第 36 卷第 4 期　　声　学　学　报　　Vol. 36, No. 4
2011 年 7 月　　ACTA　ACUSTICA　　Jul., 2011

联合宽带到达方向估计和语音特征增强的传声器阵处理方法*

李晓雪　徐　文　金丽玲　李建龙
(浙江大学信电系　杭州　310027)

2009 年 12 月 31 日收到
2010 年 3 月 15 日定稿

摘要　以提高室内混响环境下自动语音识别 (ASR) 性能为目标，讨论了一种小尺寸传声器阵处理方法。该方法采用基于旋转不变技术的信号参数估计算法 (ESPRIT) 计算宽带语音信号到达方向角，进行时延补偿；同时联合考虑阵列滤波与隐马尔可夫模型 (HMM) 识别过程，将识别输出结果反馈到前端的传声器阵处理，优化阵列滤波系数。与常规阵处理方法改善信号波形质量不同，本文通过调节阵列滤波系数降低待识别特征与训练模型之间的失配，直接提高识别过程中正确假设的概率。实验结果表明，上述方案能够有效降低会议室环境下孤立词有限词库 ASR 的错误概率，表现优于常规波束形成方法；采用全局优化进行阵列滤波设计，与局部优化算法相比，进一步改善了处理性能。

PACS 数：43.70

Microphone array processing via joint wideband angle-of-arrival estimation and speech feature enhancement

LI Xiaoxue　XU Wen　JIN Liling　LI Jianlong
(*Department of Information Science and Electronic Engineering, Zhejiang University*　Hangzhou　310027)

Received Dec. 31, 2009
Revised Mar. 15, 2010

Abstract　This paper concerns techniques of speech processing using a small-size microphone array to improve automatic speech recognition (ASR) performance in an indoor reverberant environment. The method first applies Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) to compute directions-of-arrivals of wideband speech signals and implements time-delay compensation accordingly; array signal filtering is then considered jointly with the HMM-based speech recognition procedure, whose outputs are fed back to the front end to optimize array filtering design. Different from conventional array processing aiming to enhancing signal waveform, the approach here adjusts array filtering coefficients to reduce mismatch between the features to be recognized and the training model, thus directly maximizing the likelihood of the right transcription for a selected vocabulary. Experimental data processing shows that the above approach can effectively reduce ASR error rate for an isolated word vocabulary of finite size in a meeting-room environment, superior to conventional beamforming processing; compared to a local optimization scheme, applying global optimization in array filtering design further improves the performance.

引言

当今的自动语音识别 (ASR: Automatic Speech Recognition) 技术，对纯净的语音信号已经可以达到较高的识别精度。但是在实际的工作环境下，环境噪声与混响的存在，以及其他声源的干扰，造成训练模型与待识别特征之间的失配，使得系统识别性能急

* 微软亚洲研究院大学计划项目资助

剧下降。减小甚至消除这种失配的影响，是 ASR 技术实用化必须解决的关键问题之一。空间阵列滤波是目前普遍采用的应对环境干扰的空时信号提取方法，广泛应用于雷达、声呐、通信等研究领域 [1]，而传声器阵处理作为阵列信号处理技术的一种具体应用，能够对语音信号进行方向估计和空时滤波，将目标语音从环境干扰中分离出来，并加以增强，可提高混响环境下 ASR 性能 [2] 。

然而传声器阵处理有其独特的挑战性。首先在实际应用中 (如室内会议，语音聊天，家电控制等)，传声器阵的尺寸受到很大限制。为提高小孔径阵的分辨率，需采用高分辨率谱估计技术，常用的方法包括自适应波束形成算法，如最小方差无失真响应处理 [3−4](MVDR: Minimum Variance Distortionless Response)，和基于信号子空间分解的算法，如多信号分类算法 [5](MUSIC: Multiple Signal Classification)，基于旋转不变技术的信号参数估计算法 [6](ESPRIT: Estimation of Signal Parameters via Rotational Invariance Techniques) 等。其次，语音信号的有效频率可覆盖高达三倍频程的范围，为典型的宽带信号，且表现出短时平稳性。因此相应的处理必须在宽带条件下工作，并能应对数据样本较少、统计特性时变的情况。约束最小均方误差 (CLMS: Constrained least mean-squares) 波束形成处理 [7−8] 即为具备这些能力的一种较为流行的算法。

目前基于隐马尔可夫模型 (HMM) 的传声器阵语音识别系统，大多包括先后单独操作的阵列信号处理和特征识别两个模块。其中前端的阵处理模块主要是为了进行语音增强，目的是在提取语音参数之前，尽量减少信号波形的失真。该方案基于的假设是，对波形质量得到改善的信号进行特征识别能够提高识别性能。然而语音识别系统并不是直接理解语音波形，其工作原理在于从输入信号中提取特征矢量序列，在模板库中找出最有可能产生特征观察矢量序列的模板词汇作为识别结果输出。因此仅从改善信号波形的角度进行语音增强，难以达到优化最终识别效果的目的。

针对上述问题，本文从两个方面探讨传声器阵处理的新方法。首先，首次将基于 ESPRIT 的宽带到达角估计算法 [9] 应用到声源的方向估计中，该方法不但有高的角度分辨率，而且可直接估计出多个到达信号的方位角，与 CLMS 或 MUSIC 算法相比，避免了对整个角度域的扫描计算，从而有效地降低运算量。其次，受文献 10 的启示，采用一种阵列滤波设计方法，其目的并不是为了改善信号波形质量，而是在于直接提高识别过程中正确假设的概率，从而提高系统的正确识别率。与文献 10 不同，就概率最大化这一目的，本文分别采用了局部优化和全局优化两种方案来进行滤波器系数设计。为验证所提算法的有效性，还采用会议室环境实验记录数据对各有关算法的性能进行了测试与比对。

1　传声器阵处理算法结构

一种带反馈的基于 HMM 的传声器阵有限词库语音识别系统结构如图 1 所示，语音信号先由传声器阵接收，经过阵列处理，产生增强的语音信号输送至识别器计算特征矢量，并利用 HMM 进行内容识

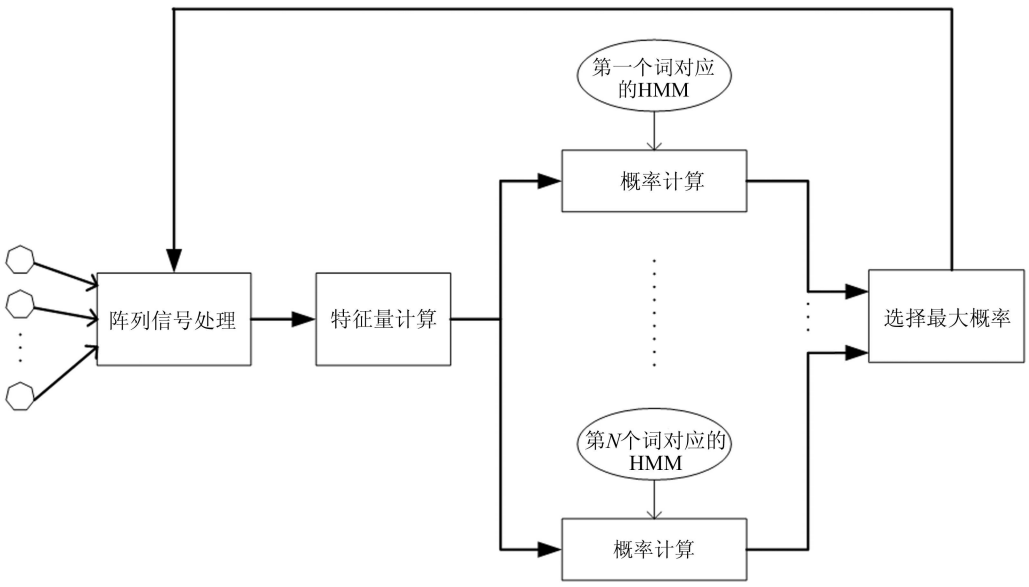

图 1　结合识别过程进行阵处理的有限词库语音识别系统结构

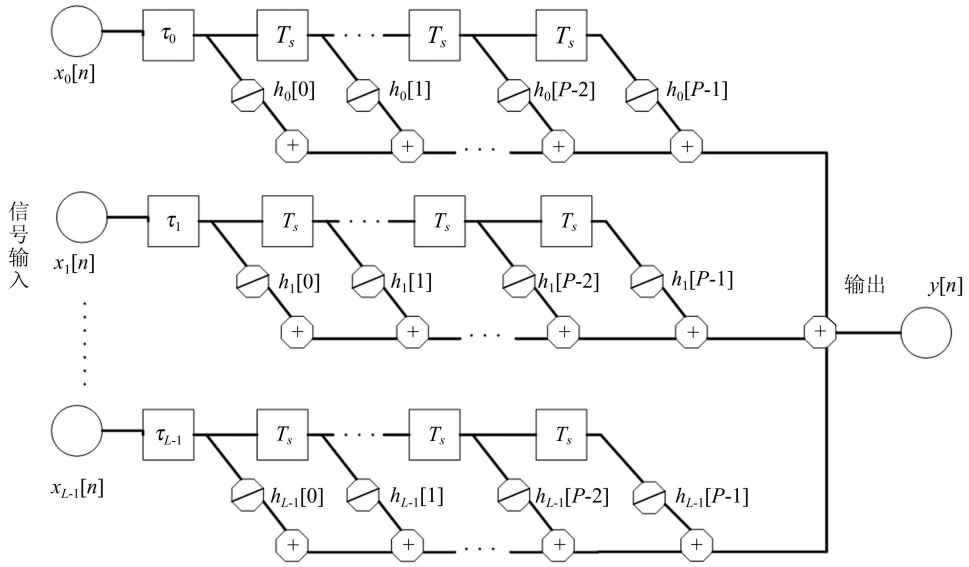

图 2 实现自适应宽带波束形成的多通道 FIR 滤波器结构

别; 训练数据的识别输出结果反馈至前端, 用于优化阵列处理参数设计。这种方法将阵列处理和识别过程结合起来, 通过训练数据的反馈, 将识别系统的统计模型考虑到阵处理中, 是一种根据期望假设最大化而非期望信号最优化的自适应处理方法, 以增强对于识别更为重要的信号分量。

阵列信号处理采用类似于 CLMS 的结构, 包括时延补偿和一组有限脉冲响应 (FIR: Finite Impulse Response) 滤波, 如图 2 所示。时延补偿的目的是为了增强特定说话人方向的信号, 与 CLMS 对方向角度进行扫描不同, 本文采用 ESPRIT 直接估计语音信号的到达方向, 并据此计算相应的各传声器阵元延时, 对接收信号进行时间对齐。滤波器组对多通道传声器接收信号进行适应其统计特性的频域滤波, 如前所述, 本文中滤波器系数设计是通过最大化信号被正确识别的概率来获取的。

图 2 中, 滤波器组由滤波阶数为 P 的 L 个通道构成, 其中 $x_l[n]$ 是第 l 个传声器的接收信号, τ_l 是第 l 个传声器通道的延时补偿量, 由第 2 节中的方向估计算法求出, T_s 是采样周期, $h_l[p]$ 是第 l 个传声器通道第 p 阶的加权系数, $y[n]$ 是最终的滤波输出。

2 基于 ESPRIT 的宽带信号到达方向角估计

对于小孔径传声器阵, 假定声源位于远场区, 抵达接收阵的是平面波, 同时受到加性噪声的干扰。将一个含 L 个阵元的均匀线阵分成两个子阵, 子阵 1 由前 $L-1$ 个阵元组成, 子阵 2 由后 L–1 个阵元组成, 用 $\boldsymbol{X}_1(z)$ 和 $\boldsymbol{X}_2(z)$ 分别表示两个子阵的接收信号 (z 变换域), $\boldsymbol{n}_1(z)$ 和 $\boldsymbol{n}_2(z)$ 代表加性噪声。假设有 d 个信号源 $S_j(z)$, 方向角用 θ_j 表示, $j=1,2,\cdots,d$(对于非相干信号源, $d\leqslant(L$–1); 对于相干信号源, $d\leqslant L/2$)。以 $a_i(z,\theta_j)$ 代表第 i 号阵元对 θ_j 方向入射信号的响应, $\varphi(z,\theta_j)$ 代表 θ_j 方向的入射信号在两子阵间的传播延时响应。ESPRIT 算法即为利用 $\varphi(z,\theta_j)$ 表征的两子阵间的旋转不变性发展而来的 [6]。

用 $\boldsymbol{a}(z,\theta_j)$ 表示子阵 1 对 θ_j 方向的入射信号的响应矢量, 即

$$\boldsymbol{a}(z,\theta_j)=[a_0(z,\theta_j),\cdots,a_{L-2}(z,\theta_j)]^{\mathrm{T}}. \tag{1}$$

各子阵的接收信号可按如下形式建模:

$$\boldsymbol{X}_1(z)=\boldsymbol{A}(z)\boldsymbol{S}(z)+\boldsymbol{n}_1(z), \tag{2}$$

$$\boldsymbol{X}_2(z)=\boldsymbol{A}(z)\boldsymbol{\Phi}(z)\boldsymbol{S}(z)+\boldsymbol{n}_2(z), \tag{3}$$

这里

$$\boldsymbol{S}(z)=[S_1(z),\cdots,S_d(z)]^{\mathrm{T}}, \tag{4}$$

$$\boldsymbol{\Phi}(z)=\mathrm{diag}[\phi(z,\theta_1),\cdots,\phi(z,\theta_d)], \tag{5}$$

$$\boldsymbol{A}(z)=[\boldsymbol{a}(z,\theta_1),\cdots,\boldsymbol{a}(z,\theta_d)], \tag{6}$$

其中 diag[·] 表示以括号中元素为对角线元素的对角阵。

令 $\boldsymbol{X}(z)=[\boldsymbol{X}_1^{\mathrm{T}}(z),\boldsymbol{X}_2^{\mathrm{T}}(z)]^{\mathrm{T}}$, 则有:

$$\boldsymbol{X}(z)=\begin{bmatrix}\boldsymbol{A}(z)\\ \boldsymbol{A}(z)\boldsymbol{\Phi}(z)\end{bmatrix}\boldsymbol{S}(z)+\begin{bmatrix}\boldsymbol{n}_1(z)\\ \boldsymbol{n}_2(z)\end{bmatrix}. \tag{7}$$

其谱密度为:

$$\boldsymbol{R}_x(z)=E\left\{\boldsymbol{X}(z)\boldsymbol{X}^{\mathrm{T}}(z^{-1})\right\}=\begin{bmatrix}\boldsymbol{A}(z)\\ \boldsymbol{A}(z)\boldsymbol{\Phi}(z)\end{bmatrix}\boldsymbol{P}(z)[\boldsymbol{A}^{\mathrm{T}}(z^{-1}),\boldsymbol{\Phi}^{\mathrm{T}}(z^{-1})\boldsymbol{A}^{\mathrm{T}}(z^{-1})]+\boldsymbol{N}_x(z), \tag{8}$$

其中 $\boldsymbol{N}_x(z)$ 为噪声谱密度， $\boldsymbol{P}(z)$ 为入射信号的谱密度：

$$\boldsymbol{P}(z)=E\left\{\boldsymbol{S}(z)\boldsymbol{S}^{\mathrm{T}}(z^{-1})\right\}. \tag{9}$$

广义信号子空间定义为矩阵 $\left[\begin{array}{c}\boldsymbol{A}(z)\\ \boldsymbol{A}(z)\boldsymbol{\Phi}(z)\end{array}\right]\boldsymbol{P}(z)$ 的列向量组张成的空间。

若入射信号互不相干， $\boldsymbol{P}(z)$ 和 $\boldsymbol{A}(z)$ 的秩均为 d, 则该信号子空间为 $2(L-1)$ 维有理向量空间的 d 维子空间。如果加性噪声是非相干噪声, 则噪声子空间构成信号子空间的正交补。因此不含加性噪声干扰的信号谱密度中, 对应于最大的 d 个非零特征值的特征向量可张成这一信号子空间。令 $[\boldsymbol{E}_1^{\mathrm{T}}(z),\boldsymbol{E}_2^{\mathrm{T}}(z)]^{\mathrm{T}}$ 代表对应的信号子空间特征向量组，存在满秩的 $d\times d$ 维矩阵 $\boldsymbol{T}(z)$ 满足：

$$\left[\begin{array}{c}\boldsymbol{E}_1(z)\\ \boldsymbol{E}_2(z)\end{array}\right]=\left[\begin{array}{c}\boldsymbol{A}(z)\\ \boldsymbol{A}(z)\boldsymbol{\Phi}(z)\end{array}\right]\boldsymbol{T}(z). \tag{10}$$

消除 $\boldsymbol{A}(z)$ 可得到如下结果：

$$\boldsymbol{E}_2(z)=\boldsymbol{E}_1(z)\boldsymbol{T}^{-1}(z)\boldsymbol{\Phi}(z)\boldsymbol{T}(z). \tag{11}$$

从 $\boldsymbol{E}_1(z)$ 到 $\boldsymbol{E}_2(z)$ 的线性变换矩阵表示为：

$$\boldsymbol{\Psi}(z)=\boldsymbol{T}^{-1}(z)\boldsymbol{\Phi}(z)\boldsymbol{T}(z), \tag{12}$$

分析其结构可得， $\boldsymbol{\Psi}(z)$ 的特征值即为对角阵 $\boldsymbol{\Phi}(z)$ 的对角元素 $\varphi(z,\theta_j)$, $j=1,\cdots,d$ 。

设信号空间有 M 个极点，加性噪声有 N 个极点，且互不重合，谱密度 $\boldsymbol{R}_x(z)$ 可用如下形式表示为：

$$\boldsymbol{R}_x(z)=\sum_{i=1}^{M+N}\left(\frac{\boldsymbol{R}_i}{1-p_iz^{-1}}+\frac{\boldsymbol{R}_i^*}{1-p_i^*z}\right)+\boldsymbol{W}(z), \tag{13}$$

其中 $\boldsymbol{R}_i$ 是极点 p_i 对应的留数矩阵， $\boldsymbol{W}(z)$ 是 FIR 谱密度矩阵。

对于窄带信号，选择 Z 域单位圆上的点 $z_j=e^{\mathrm{j}\omega_j}$, 则两子阵间的传播延时项 $\phi(z,\theta_j)=e^{\mathrm{j}\omega_jt_j}$, 其中 ω_j 为角频率， $t_j=\Delta\cdot\sin\theta_j/c$ 是信号在两子阵间的传播延时， Δ 是两子阵间位移矢量的长度。

对宽带信号，令 p_i 代表 $\boldsymbol{R}_x(z)$ 的极点，文献 11 中已经证明，第 j 个入射信号在阵元间的传播延时响应可表示成：

$$\phi(z,\theta_j)=p_i^{t_j}=p_i^{\Delta\cdot\sin\theta_j/c}. \tag{14}$$

对矩阵 $\boldsymbol{R}_x(z)$ 做关于极点的模态分解 [11], 可计算得到各极点 p_i 和相应的留数矩阵 $\boldsymbol{R}_i$ 。利用式 (10) 到式 (12) 的信号子空间分解可计算出对角阵 $\boldsymbol{\Phi}(z)$ 的对角元素 $\phi(z,\theta_j)$, $j=1,\cdots,d$, 进而可从式 (14) 中求出方向角的估计值。

3 语音识别概率最大化滤波

采用图 2 的结构对传声器阵接收信号进行处理，输出如式 (15) 所示：

$$y[n]=\sum_{l=0}^{L-1}\sum_{p=0}^{P-1}h_l[p]x_l[n-p-\tau_l]. \tag{15}$$

从滤波输出的信号 $y[n]$ 中提取语音特征矢量 $\boldsymbol{Z}=\{\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_T\}$, 下标表示帧数，是关于接收信号和滤波器参数的函数。定义一个滤波器参数矢量 $\boldsymbol{\xi}$ 包含该 FIR 滤波器中所有的系数 $h_l[p]$, $l=0,\cdots,L-1$, $p=0,\cdots,P-1$ 。基于 HMM 的语音识别系统，其识别假设的得出依照贝叶斯分类准则：

$$\widehat{w}=\arg\max_w P(\boldsymbol{Z}(\boldsymbol{\xi})|w)P(w), \tag{16}$$

其中单词的发生概率 $P(w)$ 基于语言模型得到，而后验假设概率 $P(\boldsymbol{Z}(\boldsymbol{\xi})|w)$ 的计算则基于识别系统的 HMM, 在识别过程中通常取对数似然函数进行计算。

如果能提高正确假设的后验概率，扩大正确假设与非正确假设之间的距离，则正确识别语音内容的概率有望得到提高。基于这一考虑，本文以最大化正确假设的后验概率作为最优滤波器参数矢量 $\boldsymbol{\xi}$ 的搜索准则。参数训练阶段，语音信号对应的内容 w_c 已知，因而 $P(w)$ 在计算中可略去不考虑。 $\boldsymbol{\xi}$ 的最大似然估计表达式如下：

$$\widehat{\boldsymbol{\xi}}=\arg\max_{\boldsymbol{\xi}}\log P(\boldsymbol{Z}(\boldsymbol{\xi})|w_c). \tag{17}$$

显然，通过最大化 $P(\boldsymbol{Z}(\boldsymbol{\xi})|w)$ 得到最佳的参数矢量 $\boldsymbol{\xi}$, 需要联合阵列滤波和 HMM 识别两部分进行迭代优化。针对似然函数最大化这一目标, 可分别采用局部搜索和全局搜索来进行滤波器系数优化。

3.1 局部优化

采用的局部优化算法基于文献 10 的推导，局部最优参数通过数值优化算法中的梯度下降法得到。求解梯度下降问题的一个关键是选择梯度方向, 不适当的下降方向可能抵消甚至恶化先前的优化结果，这一问题在求解高维变量的应用中尤为突出。而共轭梯度算法 [12] 沿着一组共轭的方向进行优化可以避免这种情况的发生, 是针对高维变量函数局部优化的常用方法之一。

识别中总的对数似然函数表达为：

$$L(\boldsymbol{\xi})=\log P(\boldsymbol{Z}(\boldsymbol{\xi})|w_c)=\sum_i\log P(\boldsymbol{z}_i|w_c), \tag{18}$$

其中下标 i 表示帧数。基于 HMM 计算的对数概率，通常表示为一个混合高斯模型的概率输出，即：

$$L(\boldsymbol{\xi}) = -\frac{1}{2}\sum_{i}\sum_{k=1}^{K}\{\gamma_{ik}(\boldsymbol{\xi})[\boldsymbol{z}_i(\boldsymbol{\xi})-\boldsymbol{\mu}_{ik}]^{\mathrm{T}} \boldsymbol{\Sigma}_{ik}^{-1}[\boldsymbol{z}_i(\boldsymbol{\xi})-\boldsymbol{\mu}_{ik}]+\kappa_{ik}\} \tag{19}$$

其中 γ_{ik} 是第 i 帧的特征矢量 $\boldsymbol{z}_i$ 所处的 HMM 状态其混合高斯模型的第 k 个高斯分量对应的权重，$\boldsymbol{\mu}_{ik}$ 是该高斯分量对应的均值矢量，$\boldsymbol{\Sigma}_{ik}$ 是相应的方差矩阵，而 κ_{ik} 是归一化的常量。从而可将 $L(\boldsymbol{\xi})$ 关于 $\boldsymbol{\xi}$ 的梯度表示为：

$$\nabla_{\xi}L(\boldsymbol{\xi}) = -\sum_{i}\sum_{k=1}^{K}\gamma_{ik}(\boldsymbol{\xi})\frac{\partial \boldsymbol{z}_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\boldsymbol{\Sigma}_{ik}^{-1}[\boldsymbol{z}_i(\boldsymbol{\xi})-\boldsymbol{\mu}_{ik}]. \tag{20}$$

本文采用共轭梯度法进行优化计算，由于自变量较多，有效的下降步长范围难以直接从数学计算中得出，所以将其根据数值测试结果预先设定为一个较小的量，再伴随迭代过程的进行适当加以调整。

3.2 全局优化

局部搜索算法中，函数可能只收敛到起始点附近某局部最优点处，而采用全局搜索算法，可以有效避免这一问题，使结果得到更大程度的优化，本文采用自适应模拟退火算法 [13] 对滤波器系数进行全局优化。模拟退火算法来源于固体退火原理，将固体加温至充分高，再让其徐徐冷却，升温时，固体内部粒子随温度升高变为无序状，内能增大，而徐徐冷却时粒子渐趋有序，在每个温度都达到平衡态，最后在常温时达到基态，内能减为最小。该算法的具体操作是，从初始设定的最高温度开始，每次优化先按照一定规律更新参数值 (即滤波器系数)，再选择接受或者拒绝，接受更新视为退火；在该温度下满足终止退火的条件，则进行降温操作，在次高温度下重复退火和降温，直到温度降至常温下结果收敛，或者达到设定的最大降温次数。本文采用的这种自适应模拟退火算法通过计算敏感度系数来监视和控制退火过程，设定代价函数 $J(\boldsymbol{\xi}) = -L(\boldsymbol{\xi})$ 。

4 实验结果

实验语音数据的采集是在一个 5.7 m×4.9 m 的长方形房间中进行的，高 2.8 m，房间内布置有会议桌椅、投影幕、冰箱、饮水机、空调及台式电脑等设备，构成一个典型具强混响的会议室环境。经计算，混响时间约为 330 ms 。

数据采集通过一个六通道的音频采集硬件系统，配合 PC 完成，主要包括六个同型号的全指向性电容传声器，一套放大倍数可调的多通道低噪声放大电路，和一块每通道采样频率可达 50 kHz 的 A/D 同步数据采集卡。实验中将六个传声器按照 5.2 cm 的相邻阵元中心间距排列成等间距的均匀直线阵接收语音数据。

对语音的内容识别基于 HTK[14](Hidden Markov Model Toolkit) 软件来进行。每帧提取一组 39×1 维的语音特征矢量，包含静态梅尔频率倒谱系数 (MFCC: Mel-Frequency Cepstral Coefficients) 、一阶差分 MFCC 、二阶差分 MFCC 三组的各 13 个元素。

实验中采用由常见英文单词构成的词库进行模型训练和测试，词库容量为 500 个单词，数据原始采样率为 20 kHz 。在安静的室内环境下， 12 个说话人在接近传声器处分别以正常语速阅读一遍词库中的所有单词，得到总量为 6000 个单词的训练数据。测试数据的采集分别在两种不同强度的环境噪声干扰下进行 (信噪比分别为 19 dB 和 26 dB)，5 个说话人分别在距离传声器阵 3 m 处阅读单词，采集到两种信噪比条件下的数据各 5 组 (每组 500 个单词) 。

首先考察将宽带 ESPRIT 方向估计算法运用在语音信号上的性能。在实验混响环境下，安排两个说话人分别在与传声器阵法向交 21° 和 26° 夹角的位置发音，由于语音信号自身的非平稳性，ESPRIT 算法在各短时帧内完成。针对存在较多矩阵运算的问题，采用了矩阵拆分的方法降低计算量。取信噪比为 19 dB 、持续时长为 0.6 s 的发音，短时帧帧长设定为 24 ms，相邻帧之间有 8 ms 的重叠。对每一帧数据进行独立的方向估计，对所有帧的计算结果，统计其在各角度区间的分布次数，结果如图 3 所示。显然，在说话人位置固定条件下，采用宽带 ESPRIT 算法，通过统计分布峰值的定位，完全能够正确估计

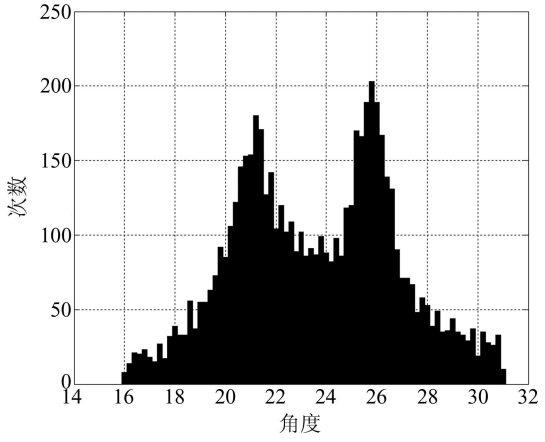

图 3 宽带 ESPRIT 计算的语音信号到达方向统计分布

出信号方向；而常规宽带波束形成算法在该尺寸的接收阵下无法分辨这一角度间隔的两个语音信号。同时，从图中可以看到，由于存在严重的噪声 / 混响干扰，大量的估计点落在实际信号角度以外的位置，可进一步结合短时帧内信噪比 / 信混比的估计加以摈除。

将经过延时补偿的多通道数据采样率降至 8 kHz, 采用第 3 节描述的优化算法搜索不同噪声强度下最优的滤波器参数。例如对信噪比为 19 dB 的英语单词 ‘nine’ 分别采用局部搜索和全局搜索两种方案进行概率最大化的滤波器参数搜索，滤波器每通道的阶数设定为 24 阶，图 4 给出了全局优化算法的迭代结果。经过模拟退火全局优化 760 多次的迭代，该似然函数从 −2200 左右上升到 −1750 左右，而经过相同次数的局部优化迭代只上升到 −2000 左右，全局优化效果更加明显，且可避免陷入局部最优点。性能改善的代价为计算量的增加，在配备 2.66 GHz 双核处理器、 2 GB 内存的个人电脑上，对于实验词库采用 Matlab 运行不超过 1000 次的全局优化迭代，计算时间最长约为 1400 s 。

实验中采集到了信噪比为 19 dB 和 26 dB 的六通道语音数据各 5 组，对受污染的多通道数据先采用宽带 ESPRIT 方向估计算法估计信号方向，并进行延时补偿。为了进行对比，在这两种不同的信噪比条件下，对 5 个说话人各收集了一组接近传声器说话的信号。近距离采集的信号在相同信噪比下受混响干扰最小，相比而言与训练模板最为匹配，所以以其识别率作为性能参考。

在不同信噪比条件下采用局部优化，全局优化两种方案分别进行滤波器系数优化，然后用各自得到的最优系数处理该信噪比下的 2500 个测试数据，进行内容识别。与未经处理的单通道数据，近距离采集的数据，以及常规波束形成处理后的数据进行比较，最终的识别结果见表 1 。显然，相比常规波束形成的方法，经过本文采用的各种优化方案处理受到噪声和混响干扰的语音数据后，语音识别的误认率明显得到了不同程度的降低，其中全局优化的效果最为显著。

前面已经指出，本文采用的方案是基于语音识别参数的后验概率最大化，而非信号波形的最优化，所以有必要分析本方案对于语音识别参数的影响。图 5 给出其中一个通道经过系数优化后的滤波器频率响应曲线。总体而言，该滤波器具有高通的性质，这一特点与环境噪声大部分集中于较低频段的事实相吻合；通带中存在一系列高低起伏的波峰与波谷，可能与语音信号短时幅度谱中存在的起伏特征有一定的关联性。其它通道的滤波器频响曲线均具有类似特性。就上述分析和识别性能得到改善的结果来说，本文提出的滤波方法在一定程度上起到了降噪的作用，图中各波峰对应的频率可理解为对于识别更为重要的分量在频域中所处的位置。信号经由滤波器滤波，关键的频率分量被加强，提取出的特征参数将与模板更为匹配，从而有效地提高信号被正确识别的可能性。

表 1　两种信噪比条件下各方法的错误识别率

	单通道	常规波束形成	局部优化	全局优化	近距离说话
19 dB	26.8%	21%	19.4%	16.6%	13.2%
26 dB	21.2%	16.8%	15.6%	13.8%	9.4%

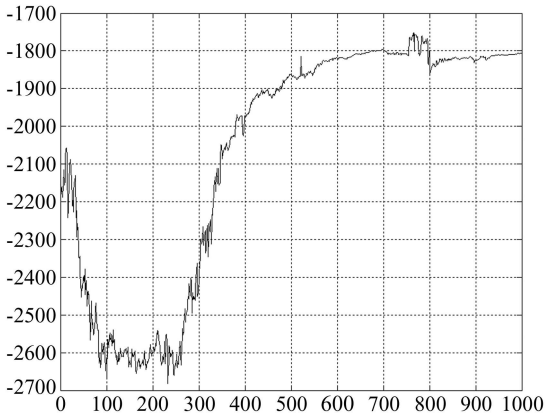

图 4　全局优化迭代过程中目标函数的变化

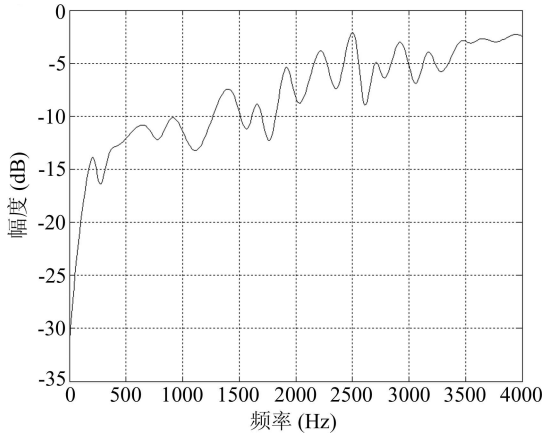

图 5　滤波器单通道的频率响应曲线图

5　结论

本文发展了一种联合基于信号到达方向估计的延时补偿和以增强语音特征为目标的多通道阵列滤波的传声器阵处理方法。该方法首次将 ESPRIT 算法应用于宽带语音信号到达方向估计，改善时延补

偿的执行效率；同时选取语音识别概率最大化作为滤波器优化准则，改善自动语音识别系统在混响环境下的可靠性；搜索过程采用全局算法，避免陷入局部最优解，进一步改善滤波器系数优化效果。针对容量为 500 个单词的词库的实验测试表明，在存在噪声和混响干扰的实际工作环境中，该方法能够加强对于识别更为重要的语音分量，提高待识别特征与模板的匹配程度，性能表现优于着眼于优化波形的常规阵处理方法。

进一步优化该联合框架下各算法的执行及将该框架拓展到更大词库的识别为未来工作的重点；针对大词库训练数据缺乏的情况，采取类似于通信系统中自适应均衡的技术方法是一种可能的方案。

参 考 文 献

1 Van Trees, Harry L. Optimum array processing. Part Ⅳ of Detection, Estimation, and Modulation Theory, New York: John Wiley, 2002: 6—12

2 Benesty J, Huang Y, Chen J. Microphone array signal processing. New York: Springer-Verlag, 2008: 1—5

3 Kayand S M, Marple S L Jr. Spectrum analysis—A modern perspective. In: Proc. IEEE, 1981; **69**(11): 1380—1419

4 Murthiand M N, Rao B D. Minimum variance distortionless response (MVDR) modeling of voiced speech. In: Proc. Int. Conf. Acoustics, Speech, Signal Processing'97, Munich, Germany, 1997; **3**: 1687—1690

5 Schmidt R O. Multiple emitter location and signal parameter spectral estimation. *IEEE Trans. Antenna and Propagation*, 1986; **34**(3): 276—280

6 Roy R, Kailath T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoustics, Speech and Signal Processing*, 1986; **37**(7): 984—995

7 Frost O L. An algorithm for linearly constrained adaptive array. In: Proc. IEEE, 1972; **68**(8): 926—935

8 Godara L, Cantoni A. Analysis of constrained LMS algorithm with application to adaptive beamforming using perturbation sequences. *IEEE Trans. Antennas and Propagation*, 1986; **34**(3): 368—379

9 Ottersten B, Kailath T. Direction-of-arrival estimation for wide-band signals using the ESPRIT algorithm. *IEEE Trans. Acoustics, Speech and Signal Processing*, 1990; **38**(2): 317—327

10 Seltzer M L, Raj B, Stern R M. Likelihood-maximizing beamforming for robust hands-free speech recognition. *IEEE Trans. Speech Audio Processing*, 2004; **12**(5): 489—498

11 Su G, Morf M. Modal decomposition signal subspace algorithms. *IEEE Trans. Acoustics, Speech and Signal Processing*, 1986; **34**(3): 585—602

12 Nocedal J, Wright S. Numerical optimization. New York: Springer-Verlag, 1999: 509—519

13 Chen S, Istepanian R, Luk B L. Digital IIR filter design using adaptive simulated annealing. *Digital Signal Processing*, 2001; **11**(3): 241—251

14 Young S, Evermann G, Gales Mark *et al.* The HTK Book, Cambridge University Engineering Department: Version 3.4, 2006