噪声谱估计算法对语音可懂度的影响

张建伟，陶 亮，周 健，王华彬
(安徽大学计算智能与信号处理教育部重点实验室，安徽合肥 230031)

摘要：噪声谱估计是单通道语音增强算法的关键步骤，当前大部分语音增强算法旨在提高语音质量，提高语音可懂度的算法却很少。在传统的单通道语音增强算法中，语音质量的提高往往是以牺牲语音的可懂度为代价的。对目前主流的几种噪声谱估计算法对语音可懂度影响进行分析。在不同噪声背景、不同信噪比情况下进行噪声谱估计，并采用谱减法对含噪语音信号作去噪处理，对比分析不同噪声、不同信噪比下增强前后语音的短时客观可懂度(Short-Time Objective Intelligibility, STOI)值，最后根据信噪比，对比分析了不同噪声环境下，语音增强前后语音能量高于噪声能量的时频块所占比例。实验表明，相比其他噪声估计算法，最小统计(Minima Statistics, MS)算法由于保留了更多的以语音能量为主的时频块，使得去噪后的语音有较高的可懂度。

关键词：噪声谱估计；谱减法；时频块；最小统计；短时客观可懂度；语音可懂度
中图分类号：TP391 **文献标识码**：A **文章编号**：1000-3630(2015)-05-0424-07
DOI 编码：10.16300/j.cnki.1000-3630.2015.05.009

Effects of noise spectrum estimation algorithms on speech intelligibility

ZHANG Jian-wei, TAO Liang, ZHOU Jian, WANG Hua-bin
(*Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education*, *Anhui University*, *Hefei* 230031, *Anhui*, *China*)

Abstract: Noise spectrum estimation is a key step in single channel speech enhancement algorithms. Most of current speech enhancement algorithms are designed to improve speech quality, however, algorithms for increasing speech intelligibility are few. The traditional speech enhancement algorithms improve speech quality, while sacrificing speech intelligibility. In this paper, classical noise spectrum estimation algorithms are evaluated for their effects on speech intelligibility. Noise spectrum is estimated in different noise environments with SNRs between −9 dB and 3 dB. The spectral subtraction is thereafter used for speech denoising. The STOI(Short-Time Objective Intelligibility) value of the enhanced speech is computed. At last, according to the signal-to-noise ratio, the proportions of speech dominated time-frequency blocks under different noise environments are analyzed. Experimental results show that, compared with other noise estimation algorithms, the minimum statistics (MS) obtains high speech intelligibility because it retains more speech dominated time-frequency blocks after speech denoising.

Key words: noise spectrum estimation; spectrum subtraction; time-frequency blocks; Minima Statistics(MS); Short-Time Objective Intelligibility(STOI); speech intelligibility

0 引 言

语音增强算法在提高语音质量方面已经取得了很大的进展[1-3]，相反，提高语音可懂度的算法却很少。Lim 首次发现，在−5~5 dB 的白噪声背景下，谱减法并未提高语音的可懂度[4]。Hu 和 Loizou 也对语音可懂度作了研究，他们采用了 8 种不同的算法，对语音增强前和增强后的可懂度进行比较，结果发现，所有算法增强后的可懂度均小于增强前的可懂度[5]。研究者发现，在传统的语音增强算法中，语音质量的提高往往是以牺牲语音的可懂度为代价的[6]。

研究者们提出了很多相关的噪声谱估计算法，而且取得了一定的效果[7-11]。Hirsch[12]提出了一种不需要进行语音端点检测的噪声谱估计方法，需要比较当前窗的功率谱和前一窗的估计噪声谱，使用一阶递归平均来更新噪声谱估计，该方法可以快速地适应变化缓慢的噪声。Martin[13]提出了一种基于最小统计(Minima Statistics, MS)的噪声谱估计方法，即在一个有限窗口内跟踪平滑含噪语音谱的最小值，然后对其按帧平滑，并乘以一个偏置补偿因子，即可获得噪声谱估计。Cohen 和 Berdugo[14]提出了

收稿日期：2014-12-15；**修回日期**：2015-03-29
基金项目：国家自然科学基金(61301219、61003131)、安徽省自然科学基金(1408085MF113)资助项目。
作者简介：张建伟(1989−)，女，山东莘县人，硕士研究生，研究方向为语音增强。
通讯作者：张建伟，E-mail: zhangjianwei.i.123@163.com

一种最小受控递归平均算法(Minima Controlled Recursive Averaging, MCRA)，该方法根据含噪语音的局部能量值与其待定时间窗内的最小值的比值确定子带中是否存在语音，如果给定帧的某个子带中存在语音，那么该子带内的噪声谱等于上一帧的噪声谱，如果不存在，则根据含噪语音的功率谱更新噪声谱。Cohen 在 2003 年提出了改进的最小控制递归平均方法(Improved Minima Controlled Recursive Averaging, IMCRA)，主要从三个方面进行了改进，即语音活跃期的最小值跟踪、语音存在概率估计、提出偏置补偿因子[15]。Sorensen 等人在 2005 年提出了一种基于连接语音时频域(Connected Time-Frequency Speech Presence Regions, Conn_freq)[16]的噪声谱估计算法，该方法可连接时频域的语音缺失段，将缩小的背景噪声留在增强后的语音中，利用人的听觉系统中的掩蔽机制，减少对语音段中噪声的感知，消除语音缺失段的噪声。

有研究者在噪声谱估计算法的基础上，提出了很多改进算法，在语音质量和可懂度方面有了一定程度的改善[17-20]。虽然这些噪声谱估计方法得到广泛应用，但是其对于增强后语音可懂度的影响则至今未见相关报道。为此，本文讨论上述 5 种不同的噪声谱估计算法对语音可懂度的影响。为尽可能排除增强过程中其他因素对可懂度的影响，增强算法采用经典的谱减法。论文首先回顾 5 种噪声谱估计方法，并将其应用于正常音的噪声谱估计。为了评价这 5 种算法对语音可懂度的影响，计算经增强后的语音可懂度，对增强前后的语音时频谱中的语音能量为主的时频块的保留情况进行分析，以探讨不同噪声谱估计方法对可懂度影响的原因。

1　噪声谱估计及算法

1.1　信号模型

设 y 表示时域含噪信号，x 表示干净语音信号，d 表示非相关加性噪声。对含噪信号作短时傅里叶变换(Short-time Fourier Transform, STFT)，$\boldsymbol{Y}(k,l)$、$\boldsymbol{X}(k,l)$、$\boldsymbol{D}(k,l)$分别是 $\boldsymbol{y}$、$\boldsymbol{x}$、$\boldsymbol{d}$ 的变换系数，我们得到时频域信号

$$\boldsymbol{Y}(k,l)=\boldsymbol{X}(k,l)+\boldsymbol{D}(k,l) \tag{1}$$

式(1)中：k 表示频带号；l 表示时帧号。

1.2　噪声谱估计算法

单通道语音增强算法都需要从含噪语音中估计噪声谱和先验信噪比，后者也建立在噪声谱估计基础上。

1.2.1 Hirsch 算法

Hirsch 提出计算所有频域子带 i 的含噪语音幅度谱 X_i 的权重和，然后按照式(2)对噪声估计进行一阶递归：

$$\hat{\boldsymbol{N}}_i(k)=(1-\alpha)*\boldsymbol{X}_i(k)+\alpha*\hat{\boldsymbol{N}}_i(k-1) \tag{2}$$

其中：α=0.85 表示平滑常数，$\boldsymbol{X}_i(k)$ 表示第 i 个子带的第 k 个频带的含噪语音幅度谱，$\hat{\boldsymbol{N}}_i(k)$ 表示第 i 个子带的第 k 个频带的噪声估计，$\boldsymbol{X}_i$ 值在纯噪声段满足瑞利分布。最后，噪声估计 $\hat{\boldsymbol{N}}_i$ 乘以一个过估计补偿因子 β，取值范围是 1.5 至 2.5。当 $(\boldsymbol{X}_i-\beta\hat{\boldsymbol{N}}_i)$ 为正值时，表示语音出现，停止递归；当 $(\boldsymbol{X}_i-\beta\hat{\boldsymbol{N}}_i)$ 为负值时，将其置零。

该算法不需要进行语音端点检测，而且可以快速适应变化缓慢的噪声，语音存在段和语音缺失段都采用公式(2)更新噪声谱，可以结合谱减法对语音作增强处理。

1.2.2 MS 算法

最小值统计的方法依赖于两点，即 (1) 语音信号和噪声从统计意义上讲是独立的；(2) 含噪语音的功率会衰减至噪声的功率水平。由于最小值总是小于平均值，因此最小值跟踪方法需要偏差补偿。为了能更快地跟踪并更新局部最小值和频谱最小值，作者把滑动窗口分为多个子窗口，在每个子窗口内更新估计噪声谱，提高了精确度[21]。

MS 算法一阶平滑估计噪声谱的规则可用式(3)表示：

$$\hat{\boldsymbol{N}}(\lambda,k)=\alpha_{\text{opt}}(\lambda,k)\hat{\boldsymbol{N}}(\lambda-1,k)+(1-\alpha_{\text{opt}}(\lambda,k))|\boldsymbol{Y}(\lambda,k)|^2 \tag{3}$$

其中：$\hat{\boldsymbol{N}}(\lambda,k)$ 表示第 λ 个搜索窗的第 k 个频带的估计噪声功率谱，$Y(\lambda,k)$ 表示第 λ 个搜索窗的第 k 个频带的含噪语音谱，即含噪语音的频域表达式，$\alpha_{\text{opt}}(\lambda,k)$ 是时频独立的平滑参数，基于最小误差准则得到。搜索窗长 D 取 150，子窗数 U 为 10，子窗长 V 为 15，实验采用来自文献[21]的算法，其他有关参数，默认为文献[21]给定的数据。

本算法无论是在语音存在段还是语音缺失段，噪声功率谱估计均跟踪平滑含噪语音谱的最小值，不采用阈值区分语音活动和语音端点，可以结合任意需要噪声谱估计的语音增强算法。

1.2.3 MCRA 算法

MCRA 算法使用一个平滑参数对功率谱的过去值取平均，其中平滑参数是通过子带中语音存在的概率来调整的。首先对输入的每一帧信号进行频域平滑：

$$\boldsymbol{S}_f(k,l)=\sum_{i=-w}^{w} b(i)\left|\boldsymbol{Y}(k-i,l)\right|^2 \tag{4}$$

其中：$b(i)$ 表示加权系数，$\boldsymbol{Y}(k-i,l)$ 表示含噪语音在时频域作短时傅里叶变换的幅度值，窗函数的长度是 $2w+1$。

其次，采用一阶递归进行时域平滑：

$$\boldsymbol{S}(k,l)=\alpha_s\boldsymbol{S}(k,l-1)+(1-\alpha_s)\boldsymbol{S}_f(k,l) \tag{5}$$

其中：$\alpha_s=0.8$ 表示平滑参数，$\boldsymbol{S}(k,l-1)$ 表示前一帧含噪语音的功率谱。

同时跟踪含噪语音功率谱的局部最小值，估计语音存在概率，最后根据式(6)、(7)中规则更新噪声谱：

$$H_0'(k,l):\hat{\boldsymbol{N}}(k,l+1)=\alpha_d\hat{\boldsymbol{N}}(k,l)+(1-\alpha_d)\left|\boldsymbol{Y}(k,l)\right|^2 \tag{6}$$

$$H_1'(k,l):\hat{\boldsymbol{N}}(k,l+1)=\hat{\boldsymbol{N}}(k,l) \tag{7}$$

其中：α_d 表示平滑参数；基于语音存在概率；H_0' 表示假设语音缺失段；H_1' 表示假设语音存在段；$\hat{\boldsymbol{N}}(k,l)$ 表示第 l 个搜索窗的的第 k 个频带。

1.2.4 IMCRA 算法

该算法是对 MCRA 算法的改进，噪声谱的更新规则不变。该算法包含两次迭代：平滑和最小值跟踪。第一次迭代是在每个频域子带内进行粗略的语音活动检测，第二次迭代是对语音缺失段的功率谱进行平滑，相对强语音信号部分并不进行平滑，使得语音活跃段的最小值跟踪具有鲁棒性。

搜索窗长 D 取 120，子窗数 U 为 8，子窗长 V 为 15，其他有关参数，默认为文献[15]给定的数据。

与 MS 算法不同的是，该算法考虑到连续窗口的相邻频域子带之间语音存在的强相关性，分别在时域和频域对含噪语音功率谱进行平滑处理。

1.2.5 连接语音时频域(Conn_freq)算法

Conn_freq 算法基于短时平滑功率谱和最小值跟踪，定义了两个语音存在检测规则，表示为

$$\boldsymbol{D}'(\lambda,k)=\begin{cases}1 & \boldsymbol{p}(\lambda,\mathrm{k})>\gamma'\boldsymbol{p}_{\min}(\lambda,\mathrm{k})\\ 0 & \boldsymbol{p}(\lambda,\mathrm{k})<\gamma'\boldsymbol{p}_{\min}(\lambda,\mathrm{k})\end{cases} \tag{8}$$

$$\boldsymbol{D}''(\lambda,k)=\begin{cases}1 & \boldsymbol{p}(\lambda,\mathrm{k})>\boldsymbol{p}_{\min}(\lambda,\mathrm{k})+\gamma''\frac{1}{K}\sum_{k=0}^{K-1}\boldsymbol{p}_{\min}(\lambda,k)\\ 0 & \boldsymbol{p}(\lambda,\mathrm{k})<\boldsymbol{p}_{\min}(\lambda,\mathrm{k})+\gamma''\frac{1}{K}\sum_{k=0}^{K-1}\boldsymbol{p}_{\min}(\lambda,k)\end{cases} \tag{9}$$

最终的语音存在检测估计为

$\boldsymbol{D}(\lambda,k)=\boldsymbol{D}'(\lambda,k)\cdot\boldsymbol{D}''(\lambda,k)$。

噪声功率谱估计为

$$\boldsymbol{P}_{\hat{N}}(\lambda,k)=\begin{cases}\boldsymbol{R}_{\min}(\lambda)\boldsymbol{p}_{\min}(\lambda,k), & if\ \boldsymbol{D}(\lambda,\mathrm{k})=1\\ \boldsymbol{P}_Y(\lambda,k), & if\ \boldsymbol{D}(\lambda,\mathrm{k})=0\end{cases} \tag{10}$$

其中：λ 表示帧号；k 表示频带；K 表示频谱的长度；$\boldsymbol{p}_{\min}(\lambda,k)$ 表示平滑功率谱最小值；γ' 和 γ'' 都是常数；$\boldsymbol{P}_Y(\lambda,k)$ 表示含噪语音功率谱；$\boldsymbol{R}_{\min}(\lambda)$ 表示补偿因子，语音缺失段进行更新，语音存在段固定不变，至于补偿因子如何更新，这里不再陈述。搜索窗长 D 取 7，子窗数 U 为 5，子窗长 V 为 8，其他有关参数，默认为文献[16]给定的数据。

该方法在连接时频域的语音缺失段，将缩小的背景噪声留在增强后的语音中，利用人的听觉系统中的掩蔽机制，减少对语音段中噪声的感知，消除语音缺失段的噪声。

1.2.6 不同算法噪声谱对比

图 1(a)和图 1(b)分别显示了 MS、MCRA、IMCRA、Hirsch 四种算法在白噪声背景下，在信噪比分别为–9 dB 和 5 dB 情况下的噪声谱估计，选取第 20 帧作为观测。图 2 显示了 Conn_freq 算法在白噪声背景下，在信噪比为–9 dB 和 5 dB 情况下的噪声谱估计。从图 2 中可以看出，Conn_freq 算法估计的噪声谱更接近真实噪声谱变化。为了更好地观察对比这 5 种算法的真实噪声谱和估计噪声谱，我们将 Conn_freq 算法的噪声谱估计图单独列出。

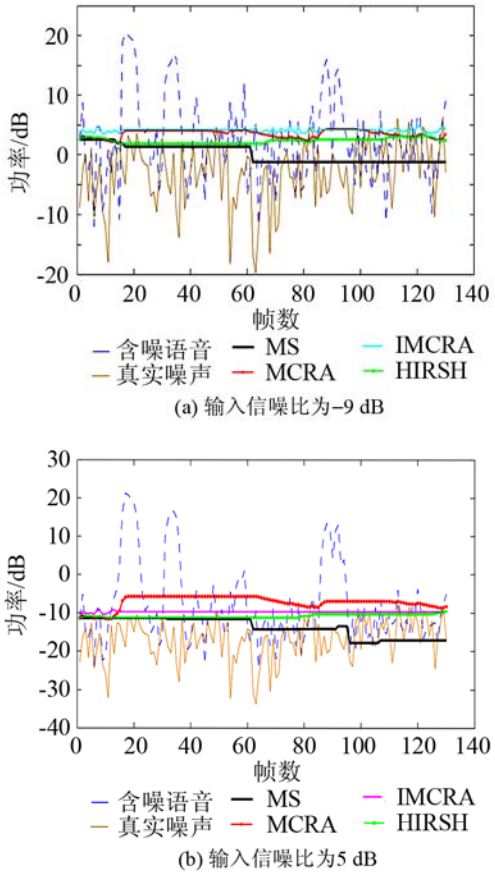


图 1 不同算法在白噪声背景下的谱估计

Fig.1 Spectrum estimations of different algorithms in the white noise environment of different SNRs

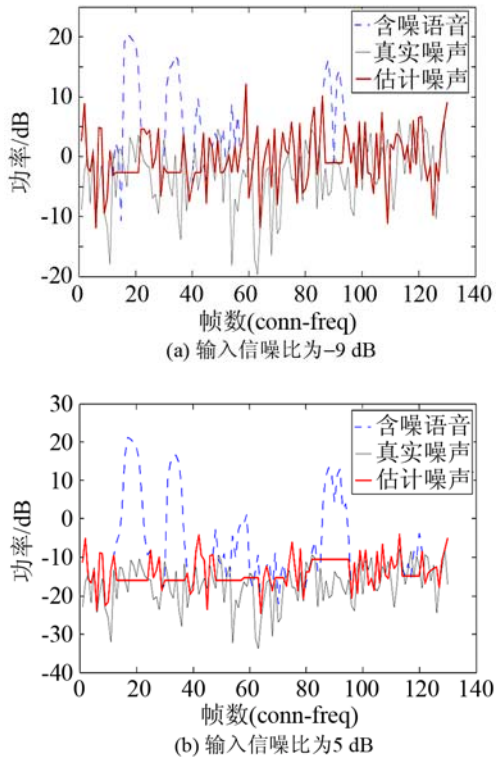


图 2　Conn_freq 算法在白噪声背景下的谱估计

Fig.2　Spectrum estimations of Conn_freq algorithm in the white noise environment of different SNRs

从图 1(a)中可以看出，在低信噪比−9 dB 的情况下，MS 算法的噪声谱估计最低，Hirsch 算法次之，IMCRA 算法和 MCRA 算法的噪声谱估计相似，只是在某些频点处，IMCRA 算法的噪声谱估计要高于 MCRA 算法。MCRA 算法和 IMCRA 算法的噪声谱估计高于 Hirsch 算法，这是因为前两种算法在语音存在段不进行噪声谱更新，而是保持前一帧的噪声谱不变，Hirsch 算法仍然采用一阶递归更新噪声谱估计。从图 1(b)中可以看出，在信噪比为 5 dB 的情况下，MS 算法的噪声谱估计还是最低，Hirsch 算法次之，MCRA 算法的噪声谱估计最高，而且超越了真实噪声谱。从图 2 中可以看出，Conn_freq 算法在信噪比分别为−9 dB 和 5 dB 时的噪声谱估计变化接近真实噪声谱，但是稍高于真实噪声谱，并未超越含噪语音谱。

2　实验仿真

实验采用来自中文语言资源联盟[22]语音数据库的干净语音共 50 句，是汉语连续音节构成的语句，每个语句有 6 个左右音节，其中男女语音各半。噪音数据采用 Noisex92 数据库[23]的三类噪声信号：White 高斯白噪声、F16 飞机驾驶舱噪声和 Babble 人群嘈杂噪声等。干净语音数据和噪声数据均为 16 kHz 采样率，混合产生信噪比在−9~3 dB 范围内的带噪语音。语音处理中，语音分帧帧长取 320 样点，帧间重叠 50%，数据加窗采用汉明窗，FFT 分析点数取 640 点。实验方法是将估计后的噪声谱用于谱减法[24]对语音作增强处理，然后从不同的角度评价增强后语音的可懂度。

谱减法是在频域将带噪语音的功率谱减去噪声的功率谱，从而得到纯净语音功率谱估计，开方后就得到语音幅度谱估计，用带噪语音的相位来近似纯净语音的相位，再采用逆傅里叶变换恢复时域信号[25]。谱减法的原理图如图 3 所示。

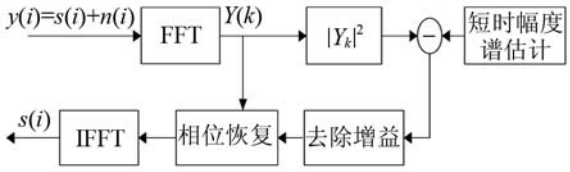


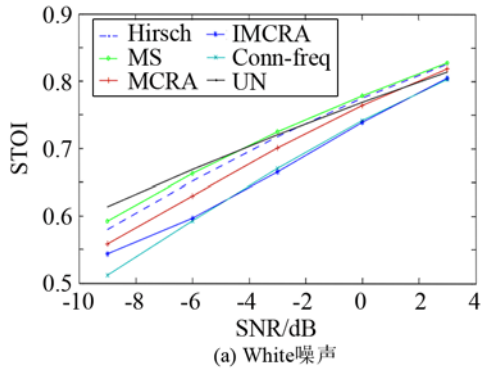
图 3　谱减法原理图

Fig.3　Principle diagram of the spectral subtraction method

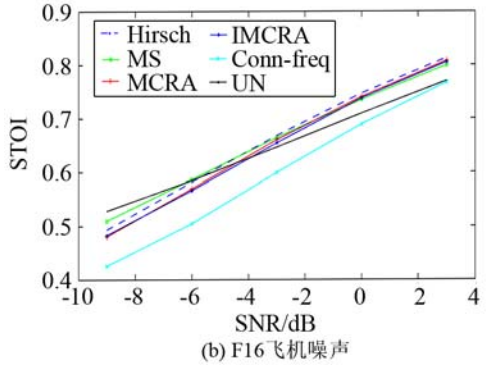
算法性能评价采用可懂度衡量指标 STOI (Short-Time Objective Intelligibility, STOI)[26]，将其用于衡量语音增强算法的可懂度性能，在 STOI 算法中，同时输入干净的语音 $x(n)$和经过增强算法重建的干净语音估计 $y(n)$，STOI 算法会给出一个(0, 1)范围内的值，STOI 值越大，表示处理后的语音的可懂度越高。图 4 显示了信噪比分别为−9、−6、−3、0、3 dB 时，在 White、F16 和 Babble 三种噪声背景下，语音增强前后的 STOI 值。

图 4 显示了不同噪声、不同信噪比环境下不同算法的 STOI 值对比，从图 4(a)可以看出，在 White 噪声背景下，MS 算法处理后的语音可懂度最高，但是在信噪比为−9、−6 dB 时仍然低于增强前的语音可懂度，也就是说，经去噪处理后，含噪语音的可懂度并未得到提高。从图 4(b)中可以看出，在 F16 噪声背景下，Conn_freq 算法处理后的语音可懂度最低，在信噪比为−3、0、3 dB 时，其他四种算法处理后的语音可懂度均得到了提高，在信噪比为−9、−6 dB 时，MS 算法处理后的语音可懂度最高，但是−9 dB 时小于增强前的语音可懂度。从图 4(c)中可以看出，在 Babble 噪声背景下，经 Conn_freq 算法处理后的语音可懂度仍是最低，MS 算法处理后的语音可懂度最高，Hirsch 算法次之，然后依次是 IMCRA 算法、MCRA 算法。

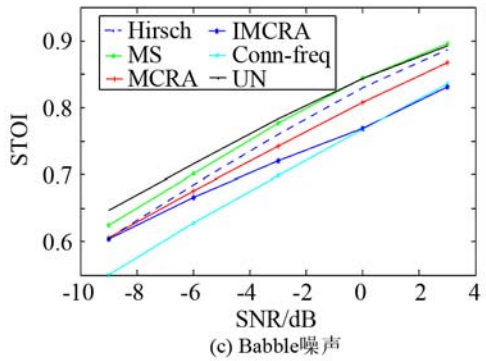
在主观听辨实验中，挑选三名听力正常测试者对增强前后的语音分别进行词语听辨测试。分别在



(a) White噪声



(b) F16飞机噪声



(c) Babble噪声

图 4 不同噪声不同信噪比环境下不同算法的 STOI 对比

Fig.4 STOIs of different algorithms versus SNRs in different noise environments

−5、0 和 5 dB 信噪比的高斯白噪声、F16 飞机噪声和 Babble 噪声背景下进行听辨实验。表 1 列出了不同算法增强后语音听辨实验中的平均词语识别率。从表 1 中可以看出，在 white −5 dB 噪声背景下，Hirsch 算法的词语识别率高于其他算法，其他情况下，采用 MS 算法增强后的语音在听辨实验中词语的平均识别率均较其他四种算法要高。

由以上分析可以得出，在 white 噪声背景下，在信噪比分别为−3、0、3 dB 时，MS 算法处理后的语音可懂度高于其他四种噪声谱估计算法和含噪语音的可懂度，而由图 1 的噪声谱估计曲线可以看出，MS 算法的噪声谱估计偏低于其他四种噪声谱估计算法。为了进一步分析五种噪声估计算法对语音可懂度的影响，下面采用语音信号增强前后的信噪比进行实验。

定义语音信号增强前的信噪比 SNR_{pre} 和增强后的信噪比 SNR_{post}，见下式：

$$SNR_{\text{pre}}=10\lg\left(\frac{\boldsymbol{X}^2(k)}{\boldsymbol{D}^2(k)}\right) \tag{11}$$

$$SNR_{\text{post}}=10\lg\left(\frac{\hat{\boldsymbol{X}}^2(k)}{\hat{\boldsymbol{D}}^2(k)}\right) \tag{12}$$

其中：$\boldsymbol{X}(k)$ 表示干净语音幅度谱；$\boldsymbol{D}(k)$ 表示噪声幅度谱；$\hat{\boldsymbol{X}}(k)$ 表示去噪后的语音幅度谱；$\hat{\boldsymbol{D}}(k)$ 表示估计的噪声幅度谱。如果 $X^2(k)/D^2(k)\geq 1$，则 $SNR_{\text{pre}}\geq 0$，表示语音信号的能量高于或等于噪声信号的能量；如果 $X^2(k)/D^2(k)<1$，则 $SNR_{\text{pre}}<0$，表示语音信号的能量低于噪声信号的能量。

文献[6]提出，当掩蔽信号过高于目标信号时，会降低目标信号的可懂度。Wang Deliang 提出的 IBM(Ideal Binary Mask)[27]实验表明，在英语含噪语音中，语音能量为主的时频块对语音可懂度的感知起关键作用，文献[28]在汉语中进行了 IBM 实验，结果表明在中文含噪语音中，语音能量为主的时频块对语音可懂度感知也起重要作用。时频块是一帧信号 FFT 后某个频率点幅度谱。

表 2 列出了 $SNR_{\text{pre}}\geq 0$ 的时频块经不同算法增强后的其信噪比仍然大于等于 0 的比例，表 3 列出

表 1 不同算法增强后的语音的词语识别率

Table 1 The world recognition rate by different algorithms

噪声	SNR/dB	词语识别率/%				
		MS	MCRA	IMCRA	Hirsch	Conn_freq
White	5	99.27	97.13	99.00	98.13	97.00
	0	99.33	97.80	98.93	98.53	97.53
	−5	93.67	86.93	91.00	96.13	93.07
Babble	5	99.53	98.37	98.67	99.00	94.00
	0	98.27	97.87	97.20	97.20	89.40
	−5	97.20	93.73	95.87	96.67	75.47
F16	5	99.40	98.53	98.20	99.37	97.53
	0	97.60	95.93	95.33	95.73	94.53
	−5	93.53	82.07	87.60	85.87	81.87

表 2　SNR_{pre}≥0 dB 的时频块经不同算法增强后的其信噪比仍然大于等于 0 的比例

Table 2　The proportions of *SNR*≥0 dB in the time-frequency blocks of SNR_{pre}≥0 dB after being enhanced by different algorithms

噪声	SNR/dB	信噪比大于等于 0 的比例/%				
		MS	MCRA	IMCRA	Hirsch	Conn_freq
White	5	0.5321	0.3770	0.3987	0.4575	0.4385
	0	0.5278	0.3720	0.3803	0.4564	0.4242
	−5	0.4975	0.3410	0.3411	0.4271	0.3717
Babble	5	0.3562	0.2614	0.3105	0.3361	0.1712
	0	0.3498	0.2524	0.2950	0.3235	0.1206
	−5	0.3382	0.2365	0.2726	0.3032	0.0754
F16	5	0.3844	0.2348	0.2445	0.3009	0.2941
	0	0.3684	0.2070	0.2129	0.2769	0.2524
	−5	0.3300	0.1670	0.1709	0.2390	0.1911

表 3　SNR_{pre}<0 dB 的时频块经不同算法增强后的其信噪比大于等于 0 的比例

Table 3　The proportions of *SNR*≥ 0 dB in the time-frequency blocks of SNR_{pre}<0 dB after being enhanced by different algorithms

噪声	SNR/dB	信噪比大于等于 0 的比例/%				
		MS	MCRA	IMCRA	Hirsch	Conn_freq
White	5	0.1911	0.0495	0.0648	0.0909	0.0348
	0	0.2282	0.0685	0.0900	0.1260	0.0427
	−5	0.2546	0.0848	0.1104	0.1550	0.0472
Babble	5	0.2117	0.1025	0.1213	0.1518	0.0873
	0	0.2514	0.1239	0.1435	0.1809	0.1000
	−5	0.2857	0.1411	0.1612	0.2052	0.1073
F16	5	0.1034	0.0142	0.0214	0.0327	0.0161
	0	0.1405	0.0237	0.0337	0.0511	0.0193
	−5	0.1728	0.0354	0.0472	0.0711	0.0211

了 SNR_{pre}<0 的时频块经不同算法增强后的其信噪比大于等于 0 的比例。

从表 2 和表 3 可以看出，不论 SNR_{pre}≥0，还是 SNR_{pre}<0，在三种噪声背景下，采用 MS 算法增强后的大于等于 0 的时频块的比例在−5~5 dB 范围内最大，这表明采用 MS 算法增强后，语音的能量大于等于噪声的能量的时频块最多，这部分语音信息没有被噪声掩盖，因此采用 MS 算法对语音作去噪处理，可以获得较高的可懂度。从表 2 中还可以看出，随着信噪比值的增大，采用同一种算法增强后的仍然大于等于 0 的时频块的比例也随之增多。从表 3 中可以看出，随着信噪比值的增大，采用同一种算法增强后的信噪比大于等于 0 的比例随之减少。

3　结 论

本文分析了 Hirsch、MS、MCRA、IMCRA 和 Conn_freq 等五种噪声谱估计算法对增强后语音可懂度的影响。详细分析了在白噪声背景下，五种算法在信噪比为−9 dB 和 5 dB 条件下的噪声谱估计，分析发现 MS 算法估计的噪声谱相比其他算法偏低。为评价算法对语音可懂度的影响，选用谱减法对含噪语音作增强处理，并对不同噪声、不同信噪比下语音增强前后的 STOI 值进行了对比，发现经 MS 算法处理后的语音可懂度高于其他算法。然后分析了增强前语音能量为主的时频块经不同算法增强后的其信噪比仍然大于等于 0 的比例和增强前噪声能量为主的时频块经不同算法增强后的其信噪比大于等于 0 的比例，通过对比发现，经 MS 算法处理后的语音中，语音的能量大于噪声的能量的时频块最多，这可能是 MS 算法相比其他噪声谱估计方法具有更高语音可懂度的原因。

参　考　文　献

[1] Yuan W, Lin J, An W, et al. Noise estimation based on time-frequency correlation for speech enhancement[J]. Applied Acoustics, 2013, **74**(5): 770-781.

[2] Lu Ching-Ta. Noise reduction using three-step gain factor and iterative-directional-median filter[J]. Applied Acoustics, 2014, **76**(1): 249-261.

[3] Ming Ji. Crookes, Danny. An iterative longest matching segment approach to speech enhancement with additive noise and channel distortion[J]. Computer Speech and Language, 2014, **28**(6): 1269-1286.

[4] Lim J. Evaluation of a correlation subtraction method for enhancing speech degraded by additive noise[J]. IEEE Transactions on Acoustics, Speech and Sinal Processing, 1978, **37**(6): 471-472.

[5] Hu Y, Loizou P. A comparative intelligibility study of sin-

gle-microphone noise reduction algorithms[J]. J. Acoust. Soc. Am., 2007, **122**(3): 1777-1786.

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

[7] McAulay R, Malpass M. Speech enhancement using a soft-decision noise suppression filter[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1980, **28**(2): 137-145.

[8] McKinley B, Whipple G. Model based speech pause detection[C]// Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on. 1997, 2: 1179-1182.

[9] Meyer J, Simmer K, Kammeyer K. Comparison of one and two-channel noise-estimation techniques[C]// Proc. 5th International Workshop on Acoustics Echo and Noise Control, IEAENC-97. 1997, 137-145.

[10] Sohn J, Kim N, Sung W. A statistical model-based voice activity detection[J]. Signal Processing Letters, IEEE, 1999, **6**(1): 1-3.

[11] Ris C, Dupont S. Assessing local noise level estimation methods: Application to noise robust ASR[J]. Speech Communication, 2001, **34**(1): 141-158.

[12] Hirsch H, Ehrlicher C. Noise estimation techniques for robust speech recognition[C]// Acoustics, Speech, and Signa Processing, 1995. ICASSP-95., 1995 International Conference on. 1995, 1: 153-156.

[13] Martin R. Spectral subtraction based on minimum statistics[C]// European Signal Processing Conference. 1994, 1: 1182-1185.

[14] Cohen I, Berdugo B. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. Signal Processing Letters, IEEE, 2002, **9**(1): 12-5.

[15] Cohen I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging[J]. IEEE Transactions on Speech and Audio Processing, 2003, **11**(5): 466-475.

[16] Sorensen K, Andersen S. Speech enhancement with natural sounding residual noise based on connected time-frequency speech presence regions[J]. EURASIP J, Applied Signal Process, 2005, **2005**(18): 2954-2964.

[17] Li N, Bao C, Xia B, et al. Speech Intelligibility Improvement Using the Constraints on Speech Distortion and Noise Over-estimation[C]// Intelligent Information Hiding and Multimedia Signal Processing, Ninth International Conference on. IEEE, 2013: 602-606.

[18] Su Y, Tsao Y, Wu J, et al. Speech enhancement using generalized maximum a posteriori spectral amplitude estimator[C]// Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013: 7467-7471.

[19] Djendi M, Scalart P. Reducing over- and under-estimation of the a priori SNR in speech enhancement techniques[J]. Digital Signal Processing, 2014, **32**(2): 124-136.

[20] Chen Y, Wu J. Forward-backward minima controlled recursive averaging to speech enhancement[C]// Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2013 IEEE Symposium on. IEEE, 2013: 49-52.

[21] Martin R. Noise power spectral density estimation based on optimal smoothing and minimal statistics[J]. IEEE Transactions on Speech and Audio Processing, 2001, **9**(5): 504-512.

[22] 中文语言资源联盟. http://www.chineseldc.org/
Chinese Linguistic Data Consortium. http://www.chineseldc.org/

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

[24] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise[C]// Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79. 1979, 4: 208-211.

[25] 张雪英. 数字语音处理及 MATLAB 仿真[M]. 北京: 电子工业出版社, 2010. 7.
ZHANG Xueying. Digital speech processing and MATLAB simulation[M]. Beijing: Publishing House of Electronics Industry, 2010, 7.

[26] Taal C, Hendriks R, Heusdens R, et al. An evaluation of objective quality measures for speech intelligibility prediction[C]// Proc. Interspeech. 2009. 2009: 1947-1950.

[27] Wang D, Kjem U, Pedersen M, et al. Speech intelligibility in background noise with ideal binary time-frequency masking[J]. J. Acoust. Soc. Am., 2009, **125**(4): 2336-2347.

[28] Zhou J, Liang R, Zhao L, et al. Whisper Intelligibility Enhancement Using a Supervised Learning Approach[J]. Circuits, Systems, and Signal Processing, 2012, **31**(6): 2061-2074.