



(12) 发明专利申请

(10) 申请公布号 CN 116129930 A

(43) 申请公布日 2023. 05. 16

(21) 申请号 202310121538.7

(22) 申请日 2023.02.15

(71) 申请人 乐鑫信息科技(上海)股份有限公司
地址 201203 上海市浦东新区自由贸易试
验区碧波路690号2号楼204室

(72) 发明人 沈小正

(74) 专利代理机构 北京市路盛律师事务所
11326
专利代理师 李海 王海凤

(51) Int. Cl.

G10L 21/0272 (2013.01)

G10L 21/0216 (2013.01)

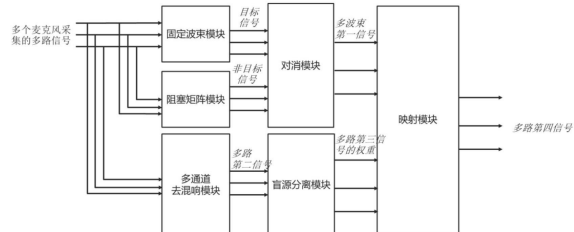
权利要求书3页 说明书11页 附图3页

(54) 发明名称

无参考回路的回声消除装置及方法

(57) 摘要

本发明提供一种无参考回路的回声消除装置,包括固定波束模块,其将语音采集单元采集的多路信号固定为多个第一波束,将多个第一波束相叠加并输出目标信号;阻塞矩阵模块,其将语音采集单元采集的多路信号输入至阻塞矩阵进行预处理以输出非目标信号;对消模块,其将目标信号和非目标信号进行对消并输出多波束第一信号;多通道去混响模块,其将语音采集单元采集的多路信号进行去混响,并输出多路第二信号;盲源分离模块,其将多路第二信号进行盲源分离,以获得多路第三信号,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定每路第三信号的权重;以及映射模块,其将权重映射至多波束第一信号,并输出映射后的多路频域第四信号。



1. 一种无参考回路的回声消除装置,其特征在于,包括:

固定波束模块,其配置为将语音采集单元采集的多路信号固定为多个第一波束,将所述多个第一波束相叠加并输出目标信号;

阻塞矩阵模块,其配置为将所述语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号;

对消模块,其配置为将所述目标信号和所述非目标信号进行对消并输出多波束第一信号;

多通道去混响模块,其配置为将所述语音采集单元采集的多路信号进行去混响,并输出多路第二信号;

盲源分离模块,其配置为将所述多路第二信号进行盲源分离,以获得多路第三信号,分别针对所述多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定所述多路第三信号中的每路第三信号的权重;以及

映射模块,其配置为将所述权重映射至所述多波束第一信号,并输出映射后的多路频域第四信号。

2. 如权利要求1所述的回声消除装置,其特征在于,其中,所述对消模块将目标信号和非目标信号进行对消并输出多波束第一信号包括:

(1) 通过如下公式对所述目标信号和所述非目标信号进行对消:

$$ERR = MIC - w * REF,$$

其中,ERR为残差信号,MIC为所述目标信号,w为滤波器参数,REF为所述非目标信号;

(2) 所述残差信号ERR包括K个单波束第一信号 B_1, B_2, \dots 直至 B_K ,分别对所述残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对所述分帧中的每一个进行傅里叶变换,以获得频域上的所述多波束第一信号 $B_{ktf}^{(1)}$,其中,k为波束的编号, $k=1, 2, \dots, K$,K为所述目标信号中的波束数目,t为对应的分帧的帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

3. 如权利要求2所述的回声消除装置,其特征在于,其中,所述盲源分离模块通过如下公式进行信噪比计算,以获得所述多路第三信号中的每路第三信号的信噪比 SNR_{ntf} :

$$SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2$$

其中, S_{ntf} 为经过多通道去混响模块在频域上输出的所述多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对所述多路第二信号进行盲源分离所获得的所述多路第三信号,其中,n为麦克风编号, $n=1, 2, \dots, N$,N为所述语音采集单元中的麦克风的数量,t为对应的分帧的帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

4. 如权利要求3所述的回声消除装置,其特征在于,其中,所述盲源分离模块通过如下公式,以确定所述多路第三信号中的每路第三信号的权重 G_{ntf} :

$$G_{ntf} = SNR_{ntf} / (1 + SNR_{ntf})$$

其中, SNR_{ntf} 为所述每路第三信号的信噪比,n为麦克风编号, $n=1, 2, \dots, N$,N为所述语音采集单元中的麦克风的数量,t为对应的分帧的帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

5. 如权利要求4所述的回声消除装置,其特征在于,其中,所述映射模块通过如下公式分别将N组所述权重 G_{ntf} 映射到K组所述多波束第一信号 $B_{ktf}^{(1)}$ 上,以得到映射后的所述多路频域第四信号E,其中:

$$E_{mtf} = G_{ntf} * B_{ktf}^{(1)},$$

其中, E_{ntf} 为所述多路频域第四信号中的第m路频域第四信号,其中 $m=1,2,\dots,M$, $M=K*N$,K为所述目标信号中的波束数目,N为所述语音采集单元中的麦克风的数量; G_{ntf} 为所述多路第三信号中的第n路第三信号的权重, $n=1,2,\dots,N$; $B_{ktf}^{(1)}$ 为所述多波束第一信号中的第k个单波束第一信号,k为波束的编号, $k=1,2,\dots,K$;t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$;

所述映射模块还配置为对所述多路频域第四信号进行傅里叶逆变换运算,以获得多路时域第四信号 e_m ,其中 $m=1,2,\dots,M$ 。

6. 如权利要求5所述的回声消除装置,其特征在于,其中,还包括:

唤醒引擎,其配置为对所述多路时域第四信号中的每路时域第四信号打分以分别获得得分,确定其中所述得分大于唤醒阈值的Z路时域第四信号,并确定所述Z路时域第四信号中能量最大的一路时域第四信号,其中,Z大于等于1;所述唤醒引擎还配置为输出所述能量最大的一路时域第四信号;以及

识别引擎,其配置为从所述唤醒引擎获取所述能量最大的一路时域第四信号,以进行语音识别并输出识别的语音。

7. 一种无参考回路的回声消除方法,其特征在于,其包括下述步骤:

将语音采集单元采集的多路信号固定为多个第一波束,将所述多个第一波束相叠加并输出目标信号;

将所述语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号;

将所述目标信号和所述非目标信号进行对消并输出多波束第一信号;

将所述语音采集单元采集的多路信号进行去混响,并输出多路第二信号;

将所述多路第二信号进行盲源分离,以获得多路第三信号,分别针对所述多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定所述多路第三信号中的每路第三信号的权重;以及

将所述权重映射至所述多波束第一信号,并输出映射后的多路频域第四信号。

8. 如权利要求7所述的回声消除方法,其特征在于,其中,所述将目标信号和非目标信号进行对消并输出多波束第一信号包括:

(1) 通过如下公式对所述目标信号和所述非目标信号进行对消:

$$ERR = MIC - w * REF,$$

其中,ERR为残差信号,MIC为所述目标信号,w为滤波器参数,REF为所述非目标信号;

所述残差信号ERR包括K个单波束第一信号 B_1, B_2, \dots 直至 B_K ,分别对所述残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对所述分帧中的每一个进行傅里叶变换,以获得频域上的所述多波束第一信号 $B_{ktf}^{(1)}$,其中,k为波束的编号, $k=1,2,\dots,K$,K

为所述目标信号中的波束数目, t 为对应的分帧的帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$ 。

9. 如权利要求8所述的回声消除方法, 其特征在于, 其中, 通过如下公式进行信噪比计算, 以获得所述多路第三信号中的每路第三信号的信噪比 SNR_{ntf} :

$$SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2,$$

其中, S_{ntf} 为经过多通道去混响模块在频域上输出的所述多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对所述多路第二信号进行盲源分离所获得的所述多路第三信号, 其中, n 为麦克风编号, $n=1, 2, \dots, N$, N 为所述语音采集单元中的麦克风的数目, t 为帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$ 。

10. 如权利要求9所述的回声消除方法, 其特征在于, 其中, 通过如下公式, 以确定所述多路第三信号中的每路第三信号的权重 G_{ntf} :

$$G_{ntf} = SNR_{ntf} / (1 + SNR_{ntf}),$$

其中, SNR_{ntf} 为所述每路第三信号的信噪比, n 为麦克风编号, $n=1, 2, \dots, N$, N 为所述语音采集单元中的麦克风的数目, t 为对应的分帧的帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$ 。

11. 如权利要求10所述的回声消除方法, 其特征在于, 其中, 通过如下公式分别将 N 组所述权重 G_{ntf} 映射到 K 组所述多波束第一信号 $B_{ktf}^{(1)}$ 上, 以得到映射后的所述多路频域第四信号 E , 其中:

$$E_{mtf} = G_{ntf} * B_{ktf}^{(1)},$$

其中, E_{mtf} 为所述多路频域第四信号中的第 m 路频域第四信号, 其中 $m=1, 2, \dots, M$, $M=K*N$, K 为所述目标信号中的波束数目, N 为所述语音采集单元中的麦克风的数目; G_{ntf} 为所述多路第三信号中的第 n 路第三信号的权重, $n=1, 2, \dots, N$; $B_{ktf}^{(1)}$ 为所述多波束第一信号中的第 k 个单波束第一信号, k 为波束的编号, $k=1, 2, \dots, K$; t 为对应的分帧的帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$;

对所述多路频域第四信号进行傅里叶逆变换运算, 以获得多路时域第四信号 e_m , 其中 $m=1, 2, \dots, M$ 。

12. 如权利要求11所述的回声消除方法, 其特征在于, 其中, 在所述输出映射后的多路时域第四信号之后, 还包括:

对所述多路时域第四信号中的每路时域第四信号打分以分别获得得分, 确定其中所述得分大于唤醒阈值的 Z 路时域第四信号, 并确定所述 Z 路时域第四信号中能量最大的一路时域第四信号, 其中, Z 大于等于1; 以及, 输出所述能量最大的一路时域第四信号; 以及

对所述输出的能量最大的一路时域第四信号进行语音识别并输出识别的语音。

无参考回路的回声消除装置及方法

技术领域

[0001] 本发明涉及远场语音交互领域,尤其涉及一种无参考回路的回声消除装置及方法。

背景技术

[0002] 近年来远场语音交互大幅提升了家电、车机、售票机的智能化程度,语音交互也是最自然的交互方式。为了获得更好的居家办公效率,会议系统被快速地部署在更多的智能设备上。无论是远场语音交互还是会议系统,回声消除技术都是其中的核心算法模块,其作用在于解决语音交互设备播放时的打断问题,以及会议系统本会场的声音传到对方会场后,喇叭播放出来之后再次进入麦克风传回本会场的问题等。回声消除一般是通过自适应滤波算法实现对回声路径的估计,然后在麦克风获得带噪语音中减去目标回声,使得语音交互更加高效、会议更加真实。

[0003] 由于智能设备自身的喇叭播放出来的声音重新传回麦克风,从而会导致用户发出的语音信号如指令等不能被清晰准确地加以识别。针对这一问题,现有技术中通常采用带有参考回路的技术方案来消除回声,即通过参考回路采集喇叭发出的参考信号,并基于此对麦克风采集到的语音信号进行回声消除。例如,中国专利CN213211700U公开了一种回声消除装置,包括:控制单元、音频信号处理单元、音频播放单元、回声消除单元、参考信号采集单元和语音信号采集单元、模数装换单元。其中,参考信号采集单元设置在音频播放单元预设距离范围内,通过将回采到的参考信号和语音信号采集单元获取的目标说话人的语音信号共同输入到回声消除单元进行回声消除。这种方式一方面依赖于参考回路,另一方面还存在容易削弱目标说话人语音的问题,尤其是在设备的音频播放单元不处于播放状态时,参考信号采集单元获得的信号主要来自目标说话人,因此,这种状态下会一定程度降低目标说话人的语音信号的清晰度,甚至出现回声消除模块完全抑制目标人声的问题。

[0004] 中国专利CN209962694U公开了一种回声消除的电路及电声设备,包括功放模块、扬声器、麦克风、回声消除模块以及滤波电路。其中,滤波电路用于从功放模块采集语音参考信号,并对其中的高频率噪音进行过滤。麦克风接收扬声器发出的回声信号和用户发出的语音信号的混合语音信号。进一步的,回声消除模块根据从滤波电路采集的语音参考信号和麦克风处采集的混合语音信号进行回声消除。这一技术方案主要通过滤波电路降低了硬件获得的参考信号的噪声,以提升回声消除的效果。然而,尽管滤波电路能够在一定程度上解决目标人声被抑制的问题,但是由于功放模块和喇叭的小型化与廉价化,滤波电路获得的回路参考信号与实际产生的非线性回声差异巨大,回声消除算法的性能大幅下降。

[0005] 中国专利CN104822001B公开了一种回声消除数据同步控制的方法和装置,包括:估计声卡延迟值;等待参考音频缓冲区队列长度近端音频缓冲区队列长度的差值大于或等于声卡延迟值所对应的音频数据长度;按音频帧从参考音频缓冲区队列和近端音频缓冲区队列队首取出数据进行回声消除;获取进行回声消除处理所生成的相对延迟值;根据所述相对延迟值调整所述声卡延迟。该方案的核心在于获得硬件参考回路与麦克风获取音

频数据的时延,通过延迟估计的方法来解决时钟抖动带来的不同步影响回声消除效果的问题。但是在外部噪声的较大的时候,很容易造成估计不准确。

[0006] 综上所述,由此可见目前的主流技术方案仍然依赖于参考回路采集参考,以对语音采集模块采集的混合人声进行回声消除,其回声消除的性能有待提高。

发明内容

[0007] 本发明针对以上问题,本申请提供一种无参考回路的回声消除装置及方法。通过麦克风阵列声学结构的设计以及算法的创新提出一种进行回声消除的新方法,本发明不需要对参考信号进行延迟估计,并且能够很好的处理非线性回声,同时也不需要进行复杂的参考回路设计。

[0008] 根据本发明的第一方面,提供一种无参考回路的回声消除装置,包括:固定波束模块,其配置为将语音采集单元采集的多路信号固定为多个第一波束,将多个第一波束相叠加并输出目标信号;阻塞矩阵模块,其配置为将语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号;对消模块,其配置为将基于语音采集单元采集的多路信号而生成的目标信号和非目标信号进行对消并输出多波束第一信号;多通道去混响模块,其配置为将语音采集单元采集的多路信号进行去混响,并输出多路第二信号;盲源分离模块,其配置为将多路第二信号进行盲源分离,输出多路第三信号,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定多路第三信号中的每路第三信号的权重;以及映射模块,其配置为将权重映射至多波束第一信号,并输出映射后的多路频域第四信号。

[0009] 作为本发明的一个实施例,对消模块将目标信号和非目标信号进行对消并输出多波束第一信号包括:(1)通过如下公式对目标信号和非目标信号进行对消: $ERR = MIC - w * REF$,其中,ERR为残差信号,NIC为目标信号,w为滤波器参数,REF为非目标信号;(2)残差信号ERR包括K个单波束第一信号 B_1 、 B_2 、 \dots 直至 B_K ,分别对残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对分帧中的每一个进行傅里叶变换,以获得频域上的多波束第一信号 $B_{ktf}^{(1)}$,其中,k为波束的编号, $k=1,2,\dots,K$,K为目标信号中的波束数目,t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$ 。

[0010] 作为本发明的一个实施例,盲源分离模块通过如下公式进行信噪比计算,以获得多路第三信号中的每路第三信号的信噪比 SNR_{ntf} : $SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2$,其中, S_{ntf} 为经过多通道去混响模块在频域上输出的多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对多路第二信号进行盲源分离所获得的多路第三信号,其中,n为麦克风编号, $n=1,2,\dots,N$,N为语音采集单元中的麦克风的数目,t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$ 。

[0011] 作为本发明的一个实施例,盲源分离模块通过如下公式,以确定多路第三信号中的每路第三信号的权重 G_{ntf} : $G_{ntf} = SNR_{ntf} / (1 + SNR_{ntf})$,其中, SNR_{ntf} 为每路第三信号的信噪比,n为麦克风编号, $n=1,2,\dots,N$,N为语音采集单元中的麦克风的数目,t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$ 。

[0012] 作为本发明的一个实施例,映射模块通过如下公式分别将N组权重 G_{ntf} 映射到K组多波束第一信号 $B_{ktf}^{(1)}$ 上,以得到映射后的多路频域第四信号E,其中:

$E_{mtf} = G_{ntf} * B_{ktf}^{(1)}$,其中, E_{mtf} 为多路频域第四信号中的第m路频域第四信号,其中 $m=1,2,\dots,M,M=K*N$,K为目标信号中的波束数目,N为语音采集单元中的麦克风的数目; G_{ntf} 为多路第三信号中的第n路第三信号的权重, $n=1,2,\dots,N$; $B_{ktf}^{(1)}$ 为多波束第一信号中的第k个单波束第一信号,k为波束的编号, $k=1,2,\dots,K$;t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$;映射模块还配置为对多路频域第四信号进行傅里叶逆变换运算,以获得多路时域第四信号 e_m ,其中 $m=1,2,\dots,M$ 。

[0013] 作为本发明的一个实施例,还包括:唤醒引擎,其配置为对多路时域第四信号中的每路时域第四信号打分以分别获得得分,确定其中得分大于唤醒阈值的Z路时域第四信号,并确定Z路时域第四信号中能量最大的一路时域第四信号,其中,Z大于等于1;唤醒引擎还配置为输出能量最大的一路时域第四信号;识别引擎,其配置为从唤醒引擎获取能量最大的一路时域第四信号,以进行语音识别并输出识别的语音。

[0014] 根据本发明的第二方面,还提供了一种无参考回路的回声消除方法,其包括下述步骤:将语音采集单元采集的多路信号固定为多个第一波束,将多个第一波束相叠加并输出目标信号;将语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号;将目标信号和非目标信号进行对消并输出多波束第一信号;将语音采集单元采集的多路信号进行去混响,并输出多路第二信号;将多路第二信号进行盲源分离,以获得多路第三信号,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定多路第三信号中的每路第三信号的权重;以及将权重映射至多波束第一信号,并输出映射后的多路频域第四信号。

[0015] 作为本发明的一个实施例,将目标信号和非目标信号进行对消并输出多波束第一信号,进一步包括:(1)通过如下公式对目标信号和非目标信号进行对消: $ERR = MIC - w * REF$,其中,ERR为残差信号,MIC为目标信号,w为滤波器参数,REF为非目标信号;残差信号ERR包括K个单波束第一信号 B_1, B_2, \dots 直至 B_K ,分别对残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对分帧中的每一个进行傅里叶变换,以获得频域上的多波束第一信号 $B_{ktf}^{(1)}$,其中,k为波束的编号, $k=1,2,\dots,K$,K为目标信号中的波束数目,t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$ 。

[0016] 作为本发明的一个实施例,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,进一步包括:通过如下公式进行信噪比计算,以获得多路第三信号中的每路第三信号的信噪比 SNR_{ntf} : $SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2$,其中, S_{ntf} 为经过多通道去混响模块在频域上输出的多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对多路第二信号进行盲源分离所获得的多路第三信号,其中,n为麦克风编号, $n=1,2,\dots,N$,N为语音采集单元中的麦克风的数目,t为帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,F$ 。

[0017] 作为本发明的一个实施例,确定多路第三信号中的每路第三信号的权重,进一步

包括:通过如下公式,以确定多路第三信号中的每路第三信号的权重 G_{ntf} : $G_{ntf} = \text{SNR}_{ntf} / (1 + \text{SNR}_{ntf})$,其中, SNR_{ntf} 为每路第三信号的信噪比, n 为麦克风编号, $n=1,2,\dots,N$, N 为语音采集单元中的麦克风的数目, t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。

[0018] 作为本发明的一个实施例,将权重映射至多波束第一信号,并输出映射后的多路频域第四信号,进一步包括:通过如下公式分别将 N 组权重 G_{ntf} 映射到 K 组多波束第一信号 $B_{ktf}^{(1)}$ 上,以得到映射后的多路频域第四信号 E ,其中: $E_{mtf} = G_{ntf} * B_{ktf}^{(1)}$,其中, E_{mtf} 为多路频域第四信号中的第 m 路频域第四信号,其中 $m=1,2,\dots,M$, $M=K*N$, K 为目标信号中的波束数目, N 为语音采集单元中的麦克风的数目; G_{ntf} 为多路第三信号中的第 n 路第三信号的权重, $n=1,2,\dots,N$; $B_{ktf}^{(1)}$ 为多波束第一信号中的第 k 个单波束第一信号, k 为波束的编号, $k=1,2,\dots,K$; t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。进一步优选地,在步骤S12中,进一步包括,对多路频域第四信号进行傅里叶逆变换运算,以获得多路时域第四信号 e_m ,其中 $m=1,2,\dots,M$ 。

[0019] 作为本发明的一个实施例,在输出映射后的多路频域第四信号之后,还包括:对多路时域第四信号中的每路时域第四信号打分以分别获得得分,确定其中得分大于唤醒阈值的 Z 路时域第四信号,并确定 Z 路时域第四信号中能量最大的一路时域第四信号,其中, Z 大于等于1;以及,输出能量最大的一路时域第四信号;以及,对输出的能量最大的一路时域第四信号进行语音识别并输出识别的语音。

[0020] 本发明利用声学结构中语音采集模块和音频播放模块的空间独立性,应用波束成形的方法,并结合统计学中的盲源分离方法,在不需要从音频播放模块处获取参考信号以及对参考信号进行延迟估计的前提下,即能够很好地消除非线性回声,从而获取清晰的目标人声,实现了无参考回路的回声消除方法。

附图说明

[0021] 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,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性的前提下,还可以根据这些附图获得其他的附图。

[0022] 图1示出根据本发明的一种无参考回路的回声消除装置的示意图;

[0023] 图2示出根据本发明的一个实施例的无参考回路的回声消除方法的流程示意图;

[0024] 图3示出根据本发明的一个实施例的回声消除装置的硬件设计的示意图;

[0025] 图4示出根据本发明的一个具体示例的回声消除装置的示意图;

[0026] 图5A示出根据本发明的一个实施例的原始带噪声数据示意图;

[0027] 图5B示出根据本发明的一个实施例输出的经过回声消除后的干净语音数据示意图。

具体实施方式

[0028] 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。

[0029] 首先介绍本发明的应用场景,本发明主要针对智能家居场景,其中智能家居所处的空间通常较大,容易产生混响声、反射声等回声干扰,而且由于通常环境中未配置有消声材料,因此在实现语音识别时会存在非常大的干扰。例如,与在车载环境中的语音识别相比,智能家居场景中的语音识别将面临更大的技术难度。

[0030] 智能家居设备,例如智能音箱或者智能电视,通常将音频播放单元和语音采集单元设置在各自独立的位置。例如,在智能音箱设备中,喇叭通常设置在音箱主体的中下部朝向底座,且在底座上设置导音锥,以将声波打在导音锥上并向空间中扩散,而作为语音采集单元的环形麦克风阵列则设置在智能音箱的顶部以方便拾音。又例如,智能电视等带屏幕设备,通常将喇叭设置在电视机的侧面,通过多个播放单元制造立体环绕的声学体验,而语音采集单元则设置在屏幕的正前方,以使用户在需要进行语音交互时站在屏幕的前方即可进行远场语音交互。因此,此类智能家居设备在声学设计上满足声源在空间上的独立性,即语音采集单元不易受到音频播放单元的干扰,从而为进行回声消除创造了有利的条件。

[0031] 具体实施例1

[0032] 如图1所示,示出了根据本发明的一种无参考回路的回声消除装置的示意图。其中,该回声消除装置包括固定波束模块、阻塞矩阵模块、对消模块、多通道去混响模块、盲源分离模块以及映射模块。

[0033] 其中,固定波束模块配置为将语音采集单元采集的多路信号固定为多个第一波束,将多个第一波束相叠加并输出目标信号。

[0034] 作为示例而非限制,语音采集单元为麦克风阵列,其包括多个麦克风。应当注意的是,本发明中的语音采集单元可以为单一麦克风阵列。

[0035] 作为示例而非限制,固定波束模块对语音采集单元采集的多路信号(例如多路麦克风信号)进行合并处理,以抑制非目标方向的干扰信号,增强目标方向的声音信号。通过调整每一路麦克风的滤波器系数,并对每一路麦克风的输出信号进行加权求和、滤波,声音信号的各个波束尽可能叠加,在目标说话人方向的信号获得相长干涉,而在另一些非目标说话人的角度的信号则获得相消干涉,最终输出期望方向的语音信号,形成多波束的目标信号。

[0036] 其中,阻塞矩阵模块配置为将语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号。

[0037] 作为示例而非限制,阻塞矩阵用于将语音采集单元采集的多路信号进行阻塞,以得到包含噪声和干扰的非目标信号。

[0038] 其中,对消模块配置为将目标信号和非目标信号进行对消并输出多波束第一信号。

[0039] 优选地,对消模块将目标信号和非目标信号进行对消并输出多波束第一信号,具体包括:(1)通过如下公式对目标信号和非目标信号进行对消: $ERR = MIC - w * REF$,其中,ERR为残差信号,MIC为目标信号,w为滤波器参数,REF为非目标信号;(2)残差信号ERR包括K个单波束第一信号 B_1 、 B_2 、...直至 B_K ,分别对残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对分帧中的每一个进行傅里叶变换,以获得频域上的多波束第一信号

$B_{ktf}^{(1)}$,其中, k 为波束的编号, $k=1,2,\dots,K$, K 为目标信号中的波束数目, t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。

[0040] 作为示例而非限制,目标信号中的波束的数量为预先设定的值。尽管波束的数量越多,最终语音处理的效果越好,但考虑到计算开销,需要根据实际情况选取一个折中的数值作为波束的预先设定数量。

[0041] 作为示例而非限制,可以对分帧中的每一个进行快速傅里叶变换(FFT)。

[0042] 作为示例而非限制,对消模块为基于自适应滤波的自适应噪声相消器。

[0043] 示例性地,可以采用广义旁瓣相消器或者传递函数广义旁瓣相消器实现上述固定波束模块、阻塞矩阵模块以及对消模块。

[0044] 示例性地,可以采用广义旁瓣相消器实现上述固定波束模块、阻塞矩阵模块以及对消模块,其中,上支路由延时求和的固定波束形成器构成,将接收到的信号投影到约束子空间,以期望只有纯净期望语音的目标信号通过;下支路由阻塞矩阵和自适应抵消器组成,将接收到的信号投影到最小方差子空间,以期望只有噪声的非目标信号通过,并在自适应滤波过程中将非目标信号与上支路的目标信号进行对消,以获得多波束第一信号。

[0045] 示例性地,可以采用传递函数旁瓣相消器实现上述固定波束模块、阻塞矩阵模块以及对消模块,其中,固定波束形成器用于接收到的信号分量对齐;阻塞矩阵用于将目标信号进行阻塞以得到噪声的非目标信号,多通道自适应噪声消除器利用噪声的非目标信号对固定波束形成器的输出中的噪声进行消除。

[0046] 其中,多通道去混响模块配置为将语音采集单元采集的多路信号进行去混响,并输出频域上的多路第二信号。

[0047] 作为示例而非限制,多通道去混响模块用于将混响的影响从声音中去除。示例性地,多通道去混响模块可以利用基于统计模型的去混响方法、基于LPC(linear predictive coding)的去混响方法、或者基于特征值分解法的去混响方法。

[0048] 作为示例而非限制,多通道去混响模块输出的多路第二信号表示为 S_{ntf} ,其中 n 为麦克风编号, $n=1,2,\dots,N$, N 为语音采集单元中的麦克风的数目, t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。

[0049] 其中,盲源分离模块配置为将多路第二信号进行盲源分离,以获得多路第三信号,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定多路第三信号中的每路第三信号的权重。作为示例而非限制,本方案采用盲源分离模块作为回声消除装置中的后处理部分,从而达到进一步抑制残余回声的作用。

[0050] 优选地,盲源分离模块通过如下公式进行信噪比计算,以获得多路第三信号中的每路第三信号的信噪比 SNR_{ntf} :

$$[0051] \quad SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2,$$

[0052] 其中, S_{ntf} 为经过多通道去混响模块在频域上输出的多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对多路第二信号进行盲源分离所获得的多路第三信号,其中, n 为麦克风编号, $n=1,2,\dots,N$, N 为语音采集单元中的麦克风的数目, t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。

[0053] 作为示例而非限制,盲源分离模块采用统计方法,对多路第二信号进行盲源分离以获得频域上的多路第三信号。盲源分离模块又可称作BSS(Blind Signal Separation)模块。示例性的,盲源分离模块可以采用的ILRMA(Independent Low-Rank Matrix Analysis)方法、IVA独立向量分析法、ICA独立成分分析法等对多路第二信号进行盲源分离。

[0054] 进一步优选地,盲源分离模块通过如下公式,以确定多路第三信号中的每路第三信号的权重 G_{ntf} :

$$[0055] \quad G_{ntf} = \text{SNR}_{ntf} / (1 + \text{SNR}_{ntf})$$

[0056] 其中, SNR_{ntf} 为多路第三信号中的每路第三信号的信噪比, n 为麦克风编号, $n=1, 2, \dots, N$, N 为语音采集单元中的麦克风的数目, t 为对应的分帧的帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$ 。

[0057] 其中,映射模块配置为将由盲源分离模块输出的权重映射至多波束第一信号,并输出映射后的多路频域第四信号。

[0058] 优选地,映射模块通过如下公式分别将从盲源分离模块输出的 N 组权重 G_{ntf} 映射到从对消模块输出的 K 组多波束第一信号 $B_{ktf}^{(1)}$ 上,以得到映射后的多路频域第四信号 E ,其中:

$$[0059] \quad E_{mtf} = G_{ntf} * B_{ktf}^{(1)},$$

[0060] 其中, E_{mtf} 为多路频域第四信号中的第 m 路频域第四信号,其中 $m=1, 2, \dots, M$, $M=K*N$, K 为目标信号中的波束数目, N 为语音采集单元中的麦克风的数目; G_{ntf} 为多路第三信号中的第 n 路第三信号的权重, $n=1, 2, \dots, N$; $B_{ktf}^{(1)}$ 为多波束第一信号中的第 k 个单波束第一信号, k 为波束的编号, $k=1, 2, \dots, K$; t 为对应的分帧的帧号, $t=1, 2, \dots, T$, f 为频点编号, $f=1, 2, \dots, F$ 。

[0061] 作为示例而非限制,由于盲源分离模块输出的信号会出现幅度缩放问题,例如盲源分离模块的输出可能和原始信号的幅度差异太大,因此映射模块通过将权重 G_{ntf} 映射到多波束第一信号上,以增强输出的频域语音信号,提高最终输出信号的信噪比。

[0062] 优选地,映射模块还配置为对多路频域第四信号进行快速傅里叶逆变换(IFFT)运算,以获得多路时域第四信号 e_m ,其中 $m=1, 2, \dots, M$ 。

[0063] 优选地,根据本发明实施例的回声消除装置还包括唤醒引擎。其中,唤醒引擎配置为对多路时域第四信号中的每路时域第四信号打分以分别获得得分,确定其中得分大于唤醒阈值的 Z 路时域第四信号,并确定 Z 路时域第四信号中能量最大的一路时域第四信号,其中, Z 大于等于1。其中,唤醒引擎还配置为输出能量最大的一路时域第四信号。

[0064] 作为示例而非限制,唤醒引擎配置为得分大于唤醒阈值的 Z 路时域第四信号分别计算能量,并确定其中能量最大的一路时域第四信号,该路时域第四信号为信噪比最高的信号。

[0065] 优选地,根据本发明实施例的回声消除装置还包括识别引擎。其中,识别引擎配置为从唤醒引擎获取能量最大的一路时域第四信号,以进行语音识别并输出识别的语音。

[0066] 作为示例而非限制,识别引擎可以是自动语音识别(ASR)引擎。

[0067] 作为示例而非限制,对于智能会议系统而言,无需将唤醒引擎输出的信号输入至识别引擎。

[0068] 根据本发明的技术方案,通过映射的方法输出多个通道的信号,可以有效提高回声消除的算法性能。另一方面,本发明的技术方案的盲源分离模块不需要通过预处理去估计盲源分离中的方差,无需额外的参数即可获得很好的分离效果。通过提升盲源分离模块输入的信噪比,以及映射模块的应用,可以获得更好的性能。

[0069] 具体实施例2

[0070] 如图2所示,示出了根据本发明的一个实施例的无参考回路的回声消除方法的流程图示意图,包括下述步骤:

[0071] 步骤S202:将语音采集单元采集的多路信号固定为多个第一波束,将多个第一波束相叠加并输出目标信号;

[0072] 步骤S204:将语音采集单元采集的多路信号输入至阻塞矩阵进行预处理,以输出非目标信号;

[0073] 步骤S206:将目标信号和非目标信号进行对消并输出多波束第一信号;

[0074] 步骤S208:将语音采集单元采集的多路信号进行去混响,并输出多路第二信号;

[0075] 步骤S210:将多路第二信号进行盲源分离,以获得多路第三信号,分别针对多路第三信号中的每路第三信号在频域上进行信噪比计算,并确定多路第三信号中的每路第三信号的权重;以及

[0076] 步骤S212:将权重映射至多波束第一信号,并输出映射后的多路频域第四信号。

[0077] 优选地,在步骤S206中,进一步包括:(1)通过如下公式对目标信号和非目标信号进行对消: $ERR = MIC - w * REF$,其中,ERR为残差信号,MIC为目标信号,w为滤波器参数,REF为非目标信号;残差信号ERR包括K个单波束第一信号 B_1, B_2, \dots 直至 B_K ,分别对残差信号中的每一个单波束第一信号进行分帧处理以获取T个分帧,并对分帧中的每一个进行傅里叶变换,以获得频域上的多波束第一信号 $B_{ktf}^{(1)}$,其中,k为波束的编号, $k=1, 2, \dots, K$,K为目标信号中的波束数目,t为对应的分帧的帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

[0078] 优选地,在步骤S208中,对输入的信号进行去混响运算,以输出频域上的多路第二信号 S_{ntf} ,其中n为麦克风编号, $n=1, 2, \dots, N$,N为语音采集单元中的麦克风的数目,t为帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

[0079] 优选地,在步骤S210中,进一步包括:通过如下公式进行信噪比计算,以获得多路第三信号中的每路第三信号的信噪比 SNR_{ntf} : $SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2$,其中, S_{ntf} 为经过多通道去混响模块在频域上输出的多路第二信号, $B_{ntf}^{(2)}$ 为经过盲源分离模块对多路第二信号进行盲源分离所获得的多路第三信号,其中,n为麦克风编号, $n=1, 2, \dots, N$,N为语音采集单元中的麦克风的数目,t为帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

[0080] 优选地,在步骤S210中,进一步包括:通过如下公式,以确定多路第三信号中的每路第三信号的权重 G_{ntf} : $G_{ntf} = SNR_{ntf} / (1 + SNR_{ntf})$,其中, SNR_{ntf} 为每路第三信号的信噪比,n为麦克风编号, $n=1, 2, \dots, N$,N为语音采集单元中的麦克风的数目,t为对应的分帧的帧号, $t=1, 2, \dots, T$,f为频点编号, $f=1, 2, \dots, F$ 。

[0081] 优选地,在步骤S212中,进一步包括:通过如下公式分别将N组权重 G_{ntf} 映射到K组

多波束第一信号 $B_{ktf}^{(1)}$ 上,以得到映射后的多路频域第四信号 E ,其中:

$E_{mtf} = G_{ntf} * B_{ktf}^{(1)}$,其中, E_{mtf} 为多路频域第四信号中的第 m 路频域第四信号,其中 $m=1,2,\dots,M,M=K*N$, K 为目标信号中的波束数目, N 为语音采集单元中的麦克风的数目; G_{ntf} 为多路第三信号中的第 n 路第三信号的权重, $n=1,2,\dots,N$; $B_{ktf}^{(1)}$ 为多波束第一信号中的第 k 个单波束第一信号, k 为波束的编号, $k=1,2,\dots,K$; t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,F$ 。进一步优选地,在步骤S12中,进一步包括,对多路频域第四信号进行傅里叶逆变换运算,以获得多路时域第四信号 e_m ,其中 $m=1,2,\dots,M$ 。

[0082] 优选地,在步骤S212之后,还包括:对多路时域第四信号中的每路时域第四信号打分以分别获取得分,确定其中得分大于唤醒阈值的 Z 路时域第四信号,并确定 Z 路时域第四信号中能量最大的一路时域第四信号,其中, Z 大于等于1;以及,输出能量最大的一路时域第四信号;以及,对输出的能量最大的一路时域第四信号进行语音识别并输出识别的语音。

[0083] 具体实施例3

[0084] 作为本发明的一个具体实施例,图3示出了本发明的一个实施例的回声消除装置的硬件设计的示意图。本发明的回声消除装置可以包括语音采集模块302、音频播放模块304、功率放大器306、数字/模拟转换器(DAC)308以及主控芯片310。其中,示例性地,语音采集模块302可以是由多个麦克风组成的麦克风阵列。麦克风阵列可设置为环形,并设置在智能设备(如音箱)的顶部,以方便对目标说话人的语音指令进行拾音。示例性地,音频播放模块304可以是喇叭,喇叭可以设置在智能设备(如音箱)柱体的中下部并朝向底座。根据本发明的回声消除装置,其中固定波束模块、阻塞矩阵模块、对消模块、多通道去混响模块、盲源分离模块、映射模块以及唤醒引擎和识别引擎均可布置在主控芯片310上。根据本发明的回声消除方法,其中方法的步骤均可由主控芯片310执行。根据本发明的实施例的功率放大器306、数字/模拟转换器DAC308可根据需要进行具体设计,本发明对此不做具体限制。

[0085] 具体实施例4

[0086] 以下结合图4示出的一个具体示例对本发明的回声消除装置及方法进行阐释。

[0087] 语音采集模块为由四个麦克风组成的麦克风阵列,其中每一路麦克风采样512帧时域语音信号,共采集四路语音信号。随后,采集的时域语音信号经过固定波束模块、阻塞矩阵模块并输入至对消模块,其中固定波束模块设定其输出的多波束第一信号的波束数目为3个,即 $K=3$ 。对消模块采用自适应滤波器进行滤波,获得增强后的目标说话人语音的时域信号 $B_1^{(1)}$ 、 $B_2^{(1)}$ 、 $B_3^{(1)}$,对消模块进一步对 $B_1^{(1)}$ 、 $B_2^{(1)}$ 、 $B_3^{(1)}$ 进行分帧共获得 T 个分帧,并且进行傅里叶变换(FFT)运算,其中傅里叶变换运算采用256个采样点,即 $F=256$ 。由此,对消模块输出的多波束第一信号为 $B_{1tf}^{(1)}$ 、 $B_{2tf}^{(1)}$ 、 $B_{3tf}^{(1)}$,其中 t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,256$ 。

[0088] 多通道去混响模块也接收由语音采集模块采集的四路语音信号,获得时域上的信号。多通道去混响模块对时域上的信号进行去混响处理,以获得频域上的多路第二信号 S_{1tf} 、 S_{2tf} 、 S_{3tf} 、 S_{4tf} ,其中 t 为对应的分帧的帧号, $t=1,2,\dots,T$, f 为频点编号, $f=1,2,\dots,256$ 。

[0089] 经过多通道去混响模块在频域上输出的多路第二信号输入至盲源分离模块,以分别获取经过盲源分离得到的多路第三信号 $B_1^{(2)}$ 、 $B_2^{(2)}$ 、 $B_3^{(2)}$ 、 $B_4^{(2)}$,并对 $B_1^{(2)}$ 、 $B_2^{(2)}$ 、 $B_3^{(2)}$ 、 $B_4^{(2)}$ 进行分帧,并且进行傅里叶变换运算,以获得 $B_{1tf}^{(2)}$ 、 $B_{2tf}^{(2)}$ 、 $B_{3tf}^{(2)}$ 、 $B_{4tf}^{(2)}$,其中t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,256$ 。

[0090] 盲源分离模块进一步在频域对多路第三信号中的每路第三信号的信噪比进行计算, $SNR_{ntf} = B_{ntf}^2 / (S_{ntf} - B_{ntf}^{(2)})^2$,并确定多路第三信号中的每路第三信号的权重 $G_{ntf}:G_{ntf} = SNR_{ntf} / (1 + SNR_{ntf})$ 其中, SNR_{ntf} 为每路第三信号的信噪比,n为麦克风编号, $n=1,2,\dots,4$,N为语音采集单元中的麦克风的数目,t为对应的分帧的帧号, $t=1,2,\dots,T$,f为频点编号, $f=1,2,\dots,256$ 。

[0091] 映射模块将盲源分离模块输出的4组权重 G_{ntf} ($n=1\cdots 4$) 映射到3组多波束第一信号 $B_{ktf}^{(1)}$ ($k=1\dots 3$) 上,以得到映射后的多路频域第四信号E,其中:

$E_{mtf} = G_{ntf} * B_{ktf}^{(1)}$,从而实现多波束第一信号在频域的缩放处理,获得频域增强语音数据 E_{mtf} ,其中 $m=1,2,\dots,M$, $M=K*N$ (M 在本示例中为12), K 为目标信号中的波束数目(在本示例中为3), N 为语音采集单元中的麦克风的数目(在本示例中为4), $M=4*3=12$,具体计算: $E_{mTF} = G_{nTF} * B_{kTF}$ 。最终,映射模块还将获得的 $M=12$ 路频域信号 E_{mtf} 进行傅里叶逆变换运算,获得12路时域信号 e_m ($m=1\cdots 12$)。

[0092] 可选地,时域信号可以进一步输入至唤醒引擎和识别引擎,以进一步进行语音识别。

[0093] 具体实施例5

[0094] 下面进一步通过实验数据验证本发明的回声消除装置以及方法的优势。

[0095] 根据本发明的一个具体实施例,语音采集模块采用环形六麦克风阵列,其中六个麦克风沿着半径为4厘米的圆周均匀分布。音频播放单元为喇叭,其设置在圆心位置,距离麦克风阵列所在平面为10厘米。根据本发明的一个实施例,设置喇叭播放85db的音乐,而目标说话人间隔8秒唤醒设备并且发出语音唤醒指令,该语音唤醒指令到达麦克风处的能量为65db。图5A示出了麦克风获取的原始带噪数据,图5B示出了经过本发明设计的回声消除装置输出的干净语音数据。在图5B中,在-20db的声学场景下,分别获取麦克风数据和经过本发明设计系统的语音数据,由此可以证明本发明的系统能够在没有参考回路的情况下,利用本发明提供的技术方案有效地消除回声,不仅能够获得具有高信噪比的语音,而且能够保证目标说话人的语音的失真很小。

[0096] 以上实施例以示例的方式给出了具体操作过程,但应理解,本发明的保护范围不限于此。

[0097] 虽然出于本发明的目的已经描述了本发明各方面的各种实施例,但是不应理解为将本发明的教导限制于这些实施例。在一个具体实施例中公开的特征并不限于该实施例,而是可以和不同实施例中公开的特征进行组合。此外,应理解,上文所述方法步骤可以顺序

执行、并行执行、合并为更少步骤、拆分为更多步骤,以不同于所述方式组合和/或省略。本领域技术人员应理解,还存在可能的更多可选实施方式和变型,可以对上述模块和构造进行各种改变和修改,而不脱离由本发明权利要求所限定的范围。

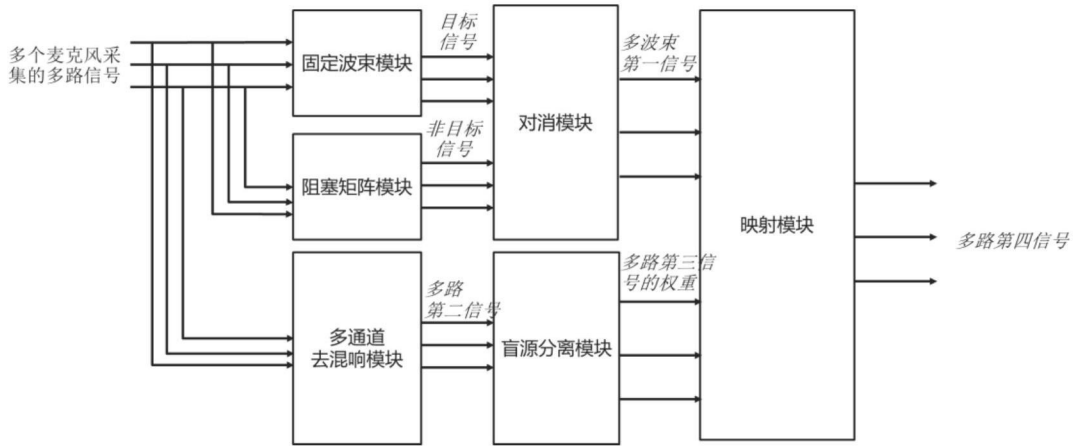


图1

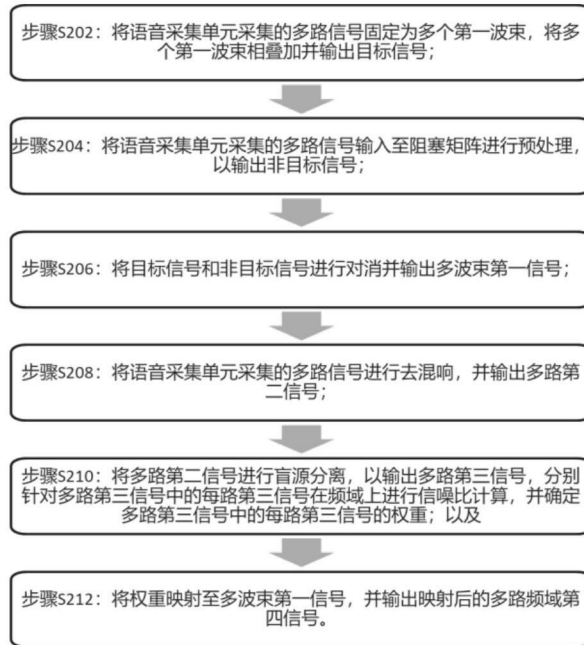


图2

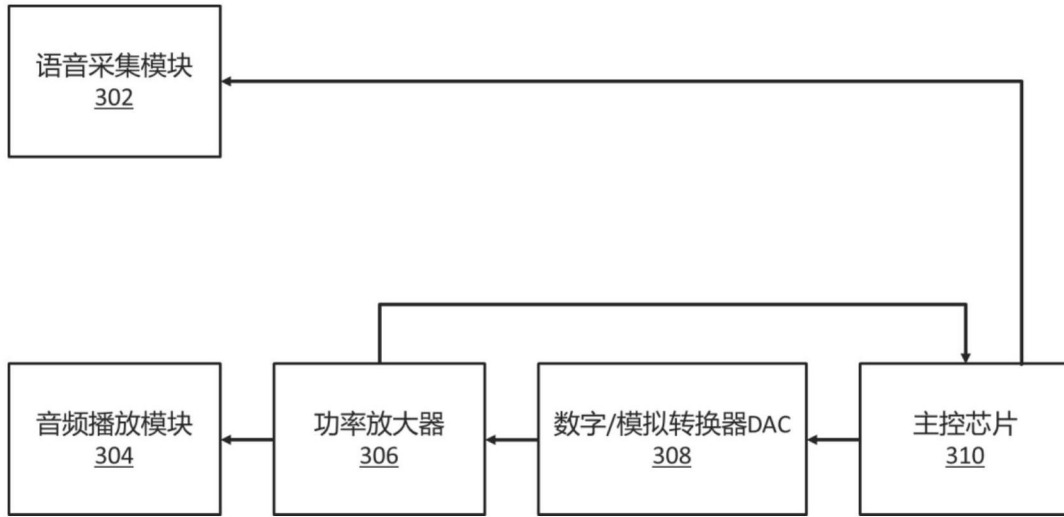


图3

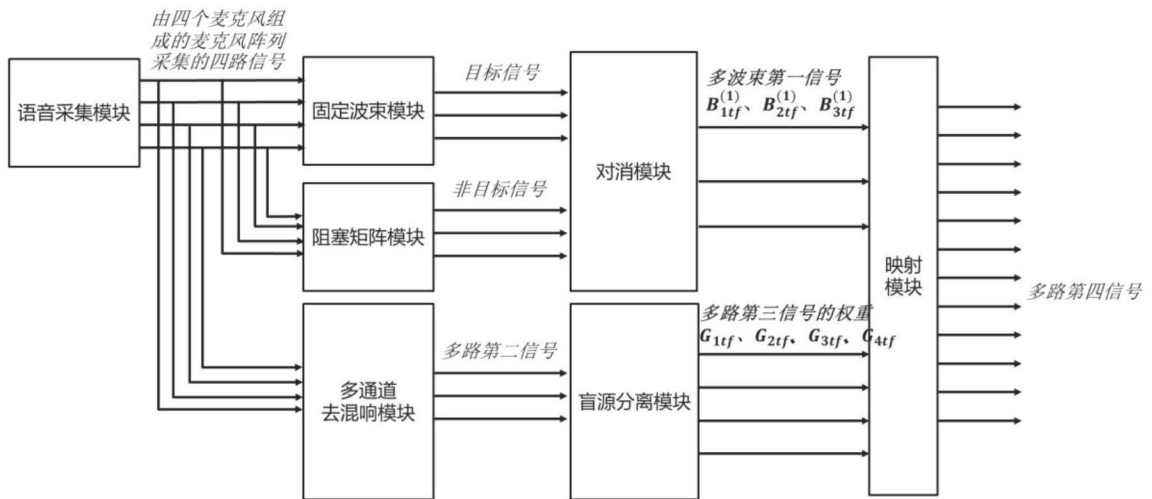


图4

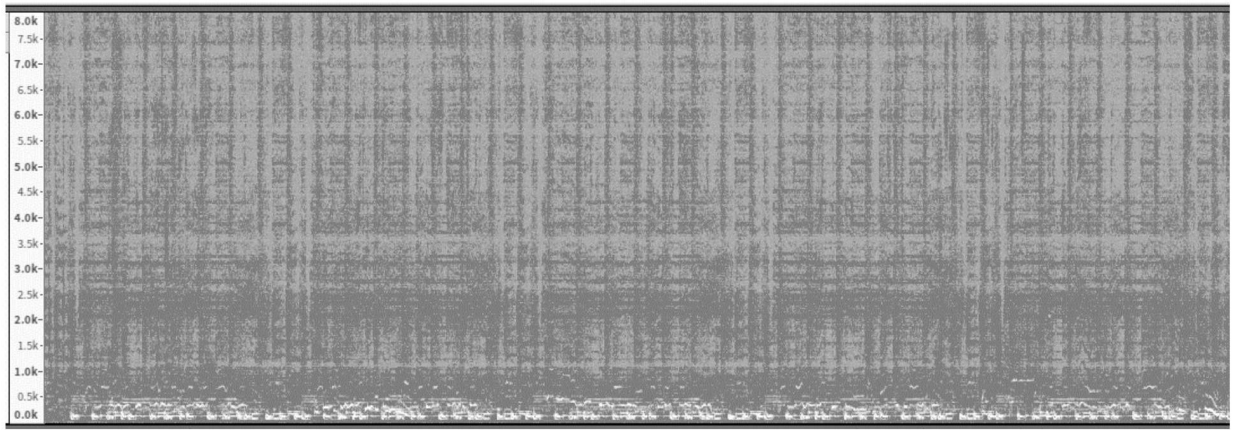


图5A

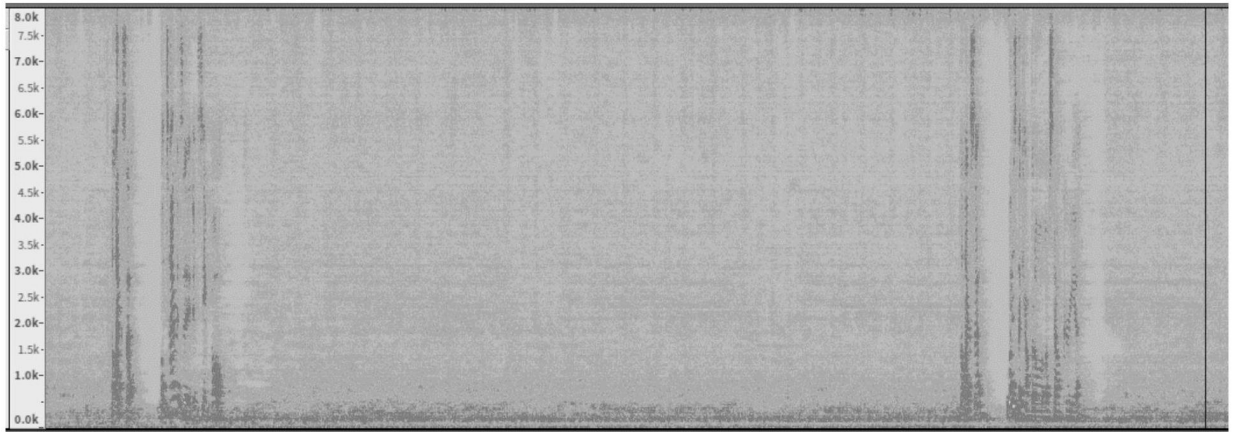


图5B