

基于 MMSE 和小波变换的语音增强算法

潘小龙, 张卫强*, 郦元宏

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 提出一种最小均方误差与小波变换相结合的改进算法来提高语音识别系统的性能. 首先通过小波变换对带噪语音进行多尺度分解, 然后对不同尺度的小波系数采用 MMSE 算法, 最后对 MMSE 处理过的小波系数进行重构得到增强语音. 实验结果表明: 与传统算法相比较, 提出的新算法去噪效果显著, 有效地提高了语音的质量.

关键词: 最小均方误差; 小波变换; 语音增强

中图分类号: TN912.35

文献标志码: A

文章编号: 1001-5132 (2016) 03-0068-04

语音识别在安静的环境中可以有较好的性能, 但现实环境中不可避免地夹杂着噪声, 对语音识别产生了较大影响, 因此对带噪语音的处理就显得特别重要. 目前, 语音增强方法主要有基于语音短时谱幅度估计法(Short Time Spectral Amplitude, STSA)、基音周期法、语音参数模型法、听觉场景分析法等.

在基于语音短时谱幅度估计法(STSA)中, 最小均方误差(Minimum Mean-Square Error, MMSE)比谱减法等传统方法能更多地利用语音和噪声统计特性的先验知识^[1], 对语音系数的先验概率进行估计, 并在此基础上按照最小均方误差的准则对语音系数进行估计. MMSE 准则中, 下信号短时间可以看成是平稳信号, 但当信号为非平稳时, 其去噪效果大大降低, 而且在低信噪比情况下, 由语音增强失真引起的包括语音频谱变形和增强后的剩余噪声对信号有较大的干扰, 从而降低语音识别系统的精度. 为解决上述问题, 人们提出了基于小波变换的语音增强算法. 鉴于小波变换在低信噪比情况下去噪效果好和处理非平稳信号的优势, 可以改善 MMSE 在低信噪比下信号的失真, 因此, 将 MMSE 与小波变换结合在一起处理语音信号可以有效提高语音识别系统的性能.

1 MMSE 语音增强

对于传统去噪算法如谱减法^[2]等进行语音增强后所产生的“音乐噪声”, MMSE 算法可以较好地解决该缺陷, 这是由于 MMSE 方法能更多地利用语音和噪声统计特性的先验知识. 以下基于 MMSE 的短时频谱分析来介绍 MMSE 原理. 语音短时频谱可用指数形式作如下表示^[3]:

$$X(k, i) = A(k, i) \exp(j\alpha(k, i)), \quad (1)$$

$$Y(k, i) = N(k, i) \exp(j\theta(k, i)), \quad (2)$$

上述公式将信号分为若干帧信号, 其中, i 代表信号为第 i 帧, $X(k, i)$ 和 $Y(k, i)$ 分别为纯净语音和带噪语音的频谱函数, $A(k, i)$ 和 $N(k, i)$ 分别为纯净语音和带噪语音的语音短时谱幅度. 对于带噪语音信号 $y(t) = x(t) + d(t)$, MMSE 目的是使纯净语音的短时谱幅度估计值与纯净语音短时谱幅度最接近, 亦即使信号失真度 $E\{(A(k, i) - \hat{A}(k, i))^2\}$ 最小.

假设各频谱分量之间相互独立, 则可以推导出语音短时谱的 MMSE 估计值为:

$$\begin{aligned} \hat{A}(k) &= E\{A(k) | Y(k)\} = \\ &= \int_0^{+\infty} a(k) p(a(k) | Y(k)) da(k) = \\ &= \left(\int_0^{+\infty} \int_0^{2\pi} a(k) p(a(k), \alpha(k)) p(Y(k) | a(k), \alpha(k)) d\alpha(k) da(k) \right) / \end{aligned}$$

收稿日期: 2014-12-04.

宁波大学学报(理工版) 网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家自然科学基金 (61271137).

第一作者: 潘小龙 (1988-), 男, 浙江台州人, 在读硕士研究生, 主要研究方向: 嵌入式系统. E-mail: feilong.21@163.com

*通信作者: 张卫强 (1963-), 男, 浙江宁波人, 副教授, 主要研究方向: 嵌入式系统. E-mail: zhangweiqiang@nbu.edu.cn

$$\alpha(k)) da(k) d\alpha(k)) / (\int_0^\infty \int_0^{2\pi} p(a(k), \alpha(k)) p(Y(k) | a(k), \alpha(k)) da(k) d\alpha(k)), \quad (3)$$

其中, $a(k)$ 为 $A(k)$ 的一帧信号所对应的谱幅度. 假设噪声谱服从零均值高斯分布, 则:

$$p(Y(k) | a(k), \alpha(k)) = \frac{\exp\{-\frac{|Y(k) - a(k)e^{j\alpha(k)}|^2}{\pi\lambda_D(k)}\}}{\pi\lambda_D(k)}. \quad (4)$$

假设语音谱服从高斯分布, 则其幅值和相位的联合分布为:

$$p(a(k), \alpha(k)) = \frac{1}{\pi\lambda_X(k)} \exp\{-\frac{a^2(k)}{\pi\lambda_X(k)}\}, \quad (5)$$

$$\lambda_X(k) = E[|X(k)|^2], \quad (6)$$

$$\lambda_D(k) = E[|D(k)|^2], \quad (7)$$

其中, $D(k)$ 为噪声的谱幅度; $\lambda_X(k)$ 为纯净语音能量期望值; $\lambda_D(k)$ 为噪声能量期望值, 可由语音间歇时静音帧估计得到, 代入式(3)可得公式如下:

$$\hat{A}(k) = \Gamma(1.5) \frac{\sqrt{v(k)}}{\gamma(k)} M(-0.5; 1; -v(k)) N(k), \quad (8)$$

其中, $\Gamma(1.5) = \sqrt{\pi}/2$, $M(a:c:x)$ 为超几何函数^[4], 则:

$$v(k) = \frac{\zeta(k)}{\zeta(k)+1} \gamma(k), \quad (9)$$

$$\zeta(k) = \frac{\lambda_X(k)}{\lambda_D(k)}, \quad (10)$$

$$\gamma(k) = \frac{|Y(k)|^2}{\lambda_D(k)}, \quad (11)$$

其中, $\zeta(k)$ 和 $\gamma(k)$ 分别为先验信噪比和后验信噪比. 可将式(8)写成如下形式:

$$\hat{A}(k) = G(k) N(k). \quad (12)$$

其中, $G(k)$ 为增益函数. 因此由以上公式可得到纯净语音谱幅度估计值, 对其添加含噪信号的相位及反傅里叶变换后就可得到增强后的语音.

2 小波变换

作为时频域分析方法, 傅里叶变换能将信号不同时刻的相同频率成分都映射到同一频率点上, 因此无法对信号某个时间点的某个频率的特征进行分析, 相当于不能表述信号的时频局部性质, 这正是非平稳信号的关键. 小波变换是空间和频率的局部变换, 因而能有效地从信号中提取信息, 克服

上述缺点. 小波去噪首先进行多尺度小波变换, 然后对小波系数进行处理, 去除噪声中的小波系数, 保留原始信号的小波系数, 对小波系数进行小波逆变换(小波重构), 最后得到去噪后的近似真实信号^[5-6]. 假设函数 $\varphi(x)$ 为平方可积函数, 即 $\varphi(x) \in L^2(R)$, 且满足以下容许性条件:

$$C_\varphi = \int_R \frac{|\varphi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (13)$$

则称 $\varphi(x)$ 为小波基函数. 对于任意实数 (a, b) , a 为小波变换的尺度因子, b 为平移因子, 对其进行离散小波变换, 此时通常定义 $a = a_0^j$, $b = ka_0^j b_0$, 其中, $j, k \in Z$. 由小波基 $\varphi(x)$ 生成的依赖于 (a, b) 的离散小波函数如下所示^[7-8]:

$$\varphi_{j,k}(x) = \varphi_a(x-b) = |a|^{-\frac{j}{2}} \varphi(a_0^{-j}x - kb_0). \quad (14)$$

则信号 $f(t)$ 以 $\varphi(x)$ 为基的小波变换为:

$$W_f(j, k) = \int_R f(t) \overline{\varphi_{j,k}(t)} dt = |a|^{-\frac{j}{2}} \int_R f(t) \overline{\varphi(a_0^{-j}x - kb_0)} dt. \quad (15)$$

在实际应用中, 通常取 $a_0 = 2$, $b_0 = 1$. 由此可见, 离散小波变换的提出, 使得小波变换在计算机的分析成为可能.

3 改进的算法

带噪声语音通过小波变换划分得到不同尺度的小波系数, 小波变换的有用信号能量主要集中在大的小波系数和特定的频率范围内, 而噪声分布于整个小波域内, 因此小波分解后有用的信号系数幅值会大于噪声的系数幅度, 可采取阈值方法来得到有用信号. 然后对各个尺度的小波系数进行 MMSE 处理, 最后对处理过的小波系数进行小波重构, 得到增强后的语音信号. 新算法的整体框架如图 1 所示.

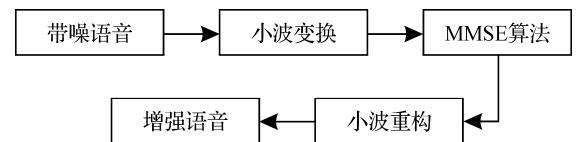


图1 语音增强模型

首先对带噪声语音进行 Mallat 小波分解, 得到不同尺度的分解系数, 在频域中对应不同频率的子带信号. 假定 $0 \sim P$ 频率空间为 F_0 , 经过一级分解后被分为 $0 \sim P/2$ 的低频子空间 F_1 和 $P/2 \sim P$ 的

高频空间 W_1 , 然后将低频空间一直分解下去, 得到 $F_n, W_n, \dots, W_2, W_1$, 它们之间的空间互不相交, 它们所对应的小波系数为 $cd1, cd2, cd3, cd4, cd5, ca5$ (其中 cdn 为高频系数, $ca5$ 为第五层的低频系数). 然后对不同频率范围、尺度的小波系数进行 MMSE 估计, 得到改进的小波系数. 最后对处理过的小波系数利用 Mallat 算法进行重构, 得到增强语音.

4 改进算法的仿真结果

运用小波变换进行去噪处理, 可选择不同小波基函数. 仿真选 Daubechies 为正交小波基, 阶次为 4, 分解层数为 5 层; 语音为在噪声环境下的“宁波”两个字的发音信号, 16 kHz 的采样率, PCM, 16 位, 单声道; MMSE 对小波系数进行最小均方误差估计.

图 2 为带噪信号经过小波去噪处理和新算法处理过的信号对比. 由图可见, 新算法中的信号幅度比小波去噪算法处理过信号幅度有所减小, 噪声部分幅度变得更为平滑, 对噪声有较好的抑制作用.

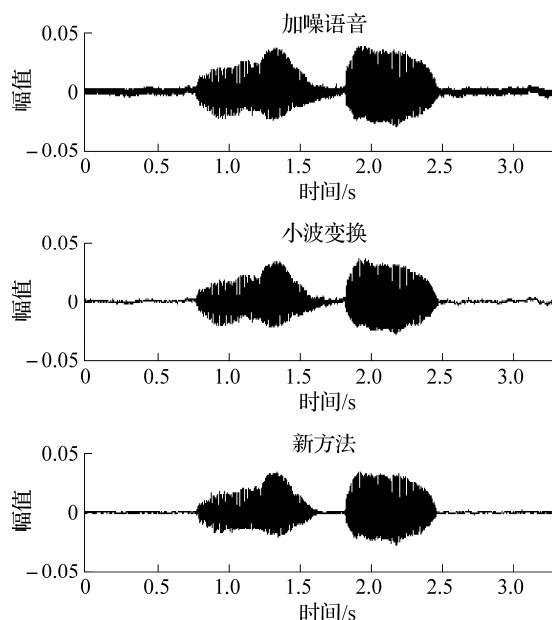


图 2 经过小波去噪和新算法处理过的信号波形

图 3~图 5 为去噪前语音信号小波系数、小波去噪后小波系数和新算法去噪后小波系数的图形. $cd1 \sim cd5$ 分别对应滤波前第 1 层高频系数到第 5 层高频系数, 其频率范围对应 8 000~16 000 Hz, 4 000~

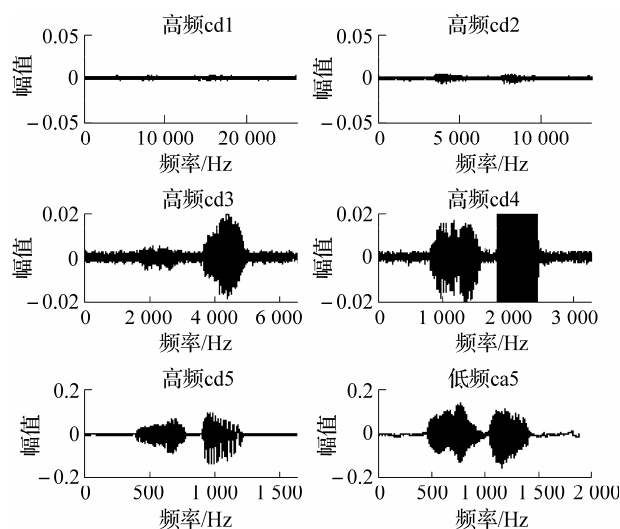


图 3 去噪前的小波系数

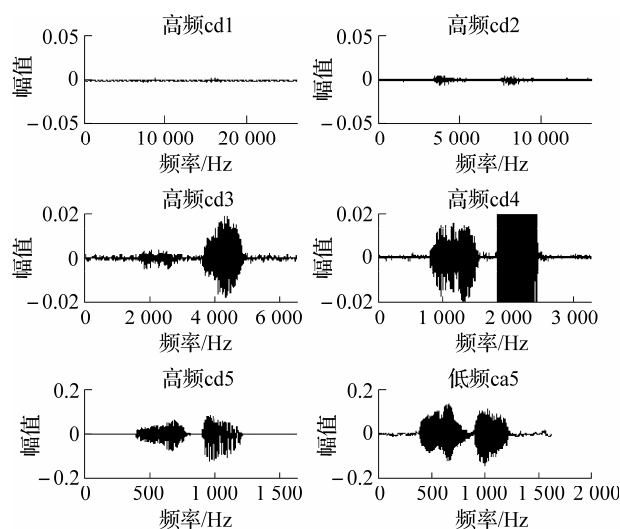


图 4 小波去噪后的小波系数

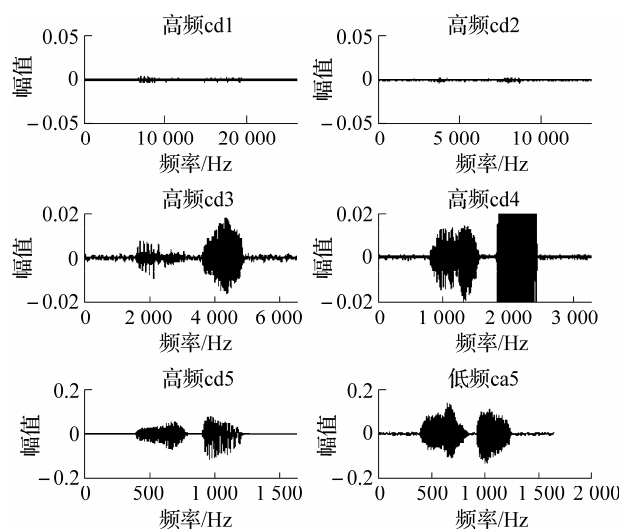


图 5 新算法去噪后的小波系数

8 000 Hz, 2 000~4 000 Hz, 1 000~2 000 Hz, 500~1 000 Hz, ca_5 为第 5 层低频系数, 频率范围为 0~500 Hz. $cd_1 \sim cd_5$ 分别对应滤波后第 1 层高频系数到第 5 层高频系数. 对比不同方法滤波的图形, 可见利用新算法滤波后的小波系数噪声明显减少, 这使得小波重构的信号对噪声有较好的抑制作用.

文中语音采集通过 Cooledit 处理软件完成, 采集 1~10 这 10 个数字的英语发音作为实验对象. 采用 16 000 Hz 的采样频率, 单声道录音, 精度为 16 位, 通过对 24 个人录音, 得到 240 个语音数据, 其中 120 个数据用来训练语音识别模型, 另外 120 个数据用来进行语音识别测试. 提取每帧信号的 24 阶 MFCC 作为语音特征参数, 噪声为高斯白噪声.

通过语音识别可以得到加噪语音、MMSE 去噪、小波变换去噪以及新方法去噪后的语音识别率, 从表 1 数据中可以得出改进方法的去噪效果较其他 2 种方法有一定的提高.

表 1 3 种方法去噪以后的识别率比较

信噪比/ db	未去噪 语音/%	MMSE 去噪/%	小波去 噪/%	改进方法 去噪/%
5	26.7	30.0	36.7	40.0
15	36.7	40.0	44.2	48.3
25	43.3	45.8	48.3	51.7

5 结论

MMSE 相比于谱减法等传统方法更多地利用

了语音和噪声统计特性的先验知识, 但前提是信号短时平稳, 这大大限制了该方法的应用. 而小波变换在非平稳和低信噪比的情况下具有较好的语音增强能力. 提出的新方法结合了 MMSE 方法和小波变换的各自优点, 实验结果表明新方法处理后得到的信号损害较小, 而且去噪效果较好, 从而提高了语音识别系统的性能, 相比于传统的去噪方法有一定的提高. 但其算法复杂度较高, 并且对波形会造成一些损害, 有待于今后继续深入研究.

参考文献:

- [1] 宁更新. 抗噪声语音识别新技术的研究[D]. 广州: 华南理工大学, 2006.
- [2] 熊燕. 抗噪声语音识别技术研究[J]. 信息科技及现代服务, 2006(7):204-205.
- [3] 方瑜. 语音增强相关问题研究[D]. 北京: 北京邮电大学, 2011.
- [4] 丁沛. 语音识别中的抗噪声技术[D]. 北京: 清华大学, 2003.
- [5] 王苏敏, 谢小云, 邓茜. 基于小波去噪的语音识别系统[J]. 数字技术与应用, 2012(5):232.
- [6] 毛艳辉. 小波去噪在语音识别预处理中的应用[D]. 上海: 上海交通大学, 2010.
- [7] 胡惠英, 吴善培. 小波去噪在语音识别中的应用[J]. 北京邮电大学学报, 1999, 22(3):31-34.
- [8] 崔晓, 张松炜. 基于小波和先验信噪比维纳滤波的语音增强[J]. 河南师范大学学报(自然科学版), 2013, 41(1):43-46.

Voice Enhancement Based on the MMSE with Wavelet Transformation

PAN Xiao-long, ZHANG Wei-qiang*, LI Yuan-hong

(Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: A new algorithm combining MMSE with wavelet transformation is put forward to improve the performance of speech recognition. Firstly, the noisy speech is processed with wavelet transformation by multi-scale decomposition, then the multi-scale wavelet coefficient is determined with MMSE algorithm. Finally the wavelet coefficient is reconstructed to enhance the speech quality. By comparing the presented algorithm with the traditional ones, as indicated by the result of MATLAB experiment, the developed algorithm is found to be of better performance in terms of speech enhancement.

Key words: MMSE; wavelet transformation; speech enhancement

(责任编辑 章践立)