文章编号：0490-6756(2009)02-0407-06

一种基于双元麦克风线性阵的语音增强方法

王三山，何培宇，段文峰，何　悦
(四川大学电子信息学院，成都 610064)

摘　要：针对现有单通道语音增强算法及传统波束形成算法的局限性，提出了一种基于双元麦克风线性阵的语音增强方法．首先利用离线设计好的优化权值对输入信号进行加权求和以实现波束形成，然后结合一种新的噪声幅度谱估计方法，采用改进的幅度谱减法进一步增强语音信号．仿真实验表明该方法简单易行并取得了较好的语音增强效果．

关键词：语音增强；麦克风阵列；波束形成；谱减法；噪声幅度谱估计
中图分类号：TN912.35　　**文献标识码**：A

A method for speech enhancement based on two-elements microphone linear array

WANG San-Shan, HE Pei-Yu, DUAN Wen-Feng, HE Yue
(College of Electronics and Information, Sichuan University, Chengdu 610064, China)

Abstract: A method for speech enhancement based on two-elements microphone linear array is presented because of the limitation of existed single channel speech enhancement algorithm and conventional beamforming. Using the optimal weights designed off-line, the beamformer output is first obtained from a weighted sum of the input signals. Then combined with a new noise amplitude spectrum estimator, the speech signal is further enhanced by modified amplitude spectral subtraction. The experiments of simulation reveal the simplicity and practicability of the proposed method as well as its good speech enhancement performance.

Key words: speech enhancement, microphone array, beamforming, spectral subtraction, noise amplitude spectrum estimation

1　引　言

在许多非手持式语音通信系统中，由于声源距麦克风较远，以及房间反射波和环境噪声等干扰，使得麦克风接收到的是质量较差的带噪语音信号[1]．噪声降低了语音的信噪比和可懂度，因此进行语音增强非常必要．

在过去几十年里，以谱减法为代表的单通道语音增强算法得到了长足的发展，但这些算法均不能捕捉信号的空间信息，它们对各个方向来的噪声都一视同仁，这无疑加大了降噪需求，容易引入较大失真[2]．此外单通道语音增强算法对噪声谱估计方法还有较强的依赖性．

相比之下，基于麦克风阵列的波束形成算法以其独特的空间滤波特性，可以有效抑制非期望方向发出的噪声[1]．但传统的波束形成算法对阵列规

收稿日期：2008-03-05
基金项目：国家自然科学基金(60472096)
作者简介：王三山(1985—)，女，布依族，贵州安龙人，硕士研究生，主要研究语音信号处理．E-mail：sanshan.1985@yahoo.com.cn
通讯作者：何培宇．E-mail：hpysbsy@163.com

模有一定要求[3],往往不适合工程应用. 另外,在实际经常出现的散射噪声环境中或是当噪声与声源来自同一方向时,单独的波束形成算法对噪声的抑制并不充分,需要引入一定的后续处理[1].

针对以上问题,我们提出一种基于双元麦克风线性阵的语音增强方法. 该方法首先采用一种对阵列规模不敏感的波束形成器设计方法[4],针对两个阵元的情况离线设计优化权值向量,然后利用这些权值对输入信号进行加权求和处理,实现波束形成,最后结合一种新的噪声幅度谱估计方法,利用改进的幅度谱减法进一步增强语音信号. 该方法只需要两个麦克风,且计算复杂度低,易于实时处理,增强效果令人满意.

2 系统描述

我们描述的系统模型如图1所示.

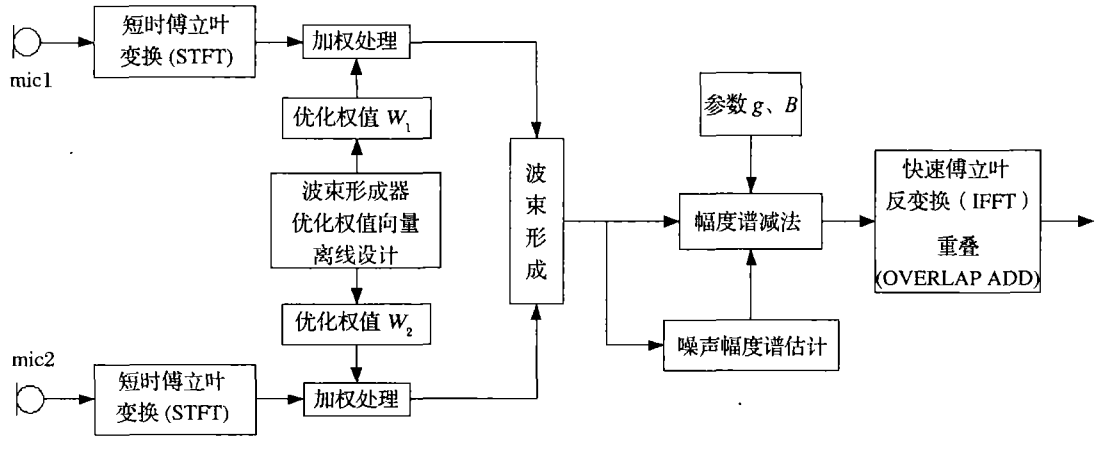

图1 语音增强系统模型

Fig. 1 Speech enhancement system configuration

2.1 波束形成器工作原理

假设麦克风线性阵的两个阵元分别位于 $p_1(x_1,y_1,z_1)$ 和 $p_2(x_2,y_2,z_2)$,则阵列信号为 $x_1(t,p_1)$ 和 $x_2(t,p_2)$,t 表示时间变量. 设两个麦克风对来波的方向增益,即本身的方向性函数已知,分别表示为 $U_1(f,r)$ 和 $U_2(f,r)$,其中 f 为频率,$r(x,y,z)$ 为声源位置,在球面坐标内映射为 $r(\varphi,\theta,\rho)$. 直角坐标和球面坐标的关系如图2所示.

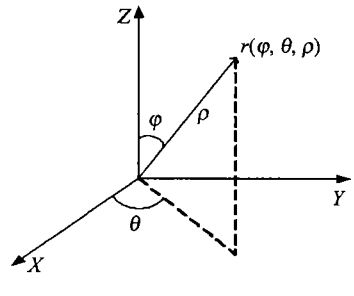

图2 直角坐标与球面坐标示意图

Fig. 2 Rectangular coordinate system and spherical coordinate system

在上述假设条件下,我们在频域上讨论波束形成器优化权值的设计. 设位于点 r 的声源信号频谱为 $S(f,r)$,则两个麦克风阵元采集到的语音信号的频谱分别为:

$$\begin{aligned} X_1(f,p_1) &= D_1(f,r)U_1(f,r)S(f,r)+N_1(f)\\ X_2(f,p_2) &= D_2(f,r)U_2(f,r)S(f,r)+N_2(f) \end{aligned} \tag{1}$$

其中,$N_1(f)$ 和 $N_2(f)$ 表示噪声,包含反射波、环境噪声以及器件噪声等. 而 $D_1(f,r)$ 和 $D_2(f,r)$ 分别表示源信号到达两个麦克风时所经历的相位延迟和幅度衰减,定义为:

$$\begin{aligned} D_1(f,r) &= \frac{e^{-j2\pi fc^{-1}\|r-p_1\|}}{\|r-p_1\|},\\ D_2(f,r) &= \frac{e^{-j2\pi fc^{-1}\|r-p_2\|}}{\|r-p_2\|} \end{aligned} \tag{2}$$

上式中 $\|\cdot\|$ 表示向量的2范数,c 表示声波在空气中的转播速度,一般取 $c=340$ m/s.

假设在频域上设计的优化权值为 $W_1(f,r)$ 和 $W_2(f,r)$,首先对两个麦克风采集到的信号分别进行加权处理,然后将加权后的两路信号相加,即得到波束形成器的输出 $Y(f,r)$,这个过程可以表示为:

$$Y(f,r)=W_1(f,r)X_1(f,p_1)+$$

$$W_2(f,r)X_2(f,p_2) \tag{3}$$

将(1)代入(3)，得到：

$$Y(f,r) = ((W_1(f,r)D_1(f,r)U_1(f,r) + W_2(f,r)D_2(f,r)U_2(f,r))S(f,r) + N(f) \tag{4}$$

其中，$N(f) = W_1(f,r)N_1(f) + W_2(f,r)N_2(f)$.

由(4)可定义波束成形函数 $B(f,r)$：

$$B(f,r) = W_1(f,r)D_1(f,r)U_1(f,r) + W_2(f,r)D_2(f,r)U_2(f,r) \tag{5}$$

不难看出，$B(f,r)$ 即是波束形成器对源信号的方向增益，它代表了波束形成器的方向性.

2.2　波束形成器优化权值向量离线设计

为了得到波束形成器的优化权值 $W_1(f,r)$ 和 $W_2(f,r)$，设计按以下步骤进行：目标波束定义、最小二乘拟合、权值向量归一化.

首先可定义目标波束函数[4] $T(\varphi,\theta,\rho,\delta)$：

$$T(\varphi,\theta,\rho,\delta) = \cos(\frac{\pi(\varphi_T-\varphi)}{\delta}) \cdot \cos(\frac{\pi(\theta_T-\theta)}{\delta}) \cdot \cos(\frac{\pi(\rho_T-\rho)}{k\delta}) \tag{6}$$

其中 $r_T(\varphi_T,\theta_T,\rho_T)$ 表示目标波束聚焦点，δ 表示波束主瓣宽度. 研究证明当 δ 小于 50 度后，拟合波束的主瓣已不能得到明显的改善，所以 $\delta=50$ 度是一个合理的取值. 另外，由于线性阵对仰角不敏感，且声源位置一般比较固定，为了简化运算，这里只考虑声源位于与阵列同一水平面某一固定位置的情况，即 $\varphi=90$ 度，$\rho=\rho_0$，ρ_0 为一定值，则(6)可修正为：

$$T(\theta) = A\cos(\frac{\pi(\theta_T-\theta)}{50}) \tag{7}$$

其中 A 为一固定增益，通常可取 $A=1$ 以简化运算.

接下来通过最小二乘法进行目标波束的拟合. 首先在工作区间均匀选取 L 个点代入(7)，则对每个频率 f 均可得到一个 $L\times1$ 维的列向量 $\boldsymbol{T}$. 同时，以矩阵形式可将(5)表示为

$$\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{W} \tag{8}$$

上式为一超定线性方程组，其中 $\boldsymbol{Z} = [D_1(f,r(l))U_1(f,r(l)),D_2(f,r(l))U_2(f,r(l))]$，$r(l)$ 表示第 l 个点对应的坐标，$\boldsymbol{W} = [W_1(f,r),W_2(f,r)]^T$ 为待设计的权值向量，$\boldsymbol{B}$ 为 $L\times1$ 维的列向量，表示真实的波束增益. 我们的设计目标即是使 $\boldsymbol{B}$ 在每个频率 f 上都尽量逼近对应的 $\boldsymbol{T}$. 这里我们采用加权最小二乘法来求解方程组，即使得加权均方误差 $\xi = |\boldsymbol{V}^T(\boldsymbol{T}-\boldsymbol{B})|^2$ 最小，其中 $\boldsymbol{V}$ 为 $L\times1$ 的列向量，表示加权系数，用以强调各点对 ξ 的不同影响. 最后得解：

$$\boldsymbol{W}_{opt} = (\boldsymbol{Z}^H\mathrm{diag}(\boldsymbol{V})\boldsymbol{Z})^{-1}\boldsymbol{Z}^H\mathrm{diag}(\boldsymbol{V})\boldsymbol{B} \tag{9}$$

为保证从目标波束聚焦点 r_T 传来的信号具有单位幅度增益和零相移，将 $\boldsymbol{W}_{opt}$ 归一化为：

$$\boldsymbol{W} = \frac{\boldsymbol{W}_{opt}}{\boldsymbol{Z}(f,r_T)\boldsymbol{W}_{opt}} \tag{10}$$

即得到所设计的波束形成器的优化权值向量 $\boldsymbol{W} = [W_1(f,r),W_2(f,r)]^T$.

由于实际设计中采用快速傅立叶变换(FFT)实现时域到频域的转换，所以应将以上讨论的连续频率变量 f 修正为离散频点变量 $k=1,2,\cdots,Nfft$，$Nfft$ 为 FFT 总频点数.

2.3　波束形成

实时处理中，首先将两个麦克风阵元分别置于 $p_1(x_1,y_1,z_1)$ 和 $p_2(x_2,y_2,z_2)$，按帧采集信号 $x_1(t,p_1)$ 和 $x_2(t,p_2)$，通过 $Nfft$ 点 FFT 对每帧信号进行短时傅立叶变换，得到 $X_1(\lambda,k,p_1)$ 和 $X_2(\lambda,k,p_2)$，λ 表示帧的序号，k 表示离散频点；然后利用离线设计的目标波束对应特定方向的波束形成器的优化权值 $W_1(k,r)$ 和 $W_2(k,r)$ 对信号的离散频谱进行加权处理；最后将两路信号相加得到波束形成器的输出，即：

$$Y(\lambda,k,r) = W_1(k,r)X_1(\lambda,k,p_1) + W_2(k,r)X_2(\lambda,k,p_2) \tag{11}$$

2.4　幅度谱减法

谱减法以其原理简单、运算量小、易于实时实现等优点在单通道语音增强系统中得到广泛应用. 该方法总是基于以下加性模型：

$$y_{noisy}(t) = s_{speech}(t) + n_{noise}(t) \tag{12}$$

其中，$y_{noisy}(t)$ 表示带噪语音信号，$s_{speech}(t)$ 表示纯净语音信号，$n_{noise}(t)$ 表示与语音不相关的噪声信号.

传统的谱减法可分为功率谱减和幅度谱减. 为了进一步减小运算量，我们采用的是幅度谱减法. 首先由(12)可得到：

$$S_{speech}(\lambda,k) = Y_{noisy}(\lambda,k) - N_{noise}(\lambda,k) \tag{13}$$

其中，$S_{speech}(\lambda,k)$、$Y_{noisy}(\lambda,k)$ 和 $N_{noise}(\lambda,k)$ 分别表示纯净语音、带噪语音和噪声的频谱. 由于人耳对语音的相位不敏感，可用带噪语音相位代替纯净语音相位，所以由(13)可得：

$$S_{speech}(\lambda,k) = Y_{noisy}(\lambda,k)\frac{|Y_{noisy}(\lambda,k)|-|\hat{N}_{noise}(\lambda,k)|}{|Y_{noisy}(\lambda,k)|} = Y_{noisy}(\lambda,k)(1-\frac{|\hat{N}_{noise}(\lambda,k)|}{|Y_{noisy}(\lambda,k)|}) = Y_{noisy}(\lambda,k)G(\lambda,k) \quad (14)$$

其中$|\hat{N}_{noise}(\lambda,k)|$为估计的噪声幅度谱,由于估计值与真实值之间总存在一定差异,在某些频点上甚至可能大于带噪语音幅度谱,所以可使$G(\lambda,k)$为:

$$G(\lambda,k) = \max(0,1-\frac{|\hat{N}_{noise}(\lambda,k)|}{|Y_{noisy}(\lambda,k)|}) \quad (15)$$

传统的谱减法容易残留烦人的音乐噪声,参考文献[5],将(15)修正为:

$$G(\lambda,k) = \max(g,1-\frac{B|\hat{N}_{noise}(\lambda,k)|}{|Y_{noisy}(\lambda,k)|}) \quad (16)$$

其中,$g(0.01 \leqslant g \leqslant 0.1)$可以避免估计的纯净语音幅度谱小于$g|Y_{noisy}(\lambda,k)|$;而$B$为一平衡因子,控制着噪声削减和语音失真之间的平衡. 引入g和B的目的是在谱峰之间引入宽带噪声来掩蔽相邻的音乐噪声,当这两个参数合理取值时,该方法可以有效地抑制音乐噪声.

我们将波束形成的输出(如(10))作为上述修正的幅度谱减法的输入. 由于在谱减法中需要对输入信号先加窗再做频域变换,所以可按下式实现两种算法的级联:

$$Y_{noisy}(\lambda,k) = F\{F^{-1}\{Y(\lambda,k,r)\}\times wnd(n)\},\quad n=1,2,\cdots,N \quad (17)$$

$F\{g\}$和$F^{-1}\{g\}$分别表示FFT和IFFT,N表示帧长,$wnd(n)$表示窗函数,常用的有汉宁窗、汉明窗以及它们的变种.

2.5 噪声幅度谱估计

在实际经常出现的非平稳噪声环境中,传统的噪声谱估计方法比如VAD等往往不能及时捕捉动态噪声信息,从而不利于实时处理. 文献[6]提出了一种基于最小值统计的噪声功率谱估计算法. 该算法无需进行端点检测,但由于采用了一定长度的数据窗来搜寻最小值,所以会引入一定的延迟. 而文献[7]提出的连续最小值跟踪算法可连续更新噪声功率谱,具有更好的动态跟踪能力. 但它的一个固有缺点是估计的噪声功率会随带噪语音功率的增加而增加,却忽略真实噪声的变化.

针对以上问题以及功率谱和幅度谱之间的简单关系,我们利用连续最小值跟踪算法估计幅度谱的局部最小值,利用语音出现概率[8]来更新估计的噪声幅度谱. 该方法中,噪声幅度谱的估计按下式进行:

$$|N_{noise}(\lambda,k)| = \alpha_d(\lambda,k)|N_{noise}(\lambda-1,k)| + (1-\alpha_d(\lambda,k))|Y_{noisy}(\lambda,k)| \quad (18)$$

不难看出,$\alpha_d(\lambda,k)$的求解是关键. 该因子按下列方式更新:

$$\alpha_d(\lambda,k) = \alpha_d + (1-\alpha_d)p(\lambda,k) \quad (19)$$

α_d为一常值,通常可取为0.85,$p(\lambda,k)$为语音出现概率. 为求取$p(\lambda,k)$,首先定义带噪语音幅度谱与对应局部最小值的比:

$$S_r(\lambda,k) = \frac{|Y_{noisy}(\lambda,k)|}{Y_{min}(\lambda,k)} \quad (20)$$

其中,$Y_{min}(\lambda,k)$为跟踪得到的幅度谱局部最小值,可按非线性准则获取:

$$Y_{min}(\lambda,k) = \begin{cases} \gamma_c Y_{min}(\lambda-1,k) - \frac{1-\gamma_c}{1-\beta_c}(|Y_{noisy}(\lambda,k)|-\beta_c|Y_{noisy}(\lambda-1,k)|), & Y_{min}(\lambda-1,k) < |Y_{noisy}(\lambda,k)| \\ |Y_{noisy}(\lambda,k)| & \text{其它} \end{cases} \quad (21)$$

γ_c和β_c为控制非线性变化的参数,通常可取$\gamma_c = 0.998$,$\beta_c = 0.96$.

接下来设定一门限函数$\delta(k)$和语音出现判决函数$I(\lambda,k)$,根据$S_r(\lambda,k)$作以下判决:

$$\begin{cases} I(\lambda,k) = 1, & S_r(\lambda,k) > \delta(k) \\ I(\lambda,k) = 0, & \text{其它} \end{cases} \quad (22)$$

其中,$I(\lambda,k) = 1$表示有语音出现,$I(\lambda,k) = 0$表示没有语音出现. $\delta(k)$可取经验值:

$$\begin{cases} \delta(k) = 2, & 0 < k \leqslant \frac{Nfft}{4} \\ \delta(k) = 5, & \frac{Nfft}{4} < k \leqslant \frac{Nfft}{2} \end{cases} \quad (23)$$

则$p(\lambda,k)$可以表示为:

$$p(\lambda,k) = \alpha_p p(\lambda,k) + (1-\alpha_p)I(\lambda,k) \quad (24)$$

其中 α_p 为一常值，通常可取为 0.5.

3 仿真实验

为评估文中提出的方法在真实环境下的性能，我们对实际办公室中采集的带噪语音进行了仿真实验. 以扬声器播放的一段纯净语音作为声源，以周围环境中的空调、计算机风扇和房间反射波等若干随机噪声作为噪声源. 两个麦克风阵元相距 10 cm，扬声器位于正对阵列中心 40 cm 的位置，且与阵列保持在同一水平面. 二者的相对位置如图3所示.

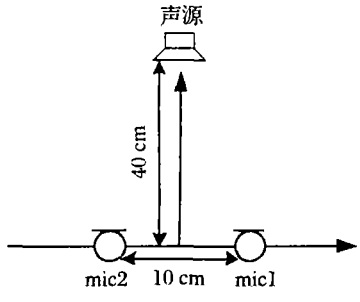

图 3　实验示意图

Fig. 3　Configuration in experiments

根据声源的位置，采用前面讨论的方法离线设计目标波束方向为 90°时波束形成器的优化权值向量，对应的方向图如图 4 所示.

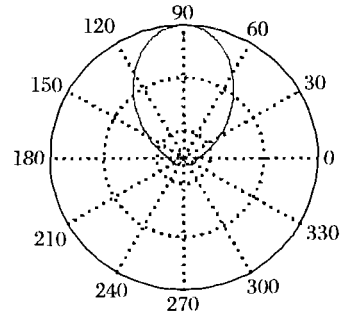

图 4　波束形成的方向图

Fig. 4　Directivity patterns after beamforming

实验中语音信号的采样频率为 16 kHz、采样精度为 16 bit、帧长为 512 点、帧间重叠率为 1/2、时域到频域的变换采用 512 点 FFT、幅度谱减法中参数 $g=0.1, B=6$. 图 5 给出采用不同的方法对带噪语音进行处理之后的信号波形.

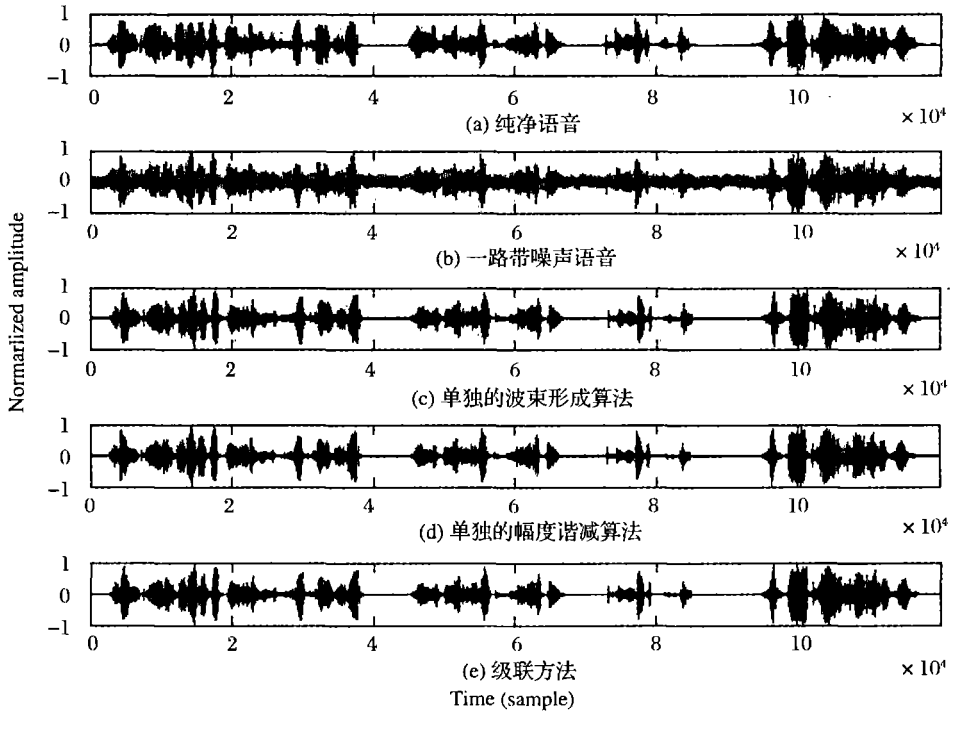

图 5　不同方法的处理结果

Fig. 5　Output of different methods

为了客观地考察不同方法对噪声的抑制,我们以降噪量来描述,这里降噪量定义为:

$$\Delta N = 10\log(P_{noise\_in}) - 10\log(P_{noise\_out}) \quad (25)$$

其中 P_{noise_in} 和 P_{noise_out} 分别表示处理前和处理后的噪声的平均功率. 这里我们利用纯噪声段来估计噪声的平均功率,然后按(25)计算得到不同方法的降噪量如表1所示.

表1 不同方法的降噪量

Tab. 1 Noise reduction of different methods

	单独的波束形成 (dB)	单独的幅度谱减法 (dB)	级联方法 (dB)
ΔN	20.26	34.85	38.06

结合表1,对处理后的各个信号进行主观试听得知:经波束形成算法单独处理后的语音比较清脆,对噪声有一定抑制但不够充分;经改进了的幅度谱减法单独处理后的结果取得了较大了降噪量,残留的音乐噪声极其微弱且对语音损伤不大,但声音略显沉闷,这主要是由于没有足够地抑制声音反射波;而经级联方法处理后的结果取得了最大的降噪量,残留的音乐噪声几乎不被察觉,且有效地消除了反射波的影响,获得了较好的声音质量.

4 结 语

我们将传统的波束形成算法和单通道谱减法作相应改进并将二者有效结合,提出了一种基于双元麦克风线性阵的语音增强方法. 该方法离线完成了相对复杂的波束形成器优化权值设计,修正了传统的幅度谱减法并在幅度谱上估计噪声. 仿真实验表明,本方法对真实环境中的带噪语音具有较好的增强效果且其原理简单、计算复杂度低,适合语音增强的实时处理.

参考文献:

[1] Brandstein M, Ward D. Microphone arrays: signal processing techniques and applications[M]. Berlin: Springer-Verlag, 2001.

[2] Ephraim Y, Cohen I. The electrical engineering handbook[M]. Boca Raton, CRC Press, 2004.

[3] Martin R, Vary P. Proceedings of IEEE digital signal processing workshop[M]. Illinois: Institute of Electrical and Electronics Engineers, 1992.

[4] Tashev I, Malvar H S. Proceedings of ICASSP[M]. Philadelphia: Institute of Electrical and Electronics Engineers, 2005.

[5] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise[J]. Proc IEEE Conf ASSP, 1979, 4(4): 208.

[6] Martin R.. Spectral subtraction based on minimum statistics [J]. Proc Eur Signal Process, 1994: 1182.

[7] Doblinger G. Proceedings of 4th european conference speech, communication and technology, EUROSPEECH'95 [M]. Madrid: Springer-Verlag, 1995.

[8] Sundarrajan R, Philipos C L. A noise-estimation algorithm for highly non-stationary environments [J]. Speech Communication, 2006, 4(48): 220

[责任编辑:李富河]