(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116665690 A
(43)申请公布日 2023.08.29

(21)申请号 202310701942.1

(22)申请日 2023.06.13

(71)申请人 京东科技信息技术有限公司
地址 100176 北京市大兴区经济技术开发区科创十一街18号院2号楼6层601

(72)发明人 彭毅 吕安超 范璐 陈蒙 何晓冬

(74)专利代理机构 北京品源专利代理有限公司 11332
专利代理师 魏亚南

(51)Int.Cl.
G10L 21/0208(2013.01)
G10L 25/30(2013.01)
G10K 11/178(2006.01)

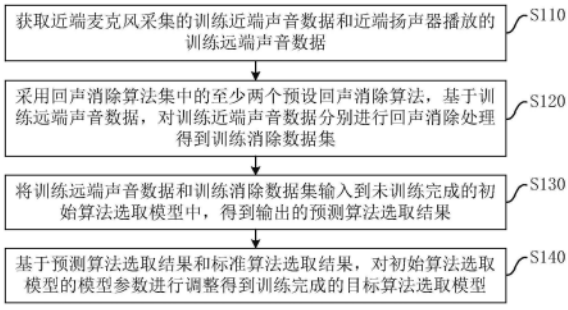
权利要求书4页 说明书16页 附图8页

(54)发明名称

算法选取模型的训练方法、回声消除方法、装置及设备

(57)摘要

本发明公开了一种算法选取模型的训练方法、回声消除方法、装置及设备。该方法包括：获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据；采用回声消除算法集中的至少两个预设回声消除算法，基于训练远端声音数据，对训练近端声音数据分别进行回声消除处理得到训练消除数据集；将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果；基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。本发明实施例中的算法选取模型能够预测出最适合当前语音场景的回声消除算法，保证了回声消除算法与当前语音场景的适配性。



CN 116665690 A

1.一种算法选取模型的训练方法,其特征在于,包括:

获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据;

采用回声消除算法集中的至少两个预设回声消除算法,基于所述训练远端声音数据,对所述训练近端声音数据分别进行回声消除处理得到训练消除数据集;

将所述训练远端声音数据和所述训练消除数据集输入到未训练完成的初始算法选取模型中,得到输出的预测算法选取结果;

基于所述预测算法选取结果和标准算法选取结果,对所述初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

2.根据权利要求1所述的方法,其特征在于,所述初始算法选取模型包括特征提取模块、拼接模块、自注意力模块、时间池化模块和输出层,其中,所述特征提取模块,用于对所述训练远端声音数据和训练消除数据集中的至少两个训练消除数据分别进行特征提取,得到训练远端语音特征和至少两个训练消除特征,所述拼接模块,用于对所述特征提取模块输出的训练远端语音特征和各所述训练消除特征进行拼接处理,得到拼接语音特征,所述自注意力模块,用于基于所述拼接模块输出的拼接语音特征,确定更新语音特征,所述时间池化模块,用于对所述自注意力模块输出的更新语音特征进行压缩,得到压缩语音特征,所述输出层,用于基于所述时间池化模块输出的压缩语音特征,输出预测算法选取结果。

3.根据权利要求1所述的方法,其特征在于,所述方法还包括:

采用预设评估算法,基于训练消除数据集中的至少一个训练消除数据,确定各所述预设回声消除算法分别对应的算法评估结果;

基于各所述算法评估结果,确定所述初始算法选取模型对应的标准算法选取结果。

4.根据权利要求1-3任一项所述的方法,其特征在于,所述基于所述预测算法选取结果和标准算法选取结果,对所述初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型,包括:

基于所述预测算法选取结果和标准算法选取结果,构建损失函数;

基于所述损失函数对所述初始算法选取模型的模型参数进行调整,得到调整后的初始算法选取模型;

基于调整后的初始算法选取模型,返回执行将所述训练远端声音数据和所述训练消除数据集输入到未训练完成的初始算法选取模型中,得到输出的预测算法选取结果的步骤;

直到所述损失函数收敛时,将调整后的初始算法选取模型作为训练完成的目标算法选取模型。

5.一种回声消除方法,其特征在于,包括:

获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据;其中,所述待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据,所述待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据;

采用回声消除算法集中的至少两个预设回声消除算法,基于所述第一待测远端声音数据,对所述第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集;

基于所述第一待测远端声音数据、所述第一待测消除数据集以及预先训练完成的目标算法选取模型,确定目标回声消除算法;

采用所述目标回声消除算法,基于所述第二待测远端声音数据,对所述第二待测近端

声音数据进行回声消除处理得到第二待测消除数据；

基于所述目标回声消除算法、所述第二待测消除数据和所述第一待测消除数据集，确定所述待测近端声音数据对应的回声消除数据；

其中，所述目标算法选取模型是采用权利要求1-4任一项所述的算法选取模型的训练方法得到的。

6.根据权利要求5所述的方法，其特征在于，所述获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据，包括：

响应于检测到消除触发指令，基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段；其中，所述第一预设语音时长小于所述第二预设语音时长；

获取近端麦克风采集的与第一待测声音时段和第二待测声音时段分别对应的第一待测近端声音数据和第二待测近端声音数据；

获取近端播放器播放的与第一待测声音时段和第二待测声音时段分别对应的第一待测远端声音数据和第二待测远端声音数据。

7.根据权利要求6所述的方法，其特征在于，所述方法还包括：

响应于检测到语音开始指令，在所述近端麦克风采集到的声音数据中存在语音数据的情况下，生成消除触发指令；和/或，

在当前采集时刻超出所述第二待测声音时段的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤；和/或，

响应于检测到静音时段，在所述静音时段对应的静音时长超过静音时长阈值的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤。

8.根据权利要求5所述的方法，其特征在于，所述回声消除算法集中包含回声消除模型，所述回声消除模型中包含输入模块、块内模块、转换模块、块外模块和输出模块，其中，所述输入模块，用于基于输入的第一待测远端声音数据和所述第一待测近端声音数据，确定第一待测语音特征，所述块内模块，用于对所述输入模块输出的第一待测语音特征进行频域特征处理，输出第一频域语音特征，所述转换模块，用于对所述块内模块输出的第一频域语音特征进行转换，得到第一转换语音特征，所述块外模块，用于对所述转换模块输出的第一转换语音特征进行时域特征处理，得到第一时域语音特征，所述输出模块，用于基于所述块外模块输出的第一时域语音特征和所述第一待测近端声音数据，输出第一待测消除数据。

9.根据权利要求8所述的方法，其特征在于，所述块内模块包括自注意力网络，所述块外模块包括长短时记忆网络。

10.根据权利要求8所述的方法，其特征在于，所述输入模块包括串联的预处理层、拼接层和卷积网络，其中，所述预处理层，用于对输入的所述第一待测远端声音数据和所述第一待测近端声音数据分别进行短时傅里叶变换，得到第一待测远端语音频谱和第一待测近端语音频谱，所述拼接层，用于对所述预处理层输出的第一待测远端语音频谱和第一待测近端语音频谱进行拼接，得到第一待测拼接语音频谱，所述卷积网络，用于对所述拼接层输出

的第一待测拼接语音频谱进行特征提取，输出第一待测语音特征。

11.根据权利要求10所述的方法，其特征在于，所述输出模块包括串联的反卷积网络、融合单元和短时傅里叶逆变换层，其中，所述反卷积网络，用于对所述块外模块输出的第一时域语音特征进行反卷积处理，得到第一反卷积语音特征，所述融合单元，用于对所述第一待测近端声音数据和所述反卷积网络输出的第一反卷积语音特征进行融合处理，得到第一待测融合语音特征，所述短时傅里叶逆变换层，用于对所述融合单元输出的第一待测融合语音特征进行短时傅里叶逆变换，输出第一待测消除数据。

12.一种算法选取模型的训练装置，其特征在于，包括：

训练声音数据获取模块，用于获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据；

训练消除数据集确定模块，用于采用回声消除算法集中的至少两个预设回声消除算法，基于所述训练远端声音数据，对所述训练近端声音数据分别进行回声消除处理得到训练消除数据集；

预测算法选取结果输出模块，用于将所述训练远端声音数据和所述训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果；

目标算法选取模型确定模块，用于基于所述预测算法选取结果和标准算法选取结果，对所述初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

13.一种回声消除装置，其特征在于，包括：

待测声音数据获取模块，用于获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据；其中，所述待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据，所述待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据；

第一待测消除数据集确定模块，用于采用回声消除算法集中的至少两个预设回声消除算法，基于所述第一待测远端声音数据，对所述第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集；

目标回声消除算法确定模块，用于基于所述第一待测远端声音数据、所述第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法；

第二待测消除数据确定模块，用于采用所述目标回声消除算法，基于所述第二待测远端声音数据，对所述第二待测近端声音数据进行回声消除处理得到第二待测消除数据；

回声消除数据确定模块，用于基于所述目标回声消除算法、所述第二待测消除数据和所述第一待测消除数据集，确定所述待测近端声音数据对应的回声消除数据；

其中，所述目标算法选取模型是采用权利要求1-4任一项所述的算法选取模型的训练方法得到的。

14.一种电子设备，其特征在于，所述电子设备包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求1-4中任一项所述的算法选取模型的训练方法，和/或，能够执行权利要求5-11中任一项所述的回声消除方法。

15.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使处理器执行时实现权利要求1-4中任一项所述的算法选取模型的训练方法,和/或,能够执行权利要求5-11中任一项所述的回声消除方法。

算法选取模型的训练方法、回声消除方法、装置及设备

技术领域

[0001] 本发明涉及人工智能技术领域，尤其涉及一种算法选取模型的训练方法、回声消除方法、装置及设备。

背景技术

[0002] 近些年，线下会议、面对面交流等线下沟通方式越来越多地转变为实时视频会议、语音聊天等线上沟通方式，因此用户对音视频的声音质量有了更加严苛的要求。

[0003] 其中，回声是影响音视频通信体验的一个重要因素，其表现为远端说话人的声音数据经过远端麦克风、近端扬声器、近端麦克风等设备、空间环境传播和反射等又传回了远端扬声器，使得远端说话人又听到了自己的声音。目前，针对回声消除开发出了很多相关算法，如自适应滤波算法和深度学习模型等。

[0004] 在实现本发明的过程中，发现现有技术中至少存在以下技术问题：

[0005] 自适应滤波算法对远端单讲场景具备很好的消除性能，但对于远近端双讲场景或存在非线性失真场景(如噪声大、延时大或扬声器质量差)的消除性能较差。而深度学习模型的消除性能与自适应滤波算法正相反。

[0006] 因此，不同的回声消除算法所适用的语音场景不同，目前还未开发出能够适用复杂语音场景的回声消除算法。

发明内容

[0007] 本发明实施例提供了一种算法选取模型的训练方法、回声消除方法、装置及设备，以解决不同的回声消除算法所适用的语音场景不同的问题，提高回声消除算法与复杂语音场景的适配性，进而提高回声消除后的语音质量。

[0008] 根据本发明一个实施例提供了一种算法选取模型的训练方法，该方法包括：

[0009] 获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据；

[0010] 采用回声消除算法集中的至少两个预设回声消除算法，基于所述训练远端声音数据，对所述训练近端声音数据分别进行回声消除处理得到训练消除数据集；

[0011] 将所述训练远端声音数据和所述训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果；

[0012] 基于所述预测算法选取结果和标准算法选取结果，对所述初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

[0013] 根据本发明一个实施例提供了一种回声消除方法，该方法包括：

[0014] 获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据；其中，所述待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据，所述待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据；

[0015] 采用回声消除算法集中的至少两个预设回声消除算法，基于所述第一待测远端声

音数据，对所述第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集；

[0016] 基于所述第一待测远端声音数据、所述第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法；

[0017] 采用所述目标回声消除算法，基于所述第二待测远端声音数据，对所述第二待测近端声音数据进行回声消除处理得到第二待测消除数据；

[0018] 基于所述目标回声消除算法、所述第二待测消除数据和所述第一待测消除数据集，确定所述待测近端声音数据对应的回声消除数据；

[0019] 其中，所述目标算法选取模型是采用本发明任一实施例所述的算法选取模型的训练方法得到的。

[0020] 根据本发明另一个实施例提供了一种算法选取模型的训练装置，该装置包括：

[0021] 训练声音数据获取模块，用于获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据；

[0022] 训练消除数据集确定模块，用于采用回声消除算法集中的至少两个预设回声消除算法，基于所述训练远端声音数据，对所述训练近端声音数据分别进行回声消除处理得到训练消除数据集；

[0023] 预测算法选取结果输出模块，用于将所述训练远端声音数据和所述训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果；

[0024] 目标算法选取模型确定模块，用于基于所述预测算法选取结果和标准算法选取结果，对所述初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

[0025] 根据本发明另一个实施例提供了一种回声消除装置，该装置包括：

[0026] 待测声音数据获取模块，用于获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据；其中，所述待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据，所述待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据；

[0027] 第一待测消除数据集确定模块，用于采用回声消除算法集中的至少两个预设回声消除算法，基于所述第一待测远端声音数据，对所述第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集；

[0028] 目标回声消除算法确定模块，用于基于所述第一待测远端声音数据、所述第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法；

[0029] 第二待测消除数据确定模块，用于采用所述目标回声消除算法，基于所述第二待测远端声音数据，对所述第二待测近端声音数据进行回声消除处理得到第二待测消除数据；

[0030] 回声消除数据确定模块，用于基于所述目标回声消除算法、所述第二待测消除数据和所述第一待测消除数据集，确定所述待测近端声音数据对应的回声消除数据；

[0031] 其中，所述目标算法选取模型是采用本发明任一实施例所述的算法选取模型的训练方法得到的。

[0032] 根据本发明另一个实施例提供了一种电子设备，该电子设备包括：

[0033] 至少一个处理器；以及

[0034] 与所述至少一个处理器通信连接的存储器；其中，

[0035] 所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行本发明任一实施例所述的算法选取模型的训练方法，和/或，能够执行本发明任一实施例所述回声消除方法。

[0036] 根据本发明另一个实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机指令，所述计算机指令用于使处理器执行时实现本发明任一实施例所述的算法选取模型的训练方法，和/或，能够执行本发明任一实施例所述回声消除方法。

[0037] 本发明实施例的技术方案，通过采用回声消除算法集中的至少两个预设回声消除算法，基于近端扬声器播放的训练远端声音数据，对近端麦克风采集的训练近端声音数据分别进行回声消除处理得到训练消除数据集，并将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果，基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型，本发明实施例训练得到的算法选取模型能够预测出最适合当前语音场景的回声消除算法，解决了不同的回声消除算法所适用的语音场景不同的问题，保证了回声消除算法与当前语音场景的适配性，进而提高了回声消除后的语音质量。

[0038] 应当理解，本部分所描述的内容并非旨在标识本发明的实施例的关键或重要特征，也不用于限制本发明的范围。本发明的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0039] 为了更清楚地说明本发明实施例中的技术方案，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。

[0040] 图1为本发明一个实施例所提供的一种算法选取模型的训练方法的流程图；

[0041] 图2为本发明一个实施例所提供的一种音视频会议的语音场景的示意图；

[0042] 图3为本发明一个实施例所提供的一种初始算法选取模型的网络架构图；

[0043] 图4为本发明一个实施例所提供的一种回声消除方法的流程图；

[0044] 图5为本发明一个实施例所提供的另一种回声消除方法的流程图；

[0045] 图6为本发明一个实施例所提供的一种回声消除方法的具体实例的流程图；

[0046] 图7为本发明一个实施例所提供的一种回声消除模型的网络架构图；

[0047] 图8为本发明一个实施例所提供的一种算法选取模型的训练装置的结构示意图；

[0048] 图9为本发明一个实施例所提供的一种回声消除装置的结构示意图；

[0049] 图10为本发明一个实施例所提供的一种电子设备的结构示意图。

具体实施方式

[0050] 为了使本技术领域的人员更好地理解本发明方案，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的实施例，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，都应当属于本发明保护的范围。

[0051] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“初始”、“目标”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0052] 图1为本发明一个实施例所提供的一种算法选取模型的训练方法的流程图，本实施例可适用于对选择回声消除算法的网络模型进行训练的情况，该方法可以由算法选取模型的训练装置来执行，该算法选取模型的训练装置可以采用硬件和/或软件的形式实现，该算法选取模型的训练装置可配置于终端设备中。如图1所示，该方法包括：

[0053] S110、获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据。

[0054] 其中，示例性的，本发明实施例可适用于音视频会议、语音游戏、智能音箱控制等语音场景。具体的，音视频会议的语音场景中包含近端扬声器、近端麦克风、远端扬声器和远端麦克风，而语音游戏、智能音箱控制等语音场景仅包含近端扬声器和近端麦克风。在语音游戏、智能音箱控制等语音场景中，近端麦克风在接收用户输入的语音信号的同时，也可能同时接收到近端扬声器播放的队友语音、背景音乐等声音信号，因此同样存在进行回声消除的需求。

[0055] 虽然在语音游戏、智能音箱控制等语音场景中不存在“远端”的概念，但为了统一描述以及与近端声音数据进行区别，在本实施例中，将不同的语音场景中近端扬声器播放的声音数据均定义为远端声音数据。

[0056] 图2为本发明一个实施例所提供的一种音视频会议的语音场景的示意图。具体的，该语音场景中包含近端扬声器、近端麦克风、远端扬声器和远端麦克风，其中，定义“近端”为主说话人所在的空间环境，“远端”为参与通话的其他人所在的空间环境。具体的，n表示对音视频流采样后的采样点索引，lpb(n)表示近端扬声器播放的远端声音数据，lpb_e(n)表示经过环境反射、空间传播等过程被近端麦克风捕捉到的声音数据，s(n)表示主说话人的语音数据，v(n)表示近端空间环境中的噪声数据。容易理解的是，近端麦克风采集到的近端声音数据mic(n)＝s(n)+lpb_e(n)+v(n)，回声消除的目的是为了去除近端声音数据mic(n)中的lpb_e(n)和v(n)，得到回声消除数据S(n)。

[0057] 以图2为例，训练近端声音数据对应图2中的mic(n)，训练远端声音数据对应图2中的lpb(n)，其中，lpb(n)可采用系统设备中的回采通路获取到。

[0058] 其中，具体的，训练近端声音数据和训练远端声音数据对应同一声音时段。其中，示例性的，训练近端声音数据和训练远端声音数据可采用开源的声音数据或真实采集到的声音数据。

[0059] S120、采用回声消除算法集中的至少两个预设回声消除算法，基于训练远端声音数据，对训练近端声音数据分别进行回声消除处理得到训练消除数据集。

[0060] 其中，具体的，回声消除算法集中包含至少两个预设回声消除算法。示例性的，回声消除算法集中包含的预设回声消除算法包括但不限于基于滤波的回声消除算法、基于变

换的回声消除算法、自适应滤波算法和回声消除模型等等，此处对回声消除算法集中的算法数量及算法类型均不作限定。

[0061] 在一个实施例中，基于滤波的回声消除算法的算法原理为将训练近端声音数据输入到滤波器中，得到输出的训练消除数据，其中，可基于训练远端声音数据的延时时间设置滤波器的截止频率。

[0062] 在一个实施例中，基于变换的回声消除算法的算法原理为利用正反变换算法，将训练远端声音数据进行反转，然后将反转后的训练远端声音数据与训练近端声音数据做抵消处理，得到训练消除数据。

[0063] 在一个实施例中，自适应滤波算法可分类为时延估计、线性滤波、双讲检测和残余回声抑制等4个模块，其算法原理为通过调整滤波器的权重向量，估计一个近似的回声路径来逼近真实的回声路径，从而得到一个模拟的训练远端声音数据，并在训练近端声音数据中除去模拟的训练远端声音数据，得到训练消除数据。

[0064] 在一个实施例中，回声消除模型的输入数据为训练远端声音数据和训练近端声音数据，输出数据为训练消除数据，此处对回声消除模型的网络架构和训练方法不作限定。

[0065] 其中，具体的，训练消除数据集中包含采用至少两个预设回声消除算法分别得到的训练近端声音数据的训练消除数据，示例性的，训练消除数据集可采用$\{S|S_1(n),S_2(n),\cdots,S_m(n)\}$表示，其中，m表示训练消除数据集中训练消除数据的总数量。

[0066] S130、将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果。

[0067] 在一个具体实施例中，本发明实施例对初始算法选取模型的网络架构不作限定。

[0068] 在另一个具体实施例中，初始算法选取模型包括特征提取模块、拼接模块、自注意力模块、时间池化模块和输出层，其中，特征提取模块，用于对训练远端声音数据和训练消除数据集中的至少两个训练消除数据分别进行特征提取，得到训练远端语音特征和至少两个训练消除特征，拼接模块，用于对特征提取模块输出的训练远端语音特征和各训练消除特征进行拼接处理，得到拼接语音特征，自注意力模块，用于基于拼接模块输出的拼接语音特征，确定更新语音特征，时间池化模块，用于对自注意力模块输出的更新语音特征进行压缩，得到压缩语音特征，输出层，用于基于时间池化模块输出的压缩语音特征，输出预测算法选取结果。

[0069] 其中，示例性的，特征提取模块可采用卷积神经网络(Convolutional Neural Networks，CNN)，此处对特征提取模块的网络架构不作限定。

[0070] 其中，示例性的，输出层包括全连接层和softmax函数，此处对输出层的网络架构不作限定。

[0071] 其中，具体的，自注意力模块针对拼接语音特征中的每个时间语音特征，使得该时间语音特征能够利用周围语音特征实现对自身的更新，得到相比于拼接语音特征更准确的更新语音特征。

[0072] 图3为本发明一个实施例所提供的一种初始算法选取模型的网络架构图。具体的，初始算法选取模型中的特征提取模块采用的是CNN网络，输出层包括全连接层(fully connected)和softmax函数。初始算法选取模型的输入数据包括训练远端声音数据lpb(n)以及训练消除数据集中的至少两个训练消除数据$(S_1(n),S_2(n),\cdots,S_m(n))$，输出数据为各

预设回声消除算法分别对应的预测结果$p_1,p_2,\cdots,p_m$。

[0073] 其中，示例性的，预测结果可以为预测选取概率、预测不选取概率或是否选取等。具体的，各预测结果之和为1。举例而言，假设回声消除算法集中包含算法1、算法2和算法3，则预测算法选取结果可以为[0.6 0.20.2]，或者，预测算法选取结果可以为[1 0 0]，其中，“1”表示选取，“0”表示不选取。此处对预测算法选取结果不作限定。

[0074] S140、基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

[0075] 在上述实施例的基础上，具体的，该方法还包括：采用预设评估算法，基于训练消除数据集中的至少一个训练消除数据，确定各预设回声消除算法分别对应的算法评估结果；基于各算法评估结果，确定初始算法选取模型对应的标准算法选取结果。

[0076] 其中，示例性的，预设评估算法包括但不限于PESQ(Perceptual evaluation of speech qualit，客观语音质量评估)算法、ERLE(echo return loss enhancement，回波损失增益)算法、CER(character error rate，字符差错率)算法和STOI(Short-Time Objective Intelligibility，短时客观可懂度)算法等中至少一种，此处对预设评估算法不作限定。

[0077] 其中，PESQ算法是一种客观的、全参考的语音质量评估方法，能够对客观语音质量评估提供一个主观MOS的预测值，计算过程包括了预处理、时间对齐、感知滤波、掩蔽效果等等。具体的，PESQ算法的算法评估结果的得分范围在-0.5～4.5之间，得分越高表示预设回声消除算法的回声消除效果越好。

[0078] 其中，ERLE算法常用于评估近端空间环境中单讲场景下的回声消除效果。示例性的，ERLE算法满足公式：

[0079] $$ERLE(dB)=10log_{10}\frac{E\{mic^2(n)\}}{E\{S^2(n)\}}$$

[0080] 其中，当训练消除数据后续需用于语音识别时，CER算法是衡量回声消除效果的重要指标，CER算法用于衡量两个序列的相似度。

[0081] 其中，STOI算法是衡量语音可懂度的重要指标，STOI算法的算法评估结果的得分范围在0-1之间，得分越高表示预设回声消除算法的回声消除效果越好。

[0082] 其中，示例性的，标准算法选取结果可用于表征预设回声消除算法对训练近端声音数据的回声消除效果，如当标准算法选取结果为各预设回声消除算法分别对应的标准选取概率时，标准选取概率越高，说明该预设回声消除算法的回声消除效果越好。

[0083] 在上述实施例的基础上，具体的，基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型，包括：基于预测算法选取结果和标准算法选取结果，构建损失函数；基于损失函数对初始算法选取模型的模型参数进行调整，得到调整后的初始算法选取模型；基于调整后的初始算法选取模型，返回执行将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果的步骤；直到损失函数收敛时，将调整后的初始算法选取模型作为训练完成的目标算法选取模型。

[0084] 其中，示例性的，损失函数的函数类型包括但不限于平方损失函数、对数损失函数、指数损失函数、均方误差损失函数、逻辑回归损失函数、Huber损失函数、交叉熵损失函

数和Kullback-Leibler散度损失函数等等。此处对损失函数的损失类型不作限定。

[0085] 在一个具体实施例中，损失函数的函数类型为交叉损失函数。示例性的，交叉损失函数L满足公式：

[0086] $$L = -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$$

[0087] 其中，y表示标准算法选取结果，$\hat{y}$表示预测算法选取结果。

[0088] 本实施例的技术方案，通过采用回声消除算法集中的至少两个预设回声消除算法，基于近端扬声器播放的训练远端声音数据，对近端麦克风采集的训练近端声音数据分别进行回声消除处理得到训练消除数据集，并将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果，基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型，本发明实施例训练得到的算法选取模型能够预测出最适合当前语音场景的回声消除算法，解决了不同的回声消除算法所适用的语音场景不同的问题，保证了回声消除算法与当前语音场景的适配性，进而提高了回声消除后的语音质量。

[0089] 图4为本发明一个实施例所提供的一种回声消除方法的流程图，本实施例可适用于对语音场景采集到的声音数据中的回声信号进行消除的情况，该方法可以由回声消除装置来执行，该回声消除装置可以采用硬件和/或软件的形式实现，该回声消除装置可配置于终端设备中。如图4所示，该方法包括：

[0090] S210、获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据。

[0091] 其中，示例性的，本发明实施例可适用于音视频会议、语音游戏、智能音箱控制等语音场景。具体的，音视频会议的语音场景中包括近端扬声器、近端麦克风、远端扬声器和远端麦克风，语音游戏、智能音箱控制等语音场景仅包括近端扬声器和近端麦克风。在语音游戏、智能音箱控制等语音场景中，近端麦克风在接收用户输入的语音信号的同时，也可能同时接收到近端扬声器播放的队友语音、背景音乐等声音信号，因此同样存在进行回声消除的需求。

[0092] 在本实施例中，待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据，待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据。

[0093] 其中，具体的，对待测近端声音数据进行划分得到第一待测近端声音数据和第二待测近端声音数据，对待测远端声音数据进行划分得到第一待测远端声音数据和第二待测远端声音数据，第一待测近端声音数据与第二待测近端声音数据不重叠，第一待测远端声音数据和第二待测远端声音数据不重叠。

[0094] S220、采用回声消除算法集中的至少两个预设回声消除算法，基于第一待测远端声音数据，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集。

[0095] 其中，具体的，回声消除算法集中包含至少两个预设回声消除算法。示例性的，回声消除算法集中包含的预设回声消除算法包括但不限于基于滤波的回声消除算法、基于变换的回声消除算法、自适应滤波算法和回声消除模型等等，此处对回声消除算法集中的算法数量及算法类型均不作限定。

[0096] 其中，具体的，第一待测消除数据集中包含采用至少两个预设回声消除算法分别得到的第一待测近端声音数据的第一待测消除数据，示例性的，第一待测消除数据集可采

用{S|S_1(n),S_2(n),…,S_m(n)}表示,其中,m表示第一待测消除数据集中训练消除数据的总数量。

[0097] S230、基于第一待测远端声音数据、第一待测消除数据集以及预先训练完成的目标算法选取模型,确定目标回声消除算法。

[0098] 其中,具体的,将第一待测远端声音数据和第一待测消除数据集输入到预先训练完成的目标算法选取模型中,得到目标算法选取结果,基于目标算法选取结果,确定目标回声消除算法。

[0099] 在本实施例中,目标算法选取模型是采用本发明上述实施例提供的算法选取模型的训练方法得到的,此处不再赘述。

[0100] 其中,示例性的,目标算法选取结果可以是每个预设回声消除算法分别对应的目标选取概率、目标不选取概率或是否选取等。具体的,当预测算法选取结果为每个预设回声消除算法分别对应的目标选取概率时,将最高的目标选取概率对应的预设回声消除算法作为目标回声消除算法,或者,将目标选取概率超过预设概率阈值的预设回声消除算法作为目标回声消除算法,如预设概率阈值可以为0.85;当预测算法选取结果为每个预设回声消除算法分别对应的目标不选取概率时,将最小的目标不选取概率对应的预设回声消除算法作为目标回声消除算法,或者,将目标选取概率不超过预设概率阈值的预设回声消除算法作为目标回声消除算法,如预设概率阈值可以为0.25;当目标算法选取结果为每个预设回声消除算法分别对应的是否选取时,将选取的预设回声消除算法作为目标回声消除算法。

[0101] S240、采用目标回声消除算法,基于第二待测远端声音数据,对第二待测近端声音数据进行回声消除处理得到第二待测消除数据。

[0102] S250、基于目标回声消除算法、第二待测消除数据和第一待测消除数据集,确定待测近端声音数据对应的回声消除数据。

[0103] 其中,具体的,将第一待测消除数据集中采用目标回声消除算法得到的第一待测消除数据作为目标待测消除数据,基于目标待测消除数据和第二待测消除数据,生成待测近端声音数据对应的回声消除数据。

[0104] 本实施例的技术方案,通过采用回声消除算法集中的至少两个预设回声消除算法,对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集,采用预先训练完成的目标算法选取模型预测目标回声消除算法,采用目标回声消除算法,基于第二待测远端声音数据,对第二待测近端声音数据进行回声消除处理得到第二待测消除数据,本发明实施例采用算法选取模型确定最适合当前语音场景的回声消除算法,解决了不同的回声消除算法所适用的语音场景不同的问题,保证了回声消除算法与当前语音场景的适配性,进而提高了回声消除后的语音质量。

[0105] 图5为本发明一个实施例所提供的另一种回声消除方法的流程图,本发明实施例对上述实施例中“获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据”进行进一步细化,如图5所示,该方法包括:

[0106] S310、响应于检测到消除触发指令,基于当前触发时刻、第一预设语音时长和第二预设语音时长,生成第一待测声音时段和第二待测声音时段。

[0107] 在本实施例中,第一预设语音时长小于第二预设语音时长。这样设置的好处在于,可以提高后续目标回声消除算法的确定流程的执行效率,以及提高了回声消除方法的整体

执行效率。

[0108] 其中，示例性的，假设当前触发时刻为8:00，第一预设语音时长和第二预设语音时长分别为1分钟和59分钟，则第一待测声音时段和第二待测声音时段分别为8:00-8:01和8:01-9:00。

[0109] 在上述实施例的基础上，具体的，该方法还包括：响应于检测到语音开始指令，在近端麦克风采集到的声音数据中存在语音数据的情况下，生成消除触发指令；和/或，在当前采集时刻超出第二待测声音时段的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤；和/或，响应于检测到静音时段，在静音时段对应的静音时长超过静音时长阈值的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤。

[0110] 在一个实施例中，当语音通话开始时，对近端麦克风采集到的声音数据进行人声识别，当识别结果中存在语音数据的情况下，生成消除触发指令。其中，语音数据用于表征人的发音器官发出的具有一定社会意义的声音信号。

[0111] 在一个实施例中，在语音场景中，近端麦克风实时采集待测近端声音数据，在当前采集时刻超出第二待测声音时段的情况下，生成消除触发指令，并返回执行S310。

[0112] 这样设置的好处在于，在实际的语音场景中，通话双方所处的语音场景发生变化的可能性很小，在一段时间内，语音场景会维持单讲场景、双讲场景或非线性失真场景不变，因此在可第二预设语音时长内保持采用同一回声消除算法进行回声消除处理。但随着语音时长的延长，语音环境发生变化的可能性会增加，本发明实施例通过对第二预设语音时长进行限定，实现每隔一段时间重新进行回声消除算法的选择的目的，使得回声消除方法能够适应语音场景变化的复杂语音场景，提高了回声消除过程的灵活性，从而进一步保证了回声消除后的语音质量。

[0113] 在一个实施例中，示例性的，静音时长阈值可以为5分钟或10分钟。

[0114] 这样设置的好处在于，静音时长越长，表征语音环境发生变化的可能性越大，本发明实施例通过对静音时长进行检测，使得回声消除方法能够适应语音场景变化的复杂语音场景，提高了回声消除过程的灵活性，进一步保证了回声消除后的语音质量。

[0115] S320、获取近端麦克风采集的与第一待测声音时段和第二待测声音时段分别对应的第一待测近端声音数据和第二待测近端声音数据。

[0116] S330、获取近端播放器播放的与第一待测声音时段和第二待测声音时段分别对应的第一待测远端声音数据和第二待测远端声音数据。

[0117] S340、采用回声消除算法集中的至少两个预设回声消除算法，基于第一待测远端声音数据，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集。

[0118] S350、基于第一待测远端声音数据、第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法。

[0119] S360、采用目标回声消除算法，基于第二待测远端声音数据，对第二待测近端声音数据进行回声消除处理得到第二待测消除数据。

[0120] S370、基于目标回声消除算法、第二待测消除数据和第一待测消除数据集，确定待测近端声音数据对应的回声消除数据。

[0121] 图6为本发明一个实施例所提供的一种回声消除方法的具体实例的流程图。具体的，在通话开始后，实时检测近端麦克风是否采集到声音数据，如果是，则基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段t1和第二待测声音时段t2。采用回声消除算法集中的至少两个预设回声消除算法，对t1的近端音频进行回声消除处理，得到至少两个第一待测消除数据($S_1(n)$,$S_2(n)$,…,$S_m(n)$)。基于t1时段的远端音频、各第一待测消除数据和目标算法选取模型，确定目标回声消除算法。判断当前采集时刻是否在t2时段内，如果否，则返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段t1和第二待测声音时段t2的步骤，如果是，则采用目标回声消除算法，对t2的近端音频进行回声消除处理，得到第二待测消除数据x(n)，相应的，t1+t2时段对应的回声消除数据$S(n)=x(n)+S_x(n)$，其中，$S_x(n)$表示第一待测消除数据集中采用目标回声消除算法得到的第一待测消除数据。

[0122] 本实施例的技术方案，通过响应于检测到消除触发指令，基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段，获取近端麦克风采集的与第一待测声音时段和第二待测声音时段分别对应的第一待测近端声音数据和第二待测近端声音数据，获取近端播放器播放的与第一待测声音时段和第二待测声音时段分别对应的第一待测远端声音数据和第二待测远端声音数据，解决了回声消除算法的选取时机单一的问题，提高了回声消除算法的选取时机的灵活性，进一步提高了回声消除过程的灵活性，从而进一步保证了回声消除后的语音质量。

[0123] 在上述实施例的基础上，具体的，回声消除算法集中包含回声消除模型，回声消除模型中包含输入模块、块内模块、转换模块、块外模块和输出模块，其中，输入模块，用于基于输入的第一待测远端声音数据和第一待测近端声音数据，确定第一待测语音特征，块内模块，用于对输入模块输出的第一待测语音特征进行频域特征处理，输出第一频域语音特征，转换模块，用于对块内模块输出的第一频域语音特征进行转换，得到第一转换语音特征，块外模块，用于对转换模块输出的第一转换语音特征进行时域特征处理，得到第一时域语音特征，输出模块，用于基于块外模块输出的第一时域语音特征和第一待测近端声音数据，输出第一待测消除数据。

[0124] 其中，具体的，回声消除模型的网络架构为双路网络结构。

[0125] 在一个具体实施例中，输入模块包括串联的预处理层、拼接层和卷积网络，其中，预处理层，用于对输入的第一待测远端声音数据和第一待测近端声音数据分别进行短时傅里叶变换(short-time Fourier transform,STFT)，得到第一待测远端语音频谱和第一待测近端语音频谱，拼接层，用于对预处理层输出的第一待测远端语音频谱和第一待测近端语音频谱进行拼接，得到第一待测拼接语音频谱，卷积网络，用于对拼接层输出的第一待测拼接语音频谱进行特征提取，输出第一待测语音特征。

[0126] 在一个具体实施例中，块内模块包括自注意力网络，块外模块包括长短时记忆网络(Long Short-Term Memory,LSTM)。

[0127] 在一个具体实施例中，输出模块包括串联的反卷积网络、融合单元和短时傅里叶逆变换层，其中，反卷积网络，用于对块外模块输出的第一时域语音特征进行反卷积处理，得到第一反卷积语音特征，融合单元，用于对第一待测近端声音数据和反卷积网络输出的第一反卷积语音特征进行融合处理，得到第一待测融合语音特征，短时傅里叶逆变换层，用

于对融合单元输出的第一待测融合语音特征进行短时傅里叶逆变换(inverse short-time Fourier transform,iSTFT),输出第一待测消除数据。

[0128] 其中,示例性的,融合单元的融合操作可以为按位相乘。

[0129] 图7为本发明一个实施例所提供的一种回声消除模型的网络架构图。具体的,回声消除模型的输入数据为第一待测远端声音数据和第一待测近端声音数据,输出数据为第一待测消除数据。其中,在块内模块采用的自注意力网络中,块内的每个频域点(Q)评估和其他频域点(K)的关系,并采用权重表示,以此权重对其他频域点对应的频域值(V)进行加权求和,得到该频域点基于自注意力机制的向量表示。“Q”中的浅色频域点表示得到“V”中浅色频域点的过程。

[0130] 这样设置的好处在于,传统的回声消除模型的块内模块同样采用长短时记忆网络,使得块内模块仅能表示某一时刻的频域语音特征,并没有时序的要求,因此长短时记忆网络只能学习有限的特征信息,建模能力有限。本发明实施例中的回声消除模型中的块内模块采用的是自注意力机制的自注意力网络,相比于长短时记忆网络,块内模块不仅能够学习频域点之间的时序信息,还能根据评估出的当前频域点与其他频域点关系的紧密程度,丰富了当前频域点的特征表示,增强了建模能力。

[0131] 图8为本发明一个实施例所提供的一种算法选取模型的训练装置的结构示意图。如图8所示,该装置包括:训练声音数据获取模块410、训练消除数据集确定模块420、预测算法选取结果输出模块430和目标算法选取模型确定模块440。

[0132] 其中,训练声音数据获取模块410,用于获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据;

[0133] 训练消除数据集确定模块420,用于采用回声消除算法集中的至少两个预设回声消除算法,基于训练远端声音数据,对训练近端声音数据分别进行回声消除处理得到训练消除数据集;

[0134] 预测算法选取结果输出模块430,用于将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中,得到输出的预测算法选取结果;

[0135] 目标算法选取模型确定模块440,用于基于预测算法选取结果和标准算法选取结果,对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型。

[0136] 本实施例的技术方案,通过采用回声消除算法集中的至少两个预设回声消除算法,基于近端扬声器播放的训练远端声音数据,对近端麦克风采集的训练近端声音数据分别进行回声消除处理得到训练消除数据集,并将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中,得到输出的预测算法选取结果,基于预测算法选取结果和标准算法选取结果,对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型,本发明实施例训练得到的算法选取模型能够预测出最适合当前语音场景的回声消除算法,解决了不同的回声消除算法所适用的语音场景不同的问题,保证了回声消除算法与当前语音场景的适配性,进而提高了回声消除后的语音质量。

[0137] 在上述实施例的基础上,具体的,初始算法选取模型包括特征提取模块、拼接模块、自注意力模块、时间池化模块和输出层,其中,特征提取模块,用于对训练远端声音数据和训练消除数据集中的至少两个训练消除数据分别进行特征提取,得到训练远端语音特征和至少两个训练消除特征,拼接模块,用于对特征提取模块输出的训练远端语音特征和各

训练消除特征进行拼接处理，得到拼接语音特征，自注意力模块，用于基于拼接模块输出的拼接语音特征，确定更新语音特征，时间池化模块，用于对自注意力模块输出的更新语音特征进行压缩，得到压缩语音特征，输出层，用于基于时间池化模块输出的压缩语音特征，输出预测算法选取结果。

[0138] 在上述实施例的基础上，具体的，该装置还包括：

[0139] 标准算法选取结果确定模块，用于采用预设评估算法，基于训练消除数据集中的至少一个训练消除数据，确定各预设回声消除算法分别对应的算法评估结果；

[0140] 基于各算法评估结果，确定初始算法选取模型对应的标准算法选取结果。

[0141] 在上述实施例的基础上，具体的，目标算法选取模型确定模块440，具体用于：

[0142] 基于预测算法选取结果和标准算法选取结果，构建损失函数；

[0143] 基于损失函数对初始算法选取模型的模型参数进行调整，得到调整后的初始算法选取模型；

[0144] 基于调整后的初始算法选取模型，返回执行将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果的步骤；

[0145] 直到损失函数收敛时，将调整后的初始算法选取模型作为训练完成的目标算法选取模型。

[0146] 本发明实施例所提供的算法选取模型的训练装置可执行本发明任意实施例所提供的算法选取模型的训练方法，具备执行方法相应的功能模块和有益效果。

[0147] 图9为本发明一个实施例所提供的一种回声消除装置的结构示意图。如图9所示，该装置包括：待测声音数据获取模块510、第一待测消除数据集确定模块520、目标回声消除算法确定模块530、第二待测消除数据确定模块540和回声消除数据确定模块550。

[0148] 其中，待测声音数据获取模块510，用于获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据；其中，待测近端声音数据包括第一待测近端声音数据和第二待测近端声音数据，待测远端声音数据包括第一待测远端声音数据和第二待测远端声音数据；

[0149] 第一待测消除数据集确定模块520，用于采用回声消除算法集中的至少两个预设回声消除算法，基于第一待测远端声音数据，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集；

[0150] 目标回声消除算法确定模块530，用于基于第一待测远端声音数据、第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法；

[0151] 第二待测消除数据确定模块540，用于采用目标回声消除算法，基于第二待测远端声音数据，对第二待测近端声音数据进行回声消除处理得到第二待测消除数据；

[0152] 回声消除数据确定模块550，用于基于目标回声消除算法、第二待测消除数据和第一待测消除数据集，确定待测近端声音数据对应的回声消除数据；

[0153] 其中，目标算法选取模型是采用本发明上述实施例提供的算法选取模型的训练方法得到的。

[0154] 本实施例的技术方案，通过采用回声消除算法集中的至少两个预设回声消除算法，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集，采用预先训练完成的目标算法选取模型预测目标回声消除算法，采用目标回声消除算法，基于第二

待测远端声音数据，对第二待测近端声音数据进行回声消除处理得到第二待测消除数据，本发明实施例采用算法选取模型确定最适合当前语音场景的回声消除算法，解决了不同的回声消除算法所适用的语音场景不同的问题，保证了回声消除算法与当前语音场景的适配性，进而提高了回声消除后的语音质量。

[0155] 在上述实施例的基础上，具体的，待测声音数据获取模块510，具体用于：

[0156] 响应于检测到消除触发指令，基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段；其中，第一预设语音时长小于第二预设语音时长；

[0157] 获取近端麦克风采集的与第一待测声音时段和第二待测声音时段分别对应的第一待测近端声音数据和第二待测近端声音数据；

[0158] 获取近端播放器播放的与第一待测声音时段和第二待测声音时段分别对应的第一待测远端声音数据和第二待测远端声音数据。

[0159] 在上述实施例的基础上，具体的，该装置还包括：

[0160] 消除触发指令生成模块，用于响应于检测到语音开始指令，在近端麦克风采集到的声音数据中存在语音数据的情况下，生成消除触发指令；和/或，

[0161] 在当前采集时刻超出第二待测声音时段的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤；和/或，

[0162] 响应于检测到静音时段，在静音时段对应的静音时长超过静音时长阈值的情况下，生成消除触发指令，并返回执行基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段的步骤。

[0163] 在上述实施例的基础上，具体的，回声消除算法集中包含回声消除模型，回声消除模型中包含输入模块、块内模块、转换模块、块外模块和输出模块，其中，输入模块，用于基于输入的第一待测远端声音数据和第一待测近端声音数据，确定第一待测语音特征，块内模块，用于对输入模块输出的第一待测语音特征进行频域特征处理，输出第一频域语音特征，转换模块，用于对块内模块输出的第一频域语音特征进行转换，得到第一转换语音特征，块外模块，用于对转换模块输出的第一转换语音特征进行时域特征处理，得到第一时域语音特征，输出模块，用于基于块外模块输出的第一时域语音特征和第一待测近端声音数据，输出第一待测消除数据。

[0164] 在上述实施例的基础上，具体的，块内模块包括自注意力网络，块外模块包括长短时记忆网络。

[0165] 在上述实施例的基础上，具体的，输入模块包括串联的预处理层、拼接层和卷积网络，其中，预处理层，用于对输入的第一待测远端声音数据和第一待测近端声音数据分别进行短时傅里叶变换，得到第一待测远端语音频谱和第一待测近端语音频谱，拼接层，用于对预处理层输出的第一待测远端语音频谱和第一待测近端语音频谱进行拼接，得到第一待测拼接语音频谱，卷积网络，用于对拼接层输出的第一待测拼接语音频谱进行特征提取，输出第一待测语音特征。

[0166] 在上述实施例的基础上，具体的，输出模块包括串联的反卷积网络、融合单元和短时傅里叶逆变换层，其中，反卷积网络，用于对块外模块输出的第一时域语音特征进行反卷

积处理，得到第一反卷积语音特征，融合单元，用于对第一待测近端声音数据和反卷积网络输出的第一反卷积语音特征进行融合处理，得到第一待测融合语音特征，短时傅里叶逆变换层，用于对融合单元输出的第一待测融合语音特征进行短时傅里叶逆变换，输出第一待测消除数据。

[0167] 本发明实施例所提供的回声消除装置可执行本发明任意实施例所提供的回声消除方法，具备执行方法相应的功能模块和有益效果。

[0168] 图10为本发明一个实施例所提供的一种电子设备的结构示意图。电子设备10旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备(如头盔、眼镜、手表等)和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本发明的实现。

[0169] 如图10所示，电子设备10包括至少一个处理器11，以及与至少一个处理器11通信连接的存储器，如只读存储器(ROM)12、随机访问存储器(RAM)13等，其中，存储器存储有可被至少一个处理器执行的计算机程序，处理器11可以根据存储在只读存储器(ROM)12中的计算机程序或者从存储单元18加载到随机访问存储器(RAM)13中的计算机程序，来执行各种适当的动作和处理。在RAM 13中，还可存储电子设备10操作所需的各种程序和数据。处理器11、ROM 12以及RAM 13通过总线14彼此相连。输入/输出(I/O)接口15也连接至总线14。

[0170] 电子设备10中的多个部件连接至I/O接口15，包括：输入单元16，例如键盘、鼠标等；输出单元17，例如各种类型的显示器、扬声器等；存储单元18，例如磁盘、光盘等；以及通信单元19，例如网卡、调制解调器、无线通信收发机等。通信单元19允许电子设备10通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0171] 处理器11可以是各种具有处理和计算能力的通用和/或专用处理组件。处理器11的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的处理器、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。处理器11执行上文所描述的各个方法和处理，例如上述实施例提供的算法选取模型的训练方法，和/或，回声消除方法。

[0172] 在一些实施例中，上述实施例提供的算法选取模型的训练方法，和/或，回声消除方法可被实现为计算机程序，其被有形地包含于计算机可读存储介质，例如存储单元18。在一些实施例中，计算机程序的部分或者全部可以经由ROM 12和/或通信单元19而被载入和/或安装到电子设备10上。当计算机程序加载到RAM 13并由处理器11执行时，可以执行上文描述的算法选取模型的训练方法，和/或，回声消除方法中的一个或多个步骤。备选地，在其他实施例中，处理器11可以通过其他任何适当的方式(例如，借助于固件)而被配置为执行算法选取模型的训练方法，和/或，回声消除方法。

[0173] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器

可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0174] 用于实施本发明的方法的计算机程序可以采用一个或多个编程语言的任何组合来编写。这些计算机程序可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器，使得计算机程序当由处理器执行时使流程图和/或框图中所规定的功能/操作被实施。计算机程序可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0175] 在本发明的上下文中，计算机可读存储介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的计算机程序。计算机可读存储介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。备选地，计算机可读存储介质可以是机器可读信号介质。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0176] 为了提供与用户的交互，可以在电子设备上实施此处描述的系统和技术，该电子设备具有：用于向用户显示信息的显示装置(例如，CRT(阴极射线管)或者LCD(液晶显示器)监视器)；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给电子设备。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0177] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)、区块链网络和互联网。

[0178] 计算系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器，又称为云计算服务器或云主机，是云计算服务体系中的一项主机产品，以解决了传统物理主机与VPS服务中，存在的管理难度大，业务扩展性弱的缺陷。

[0179] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本发明中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本发明的技术方案所期望的结果，本文在此不进行限制。

[0180] 上述具体实施方式，并不构成对本发明保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本发明

的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

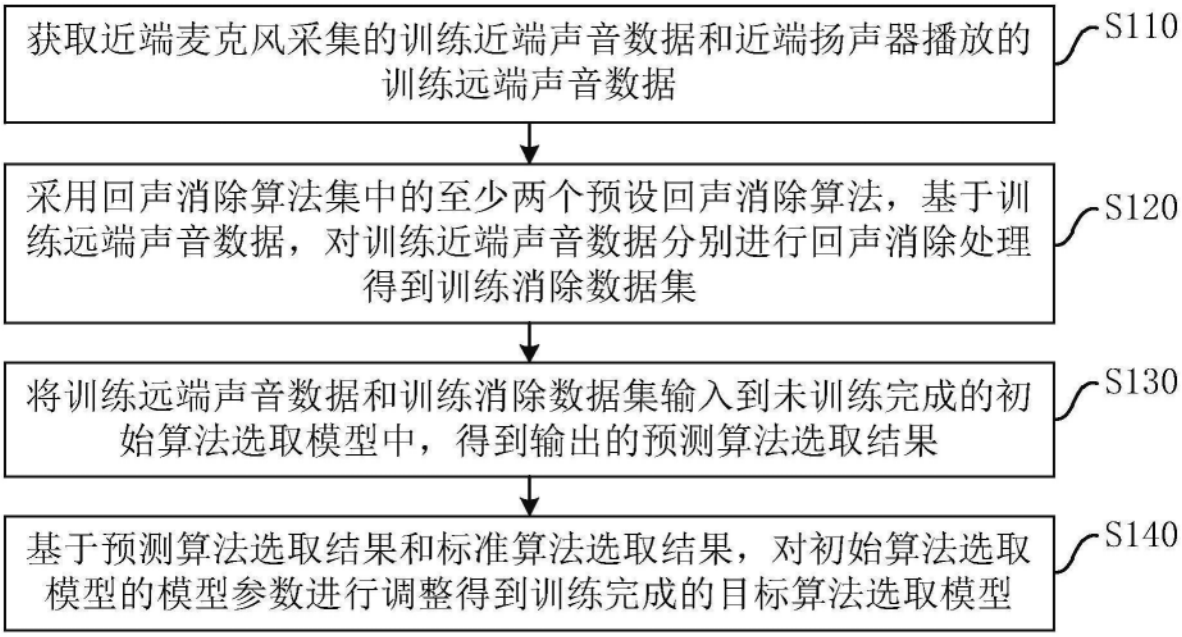

获取近端麦克风采集的训练近端声音数据和近端扬声器播放的训练远端声音数据
S110
采用回声消除算法集中的至少两个预设回声消除算法，基于训练远端声音数据，对训练近端声音数据分别进行回声消除处理得到训练消除数据集
S120
将训练远端声音数据和训练消除数据集输入到未训练完成的初始算法选取模型中，得到输出的预测算法选取结果
S130
基于预测算法选取结果和标准算法选取结果，对初始算法选取模型的模型参数进行调整得到训练完成的目标算法选取模型
S140

图1

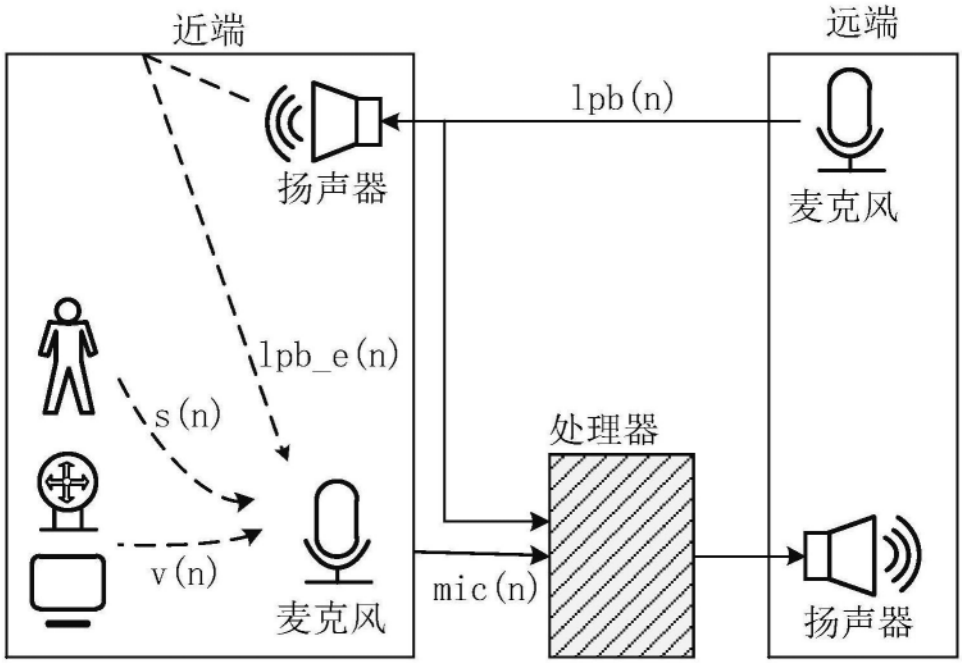

近端
远端
lpb(n)
扬声器
麦克风
lpb_e(n)
s(n)
处理器
v(n)
mic(n)
麦克风
扬声器

图2

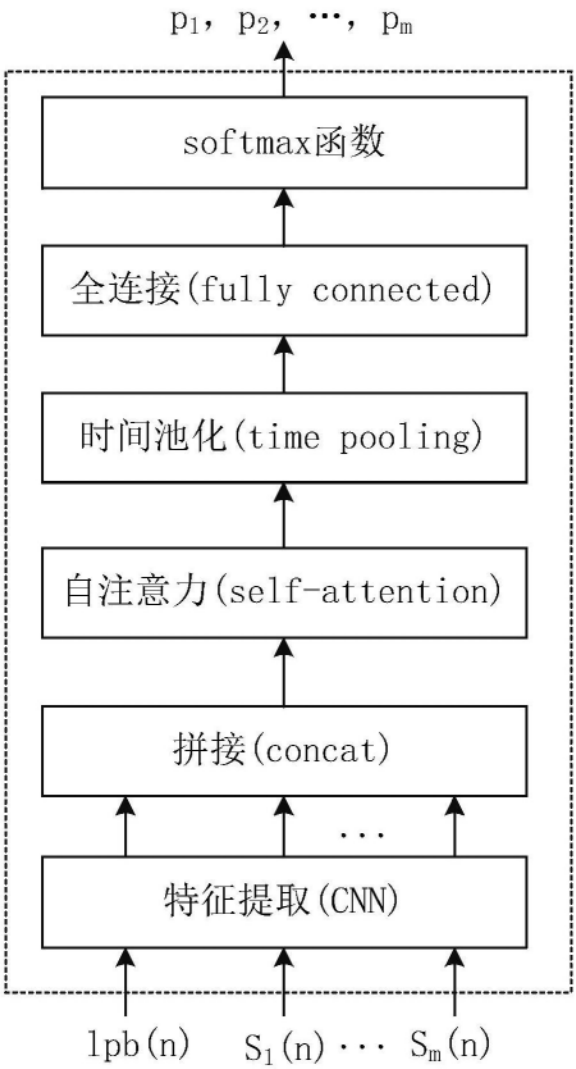

p_1, p_2, …, p_m
softmax函数
全连接(fully connected)
时间池化(time pooling)
自注意力(self-attention)
拼接(concat)
…
特征提取(CNN)
lpb(n)
S_1(n) … S_m(n)

图3

获取近端麦克风采集的待测近端声音数据以及近端扬声器播放的待测远端声音数据
S210
采用回声消除算法集中的至少两个预设回声消除算法，基于第一待测远端声音数据，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集
S220
基于第一待测远端声音数据、第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法
S230
采用目标回声消除算法，基于第二待测远端声音数据，对第二待测近端声音数据进行回声消除处理得到第二待测消除数据
S240
基于目标回声消除算法、第二待测消除数据和第一待测消除数据集，确定待测近端声音数据对应的回声消除数据
S250

图4

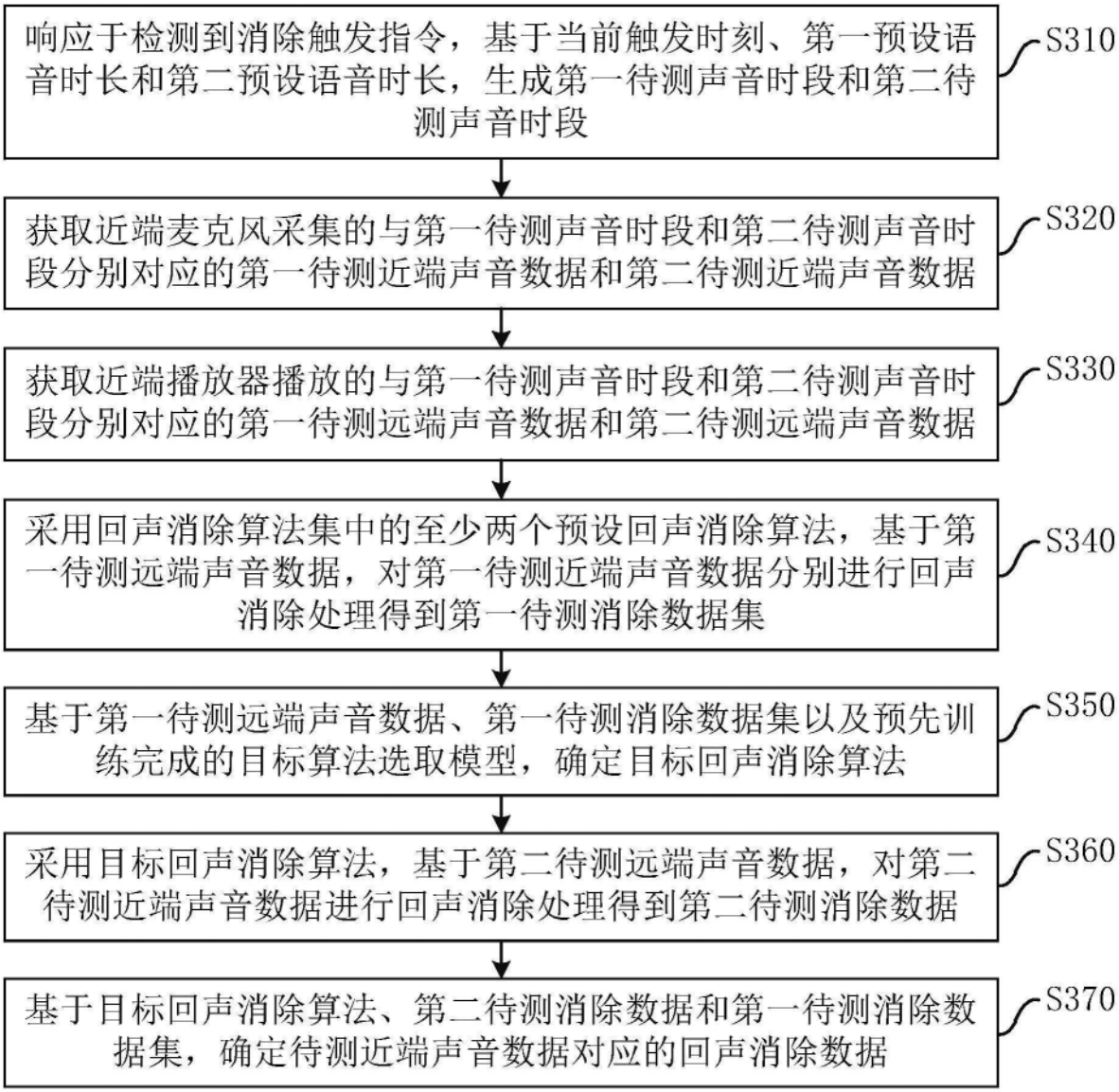

响应于检测到消除触发指令，基于当前触发时刻、第一预设语音时长和第二预设语音时长，生成第一待测声音时段和第二待测声音时段
S310
获取近端麦克风采集的与第一待测声音时段和第二待测声音时段分别对应的第一待测近端声音数据和第二待测近端声音数据
S320
获取近端播放器播放的与第一待测声音时段和第二待测声音时段分别对应的第一待测远端声音数据和第二待测远端声音数据
S330
采用回声消除算法集中的至少两个预设回声消除算法，基于第一待测远端声音数据，对第一待测近端声音数据分别进行回声消除处理得到第一待测消除数据集
S340
基于第一待测远端声音数据、第一待测消除数据集以及预先训练完成的目标算法选取模型，确定目标回声消除算法
S350
采用目标回声消除算法，基于第二待测远端声音数据，对第二待测近端声音数据进行回声消除处理得到第二待测消除数据
S360
基于目标回声消除算法、第二待测消除数据和第一待测消除数据集，确定待测近端声音数据对应的回声消除数据
S370

图5

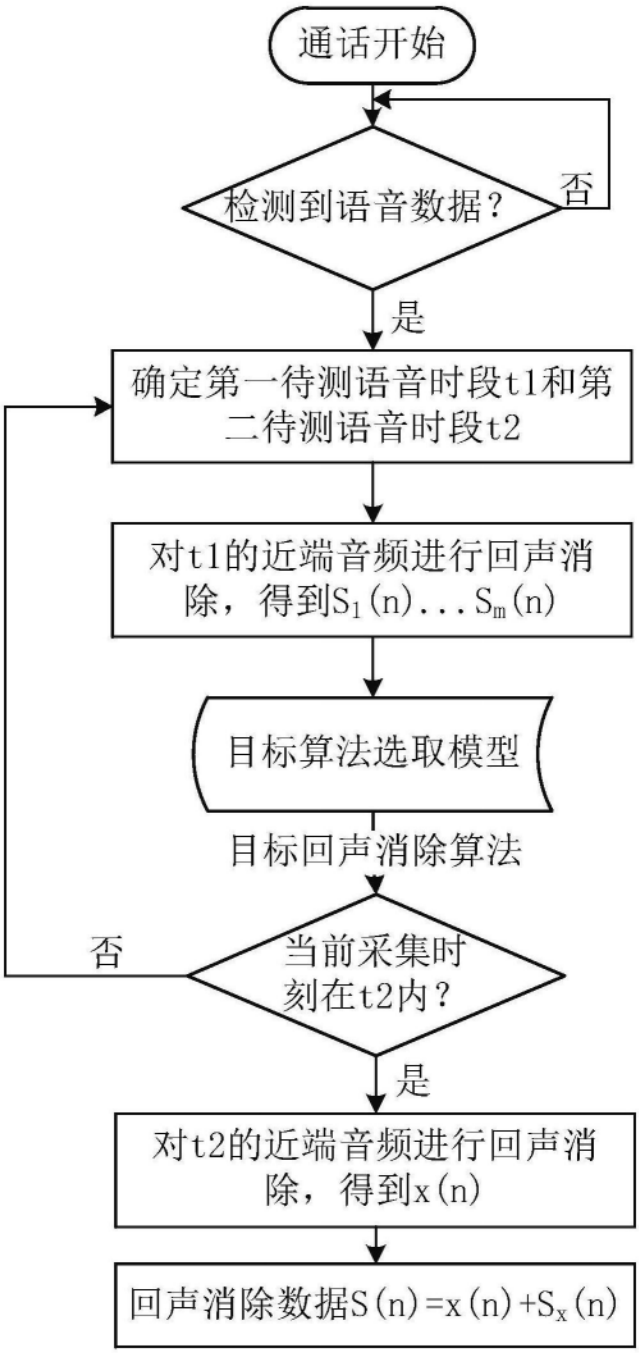

通话开始
检测到语音数据？
否
是
确定第一待测语音时段t1和第二待测语音时段t2
对t1的近端音频进行回声消除，得到$S_1(n)...S_m(n)$
目标算法选取模型
目标回声消除算法
否
当前采集时刻在t2内？
是
对t2的近端音频进行回声消除，得到x(n)
回声消除数据$S(n)=x(n)+S_x(n)$

图6

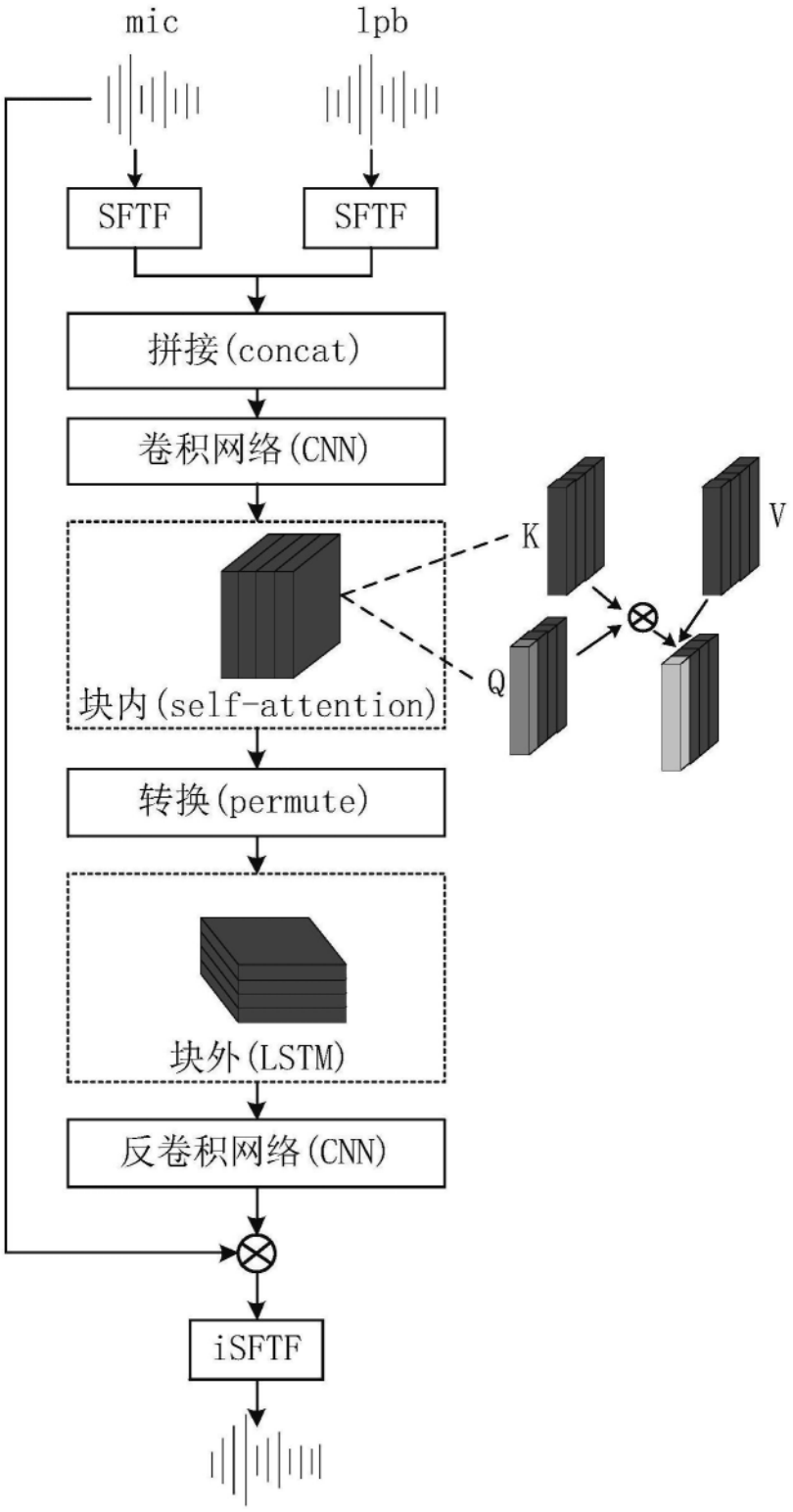

mic
lpb
SFTF
SFTF
拼接(concat)
卷积网络(CNN)
K
V
Q
块内(self-attention)
转换(permute)
块外(LSTM)
反卷积网络(CNN)
iSFTF

图7

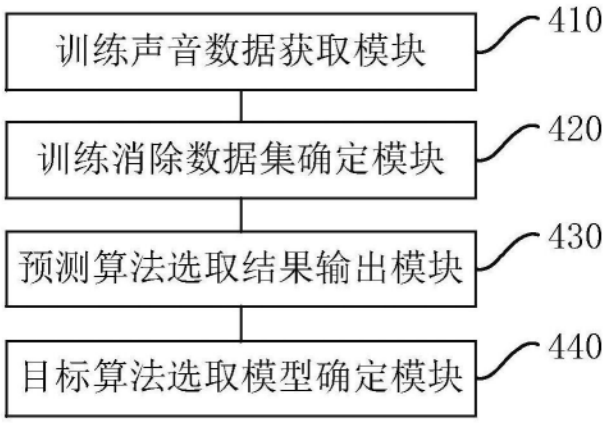

训练声音数据获取模块
410
训练消除数据集确定模块
420
预测算法选取结果输出模块
430
目标算法选取模型确定模块
440

图8

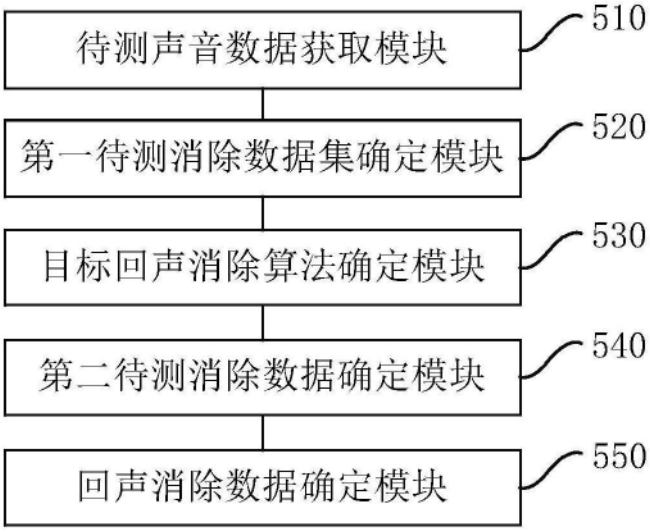

待测声音数据获取模块
510
第一待测消除数据集确定模块
520
目标回声消除算法确定模块
530
第二待测消除数据确定模块
540
回声消除数据确定模块
550

图9

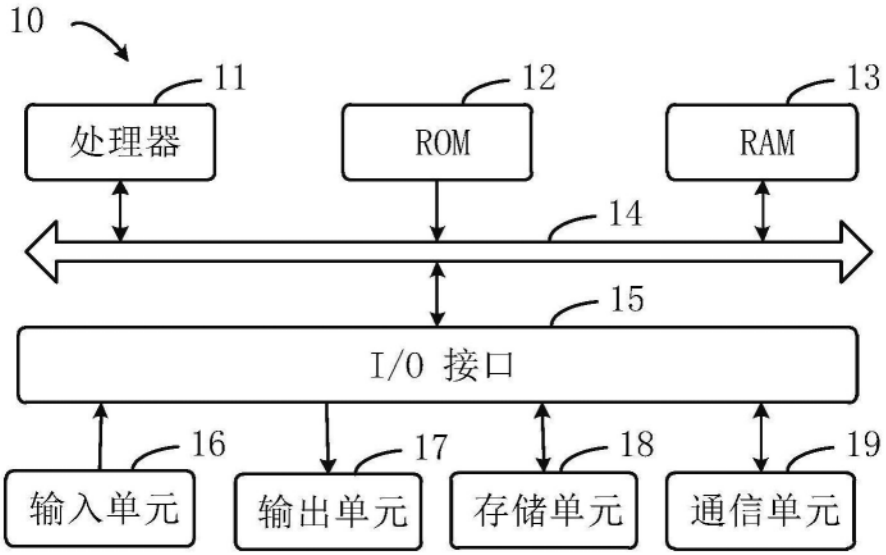


图10