

基于非均匀谱压缩特征的模型补偿新算法

宁更新 韦岗 孔祥祝

(华南理工大学电子与信息学院 广州 510640)

摘要: 在信噪比依赖的非均匀谱压缩(SNSC)鲁棒语音特征提取技术和 VTS 算法的基础上, 该文提出了一种新的 MC-SNSC 模型补偿算法。SNSC 技术是一种根据人类听觉对声音强度-响度感知转化关系的谱幅度变化操作和噪声抑制技术。基于对数谱域的噪声以及 SNSC 特征提取对语音信号特征所产生的失配函数, 推导出了 MC-SNSC 模型补偿算法。实验证明使用这一新算法, 识别率比当前较理想的 VTS 和 PMC 算法有很明显的提升, 算法的复杂度较 VTS 等算法仅有轻微的增加。

关键词: 语音识别; 模型补偿; 非均匀谱压缩

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)06-1384-05

Novel Model Compensation Based on Non-uniform Spectral Compression Features

Ning Geng-xin Wei Gang Kong Xiang-zhu

(School of Electronic & Information Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: A novel model compensation method is proposed, which integrates the Vector Taylor Series (VTS) approach with a robust feature extraction technique called SNR-dependent Non-uniform Spectral Compression (SNSC). The SNSC method is a spectral operation of magnitude transformation which resembles the human intensity-to-loudness conversion process and de-emphasizes noisy bands. Based on this mismatch function, which models the effect of the noise onto the clean speech in the Log-spectral domain together with the SNSC, a new model compensation procedure is derived. By adopting this novel model compensation approach, significant improvement over the PMC and VTS method can be found in different additive noisy environments at the expense of slight increase in computational complexity.

Key words: Speech recognition; Model compensation; Non-uniform spectral compression

1 引言

近年来, 人们对在噪声环境下的鲁棒语音识别技术起了广泛的兴趣。然而当识别器的训练环境与应用环境不同时, 其性能会急剧下降。为了解决这一问题, 各种技术方法陆续地被提出来, 这些技术方法主要分为 3 大类: (1) 语音鲁棒特征的表达和提取技术^[1], (2) 语音增强技术^[2]和(3)模型补偿技术^[3, 4]。关于这些方法的详细回顾参见文献[5]。本文的重点将放在基于鲁棒特征提取 SNSC(SNR-dependent Non-uniform Spectral Compression^[6])技术的新颖模型补偿算法。

文献[6]给出了一种使用 SNSC 方案的抗噪特征。它根据一子带信噪比依赖的压缩因子来压缩带噪子带语音功率谱。文献[6]证明采用 SNSC 技术提取的语音特征识别效果比用 MFCC 等特征有明显优越性。在 SNSC 中, 谱压缩是在线性谱域进行的, 压缩后的功率谱为

$$\tilde{Y}_k = Y_k^{\alpha_k}, \quad 0 \leq \alpha_k \leq 1 \quad (1)$$

其中 Y_k 是原始带噪语音功率谱, α_k 是谱压缩因子并且它是信

噪比依赖的, 下标 k 表示由第 k 个 Mel 尺度滤波器的输出, 即第 k 个子带。由于 α_k 是信噪比依赖的, 所以在特定噪声和信噪比条件下的模型训练就需要进行估计。因此在特定环境下训练出来的模型也只适用识别特定条件下的任务。为了克服这一问题, 本文提出一种新的 MC-SNSC (Model Compensation approach based on SNR-dependent Non-uniform Spectral Compression) 算法。该算法只需修正纯净语音模型, 产生出噪声补偿与压缩的模型来识别带噪语音。

2 非均匀谱压缩 SNSC 技术

依据人类听觉感知对噪声条件下的语音响度掩蔽效应, SNSC 技术对信噪比低的语音子带分量进行大的压缩。图 1 为基于 SNSC 技术的静态语音 MFCC 特征的提取流程图。当得到通过 Mel 频率尺度滤波器组的能量输出后, 子带能量谱就依照式(1)进行压缩。SNSC 技术的核心在于得到信噪比依赖的谱压缩因子 α_k 的映射。本文对文献[6]中的压缩映射方法作了一些的修改和简化, α_k 定义如下

$$\alpha_k = (1 - A_o) [1 - e^{-[\log(Y_k / N_k) - \beta] / \gamma}] u(\log(Y_k / N_k) - \beta) + A_o \quad (2)$$

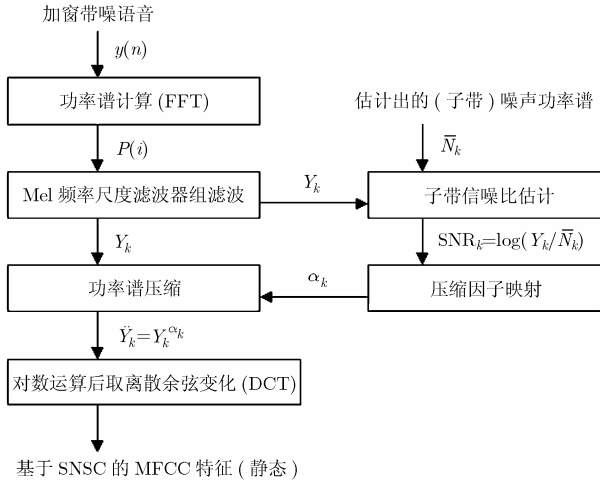


图 1 SNSC 方案流程图

其中 A_o 为压缩因子的下边界, β 为 SNR 的阈值, 表示正好听到的阈值, γ 用来调节压缩因子的变化快慢, $u(\cdot)$ 是单位阶跃函数。信噪比越小, 压缩因子也就相应地越小, 这样就可以减小带噪子带的语音的贡献, 也即减小不稳定子带信号的对识别的影响, 从而提高子带的信噪比。对式(2)的更详细解释可以参见文献[6]。值得注意的是在本文中信噪比的定义为 $\log(Y_k/N_k)$, 与文献[6]定义不同, 但这本质上没有改变 SNSC 技术的核心思想, 而且更有利于模型补偿算法的实现。

在文章中采用以下假设: (1) 噪声是平稳加性噪声, 噪声与语音信号是相互独立的, (2) 每个 Mel 子带的信号是相互对立的, 且 (3) 每个子带的对数频谱域功率分布被认为是混合高斯分布, 加性噪声不影响混合高斯分布的权值系数。

这样在对数谱域的压缩后的带噪语音谱和压缩因子分别为

$$\tilde{Y}_k^l = \alpha_k Y_k^l = \alpha_k \log(e^{X_k^l} + e^{N_k^l}) \quad (3)$$

$$\alpha_k = (1 - A_o) [1 - e^{-(Y_k^l - N_k^l - \beta)/\gamma}] \cdot u(Y_k^l - N_k^l - \beta) + A_o \quad (4)$$

式(3)描述了压缩后的带噪语音对数谱与纯净语音及噪声对数谱之间的失配关系, 亦即本文用来做模型补偿的失配函数。

纵贯全文我们定义和使用一致的域标号。上标 ‘ l ’ 表示对数功率谱域, 没有上标的则表示线性功率谱域。估计出的噪声模型参数用 ‘ \sim ’ 标记, 补偿出来的带噪语音模型参数用 ‘ $\hat{\cdot}$ ’ 标记。

3 模型补偿 MC-SNSC 算法

图 2 为基于 SNSC 特征提取技术和 MC-SNSC 补偿算法的识别系统流程图。在特征提取阶段, SNSC 方法采用压缩来压缩子带信号能量, 压缩因子的度量有各个子带的信噪比决定。然后将纯净语音模型和估计出的噪声模型通过使用 MC-SNSC 算法产生一个带噪语音特征模型。

在 MC-SNSC 方法中, 连续的函数被用来估计动态特征参数 $\left[\Delta \tilde{Y}_k^l = \frac{\sum_{m=-3}^3 m \Delta \tilde{Y}_k^l(t+m)}{\sum_{m=-3}^3 m^2} \right]$ 。这样在估计中, 动态特征失配函数就等于静态特征失配函数的一阶导数, 从而得到动态特征的失配函数式(5)。

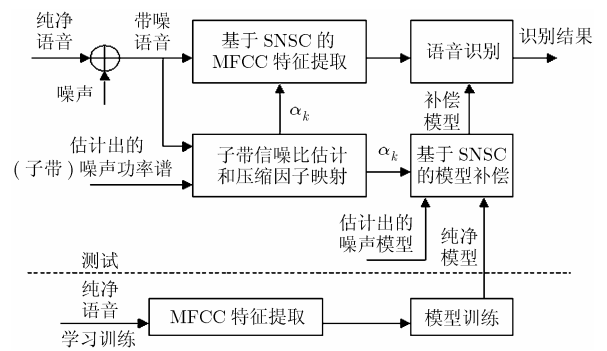


图 2 基于 SNSC 和 MC-SNSC 的识别系统流程图

$$\begin{aligned} \Delta \tilde{Y}_k^l &= \Delta(\alpha_k Y_k^l) = \frac{d(\alpha_k Y_k^l)}{dt} \\ &= A_o \frac{\partial Y_k^l}{\partial t} + (1 - A_o) \left[\left(1 - e^{-(Y_k^l - N_k^l - \beta)/\gamma} \right) \frac{\partial Y_k^l}{\partial t} \right. \\ &\quad \left. + \frac{1}{\gamma} e^{-(Y_k^l - N_k^l - \beta)/\gamma} Y_k^l \left(\frac{\partial Y_k^l}{\partial t} - \frac{\partial N_k^l}{\partial t} \right) \right] \cdot u(Y_k^l - N_k^l - \beta) \\ &\quad + \left(1 - e^{-(Y_k^l - N_k^l - \beta)/\gamma} \right) Y_k^l \cdot \delta(Y_k^l - N_k^l - \beta) \\ &\quad \cdot \left(\frac{\partial Y_k^l}{\partial t} - \frac{\partial N_k^l}{\partial t} \right) \end{aligned} \quad (5)$$

其中 $\delta(\cdot)$ 是迪拉克函数(Dirac delta function)。依据失配函数式(3)和式(5), 以及 Gauss-Hermite 数字积分, 可以推导出下面一系列模型补偿公式。

3.1 均值补偿

采用失配函数式(3), 在对数谱域的压缩带噪语音特征的统计均值为

$$\begin{aligned} \hat{\mu}_{\tilde{Y}_k^l} &= \hat{\mu}_{(\alpha Y)_k^l} = E \{ \alpha_k Y_k^l \} \\ &= (1 - A_o) E \{ Y_k^l u(Y_k^l - N_k^l - \beta) \} - E \left\{ e^{-(Y_k^l - N_k^l - \beta)/\gamma} \right. \\ &\quad \left. \cdot Y_k^l u(Y_k^l - N_k^l - \beta) \right\} + A_o E \{ Y_k^l \} \end{aligned} \quad (6)$$

为了便于表达本文定义, 令 $g(\gamma) = E \{ e^{-(Y_k^l - N_k^l - \beta)/\gamma} Y_k^l u(Y_k^l - N_k^l - \beta) \}$, 那么

$$\hat{\mu}_{\tilde{Y}_k^l} = (1 - A_o) [g(\infty) - g(\gamma)] + A_o \hat{\mu}_{Y_k^l} \quad (7)$$

其中 $g(\gamma) = e^{-[\Phi + \Psi / (2\gamma)] / \gamma} \left[\frac{\hat{\Sigma}_{Y_{kk}}^l}{\sqrt{2\pi\Psi}} e^{-[\Phi + \Psi / (2\gamma)]^2 / (2\Psi)} + \Omega \text{Sum}(\gamma) \right]$,

$\text{Sum}(\gamma) \cong \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \sum_{i=1}^n \omega_i \text{erf} \left[\sqrt{\frac{\hat{\Sigma}_{N_{kk}}^l}{\hat{\Sigma}_{Y_{kk}}^l}} t_i + \frac{\Phi + \Psi / \gamma}{\sqrt{2\hat{\Sigma}_{Y_{kk}}^l}} \right]$, $\Phi = \hat{\mu}_{N_k^l} - \hat{\mu}_{Y_k^l} + \beta$, $\Psi \cong \hat{\Sigma}_{N_{kk}}^l + \hat{\Sigma}_{Y_{kk}}^l$, $\Omega \cong \hat{\mu}_{Y_k^l} - \hat{\Sigma}_{Y_{kk}}^l / \gamma$, $\text{erf}(\cdot)$ 是误差函数, t_i 和 ω_i 分别是 n 阶 Gauss-Hermite 数字积分的横坐标和权值, 由 Hermite 多项式 $H_n(t)$ [7] 的根给定。

由于噪声是平稳的, 那么噪声信号动态特征的均值就为零。故依据失配函数式(5), 动态特征的均值补偿算法为

$$\begin{aligned} \hat{\mu}_{\Delta \tilde{Y}_k^l} &= \hat{\mu}_{\Delta(\alpha Y)_k^l} = E \{ \Delta(\alpha_k Y_k^l) \} \\ &= A_o \cdot \hat{\mu}_{\Delta Y_k^l} + (1 - A_o) \cdot \hat{\mu}_{\Delta Y_k^l} \cdot E \left\{ 1 - e^{-(Y_k^l - N_k^l - \beta)/\gamma} \right. \\ &\quad \left. \cdot Y_k^l u(Y_k^l - N_k^l - \beta) \right\} + \frac{1}{\gamma} g(\gamma) \end{aligned} \quad (8)$$

定义 $h(\gamma) = E\left\{e^{-(Y_k^l - N_k^l - \beta)/\gamma} \cdot u(Y_k^l - N_k^l - \beta)\right\}$, 那么

$$\hat{\mu}_{\Delta Y_k}^l = \left[A_o + (1 - A_o) \left[h(\infty) - h(\gamma) + \frac{1}{\gamma} g(\gamma) \right] \right] \hat{\mu}_{\Delta Y_k}^l \quad (9)$$

其中 $h(\gamma) = e^{-[\Phi + \Psi/(2\gamma)]/\gamma} \text{Sum}(\gamma)$ 。

3.2 方差补偿

同样根据式(3), 可以得出对数谱域的压缩噪声语音模型的静态方差补偿算法。其方差矩阵的对角线分量可以由下式计算:

$$\begin{aligned} \hat{\Sigma}_{Y_{kk}}^l &= E\left\{(\alpha_k Y_k^l)^2\right\} \\ &= (1 - A_o^2) E\left\{(Y_k^l)^2 u(Y_k^l - N_k^l - \beta)\right\} - 2(1 - A_o) \\ &\quad \cdot E\left\{e^{-(Y_k^l - N_k^l - \beta)/\gamma} (Y_k^l)^2 u(Y_k^l - N_k^l - \beta)\right\} + (1 - A_o)^2 \\ &\quad \cdot E\left\{e^{-2(Y_k^l - N_k^l - \beta)/\gamma} \cdot (Y_k^l)^2 \cdot u(Y_k^l - N_k^l - \beta)\right\} \\ &\quad + A_o^2 \left((\hat{\mu}_{Y_k}^l)^2 + \hat{\Sigma}_{Y_{kk}}^l \right) - (\hat{\mu}_{Y_k}^l)^2 \end{aligned} \quad (10)$$

定义 $f(\gamma) = E\left\{e^{-(Y_k^l - N_k^l - \beta)/\gamma} (Y_k^l)^2 u(Y_k^l - N_k^l - \beta)\right\}$, 那么

$$\begin{aligned} \hat{\Sigma}_{Y_{kk}}^l &= (1 - A_o^2) f(\infty) - 2(1 - A_o) f(\gamma) + (1 - A_o)^2 f(\gamma/2) \\ &\quad + A_o^2 \left((\hat{\mu}_{Y_k}^l)^2 + \hat{\Sigma}_{Y_{kk}}^l \right) - (\hat{\mu}_{Y_k}^l)^2 \end{aligned} \quad (11)$$

其中

$$\begin{aligned} f(\gamma) &= e^{-[\Phi + \Psi/(2\gamma)]/\gamma} \left[\frac{\hat{\Sigma}_{Y_{kk}}^l}{\sqrt{2\pi\Psi}} (\hat{\Sigma}_{Y_{kk}}^l \Phi / \Psi + 2\hat{\mu}_{Y_k}^l - \hat{\Sigma}_{Y_{kk}}^l / \gamma) \right. \\ &\quad \left. \cdot e^{-[\Phi + \Psi/(2\gamma)]^2 / (2\Psi)} + (\hat{\Sigma}_{Y_{kk}}^l + \Omega^2) \text{Sum}(\gamma) \right] \circ \end{aligned}$$

对于方差矩阵的非对角线分量的计算, 由于其中含有两重的 Hermite-Gauss 数字积分。为了降低算法的运算复杂度, 我们采用一近似的算法来获得其非对角线分量:

$$\hat{\Sigma}_{Y_{kl}}^l \approx \lambda_{kl} E\{\alpha_k\} E\{\alpha_l\} \hat{\Sigma}_{Y_{kk}}^l \quad (12)$$

其中 $\lambda_{kl} = \lambda_{lk} = \sqrt{\rho_{kl}\rho_{ll}}$, $\rho_{kl} = \hat{\Sigma}_{Y_{kl}}^l / \hat{\Sigma}_{Y_{kk}}^l$, 且 λ_{kl} 是一个调节因子, 调节非对角线分量使其能跟与之对应的对角线分量有相同的变化步调。

类似于对静态特征方差的估计, 使用失配函数式(5), 就可以估计出动态特征的方差补偿算法, 相对应的方差矩阵的对角线分量为

$$\begin{aligned} \hat{\Sigma}_{\Delta Y_{kl}}^l &= \left[A_o^2 + (1 - A_o^2) h(\infty) - 2(1 - A_o) h(\gamma) + (1 - A_o)^2 \right. \\ &\quad \cdot h(\gamma/2) + \frac{2}{\gamma} (1 - A_o) g(\gamma) - \frac{2}{\gamma} (1 - A_o)^2 \cdot g(\gamma/2) \\ &\quad \left. + \frac{1}{\gamma^2} (1 - A_o)^2 f(\gamma/2) \right] \left(\hat{\Sigma}_{\Delta Y_{kk}}^l + (\hat{\mu}_{\Delta Y_k}^l)^2 \right) + \frac{1}{\gamma^2} (1 - A_o)^2 \\ &\quad \cdot f(\gamma/2) \hat{\Sigma}_{\Delta N_{kk}}^l - (\hat{\mu}_{\Delta Y_k}^l)^2 \end{aligned} \quad (13)$$

非对角线分量为

$$\hat{\Sigma}_{\Delta Y_{kl}}^l \approx \lambda'_{kl} E\{\alpha_k\} E\{\alpha_l\} \hat{\Sigma}_{\Delta Y_{kk}}^l \quad (14)$$

其中 $\lambda'_{kl} = \lambda'_{lk} = \sqrt{\rho'_{kl}\rho'_{ll}}$, $\rho'_{kl} = \hat{\Sigma}_{\Delta Y_{kl}}^l / \hat{\Sigma}_{\Delta Y_{kk}}^l$ 。

3.3 非压缩带噪语音模型补偿的解决方法

上述的补偿公式中可见 MC-SNSC 算法需要一非压缩带噪语音的补偿模型 ($\hat{\mu}_{Y_k}^l, \hat{\Sigma}_{Y_{kk}}^l$) 及 ($\hat{\mu}_{\Delta Y_k}^l, \hat{\Sigma}_{\Delta Y_{kk}}^l$)。对于非压缩带噪语音的模型补偿的解决方法有很多如 PMC, VTS 等。文献[4]中研究了多种非压缩带噪语音的模型补偿算法, 发现 VTS 是一种比较理想的解决方法。因此本文采用一阶 VTS 方法 (VTS-1)^[3] 作为非压缩带噪语音模型补偿方案。

以上所述即是 MC-SNSC 算法, 子带的 SNSC 特征提取技术可以降低带噪语音段的信号对识别的影响, 亦即相当于提高子带的信噪比, 提高所提取语音特征的鲁棒性。新的 MC-SNSC 算法使用背景噪声模型和非压缩带噪语音模型 (如由 VTS、PMC 得到) 产生出一个新的压缩带噪语音模型。MC-SNSC 算法比较准确反映出压缩带噪语音信号同纯净语音信号及噪声信号的失配关系, 从而使识别工作能够用于不同的环境。

4 算法评估

实验采用基于孤立字的 6 状态 HMM 来做识别器。每个状态有 4 个高斯密度函数。选取 24 个倒谱系数 (12 个静态特征, 12 个动态特征) 作为识别矢量。训练阶段, 用纯净语音训练出纯净语音模型。在识别阶段, 使用纯净语音模型作为基本模型来识别。选用 TI-digits46 数据库中的 TI-20 作为算法评估语音库。TI-20 数据库由训练集和测试集组成, 他们都是由 16 个人 (8 男 8 女) 所说 20 个英文单词的话语。这 20 个英文单词包括数字 '0' 到 '9' 和 10 个附加命令如 'go', 'help' 等。我们选用训练库中的每个单词每人 2 遍话语来做训练集共 640 句; 测试库中所有 5082 句话作测试集。算法分析窗口的长度为 32ms, 帧速率为 9.6ms/帧。选取 NOISEX-92 中的 White, Pink 和 Factory 噪声作为评估的环境噪声。使用 200 帧非重叠的噪声来估计噪声模型。全局信噪比定义为

$$\text{SNR}_{\text{global}} = 10 \lg \left(\frac{\sum_{m=1}^H \sum_{k=0}^{L/2} P_m(k)}{\sum_{k=0}^{L/2} g^2 \bar{N}(k)} \right) \quad (15)$$

其中 $P_m(k)$ 是第 m 帧的纯净语音功率谱, $\bar{N}(k)$ 是估计的噪声能量平均谱, H 是每句的语音帧数, L 是 FFT 的长度, g 是缩放因子让所加的噪声符合指定的全局信噪比。带噪语音由式 (15) 生成。

$$y(i) = x(i) + g \cdot n(i) \quad (16)$$

其中 $y(i)$ 是带噪语音, $x(i)$ 和 $n(i)$ 分别是纯净语音和噪声。

为了比较 MC-SNSC 算法的性能, 采用 5 种方法识别: 2 种失配情况下的识别, Log-normal PMC 算法, VTS-1 算法和 MC-SNSC 算法。其中对于 Log-Normal PMC 的动态特征采用粗略的补偿方法, 只对其均值进行补偿^[4]。

$$\hat{\mu}_{\Delta Y_k}^l = \left(e^{\mu_{Y_k}^l} / e^{\hat{\mu}_{Y_k}^l} \right) \mu_{\Delta Y_k}^l \quad (17)$$

表 1 列出的是在不同噪声环境和信噪比条件下的各种算法的识别率。比较两种失配算法, 可以发现基于 SNSC 的 MFCC 特征比传统的 MFCC 在失配情况下有更好的识别率。

并且可以看出使用补偿算法的识别率比 2 种失配情况下有很大的提高。在 -5dB 条件下, MC-SNSC 的对各种噪声语音的平均识别率比 Log-normal PMC 和 VTS 有绝对的 18.6% 和 7.4% 增加, 在 0dB 情况下绝对增加值分别为 13.4% 和 5%。在信噪比 $5\text{--}10\text{dB}$ 下, MC-SNSC 的识别率比其他两种算法仍然有性能上的提高。

表 2 列出了更新单个高斯密度分布的静态和动态特征参数的乘法、除法、指数运算以及对数预算的次数。其中 N 和 M 分别表示在倒谱域和对数谱域特征的维数。从表中可看出文中所提几种方法运算复杂度相当。MC-SNSC 算法的复杂度比其他算法只有轻微的增加。

实验证明了我们的 MC-SNSC 算法可以处理在不同加性噪声环境下的语音识别任务, 并且能够取得比较好的识别效

果。性能的提升归功于新的特征的提取, 以及相对应比较准确的模型补偿方法的应用。通过使用这种方法, 识别率比目前的 PMC 和 VTS-1 算法有明显的提升。

5 结束语

本文提出了一个新颖的 MC-SNSC 模型补偿方法, 该方法整合了鲁棒特征提取 SNSC 技术和 VTS 模型补偿方法。文中给出了反映加性噪声和谱压缩影响的语音特征失配函数, 以及在此基础上推导出的一系列模型补偿算法。实验结果证明 MC-SNSC 方法可以处理不同加性噪声环境下的语音识别任务, 并且其识别性能比起当前流行的 VTS, PMC 等方法也有大幅度的提高。而 MC-SNSC 的算法复杂度与 VTS 等算法相当。

表 1 不同环境下各种方法的识别率(%)比较($\gamma = 0.9, A_p = 0.8, \beta = -0.3$)

环境	SNR(dB)	基于 MFCC 失配	基于 SNSC- MFCC 失配	Log-normal PMC	VTS-1	MC-SNSC
White 噪声	30	98.31	96.78	98.57	97.86	98.49
	10	51.89	70.89	93.97	94.49	97.10
	5	27.10	53.80	85.03	89.96	94.96
	0	14.28	32.56	73.41	81.60	89.39
	-5	6.74	15.95	51.61	60.29	71.26
Pink 噪声	30	98.80	97.06	98.76	98.37	98.35
	10	67.70	71.16	95.50	95.81	98.07
	5	29.80	47.64	90.29	93.47	95.97
	0	9.28	21.35	73.73	86.90	90.77
	-5	5.24	9.64	50.85	67.53	73.01
Factory 噪声	30	98.84	97.27	98.88	98.71	98.51
	10	75.31	75.50	96.15	96.26	97.62
	5	37.02	49.72	91.49	92.66	95.38
	0	11.83	25.98	80.46	84.50	87.69
	-5	5.22	12.52	54.04	62.28	68.03

表 2 文中所提及算法的复杂度

方法	运算复杂度(操作数)	$M=24, N=12, n=4$
Log-normal PMC	$2MN(2M+N+3)+6M^2+8M$	42950
VTS	$2MN(2M+N+3)+12M^2+13M$	46900
MC-SNSC	$2MN(2M+N+3)+16M^2+(3n+69)M$	51100

参考文献

- [1] 罗宇, 杜利民. 基于单高斯模型集的汉语美子带特征重建算法. 电子学报, 2004, 32(10): 1654–1657.
Luo Yu and Du Li-min. Single Gauss model set based MAP data imputation method for Mel-frequency filter-bank vectors of chinese speech. *Acta Electronica Sinica*, 2004, 32(10): 1654–1657.
- [2] Ding Pei and Cao Z G. An efficient robust ASR system based on the combination of speech enhancement and HMM adaptation. *Chinese Journal of Electronics*, 2002, 11(3): 422–425.
- [3] Acero A, Deng L, Kristjansson T, and Zhang J. HMM adaptation using vector Taylor series for noise speech recognition. in Proc. ICSLP'2000, Beijing, China, Oct. 2000: 869–872.
- [4] Hung J W, Shen J L, and Lee L S. New approach for domain transformation and parameter combination for improved accuracy in parallel model combination (PMC) techniques. *IEEE Trans. on Speech and Audio Processing*, 2001, 9(8):

- 842-854.
- [5] Gong Y. Speech recognition in noisy environments: A survey. *Speech Communication*, 1995, 16(3): 261-291.
- [6] Chu K K and Leung S H. SNR-dependent non-uniform spectral compression for noisy speech recognition. In Proc. ICASSP'04, Montreal, Canada, May 2004: 973-976.
- [7] Abramowitz M and Stegun I A. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications Inc., 1972.
- 宁更新: 男, 1981 年生, 博士, 研究方向为通信与语音信号处理等.
- 韦 岗: 男, 1963 年生, 教授, 博士生导师, 研究方向为现代通信理论与技术、多媒体信息处理、模式识别、神经网络等.
- 孔祥祝: 男, 1979 年生, 硕士, 从事语音识别和移动通信等方面的研究.