

基于盲波束形成的麦克风阵列语音增强方法

王冬霞 殷福亮 金乃高
(大连理工大学电子与信息工程学院 大连 116024)

摘要: 考虑到语音信号方向向量估计误差对传统波束形成语音增强性能的影响, 该文提出一种盲波束形成语音增强方法。由于采用阵列四阶互累积量和线性约束最小方差波束形成器相结合, 使得该方法对语音信号方向向量误差具有一定韧性。此外, 采用多通道后置滤波去除盲波束形成器输出端的残留噪声。仿真结果表明, 在语音信号波达方向等先验信息未知的情况下, 该文提出的盲波束形成语音增强方法仍具有较好的噪声抑制性能。

关键词: 麦克风阵列; 语音增强; 盲波束形成; 四阶互累积量

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)10-2321-04

Blind Beamforming for Microphone Array Based Speech Enhancement

Wang Dong-xia Yin Fu-liang Jin Nai-gao

(School of Electronics and Information Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Considering the bad effect of steering vector errors on speech enhancement performance, a new blind beamforming based speech enhancement method is introduced in this paper. The blind beamforming combines fourth-order cross cumulant and Linear Constrained Minimum Variance (LCMV) beamformer, which makes it robust to the steering vector. A multi-channel post-processor is then used to suppress the residual nonstationary noise. Simulations results demonstrate that the proposed method exhibits good noise suppression performance even with the unknown prior knowledge about direction of arrive with respect to speech signal.

Key words: Microphone array; Speech enhancement; Blind beamforming; Fourth-order cross cumulant

1 引言

在一些非手持式智能语音通信系统中, 由于麦克风与说话人距离较远, 麦克风接收的信号除语音直达波外, 还包括其反射波以及环境噪声。所有这些使得麦克风接收的语音信号质量较差。因此, 需要进行有效的噪声抑制, 以增强带噪语音信号质量。

麦克风阵列取代单麦克风, 以其独特的空间滤波特性, 得到了国内外学者的广泛关注, 并逐渐成为语音增强方法的研究热点^[1-4]。然而, 大多数的麦克风阵列语音增强方法基于传统波束形成, 即依赖于期望信号波达方向(DOA)和阵列拓扑结构等先验知识。但是, 在一些复杂声学环境中, DOA的准确估计具有一定困难。在这种情况下, 基于传统波束形成的语音增强方法, 其噪声抑制性能会有所退化。而盲波束形成方法可以在期望信号 DOA 等先验信息未知的情况下, 仅根据各个阵元的测量数据, 利用信号的统计独立特性或恒模特性, 波束形成重构出期望信号^[5-7]。

目前, 基于盲波束形成的麦克风阵列语音增强方法研究较少。日本学者Asano提出一种基于四阶累积量的麦克风阵列语音增强方法, 并取得较好的噪声抑制效果^[8]。但是, 该方法采用的是瞬时混合声学模型, 且仅适用于语音中的长平

稳部分(如元音)。而实际非手持环境下(室外或室内)的声学模型通常表现为卷积混合形式, 且其中的语音信号也不仅仅包括元音成分。

为此, 本文在文献[5]的基础上, 提出一种基于盲波束形成的麦克风阵列语音增强方法。该方法采用阵列四阶互累积量和线性约束最小方差波束形成器(LCMV)相结合, 在语音信号波达方向(DOA)及阵列拓扑结构等先验信息未知情况下, 抑制实际声学环境下的高斯干扰噪声。

2 声学模型

令 $s(n)$ 表示语音信号, $i_j(n)$ 表示方向干扰噪声, $h_{ks}(n)$ 和 $h_{kj}(n)$ 分别表示语音源和第 j 个干扰源与第 k 个麦克风之间的声学通道传递函数, 则第 k 个麦克风接收信号可以表示为

$$x_k(n) = s(n) * h_{ks}(n) + \sum_{j=1}^J i_j(n) * h_{kj}(n) + v_k(n) \quad (1)$$

式中“*”表示线性卷积, $v_k(n)$ 表示第 k 个麦克风接收的背景噪声。为简化分析又不失一般性, 令声学通道转移函数为

$$h_{kp}(n) = \alpha_{kp} \delta(n - \tau_{kp}(\theta_p)) \quad (2)$$

式中 α_{kp} 表示声源传播的幅度衰减因子, $\tau_{kp}(\theta_p)$ 表示第 k 个麦克风接收第 p 个声源信号的时延, 它是声源波达方向 θ_p 为变量的函数。将式(2)代入式(1), 得到

2006-03-14 收到, 2006-08-07 改回

国家自然科学基金(60372082/60172073)和教育部跨世纪优秀人才基金资助课题

$$x_k(n) = a_{k_s}s(n - \tau_{k_s}(\theta_s)) + \sum_{j=1}^J a_{k_j}i_j(n - \tau_{k_j}(\theta_j)) + v_k(n) \quad (3)$$

式(3)为简单的卷积声学模型(如室外录音情况),区别于文献[8]中的瞬时声学模型。本文目的是,在式(3)所选的卷积声学模型且语音信号波达方向未知情况下,尽可能地抑制干扰噪声信号,恢复出纯净语音信号。

3 盲波束形成语音增强方法

本文提出的盲波束形成语音增强方法系统结构如图1所示。图中STFT表示阵列信号的短时傅里叶变换,ISTFT表示对应的逆变换。下面,对该盲波束形成语音增强方法进行详细阐述。

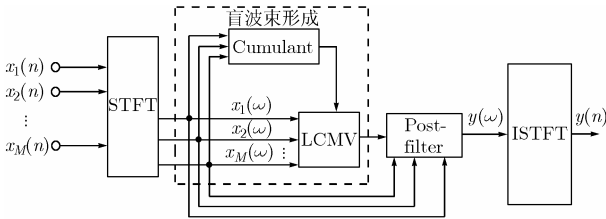


图1 基于盲波束形成的麦克风阵列语音增强方法原理框图

3.1 语音信号的广义方向矢量

设 $S(\omega, m)$, $I_j(\omega, m)$ 和 $V_k(\omega, m)$ 分别表示 $s(n)$, $i_j(n)$ 和 $v_k(n)$ 所对应的第 m 帧信号短时谱, 对式(3)进行短时傅里叶变换, 得

$$\begin{aligned} X_k(\omega, m) &= a_k e^{-i\omega\tau_{k_s}(\theta_s)} S(\omega, m) + \sum_{j=1}^J b_{k_j} e^{-i\omega\tau_{k_j}(\theta_j)} I_j(\omega, m) \\ &\quad + V_k(\omega, m) \\ &= \alpha_k(\omega, \theta_s) S(\omega, m) + \sum_{j=1}^J \beta_{k_j}(\omega, \theta_j) I_j(\omega, m) \\ &\quad + V_k(\omega, m) \end{aligned} \quad (4)$$

式中 $\alpha_k(\omega, \theta_s)$ 与 $\beta_{k_j}(\omega, \theta_j)$ 分别为第 k 个麦克风接收的语音信号方向响应与干扰噪声方向响应。于是, 阵列接收信号用矩阵形式表示为

$$\begin{bmatrix} X_1(\omega, m) \\ \vdots \\ X_M(\omega, m) \end{bmatrix} = \begin{bmatrix} \alpha_1(\omega, \theta_s) \\ \vdots \\ \alpha_M(\omega, \theta_s) \end{bmatrix} S(\omega, m) + \begin{bmatrix} \sum_{j=1}^J \beta_{1,j}(\omega, \theta_j) I_j(\omega, m) \\ \vdots \\ \sum_{j=1}^J \beta_{M,j}(\omega, \theta_j) I_j(\omega, m) \end{bmatrix} + \begin{bmatrix} V_1(\omega, m) \\ \vdots \\ V_M(\omega, m) \end{bmatrix} \quad (5)$$

即

$$\begin{aligned} \mathbf{X}(\omega, m) &= \boldsymbol{\alpha}(\omega, \theta_s) S(\omega, m) + \boldsymbol{\beta}(\omega, \theta_j) \mathbf{I}_j(\omega, m) \\ &\quad + \mathbf{V}(\omega, m) \end{aligned} \quad (6)$$

式中 $\mathbf{X}(\omega, m)$ 表示阵列接收信号矢量, $\boldsymbol{\alpha}(\omega, \theta_s)$ 和 $\boldsymbol{\beta}(\omega, \theta_j)$

分别表示阵列接收信号中的语音信号方向矢量和干扰噪声方向矢量。假设语音信号与高斯干扰噪声及背景噪声统计独立, 并考虑各频率点处高斯信号高阶累积量为零的特性, 构建四阶零延时互累积量为

$$\begin{aligned} C_k(\omega, m) &= \text{cum} \{ X_1(\omega, m), X_1^*(\omega, m), X_1^*(\omega, m), X_k(\omega, m) \} \\ &= \text{cum} \left\{ \alpha_1(\omega, \theta_s) S(\omega, m) + \sum_{j=1}^J \beta_{1,j}(\omega, \theta_j) I_j(\omega, m) \right. \\ &\quad + V_1(\omega, m), \alpha_1^*(\omega, \theta_s) S^*(\omega, m) + \sum_{j=1}^J \beta_{1,j}^*(\omega, \theta_j) I_j^*(\omega, m) \\ &\quad + V_1^*(\omega, m), \alpha_1^*(\omega, \theta_s) S^*(\omega, m) + \sum_{j=1}^J \beta_{1,j}^*(\omega, \theta_j) I_j^*(\omega, m) \\ &\quad + V_1^*(\omega, m), \alpha_k(\omega, \theta_s) S(\omega, m) + \sum_{j=1}^J \beta_{k,j}(\omega, \theta_j) I_j(\omega, m) \\ &\quad \left. + V_k(\omega, m) \right\} \\ &= |\alpha_1(\omega, \theta_s)|^2 \alpha_1^*(\omega, \theta_s) \gamma_{4s} \alpha_k(\omega, \theta_s) = \zeta \alpha_k(\omega, \theta_s) \end{aligned} \quad (7)$$

式中“*”表示共轭运算符, “cum”表示累积量操作符, 标量因子 $\zeta = |\alpha_1(\omega, \theta_s)|^2 \alpha_1^*(\omega, \theta_s) \gamma_{4s}$, 其中 γ_{4s} 为语音信号的零延时四阶互累积量。于是, 阵列累积量矢量表达式为

$$\mathbf{C}(\omega, m) = \zeta [\alpha_1(\omega, \theta_s), \dots, \alpha_M(\omega, \theta_s)]^T = \zeta \boldsymbol{\alpha}(\omega, \theta_s) \quad (8)$$

式中“T”表示转置操作符。

由式(8)可以看到, $\mathbf{C}(\omega, m)$ 与 $\boldsymbol{\alpha}(\omega, \theta_s)$ 仅相差一个标量因子, 即与语音信号方向矢量在方向上一致^[5]。也就是说, 在语音信号方向矢量未知的情况下, 通过计算阵列信号的四阶互累积量便可以得到语音信号方向矢量。与语音信号方向矢量 $\boldsymbol{\alpha}(\omega, \theta_s)$ 相对应, 这里称 $\mathbf{C}(\omega, m)$ 为语音信号的广义方向矢量。

3.2 盲波束形成

与文献[5]中采用的最小方差无失真响应(MVDR)波束形成器相比, 线性约束最小方差波束形成器(LCMV)是常用的传统波束形成噪声抑制方法。它在保持语音信号波达方向信号功率不变的情况下, 使阵列输出信号的方差最小。其频率响应为^[1]

$$\mathbf{w}(\omega, m) = g \frac{\mathbf{R}_x^{-1}(\omega, m) \boldsymbol{\alpha}(\omega, m)}{\boldsymbol{\alpha}^H(\omega, m) \mathbf{R}_x^{-1}(\omega, m) \boldsymbol{\alpha}(\omega, m)} \quad (9)$$

式中 g 为 Lagrange 乘子, $\mathbf{R}_x(\omega, m)$ 为阵列输入信号矢量 $\mathbf{X}(\omega, m)$ 的相关矩阵, 上标“H”表示共轭转置操作符。上标“-1”表示矩阵求逆运算操作符。根据式(8), 用 $\mathbf{C}(\omega, m)$ 取代式(9)中的 $\boldsymbol{\alpha}(\omega, \theta_s)$, 得到盲波束形成器的频率响应为

$$\mathbf{w}(\omega, m) = g \frac{\mathbf{R}_x^{-1}(\omega, m) \mathbf{C}(\omega, m)}{\mathbf{C}^H(\omega, m) \mathbf{R}_x^{-1}(\omega, m) \mathbf{C}(\omega, m)} \quad (10)$$

其中相关矩阵 $\mathbf{R}_x(\omega, m)$ 采用平滑方式得到, 即

$$\mathbf{R}_x(\omega, m) = (1 - \gamma) \mathbf{R}_x(\omega, m - 1) + \gamma \mathbf{X}(\omega, m) \mathbf{X}^H(\omega, m) \quad (11)$$

式中 γ 为遗忘因子。

于是, 盲波束形成器输出信号 $Z(\omega, m)$ 为

$$Z(\omega, m) = \mathbf{w}^H(\omega, m)\mathbf{X}(\omega, m) \quad (12)$$

3.3 后置滤波

为进一步抑制盲波束形成输出端的残留噪声, 采用多通道后置滤波器对其进行后置处理。多通道后置滤波器的频率响应表达式为^[2]

$$H(\omega, m) = \frac{2}{(M-1)(M-2)} \sum_{i=0}^{M-3} \sum_{j=i+1}^{M-2} P_{x_i x_j}(\omega, m) / P_Z(\omega, m) \quad (13)$$

式中 $P_{x_i x_j}(\omega)$ 及 $P_Z(\omega)$ 表示互功率谱与自功率谱。为了更好地抑制噪声, 分别采用平滑方式得到, 即

$$\left. \begin{aligned} P_{x_i x_j}(\omega, m) &= (1 - \beta_1)P_{x_i x_j}(\omega, m-1) + \beta_1 X_i(\omega, m)X_j^*(\omega, m) \\ P_Z(\omega, m) &= (1 - \beta_2)P_Z(\omega, m-1) + \beta_2 |Z(\omega, m)|^2 \end{aligned} \right\} \quad (14)$$

式中 β_1, β_2 为遗忘因子 ($0 < \beta_1, \beta_2 < 1$)。于是, 后置滤波器输出端的语音为

$$Y(\omega, m) = H(\omega, m)Z(\omega, m) \quad (15)$$

最后, 整个语音增强系统的输出信号为

$$y(n) = \mathbf{F}^{-1} \left\{ \sum_m Y(\omega, m) \right\} \quad (16)$$

式中 “ \mathbf{F}^{-1} ” 表示反傅里叶变换操作符。

4 仿真实验与结果

4.1 实验环境

选取间距为 5cm 的 4 个麦克风组成线性麦克风阵列。语音源入射方向为 60° , 干扰噪声源入射方向为 30° 。语音文本“第一课认识新同学”由男生朗读, 信号采样率为 16kHz, 每个样值用 16 位二进制进行编码。实验选取帧长为 256, 盲波束形成算法中所有遗忘因子取值 0.7。在语音文本中加入 Babble 干扰噪声或随机产生的零均值高斯背景噪声^[9], 并参照式(3)组成不同输入信噪比的麦克风接收信号。

为验证本文方法的噪声抑制性能, 实验过程中分别选取了 Asano 方法(方法 1)^[8]及基于传统波束形成的麦克风阵列语音增强方法(方法 2)^[4]与其做性能比较。

4.2 实验结果

图 2 与图 3 分别显示 0dB 方向干扰噪声及非方向噪声环境下, 各语音信号的波形。从图中看到, 在语音信号方向向量未知情况下, 本文方法在两种噪声类型的声学环境下, 均取得了较好的噪声抑制效果。

图 4 显示采用不同声学模型的两麦克风阵列语音增强方法噪声抑制性能比较。可以看出, 在语音信号 DOA 等先验知识未知的情况下, 两种麦克风阵列语音增强方法均具有一定的噪声抑制能力。比较两种方法的噪声抑制曲线可以看到, 本文方法的噪声抑制性能优于 Asano 方法(方法 1)^[11]。其原因是, 一方面 Asano 方法采用的瞬时声学模型并不适合实际声学环境, 另一方面本文对盲波束形成输出端的残留噪声进行了后置处理。

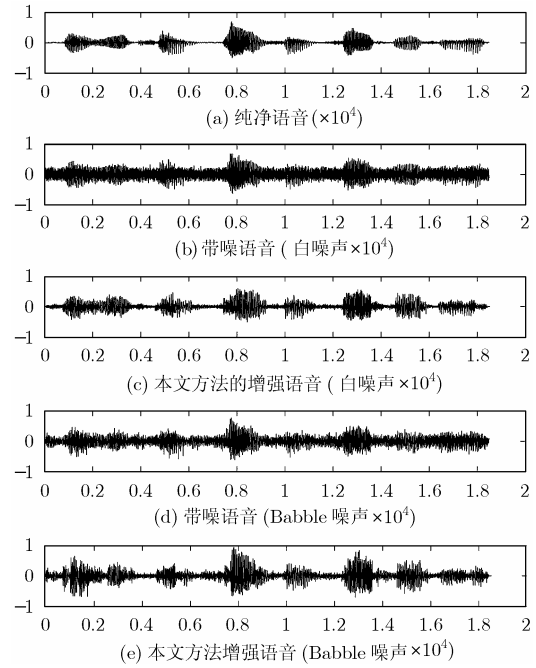


图 2 方向干扰噪声环境下, 各语音信号波形

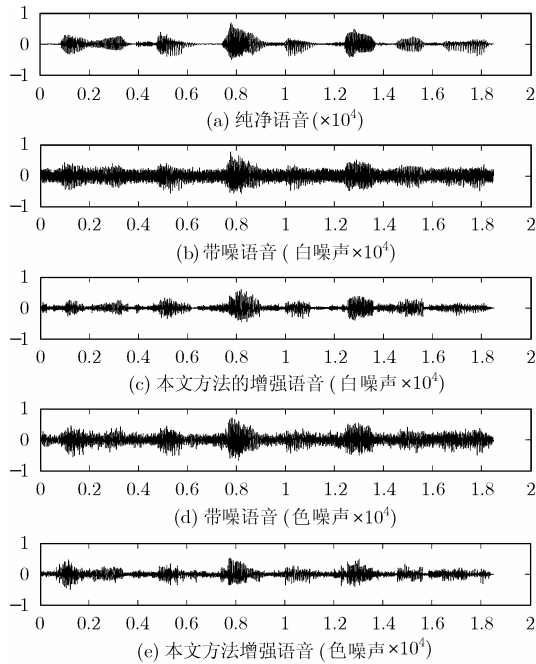


图 3 非方向背景噪声环境下, 各语音信号波形

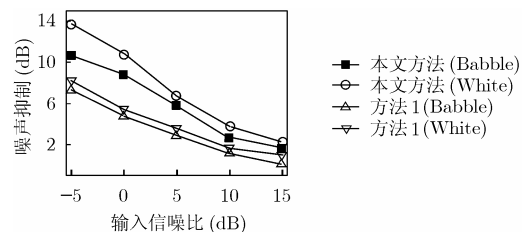


图 4 DOA 等先验知识未知情况下, 各方法的噪声抑制性能曲线

当不存在语音信号方向向量误差情况下, 基于两种不同

波束形成麦克风阵列语音增强方法的噪声抑制性能比较如图 5 所示。可以看到, 在所选声学模型情况下, 基于传统波束形成的语音增强方法(方法 2)^[4], 其噪声抑制性能好于盲波束形成语音增强方法, 而且白噪声情况明显好于色噪声环境(Babble)。

图 6 显示当语音信号方向矢量存在误差情况下, 基于不同波束形成麦克风阵列语音增强方法的噪声抑制性能。可以看到在两种噪声类型情况下, 本文盲波束形成语音增强方法的噪声抑制性能强于基于传统波束形成的麦克风阵列语音增强方法(方法 2), 且随着输入信噪比的降低而增强。

比较图 5 与图 6, 基于不同波束形成的麦克风语音增强方法均有各自特点。然而, 在所选声学模型情况下, 当语音信号方向向量估计误差较大时, 盲波束形成语音增强方法优于基于传统波束形成的麦克风阵列语音增强方法。

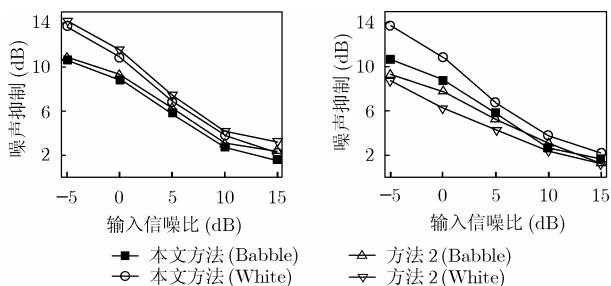


图 5 无方向误差存在情况下, 各方法的噪声抑制性能曲线

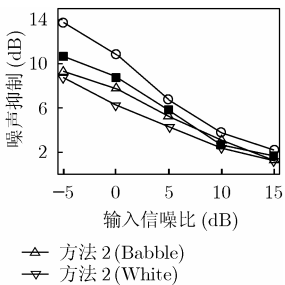


图 6 方向误差存在情况下(3°), 各方法的噪声抑制性能曲线

图 7 显示输入信噪比为 0dB 时, 各语音增强方法随语音方向估计误差的不同, 其噪声抑制能力的变化情况。可以看到, 由于不需要语音方向矢量信息, 基于盲波束形成的语音增强方法, 其噪声抑制因子几乎保持不变。而基于传统波束形成的语音增强方法, 则随着方向估计误差的增加, 其噪声抑制能力降低。同时, 由于仿真实验所选采样率的原因, 传统波束形成方法在一些不同方向估计误差处, 其噪声抑制几乎相同。与传统波束形成方法相比, 语音方向矢量误差越大, 本文方法的噪声抑制性能越明显。

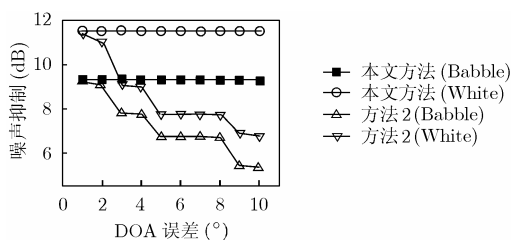


图 7 不同方向误差存在情况下(1°-10°), 各方法的噪声抑制性能曲线

5 结束语

考虑到复杂声场环境下, 语音信号方向向量估计误差对传统波束形成语音增强性能的影响, 本文提出一种基于盲波束形成的麦克风阵列语音增强方法。该方法采用基于四阶互累积量的盲波束形成器和多通道后置滤波器结合起来抑制高斯干扰噪声。仿真实验结果表明, 本文方法在语音信号波达方向等先验信息未知的情况下, 仍具有较好的噪声抑制性能。

参考文献

- [1] Griffiths L J and Jim C W. An alternative approach to linearly constrained adaptive beamforming [J]. *IEEE Trans. on Antennas and Propagation*, 1981, 30(1): 27-34.
- [2] Zelinski R. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms [A]. *IEEE International Conference on Acoustics, Speech and Signal Processing* [C], New York, 1988, 5: 2578-2581.
- [3] Gannot S and Cohen I. Speech enhancement based on the general transfer function GSC and postfiltering [J]. *IEEE Trans. on Speech and Audio Processing*, 2004, 12(6): 561-571.
- [4] Wang Dongxia and Yin Fuliang. A subband adaptive learning algorithm for microphone array based speech enhancement [A]. *Lecture Notes in Computer Science* [C], Springer-Verlag GmbH, 2005, 3497: 592-597.
- [5] Dogan C M and Mendel M J. Cumulant-based blind optimum beamforming [J]. *IEEE Trans. on Aerospace and Electronic Systems*, 1994, 30(3): 722-740.
- [6] Cardoso F J and Souloumiac A. Blind beamforming for non-Gaussian signals [J]. *IEE Proceeding on Radar and Signal Processing*, 1993, 140(6): 362-370.
- [7] Gonen E and Mendel M J. Applications of cumulants to array processing — Part III: Blind beamforming for coherent signals [J]. *IEEE Trans. on Signal Processing*, 1997, 45(9): 2252-2264.
- [8] Asano F. Speech enhancement using array signal processing with 4th-order cumulant [A]. *Proceedings of Fourth International Conference on Signal Processing* [C], Beijing, 1998, 1: 415-418.
- [9] 殷福亮, 宋爱军. 数字信号处理C语言程序集[M]. 沈阳: 辽宁科学技术出版社, 1997: 30-34.

王冬霞: 女, 1976 年生, 博士生, 讲师, 研究方向为语音信号处理和阵列信号处理的理论与应用。

殷福亮: 男, 1962 年生, 教授, 博士生导师, 主要研究方向为数字信号处理、语音处理、图像处理、现代通信技术。

金乃高: 男, 1977 年生, 博士生, 研究方向为语音信号处理、图像处理、信息融合。