第45卷 第8期 西 安 交 通 大 学 学 报 Vol. 45 No. 8
2011年8月 JOURNAL OF XI'AN JIAOTONG UNIVERSITY Aug. 2011

失真控制下的短时谱估计语音增强算法

刘晓明，班超帆，冯晓荣
（重庆大学通信工程学院，400030，重庆）

摘要：针对传统谱估计增强算法易产生语音畸变、导致语音清晰度低的问题，提出了一种失真控制下的短时谱估计语音增强的新算法. 该算法首先引入语音畸变的客观度量参数，并根据这一参数得到抑制语音畸变的约束条件，然后结合人耳听觉掩蔽特性和无语音概率参数，修正最小均方误差对数谱估计函数，最后联立约束条件和估计函数，得到增强后的语音，从而实现了在噪声抑制和语音畸变之间的折中，改善了语音增强的效果. 主观试听和客观测试结果均表明，与其他谱减法相比，在相同的信噪比和去噪度条件下，新算法的语音畸变度最小且几乎察觉不到音乐噪声.

关键词：语音增强；噪声抑制；语音清晰度；失真控制
中图分类号：TN912.35 **文献标志码**：A **文章编号**：0253-987X(2011)08-0078-07

A Short-Time Spectrum Estimation Algorithm of Speech Enhancement under the Distortion Control

LIU Xiaoming, BAN Chaofan, FENG Xiaorong
(College of Communication Engineering, Chongqing University, Chongqing 400030 China)

Abstract: A novel short-time spectrum estimation algorithm of speech enhancement based on distortion control is presented to overcome the speech distortion and low speech intelligibility in conventional method. An objective measurement parameter of speech distortion is introduced in the algorithm and a constraint condition is generated according to the parameter. Then, a modified log spectral amplitude estimation algorithm is proposed based on the speech absence probability and masking properties of the human auditory system. Finally, the enhanced speech is obtained by relating the constraint condition with the estimation algorithm. The new algorithm gets a better performance and the trade-off between noise reduction and speech distortions is achieved. Tests on objective measurements combined with informal subjective listening show that the proposed algorithm has a better performance of speech intelligibility without any perceptional musicality than other spectral subtraction algorithms under the same level of SNR and noise reduction.

Keywords: speech enhancement; noise reduction; speech intelligibility; distortion control

语音处理系统的性能在有噪环境中会急剧下降，一种有效的预处理技术是语音增强. 目前常用的语音增强算法是基于短时谱估计的增强算法. 此类算法运算简单，物理意义清晰，近几十年来国内外学者对其进行了深入的研究，并提出了各种算法，如功率谱算法(power spectral, PS)[1]、最小频谱均方误差谱算法(minimum mean square error-short time spectral amplitud, MMSE-STSA)[2]和最小对数谱均方误差谱算法(minimum mean square error-log spectral amplitude, MMSE-LSA)[3]等.

上述谱估计算法着眼于如何更好地从带噪语音中估计出噪声和语音的功率谱密度，从而能更好地跟踪噪声和语音的变化，在去除噪声和减小语音畸变之间折中，并取得了一定的效果.

收稿日期：2011-01-04. 作者简介：刘晓明(1963—)，男，教授. 基金项目：国家自然科学基金资助项目(51035008).
网络出版时间：2011-05-24 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20110524.1151.004.html

本文从另外的角度来分析如何在减小噪声的同时提高去噪语音的清晰度，减小语音畸变. 前面提到的语音增强算法都是通过代价函数，使纯净语音和估计语音的幅度谱或者功率谱的均方误差最小，从而得到估计语音的幅度估计. 但是，纯净语音和估计语音的差值有正负之分，正的差值对应衰减失真，负的差值对应放大失真. 人耳对放大失真和衰减失真的反应是不一样的，简单地使用均方误差这一度量将混淆两者之间的差别. 本文对估计差的正值和负值进行分析和控制，对赵晓群等人提出的谱减语音增强算法[4]进行修正，期望能在去噪度、残留音乐噪声和语音畸变之间均衡. 客观测试和主观试听结果都表明，本文算法在非平稳噪声干扰下的性能要优于传统的增强算法，语音畸变度最小，特别是在低信噪比输入情况下效果尤为显著.

1　短时谱估计算法

本文主要讨论时域和频域都与语音信号完全重叠的宽带加性噪声. 假设带噪语音的时域表示为

$$y(t)=x(t)+d(t) \tag{1}$$

式中：$x(t)$、$d(t)$和 $y(t)$分别为在 t 时刻的纯净语音、加性噪声和带噪语音信号. 由于短时谱估计是按帧进行的，将式(1)写成帧的形式为

$$y(n,i)=x(n,i)+d(n,i),\quad 0\leqslant i\leqslant N-1 \tag{2}$$

式中：$x(n,i)$、$d(n,i)$和 $y(n,i)$分别为第 n 帧中第 i 个样点的纯净语音、加性噪声和带噪语音信号；N 为帧长. 对式(2)表示的每帧信号进行离散傅里叶变换，得到相应的频域表示

$$Y(n,k)=X(n,k)+D(n,k) \tag{3}$$

式中：$X(n,k)$、$D(n,k)$和 $Y(n,k)$分别为第 n 帧、第 k 个频点的纯净语音、加性噪声和带噪语音信号的频谱. 利用语音在 10～30 ms 分析帧内的短时平稳性，以及人耳对语音相位失真感知不灵敏的特性[5]，只对纯净语音的谱幅度进行估计，而把带噪语音的相位直接作为增强语音的相位. 因此，在不会引起混淆的情况下，$Y(n,k)$、$X(n,k)$、$D(n,k)$分别表示第 n 帧第 k 个频点的带噪语音、纯净语音和噪声的幅度谱. 用 $\hat{X}(n,k)$和 $G(n,k)$分别表示第 n 帧第 k 个频点的估计的语音幅度谱和谱估计函数，最终短时谱估计算法可写成如下形式

$$\hat{X}(n,k)=G(n,k)\,Y(n,k) \tag{4}$$

2　基于两类失真控制的语音信号谱估计函数

2.1　语音畸变的客观度量

对于一段完整的语音信号，第 k 个频点的纯净语音的幅度谱 $X(k)$和估计语音的幅度谱 $\hat{X}(k)$的关系表示为：$X(k)=\hat{X}(k)+R(k)$，其中 $R(k)$为第 k 个频点残差信号的幅度谱. 语音增强算法产生语音畸变的实质是[6]在某些频点上 $X(k)>\hat{X}(k)$，而在另外一些频点上 $X(k)<\hat{X}(k)$. 这就导致估计语音整个频谱的包络与纯净语音不一样，引起语音的一些重要的声学特性(比如元音的共振峰)发生改变，于是产生语音畸变，语音的清晰度也不高.

衡量语音畸变的大小最直接的参量是相关系数. 定义 $X(k)$和 $\hat{X}(k)$的相关系数为

$$\rho(k)=\frac{\operatorname{cov}(X(k),\hat{X}(k))}{\sigma_{X(k)}\sigma_{\hat{X}(k)}}=\frac{E[X(k)\hat{X}(k)]-\mu_{X(k)}\mu_{\hat{X}(k)}}{\sigma_{X(k)}\sigma_{\hat{X}(k)}}=\frac{\sigma^2_{X(k)}+\mu^2_{X(k)}-\mu_{X(k)}\mu_{\hat{X}(k)}}{\sigma_{X(k)}(\sigma^2_{X(k)}+\mu^2_{X(k)}+\sigma^2_{R(k)}+\mu^2_{R(k)}-\mu^2_{\hat{X}(k)})^{1/2}} \tag{5}$$

式中：$\mu_{X(k)}$、$\mu_{\hat{X}}(k)$和 $\mu_{R(k)}$分别为 $X(k)$、$\hat{X}(k)$和 $R(k)$的期望；$\sigma^2_{X(k)}$、$\sigma^2_{\hat{X}(k)}$和 $\sigma^2_{R(k)}$分别为 $X(k)$、$\hat{X}(k)$和 $R(k)$的方差. 对于所有的频点 k 都有 $X(k)=\alpha\hat{X}(k)$(α 为常数)时，就没有语音畸变，此时 $\rho(k)=1$，否则 $\hat{X}(k)$包含语音畸变，此时 $0<\rho(k)<1$，且 $\rho(k)$越趋近于 0，语音畸变越严重. 改进语音增强算法时，希望各个频点的 $\rho(k)$都尽可能地接近 1，这样即使不同的频点上 $R(k)$正负性不一样，总的纯净语音和估计语音的频谱包络形状相差也不会很大，语音畸变小.

由于 $\rho(k)$定义复杂且不便与语音增强算法结合，本文引入频域的分段的信噪比度量，并把它定义为信号与残差信号的比例(R_{ESI})，第 k 个频点的 R_{ESI} 定义式如下

$$R_{\mathrm{ESI}}(k)=\frac{\sum_{n=1}^{n}X^2(n,k)}{\sum_{n=1}^{n}(X(n,k)-\hat{X}(n,k))^2}=\frac{X^2(k)}{(X(k)-\hat{X}(k))^2} \tag{6}$$

式中：n 为帧数.

本文采用 $R_{ESI}(k)$ 作为语音清晰度的度量，下面讨论 $\rho(k)$ 和 $R_{ESI}(k)$ 的关系.

(1) 当 $\mu_{X(k)}=\mu_{\hat{X}(k)}=0$ 时，$R_{ESI}(k)=\frac{\sigma^2_{\hat{X}(k)}}{\sigma^2_{R(k)}}$，$\rho^2(k)=\frac{\sigma^2_{\hat{X}(k)}}{\sigma^2_{\hat{X}(k)}+\sigma^2_{R(k)}}$，得到 $\rho(k)$ 与 $R_{ESI}(k)$ 的关系式

$$R_{ESI}(k)=\frac{\rho^2(k)}{\rho^2(k)+1}=1/(1+\frac{1}{\rho^2(k)}) \quad (7)$$

由式(7)可知，$\rho(k)$ 和 $R_{ESI}(k)$ 的变化趋势是一致的，即：当 $\rho(k)$ 趋近于 0 时，$R_{ESI}(k)$ 减小，趋近于 0；当 $\rho(k)$ 趋近于 1 时，$R_{ESI}(k)$ 增大，趋近于 1/2.

(2)当 $\mu_{X(k)}\neq 0$ 或 $\mu_{\hat{X}(k)}\neq 0$ 时，由于语音和噪声的期望和方差的随机性，无法确定 $\rho(k)$ 和 $R_{ESI}(k)$ 的函数关系，但是根据文献[7]的实验结果，$R_{ESI}(k)$ 与 $\rho(k)$ 相关，其相关系数为 0.81. 所以，语音畸变严重时，各频点的 $\rho(k)$ 趋近于 0，$R_{ESI}(k)$ 也趋近于 0；反之，各频点的 $\rho(k)$ 趋近于 1，$R_{ESI}(k)$ 趋近于正无穷.

通过以上证明可知，$R_{ESI}(k)$ 是一种与语音畸变密切相关的客观度量，本文利用 $R_{ESI}(k)$ 来研究语音清晰度是可行的.

2.2 两类失真与语音清晰度

要抑制语音畸变，就要考虑如何尽可能地保证各个频点的 $\rho(k)$ 都接近 1，即 $R_{ESI}(k)$ 在整个频谱上恒大于某一门限. 本小节首先讨论这一问题.

用第 k 个频点的噪声幅度谱的平方 $D^2(k)$ 除以式(6)的分子分母，得到下式

$$R_{ESI}(k)=\frac{R_{CLN}(k)}{(R_{CLN}(k)^{1/2}-R_{ENH}(k)^{1/2})^2} \quad (8)$$

式中：$R_{CLN}(k)=\frac{X^2(k)}{D^2(k)}$ 为纯净语音信噪比；$R_{ENH}(k)=\frac{\hat{X}^2(k)}{D^2(k)}$ 为增强语音信噪比.

在 $R_{CLN}(k)$ 固定的情况下，式(8)可以看作是以 $R_{ENH}(k)$ 为自变量，$R_{ESI}(k)$ 为因变量的函数，其定义域为 $\{R_{ENH}(k)\,|\,R_{ENH}(k)\in R_{ENH}(R)\neq R_{CLN}(k)\}$. 由式(8)绘出在不同 $R_{CLN}(k)$ 条件下，$R_{ESI}(k)$ 与 $R_{ENH}(k)$ 的变化曲线，如图 1 所示. 由图 1 可知：选择某个 $R_{CLN}(k)$ 后，函数曲线可分成 3 个区域：

(1)区域 I：$R_{ENH}(k)<R_{CLN}(k)$；

(2)区域 II：$R_{CLN}(k)<R_{ENH}(k)<4R_{CLN}(k)$；

(3)区域 III：$R_{ENH}(k)\geqslant 4R_{CLN}(k)$.

由 $R_{ENH}(k)$ 和 $R_{CLN}(k)$ 的定义可知，$R_{CLN}(k)<R_{ENH}(k)$ 等价于 $X(k)<\hat{X}(k)$，$R_{ENH}(k)\geqslant 4R_{CLN}(k)$ 等价于 $\hat{X}(k)\geqslant 2X(k)$，所以可将上述 3 个区域改写为：

(1)区域 I：当 $\hat{X}(k)<X(k)$ 时，$\hat{X}(k)$ 包含衰减失真且 $R_{ESI}(k)>0$ dB；

(2)区域 II：当 $X(k)<\hat{X}(k)<2X(k)$ 时，$\hat{X}(k)$ 包含放大失真且 $R_{ESI}(k)>0$ dB；

(3)区域 III：当 $\hat{X}(k)\geqslant 2X(k)$ 时，$\hat{X}(k)$ 包含放大失真且 $R_{ESI}(k)\leqslant 0$ dB.

区域 III 中，$R_{ESI}(k)<0$ dB，由 2.1 小节的讨论可知，R_{ESI} 的值很低时，语音畸变严重，所以要抑制这类放大失真，减小语音畸变. 在区域 I+II 中，$\hat{X}(k)$ 包含衰减失真和另外一部分放大失真，此时 $R_{ESI}(k)>0$ dB，所以可以保留这 2 种失真.

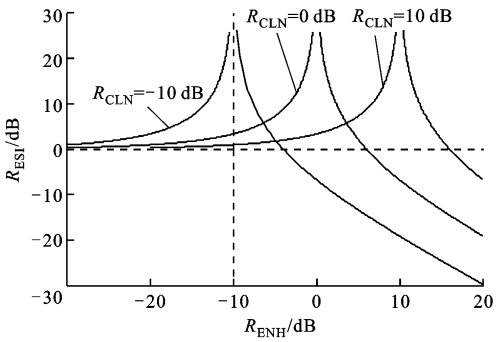


图 1 不同信噪比条件下 $R_{ESI}(k)$ 与 $R_{ENH}(k)$ 的关系

文献[8]也通过严格的实验证明：在信噪比为 0 dB条件下，使用 $\hat{X}(k)<X(k)$ 作为约束条件，即使是非常伤害语音清晰度的幅度谱减算法，增强后的语音的单词识别率也能从 24.8%提高到 90.1%.

从另外一个角度分析：去噪度和语音畸变度不可能同时优化. 语音畸变较小时，残留的噪声显著；如果语音畸变较大，则残留噪声较少. 所以综合考虑两者之间的平衡，最后选择了 $\hat{X}(k)<2X(k)$（即保证 $R_{ESI(k}(k)>0$ dB).

综上所述，如果加上合适的限制，使 $R_{ESI}(k)$ 大于 0 dB，就能在不降低去噪度的前提下抑制语音畸变. 因此，可以基于此结论引入一个判决式来提高语音清晰度

$$X_C(n,k)=\begin{cases}\hat{X}(n,k), & \hat{X}(n,k)<2X(n,k)\\ 0, & \text{其他}\end{cases} \quad (9)$$

由于式(9)在做判决时要使用纯净语音的幅度谱，而在实际的去噪过程中不可能得到纯净的语音谱，因此不等式 $\hat{X}(k)<2X(k)$ 两边同时平方，并带入式(3)、(4)（这里带入的是第 $n-1$ 帧第 k 个频点

的估计函数$G(n-1,k)$)得

$$R_{\mathrm{CLN}}(n,k)>\frac{G^2(n-1,k)}{4-G^2(n-1,k)} \tag{10}$$

文献[9]给出了利用噪声频谱方差 $\sigma^2_{D(n,k)}$ 得到带噪语音的后验信噪比

$$\gamma_{X(n,k)}=\frac{Y^2(n,k)}{\sigma^2_{D(n,k)}} \tag{11}$$

将式(11)代入式(10)得

$$\gamma_{(n,k)}>\frac{G^2(n-1,k)}{4-G^2(n-1,k)}+\frac{D^2(n,k)}{\delta^2_{D(n,k)}} \tag{12}$$

式中：噪声幅度谱 $D(n,k)$ 可利用相邻几帧的谱估计得到. 最后，将约束条件表示为短时谱估计函数的形式　　$G_{\mathrm{C}}=$

$$\begin{cases}\hat{X}(n,k), & \gamma_{X(n,k)}>\dfrac{G^2(n-1,k)}{4-G^2(n-1,k)}+\dfrac{D^2(n,k)}{\sigma^2_{D(n,k)}}\\ 0, & \text{其他}\end{cases} \tag{13}$$

2.3　失真控制下的语音谱估计函数

改进语音增强算法的目的有 2 个：一是要提高语音清晰度，抑制语音畸变；二是要更好地去除噪声，提高语音质量. 但是，使用传统语音增强算法去噪时会在估计语音中引入音乐噪声，这是因为噪声和语音是随机变化的，而语音增强算法使用各种语音或者噪声的统计参量作为输入变量，算法跟踪速度固定或者有限，不能随着语音或噪声突变，因此去噪后的语音包含一定节奏感的残余噪声，即音乐噪声. 包含音乐噪声的估计语音会使收听者感到不适，容易产生听觉疲劳.

本文引入人耳听觉掩蔽效应，并将带噪语音分成 3 种情况分别进行处理，使算法模型在抑制音乐噪声时更加符合实际. 为了使谱估计更准确，根据无语音概率参数(SAP)把语音的第 n 帧第 k 频点分量的状态分为无语音 $H_0(n,k)$ 和有语音 $H_1(n,k)$ 两种状态，同时引入人耳听觉掩蔽效应，又将有语音的状态 $H_1(n,k)$ 细分为噪声未被语音掩蔽的状态 $H_{1,0}(n,k)$ 和噪声被语音掩蔽的状态 $H_{1,1}(n,k)$[10]，各个状态分别表示为

$$H_0(n,k):Y(n,k)=D(n,k) \tag{14}$$

$$H_1(n,k):\begin{cases}H_{1,0}(n,k):Y(n,k)=X(n,k)+D(n,k),\ D(n,k)>T(n,k)\\ H_{1,1}(n,k):Y(n,k)=X(n,k)+D(n,k),\ D(n,k)<T(n,k)\end{cases} \tag{15}$$

式中：$T(n,k)$ 是纯净语音第 n 帧第 k 频点分量的听觉掩蔽阈值. 本文采用 Johnston 提出的算法计算掩蔽阈值 $T(n,k)$[11].

现有的谱减语音增强算法中，最小对数谱均方误差谱减算法(MMSE-LSA)，更符合人耳的主观听觉特性，具有良好的噪声抑制性能[12]，其估计函数如下

$$G_{\mathrm{MMSE\text{-}LS}}=\frac{\xi_{X(n,k)}}{1+\xi_{X(n,k)}}\exp\left\{\frac{1}{2}\int_{v(n,k)}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right\} \tag{16}$$

式中：$v(n,k)=\dfrac{\xi_{X(n,k)}}{1+\xi_{X(n,k)}}$；$\xi_{X(n,k)}$ 是带噪语音的先验信噪比，由下式[4]计算得出

$$\left.\begin{aligned}\xi'_{X(n,k)}&=\beta\frac{\hat{X}^2(n-1,k)}{\delta^2_{N-1}(k)}+(1-\beta)\max(\gamma_{X(n,k)}-1,0)\\ \xi_{X(n,k)}&=\max(\xi'_{X(n,k)},\xi_{\min})\end{aligned}\right\} \tag{17}$$

式中：β 和 $\xi_{\min}$ 通常分别为 0.98 和 −15 dB.

本文算法以最小对数谱均方误差谱减算法为基础，对其进行修正与改进，使谱估计更加准确. 根据 SPA 参数和听力阈值将带噪语音分为 3 个状态，每个状态使用不同的估计函数，同时每种状态都使用 2.2 节引入的对 2 种失真的增益控制，以提高语音清晰度，减小语音畸变.

(1)有语音且噪声没有被掩蔽的状态 $H_{1,0}(n,k)$. 人耳能够感受到噪声，采用最小对数谱均方误差谱估计算法去噪，同时加上对两类失真的控制，抑制部分放大失真，保留衰减失真，其条件估计函数表示为

$$G=G_{\mathrm{C}}G_{\mathrm{MMSE\text{-}LS}}=\frac{\exp\{E[\ln\hat{Y}_{1,1}(n,k)]\}}{Y(n,k)}G_{\mathrm{C}}=G_{\mathrm{C}}\frac{\xi_{X(n,k)}}{1+\xi_{X(n,k)}}\exp\left\{\frac{1}{2}\int_{v(n,k)}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right\} \tag{18}$$

式中：$\hat{Y}_{1,1}(n,k)$ 表示在$(Y(n,k),H_1(n,k),H_{1,0}(n,k))$状态下对语音的幅度谱的估计.

(2)有语音且噪声被掩蔽的状态 $H_{1,1}(n,k)$. 此状态下有噪声，但人耳因为听力掩蔽效应感觉不到噪声的存在，因此无需对带噪语音进行减谱处理，只需要抑制部分放大失真，提高语音清晰度.

$$G=\frac{\exp\{E[\ln\hat{Y}_{1,0}(n,k)]\}}{Y(n,k)}=G_{\mathrm{C}} \tag{19}$$

式中：$\hat{Y}_{1,0}(n,k)$ 表示在$(Y(n,k),H_1(n,k),H_{1,1}(n,k))$状态下对语音的幅度谱的估计.

(3)无语音状态 $H_0(n,k)$. 此状态下只有噪声. 由于在语音谱中保留少量的宽带噪声，可以在听觉

上掩蔽音乐噪声，使去噪后的语音更自然，因此后验谱估计函数取一个固定的经验值 $G_{\min}$，再引入式(13)的约束条件，抑制语音畸变

$$G=\frac{\exp\{E[\ln\hat{Y}_0(n,k)]\}}{Y(n,k)}=G_{\min}G_{\mathrm{C}} \tag{20}$$

式中：$\hat{Y}_0(n,k)$表示在$(Y(n,k),H_0(n,k))$状态下对语音的幅度谱的估计.

通过以上 3 个状态的分析，本文改进的估计函数定义为

$$G_{\text{LS-ADS}}=\frac{\exp\{E[\ln\hat{Y}(n,k)]\}}{Y(n,k)} \tag{21}$$

式中：$E[\ln\hat{Y}(n,k)]$可表示为

$$E[\ln\hat{Y}_{1,1}(n,k)]=$$
$$E[\ln\hat{Y}_{1,1}(n,k)]p[H_1(n,k),H_{1,1}(n,k)\mid Y(n,k)]+$$
$$E[\ln\hat{Y}_{1,0}(n,k)]p[H_1(n,k),H_{1,0}(n,k)\mid Y(n,k)]+$$
$$E[\ln\hat{Y}_0(n,k)]p[H_0(n,k)\mid Y(n,k)] \tag{22}$$

其中 $p[H_1(n,k),H_{1,1}(n,k)\mid Y(n,k)]=P_1(n,k)$ 表示后验语音出现且被掩蔽的概率，$p[H_1(n,k),H_{1,0}(n,k)\mid Y(n,k)]=P_0(n,k)$表示后验语音出现且未被掩蔽的概率，$p[H_0(n,k)\mid Y(n,k)]=1-p[H_1(n,k)\mid Y(n,k)]=1-P_1(n,k)-P_0(n,k)$表示后验无语音概率. $P_1(n,k)$与 $P_0(n,k)$可由文献[4]计算得出.

联合式(17)～式(21)，得到最后的短时谱估计函数

$$G_{\text{LS-ADS}}=$$
$$(G_{\mathrm{C}}G_{\text{MMSE-LS}})^{P_0(n,k)}G_{\mathrm{C}}{}^{P_1(n,k)}(G_{\min}G_{\mathrm{C}})^{1-P_1(n,k)-P_0(n,k)}=$$
$$G_{\mathrm{C}}G_{\text{MMSE-LS}}{}^{P_0(n,k)}G_{\min}{}^{1-P_1(n,k)-P_0(n,k)} \tag{23}$$

3 实验结果与分析

语音增强算法有 2 个指标：一是主观度量，提高语音清晰度，使听者乐于接受，不易产生听觉疲劳；二是客观度量，提高语音质量，以信噪比和某些语音特征参数的提高程度为衡量准则. 对低信噪比环境下的带噪语音，信噪比的提高已经不能准确地反映算法性能，重要的是要使增强后的语音听起来清晰舒适，可懂度高，语音的主观度量显得更加重要.

为了测试算法性能，从 Noisex 92 噪声库中选取白噪声(White)、汽车噪声(Volvo)、飞机驾驶舱噪声(F16)和工厂噪声(Factory)，与取自标准语音库的纯净语音混合生成不同信噪比(10、5、0 和－5 dB)的带噪语音. 语音信号经过 8 kHz 采样，16 bit 量化，窗长 256 点的汉明窗分帧(每帧 20 ms，帧间重叠 50%).

实验结果从主观和客观 2 个方面进行评价. 客观评价采用分段信噪比(R_{seg})，主观评价采用语谱图. 为方便评测比较，客观测试指标给出的是语音增强前后的提高量. R_{seg}的定义如下

$$R_{\text{seg}}=\frac{1}{L}\sum_{i=0}^{L-1}10\log\left\{\frac{\sum_{n=0}^{M}x(iM+n)^2}{\sum_{n=0}^{M}[x(iM+n)-\hat{x}(iM+n)]^2}\right\} \tag{24}$$

式中：L 表示带噪语音信号帧的个数；M 是每帧未重叠的点数.

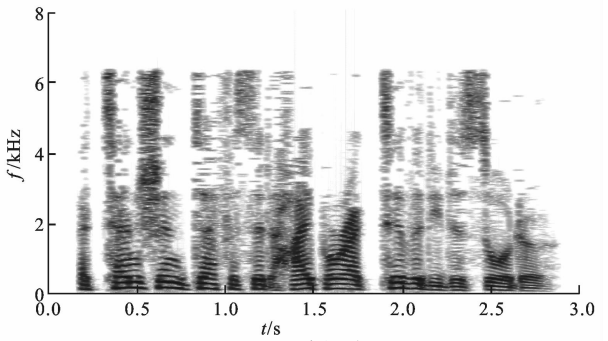
表 1 给出了 3 种算法在不同噪声和不同信噪比条件下的客观测试结果. 由表 1 可见，对于不同的噪声类型以及它们对应的不同的信噪比，使用本文算法增强后，语音的信噪比都有不同程度的提高，这说明算法确实能有效地抑制各种类型的背景噪声. 将表 1 中列举的 3 种算法进行横向对比，在带噪语音信噪比较低时，与其他 2 种算法相比，本文算法大幅度提高了估计语音的信噪比，而在信噪比较高时，MMSE-LSA 算法所得到的估计语音的信噪比比本文算法高，这是由于本文算法保留了低于语音掩蔽阈值的噪声，考虑了残留噪声和语音畸变之间的折中，对语音信号的损伤最小. 下面通过比较语谱图进一步证明了本文算法在主观测试上的优势.

表 1 各种噪声类型下增强语音分段信噪比变化量

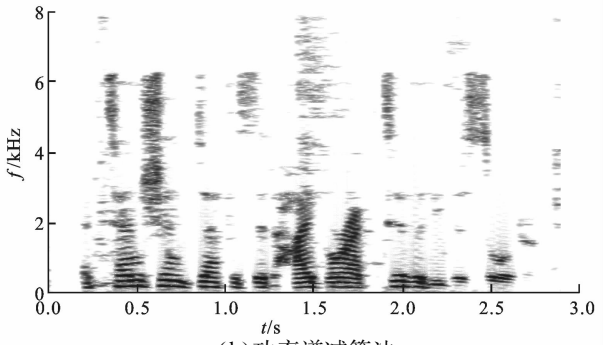
噪声类型	信噪比/dB	ΔR_{seg}/dB		
		MMSE-LSA 算法	PS 算法	本文算法
White	－5	11.92	3.87	14.10
	0	9.69	3.01	10.53
	5	7.28	2.61	6.80
	10	5.31	2.35	3.64
Volvo	－5	10.98	3.64	13.27
	0	9.37	3.53	10.98
	5	7.39	2.57	6.55
	10	5.68	2.49	3.97
E16	－5	9.68	2.87	12.14
	0	8.24	2.71	9.29
	5	6.33	2.20	6.15
	10	3.89	1.68	2.04
Factory	－5	12.19	5.46	13.89
	0	10.56	5.04	10.63
	5	7.47	4.72	6.89
	10	4.57	3.41	3.32

注：ΔR_{seg}为输出分段信噪比与输入分段信噪比之差.

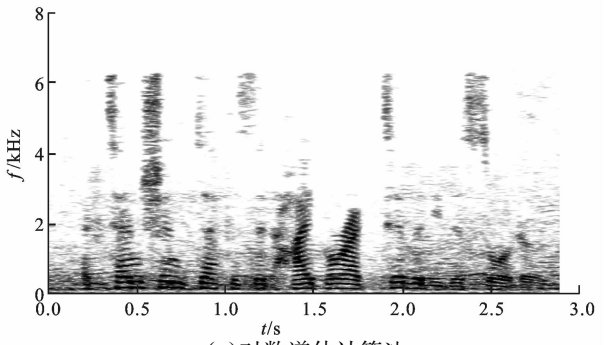
在纯净语音上叠加工厂噪声(信噪比为 0 dB)，分别使用 3 种算法进行语音增强. 图 2 是带噪语音



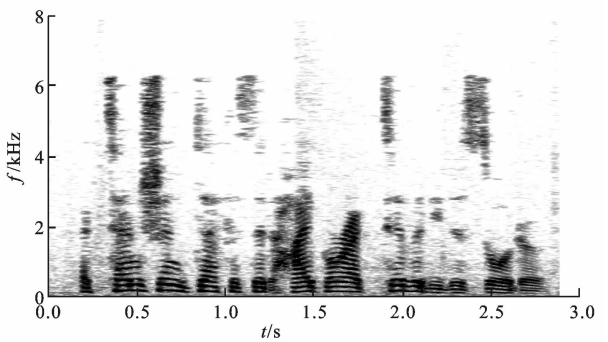
(a)纯净语音



(b)功率谱减算法



(c)对数谱估计算法



(d)本文算法

图 2　带噪语音经过几种算法增强后的语谱图

经过各算法增强后的语谱图. 由图 2 可见，功率谱减增强后的语音去掉了部分噪声，但是也损伤了很多语音元素，尤其是在音节的起始段. 起始段中大部分是清音，损伤后严重影响语音的可懂度. 对数谱减的语音增强考虑了人耳的听觉特性，语谱图中高频部分的语音保留较多，可懂度提高，但同时有许多残留孤立点，这些点就是谱减法中常出现的音乐噪声，且主观试听发现存在语音畸变. 这是因为：①对数谱减算法没有考虑估计值和真实值误差的正负性对语音清晰度的影响；②只使用了简单的语音和噪声假设模型，不能很好地跟踪噪声和语音的变化. 本文提出的算法引入了对两种失真的控制，并改进了先验无语音概率的估计，增强后的语谱图中高频部分的语音保留较多，孤立点大大减少，残留的孤立点几乎全部分布在高能量的语音元素周围，主观听力测试中不是特别明显，且主观试听表明增强后语音失真最小，清晰度最高.

4　结束语

本文引入语音畸变的客观度量参数，提出了对 2 种失真的控制方法，抑制了语音畸变. 同时，对估计函数进行软判决修正，利用 SPA 参数和听觉掩蔽阈值来实时更新各帧的估计函数，使谱估计更符合实际语音和噪声模型，改善了传统的 MMSE 估计中残留较多音乐噪声的问题，在噪声抑制和语音畸变之间得到了较好的折中，进一步改善了语音增强的效果. 客观测试和主观试听结果都表明，本文算法在非平稳噪声干扰下的性能要优于传统的谱减算法，在低信噪比输入条件下效果尤为显著.

参考文献：

[1] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1979, 27(2): 113-120.

[2] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean square error short-time spectral amplitude estimator [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984, 2(6): 1109-1121.

[3] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean square error log spectral amplitude estimator [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, 3(2): 443-445.

[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 Xiaoqun, HUANG Xiaoshan. Improved speech enhancement based on spectral subtraction and auditory masking effect [J]. Journal of Communication, 2008, 29(9): 73-80.

[5] 赵力. 语音信号处理[M]第2版. 北京：机械工业出版社，2009：31-32.

[6] MA Jianfen，LOIZOU P C. SNR loss：a new objective measure for predicting the intelligibility of noise-suppressed speech [EB/OL]. (2009-12-02) [2010-10-24]. http：// dx. doi. org/10.1016/j. specom. 2010.10.005.

[7] MA Jianfen，HU Yi，LOIZOU P C. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions [J]. Acoust Soc Amer，2009，125(5)：3387-3405.

[8] LOIZOU P C，KIM P. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions [J]. IEEE Trans Audio Speech Lang Process，2011，19 (1)：47-56.

[9] MARTIN R. Noise power spectral density estimation based on optimal smoothing and minimum statistics [J]. IEEE Trans on Acoustics，Speech，and Signal Processing，2001，9(5)：504-512.

[10] 朴凡亮，王为民，戴启军，等. 基于噪声被掩蔽概率的优化语音增强方法[J]. 电子与信息学报，2005，27(5)：753-756.
PU Fanliang，WANG Weiming，DAI Qijun，et al. Optimizing speech enhancement based on noise marked probability [J]. Journal of Electronics and Information Technology，2005，27(5)：753-756.

[11] JOHSTOM J D. Transform coding of audio signals using perceptual noise criteria [J]. IEEE Journal on Selected Areas in Communications，1988，6(2)：314-323.

[12] PORTER J E，BOLL F S. Optimal estimators for spectral restoration of speech [C]// Proceedings of International Conference on Acoustics，Speech，and Signal Processing. Piscataway，NJ，USA：IEEE，1984：53-56.

【本刊相关文献链接】

n 维空间随机矩阵变换的音频置乱算法. 西安交通大学学报，2010，44(4)：13-17.

分组网络环境下的实时语音质量客观评价. 西安交通大学学报，2006，40(8)：936-939.

基于心理声学模型的高性能语音质量评价算法. 西安交通大学学报，2006，40(4)：437-440.

基于模糊多类支持向量机的语音质量客观评价. 西安交通大学学报，2006，40(2)：199-202.

语音信号中相位信息的听觉感知研究. 西安交通大学学报，2003，37(12)：1288-1291.

一种基于神经网络的小波域音频水印算法. 西安交通大学学报，2003，37(4)：355-358.

基于局部余弦变换的2.4 kb/s低比特率语音编码. 西安交通大学学报，2003，37(4)：388-391.

声带振动功能模式识别的研究. 西安交通大学学报，2002，36(12)：1258-1261.

基于统计模型实现语音信号有声/无声检测的研究. 西安交通大学学报，2002，36(8)：839-842.

Internet环境下服务可控的网络语音体系结构研究. 西安交通大学学报，2002，36(4)：402-405.

含噪语音信号中噪声参数的一种估计方法. 西安交通大学学报，2001，35(10)：1096-1097.

抽取音频数据特征的快速离散余弦变换方法. 西安交通大学学报，2001，35(8)：854-857.

（编辑 刘杨）

【文摘预登】

H.264中全零块检测的线性分类器算法

王萍，邓妍，黄华
（西安交通大学电子与信息工程学院，710049，西安）

针对H.264视频编码中传统全零块检测方法检测率较低的问题，提出了一种全零块检测的线性分类器算法. 根据编码特性选择5个变量作为全零块检测的特征，基于参考块的全零块情况，设计了2个不同的线性分类器来区分全零块和非全零块，然后选取代表不同运动程度的视频样本，利用Fisher准则训练得到2个分类器的权系数和不同量化参数下全零块检测的阈值；最后用最小二乘原理将各个量化参数下的阈值拟合成量化参数的二次多项式，从而得到最终的全零块检测的线性分类器. 实验结果表明，新算法在基本不降低视频质量的同时，能获得比其他现有算法更好的检测率，并且有效地缩短了编码时间.