



(12)发明专利

(10)授权公告号 CN 104717587 B

(45)授权公告日 2019.07.12

(21)申请号 201410773717.X

(22)申请日 2014.12.12

(65)同一申请的已公布的文献号  
申请公布号 CN 104717587 A

(43)申请公布日 2015.06.17

(30)优先权数据  
13197139.2 2013.12.13 EP

(73)专利权人 GN奈康有限公司  
地址 丹麦巴勒鲁普

(72)发明人 拉斯穆斯·孔斯格德·奥尔森

(74)专利代理机构 北京康信知识产权代理有限  
责任公司 11240

代理人 余刚 吴孟秋

(51)Int.Cl.

H04R 3/00(2006.01)

H04R 1/10(2006.01)

(56)对比文件

CN 101192411 A,2008.06.04,

CN 103026733 A,2013.04.03,

WO 2007137364 A1,2007.12.06,

WO 2010022456 A1,2010.03.04,

CN 103348686 A,2013.10.09,

US 2009208028 A1,2009.08.20,

US 2011129097 A1,2011.06.02,

审查员 罗丽

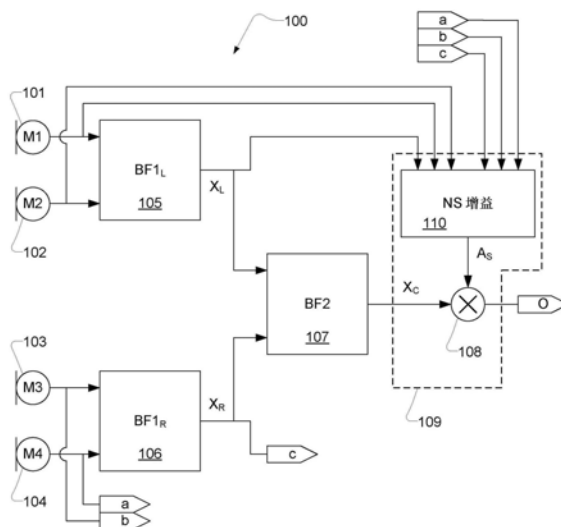
权利要求书2页 说明书12页 附图3页

(54)发明名称

用于音频信号处理的耳机和方法

(57)摘要

本发明涉及用于音频信号处理的耳机和方法。被配置为处理来自多个麦克风的音频信号的耳机和方法,包括:输出第一对麦克风信号的第一对麦克风和输出第二对麦克风信号的第二对麦克风;第一近场束形成器和第二近场束形成器,均被配置为接收一对麦克风信号并且当以相应的束形成器输出的相应的束形成信号测量时适配相应的麦克风对的空间灵敏度;其中,空间灵敏度适于抑制与期望的信号有关的噪声;第三束形成器,被配置成将从第一束形成器和第二束形成器输出的信号动态地组合为组合信号;其中,信号被组合从而使得组合信号中的信号能量被最小化而同时保持期望的信号;及降噪单元,被配置为处理来自第三束形成器的组合信号并且输出组合信号,以噪声降低。



1. 一种被配置为处理来自第一对麦克风和第二对麦克风的音频信号的耳机,所述第一对麦克风和所述第二对麦克风设置在在正常位置对准朝向佩戴所述耳机的用户的嘴巴的相应的第一和第二端射配置中,所述耳机包括:

第一对麦克风(101,102)和第二对麦克风(103,104),所述第一对麦克风(101,102)输出第一对麦克风信号并且所述第二对麦克风(103,104)输出第二对麦克风信号;其中,所述第一对麦克风被设置有第一相互距离并且所述第二对麦克风被设置有第二相互距离,并且其中,至少在所述耳机处于正常操作时,所述第一对麦克风被设置成距所述第二对麦克风的距离大于所述第一相互距离和所述第二相互距离;

第一束形成器(105)和第二束形成器(106),被配置为分别接收所述第一对麦克风信号和所述第二对麦克风信号并且执行聚焦在佩戴所述耳机的用户的嘴巴上的相应的近场束形成;

第三束形成器(107),被配置为通过加权将从所述第一束形成器(105)和所述第二束形成器(106)输出的束形成信号( $X_L;X_R$ )动态地组合成组合信号( $X_C$ );其中,所述第三束形成器计算所述束形成信号( $X_L;X_R$ )的相应的噪声等级并被配置为输出具有最低噪声等级的所述组合信号( $X_C$ );

降噪单元(109),被配置为通过时变滤波器过滤来自所述第三束形成器(107)的所述组合信号( $X_C$ )。

2. 根据权利要求1所述的耳机,

其中,所述降噪单元(109)被配置为响应于噪声抑制增益( $A_L;A_R$ )对来自所述第三束形成器(107)的所述组合信号( $X_C$ )执行噪声抑制;以及

其中,根据麦克风对信号的麦克风信号中的一个或多个麦克风信号和/或束形成信号( $X_L;X_R$ )中的一个或多个来评估所述噪声抑制增益( $A_L;A_R$ )。

3. 根据权利要求1或2所述的耳机,包括:

第一控制分支,根据所述第一对麦克风信号和/或来自所述第一束形成器的信号合成第一噪声抑制增益( $A_L$ );

第二控制分支,根据所述第二对麦克风信号和/或来自所述第二束形成器的信号合成第二噪声抑制增益( $A_R$ );

选择器,被配置为动态地选择和/或输出所述第一噪声抑制增益( $A_L$ )或所述第二噪声抑制增益( $A_R$ );

其中,所述降噪单元被配置为响应于从所述选择器中选择和/或输出的噪声抑制增益( $A_S$ )过滤来自所述第三束形成器的所述组合信号。

4. 根据权利要求3所述的耳机,

其中,所述选择器被配置为响应于第一信号质量指示符( $P_L$ )和第二信号质量指示符( $P_R$ )进行操作;以及

其中,根据相应的束形成信号( $X_L;X_R$ )合成所述第一信号质量指示符( $P_L$ )和所述第二信号质量指示符( $P_R$ )。

5. 根据权利要求3所述的耳机,

其中,束形成信号( $X_L;X_R$ )被处理以响应于相应的噪声抑制增益( $A_L;A_R$ )而降低噪声,并且然后被输入至评估器(203,208),所述评估器被配置为将信号质量指示符( $P_L;P_R$ )输出至

所述选择器(204)并且从而控制选择;以及

其中,在语音活动性被检测为不存在的时间间隔期间,所述评估器(203,208)根据最小功率标准响应于相应的噪声抑制增益( $A_L;A_R$ )评估所述束形成信号( $X_L;X_R$ )。

6.根据权利要求2所述的耳机,其中,所述噪声抑制增益( $A_L;A_R$ )被计算以通过预定的固定因数降低噪声。

7.根据权利要求1或2所述的耳机,其中,所述第一束形成器或所述第二束形成器中的至少一个被配置为包括:

第一级,根据输入信号生成求和信号和差分信号,且使所述输入信号中的至少一个相对于期望的信号与所述输入信号中的另一个相位和/或振幅对准;以及

第二级,过滤所述差分信号并生成滤波信号;

其中,根据所述求和信号与所述滤波信号之间的差生成所述束形成信号( $X_L;X_R$ );以及

其中,过滤适于使用最小均方技术来使所述束形成信号( $X_L;X_R$ )的功率最小化。

8.根据权利要求1或2所述的耳机,其中所述第三束形成器被配置为相对于与所述麦克风的的空间位置相关的预定的空间位置具有固定的灵敏度。

9.根据权利要求1或2所述的耳机,其中,所述麦克风输出数字信号;

其中,所述耳机执行多频带内的所述数字信号到时频表示的变换;并且

其中,所述耳机执行至少所述组合信号到时域表示的逆变换。

10.根据权利要求1或2所述的耳机,其中,所述麦克风输出模拟信号;

其中,所述耳机对所述模拟信号执行模数转换以提供数字信号;

其中,所述耳机执行多个频带内的所述数字信号到时频表示的变换;并且

其中,所述耳机执行至少所述组合信号到时域表示的逆变换。

11.一种用于处理来自设置在耳机中的多个麦克风的音频信号的方法,包括:

分别从第一对麦克风(101,102)和第二对麦克风(103,104)接收第一对麦克风信号和第二对麦克风信号;其中,所述第一对麦克风设置有第一相互距离并且所述第二对麦克风设置有第二相互距离,并且其中,所述第一对麦克风被布置成至少在所述耳机处于正常操作时距所述第二对麦克风大于所述第一相互距离和所述第二相互距离的距离;

对所述第一对麦克风信号和所述第二对麦克风信号执行第一近场束形成和第二近场束形成并在正常位置聚焦在佩戴耳机的用户的嘴巴上以输出相应的束形成信号( $X_L;X_R$ );

执行第三束形成以通过加权将从所述第一近场束形成和所述第二近场束形成输出的束形成信号( $X_L;X_R$ )动态地组合成组合信号( $X_C$ );其中,所述第三束形成计算所述束形成信号( $X_L;X_R$ )的相应的噪声等级并输出具有最低噪声等级的所述组合信号( $X_C$ );

通过时变滤波器过滤来自所述第三束形成(107)的所述组合信号( $X_C$ )来执行噪声降低。

12.一种包括程序代码方法的计算机可读介质,当所述程序代码方法在数据处理系统上被执行时,适于使得所述数据处理系统执行根据权利要求11所述的方法的步骤。

## 用于音频信号处理的耳机和方法

### 技术领域

[0001] 本发明涉及束形成器的技术领域,更具体地涉及用于音频信号处理的耳机和方法。

### 背景技术

[0002] 已经发现,使用多个麦克风和使用束形成技术提供了比单个麦克风或者非束形成系统卓越的音频信号再现。多个麦克风位于不同的位置并且允许所谓的空间采样,从而能够消除与诸如人的声音的期望的信号干涉的噪声,这也被称为束形成、空间滤波、或者噪声消除。随后的时变后置滤波器通常被应用为进一步将人的声音从(背景)噪声信号区分出来的装置。

[0003] 耳机、助听器、膝上型电脑、以及其他电子消费设备中通常包含多个麦克风和束形成技术的使用。

[0004] 束形成器的技术领域已经进行了大量的研究;然而,它们的质量和配置还没有被充分的开发。

[0005] 相关的现有技术

[0006] US 2012/0020485公开了一种音频信号处理方法,该音频信号处理方法评估由第一对麦克风所接收的第一声音分量相对于第一对麦克风的到达方向的第一指示;并且评估由第二对麦克风所接收的第二声音分量相对于第二对麦克风的到达方向的第二指示。在使用该方法的设备的正常操作过程中,第一对麦克风和第二对麦克风被设置在人的头部的相应侧处。该方法还包括基于第一方向指示和第二方向指示控制音频信号的增益以产生输出信号。

### 发明内容

[0007] 提供一种被配置为处理来自多个麦克风的音频信号的诸如耳机的装置,包括:第一对麦克风和第二对麦克风,第一对麦克风输出第一对麦克风信号并且第二对麦克风输出第二对麦克风信号;其中,第一对麦克风设置有第一相互距离并且第二对麦克风设置有第二相互距离,并且其中,至少在该装置处于正常操作时,第一对麦克风被设置成距所述第二对麦克风大于所述第一相互距离和第二相互距离的距离;第一束形成器和第二束形成器,均被配置为接收一对麦克风信号并且当以从相应束形成器输出的相应束形成信号测量适配相应的麦克风对的空间灵敏度;其中,空间灵敏度适于抑制与期望信号有关的噪声;第三束形成器,被配置为将从第一束形成器和第二束形成器输出的信号动态地组合成组合信号;其中,信号被组合为使得组合信号中的噪声能量被最小化而同时保持期望的信号;以及降噪单元,降噪单元被配置为处理来自第三束形成器的组合信号并且输出组合信号以使得噪声降低。

[0008] 因此,在具有处理麦克风信号的第一束形成器和第二束形成器的第一束形成级和具有处理从第一阶段输出的信号的第三束形成器的第二级中提供束形成。第一束形成级用

于通过适配相应的麦克风对的空间灵敏度来局部增强或者强化相对于麦克风信号的期望的信号。例如,通过调整束形成器系数来控制束形成器空值的空间配置而适配空间灵敏度,调整束形成器系数可包括调整束形成器系数以使得束形成器获得全方向特征,其有助于避免放大诸如风噪声的不相关的(麦克风之间的)噪声。第一束形成级的效力取决于每个麦克风对中的麦克风均紧紧地靠近彼此的假设(出于下列说明的原因)。

[0009] 除在捕获希望信号时的这样的局部优化之外,噪声分量的等级在第一束形成信号与第二束形成信号之间可显著地变化。这可能是由于麦克风处的不同等级,例如,风湍流是高度局部现象,和来自头戴式设备中的用户头部的音响阴影效果。而且,第一束形成器和第二束形成器可能不能够根据麦克风对的相对位置、关注的信号、以及干扰噪声等地消除噪声。

[0010] 因此,第三束形成器被配置为接收已经经过第一级束形成器的局部优化的信号,通过所述局部优化,尽可能地将期望的信号隔离。通过动态地组合来自左手侧和右手侧的信号,可以选择或强化来自最有利定位的麦克风对的空间上受控的信号。

[0011] 以这种方式处理麦克风信号,通过降噪单元(如所述要求的,当其被配置为处理来自第三束形成器的组合信号时)提高噪声抑制的效果。这部分是由于所观察到的在这样的两级束形成之后期望的信号更清晰并且由此使得噪声抑制更为有效。而且,两级束形成器方法利用已知的双麦克风束形成器实现了紧密隔开的麦克风和非紧密隔开的麦克风的束形成的组合效益(benefit,好处)。第三束形成器可通过输入信号的线性或者非线性权重组合其输入信号。

[0012] 诸如耳机、助听器、或者通过麦克风拾取音频信号的另一装置的装置可被配置为由人佩戴,使得第一对麦克风设置在人头部的左手侧,而第二对麦克风设置在人头部的右手侧。通常,两对麦克风位于耳机的耳罩上,眼镜架或者横梁或者位于人头部的相应侧的其他突起。麦克风至少大致被布置成所谓的端射配置(end-fire configuration)。可替代地或者此外,麦克风可被布置成较宽的配置。

[0013] 通过设置麦克风,使得对内麦克风(intra-pair microphones)至少在耳机处于正常操作且对内麦克风为指向佩戴耳机的用户的嘴巴的端射配置时比对间麦克风(inter-pair microphones)位于更靠近,第一束形成器和第二束形成器可利用所谓的近场效应的优点在低频比更高频更高地提高信噪比,并且可以消除更多的更高频噪声,从而避免空间混淆现象。信噪比的提高可以高达15dB。此外,第三束形成器可利用不同麦克风对暴露的不同局部噪声等级。当麦克风对位于人的头部的不同侧时,头部可形成降低人头部的一侧上的噪声等级的风影和/或声影。本发明的主要优点在于,设计对所有麦克风输入进行操作的单一适配性束形成器的高度复杂的问题被分解成三个简单的、鲁棒、非常易于理解的双麦克风束形成器。

[0014] 总之,可以选择具有不同特征的不同类型的麦克风。

[0015] 通常,期望的信号是表示来自麦克风附近的讲话者的语音或者从相对于麦克风的方位的特定方向出现的语音的信号。期望的信号被表征为从相对于麦克风的空间位置的具有预定空间位置的一个或者多个声源发出。因为使用多个麦克风来拾取期望的信号,所以期望的信号由麦克风信号和/或束形成信号之间的预定相差和/或振幅差来表征。期望的信号还可以由预定的时间特征和/或预定的相频/幅频特征来表征。

[0016] 噪声信号或者简单的噪声可包括由以足够高的风速出现并且作用于麦克风隔膜上的风而引起的湍流声音(turbulence sound)。噪声还可包括诸如来自机器的音调的背景声音、来自物品卡嗒或者叮当的声音、来自彼此之间谈话的人的声音等。在一些定义中,噪声由从位于不同于期望的信号的其他位置的一个或者多个声源发出来表征。

[0017] 第一束形成器和第二束形成器逐渐地或者以以下步骤来适配方向灵敏度,所述步骤包括至少根据以下特征组近似灵敏度:全方向性、双向性、心型、子心型、超心型(hypercardioid)、超级心形(supercardioid)、或者猎枪型(shotgun)。方向灵敏度可以逐渐在全方向、双向性、以及心型特征之间逐渐改变。第一束形成器可以如W0 2009/132646中所公开地配置,对于尤其结合其图1所公开的所有内容,通过引用结合于此。

[0018] 第三束形成器可根据从噪声功率评估的系数组合来自第一束形成器和第二束形成器的信号。在来自第一束形成器的信号的噪声功率高于来自第二束形成器的信号的噪声功率的情况下,来自第二束形成器的信号高于来自第一束形成器的信号被加权,反之亦然。当语音被检测为不存在时,可以评估信号的噪声等级。

[0019] 在端射对的情况下,根据希望的方向灵敏度,第一对麦克风之间的第一相互距离和第二对麦克风之间的第二相互距离短于所关注的最小波长。在波长比所关注的波长更短的频率或者以上,由于空间混淆现象的效应,抑制或者消除噪声的能力将降低。麦克风对之间的距离可对应于人的两耳之间的直线距离(约为18cm-22cm)。对于高达4KHz的所关注的带宽,第一相互距离和第二相互距离可以为约10mm、20mm、或者40mm。

[0020] 通常,该装置可以在时域或者时频域内执行信号处理。在后者情况下,在运行基础上,对预定时间段的信号块执行时间到频率的变换。在时频域中,信号被表示为多个频率区中的时域采样。因此,对经过时频域处理的信号执行频率到时间的重构。

[0021] 在一些实施方式中,降噪单元被配置为响应于噪声抑制系数对来自第三束形成器的组合信号执行噪声抑制;并且从麦克风信号和/或束形成信号评估噪声抑制系数。降噪单元被配置为时域或时频域中的时变滤波器。噪声抑制系数可以随时间改变并确定时变滤波。

[0022] 噪声抑制系数可包括从第一组麦克风信号以及从束形成信号评估的第一系数。可替代地或者此外,噪声抑制系数可包括从第二组麦克风信号以及从束形成信号评估的第二系数。从第一系数和第二系数可对噪声抑制系数进行组合。

[0023] 噪声抑制系数可以是时频域内的乘法器的增益因数或者时域滤波器的滤波器系数。

[0024] 在一些实施方式中,所述装置包括:第一控制分支,从第一对麦克风信号和/或第一束形成器合成第一噪声抑制增益;第二控制分支,从第二对麦克风信号和/或第二束形成器合成第二噪声抑制增益;以及选择器,被配置为动态地选择和/或输出第一噪声抑制增益或第二噪声抑制增益;其中,降噪单元被配置为响应于从选择器所选择和/或输出的噪声抑制增益处理来自第三束形成器的组合信号。

[0025] 因此,可以动态地选择第一噪声抑制增益或者第二噪声抑制增益,从而使得其是根据从由相应束形成器输出的相应束形成信号和相应噪声抑制增益所评估的信号质量测量。这是有利的,因为可以在第一和第二降噪增益并不是等同有利的条件下计算第一降噪增益和第二降噪增益。结果,噪声不能被等同地很好地抑制和/或期望的信号不能被等同地

很好地保持。例如,用于计算第一噪声抑制增益的机制可访问使其自身更容易区分噪声信号与期望信号的信号。该情形可以根据如下情况而出现,在所述情况中,由于引起较少风声噪声或者背景噪声的用户头部阴影,在第一束形成器输入处噪声不大。在由第一噪声抑制计算所采用的空间线索(spatial cue)更容易区分的情况也可能出现这种情形。

[0026] 滞后量或者阈值可被应用并用作是否启用选择器的标准。因此,当评估噪声能级低于预定滞后量或者阈值时,可以禁用开关。滞后量或者阈值可以在约1dB至约3dB的范围内。因此,可以在(1)实现最低输出噪声等级与(2)使诸如语音信号的期望信号的失真最小化之间进行权衡。

[0027] 在一些实施方式中,选择器被配置为响应于第一信号质量指示符和第二信号质量指示符进行操作;从响应于相应的降噪增益被处理为降低噪声的相应束形成信号合成信号质量指示符。

[0028] 就噪声抑制方面而言,信号质量的重要方面是信噪比。例如,参考图2,当使用束形成的、降噪信号作为信号质量评估的输入时,通过 $X_L$ 和 $X_R$ 影响信噪比。例如,如果 $X_L$ 的信噪比大于 $X_R$ 的信噪比,则在 $A_L$ 和 $A_R$ 将降低噪声分量降低相同因数的情况下, $A_L X_L$ 的信噪比将大于 $A_R X_R$ 的信噪比。

[0029] 而且,信号质量评估受到 $A_L$ 和 $A_R$ 的质量影响。在一些情况下,在头部的一侧,更容易从噪声中区分出语音。原因在于,用户头部可以在用户头部的下风侧上保护麦克风免受风。另一原因在于,在用户头部的下风侧上可更为清楚地区分噪声抑制计算所采用的空间线索。

[0030] 可以根据相应的降噪增益 $A_L; A_R$ 与相应的束形成信号 $X_L; X_R$ 的均方乘积来计算信号质量指示符 $P_L; P_R$ 。可以计算每个频带的信号质量指示符或者累积所有频带的信号质量指示符。

[0031] 在一些实施方式中,响应于降噪增益被处理以降低噪声的束形成信号被输入到评估器中,所述评估器被配置为将控制信号输出到选择器并且由此控制选择;并且在语音活动性被检测被不存在时的时间间隔期间,评估器根据最小功率的标准评估响应于相应的降噪增益被处理以降低噪声的束形成信号。

[0032] 因此,可以根据在人的头部的相应侧处的噪声条件(例如,噪声功率)的评估执行对相应噪声抑制增益的选择。

[0033] 用作选择标准的左和右束形成的降噪信号的最小噪声功率将多个质量参数组合成简单的计算。如上所述,当麦克风输入通过对准滤波器被对准时,噪声功率是信噪比的相似测量,但计算更为简单。

[0034] 当执行降噪时,存在引入使语音质量劣化的语音处理人为现象的风险。最小噪声功率标准中所使用的噪声功率测量在许多情况下选择更高的语音质量。当该标准基于最小功率时,优选更容易检测语音分量的所有部分,尤其是低等级部分的信号,从而导致语音处理认为现象的更少的可听事件。语音活动性检测器可输出指示是否检测到语音活动性的信号。当一个或者多个麦克风信号和/或束形成信号的振幅或者峰值大小或者功率水平超过预定或者时变阈值时,可以检测到语音活动性。阈值等级可被适配成评估的噪声等级。

[0035] 在一些实施方式中,计算噪声抑制系数以通过预定的固定因数降低噪声。

[0036] 例如,预定因数可以是13dB、6dB、10dB、15dB或者另一因数。这可以通过将噪声抑

制增益限制为预定因数来实现。

[0037] 作为一个实例,第一束形成器和第二束形成器输出处的评估的噪声等级可以分别为如-30dB和-20dB;固定因数可以为如10dB;因此,噪声抑制之后的评估的噪声等级则分别为-40dB和-30dB。

[0038] 例如,可以使用噪声抑制增益选择模块之前的信号链中任一点处的麦克风上的对准滤波器/增益在朝向所关注的信号的等级匹配左和右信号束形成信号。作为使用固定噪声抑制因数和等级匹配的左和右信道的有益结果,噪声功率计算被调节成用作反映更程度的左和右束形成器输出的信噪比的左和右信号质量测量。

[0039] 在一些实施方式中,第一束形成器和第二束形成器中的至少一个被配置为包括:第一级,从输入信号生成求和信号和差分信号,且使输入信号中的至少一个相对于期望的信号与输入信号中的另一个相位和/或振幅对准;以及第二级,过滤差分信号并且生成滤波信号;其中,从求和信号与滤波信号之间的差生成束形成输出信号;并且其中,滤波器适于使用最小均方技术来使得束形成输出信号的功率最小化。

[0040] 因此,第一束形成器和/或第二束形成器选择性并且适配性地消除来自特定方向的声音。

[0041] 滤波器可具有低通特性以相对于高频分量增强低频分量。滤波器可以是低音增强滤波器(bass-boost filter)。

[0042] 这样的束形成器配置为如WO 2009/132646中所公开的,对于其所公开的全部内容,通过引用结合于此。

[0043] 在一些实施方式中,第三束形成器被配置为相对于与麦克风的空間位置相关的预定的空间位置具有固定的灵敏度。

[0044] 固定灵敏度意味着第三束形成器相对于从预定空间位置处的声源发出的声音施加固定的频率响应。

[0045] 预定位置相对于第一组麦克风和第二组麦克风的空間位置和方位以预定方式定位。当人按照正常方式佩戴该装置时,预定空间优选为关于人的嘴巴居中。

[0046] 第三束形成器的束形成系数可被约束为对固定增益求和,例如,对空间位置的单位增益。就增益不可适配这种意义而言,增益是固定的。然而,可以结合校准调整增益或者将增益调整为优选设置。

[0047] 第三束形成器可通过线性组合对输入信号进行组合。可替代地,可以通过非线性组合对信号进行组合。

[0048] 在一些实施方式中,麦克风输出数字信号;该装置在多频带内执行数字信号到时频表示的变换;并且该装置执行至少组合信号到时域表示的逆变换。

[0049] 可以通过施加给预定时间段的信号块的快速傅里叶变换FFT执行该变换。这种变换可包括应用Hann窗口或者另一类型的窗口。经由逆快速傅里叶变换IFFT可从时频表示重构时域信号。

[0050] 预定时间段内的信号块可具有50%重叠的8ms时限,这意味着,每隔4ms计算变化、适配更新、降噪更新、以及时域信号重构。然而,其他时间段和/或更新间隔也是可以的。数字信号可以以多次过采样速率的一位信号、两位信号或者三位信号、或者8位、10位、12位、16位或者24位信号。



[0051] 在可替代的实现方式/实施方式中,系统的全部或者部分直接在时域内操作。例如,通过FIR或者IIR滤波、在频域中计算的噪声抑制滤波器系数可将噪声抑制应用至时域信号。

[0052] 在一些实施方式中,麦克风输出模拟信号;装置执行模拟信号的模数转换以提供数字信号;装置执行多频带内数字信号到时频表示的变换;并且装置执行至少组合信号到时域表示的逆变换。

[0053] 在一些实施方式中,当人使用该装置时,麦克风组中的至少一对麦克风被设置成定位朝向期望人的嘴巴所期望处于的位置的端射配置。例如,对于耳机或者助听器,这样的配置已经显示具有良好的噪声消除和抑制功能。

[0054] 还提供一种用于处理来自多个麦克风的音频信号的方法,包括:分别从第一对麦克风和第二对麦克风接收第一对麦克风信号和第二对麦克风信号;其中,第一对麦克风设置有第一相互距离并且第二对麦克风设置有第二相互距离,并且其中,第一对麦克风被布置成在至少该装置处于正常操作时距第二对麦克风大于第一相互距离和第二相互距离的距离;对第一对麦克风信号和第二对麦克风信号执行第一束形成和第二束形成以输出相应的束形成信号;当以相应的束形成信号测量时通过相应的麦克风对适配空间灵敏度,从而使空间灵敏度被适于抑制与期望的信号有关的噪声;执行第三束形成,以将从第一束形成和第二束形成输出的信号组合为组合信号;其中,信号被组合成使得自合信号中的噪声能量被最小化而同时保持期望的信号;以及执行降噪以处理来自第三束形成器的组合信号并且输出组合信号,从而降低噪声。

[0055] 例如,还提供了一种存储在诸如DVD的计算机可读介质上的计算机程序产品,包括程序代码方法,当所述程序代码方法在数据处理系统上被执行时,适于使数据处理系统执行所述方法中的步骤。

[0056] 例如,还提供了一种计算机信号,例如,下载信号,包含在载波中并且表示指令序列,当由处理器执行所述指令序列时,使得处理器执行方法中的步骤。

[0057] 此处,在下文中,术语“处理装置”和“处理单元”旨在包括适于执行本文所描述的功能的任何电路和/或设备。具体地,上述术语包括通用或者专用可编程微处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、可编程逻辑阵列(PLA)、场可编程门阵列(FPGA)、专用电子电路等、或者其组合。

## 附图说明

[0058] 将参考附图,通过本发明的实施方式的以下示出的非限制性的细节描述进一步阐述本发明的上述目标和/或其他目标、特性以及优点,其中:

[0059] 图1示出了信号处理器的框图;

[0060] 图2示出了信号处理器的更为详细的框图;以及

[0061] 图3示出了具有多个麦克风的装置的不同配置。

## 具体实施方式

[0062] 在下列描述中,将参考通过例如示例示出如何实现本发明的附图。

[0063] 图1示出了信号处理器以及第一对麦克风和第二对麦克风的框图。第一组麦克风

101和102以及第二组麦克风103和104设置有麦克风间的对内距离,所述对内距离与麦克风对之间的麦克风对间距离相比相对较短。以参考标号100表示信号处理器。

[0064] 第一对麦克风101和102输出被输入到第一束形成器105的第一麦克风信号对,而第二对麦克风103和104输出被输入到第二束形成器106的第二麦克风信号对。第一束形成器105和第二束形成器106输出相应的输出信号 $X_L$ 和 $X_R$ 。

[0065] 第一束形成器105和第二束形成器106均被配置成适配其空间灵敏度。空间灵敏度被适配成消除或者抑制与期望的信号相关的噪声。第一束形成器和第二束形成器可以如W0 2009/132646中所公开的那样配置。

[0066] 第三束形成器107被配置为将从第一束形成器105和第二束形成器106输出的信号 $X_L$ ;  $X_R$ 动态地组合为组合信号 $X_C$ 。可以下列表达式表达组合信号 $X_C$ :

$$[0067] \quad X_C = G_L X_L + G_R X_R$$

[0068] 其中, $G_L$ 和 $G_R$ 分别表示来自接收 $X_L$ 的第一输入和来自接收 $X_R$ 的第二输入的转移函数(transfer function)。上述表达式取决于频域表示; $X_L$ 和 $X_R$ 是复数。对于时域表示,存在等同表示。第三束形成器被配置为动态地调整实数或者复数 $G_L$ 和 $G_R$ 以输出具有最低噪声等级的 $X_C$ 而同时保持期望的信号。

[0069] 下列表达式是如何计算实数 $G_L$ 、 $G_R$ 的实例:

$$[0070] \quad \hat{G}_L = \frac{\langle |X_R|^2 \rangle - \text{Re} \langle X_L X_R^* \rangle}{\langle |X_L - X_R|^2 \rangle}$$

$$[0071] \quad \hat{G}_R = \hat{G}_L - 1$$

[0072] 其中,Re是复数的实部,\*、 $\langle \cdot \rangle$ 、以及 $|\cdot|$ 分别表示复共轭、时间间隔的平均以及绝对值。

[0073] 实数 $\hat{G}_L$ 和 $\hat{G}_R$ 的上述表达式是受以下约束的均方成本函数(mean squares cost function)的解:

$$[0074] \quad \hat{G}_L = \underset{G_L}{\text{arg min}} \langle |X_C|^2 \rangle$$

[0075] 受限于:

$$[0076] \quad \hat{G}_L + \hat{G}_R = 1$$

[0077] 即, $X_C$ 的均方被最小化为收到约束的实数 $G_L$ 的函数。该约束确保期望的信号优于来自至少一些其他位置的信号。

[0078] 在一些实施方式中,匹配滤波器被插入在麦克风和第一级的束形成器(即,实施方式中所示的第一和第二束形成器)的输入之间。从而过滤输入到第一束形成器和第二束形成器的信号,从而,在所有输入中,期望的信号分量充分地相同(即,相对于相位和振幅)。滤波器补偿期望信号至麦克风的声通路中的变化以及麦克风灵敏度的变化或者其他变化。这样的匹配滤波器还可以由对准滤波器并且匹配可以由对准表示。由于相对于期望源的输入对准,第一束形成器和第二束形成器的输出的期望的信号分量由于内置约束而几乎相同(例如,W0 2009/132646中所描述的)。即,第三束形成器的输入相对于期望的信号分量充分

相同。因此， $\hat{G}_L + \hat{G}_R = 1$ 约束导致第三束形成器的输出和输入相对于期望的信号充分相同。

[0079] 可选择输入中的一个作为用于麦克风对准的参考。例如，对准滤波器中的一个可被配置为产生全通特征；其他对准滤波器被相应地配置。结果，第一束形成器中的每个的输出相对于期望的信号充分相似并且还参考输入相似。

[0080] 通过假定并且补偿原始期望信号与麦克风之间的已知的声学关系并且通过使用灵敏度变化非常小的麦克风可预先配置麦克风对准滤波器。可以在生产时的校准步骤中评估麦克风灵敏度。可以在设备运行时评估麦克风对准滤波器；例如，当被语音或者噪声活动性检测器激活时，通过最小平方技术评估对准滤波器。

[0081] 通过将麦克风对准滤波器直接集成到一个或者多个束形成器的计算中，或者可替代地，在第一束形成器和第二束形成器输出处，可以等同地实现束形成器相对于期望的信号约束。

[0082] 当以这种方式组合输入信号 ( $X_L; X_R$ ) 时，表现出最低噪声等级的输入信号相对于其他的输入信号被强化。

[0083] 用于计算 $G_L$ 和 $G_R$ 的上述表达式至少在一定程度上限制对期望的信号的影响并且充分地发挥作用而不需要任何的音活动性检测器VAD。

[0084] 下列表达式是可选的并且需要稍微少的资源来进行计算少，但是，结合语音活动性检测器VAD使用比较有利：

$$[0085] \quad \tilde{G}_L = \frac{\langle |X_R|^2 \rangle}{\langle |X_R|^2 \rangle + \langle |X_L|^2 \rangle}$$

$$[0086] \quad \tilde{G}_R = \tilde{G}_L - 1$$

[0087] 其中， $X_R$ 和 $X_L$ 是相应信号的复数表示。该表达式收到相似的最小化并且受如上所述的约束，但是，假定 $X_R$ 和 $X_L$ 中的噪声分量不相关。在这种情况下，语音活动性检测器被应用于丢弃 $X_R$ 和 $X_L$ 的其中为了评估 $G_L$ 和 $G_R$ 而存在语音的信号部分。对于多麦克风收，在US7206421B1中公开了这样的加权规则。

[0088] 为了更为鲁棒的性能， $G_L$ 和 $G_R$ 可被进一步约束为如0与1之间的间隔。

[0089] 通常，应注意，发射期望的信号的源的评估位置可被预先配置并且被锁定至相对于麦克风位置的期望位置。这可以是耳机的情况，其中，当耳机被佩戴在正常位置时，可以充分地限定人的嘴巴的位置。在其他情况下，该装置可包括根据例如来自一个、两个或者多个麦克风对或者两组以上麦克风的信号的相差和/或振幅差评估期望的信号的源的位置的跟踪器。这可以是用于例如汽车中的通信设备的喇叭扩音器或者免提装置的情况。

[0090] 将来自束形成信号 $X_L$ 和 $X_R$ 的组合信号 $X_C$ 输入到计算噪声抑制增益 $A_S$ 的噪声抑制单元109中。此外，噪声抑制单元109在计算噪声抑制增益 $A_S$ 的过程中可包括来自一个或者多个麦克风101、102、103、104的麦克风信号。来自M3和M4的信号和从束形成器106输出的信号 $X_R$ 被标记为‘a’、‘b’、以及‘c’并且如相应标签所指示地被输入到噪声抑制单元109中。

[0091] 下面进一步描述噪声抑制增益 $A_S$ 的计算。

[0092] 在所示的实施方式中，通过乘法器108将噪声抑制增益 $A_S$ 施加到组合信号 $X_C$ 。从乘

法器输出的信号是包括由麦克风拾取的束形成和噪声抑制信号分量的再现音频信号。标签‘0’表示来自信号处理器的输出。该输出经受进一步的信号处理、放大和/或传输。

[0093] 图2示出了信号处理器的更为详细的框图。其示出了将噪声抑制增益 $A_S$ 选择为第一或左噪声抑制增益 $A_L$ 或者第二或右噪声抑制增益 $A_R$ 。左噪声抑制增益 $A_L$ 根据束形成信号 $X_L$ 和/或麦克风信号 $x_{m1}$ 和/或 $x_{m2}$ 而计算。对应地,右噪声抑制增益 $A_R$ 根据束形成信号 $X_R$ 和/或麦克风信号 $x_{m3}$ 和/或 $x_{m4}$ 而计算。

[0094]  $A_L$ 经由乘法器205被施加到 $X_L$ 并且 $A_R$ 经由乘法器209被施加到 $X_R$ 。乘法器105和209的相应输出被输入到相应的信号质量评估器203和208。输入可被解释为左降噪的束形成信号和右降噪的束形成信号。

[0095] 信号质量评估器203和208可根据信噪比的标准评估从乘法器205和209输出的信号的信号质量。可替代地,可以根据当语音活动性被检测为不存在的时间间隔期间的噪声信号功率标准来评估信号质量。这可以通过应用麦克风对准滤波器以使期望的信号分量在所有的束形成器输入和输出处充分相同来促进。在这种情况下,信噪比和噪声功率是信号质量的相似的测量。信号质量评估器经由选择器204输出选择 $A_L$ 或者 $A_R$ 的信号 $P_L$ 和 $P_R$ 。从选择器输出的 $A_S$ 表示所选择的噪声抑制增益并且其经由乘法器108被施加到 $X_C$ 。

[0096] 信号 $P_L$ 和 $P_R$ 以及由此的信号质量评估器203和208可被定义为有关作为输入接收的信号的噪声分量的功率计算。例如, $P_L$ 可被定义为仅是噪声间隔过程中的束形成的降噪输入的均方。可以对合适的时间间隔(例如,100ms或者1s)并且可以对合适的频率间隔(例如,0Hz-8000Hz)执行平均。

[0097] 当 $P_L$ 小于 $P_R$ 时,选择器204可被配置为选择 $A_L$ ,并且当 $P_L$ 大于 $P_R$ 时,选择器204可被配置为相反地选择 $A_R$ 。语音活动性检测器202和207分别将表示是否检测到语音的信号输出到信号质量评估器203和208。

[0098] 单输入型的语音活动性检测器VAD可被配置为通过接收输入信号并且计算输入信号的缓慢变化的量级平均值而评估噪声本底等级 $N$ 。当信号的大小暂时超过所评估的噪声本底为如10dB的预定因数时,比较器可输出表示存在语音信号的信号。当检测存在语音时,VAD可禁用噪声本底评估。当噪声处于准稳态并且当语音的大小充分超过评估的噪声本底时,语音检测器工作。这样的语音活动性检测器可以限带信号或多频带操作以生成从多频带聚合的语音活动性信号。当语音活动性检测器以多频带操作时,其可输出用于相应的多频带的多个语音活动性信号。

[0099] 多输入型的语音活动性检测器VAD可被配置为计算表示多个信号之间的相关性的信号。例如,由于嘴巴比噪声源更靠近于麦克风,所以语音信号可表现出麦克风之间更高等级的相关性。其他类型的语音活动性检测器基于计算诸如方向性和接近度的空间特征或者线索以及将信号分解成代码本时间/频率配置文件的字典方法(dictionary approach)。

[0100] 从下列表达式可以计算指定的 $G_{NS}$ 、或 $A_L$ 、或 $A_R$ 的噪声抑制增益:

$$[0101] \quad G_{NS} = \frac{|X|^2}{|X|^2 + P_N F}$$

[0102] 其中, $P_N$ 是时间实例 $t$ 时的评估噪声本底等级的平方; $|X|^2$ 是时间实例 $t$ 时的输入信号的平方;并且 $F$ 是因数,例如,因数10。噪声抑制增益如果在频域内被应用,噪声抑制增益则影响经由乘法器的输入信号。

[0103] 因此,一方面,如果噪声本底等级非常低,则当明显存在语音时, $G_{NS}$ 变成1。另一方面,如果不存在语音或者噪声等级上升时, $G_{NS}$ 移至小于1的值并且由此抑制输入信号。因数 $F$ 被选择为设置应如何积极地抑制输入信号。

[0104] 关于语音活动性检测器和噪声抑制增益的以上描述,其输入信号可以是第一束形成器和/或第二束形成器和/或第三束形成器输出的信号和/或任何麦克风信号。

[0105] 通常,评估信号与噪声关系的方式基于跟踪噪声本底,其中,语音或者嘈杂的语音通过明显超过噪声本底等级的信号部分被识别。例如,再生等级例如通过如[R.Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," Trans.on Speech and Audio Processing, Vol.9, No.5, July 2001]中的最小统计学来评估,其中,适配性地评估最小信号等级。

[0106] 识别信号和噪声部分的其他方式基于计算诸如方向性和接近度[O.Yilmaz and S.Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking", IEEE Transactions on Signal Processing, Vol.52, No.7, pages 1830-1847, July 2004]或者相干性[K.Simmer et al., "Post-filtering techniques." Microphone Arrays. Springer Berlin Heidelberg, 2001. 39-60]的多麦克风/空间特征。也可以应用将信号分解成代码本时间/频率分布的字典方法[M.Schmidt and R.Olsson: "Single-channel speech separation using sparse non-negative matrix factorization," Interspeech, 2006]。

[0107] 通常,可以如[Y.Ephraim and D.Malah, "Speech enhancement using optimal non-linear spectral amplitude estimation," in Proc.IEEE Int.Conf.Acoust.Speech Signal Processing, 1983, pp.1118-1121]中描述的那样或者如有关噪声抑制技术的文献中的其他地方所描述的那样来实现噪声抑制。通常,对信号应用时变滤波器。通常在表示多个频带的信号的频率变换域/滤波器组中实施分析和/或滤波。在每个表示的频率中,根据评估的期望信号与噪声分量的关系计算时变增益,例如,当评估的信噪比超过预定的适配性或者固定阈值时,增益被控制为趋于1。相反,当评估的信噪比不超过阈值时,增益被设置成小于1的值。指定 'x' 和 'y' 的标签连接相应的信号:x到x以及y到y。

[0108] 图3示出了具有多个麦克风的装置的不同配置。在左手侧,具有眼睛框(bow) 306的眼镜架303被配置有两组麦克风304和305。在右手侧,柔性颈带307被配置有两组麦克风308和309。参考标号301表示佩戴眼镜架303的人的头部,并且参考标号302表示佩戴颈带307的人的头部。

[0109] 麦克风可被布置成所谓的端射配置,其中,相应对或者相应组的麦克风中的麦克风位于与期望的信号的源的位置交叉或者靠近于期望的信号的源的位置的线上。该位置可以是人的嘴巴开口的位置或者靠近人的嘴巴开口的位置。在端射配置中,麦克风对中的麦克风位于与期望的信号的源的位置交叉的直线上。据发现,当该装置是耳机、助听器等时,这样的配置适合用于有效地抑制或者消除来自位于其他地方的源的噪声。

[0110] 在可替代的配置中,使用用于麦克风位置的所谓的较宽配置(broadside configuration)。在较宽配置中,麦克风对中的麦克风位于距期望的信号的源的位置成相等的距离的直线上。

[0111] 在可替代的配置中,麦克风对中的麦克风位于相对于从麦克风对至期望的信号的



源的位置的方向以例如 $5^\circ$ 、 $10^\circ$ 、 $45^\circ$ 倾斜的直线上,从而提供实际上更为适用的配置。

[0112] 通常,在上面中,假定了使用输出数字信号的所谓的数字麦克风。然而,可以使用结合模数转换器或者从声场到采样域的任何其他换能器的模拟麦克风。麦克风通常被包括在直径范围从通常3mm至5mm或者6mm的所谓封壳中。

[0113] 通常,束形成器可从一对以上的麦克风接收信号。束形成器,例如,第一级束形成器,可从3个、4个或者多个麦克风接收麦克风信号。第一级可包括除第一束形成器和第二束形成器之外的束形成器;例如,第一级可包括3个、4个或者多个束形成器。

[0114] 应该注意的是,在助听器 and 助听装置中,相比于在耳机中采用的近场束形成,束形成被配置用于远场束形成。

[0115] 此外,束形成不能产生净的正效应,除非背景噪声充分超过麦克风噪声。这是因为束形成器的所谓的白色噪声增益 (white-noise-gain), 其中, (输入之间) 不相关的噪声, 诸如麦克风噪声、风噪声和量子化噪声被束形成器放大。

[0116] 对于朝向远场源的有效束形成,在低频,需要大约30dB的净空 (headroom), 而对于朝向近场源的束形成,大约15dB的显著较低的净空就足够了。

[0117] 因此,有时,当背景噪声不够高时,在一个频率范围内,束形成在该范围内必须被禁用以避免噪声的净放大。

[0118] 由于当源处于近场时严格的净空要求,在低频时的大部分时间,通常必须禁用远场束形成器。

[0119] 相反,朝向近场源束形成的近场束形成器在大多数时间的运行通常是不受阻的。结果,当第一束形成器和第二束形成器被配置为近场束形成器时,第三束形成器不可思议地更有效地操作。因此,由于第一束形成器和第二束形成器在大多数时间的运行不受阻,所以,第一束形成器的输出和第二束形成器的输出之间的信噪比中存在显著差异的可能性更高。因此,由于第三束形成器选择性地组合第一束形成器的输出和第二束形成器的输出,所以,信噪比被显著改善。这是因为(具有近场束形成器的)麦克风将不会经常(像远场束形成器那样)引起第一束形成器和第二束形成器被有效地禁用。

[0120] 主要优点在于所要求的耳机和方法组合了朝向近场源(其为用户的嘴巴)的端射阵列束形成的优点,具有用户的头部的噪声和风屏蔽效果的好处,而达到噪声抑制的不可预料的水平。该极大地提高了例如在室外环境中拾取的语音信号的质量,从而提高了在例如电话呼叫的远程端处的语音理解 (speech comprehension)。

[0121] 用于耳机的束形成器(即,近场束形成器)被配置为在距第一对麦克风和/或第二对麦克风的距离小于 $25\text{cm} \pm 10\%$ 或小于约 $20\text{cm} \pm 10\%$ 或小于约 $18\text{cm} \pm 10\%$ 的范围内空间聚焦 (focus) 在源(诸如用户的嘴巴)上。据此,第一对麦克风设置有第一相互距离,第二对麦克风设置有第二相互距离。第一相互距离和/或第二相互距离处于大约 $5\text{mm} \pm 10\%$ 至大约 $20\text{mm} \pm 10\%$ 或大约 $35\text{mm} \pm 10\%$ ,例如,大约10mm或15mm的范围内。

[0122] 聚焦在佩戴耳机的用户的嘴巴上的近场束形成器意味着束形成器被聚焦在用户的嘴巴的开口的位置上或靠近用户的嘴巴,例如,在嘴巴前面几厘米,诸如2、3、4、5、10或15cm。

[0123] 可以通过以频域(复数)表示的以下表达式更详细地描述普通的理想化的双麦克风束形成器:

[0124]  $Z = (X_1 - \Delta_2 \cdot X_2) \cdot EQ$

[0125] 其中,  $X_1$ 和 $X_2$ 是分别来自端射麦克风配置中的前面和后面麦克风的麦克风信号;  $\Delta_2$ 是确定束形成器的方向特征(例如,心型(cardiod)或双向)的时间延迟(相位修改);EQ确定束形成器的输出处的频率特征; $Z$ 为束形成输出,假设由所述表达式表示的束形成器从匹配的麦克风接收其输入。

[0126] 现在研究束形成器对所关注的源的响应。在其后续中,  $X_1$ 和 $X_2$ 由来自共用源的共用源信号 $S$ 和从共用源至麦克风的相应的转移函数 $B_1$ 和 $B_2$ 表达:

[0127]  $X_1 = B_1 \cdot S$

[0128]  $X_2 = B_2 \cdot S$

[0129] 没有一般性损耗,我们指定束形成器应该表现出对作为第一麦克风的源的相同的响应。

[0130]  $Z = B_1 \cdot S$

[0131] 然后

[0132] 
$$EQ = \frac{1}{\left(1 - \Delta_2 \cdot \left(\frac{B_2}{B_1}\right)\right)}$$

[0133] 其产生了用于远场束形成器的如下的:

[0134] 
$$\left|\frac{B_2}{B_1}\right| \cong 1$$

[0135] 这是因为源处于近场中。从以下的表达式可以看出,由于分母接近于0,所以,对于低频,EQ增大。由此产生了非常高的麦克风噪声增益。

[0136] 用于远场束形成器的EQ因此可以以如下方式表达:

[0137] 
$$EQ_{FF} = \frac{1}{(1 - \Delta_2 \cdot \Delta_{12})}$$

[0138] 其中,  $\Delta_{12}$ 是时间延迟(即,相位修改)。

[0139] 对于近场束形成器,从近场源至麦克风对中的麦克风的一个的转移函数 $B_2$ 和从近场至麦克风对中的麦克风的另一个的转移函数 $B_1$ 之间的比率的绝对值等于常数 $a$ (以频域符号或复数符号),即:

[0140] 
$$\left|\frac{B_2}{B_1}\right| = a$$

[0141] 其中, $a$ 的值小于1而大于0,  $0 < a < 1$ 。 $a$ 的值取决于从用户的嘴巴到一对麦克风的分的路径。麦克风对的端射配置给出了相对低的 $a$ 值。 $a$ 的值可以例如约为 $0.7 \pm 10\%$ 或者在0.4至0.9的范围内。 $a$ 的值可以大约是那样的值或处于所关注的频率范围,例如,从大约500Hz  $\pm 10\%$ 或800Hz  $\pm 10\%$ 至大约4KHz  $\pm 10\%$ 或8KHz  $\pm 10\%$ 的频率范围或更宽或更窄的频率范围,的范围内。从以上表达式可以看出,由于 $a$ ,  $EQ_{NF}$ 小于低频处的 $EQ_{FF}$ 。由此产生了更低的麦克风噪声增益并因此产生了束形成器将提高信噪比的更宽的背景噪声。

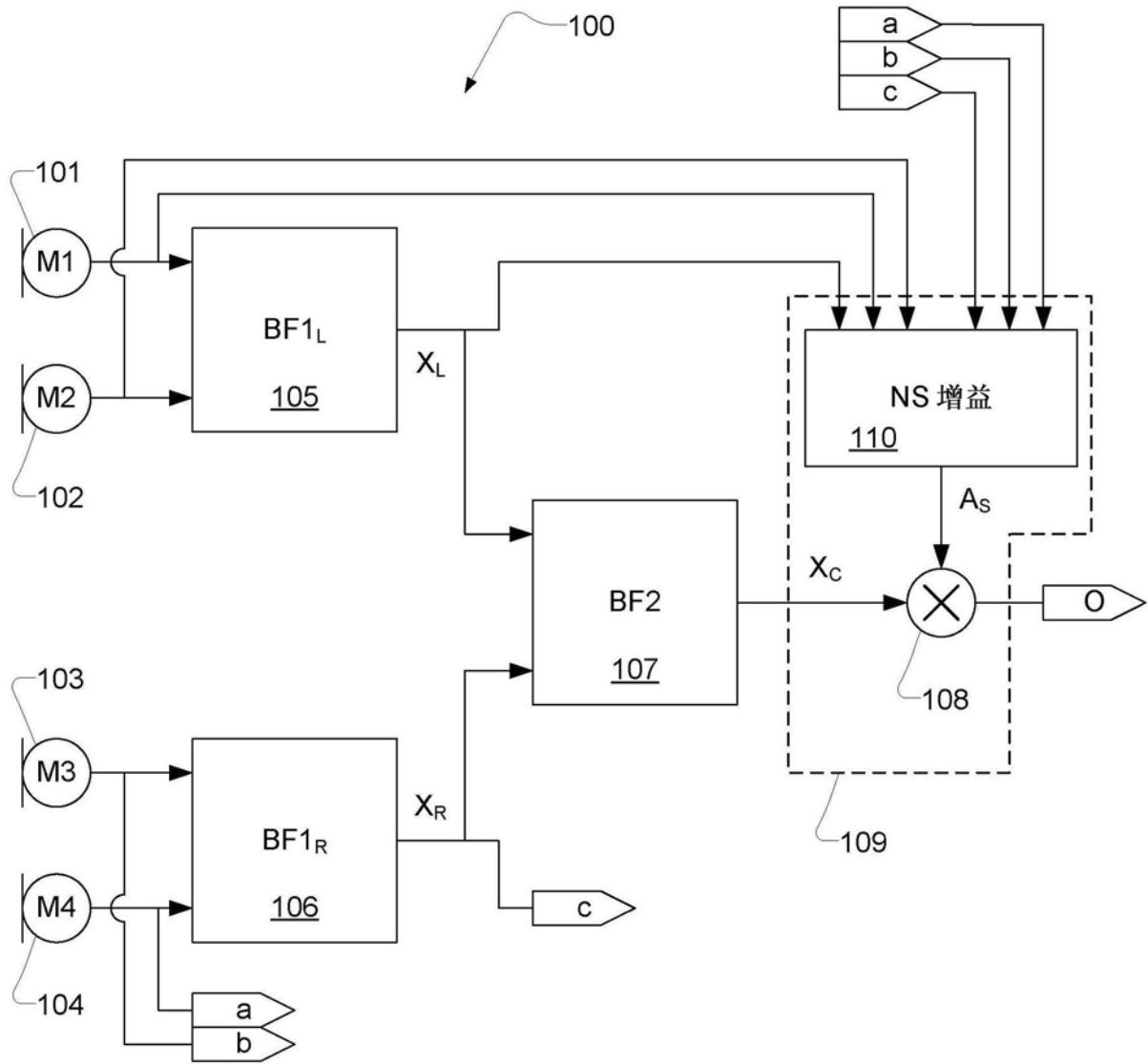


图1



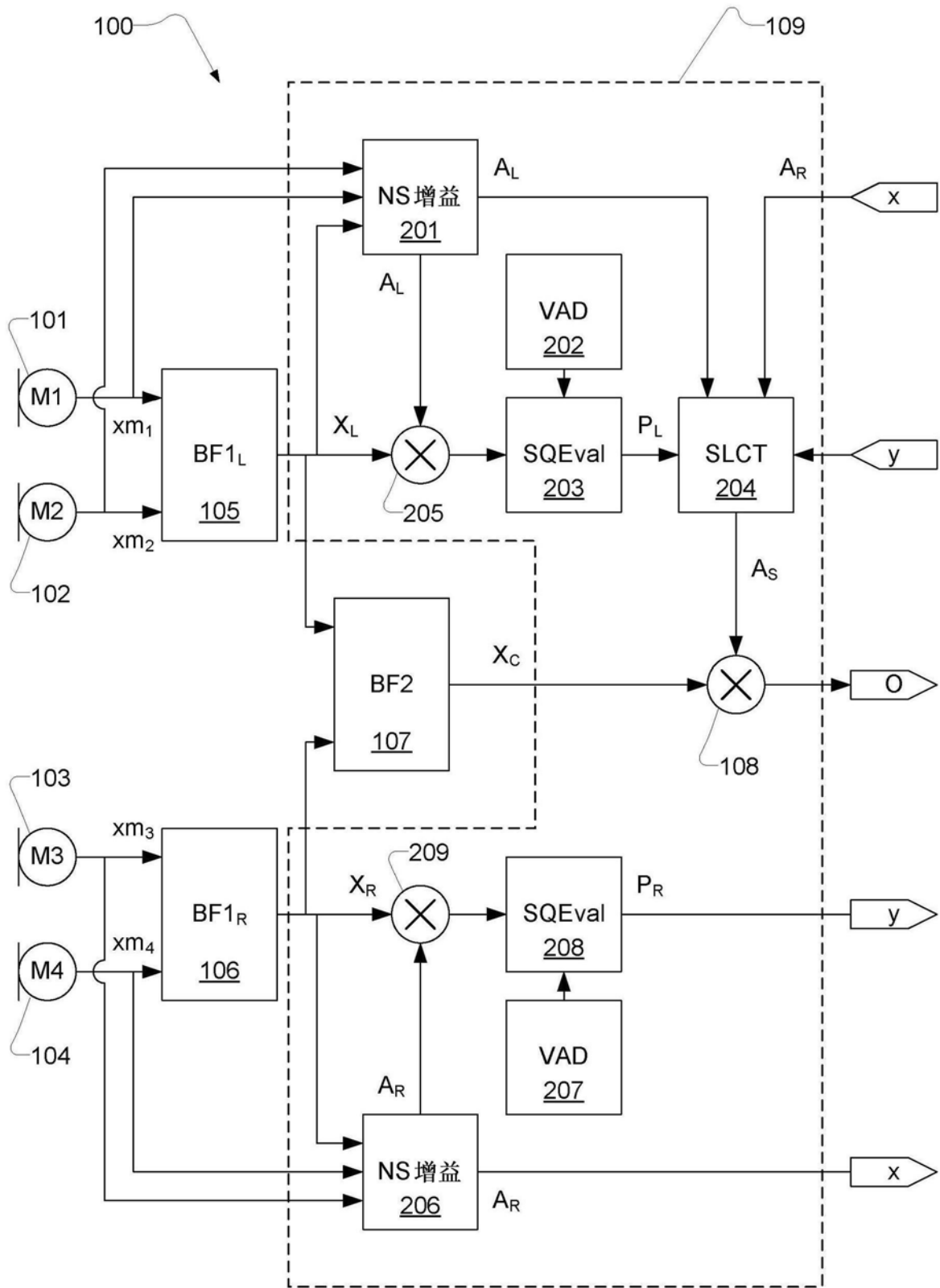


图2

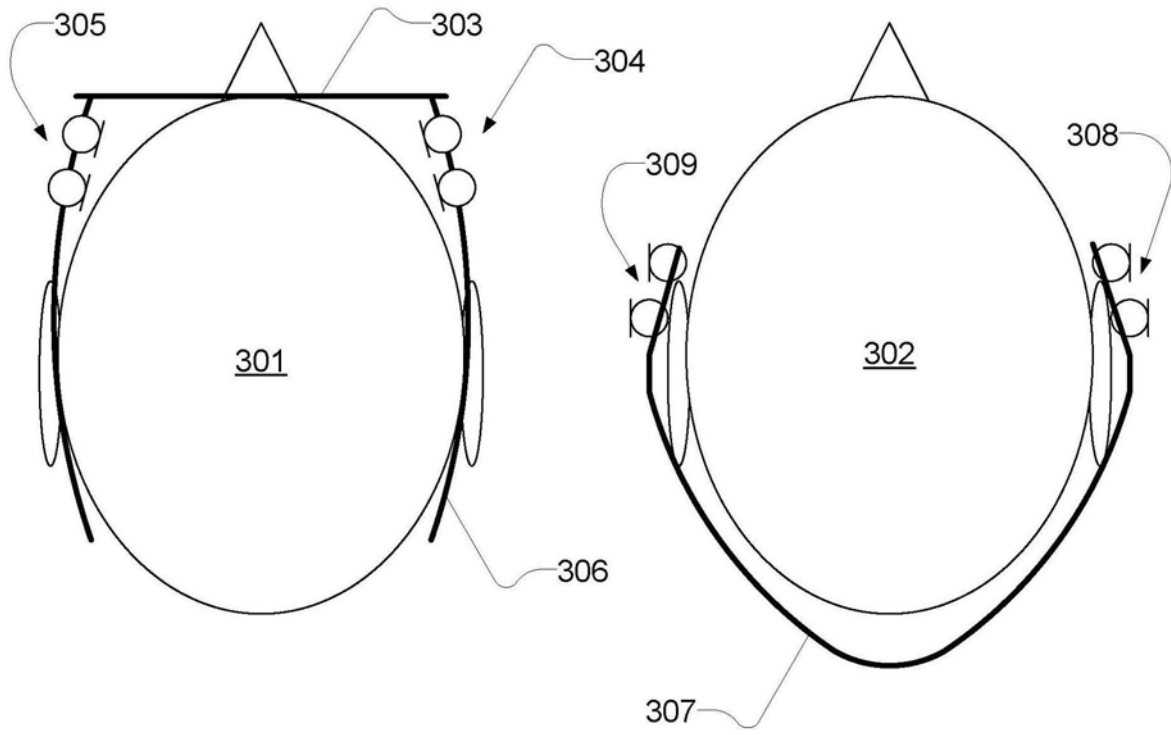


图3