计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2017,26(11):124−131 [doi: 10.15888/j.cnki.csa.006033]

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

基于改进能熵比的维纳滤波语音增强算法①

王　帅[1,2], 蒲宝明[2], 李相泽[3], 张笑东[1,2], 姚恺丰[4]

[1](中国科学院大学, 北京 100049)
[2](中国科学院 沈阳计算技术研究所, 沈阳 110168)
[3](东北大学 计算机科学与工程学院, 沈阳 110819)
[4](国家电网公司东北分部 国网东北电力调控分中心, 沈阳 110180)

摘　要: 为了提高低信噪比环境下语音增强的效果、算法的鲁棒性. 在基于维纳滤波算法的基础上, 结合基于频域特征的语音端点检查算法, 提出了一种新的语音增强算法. 端点检测算法使用小波包 ERB 子带的谱熵和改进的频域能量的能熵比法. 其中, 小波包 ERB 子带的谱熵考虑了人耳听觉掩蔽模型和语音与噪声信号之间的频率分布之间的不同; 频域能量利用了有语音帧和无语音帧的能量不同. 维纳滤波算法实时采集语音数据并使用新的参数来区别无语音段和有语音段, 并在无语音段平滑更新噪声谱. 实验结果表明, 该端点检测算法能够很好的区分有语音段和无语音段, 这就使得在低信噪比的情况下语音增强效果得到了提升, 同时算法的鲁棒性和实时性也得到了保障. 在与其他两种算法对比中, 得到了更好的语音增强效果.

关键词: 维纳滤波; 语音增强; 小波包 ERB 子带; 能熵比; 人耳掩蔽模型

引用格式: 王帅,蒲宝明,李相泽,张笑东,姚恺丰.基于改进能熵比的维纳滤波语音增强算法.计算机系统应用,2017,26(11):124–131. http://www.c-s-a.org.cn/1003-3254/6033.html

Speech Enhancement Algorithm Using Wiener Filtering Based on Improved Energy to Entropy Ratio

WANG Shuai[1,2], PU Bao-Ming[2], LI Xiang-Ze[3], ZHANG Xiao-Dong[1,2], YAO Kai-Feng[4]

[1](University of Chinese Academy of Sciences, Beijing 100049, China)
[2](Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)
[3](School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China)
[4](Northeast Branch of State Grid, Shenyang 110180, China)

Abstract: In order to achieve the improved effectiveness of speech enhancement under low-SNR circumstance and the robustness of the algorithm, this paper puts forward a new speech enhancement algorithm which is based on wiener filtering algorithm combined with speech endpoint detection algorithm on account of frequency domain features. The endpoint detection algorithm adopts the ratio between spectrum entropy for wavelet packet ERB sub-band and energy entropy for improvement of frequency domain. Therein, the spectral entropy of wavelet packet ERB sub-band takes masking properties of human auditory and the difference between speech and noise signal frequency distribution into account; the frequency-domain energy takes advantages of the energy difference between voice-frames and non- voice-frames. In addition, the wiener filtering algorithm acquits real time data and uses the new parameters to distinguish voice segment and no-voice segment where noise spectrum is updated smoothly. At last, the experimental results demonstrate that the endpoint detection algorithm can be able to effectively distinguish between speech segments and no speech segments, leading to the improvement of speech enhancement in the case of low SNR and the guarantee of robustness as well as real-time of the algorithm. In contrast with the other two algorithms, the new approach to speech enhancement has a better effect.

① 收稿时间: 2017-02-15; 修改时间: 2017-03-02; 采用时间: 2017-03-06

Key words: Wiener filtering; speech enhancement; wavelet packet equivalent rectangular bandwidth sub-band; energy to entropy ratio; masking properties of human auditory

语音增强的作用是从带噪语音信号中提取尽可能纯净的原始语音. 作为关键的语音处理技术, 语音增强在人工耳蜗、语音识别等应用中起到重要的作用. 语音增强算法[1-3]主要可以分为三类: (1) 谱减法: 这是最容易实现的语音增强算法. 因为噪声是加性的, 因此当只有噪声的时候, 可以估计和更新噪声谱, 然后从带噪信号中将噪声减去. 这种假设的前提是噪声是平稳的, 即噪声在无语音段的时候变化不大. 但是这种方法容易导致产生音乐噪声而引起语音的失真. (2) 基于统计模型的算法: 给定带噪信号的一系列测量参数, 例如基于傅里叶变换的系数, 我们系统对需要的参数找到一个线性 (或者非线性) 估计, 也就是纯净信号的一种变换系数. 维纳算法[11]和最小均方误差算法就属于这一类. (3) 子空间算法: 和前面两种算法不同, 子空间算法主要来源于线性代数理论. 在欧氏空间中, 纯净信号的分布可能只在带噪语音信号的子空间中. 因此, 如果能找到一种方法能够将带噪信号的向量空间分解成两个子空间, 其中一个子空间主要包括纯净信号, 另一个子空间主要包括噪声信号, 这样就可以简单的通过清除带噪信号向量空间中“噪声子空间”的部分, 来达到估计纯净信号的目的. 将带噪信号向量空间分解为“信号”和“噪声”子空间能够通过线性代数中的正交矩阵分解技术来实现, 特别是奇异值分解或者特征向量/特征值分解.

由于项目需要在 DSP 中实现实时的音频信号采集并进行语音增强. 经过长时间的学习分析, 选定了实现简单和去噪效果好的基于维纳滤波的语音增强算法. 维纳滤波算法中其中有比较重要的一步是进行端点检测, 端点检测的关键在于获取到区分语音段和非语音段的参数.

本研究根据语音的频域特性, 提出了基于小波包 ERB 子带的谱熵和改进的频域能量结合的新能熵系数的维纳滤波算法. 因为在差值频域能量计算中使用了噪声功率谱, 为了使噪声功率谱能够在非语音段更新, 使用前面的端点检测结果, 在非语音段对噪声功率谱进行更新. 试验表明, 本研究提出的语音增强算法能够在不同的类型和大小的噪声环境下都能够有比较好的语音质量的提升, 并具有比较好的鲁棒性.

1 基本原理

1.1 维纳滤波算法

设带有噪声的语音为:

$$y(n)=s(n)+d(n) \tag{1}$$

其中, $s(n)$为纯净语音信号; $d(n)$为噪声信号. 需要设计这样一个滤波器$h(n)$, 当输入为$y(n)$时, 滤波器的输出为:

$$\hat{s}(n)=y(n)*h(n) \tag{2}$$

$\hat{s}(n)$是对纯净语音信号$s(n)$的估计, 它使用最小均方误差准则使$s(n)$和$\hat{s}(n)$的均方误差 $\varepsilon=E\{[s(n)-\hat{s}(n)]^2\}$达到最小, 进而求出滤波器 $h(n)$.

由正交性原理, 效果最好的$h(n)$要对于所有的 m, 公式 (3) 都要成立.

$$E\{[s(n)-\hat{s}(n)]\cdot y(n-m)\}=0 \tag{3}$$

把$\hat{s}(n)=y(n)*h(n)$带入式 (3), 对等式两边做傅里叶变换. 而且纯净语音信号$s(n)$和噪声信号$d(n)$不相关, 因此推导出:

$$H(k)=\frac{P_s(k)}{P_s(k)+P_d(k)} \tag{4}$$

式中, $P_s(k)$为$s(n)$的功率谱密度; $P_d(k)$为噪声信号$d(n)$功率谱密度. 因为语音是短时平稳的信号, 并且语音的功率谱我们也没办法求得, 所以把式 (4) 改写为:

$$H(k)=\frac{E\left[|S(k)|^2\right]}{E\left[|S(k)|^2\right]+P_d(k)} \tag{5}$$

上式分子分母同除$P_d(k)$得:

$$H(k)=\frac{\varepsilon(k)}{1+\varepsilon(k)},H(k)=1-\frac{1}{\gamma(k)} \tag{6}$$

其中, $\varepsilon(k)=\frac{E\left[|S(k)|^2\right]}{P_d(k)}$是先验信噪比; $\gamma(k)=\frac{E\left[|Y(k)|^2\right]}{P_d(k)}$是后验信噪比. 引入平滑参数 α, 得到:

$$\begin{aligned}\varepsilon_i(k)&=\alpha\varepsilon_i(k)+(1-\alpha)\varepsilon_i(k)\\&=\alpha\varepsilon_i(k)+(1-\alpha)(\gamma_i(k)-1)\\&\approx\alpha\varepsilon_{i-1}(k)+(1-\alpha)(\gamma_i(k)-1)\end{aligned} \tag{7}$$

因此, 可以使用第 i-1 帧的先验信噪比以及第 i 帧的后验信噪比求出第 i 帧的先验信噪比, 这样就能够求出本帧的维纳滤波器传递函数:

$$H_i(k)=\frac{\widehat{\varepsilon_i}(k)}{1+\widehat{\varepsilon_i}(k)} \tag{8}$$

由:

$$\widehat{S}(n)=H(k)\cdot Y(k) \tag{9}$$

我们可以求得$s(n)$的估计$\hat{s}(n)$的傅里叶变换, 那么我们对$\widehat{S}(n)$进行反傅里叶变换, 即可求得$\hat{s}(n)$, 也就得到了增强后的语音信号.

1.2 小波包 ERB 临界频带谱熵

对语音信号$s(n)$分帧加窗, 进行快速傅里叶变换. 得到每帧信号的频谱$S_i(f_m)$. 对每帧内的频谱分量进行归一化处理, 即得到每个频谱分量的概率密度函数.

$$P_m=S_i(f_m)/\sum_{k=0}^{N-1}S_i(f_k),(m=1,2,\cdots,N) \tag{10}$$

则第 i 帧的谱熵为:

$$H_i=-\sum_{m=1}^{N}P_m\log P_m \tag{11}$$

从谱熵的定义可以看出, 谱熵反映了信源在频域幅值分布的“无序性”. 对于噪声来说 它的归一化谱概率密度函数分布比较均匀, 所以它的谱值就大; 然而对于语音信号来说. 由于频谱具有共振峰频谱特性, 它的归一化频谱密度函数分布不均匀, 是的语音的谱熵一般来说都低于噪声的谱熵. 因此, 利用这个特性能够从带噪语音中提取出语音的端点.

为了消除每帧信号 FFT 后的谱线幅值受到噪声的影响, 把每条谱线的谱熵改为子带的谱熵, 并且经过证明, 这样能够提高语音信号和噪声信号的区分度. 又根据人耳掩蔽模型[8,14], 这样能够更好的模拟人类听觉, 进而引入 ERB-scales 临界频带[5,6], 把人耳可听的声音分为 24 个频带. 使用多尺度一维离散小波变换, 对每帧的语音进行 8 层分解. 中心频率和 ERB-scales 关系由下式给出:

$$\mathrm{ERB}=24.7\left(\frac{4.37f_c}{1000}+1\right) \tag{12}$$

根据 ERB-scales 的中心频率得到以下小波包 ERB 子带如图 1 所示.

在小波变换中, 因为 Daubechies 小波具有很好的正交性, 可以提供更加有效的分解和重构, 所以本研究使用了 Daubechies 小波中的 db2 小波作为母函数. 对应的 ERB 坐标的小波包频带划分和小波包 ERB 频带划分如表 1 所示.

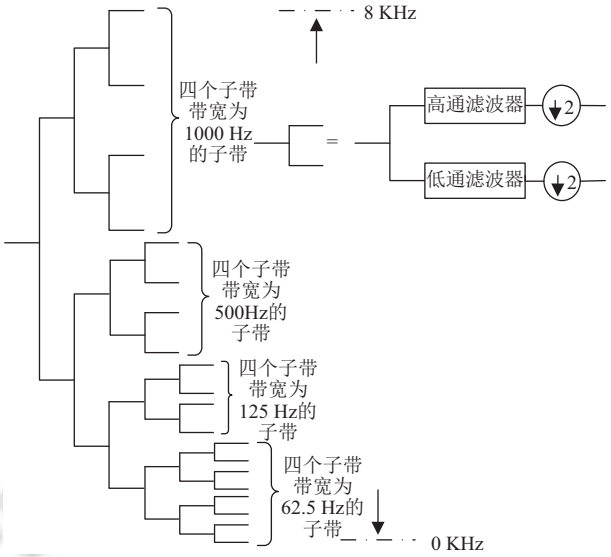

图 1 小波包 ERB 子带

表 1 频带划分

序号	ERB子带	小波包ERB子带	序号	ERB子带	小波包ERB子带
24	6917.58	8000	12	1086.66	1000
23	5977.56	7000	11	913.62	875
22	5161.17	6000	10	763.35	750
21	4452.17	5000	9	632.83	625
20	3836.44	4000	8	519.49	500
19	3301.7	3500	7	421.06	437.5
18	2837.29	3000	6	335.57	375
17	2433.98	2500	5	261.33	312.5
16	2083.71	2000	4	196.85	250
15	1779.52	1750	3	140.86	187.5
14	1515.35	1500	2	92.23	125
13	1285.92	1250	1	50	62.5

把每一帧音频文件进行小波包分解, 并对每个频带求概率密度得到:

$$P_c=S_i(f_c)/\sum_{k=0}^{L-1}S_i(f_k),\quad c=1,2,\cdots,L \tag{13}$$

则 ERB 临界频带谱熵为:

$$H_i\_CB=-\sum_{m=1}^{L}P_c\log P_c \tag{14}$$

图 2 是使用 ERB 临界频带谱熵进行端点检测的结果. ERB 临界频带谱熵能够很好的体现人耳的听觉感知特性.

1.3 改进的频域能量

带噪语音段和无语音段之间的能量差异能够用于语音端点的检测, 当前主要使用语音波形时域信号的短时能量. 短时能量能够反映音频信号在分帧后每一

帧的能量, 在信噪比比较高的时候对有语音段和无语音段的区分度比较好.

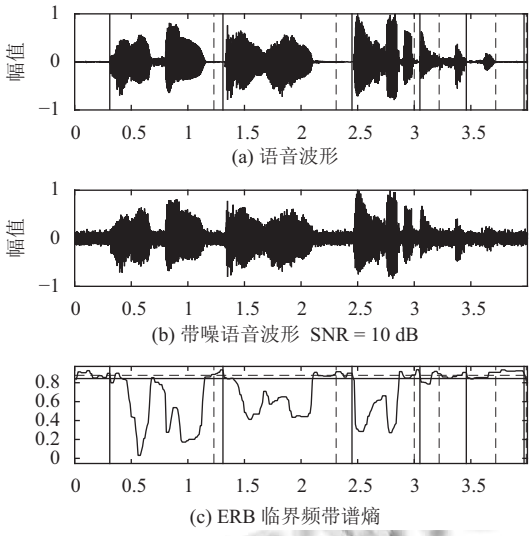

图 2 ERB 临界频带谱熵进行端点检测

对语音信号$s(n)$分帧加窗, 求其短时能量:

$$E_i = \sum_{n=1}^{N} s_n^2 \tag{15}$$

但在低信噪比情况下, 纯噪声和带噪语音的短时能量差别变小, 这时基于短时能量的语音端点检测几乎失效, 所以我们引入改进的频域能量[12]:

$$P_E(n) = \sum_{i=0}^{N-1} \left(P^{(i)}(n) - P_{noise}^{(i)} \right) \tag{16}$$

即带噪语音信号与背景噪声在各频率分量功率谱幅值的差值之和, 这样能够减小噪声对短时能量计算的影响. 由于噪声信号是非平稳的, 所以本文提出了在无语音段去更新噪声谱, 这样就进一步减小了噪声的非平稳性对端点检测的影响. 最终得到的差值频域能量更好的适应低信噪比的环境.

进一步引入改进的能量计算关系:

$$P_D(n) = \log_{10}(1 + P_E(n)/a) \tag{17}$$

式中, a 是个常数, 本文中取 a=2.3. 由于引进了 a, 这样使得当 a 存在时, $P_E(n)$的幅值有剧烈的变化时在$P_D(n)$中缓和. 所以选择适当的 a, 有助于区分噪声和轻音.

图 3 反映了改进的频域能量的检测结果. 信噪比高的时候效果非常好.

1.4 改进的能量熵系数

文献[4]中提出了结合语音信号的谱熵和短时能量的测量参数 EE-Feature 来进行端点检测, 如下式所示:

$$EEF_i = \sqrt{1 + |(E_i - C_E) \cdot (H_i - C_H)|} \tag{18}$$

其中, E_i 和 H_i 第 i 帧的短时能量和谱熵, C_E 和 C_H 是语音前十帧的平均值.

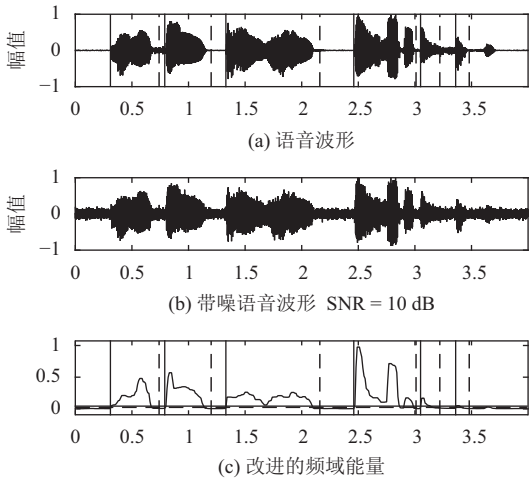

图 3 频域能量进行端点检测

由图 2 和图 3 中可以看出, 在音频信号的有语音段能量是向突起的, 而谱熵相反在有语音段向下凹. 这说明有语音段的能量大, 而谱熵[7]小. 所以把能量值除以谱熵的值, 能够更突出有语音段的数值, 噪声段的值更小, 进一步拉开了有语音段和噪声段的数值差. 这样更容易的检测出语音端点.

本研究结合 ERB 临界频带谱熵和改进的差值频域能量进一步改进得到一个能熵比系数.

$$EEF_i = \sqrt{1 + \left| (P_D(i) - C_{P_D}) / \left(H_{CB}(i) - C_{H_{CB}} \right) \right|} \tag{19}$$

其中, $P_D(i)$是第 i 帧的改进的差值频域能量, $H_{CB}(i)$是第 i 帧的 ERB 临界频带谱熵, C_{P_D}是背景噪声的差值频域能量, $C_{H_{CB}}$是背景噪声的 ERB 临界频带谱熵.

由图 4, 可以看出断点检测效果比单独的 ERB 临界频带谱熵和差值频域能量都更加准确. 噪声和语音段区分的十分明显.

2 改进后的语音增强算法

使用传统的端点检测算法, 使得维纳滤波算法进行语音增强的效果不佳. 本研究引入 1.3 节提到的改进的能熵比法. 改进后的算法能够更好的进行端点检测[9,10], 增强后的语音信号也得到了相应的提高, 对语音信号进行语音增强的具体步骤如下:

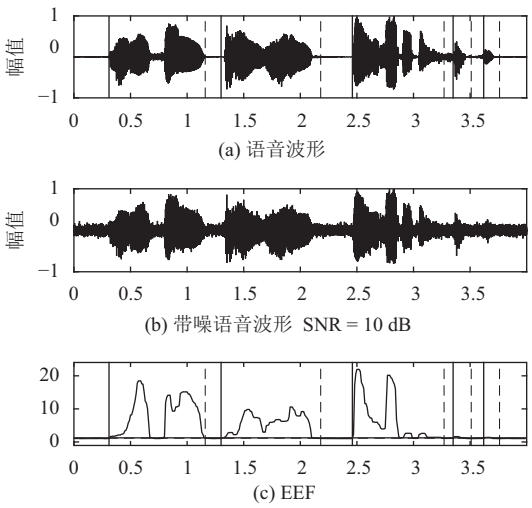

图 4 改进的能熵比进行端点检测

(1) 带噪语音为$x(n)$, 经过分帧后得到$x_i(n)$, 其中N为每一帧的采样点数, i 范围为 1 到$\frac{n}{N/2}-1$, 相邻之间有 80 个采样点的重合.

(2) 取语音前 120 ms 作为无语音帧进行噪声功率谱的估计.

(3) 对第 i 帧做快速傅里叶变换, 并求得带噪语音功率谱.

(4) 由公式 (7) 更新先验信噪比.

(5) 使用改进的 EEF 进行端点检测, 并在无语音段更新噪声功率谱.

(6) 使用公式 (8) 计算出维纳滤波数字滤波器 $h(n)$ 的傅里叶变换. 在使用公式 (9) 计算的到纯净信号的估计$\hat{s}(n)$的傅里叶变换$\hat{S}(n)$.

(7) 对$\hat{S}(n)$进行傅里叶反变换, 就得到了经过滤波的第 i 帧信号. 继续第 3 步, 直到$i > \frac{n}{N/2}-1$.

(8) 输出增强后的语音信号.

算法流程图如图 5 所示.

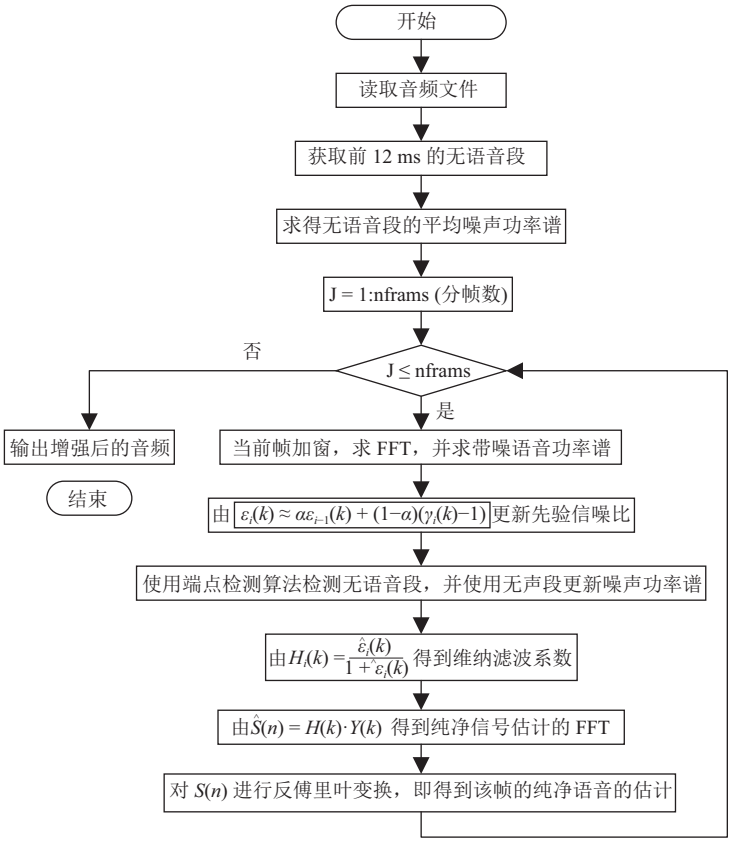

图 5 算法流程图

3 实验仿真

3.1 测试数据

实验中使用到的测试数据, 使用的是用录音笔录制的. 语音录制在双壁隔音室中一共有 30 个语句, 6 个人进行录制的. 采样率 8 kHz, 16 bit 量化. 实验中使用帧长 160 进行分帧, 相邻帧重叠 80, 使用 Hamming 窗对信号加窗. 进行端点检测的 EEF 阈值设置为前 120 ms

的平均 EEF 值的 1.1 倍.

带噪语音通过录制的纯净语音和噪声进行合成得到的. 背景噪声有 babble(说话噪声) 和高斯白噪声. 通过将不同幅度的 babble 和白噪声和纯净语音信号相叠加, 来模拟不同的信噪比的带噪语音信号. 通过改变噪声的方差σ^2来对带噪语音进行信噪比的控制.

为了证明算法的增强效果, 本文选择基于对数谱距离、奇异谱分析和基于改进的 EEF 的维纳滤波语音增强算法进行大量的实验对比分析.

3.2 实验结果

3.2.1 不同噪声实验效果

如图 6 所示是本文的方法消除信噪比为 1 的高斯白噪声的结果. 该例子中, 采用的语音数据是一段男声汉语语音, 内容为: “蓝天, 白云, 碧绿的大海”, 数据长度 32000.从图中带噪语音信号和增强后的信号的对比中可以看出对白噪声有很好的抑制效果. 增强后的信噪比为 8.6803, 信噪比提高了 7.68.

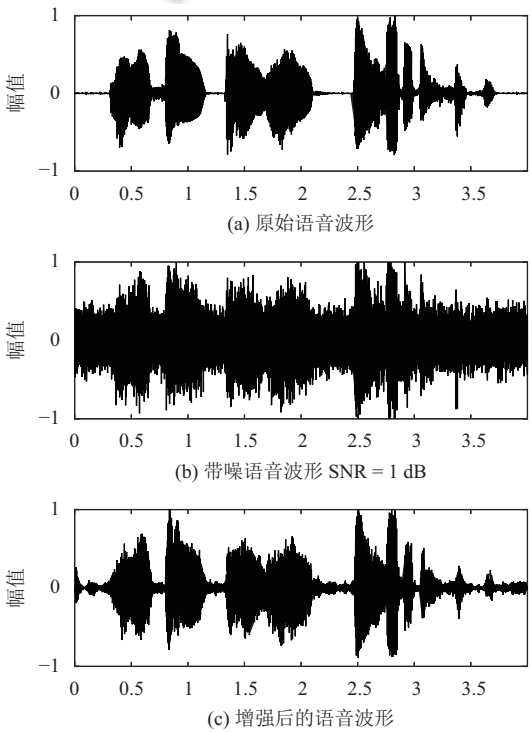

图 6 高斯白噪声增强结果

如图 7 所示是本文的方法消除信噪比为 1 的 babble 噪声的结果. 由图可知, 本文算法对 babble 噪声在大部分情况下有很好的消除效果. 在 babble 噪声比较大的情况下具有一定的消除效果. 由于 babble 噪声的不稳定性, 本文算对高斯白噪声的增强效果好于对 babble 噪声的效果. 因此算法对平稳的噪声增强效果更好. 这和维纳滤波算法的原理有直接的关系.

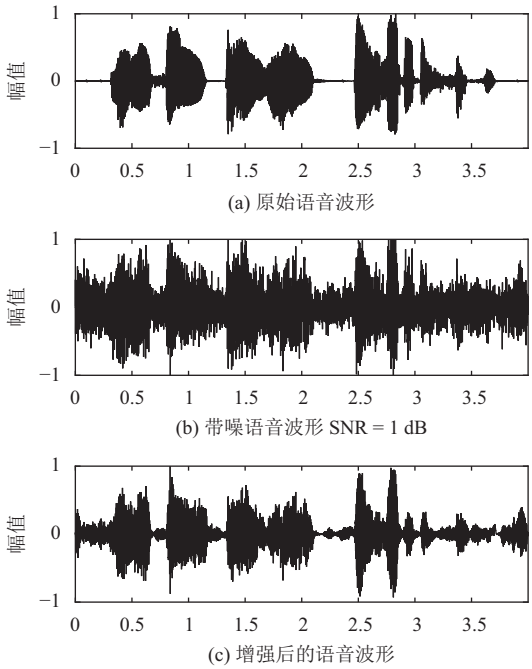

图 7 Babble 噪声增强效果

3.2.2 不同噪声强度下实验效果

如表 2 所示是本文的方法消除信噪比为 1、5 和 10 的高斯白噪声的结果. 对不同信噪比的语音都有比较好的增强效果. 对比奇异谱分析方法都能提高 0.5 左右的信噪比.

表 2 不同噪声强度增强结果

带噪语音信噪比	增强后信噪比	改善
1	8.6803	7.6803
5	11.5432	6.5432
10	14.9779	4.9779

3.2.3 不同噪声和不同算法下实验效果

为了检验算法的有效性, 在不同信噪比下使用本文的算法与基于对数谱距离的语音增强算法、奇异谱分析[13]语音增强算法进行了实验对比.

将上述三种算法分别在不同信噪比下的带噪语音进行大量的实验, 来检验算法的去噪效果. 如图 8 所示是三种方法消除信噪比为 1 的高斯白噪声的结果. 从图中带噪语音信号和增强后的信号的对比中可以看出本文中的算法对白噪声有很好的抑制效果. 如图 9 所示是三种方法消除信噪比为 1 的 babble 噪声的结果.

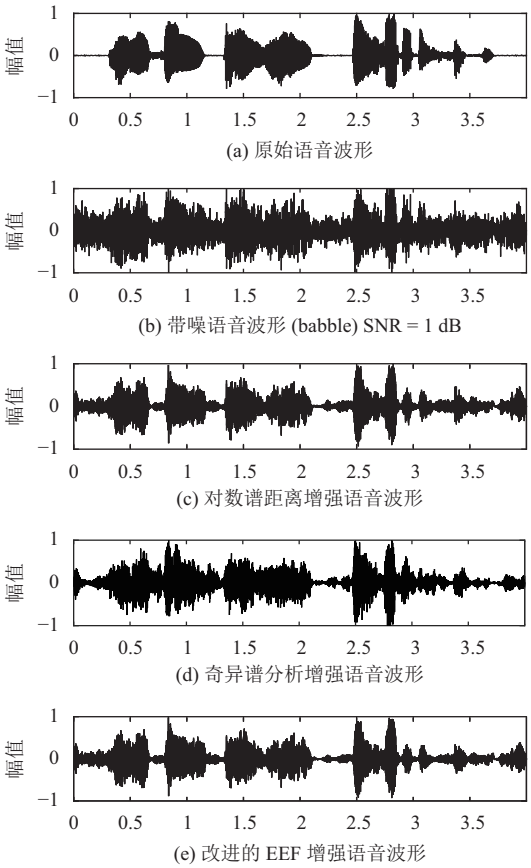

图 8 三种算法 Babble 噪声增强效果对比

由图可知, 本文算法对 babble 噪声在大部分情况下有很好的消除效果. 实验证明采用本文提出的基于改进的能熵比的维纳滤波语音增强算法后语音的信噪比的到了很大的提高, 并且对增强后的语音进行主观评价, 语音质量得到了很大的提高.

为了对本文方法做出更客观的评价, 本文使用分段信噪比测度来对增强后的语音质量进行衡量. 分段信噪比 (SNRseg)[15]定义为:

$$\mathrm{SNRseg}=\frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n))^2} \tag{20}$$

由表 2 可知, 当输入的信噪比相同时本文提出的改进算法具有更好的效果, 输出信噪比较高. 与其他算法相比更能够适合低信噪比的信号进行增强. 奇异谱分析算法在奇异值较大的地方会把噪声当作信号, 导致增强效果不佳. 对数谱距离算法也是基于维纳滤波算法的, 对端点的检测没有本文的改进 EEF 算法更加的精准, 因此没有本文算法语音增强效果好.

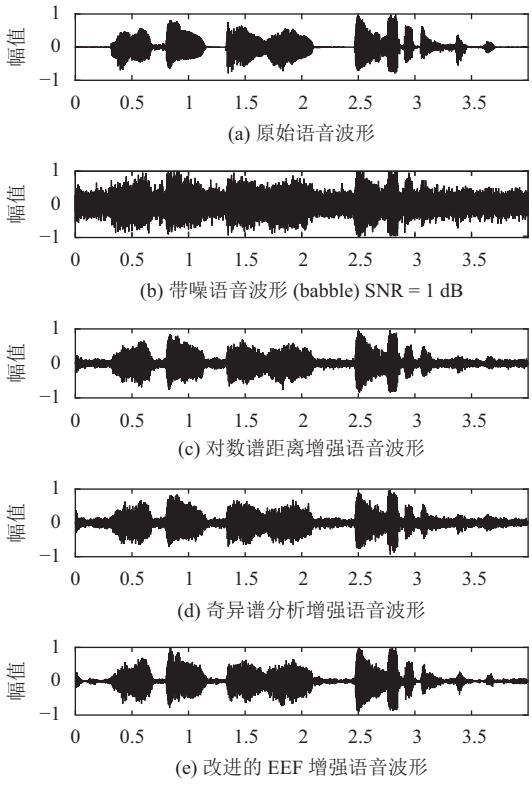

图 9 三种算法高斯白噪声增强结果对比

表 3 不同噪声强度增强结果

输入信噪比(dB)	输出信噪比(dB)		
	对数谱距离	奇异谱分析	改进的EEF
–10	1.6854	1.5876	2.1850
–5	3.5976	3.5480	4.2325
0	6.8342	7.5637	7.9789
5	9.8724	10.2556	11.5432
10	12.6739	12.8840	14.9779

4 结语

根据大量的文献阅读, 本文结合人耳听觉掩蔽 ERB 模型和语音信号的频域特性. 得到了 ERB 临界频带谱熵和改进的频域能量. 本文提出了结合这两个参数得到的改进的能熵比法 (EEF), 进一步提出了基于改进的能熵比法的维纳滤波语音增强新算法. 不同的信噪比和检测方法下进行比较实验, 本文的算法具有比较好的鲁棒性和去噪效果. 在信噪比低的情况下也能得到较好的结果. 算法最终提高了带噪语音的信噪比, 减少了信号的失真. 因为该算法使用小波包分解算法, 影响了算法的实时性. 下一步工作要对 ERB 频带使用计算量小的分解方法, 提高算法的实时性, 减少计算量. 目前想到的解决方法是使用 FFT 变换得到信号的频谱, 然后把得到的频谱按照 ERB 频带进行重新统计.

这样就避免了使用小波变换, 减少了计算复杂度.

参考文献

1 Weiss MR, Aschkenasy E, Parsons TW. Study and development of the INTEL technique for improving speech intelligibility. Final Technical Report, 1975.

2 McAulay R, Malpass M. Speech enhancement using a soft-decision noise suppression filter. IEEE Trans. on Acoustics, Speech, and Signal Processing, 1980, 28(2): 137–145. [doi: 10.1109/TASSP.1980.1163394]

3 Dendrinos M, Bakamidis S, Carayannis G. Speech enhancement from noise: A regenerative approach. Speech Communication, 1991, 10(1): 45–57. [doi: 10.1016/0167-6393(91)90027-Q]

4 Huang LS, Yang CH. A novel approach to robust speech endpoint detection in car environments. Proc. of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Istanbul, Turkey. 2000, 3. 1751–1754.

5 Sahu PK, Biswas A, Bhowmick A, *et al*. Auditory ERB like admissible wavelet packet features for TIMIT phoneme recognition. Engineering Science and Technology, An International Journal, 2014, 17(3): 145–151. [doi: 10.1016/j.jestch.2014.04.004]

6 Smith JO, Abel JS. Bark and ERB bilinear transforms. IEEE Trans. on Speech and Audio Processing, 1999, 7(6): 697–708. [doi: 10.1109/89.799695]

7 孙炯宁, 傅德胜, 徐永华. 基于熵和能量的语音端点检测算法. 计算机工程与设计, 2005, 26(12): 3429–3431. [doi: 10.3969/j.issn.1000-7024.2005.12.085]

8 张春雷, 曾向阳, 王曙光. 基于临界带功率谱方差的端点检测. 声学技术, 2012, 31(2): 204–208.

9 尹晨晓, 郭英, 张碧锋, 等. 基于 Bark 小波的语音端点检测算法. 计算机工程, 2011, 37(12): 276–278.

10 张雪英, 李寿永. 小波包滤波器用于语音识别前端处理. 电子测量与仪器学报, 2002, (增刊): 953–956.

11 李战明, 尚丰. 一种基于语音端点检测的维纳滤波语音增强算法. 电子设计工程, 2016, 24(2): 42–44.

12 郭逾, 张二华, 刘驰. 一种基于频域特征和过渡段判决的端点检测算法. 山东大学学报 (工学版), 2016, 46(2): 57–63. [doi: 10.6040/j.issn.1672-3961.2.2015.147]

13 靳立燕, 陈莉, 樊泰亭, 等. 基于奇异谱分析和维纳滤波的语音去噪算法. 计算机应用, 2015, 35(8): 2336–2340. [doi: 10.11772/j.issn.1001-9081.2015.08.2336]

14 宫云梅, 赵晓群, 史仍辉. 基于语音存在概率和听觉掩蔽特性的语音增强算法. 计算机应用, 2008, 28(11): 2981–2983, 2986.

15 Loizou PC. 语音增强: 理论与实践. 高毅, 译. 北京: 电子科技大学出版社, 2012: 501–503.