计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(12):187−191 [doi: 10.15888/j.cnki.csa.006657]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

低信噪比下高可懂度语音增强算法①

刘　鹏

(山西工程技术学院 信息工程与自动化系, 阳泉 045000)
通讯作者: 刘　鹏, E-mail: liupeng@sxit.edu.cn

摘　要: 提出了低信噪比下高可懂度的基于分段信噪比相对均方根 (RMS) 的语音增强子空间算法. 现有的多数语音增强算法在低信噪比的恶劣条件下, 改善带噪语音质量的同时通常会伴有语音可懂度的降低. 一个重要原因是这些算法大都仅基于最小均方误差 (MMSE) 来抑制语音失真, 却忽略了语音增强算法所导致的语音失真对差异类型语音分段的可懂度影响程度不同. 为了改进这一缺点, 提出了基于短时信噪比 RMS 对语音分段进行分类, 然后调整处于信噪比中均方根语音分段的增益矩阵分量, 来减小语音失真对增强语音可懂度的影响. 客观评价实验说明, 改进算法可以改善增强语音可懂度归一化协方差评价法 (NCM) 的评测值. 主观试听实验说明, 改进算法的确提升了增强后语音的可懂度.

关键词: 子空间; 语音可懂度; 语音分段; 均方根; 增益矩阵; 客观评价; 主观试听

引用格式: 刘鹏.低信噪比下高可懂度语音增强算法.计算机系统应用,2018,27(12):187–191. http://www.c-s-a.org.cn/1003-3254/6657.html

High Intelligibility Speech-Enhancement Algorithm Under Low SNR Condition

LIU Peng

(Department of Information Engineering and Automation, Shanxi Institute of Technology, Yangquan 045000, China)

Abstract: A higher intelligibility subspace speech-enhancement algorithm based on the relative Root Mean Square (RMS) of speech segmental Signal-to-Noise Ratio (SNR) with low SNR is proposed. Under harsh conditions of low SNR, an improvement of noisy speech quality based on the majority existing speech-enhancement algorithms is often accompanied by a decrease in speech intelligibility. One important reason is that these algorithms only use Minimum Mean Square Error (MMSE) to constrain speech distortions but ignore that speech distortions caused by speech enhancement algorithms have different intelligibility influences on different speech segments. In order to overcome this disadvantage, the RMS of short-time segmental SNR was used to classify speech segments. Then the gain matrix components of middle-level RMS segments were modified to reduce the influence of speech distortion on enhanced speech intelligibility. Objective evaluation shows that the improved algorithm can improve enhanced speech intelligibility Normalized Covariance Metric (NCM) evaluation values. Subjective audition shows that the proposed algorithm does improve the enhanced speech intelligibility.

Key words: subspace; speech intelligibility; speech segment; root-mean-square; gain matrix; objective evaluation; subjective audition

语音增强算法的评估表明, 语音增强算法仅能通过抑制背景噪声来增强带噪语音的听觉舒适度以改善语音的质量, 但却无法显著提高带噪语音的可懂度, 大多仅可以保持语音的可懂度[1,2]. 事实上, 在低信噪比的

① 收稿时间: 2018-04-23; 修改时间: 2018-05-14; 采用时间: 2018-05-23; csa 在线出版时间: 2018-12-03

恶劣条件下, 改善带噪语音质量的同时经常会伴有语音可懂度的降低. 这是由于在抑制背景噪声的过程中导致原有纯净语音信号发生了较大失真, 造成了语音可懂度信息的丢失, 影响了听者的正确理解[3]. 现有的语音增强算法大都只使用最小均方误差 (MMSE) 来降低语音失真[4], 却忽略了语音增强算法所导致的语音失真对差异类型语音分段的可懂度影响程度不同.

Chen F, Loizou PC 等学者基于信噪比相对均方根 (Root-Mean-Square, RMS) 对短时语音分段进行了分类研究得到: 高均方根片段 (短时信噪比不小于整体均方根的片段)、中均方根片段 (短时信噪比小于整体均方根但不小于–10 dB 整体均方根的片段) 和低均方根片段 (短时信噪比小于–10 dB 整体均方根但不小于–30 dB 整体均方根的片段). 研究表明, 中均方根分段包含大多数辅音-元音边界, 更准确地模拟了语音可懂度[5]. Wang L, Chen F 等学者利用 RMS 对语音信号进行分割, 评估了基于 RMS 分割的语音信号边界如何影响语言可懂度预测的表现[6]. Guan T, Chu GX 等学者将语音增强算法处理后的语音按照信噪比相对均方根分段研究后发现: 语音增强算法所导致的语音失真对中均方根分段的可懂度影响更为严重, 而这正是导致增强后语音可懂度下降的一个重要原因[7].

本文在子空间语音增强算法的基础上进行改进, 提出了基于 RMS 分段的低信噪比下高可懂度子空间语音增强算法. 该算法借助先验信噪比 RMS 对带噪语音的短时分段进行了分类增强, 通过调整处于信噪比中均方根语音分段的增益矩阵分量来进一步减小中均方根分段的语音失真, 降低了语音失真对增强语音可懂度的影响, 从而在低信噪比条件下实现了增强后语音可懂度的提高.

1 子空间增强算法

假定纯净语音信号为 x, 带噪语音 y 与加性噪声 d 互不相关, 即有 $y=x+d$, 其中 y, x 和 d 都是 K 维信号矢量. 令 $\hat{x}$ 为增强语音, H 为在语音信号最小失真情况下的线性最优估计器, 其维数为 $K\times K$. 则有 $\hat{x}=H\cdot y$, 且该估计器的误差信号 ε 为:

$$\varepsilon=\hat{x}-x=(\boldsymbol{H}-\boldsymbol{I})\cdot x+\boldsymbol{H}\cdot d=\varepsilon_x+\varepsilon_d \tag{1}$$

其中, ε_x 和 ε_d 分别表示语音信号的失真和残留噪声. ε_x 的能量表示为:

$$\overline{\varepsilon_x^2}=\mathrm{E}\left[\varepsilon_x^{\mathrm{T}}\varepsilon_x\right]=tr\left(\mathrm{E}\left[\varepsilon_x\varepsilon_x^{\mathrm{T}}\right]\right) \tag{2}$$

定义

$$\sum\triangleq\boldsymbol{R}_d^{-1}\boldsymbol{R}_x=\boldsymbol{R}_d^{-1}\left(\boldsymbol{R}_y-\boldsymbol{R}_d\right)=\boldsymbol{R}_d^{-1}\boldsymbol{R}_y-\boldsymbol{I} \tag{3}$$

其中的 $\boldsymbol{R}$ 用于代表相应信号矢量的协方差矩阵. 求解频域约束条件下的方程 (4), 即得线性最优估计器 $\boldsymbol{H}_{\mathrm{opt}}$.

$$\begin{aligned}&\min_H\overline{\boldsymbol{\varepsilon}_x^2}\\&\text{subject to: }\begin{matrix}\mathrm{E}\left\{|v_k^{\mathrm{T}}\varepsilon_d|^2\right\}\le\alpha_k,\ \lambda_\Sigma(k)>0\\ \mathrm{E}\left\{|v_k^{\mathrm{T}}\varepsilon_d|^2\right\}=0,\ \lambda_\Sigma(k)\le 0\end{matrix}\end{aligned} \tag{4}$$

公式 (4) 中, α_k 为正常数.

经过矩阵特征值分解及公式化简[8,9], 求解出约束方程 (4) 的解为:

$$\boldsymbol{H}_{\mathrm{opt}}=\boldsymbol{V}^{-\mathrm{T}}\Lambda'_\Sigma\left(\Lambda'_\Sigma+\mu(k,m)\boldsymbol{I}\right)^{-1}\boldsymbol{V}^{\mathrm{T}} \tag{5}$$

其中, $\mu(k,m)$ 为短时帧 m 的第 k 个谱分量的 Lagrange 乘数, $\boldsymbol{V}$ 是矩阵 Σ 的特征向量矩阵, Λ'_Σ 是由矩阵 Σ 的非负特征值构成的矩阵 (负值以零代换), 即对于第 m 帧, 第 k 个谱分量有:

$$\lambda'_\Sigma(k,m)=\max\left\{\lambda_\Sigma(k,m),0\right\} \tag{6}$$

因此, 第 m 帧的增益矩阵为:

$$G(m)\triangleq\Lambda'_\Sigma\left(\Lambda'_\Sigma+\mu(m)I\right)^{-1} \tag{7}$$

$G(m)$ 的第 k 个对角元素 $g(k,m)$ 表示为:

$$g(k,m)=\frac{\lambda_\Sigma^{'(k,m)}}{\lambda_\Sigma^{'(k,m)}+\mu(k,m)},\ k=1,2,\cdots,\mathrm{K} \tag{8}$$

Lagrange 乘数 $\mu(k,m)$ 由下式确定:

$$\mu(k,m)=\begin{cases}\mu_0-\mathrm{SNR}_{\mathrm{dB}}^{(k,m)}/s_0, & -5<\mathrm{SNR}_{\mathrm{dB}}^{(k,m)}<20\\ 1, & \mathrm{SNR}_{\mathrm{dB}}^{(k,m)}\ge 20\\ 5, & \mathrm{SNR}_{\mathrm{dB}}^{(k,m)}\le -5\end{cases} \tag{9}$$

μ_0 和 s_0 是由实验确定的常数, 实验中 μ_0=4.2, s_0=6.25. 帧 m 的第 k 个谱分量的信噪比 $SNR_{\mathrm{dB}}^{(k,m)}$ 借助相应后验信噪比 $\gamma(k,m)$ 作为其估计值, 即 $SNR_{\mathrm{dB}}^{(k,m)}=10\lg\gamma(k,m)$, 且后验信噪比 $\gamma(k,m)$ 可由公式 (10) 求出.

$$\gamma(k,m)=\frac{\mathrm{tr}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{R}_x\boldsymbol{V}\right)}{\mathrm{tr}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{R}_d\boldsymbol{V}\right)}=\frac{\sum_{k=1}^{K}\lambda_\Sigma^{'(k,m)}}{K}=\lambda_\Sigma^{'(k,m)} \tag{10}$$

因此, 按照子空间算法增强后的语音为:

$$\widehat{\boldsymbol{x}} = \boldsymbol{H}_{\text{opt}} \cdot y = \boldsymbol{V}^{-\text{T}} G \boldsymbol{V}^{\text{T}} \cdot y \tag{11}$$

2 子空间算法改进

基于先验信噪比相对均方根对短时语音分段按照如下公式确定类型:

$$\begin{cases} \text{H}-\text{level}: 10\lg\dfrac{\xi(m)}{\xi_{\text{RMS}}} \geq 0\ \text{dB} \\ \text{M}-\text{level}: -10\ \text{dB} \leq 10\lg\dfrac{\xi(m)}{\xi_{\text{RMS}}} < 0\ \text{dB} \\ \text{L}-\text{level}: -30\ \text{dB} \leq 10\lg\dfrac{\xi(m)}{\xi_{\text{RMS}}} < -10\ \text{dB} \end{cases} \tag{12}$$

借助公式 (12) 可以实现基于短时先验信噪比的 RMS 语音段分类, 进而筛选出受语音失真可懂度影响更为严重的中均方根分段 (对应公式中的 M-level). 其中, $\xi(m)$代表帧 m 的先验信噪比, ξ_{RMS}代表含噪语音短时分段的先验信噪比相对均方根, 其计算公式如下:

$$\xi_{\text{RMS}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}|\xi(m)|^2} \tag{13}$$

令 $\xi(k, m)$ 为帧 m 第 k 个谱分量的先验信噪比, 可借助“直接判决”法[10]和公式推导[11]依据下式确定其值:

$$\begin{cases} \xi(k,m) = \alpha \cdot \sqrt{g(k,m-1)} \cdot \gamma(k,m-1) \\ \qquad + (1-\alpha)\cdot\max[\gamma(k,m)-1, 0],\ m>1 \\ \xi(k,m) = \alpha + (1-\alpha)\cdot\max[\gamma(k,m)-1, 0],\ m=1 \end{cases} \tag{14}$$

其中, α 为平滑系数, 通常在 0.8 至 1 区间取值, 改进算法中其取值为 0.98. 公式 (14) 表明, 语音增强过程中语音分段的先验信噪比可由增益矩阵和后验信噪比估计得出. 公式 (14) 中第 m–1 帧第 k 个谱分量的增益矩阵元素$g(k,m-1)$和后验信噪比$\gamma(k,m-1)$可分别通过公式 (8) 和公式 (10) 求出.

相关研究表明[12], 低信噪比 (信噪比小于零) 的条件下,信噪比和增益矩阵的估计值高于其真实值也是增强后语音可懂度降低的一个重要原因. 对带噪语音进行短时分段处理后, 由于原语音增强算法中语音失真对中均方根分段的可懂度影响更为严重, 因此可以通过文献[12]提出的人工引入偏差的方法来调整增益函数, 调整公式 (12) 中对应的中均方根区域 (M-level) 的增益函数值, 具体依照公式 (15) 对增强算法的增益矩阵分量进行调整, 来进一步减小低信噪比条件下中均方根分段的语音失真, 从而有效提高增强语音的可懂度.

$$g'(k,m) = (1-b(k,m))\cdot g(k,m) + b(k,m)$$
$$\begin{cases} 0 < b(k,m) < 1,\ -10\ \text{dB} \leq 10\lg\dfrac{\xi(m)}{\xi_{\text{RMS}}} < 0\ \text{dB} \\ b(k,m) = 0,\ \text{any else} \end{cases} \tag{15}$$

公式 (15) 中, $\boldsymbol{b}(k, m)$ 为增益调整系数, 实验中当 $-10\ \text{dB} \leq 10\lg\dfrac{\xi(m)}{\xi_{\text{RMS}}} < 0\ \text{dB}$时将$b(k,m)$在区间[0.1, 0.9]分别以步长 0.1 取值发现, $b(k,m) = 0.2$所得到的效果最好. $G'(M)$ 为基于短时先验信噪比 RMS 分类调整后的增益矩阵, 可由公式 (16) 求出.

$$G'(m) = \text{diag}(g'(1,m), g'(2,m), \cdots, g'(\text{K},m)) \tag{16}$$

因此, 依据改进算法, 具体的实施步骤如下:

(1) 按照子空间增强算法计算得到原有增益矩阵 G;

(2) 依据公式 (12) 将语音分段基于短时先验信噪比 RMS 进行分类, 筛选出受语音失真可懂度影响更为严重的中均方根分段 (M-level);

(3) 根据公式 (15) 确定增益调整系数 $\boldsymbol{b}(k, m)$, 进而通过公式 (16) 得到调整后的增益矩阵 $\boldsymbol{G}'$.

(4) 最后, 改进增强后的语音为: $\widehat{\boldsymbol{x}'} = \boldsymbol{V}^{-\text{T}}\boldsymbol{G}'\boldsymbol{V}^{\text{T}} \cdot y$

3 实验结果与分析

为了研究改进算法对带噪语音可懂度的提升效果, 在 Matlab 平台开展模拟实验. 背景噪声来源于 NOISEX-92 中的 babble, car, street 和 train, 纯净语音材料来源于“普通话言语测听材料 MSTMs”[13]. 实验中选取 MSTMs 中语句测试表的 60 个句子, 按照选定的信噪比加入同一类噪声, 再通过选定的方式处理后获得一个测试条件 (condition). 对带噪语音的增强处理方式有: 加噪未处理, 原算法处理和改进算法处理. 实验中语音可懂度的评价分别选用了客观评价法和主观试听法. 语音可懂度客观评价和主观试听均在 4 种噪声 (babble, car, street 和 train)、3 种低信噪比 (–5 dB、–10 dB 和–15 dB) 和 3 种处理方式的条件下进行, 分别产生了 36 个测试条件. 实验中信号的采样频率统一为 8 kHz, 量化精度为 16 bit, 改进算法中带噪语音按照 16 ms 进行短时分段处理.

3.1 客观评价结果与分析

语音可懂度客观评价选用归一化协方差 (Normalized Covariance Metric, NCM) 评价法[14]. 相关研究说明[15], 归一化协方差 (NCM) 法与主观试听的相关度 r=0.89,

其预测的标准偏差 σ_e=0.07, 优于 PESQ[14](r=0.79, σ_e=0.11) 等其它客观方法. 实验中把选取的 MSTMs 中 60 个日常句子的归一化协方差 NCM 平均值分别作为相应测试条件下语音可懂度的客观评价值. 表 1~表 3 给出了实验中语音可懂度的 NCM 评价结果.

表 1 信噪比 *SNR* = –5 dB, 不同条件下语音的 NCM 值

语音类型	NCM 值 *SNR* = –5 dB			
	Babble	Car	Street	Train
加噪未增强	0.44	0.47	0.52	0.49
原算法增强	0.64	0.82	0.68	0.70
改进算法增强	0.69	0.83	0.71	0.78

表 2 信噪比 *SNR* = –10 dB, 不同条件下语音的 NCM 值

语音类型	NCM 值 *SNR* = –10 dB			
	Babble	Car	Street	Train
加噪未增强	0.30	0.32	0.38	0.34
原算法增强	0.38	0.60	0.47	0.47
改进算法增强	0.49	0.73	0.57	0.61

表 3 信噪比 *SNR* = –15 dB, 不同条件下语音的 NCM 值

语音类型	NCM 值 *SNR* = –15 dB			
	Babble	Car	Street	Train
加噪未增强	0.19	0.20	0.27	0.23
原算法增强	0.15	0.29	0.26	0.26
改进算法增强	0.31	0.44	0.40	0.45

归一化协方差 (NCM) 评测值与主观试听可懂度正相关, 因此处理后的带噪语音 NCM 值越大说明其主观可懂度越高. 从表 1~表 3 语音 NCM 测试值的对比可以看出: 改进算法由于对增益矩阵进行了调整, 进一步减小了低信噪比条件下中均方根分段的语音失真, 而这种失真对语音整体的可懂度具有较大影响, 所以相较于其它两种对带噪语音的处理 (加噪未增强和原算法增强), 改进算法增强提高了增强后带噪语音的可懂度.

3.2 主观试听结果与分析

可懂度主观试听实验招募了 27 名在校大学生作为试听对象. 为了防止重复试听所导致的人为记忆对测试结果的影响, 试听采取 3 人分组, 每组只对选定的信噪比条件下的单一处理方式语音进行试听, 测试条件下的可懂度主观试听值为试听中 3 人准确识别率的均值. 表 4~表 6 给出了实验中可懂度主观试听的评价结果.

表 4 信噪比 *SNR* = –5 dB, 不同条件下语音的主观试听值

语音类型	主观试听值 *SNR* = –5 dB			
	Babble	Car	Street	Train
加噪未增强	0.55	0.64	0.75	0.65
原算法增强	0.78	0.92	0.85	0.89
改进算法增强	0.88	0.95	0.92	0.94

表 5 信噪比 *SNR* = –10 dB, 不同条件下语音的主观试听值

语音类型	主观试听值 *SNR* = –10 dB			
	Babble	Car	Street	Train
加噪未增强	0.39	0.48	0.56	0.50
原算法增强	0.43	0.61	0.54	0.54
改进算法增强	0.55	0.69	0.67	0.68

表 6 信噪比 *SNR* = –15 dB, 不同条件下语音的主观试听值

语音类型	主观试听值 *SNR* = –15 dB			
	Babble	Car	Street	Train
加噪未增强	0.28	0.33	0.44	0.38
原算法增强	0.23	0.33	0.41	0.40
改进算法增强	0.32	0.51	0.49	0.52

由于语音增强算法所导致的语音失真对中均方根分段的可懂度影响更为严重, 在低信噪比的恶劣条件下对语音整体可懂度影响很大, 调整中均方根分段的增益分量后, 增强语音的主观试听清晰度得到改善. 因此, 改进算法将带噪语音基于短时分段信噪比均方根分类增强, 实现了低信噪比条件下增强语音可懂度的提高.

4 结论语

本文在子空间语音增强算法的基础上提出了低信噪比条件下基于短时分段信噪比 RMS 分类增强的改进算法. 该算法基于短时信噪比 RMS 判断语音分段类型, 然后针对中均方根分段适当调整增益矩阵分量, 改进了现有算法单纯基于最小均方误差 (MMSE) 来抑制语音失真却忽略了失真对差异类型语音分段的影响程度不同这一不足, 进一步降低了低信噪比条件下语音失真对降噪后语音可懂度的影响. 在模拟实验中, 选取 NCM 评价法和主观试听法分别对改进算法的语音可懂度性能开展了客观和主观对比实验验证. 结果表明, 改进算法有效提高了低信噪比条件下增强语音的可懂度. 但值得注意的是, 本文所提出的子空间改进算法相较于原算法多增加了一个后置滤波的过程, 这将一定程度上增加算法的复杂度. 因此, 在非低信噪比的条件

下 (信噪比大于零), 由于原有算法导致的语音失真对可懂度影响并不严重, 此时不适合使用本文所提出的改进算法.

参考文献

1 Loizou PC, 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. [doi: 10.1109/TASL.2010.2045180]

2 Hu Y, Loizou PC. A comparative intelligibility study of single-microphone noise reduction algorithms. Journal of the Acoustical Society of America, 2007, 22(3): 1777.

3 Kim G, Loizou PC. Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms. The Journal of the Acoustical Society of America, 2011, 130(3): 1581–1596. [doi: 10.1121/1.3619790]

4 Loizou PC. Speech enhancement: Theory and practice[M]. 2nd ed. Boca Raton, Florida: CRC Press LLC, 2013.

5 Chen F, Loizou P C. Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise. The Journal of the Acoustical Society of America, 2012, 131(5): 4104–4113. [doi: 10.1121/1.3695401]

6 Wang L, Chen F, Lai YH. Segmental contribution to predicting speech intelligibility in noisy conditions. IEEE Second International Conference on Multimedia Big Data. IEEE. 2016. 476–480.

7 Guan T, Chu GX, Tsao Y, *et al*. Assessing the perceptual contributions of level-dependent segments to sentence intelligibility. The Journal of the Acoustical Society of America, 2016, 140(5): 3745–3754. [doi: 10.1121/1.4967453]

8 陈国明, 赵力, 邹采荣. 窄带噪声下的子空间语音增强算法. 应用科学学报, 2007, 25(3): 243–246. [doi: 10.3969/j.issn.0255-8297.2007.03.005]

9 贾海蓉, 张雪英, 白静. 联合听觉掩蔽效应的子空间语音增强算法. 计算机工程, 2011, 37(8): 259–261. [doi: 10.3969/j.issn.1000-3428.2011.08.090]

10 Lu Y, Loizou PC. Speech enhancement by combining statistical estimators of speech and noise. Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, TX, USA. 2010. 4754–4757.

11 刘鹏, 马建芬. 具有较高可懂度的子空间语音增强算法. 计算机工程与设计, 2013, 34(7): 2619–2622. [doi: 10.3969/j.issn.1000-7024.2013.07.068]

12 Chen F, Loizou PC. Impact of SNR and gain-function over- and under-estimation on speech intelligibility. Speech Communication, 2012, 54(2): 272–281. [doi: 10.1016/j.specom.2011.09.002]

13 张华, 王硕, 陈静, 等. 普通话言语测听材料的研发与应用. 国际耳鼻咽喉头颈外科杂志, 2016, 40(6): 355–361. [doi: 10.3760/cma.j.issn.1673-4106.2016.06.009]

14 Hu Y, Loizou PC. Evaluation of objective quality measures for speech enhancement. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(1): 229–238. [doi: 10.1109/TASL.2007.911054]

15 Ma J F, Hu Y, Loizou PC. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. Journal of the Acoustical Society of America, 2009, 125(5): 3387–3405. [doi: 10.1121/1.3097493]