

一种基于二阶盲信号分离的多路声回波抑制模型

何培宇, 周激流, 夏秀渝, 王永德, 赵 刚

(四川大学电子信息学院, 四川成都 610064)

摘 要: 本文提出了一个基于二阶盲信号分离的多路声回波抑制模型. 该模型回避了多路声回波对消中因声回波源信号间的强互相关性所致的固有的解的非唯一性问题, 而是充分利用了这种互相关性来去除声回波. 模型仅添加一个辅助麦克风并巧妙置位即可对各路麦克风信号中的多路声回波进行有效的分离和抑制. 为了实时处理的目的, 提出了一个计算复杂度低且收敛稳健的二阶频域盲信号分离算法来检验该模型. 实验结果充分确认了提出模型的有效性.

关键词: 多路声回波抑制; 盲信号分离; 二阶统计特性; 互相关性

中图分类号: TN912.3 文献标识码: A 文章编号: 0372-2112(2006)11-2109-06

A New Multi-Channel Acoustic Echo Suppression Model Based on Second Order Blind Signal Separation

HE Pei yu, ZHOU Ji liu, XIA Xiur yu, WANG Yong de, ZHAO Gang

(School of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan 610064, China)

Abstract: A multi-channel acoustic echo suppression model (MCAESM) via second-order blind signal separation (BSS) is proposed. It avoids the inherent non-uniqueness problem in multi-channel acoustic echo cancellation due to strongly cross-correlated acoustic echo source signals. On the contrary, it makes full use of the cross-correlativity to remove these acoustic echoes. In this model, the multi-channel acoustic echoes from each microphone signal can be effectively separated and suppressed by employing only one additional well-positioned microphone. For the purpose of real-time processing, a new frequency domain BSS algorithm based on second-order statistics, with less computational complexity and stronger robustness, is proposed to verify the MCAESM. Experimental results confirm the effectiveness of the proposed model.

Key words: multi-channel acoustic echo suppression; blind signal separation; second-order statistics; cross-correlativity

1 引言

电视电话会议采用典型的全双工多通道音频系统来增强声音的真实感与现场感. 消除扬声器与麦克风空间声耦合所致多路声回波是系统语音通信质量的根本保障. 图 1 示出

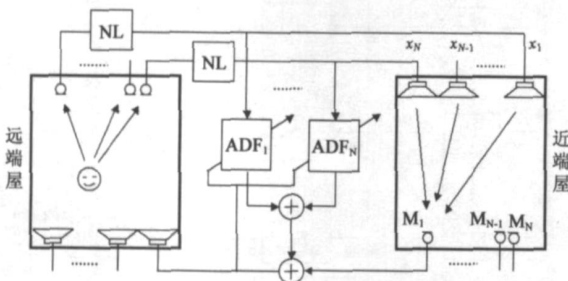


图 1 多路声回波对消构成图

了一个多路声回波对消器(MCAEC)的基本结构. 图中, ADF_1, \dots, ADF_N 表示用于多路声通道(多个扬声器——房间——一个麦克风, 简记为 MLRM)辨识的 N 个自适应滤波器. 图 1 只

示意了近端屋中一个麦克风的 N 路声回波对消结构, 近端屋的其它麦克风以及远端屋各麦克风对应的 N 路声回波对消结构均相同. 这里略去. 多路声通道 MLRM 的自适应辨识是多路声回波对消的关键. 然而, 自适应滤波器对 MLRM 辨识的解的非唯一性问题(NUP)是影响 MCAEC 性能的症结所在^[1,2]. NUP 的存在, 可能导致自适应辨识滤波器与房间多路声通道的严重失配, 进而降低 MCAEC 的性能. 导致 NUP 的根本原因在于声回波源信号(x_1, x_2, \dots, x_N)间存在强互相关性. 对 x_1, x_2, \dots, x_N 进行适度的非线性预处理(图 1 中, 以 NL 示出)以削弱它们间的互相关性的方法是目前广泛采用的较为有效的措施^[3,4], 但是这种方法是适当降低音质为代价的. 本文提出的基于语音盲信号分离的多路声回波抑制模型不需要直接辨识 MLRM, 更不必进行去互相关的非线性预处理, 而是充分利用了这种互相关性, 从麦克风信号中分离出并抑制掉多路声回波. 从房间硬件配置来讲, 只需要添加一个辅助麦克风并进行巧妙的定位, 提出的模型便产生功效. 鉴于基于高阶统计特性(HOS)类的盲信号分离方法^[5-8]计算复杂度极高难于实时

应用的事实,本文还提出了一个计算复杂度低,便于实时实现的基于二阶统计特性(SOS)的盲信号分离(BSS)频域自适应算法.实验结果证实了提出模型的有效性和可行性.

2 基于盲信号分离的多路声回波抑制模型 MCAESM

图 2 是一个新的基于 BSS 的多路声回波抑制模型 MCAESM.为描述方便,取 $N = 2$.图中, g_{11} 和 g_{21} 分别表示远端屋中声源信号 s_f 位置到两个麦克风的房间冲激响应(RIRs); x_1 和 x_2 是来自远端屋的两路立体声信号,即声回波源信号; s_f 和 s_n 分别表示远端语音信号和近端语音信号; h_{12} 、 h'_{11} 和 h'_{12} 分别表示从近端声源信号 s_n 、 x_1 和 x_2 位置到麦克风 M_1 的三路房间冲激响应.比较图 1 与图 2 的近端屋配置,可以看出,图 2 仅添置了一个附加麦克风 M_a 用于近端屋多路声回波抑制的辅助信号拾取.图 2 中, BSS 用于将麦克风 M_1 拾取的卷积混合信号分离成为多路声回波 y_1 和近端语音信号 y_2 .仅借助两个麦克风来分离多路声回波与近端语音信号是基于 $x_1x_2 \dots x_N$ 彼此间互相关的考虑.因为

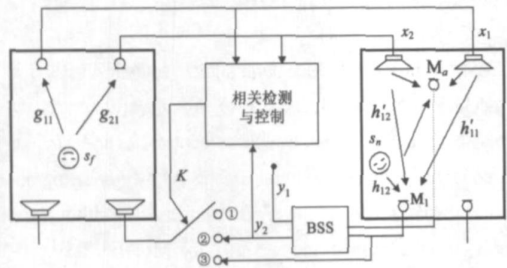


图 2 基于 BSS 的多路声回波抑制模型 ($N=2$)

$$x = g * s_f \tag{1}$$

其中

$$x = (x_1 \ x_2 \ \dots \ x_N)^T \tag{2}$$

$$g = (g_{11} \ g_{21} \ \dots \ g_{N1})^T \tag{3}$$

* 表示线性卷积运算.显然, $x_1x_2 \dots x_N$ 源于远端的同一信号 s_f .又因为 s_f 和 s_n 是相互独立的两个声源信号,因此无论要消除多少路声回波,提出模型中对于近端屋每一路麦克风信号的盲分离处理部分始终属于两路混合输入和两路盲分离输出情形,简记为 2×2 BSS.于是近端屋任意一个麦克风 M_i 和辅助麦克风 M_a 的拾取信号 x_{M_i} 和 x_{M_a} 均可表示为来自两源 s_f 和 s_n 的线性卷积混合信号

$$\begin{pmatrix} x_{M_i} \\ x_{M_a} \end{pmatrix} = \begin{pmatrix} \tilde{h}_{i1} & h_{i2} \\ \tilde{h}_{a1} & h_{a2} \end{pmatrix} * \begin{pmatrix} s_f \\ s_n \end{pmatrix} \tag{4}$$

这里

$$\begin{pmatrix} \tilde{h}_{i1} \\ \tilde{h}_{a1} \end{pmatrix} = \begin{pmatrix} h'_{i1} & h'_{i2} & \dots & h'_{iN} \\ h'_{a1} & h'_{a2} & \dots & h'_{aN} \end{pmatrix} * (g_{11} \ g_{21} \ \dots \ g_{N1})^T \tag{5}$$

\tilde{h}_{i1} 和 \tilde{h}_{a1} 分别表示远端声源 s_f 到近端屋第 i 个麦克风 M_i 和辅助麦克风 M_a 的等效房间冲激响应;同理, h_{i2} 和 h_{a2} 分别表示近端声源 s_n 到 M_i 和 M_a 的房间冲激响应. $i = 1, 2, \dots, N$.

用 p 表示 x_1 和/或 x_2 的功率; ρ 表示 x_1 或 x_2 与盲分离语音输出 y_2 间的相关系数.下面以两声道情形为例,说明图 2 所示 MCAESM 对第一路麦克风 M_1 的声回波抑制的操作步骤.

(1) 如果 $\rho \geq c_1$,我们检测 ρ ;否则将控制开关 K 置于图示位置 ③,这表明近端屋没有声回波,于是送至远端屋的近端信号直接取自麦克风 M_1 .

(2) 如果 $\rho \geq c_2$,控制开关 K 被直接置于图示位置 ①,这意味着没有近端语音存在,此时通过断路即可直接抑制多路声回波;否则将控制开关 K 置于图示位置 ②,这表明,在近端屋声回波和近端语音均存在.在连接端 ②,被送往远端的信号恰好是 BSS 分离出的近端语音信号.

对第二路麦克风 M_2 的声回波抑制,机理同上.以上, c_1 和 c_2 是两个小于 1 的阈值常数,具体取值由实际系统确定.

3 二阶频域 BSS 算法 SOSFBSS

MCAESM 多路声回波抑制的核心信号处理部分在于 BSS,构建一个简单易行的 BSS 算法是 MCAESM 迈向实时处理的关键一步.本文基于二阶统计特性构建了一个频域盲分离 SOSFBSS 算法.以 $M \times M$ BSS 为例,下面简要导出 SOSFBSS 算法. $M \times M$ BSS 的卷积混合输入和解卷积混合输出的离散时

$$z(n) = \sum_{l=0}^{\infty} A(l) s(n-l) \tag{6}$$

$$y(n) = \sum_{l=0}^{L_f-1} B(l) z(n-l) \tag{7}$$

其中, $s(n) = [s_1(n) \ s_2(n) \ \dots \ s_M(n)]^T$, $z(n) = [z_1(n) \ z_2(n) \ \dots \ z_M(n)]^T$ 及 $y(n) = [y_1(n) \ y_2(n) \ \dots \ y_M(n)]^T$ 分别表示 M 维源信号矢量、 M 维麦克风信号矢量及 M 维盲分离信号矢量; $A(n)$ 及 $B(n)$ 分别表示 $M \times M$ 维的离散时域卷积混合及解卷积混合矩阵; n 表示离散时间指数.在频域中,时域中的卷积混合被变为频域中的瞬时混合,即

$$Z(\omega_k, b) = \underline{A}(\omega_k, b) S(\omega_k, b) \tag{8}$$

$$Y(\omega_k, b) = \underline{B}(\omega_k, b) Z(\omega_k, b) \tag{9}$$

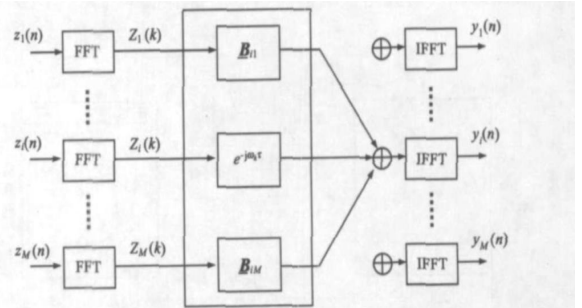


图 3 频域 $M \times M$ BSS 解卷积混合模型

其中, $\omega_k = 2\pi(k-1)/L$, k 表示离散频域指数,且 $k = 1, 2, \dots, L$; b 表示进入离散傅立叶变换处理的离散时域混合信号的帧指数.图 3 给出了频域 $M \times M$ BSS 解卷积混合模型图.图中, $z_1(n), \dots, z_M(n)$ 分别表示时域的 M 个麦克风卷积混合离散时间信号, $Z_1(k), \dots, Z_M(k)$ 分别表示与之对应的频域离散信号.由于时域中的卷积混合在频域中变成瞬时混合,所以第 i

路解混合输出 $y_i(n)$ 可以由图示方框中的 $M-1$ 个频域自适应滤波器解出。由于语音信号是短时平稳信号, 大量文献^[9-12]表明 BSS 基于二阶统计特性 (SOS) 便可分离卷积混合信号, 于是我们基于分离输出的互功率谱构造目标函数, 即

$$J(b) = \sum_{k=1}^L \sum_{m=1}^M \sum_{l=1}^M |(\mathbf{R}_y(\omega_k, b))_{ml}|^2 - \sum_{k=1}^L \sum_{m=1}^M |(\mathbf{R}_y(\omega_k, b))_{mm}|^2$$

$$= \sum_{k=1}^L \{ \text{tr}[\mathbf{R}_y(\omega_k, b) \mathbf{R}_y^H(\omega_k, b)] - \text{tr}[\text{diag}(\mathbf{R}_y(\omega_k, b)) \text{diag}(\mathbf{R}_y^H(\omega_k, b))] \} \quad (10)$$

其中 $\mathbf{R}_y(\omega_k, b) = \langle \mathbf{Y}(\omega_k, b) \mathbf{Y}^H(\omega_k, b) \rangle$ (11)
 $\text{tr}[\cdot]$ 表示求矩阵的迹, $\langle \cdot \rangle$ 表示求矩阵各元素的统计平均值, $(\cdot)^H$ 表示对矢量求转置共轭。由于源信号是彼此独立的, 所以目标函数应该达到最小, 于是解混合矩阵可由最速下降法迭代获得^[13], 即

$$\underline{\mathbf{B}}(\omega_k, b+1) = \underline{\mathbf{B}}(\omega_k, b) + \mu [- \frac{\partial J(b)}{\partial \underline{\mathbf{B}}^*(\omega_k, b)}]$$

$$= \frac{1}{2} \left[\frac{\partial J(b)}{\partial \text{Re}[\underline{\mathbf{B}}(\omega_k, b)]} + j \frac{\partial J(b)}{\partial \text{Im}[\underline{\mathbf{B}}(\omega_k, b)]} \right]$$

$$= 2 [\mathbf{R}_y(\omega_k, b) - \text{diag}(\mathbf{R}_y(\omega_k, b))] \mathbf{B} \mathbf{R}_z(\omega_k, b) \quad (12)$$

式中

$$\mathbf{R}_z(\omega_k, b) = \langle \mathbf{Z}(\omega_k, b) \mathbf{Z}^H(\omega_k, b) \rangle \quad (13)$$

我们采用重叠-存储 (Overlap Save) 方法^[14]来精确实现式 (7) 所示的线性卷积。同时, 我们对频域解混合滤波矩阵 $\underline{\mathbf{B}}(\omega_k, b)$ 的自适应迭代求解采用如下两个限制 C_1 和 C_2 , 即

(1) 由于频域中, 对应频点 ω_k 的各个解混合滤波矩阵 $\underline{\mathbf{B}}(\omega_k, b)$ 的对角元素均置为常数 $e^{-j\omega_k \tau}$ 或取值为 1 ($\tau = 0$ 时), 所以对应频点 ω_k 的各个梯度矩阵 $\frac{\partial J(b)}{\partial \underline{\mathbf{B}}^*(\omega_k, b)}$ 的对角元素必须强迫置零, 即

$$C_1 \left\{ \frac{\partial J(b)}{\partial \underline{\mathbf{B}}^*(\omega_k, b)} \right\} = \frac{\partial J(b)}{\partial \underline{\mathbf{B}}^*(\omega_k, b)} - \text{diag} \left\{ \frac{\partial J(b)}{\partial \underline{\mathbf{B}}^*(\omega_k, b)} \right\} \quad (14)$$

(2) 为了使得时域中的解混合滤波矩阵 $\mathbf{B}(n)$ 的非对角元素的冲激响应序列当 $n \geq L_f$ 时, $b_{ij}(n) = 0$, 其中, $i \neq j, i, j = 1, \dots, M, L_f$ 是每次进入频域处理的混合信号序列的帧长度。设 L 为 Overlap Save 方法中快速傅立叶算法 FFT 的频点数, 也是各个频域滤波器的长度, 通常取 $L = 2L_f$ 。于是我们采用如下的限制^[15]。

$$C_2 \left\{ \frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega, b)} \right\} = \mathbf{F} \mathbf{D} \mathbf{F}^{-1} \left\{ \frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega, b)} \right\} \quad (15)$$

其中 $\frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega, b)} = \left(\frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega_1, b)}, \frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega_2, b)}, \dots, \frac{\partial J(b)}{\partial \underline{\mathbf{B}}_{ml}(\omega_L, b)} \right)$ (16)

式 (15) 中, \mathbf{D} 是一个 $L \times L$ 维的对角矩阵, 其对角元素 $d_{ii} = 1$

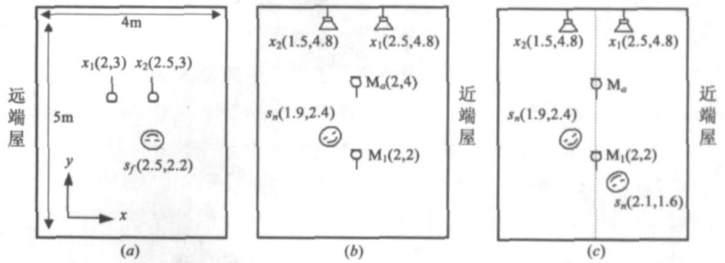


图 4 实验设置 (远近端屋: 长 5m、宽 4m、高 3m)

($i \leq L_f, d_{ii} = 0 (i > L_f)$). \mathbf{F} 和 \mathbf{F}^{-1} 分别是离散傅立叶变换矩阵和反变换矩阵。在频域中进行盲分离的不足是, 用于重建时域盲分离输出的频率分量可能来自不同的源信号。这是因为频域中的 BSS 解瞬时混合输出在各个频点上存在位置排列的模糊性。这种模糊性常致使解混合滤波器谱的不平滑。因此在时域中, 当 $n > L_f$ 时, 各解混合滤波器系数 $b_{ij}(n)$ 被强迫置零等效于在频域中借助一辛格函数的卷积运算来平滑其频谱。因此通过限制 C_2 能够有效地解决频域中各频点上 BSS 输出位置排列的模糊性问题。

考虑到高性能的 DSP 一般都具有硬件乘加单元, 以两个复数的一次乘运算作为计算复杂度 (CC) 的基本度量单位足以反映实时处理的运算量。对提出的 SOSFBSS 算法, 容易得出 2×2 BSS 中处理每帧输入数据的计算复杂度为 $CC = 4(7 + \log_2 L)L$ 。若考虑一个采样间隔内 SOSFBSS 的计算量并设 $L = 2L_f$, 则 $CC = 8(7 + \log_2 L)$ 。显而易见, 将提出的 SOSFBSS 算法用于 MCAESM 是简单易行的。

4 实验研究

4.1 信号间的相关性研究

本节将上述构建的 SOSFBSS 算法用于提出模型 MCAESM 的麦克风卷积混合信号分离, 并用以实验研究图 2 各路信号间的相关性, 以确认提出模型的有效性和可行性。定义两个平稳随机过程 $X(t)$ 和 $Y(t)$ 的相关系数为

$$\rho_{XY}(\tau) = R_{XY}(\tau) / \sigma_X \sigma_Y \quad (17)$$

这里, $R_{XY}(\tau)$ 是 $X(t)$ 和 $Y(t)$ 之间时滞为 τ 的相关函数; σ_X 和 σ_Y 分别是 $X(t)$ 和 $Y(t)$ 的标准差。图 4 是我们的实验设置。

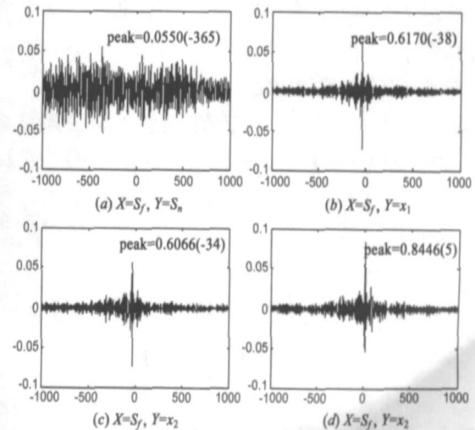


图 5 各路信号间的相关系数 $\rho_{XY}(\tau)$ 与时滞 τ 的关系

远端屋房间冲激响应 g_{11} 和 g_{21} 以及近端屋房间冲激响应 h'_{11} 、 h'_{12} 和 h_{12} 是在一个 4 米宽、5 米长、3 米高且混响时间为 $T_{60} = 0.24$ 秒的房间中测定的。房间中各声源、扬声器及麦克风的高度均置为 1.5 米，其平面位置由横纵坐标示出。远端声源 s_f 和近端声源 s_n 是两个独立的语音，它们分别是采样率 $f_s = 11025\text{Hz}$ 的英语和汉语录音。因此，远端来的立体声 x_1 和 x_2 以及麦克风信号 x_{M_1} 和 x_{M_a} 由测得的长度为 1024 的冲激响应 RIRs 经卷积混合而成。图 5 描绘了近端屋设置为图 4(b) 时以上各路信号间的相关系数 $\rho_{xy}(\tau, t)$ ，其中 t 是短时平稳语音信号的连续时间变量。从图中 (a) 可以看出， s_n 和 s_f 的相关系数 ρ 几乎为零，在时滞 $\tau = -365$ 处，取得的峰值仅为 0.0550，这是因为 s_n 和 s_f 彼此独立的缘故。正如预料，图 (b)、(c) 和 (d) 充分地反映了 s_f 和 x_1 、 s_f 和 x_2 以及 x_1 和 x_2 彼此间具有很强的相关性，相关系数的峰值分别为 0.6170, 0.6066 和 0.8446，均出现在零时滞的附近。由此可以断定，对两个多声源 ($s_n, x_1, x_2, \dots, x_N$) 卷积混合信号实施盲分离是可行的，分离输出必定是近端语音信号和与远端声源相关的声回波成分。

为了决定相关系数的门限 c_2 ，我们进一步探讨在不同的卷积混合条件下，两路盲信号分离输出信号 y_2 和 y_1 之间的相关性、 y_2 与声回波源信号 x_2 或 x_1 间的相关性。实验结果如图 6 所示。图 6(a) 表示近端麦克风卷积混合信号由近端语音信号 s_n 和声回波源信号 x_1 及 x_2 卷积混合而成，经过盲分离后的输出信号 y_2 和 y_1 几乎不相关 ($\rho_{\max} = 0.0766$)；图 6(b) 表示近端麦克风卷积混合信号仅源于近端语音信号 s_n 和一路声回波源信号 x_2 ，经过盲分离后的输出信号 y_2 和 y_1 几乎不相关 ($\rho_{\max} = 0.0764$)。这表明 SOSFBSS 算法不仅能分离单路声回波源信号和近端语音的卷积混合信号，并且也能分离相关的多路声回波源信号和近端语音的卷积混合信号，试听效果也几乎无异。图 6(c) 表明，当近端语音 s_n 及声回波源信号 x_1 和 x_2 均混入麦克风时，近端语音分离输出 y_2 和 x_2 几乎不相关 ($\rho_{\max} = 0.0709$)。同图 6(d) 相比，在麦克风信号只源于 x_1 和 x_2 时，SOSFBSS 算法分离输出 y_2 和 x_2 有较强的相关性 ($\rho_{\max} = 0.2371$)。从这种差异，我们就能决定门限 c_2 ，比如，设 $c_2 = 0.1500$ 。因此，对声回波源信号功率 $p \geq c_1$ 的情形，当 x_1 或 x_2 与盲分离语音输出 y_2 间的相关系数 $\rho \geq c_2$ 时，将软件开关 K 置于图 2①端即可以直接消除多路声回波；否则，将软件开关 K 置于图 2②端，让 BSS 分离输出的近端语音前往远端。

4.2 辅助麦克风的定位研究

在图 4(c) 中，我们通过改变辅助麦克风 M_a 的位置来研究提出的 SOSFBSS 算法对多路声回波抑制的功效。图 4(c) 中，保持 M_1 不动，将近端声源定位于 (1.9, 2.4) 或 (2.1, 1.6)。让 M_a 沿虚线上下移动并用 d 表示 M_a 与 M_1 间的垂直距离，且 $d > 0$ 表示 M_a 在 M_1 的上方， $d < 0$ 表示 M_a 在 M_1 的下方。改变 d 的大小，我们通过测定声回波抑制比来考查 M_a 的可行性位置。这里定义声回波抑制比为

$$ESR = 10 \log(P_{y_2}/P_{x_{M_1}}) \quad (18)$$

其中， P_{y_2} 、 $P_{x_{M_1}}$ 分别表示图 2 中 BSS 输出 y_2 与麦克风输入信号 x_{M_1} 的平均功率。ESR 越小表示多路声回波被抑制得越

多，反之亦然。测量时，我们先让近端语音与多路回声均存在，首先对 BSS 进行自适应训练，待 BSS 收敛后，固定各滤波器系数，在关闭近端语音仅保留多路回声的条件下来测定 P_{y_2} 和 $P_{x_{M_1}}$ 。图 7 描绘了多路声回波抑制比 ESR 与 d 关系曲线。图中，三角连线是近端声源置于 (1.9, 2.4) 的条件下测得的结果，圆圈连线是近端声源置于 (2.1, 1.6) 的条件下测得的结果。

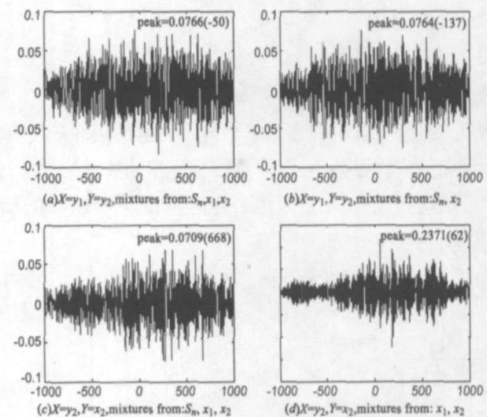


图 6 各路信号间的相关系数 $\rho_{xy}(\tau)$ 与时滞 τ 的关系

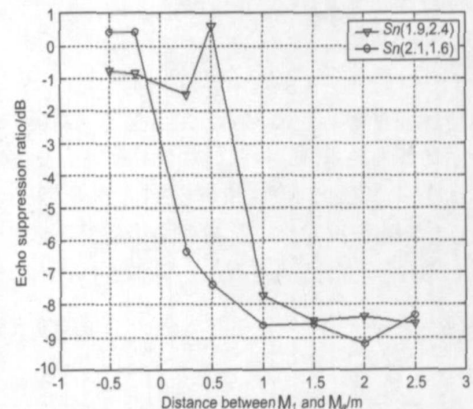


图 7 多路声回波抑制比 ESR 与两麦克风间距离 d 的关系曲线

从图 7 可以看出，随着麦克风 M_a 沿虚线移动与固定麦克风 M_1 的垂直距离 d 从 -0.5m 到 2.5m 的增大，两种实验条件下的 ESR 均呈现总体减小的趋势并在 $d = 1\text{m}$ 附近明显减小；当 $d > 1.5\text{m}$ 以后，多路声回波抑制比 ESR 基本上保持在 -8.5dB 左右。从图中还可以看出，ESR 在两个极端距离下相差高达 9dB 左右，这充分表明辅助麦克风 M_a 置放位置对多路声回波抑制的重要性。它离扬声器越近，多路声回波越强，作为辅助麦克风的主导信号的地位越明确，实质作用与 BSS 前端的波束形成对语音信号的预处理是一样的。因此在辅助麦克风 M_a 靠近扬声器时，多路声回波抑制得到明显改善。对于近端声源置于 (1.9, 2.4) 的情形，当 $d = 0.5\text{m}$ 时， $ESR = 0.5923\text{dB}$ ，这表明进入主麦克风 M_1 的多路声回波非但没有被削弱，反而略有增强。最主要的原因是当 $d = 0.5\text{m}$ 时，辅助麦克风 M_a 离近端源的距离 (0.1414m) 比主麦克风 M_1 离近端源的距离 (0.4123m) 还小，此时麦克风 M_a 所拾取的近端语音信

号明显强于主麦克风 M_1 对 S_n 的拾取,这就导致 BSS 处理在我们期待的近端语音分离输出端(如图 2 ②端所示)强化多路声回波而弱化近端语音。用同样的理由我们可以解释近端声源置于 $(2, 1, 1.6)$ 时, $d = -0.5\text{m}$ 和 $d = -0.25\text{m}$ 对应的 $ESR > 0\text{dB}$ 的现象。由于现实中近端声源在主麦克风附近的位置具有随意性,因此辅助麦克风 M_a 的放置应遵循拾取的多路声回波较近端语音强的原则,即置于靠近辅助麦克风和远离主麦克风的位置,这样才能保证 BSS 将多路声回波与近端语音分开并确保分离出的近端语音在期待的端口输出,最终抑制掉多路声回波。

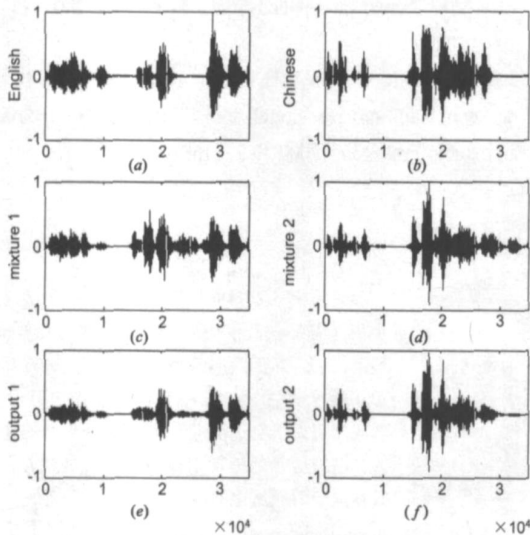


图 8 基于 SOSFBSS 算法的多路声回波抑制实验结果 $M_1(2,2), M_a(2,4), d=2\text{m}$

现以图 4(a) 和 (b) 的实验设置为例,让 M_a 置于 $(2, 4)$ 处,即 $d = 2\text{m}$ 。图 8 给出了一组基于 SOSFBSS 算法的多路声回波抑制实验结果图。图 8(a) 和 (b) 分别是近端英语和远端汉语的语音信号波形图; (c) 和 (d) 分别是主麦克风 M_1 和辅助麦

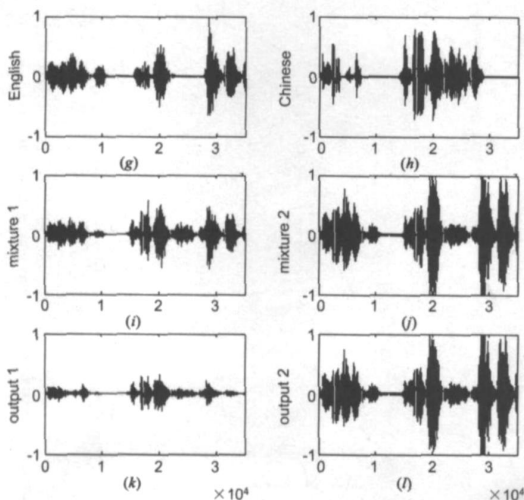


图 9 基于 SOSFBSS 算法的多路声回波抑制实验结果 $M_1(2,2), M_a(2,2.5), d=0.5\text{m}$

克风 M_a 拾取的语音卷积混合信号波形图; (e) 是来自图 2 BSS 近端语音盲分离输出 y_2 的波形图; (f) 是来自图 2 BSS 声回波分离输出 y_1 的波形图。比较 BSS 分离前后的各路语音波形,我们可以看出, BSS 输出信号 y_2 的波形基本与近端语音信号波形一致,而 BSS 输出信号 y_1 的波形大致与远端语音信号波形相似。当然, BSS 输出 y_1 如何并不重要,最重要的是 BSS ②端输出应尽量保持近端语音的成分并极大地抑制混入主麦克风的多路回声信号。保持同样的实验设置,让 M_a 置于 $(2, 2.5)$ 处,即 $d = 0.5\text{m}$,图 9 给出了一组与图 8 对应的实验结果。图 8(a)(b)(c) 与图 9(g)(h)(i) 完全相同。比较图 8(d) 和图 9(j) 所示的辅助麦克风 M_a 拾取的语音卷积混合信号波形图,我们可以看出,后者的声回波成分明显减弱。图 9(k) 与 (l) 分别表示对应的 BSS 输出信号 y_1 和 y_2 。比较图 9(g) 与 (k) 及图 9(h) 与 (l),显然,此时的 BSS 两端输出 y_1 和 y_2 与我们期待的盲分离输出相差甚远,对应的语音波形间已找不到相似之处。无论从语音波形图、多路声回波抑制比 ESR 、还是从主观试听来对比这两种实验条件下的实验结果,辅助麦克风 $d = 2.0\text{m}$ 的设置均优于 $d = 0.5\text{m}$ 的设置,即前者的多路声回波抑制效果明显优于后者,改善高达 9dB 。

5 结论

提出的多路声回波抑制模型实验表明是有效的。该模型彻底避免了常规的多路声回波抵消方法中自适应滤波器解的非唯一性问题。常规的模型中,由于各路声回波源信号间存在互相关性使得自适应滤波器不能唯一地辨识多路声回波通道从而造成多路声回波抵消性能易于恶化的不良后果。提出的模型恰恰利用了这种互相关性,仅在房间中添加一个辅助麦克风就可以对各个主麦克风构成两个独立声源、两个麦克风、两路盲信号分离输出的回波抑制模型。整个模型的功效就建立在对各种信号的相关性检测,开关控制及盲信号分离的操作上。各种信号的相关性检测及开关控制可以借鉴常规的多路声回波抵消模型的操作^[19]。因此,提出模型的性能主要取决于盲信号分离性能, BSS 性能越好,多路声回波抑制效果就越好。为实时处理的目的,本文还提出了一个的基于二阶统计特性的 SOSFBSS 算法用于模型的有效性及可行性检验,该算法计算复杂度低且收敛稳健。特别地,实验研究表明,对辅助麦克风的巧妙定位能使多路声回波抑制性能得以显著改善。

参考文献:

- [1] M M Sondhi, D R Morgan. Stereophonic acoustic echo cancellation—an overview of the fundamental problem[J]. IEEE SP Lett, 1995, 2(8): 148–151.
- [2] J Benesty, D R Morgan, M M Sondhi. A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation[J]. IEEE Trans on Speech and Audio Processing, 1998, 6(2): 156–165.
- [3] D R Morgan, et al. Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation[J]. IEEE Trans on Speech and Audio Processing, 2001, 9(6): 686–696.

- [4] T Gansler, J Benesty. An adaptive nonlinearity solution to the uniqueness problem of stereophonic echo cancellation[A]. Proceedings of ICASSP 02[C]. 2002. 1885– 1888.
- [5] P Comon. Independent component analysis[A]. Proc International Workshop on Higher Order Statistics (HOS ' 91) [C]. Chamrousse, France: 1991. 111– 120.
- [6] J F Cardoso. Iterative techniques for blind source separation using only fourth order cumulants[A]. Proc 6th European Signal Processing Conference (EU SIPC O ' 92) [C]. Brussels, Belgium: 1992. 739– 742.
- [7] D Yellin, E Weinstein. Criteria for multichannel signal separation[J]. IEEE Trans on Signal Processing, 1994, 42(8): 2158– 2168.
- [8] B A Pearlmutter, L C Parra. Maximum likelihood blind source separation: A context sensitive generalization of ICA[A]. Proc Advances in Neural Information Processing Systems (NIPS ' 96) [C]. Denver, Colo, USA: M I T Press, 1996. 613– 619.
- [9] E Weinstein, et al. Multi channel signal separation by decorrelation[J]. IEEE Trans on Speech and Audio Processing, 1993, 4(1): 405– 413.
- [10] L Parra, et al. Convolutional blind source separation based on multiple decorrelation[A]. Proc. of NNSP98[C]. Cambridge, UK: 1998. 23– 32.
- [11] H Buchner, R Aichner, W Kellermann. Blind source separation for convolutional mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity[A]. Proc. Int. Workshop Acoustic Echo Noise Control (IWAENC) [C]. Kyoto, Japan: 2003. 275– 278.
- [12] H Buchner, R Aichner, W Kellermann. A generalization of blind source separation algorithms for convolutional mixtures based on second order statistics[J]. IEEE Trans on Speech and Audio Processing, 2005, 13(1): 120– 134.
- [13] 何培宇. 面向实时处理的自适应声学语音信号处理[D]. 四川大学, 2005.
- [14] Simon Haykin. Adaptive Filter Theory, Third Edition[M]. Englewood Cliffs, NJ: Prentice Hall, 1996. 445– 458.
- [15] Bin Yin, P C W Sommen, Peiyu He. Exploiting acoustic similarity of propagating paths for audio signal separation[J]. EURASIP Journal on Applied Signal Processing, 2003, (11): 1091– 1109.
- [16] J Benesty, D R Morgan, J H Cho. A new class of doubletalk detectors based on cross correlation[J]. IEEE Tran on Speech and Audio Processing, 2000, 8(2): 168– 172.

作者简介:



何培宇 女, 1963 年出生于四川茂汶, 1986 年获清华大学学士学位, 1989 年获四川大学硕士学位, 2005 年获四川大学博士学位, 2000 年~2001 年于荷兰 Eindhoven 工业大学作访问学者。现任四川大学电子信息学院教授, 主要研究方向为自适应信号处理理论与 DSP 实时处理技术及其在通信等领域中的应用。

E-mail: hpeysy@163.com