(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110428811 B
(45)授权公告日 2021.09.07

(21)申请号 201910876122.X

(22)申请日 2019.09.17

(65)同一申请的已公布的文献号
申请公布号 CN 110428811 A

(43)申请公布日 2019.11.08

(73)专利权人 北京声智科技有限公司
地址 100080 北京市海淀区北四环西路67号3层306室

(72)发明人 陈孝良 靳源 冯大航 常乐

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 赵焕

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/22(2006.01)
G10L 15/26(2006.01)

(56)对比文件
CN 105723362 A,2016.06.29
CN 102376182 A,2012.03.14
CN 109753665 A,2019.05.14
CN 107871506 A,2018.04.03
CN 106971743 A,2017.07.21
US 7209882 B1,2007.04.24
CN 107657947 A,2018.02.02
郭慧阳.基于深度学习的说话人识别技术研究.《中国优秀硕士学位论文全文数据库(电子期刊)》.2019,
Josef Chaloupka et al..Phone Speech Detection and Recognition in the Task of Historical Radio Broadcast Transcription.《IEEE》.2015,

审查员 宋玉香

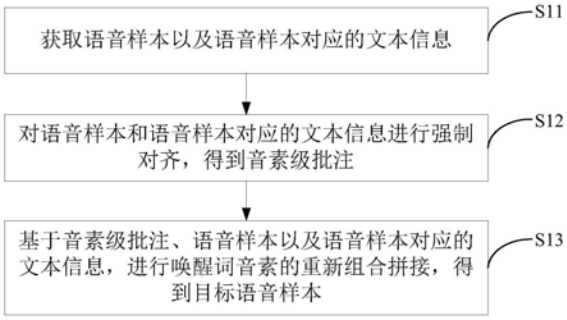
权利要求书3页 说明书10页 附图2页

(54)发明名称

一种数据处理方法、装置及电子设备

(57)摘要

本发明提供了一种数据处理方法、装置及电子设备，获取语音样本以及所述语音样本对应的文本信息；对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。通过本发明中对语音样本的唤醒词音素的重组，就可以得到来自不同语音样本的唤醒词音素的组合，这样，就可以得到数量较多的语音样本。



CN 110428811 B

1.一种数据处理方法,其特征在于,包括:

获取语音样本以及所述语音样本对应的文本信息;所述语音样本为多个;

对所述语音样本和所述语音样本对应的文本信息进行强制对齐,得到音素级批注;

基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息,进行唤醒词音素的重新组合拼接,得到目标语音样本;

所述基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息,进行唤醒词音素的重新组合拼接,得到目标语音样本,包括:

基于所述语音样本对应的文本信息,筛选出包括唤醒词的语音样本,并作为正例样本;

从所述正例样本中截取出所述唤醒词的唤醒词音素,并作为音素语音单元;

依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素,并按照所述前后连接顺序进行拼接重组,得到所述目标语音样本。

2.根据权利要求1所述的数据处理方法,其特征在于,所述对所述语音样本和所述语音样本对应的文本信息进行强制对齐,得到音素级批注,包括:

提取所述语音样本的音频特征;

基于所述语音样本对应的文本信息以及音频特征,对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐,得到所述音素级批注。

3.根据权利要求1所述的数据处理方法,其特征在于,所述依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素,包括:

筛选出满足预设条件的音素语音单元,并作为目标音素语音单元;所述预设条件包括同一用户或具有相同属性值的用户;所述属性值包括年龄段和性别;

依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

4.根据权利要求1或3所述的数据处理方法,其特征在于,在按照所述前后连接顺序进行拼接重组,得到所述目标语音样本之后,还包括:

对所述目标语音样本的韵律特征进行调整。

5.根据权利要求1所述的数据处理方法,其特征在于,在按照所述前后连接顺序进行拼接重组,得到所述目标语音样本之后,还包括:

将所述语音样本中的非正例样本作为反例样本;

将所述正例样本的唤醒词音素和反例样本的音素进行拆分,分别得到正例音素组和反例音素组;所述正例音素组和所述反例音素组中均包括预设数量且是连续的多个音素。

6.根据权利要求5所述的数据处理方法,其特征在于,依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素,并按照所述前后连接顺序进行拼接重组,得到所述目标语音样本,包括:

依据所述唤醒词的唤醒词音素之间的前后连接顺序,从目标数据中依次随机选取出所述唤醒词的每一个唤醒词音素,并按照所述前后连接顺序进行拼接重组,得到所述目标语音样本;所述目标数据包括所述音素语音单元和所述反例样本中的正例音素组。

7.根据权利要求6所述的数据处理方法,其特征在于,还包括:

将所述正例样本、所述目标语音样本、以及将反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本按照预设比例进行混合，并训练得到语音唤醒模型；

基于所述语音唤醒模型的调整原则，调整所述正例样本、所述目标语音样本以及所述新的反例样本的比例，并重新训练所述语音唤醒模型。

8.一种数据处理装置，其特征在于，包括：

样本获取模块，用于获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

对齐模块，用于对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

样本处理模块，用于基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本；

所述样本处理模块包括：

样本筛选子模块，用于基于所述语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本；

截取子模块，用于从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

样本拼接子模块，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

9.根据权利要求8所述的数据处理装置，其特征在于，所述对齐模块包括：

特征提取子模块，用于提取所述语音样本的音频特征；

对齐子模块，用于基于所述语音样本对应的文本信息以及音频特征，对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐，得到所述音素级批注。

10.根据权利要求8所述的数据处理装置，其特征在于，所述样本拼接子模块包括：

筛选单元，用于筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件包括同一用户或具有相同属性值的用户；所述属性值包括年龄段和性别；

选取单元，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

11.一种电子设备，其特征在于，包括：存储器和处理器；

其中，所述存储器用于存储程序；

处理器调用程序并用于：

获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本；

所述基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本，包括：

基于所述语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本；

从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

一种数据处理方法、装置及电子设备

技术领域

[0001] 本发明涉及语音数据的处理领域，更具体的说，涉及一种数据处理方法、装置及电子设备。

背景技术

[0002] 语音唤醒技术框架主要分为端到端与基于隐马尔可夫模型两种，这两种唤醒系统均需要建立一个性能较优的声学模型。

[0003] 在建立声学模型之前，需要获取较多的训练数据，如何得到较多的训练数据是本领域技术人员亟需解决的技术问题。

发明内容

[0004] 有鉴于此，本发明提供一种数据处理方法、装置及电子设备，以解决亟需获取较多的用于训练得到声学模型的训练数据的问题。

[0005] 为解决上述技术问题，本发明采用了如下技术方案：

[0006] 一种数据处理方法，包括：

[0007] 获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

[0008] 对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

[0009] 基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。

[0010] 优选地，所述对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注，包括：

[0011] 提取所述语音样本的音频特征；

[0012] 基于所述语音样本对应的文本信息以及音频特征，对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐，得到所述音素级批注。

[0013] 优选地，基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本，包括：

[0014] 基于所述语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本；

[0015] 从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

[0016] 依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

[0017] 优选地，所述依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，包括：

[0018] 筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件

包括同一用户或具有相同属性值的用户;所述属性值包括年龄段和性别;

[0019] 依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

[0020] 优选地,在按照所述前后连接顺序进行拼接重组,得到所述目标语音样本之后,还包括:

[0021] 对所述目标语音样本的韵律特征进行调整。

[0022] 优选地,在按照所述前后连接顺序进行拼接重组,得到所述目标语音样本之后,还包括:

[0023] 将所述语音样本中的非正例样本作为反例样本;

[0024] 将所述正例样本的唤醒词音素和反例样本的音素进行拆分,分别得到正例音素组和反例音素组;所述正例音素组和所述反例音素组中均包括预设数量且是连续的多个音素;

[0025] 将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据。

[0026] 优选地,依据所述唤醒词的唤醒词音素之间的前后连接顺序,从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素,并按照所述前后连接顺序进行拼接重组,得到所述目标语音样本,包括:

[0027] 依据所述唤醒词的唤醒词音素之间的前后连接顺序,从目标数据中依次随机选取出所述唤醒词的每一个唤醒词音素,并按照所述前后连接顺序进行拼接重组,得到所述目标语音样本;所述目标数据包括所述音素语音单元和所述反例音素组中的正例音素组。

[0028] 优选地,还包括:

[0029] 将所述目标样本、所述目标语音样本、以及将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本按照预设比例进行混合,并训练得到语音唤醒模型;

[0030] 基于所述语音唤醒模型的调整原则,调整所述目标样本、所述目标语音样本以及所述新的反例样本的比例,并重新训练所述语音唤醒模型。

[0031] 一种数据处理装置,包括:

[0032] 样本获取模块,用于获取语音样本以及所述语音样本对应的文本信息;所述语音样本为多个;

[0033] 对齐模块,用于对所述语音样本和所述语音样本对应的文本信息进行强制对齐,得到音素级批注;

[0034] 样本处理模块,用于基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息,进行唤醒词音素的重新组合拼接,得到目标语音样本。

[0035] 优选地,所述对齐模块包括:

[0036] 特征提取子模块,用于提取所述语音样本的音频特征;

[0037] 对齐子模块,用于基于所述语音样本对应的文本信息以及音频特征,对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐,得到所述音素级批注。

[0038] 优选地,所述样本处理模块包括:

[0039] 样本筛选子模块,用于基于所述语音样本对应的文本信息,筛选出包括唤醒词的语音样本,并作为正例样本;

[0040] 截取子模块，用于从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

[0041] 样本拼接子模块，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

[0042] 优选地，所述样本拼接子模块包括：

[0043] 筛选单元，用于筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件包括同一用户或具有相同属性值的用户；所述属性值包括年龄段和性别；

[0044] 选取单元，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

[0045] 一种电子设备，包括：存储器和处理器；

[0046] 其中，所述存储器用于存储程序；

[0047] 处理器调用程序并用于：

[0048] 获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

[0049] 对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

[0050] 基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。

[0051] 相较于现有技术，本发明具有以下有益效果：

[0052] 本发明提供了一种数据处理方法、装置及电子设备，获取语音样本以及所述语音样本对应的文本信息；对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。通过本发明中对语音样本的唤醒词音素的重组，就可以得到来自不同语音样本的唤醒词音素的组合，这样，就可以得到数量较多的语音样本。

附图说明

[0053] 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，还可以根据提供的附图获得其他的附图。

[0054] 图1为本发明实施例提供的一种数据处理方法的方法流程图；

[0055] 图2为本发明实施例提供的另一种数据处理方法的方法流程图；

[0056] 图3为本发明实施例提供的再一种数据处理方法的方法流程图；

[0057] 图4为本发明实施例提供的一种数据处理装置的结构示意图。

具体实施方式

[0058] 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，都属于本发明保护的范围。

[0059] 本发明实施例提供了一种数据处理方法，主要涉及的是语音信息的处理，参照图1，数据处理方法可以包括：

[0060] S11、获取语音样本以及所述语音样本对应的文本信息。

[0061] 所述语音样本为多个，可以通过网络爬虫爬取语音信息，也可以直接使用以往训练声学模型的数据，将这些数据作为语音样本。

[0062] 在获取到语音样本之后，为了语音识别准确性，可以人工将语音转换成文本，得到语音样本对应的文本信息。此外，若是对识别结果的准确性要求较低，此时可以采用机器学习、加权有限状态转换器WFST模型等方式进行语音识别。

[0063] 本实施例中，主要是用于训练设备唤醒的声学模型，因此，在得到语音样本之后，将包括唤醒词的样本作为正例，将不包括唤醒词的样本作为反例。其中，唤醒词可以是“天猫精灵”“小爱同学”等等。

[0064] S12、对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注。

[0065] 由于获取的语音样本的数量总是有限的，为能达到训练用的数据量，增强数据多样性，提升模型泛化能力，本实施例通过合成与唤醒词相关的数据来增加数据。首先，需要对语音样本进行强制对齐。强制对齐可以使用DNN-HMM模型进行对齐。

[0066] 本发明的另一实现方式中，步骤S12可以包括：

[0067] 1)提取所述语音样本的音频特征。

[0068] 提取语音样本的音频特征，音频特征可以是梅尔倒谱系数MFCC(Mel-scaleFrequency Cepstral Coefficients)特征或FBANK(Mel-Frequency Cepstral Coefficients)特征。

[0069] 2)基于所述语音样本对应的文本信息以及音频特征，对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐，得到所述音素级批注。

[0070] 将所述语音样本对应的文本信息以及音频特征，输入到DNN-HMM模型中，即可进行强制对齐。

[0071] 需要说明的是，本实施例中的强制对齐的采用的音素级对齐，这样可以给后期音素拼接提供基础。

[0072] S13、基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。

[0073] 步骤S12中进行了音素级强制对齐，进而可以将唤醒词的音素进行再组合，这样就可以得到较多的目标语音样本。

[0074] 本实施例中，获取语音样本以及所述语音样本对应的文本信息；对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。通过本发明中对语音样本的唤醒词音素的重组，就可以得到来自不同语音样本的唤醒词音素的组合，这样，就可以得到数量较多的语音样本。

[0075] 本发明的另一实现方式中，介绍步骤S13的具体实现方式，本实施例中，针对正例

和反例的处理方法不同，首先介绍正例的处理方式，参照图2，可以包括：

[0076] S21、基于所述目标语音样本对应的文本信息，筛选出包括唤醒词的目标语音样本，并作为正例样本。

[0077] 在区分正例和反例时，依据的是目标语音样本对应的文本信息，若包括唤醒词，则认为是正例，若不包括唤醒词，则认为是反例。

[0078] S22、从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元。

[0079] 利用通过DNN-HMM模型得到的音素级标注，将包含唤醒词音素的语音段截取出来得到对应的音素语音单元，建立语音库。

[0080] S23、依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

[0081] 具体的，唤醒词对应的唤醒词音素之间是有前后连接顺序的。举例来说，“小爱同学”的唤醒词音素分别为：

[0082] <sil><x><i><ao><a><i><t><o><ng><x><u><e><sil>，其中，<sil>表示静音。每一音素可以称为一基元。

[0083] 由于每一组的语音样本的音素语音单元中包括各种不同的唤醒词的音素语音单元，如同时包括男人、女人、小孩和老人说的“小爱同学”的音素语音单元，并且这些音素语音单元已经进行过强制对齐。

[0084] 即<sil><x><i><ao><a><i><t><o><ng><x><u><e><sil>会有多组，然后从这些组中随机选取第一个<sil>，然后随机选取出第二个<x>，直至选取到最后一个<sil>，然后进行顺序拼接。由于“小爱同学”一共涉及13个音素，则通过随机拼接可以产生13^{13}个“小爱同学”的语音，大大增加了训练数据的数量。

[0085] 但是需要注意的是，由于是随机进行选取拼接，则会出现两个不协调的音素拼接在一起，如小孩说的<x>，和老人说的<i>，这样会导致语音的不自然，此时需要通过基音同步叠加技术PSOLA对拼接的韵律特征进行调整，即对所述目标语音样本的韵律特征进行调整。

[0086] 另外，为了避免语音拼接的不自然，还可以仅在同一人或同一类人(处于同一年龄段和同一性别)对应的音素语音单元中筛选出音素语音单元，并进行拼接。具体的：

[0087] 筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件包括同一用户或具有相同属性值的用户；所述属性值包括年龄段和性别，依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

[0088] 相比于上述实施例中的随机拼接，本实施例中，从同一人或同一类人中选取音素语音单元，由于同一人或同一类人的语音近似，进而拼接得到的目标语音样本的语音更加自然。

[0089] 上述介绍的是对正例的处理过程，现介绍对反例的处理过程，具体的，参照图3，步骤S13可以包括：

[0090] S31、将所述语音样本中的非正例样本作为反例样本。

[0091] S32、将所述正例样本的唤醒词音素和反例样本的音素进行拆分，分别得到正例音

素组和反例音素组。

[0092] 所述正例音素组和所述反例音素组中均包括预设数量且是连续的多个音素。

[0093] 其中，预设数量可以依据实际情况而定，优选可以是3个。

[0094] S33、将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据。

[0095] 具体的，在实际场景中，会有语音非常近似唤醒词，例如唤醒词是小明，这个反例的集里主要是肖明，小米等等，听起来和唤醒词有点像的词，为了防止误唤醒，执行后续操作。

[0096] 逐帧对通过DNN-HMM模型得到的音素级标注进行修正，当前后帧或当前帧与唤醒词的音素不一致，标记当前帧为filler，否则标注不变。

[0097] 例如“小爱同学”唤醒词：

[0098] <sil><x><i><ao><a><i><t><o><ng><x><u><e><sil>

[0099] 三个音素为一组<sil><x><i>、<x><i><ao>、<i><ao><a>、…、<u><e><sil>，即为正例音素组。

[0100] ASR数据“小孩子”

[0101] <sil><x><i><ao><h><a><i><z><i><sil>，即为反例音素组。

[0102] 查找发现小孩子中的<sil><x><i>、<x><i><ao>出现在正例音素组中，则保留。

[0103] 但是<h><a><i>由于<a>前面的<h>不符合“爱”字前后文信息，即没出现在正例音素组中，则修改为:filler，其中，filler即为设定的修正标签，也是本实施例中的预设数据。

[0104] 在对每一反例样本修正后，可以将反例样本中的正例音素组以及上述的音素语音单元组合，得到大量的音素语音单元，也可以称为目标数据，然后基于目标数据进行上述的“依次随机选取出所述唤醒词的每一个唤醒词音素”操作。由于把反例的数据也进一步用于唤醒词音素的拼接，进一步增加了拼接得到的数据量。

[0105] 通过上述的实施例得到大量的训练数据之后，在训练中根据词库量合理分配正向集和反向集的比例，然后将拼接生成的唤醒词数据与原有的录制唤醒词数据以一定比例混合训练语音唤醒模型。在进行语音唤醒模型的训练时，

[0106] 可以将所述目标样本、所述目标语音样本、以及将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本按照预设比例进行混合，并训练得到语音唤醒模型。

[0107] 需要说明的是，在训练唤醒模型时，唤醒词音素、filler、sil中哪个占总体比例大，那么唤醒模型就更偏向哪个。笼统的说就是唤醒词多就唤醒率高、误唤醒率也高。filler和sil多就唤醒率低、误唤醒率也低。

[0108] 在唤醒准确率较低，或者是虚警率高时，调整所述新正例样本和所述新反例样本的数据内容，如可以增大反例比例，在召回率低就调高正例比例。反例比例一般都在75％～85％之间。

[0109] 为提高经训练的模型的各项指标可以先预设一个比例，例如正例和反例(即将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本)分别占20％和80％，正例数据中，原生语音数据(即语音样本)占比较高，可以为50％-100％，经过上述拼接方法得到的拼接数据(即目标语音样本)的比例为0％-50％。在训练中

分别调整正例反例的比例，以及调整正例中拼接数据的比例，每次调整中只调整上述二个比例中的一个，预设一个调整幅度，例如0.1%-0.5%，记录随着比例调整后训练完成的模型的召回率、准确率、虚警率的变化情形，并以此确定训练数据的比例。调整原则即为召回率、准确率和/或虚警率。

[0110] 本实施例中，通过语音拼接重组数据，实现数据增强，相比于现有的数据扩充技术如加噪声、加混响、变速、变调等方式相比，现有的变化的还不够广泛，而本实施例可以提高唤醒声学模型训练人群泛化能力，训练出的模型唤醒识别率更高。

[0111] 可选的，在上述数据处理方法的实施例的基础上，本发明的另一实施例提供了一种数据处理装置，参照图4，可以包括：

[0112] 样本获取模块101，用于获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

[0113] 对齐模块102，用于对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

[0114] 样本处理模块103，用于基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。

[0115] 本发明的另一实现方式中，所述对齐模块包括：

[0116] 特征提取子模块，用于提取所述语音样本的音频特征；

[0117] 对齐子模块，用于基于所述语音样本对应的文本信息以及音频特征，对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐，得到所述音素级批注。

[0118] 本实施例中，获取语音样本以及所述语音样本对应的文本信息；对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。通过本发明中对语音样本的唤醒词音素的重组，就可以得到来自不同语音样本的唤醒词音素的组合，这样，就可以得到数量较多的语音样本。

[0119] 需要说明的是，本实施例中的各个模块和子模块的工作过程，请参照上述实施例中的相应说明，在此不再赘述。

[0120] 可选的，在上述任一数据处理装置的实施例的基础上，所述样本处理模块103可以包括：

[0121] 样本筛选子模块，用于基于所述语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本；

[0122] 截取子模块，用于从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

[0123] 样本拼接子模块，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

[0124] 进一步，所述样本拼接子模块包括：

[0125] 筛选单元，用于筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件包括同一用户或具有相同属性值的用户；所述属性值包括年龄段和性别；

[0126] 选取单元，用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述目标

音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

[0127] 进一步，还包括：

[0128] 特征调整子模块，用于对所述目标语音样本的韵律特征进行调整。

[0129] 本发明的另一实现方式中，所述样本处理模块103还可以包括：

[0130] 样本确定子模块，用于将所述语音样本中的非正例样本作为反例样本；

[0131] 样本拆分子模块，用于将所述正例样本的唤醒词音素和反例样本的音素进行拆分，分别得到正例音素组和反例音素组；所述正例音素组和所述反例音素组中均包括预设数量且是连续的多个音素；

[0132] 数据设置子模块，用于将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据。

[0133] 进一步，样本拼接子模块用于依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本时，具体用于：

[0134] 依据所述唤醒词的唤醒词音素之间的前后连接顺序，从目标数据中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本；所述目标数据包括所述音素语音单元和所述反例音素组中的正例音素组。

[0135] 进一步，还包括：

[0136] 数据训练模块，用于将所述目标样本、所述目标语音样本、以及将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本按照预设比例进行混合，并训练得到语音唤醒模型；

[0137] 数据调整模块，用于基于所述语音唤醒模型的调整原则，调整所述目标样本、所述目标语音样本以及所述新的反例样本的比例，并重新训练所述语音唤醒模型。

[0138] 本实施例中，通过语音拼接重组数据，实现数据增强，相比于现有的数据扩充技术如加噪声、加混响、变速、变调等方式相比，现有的变化的还不够广泛，而本实施例可以提高唤醒声学模型训练人群泛化能力，训练出的模型唤醒识别率更高。

[0139] 需要说明的是，本实施例中的各个模块和子模块的工作过程，请参照上述实施例中的相应说明，在此不再赘述。

[0140] 可选的，在上述数据处理方法及装置的实施例的基础上，本发明的另一实施例提供了一种电子设备，包括：存储器和处理器；

[0141] 其中，所述存储器用于存储程序；

[0142] 处理器调用程序并用于：

[0143] 获取语音样本以及所述语音样本对应的文本信息；所述语音样本为多个；

[0144] 对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；

[0145] 基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。

[0146] 进一步，所述对所述语音样本和所述语音样本对应的文本信息进行强制对齐，得到音素级批注，包括：

[0147] 提取所述语音样本的音频特征；

[0148] 基于所述语音样本对应的文本信息以及音频特征，对所述语音样本和所述语音样本对应的文本信息进行音素级强制对齐，得到所述音素级批注。

[0149] 进一步，基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本，包括：

[0150] 基于所述语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本；

[0151] 从所述正例样本中截取出所述唤醒词的唤醒词音素，并作为音素语音单元；

[0152] 依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本。

[0153] 进一步，所述依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，包括：

[0154] 筛选出满足预设条件的音素语音单元，并作为目标音素语音单元；所述预设条件包括同一用户或具有相同属性值的用户；所述属性值包括年龄段和性别；

[0155] 依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述目标音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素。

[0156] 进一步，在按照所述前后连接顺序进行拼接重组，得到所述目标语音样本之后，还包括：

[0157] 对所述目标语音样本的韵律特征进行调整。

[0158] 进一步，在按照所述前后连接顺序进行拼接重组，得到所述目标语音样本之后，还包括：

[0159] 将所述语音样本中的非正例样本作为反例样本；

[0160] 将所述正例样本的唤醒词音素和反例样本的音素进行拆分，分别得到正例音素组和反例音素组；所述正例音素组和所述反例音素组中均包括预设数量且是连续的多个音素；

[0161] 将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据。

[0162] 进一步，依据所述唤醒词的唤醒词音素之间的前后连接顺序，从所述音素语音单元中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本，包括：

[0163] 依据所述唤醒词的唤醒词音素之间的前后连接顺序，从目标数据中依次随机选取出所述唤醒词的每一个唤醒词音素，并按照所述前后连接顺序进行拼接重组，得到所述目标语音样本；所述目标数据包括所述音素语音单元和所述反例音素组中的正例音素组。

[0164] 进一步，还包括：

[0165] 将所述目标样本、所述目标语音样本、以及将所述反例音素组中非所述正例音素组的反例音素组设置为预设数据之后得到的新的反例样本按照预设比例进行混合，并训练得到语音唤醒模型；

[0166] 基于所述语音唤醒模型的调整原则，调整所述目标样本、所述目标语音样本以及所述新的反例样本的比例，并重新训练所述语音唤醒模型。

[0167] 本实施例中，获取语音样本以及所述语音样本对应的文本信息；对所述语音样本

和所述语音样本对应的文本信息进行强制对齐，得到音素级批注；基于所述音素级批注、所述语音样本以及所述语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本。通过本发明中对语音样本的唤醒词音素的重组，就可以得到来自不同语音样本的唤醒词音素的组合，这样，就可以得到数量较多的语音样本。

[0168] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下，在其它实施例中实现。因此，本发明将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

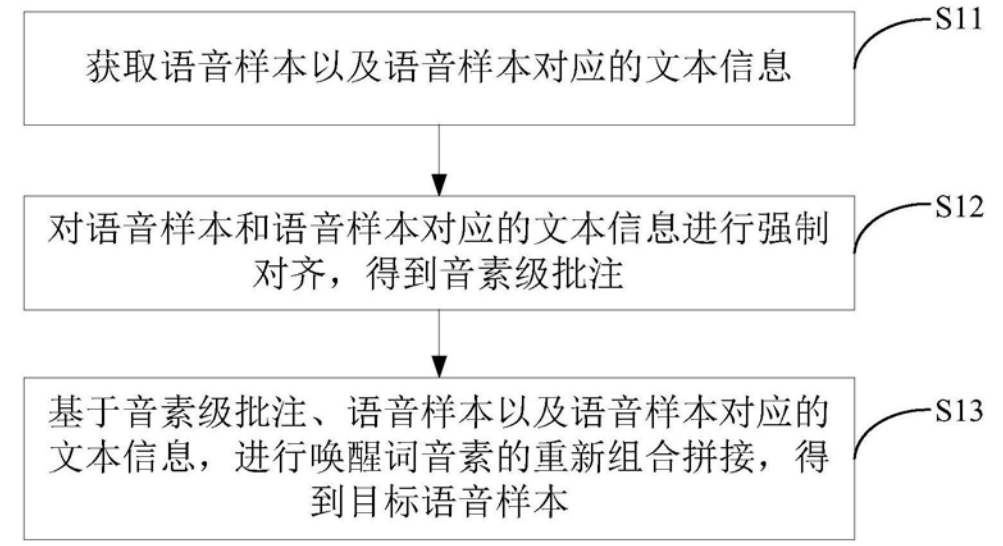

获取语音样本以及语音样本对应的文本信息
S11
对语音样本和语音样本对应的文本信息进行强制对齐，得到音素级批注
S12
基于音素级批注、语音样本以及语音样本对应的文本信息，进行唤醒词音素的重新组合拼接，得到目标语音样本
S13

图1

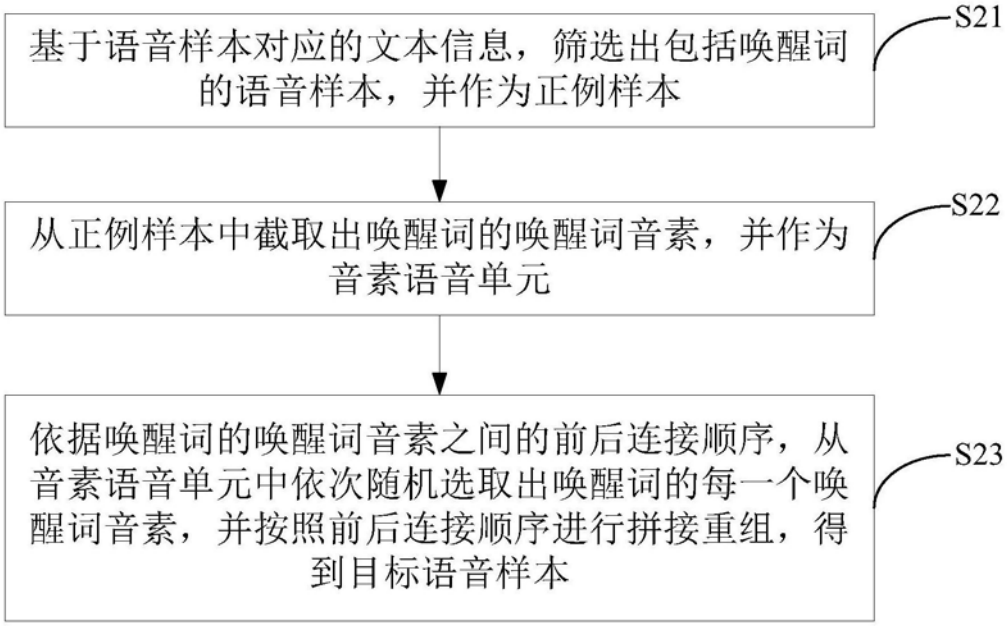

基于语音样本对应的文本信息，筛选出包括唤醒词的语音样本，并作为正例样本
S21
从正例样本中截取出唤醒词的唤醒词音素，并作为音素语音单元
S22
依据唤醒词的唤醒词音素之间的前后连接顺序，从音素语音单元中依次随机选取出唤醒词的每一个唤醒词音素，并按照前后连接顺序进行拼接重组，得到目标语音样本
S23

图2

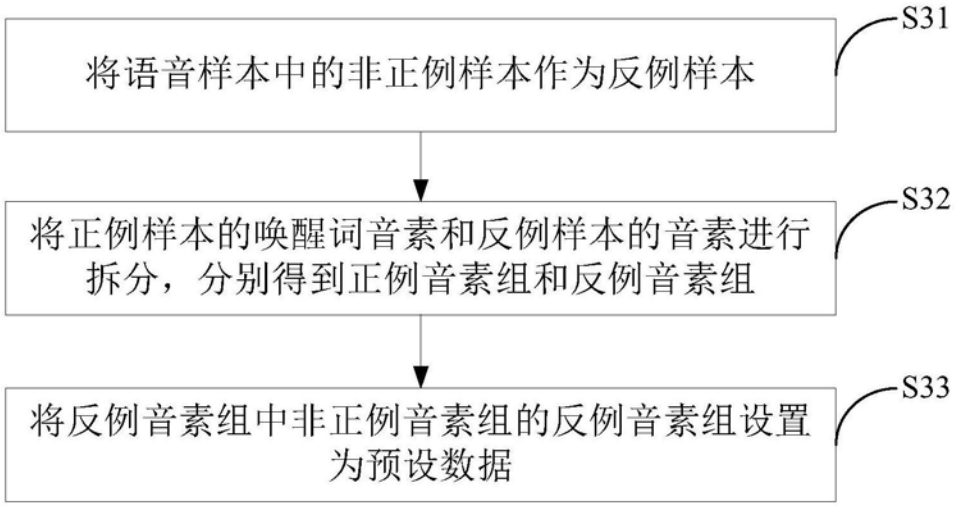

S31
将语音样本中的非正例样本作为反例样本
S32
将正例样本的唤醒词音素和反例样本的音素进行拆分，分别得到正例音素组和反例音素组
S33
将反例音素组中非正例音素组的反例音素组设置为预设数据

图3

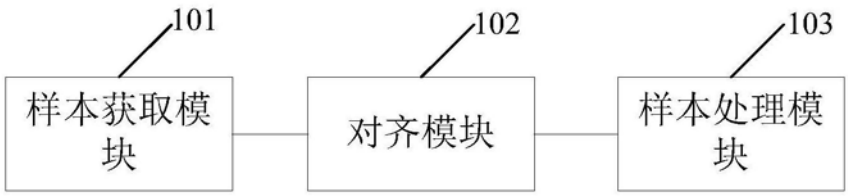

101
样本获取模块
102
对齐模块
103
样本处理模块

图4