

一种有效的语音盲信号分离简化混合模型

何培宇¹, 殷 斌², P. C. W. Sommen²

(1. 四川大学电子信息学院, 四川成都 610064; 2. Eindhoven University of Technology, Eindhoven, The Netherlands)

摘 要: 本文针对语音盲信号分离问题, 提出了一种新的简化混合模型. 基于从同一声源到距离靠近 (距离大约在几十厘米的范围内) 的两个麦克风的房间冲激响应具有相似性这一特点, 该模型只需利用这两个冲激响应的差异信息, 因而简化了盲源分离的计算复杂度. 自行设计实现的语音盲信号分离实时 DSP 实验系统证实, 该模型有效可行, 特别是通过引入时间延迟来解决非最小相位传递函数的求逆问题使得盲信号分离性能得到显著的改善.

关键词: 盲信号分离; 简化混合模型; 非最小相位系统求逆; 实时 DSP 系统

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2002) 10-1438-03

An Effective Simplified Mixing Model in Audio Blind Signal Separation

HE Pei-yu¹, YIN Bin², P. C. W. Sommen²

(1. Faculty of Electronics Information, Sichuan University, Chengdu, Sichuan 610064, China;

2. Eindhoven University of Technology, Eindhoven, The Netherlands)

Abstract: A novel simplified mixing model (SMM) is proposed for audio blind signal separation (ABSS). The SMM uses the fact that the room impulse responses from a single source to two closely spaced (approximately several 10 cms apart from each other) microphones are very similar. Only the difference between these two impulse responses is needed. As a result, the computational complexity of ABSS is reduced to some extent. A real-time blind signal separation DSP experimental system has been designed and implemented based on the SMM. The real-time experimental results have shown that the proposed simplified mixing model is effective and feasible. In particular, the blind signal separation performance can be improved considerably by introducing a time delay to solve the inverse problem of a non-minimum phase transfer function.

Key words: blind signal separation; simplified mixing model; inverse of a non-minimum phase system; real-time DSP system

1 引言

盲信号分离(BSS)是指在对此独立的源信号的混合过程及各源信号本身均未知的情况下, 如何从混合信号中分离出这些源信号. 在声学信号处理中, 当多个讲话者同时讲话时, BSS可以用来从多个麦克风混合语音信号中提炼出单个的语音信号. 这将在明晰的电信会议系统终端中得到日益增长的应用^[1].

对于非平稳的源信号, 文献[2, 3]表明利用二阶统计特性便足以分离混合信号. 与利用高阶统计特性的BSS相比, 其计算复杂度得以降低. 即使如此, BSS的复杂运算仍然是实时盲信号分离的主要障碍. 基于对语音盲信号分离可行性的考虑, 我们提出了一个简化的混合模型(SMM). 这个混合模型基于从同一源到两个靠得很近 (通常为几十厘米的距离) 的麦克风的房间冲激响应(RIR)具有相似性这一特点, 只用这两个冲激响应的差异(DIR)信息来构成SMM的核心要素, 因此降低了盲分离的计算复杂度.

鉴于对该SMM有效性、实时可行性的考察, 我们设计实现了一个语音盲信号分离DSP实时实验系统. 以LabVIEW虚拟仪器作为控制显示界面极大地方便了对SMM有效性、可行性的研究. 该系统首先对两个冲激响应函数的差异进行自适应实时测量, 然后用测得的结果来求解SMM, 从而分离出独立的源信号. 实时实验结果证实了该SMM的有效性及可行性. 特别是对该SMM引入时间延迟后, 解决了非最小相位求

逆问题, 因而使得盲分离性能得以极大的改善.

2 盲信号分离简化混合模型

在一个混响环境中, 语音的混合过程通常可以用如下数学模型来表示

$$\begin{pmatrix} x_1(t) \\ \dots \\ x_n(t) \end{pmatrix} = \begin{pmatrix} h_{11}(t) & h_{21}(t) & \dots & h_{n1}(t) \\ \dots & \dots & \ddots & \dots \\ h_{1n}(t) & h_{2n}(t) & \dots & h_{nn}(t) \end{pmatrix} * \begin{pmatrix} S_1(t) \\ \dots \\ S_n(t) \end{pmatrix} \quad (1)$$

此处, S_i 及 x_j 分别表示第 i 个声源信号和第 j 个麦克风拾取的混合信号. h_{ij} 表示第 i 个声源位置 (L_i) 到第 j 个麦克风位置 (M_j) 的房间冲激响应, 它可以近似地用 FIR 滤波器来表示. 符号 $*$ 代表线性卷积. 这里, 我们假设声源及麦克风的数量相等, 设为 n . 因而声源信号就可以在获得房间冲激响应的估计后, 通过求逆混合过程获得 (前向模型方法), 或者通过直接的解混合过程模型求得 (后向模型方法)^[3]. 通常, RIR 长达数百毫秒. 若采用 8kHz 的采样率, 则 FIR 滤波器的权数长达数千个. 要估计诸 h_{ij} 数千个权系数, 其运算复杂度便可想而知了. 进而, BSS 的实时实现将面临更大的挑战. 鉴于此, 我们将上述混合模型简化如下

$$\begin{pmatrix} x_1(t) \\ \dots \\ x_n(t) \end{pmatrix} = \begin{pmatrix} (t) & h_{21}(t) & \dots & h_{n1}(t) \\ \dots & \dots & \ddots & \dots \\ h_{1n}(t) & h_{2n}(t) & \dots & (t) \end{pmatrix} * \begin{pmatrix} \bar{S}_1(t) \\ \dots \\ \bar{S}_n(t) \end{pmatrix} \quad (2)$$

其中, $\bar{s}_i = h_{ii} * s_i$ (3)

$$h_{ij} = h_{ij} * h_{ii} \quad i, j = 1, 2, \dots, n \quad (4)$$

h_{ij} 表示从同一 L_i 到 M_i 及 M_j 的两个冲激响应的 DIR. 当两个麦克风 M_i 及 M_j 间的距离靠得较近时, 因 h_{ii} 和 h_{ij} 间存在相似性, 故我们可以用较少权数, 或者说主要能量都集中在前有限个权系数的 h_{ij} 来补偿这两个冲激响应的差异, 如式 (4) 所示. 这种假设将被后面的实验结果证实是可行的. 我们所要做的就是首先估计 DIR, 然后通过求逆这个 SMM 得到源信号. 从式 (2) 可知, 各混合信号 x_j 仅仅通过差异信息 h_{ij} 和诸独立的到达声源信号 \bar{s}_i 联系起来, 而不是通过常规的冲激响应联系起来. 从这个角度来看, 我们提出的 SMM 的计算复杂度得以降低到某种程度. 图 1 描述了 $n = 2$ 时的简化混合模型.

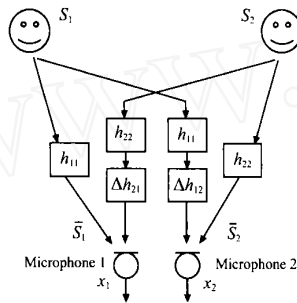


图 1 简化混合模型 SMM

SMM 有如下三个主要优点: 待估的滤波器数由 n^2 降至 $n(n-1)$; 盲分离出的源信号是到达麦克风前端的源信号 \bar{s}_i 而不是源端的 s_i , 这种分离结果常常听起来更为自然; 当两个麦克风靠得很近时, 两者的冲激响应差异更接近于一个仅有时间延迟的函数. 于是只需要更少的权系数来描述这个差异. 这一情形将明显地降低计算复杂度, 使实时分离成为可能.

3 语音盲信号分离 DSP 实验系统

3.1 基本方案

由于 DSP 处理离散时间序列, 对 $n = 2$ 的情形 (两个声源, 两个麦克风), SMM 在离散时间域的表达式如下

$$\begin{bmatrix} x_1(m) \\ x_2(m) \end{bmatrix} = \begin{bmatrix} 1 & h_{21} \\ h_{12} & 1 \end{bmatrix} * \begin{bmatrix} \bar{s}_1(m) \\ \bar{s}_2(m) \end{bmatrix} \quad (5)$$

这里, m 是离散时间指数. 为了获得 \bar{s}_1 的估计 \hat{s}_1 及 \bar{s}_2 的估计 \hat{s}_2 , 我们将式 (5) 的卷积运算变为离散傅立叶变换 (DFT) 中的乘积运算, 表示如下

$$\begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} \cong \begin{bmatrix} 1 & H_{21}(k) \\ H_{12}(k) & 1 \end{bmatrix} \begin{bmatrix} \bar{S}_1(k) \\ \bar{S}_2(k) \end{bmatrix} \quad (6)$$

因此, 频域中的 \hat{s}_1 及 \hat{s}_2 能被导出

$$\begin{bmatrix} \hat{S}_1(k) \\ \hat{S}_2(k) \end{bmatrix} \cong F(k) \begin{bmatrix} 1 & -H_{21}(k) \\ -H_{12}(k) & 1 \end{bmatrix} \begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} \quad (7)$$

此处 $F(k) = \frac{1}{1 - H_{12}(k)H_{21}(k)}$ (8)
 k 表示离散频点指数.

图 2 和图 3 描述了基于二阶 SMM 的 BSS 实现方案. 我们采用重叠-存贮 (Overlap-Save) 方法来精确实现式 (5) 所示的线性卷积, 用频域自适应滤波器 (FDAF) 的快速 LMS 算法 (fast LMS algorithm)^[4] 来实时估计 H_{12} 及 H_{21} , 然后将估计结果拷贝到 BSS 结构中, 最后进行实时盲信号分离. 图 2 描述了频域 H_{12} 的估计测量方案. 这里, 一个均方误差被定义为

$$MSE_e = 10 \log(P_e / P_d) \quad (9)$$

P_e 及 P_d 分别表示误差输出 e 及期待信号输入 d 的功率. 一个类似的测量方法及 MSE_e 定义被用于 H_{21} 的估计情形.

3.2 灵活的 DSP 实时 BSS 实验系统 (ES)

该 ES 主要由数字信号处理器 DSP (德州仪器, TMS320c40), 虚拟仪器 LabVIEW (国家仪器) 与 A/D 与 D/A 变换器等相关软硬件构成.

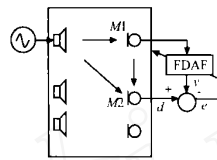


图 2 自适应 DIR 测量

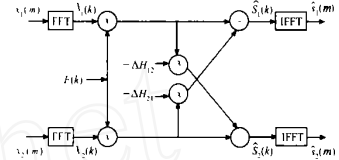


图 3 二阶 BSS 结构

图 4 给出了该 ES 的虚拟仪器 LabVIEW 图形用户界面. 该控制观察虚拟面板包括控制与指示两部分. 控制部分涉及到 FDAF 状态及参数的选择和调整. 例如, 自适应步长大小调整, 滤波器复位选择, 输入信号归一化选择及 FDAF 自适应梯度限制和非限制选择等. 此外, 在运行 ES 之前, 在这个控制面板上, 还可以根据需要进行输入采样率的大小, 快速傅立叶变换的长度, 声学路径及辨识 DIR 的 FDAF 长度等. 至于指示部分, 各种测试结果, 例如, DIR 及其相应的频谱图, FDAF 学习误差及其相应的频谱图, FDAF 的学习曲线等, 在测量期间均能实时地显示在虚拟的示波器屏幕上. BSS 的各种控制按钮也被设计来方便实验试听. 借助控制面板上的两个专用按钮和一个专用试听音响, 我们可以实时试听比较分离前的混合信号 $x_1(t)$ (或 $x_2(t)$) 和分离后的源独立信号的估计输出 $\hat{s}_1(t)$ (或 $\hat{s}_2(t)$). 该 ES 的另外一个重要功能就是能将 h_{12} (或 h_{21}) 的估计结果存成文件以便于离线后的算法研究.

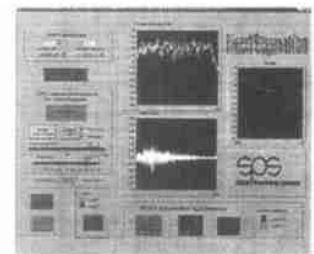


图 4 ES 的控制与观察图形用户界面

从公式 (4) 可以看出, DIR 与 RIR 的逆有关, 而多数情况下房间冲激响应是非最小相位的^[5]. 考虑到一个非最小相位系统的稳定的逆必定是非因果的, 我们引入一个时间延迟到 DIR 的实时测量及相应的 BSS 结构中. 该时间延迟在控制面板上大小可调, 便于实时测量及盲分离中寻求更好的效果.

4 实验结果

一个长、宽、高分别为 4.8 米、4.8 米及 2.8 米的房间被用来进行实时 BSS 实验. 实验室墙面上挂有各种面板, 落地式窗帘正反两面也具有不同的质地. 面板位置及窗帘两面均可调节以改变房屋的声学特性. 为了解决非因果性问题, 在测量 h_{12} 时, 一个时间延迟 被用来产生 $h_{12}(t - \tau)$. 同理, 在测量 h_{21} 时, 被引入来产生 $h_{21}(t - \tau)$. 在 BSS 实验中, 盲信号分离增益 (SG) 被定义如下

$$SG = (SG_1 + SG_2) / 2 \quad (10)$$

这里, $SG_i = 10 \log \left(\frac{P_{s_i} / P_{x_i}}{P_{s_j} / P_{x_j}} \right)$, $i, j, i, j = 1, 2$. P_{x_i} , P_{x_j} , P_{s_i} 及 P_{s_j} 分别表示 x_i , x_j , s_i 及 s_j 的功率. 一个测量均方误差为

$$MSE = (MSE_1 + MSE_2) / 2 \quad (11)$$

SG_1 是在 s_1 为白噪声, $s_2 = 0$ 的情况下被计算出来的, 对 SG_2 , 反之亦然.

图 5 画出了在不同的实验设置下, 分离增益 SG 及其对应的 DIR 测量均方误差 MSE 随着两个麦克风 $M1$ 和 $M2$ 间距离 w 的增大而变化的曲线图. 这里, d 表示两个声源 $L1$ 和 $L2$ 到两个麦克风 $M1$ 和 $M2$ 连线的垂直距离. T 表示时间延迟对应的采样点数. 在这个实验设置中, 我们将两个声源的距的两侧. 比较图 5(a) 和 (b), 离固定为 90cm 宽, 将两个麦克风 $M1$ 及 $M2$ 以等距离 ($w/2$) 分别置放于两个声源 $L1$ 及 $L2$ 中轴线以及 (c) 和 (d), 可以得知, 均方误差 MSE 越小, 分离增益 SG 越高. 理由显而易见, MSE 愈小就意味着 DIR 学习得愈好. 从这些曲线还可以得知, 随着时间延迟的引入, MSE 及 SG 均得到显著的改善. 图 6 画出了一组当 $w = 100\text{cm}$ 时, DIR 的测量结果 $h_{12} |_{T=0}$ 和 $h_{12} |_{T=400}$. 比较图 6(a) 和 (c), 以及 (b) 和 (d), 我们能够判断, 随着 d 的增加, RIR 的非最小相位变得越来越严重. 在这种情况下, 它的逆系统的非因果部分对 DIR 测量精度的提高起着举足轻重的作用. 这一事

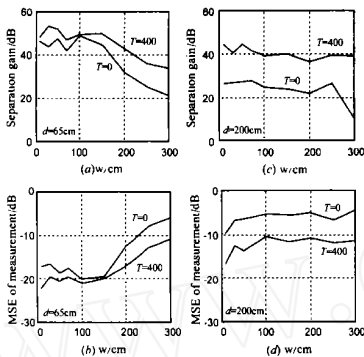


图 5 SG 及相应的 DIR 测量均方误差 MSE 与两个麦克风 $M1$ 及 $M2$ 间距离 w 的关系曲线. 设采样率为 8kHz, FFT 长度为 4096, 输入输出数据块长度为 2048, DIR 测量长度为 2048.

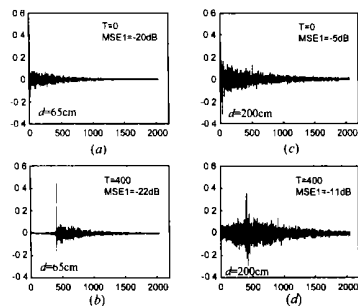


图 6 在 $w = 100\text{cm}$ 时的 $h_{12} |_{T=0}$ 及 $h_{12} |_{T=400}$

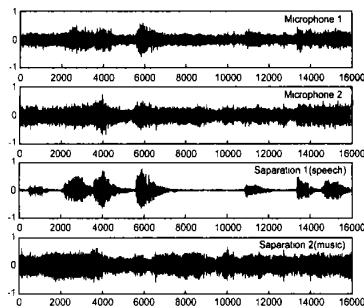


图 7 混合输入和分离输出

实也能清楚地从图 5 得到证实. 可以看出, 在 $d = 200\text{cm}$ 的情况下, 引入 400 点 ($T = 400$) 的时间延迟后, 分离增益 SG 有大约 15dB 的改善. 从图 6(a) 及 (b) 还可以看出, 对这个特定的实验房屋, 较好的分离增益主要是在小于 150cm 这样一个范围内取得.

图 7 给出了一组 $w = 30\text{cm}$, $d = 65\text{cm}$ 的实时 BSS 实验结果. 从麦克风拾取的两路混合信号可以看出, 音乐信号 (music) 已严重地淹没了语音信号 (speech). 作为实时 BSS 分离结果, 语音信号已清楚地从混合信号中分离出来. 同时, 分离出的实时语音及音乐能从专用试听音响清楚地听到.

5 结论

提出的简化混合模型 SMM 用于语音盲信号的分离, 经实时实验系统证实是有效可行的. 特别是当两个麦克风的距离靠得较近时, SMM 展示出潜在的实际应用可行性. 此外, 时间延迟的引入解决了非最小相位系统的求逆问题. 实验结果确信了解决如此逆问题的可行方法. 进一步地, 该实时 DSP 实验系统的建立无疑将加速 BSS 从理论研究到实际应用的过渡.

参考文献:

- [1] D W E Schobben, et al. Transparent communication [A]. Proc. of 1998 IEEE Benelux Signal Processing Chapter Symposium [C]. 1998. 171 - 174.
- [2] E Weinstein, et al. Multi-channel signal separation by decorrelation [J]. IEEE Trans. Speech and Audio Processing, 1993, 4(1): 405 - 413.
- [3] L Parra, et al. Convolutional blind source separation based on multiple decorrelation [A]. in Proc. of NNSP98 [C]. Cambridge, UK: NNSP, 1998.
- [4] Simon Haykin. Adaptive Filter Theory [M]. New York: Prentice-Hall, Inc., 1996. 445 - 458.
- [5] T-W Lee, et al. Blind source separation of real world signals [A]. Proc. of ICNN 97 [C]. Houston, USA: ICNN, 1997. 2129 - 2134.

作者简介:



何培宇 女, 1963 年生于四川茂县, 1986 年毕业于清华大学无线电系, 获工学学士学位; 1989 年毕业于四川大学无线电系并获理学硕士学位, 2000 年 4 月至 2001 年 4 月在荷兰 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在通信中的应用, 如自适应均衡器、全数字化高速 Modem 的 DSP 实现、盲信号分离及声回波抵消等.

的 DSP 实现、盲信号分离及声回波抵消等.