基于听觉感知特性的信号子空间麦克风阵列语音增强算法

程 宁 [1] 刘文举 [1]

摘 要 针对麦克风阵列信号子空间语音增强算法的不足, 结合人耳的听觉掩蔽效应, 提出了改进的信号子空间算法. 提出了通过置信度判断来确定噪声子空间维度的方法, 在噪声子空间上, 通过条件概率的方法估计出噪声功率谱. 在此基础上, 结合人耳的听觉掩蔽效应给出了线性滤波器的一种合理估计. 实验结果表明所提的方法相对于传统算法, 更有效地抑制了噪声, 在多项语音质量评价指标上都有明显的改进.

关键词 语音增强, 信号子空间, 麦克风阵列, 听觉掩蔽效应, 特征值分解
中图分类号 TP391

Perceptual Properties Based Signal Subspace Microphone Array Speech Enhancement Algorithm

CHENG Ning[1] LIU Wen-Ju[1]

Abstract To aim at the drawbacks of the conventional subspace microphone array speech enhancement method, some improvements are proposed by using masking properties of human ears. Confidence function is used to determine the noise subspace dimension. In noise subspace, the noise power spectrum is estimated by the conditional likelihood. Then, the masking properties are incorporated into the subspace method to estimate the linear filter. Experiments show that compared with conventional algorithms, the proposed approach suppresses noise more effectively and obtains a significant improvement on objective speech quality measures.

Key words Speech enhancement, signal subspace, microphone array, masking properties, eigen-decomposition

麦克风阵列语音增强算法近年来得到了广泛的研究. 其中, 信号子空间算法具有出色的消除加性宽带噪声的能力[1−4]. 信号子空间算法将带噪信号空间分解为信号子空间 (包含目标信号和噪声) 和噪声子空间 (只包含噪声), 并在信号子空间中估计出目标信号. 信号子空间算法的核心在于合理地估计线性滤波器, 其要点之一是准确地估计信号子空间维度和噪声功率谱. 对信号子空间语音增强算法的研究已证明该算法具有很好的语音增强性能[1−4]. 尽管信号子空间算法性能优越, 但想要完全消除噪声, 依然具有相当的难度. 通常, 信号子空间算法消噪以后, 增强语音中依然会存在一定的残余噪声, 这些噪声降低了语音的感知质量. 为了尽量减少残余噪声对目标语音的影响, 人们在大量的实验基础上发现人耳的听觉掩蔽效应能够用来实现这一目标. 人耳的听觉掩蔽效应是指: 在通常情况下, 目标语音信号是强信号, 而背景噪声相对较弱, 这样人耳听觉系统会根据具体的目标语音信号确定频域上的听觉掩蔽阈值, 如果使滤波后的残余噪声限制在人耳的听觉掩蔽阈值之下, 那么该噪声就不会被人耳感知. 经过多年来的研究, 这一听觉效应被有效地应用在了语音增强算法中[4−6]. 只要将增强后的语音中残余噪声的量限制在一定的范围内, 就能使其在目标语音的掩蔽下不被人耳感知, 从而实现对目标语音的增强.

本文在改进了原有的信号子空间算法的基础上, 将人耳的听觉掩蔽效应应用在信号子空间算法中, 通过限制增强后语音中残余噪声的水平, 得到了一种新的算法, 使得增强后的语音具有更好的质量. 本文首先改进了信号子空间算法, 利用噪声子空间中特征值应该相等的特点, 用置信度判断其是否相等的方法来确定噪声子空间的维度, 根据噪声子空间中噪声功率谱应小于信号子空间中带噪信号功率谱的特点, 用条件概率的方法对噪声功率谱做出估计. 接着利用人耳的听觉掩蔽特性, 根据目标语音强度估计出相应的听觉掩蔽阈值, 通过将增强语音中残余噪声的水平限制在该阈值以下, 合理地估计了线性滤波器中的拉格朗日乘子, 给出了线性滤波器的

收稿日期 2008-08-25 收修改稿日期 2009-03-13
Received August 25, 2008; in revised form March 13, 2009
国家重点基础研究发展计划 (973 计划) (2004CB318105), 国家高技术研究发展计划 (863 计划) (20060101Z4073, 2006AA01Z194), 国家自然科学基金 (90820011, 60675026, 60121302) 资助
Supported by National Basic Research Program of China (973 Program) (2004CB318105), National High Technology Research and Development Program of China (863 Program) (20060101Z4073, 2006AA01Z194), and National Natural Science Foundation of China (90820011, 60675026, 60121302)
1. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190
DOI: 10.3724/SP.J.1004.2009.01481

一种新的估计方式, 得到了一种新的基于听觉掩蔽效应的信号子空间麦克风阵列语音增强算法.

1 算法描述

1.1 信号子空间算法描述

信号子空间算法的原理在于通过特征值分解的方法将带噪信号空间分解成两个子空间: 信号子空间 (包含目标信号和噪声) 和噪声子空间 (只包含噪声), 然后在信号子空间上恢复出目标信号. 这样做的原因在于: 语音信号能够被建模成一些基向量的线性组合. 通常, 纯净语音信号功率谱矩阵的一些特征值非常接近于零, 这表明纯净语音信号的能量只分布在某些基向量上. 信号子空间算法的噪声假设为白噪声 (有色噪声可通过预白化的方法予以白化[2−3]), 白噪声的所有特征值都是正的, 噪声能量分布在带噪信号的所有基向量上. 所以, 由带噪信号的基所组成的空间可分解成一个信号子空间 (包含目标信号和噪声) 和一个噪声子空间 (只包含噪声). 相应的, 在信号子空间上就可以恢复出目标信号, 而噪声子空间由于不包含目标信号则可以不用考虑.

假设由 L 个麦克风组成的阵列上接收到的带噪语音信号向量的频域表示为: $\boldsymbol{X} = [X_1, X_2, \cdots, X_L]^{\mathrm{H}}$, 目标语音向量为 $\boldsymbol{S}$, $\boldsymbol{N}$ 是噪声信号向量, 有如下表达式成立:

$$\boldsymbol{X} = \boldsymbol{S} + \boldsymbol{N} \tag{1}$$

设 R_X 为带噪信号的功率谱矩阵, R_S 为目标信号的功率谱矩阵, R_N 为噪声信号的功率谱矩阵. 在目标信号与噪声信号不相关的假设下, 有:

$$R_X = R_S + R_N \tag{2}$$

目标信号功率谱矩阵的特征值分解可表述如下:

$$R_S = U\Lambda_S U^{\mathrm{H}} \tag{3}$$

其中, $\Lambda_S = \mathrm{diag}\{\lambda_{S_1}, \cdots, \lambda_{S_Q}, 0, \cdots, 0\}$ 为特征值降序排列的特征值矩阵, U 为对应的特征向量矩阵, $Q \le L$ 为矩阵的秩.

假设噪声为白噪声且功率谱为 σ_N^2, 则有:

$$R_X = U\Lambda_X U^{\mathrm{H}} = U(\Lambda_S + \sigma_N^2 I)U^{\mathrm{H}} \tag{4}$$

其中, $\Lambda_X = \mathrm{diag}\{\lambda_{X_1}, \cdots, \lambda_{X_L}) = \mathrm{diag}\{\lambda_{S_1} + \sigma_N^2, \cdots, \lambda_{S_Q} + \sigma_N^2, \sigma_N^2, \cdots, \sigma_N^2\}$ 为特征值降序排列的特征值矩阵.

R_X 的特征值为 λ_{X_i}, R_S 的特征值为 λ_{S_i}, 则 λ_{X_i} 和 λ_{S_i} 存在如下的关系:

$$\lambda_{S_i} = \begin{cases} \lambda_{X_i} - \sigma_N^2, & \text{若 } i = 1, \cdots, Q \\ 0, & \text{若 } i = Q+1, \cdots, L \end{cases} \tag{5}$$

设 H 为一线性滤波器, 可得到目标信号的估计如下:

$$\hat{\boldsymbol{S}} = H\boldsymbol{X} \tag{6}$$

事实上, 由线性滤波器恢复的目标语音的质量主要表现在两个方面: 1) 目标语音的畸变; 2) 残余噪声的大小. 在文献 [1] 中, Ephaim 等在将噪声限制在一定的范围内的条件下, 通过极小化语音畸变, 得到了 H 的表达式如下:

$$H = UGU^{\mathrm{H}} = U \begin{bmatrix} G_1 & 0 \\ 0 & 0 \end{bmatrix} U^{\mathrm{H}} \tag{7}$$

其中, G_1 为 $Q \times Q$ 的满秩矩阵.

G 可表述如下:

$$G = \Lambda_S(\Lambda_S + \sigma_N^2\Lambda_\mu)^{-1} \tag{8}$$

其中, $\Lambda_\mu = \mathrm{diag}\{\mu_1, \cdots, \mu_L\}$ 为拉格朗日乘子矩阵.

G 为对角矩阵, 对角线元素 g_i 可表述如下:

$$g_i = \begin{cases} \dfrac{\lambda_{S_i}}{\lambda_{S_i} + \mu_i\sigma_N^2}, & \text{若 } i = 1, \cdots, Q \\ 0, & \text{若 } i = Q+1, \cdots, L \end{cases} \tag{9}$$

1.2 改进的子空间维度估计和噪声估计方法

为了估计出 g_i, 首先需要估计噪声子空间维度 $L-Q$ 和噪声功率谱 σ_N^2. 传统的信号子空间算法确定子空间维度的方法通常是设一个固定阈值, 信号子空间的维度就是大于该阈值的特征值的个数. 这种确定子空间维度的方法在实际应用中效果较差, 因为阈值的设定具有较大的人为性, 而且通常不能随着信号的改变而自适应的调整. 这导致了子空间维度估计常出现较大误差, 降低了信号子空间算法的性能.

为了能够准确地计算信号子空间维度, 一种合理的方法就是利用噪声子空间本身的特点, 即白噪声子空间上噪声功率谱应该相等. 由于 Λ_X 中的特征值是降序排列的, 先假设噪声子空间维度是 $L-Q=1$, 然后依次增加噪声子空间的维度值 $L-Q$, 测试 Λ_X 中最后 $L-Q$ 维特征值是否相等, 取符合相等条件的最大的 $L-Q$ 值为噪声子空间的维度, 这样就可以较为准确地估计出噪声子空间维度 $L-Q$, 进而估计出信号子空间维度 Q.

利用这一思想, 本文提出了采用条件假设来估计噪声子空间维度的方法, 提出原假设和对立假设如下:

H_0: $\lambda_{X_{Q+1}} = \lambda_{X_{Q+2}} = \cdots = \lambda_{X_L}$,

H_1: 在这 $L-Q$ 个特征值中至少有两个特征值不同.

假设特征值服从高斯分布[7], 则分布函数可表述如下:

$$F(H_0)=(2\pi)^{-\frac{L-Q}{2}}\left(\frac{1}{L-Q}\sum_{i=Q+1}^{L}\lambda_{X_i}\right)^{-\frac{L-Q}{2}}\times \exp\left\{-\frac{1}{2}\mathrm{tr}[\Lambda_m]\right\} \tag{10}$$

$$F(H_1)=(2\pi)^{-\frac{L-Q}{2}}\left(\prod_{i=Q+1}^{L}\lambda_{X_i}\right)^{-\frac{1}{2}}\times \exp\left\{-\frac{1}{2}\mathrm{tr}[\Lambda_m]\right\} \tag{11}$$

其中, $\Lambda_m=\mathrm{diag}\{\lambda_{X_{Q+1}},\cdots,\lambda_{X_L}\}$. 令: $\overline{\lambda}=\frac{1}{L-Q}\sum_{i=Q+1}^{L}\lambda_{X_i}$, $\lambda_{X_i}=\overline{\lambda}+h_i$, h_i 为 λ_{X_i} 相对于 $\overline{\lambda}$ 的偏差. 有:

$$\begin{aligned}-2\ln\left[\frac{F(H_0)}{F(H_1)}\right]&=-\ln\prod_{i=Q+1}^{L}\lambda_{X_i}+(L-Q)\ln\overline{\lambda}=\\&-\sum_{i=Q+1}^{L}\ln\left(\frac{\lambda_{X_i}}{\overline{\lambda}}\right)=-\sum_{i=Q+1}^{L}\ln\left(1+\frac{h_i}{\overline{\lambda}}\right)=\\&-\sum_{i=Q+1}^{L}\ln\left(\frac{h_i}{\overline{\lambda}}-\frac{h_i^2}{2\overline{\lambda}^2}+\cdots\right)\approx\\&-\sum_{i=Q+1}^{L}\frac{h_i}{\overline{\lambda}}+\sum_{i=Q+1}^{L}\frac{h_i^2}{2\overline{\lambda}^2}=\sum_{i=Q+1}^{L}\frac{h_i^2}{2\overline{\lambda}^2}\end{aligned} \tag{12}$$

由文献 [7], h_i 近似地服从均值为零、方差为 $2\overline{\lambda}^2$ 的高斯分布. 所以, 由文献 [8], $-2\ln[F(H_0)/F(H_1)]$ 近似地服从自由度为 $\theta=L-Q-1$ 的卡方分布. 确定置信度 α, 取满足 $-2\ln[F(H_0)/F(H_1)]\geq\chi^2_{\theta,\alpha}$ 的最大 $L-Q$ 值为噪声子空间维度, 进而得到信号子空间维度 Q.

在噪声子空间上需要估计出噪声功率谱. 为准确地估计噪声功率谱, 考虑到噪声子空间中的噪声功率谱要小于信号子空间中的带噪信号功率谱的特点, 本文用条件概率来估计噪声功率谱:
令

$$\overline{\lambda}_N=\frac{1}{L-Q}\sum_{i=Q+1}^{L}\lambda_{X_i},\ \overline{\lambda}_{S+N}=\frac{1}{Q}\sum_{i=1}^{Q}\lambda_{X_i}$$

$\overline{\lambda}_N$ 应取小于 $\overline{\lambda}_{S+N}$ 的值, 所以本文用条件概率给出噪声功率谱估计如下:

$$\begin{aligned}\hat{\sigma}_N^2&=\mathrm{E}[\overline{\lambda}_N|\overline{\lambda}_N<\overline{\lambda}_{S+N}]=\\&\frac{\int_0^{\overline{\lambda}_{S+N}}x f_{\overline{\lambda}_N}(x)\mathrm{d}x}{\int_0^{\overline{\lambda}_{S+N}}f_{\overline{\lambda}_{S+N}}(x)\mathrm{d}x}=\\&\frac{\int_0^{\overline{\lambda}_{S+N}}\frac{x^2}{\sqrt{2\pi\overline{\lambda}_N}}\mathrm{e}^{-\frac{x^2}{2\overline{\lambda}_N}}\mathrm{d}x}{\int_0^{\overline{\lambda}_{S+N}}\frac{x}{\sqrt{2\pi\overline{\lambda}_{S+N}}}\mathrm{e}^{-\frac{x^2}{2\overline{\lambda}_{S+N}}}\mathrm{d}x}=\\&\frac{\sqrt{2\pi}\,\overline{\lambda}_N\left(1-\mathrm{e}^{-\frac{\overline{\lambda}_{S+N}^2}{2\overline{\lambda}_N}}\right)^{\frac{1}{2}}-\sqrt{\overline{\lambda}_N}\overline{\lambda}_{S+N}\mathrm{e}^{-\frac{\overline{\lambda}_{S+N}^2}{2\overline{\lambda}_N}}}{\sqrt{\overline{\lambda}_{S+N}}\left(1-\mathrm{e}^{-\frac{\overline{\lambda}_{S+N}}{2}}\right)}\end{aligned} \tag{13}$$

噪声子空间维度的过估计或欠估计会导致噪声功率谱的估计误差, 这一误差可以用一个补偿因子来解决[9].

$$\tilde{\sigma}_N^2=\frac{1}{B(Q)}\hat{\sigma}_N^2 \tag{14}$$

其中, $B(Q)=\frac{\mathrm{E}[\hat{\sigma}_N^2]}{\overline{\sigma}_N^2}$ 为补偿因子, $\overline{\sigma}_N^2$ 为预估噪声功率谱.

1.3　基于听觉掩蔽效应的线性滤波器估计

在估计出信号子空间维度和噪声功率谱后, 为了能得到线性滤波器 H 的合理估计, 接下来, 需要对式 (9) 中拉格朗日乘子 μ_i 给出合理的估计. 根据文献 [1] 中的分析, μ_i 对于目标语音的估计有很重要的影响, 它表示的是残余噪声和信号畸变之间的折中. 随着它的增加, 残余噪声将会减少而信号畸变将会增加. Ephaim 等[1] 给出了一种简单选择 μ_i 的方法, 即取一个固定的 μ_i 值, 由于固定的 μ_i 值不能随信号的改变而相应地调整, 这样的选择显然不是最优的.

本文利用人耳的听觉掩蔽效应, 通过限制增强语音中的残余噪声水平, 给出了一种 μ_i 值的估计方法.

首先需要根据目标语音的强度来计算人耳基膜上的听觉掩蔽阈值, 在该阈值之下的噪声能够被目标语音所掩蔽.

估计目标语音的强度需要用到信号子空间的基向量, 所以根据信号子空间维度 Q, 将特征向量矩阵 U 分解为两个子矩阵: $U=[U_1;U_2]$, 其中, $U_1\in\mathbf{C}^{L\times Q}$ 为信号子空间的基, $U_2\in\mathbf{C}^{L\times(L-Q)}$ 为噪声子空间的基.

人耳听觉频率范围是 0 到 15 500 Hz, 覆盖了 24 个关键子频带, 需要在每个子频带中计算听觉掩蔽

阈值. 文献 [5] 中给出了各个关键子频带的上下限频率值.

首先, 计算表征人耳基膜上能量的激励能量值 $\boldsymbol{C}(j,b)$:

$$\boldsymbol{C}(j,b)=SF(j)*\boldsymbol{E}(j,b) \tag{15}$$

其中, $\boldsymbol{E}(j,b)$ 表示的是第 j 个子频带内第 b 个频点上的能量, $SF(j)$ 是表达人耳基膜特性的传播函数, $j=1,\cdots,J$, J 是实际使用的子频带个数. 因为通常用到的带噪语音信号的最高有效频率都在 15 500 Hz 以下, 所以用到的关键子频带的个数通常要小于 24 个, 这样一来, J 的值就需要根据具体的语音数据的最高频率来确定, 通常 J 的值都会小于 24.

传播函数 $SF(j)$ 计算如下[6]:

$$\begin{aligned}SF(j)=&15.81+7.5(j+0.474)-\\&17.5\sqrt{1+(j+0.474)^2}\end{aligned} \tag{16}$$

频点能量 $\boldsymbol{E}(j,b)$ 则可根据目标语音信号的特征值和特征向量计算出来[4]:

$$\boldsymbol{E}(j,b)=\frac{1}{L}\sum_{i=1}^{Q}\lambda_{S_i}|\boldsymbol{U}_{1,i}|^2 \tag{17}$$

其中, $\lambda_{S_i}=\lambda_{X_i}-\tilde{\sigma}_N^2$ 为目标信号功率谱矩阵的特征值估计, $\boldsymbol{U}_{1,i}$ 为信号子空间的第 i 个基.

听觉掩蔽阈值 $\boldsymbol{C}_{\rm thr}$ 可由激励能量值 $\boldsymbol{C}(j,b)$ 计算出来:

$$\boldsymbol{C}_{\rm thr}=10^{\log_{10}|\boldsymbol{C}(j,b)|-\frac{|O(j)|}{20}-\frac{|\tilde{\sigma}_N^2|}{5}} \tag{18}$$

其中, $O(j)$ 是偏移量[5], $\tilde{\sigma}_N^2$ 是噪声功率谱.

为了能够在特征值域上应用听觉掩蔽效应, 需要将听觉掩蔽阈值 $\boldsymbol{C}_{\rm thr}$ 映射到特征值域上, Jabloun[4] 根据频域到特征值域的变换关系, 给出听觉掩蔽阈值 $\boldsymbol{C}_{\rm thr}$ 到特征值域上的映射如下:

$$\boldsymbol{\theta}=|U_1^{\rm H}|^2\boldsymbol{C}_{\rm thr} \tag{19}$$

其中, $\boldsymbol{\theta}=[\theta_1,\cdots,\theta_Q]^{\rm H}$ 为特征值域的掩蔽能量, 在掩蔽能量之下的噪声将会被目标语音掩蔽掉.

接下来, 需要计算增强后语音中的残余噪声能量, 以使其低于掩蔽能量值而被目标语音掩蔽. 残余噪声 $\hat{\boldsymbol{N}}$ 可由带噪输入信号中的噪声线性滤波后得到, 即: $\hat{\boldsymbol{N}}=H\boldsymbol{N}$. 计算残余噪声 $\hat{\boldsymbol{N}}$ 的功率谱矩阵如下:

$$\begin{aligned}R_{\hat{\boldsymbol{N}}}=&{\rm E}\{\hat{\boldsymbol{N}}\hat{\boldsymbol{N}}^{\rm H}\}=\\&{\rm E}\{H\boldsymbol{N}\boldsymbol{N}^{\rm H}H^{\rm H}\}=\\&HR_NH^{\rm H}=\\&UGU^{\rm H}(\tilde{\sigma}_N^2I)UG^{\rm H}U^{\rm H}=\\&U\Lambda_{\hat{\boldsymbol{N}}}U^{\rm H}\end{aligned} \tag{20}$$

其中, $\Lambda_{\hat{N}}=\Lambda_S(\Lambda_S+\tilde{\sigma}_N^2\Lambda_\mu)^{-1}\tilde{\sigma}_N^2I[\Lambda_S(\Lambda_S+\tilde{\sigma}_N^2\Lambda_\mu)^{-1}]^{\rm H}$ 为对角矩阵, 其第 i 个对角元素为:

$$\lambda_{\hat{N}_i}=\left(\frac{\lambda_{S_i}}{\lambda_{S_i}+\mu_i\tilde{\sigma}_N^2}\right)^2\tilde{\sigma}_N^2 \tag{21}$$

为掩蔽噪声, 应使 $\lambda_{\hat{N}_i}\le\theta_i$, 可得:

$$\mu_i\ge\frac{\lambda_{S_i}\left(\tilde{\sigma}_N-\theta_i^{\frac{1}{2}}\right)}{\tilde{\sigma}_N^2\cdot\theta_i^{\frac{1}{2}}} \tag{22}$$

考虑到应使 $\mu_i\ge0$, 本文取:

$$\mu_i=\begin{cases}\dfrac{\lambda_{S_i}\left(\tilde{\sigma}_N-\theta_i^{\frac{1}{2}}\right)}{\tilde{\sigma}_N^2\cdot\theta_i^{\frac{1}{2}}}, & \text{若 } \tilde{\sigma}_i-\theta_i^{\frac{1}{2}}\ge0\\ 0, & \text{若 } \tilde{\sigma}_i-\theta_i^{\frac{1}{2}}<0\end{cases} \tag{23}$$

将式 (23) 代入到式 (9) 中, 得到对角矩阵 G 的对角线元素 g_i 的估计如下:

$$g_i=\begin{cases}\dfrac{1}{1+\max(\tilde{\sigma}_N/\theta_i^{\frac{1}{2}}-1,0)}, & \text{若 } i=1,\cdots,Q\\ 0, & \text{若 } i=Q+1,\cdots,L\end{cases} \tag{24}$$

将 g_i 值代入到式 (7) 中, 即可得到所需的线性滤波器 H.

1.4 本文算法的实现

图 1 所示为本文算法实现的总体流程.

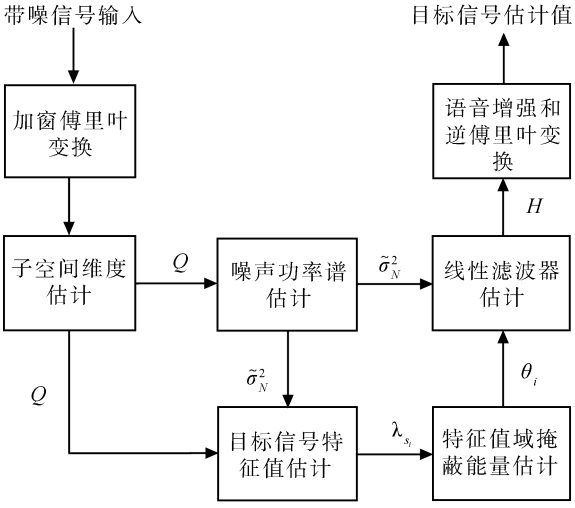


图 1 本文算法实现的流程图

Fig. 1 The flowchart of the proposed algorithm

1) 将带噪输入信号进行加窗傅里叶变换到频域后, 对其功率谱矩阵 R_X 进行特征值分解, 得到特征值 λ_{X_i} 和特征向量矩阵 U.

2) 从 1 开始, 依次递增噪声子空间维度 $L-Q$ 的值并测试 Λ_X 中最后 $L-Q$ 维特征值是否相等, 以此来确定噪声子空间的维度. 具体的方法是用假设检验的方法, 取满足表达式 $-2\ln[F(H_0)/F(H_1)] \geq \chi^2_{\theta,\alpha}$ 的最大 $L-Q$ 值为噪声子空间维度, 进而得到信号子空间维度 Q. 根据 Q 将特征向量矩阵 U 分为两个子矩阵 $U=[U_1;U_2]$. 其中, $U_1 \in \mathbf{C}^{L\times Q}$ 为信号子空间的基, $U_2 \in \mathbf{C}^{L\times(L-Q)}$ 为噪声子空间的基.

3) 根据噪声子空间中的噪声功率谱要小于信号子空间中的带噪信号功率谱的特点, 应用条件概率表达式 (13) 计算噪声功率谱 $\hat{\sigma}_N^2$, 并用补偿式 (14) 补偿由子空间估计误差带来的噪声功率谱误差, 得到噪声功率谱更新估计值 $\tilde{\sigma}_N^2$.

4) 用式 (5) 估计目标信号功率谱矩阵特征值, 用式 (15) 计算得到表征人耳基膜上能量的激励能量值 $\boldsymbol{C}(j,b)$.

5) 用式 (18) 计算得到听觉掩蔽阈值 $\boldsymbol{C}_{\rm thr}$, 用式 (19) 将 $\boldsymbol{C}_{\rm thr}$ 映射到特征值域得到特征值域的掩蔽能量 $\boldsymbol{\theta}=[\theta_1,\cdots,\theta_Q]^{\rm H}$.

6) 根据听觉掩蔽效应, 在特征值域上, 由式 (23) 计算出拉格朗日乘子 μ_i, 并由式 (24) 计算得到 g_i.

7) 将 g_i 代入到式 (7) 中, 得到线性滤波器 H, 最后由式 (6) 计算得到目标信号的估计值 $\hat{\boldsymbol{S}}$.

2　语音增强实验结果及分析

本文实验采用了卡内基梅隆 (Carnegie Mellon University, CMU) 麦克风阵列数据库[10]. 数据是由一个 8 麦克风的等间距线性麦克风阵列所采集. 麦克风之间的距离是 7 cm. 数据采样率为 16 kHz. 噪声为计算机风扇噪声. 数据集收集了 10 个人, 每人 13 句话, 共 130 句. 输入数据采用了加窗分帧傅里叶变换, 帧长 25 ms, 帧移为 10 ms. 所采用的窗函数为海明窗函数. 实验采用的客观语音质量评价标准包括: 分段信噪比增强 (Segmental signal-to-noise ratio enhancement, SSNRE), 增强信号的分段信噪比相对于输入信号分段信噪比的增加, IS (Itakura-Saito) 距离, 对数域比例距离 (Log-area-ratio, LAR) 和对数域似然距离 (Log-likelihood ratio, LLR)[11]. 其中, SSNRE 越大, IS、LAR 和 LLR 越小, 语音质量越好. 本文取置信度 $\alpha=0.9$ 来估计噪声子空间的维度.

为测试本文所提子空间维度估计方法和噪声估计方法的有效性, 本文在 Ephaim 算法的基础上, 用所提的子空间维度估计方法和噪声估计方法代替原算法中的用设定阈值来估计子空间维度和在静音段估计噪声的方法进行了实验, 实验结果如表 1 所示.

表 1　Ephaim 算法在 CMU 数据库上的平均测试结果

Table 1　Experimental results of the Ephaim algorithm on CMU database

算法	SSNRE (dB)	IS	LAR	LLR
Ephaim 算法	4.47	8.68	11.41	1.00
本文方法	5.11	7.15	11.39	0.90

表 1 中的 "本文方法" 是 Ephaim 算法用本文所提的子空间维度估计方法和噪声估计方法估计子空间维度和噪声后取得的结果. 可以看出, 本文方法在 4 项评价指标上都提高了 Ephaim 算法的性能. 本文方法在各项评价指标上的改进分别为: 分段信噪比增强 14.3 %, IS 距离 17.6 %, LAR 指标 0.2 %, LLR 指标 10 %.

为测试本文将听觉掩蔽效应引入到信号子空间算法中的效果, 本文在 Ephaim 算法的基础上, 将该算法在拉格朗日乘子 μ_i 分别取 2 (Ephaim 推荐) 和式 (23) 中的值时的实验结果进行了比较 (子空间维度估计方法和噪声估计方法都采用 Ephaim 算法原有的方法), 实验结果如表 2 所示.

表 2　Ephaim 算法在 CMU 数据库上 μ_i 取不同值时的平均测试结果

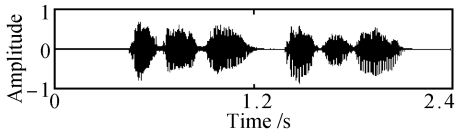
Table 2　Experimental results of the Ephaim algorithm on CMU database with different μ_i

Ephaim 算法	SSNRE (dB)	IS	LAR	LLR
$\mu_i=2$	4.47	8.68	11.41	1.00
本文 μ_i	7.73	2.35	6.40	0.86

可以看到, 当 μ_i 取本文的值时, Ephaim 算法效果有明显的改进. 在各项评价指标上的改进分别为: 分段信噪比增强 72.9 %, IS 距离 72.9 %, LAR 指标 43.9 %, LLR 指标 14 %. 由此可见, 将听觉掩蔽效应引入到信号子空间算法中的确能改进信号子空间算法的性能, 而本文所提的 μ_i 值对于提升算法性能有很大效果. 由此也可以看出, 拉格朗日乘子 μ_i 的估计对于信号子空间算法的性能具有较大的影响. 合理的 μ_i 值估计能够极大地提高算法性能.

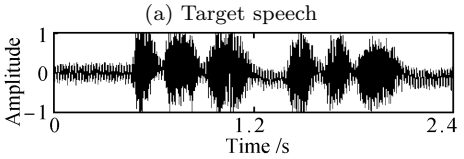
由图 2 可以很容易地看出本文所提的 μ_i 值使得 Ephaim 算法消除了更多的噪声, 更重要的是使得恢复的目标语音波形具有更小的畸变. 由此可以看出, μ_i 值的合理估计对于减少增强语音的畸变有很重要的影响.

为评估本文算法的整体效果, 本文将所提算法与 McCowan 算法[12] 和 Ephaim 算法[1] 进行了比较, 实验结果如表 3 所示.



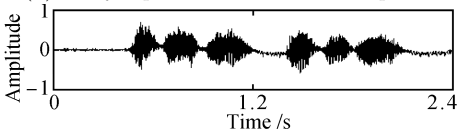
(a) 目标语音

(a) Target speech



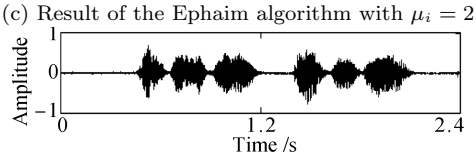
(b) 中间麦克风的带噪输入

(b) Noisy input of the central microphone



(c) Ephaim 算法取 $\mu_i = 2$ 的实验结果

(c) Result of the Ephaim algorithm with $\mu_i = 2$



(d) Ephaim 算法取本文所提的 μ_i 的实验结果

(d) Result of the Ephaim algorithm with the proposed μ_i

图 2 语句 “beeoer” 的波形图

Fig. 2 The waveforms of the utterance “beeoer”

表 3 算法在 CMU 数据库上的平均测试结果

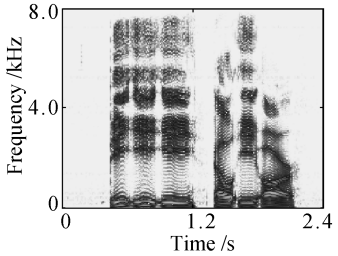
Table 3 Experimental results on CMU database

	SSNRE (dB)	IS	LAR	LLR
带噪输入	—	2.16	8.70	0.68
McCowan	5.07	26.48	11.31	1.43
Ephaim 算法	4.47	8.68	11.41	1.00
本文算法	10.40	2.59	5.27	0.57

从表 3 中可以看到, 本文算法在各项评价指标上都比比较算法要好. 相对于比较算法中的最好算法, 本文算法在各项评价指标上的改进分别为: 分段信噪比增强 5.33 dB, IS 距离 70.2 %, LAR 指标 53.4 %, LLR 指标 43 %.

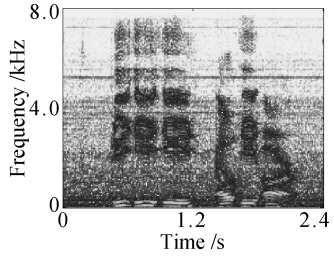
从图 3 中可以看出, McCowan 算法具有较大的目标信号失真, 且对于噪声能量集中的低频噪声消噪效果较差. Ephaim 算法比 McCowan 算法有较小的目标信号失真, 但低频消噪效果也较差. 相对于比较算法, 本文算法在不增加目标信号失真的情况下更好地消除了噪声能量集中的低频噪声.

在进行了语音客观质量评价后, 为进一步验证本文算法增强语音给人的听觉效果, 本文又进行了主观听觉评价实验. 本文在 CMU 数据库中选取了 20 句带噪语音 (10 个人每人两句), 用本文的算法和比较算法分别进行处理得到增强后的语音. 一共有



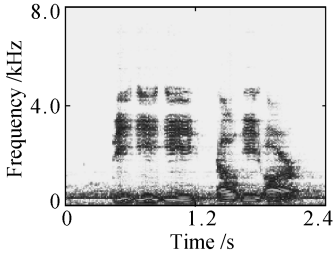
(a) 目标语音

(a) Target speech



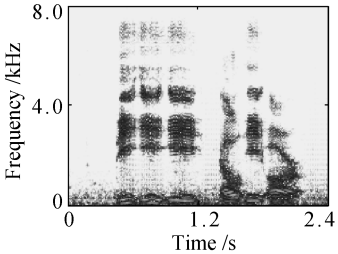
(b) 中间麦克风的带噪输入

(b) Noisy input of the central microphone



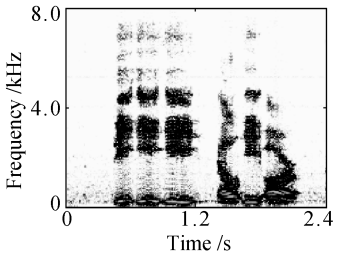
(c) McCowan 算法增强结果

(c) Result of the McCowan algorithm



(d) Ephaim 算法增强结果

(d) Result of the Ephaim algorithm



(e) 本文算法增强结果

(e) Result of the proposed algorithm

图 3 语句 “beeoer” 的语谱图

Fig. 3 The spectrograms of the utterance “beeoer”

5 个人参加了该听觉实验. 每个人都要求在增强后的语音中按给定的标准选择自己认为最理想的语音. 标准有两个: 1) 噪声最小; 2) 语音失真最小. 对于标准 1 和 2 分别统计在该标准下各个算法结果被选中的百分比. 实验结果如表 4 所示.

表 4　主观听觉实验的实验结果

Table 4　Subjective auditory experimental results

	噪声最小 (%)	语音失真最小 (%)
McCowan	11	12
Ephaim 算法	8	27
本文算法	81	61

从表 4 中可以看出, 相对于比较算法的增强语音结果而言, 本文所提算法增强后的语音包含的噪声最少, 与目标语音也最为接近.

最后, 本文对算法的计算复杂度进行简单的分析. 本文所提的算法是对信号子空间算法的改进. 信号子空间算法的计算量主要集中在对于矩阵特征值和特征向量的求解以及矩阵乘法上. 对于大小为 $n \times n$ 的矩阵而言, 进行矩阵特征值和特征向量求解的算法复杂度和矩阵乘法的算法复杂度都是 $\mathrm{O}(n^3)$, 所以信号子空间算法的复杂度是 $\mathrm{O}(n^3)$. 相对于传统的信号子空间算法如 Ephaim 算法而言, 本文只在求取信号子空间维度、噪声功率谱和拉格朗日乘子时增加了两个计算复杂度为 $\mathrm{O}(n^2)$ 的操作, 所以本文算法的复杂度也是 $\mathrm{O}(n^3)$, 与其他信号子空间算法在同一个量级上. 而 McCowan 算法由于不存在矩阵运算, 计算量主要集中在对信号的功率谱、场函数和后滤波器的估计上, 其算法复杂度在 $\mathrm{O}(n^2)$ 量级.

3　结论

本文首先改进了信号子空间算法, 用置信度判断噪声子空间中特征值是否相等来确定噪声子空间维度, 根据噪声子空间中噪声功率谱小于信号子空间中带噪信号功率谱的特点, 在噪声子空间上, 用条件概率估计出噪声功率谱. 在此基础上, 结合人耳的听觉掩蔽效应, 在合理地估计了线性滤波器中拉格朗日乘子的基础上, 给出了线性滤波器的一种新的估计方式, 得到了一种新的基于听觉掩蔽效应的信号子空间麦克风阵列语音增强算法. 实验结果表明, 本文算法相对于传统算法, 有更好的消噪效果, 在多项客观语音质量评价指标上都有明显的改进, 在主观听觉实验中也取得了更好的听觉实验结果.

References

1 Ephaim Y, Van Trees H L. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 1995, **3**(4): 251−266

2 Hansen P C, Jensen S H. Prewhitening for rank-deficient noise in subspace methods for noise reduction. *IEEE Transactions on Signal Processing*, 2005, **53**(10): 3718−3726

3 You C H, Rahardja S, Koh S N. Audible noise reduction in eigendomain for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, **15**(6): 1753−1765

4 Jabloun F, Champagne B. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 2003, **11**(6): 700−708

5 Virag N. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, 1999, **7**(2): 126−137

6 Udrea R M, Vizireanu N D, Ciochina S. An improved spectral subtraction method for speech enhancement using a perceptual weighting filter. *Digital Signal Processing*, 2008, **18**(4): 581−587

7 Anderson T W. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 1963, **34**(1): 122−148

8 Chen Xi-Ru. *Probability and Mathematical Statistics*. Hefei: University of Science and Technology of China Press, 2004. 102
(陈希孺. 概率论与数理统计. 合肥: 中国科学技术大学出版社, 2004. 102)

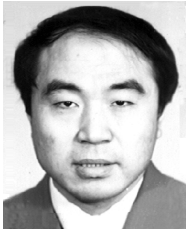
9 Hendriks R C, Jensen J, Heusdens R. Noise tracking using DFT domain subspace decompositions. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, **16**(3): 541−553

10 Sullivan T. CMU microphone array database [Online], available: http://www.speech.cs.cmu.edu/databases/micarray, August 12, 2008

11 Hansen J H L, Pellom B. An effective evaluation protocol for speech enhancement algorithms. In: Proceedings of the 5th International Conference on Spoken Language Processing. Sydney, Australia: ISCA, 1998. 2819−2822

12 McCowan I A, Bourlard H. Microphone array post-filter based on noise field coherence. *IEEE Transactions on Speech and Audio Processing*, 2003, **11**(6): 709−716

程 宁　中国科学院自动化研究所博士研究生. 主要研究方向为语音信号处理. 本文通信作者.
E-mail: kinchengning@hotmail.com
(**CHENG Ning**　Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. His main research interest is speech signal processing. Corresponding author of this paper.)



刘文举　中国科学院自动化研究所副研究员. 主要研究方向为语音识别, 语音信号处理. E-mail: lwj@nlpr.ia.ac.cn
(**LIU Wen-Ju**　Associate professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers speech recognition and speech signal processing.)