

基于超高斯激励的噪声顽健语音线性预测分析算法

周彬, 邹霞, 张雄伟, 赵改华

(解放军理工大学 指挥信息系统学院, 江苏 南京 210007)

摘要: 针对传统的语音信号线性预测分析算法在噪声环境下性能恶化的问题, 提出了一种新的基于超高斯激励的噪声顽健线性预测算法。该算法采用具有超高斯特性的学生 t 分布对语音信号线性预测激励建模, 并显式地考虑环境噪声的影响, 从而构建语音信号线性预测分析的概率图模型。在此基础上, 利用变分贝叶斯的方法求解模型参数的近似后验分布, 进而实现对带噪语音线性预测系数的最优估计。实验结果表明, 该算法能够有效提高噪声环境下语音信号线性预测分析的顽健性。

关键词: 语音线性预测; 噪声顽健; 超高斯激励; 变分贝叶斯

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-436X(2013)05-0062-09

Noise-robust linear prediction analysis of speech based on super-Gaussian excitation

ZHOU Bin, ZOU Xia, ZHANG Xiong-wei, ZHAO Gai-hua

(College of Command Information Systems, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: To overcome the problem that the performance of the traditional linear prediction (LP) analysis of speech deteriorates significantly in the presence of background noise, a novel algorithm for robust LP analysis of speech based on super-Gaussian excitation was proposed. The excitation noise of LP was modeled as a Student- t distribution, which was shown to be super-Gaussian. Then a novel probabilistic graphical model for robust LP analysis of speech was built by incorporating the effect of additive noise explicitly. Furthermore, variational Bayesian inference was adopted to approximate the intractable posterior distributions of the model parameters, based on which the LP coefficients of the noisy speech were estimated iteratively. The experimental results show that the developed algorithm performs well in terms of LP coefficients estimation of speech and is much more robust to ambient noise than several other algorithms.

Key words: linear prediction of speech; noise-robust; super-Gaussian excitation; variational Bayes

1 引言

语音信号的线性预测分析是现代语音信号处理中的核心技术之一, 其广泛应用于语音压缩编码、语音识别等领域。根据语音信号的产生机理和线性预测理论, 语音信号的产生过程可以用源—滤波模型表示。在经典的语音信号源—滤波模型中, 声源激励一般采用准周期脉冲串或者高斯白噪声近似表示, 而声道脉冲响应则采用全极点滤波器表示, 从而构成了语音信号的线性预测(LP)模型, 也

被称为自回归(AR, autoregression)模型。

线性预测系数的提取是语音信号线性预测分析的核心问题。传统的 LP 分析方法, 例如目前低速率语音编码中常用的自相关解法^[1], 虽然对纯净语音信号取得了较好的效果, 但是在噪声环境下其性能会急剧恶化。解决这一问题的常规方法是通过前端语音增强算法对带噪语音进行噪声抑制, 然后再对增强后的语音进行参数提取, 常用的前端增强算法包括对数幅度谱最小均方误差估计算法^[2]、超高斯语音模型短时谱估计算法^[3]、基于人耳掩蔽效

收稿日期: 2012-07-18; 修回日期: 2013-01-19

基金项目: 江苏省自然科学基金资助项目(BK2012510); 国家博士后科研基金资助项目(20090461424)

Foundation Items: The Natural Science Foundation of Jiangsu Province(BK2012510); The National Postdoctoral Foundation of China (20090461424)

应的谱减法^[4]等。但是，由于前端增强算法通常存在残余噪声并引入语音失真，采用这一方法难以有效地提高线性预测参数提取的准确度。因此，研究对噪声顽健的语音信号 LP 分析算法，从而提高线性预测参数提取的准确度具有重要意义。

文献[5]最早提出针对带噪语音的顽健线性预测分析，建立了带噪语音的全极点模型，并提出基于非因果维纳滤波的参数迭代优化算法。在此基础上，文献[6]提出利用 Gibbs 抽样的方法实现对 LP 系数的最小均方误差(MMSE)估计的算法，从而改进线性预测系数提取的准确度。文献[7]进一步提出利用线性预测系数和预测激励的先验信息，得到预测系数的最大后验概率(MAP)估计，以此提高参数提取的顽健性。文献[8]在 AR 模型的基础上，构建了针对带噪语音的状态空间方程，并采用变分贝叶斯的方法求解模型参数。需要注意的是，上述传统算法中都假设预测激励服从高斯分布，由于高斯分布的线性变换性质，这对应于原始语音信号也服从高斯分布^[9]。然而，实际的语音信号并不服从高斯分布。研究表明，语音信号在时域和频域服从超高斯分布，具有重拖尾和尖峰特性^[10]。同时，文献[9]的研究也表明，与高斯激励 AR 模型相比，超高斯激励 AR 模型具有更好的顽健性。文献[11]提出一种语音信号稀疏线性预测算法，利用语音信号预测残差的稀疏性进而改进参数提取的顽健性，实际上等价于一种超高斯激励贝叶斯线性预测模型，即在假设预测激励服从 Laplace 分布条件下对 LP 参数的最大后验概率估计。这些算法相对于传统的基于高斯激励 AR 模型的线性预测算法，在一定程度上提高了顽健性，但是由于没有显式地考虑环境噪声的影响，因而在噪声强度较大时仍存在性能恶化的问题。

针对上述问题，本文提出一种新的基于超高斯激励模型的带噪语音线性预测分析算法。采用一种具有超高斯特性的学生 t 分布对语音信号预测残差进行建模，同时显式地考虑加性噪声对信号的影响，构建噪声环境下语音线性预测分析的概率图模型。由于超高斯分布和加性噪声的引入，使得模型的复杂度提高，准确的模型参数推理难以实现，因此，本文利用变分贝叶斯的方法求解模型参数的近似后验分布，进而实现最大后验概率意义下的线性预测系数提取。实验结果表明，本文算法能够更准确地估计带噪语音的线性预测系数，从而有效地提

高噪声环境下语音信号线性预测分析的顽健性。

2 传统的语音信号线性预测分析算法

根据语音信号的产生机理和线性预测理论，语音信号可以用自回归模型近似表示，即，当前时刻取样值 $s(n)$ 可以通过过去若干时刻取样值的线性组合来表示，即

$$s(n) = \sum_{i=1}^p a_i s(n-i) + \varepsilon(n) = \mathbf{a}^T \mathbf{s}_{n-1} + \varepsilon(n) \quad (1)$$

其中， p 为线性预测阶数， $\mathbf{a} = \{a_1, a_2, \dots, a_p\}$ 为线性预测系数， $\mathbf{s}_{n-1} = \{s(n-1), s(n-2), \dots, s(n-p)\}$ 为 n 时刻之前的 p 个取样值组成的矢量， $\varepsilon(n)$ 为预测残差。在均方预测误差 $\varepsilon(n)$ 最小化的准则下，线性预测系数 \mathbf{a} 可通过 LP 正则方程求解^[1]。

$$\sum_{i=1}^p a_i \mathbf{R}(i, j) = \mathbf{R}(0, j), j = 1, 2, \dots, p \quad (2)$$

其中， \mathbf{R} 为语音信号 $s(n)$ 的自相关矩阵或自协方差矩阵，分别对应线性预测分析的自相关解法和自协方差解法。

上述求解线性预测系数的问题可以从概率估计的角度进行等价描述，即正则方程式(2)的求解等价于在假设预测误差 $\varepsilon(n)$ 为高斯白噪声的条件下，求解 \mathbf{a} 的最大似然(ML)估计，则

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} p(\mathbf{s} | \mathbf{a}) \quad (3)$$

其中， $\mathbf{s} = \{s(1), s(2), \dots, s(N)\}$ 表示当前帧语音信号， N 为帧长度， $p(\mathbf{s} | \mathbf{a})$ 表示给定线性预测系数 \mathbf{a} 时， \mathbf{s} 的条件概率密度函数，通常被称为似然分布函数。假设预测残差 $\varepsilon(n)$ 服从准确度为 τ (即方差为 τ^{-1}) 的零均值高斯分布，则 $p(\mathbf{s} | \mathbf{a})$ 可表示为

$$p(\mathbf{s} | \mathbf{a}) = \prod_{n=1}^N \text{Gau}(s(n); \mathbf{a}^T \mathbf{s}_{n-1}, \tau^{-1}) \quad (4)$$

若进一步假设 \mathbf{a} 的先验分布为 $p(\mathbf{a})$ ，则 \mathbf{a} 的最大后验概率(MAP)估计为^[5]

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} p(\mathbf{a} | \mathbf{s}) = \arg \max_{\mathbf{a}} \{p(\mathbf{s} | \mathbf{a}) p(\mathbf{a})\} \quad (5)$$

在假设线性预测残差为高斯白噪声的条件下，根据高斯分布的线性变换性质，原始语音也服从高斯分布。然而研究表明，实际的语音信号并不服从高斯分布，而更趋近于超高斯分布^[10]。因此，近年来学者提出了基于超高斯激励的线性预测分析算法，如文献[11]，理论分析和实验结果都表明其相对

于传统高斯激励线性预测模型具有更好的顽健性。

3 本文模型和算法

本节针对带噪语音的线性预测系数提取问题，提出了一种新的基于超高斯激励的线性预测分析模型及其求解算法。首先引入具有超高斯特性的学生 t 分布对语音信号线性预测残差建模，并考虑环境噪声的影响，建立带噪语音线性预测分析的概率图模型，进而利用变分贝叶斯的方法求解得到线性预测系数的最优估计。

3.1 学生 t 分布

本节首先引入一种超高斯分布模型—学生 t 分布。假设随机变量 x 服从学生 t 分布，其概率密度函数为^[12]

$$p(x) = St(x; \mu, \lambda, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(\frac{\lambda}{\nu\pi}\right)^{1/2} \left(1 + \frac{\lambda}{\nu}(x-\mu)^2\right)^{-(\nu+1)/2} \quad (6)$$

其中， μ 、 λ 、 ν 分别表示学生 t 分布的均值、准确度和自由度。

由于直接使用式(6)过于复杂，为了便于推导计算，可以将学生 t 分布用高斯尺度混合 (GSM, Gaussian scale mixture)模型表示^[13]，通过引入辅助隐变量 τ ，则学生 t 分布可视为联合分布 $p(x, \tau | \mu, \lambda, \nu)$ 的边缘分布，即

$$St(x; \mu, \lambda, \nu) = \int_0^\infty p(x, \tau | \mu, \lambda, \nu) d\tau = \int_0^\infty p(x | \mu, \tau) p(\tau | \lambda, \nu) d\tau \quad (7)$$

其中， $p(x|\mu, \tau)$ 为均值为 μ 、方差为 τ^{-1} 的条件高斯分布。

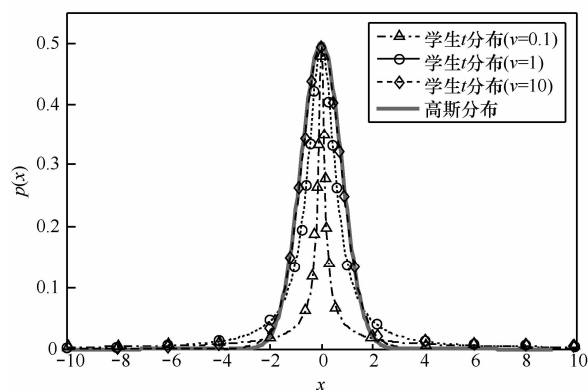
$$p(x | \mu, \tau) = Gau(x; \mu, \tau^{-1}) = \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}(x-\mu)^2\right\} \quad (8)$$

而 $p(\tau|\lambda, \nu)$ 是参数分别为 $a=\nu/2$ 、 $b=\lambda^{-1}\nu/2$ 的 Gamma 分布，即

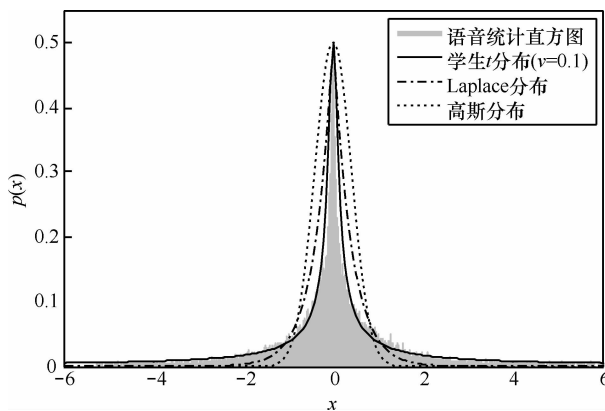
$$p(\tau | \lambda, \nu) = Gam\left(\tau; \frac{\nu}{2}, \frac{\nu}{2\lambda}\right) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\nu}{2\lambda}\right)^{\nu/2} \tau^{\nu/2-1} \exp\left(-\frac{\nu}{2\lambda}\tau\right) \quad (9)$$

图 1 给出了学生 t 分布与高斯分布、Laplace 分布以及语音信号时域统计直方图的比较。为了便于观测比较，图中各个分布的均值均设置为 0，峰值概率(mode)设置为 0.5，这些设置不会影响分布函数

的基本特性。图 1(a)给出学生 t 分布与高斯分布的比较，从图 1(a)中可以看出，当 ν 较大时，学生 t 分布趋近于高斯分布；反之，当 ν 较小时，学生 t 分布具有超高斯特性，相对于高斯分布其峭度更大，具有重拖尾、尖峰特性，且自由度参数 ν 控制学生 t 分布的尖锐程度(sharpness)。图 1(b)给出学生 t 分布、Laplace 分布、高斯分布与实际语音统计特性的对比，其中，阴影部分表示语音统计直方图，是通过汉语语料库中的语音信号时域幅度统计得到的。从图 1(b)中可以看出，与高斯分布相比，具有超高斯特性的 Laplace 分布和学生 t 分布能够更好地表征实际的语音分布。进一步的比较可以看出，学生 t 分布比 Laplace 分布对实际语音分布的拟合程度更好，其原因在于 Laplace 分布仅有一个准确度参数，而学生 t 分布可以通过自由度参数 ν 控制其尖锐程度，因而能够更灵活地拟合实际的语音先验分布。基于这一考虑，本文提出采用学生 t 分布对语音信号线性预测激励建模，以克服传统高斯激励模型噪声顽健性不足的问题。



(a) 学生 t 分布与高斯分布的比较



(b) 不同分布与实际语音统计特性的对比

图 1 学生 t 分布及其与高斯分布、Laplace 分布和语音统计直方图的比较

3.2 基于学生 t 分布的线性预测模型

在 3.1 节分析学生 t 分布的超高斯特性的基础上, 本节利用学生 t 分布对语音信号预测激励建模, 并构建完整的带噪语音线性预测模型。假设语音信号 $s(n)$ 可以用 p 阶 AR 模型表示, 且线性预测残差 $\varepsilon(n)$ 服从均值为 0, 准确度和自由度分别为 λ 和 ν 的学生 t 分布, 即

$$p(\varepsilon(n)) = St(\varepsilon(n); 0, \lambda, \nu) \quad (10)$$

则语音信号 $s(n)$ 的条件概率密度函数为

$$p(s(n) | \mathbf{a}, \lambda, \nu) = St(s(n); \mathbf{a}^T \mathbf{s}_{n-1}, \lambda, \nu) \quad (11)$$

利用式 (7) 给出学生 t 分布的高斯尺度混合模型表示, 引入隐变量 τ , 则 $s(n)$ 的条件概率密度函数可表示为

$$p(s(n) | \mathbf{a}, \tau) = Gau(s(n); \mathbf{a}^T \mathbf{s}_{n-1}, \tau^{-1}) \quad (12)$$

$$p(\tau | \lambda, \nu) = Gam\left(\tau; \frac{\nu}{2}, \frac{\nu}{2\lambda}\right) \quad (13)$$

假设在加性噪声环境下, 带噪语音表示为 $y(n) = s(n) + r(n)$, 其中, 加性噪声 $r(n)$ 服从方差为 η^{-1} 的高斯分布, 即

$$p(r(n)) = Gau(r(n); 0, \eta^{-1}) \quad (14)$$

则带噪语音 $y(n)$ 的条件概率密度函数为

$$p(y(n) | s(n), \eta) = Gau(y(n); s(n), \eta^{-1}) \quad (15)$$

为了实现完全贝叶斯估计, 还需要假设模型参数的先验分布。一般地, 在难以得到准确具体的参数先验知识的条件下, 通常假设其先验分布为无信息先验(non-informative priors), 以尽可能减少不准确的先验假设对后验分布的影响^[12]。同时为了满足共轭先验(conjugate priors)^[12]以便于数学推导, 本文中对模型参数的先验分布做如下假设: 假设线性预测系数 \mathbf{a} 服从 p 维高斯先验分布, 则

$$p(\mathbf{a}) = Gau(\mathbf{a}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a^{-1}) \quad (16)$$

且预测残差的准确度 λ 和加性高斯噪声的准确度 η 分别服从先验逆 Gamma 分布和先验 Gamma 分布, 则

$$p(\lambda) = IGam(\lambda; a_\lambda, b_\lambda) = \frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \lambda^{-(a_\lambda+1)} \exp\left(-\frac{b_\lambda}{\lambda}\right) \quad (17)$$

$$p(\eta) = Gam(\eta; a_\eta, b_\eta) = \frac{b_\eta^{a_\eta}}{\Gamma(a_\eta)} \eta^{a_\eta-1} \exp(-b_\eta \eta) \quad (18)$$

并通过参数调整使得 $p(\mathbf{a})$ 、 $p(\lambda)$ 和 $p(\eta)$ 均满足无信息先验性质。此外, 为了实现预测残差的超高斯约束, 预测残差自由度 ν 固定为一个经验值 $\nu=0.1$ 。

图 2 给出了本文算法的概率图模型。图 2 中描述了各个变量间的依赖关系, 其中, 实心圆表示可观测变量, 即带噪语音 $\mathbf{y} = \{y(1), y(2), \dots, y(N)\}$, 空心圆表示隐含变量和参数变量, 包括纯净语音 \mathbf{s} 、线性预测系数 \mathbf{a} 、加性噪声的准确度 η 、预测残差准确度 λ 以及隐变量 τ , 圆点表示模型的超参数。下文为了简洁表示, 将待估计的隐含变量和模型参数合并表示为 $\Theta = \{\mathbf{s}, \mathbf{a}, \tau, \eta, \lambda\}$, 同时分别用 s_n 、 y_n 、 ε_n 代替表示 $s(n)$ 、 $y(n)$ 、 $\varepsilon(n)$ 。根据贝叶斯法则, \mathbf{y} 与 Θ 的联合概率密度函数为

$$p(\mathbf{y}, \Theta) \propto \prod_{n=1}^N p(y_n | s_n, \eta) \cdot \prod_{n=1}^N \{p(s_n | \mathbf{a}, \tau) p(\tau | \lambda, \nu)\} p(\mathbf{a}) p(\eta) p(\lambda) \quad (19)$$

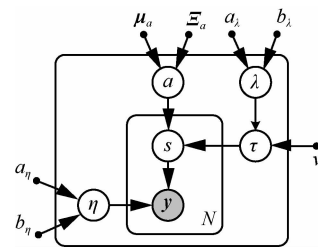


图 2 本文算法的概率模型

根据所建立的概率图模型, 语音信号线性预测系数的求解转化为一个贝叶斯估计问题, 即在给定带噪语音 \mathbf{y} 的条件下, 求解 \mathbf{a} 的最大后验概率估计为

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} p(\mathbf{a} | \mathbf{y}) \quad (20)$$

由于式(20)中 $p(\mathbf{a} | \mathbf{y})$ 的解析解难以求得, 因此难以实现准确的贝叶斯估计。针对这一问题, 本文提出采用变分贝叶斯估计进行近似求解。

3.3 模型参数求解算法

在 3.2 节构建的本文模型的基础上, 本节推导基于变分贝叶斯估计的模型参数求解算法。变分贝叶斯估计是一种近似求解方法, 它通过一个假设的后验概率分布 $q(\Theta)$, 在 KL 散度最小的准则下, 逼近真实的后验分布 $p(\Theta | \mathbf{y})$, 即

$$q(\Theta) = \arg \min_{q(\Theta)} \{KL(q(\Theta) \| p(\Theta | \mathbf{y}))\} \quad (21)$$

其中, KL 散度用于衡量 2 个分布之间的距离, 其

定义为

$$KL(q \parallel p) = \int q(\theta) \log \frac{q(\theta)}{p(\theta|y)} d\theta \quad (22)$$

为便于求解, 通常假设 θ 中各个变量满足相互后验独立^[9], 因此本文中假设 $q(\theta)$ 可分解为

$$q(\theta) = q(s)q(a)q(\eta)q(\tau)q(\lambda) \quad (23)$$

在这一条件下, 通过最小化 $q(\theta)$ 和 $p(\theta|y)$ 之间的 KL 散度, 可得到其最优解的表达式为^[9]

$$\begin{aligned} \log q^*(\theta_j) &= E_{q(\theta/\theta_j)} \{ \log p(y, \theta) \} \\ &= \langle \log p(y, \theta) \rangle_{q(\theta/\theta_j)} \end{aligned} \quad (24)$$

其中, θ_j 表示参数集 θ 中的第 j 个参数, $q(\theta/\theta_j)$ 表示集合 θ 中除 θ_j 以外的其余参数的近似后验分布, $\langle \cdot \rangle_q$ 表示关于分布 q 的期望。

将式(12)~式(19)代入式(24), 依次求解即可得到 θ 中各元素的近似后验分布(具体的推导过程在附录 1 中给出)。

$$q^*(a) = Gau(a; \mu_a^*, \Sigma_a^{*-1}) \quad (25)$$

其中, $\Sigma_a^* = \frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N \langle s_{n-1} s_{n-1}^T \rangle_{q(s)} + \Sigma_a$, $\mu_a^* = \Sigma_a^{*-1} \cdot$

$$\left[\frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N \langle s_n s_{n-1}^T \rangle_{q(s)} + \Sigma_a \mu_a \right].$$

$$q^*(\eta) = Gam(\eta; a_\eta^*, b_\eta^*) \quad (26)$$

其中, $a_\eta^* = a_\eta + N/2$, $b_\eta^* = b_\eta + \frac{1}{2} \sum_{n=1}^N y_n^2 + \frac{1}{2} \sum_{n=1}^N \langle s_n^2 \rangle_{q(s)} -$

$$\sum_{n=1}^N y_n \langle s_n \rangle_{q(s)}.$$

$$q^*(\lambda) = IGam(\lambda; a_\lambda^*, b_\lambda^*) \quad (27)$$

其中, $a_\lambda^* = a_\lambda + \frac{Nv}{2}$, $b_\lambda^* = b_\lambda + \frac{Nva_\tau^*}{2b_\tau^*}$.

$$q^*(\tau) = Gam(\tau; a_\tau^*, b_\tau^*) \quad (28)$$

其中, $a_\tau^* = \frac{N+v}{2}$, $b_\tau^* = \frac{va_\lambda^*}{2b_\lambda^*} + \frac{1}{2} \sum_{n=1}^N \left\{ \langle s_n^2 \rangle_{q(s)} -$

$$2 \mu_a^{*T} \langle s_n s_{n-1} \rangle_{q(s)} + \mu_a^{*T} \langle s_{n-1} s_{n-1}^T \rangle_{q(s)} \mu_a^* + \text{Tr} \left(\langle s_{n-1} s_{n-1}^T \rangle_{q(s)} \Sigma_a^{*-1} \right) \right\}.$$

$$q^*(s) = Gau(s; \kappa^*, A^{*-1}) \quad (29)$$

其中, $A^* = \frac{a_\tau^*}{b_\tau^*} A^T A + \frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N B_{n-1} + \frac{a_\eta^*}{b_\eta^*} I$, $\kappa^* = \frac{a_\eta^*}{b_\eta^*}$.

$A^{*-1} y$ 。

根据式(29)可知, $q^*(s)$ 满足均值为 κ^* 、协方差矩阵为 A^{*-1} 的高斯分布, 因此对于任意 2 个语音信号取样点 s_m 和 s_n , 其乘积 $s_m s_n$ 的后验期望表示为

$$\begin{aligned} \langle s_m s_n \rangle_{q(s)} &= \kappa^* (N-m+1) \kappa^* (N-n+1) + \\ &v_{N-m+1, N-n+1} \end{aligned} \quad (30)$$

其中, $v_{N-m+1, N-n+1}$ 表示矩阵 A^* 中位于 $(N-m+1, N-n+1)$ 的元素。由此, 式(25)~式(28)中的 $\langle s_n^2 \rangle_{q(s)}$ 、 $\langle s_n s_{n-1} \rangle_{q(s)}$ 、 $\langle s_{n-1} s_{n-1}^T \rangle_{q(s)}$ 可通过式(30)计算得到。

由于式(25)~式(29)相互依赖, 因此需要通过迭代实现。理论研究证明, 在变分贝叶斯框架内, 该迭代过程将保证收敛^[14]。综上, 本文算法的基本流程总结如下。

参数初始化: $a_\eta^* = b_\eta^* = 1$, $a_\lambda^* = b_\lambda^* = 1$, μ_a^* 初始值为带噪语音的 LPC 系数, $\Sigma_a^* = I$, b_τ^* 的初始值为带噪语音线性预测残差的方差, $a_\tau^* = 1$ 。

while $i < M$ (i 为迭代序号, M 为最大迭代次数)

更新式(29)中 $q^*(s)$ 的后验参数 A^* 和 κ^* ;

根据式(30)计算 $\langle s_n^2 \rangle_{q(s)}$ 、 $\langle s_n s_{n-1} \rangle_{q(s)}$ 、

$\langle s_{n-1} s_{n-1}^T \rangle_{q(s)}$;

依次计算更新式(25)~式(28)中的后验参数 μ_a^* 、 Σ_a^* 、 a_η^* 、 b_η^* 、 a_τ^* 、 b_τ^* 、 a_λ^* 、 b_λ^* ;

end while

输出线性预测系数 $a = \mu_a^*$ 。

4 实验结果与分析

在仿真实验中, 基于 MATLAB 实现本文提出的顽健线性预测分析算法, 并与现有的 2 种算法进行比较, 包括采用对数谱幅度估计前端语音增强^[2]的 LPC 自相关求解算法、文献[7]中提出的基于高斯激励 AR 模型的变分贝叶斯估计算法, 下文中分别用 LSA(log-spectral amplitude)和 GAR(Gaussian-AR)表示这 2 种算法。考虑到本文算法所提取的 LP 系数对应的语音合成器难以在理论上保证其稳定性, 在下面的实验中, 若判定当前 LP 系数不满足稳定性条件, 则采用传统的自相关求解结果(即算法迭代初始值)替换当前结果。由于实验过程中产生不稳定结果的比例很低(小于 1%), 且不稳定的偏差较小, 因此这一处理对算法总体性能的影响是很小

的。实验中的具体参数设置为：语音分析帧长为 256，帧移为 128；线性预测阶数 $p=10$ 。

实验中的纯净语音选取采样率为 8kHz 的标准汉语语音，其中，男女声各 12 句，每句时间长度约为 4s。从标准噪声库 Noisex92 中选取白噪声、M109 坦克噪声、F16 飞机噪声作为原始噪声，并将采样率转换为 8kHz。通过 MATLAB 对噪声信号和纯净语音进行加性混和，信噪比分别为 0dB、5dB、10dB 和 15dB。

由于线性预测系数对应的频域表示为信号的谱包络，因此本实验中首先通过对比语音信号谱包络来比较预测系数提取的性能。图 3 给出了在信噪比为 5dB 的 3 种噪声环境下，对一帧语音(语音内容为女声元音“wu”)分别采用上述 3 种算法提取得到的语音短时谱包络，并与纯净语音的短时谱包络进行比较。从图中可以看出，对数谱幅度估计算法^[2]能够恢复低频部分的谱包络结构，但高频部分的共振峰仍存在较严重的失真。文献[7]提出的高斯激励 AR 模型贝叶斯估计算法能够在一定程度上改进高频共振峰估计效果。而本文算法相对于其他 2 种算法有较大的提高，基本能够恢复语音的谱包络结构，即使对于高频部分的共振峰仍然能够较好地恢复。同时可以看出，不同噪声条件下的参数提取性能存在差异。具体地，在坦克噪声条件下，共振

峰的恢复性能最好，分析其原因在于坦克噪声较集中于低频部分，而语音低频部分的共振峰能量较高，因而不易被噪声污染。相反，由于飞机噪声更多地集中于高频，因而对语音高频部分的共振峰结构影响较大。而对于高斯白噪声，其在整个频谱范围均对语音共振峰带来较严重的影响，因而比其他 2 种噪声条件下的共振峰失真更为严重。在下文中的定量分析实验中也验证了这一分析。

为了定量地比较算法的性能，采用对数似然比 (LLR, log-likelihood ratio)失真测度衡量 3 种算法提取语音预测系数的准确度。LLR 是一种常用的衡量线性预测系数匹配特性的失真测度，其计算公式为^[15]

$$D_{LLR}(a_d, a_r) = \log \left(\frac{a_d^T R_r a_d}{a_r^T R_r a_r} \right) \quad (31)$$

其中， a_r 为参考 LP 系数， R_r 为参考自相关矩阵， a_r 和 R_r 均由纯净语音中计算得到，而 a_d 为从带噪语音中估计得到的待测 LP 系数。

表 1 给出了不同噪声环境下，3 种算法提取 LP 系数的平均 LLR 测度比较结果。在计算过程中，通过 VAD 检测仅对所有语音帧的 LLR 取平均得到最终的平均 LLR 测度值。从表 1 中结果可以看出，相对于 LSA 前端增强算法和高斯激励模型变分贝叶斯估计算法，本文算法的平均 LLR 测度值较小，表明线性预测参数提取的失真度较小，尤其是在高斯白

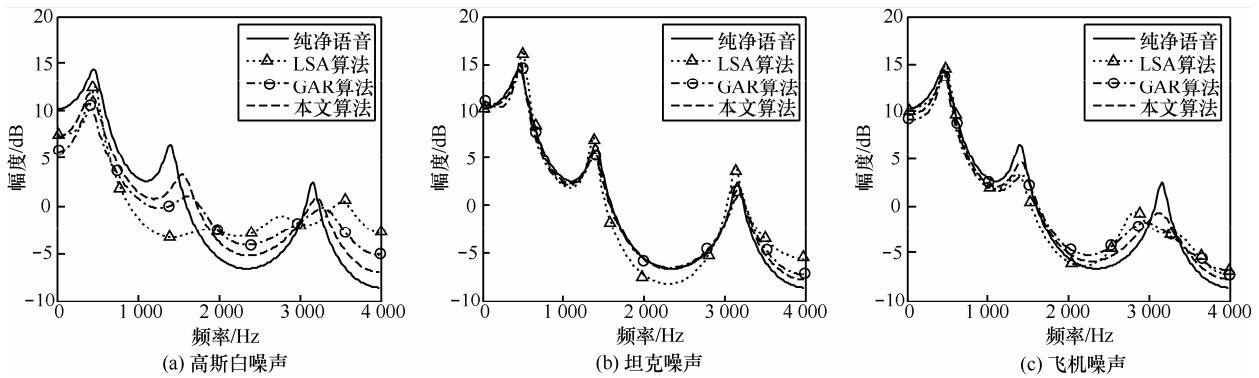


图 3 3 种算法提取的语音短时谱包络

表 1

3 种算法的平均 LLR 测度比较结果

输入 信噪比/dB	高斯白噪声				M109 坦克噪声				F16 飞机噪声			
	带噪 语音	LSA 算法	GAR 算法	本文 算法	带噪 语音	LSA 算法	GAR 算法	本文 算法	带噪 语音	LSA 算法	GAR 算法	本文 算法
0	1.776	1.653	1.616	1.536	0.656	0.664	0.623	0.582	1.101	0.960	0.877	0.834
5	1.381	1.195	1.162	1.083	0.441	0.429	0.399	0.378	0.803	0.746	0.695	0.656
10	1.015	0.855	0.801	0.754	0.273	0.248	0.222	0.191	0.550	0.463	0.438	0.392
15	0.699	0.610	0.588	0.527	0.154	0.148	0.147	0.132	0.348	0.280	0.274	0.256

噪声条件下改进幅度较大。需要注意的是，与白噪声和飞机噪声的条件下相比，3种算法在 M109 坦克噪声条件下的 LLR 改进量均比较小，其原因在于坦克噪声条件下的带噪语音 LLR 测度本身就比较小，因此进一步提高比较有限，这一结果与前述对语音短时谱包络的实验结果及分析是一致的。

为了进一步比较 3 种参数提取算法对语音信号的影响，利用所提取的语音参数通过 LPC 滤波器合成得到语音信号，通过比较合成语音的 PESQ 得分衡量参数提取的综合性能。图 4 给出了 3 种算法增强后语音的感知语音质量(PESQ)评估比较结果。从中可以看出，与其他 2 种算法相比，本文所提算法的合成语音具有更好的语音质量，非正式的主观听觉测试与上述结果相一致。

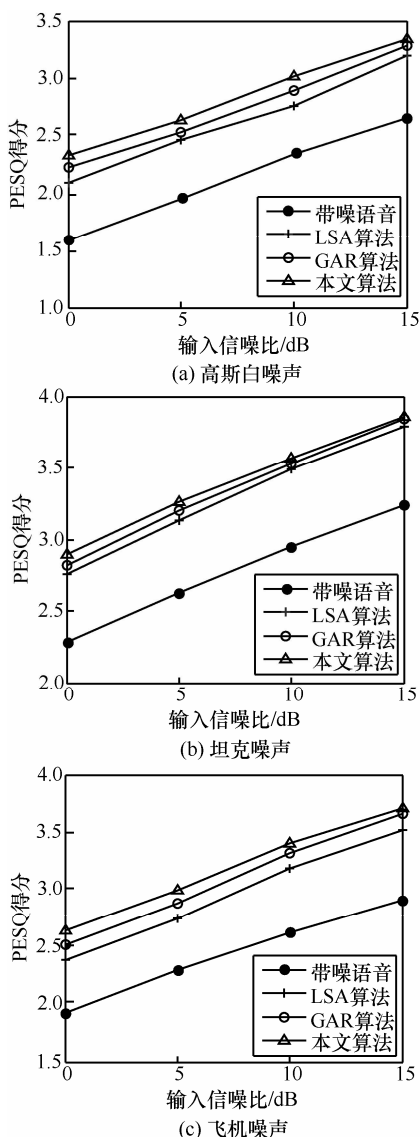


图 4 3 种算法的 PESQ 评估结果

此外，在计算复杂度方面，传统的 LPC 自相关解法需要 p^3+Np 量级的乘法和加法运算 (p 为线性预测阶数, N 为语音分析帧长度)，通过高效的 Durbin 算法能够降为 p^2+Np 量级^[1]。但前端语音增强预处理会增加额外的复杂度，例如实验中采用复杂度较低的 LSA 增强算法，其增加的复杂度为 MbN 量级。本文算法与 GAR 算法的复杂度相近，两者都需要通过循环迭代实现，且单次迭代的计算量均为 $p^3+Np^2+N^2$ 量级。在实验观测中，本文算法通常能够在 10 次迭代后收敛，对于一段长度为 4s 的语音进行处理，其平均运行时间为 39s；而在相同条件下，传统的 LSA 与 LPC 相结合的算法和 GAR 算法的平均运行时间分别为 1.2s 和 34s。由此可见，本文算法在计算复杂度方面较传统算法有所增加。深入分析后可以看出，本文算法的计算量主要体现在矩阵求逆运算上，由于矩阵不满足托普利兹性质，因此目前尚没有高效的快速求解算法。

5 结束语

本文提出了一种基于超高斯激励 AR 模型的噪声顽健语音信号线性预测分析算法。针对传统高斯激励 AR 模型的噪声顽健性问题，本文采用超高斯的学生 t 分布对线性预测残差建模，并显式地考虑加性噪声的影响，从而构建了噪声顽健语音信号线性预测分析的概率图模型。针对模型复杂度较高、准确的模型参数推理难以实现的问题，本文利用变分贝叶斯估计推导得到模型参数的近似后验分布，并通过迭代算法实现顽健语音参数提取。实验结果表明，与传统的高斯激励 AR 模型相比，本文算法在多种背景噪声条件下均能够有效提高语音线性预测系数提取的顽健性。在下一步的工作中，将考虑在本文所提出的算法框架内，更好地对线性预测系数的先验信息进行建模，从而期望进一步提高参数提取的准确度。

附录 1 近似后验分布的推导

A. 加性噪声准确度 η 的近似后验分布 $q^*(\eta)$

$$\begin{aligned}
 \log q^*(\eta) &= E_{q(\Theta/\eta)} \{ \log p(y, \Theta) \} \\
 &= \langle \log p(y|s, \eta) \rangle_{q(s)} + \log p(\eta) + C \\
 &= \frac{N}{2} \log \eta - \frac{\eta}{2} \left\langle \sum_{n=1}^N (y_n - s_n)^2 \right\rangle_{q(s)} + (a_\eta - 1) \log \eta - b_\eta \eta + C \\
 &= \left(a_\eta + \frac{N}{2} - 1 \right) \log \eta - \eta \cdot
 \end{aligned}$$

$$\left[b_\eta + \frac{1}{2} \sum_{n=1}^N y_n^2 + \frac{1}{2} \sum_{n=1}^N \langle s_n^2 \rangle_{q(s)} - \sum_{n=1}^N y_n \langle s_n \rangle_{q(s)} \right] + C \quad (32)$$

因此,

$$q^*(\eta) = \text{Gam}(\eta; a_\eta^*, b_\eta^*) \quad (33)$$

其中, $a_\eta^* = a_\eta + N/2$, $b_\eta^* = b_\eta + \frac{1}{2} \sum_{n=1}^N y_n^2 + \frac{1}{2} \sum_{n=1}^N \langle s_n^2 \rangle_{q(s)} - \sum_{n=1}^N y_n \langle s_n \rangle_{q(s)}$ 。

B. 预测残差准确度 λ 的近似后验概率 $q^*(\lambda)$

$$\begin{aligned} \log q^*(\lambda) &= E_{q(\theta, \lambda)} \{ \log p(\mathbf{y}, \Theta) \} \\ &= \langle N \log p(\tau | \lambda, \nu) \rangle_{q(\tau)} + \log p(\lambda) + C \\ &= E_\tau \left\{ -\frac{N\nu}{2} \tau^{-1} - \frac{N\nu}{2} \log \lambda \right\} - (a_\lambda + 1) \log \lambda - b_\lambda \lambda^{-1} + C \\ &= - \left(a_\lambda + \frac{N\nu}{2} + 1 \right) \log \lambda - \lambda^{-1} \left[b_\lambda + \frac{N\nu}{2} \langle \tau \rangle_{q(\tau)} \right] + C \end{aligned} \quad (34)$$

根据式(41), τ 后验服从 Gamma 分布, 则 $\langle \tau \rangle_{q(\tau)} = a_\tau^* / b_\tau^*$, 因此,

$$q^*(\lambda) = \text{IGam}(\lambda; a_\lambda^*, b_\lambda^*) \quad (35)$$

其中, $a_\lambda^* = a_\lambda + \frac{N\nu}{2}$, $b_\lambda^* = b_\lambda + \frac{N\nu a_\tau^*}{2b_\tau^*}$ 。

C. 线性预测系数 \mathbf{a} 的近似后验概率 $q^*(\mathbf{a})$

$$\begin{aligned} \log q^*(\mathbf{a}) &= \langle \log p(\mathbf{y}, \Theta) \rangle_{q(\theta, \mathbf{a})} \\ &= \langle \log p(\mathbf{s} | \mathbf{a}, \tau) \rangle_{q(\mathbf{s}, \tau)} + \log p(\mathbf{a}) + C \\ &= -\frac{1}{2} \langle \tau \rangle_{q(\tau)} \left\langle \sum_{n=1}^N (s_n - \mathbf{a}^T \mathbf{s}_{n-1})^2 \right\rangle_{q(\mathbf{s})} - \\ &\quad \frac{1}{2} \left[(\mathbf{a} - \boldsymbol{\mu})^T \boldsymbol{\Sigma} (\mathbf{a} - \boldsymbol{\mu}) \right] + C \end{aligned} \quad (36)$$

将式(36)化简后得到

$$\begin{aligned} \log q^*(\mathbf{a}) &= -\frac{1}{2} \mathbf{a}^T \left[\langle \tau \rangle_{q(\tau)} \sum_{n=1}^N \langle \mathbf{s}_{n-1} \mathbf{s}_{n-1}^T \rangle_{q(\mathbf{s})} + \boldsymbol{\Sigma}_a \right] \mathbf{a} + \mathbf{a}^T \cdot \\ &\quad \left[\langle \tau \rangle_{q(\tau)} \sum_{n=1}^N \langle \mathbf{s}_n \mathbf{s}_{n-1}^T \rangle_{q(\mathbf{s})} + \boldsymbol{\Sigma}_a \boldsymbol{\mu}_a \right] + C \end{aligned} \quad (37)$$

根据式(41), τ 后验服从 Gamma 分布, 则 $\langle \tau \rangle_{q(\tau)} = a_\tau^* / b_\tau^*$, 因此, $q^*(\mathbf{a})$ 为高斯分布且其可表示为

$$q^*(\mathbf{a}) = \text{Gau}(\mathbf{a}; \boldsymbol{\mu}_a^*, \boldsymbol{\Sigma}_a^{*-1}) \quad (38)$$

其中, $\boldsymbol{\Sigma}_a^* = \frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N \langle \mathbf{s}_{n-1} \mathbf{s}_{n-1}^T \rangle_{q(\mathbf{s})} + \boldsymbol{\Sigma}_a$, $\boldsymbol{\mu}_a^* = \boldsymbol{\Sigma}_a^{*-1} \cdot$

$$\left[\frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N \langle \mathbf{s}_n \mathbf{s}_{n-1}^T \rangle_{q(\mathbf{s})} + \boldsymbol{\Sigma}_a \boldsymbol{\mu}_a \right]$$

D. 隐变量 τ 的近似后验分布 $q^*(\tau)$

$$\begin{aligned} \log q^*(\tau) &= E_{q(\theta, \tau)} \{ \log p(\mathbf{y}, \Theta) \} \\ &= \langle \log p(\mathbf{s} | \mathbf{a}, \tau) \rangle_{q(\mathbf{s}, \mathbf{a})} + \langle \log p(\tau | \lambda, \nu) \rangle_{q(\lambda)} + C \\ &= \frac{N}{2} \log \tau + \left\langle \sum_{n=1}^N \left(-\frac{\tau}{2} (s_n - \mathbf{a}^T \mathbf{s}_{n-1})^2 \right) \right\rangle_{q(\mathbf{s}, \mathbf{a})} + \\ &\quad \left(\frac{\nu}{2} - 1 \right) \log \tau - \frac{\nu}{2 \langle \lambda \rangle_{q(\lambda)}} \tau + C \\ &= \left(\frac{\nu + N}{2} - 1 \right) \log \tau - \\ &\quad \left(\frac{\nu}{2 \langle \lambda \rangle_{q(\lambda)}} + \frac{1}{2} \left\langle \sum_{n=1}^N (s_n - \mathbf{a}^T \mathbf{s}_{n-1})^2 \right\rangle_{q(\mathbf{s}, \mathbf{a})} \right) \tau + C \end{aligned} \quad (39)$$

根据式(38), \mathbf{a} 后验服从多元高斯分布, 则式(39)化简后得到

$$\begin{aligned} \log q^*(\tau) &= \left(\frac{\nu + N}{2} - 1 \right) \log \tau - \tau \left(\frac{\nu}{2 \langle \lambda \rangle_{q(\lambda)}} + \frac{1}{2} \sum_{i=1}^N \left\langle \langle s_i^2 \rangle_{q(\mathbf{s})} - \right. \right. \\ &\quad \left. \left. 2 \boldsymbol{\mu}_a^{*T} \langle \mathbf{s}_i \mathbf{s}_{i-1} \rangle_{q(\mathbf{s})} + \boldsymbol{\mu}_a^{*T} \langle \mathbf{s}_{i-1} \mathbf{s}_{i-1}^T \rangle_{q(\mathbf{s})} \boldsymbol{\mu}_a^* + \text{Tr} \left(\langle \mathbf{s}_{i-1} \mathbf{s}_{i-1}^T \rangle_{q(\mathbf{s})} \boldsymbol{\Sigma}_a^{*-1} \right) \right\rangle \right) + C \end{aligned} \quad (40)$$

根据式(35), λ 后验服从逆 Gamma 分布, 则 $\langle \lambda \rangle_{q(\lambda)} = b_\lambda^* / a_\lambda^*$, 因此,

$$q^*(\tau) = \text{Gam}(\tau; a_\tau^*, b_\tau^*) \quad (41)$$

其中, $a_\tau^* = \frac{N + \nu}{2}$, $b_\tau^* = \frac{\nu a_\lambda^*}{2 b_\lambda^*} + \frac{1}{2} \sum_{i=1}^N \left\{ \langle s_i^2 \rangle_{q(\mathbf{s})} - 2 \boldsymbol{\mu}_a^{*T} \langle \mathbf{s}_i \mathbf{s}_{i-1} \rangle_{q(\mathbf{s})} + \boldsymbol{\mu}_a^{*T} \langle \mathbf{s}_{i-1} \mathbf{s}_{i-1}^T \rangle_{q(\mathbf{s})} \boldsymbol{\mu}_a^* + \text{Tr} \left(\langle \mathbf{s}_{i-1} \mathbf{s}_{i-1}^T \rangle_{q(\mathbf{s})} \boldsymbol{\Sigma}_a^{*-1} \right) \right\}$, $\text{Tr}(\cdot)$ 表示矩阵的迹。

E. 纯净语音 \mathbf{s} 的近似后验概率 $q^*(\mathbf{s})$

$$\begin{aligned} \log q^*(\mathbf{s}) &= E_{q(\theta, \mathbf{s})} \{ \log p(\mathbf{y}, \Theta) \} \\ &= \langle \log p(\mathbf{y} | \mathbf{s}, \eta) \rangle_{q(\eta)} + \langle \log p(\mathbf{s} | \mathbf{a}, \tau) \rangle_{q(\mathbf{a}, \tau)} + C \\ &= -\frac{1}{2} \langle \eta \rangle_{q(\eta)} \sum_{n=1}^N (y_n - s_n)^2 - \left\langle \frac{\tau}{2} \sum_{n=1}^N (s_n - \mathbf{a}^T \mathbf{s}_{n-1})^2 \right\rangle_{q(\mathbf{a}, \tau)} + C \end{aligned} \quad (42)$$

根据式(33)和式(41), η 和 τ 后验服从 Gamma 分布, 则 $\langle \eta \rangle_{q(\eta)} = a_\eta^* / b_\eta^*$, $\langle \tau \rangle_{q(\tau)} = a_\tau^* / b_\tau^*$, 并根据式(38), \mathbf{a} 后验服从多元高斯分布, 则对式(42)化简后得到

$$\log q^*(\mathbf{s}) = -\frac{1}{2} \mathbf{s}^T \left[\frac{a_\tau^*}{b_\tau^*} \mathbf{A}^T \mathbf{A} + \frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N \mathbf{B}_{n-1} + \frac{a_\eta^*}{b_\eta^*} \mathbf{I} \right] \mathbf{s} + \mathbf{s}^T \left[\frac{a_\eta^*}{b_\eta^*} \mathbf{y} \right] + C \quad (43)$$

其中, \mathbf{A} 是首行元素为 $[1, -\boldsymbol{\mu}^{*T}, 0, \dots, 0]$ 的 N 阶托普利兹矩阵, N 阶矩阵 \mathbf{B}_n 仅在 $(N-n+1, N-n+1)$ 处具有非零 p 阶子矩阵 $\boldsymbol{\Sigma}_a^{*-1}$, 其余元素均为 0。

因此,

$$q^*(\mathbf{s}) = \text{Gau}(\mathbf{s}; \boldsymbol{\kappa}^*, \mathbf{A}^{*-1}) \quad (44)$$

其中, $A^* = \frac{a_\tau^*}{b_\tau^*} A^T A + \frac{a_\tau^*}{b_\tau^*} \sum_{n=1}^N B_{n-1} + \frac{a_\tau^*}{b_\tau^*} I$, $\kappa^* = \frac{a_\tau^*}{b_\tau^*} A^{*T} y$ 。

附录 2 本文中使用的标准分布及其统计量

A. 多元高斯分布:

$$Gau(x; \mu, \Sigma^{-1}) = \left(\frac{\Sigma}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma(x-\mu)\right\}, \langle x \rangle_{Gau} = \mu$$

B. Gamma 分布:

$$Gam(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \langle x \rangle_{Gam} = a/b$$

C. 逆 Gamma 分布:

$$IGam(x; a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp\left(-\frac{b}{x}\right), \langle x \rangle_{IGam} = b/a$$

参考文献:

[1] QUITERTI T F. Discrete-Time Speech Signal Processing: Principles and Practice[M]. New Jersey: Prentice Hall PTR, 2001.

[2] BORGSTROM J B, ALWAN A. Log-spectral amplitude estimation with generalized gamma distributions for speech enhancement[A]. Proceedings of IEEE ICASSP[C]. Prague, Czech, 2011. 4756-4759.

[3] MARTIN R. Speech enhancement based on minimum mean square error estimation and supergaussian priors[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5):845-856.

[4] 赵晓群, 黄小珊. 改进的基于人耳掩蔽效应谱减语音增强算法[J]. 通信学报, 2008, 29(9):73-80.

ZHAO X Q, HUANG X S. Improved speech enhancement based on spectral subtraction and auditory masking effect[J]. Journal on Communications, 2008, 29(9):73-80.

[5] LIM J, OPPENHEIM A. All-pole modeling of degraded speech[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, 26(3):197-210.

[6] MORRIS R W, ARROWOOD J A, CLEMENTS M A. Comparison of autoregressive parameter estimation algorithms for speech processing and recognition[A]. Proceedings of IEEE ICASSP[C]. Philadelphia, USA, 2005. 229-232.

[7] YOSHIOKA T, MIYOSHI M. Adaptive suppression of non-stationary noise by using the variational Bayesian method[A]. Proceedings of IEEE ICASSP[C]. Las Vegas, USA, 2008. 4889-4892.

[8] HUANG Q, YANG J, XUE Y. A novel AR model based speech

enhancement approach within variational Bayesian framework[J]. Chinese Journal of Electronics, 2007, 16(3):499-502.

[9] CHRISTMAS J, EVERSON R. Robust autoregression: student-*t* innovations using variational Bayes[J]. IEEE Transactions on Signal Processing, 2011, (59)1:48-57.

[10] GAZOR S, ZHANG W. Speech probability distribution[J]. IEEE Signal Processing Letters, 2003, 10(7):204-207.

[11] GIACOBELLO D, CHRISTENSEN M G, MURTHI M N, et al. Sparse linear prediction and its applications to speech processing[J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(5):1644-1657.

[12] BISHOP C. Pattern Analysis and Machine Learning[M]. New York: Springer, 2006.

[13] TZIKAS D, LIKAS A, GALATSANOS N. Variational Bayesian sparse kernel-based blind deconvolution with Student's-*t* priors[J]. IEEE Transactions on Image Processing, 2009, 18(4):753-764.

[14] MAINA C, WALSH J. Joint speech enhancement and speaker identification using approximate Bayesian inference[J]. IEEE Transactions on Audio, Speech and Language Processing, 2011, 19(6):1517-1529.

[15] HU Y, LOIZOU P C. Evaluation of objective quality measures for speech enhancement[J]. IEEE Transactions on Audio, Speech and Language Processing, 2008, 16(1):229-238.

作者简介:



周彬 (1986-), 男, 河南济源人, 解放军理工大学博士生, 主要研究方向为语音增强、低速率语音编码。

邹霞 (1979-), 男, 湖北公安人, 博士, 解放军理工大学副教授, 主要研究方向为低速率语音编码、语音增强。

张雄伟 (1965-), 男, 浙江嘉兴人, 解放军理工大学教授、博士生导师, 主要研究方向为数字通信、多媒体信息处理。

赵改华 (1987-), 女, 河南漯河人, 解放军理工大学硕士生, 主要研究方向为语音增强。