基于帧间相关性的最大后验估计语音增强算法

欧世峰, 赵晓晖
(吉林大学通信工程学院信息科学实验室, 吉林长春 130012)

摘 要: 通过讨论纯净语音分量的概率分布特征以及相邻分量间的统计相关特性, 在自适应 K-L 变换(KLT, Karhunen-Loève Transform)域给出了一种新的语音信号统计模型, 然后基于该信号模型, 利用最大后验(MAP, Maximum a Posterior)估计理论提出了一种新型的单通道语音增强算法. 该算法充分考虑到在 KLT 域相邻时刻语音分量间存在的相关信息, 利用信号的高斯模型假设条件, 以联合概率密度函数的形式将这种相关信息融合到 MAP 中, 获得纯净语音分量的估计. 算法不仅结构简单利于实现, 且有效地避免了传统算法对语音分量估计的不足. 仿真结果表明本文算法在客观和主观测试中都具有较好的语音增强效果.

关键词: 语音增强; 自适应 KLT; 相关; 最大后验估计
中图分类号: TN912 文献标识码: A 文章编号: 0372-2112 (2007) 10-2007-07

MAP Estimation for Noisy Speech Enhancement Based on Inter-Frame Correlation

OU Shi-feng, ZHAO Xiao-hui
(Laboratory of Information Science, College of Communication Engineering, Jilin University, Changchun, Jilin 130012, China)

Abstract: This paper investigates the probability distribution of speech components as well as the correlation characteristic between adjacent components, and presents a new speech model for enhancing noisy speech in adaptive KLT domain. Based on this model, a novel speech enhancement algorithm using MAP estimation is proposed, which incorporates the inter-frame correlation information as a form of joint probability density function into MAP under Gaussian model assumption for speech and noise components. The obtained estimation result keeps simple and avoids deficiency of classic approaches in enhancing noisy speech. In simulations with speech signals degraded by various noises, the proposed algorithm shows improved performance for a number of objective and subjective measures.

Key words: speech enhancement; adaptive Karhunen-Loève transformation; correlation; maximum a posterior estimation

1 引言

语音增强技术是噪声背景下进行语音识别和参数编码不可缺少的一部份, 它在解决语音噪声污染问题、改进语音通信质量、提高语音可懂度等方面发挥着重要的作用. 目前, 基于单通道输入的语音增强算法正在被广泛地应用于当今的通信系统中, 这些方法简单有效且利于硬件实现, 代表性的算法有谱减法、卡尔曼滤波法、维纳滤波算法等[1-3], 其中基于信号统计模型的语音增强算法由于其去噪性能的优越性, 长期以来都是国内外学者研究的热点[4-6]. 该类算法多利用高斯模型作为假设分布条件, 通过离散傅立叶变换(DFT, Discrete Fourier Transform)在频域中设计有效的算法来获取对纯净语音频谱各个分量的估计. 通常所用的估计方法有最小均方误差(MMSE, Minimum Mean-Square Error)估计、MAP 估计等. 相对于 MMSE 估计, 由于 MAP 估计方法可以避免算法中复杂的指数运算, 且具有较为出色的语音增强效果, 近年来得到了广泛的研究与应用[7,8]. 随着信号处理技术的发展, 统计模型语音增强算法在其他变换域中的研究也得到了人们的普遍关注. Soon 利用离散余弦变换(DCT, Discrete Cosine Transform)结合信号的高斯模型假设得到了一种维纳滤波算法[10]. Gazor 通过对语音信号进行 DCT 和 KLT 变换, 利用 MMSE 估计和最大似然(ML, Maximum Likelihood)估计分别得到了一种非线性滤波算法[11]. 相对于 DFT 变换, 由于 DCT(或 KLT)变换具有信号合成效果好、能量高度压缩, 且无需对语音相位进行估计等优点[10,12], 这些算法在保持低运算复杂度的情况下均取得了较为理想的语音增强效果.

收稿日期: 2005-11-01; 修回日期: 2007-06-26

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

但在语音信号的处理中, 帧间重叠的使用以及语音信号自身的相关特性, 导致了相邻时刻间语音分量存在着较强的相关性[13], 而且由于计算中语音帧的实际使用长度有限, 频率域相邻点间的语音分量也可能是统计相关的[14]. 上述统计模型语音增强算法大都假设在变换域中语音分量间为独立同分布信号, 它们在对纯净语音分量的估计中只利用了当前帧带噪语音分量对其的约束关系, 在一定程度上这固然简化了信号模型, 但却没有考虑相邻帧语音分量间存在的相关信息, 因而往往与实际情况不符. 本文详细分析了自适应KLT 变换域中纯净语音分量的统计模型, 并验证了相邻帧间语音分量的相关性, 然后将这种帧间相关信息以联合概率密度函数的形式融合到MAP 估计中, 利用语音与噪声的高斯模型假设条件, 推导出了一种新型的单通道语音增强算法, 最后通过仿真实验表明了本文算法的有效性能.

2 自适应 KLT 语音增强算法

用 $\boldsymbol{y}(n)$, $\boldsymbol{x}(n)$和 $\boldsymbol{v}(n)$分别表示 n 时刻K 维带噪语音、纯净语音和加性噪声向量, 假设纯净语音信号与噪声互不相关, 有

$$\boldsymbol{y}(n)=\boldsymbol{x}(n)+\boldsymbol{v}(n),\ E\{\boldsymbol{x}(i)\boldsymbol{v}^{\mathrm{T}}(j)\}=\boldsymbol{0},\ \forall i,j \tag{1}$$

运用矩阵 $\boldsymbol{U}^{\mathrm{T}}(n)$对其分别进行 KLT 变换, 可得

$$\begin{aligned}&\boldsymbol{Y}(n)=\boldsymbol{U}^{\mathrm{T}}(n)\boldsymbol{y}(n),\ \boldsymbol{X}(n)=\boldsymbol{U}^{\mathrm{T}}(n)\boldsymbol{x}(n),\\&\boldsymbol{V}(n)=\boldsymbol{U}^{\mathrm{T}}(n)\boldsymbol{v}(n)\end{aligned} \tag{2}$$

其中 $\boldsymbol{Y}(n)$, $\boldsymbol{X}(n)$与 $\boldsymbol{V}(n)$分别表示信号 $\boldsymbol{y}(n)$, $\boldsymbol{x}(n)$和 $\boldsymbol{v}(n)$在 n 时刻的 KLT 变换向量, $\boldsymbol{U}^{\mathrm{T}}(n)$为 KLT 变换矩阵. 由于 $\boldsymbol{x}(n)$与 $\boldsymbol{v}(n)$实际上并不能获知, 计算中可将通过带噪语音 $\boldsymbol{y}(n)$获得的变换矩阵 $\boldsymbol{U}^{\mathrm{T}}(n)$近似为信号 $\boldsymbol{x}(n)$与 $\boldsymbol{v}(n)$的 KLT 变换矩阵[9].

实现式(2)变换需要对 KLT 变换矩阵 $\boldsymbol{U}^{\mathrm{T}}(n)$的精确估计. 一般所用的方法是通过批处理技术计算信号的自相关矩阵, 然后利用特征分解直接获取 $\boldsymbol{U}^{\mathrm{T}}(n)$的估计[15]. 但该方法计算量较大, 且也不能实时地跟踪语音信号的非平稳变化特性. 考虑算法的估计精度和信号的非平稳特性, 文献[9]采用一种自适应迭代算法来逐个估计信号的 KLT 变换向量. 定义能量约束函数 $J\{\boldsymbol{u}(n)\}$

$$J\{\boldsymbol{u}(n)\}=\sum_{i=1}^{n}\beta^{n-i}\,\|\boldsymbol{y}(i)-\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{T}}(i-1)\boldsymbol{y}(i)\|^2 \tag{3}$$

其中 $\boldsymbol{u}(n)$为一 K 维列向量, $0\leqslant\beta\leqslant1$ 为遗忘因子. 经证明 $J\{\boldsymbol{u}(n)\}$具有唯一的全局最小值, 且此最小值恰好收敛于自相关矩阵 $\boldsymbol{R}_y(n)=\sum_{i=1}^{n}\beta^{n-i}\boldsymbol{y}(i)\boldsymbol{y}^{\mathrm{T}}(i)$的主 KLT 变换向量上. 因此, 通过最小化式(3)的能量约束函数, 利用投影压缩技术可以逐个获取信号的 KLT 变换向量, 其具体迭代步骤可参考文献[9].

前文式(1)中的加性噪声模型经上述自适应 KLT 变换后可以写成

$$Y_k(n)=X_k(n)+V_k(n),\ E\{X_k(i)V_k(j)\}=0,\ \forall i,j \tag{4}$$

其中 $Y_k(n)$, $X_k(n)$与 $V_k(n)$分别表示向量 $\boldsymbol{Y}(n)$, $\boldsymbol{X}(n)$与 $\boldsymbol{V}(n)$的第 k 个KLT 变换分量($k=0,1,\cdots,K-1$). 语音增强的任务即是设计算法从这些带噪语音分量中估计出纯净语音分量 $X_k(n)$, 然后再通过 KLT 反变换 $\boldsymbol{U}(N)$实现对原始语音信号 $\boldsymbol{x}(n)$的恢复. 假设语音分量 $X_k(n)$为独立同分布信号, 其数值大小只依赖于当前帧带噪语音分量 $Y_k(n)$, 则可得文献[9]中算法对 $X_k(n)$的估计为

$$\hat{X}_k(n)=\frac{\sigma^2_{k,X}}{\sigma^2_{k,X}+\sigma^2_{k,V}}Y_k(n) \tag{5}$$

其中 $\sigma^2_{k,X}=E\{X_k(n)\}$, $\sigma^2_{k,V}=E\{V_k(n)\}$分别表示 $X_k(n)$与 $V_k(n)$的方差.

3 语音分量统计模型

在给出本文算法对语音分量 $X_k(n)$的估计之前, 本节将首先讨论自适应KLT 变换域中纯净语音分量$X_k(n)$的概率分布特征以及在相邻时刻间(即相邻帧间$[X_k(n), X_k(n+1)]$)和相邻频点间$[X_k(n), X_{k+1}(n)]$语音分量的统计相关特性, 并根据讨论结果归纳出纯净语音分量 $X_k(n)$在自适应KLT 变换域中的统计模型.

考虑取自 http://www.dailywav.com 的 4 段纯净语音信号(2 段男声, 2 段女声), 去除静音部分并进行8kHz 采样后, 运用文献[9]中的迭代算法对其进行自适应 KLT 变换. 参数选取采用文献[9]中的仿真条件, 语音帧长度为$K=20$, 帧间重叠 $K-h$ 个采样点, 参数 h 表示每次信号迭代中更新的采样点数. 图 1 给出的是自适应 KLT 变换域中纯净语音分量 $X_k(n)$的概率分布曲线(不失一般性, 该图综合了上述 4 段语音的平均分布结果). 同时, 图中还相应地给出了高斯和拉普拉斯两种常用模型的分布曲线. 从图 1 结果中可以看出, 两种模型的概率分布曲线均只能部分准确地描绘语音分量 $X_k(n)$的实际分布特征: 在幅值较小的区域即零值附近, $X_k(n)$的分布曲线趋近于拉普拉斯分布, 而在幅值较大的区域, 这一曲线则更加符合高斯模型分布. 众所周知, 语音信号的能量主要集中在幅值较大的分量上, 符合高斯模型分布的那部分分量代表着语音信号的大部分特征, 而且在大量背景噪声干扰的情况下, 零值附近的语音分量基本上已完全湮没于噪声信号中, 语音增强技术大都依赖信号模型对那些代表语音信号大部分特征的幅值较大的分量进行恢复. 因此, 考虑以上两

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

点因素, 在自适应KLT 变换域, 本文假设纯净语音序列 $\{X_k(n),\ n=1, 2, \cdots\}$ 为零均值的高斯分布过程.

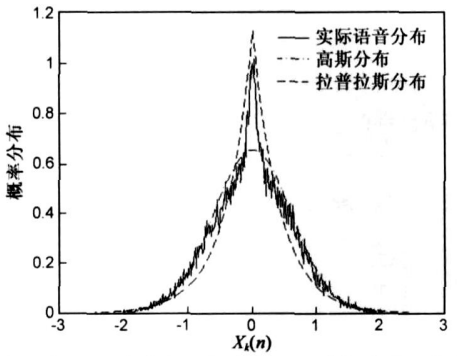

图 1　自适应 KLT 域纯净语音分量 $X_k(n)$ 的概率分布曲线 (k=4)

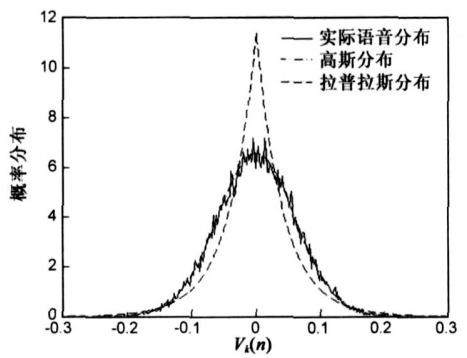

图 2　自适应 KLT 域噪声分量 $V_k(n)$ 的概率分布曲线 (k=1)

图 2 给出的是自适应 KLT 变换域中噪声信号分量 $V_k(n)$ 的概率分布曲线, 噪声信号类型为白噪声. 从图中不难看出, 在整个分布区域上, 噪声分量 $V_k(n)$ 都严格服从于零均值的高斯模型分布.

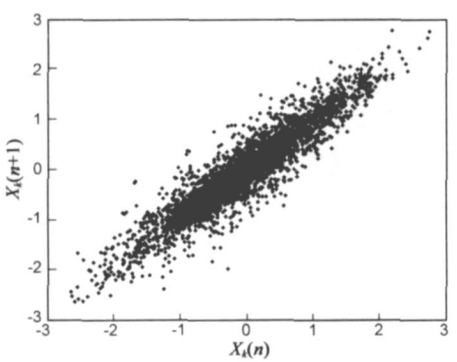

图 3　相邻时刻间纯净语音分量 $[X_k(n), X_k(n+1)]$ 的幅值分散图 (k=2)

为研究自适应 KLT 变换域中纯净语音信号相邻分量间的统计相关特性, 图 3、图 4 分别给出了相邻时刻和相邻频点间语音分量 $[X_k(n),\ X_k(n+1)]$ 以及 $[X_k(n), X_{k+1}(n)]$ 的幅值分散图, 两图所用信号为同一段去除静音部分的纯净语音, 参数选择仍为文献[9]中仿真条件. 它们的横轴均为 $X_k(n)$, 纵轴分别为 $X_k(n+1)$ 和 $X_{k+1}(n)$. 对比图 3、图 4 可以看出, 相邻时刻间语音分量的幅值分布具有较强的约束关系, $X_k(n)$ 数值的大小在很大程度上要影响着 $X_k(n+1)$ 的分布; 而相邻频点间语音分量的幅值则呈现出无规则随机分布状态, 其纵、横轴数据间的相关因素并不明显.

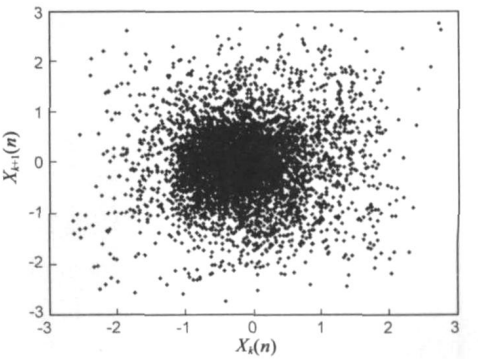

图 4　相邻频率点间纯净语音分量 $[X_k(n), X_{k+1}(n)]$ 的幅值分散图 (k=2)

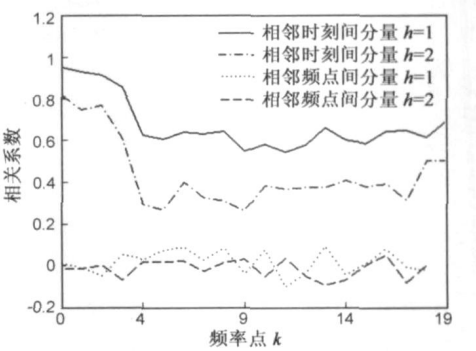

图 5　相关系数随参数 k 的变化曲线图

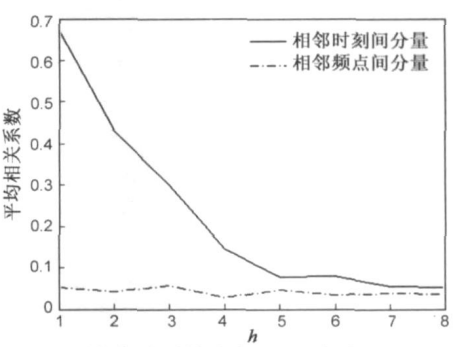

图 6　平均相关系数随参数 h 的变化曲线图

为准确衡量语音信号分量 $[X_k(n), X_k(n+1)]$ 以及 $[X_k(n), X_{k+1}(n)]$ 间相关特性的大小, 定义它们的相关系数 $\rho_t(k)$ 和 $\rho_f(k)$ 分别为

$$\rho_t(k) = \frac{N}{N-1}\frac{\sum_{n=1}^{N-1}\{X_k(n)\cdot X_k(n+1)\}}{\sum_{n=1}^{N}X_k^2(n)} \tag{6}$$

$$\rho_f(k) = \frac{\sum_{n=1}^{N}\{X_k(n)\cdot X_{k+1}(n)\}}{\{\sum_{n=1}^{N}X_k^2(n)\cdot\sum_{n=1}^{N}X_{k+1}^2(n)\}^{0.5}} \tag{7}$$

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中 N 表示语音帧总数. 图 5 给出的是帧间重叠样点数 $K-h$ 分别为 19 和 18, 即每次信号迭代更新采样点数 h 为 1 和 2 时, 不同分量间相关系数随频率点 k 变化的曲线; 通过在所有频率点上对相关系数 $\rho_T(k)$ 和 $\rho_F(k)$ 进行平均, 得到图 6 中平均相关系数 $\bar{\rho}_T$ 以及 $\bar{\rho}_F$ 随着参数 h 的变化曲线, $\bar{\rho}_T$ 和 $\bar{\rho}_F$ 的计算公式分别表示如下

$$\bar{\rho}_T = \frac{1}{K}\sum_{k=0}^{K-1} |\rho_T(k)| \tag{8}$$

$$\bar{\rho}_F = \frac{1}{K-1}\sum_{k=0}^{K-2} |\rho_F(k)| \tag{9}$$

结合图 5、图 6 可以看出, 自适应 KLT 域中相邻时刻间的语音分量并不是相互独立的, 它们在每个频率点上都存在着明显的依赖关系, 其平均相关系数 $\bar{\rho}_T$ 的大小取决于参数 h, 当 h 逐渐减小, 即帧间叠接样点数 $K-h$ 逐渐增加时, $\bar{\rho}_T$ 的数值会得到明显的增加. 相对而言, 相邻频点间语音分量的相关系数则是随机地分布于零点附近, 数值较小且其平均相关系数 $\bar{\rho}_F$ 的大小并没有因 h 的改变而有所变化, 这是由于在频率域中, 不同频率成份之间一般是不存在相互关系的, 而且, 如果分量 $[X_k(n), X_{k+1}(n)]$ 间存在着一定程度的相关因素, 那么当帧间重叠样点数 $K-h$ 有所增加或减小时, 其平均相关系数 $\bar{\rho}_F$ 的数值理应产生相应的变化, 而这恰恰与图 6 的实验结果相违背, 因此考虑到这两点因素, 本文仍假设相邻频点间的语音分量是彼此独立的.

综合以上讨论结果, 本文假设自适应 KLT 变换域中纯净语音分量 $X_k(n)$ 的统计模型如下:

(1) 相邻频点间语音分量 $X_k(n)$ 与 $X_{k+1}(n)$ 统计上相互独立;

(2) 相邻时刻间语音分量 $X_k(n)$ 与 $X_k(n+1)$ 高度相关, 其幅值间存在着较强的相关性;

(3) 语音分量序列 $\{X_k(n),\ n=1, 2, \cdots\}$ 以及噪声分量序列 $\{V_k(n),\ n=1, 2, \cdots\}$ 均为零均值的高斯分布过程.

4 帧间相关 MAP 语音增强算法

基于前节中 $X_k(n)$ 的统计模型, 本文综合考虑相邻时刻间纯净语音分量的相关信息以及本帧带噪语音对纯净语音的约束关系, 尝试利用分量 $X_k(n-1)$ 和 $Y_k(n)$ 来联合获取当前帧纯净语音分量 $X_k(n)$ 的估计 $\hat{X}_k\{n|Y_k(n), X_k(n-1)\}$. 为方便起见, 本节在以下的公式推导中省略了时刻与频率标记 n 和 k, 并用符号 $\tilde{X}$ 来表示前一帧语音分量 $X_k(n-1)$.

定义 $\hat{X}$ 的估计误差 ε 及代价函数 $d(\varepsilon)$ 分别为 $\varepsilon = X-\hat{X}$ 和 $d(\varepsilon) = d(X, \hat{X})$, 则估计误差 ε 的贝叶斯风险函数 R 为[16]

$$\begin{aligned} R &= E\{d(X,\hat{X})\} = \iint d(X,\hat{X})P(X,z)\,dz\,dX \\ &= \int\left\{\int d(X,\hat{X})\,P(X|z)\,dX\right\}P(z)\,dz \end{aligned} \tag{10}$$

其中 $z=[Y,\tilde{X}]$, $P(X,z)$ 与 $P(X,z)$ 分别表示 (X,z) 的联合概率密度函数和 X 的条件概率密度函数. 通过最小化上式的贝叶斯风险函数, 可实现代价函数 $d(\varepsilon)$ 条件下 X 的估计. 令 $d(X,\hat{X}) = (X-\hat{X})^2$, 即当 $R=E(\varepsilon^2)$ 为 ε 的均方误差函数时, 可得到 X 的 MMSE 估计 $\hat{X} = \arg\min_X R = E(X|z)$, 但该估计要涉及到复杂的积分和指数运算, 计算量要求较大不利于实现, 为此本文考虑 $d(\varepsilon)$ 的另一种函数形式

$$d(\varepsilon) = \begin{cases} 0 & |\varepsilon| < \delta \\ 1 & |\varepsilon| > \delta \end{cases} \tag{11}$$

其中 $\delta>0$ 表示一任意小正数. 上式即是所谓的"hit-or-miss"函数, 其具体性质可参考文献[16]. 将 $d(\varepsilon) = d(X,\hat{X})$ 代入式(10), 通过最小化 R, 可得到 X 的 MAP 估计为

$$\hat{X} = \arg\min_X R = \arg\max_X P(X|Y,\tilde{X}) \tag{12}$$

其中 $P(X|Y,\tilde{X})$ 表示 X 在条件 $(Y,\tilde{X})$ 下的概率密度函数. 出于简化计算的目的, 本文只考虑相邻两帧中纯净语音信号的相关信息, 仍然假设其中的带噪语音是相互独立的, 利用贝叶斯定理对 $P(X|Y,\tilde{X})$ 进行化简, 可得

$$\begin{aligned} P(X|Y,\tilde{X}) &= \frac{P(Y,\tilde{X}|X)P(X)}{P(Y,\tilde{X})} \\ &= \frac{P(Y|X)P(X,\tilde{X})}{P(Y,\tilde{X})} \end{aligned} \tag{13}$$

将上式代入式(12), 有

$$\hat{X} = \arg\max_X\{P(Y|X)P(X,\tilde{X})\} \tag{14}$$

式(13)分母中的联合概率密度函数 $P(Y|\tilde{X})$ 相对于 X 为一常数, 故在上式中可以将其省略. 根据前文的假设条件, 令纯净语音分量 X 与噪声分量 V 皆为零均值的高斯分布过程, 其概率密度函数 $P(X)$, $P(V)$ 可分别写为

$$P(X) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(\frac{-X^2}{2\sigma_X^2}\right) \tag{15}$$

$$P(V) = \frac{1}{\sqrt{2\pi}\sigma_V}\exp\left(\frac{-V^2}{2\sigma_V^2}\right) \tag{16}$$

其中 σ_X^2, σ_V^2 分别表示 X 与 Y 的方差. 在式(4)的加性噪声模型下, 纯净语音分量与噪声分量互不相关, 且又都服从高斯分布, 故 X 与 V 相互独立, 结合以上两式得带噪语音分量 Y 在条件 X 的概率密度函数 $P(Y|X)$ 为

$$P(Y|X) = \frac{2}{\sqrt{2\pi}\sigma_V}\exp\left(\frac{-(Y-X)^2}{2\sigma_V^2}\right) \tag{17}$$

定义向量 $\boldsymbol{Q}=[X,\tilde{X}]^T$, 计算其协方差矩阵 $\boldsymbol{R}$ 为

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\boldsymbol{R} = E[\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}] = \begin{bmatrix} E[X^2] & E[X\tilde{X}] \\ E[X\tilde{X}] & E[\tilde{X}^2] \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_{\tilde{X}} \\ \rho\sigma_X\sigma_{\tilde{X}} & \sigma_{\tilde{X}}^2 \end{bmatrix} \tag{18}$$

其中 ρ 为 X 与 $\tilde{X}$ 的相关系数, $\sigma_{\tilde{X}}^2$ 为变量 $\tilde{X}$ 的方差. 由于相邻帧中纯净语音分量的能量变化比较缓慢, 且 X 与 $\tilde{X}$ 均服从相同参数的高斯分布, 实际计算中 $\sigma_{\tilde{X}}^2$ 可近似等于 σ_X^2. 根据概率理论, 可得 X 与 $\tilde{X}$ 的联合概率密度函数 $P(X, \tilde{X})$ 为

$$P(X, \tilde{X}) = \frac{1}{2\pi |\boldsymbol{R}|^{0.5}} \exp\left(-\frac{1}{2}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Q}\right) \tag{19}$$

将式(17)和式(19)代入式(14), 最终可得本文算法对纯净语音分量 X 的估计为

$$\begin{aligned} \hat{X} &= \arg\max_X \{\ln[P(Y|X)P(\tilde{X}, X)]\} \\ &= \frac{\sigma_V^2\rho\tilde{X} + \sigma_X^2(1-\rho^2)Y}{\sigma_V^2 + \sigma_X^2(1-\rho^2)} \end{aligned} \tag{20}$$

将上式估计结果同式(5)比较不难看出, 如不考虑存在于相邻帧间语音分量的相关信息, 仍假设 X 与 $\tilde{X}$ 是统计独立的, 则此时它们之间的相关系数 $\rho = 0$, 式(20)估计即退化为文献[9]中算法.

实现本文算法需要参数 ρ, σ_V^2 和 σ_X^2 的先验知识, 对于 σ_V^2, σ_X^2 本文采用文献[9]中算法对其进行估计, 具体过程在此不再加以详述. 对于相关系数 ρ, 由于 $\tilde{X}$ 与 V 相互独立, 有 $E[X\tilde{X}] = E[Y\tilde{X}]$, 故 ρ 的计算公式可以写为

$$\rho = \frac{E[X\tilde{X}]}{E[X^2]} = \frac{E[Y\tilde{X}]}{E[X^2]} \tag{21}$$

这里的 $E[Y\tilde{X}]$ 可通过下式对其加以估计

$$E[Y\tilde{X}] = \alpha\tilde{E}[Y\tilde{X}] + (1+\alpha)Y\tilde{X} \tag{22}$$

其中 $\tilde{E}[Y\tilde{X}]$ 表示算法在前一帧中对 $E[Y\tilde{X}]$ 的估计, α 为控制系数, 在本文仿真中其值选取为 0.95.

5　仿真实验与结果分析

为验证本文提出算法的语音增强性能和效果, 仿真中将其输出语音与文献[9]中算法进行了对比. 实验所用语音数据为四段纯净的语音信号, 背景噪声选择四种不同类型的噪声信号, 依次为白噪声、F16 驾驶舱内噪声(F16)、驱逐舰引擎噪声(Destroyer)以及海盗号驾驶舱内噪声(Buccaneer), 它们均取自 http://spib.ece.rice.edu/. 噪声与语音信号的采样频率均为 8kHz. 将不同噪声信号叠加到纯净语音信号上, 分别产生输入信噪比为 0dB、5dB 和 10dB 的带噪语音信号. 考虑运算复杂度和算法性能的矛盾关系, 仿真中参数选取同文献[9]: 语音帧长度为 $K = 20$, 帧间重叠 $K-1$ 个采样点. 由于无法直接获取前帧中纯净语音分量 $\tilde{X}$, 实际处理中采用经本文算法处理后的前一帧语音分量来近似式(20)中的 $\tilde{X}$.

首先, 将通过文献[9]中算法和本文算法处理后的语音信号在语谱图上进行对比. 图 7 给出的是白噪声背景下几种语音信号的语谱图, 其中带噪语音信号的输入 *SNR* 为 0*dB*. 从四幅仿真结果图中可以看出: 两种算法均可以有效地抑制背景噪声, 但相对而言, 本文提出算法的噪声消除性能要更优一些; 文献[9]算法输出语音中残留着较多的"音乐噪声", 而且这些"音乐噪声"随机地分布于信号的语谱之中, 故将会对人们的主观接受程度产生较大的影响; 本文算法输出语音在消除大部分噪声干扰的同时, 并没有产生明显的"音乐噪声", 而且由于相关系数 ρ 所起到的平滑作用, 本文算法在语谱图中对语音频率分量的损伤程度要略小于文献[9]中算法.

为了准确衡量两种算法输出语音信号的失真程度以及残余噪声的大小, 采用帧间信噪比(SegSNR)与对数谱距离(LSD)两种测量标准来评估算法的去噪性能[17], 在对于它们的计算过程中, 本文只考虑其 SNR 大于-10dB 而小于 35dB 的语音帧. 表 1、表 2 分别给出的是经两种算法处理后语音信号 SegSNR 以及 LSD 的对比. 从中不难看出, 在各种噪声背景和不同输入 SNR 下, 本文算法使用两种评价标准的结果都要好于文献

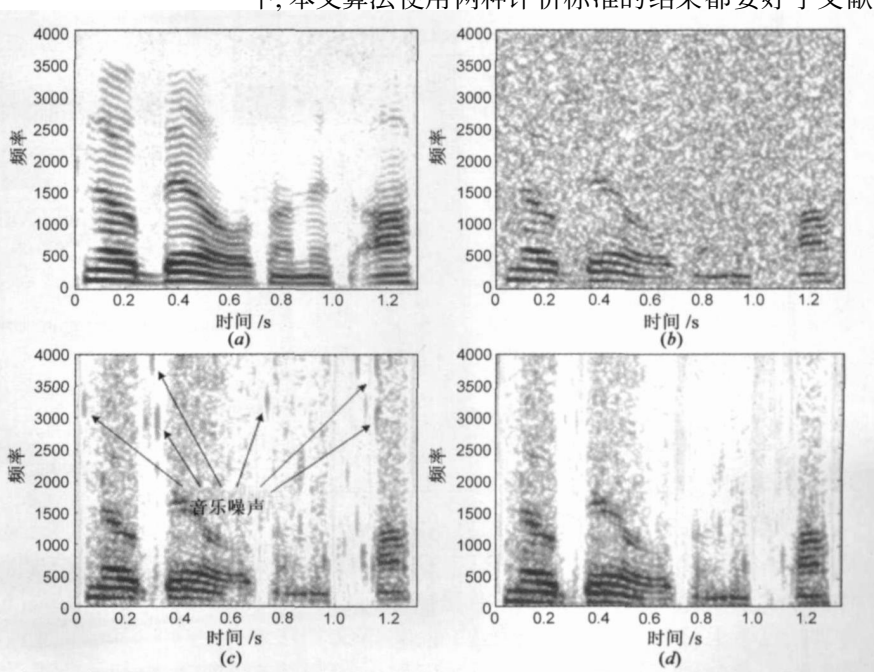

图 7　白噪声背景下几种语音信号语谱图(输入 SNR=0dB), 图(*a*)~(*d*)依次为: 纯净语音、带噪语音、文献[9]中算法增强后的语音和本文算法增强后的语音

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

[9] 中算法.

表 1 噪声背景下两种算法输出语音 SegSNR(dB) 对比表

噪声类型	输入 SNR	输入 SegSNR	输出 SegSNR	
			[9] 中算法	本文算法
白噪声	0dB	-0.68	4.42	4.65
	5dB	2.06	6.70	7.18
	10dB	6.50	9.48	9.99
F16	0dB	-0.36	3.32	3.39
	5dB	2.99	5.35	5.51
	10dB	6.73	8.08	8.25
Destroyer	0dB	-0.65	3.40	3.74
	5dB	2.73	5.66	5.99
	10dB	6.55	8.40	8.72
Buccaneer	0dB	-0.48	3.85	4.07
	5dB	2.88	6.10	6.34
	10dB	6.57	8.76	8.99

表 2 噪声背景下两种算法输出语音 LSD(dB) 对比表

噪声类型	输入 SNR	输入 LSD	输出 LSD	
			[9] 中算法	本文算法
白噪声	0dB	6.73	5.34	5.14
	5dB	5.77	4.08	3.95
	10dB	4.84	3.51	3.40
F16	0dB	5.75	5.49	5.33
	5dB	4.95	4.64	4.45
	10dB	4.24	3.74	3.65
Destroyer	0dB	5.89	5.33	5.19
	5dB	5.01	4.19	4.09
	10dB	4.23	3.49	3.40
Buccaneer	0dB	6.42	5.71	5.45
	5dB	5.44	4.46	4.31
	10dB	4.74	3.86	3.76

表 3 测试中本文算法输出语音信号的被认可度表

噪声类型	输入 SNR	与其他语音的对比	
		未处理语音	[9] 输出语音
白噪声	0dB	80%	83%
	5dB	85%	79%
	10dB	96%	78%
F16	0dB	76%	77%
	5dB	83%	75%
	10dB	88%	75%
Destroyer	0dB	83%	82%
	5dB	89%	77%
	10dB	95%	75%
Buccaneer	0dB	85%	81%
	5dB	93%	82%
	10dB	96%	78%

主观测试选用文献[9] 中所用方法, 25 位试听者均是来自通信工程学院的学生, 他们中没有人接触过语音信号处理领域中的问题. 测试语音在不同信噪比下分别经过了上述两种增强算法的处理, 将本文算法增强后的语音分别与原始带噪语音和文献[9] 中算法增强后的语音进行对比, 要求试听者从中选择自己所偏爱的一种, 从而得到本文语音增强算法在不同 SNR 与噪声背景下的被认可百分比例. 实验结果如表 3 所示, 从中可以看出, 本文算法在总体上被测试者接受的比例要明显高于文献[9] 中算法. 在较低信噪比处, 试听者对未经处理语音认可的人数占有少许比例, 这主要是因为算法本身对语音信号存在着损伤, 进而带来语音信号在一定程度上的失真造成的, 但随着输入 SNR 的提高, 这种失真的程度逐渐减少, 本文算法的被认可比例得到显著提高.

在计算复杂度方面, 相对于文献[9] 中算法, 本文算法主要增加了对相关系数进行估计的计算, 运算量增加不大, 二者处于同一量级.

6 结束语

基于相邻帧间纯净语音分量的相关信息和 MAP 估计, 本文提出了一种新型的单通道语音增强算法. 它通过对信号分量的高斯模型假设, 在自适应 KLT 域将相邻两帧间语音分量的相关信息融合到 MAP 方法中, 来获取纯净语音分量的估计, 算法简单有效. 仿真实验利用四种背景噪声对本文算法性能进行了评估, 结果显示, 相对于文献[9] 中提出的高性能语音增强算法, 本文算法在主观和客观测试中都具有更好的语音增强效果.

参考文献:

[1] S F Boll. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans Acoust, Speech, Signal Process, 1979, 27(2): 113-120.

[2] S Gannot, D Burshtein, E Weinstein. Iterative and sequential Kalman filter-based speech enhancement algorithms[J]. IEEE Trans Speech Audio Process, 1998, 6(4): 373-385.

[3] I Y Soon, S N Koh. Speech enhancement using 2-D Fourier transform[J]. IEEE Trans Speech Audio Process, 2003, 11(6): 717-724.

[4] Y Ephraim, D Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. IEEE Trans Acoust, Speech, Signal Process, 1984, 32(6): 1109-1121.

[5] R Martin. Noise power spectral density estimation base on optimal smoothing and minimum statistics[J]. IEEE Trans Speech Audio Process, 2001, 9(4): 504-512.

[6] C H You, S N Koh, S Rahardja. Beta-order MMSE spectral amplitude estimation for speech enhancement[J]. IEEE Trans Speech Audio Process, 2005, 13(4): 475-486.

[7] P Wolfe, S Godsill. Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement[A]. In Proc 11th IEEE Workshop on Statistical Signal Process[C], Singa-

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

pore: IEEE, 2001. 496– 499.

[8] T H Dat, K Takeda, F Itakura. Generalized Gamma modeling of speech and its online estimation for speech enhancement[A]. In Proc. IEEE ICASSP-05[C]. Philadelphia, USA: IEEE, 2005. 181– 184.

[9] A Rezayee, S Gazor. An adaptive KLT approach for speech enhancement[J]. IEEE Trans Speech Audio Process, 2001, 9(2): 87– 95.

[10] I Y Soon, S N Koh, C K Yeo. Noisy speech enhancement using discrete cosine transform[J]. Speech Commun, 1998, 24(3): 249– 257.

[11] S Gazor. Employing Laplacian-Gaussian densities for speech enhancement[A]. In Proc. IEEE ICASSP-04[C]. Montreal, Canada: IEEE, 2004. 297– 300.

[12] J H CHANG. Warped discrete cosine transform-based noisy speech enhancement[J]. IEEE Trans. Circuits and Sys. Ⅱ, Exp. Briefs, 2005, 52(9): 535– 539.

[13] I Cohen. Relaxed statistical model for speech enhancement and a priori SNR estimation[J]. IEEE Trans Speech Audio Process., 2005, 13(5): 870– 881.

[14] C Li, S V Andersen. Inter-frequency dependency in MMSE speech enhancement[A]. in Proc the 6th Nordic Signal Process Symposium[C]. Espoo, Finland: IEEE, 2004. 200– 203.

[15] Y Ephraim, H L Van. A signal subspace approach for speech enhancement [J]. IEEE Trans Speech Audio Process, 1995, 3(4): 251– 266.

[16] S Kay. Fundamentals of Statistical Signal Processing: Estimation Theory[M]. Upper Saddle River, New Jersey: Prentice-Hall, 1993.

[17] S R Quackenbush, T P Barnwell, M A Clements. Objective Measures of Speech Quality[M]. Englewood Cliffs, New Jersey: Prentice-Hall, 1988.

作者简介:

欧世峰　男, 1979 年 12 月出生于山东巨野, 吉林大学通信工程学院博士研究生. 主要研究方向为语音信号处理与盲信号处理.

E-mail: ousfeng@126.com

赵晓晖　男, 1957 年 11 月出生于北京, 吉林大学通信工程学院教授, 博士生导师, 国内外发表学术论文 80 余篇. 主要研究方向为自适应信号处理理论及其在通信中的应用.

E-mail: xhzhao@jlu.edu.cn

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net