



(12) 发明专利

(10) 授权公告号 CN 102509552 B

(45) 授权公告日 2013. 09. 11

(21) 申请号 201110324746. 4

审查员 辛杰

(22) 申请日 2011. 10. 21

(73) 专利权人 浙江大学

地址 310027 浙江省杭州市西湖区浙大路
38 号

(72) 发明人 钱运涛 杨立春

(74) 专利代理机构 杭州天正专利事务所有限公
司 33201

代理人 黄美娟 王兵

(51) Int. Cl.

G10L 21/0208(2013. 01)

H04R 3/04(2006. 01)

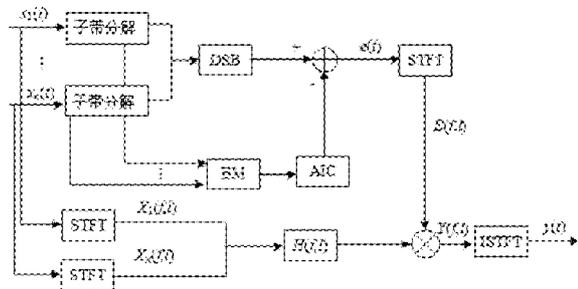
权利要求书3页 说明书8页 附图2页

(54) 发明名称

一种基于联合抑制的麦克风阵列语音增强方法

(57) 摘要

一种基于联合抑制的麦克风阵列语音增强方法,包括:构建麦克风阵列、以接收外界信号;分析信号,获取目标声源相对麦克风阵列中不同阵元信号相对基准阵元信号的时延;根据时延,分别对与两个麦克风所对应的数字信号进行时延补偿,获得补偿后的信号;对补偿后的阵列信号分别进行子带分解,然后再每个子带上进行固定波束形成;同时分别在每个子带上使用阻塞矩阵获取每个子带上的噪音参考信号;然后在每个子带上分别通过自适应滤波处理算法,将相应子带上固定波速形成器中的与目标语音无关的噪音去除并进行子带合并,从而形成初步增益信号;同时利用前面补偿后的任意两个阵元信号,通过递推互功率谱密度获得抑制与目标语音相关噪音信号的滤波器,进而结合初步增益信号得到最终的目标语音信号。



1. 一种基于联合抑制的麦克风阵列语音增强方法,包括以下步骤:

步骤 1,构建麦克风阵列、以接收外界模拟信号,由两个以上全指向性麦克风组成线性、圆形或球形阵列;将接收到的模拟信号转换为待处理的数字信号;

步骤 2,分析所述的数字信号,获取目标语音相对阵列中每个阵元的相对时延;根据时延,以阵列中某个阵元接收的语音信号作为参考,对其他阵元接收到的数字信号进行时延补偿,获得补偿后的信号;利用子带分解分别将每个阵元的时延补偿信号分解成若干个子带信号,然后分别对每个子带信号进行固定波束形成,以获得每个频带范围内目标声源方向的初步增益;包括以下子步骤:

2.1,分别获取目标声源到达基准麦克风和其他麦克风的相对时延,第 i 个麦克风接受的信号相对基准麦克风 1 的信号时延,其时延值通过公式 1 来估计:

$$R_{x_1 x_i} = \frac{1}{T - \tau} \int_{\tau}^T x b_1(t) x b_i(t - \tau) dt \dots\dots \text{公式 1}$$

其中: T 为麦克风阵列采集语音信号的实际时间, $x b_1(t)$, $x b_i(t)$ 分别对应的是 t 时刻第一麦克风,即基准麦克风,第 i 麦克风所对应的数字信号, $R_{x_1 x_i}$ 表示 $x b_1(t)$, $x b_i(t)$ 的互相关函数,其最大值对应的 τ 就是两个麦克风相对时延值;

2.2,分别对 $x b_2(t)$, \dots , $x b_n(t)$ 进行时延校正,获得 $x b_2(t)$ 的校正后信号 $x_2(t)$,依次类推, $x b_n(t)$ 的校正后信号 $x_n(t)$;为了统一书写格式,把基准麦克风信号 $x b_1(t)$ 以 $x_1(t)$ 代替;

2.3,利用子带分解方法分别把上述时延补偿后的信号分解成 k 个子带: $x_1(t)$ 分解成 $x_{-1 \text{sub-}1}(t)$, $x_{-1 \text{sub-}2}(t)$, $\dots\dots$, $x_{-1 \text{sub-}k}(t)$; $x_2(t)$ 到 $x_n(t)$ 依次类推;

2.4,对每个子带分别进行固定波束形成,第 i 个子带的固定波束形成获得的信号 $y b_{\text{sub-}i}(t)$ 由公式 2 计算:

$$y b_{\text{sub-}i}(t) = \frac{1}{n} \sum_{m=1}^n x_{-m \text{sub-}i}(t) \dots\dots \text{公式 2}$$

步骤 3,利用阻塞矩阵分别获得每个子带内的噪音参考信号;

步骤 4,利用变步长归一化 LMS 算法对步骤 2 的每个子带进行自适应滤波处理,处理结束后进行子带合并处理,获得输出信号 e ;

步骤 5,利用阵列中两个阵元补偿后的信号进行递推噪音互功率谱密度估计 (Iterative noise Cross Power Spectral Density, ICPSD) 算法,获得去除相关噪音的滤波器 H ,结合步骤 4 获得的增强信号 e ,进而获得最终的目标语音输出 y ;包括以下子步骤:

5.1、对步骤 2 中得到的信号 $x_1(t)$, $x_n(t)$ 分别进行短时傅里叶变换并进行分帧和加窗操作得到 $X_1(f, l)$, $X_n(f, l)$,其中 f 表示频率因子, l 代表第几帧;用 $P_{X_1 X_n}$ 和 $P_{N_1 N_n}$ 分别表示为第 1 和第 n 两个阵元接收信号的互功率谱密度 (Cross Power Spectral Density, CPSD) 以及两个阵元接收信号的估计噪音的互功率谱密度; $P_{X_1 X_1}$ 和 $P_{X_n X_n}$ 分别为这两个阵元信号的功率谱密度 (Power Spectral Density, PSD);则滤波器 $H(f, l)$ 可由公式 3 计算:

$$H(f, l) = \frac{|P_{X_1 X_n}(f, l)|}{\sqrt{P_{X_1 X_1} * P_{X_n X_n}}} * \frac{R(f, l)}{R(f, l) + 1} \dots\dots \text{公式 3}$$

5. 2、公式 3 中的 $R(f, l)$ 由公式 4 定义：

$$R(f, l) = \lambda_d H(f, l-1)^2 \cdot \frac{|X_1(f, l-1) \cdot X_n(f, l-1)^*|}{|P_{N_1 N_n}(f, l-1)|} + (1 - \lambda_d) R_p(f, l)$$

…公式 4

5. 3、公式 4 中的 $R_p(f, l)$ 由公式 5 定义：

$$R_p(f, l) = \max\left(\frac{|X_1(f, l) \cdot X_n(f, l)^*|}{|P_{N_1 N_n}(f, l)|} - 1, 0\right) \dots\dots \text{公式 5}$$

5. 4、估计噪声互功率谱密度由公式 6 计算：

$$P_{N_1 N_n}(f, l) = \lambda_n P_{N_1 N_n}(f, l-1) + (1 - \lambda_n) |X_1(f, l) \cdot X_n(f, l)^*| \cdot (1 - H(f, l-1)) \dots\dots \text{公式 6}$$

上面各式中的 λ_n , λ_d 为标量, 其值为接近 1 的小数, 这两个标量的目的是使噪声功率谱跟踪速度变慢, 防止在语音段噪声错误的跟踪到语音值, 同时在噪声段也能正确估计出噪声谱; $(\cdot)^*$ 为复共轭; $|\cdot|$ 表示取模运算; \cdot 表示点乘运算; $\max(\cdot)$ 表示取最大值运算;

5. 5、语音功率谱密度 $P_{X_1 X_1}$ 、 $P_{X_n X_n}$ 和互功率谱密度 $P_{X_1 X_n}$ 分别通过下式求得：

$$P_{X_1 X_1}(f, l) = \lambda_x P_{X_1 X_1}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_1(f, l)^*| \dots\dots \text{公式 7}$$

$$P_{X_n X_n}(f, l) = \lambda_x P_{X_n X_n}(f, l-1) + (1 - \lambda_x) |X_n(f, l) \cdot X_n(f, l)^*| \dots\dots \text{公式 8}$$

$$P_{X_1 X_n}(f, l) = \lambda_x P_{X_1 X_n}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_n(f, l)^*| \dots\dots \text{公式 9}$$

为了能让语音互功率谱能在语音段迅速跟踪到真实语音信息, 上式中标量 $\lambda_x \in [0.6, 0.7]$;

5. 6、将步骤 4 获得的增强信号 $e(t)$ 进行短时傅里叶变换得到 $E(f, l)$, 则获得最终的语音增强频域信号 $Y(f, l)$ 由公式 10 计算得到：

$$Y(f, l) = E(f, l) \cdot H(f, l) \dots\dots \text{公式 10}$$

5. 7、公式 10 计算得到的 $Y(f, l)$ 进行短时反傅里叶变换得到最终的语音增强信号 $y(t)$ 。

2. 根据权利要求 1 所述的一种基于联合抑制的麦克风阵列语音增强方法, 其特征在于: 步骤 1 中的阵列可以从间距为 1cm 左右的致密型阵列到阵元间距大于 4cm 的大尺度常规阵列, 每个阵元接收到的模拟信号使用音频采集卡对阵列接收到的模拟信号进行 A/D 转换, 形成待处理的数字信号。

3. 根据权利要求 1 所述的一种基于联合抑制的麦克风阵列语音增强方法, 其特征在于: 步骤 3 中, 通过公式 11 获取每个子带中 n 个元素的参考噪声向量 $N_{\text{sub-1}}, N_{\text{sub-2}}, \dots, N_{\text{sub-k}}$, 其中第 i 个子带的参考噪声向量为:

$$N_{\text{sub-i}} = B * [x_{\text{sub-i}}(t), \dots, x_{\text{sub-i}}(t)]^T \dots\dots \text{公式 11}$$

公式 11 中 B 为阻塞矩阵 (Blocking Matrix, BM), 它由公式 12 确定:

$$B = \begin{bmatrix} n-1, -1, \dots, -1 \\ -1, n-1, \dots, -1 \\ \dots\dots\dots \\ -1, -1, \dots, n-1 \end{bmatrix} \dots\dots \text{公式 12}$$

公式 12 中的 n 表示阵列中的阵元数, $n \geq 2$ 。

4. 根据权利要求 3 所述的一种基于联合抑制的麦克风阵列语音增强方法, 其特征在于: 步骤 4 中, 通过归一化最小均方 (Normalized Least Mean Squares, NLMS) 自适应滤波器算法实现与目标语音不相关的噪音抑制, 包含以下步骤:

4.1、假定 NLMS 滤波器长度为 M , 则子带 i 第 m 个参考噪音序列 $V_{\text{sub-}i}(m)$ 和权系数向量 $W_{\text{sub-}i}(m)$ 可分别表示为:

$$V_{\text{sub-}i}(m) = [N_{\text{sub-}i}(m), N_{\text{sub-}i}(m-1), \dots, N_{\text{sub-}i}(m-M+1)]^T \dots\dots \text{公式 13}$$

$$W_{\text{sub-}i}(m) = [w_0(m), w_1(m), \dots, w_n(m)]^T \dots\dots \text{公式 14}$$

公式 14 中的 $w_0(m), w_1(m), \dots, w_n(m)$ 的初始值均为 0;

4.2、估计的噪音信号 $d_{\text{sub-}i}$ 为:

$$d_{\text{sub-}i}(m) = V_{\text{sub-}i}^T(m) W_{\text{sub-}i}(m) = \sum_{j=0}^{M-1} w_j(m) N_{\text{sub-}i}(m-j) \dots\dots \text{公式 15}$$

4.3、NLMS 输出信号即误差信号为期望信号减去估计的噪音信号:

$$e_{\text{sub-}i}(m) = y_{\text{sub-}i}(m) - d_{\text{sub-}i}(m) \dots\dots \text{公式 16}$$

4.4、下一个权系数向量由下式计算:

$$W_{\text{sub-}i}(m+1) = W_{\text{sub-}i}(m) + 2\mu e_{\text{sub-}i}(m) / (V_{\text{sub-}i}^T(m) V_{\text{sub-}i}(m)) \dots\dots \text{公式 17}$$

其中 μ 为步长因子, 初始阶段让其从一个较大的值逐步变化到一个稳定的较小值, 用以加速滤波器初始阶段的收敛性; $V_{\text{sub-}i}^T(m) V_{\text{sub-}i}(m)$ 为归一化系数, 用以提高 LMS 算法的收敛性;

4.5、最后将获得的 k 个子带输出信号进行子带合并, 得到输出信号 e 。

一种基于联合抑制的麦克风阵列语音增强方法

技术领域

[0001] 本发明涉及语音增强的技术领域,具体就是,本发明涉及的是面向从二元麦克风小尺度阵列到多元麦克风大尺度阵列的目标语音增强算法。

背景技术

[0002] 当前从移动通讯、电话会议、笔记本语音交流、汽车通话到电视互动等应用无一例外的面临着各种类型噪音如混响、回声和其他环境噪音的干扰,这给通讯双方造成了不便,因而噪音抑制问题自然而然的被提出并被广泛研究。目前噪音抑制或者说是语音增强技术主要有两个方向,一个是面向单麦克风的单道语音增强技术,其优势是成本低,但处理后的目标语音损伤较大,同时对非平稳噪音抑制能力有限;另外一个方向是麦克风阵列语音增强技术,相对单道麦克风语音增强技术,其处理后目标语音损伤小,且能够抑制非平稳噪音,因而近年来被广泛关注和研究。

[0003] 麦克风阵列语音增强算法一般采用波束形成噪音抑制方法,其思想是期望在阵列中心到目标语音间形成一个波束,而波束外的信号被抑制掉,然后对波束内渗入的噪音采用单道噪音抑制算法进行处理,得到最终的目标语音输出。麦克风阵列波束形成算法根据估计和跟踪噪音过程的不同可分为固定波束形成算法和自适应波束形成算法,固定波束形成算法包括延迟求和波束形成技术和超指向性波束形成技术等;自适应波束形成算法包括最小均方自适应滤波器、递归最小二乘自适应滤波器以及基于它们的改进算法。固定波束形成算法的缺陷是抑制效果有限或条件要求较高,如延迟求和波束形成器需要大量的阵元条件,或敏感,如超指向性波束形成技术对不相关噪音以及麦克风阵列特性(增益、位置、相位)误差敏感;而自适应波束形成器面临最大的问题是收敛性问题,收敛性与噪音复杂度以及噪音与目标语音的相关性有关,另外还需要语音活动检测技术支持。一般实际应用和研究时,往往采用固定波束形成器结合自适应波束形成器的方法,如广义旁瓣抵消器(Generalized Sidelobe Canceller, GSC)结构和差分麦克风阵列结构。但差分麦克风阵列面向的是微阵列语音增强技术,且要求目标语音为端射方向,否则目标语音会有变形。

[0004] 而对于已有公开的麦克风阵列噪音抑制技术,普遍存在算法复杂度高,计算量大或仅对某一阵列环境的语音增强应用等特点,造成普适性不够。如2009年1月21日公开的专利CN101351058,是面向的圆形麦克风阵列,仅能对特定应用有效,另外它期望通过获得某个弧度范围的入射方向推算其他弧度访问的入射方向,由于入射方向估计难度较大,误差不可避免,这样推算其他弧度内的入射角误差可能会更大,这些都会对最终语音增强造成不利影响;而2009年2月18日公开的专利CN101369427,利用GSC结构,采用功率谱密度估计代替时域或频域信号,进行自适应估计,获得估计噪声功率谱,进而通过相减获得波束形成信号功率谱密度,然后进行后置滤波,该算法波束形成算法中由于需要计算功率谱密度,当阵元数较多时,计算量相对其他方法,计算量较大,另外波束形成中通过减去噪音功率谱密度,与单道谱减法缺点一样,容易引起乐噪。

发明内容

[0005] 当前麦克风阵列波束形成算法中采用的自适应滤波器对与目标声源不相关的噪声有很好的抑制效果,但对与目标语音相关性的噪声则效果较差,对这些残留的相关噪声抑制一般采用单道后置滤波器进行处理,例如谱减法、维纳滤波器等,这些方法一般都需要利用语音活动检测的方法,用非语音段的噪声代替或估计语音段方法,因而会造成目标声源变形或形成乐噪。为了克服这些缺点,本发明提出一种与已有的阵列语音增强技术相比,噪声抑制效果更好、处理后的语音质量更高、且算法稳定并适合各种尺度和类型的麦克风阵列的语音增强方法。

[0006] 一种基于联合抑制的麦克风阵列语音增强方法,包括以下步骤:

[0007] 步骤 1,构建麦克风阵列、以接收外界模拟信号,由二到多个以上全指向性麦克风组成线性、圆形或球形阵列;将接收到的模拟信号转换为待处理的数字信号;

[0008] 步骤 2,分析所述的数字信号,获取目标语音相对阵列中每个阵元的相对时延;根据时延,以阵列中某个阵元接收的语音信号作为参考,对其他阵元接收到的数字信号进行时延补偿,获得补偿后的信号;利用子带分解分别将每个阵元的时延补偿信号分解成若干个子带信号,然后分别对每个子带信号进行固定波束形成,以获得每个频带范围内目标声源方向的初步增益;

[0009] 步骤 3,利用阻塞矩阵分别获得每个子带内的噪声参考信号;

[0010] 步骤 4,利用变步长归一化 LMS 算法对步骤 4 的每个子带进行自适应滤波处理,处理结束后进行子带合并处理,获得输出信号 e。

[0011] 步骤 5,利用阵列中两个阵元补偿后的信号进行改进的递推噪声互功率谱密度估计 (Iterative noise Cross Power Spectral Density, ICPSD) 算法,获得去除相关噪声的滤波器 H,结合步骤 4 获得的增强信号 e,进而获得最终的目标语音输出 y。

[0012] 进一步,步骤 1 中的阵列可以从间距为 1cm 左右的致密型阵列到阵元间距大于 4cm 的大尺度常规阵列,每个阵元接收到的模拟信号使用音频采集卡对阵列接收到的模拟信号进行 A/D 转换,形成待处理的数字信号。

[0013] 步骤 2 包括以下子步骤:

[0014] 2.1,分别获取目标声源到达基准麦克风和其他麦克风的相对时延,如第 i 个麦克风接受的信号相对基准麦克风 1 的信号时延,其时延值通过公式 1 来估计:

$$[0015] \quad R_{x_1 x_i} = \frac{1}{T - \tau} \int_{\tau}^T x_{b_1}(t) x_{b_i}(t - \tau) dt \dots \dots \text{公式 1}$$

[0016] 其中: T 为麦克风阵列采集语音信号的实际时间, $x_{b_1}(t)$, $x_{b_i}(t)$ 分别对应的是 t 时刻第一麦克风 (基准麦克风), 第 i 麦克风所对应的数字信号, $R_{x_1 x_i}$ 表示 $x_{b_1}(t)$, $x_{b_i}(t)$ 的互相关函数,其最大值对应的 τ 就是两个麦克风相对时延值;

[0017] 2.2,分别对 $x_{b_2}(t)$, \dots , $x_{b_n}(t)$ 进行时延校正,获得 $x_{b_2}(t)$ 的校正后信号 $x_2(t)$,依次类推, $x_{b_n}(t)$ 的校正后信号 $x_n(t)$;为了统一书写格式,把基准麦克风信号 $x_{b_1}(t)$ 以 $x_1(t)$ 代替;

[0018] 2.3,利用子带分解方法分别把上述时延补偿后的信号分解成 k 个子带:例如 $x_1(t)$ 分解成 $x_{-1_{\text{sub}-1}}(t)$, $x_{-1_{\text{sub}-2}}(t)$, \dots , $x_{-1_{\text{sub}-k}}(t)$; $x_2(t)$ 到 $x_n(t)$ 依次类推。

[0019] 2.4,对每个子带分别进行固定波束形成,如第 i 个子带的固定波束形成获得的信

号 $y_{b_{sub-i}}(t)$ 由公式 2 计算：

$$[0020] \quad y_{b_{sub-i}}(t) = \frac{1}{n} \sum_{m=1}^n x_{-m_{sub-i}}(t) \dots \dots \text{公式 2}$$

[0021] 进一步,步骤 3 中,通过公式 3 获取每个子带中 n 个元素的参考噪声向量 N_{sub-1} , $N_{sub-2} \dots \dots, N_{sub-k}$, 其中第 i 个子带的参考噪声向量为：

$$[0022] \quad N_{sub-i} = B * [x_{-1_{sub-i}}(t), \dots, x_{-n_{sub-i}}(t)]^T \dots \dots \text{公式 3}$$

[0023] 公式 3 中 B 为阻塞矩阵 (Blocking Matrix, BM), 它由公式 4 确定：

$$[0024] \quad B = \begin{bmatrix} n-1, -1, \dots, -1 \\ -1, n-1, \dots, -1 \\ \dots \dots \dots \\ -1, -1, \dots, n-1 \end{bmatrix} \dots \dots \text{公式 4}$$

[0025] 公式 4 中的 n 表示阵列中的阵元数, $n \geq 2$ ；

[0026] 进一步,步骤 4 中,通过归一化最小均方 (Normalized Least Mean Squares, NLMS) 自适应滤波器算法实现与目标语音不相关的噪声抑制,包含以下步骤：

[0027] 4.1、假定 NLMS 滤波器长度为 M , 则子带 i 第 m 个参考噪声序列 $V_{sub-i}(m)$ 和权系数向量 $W_{sub-i}(m)$ 可分别表示为：

$$[0028] \quad V_{sub-i}(m) = [N_{sub-i}(m), N_{sub-i}(m-1), \dots, N_{sub-i}(m-M+1)]^T \dots \dots \text{公式 5}$$

$$[0029] \quad W_{sub-i}(m) = [w_0(m), w_1(m), \dots, w_n(m)]^T \dots \dots \text{公式 6}$$

[0030] 公式 6 中的 $w_0(m), w_1(m), \dots, w_n(m)$ 的初始值均为 0。

[0031] 4.2、估计的噪声信号 d_{sub-i} 为：

$$[0032] \quad d_{sub-i}(m) = V_{sub-i}^T(m) W_{sub-i}(m) = \sum_{j=0}^{M-1} w_j(m) v_{sub-i}(m-1) \dots \dots \text{公式 7}$$

[0033] 4.3、NLMS 输出信号即误差信号为期望信号减去估计的噪声信号：

$$[0034] \quad e_{sub-i}(m) = y_{b_{sub-i}}(m) - d_{sub-i}(m) \dots \dots \text{公式 8}$$

[0035] 4.4、下一个权系数向量由下式计算：

$$[0036] \quad W_{sub-i}(m+1) = W_{sub-i}(m) + 2\mu e_{sub-i}(m) / (V_{sub-i}^T(m) V_{sub-i}(m)) \dots \dots \text{公式 9}$$

[0037] 其中 μ 为步长因子, 初始阶段让其从一个较大的值逐步变化到一个稳定的较小值, 用以加速滤波器初始阶段的收敛性; $V^T V$ 为归一化系数, 用以提高 LMS 算法的收敛性。

[0038] 4.5、最后将获得的 k 个子带输出信号进行子带合并, 得到输出信号 e 。

[0039] 进一步,步骤 5 中包括以下子步骤：

[0040] 5.1、对步骤 2 中得到的信号 $x_1(t), x_n(t)$ 分别进行短时傅里叶变换并进行分帧和加窗操作得到 $X_1(f, l), X_n(f, l)$, 其中 f 表示频率因子, l 代表第几帧。用 $P_{X_1 X_n}$ 和 $P_{N_1 N_n}$ 分别表示为第 1 和第 n 两个阵元接收信号的互功率谱密度 (Cross Power Spectral Density, CPSD) 以及两个阵元接收信号的估计噪声的 CPSD; $P_{X_1 X_1}$ 和 $P_{X_n X_n}$ 分别为这两个阵元信号的功率谱密度 (Power Spectral Density, PSD)。则滤波器 $H(f, l)$ 可由公式 10 计算：

$$[0041] \quad H(f, l) = \frac{|P_{X_1 X_n}(f, l)|}{\sqrt{P_{X_1 X_1} * P_{X_n X_n}}} * \frac{R(f, l)}{R(f, l) + 1} \dots \dots \text{公式 10}$$

[0042] 5.2、公式 10 中的 $R(f, l)$ 由公式 11 定义：

$$[0043] \quad R(f, l) = \lambda_d H(f, l-1)^2 \cdot \frac{|X_1(f, l-1) \cdot X_n(f, l-1)^*|}{|P_{N_1 N_n}(f, l-1)|} + (1 - \lambda_d) R_p(f, l) \dots \text{公式 11}$$

[0044] 5.3、公式 11 中的 $R_p(f, l)$ 由公式 12 定义：

$$[0045] \quad R_p(f, l) = \max\left(\frac{|X_1(f, l) \cdot X_n(f, l)^*|}{|P_{N_1 N_n}(f, l)|} - 1, 0\right) \dots \text{公式 12}$$

[0046] 5.4、估计噪声互功率谱密度由公式 13 计算：

$$[0047] \quad P_{N_1 N_n}(f, l) = \lambda_n P_{N_1 N_n}(f, l-1) + (1 - \lambda_n) |X_1(f, l) \cdot X_n(f, l)^*| \cdot (1 - H(f, l-1))$$

[0048] ……公式 13

[0049] 上面各式中的 λ_n , λ_d 为标量, 其值为接近 1 的小数, 这两个标量的目的是使噪声功率谱跟踪速度变慢, 防止在语音段噪声错误的跟踪到语音值, 同时在噪音段也能正确正确估计出噪声谱; $(\cdot)^*$ 为复共轭; $|\cdot|$ 表示取模运算; \cdot 表示点乘运算; $\max(\cdot)$ 表示取最大值运算。

[0050] 5.5、语音功率谱密度 $P_{X_1 X_1}$ 、 $P_{X_n X_n}$ 和互功率谱密度 $P_{X_1 X_n}$ 分别通过下式求得：

$$[0051] \quad P_{X_1 X_1}(f, l) = \lambda_x P_{X_1 X_1}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_1(f, l)^*| \dots \text{公式 14}$$

$$[0052] \quad P_{X_n X_n}(f, l) = \lambda_x P_{X_n X_n}(f, l-1) + (1 - \lambda_x) |X_n(f, l) \cdot X_n(f, l)^*| \dots \text{公式 15}$$

$$[0053] \quad P_{X_1 X_n}(f, l) = \lambda_x P_{X_1 X_n}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_n(f, l)^*| \dots \text{公式 16}$$

[0054] 为了能让语音互功率谱能在语音段迅速跟踪到真实语音信息, 上式中标量 $\lambda_x \in [0.6, 0.7]$ 。

[0055] 5.6、将步骤 4 获得的增强信号 $e(t)$ 进行短时傅里叶变换得到 $E(f, l)$, 则获得最终的语音增强频域信号 $Y(f, l)$ 由公式 17 计算得到：

$$[0056] \quad Y(f, l) = E(f, l) \cdot H(f, l) \dots \text{公式 17}$$

[0057] 5.7、公式 17 计算得到的 $Y(f, l)$ 进行短时反傅里叶变换得到最终的语音增强信号 $y(t)$ 。

[0058] 本发明中, 对于与目标语音不相关的噪声采用广义旁瓣抵消器结构进行处理。本发明中, 由于采用子带, 相当于降采样, 计算量相应降低了, 同时加速了自适应滤波器的收敛性。

[0059] 本发明中对于与目标语音相关的噪声, 利用基于阵列的互功率谱递推计算方法, 仅对两个阵元进行, 计算复杂度能够被控制; 另外相对传统的其他功率谱密度估计, 本算法采用的是递推算法, 无需语音活动检测方法支持, 由此得到可以抑制相关性噪声的滤波器, 进而得到最终的目标语音输出。

[0060] 本发明的优点是: 采用了阵列技术分别对与目标语音相关的噪声和不相关的噪声分别进行抑制, 相对已有的阵列处理技术, 噪声抑制效果好, 处理后的语音质量高, 算法稳定并且适合各种尺度和类型的阵列环境等, 因而应用前景广阔。

附图说明

[0061] 图 1 整个算法处理结构图。

[0062] 图 2 基于子带分解的广义旁瓣结构的语音抑制处理流程图。

[0063] 图 3 基于递推互功率谱密度滤波器处理流程图。

具体实施方式

[0064] 参照附图,进一步说明本发明:

[0065] 一种基于联合抑制的麦克风阵列语音增强方法,包括以下步骤:

[0066] 步骤 1,构建麦克风阵列、以接收外界模拟信号,由二到多个以上全指向性麦克风组成线性、圆形或球形阵列;将接收到的模拟信号转换为待处理的数字信号;

[0067] 步骤 2,分析所述的数字信号,获取目标语音相对阵列中每个阵元的相对时延;根据时延,以阵列中某个阵元接收的语音信号作为参考,对其他阵元接收到的数字信号进行时延补偿,获得补偿后的信号;利用子带分解分别将每个阵元的时延补偿信号分解成若干个子带信号,然后分别对每个子带信号进行固定波束形成,以获得每个频带范围内目标声源方向的初步增益;

[0068] 步骤 3,利用阻塞矩阵分别获得每个子带内的噪音参考信号;

[0069] 步骤 4,利用变步长归一化 LMS 算法对步骤 4 的每个子带进行自适应滤波处理,处理结束后进行子带合并处理,获得输出信号 e。

[0070] 步骤 5,利用阵列中两个阵元补偿后的信号进行改进的递推噪音互功率谱密度估计 (Iterative noise Cross Power Spectral Density, ICPSD) 算法,获得去除相关噪音的滤波器 H,结合步骤 4 获得的增强信号 e,进而获得最终的目标语音输出 y。

[0071] 步骤 1 中的阵列可以从间距为 1cm 左右的致密型阵列到阵元间距大于 4cm 的大尺度常规阵列,每个阵元接收到的模拟信号使用音频采集卡对阵列接收到的模拟信号进行 A/D 转换,形成待处理的数字信号。

[0072] 步骤 2 包括以下子步骤:

[0073] 2.1,分别获取目标声源到达基准麦克风和其他麦克风的相对时延,如第 i 个麦克风接受的信号相对基准麦克风 1 的信号时延,其时延值通过公式 1 来估计:

$$[0074] \quad R_{x_1 x_i} = \frac{1}{T - \tau} \int_{\tau}^T x_{b_1}(t) x_{b_i}(t - \tau) dt \dots \dots \text{公式 1}$$

[0075] 其中:T 为麦克风阵列采集语音信号的实际时间, $x_{b_1}(t)$, $x_{b_i}(t)$ 分别对应的是 t 时刻第一麦克风(基准麦克风),第 i 麦克风所对应的数字信号, $R_{x_1 x_i}$ 表示 $x_{b_1}(t)$, $x_{b_i}(t)$ 的互相关函数,其最大值对应的 τ 就是两个麦克风相对时延值;

[0076] 2.2,分别对 $x_{b_2}(t)$, \dots , $x_{b_n}(t)$ 进行时延校正,获得 $x_{b_2}(t)$ 的校正后信号 $x_2(t)$,依次类推, $x_{b_n}(t)$ 的校正后信号 $x_n(t)$;为了统一书写格式,把基准麦克风信号 $x_{b_1}(t)$ 以 $x_1(t)$ 代替;

[0077] 2.3,利用子带分解方法分别把上述时延补偿后的信号分解成 k 个子带:例如 $x_1(t)$ 分解成 $x_{-1_{\text{sub-1}}}(t)$, $x_{-1_{\text{sub-2}}}(t)$, \dots , $x_{-1_{\text{sub-k}}}(t)$; $x_2(t)$ 到 $x_n(t)$ 依次类推。

[0078] 2.4,对每个子带分别进行固定波束形成,如第 i 个子带的固定波束形成获得的信号 $y_{b_{\text{sub-i}}}(t)$ 由公式 2 计算:

$$[0079] \quad y_{b_{\text{sub-i}}}(t) = \frac{1}{n} \sum_{m=1}^n x_{-m_{\text{sub-i}}}(t) \dots \dots \text{公式 2}$$

[0080] 步骤 3 中,通过公式 3 获取每个子带中 n 个元素的参考噪音向量 $N_{\text{sub-1}}$, $N_{\text{sub-2}}$ \dots , $N_{\text{sub-k}}$,其中第 i 个子带的参考噪音向量为:

[0081] $N_{sub-i} = B * [x_{-1_{sub-i}}(t), \dots, x_{-n_{sub-i}}(t)]^T \dots \dots$ 公式 3

[0082] 公式 3 中 B 为阻塞矩阵 (Blocking Matrix, BM), 它由公式 4 确定 :

[0083]
$$B = \begin{bmatrix} n-1, -1, \dots, -1 \\ -1, n-1, \dots, -1 \\ \dots \dots \dots \\ -1, -1, \dots, n-1 \end{bmatrix} \dots \dots$$
 公式 4

[0084] 公式 4 中的 n 表示阵列中的阵元数, $n \geq 2$;

[0085] 步骤 4 中, 通过归一化最小均方 (Normalized Least Mean Squares, NLMS) 自适应滤波器算法实现与目标语音不相关的噪音抑制, 包含以下步骤 :

[0086] 4. 1、假定 NLMS 滤波器长度为 M, 则子带 i 第 m 个参考噪音序列 $V_{sub-i}(m)$ 和权系数向量 $W_{sub-i}(m)$ 可分别表示为 :

[0087] $V_{sub-i}(m) = [N_{sub-i}(m), N_{sub-i}(m-1), \dots, N_{sub-i}(m-M+1)]^T \dots \dots$ 公式 5

[0088] $w_{sub-i}(m) = [w_0(m), w_1(m), \dots, w_n(m)]^T \dots \dots$ 公式 6

[0089] 公式 6 中的 $w_0(m), w_1(m), \dots, w_n(m)$ 的初始值均为 0。

[0090] 4. 2、估计的噪音信号 d_{sub-i} 为 :

[0091]
$$d_{sub-i}(m) = V_{sub-i}^T(m)W_{sub-i}(m) = \sum_{j=0}^{M-1} w_j(m)v_{sub-i}(m-1) \dots \dots$$
 公式 7

[0092] 4. 3、NLMS 输出信号即误差信号为期望信号减去估计的噪音信号 :

[0093] $e_{sub-i}(m) = y_{sub-i}(m) - d_{sub-i}(m) \dots \dots$ 公式 8

[0094] 4. 4、下一个权系数向量由下式计算 :

[0095] $W_{sub-i}(m+1) = W_{sub-i}(m) + 2\mu e_{sub-i}(m) / (V_{sub-i}^T(m)V_{sub-i}(m)) \dots \dots$ 公式 9

[0096] 其中 μ 为步长因子, 初始阶段让其从一个较大的值逐步变化到一个稳定的较小值, 用以加速滤波器初始阶段的收敛性 ; $V^T V$ 为归一化系数, 用以提高 LMS 算法的收敛性。

[0097] 4. 5、最后将获得的 k 个子带输出信号进行子带合并, 得到输出信号 e。

[0098] 步骤 5 中包括以下子步骤 :

[0099] 5. 1、对步骤 2 中得到的信号 $x_1(t), x_n(t)$ 分别进行短时傅里叶变换并进行分帧和加窗操作得到 $X_1(f, l), X_n(f, l)$, 其中 f 表示频率因子, l 代表第几帧。用 $P_{X_1X_n}$ 和 $P_{N_1N_n}$ 分别表示为第 1 和第 n 两个阵元接收信号的互功率谱密度 (Cross Power Spectral Density, CPSD) 以及两个阵元接收信号的估计噪音的 CPSD ; $P_{X_1X_1}$ 和 $P_{X_nX_n}$ 分别为这两个阵元信号的功率谱密度 (Power Spectral Density, PSD)。则滤波器 $H(f, l)$ 可由公式 10 计算 :

[0100]
$$H(f, l) = \frac{|P_{X_1X_n}(f, l)|}{\sqrt{P_{X_1X_1} * P_{X_nX_n}}} * \frac{R(f, l)}{R(f, l) + 1} \dots \dots$$
 公式 10

[0101] 5. 2、公式 10 中的 $R(f, l)$ 由公式 11 定义 :

[0102]
$$R(f, l) = \lambda_d H(f, l-1)^2 * \frac{|X_1(f, l-1) * X_n(f, l-1)^*|}{|P_{N_1N_n}(f, l-1)|} + (1 - \lambda_d) R_p(f, l) \dots \dots$$
 公式 11

[0103] 5. 3、公式 11 中的 $R_p(f, l)$ 由公式 12 定义 :

[0104]
$$R_p(f, l) = \max\left(\frac{|X_1(f, l) * X_n(f, l)^*|}{|P_{N_1N_n}(f, l)|} - 1, 0\right) \dots \dots$$
 公式 12

[0105] 5.4、估计噪声互功率谱密度由公式 13 计算：

$$[0106] \quad P_{N_1N_n}(f, l) = \lambda_n P_{N_1N_n}(f, l-1) + (1 - \lambda_n) |X_1(f, l) \cdot X_n(f, l)^*| \cdot (1 - H(f, l-1))$$

[0107] ……公式 13

[0108] 上面各式中的 λ_n, λ_d 为标量,其值为接近 1 的小数,这两个标量的目的是使噪声功率谱跟踪速度变慢,防止在语音段噪声错误的跟踪到语音值,同时在噪声段也能正确估计出噪声谱; $(\cdot)^*$ 为复共轭; $|\cdot|$ 表示取模运算; \cdot 表示点乘运算; $\max(\cdot)$ 表示取最大值运算。

[0109] 5.5、语音功率谱密度 $P_{X_1X_1}$ 、 $P_{X_nX_n}$ 和互功率谱密度 $P_{X_1X_n}$ 分别通过下式求得：

$$[0110] \quad P_{X_1X_1}(f, l) = \lambda_x P_{X_1X_1}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_1(f, l)^*| \dots \text{公式 14}$$

$$[0111] \quad P_{X_nX_n}(f, l) = \lambda_x P_{X_nX_n}(f, l-1) + (1 - \lambda_x) |X_n(f, l) \cdot X_n(f, l)^*| \dots \text{公式 15}$$

$$[0112] \quad P_{X_1X_n}(f, l) = \lambda_x P_{X_1X_n}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_n(f, l)^*| \dots \text{公式 16}$$

[0113] 为了能让语音互功率谱能在语音段迅速跟踪到真实语音信息,上式中标量 $\lambda_x \in [0.6, 0.7]$ 。

[0114] 5.6、将步骤 4 获得的增强信号 $e(t)$ 进行短时傅里叶变换得到 $E(f, l)$, 则获得最终的语音增强频域信号 $Y(f, l)$ 由公式 17 计算得到：

$$[0115] \quad Y(f, l) = E(f, l) \cdot H(f, l) \dots \text{公式 17}$$

[0116] 5.7、公式 17 计算得到的 $Y(f, l)$ 进行短时反傅里叶变换得到最终的语音增强信号 $y(t)$ 。

[0117] 本发明有两条支路同时进行,上面一条支路是基于 GSC 结构的波束形成算法,主要是在时域内抑制与目标语音不相关噪声干扰;下面一条支路是基于递推的互功率谱密度滤波器算法,主要目的是在频域内抑制与目标语音相关的噪声干扰。

[0118] 在整个算法的预处理阶段,首先对每个输入信号进行时延补偿,得到补偿后的时域信号；

[0119] 在 GSC 结构中,对预处理的时域信号先进行子带分解;然后再在每个子带上进行基于 GSC 结构算法的噪声抑制。

[0120] 进一步每个子带上的 GSC 结构又分成两个支路:固定波束形成器支路和自适应干扰抵消 (Adaptive Interference Cancellation, AIC) 加阻塞矩阵 (Blocking Matrix, BM) 支路。

[0121] 固定波束形成支路采用延迟求和波束形成 (Delay and Sum Beamformer, DSB) 算法：

$$[0122] \quad yb_{sub-i}(t) = \frac{1}{n} \sum_{m=1}^n x_{-m_{sub-i}}(t)$$

[0123] 阻塞矩阵目的是产生参考噪声信号,以便自适应滤波器可以抑制 DSB 输出中的噪声信号,阻塞矩阵由下式得出：

$$[0124] \quad B = \begin{bmatrix} n-1, -1, \dots, -1 \\ -1, n-1, \dots, -1 \\ \dots \dots \dots \\ -1, -1, \dots, n-1 \end{bmatrix}$$

[0125] 自适应干扰抵消采用变步长归一化最小均方 (Variable Step-size Normalized Least Mean Square, VSNLMS) 自适应滤波器实现, 变步长指的是在开始阶段使用较大的步长用以加速滤波器收敛, 然后逐步采用较小的步长以降低算法的稳态误差。归一化权系数更新过程由下式计算:

$$[0126] \quad W_{sub-i}(m+1) = W_{sub-i}(m) + 2\mu e_{sub-i}(m) / (V_{sub-i}^T(m) V_{sub-i}(m))$$

[0127] 滤波器期望输出信号为 DSB 输出信号; 参考噪声信号由阻塞矩阵获得。

[0128] 则最终的滤波器输出信号为:

$$[0129] \quad e_{sub-i}(m) = y_{sub-i}(m) - d_{sub-i}(m)$$

[0130] 最后对各个子带滤波器输出信号进行子带合并, 获得全带输出信号 $e(t)$ 。

[0131] 算法下面一条支路为递推互功率谱密度滤波器算法, 如图 3 所示, 先选取经预处理后的任意两个麦克风输入信号, 这里选取第一和第 n 个信号说明, 在经过短时傅里叶变化并经过分帧加窗后, 得到 $X_1(f, l)$, $X_n(f, l)$, 其中 f 表示频率因子, l 代表第几帧。然后分别计算它们的功率谱密度 PSD 和互功率谱密度 CPSD。

$$[0132] \quad P_{X_1X_1}(f, l) = \lambda_x P_{X_1X_1}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_1(f, l)^*|$$

$$[0133] \quad P_{X_nX_n}(f, l) = \lambda_x P_{X_nX_n}(f, l-1) + (1 - \lambda_x) |X_n(f, l) \cdot X_n(f, l)^*|$$

$$[0134] \quad P_{X_1X_n}(f, l) = \lambda_x P_{X_1X_n}(f, l-1) + (1 - \lambda_x) |X_1(f, l) \cdot X_n(f, l)^*|$$

[0135] 其中 $P_{X_1X_1}$ 、 $P_{X_nX_n}$ 分别对应 $X_1(f, l)$, $X_n(f, l)$ 的 PSD, 它们的互功率谱密度为 $P_{X_1X_n}$ 。

[0136] 噪声 CPSD 初始值为 $P_{X_1X_1}$, 并由下式递推计算:

$$[0137] \quad P_{N_1N_n}(f, l) = \lambda_n P_{N_1N_n}(f, l-1) + (1 - \lambda_n) |X_1(f, l) \cdot X_n(f, l)^*| \cdot (1 - H(f, l-1))$$

[0138] 进而得到抑制相关噪声的滤波器为:

$$[0139] \quad |H(f, l)| = \frac{|P_{X_1X_n}(f, l)|}{\sqrt{P_{X_1X_1} P_{X_nX_n}}} \cdot \frac{R(f, l)}{R(f, l) + 1}$$

[0140] 最后, 对 GSC 结构的输出信号进行短时傅里叶变换, 并分帧加窗后表示为 $E(f, l)$ 。这样得到最终的频域输出信号为:

$$[0141] \quad Y(f, l) = E(f, l) \cdot H(f, l)$$

[0142] 把 $Y(f, l)$ 还原到时域得到最终的语音增强信号 $y(t)$ 。

[0143] 本说明书实施例所述的内容仅仅是对发明构思的实现形式的列举, 本发明的保护范围不应当被视为仅限于实施例所陈述的具体形式, 本发明的保护范围也及于本领域技术人员根据本发明构思所能够想到的等同技术手段。

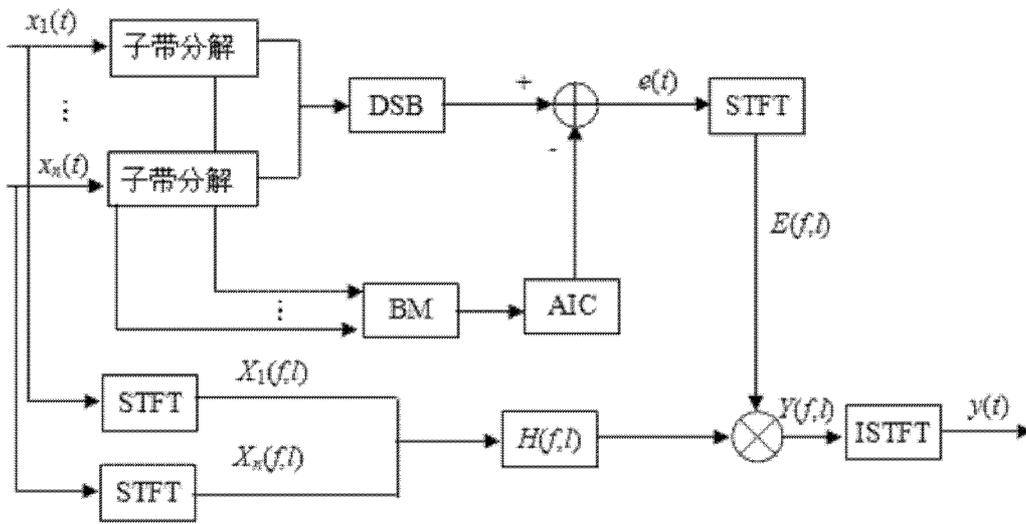


图 1

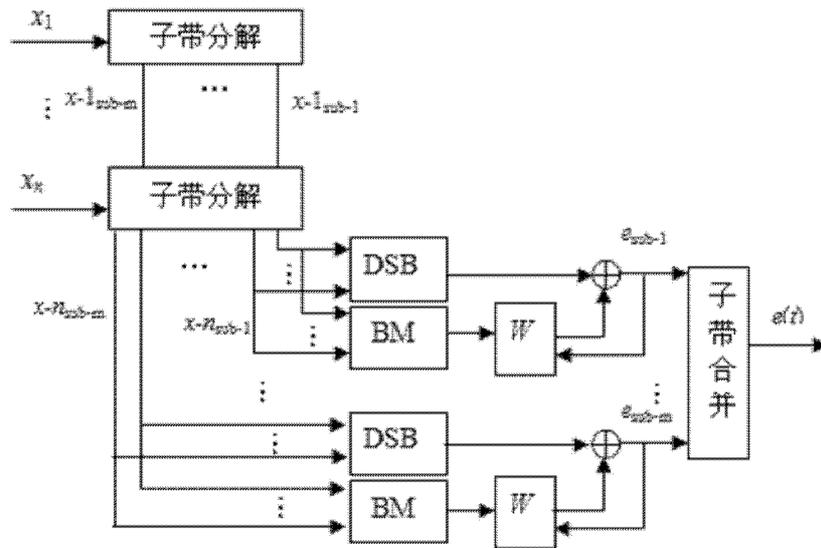


图 2

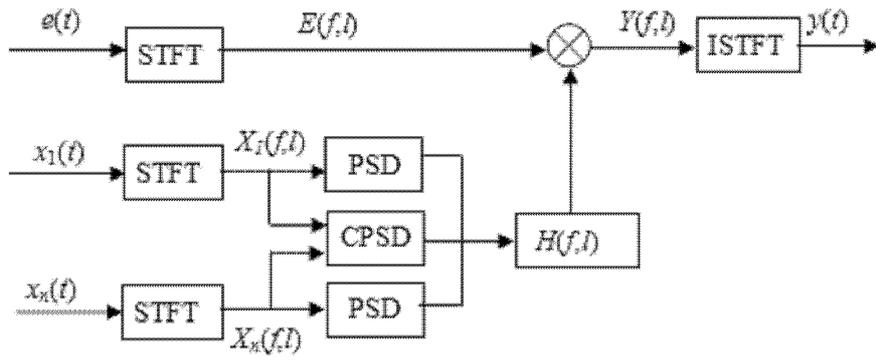


图 3