2013.01)i; G10L 21/02(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））

CNPAT,CNKI,EPODOC,WPI,USTXT,EPTXT,WOTXT: 多语种，多语言，混合语种，混合语言，多方言，混合方言，多种语种，多种语言，两种语言，两种语种，双语，增强，音色，语音，模型，建模，multilingual, multi+, two, mix+, hybrid+, language?, model+, enhanc+, sound, speech, voice, wave, quality, tone

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 110827805 A (苏州思必驰信息科技有限公司) 2020年 2月 21日 (2020-02-21) 说明书第[0053]-[0097]段，图1-6	1-20
A	CN 108711420 A (北京猎户星空科技有限公司) 2018年 10月 26日 (2018-10-26) 全文	1-20
A	CN 111816169 A (苏州思必驰信息科技有限公司) 2020年 10月 23日 (2020-10-23) 全文	1-20
A	CN 112397051 A (武汉TCL集团工业研究院有限公司) 2021年 2月 23日 (2021-02-23) 全文	1-20
A	CN 109616096 A (北京智能管家科技有限公司) 2019年 4月 12日 (2019-04-12) 全文	1-20
A	CN 107481713 A (清华大学 等) 2017年 12月 15日 (2017-12-15) 全文	1-20
A	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) 全文	1-20

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2021年 11月 3日	2021年 11月 19日

ISA/CN的名称和邮寄地址	受权官员
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	耿娜
传真号 (86-10)62019451	电话号码 86-(10)-53962385

国际检索报告

国际申请号
PCT/CN2021/096668

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 108831481 A (平安科技深圳有限公司) 2018年 11月 16日 (2018-11-16) 全文	1-20
A	US 2020160836 A1 (GOOGLE L.L.C.) 2020年 5月 21日 (2020-05-21) 全文	1-20

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2021/096668

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	110827805	A	2020年 2月 21日	无			
CN	108711420	A	2018年 10月 26日	CN	108711420	B	2021年 7月 9日
CN	111816169	A	2020年 10月 23日	无			
CN	112397051	A	2021年 2月 23日	无			
CN	109616096	A	2019年 4月 12日	无			
CN	107481713	A	2017年 12月 15日	CN	107481713	B	2020年 6月 2日
CN	105845125	A	2016年 8月 10日	WO	2017197809	A1	2017年 11月 23日
				US	2019213995	A1	2019年 7月 11日
				US	10789938	B2	2020年 9月 29日
				CN	105845125	B	2019年 5月 3日
CN	108831481	A	2018年 11月 16日	WO	2020024352	A1	2020年 2月 6日
US	2020160836	A1	2020年 5月 21日	无			