



(12) 发明专利

(10) 授权公告号 CN 102074246 B

(45) 授权公告日 2012. 12. 19

(21) 申请号 201110001041. 9

CN 1719516 A, 2006. 01. 11, 全文.

(22) 申请日 2011. 01. 05

审查员 王馨宁

(73) 专利权人 瑞声声学科技(深圳)有限公司
地址 518057 广东省深圳市南山区高新技术产业园北区新西路 18 号

专利权人 瑞声声学科技研发(南京)有限公司
瑞声光电科技(常州)有限公司

(72) 发明人 叶利剑

(51) Int. Cl.

G10L 21/02 (2006. 01)

G10L 19/14 (2006. 01)

(56) 对比文件

CN 101763858 A, 2010. 06. 30, 全文.

US 2009/0287482 A1, 2009. 11. 19, 全文.

CN 101510426 A, 2009. 08. 19, 全文.

US 2008/0175408 A1, 2008. 07. 24, 全文.

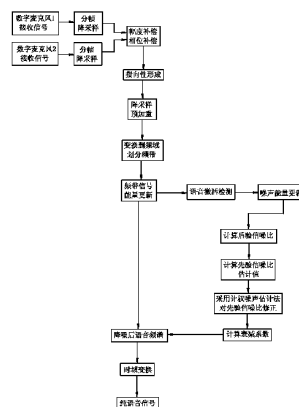
权利要求书 5 页 说明书 14 页 附图 2 页

(54) 发明名称

基于双麦克风语音增强装置及方法

(57) 摘要

本发明提供了一种基于双麦克风语音增强装置,包括数字麦克风阵列模块和与所述数字麦克风阵列模块电连接的信号处理集成芯片。其中,所述信号处理集成芯片内设有:脉冲密度调制解码器模块、指向性形成模块、语音增强处理模块和输出模块。本发明还提供了一种基于本发明双麦克风语音增强装置的语音增强方法。与相关技术相比,本发明的基于双麦克风语音增强装置中将解码、降噪等芯片集成到一块信号处理芯片中,集成度高。两个数字麦克风不同方向进声,采用指向性形成抑制波束外的背景噪声。其语音增强方法简单,节约开发成本。



1. 一种基于双麦克风语音增强装置,包括数字麦克风阵列模块和与所述数字麦克风阵列模块电连接的信号处理集成芯片,所述数字麦克风阵列模块包括用于接收时域带噪语音数字信号并将接收的时域带噪语音数字信号输出的第一数字麦克风和第二数字麦克风,其特征在于:所述信号处理集成芯片内设有:

第一脉冲密度调制解码器模块,其包括滤波器,用于对时域带噪语音数字信号进行分帧、降采样及滤波处理,得到初步降采样率数字信号;

指向性形成模块,用于将所述初步降采样率数字信号进行指向性形成,得到初步降噪的指向性输出信号;

第二脉冲密度调制解码器模块,用于对所述初步降噪的指向性输出信号进行分帧、降采样、滤波和加重处理,得到第二降采样率数字信号;

语音增强处理模块,用于对所述第二降采样率数字信号进行语音增强处理,得到去除噪声的语音信号;

输出模块,用于输出所述去除噪声的语音信号。

2. 根据权利要求1所述的基于双麦克风语音增强装置,其特征在于:所述第一脉冲密度调制解码器模块包括N级梳状滤波器。

3. 根据权利要求2所述的基于双麦克风语音增强装置,其特征在于:所述第二脉冲密度调制解码器模块包括第一半带滤波器、第二半带滤波器、低通滤波器和高通滤波器。

4. 根据权利要求3所述的基于双麦克风语音增强装置,其特征在于:该装置还包括设有内置空间的壳体,所述壳体内设有挡板,所述挡板将所述内置空间隔成至少三个互不相通的收容空间,所述第一数字麦克风、第二数字麦克风和所述信号处理集成芯片分别置于三个所述收容空间内,收容所述第一数字麦克风和第二数字麦克风的收容空间分别设有第一进声孔和第二进声孔且所述第一进声孔和第二进声孔位于所述壳体的相对的面上。

5. 一种基于如权利要求4所述的基于双麦克风语音增强装置的语音增强方法,其特征在于:该方法包括如下步骤:

步骤S001:通过所述第一数字麦克风经所述第一进声孔接收第一时域带噪语音数字信号;通过所述第二数字麦克风经所述第二进声孔接收第二时域带噪语音数字信号,并将接收到的所述第一时域带噪语音数字信号和第二时域带噪语音数字信号传送给所述第一脉冲密度调制解码器模块;

步骤S002:通过所述第一脉冲密度调制解码器模块对所述第一时域带噪语音数字信号和第二时域带噪语音数字信号进行分帧处理,并利用所述N级梳状滤波器对其进行降采样及滤波处理,得到第一初步降采样率数字信号和第二初步降采样率数字信号;

N级梳状滤波器的传递函数形式如下:

$$H(z) = \left(\frac{1}{r} \sum_{i=0}^{r-1} z^{-i} \right)^N, \quad N \text{ 为级联阶数, } r \text{ 为抽取比例};$$

$$x'_{-1}(n, m) = x_1(rn, m)$$

$$x'_{-2}(n, m) = x_2(rn, m), \quad x_1, x_2 \text{ 分别为第一、第二时域带噪语音数字信号};$$

$$x'_{-1}, x'_{-2} \text{ 分别为第一、第二初步降采样率数字信号};$$

$$n \text{ 表示点数, } m \text{ 表示帧数};$$

步骤 S003 :利用所述指向性形成模块接收所述第一脉冲密度调制解码器模块输出的所述第一初步降采样率数字信号和第二初步降采样率数字信号,并对其进行幅度补偿和相位补偿处理,得到第一初步降噪指向性信号和第二初步降噪指向性信号,并将所述第一、第二初步降噪指向性信号相减得到初步降噪指向性输出信号;

经过幅度补偿,形式如下:

$$x_1^* = a_m x_1$$

$$x_2^* = a_m x_2$$

其中, x_1^* 、 x_2^* 为分别为第一幅度补偿初步降噪指向性信号和第二幅度补偿初步降噪指向性信号, a_m 为常数,其由所述第一数字麦克风和第二数字麦克风的灵敏度差异决定;

所述相位补偿的形式如下:

$$x_1^{**}(n, m) = x_1^*(n - [f_s \frac{d}{c_s}], m)$$

$$x_2^{**}(n, m) = x_2^*(n - [f_s \frac{d}{c_s}], m)$$

其中, x_1^{**} 、 x_2^{**} 分别为第一相位补偿初步降噪指向性信号和第二相位补偿初步降噪指向性信号, m 为帧数, n 为点数, f_s 为采样率, d 为所述第一数字麦克风与第二数字麦克风的距离, c_s 为声速;

所述初步降噪指向性输出信号如下:

$$y(n, m) = x_1^{**}(n, m) - x_2^{**}(n, m)$$

步骤 S004 :通过所述第二脉冲密度调制解码器模块对所述初步降噪指向性输出信号进行分帧、降采样、滤波和预加重处理,得到第二降采样率数字信号;

对所述初步降噪指向性输出信号进行分帧后的信号设为 $y(m)$, m 表示帧数,降采样及滤波处理分为三步:

第一步:先通过所述第一半带滤波器按 2 : 1 的比例对所述初步降噪指向性输出信号抽取采样;

第二步:再通过所述第二半带滤波器按 2 : 1 的比例对上述第一步采样后的指向性输出信号再抽取采样;

第三步:然后通过所述低通滤波器按 3 : 1 的比例对上述第二步采样后的指向性输出信号继续抽取采样;

通过所述高通滤波器对上述第三步采样后的指向性输出信号进行预加重处理,得到第二降采样率数字信号;

步骤 S005 :通过所述语音增强处理模块对所述第二降采样率数字信号进行语音增强处理,得到去除背景噪声的语音信号,包括如下步骤:

步骤 A001: 将第二降采样率数字信号经短时傅里叶变换成频域信号,并将该频域信号划分为若干频带,计算各个频带的能量并进行平滑处理;

短时傅里叶变换如下：

$$X(f, m) = \frac{2}{M} \sum_{n=0}^{M-1} \text{win}(n-m) \times x(m) e^{-2\pi j f \frac{n}{M}}$$

其中，M 为短时傅里叶变换的计算长度，f 表示频率值，X 表示频域信号，x 表示第二降采样率数字信号；

汉明窗函数定义如下：

$$\text{win}(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi * n/M) & 0 \leq n \leq M-1 \\ 0 & \text{其余 } n \end{cases}$$

采用如下方法计算每个频带的能量并进行平滑：

$$E(m, k) = |X(m, k)|^2 \quad 0 \leq k \leq N-1$$

$$Y_E(m, k) = \alpha Y_E(m-1, k) + (1-\alpha) E(m, k) \quad 0 \leq k \leq N-1$$

其中， $Y_E(m, k)$ 表示经平滑后的每个频带区间的能量，m 表示当前帧的序号，k 表示当前的子带的序号， α 为平滑因子，此处 $\alpha = 0.75$ ；N 为选取的频带总数， $E(m, k)$ 表示频带能量值， $X(m, k)$ 表示第 m 帧第 k 个频带的频域信号；

步骤 A002: 用计算器计算先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ ；

设初始的噪声能量估计值 $V(0, k) = 0$ ，初始的先验信噪比估计值 $\hat{SNR}_{prior}(0, k) = 0$ ；

由各频带的能量 $Y_E(m, k)$ 及前一帧得到的噪声能量估计值 $V(m-1, k)$ ，计算各频带当前帧的后验信噪比 $SNR_{post}(m, k) = \frac{Y_E(m, k)}{V(m-1, k)}$ ，并由前一帧的先验信噪比估计值

$\hat{SNR}_{prior}(m-1, k)$ 得到当前帧的先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ ；

$$\hat{SNR}_{prior}(m, k) = \beta \frac{|\hat{S}(m-1, k)|^2}{V(m, k)} + (1-\beta) \max[0, SNR_{post}(m, k) - 1]$$

$\hat{S}(m-1, k)$ 表示前一帧最后得到的增强后语音信号， β 为第一平滑系数， $V(m, k)$ 为噪声能量估计值；

步骤 A003: 对先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ 进行修正，得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ ；

先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ ，乘以计权因子去 q_0 ，得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ ，计权因子计算如下：

$$q_{\theta} = \begin{cases} 1 & \hat{SNR}_{prior}(m, k) < \gamma_1 \\ \frac{1}{\gamma_1 - \gamma_2} (\hat{SNR}_{prior}(m, k) - \gamma_2) & \gamma_1 < \hat{SNR}_{prior}(m, k) < \theta_z \\ 0 & \hat{SNR}_{prior}(m, k) > \theta_z \end{cases}$$

其中 γ_1 取 1.5, γ_2 取 200, θ_z 取 20;

步骤 A004: 根据修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$, 计算各个频带的衰减增益 $q(m, k)$:

$$q(m, k) = \begin{cases} 1 - \frac{a - 0.5 \cdot \lg[\hat{SNR}'_{prior}(m, k)]}{SNR_{post}} \end{cases}$$

其中, 对于不同频带, a 为不同的常数;

步骤 A005: 将当前帧的各频带的频域信号 $X(m, k)$, 乘以相应频带的衰减增益, 得到该频带的增强后的语音信号 $\hat{S}(k)$:

$$\hat{S}(m, k) = q(m, k) * X(m, k) \quad 0 \leq k \leq N-1$$

步骤 A006: 利用修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ 对当前帧进行判断, 判断当前帧是否是噪声并根据判决结果对各频带的噪声能量估计值进行更新:

如果判断为噪声, 则噪声能量估计值等于前一帧的值 $V(m, k) = V(m-1, k)$, 否则采用频带能量 $Y_E(m, k)$, 对噪声能量估计值 $V(m, k)$ 进行更新, 并将此噪声能量估计值 $V(m, k)$, 用于下一帧的步骤 A002 中, 进行后验信噪比的估计:

$$V(m, k) = \mu V(m-1, k) + (1 - \mu) \cdot Y_E(m, k)$$

μ 表示第二平滑系数;

步骤 A007: 将增强后的语音信号 $\hat{S}(k)$ 用短时傅里叶变换变换成时域信号, 将该时域信号经过低通滤波器, $H(z) = 1 + bz^{-1}$, 其中 b 为一般常数, 进行去加重处理, 得到去除背景噪声的语音信号;

步骤 S006: 通过所述输出模块接收所述语音增强处理模块输出的所述去除背景噪声的语音信号并输出。

6. 根据权利要求 5 所述的基于双麦克风语音增强装置的语音增强方法, 其特征在于, 步骤 A004 中还包括采用门限判决对所述衰减增益进行调整;

首先设定一阈值、调整值 q_{mod} 和门限值 q_{floor} ;

其次以当前帧修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ 为判断, 小于所述阈值的衰减增益均乘以调整值 q_{mod} , 从而进一步抑制噪声;

接着将所有小于等于门限值的衰减增益调整到门限值 q_{floor} , 方法如下:

$$q'(m,k) = \begin{cases} q_{\text{mod}} q(m,k), & \widehat{SNR}'_{\text{prior}}(m,k) < \theta_G \\ q(m,k), & \text{否则} \end{cases}$$

$$q'(m,k) = \begin{cases} q(m,k), & q(k) > q_{\text{floor}} \\ q_{\text{floor}}, & \text{否则} \end{cases}, \quad \text{其中 } q_{\text{mod}} = 0.1, \theta_G = 1.2, q_{\text{floor}} = 0.01。$$

基于双麦克风语音增强装置及方法

【技术领域】

[0001] 本发明涉及一种语音增强装置及方法,尤其涉及一种基于双麦克风语音增强装置及方法。

【背景技术】

[0002] 随着无线通讯的发展,全球移动电话用户越来越多,用户对移动电话的要求已不仅满足于通话,而且要能够提供高质量的通话效果,尤其是目前移动多媒体技术的发展,移动电话的通话质量更显重要。

[0003] 由于大量环境噪声的存在,手机等通讯设备的麦克风采集到的语音信号普遍信噪比不够高,特别是在街道汽车等高噪声环境中,需要提高音量才能使对方听清。所以需要通过语音增强的方法来提升输入语音的信噪比,改善通讯质量。相关技术单通道语音增强方法降噪效果有限,并且会对语音造成较大失真,而双麦克风语音增强装置则可以有效的提高信噪比,相关技术的双麦克风语音增强装置包括麦克风和多种信号处理芯片。

[0004] 然而,相关技术的双麦克风语音增强装置中的多种信号处理芯片是分开设置的,这种结构使得相关技术的双麦克风语音增强装置的语音增强方法较为复杂,开发成本高。

[0005] 因此,有必要提出一种新的基于双麦克风语音增强装置来解决上述问题。

【发明内容】

[0006] 本发明需解决的技术问题是提供一种能优化语音增强方法的基于双麦克风语音增强装置。

[0007] 根据上述的技术问题,设计了一种基于双麦克风语音增强装置,其目的是这样实现的:一种基于双麦克风语音增强装置,包括数字麦克风阵列模块和与所述数字麦克风阵列模块电连接的信号处理集成芯片,所述数字麦克风阵列模块包括用于接收时域带噪语音数字信号并将接收的时域带噪语音数字信号输出的第一数字麦克风和第二数字麦克风。其中,所述信号处理集成芯片内设有:

[0008] 第一脉冲密度调制解码器模块,其包括滤波器,用于对时域带噪语音数字信号进行分帧、降采样及滤波处理,得到初步降采样率数字信号;

[0009] 指向性形成模块,用于将所述初步降采样率数字信号进行指向性形成,得到初步降噪的指向性输出信号;

[0010] 第二脉冲密度调制解码器模块,用于对所述初步降噪的指向性输出信号进行分帧、降采样、滤波和加重处理,得到第二降采样率数字信号;

[0011] 语音增强处理模块,用于对所述第二降采样率数字信号进行语音增强处理,得到去除噪声的语音信号;

[0012] 输出模块,用于输出所述去除噪声的语音信号。

[0013] 优选的,所述第一脉冲密度调制解码器模块包括N级梳状滤波器。

[0014] 优选的,所述第二脉冲密度调制解码器模块包括第一半带滤波器、第二半带滤波

器、低通滤波器和高通滤波器。

[0015] 优选的,该装置还包括设有内置空间的壳体,所述壳体内设有挡板,所述挡板将所述内置空间隔成至少三个互不相通的收容空间,所述第一数字麦克风、第二数字麦克风和所述信号处理集成芯片分别置于三个所述收容空间内,收容所述第一数字麦克风和第二数字麦克风的收容空间分别设有第一进声孔和第二进声孔且所述第一进声孔和第二进声孔位于所述壳体的相对的面上。

[0016] 本发明还提供一种基于本发明所述的基于双麦克风语音增强装置的语音增强方法,该方法包括如下步骤:

[0017] 步骤 S001:通过所述第一数字麦克风经所述第一进声孔接收第一时域带噪语音数字信号;通过所述第二数字麦克风经所述第二进声孔接收第二时域带噪语音数字信号,并将接收到的所述第一时域带噪语音数字信号和第二时域带噪语音数字信号传送给所述第一脉冲密度调制解码器模块。

[0018] 步骤 S002:通过所述第一脉冲密度调制解码器模块对所述第一时域带噪语音数字信号和第二时域带噪语音数字信号进行分帧处理,并利用所述 N 级梳状滤波器对其进行降采样及滤波处理,得到第一初步降采样率数字信号和第二初步降采样率数字信号。

[0019] N 级梳状滤波器的传递函数形式如下:

$$[0020] \quad H(z) = \left(\frac{1}{r} \sum_{i=0}^{r-1} z^{-i} \right)^N, \quad N \text{ 为级联阶数, } r \text{ 为抽取比例};$$

$$[0021] \quad x'_1(n, m) = x_1(rn, m)$$

$$[0022] \quad x'_2(n, m) = x_2(rn, m), \quad x_1, x_2 \text{ 分别为第一、第二时域带噪语音数字信号};$$

$$[0023] \quad x'_1, x'_2 \text{ 分别为第一、第二初步降采样率数字信号 } n \text{ 表示点数, } m \text{ 表示帧数}.$$

[0024] 步骤 S003:利用所述指向性形成模块接收所述第一脉冲密度调制解码器模块输出的所述第一初步降采样率数字信号和第二初步降采样率数字信号,并对其进行幅度补偿和相位补偿处理,得到第一初步降噪指向性信号和第二初步降噪指向性信号,并将所述第一、第二初步降噪指向性信号相减得到初步降噪指向性输出信号。

[0025] 所述幅度补偿,形式如下:

$$[0026] \quad x_1^* = a_m x_1'$$

$$[0027] \quad x_2^* = a_m x_2'$$

[0028] 其中, x_1^* 、 x_2^* 为分别为第一幅度补偿初步降噪指向性信号和第二幅度补偿初步降噪指向性信号, a_m 为常数,其由所述第一数字麦克风和第二数字麦克风的灵敏度差异决定。

[0029] 所述相位补偿的形式如下:

$$[0030] \quad x_1^{**}(n, m) = x_1^*(n - [f_s \frac{d}{c_s}], m)$$

$$[0031] \quad x_2^{**}(n, m) = x_2^*(n - [f_s \frac{d}{c_s}], m)$$

[0032] 其中, x_1^{**}, x_2^{**} 分别为第一相位补偿初步降噪指向性信号和第二相位补偿初步降噪指向性信号, m 为帧数, n 为点数, f_s 为采样率, d 为所述第一数字麦克风与第二数字麦克风的距离, c_s 为声速。

[0033] 所述初步降噪指向性输出信号如下:

$$[0034] \quad y(n, m) = x_1^{**}(n, m) - x_2^{**}(n, m)$$

[0035] 步骤 S004: 通过所述第二脉冲密度调制解码器模块对所述初步降噪指向性输出信号进行分帧、降采样、滤波和预加重处理, 得到第二降采样率数字信号。

[0036] 对所述初步降噪指向性输出信号进行分帧后的信号设为 $y(m)$, m 表示帧数, 降采样及滤波处理分为三步:

[0037] 第一步: 先通过所述第一半带滤波器按 2 : 1 的比例对所述初步降噪指向性输出信号抽取采样。

[0038] 第二步: 再通过所述第二半带滤波器按 2 : 1 的比例对上述第一步采样后的指向性输出信号再抽取采样。

[0039] 第三步: 然后通过所述低通滤波器按 3 : 1 的比例对上述第二步采样后的指向性输出信号继续抽取采样。

[0040] 通过所述高通滤波器对上述第三步采样后的指向性输出信号进行预加重处理, 得到第二降采样率数字信号。

[0041] 步骤 S005: 通过所述语音增强处理模块对所述第二降采样率数字信号进行语音增强处理, 得到去除背景噪声的语音信号, 包括如下步骤:

[0042] 步骤 A001: 将第二降采样率数字信号经短时傅里叶变换成频域信号, 并将该频域信号划分为若干频带, 计算各个频带的能量并进行平滑处理。短时傅里叶变换如下:

$$[0043] \quad X(f, m) = \frac{2}{M} \sum_{n=0}^{M-1} w \sin(n - m) \times x(m) e^{-2\pi j f \frac{n}{M}}$$

[0044] 其中, M 为短时傅里叶变换的计算长度, f 表示频率值, X 表示频域信号, x 表示第二降采样率数字信号。

[0045] 汉明窗函数定义如下:

$$[0046] \quad \text{win}(n) = \{$$

$$[0047] \quad 0.54 - 0.46 \cos(2\pi * n/M) \quad 0 \leq n \leq M-1$$

$$[0048] \quad 0 \quad \text{其余 } n$$

$$[0049] \quad \}$$

[0050] 采用如下方法计算每个频带的能量并进行平滑:

$$[0051] \quad E(m, k) = |X(m, k)|^2 \quad 0 \leq k \leq N-1$$

$$[0052] \quad Y_E(m, k) = \alpha Y_E(m-1, k) + (1-\alpha) E(m, k) \quad 0 \leq k \leq N-1$$

[0053] 其中, $Y_E(m, k)$ 表示经平滑后的每个频带区间的能量, m 表示当前帧的序号, k 表示当前的子带的序号, α 为平滑因子, 此处 $\alpha = 0.75$; N 为选取的频带总数, $E(m, k)$ 表示频带

能量值, $X(m, k)$ 表示第 m 帧第 k 个频带的频域信号。

[0054] 步骤 A002: 用计算器计算先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$;

[0055] 设初始的噪声能量估计值 $V(0, k)=0$, 初始的先验信噪比估计值 $\hat{SNR}_{prior}(0, k)=0$;

[0056] 由各频带的能量 $Y_E(m, k)$ 及当前帧的前一帧得到的噪声能量估计值 $V(m-1, k)$, 计算各频带当前帧的后验信噪比 $SNR_{post}(m, k) = \frac{Y_E(m, k)}{V(m-1, k)}$, 并由前一帧的先验信噪比估

计值 $\hat{SNR}_{prior}(m-1, k)$ 得到当前帧的先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$;

$$[0057] \quad \hat{SNR}_{prior}(m, k) = \beta \frac{|\hat{S}(m-1, k)|^2}{V(m, k)} + (1 - \beta) \max[0, SNR_{post}(m, k) - 1]$$

[0058] $\hat{S}(m-1, k)$ 表示前一帧最后得到的增强后语音信号, β 为第一平滑系数, $V(m, k)$ 为噪声能量估计值。

[0059] 步骤 A003: 对先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ 进行修正, 得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$;

[0060] 先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$, 乘以计权因子去 q_θ , 得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$, 计权因子计算如下:

$$[0061] \quad q_\theta = \begin{cases} 1 & \hat{SNR}_{prior}(m, k) < \gamma_1 \\ \frac{1}{\gamma_1 - \gamma_2} (\hat{SNR}_{prior}(m, k) - \gamma_2) & \gamma_1 < \hat{SNR}_{prior}(m, k) < \theta_z \\ 0 & \hat{SNR}_{prior}(m, k) > \theta_z \end{cases}$$

[0062] 其中 γ_1 取 1.5, γ_2 取 200, θ_z 取 20。

[0063] 步骤 A004: 根据修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$, 计算各个频带的衰减增益 $q(m, k)$:

$$[0064] \quad q(m, k) = \begin{cases} a - 0.5 \cdot \lg[\hat{SNR}'_{prior}(m, k)] \\ SNR_{post} \end{cases}$$

[0065] 其中, 对于不同频带, a 为不同的常数。

[0066] 步骤 A005: 将当前帧的各频带的频域信号 $X(m, k)$, 乘以相应频带的衰减增益, 得

到该频带的增强后的语音信号 $\hat{S}(k)$;

[0067] $\hat{S}(m,k) = q(m,k) * X(m,k) \quad 0 \leq k \leq N-1$

[0068] 步骤 A006: 利用修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m,k)$ 对当前帧进行判断, 判断当前帧是否是噪声并根据判决结果对各频带的噪声能量估计值进行更新;

[0069] 如果判断为噪声, 则噪声能量估计值等于前一帧的值 $V(m,k) = V(m-1,k)$, 否则采用频带能量 $Y_E(m,k)$, 对噪声能量估计值 $V(m,k)$ 进行更新, 并将此噪声能量估计值 $V(m,k)$, 用于下一帧的步骤 A002 中, 进行后验信噪比的估计;

[0070] $V(m,k) = \mu V(m-1,k) + (1-\mu) \cdot Y_E(m,k)$

[0071] μ 表示第二平滑系数。

[0072] 步骤 A007: 将增强后的语音信号 $\hat{S}(k)$ 用短时傅里叶变换变换成时域信号, 将该时域信号经过低通滤波器, $H(z) = 1 + bz^{-1}$, 其中 b 为一般常数, 进行去加重处理, 得到去除背景噪声的语音信号。

[0073] 步骤 S006: 通过所述输出模块接收所述语音增强处理模块输出的所述去除背景噪声的语音信号并输出。

[0074] 优选的, 步骤 A004 中还包括采用门限判决对所述衰减增益进行调整;

[0075] 首先设定一阈值、调整值 q_{mod} 和门限值 q_{floor} ;

[0076] 其次以当前帧修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m,k)$ 为判断, 小于所述阈值的衰减增益均乘以调整值 q_{mod} , 从而进一步抑制噪声。

[0077] 接着将所有小于等于门限值的衰减增益调整到门限值 q_{floor} , 方法如下:

[0078]

$$q'(m,k) = \begin{cases} q_{mod}q(m,k), & \hat{SNR}'_{prior}(m,k) < \theta_G \\ q(m,k), & \text{否则} \end{cases}$$

[0079] $q'(m,k) = \begin{cases} q(m,k), & q(k) > q_{floor} \\ q_{floor}, & \text{否则} \end{cases}$ 其中 $q_{mod} = 0.1$, $\theta_G = 1.2$, $q_{floor} = 0.01$ 。

【附图说明】

[0080] 图 1 为本发明基于双麦克风语音增强装置的方法的流程示意图。

[0081] 图 2 为本发明基于双麦克风语音增强装置的结构框图。

[0082] 图 3 为本发明基于双麦克风语音增强装置的方法的非线性计权曲线图。

【具体实施方式】

[0083] 下面结合附图和实施方式对本发明作进一步说明。

[0084] 如图 1-3 所示, 一种基于双麦克风语音增强装置, 包括数字麦克风阵列模块和与所述数字麦克风阵列模块电连接的信号处理集成芯片 3, 所述数字麦克风阵列模块包括用于接收时域带噪语音数字信号并将接收的时域带噪语音数字信号输出的第一数字麦克风 1

和第二数字麦克风 2。其中,所述信号处理集成芯片 3 内设有:

[0085] 第一脉冲密度调制解码器模块,其包括滤波器,用于对时域带噪语音数字信号进行分帧、降采样及滤波处理,得到初步降采样率数字信号;

[0086] 指向性形成模块,用于将所述初步降采样率数字信号进行指向性形成,得到初步降噪的指向性输出信号;

[0087] 第二脉冲密度调制解码器模块,用于对所述初步降噪的指向性输出信号进行分帧、降采样、滤波和加重处理,得到第二降采样率数字信号;

[0088] 语音增强处理模块,用于对所述第二降采样率数字信号进行语音增强处理,得到去除噪声的语音信号;

[0089] 输出模块,用于输出所述去除噪声的语音信号。

[0090] 优选的,所述第一脉冲密度调制解码器模块包括 N 级梳状滤波器。

[0091] 优选的,所述第二脉冲密度调制解码器模块包括第一半带滤波器、第二半带滤波器、低通滤波器和高通滤波器。

[0092] 优选的,该装置还包括设有内置空间的壳体 5,壳体 5 内设有挡板 4,挡板 4 将所述内置空间隔成至少三个互不相通的收容空间(示标号),所述第一数字麦克风 1、第二数字麦克风 2 和所述信号处理集成芯片 3 分别置于三个所述收容空间内,收容第一数字麦克风 1 和第二数字麦克风 2 的收容空间分别设有第一进声孔 11 和第二进声孔 21 且第一进声孔 11 和第二进声孔 21 位于所述壳体的相对的面上。

[0093] 本发明还提供一种基于本发明的双麦克风语音增强装置的语音增强方法,详细而言,该方法中的降噪算法基本步骤如下:

[0094] 1、对第一数字麦克风和第二数字麦克风分别接收到的第一时域带噪语音数字信号和第二时域带噪语音数字信号进过分帧,降采样滤波处理,得到初步降采样率数字信号。

[0095] 2、初步降采样率数字信号通过指向性模块对其幅度补偿及相位补偿处理,得到初步降噪指向性输出信号。

[0096] 3、初步降噪指向性输出信号经过多级降采样滤波处理,得到第二降采样率数字信号。

[0097] 4、将第二降采样率数字信号经过加窗、预加重、短时傅里叶变换处理,得到频域内的第二降采样率数字信号并将其划分为若干频带,计算各个频带能量并对各个频带能量进行平滑处理,得到经平滑处理后的每个频带区间的能量及噪声能量估计值。

[0098] 5、根据每个频带区间的能量及噪声能量估计值,计算各个频带当前帧的后验信噪比,并由当前帧的前一帧的先验信噪比估计值得到当前帧的先验信噪比估计值。

[0099] 6、对得到的先验信噪比估计值采用计权噪声估计法进行修正。

[0100] 7、根据得到的修正先验信噪比估计值,计算各个频带的衰减增益因子。

[0101] 8、采用门限判决对增益系数进行调整。

[0102] 9、用得到的衰减增益对各个频带的信号频谱进行处理。

[0103] 10、由修正的先验信噪比估计值对当前帧的信号进行判断,判断此信号是否是噪声。

[0104] 11、根据噪声判决结果对各频带的噪声估计值进行更新。

[0105] 12、将处理后的频域信号变换到时间域,再对变换到时间域的信号进行去加重处

理,得到输出信号。

[0106] 下面以具体实例对该语音增强方法进行说明:

[0107] 步骤 S001:利用第一数字麦克风从第一进声孔接收第一时域带噪语音数字信号 $x_1(n)$;利用第二数字麦克风从第二进声孔接收第二时域带噪语音数字信号 $x_2(n)$,并将所述接收到的第一、第二时域带噪语音数字信号 $x_1(n)$, $x_2(n)$ 传送给脉冲密度调制解码器模块;

[0108] 步骤 S002:通过脉冲密度调制解码器模块对第一、第二时域带噪语音数字信号 $x_1(n)$, $x_2(n)$ 按 16 : 1 的信号抽取比例进行分帧处理,分帧后的第一、第二时域带噪语音数字信号的采样率为 3.072MHz,再经降采样后其输出频率为 192kHz;

[0109] 所谓对第一、第二时域带噪语音数字信号 $x_1(n)$, $x_2(n)$ 进行分帧,是指将第一、第二时域带噪语音数字信号 $x_1(n)$, $x_2(n)$ 以帧为单位等分成若干带噪信号单元。所述带噪信号单元由采样点组成,本发明中数字麦克风采样频率为 3.072MHz,采用 64 倍降采样率,根据短时谱分析的需要,帧长设定成 10~35ms 之间,本实施方式以 16ms 分帧,即一帧带噪信号单元设有 49152 个采样点。当然,任意一帧带噪信号单元都具有一定的帧长,本发明中任意帧的帧长为 49152。

[0110] 为了防止相邻两帧的带噪信号单元间的块效应,在分帧时要使相邻两帧的带噪信号单元之间有一定的混叠部分,即,本帧数据中有 D 个数据为前一帧数据的部分数据,其中混叠部分描述如下:

$$[0111] \quad s(n)=d(m, D+n) \quad 0 \leq n < L, i = 1, 2$$

[0112] 其中 x_i 表示输入带噪语音信号, i 取 1 和 2 分别表示两路信号

$$[0113] \quad d_i(m, n)=d_i(m-1, L+n) \quad 0 \leq n < D$$

[0114] 其中, d_i 表示当前帧的 49152 点采样信号,因为任意一帧的长度为 49152,重叠率为 75%,所以重叠部分的采样点个数 $D=36864$ 。相邻帧的带噪信号单元的第一个采样点相隔的距离 $L=49152-36864 = 12288$ 。

[0115] 本发明相邻两帧的带噪信号单元之间可以具有 50%~75% 的重叠率。本实施方式选取相邻两帧的带噪信号单元之间具有 75% 的重叠率,即以本帧的前 75% (36864 点) 的带噪信号单元和本帧的前一帧的后 75% (36864 点) 的带噪语音信号单元一致。

[0116] 降采样滤波器采用 5 级的梳状滤波器,梳状滤波器的传递函数形式如下:

$$[0117] \quad H(z)=\left(\frac{1}{r} \sum_{i=0}^{r-1} z^{-i}\right)^N$$

[0118] N 为级联阶数取 5, r 为抽取比例取 16

$$[0119] \quad x'_{-1}(n, m) = x_1(16n, m)$$

$$[0120] \quad x'_{-2}(n, m) = x_2(16n, m)$$

[0121] 得到第一、第二初步降采样的信号 x'_{-1} , x'_{-2} , 采样率降低为 192kHz, 其中, n 表示点数, m 表示帧数。

[0122] 步骤 S003:利用指向性形成模块接收脉冲密度调制解码器模块输出的第一、第二初步降采样信号 x'_{-1} , x'_{-2} , 并对第一、第二初步降采样信号 x'_{-1} , x'_{-2} 进行幅度补偿和相位补偿,得到第一、第二初步降噪指向性信号 x_1^{**}, x_2^{**} , 并将第一、第二初步降指向性信号

x_1^{**}, x_2^{**} 相减, 得到初步降噪指向性输出信号 $y(n, m)$ 。

[0123] 所述幅度补偿, 形式如下:

$$[0124] \quad x_1^* = a_m x_1^{'}$$

$$[0125] \quad x_2^* = a_m x_2^{'}$$

[0126] 其中 a_m 为常数, 其由第一数字麦克风和第二数字麦克风的灵敏度差异决定。

[0127] 所述相位补偿的形式如下:

$$[0128] \quad x_1^{**}(n, m) = x_1^*(n - \lfloor f_s \frac{d}{c_s} \rfloor, m)$$

$$[0129] \quad x_2^{**}(n, m) = x_2^*(n - \lfloor f_s \frac{d}{c_s} \rfloor, m)$$

[0130] 其中 m 为帧数, n 为点数, f_s 为采样率, 其值取 192kHz, d 为第一数字麦克风和第二数字麦克风的距离, c_s 为声速。

[0131] 所述初步降噪指向性输出信号 $y(n, m)$ 如下:

$$[0132] \quad y(n, m) = x_1^{**}(n, m) - x_2^{**}(n, m)$$

[0133] 步骤 S004: 利用第二脉冲密度调制解码器模块对初步降噪指向性输出信号进行分帧、降采样、滤波和预加重处理, 得到第二降采样率数字信号, 具体为:

[0134] 对所述初步降噪信号进行分帧处理后的信号为 $y(m)$, m 表示帧数, 降采样及滤波分为三步:

[0135] 第一步: 先通过所述第一半带滤波器按 2 : 1 的比例对所述初步降噪指向性输出信号抽取采样, 使其采样率降至 96kHz;

[0136] 第二步: 再通过所述第二半带滤波器按 2 : 1 的比例对上述第一步采样后的指向性输出信号再抽取采样, 使其采样率降至 48kHz;

[0137] 第三步: 然后通过所述低通滤波器按 3 : 1 的比例对上述第二步采样后的指向性输出信号继续抽取采样, 使其采样率降至 16kHz;

[0138] 第一半带滤波器为 14 阶有限冲激响应滤波器:

$$[0139] \quad y_1^{'}(n, m) = \sum_{i=0}^{14} a_1(i) y(n-i, m)$$

$$[0140] \quad y_1(n, m) = y_1^{'}(2n, m)$$

[0141] 第二半带滤波器为 190 阶有限冲激响应滤波器:

$$[0142] \quad y_2^{'}(n, m) = \sum_{i=0}^{190} a_2(i) y_1(n-i, m)$$

$$[0143] \quad y_2(n, m) = y_2^{'}(2n, m)$$

[0144] 低通滤波器为 86 阶有限冲激响应滤波器:

$$[0145] \quad y_3^{'}(n, m) = \sum_{i=0}^{86} a_3(i) y_2(n-i, m)$$

$$[0146] \quad y_3(n, m) = y_3^{'}(3n, m)$$

[0147] 通过所述高通滤波器对上述第三步采样后的指向性输出信号进行预加重处理,得到第二降采样率数字信号。

[0148] 高通滤波器其形式如下:

[0149] $H(z) = 1 - \alpha z^{-1}$, 其中常数 $\alpha = 0.9325$

[0150] 步骤 S005: 利用语音增强处理模块对第二降采样率数字信号进行语音增强处理, 形成去除背景噪声的语音信号;

[0151] 由于语音信号是短时平稳的, 所以对语音信号进行分帧处理是可行的, 但分帧又会带来帧信号边界处的不连续而造成的频率泄露。所以要对帧信号进行短时傅里叶变换(STFT)。短时傅里叶变换可以理解为对帧信号先加窗再做傅里叶变换。加窗函数的目的就是为在做短时傅里叶变换时, 减少帧信号边界处的不连续造成的频率泄露。本实施方式中使用了一个长度等于帧长 256 点的汉明窗, 此汉明窗可以有效的降低吉布斯效应的震荡程度, 具体如下:

[0152] 步骤 A001: 将第二降采样率数字信号经短时傅里叶变换成频域信号, 并将该频域信号划分为若干频带, 计算各个频带的能量并对其进行平滑处理。

[0153] 短时傅里叶变换如下:

$$[0154] \quad X(f, m) = \frac{2}{M} \sum_{n=0}^{M-1} \text{win}(n-m) \times x(m) e^{-2\pi j f \frac{n}{M}}$$

[0155] 其中, M 为短时傅利叶变换的计算长度, f 表示频率值, X 表示频域信号, x 表示第二降采样率数字信号;

[0156] 汉明窗函数定义如下:

[0157] $\text{win}(n) = \{$

[0158] $0.54 - 0.46 \cos(2 * \pi * n / M)$ $0 \leq n \leq M-1$

[0159] 0 其余 n

[0160] $\}$

[0161] 则短时傅里叶变换如下

$$[0162] \quad Y(f, m) = \frac{2}{M} \sum_{n=0}^{M-1} \text{win}(n-m) \times y(m) e^{-2\pi j f \frac{n}{M}}$$

[0163] 其中, M=256, 为短时傅利叶变换的计算长度, m 表示第 m 帧信号。

[0164] 这样就将当前帧的带噪语音信号 y 从时域变换到为频域信号 Y。

[0165] 变换到频域后的带噪语音信号包括语音信号和噪声信号, 其以帧为单位划分为若干频带。

[0166] 接着对频率为 8kHz 以下的变换到频域后的带噪语音的信号进行频带划分, 之后的信号处理均在各个频带中进行, 这样既可以减少运算复杂度, 又可以针对不同的频带做不同的处理, 得到更好的效果。

[0167] 本发明中的信号共划分为 30 个频带。具体见表 1。

[0168] 表 1 23 个频带划分

[0169]

频带数	起始频率 (Hz)	截止频率 (Hz)
1	62.5	93.75
2	125	156.25
3	187.5	218.75
4	250	281.25
5	312.5	343.75
6	375	406.25
7	437.5	468.75
8	500	531.25
9	562.5	593.75
10	625	656.25
11	687.5	718.75
12	750	781.25
13	812.5	906.25
14	937.5	1062.5
15	1093.75	1250
16	1281.25	1468.75
17	1500	1718.75
18	1750	2000
19	2031.25	2312.5
20	2343.75	2687.5
21	2718.75	3125
22	3156.25	3687.5
23	3718.75	3968.75

[0170]

24	4000	4312.5
25	4343.75	4687.5
26	4718.75	5156.5
27	5187.5	5718.75
28	5750	6250
29	6281.25	6875
30	6936.25	7968.75

[0171] 采用如下方法计算每个频带的能量并对其进行平滑处理：

$$[0172] \quad E(m, k) = |X(m, k)|^2 \quad 0 \leq k \leq N-1$$

$$[0173] \quad Y_E(m, k) = \alpha Y_E(m-1, k) + (1-\alpha) E(m, k) \quad 0 \leq k \leq N-1$$

[0174] 其中, $Y_E(m, k)$ 表示经平滑处理后的每个频带区间的能量, m 表示当前帧的序号, k 表示当前的子带的序号, α 为平滑因子, 此处 $\alpha=0.75$, N 为选取的频带总数, $E(m, k)$ 表示频带能量值, $X(m, k)$ 表示第 m 帧第 k 个频带的频域信号。

[0175] 步骤 A002: 用计算器计算先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$

[0176] 计算当前帧信号的后验信噪比, 具体如下:

[0177] 设初始的噪声能量估计值 $V(0, k)=0$, 初始的先验信噪比估计值 $\hat{SNR}_{prior}(0, k)=0$, 由各频带的能量 $Y_E(m, k)$ 及当前帧的前一帧得到的噪声能量估计值 $V(m-1, k)$, 计算各频带当前帧的后验信噪比:

$$[0178] \quad SNR_{post}(m, k) = \frac{Y_E(m, k)}{V(m-1, k)}$$

[0179] 然后基于 Ephraim-Malah (伊弗雷-马拉) 的先验信噪比估计公式, 计算当前帧的先验信噪比估计值:

$$[0180] \quad \hat{SNR}_{prior}(m, k) = \beta \frac{|\hat{S}(m-1, k)|^2}{V(m, k)} + (1-\beta) \max[0, SNR_{post}(m, k) - 1]$$

[0181] $\hat{S}(m-1, k)$ 表示前一帧最后得到的增强后语音信号, β 为第一平滑系数, $V(m, k)$ 为噪声能量估计值。

[0182] 步骤 A003: 对先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ 进行修正, 得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ 。

[0183] 基于 Ephraim-Malah (伊弗雷 - 马拉) 的先验信噪比估计值, 会出现对噪声的过估计, 即对于高信噪比的情况下, 得到的信噪比估计值偏高, 造成增强后的语音信号失真, 本实施方式中采用计权噪声估计方法对先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$ 进行修正。

[0184] 先验信噪比估计值 $\hat{SNR}_{prior}(m, k)$, 乘以计权因子去 q_θ , 得到修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ 。

[0185] 计权因子计算如下:

$$[0186] \quad q_\theta = \begin{cases} 1 & \hat{SNR}_{prior}(m, k) < \gamma_1 \\ \frac{1}{\gamma_1 - \gamma_2} (\hat{SNR}_{prior}(m, k) - \gamma_2) & \gamma_1 < \hat{SNR}_{prior}(m, k) < \theta_z \\ 0 & \hat{SNR}_{prior}(m, k) > \theta_z \end{cases}$$

[0187] 其中 γ_1 取 1.5, γ_2 取 200, θ_z 取 20。

[0188] 步骤 A004: 根据修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$, 计算各个频带的衰减增益 $q(m, k)$ 。

[0189] 各个频带的衰减增益 $q(m, k)$ 基于先验信噪比估计值采用频谱相减的方法得到衰减因子, 其具体公式如下

$$[0190] \quad q(k) = \begin{cases} 1 - \frac{\alpha - 0.5 \cdot \lg[\hat{SNR}'_{prior}(k)]}{SNR_{post}} \end{cases}$$

[0191] 其中, 对于不同频带, α 为不同的常数。

[0192] 这里考虑到噪声主要集中在较低的频带, 因此对于中低频段和高频, α 取不同值。

[0193] 本发明中对于 $k \leq 14$ 的频带, 即频率为 1.2kHz 以下的信号, $\alpha = 8.89$,

[0194] 对于 $14 < k \leq 18$ 的频带, 即频率为 1.2~2kHz 之间的信号, $\alpha = 6.44$,

[0195] 对于 $18 < k \leq 23$ 的频带, 即频率为 2~4kHz 以上的信号, $\alpha = 6.21$

[0196] 对于 $k > 23$ 的频带, 即频率为 4kHz 以上的信号, $\alpha = 5.37$ 。

[0197] 采用门限判决对所述衰减增益进行调整:

[0198] 首先, 设定一阈值、调整值 q_{mod} 和门限值 q_{floor} ;

[0199] 其次, 以当前帧修正后的先验信噪比估计值 $\hat{SNR}'_{prior}(m, k)$ 为判断, 小于所述阈值的衰减增益均乘以调整值 q_{mod} , 从而进一步抑制噪声;

[0200] 接着将所有小于等于门限值的衰减增益调整到门限值 q_{floor} , 这样做可以避免一定的语音失真。具体方法如下:

[0201]

$$q'(m, k) = \begin{cases} q_{mod} q(m, k), & \hat{SNR}'_{prior}(m, k) < \theta_G \\ q(m, k), & \text{否则} \end{cases}$$

[0202]

$$q'(m, k) = \begin{cases} q(m, k), & q(k) > q_{\text{floor}} \\ q_{\text{floor}}, & \text{否则} \end{cases}$$

[0203] 其中 $q_{\text{mod}} = 0.1$, $\theta_g = 1.2$, $q_{\text{floor}} = 0.01$ 。

[0204] 步骤 A005: 将当前帧的各频带的频域信号 $X(m, k)$, 乘以相对应频带的衰减增益, 得到该频带的增强后的语音信号 $\hat{S}(k)$:

$$[0205] \quad \hat{S}(m, k) = q(m, k) * X(m, k) \quad 0 \leq k \leq N-1$$

[0206] 步骤 A006: 利用修正后的先验信噪比估计值 $\hat{SNR}'_{\text{prior}}(m, k)$ 对当前帧进行判断, 判断当前帧是否是噪声并根据判决结果对各频带的噪声能量估计值进行更新, 具体为:

[0207] 本发明中, 每个频带的噪声能量的判决与更新采用了基于先验信噪比的语音激活检测 (VAD) 方法。首先, 判断当前帧是否是纯噪声信号:

$$[0208] \quad VAD(m) = \sum_{k=1}^N \left[\frac{\gamma(m, k) \zeta(m, k)}{1 + \zeta(m, k)} - \lg(1 + \zeta(m, k)) \right]$$

[0209] 其中 $\gamma(m, k) = \min[\text{SNR}_{\text{post}}(m, k), 40]$,

$$[0210] \quad \zeta(m, k) = \max[\hat{SNR}_{\text{prior}}(m, k), 10^{-25}]$$

[0211] 对 $VAD(m)$ 进行判断, 并进行噪声更新, 具体如下:

$$[0212] \quad V(m, k) = \begin{cases} \mu V(m-1, k) + (1-\mu) E(m, k) & VAD(m) < \eta \\ V(m-1, k) & VAD(m) \geq \eta \end{cases}$$

[0213] 其中 η 为噪声更新判决因子, 本发明中取 $\eta = 0.01$ 。

[0214] μ 为平滑因子, 这里取 $\mu = 0.9$ 。

[0215] 如果判断为噪声, 则噪声能量估计值等于前一帧的值 $V(m, k) = V(m-1, k)$, 否则采用频带能量 $Y_E(m, k)$ 对噪声能量估计值 $V(m, k)$ 进行更新, 并将此噪声能量估计值 $V(m, k)$ 用于当前帧的下一帧的步骤 A002 中, 进行后验信噪比的估计:

$$[0216] \quad V(m, k) = \mu V(m-1, k) + (1-\mu) \cdot Y_E(m, k)$$

[0217] 其中, μ 表示第二平滑系数。

[0218] 步骤 A007: 将增强后的语音信号 $\hat{S}(k)$ 用短时傅里叶变换变换成时域信号, 将该时域信号经过低通滤波器 $H(z) = 1 + bz^{-1}$, 其中 b 为一般常数, 进行去加重处理, 得到去除背景噪声的语音信号并输出。

[0219] 首先, 通过逆快速傅里叶变换 (FFT) 把频域的语音谱变换到时间域, 得到增强后的时域语音信号。

[0220] 时域的变换用通用的逆离散傅利叶变换 (IDFT) 实现。

$$[0221] \quad s(m, n) = \frac{1}{2} * \sum_{n=0}^{M-1} \hat{S}(f) e^{j2\pi nf/M} \quad 0 \leq k \leq M-1$$

[0222] 其中, $M=256$, 为帧长, s 为变换到时域后的全频带增强后的语音信号。

[0223] 其次, 对增强后的时域语音信号进行去加重处理。

[0224] 和上述的预加重处理相反,这里将增强后的时域语音信号通过一个低通滤波器,最大程度的将其还原成原有的信号。滤波器的频响如下:

[0225] $H(z) = 1 + \alpha z^{-1}$, 其中 $\alpha = 0.9$ 。

[0226] 再次,将增强后的语音信号的相邻帧的重叠部分进行相加操作。

[0227] 具体的重叠部分相加可以用如下的方法来表示:

$$[0228] \quad s'(n) = \begin{cases} s(m, n) + s(m-1, n+L) & 0 \leq n < M-L \\ s(m, n) & M-L \leq n < M \end{cases}$$

[0229] $L=64$ 为相邻的帧信号开始处的距离, $M=256$, 为帧长, s' 为去除背景噪声的语音信号。

[0230] 步骤 S006: 利用输出模块接收语音增强处理模块输出去除背景噪声的语音信号。

[0231] 与相关技术比较,本发明的基于双麦克风语音增强装置将解码、降噪等芯片集成到一块信号处理芯片中,集成度高。本发明使用双麦克风阵列,两个数字麦克风不同方向进声,采用指向性形成抑制波束外的背景噪声,并通过从语音增强算法,进一步衰减各类型的背景噪声,采用数字麦克风及相关解码算法,省去了模数转换芯片。数字麦克风采集得到的信号直接进行降噪处理,节省了系统资源。语音增强算法大大提高了对残余噪声的衰减,保证了语音可懂度。传统的 ECM 麦克风,信号输入后需经过模数转换的过程,占用了系统的资源。而本发明采用的数字麦克风,采集到的信号直接是数字 pdm 信号,这样无需模数转换,只需要简单的解码即可直接进行降噪处理,大大提高了系统性能。

[0232] 以上所述的仅是本发明的实施方式,在此应当指出,对于本领域的普通技术人员来说,在不脱离本发明创造构思的前提下,还可以做出改进,但这些均属于本发明的保护范围。

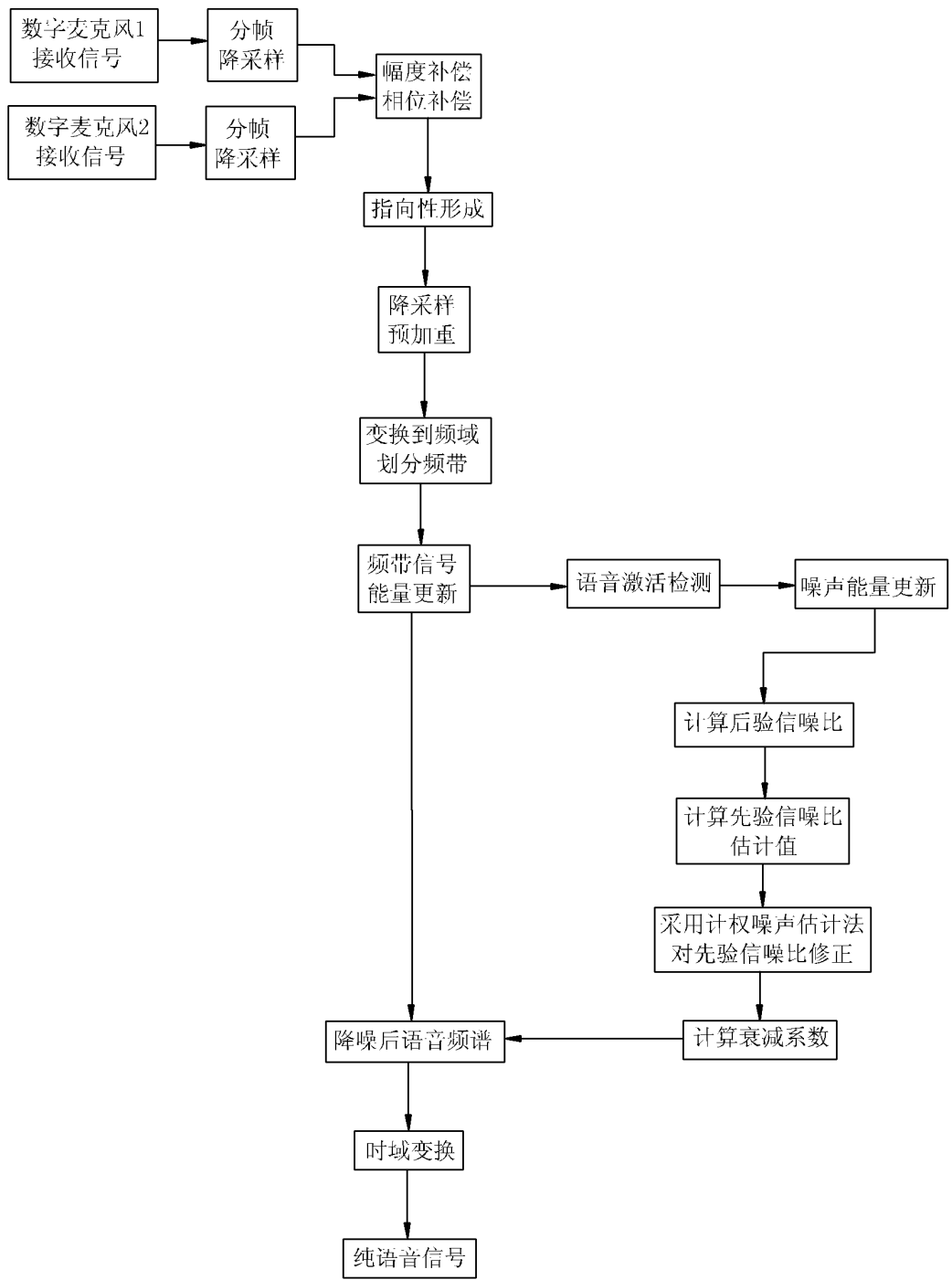


图 1

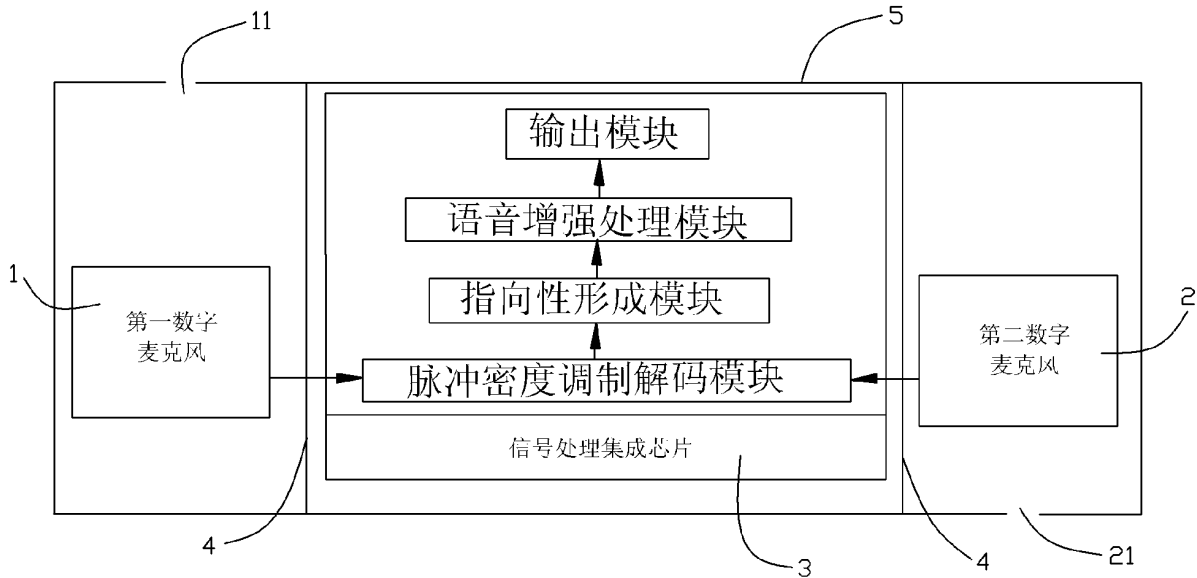


图 2

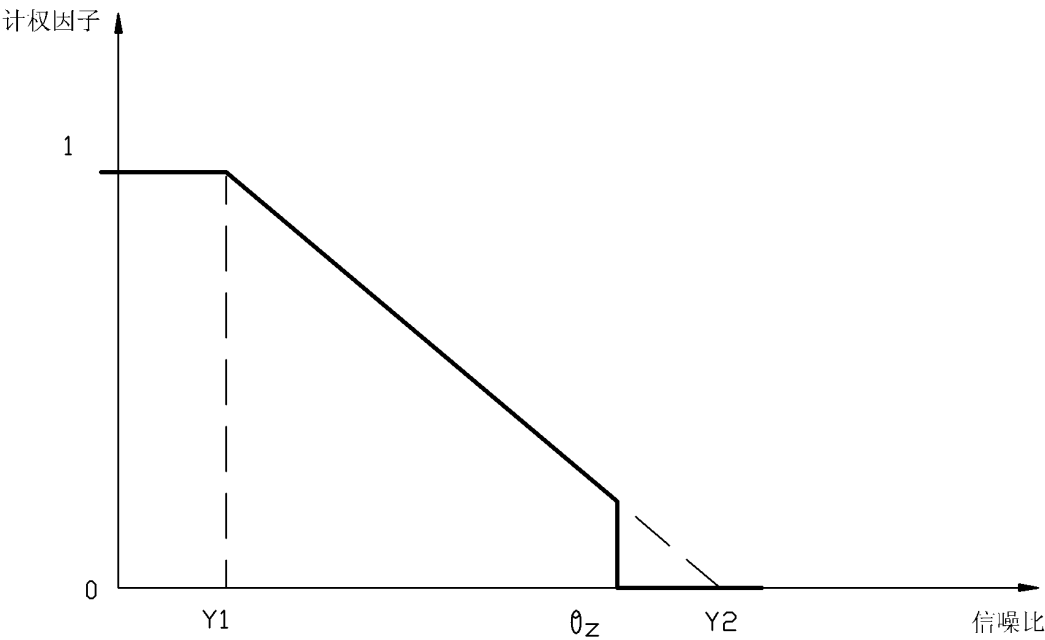


图 3