

基于高斯混合模型的压缩域语音增强方法

梁 岩, 鲍长春, 夏丙寅, 何玉文, 周 璇, 李 娜

(北京工业大学电子信息与控制工程学院, 北京 100124)

摘 要: 为了有效利用纯净语音导抗谱频率参数(ISFs)的先验知识, 本文针对 ITU-T G. 722.2 宽带语音编码标准提出了一种基于高斯混合模型的压缩域语音增强方法. 首先, 将含噪语音、纯净语音的导抗谱频率参数, 以及对应的增益调整因子构成特征矢量, 并利用高斯混合模型拟合其概率密度; 然后, 在最小均方误差 (MMSE) 准则下对纯净语音的特征参数进行最优贝叶斯估计. 为了兼容编码器中的非连续性传输模式, 当处理信号为非语音信息时, 算法在保持噪声帧谱包络参数不变的前提下, 按固定比例调整对数帧能量; 且若出现帧擦除情况, 算法不调整接收到的码流, 并按正常帧处理方式调整恢复后的参数以更新相关历史. 本文采用 ITU-T G. 160 标准进行了性能测试, 结果表明, 与参考方法相比, 所提方法在保证信噪比提高能力的同时, 可以达到更大的噪声衰减量, 且增强语音的客观质量更优.

关键词: 语音增强; 参数域; 高斯混合模型; 贝叶斯估计; 非连续性传输; 帧擦除

中图分类号: TN912. 35 **文献标识码:** A **文章编号:** 0372-2112 (2012)10-2031-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2012.10.022

Compressed Domain Speech Enhancement Based on Gaussian Mixture Model

LIANG Yan, BAO Chang-chun, XIA Bing-yin, HE Yu-wen, ZHOU Xuan, LI Na

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: A Gaussian Mixture Model (GMM) based speech enhancement method in compressed domain used for ITU-T G. 722.2 wideband speech codec is proposed to take full advantage of the prior knowledge of the Immittance Spectral Frequencies (ISFs) for the clean speech. Firstly, GMM is adopted to model the joint probability density of feature vectors which are composed by the ISFs of noisy speech and clean speech with the corresponding gain scaling factor. Secondly, an optimal Bayesian estimation of feature parameters derived from clean speech is obtained under the minimum mean square error (MMSE) criterion. To be compatible with the DTX (Discontinuous Transmission) mode, the logarithmic energy is attenuated and the ISFs remain when a SID (Silence Insertion Descriptor) frame is received. Furthermore, if an erased frame is received, the bit stream is unchanged and the proposed method is performed on the recovered parameters for the memory update. The evaluation is conducted under the ITU-T G. 160. The results indicate that, comparing with the reference method, the proposed method can produce larger amount of noise level reduction with better objective speech quality, while the SNR improvement remains acceptable.

Key words: speech enhancement; compressed domain; Gaussian Mixture Model; Bayesian estimation; DTX; frame erasure

1 引言

应用于移动通信网络中的语音增强技术是语音信号处理领域的研究热点之一. 实际应用中, 语音增强算法可以分为以语音时域采样点为研究对象的线性域语音增强算法^[1,2], 以及侧重于语音编码参数调整的压缩域语音增强算法. 同线性域算法相比, 压缩域语音增强算法避免了由于语音的重新编码而引入的语音质量下降、计算资源消耗以及额外延时等问题, 其应用前景更为广阔.

现有的压缩域语音增强算法多数着眼于对码激励线性预测 (Code Excited Linear Prediction, CELP) 语音编码模型的参数进行调整. 该模型的参数主要有固定码书增益、自适应码书增益、线性预测编码系数、基音以及固定码书等.

现有的多数压缩域语音增强算法均集中于码书增益的调整. 2000 年, R Chandran 等人提出了一种固定码书增益的调整方法^[3], 调整准则类似于功率谱减准则, 并依据能量准则进一步调整自适应码书增益, 该算法需利用时域信息以获得背景噪声的功率. 文献[4,5]中的

算法借鉴了最小统计量噪声估计算法^[6]的思想,将含噪语音固定码书增益的最小值近似作为背景噪声的固定码书增益,并依据类似的维纳滤波准则调整接收的固定码书增益.文献[7]的算法首先需对含噪语音的码流解码,并利用线性域增强技术得到预增强语音,然后根据该预增强语音和含噪语音计算增益调整因子,最后利用该因子联合调整码书增益.

一些学者针对噪声背景下线性预测系数的调整进行了研究.文献[8]引入了线性预测编码(Linear Predictive Coding, LPC)系数关于信噪比参数的高阶导数,并近似估计纯净语音的谱包络参数.文献[9]利用含噪语音的自相关矩阵近似估计背景噪声的自相关矩阵,并进一步估计纯净语音的 LPC 参数.两种算法运用了较多的矩阵运算,复杂度偏高. Harding 等人^[10]利用高斯混合模型拟合纯净语音、含噪语音 LPC 参数的联合概率密度,并在最小均方误差准则下估计纯净语音的 LPC 参数,原理简单且复杂度低.

虽然学者在压缩域语音增强领域进行了一系列尝试,但至今仍未出现一套完整的针对移动网络中语音编码标准的压缩域语音增强方法.本文针对通信网络中的 ITU-T G. 722.2^[11]宽带语音编码标准,研究了一套基于高斯混合模型的压缩域语音增强算法,以明显抑制背景噪声,同时提高语音客观质量,且算法兼容编码器的 DTX 模式及帧擦除情况.

2 ITU-T G. 722.2 语音编码器

ITU-T G. 722.2 是一套自适应多码率宽带语音编码标准.它针对 7kHz 带宽、16kHz 采样的语音信号,提供了介于 6.6kbps 和 23.85kbps 之间的 9 种编码速率以满足不同传输信道的要求,同时提供了 DTX 模式以提高信道利用率,且引入了帧擦除掩蔽(Frame Erasure Concealment, FEC)算法减少由丢帧而造成的合成语音质量下降问题.

ITU-T G. 722.2 编码器采用了 CELP 语音编码模型,它主要由短时线性预测分析、自适应码书搜索、长时预测后滤波、固定码书搜索以及码书增益计算等模块构成,在 23.85kbps 码速率模式下还包括高频子带增益计算模块.传输的编码参数主要有: ISF (Immittance Spectral Frequency) 参数,用于在解码端恢复语音的谱包络信息;基音、长时预测滤波器索引和自适应码书增益,用于恢复激励信号的周期成分;固定码书索引和固定码书增益,用于恢复激励信号的随机成分.

对比研究噪声背景下各编码参数随信噪比的变化规律,可以发现:固定码书增益与背景噪声功率有近似正比例的联系;白噪声背景下,语音线性预测 (Linear Prediction, LP) 谱共振峰的峰-谷对比减小,同时高频处

的谱变得平坦且能量有被抬高的趋势;有色噪声对语音的影响则重点体现于码书增益和固定码书.

基于上述特点,本文构建了一套针对 ITU-T G. 722.2 编码器的压缩域语音增强算法框架:首先,采用对数 LP 谱能量测度将背景噪声分为白噪声和有色噪声两大类;然后,利用能量准则联合调整码书增益;最后,在白噪声背景下进一步估计纯净语音的 ISF 参数,同时在有色噪声背景下对激励信号进行梳状滤波以增强谐波结构,并抑制谐波间噪声.

3 激励能量以及谱包络能量的联系

语音产生的“源-滤波器”模型如图 1 所示.

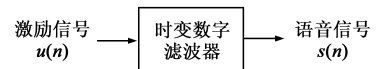


图1 语音产生的“源-滤波器”模型

图 1 中,时变数字滤波器的传递函数为:

$$H(z) = \frac{G}{1 - \sum_{i=1}^p \alpha_i z^{-i}} \quad (1)$$

其中, $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$ 是 LPC 系数, P 为滤波器阶数, G 是激励信号的增益,它与 CELP 编码器中各激励信号的关系如下:

$$G = \sqrt{\sum_{n=0}^{M-1} [g_c c(n) + g_p v(n)]^2} \quad (2)$$

其中, g_c 、 g_p 分别对应固定码书激励和自适应码书激励的增益, $c(n)$ 、 $v(n)$ 分别为固定码书激励以及自适应码书激励, M 为激励信号的帧长.

根据随机信号处理的相关知识,语音信号 $s(n)$ 的功率值 P_s 可以依据下式计算:

$$P_s = |H(z)|^2 P_u \quad (3)$$

其中, P_u 为激励信号 $u(n)$ 的功率值,其取值为 1. 将式 (1) 和式 (2) 带入, P_s 进一步推导为:

$$P_s = \left| \frac{1}{1 - \sum_{i=1}^p \alpha_i z^{-i}} \right|^2 \left\{ \sum_{n=0}^{M-1} [g_c c(n) + g_p v(n)]^2 \right\} \quad (4)$$

由式 (4) 可知,语音信号的能量由两部分构成,分别是激励能量和谱包络能量.因此,调整激励能量可以抵消由于谱包络调整而引入的能量失真.本文通过调整 CELP 编码器的码书增益来实现激励能量的调整,增益调整因子 λ 为:

$$\lambda = \frac{\sum_n [g_{cx} c_x(n) + g_{px} v_x(n)]^2}{\sum_n [g_{cy} c_y(n) + g_{py} v_y(n)]^2} \quad (5)$$

其中, g_{cx} 、 g_{px} 分别是纯净语音的固定码书增益和自适应码书增益, g_{cy} 、 g_{py} 分别是含噪语音的固定码书增益和自

适应码书增益, $c_x(n)$ 、 $v_x(n)$ 分别为纯净语音的固定码书激励以及自适应码书激励, $c_y(n)$ 、 $v_y(n)$ 分别为含噪语音的固定码书激励以及自适应码书激励。

在本文方法中,首先,将含噪语音、纯净语音的 ISF 参数,以及对应的增益调整因子构成特征矢量,并利用高斯混合模型拟合其概率密度;然后,利用 MMSE 贝叶斯估计器对纯净语音的 ISF 参数和增益调整因子进行估计。假设含噪语音的 ISF 参数矢量为 \mathbf{y} , 纯净语音的 ISF 参数矢量为 \mathbf{x} , 则本方法中高斯混合模型的特征矢量 $\mathbf{z} = (\mathbf{y}, \mathbf{x}, \lambda)$ 如图 2 所示。

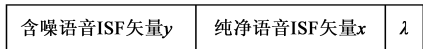


图2 GMM的特征矢量构成

4 纯净语音 ISF 参数的贝叶斯估计

实际应用中,背景噪声的引入导致无法获得纯净语音的 ISF 参数矢量 \mathbf{x} , 而只能得到含噪语音的 ISF 参数矢量 \mathbf{y} 。贝叶斯估计方法可以在含噪语音 ISF 参数矢量 \mathbf{y} 已知的前提下,依据 \mathbf{x} 的先验统计知识获得参数 \mathbf{x} 的估计值 $\hat{\mathbf{x}}$ 。使得代价函数 $C(\mathbf{x}, \hat{\mathbf{x}})$ 期望值 η_0 取值最小的 $\hat{\mathbf{x}}$ 为最优估计值^[12], 其中, η_0 的计算公式如下:

$$\eta_0 = \iint C(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (6)$$

式中, \mathbf{x} , \mathbf{y} 分别对应纯净语音、含噪语音特征集中的一个样本。依据条件概率和联合概率的关系,式(6)等价于:

$$\eta_0 = \int \{ \int C(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{y}) d\mathbf{x} \} p(\mathbf{y}) d\mathbf{y} \quad (7)$$

概率密度 $p(\mathbf{y})$ 具有非负性,因此,令 η_0 取值最小等价于令被积函数 η_1 取值最小, η_1 表示为:

$$\eta_1 = \int C(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{y}) d\mathbf{x} \quad (8)$$

采用欧式距离作为代价函数,即:

$$C(\mathbf{x}, \hat{\mathbf{x}}) = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) \quad (9)$$

求取 η_1 对 $\hat{\mathbf{x}}$ 的偏导函数并令其取值为 0, 可得到最小均方误差准则下的参数最优估计 $\hat{\mathbf{x}}_{opt}$:

$$\hat{\mathbf{x}}_{opt} = \int \mathbf{x} p(\mathbf{x} | \mathbf{y}) d\mathbf{x} \quad (10)$$

5 高斯混合模型

由式(10)可知,条件概率 $p(\mathbf{x} | \mathbf{y})$ 的准确性直接影响参数估计的准确度。所提算法利用高斯混合模型^[12]实现对概率密度函数的准确拟合。

高斯混合模型的理论基础是概率论的中心极限定理:在相当一般的条件下,当独立随机变量的个数增加时,其和的分布趋近于高斯分布。只要随机变量的个数充分大,就可以采用高斯分布近似逼近该随机变量的

分布情况。

假设 t 是具有任意分布的 N 维随机变量,其概率密度函数可由 K 个高斯概率密度函数的线性组合进行逼近,即:

$$P(t) = \sum_{i=1}^K w_i p_i(t) \quad (11)$$

其中, w_i 是第 i 个概率密度函数的权重,且满足:

$$\sum_{i=1}^K w_i = 1, \quad w_i \geq 0 \quad (12)$$

式(11)中的 $p_i(t)$ 为第 i 个高斯概率密度函数。为了有效拟合随机变量的概率密度函数,需要利用大量的数据进行训练以提取高斯混合模型的参数集。目前,估计高斯混合模型参数集最为成熟的算法是期望最大(Expectation and Maximization, EM)算法^[12]。

6 基于 GMM 的压缩域语音增强方法

图 3 所示为本文所提的基于 GMM 的压缩域语音增强方法的原理框图。

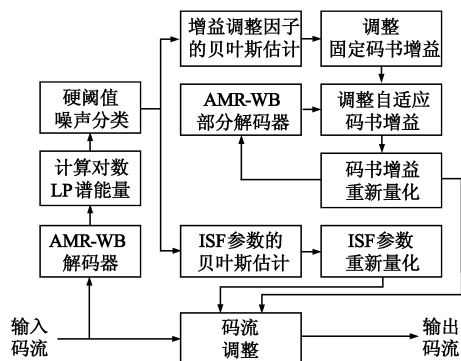


图3 基于GMM的压缩域语音增强方法的原理框图

算法包含 3 个主要模块。首先,利用对数 LP 谱能量特征将背景噪声分为有色噪声和白噪声两类;其次,在 MMSE 准则下估计纯净语音的 ISF 参数和增益调整因子;最后,利用调整因子对码书增益进行联合调整。

6.1 基于对数 LP 谱能量的噪声分类

不同噪声背景下,算法对激励信号的调整方法不同。本文采用对数 LP 谱能量测度对背景噪声类别进行区分,并分为白噪声和有色噪声两大类。选择这一测度主要考虑了参数的计算复杂度因素,因为多数的分类特征参数需在时域信号上提取,而这需解码器对码流进行全解码才可获得时域信号,这势必会增加计算负载。

对数 LP 谱能量和语音的短时合成滤波器密切相关,该滤波器的频率响应 $H(e^{j\omega})$ 可以表示为:

$$H(e^{j\omega}) = \frac{1}{1 - \sum_{i=1}^n \alpha_i e^{-j\omega i}} \quad (13)$$

对数 LP 谱能量 E_{LP} 的计算公式如下:

$$E_{LP} = \int_{-\pi}^{\pi} \log(|H(e^{j\omega})|^2) d\omega \quad (14)$$

图4给出了不同背景噪声对数LP谱能量的分布直方图。可以看出,白噪声、有色噪声(包括车内噪声、街道噪声)的对数LP谱能量的分布区间存在明显的区别,采用简单的硬阈值法即可实现背景噪声的分类。

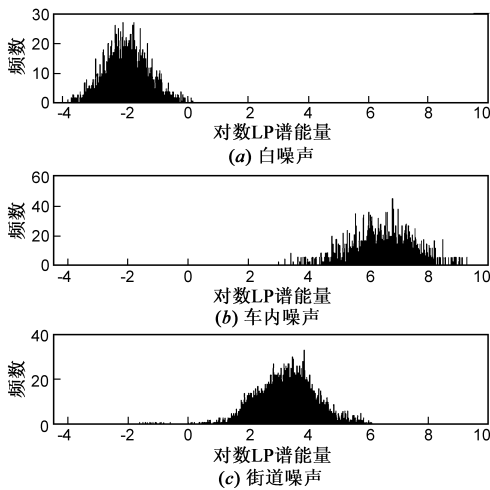


图4 不同噪声的对数LP谱能量分布直方图

6.2 基于高斯混合模型的 ISF 参数估计

根据前面的介绍,纯净语音 ISF 参数矢量 \mathbf{x} 的 MMSE 贝叶斯估计器 $\hat{\mathbf{x}}_{\text{opt}}$ 如式(10)所示. 高斯混合模型利用 K 个高斯概率密度的线性组合拟合随机变量的分布,式(10)可以进一步表示为:

$$\hat{\mathbf{x}}_{\text{opt}} = \sum_{i=1}^K p(\mathbf{z}_i | \mathbf{y}) E[\mathbf{x} | \mathbf{y}, \mathbf{z}_i] \quad (15)$$

其中, $\mathbf{z} = (\mathbf{y}, \mathbf{x}, \lambda)$ 是高斯混合模型的特征矢量, \mathbf{y} 为含噪语音的 ISF 参数矢量, λ 为增益调整因子。

假设变量 \mathbf{z} 的协方差矩阵为对角阵,即:

$$\Sigma_{\mathbf{x}, \mathbf{y}_i} = 0 \quad (16)$$

式中, $\Sigma_{\mathbf{x}, \mathbf{y}_i}$ 是第 i 个高斯基函数中变量 \mathbf{x} 、 \mathbf{y} 的互协方差矩阵. 则式(15)中的条件概率 $E[\mathbf{x} | \mathbf{y}, \mathbf{z}_i]$ 可表示为:

$$E[\mathbf{x} | \mathbf{y}, \mathbf{z}_i] = \mu_{\mathbf{x}_i} \quad (17)$$

其中, $\mu_{\mathbf{x}_i}$ 是第 i 个高斯基函数变量 \mathbf{x} 的均值矢量。

变量 \mathbf{x} 、 \mathbf{y} 和 \mathbf{z} 的均值矢量和协方差矩阵间的关系如下:

$$\mu_i = \begin{pmatrix} \mu_{\mathbf{y}_i} \\ \mu_{\mathbf{x}_i} \\ \mu_{\lambda_i} \end{pmatrix}, \Sigma_i = \begin{bmatrix} \Sigma_{\mathbf{y}, \mathbf{y}_i} & 0 & 0 \\ 0 & \Sigma_{\mathbf{x}, \mathbf{x}_i} & 0 \\ 0 & 0 & \Sigma_{\lambda_i} \end{bmatrix}$$

其中, μ_i 和 Σ_i 分别是第 i 个高斯基函数变量 \mathbf{z} 的均值矢量和协方差矩阵; μ_{λ_i} 和 Σ_{λ_i} 分别是增益调整因子 λ 的均值和方差; $\Sigma_{\mathbf{x}, \mathbf{x}_i}$ 和 $\Sigma_{\mathbf{y}, \mathbf{y}_i}$ 分别是变量 \mathbf{x} 和 \mathbf{y} 的协方差矩阵; $\mu_{\mathbf{y}_i}$ 是变量 \mathbf{y} 的均值矢量。

由此可知,式(15)可进一步化简为:

$$\hat{\mathbf{x}}_{\text{opt}} = \sum_{i=1}^K p(\mathbf{z}_i | \mathbf{y}) \mu_{\mathbf{x}_i} \quad (18)$$

式中,后验概率 $p(\mathbf{z}_i | \mathbf{y})$ 的计算公式为:

$$p(\mathbf{z}_i | \mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{z}_i)}{p(\mathbf{y})} = \frac{p(\mathbf{y}, \mathbf{z}_i)}{\sum_{j=1}^K p(\mathbf{y}, \mathbf{z}_j)} \quad (19)$$

其中, $p(\mathbf{y}, \mathbf{z}_i)$ 的计算公式为:

$$p(\mathbf{y}, \mathbf{z}_i) = \frac{w_i \exp\left[-\frac{1}{2}(\mathbf{y} - \mu_{\mathbf{y}_i})^T (\Sigma_{\mathbf{y}, \mathbf{y}_i})^{-1} (\mathbf{y} - \mu_{\mathbf{y}_i})\right]}{(2\pi)^{N/2} |\Sigma_{\mathbf{y}, \mathbf{y}_i}|^{1/2}} \quad (20)$$

6.3 基于 GMM 的增益调整因子估计

类似于 ISF 参数的估计方法,增益调整因子 λ 的 MMSE 贝叶斯估计器 $\hat{\lambda}$ 为:

$$\hat{\lambda} = \sum_{i=1}^K p(\mathbf{z}_i | \mathbf{y}) \mu_{\lambda_i} \quad (21)$$

6.4 码书增益的联合调整

采用能量准则联合调整含噪语音的固定码书增益和自适应码书增益. 增强语音固定码书激励能量与调整后的含噪语音固定码书激励能量相等,可得到增强语音的固定码书增益估计,即:

$$\lambda(m) \times \left[(g_c(m))^2 \sum_n c^2(n) \right] = (g'_c(m))^2 \sum_n c'^2(n) \quad (22)$$

其中, $g_c(m)$ 、 $g'_c(m)$ 分别对应含噪语音、增强语音第 m 个子帧的固定码书增益; $c(n)$ 和 $c'(n)$ 分别是含噪语音和增强语音当前子帧的固定码书激励。

由式(22)可得 $g'_c(m)$ 的计算公式:

$$g'_c(m) = g_c(m) \times \sqrt{\lambda(m) \times \frac{\sum_n c^2(n)}{\sum_n c'^2(n)}} \quad (23)$$

调整固定码书增益后,需进一步调整含噪语音的自适应码书增益,以避免增强语音的能量损失. 算法中,自适应码书增益调整准则是令增强语音总激励能量同调整后的含噪语音总激励能量相等,即:

$$\lambda(m) \sum_n [g_c(m)c(n) + g_p(m)v(n)]^2 = \sum_n [g'_c(m)c'(n) + g'_p(m)v'(n)]^2 \quad (24)$$

式中, $g_p(m)$ 和 $g'_p(m)$ 分别为含噪语音、增强语音第 m 个子帧的自适应码书增益; $v(n)$ 和 $v'(n)$ 分别是含噪语音和增强语音当前子帧的自适应码书激励. 求解式(24),若方程有解,较大的正数解即为增强语音的自适应码书增益;若方程无解,算法不对语音的自适应码书增益进行调整。

6.5 有色噪声背景下激励信号的梳状滤波

有色噪声背景下,引入梳状滤波来抑制激励信号

中残留的背景噪声.梳状滤波器具有增强谐波结构、抑制谐波间噪声的作用,其传递函数可以表示为^[13]:

$$H(z) = \frac{1-b}{1+a} \times \frac{1+az^{-T}}{1-bz^{-T}} \quad (25)$$

式中, T 为信号周期, a 、 b 是滤波器系数,其取值决定了滤波器的频率响应.图 5 给出了 a 、 b 取值不同时梳状滤波器的频率响应.

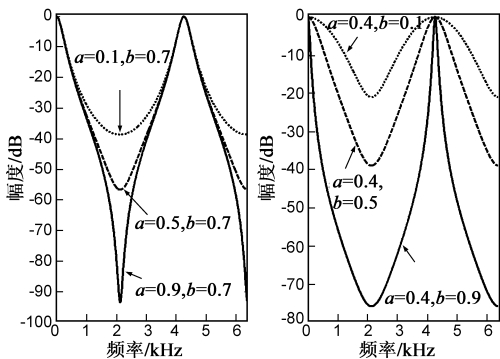


图5 梳状滤波器的频率响应

可以看出,系数 a 取值越大,滤波器波谷的平滑程度越差,对波谷的抑制作用越明显;而系数 b 的取值越大,滤波器的波峰结构越尖锐,对波峰附近区域的抑制作用越大.

算法中,滤波器系数依据语音信号的浊音度以及信噪比参数进行自适应调整.浊音度参数表征了语音信号的类浊音性,取值越高,信号的谐波结构越明显.梳状滤波器系数的调整准则如下:若信号的信噪比较低,系数 a 的取值应偏大,明显抑制谐波间噪声的同时,尽可能多地提取激励信号的谐波成分;若语音信号的浊音度较强,应适当降低系数 b 的取值,在保证谐波结构的同时抑制谐波间噪声.

7 DTX 模式下的压缩域语音增强方法

为了提高信道利用率,ITU-T G.722.2 编码器引入 DTX^[14,15] 技术以进一步降低信源编码的传输码率.若处理信号为语音,按正常模式编码并传输码流参数;若为非语音信号(静音或噪声),传输的编码参数有加权平均 ISF 参数 f^{mean} 以及加权对数帧能量 E_{\log}^{mean} ,分别表征背景噪声的谱特征和能量信息.针对编码器的 DTX 模式,本文所提压缩域增强方法的原理框图如图 6 所示.

针对非语音帧的编码参数,本文方法的调整原则为:不修正加权平均 ISF 参数的同时,按照式(26)调整

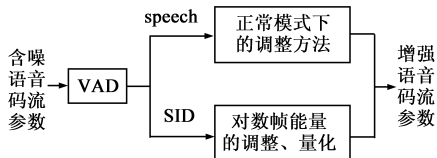


图6 针对DTX模式本文方法的原理框图

加权对数帧能量:

$$\hat{E}_{\log}(n) = 0.1 \times E_{\log}^{\text{mean}}(n) \quad (26)$$

式中, $\hat{E}_{\log}(n)$ 为第 n 帧调整后的噪声帧能量.

8 性能测试

测试项目均在 ITU-T G.160^[16] 标准的要求下进行.噪声类型包含白噪声、有色噪声(车内噪声及街道噪声)两类.文献[5]中的算法作为参考算法也参与了性能测试,简记为 REF.

实验中,语音信号均选自 NTT 标准语音库中文字子库,每段测试语音的时长均为 8s,且在测试前经过 -26dB 电平调整.高斯白噪声信号由 Cool Edit 软件获得;有色噪声信号(车内噪声、街道信号)选自 ITU-T 噪声库.所有信号的采样频率均为 16kHz.

高斯混合模型训练阶段,纯净语音信号选自 NTT 数据库中文字子库,各序列时长为 8s.背景噪声有白噪声、车内噪声及街道噪声,且包含 6dB、12dB 和 18dB 三种信噪比条件.每种噪声背景下的训练语音序列总数为 96,各信噪比条件下均含 32 个序列.高斯混合模型的高斯基个数为 32.

由于算法的处理对象为编码码流,为了排除由语音编码引入的干扰,测试所需的参考信号(含噪语音和纯净语音)均需经过 ITU-T G.722.2 编码器的编、解码处理.

8.1 信号电平衰减测试

在进行白噪声背景下的性能测试之前,首先要测试算法的期望噪声衰减量 Q_m .图 7 所示为针对正常模式的本文方法、DTX 模式所提方法以及 REF 算法期望噪声衰减量的对比.图中的横坐标表征 ITU-T G.722.2 编码器的 9 种编码模式,模式号越大,编码速率越高.可以看出,正常模式下本文方法的 Q_m 测度明显高于参考算法,说明本文算法对背景噪声的抑制效果更为明显.同时,本文所提针对 DTX 模式下的方法,其噪声抑制能力最优.

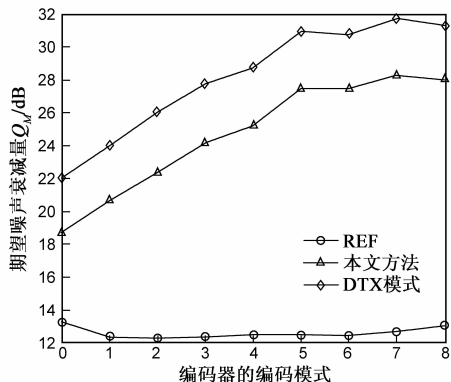


图7 期望噪声衰减量对比

信号电平衰减测试中, Q_{n1} 、 Q_{n2} 测度表征算法的实际噪声衰减量, Q_s 测度表征算法对语音信号电平的衰减. 测试结果如图 8 所示. 可以看出, 在非 DTX 模式下, 本文方法对噪声的抑制作用明显高于参考算法, 且高编码模式下对语音电平的衰减较低. 低速率模式下本文算法的 Q_s 偏高, 说明算法估计得到的 ISF 参数与纯净语音 ISF 的匹配程度仍有待提高. 同时, 本文所提针对 DTX 模式的方法, 与非 DTX 模式的方法相比, 其噪声抑制程度偏高, 语音电平衰减程度相当.

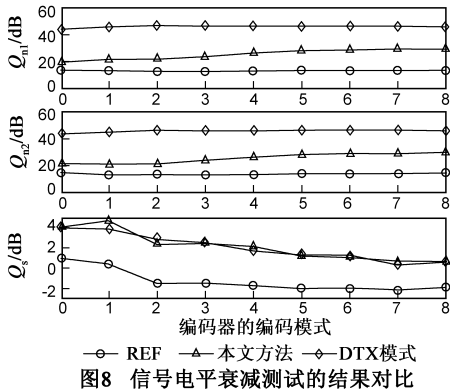


图8 信号电平衰减测试的结果对比

8.2 收敛性测试

收敛性测试主要检测算法能否在较短时间内明显抑制背景噪声, 测试结果如表 1, “—”表示算法不收敛. 可以看到, 除个别的码速率模式下参考算法不收敛, 三种方法的收敛时间均比较接近.

表 1 收敛性测试的结果对比

	/s	编码器的编码模式								
		0	1	2	3	4	5	6	7	8
本文算法	t_1	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
	t_2	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
	t_3	0	0	0	0.1	0	0	0	0	0
DTX 模式	t_1	1.4	1.4	1.5	1.5	1.5	1.5	1.4	1.4	1.5
	t_2	1.4	1.4	1.5	1.4	1.4	1.4	1.4	1.4	1.4
	t_3	0	0	0	0.1	0	0	0	0	0
参考算法	t_1	1.3	1.3	1.4	1.5	1.5	1.6	1.5	1.6	1.5
	t_2	1.0	1.1	—	—	1.3	1.2	—	—	—
	t_3	0	0	0	0	0	2.6	0	0	0

8.3 语音客观质量测试

选用 PESQ^[17] 提高 (PESQ Improvement, PESQI) 测度作为语音客观质量测试的指标, 测试结果如图 9 所示. 可以看出, 在各信噪比情况下, 本文所提方法的 PESQI 值均为正, 且除个别编码速率外, 其 PESQI 测度明显高于 REF 算法. 同时, 本文所提的针对 DTX 情况下的方法并不会明显降低合成语音的质量, 其性能接近正常模

式下的处理方法.

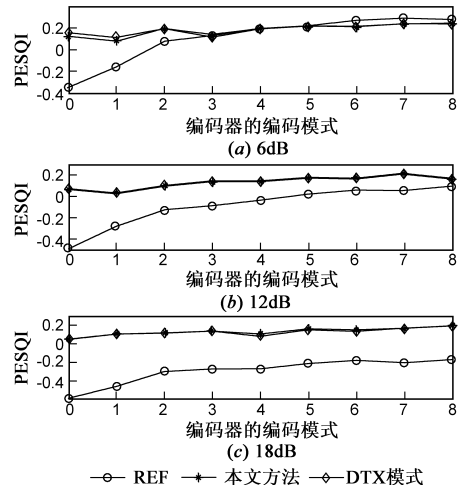


图9 不同信噪比条件下的PESQI结果对比

8.4 有色噪声背景下的性能测试

有色噪声背景下的测试指标包括 SNRI (SNR Improvement), 评价算法的信噪比提高能力; TNL (Total Noise Level Reduction), 评测算法的噪声抑制能力; DSN (SNRI-to-NPLR Difference, NPLR: Noise Power Level Reduction), 检测算法对语音信号的衰减作用. 图 10 所示为测试结果对比. 可以看出, 本文算法虽然信噪比提高能力稍差于参考方法, 但能够提供更大的噪声衰减量, 且对语音成分的损失与参考方法基本相当. 同时, 针对 DTX 模式的方法性能接近于正常模式下所提方法的性能.

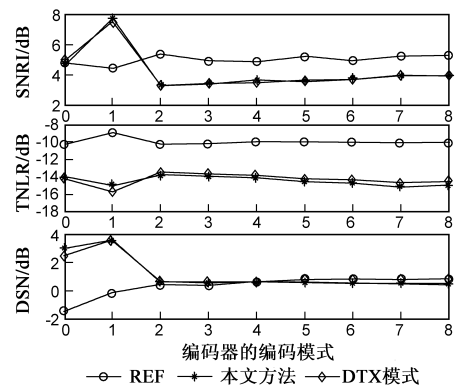


图10 有色噪声背景下的性能测试结果对比

9 帧擦除情况下的压缩域语音增强方法

当码流参数发生帧擦除情况时, ITU-T G.722.2 编码器的码流格式会发生变化, 并在解码端利用 FEC^[18] 算法恢复丢失的码流参数.

综合考虑解码端 FEC 算法对码流信息的利用情况以及本文方法的重点调整对象, 当帧擦除情况出现时, 所提方法并不调整码流参数, 也不修正当前帧的帧类型. 需要强调的是, 即使不对码流参数进行调整, 仍需

利用 FEC 算法恢复含噪语音的各码流参数,并按照正常情况下的方法调整编码参数,以达到更新相关历史的目的.帧擦除情况下本文方法的原理框图如图 11 所示.

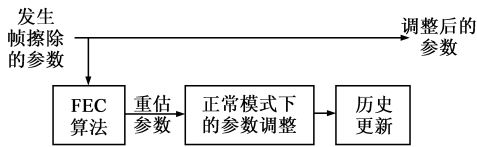


图11 帧擦除情况下本文方法的原理框图

选用 PESQI 测度检测帧擦除情况下本文所提方法的性能.纯净语音选自 NTT 数据库中中文子库,车内噪声、街道噪声选自 ITU-T 噪声库,白噪声由 Cool Edit 生成.信噪比条件包含 6dB、12dB 和 18dB,各种噪声背景每种信噪比条件下各包含 2 个测试序列.各语音信号均按编码模式 7(速率为 23.05kbps)编码后,以 3% 的错误率对码流进行帧擦除.测试结果如表 2 所示.

表 2 帧擦除情况下算法的性能测试结果

	PESQI		
	6dB	12dB	18dB
白噪声	0.314	0.239	0.254
	0.203	0.141	0.123
车内噪声	0.036	0.007	-0.007
	-0.03	0.048	-0.089
街道噪声	0.007	-0.031	0.004
	0.073	0.029	0.099
平均结果	0.301	0.22	0.19

可以看出,当帧擦除率为 3% 时,本文所提方法在白噪声背景下可以明显提高增强语音的客观质量;而在有色噪声背景下,除个别语句外,本文方法均可以适当提高增强语音的客观质量.

10 结论

为了有效地利用纯净语音谱包络参数的先验信息,本文利用高斯混合模型拟合纯净语音和含噪语音特征参数的联合概率密度,并在最小均方误差准则下对纯净语音的 ISF 参数进行贝叶斯估计,同时估计增益调整因子,然后利用该增益调整因子进一步联合调整固定码书增益和自适应码书增益.并且,算法可以兼容编码器的 DTX 模式以及传输网络中的帧擦除情况.利用 ITU-T G.160 标准对本文算法进行性能测试.结果表明,与参考算法相比,本文算法具有一定的信噪比提高能力,能够更有效的抑制背景噪声,且对语音信号电平的影响与参考算法相当,同时在白噪声背景下明显提高语音的客观质量.

参考文献

- [1] Ephraim Y. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1985, 33 (2): 443 - 445.
- [2] Ephraim Y. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984, 32(6): 1109 - 1121.
- [3] R Chandran, D J Marchok. Compressed domain noise reduction and echo suppression for network speech enhancement [A]. Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems [C]. MI, USA: IEEE Press, 2000. 10 - 13.
- [4] H Taddei, C Beaugeant, M de Meuleneire. Noise reduction on speech codec parameters [A]. Proceedings of IEEE International Conference Acoustic Speech Signal Process (ICASSP) [C]. NJ, USA: IEEE Press, 2004. 497 - 500.
- [5] N Duetsch, H Taddei, C Beaugeant, T Fingscheidt. Noise reduction on speech codec parameters [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04) [C]. USA: IEEE Press, 2004, Vol. 1. 497 - 500.
- [6] H M Goodarzi, S Seyedtabaii. Speech enhancement using spectral subtraction based on a modified noise minimum statistics estimation [A]. Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC [C]. Seoul, South Korea, 2009. 1339 - 1343.
- [7] R A Sukkar, R C Younce, Peng Zhang. Method and Apparatus for Voice Quality Enhancement [P]. USA: US2006215683-A1, 2006 - 09 - 28.
- [8] Cuntai Guan, Yongbin Chen, Boxiu Wu. Direct modulation on LPC coefficients with application to speech enhancement and improving the performance of speech recognition in noise [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [C]. Minneapolis, MN, USA: IEEE Press, 1993, Vol. 2. 107 - 110.
- [9] E T Fapi, C Beaugeant, H Taddei, D Pastor. Noise reduction within network through modification of LPC parameters [A]. Proceedings of the 7th International ITG Conference on Source and Channel Coding (SCC) [C]. Ulm, Germany, 2008. 1 - 6.
- [10] P Harding, B Milner. Speech enhancement by reconstruction from cleaned acoustic features [A]. Interspeech 2011 [C]. Florence, Italy, 2011. 1189 - 1192.
- [11] ITU-T Recommendation G.722.2. Wideband Coding of Speech at Around 16kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB) [S].
- [12] 刘鑫. 宽带音频的非线性频带扩展技术研究[D]. 北京: 北京工业大学, 2011. 21 - 34.

Liu Xin. Nonlinear Bandwidth Extending for Wideband Audio [D]. Beijing: Beijing University of Technology, 2011. 21 – 34. (in Chinese)

- [13] 胡广书. 数字信号处理导论[M]. 北京: 清华大学出版社, 2005. 294 – 296.
- [14] ITU-T Recommendation G. 722. 2 Annex A. Comfort Noise Aspects [S].
- [15] ITU-T Recommendation G. 722. 2 Annex B. Source Controlled Rate Operation [S].
- [16] ITU-T Recommendation G. 160. Voice Enhancement Devices for Mobile Networks [S].
- [17] ITU-T Recommendation P. 862. Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow Band Telephone Networks and Speech Codes [S].
- [18] ITU-T Recommendation G. 722. 2 Appendix I. Error Concealment of Erroneous or Lost Frames [S].

作者简介



梁岩 女, 1986年10月出生于山西朔州. 2009年进入北京工业大学攻读硕士学位. 主要研究方向: 压缩域语音增强.

E-mail: liangyan861003@163.com



鲍长春 男, 1965年出生于内蒙古赤峰市. 博士, 教授, 博士生导师. 主要研究领域为语音与音频信号处理.

E-mail: chcbao@bjut.edu.cn