第 39 卷第 4 期　　声　学　学　报　　Vol. 39, No. 4
2014 年 7 月　　ACTA　ACUSTICA　　Jul., 2014

提高耳语音可懂度的非对称压缩语音增强方法*

周　健[1,2]　郑文明[3]　王青云[2]　赵　力[2]

(1　安徽大学计算智能与信号处理教育部重点实验室　合肥　230031)

(2　东南大学水声信号处理教育部重点实验室　南京　210096)

(3　东南大学儿童发展与学习科学教育部重点实验室　南京　210096))

2012 年 12 月 28 日收到

2013 年 5 月 6 日定稿

摘要　提出两种基于非对称代价函数的耳语音增强算法,将语音增强过程中的放大失真和压缩失真区分对待,Modified Itakura-Saito (MIS) 算法对放大失真给予更多的惩罚, 而 Kullback-Leibler (KL) 算法则对压缩失真给予更多的惩罚. 实验结果表明, 在低于 −6 dB 的低信噪比情况中, 经 MIS 算法增强后的耳语音的可懂度相比传统算法有显著提高; 而 KL 算法则获得了同最小均方误差语音增强算法近似的可懂度提高效果, 证实了耳语音中的放大失真和压缩失真对于耳语音可懂度的影响并不相同, 低信噪比时较大的压缩失真有助于提高耳语音可懂度, 而高信噪比时的压缩失真对耳语音可懂度影响较小。

PACS 数: 43.72

An asymmetric attenuated speech enhancement approach for improving intelligibility of noisy whisper

ZHOU Jian[1,2]　ZHENG Wenming[3]　WANG Qingyun[2]　ZHAO Li[2]

(1　*Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University*　Hefei　230031)

(2　*Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University*　Nanjing　210096)

(3　*Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University*　Nanjing　210096)

Received Dec. 28, 2012

Revised May 6, 2013

Abstract　Two asymmetric cost function for whispered speech enhancement methods are proposed. The cost of the amplification distortion and the attenuation distortion are different in both methods. The Modified Itakura-Saito (MIS) distance function gives more penalties to speech amplification distortion while the Kullback-Leibler (KL) divergence function gives more penalties to speech attenuation distortion. The experimental results show that the MIS method gains larger intelligibility improvement of the whispered speech than the conventional speech enhancement algorithms in much lower Signal to Noise Ratio (SNR) less than −6 dB, and the KL method has similar intelligibility improvement performance to the Minimum Mean Square Error (MMSE) speech enhancement method. The results confirm that the amplification distortion and the attenuation distortion have different effects on the intelligibility of the enhanced whisper. Specifically, larger attenuation distortion can improve speech intelligibility in lower SNR condition and it has a little influence on speech intelligibility in high SNR condition.

* 国家自然科学基金 (61301295, 61231002, 61273266, 61003131) 、安徽省自然科学基金 (1308085QF100, 1408085MF113) 和安徽大学博士科研启动经费资助

引言

耳语是人类之间一种特殊的言语交流方式。耳语与正常音发音不同，人发耳语音时声带不振动，因此耳语信号中基音频率缺失，且声能较正常音低约 20 dB 左右。初期的耳语音研究主要停留在语音基础研究和基础医学需要[1]，但是随着科学技术的发展，耳语音的研究逐渐走向实际应用，现有的基于耳语音的应用研究包括耳语音的识别、耳语音转换为正常音、耳语音的情感分析等等[2−5]。耳语音的相关研究成果可以应用于移动通讯、远程电话客户服务、安全场所的身份识别、犯罪鉴定等多个方面。对于喉部切除的失音患者，如能将其发出的气声自动识别出来，无需电子喉就能转换为正常音，对于每年上万人数量增长的失音患者来说，提供了一种更容易被接受的语言交流方式。耳语音研究已日益引起国内外研究机构和科研人员的重视。

现有语音增强方法主要以提高语音感知质量为目的。由于忽视了可懂度指标，这些算法并不能提高语音的可懂度[6−7]。我们采用常见的 4 类语音增强算法 (谱减法、幅度谱估计法、子空间法、维纳滤波法) 进行耳语音可懂度提高实验，实验结果表明，这些算法同样无法提高耳语音可懂度。分析不难发现，基于均方差准则的语音增强算法对于语音压缩失真和放大失真不加区分地给予相同的惩罚力度。比如维纳滤波法采用语音估计谱与真实谱的均方误差最小化方法来计算语音估计谱; 短时谱幅度估计方法采用语音估计谱幅度与真实谱幅度的均方误差最小化方法来获得语音估计谱幅度; 而对数谱幅度估计方法则是将估计谱幅度的对数与真实谱幅度的对数的均方误差最小化来求解谱幅度等等。然而，均方误差准则是一个误差的平方函数，估计谱正向或负向偏离真实谱时均方误差函数赋予的惩罚值是相同的，没有对正向偏离和负向偏离加以区分。文献 8 研究表明：放大失真会降低增强后语音的可懂度，而适当的压缩失真则对语音可懂度没有太大影响。

本文提出两种非对称语音增强方法，其中 Kullback-Leibler (KL) 散度代价函数对语音压缩失真给予较大的惩罚，而 Modified Itakura-Saito (MIS) 代价函数对语音放大失真给予较大的惩罚，对压缩失真则给予较小的惩罚。此外，鉴于现有的经典噪声谱估计方法 (比如最小统计 (Minimum Statistics, MS) 、基于最小控制的递归平均 (Minima Controlled Recursive Algorithm, MCRA)) 在语音出现段停止对噪声谱进行更新的缺点，本文还提出了一种在语音段仍然能实时动态更新噪声谱估计的方法。

1 基于非对称代价函数的语音谱估计方法

本文讨论两种非对称代价函数，它们分别来源于 Kullback-Leibler (KL) 散度[16] 和 Itakura-Saito (IS) 距离[17]。设经过分帧后第 l 帧含噪语音为 $y_l(n) = x_l(n) + d_l(n)$，其中 $x_l(n)$ 和 $d_l(n)$ 分别表示不相关的干净耳语音和噪声。令 $Y(\omega_{k,l})$, $X(\omega_{k,l})$ 和 $D(\omega_{k,l})$ 分别表示 $y_l(n)$, $x_l(n)$, $d_l(n)$ 的短时傅里叶频谱，相应的幅度分别为 $Y_{k,l}$, $X_{k,l}$ 和 $D_{k,l}$, $\widehat{X}_{k,l}$ 表示干净语音谱幅度估计。下面首先给出两种非对称代价函数定义。

基于 KL 散度的非对称代价函数为：

$$d_{\mathrm{KL}}(\widehat{X}_{k,l}, X_{k,l}) = \widehat{X}_{k,l} \ln \frac{\widehat{X}_{k,l}}{X_{k,l}} - \widehat{X}_{k,l} + X_{k,l}. \quad (1)$$

基于修正 IS (Modified Itakura-Saito, MIS) 距离的非对称代价函数为：

$$d_{\mathrm{MIS}}(\widehat{X}_{k,l}, X_{k,l}) = \mathrm{e}^{(X_{k,l} - \widehat{X}_{k,l})} - (X_{k,l} - \widehat{X}_{k,l}) - 1. \quad (2)$$

与最小均方误差函数不同，式 (1) 对压缩失真给予较多的惩罚代价，而式 (2) 对放大失真给予较多的惩罚代价。图 1 显示了 $d_{\mathrm{KL}}(X, X_0)$, $d_{\mathrm{MIS}}(X, X_0)$ 和均方误差代价函数 $d_{SE}(X, X_0)(X_0 = 5)$ 等三种代价函数曲线。从图 1 可以看出，均方误差代价函数 $d_{\mathrm{SE}}(X, X_0)$ 对于 X_0 的过估计值和欠估计值采用相同的惩罚值，而 $d_{\mathrm{KL}}(X, X_0)$ 和 $d_{\mathrm{MIS}}(X, X_0)$ 对于 X_0 的过估计值和欠估计值采用了不同的惩罚值。其中 $d_{\mathrm{KL}}(X, X_0)$ 代价函数对于 $X < X_0$ 的估计子具有更大的惩罚值，与此相反，$d_{\mathrm{MIS}}(X, X_0)$ 代价函数对于 $X > X_0$ 的估计子具有更大的惩罚值。

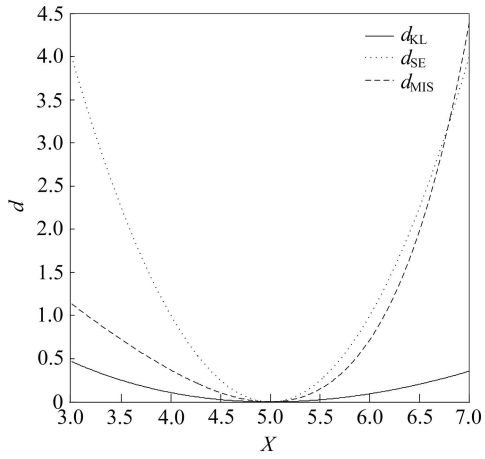

图 1 三类不同的代价函数

令 $\lambda_x(k,l)=E\{X_{k,l}^2\}$, $\lambda_d(k,l)=E\{D_{k,l}^2\}$ 分别表示第 l 帧纯净耳语音及噪声在数字频率 k 处的谱方差。式 (1) 的贝叶斯风险期望值为 [18]:

$$\begin{aligned}\Re_{\mathrm{KL}_B}=E\{d_{\mathrm{KL}}(\widehat{X}_{k,l},X_{k,l})\}=\iint d_{\mathrm{KL}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l},Y(\omega_{k,l}))\mathrm{d}X_{k,l}\mathrm{d}Y(\omega_{k,l})=\\ \int\left[\int d_{\mathrm{KL}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l}\right]p(Y(\omega_{k,l}))\mathrm{d}Y(\omega_{k,l}),\end{aligned}\tag{3}$$

在 $p(Y(\omega_{k,l}))$ 已知条件下，要使 $\Re_{\mathrm{KL}_B}$ 最小，只要使式 (3) 的内层积分最小。令

$$\Phi_{\mathrm{KL}}=\int d_{\mathrm{KL}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l},$$

并对 $\widehat{X}_{k,l}$ 求偏导，且令偏导值为 0, 解得：

$$\widehat{X}_{k,l}=\left\{\frac{\xi_{k,l}}{\xi_{k,l}+1}\exp\left\{\frac{1}{2}\int_{v_{k,l}}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right\}\right\}Y_{k,l},\tag{4}$$

式 (4) 中，

$$\xi_{k,l}=\frac{\lambda_x(k,l)}{\lambda_d(k,l)},\ \nu_{k,l}=\frac{\gamma_{k,l}\xi_{k,l}}{1+\xi_{k,l}},\ \gamma_{k,l}=\frac{|Y_{k,l}|^2}{\lambda_d(k,l)}.$$

式 (4) 和文献 9 中的对数谱估计式相同，这表明对数谱估计具有非对称压缩特性。

当误差代价函数取

$$d_{\mathrm{MIS}}(\widehat{X}_{k,l},X_{k,l})=\mathrm{e}^{(\widehat{X}_{k,l}-X_{k,l})}-(\widehat{X}_{k,l}-X_{k,l})-1$$

时，其对应的风险代价为 [18]:

$$\begin{aligned}\Re_{\mathrm{MIS}_B}=E\{d_{\mathrm{MIS}}(\widehat{X}_{k,l},X_{k,l})\}=\iint d_{\mathrm{MIS}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l},Y(\omega_{k,l}))\mathrm{d}X_{k,l}\mathrm{d}Y(\omega_{k,l})=\\ \int\left[\int d_{\mathrm{MIS}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l}\right]p(Y(\omega_{k,l}))\mathrm{d}Y(\omega_{k,l}),\end{aligned}\tag{5}$$

设

$$\Phi_{\mathrm{MIS}}=\int d_{\mathrm{MIS}}(\widehat{X}_{k,l},X_{k,l})p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l},\ \text{令}\ \frac{\partial\Phi_{\mathrm{MIS}}}{\partial\widehat{X}_{k,l}}=0,$$

解得：

$$\widehat{X}_{k,l}=-\ln\int_0^{\infty}\mathrm{e}^{-X_{k,l}}p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l},\tag{6}$$

根据贝叶斯准则，

$$\int_0^{\infty}\mathrm{e}^{-X_{k,l}}p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l}=\frac{\int_0^{\infty}\mathrm{e}^{-X_{k,l}}\int_0^{2\pi}p(Y(\omega_{k,l})|X_{k,l},\theta_{k,l})p(X_{k,l},\theta_{k,l})\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}}{\int_0^{\infty}\int_0^{2\pi}p(Y(\omega_{k,l})|X_{k,l},\theta_{k,l})p(X_{k,l},\theta_{k,l})\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}}=$$

$$\frac{\int_0^{\infty}\mathrm{e}^{-X_{k,l}}\int_0^{2\pi}\frac{1}{\pi\lambda_d(k,l)}\exp\left\{-\frac{1}{\lambda_d(k,l)}|Y(\omega_{k,l})-X(\omega_{k,l})|^2\right\}\frac{X_{k,l}}{\pi\lambda_x(k,l)}\exp\left\{-\frac{X_{k,l}^2}{\lambda_x(k,l)}\right\}\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}}{\int_0^{\infty}\int_0^{2\pi}\frac{1}{\pi\lambda_d(k,l)}\exp\left\{-\frac{1}{\lambda_d(k,l)}|Y(\omega_{k,l})-X(\omega_{k,l})|^2\right\}\frac{X_{k,l}}{\pi\lambda_x(k,l)}\exp\left\{-\frac{X_{k,l}^2}{\lambda_x(k,l)}\right\}\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}}=$$

$$\frac{\int_0^{\infty}X_{k,l}\mathrm{e}^{-X_{k,l}}\exp\left\{-\left(\frac{1}{\lambda_x(k,l)}+\frac{1}{\lambda_d(k,l)}\right)X_{k,l}^2\right\}\int_0^{2\pi}\exp\left(2\mathrm{Re}\left(\frac{X_{k,l}Y_{k,l}}{\lambda_d(k,l)}\mathrm{e}^{-\mathrm{j}\theta_{X_{k,l}}}\right)\right)\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}}{\int_0^{\infty}X_{k,l}\exp\left\{-\left(\frac{1}{\lambda_x(k,l)}+\frac{1}{\lambda_d(k,l)}\right)X_{k,l}^2\right\}\int_0^{2\pi}\exp\left(2\mathrm{Re}\left(\frac{X_{k,l}Y_{k,l}}{\lambda_d(k,l)}\mathrm{e}^{-\mathrm{j}\theta_{X_{k,l}}}\right)\right)\mathrm{d}\theta_{k,l}\mathrm{d}X_{k,l}},$$

令

$$\frac{1}{\lambda(k,l)}=\frac{1}{\lambda_d(k,l)}+\frac{1}{\lambda_x(k,l)},$$

则

$$\int_0^\infty \mathrm{e}^{-X_{k,l}}p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l}=\frac{\int_0^\infty X_{k,l}\mathrm{e}^{-X_{k,l}}\exp\left\{-\left(\frac{X_{k,l}^2}{\lambda(k,l)}\right)\right\}\mathrm{I}_0\left(2X_{k,l}\sqrt{\frac{v_{k,l}}{\lambda(k,l)}}\right)\mathrm{d}X_{k,l}}{\int_0^\infty X_{k,l}\exp\left\{-\left(\frac{X_{k,l}^2}{\lambda(k,l)}\right)\right\}\mathrm{I}_0\left(2X_{k,l}\sqrt{\frac{v_{k,l}}{\lambda(k,l)}}\right)\mathrm{d}X_{k,l}}=$$

$$\frac{\int_0^\infty\sqrt{\left(\frac{\pi X_{k,l}}{2}\right)}[\mathrm{I}_{-0.5}(X_{k,l})-\mathrm{I}_{0.5}(X_{k,l})]X_{k,l}\exp\left\{-\frac{X_{k,l}^2}{\lambda(k,l)}\right\}\mathrm{I}_0\left(2X_{k,l}\sqrt{\frac{v_{k,l}}{\lambda(k,l)}}\right)}{\int_0^\infty X_{k,l}\exp\left\{-\left(\frac{X_{k,l}^2}{\lambda(k,l)}\right)\right\}\mathrm{I}_0\left(2X_{k,l}\sqrt{\frac{v_{k,l}}{\lambda(k,l)}}\right)\mathrm{d}X_{k,l}}=$$

$$\exp(-v_{k,l})\sum_{m=0}^\infty\frac{1}{m!}(v_{k,l})^m\mathrm{F}\left(-m,-m,\frac{1}{2};\frac{\lambda(k,l)}{4v_{k,l}}\right)-$$

$$\exp(-v_{k,l})\sqrt{\lambda(k,l)}\sum_{m=0}^\infty\frac{\Gamma(m+1.5)}{m!\,\Gamma(m+1)}(v_{k,l})^m\mathrm{F}\left(-m,-m,\frac{3}{2};\frac{\lambda(k,l)}{4v_{k,l}}\right),$$

上式中， $\mathrm{I}_v(\cdot)$ 是第一类修正贝塞尔函数， $\mathrm{F}(a,b,c;x)$ 是高斯超几何函数， $\Gamma(\cdot)$ 是伽马函数。

根据 $\sqrt{\lambda(k,l)}=\sqrt{v_{k,l}}Y_{k,l}/\gamma_{k,l}$, 上式可以写为：

$$\begin{aligned}\int_0^\infty \mathrm{e}^{-X_{k,l}}p(X_{k,l}|Y(\omega_{k,l}))\mathrm{d}X_{k,l}&=\exp(-v_{k,l})\sum_{m=0}^\infty\frac{1}{m!}(v_{k,l})^m\mathrm{F}\left(-m,-m,\frac{1}{2};\frac{Y_{k,l}^2}{4\gamma_{k,l}^2}\right)-\\&\exp(-v_{k,l})\frac{\sqrt{v_{k,l}}}{\gamma_{k,l}}Y_{k,l}\sum_{m=0}^\infty\frac{\Gamma(m+1.5)}{m!\,\Gamma(m+1)}(v_{k,l})^m\mathrm{F}\left(-m,-m,\frac{3}{2};\frac{Y_{k,l}^2}{4\gamma_{k,l}^2}\right).\end{aligned}\tag{7}$$

3 改进的噪声谱估计方法

在进行噪声谱估计时，目前应用较多的是最小统计方法 [19] 和最小控制递归平均方法 [20]。这两个算法在语音出现阶段保持噪声谱和前一帧的噪声谱不变。然而，在非平稳噪声背景下语音出现时噪声谱仍然具有动态变化特点，采用这两个算法可能会使某些语音段出现噪声谱的欠估计，并导致增强后的语音中残留了更多的噪声，降低了语音可懂度。为此，本节将介绍一种修正的基于 MCRA 噪声谱估计方法，该方法在语音出现时仍然对噪声谱进行动态更新，以降低语音段噪声谱的欠估计程度。

假设第 l 帧含噪语音和噪声在数字频率 k 处的频谱幅度 $Y(k,l)$ 和 $D(k,l)$ 分别服从高斯分布，$H_0(k,l)$ 表示时频点 (k,l) 处无语音出现， $H_1(k,l)$ 表示时频点 (k,l) 处语音出现，因此无语音出现及有语音出现的含噪语音谱的概率密度函数分别为：

$$f(Y(k,l)|H_0(k,l))=\frac{1}{\pi\lambda_d(k,l)}\exp\left\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)}\right\},$$

$$f(Y(k,l)|H_1(k,l))=\frac{1}{\pi(\lambda_x(k,l)+\lambda_d(k,l))}$$

$$\exp\left\{-\frac{|Y(k,l)|^2}{\lambda_x(k,l)+\lambda_d(k,l)}\right\},$$

由于

$$\gamma_{k,l}=\frac{|Y(k,l)|^2}{\lambda_d(k,l)},\ \xi_{k,l}=\frac{\lambda_x(k,l)}{\lambda_d(k,l)},$$

因此 [10],

$$f(\gamma_{k,l}|H_0(k,l))=\mathrm{e}^{-\gamma_{k,l}}\mu(\gamma_{k,l}),$$

$$f(\gamma_{k,l}|H_1(k,l))=\frac{1}{1+\xi_{k,l}}\exp\left\{-\frac{\gamma_{k,l}}{1+\xi_{k,l}}\right\}\mu(\gamma_{k,l}),$$

其中， $\mu(.)$ 是单位阶跃函数。

令 $p(k,l)=P(H_1(k,l)|\gamma_{k,l})$, 则有 [10]:

$$p(k,l)=\left\{1+\frac{q(k,l)}{1-q(k,l)(1+\xi_{k,l})\exp(-v_{k,l})}\right\}^{-1},$$

其中，$q(k,l)=P(H_0(k,l))$。

当时频点 $(k,l+1)$ 无语音出现时，噪声谱采用如下迭代式进行更新：

$$\overline{\lambda}_d(k,l+1)=\alpha_d\overline{\lambda}_d(k,l)+(1-\alpha_d)|Y(k,l+1)|^2 \quad (8)$$

当时频点 $(k,l+1)$ 出现语音时，噪声谱采用如下迭代式进行更新：

$$\overline{\lambda}_d(k,l+1)=\sum_{i=-w}^{i=w}b(i)\overline{\lambda}_d(i,l), \quad (9)$$

其中，$\sum_{i=-w}^{i=w}b(i)=1$。结合语音出现概率，噪声谱更新式可以表示为：

$$\begin{aligned}&\overline{\lambda}_d(k,l+1)=\\&(\alpha_d\overline{\lambda}_d(k,l)+(1-\alpha_d)|Y(k,l+1)|^2)(1-p(k,l))+\\&\left(\sum_{i=-w}^{i=w}b(i)\overline{\lambda}_d(i,l)\right)p(k,l),\end{aligned} \quad (10)$$

式 (10) 是噪声谱的有偏估计，通常采用补偿因子 β 进行偏差补偿，最终的噪声谱估计式为：

$$\widehat{\lambda}_d(k,l)=\beta\overline{\lambda}_d(k,l),$$

其中：

$$\beta=\frac{\lambda_d(k,l)}{E\{\overline{\lambda}_d(k,l)\}}.$$

4　可懂度提高实验及讨论

4.1　噪声谱估计

本文的噪声谱估方法在语音未出现时，噪声谱估计性能同 MCRA 方法，能够无偏地估计噪声谱，但在语音出现时，MCRA 方法停止噪声谱估计，语音段的噪声谱采用语音出现前一寂静帧的噪声谱估计，而本文方法在有声段仍然能够较好地动态更新噪声谱估计。

图 2(a) 显示了本文提出的噪声谱估计方法以及采用 MS 和 MCRA 方法估计的噪声谱 (DFT 频率轴 $k=16$)。从图 2 可以看出，MS 方法估计的噪声谱要偏低于真实的噪声谱 (即经过平滑后的功率谱)。图 2(b) 是将图 2(a) 中的噪声谱替换为含噪语音谱之后的结果，从图 2(b) 可以看出，采用本文提出的噪声谱估计算法虽然得到了偏高的噪声谱估计，但是并没有超越语音谱，因此，在进行谱增强时，不会造成较严重的语音压缩失真。

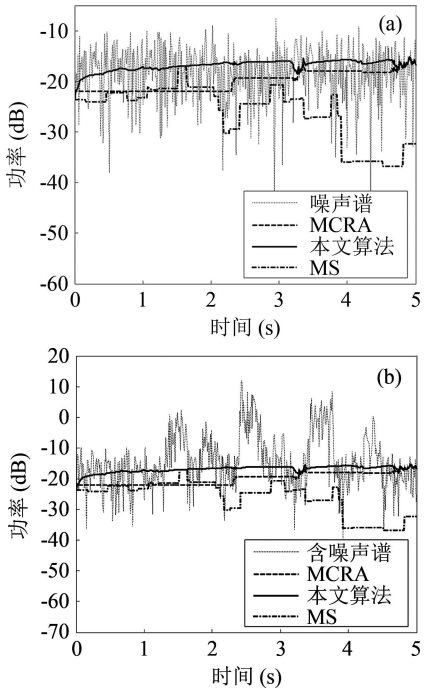

图 2　本文方法同 MS 方法和 MCRA 方法的噪声谱估计比较

4.2　耳语音可懂度提高实验

在采用本文提出的噪声谱估计方法提取噪声谱后，我们将第 2 节导出的基于非对称代价函数的语音增强算法应用于耳语音增强。为了评估算法的性能，我们采用 50 句内容不同的语料用于产生耳语音。每句分别由 3 男 3 女在安静环境中录制。耳语波形采用 16 kHz 采样率，每个采样点 16 bit, PCM 存储。

每句耳语音和噪声混合成信噪比分别为 −15 dB, −12 dB, −9 dB, −6 dB, 0 dB, 3 dB, 6 dB 的含噪耳语音。我们采用 NOISEX-92 数据库中的高斯白噪声、F16 飞机噪声、Babble 噪声和 M109 坦克噪声等四类噪声 [11]。在合成含噪语音时，从噪声中随机截取一段与耳语音长度相同的噪声段，合成指定信噪比的含噪耳语音。对增强前后的耳语音，我们采用主观和客观相结合方式评价增强前后耳语音的可懂度，其中客观评价采用短时客观可懂度 (Short-Time Objective Intelligibility, STOI) 指数 [12]，主观评价采用清晰度测试方式。

为了比较不同代价函数对于耳语音可懂度提高的影响，我们还选取 MMSE 方法 [13] 和 Loizou 教授提出的 IS 算法 [14]，其中 MMSE 算法对压缩失真和放大失真不加区分，IS 算法同 KL 算法一样，对于压缩失真给予较大的惩罚，但 KL 算法对压缩失真给予的惩罚力度相比 IS 算法要小，而本文提出的 MIS 算法对放大失真给予较大的惩罚。图 3 给出了四种噪声环境下不同算法增强后的耳语音可懂度平均 STOI 值。

从图 3 可以看出，在 −15 dB 和 −12 dB 的高斯噪声环境下，经 MIS 算法增强的耳语音的可懂度比 IS, KL 和 MMSE 算法要高。而且在 −15 dB 至 −6 dB 信噪比范围内，在 F16 噪声、Babble 噪声和 M109 噪声背景下，经 MIS 算法增强后耳语音的 STOI 值远大于含噪耳语音，实验结果表明 MIS 算法在低信噪比时能够有效提高含噪耳语音可懂度。但当信噪比增大时，比如信噪比大于 0 dB 时，经 MIS 算法增强后的耳语音的 STOI 值相比含噪耳语音则没有显著提高，在高斯噪声环境和 Babble 噪声环境下甚至会逐渐低于含噪耳语音的 STOI 值。这表明当信噪比较高时，使用 MIS 算法则无法提高耳语音的可懂度。

另外，在清晰度测试中，我们组织 3 男 3 女进行增强前后耳语音试听。实验选择 Babble, F16 和高斯噪声环境，信噪比选择 −12 dB, −6 dB, 0 dB, 6 dB 。实验结果如表 1 所示。

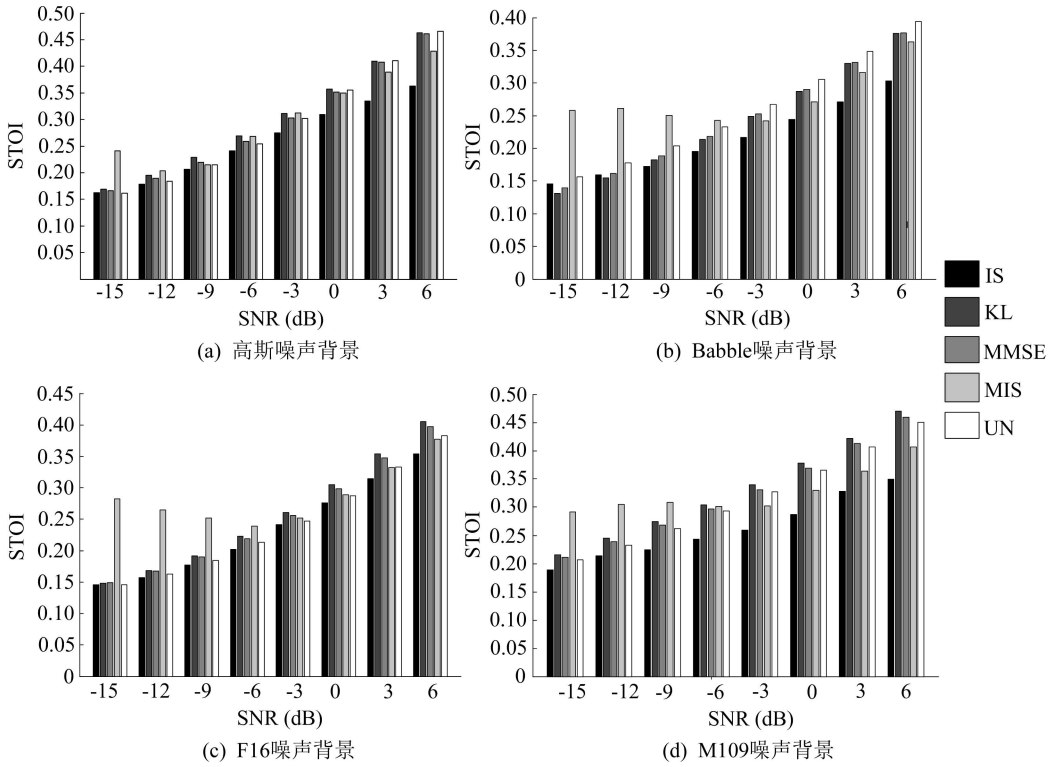

图 3 不同背景噪声下语音增强算法的 STOI 指标比较

表 1 主观听辨实验结果

SNR (dB)	噪声类型	主观听辨词语识别率 (%)				
		IS	KL	MMSE	MIS	UN
−12	Babble	19.10	18.55	19.44	**31.33**	21.34
	F16	18.88	20.16	20.11	**31.80**	19.48
	高斯	21.36	23.38	22.73	**24.40**	22.09
−6	Babble	33.16	36.31	37.16	**41.32**	39.64
	F16	34.29	37.86	37.23	**40.61**	36.21
	高斯	41.05	44.71	44.03	**45.59**	43.30
0	Babble	51.24	60.38	60.98	56.89	64.18
	F16	58.04	64.07	62.77	60.61	60.31
	高斯	64.91	74.05	73.82	73.37	74.57
6	Babble	66.64	82.76	82.87	79.88	86.75
	F16	78.19	89.19	87.47	83.50	89.28
	高斯	79.75	92.17	91.29	92.18	93.24

从表 1 可以看出，在 −12 dB 和 −6 dB 时 MIS 算法均有效提高了增强后耳语音词语识别率。例如，在 −12 dB Babble 环境下，经 MIS 算法增强后的耳语音可懂度相比未去噪的耳语音提高了约 9.9% 。而 KL 算法则在不同的测试环境中均未有效提高耳语音可懂度。

MIS 算法在信噪比很低的情况下能提高增强后的耳语音可懂度，而在信噪比较高时无法提高增强后的耳语音的可懂度，原因是在信噪比很低时，噪声占主导，此时若产生较大的压缩失真，则实际上会压缩掉更多的噪声成分，使得增强后的语音可懂度提高；而在较高信噪比时，含噪耳语音中语音的能量占主导，较大的压缩操作则会造成语音频谱成分压缩失真，从而降低了耳语音可懂度，从主观及客观评价结果可以看出，在信噪比较高时，虽然这种压缩失真会降低耳语音可懂度，但是影响不明显。

实验中，经 MIS 算法增强后的耳语音的 STOI 值均比 IS 算法要大， IS 算法在所有的信噪比条件下增强后的耳语音的 STOI 值均要低于未增强的含噪耳语音的 STOI 值，且在较低的信噪比条件下，两者不明显，信噪比越高，IS 和 MIS 所增强的耳语音的 STOI 值的差距越大。这是因为在过低的信噪比条件下， IS 虽然产生放大失真，但是这种放大失真在很低信噪比时对语音的放大至多不会超过含噪耳语音，即不可能比含噪声耳语音中出现的放大失真还要大，所以此时 IS 所获得的 STOI 值基本上和含噪耳语音相同，而当信噪比逐渐增大时，相比 MMSE 而言， IS 的放大失真引入了更多的噪声，从而降低了增强后的耳语音的可懂度。这进一步证明，放大失真对可懂度的影响比压缩失真要显著。

此外，从图 3 和表 1 还可以看出：KL 算法与 MMSE 算法在提高耳语音可懂度方面性能相同，且均未有效提高耳语音可懂度。这是由于 KL 方法虽然在一定程度上鼓励放大失真，但是这种鼓励放大的力度非常小，这从图 1 可以看出，即在干净谱的一定偏差邻域范围内，KL 对于放大失真的鼓励几乎可以忽略不计。KL 算法的实验结果也解释了为何基于对数谱估计的语音增强算法无法提高耳语音可懂度 (注意 KL 算法和 logMMSE[9] 算法具有相同的谱估计增益函数) 。

图 4 显示了 −6 dB 高斯噪声环境经不同算法增强后的耳语音的时域波形，其中图 4(a) 是一段干净耳语音，图 4(b) 是含噪耳语音，信噪比为 −6 dB，图 4(c) 是经 MIS 算法增强后的耳语音，图 4(d) 是经谱减法增强后的耳语音，图 4(e) 是经 IS 算法 [14] 增强后的耳语音，图 4(f) 是经 MMSE 算法增强后的耳语音，图 4(g) 是经 KL 算法增强后的耳语音，图 4(h) 是经维纳滤波算法 [15] 增强后的耳语音。从图 4 可以看出，经 MIS 算法增强后的耳语音中的残留噪声更少。

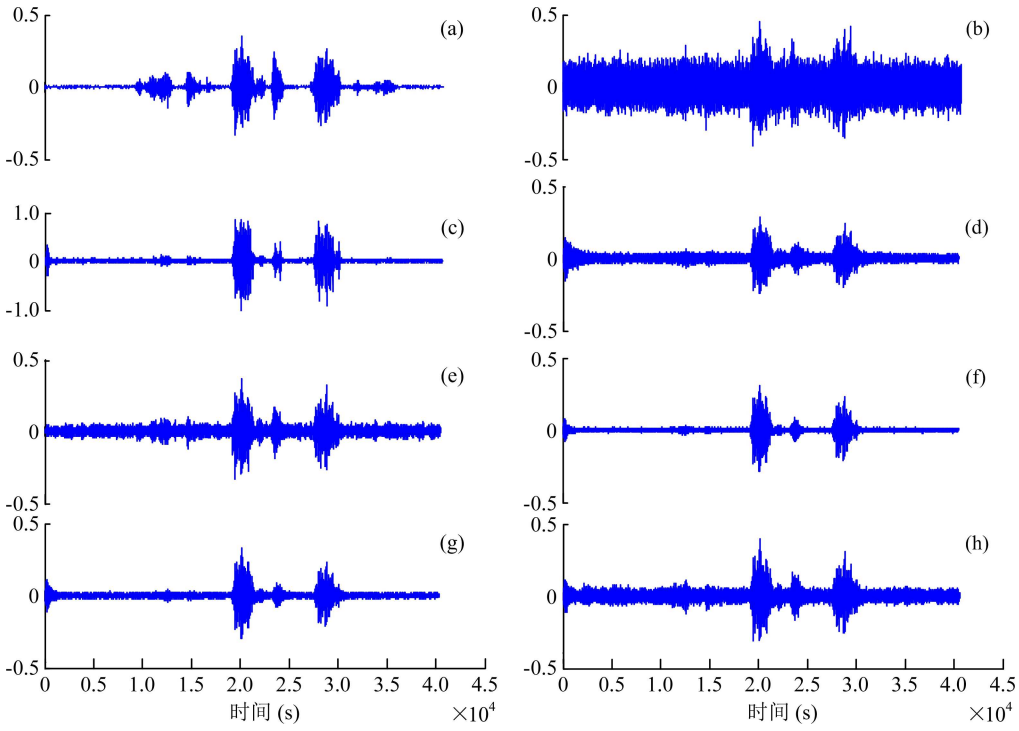

图 4　−6 dB 高斯噪声环境下不同算法增强后的耳语音

5 结论

传统的语音增强算法以提高语音听觉舒适度和愉悦度为目的，且谱域增益函数利用均方差作为准则，没有考虑语音放大失真和压缩失真对语音可懂度的影响；另外，传统的噪声谱估计在语音活动段维持噪声谱不变，语音段噪声谱欠估计会使得增强后的语音出现更多的残留，降低了增强后的耳语音可懂度。

本文提出了两种基于非对称代价函数的耳语音增强算法，它们对于压缩失真和放大失真区分对待，其中 MIS 算法对于放大失真给予较大的惩罚，而 KL 算法则适度鼓励放大失真。算法推导结果证明 KL 算法导出的谱增益函数与对数谱估计法导出的增益函数相同，这表明基于均方差的对数谱增益函数在一定程度上鼓励了放大失真，因而无法提高耳语音可懂度。本文还对传统的 MCRA 噪声谱估计方法进行了改进，使得语音出现部分噪声谱仍然能动态更新。

实验结果表明，信噪比小于 −6 dB 时通过 MIS 算法增强后的耳语音的可懂度相比传统算法均有显著提高，而且其增强后的耳语音的可懂度要大于未经处理的含噪耳语音；而 KL 算法则获得了同 MMSE 算法近似的可懂度提高效果，进一步证实耳语音中的放大失真和压缩失真对于耳语音可懂度的影响并不相同，低信噪比时较大的压缩失真有助于提高耳语音可懂度，而高信噪比时压缩失真对耳语音可懂度影响较小。

参 考 文 献

1 Tartter V C. Identifiability of vowels and speakers from whispered syllables. *Attention, Perception, & Psychophysics*, 1991; **49**(4): 365—372

2 王敏，赵鹤鸣. 基于多带解调分析和瞬时频率估计的耳语音话者识别. 声学学报， 2010; **35**(4): 471—476

3 陶智，赵鹤鸣，谈雪丹，顾济华，张晓俊，吴迪. 采用扩展型双线性变换法将耳语音转换为正常语音的研究. 声学学报， 2012; **37**(6): 651—658

4 顾晓江，赵鹤鸣，吕岗. 模型与特征混合补偿法及其在耳语说话人识别中的应用. 声学学报， 2012; **37**(2): 198—203

5 Jin Yun, Zhao Yan, Huang Chengwei, Zhao Li. Study on the emotion recognition of whispered speech. In: Zhou Shangming, Wang Wenwu ed. GCIS2009, Proceedings of WRI Global Congress on Intelligent Systems, Xiamen, China, 2009, Piscataway, NJ: IEEE, 2009: 242—246

6 Li Junfeng, Yang Lin, Zhang Jianping, Yan Yonghong. Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English. *The Journal of the Acoustical Society of America*, 2011; **129**(5): 3291—3301

7 杨琳，张建平，颜永红. 单通道语音增强算法对汉语语音可懂度影响的研究. 声学学报， 2010; **35**(2): 248—253

8 Loizou P C, Kim G. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011; **19**(1): 47—56

9 Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1985; **33**(2): 443—445

10 Cohen I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, 2003; **11**(5): 466—475

11 Varga A, Steeneken H J M. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, 1993; **12**(3): 247—251

12 Taal C, Hendriks R, Heusdens R *et al.* An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011; **19**(7): 2125—2136

13 Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1984; **32**(6): 1109—1121

14 Loizou P C. Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum. *IEEE Transactions on Speech and Audio Processing*, 2005; **13**(5): 857—869

15 Scalart P, Filho J V. Speech enhancement based on a priori signal to noise estimation. In: ICASSP96, IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, Georgia, 1996, Piscataway, NJ: IEEE, 1996: 629—632

16 Kullback S, Leibler R. On Information and sufficiency. *The Annals of Mathematical Statistics*, 1951; **22**(1): 79—86

17 Fumitada I, Shuji S. A statistical method for estimation of speech spectral denstity and formant frequencies. *Electronics and Communications*, 1970; **53-A**: 36—43

18 Hazewinkel M. Encyclopedia of mathematics. Berlin, German: 2001: 287—352

19 Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, 2001; **9**(5): 504—512

20 Cohen I. Noise spectrum estimation in adverse environment: improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, 2003; **11**(5): 466—475