

结合 EEMD 与 K -SVD 字典训练的语音增强算法

甘振业, 陈 浩, 杨鸿武

(西北师范大学 物理与电子工程学院, 兰州 730070)

摘 要: 该文提出一种总体平均经验模态分解(ensemble empirical mode decomposition, EEMD)方法与 K 奇异值分解(K -singular value decomposition, K -SVD)字典算法相结合的语音增强算法。将带噪语音通过 EEMD 分解得到各本征模式分量(intrinsic mode function, IMF), 对各 IMF 分量进行互相关和自相关分析, 去除噪声 IMF 分量, 并将过渡 IMF 分量再次进行 EEMD 分解, 去除其中的噪声 IMF 分量。将过渡 IMF 分量和剩余的 IMF 分量叠加, 得到预降噪的带噪语音。利用纯净语音, 通过 K -SVD 字典训练算法得到过完备字典。对预降噪的带噪语音通过过完备字典进行稀疏表示, 稀疏系数重构出纯净语音。实验结果表明: 在低信噪比和高信噪比情况下, 该算法的去噪效果明显优于传统的谱减法、小波阈值去噪法和 K -SVD 字典训练。

关键词: 语音增强; 总体平均经验模态分解; K 奇异值分解; 相关性

中图分类号: TN912.35

文献标志码: A

文章编号: 1000-0054(2017)03-0286-07

DOI: 10.16511/j.cnki.qhdxxb.2017.26.011

Speech enhancement algorithm that combines EEMD and K -SVD dictionary training

GAN Zhenye, CHEN Hao, YANG Hongwu

(College of Physics and Electronic Engineering,
Northwest Normal University, Lanzhou 730070, China)

Abstract: This paper presents a speech enhancement algorithm that combines the ensemble empirical mode decomposition (EEMD) algorithm and the K -singular value decomposition (K -SVD) dictionary-training algorithm. The EEMD algorithm is used to obtain the intrinsic mode function (IMF) components from noisy speech. The cross-correlations and autocorrelations of each IMF are calculated from the IMF components to filter out the noisy IMF components. The transition IMF components are again decomposed with EEMD to further remove the noisy component. The remained IMFs and transition IMFs are superimposed to generate the de-noised speech. An over-complete dictionary is then trained from the clean speech by the K -SVD dictionary training algorithm. The de-noised speech is then sparse decomposed with the over-complete dictionary to obtain the enhanced speech by recovering the speech signal from sparse coefficient vectors. Tests show that the algorithm

achieves better de-noising than the traditional spectral subtraction, wavelet threshold de-noising and K -SVD dictionary-training algorithms for both low signal-to-noise ratio (SNR) and high SNR environments.

Key words: speech enhancement; ensemble empirical mode decomposition (EEMD); K -singular value decomposition (K -SVD); correlation

语音是人类最方便的通信方式, 在处理语音信号时, 通常采用纯净语音信号。实际采集的语音信号中含有一定的噪声, 影响到语音的可懂度及音质。因此, 有效去除带噪语音信号中的噪声成分具有非常重要的理论意义和实用价值。语音增强是解决噪声污染语音信号的重要方法。

语音增强算法现已比较成熟。传统语音增强算法主要有 Wiener 滤波法^[1]、谱减法^[2]、小波变换^[3]及这些算法的改进算法。现代语音增强算法主要有主成分分析^[4]、非负矩阵分解、压缩感知理论及这些算法的改进算法等。其中, 比较具有代表性的是近年来兴起的压缩感知理论^[5-6]。文^[7]中提出的 K 奇异值分解(K -singular value decomposition, K -SVD)字典训练算法属于基本压缩感知理论的改进算法, 该算法通过为信号构造过完备冗余字典, 能够更好地在自适应的稀疏基上稀疏表示语音信号, 从而更加精确地重构纯净语音信号。但 K -SVD 字典训练算法在低信噪比的情况下, 却不能较好地稀疏表示带噪语音信号。文^[8]提出了经验模态分解(empirical mode decomposition, EMD), 可根据信

收稿日期: 2016-06-23

基金项目: 国家自然科学基金资助项目

(61263036, 61262055, 11664036);

甘肃省杰出青年基金资助项目(1210RJDA007)

作者简介: 甘振业(1977—), 男, 副教授。

通信作者: 杨鸿武, 教授, E-mail: yanghw@nwnu.edu.cn

号的局部时变特征进行自适应时频分解,非常适合对非平稳信号的分析,但 EMD 存在模态混叠问题。为解决此问题,文[9]提出了总体平均经验模态分解(ensemble empirical mode decomposition, EEMD),向待分解的信号中加入 Gauss 白噪声来平滑脉冲干扰等异常事件,解决模态混叠问题。带噪语音通过 EEMD 分解可得到高频噪声本征模式分量(intrinsic mode function, IMF),但其余的 IMF 分量中仍含有少量的噪声成分。结合上述算法,通过预降噪提高语音信号的稀疏性,从而准确重构纯净语音,是有效去除噪声的关键。

本文针对传统语音增强算法的缺陷及 K-SVD 算法在低信噪比的情况下不能较好地稀疏表示带噪语音信号的问题,提出一种 EEMD 与 K-SVD 字典训练相结合的语音增强算法。该算法利用 EEMD 分解去除噪声 IMF 分量,并利用 EEMD 进一步分解过渡 IMF 分量^[10],去除其中的噪声成分。去噪后的过渡 IMF 分量与剩余的 IMF 分量叠加得到新的带噪语音。采用纯净语音信号通过 K-SVD 训练过完备字典,利用正交匹配追踪算法(orthogonal matching pursuit, OMP)^[10]将新的带噪语音在过完备字典上进行稀疏分解,通过稀疏系数重构出去噪后的语音,达到语音增强的目的。

1 EEMD 算法

EEMD 算法^[9]是 EMD 算法的改进算法,主要解决 EMD 算法中产生模态混叠的问题。EEMD 算法利用了 Gauss 白噪声具有频率均匀分布的统计特性。当信号加入白噪声后,使得信号在不同尺度上具有连续性,从而减小模态混叠的程度。EEMD 算法的分解步骤如下:

步骤 1 在原始带噪语音信号 $x(t)$ 中加入 100 次标准差为原始信号标准差 0.1~0.4 倍的随机正态分布 Gauss 白噪声。

步骤 2 对步骤 1 得到的信号进行 EMD 分解。

步骤 3 重复步骤 1 和步骤 2 N 次。每次在信号中加入新的随机正态分布 Gauss 白噪声序列。

步骤 4 求出所有迭代所获得的 IMF 的平均值,作为分析信号的 IMF 组合。

最终, EEMD 分解的结果为

$$x(t) = \sum_j c_j(t) + r(t), \quad (1)$$

其中: $c_j(t)$ ($j=1, 2, \dots, N$) 代表从高频到低频的 IMF 分量; $r(t)$ 为残余分量,代表信号的平均趋势。

2 K-SVD 字典训练算法

K-SVD^[7]算法又称为广义 K -均值算法,该算法能够在获取字典的同时获得稀疏系数矩阵,主要包括稀疏编码和字典更新 2 个步骤:

步骤 1 字典 D 已知,求解信号的稀疏系数;

步骤 2 根据得到的稀疏系数,更新字典中的原子,常用的方法有 MOD(method of optimal directions)^[11]和 K-SVD,二者在字典更新时的原子更新机制不同:MOD 是对字典中所有的原子进行更新,而 K-SVD 只对某个原子进行更新。

假设要进行稀疏表示的纯净语音为 Y ,稀疏表示使用的字典为 D ,相应的系数矩阵为 M 。K-SVD 的目标函数为

$$\min_{D, M} \sum_i \|m_i\|_0 \\ \text{s. t. } \|Y - DM\|_F^2 \leq \epsilon. \quad (2)$$

其中: $\|\cdot\|_F$ 表示 Frobenius 范数, $\|\cdot\|_0$ 表示 0 范数, ϵ 表示逼近残差。

过完备字典的训练步骤如下:

步骤 1 挑选一定数目的训练样本作为初始字典 D 。

步骤 2 稀疏编码,这个阶段采用追踪算法进行计算。常用的追踪算法有 BP、MP、OMP 等,这些算法可以对样本信号进行稀疏表示,计算相应的系数。

步骤 3 将字典中的原子逐个进行更新,假设 E_k 为去掉第 K 个原子以后的稀疏表示误差:

$$E_k = Y - \sum_{j \neq k} d_j m_j. \quad (3)$$

其中 m_j 表示矩阵 M 的列向量。

步骤 4 定义 $\omega_k = \{i | 1 \leq i \leq K, m_T^k(i) \neq 0\}$ 表示用到原子 d_k 的样本 Y 的索引,以防止结果发散。定义 Ω_k 为 $N \times \omega_k$ 的矩阵,其中 $(\omega_k(i), i)$ 是 1,其他都是 0,则式(3)等价于

$$\|E_k \Omega_k - d_k m_T^k \Omega_k\|_F^2 = \|E_k^R - d_k m_R^k\|_F^2. \quad (4)$$

其中: m_T^k 表示矩阵 m 的第 k 行, $m_R^k = m_T^k \Omega_k$ 表示 m_T^k 去掉 0 值项的行矢量, $E_k^R = E_k \Omega_k$ 表示样值稀疏分解用到原子 d_k 的误差列。

步骤 5 将 E_k^R 通过 SVD 分解得到 $E_k^R = U V^T$ 。分解的 U 的第一列更新原来字典中的 d_k , 矩阵 V 的第一列与 $\Delta(1, 1)$ 的乘积更新系数向量 m_j 。

重复步骤 2 到步骤 5,直到满足终止条件,即可得到纯净语音 Y 的过完备字典。

3 EEMD 与 K-SVD 结合的语音增强算法

3.1 算法的基本思想

EEMD 和 K-SVD 字典训练算法在语音增强方面各有优劣。带噪语音信号经过 EEMD 分解之后,可以得到几乎为纯噪声的 IMF 分量(第一 IMF 分量),去除该噪声 IMF 分量对恢复原始语音信号几乎没有损失。其余 IMF 分量仍含有少量的噪声,但此时带噪语音信号稀疏性增强,能更好地通过过完备字典稀疏表示,通过稀疏系数重构得到的纯净语音信号更加精确。算法的流程如图 1 所示。

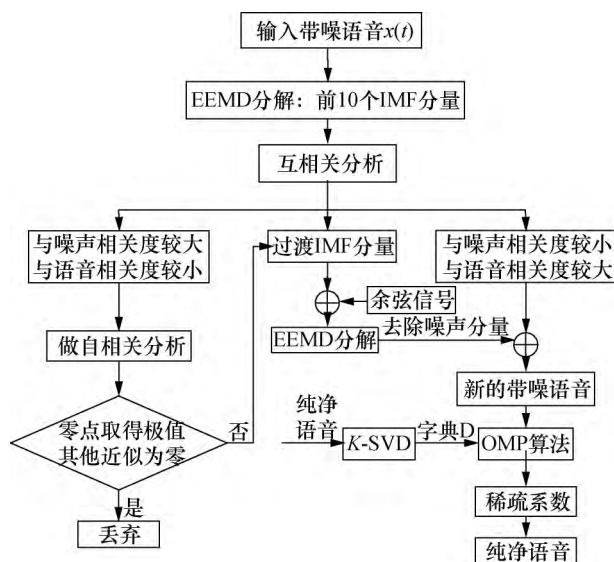


图 1 算法流程

3.2 增强算法的基本过程

带噪语音信号通过 EEMD 分解得到各 IMF 分量(这里取前 10 层的 IMF 分量,10 层之后的分量对信号重构影响极小,故此忽略)。对各 IMF 分量做互相关分析,与噪声相关性强而与语音相关性弱的 IMF 分量有可能为噪声 IMF 分量,对该分量进行自相关分析,若该 IMF 分量的自相关函数在零点取得极值而在其他各点为零或近似为零,则可判定该 IMF 分量为噪声分量,直接舍弃。

通常情况下,第二或第三 IMF 分量既含有大量的语音信息,又包含大量噪声,这种 IMF 分量称为过渡模态分量^[12]。文[12]指出,将过渡 IMF 的第一次 EMD 分解处理得到的结果与一个合适的余弦信号相加,并对相加得到的混合信号再次进行 EMD 处理,就可以有效去除过渡 IMF 中的混叠噪声。因此,对过渡模态分量再次进行 EEMD 分解,去除几乎为纯噪声的第一 IMF 分量,其余 IMF 分

量组成新的 IMF 分量,然后将去除噪声后的所有 IMF 分量叠加得到新的带噪语音。

$$\min_x \|x\|_0$$

$$\text{s. t. } \|y - Dx\|_2 \leq \epsilon. \quad (5)$$

其中: y 表示要进行稀疏表示的新的带噪语音信号, D 表示 K-SVD 训练得到的过完备字典, x 表示 y 在 D 上分解得到的稀疏系数, ϵ 表示稀疏表示的语音信号和带噪语音信号之间的残差。

可以采用 OMP 算法对式(5)进行求解,得到新的带噪语音信号的过完备稀疏表,过程如下:

步骤 1 初始化,残差 $r(0) = y$, 原子集合矩阵 $\Lambda_0 = \phi$, 字典为 D , 迭代序号 $k = 1$ 。

步骤 2 计算残差 $r(i)$ 与字典 D 的内积,记录内积最大值位置对应的原子 d_j 下标 $\lambda_i = \arg \max_{j=1,2,\dots,N} |\langle r(i), d_j \rangle|$, $k = k + 1$ 。

步骤 3 扩充原子集合矩阵 $\Lambda_i = [\Lambda_{i-1}, \phi_{\lambda_i}]$ 。

步骤 4 稀疏系数 $x_i = \arg \min \| \Lambda_{i-1}, d_{\lambda_i} \|_2$, 计算信号估计值 $y_i = \Lambda_i x_i$, 更新残差 $r(i) = y - y_i$ 。满足 $\|r(i)\|_2 \leq \epsilon$, 转至步骤 2 继续迭代,否则终止。

通过上述步骤得到的最优解即为 \hat{x} , 若有 n 个非零系数,则这 n 个非零系数与相应的原子相乘就可以得到增强之后的语音信号 y

$$y = D\hat{x} = \sum_{j=1}^n x_j d_i. \quad (6)$$

对于 ϵ 的设置,文[13]指出,对于当已知噪声方差为 σ^2 的情况下,在进行原子选择时通过设定与噪声方差成比例的恰当的阈值 ϵ ,就可以从带噪信号中选出与纯净语音相关的系数,舍弃与噪声有关的系数。实验中设置

$$\epsilon = k\sigma^2. \quad (7)$$

实验中 k 取值通常在 1.05~1.15 之间,本文设定 k 取值 1.08,能够得到最好的重构效果。

4 结果与分析

为了验证本文算法在语音增强方面的有效性,对不同信噪比的带噪语音分别采用本文算法、谱减法、小波阈值去噪及 K-SVD 算法进行去噪处理,对比分析实验结果。

实验语音样本采用在专业录音棚内录制的标准普通话。单字和词语的录音人为 8 人(4 男 4 女),每个人分别录制 8 个单字和 8 个词语的语音样本;语句的录音人为 4 人(2 男 2 女),每人分别录制 4 个语句的语音样本。语音信号以单声道、16 kHz 采样率、16 位量化精度保存为 WAVE 文件。对纯净

语音信号,分别加入白噪声、车内噪声、餐厅内噪声、粉红噪声和工厂车间噪声,加入的白噪声及有色噪声均来自 NOIZEX-92 噪声库。对所有纯净语音样本分别加入不同强度的白噪声,指定输出带噪语音信噪比分别为-10、-5、0、5、10 dB,并计算其分段信噪比(segmental signal-to-noise ratio, SegSNR):

$$\text{SegSNR} = \frac{10}{L} \sum_{l=0}^{L-1} \log_{10} \left(\frac{\sum_{m=0}^{M-1} x_l^2(m)}{\sum_{m=0}^{M-1} (s_l(m) - x_l(m))^2} \right). \quad (8)$$

其中: $s_l(m)$ 表示纯净语音, $x_l(m)$ 表示带噪语音, M 表示帧长, L 表示总帧数。

对加入噪声的语音进行 EEMD 分解,并求解出各 IMF 分量与纯净语音和噪声的互相关系数。图 2 和 3 分别表示纯净语音“灯”加入不同强度噪声、指定输出信噪比分别为-5 和 5 dB 时,各 IMF 分量与纯净语音以及噪声的互相关系数。

可以看出,无论是在高信噪比还是在低信噪比情况下,第 1 个 IMF 分量(记为 IMF1)与噪声的相关度较大,而与纯净语音的相关度较小。第 6—10 IMF 分量与纯净语音及噪声的相关度都很小,可以忽略不计。

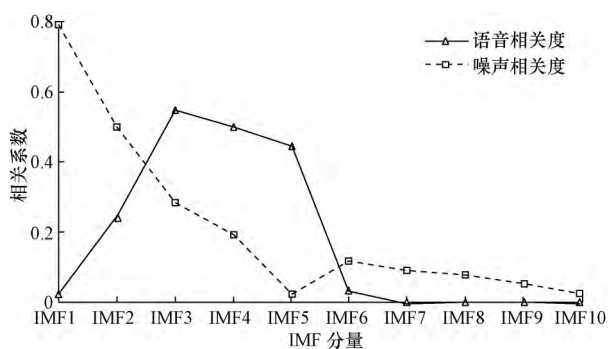


图 2 加噪输出信噪比为-5 dB 时各 IMF 分量的相关系数

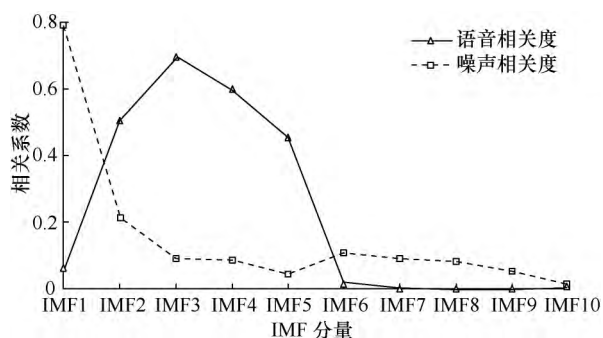


图 3 加噪输出信噪比 5 dB 时各 IMF 分量的相关系数

对 2 种信噪比情况下 EEMD 分解的 IMF1 分量做自相关分析,如图 4 和 5 所示。

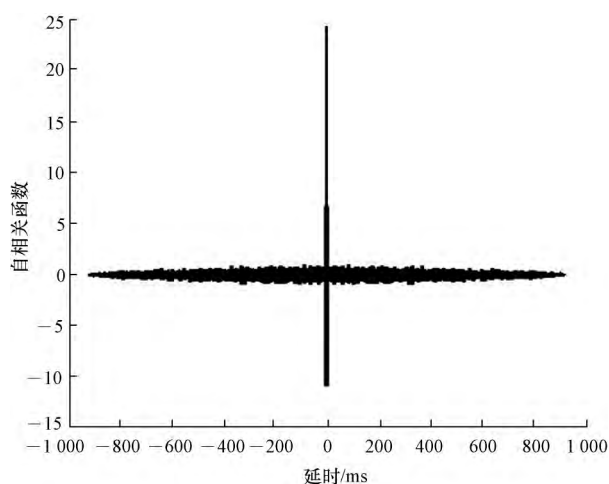


图 4 加噪输出信噪比-5 dB 时 IMF1 的自相关性分析

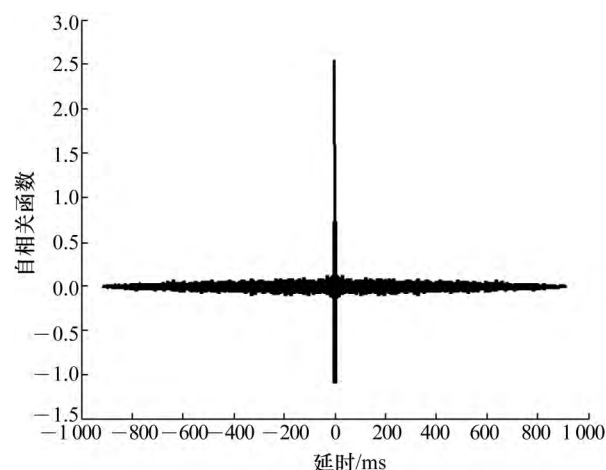


图 5 加噪输出信噪比 5 dB 时 IMF1 的自相关性分析

可以看出,2 种信噪比情况下 EEMD 分解的 IMF1 分量在零点处取得极大值,而其余各点基本接近于零。可以判定这 2 种信噪比情况下 EEMD 分解的 IMF1 分量均为噪声 IMF 分量,可直接舍弃。IMF2 分量与噪声及语音的相关度几乎相当,说明 IMF2 分量中既包含大量语音,同时也含有大量噪声。因此,对 IMF2 分量再次进行 EEMD 分解,去除第 1 噪声 IMF 分量,其余 IMF 分量重新组合。将新组合的 IMF2 分量同 IMF3、IMF4 和 IMF5 分量组成新的带噪语音。利用纯净语音,通过 K-SVD 字典训练得到过完备字典。新的带噪语音通过过完备字典进行稀疏表示,通过稀疏系数重构得到增强后的语音。

为了验证本文方法的降噪性能,将本文算法分别与谱减法、小波阈值去噪和 K-SVD 字典训练算

法的降噪结果进行对比。图 6 为不同信噪比情况下各算法对带噪语音增强后的平均分段信噪比。图 7 为不同信噪比情况下 4 种算法增强处理之后的平均感知语音质量评价 (perceptual evaluation of speech quality, PESQ), 图 8 和 9 分别给出了加入白噪声和色噪声, 信噪比为 -5 dB 下 4 种算法增强后的波形。从中可以看出, 本文算法要优于其他几种算法。

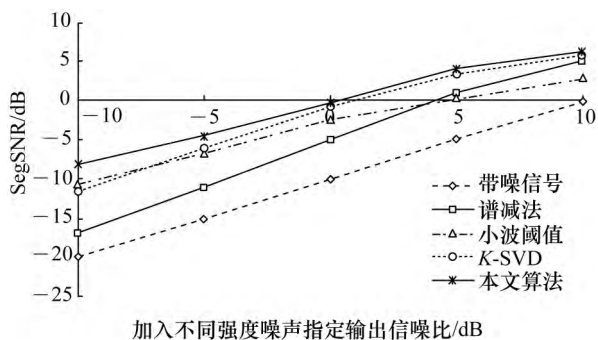


图 6 各算法不同信噪比下增强处理后的信噪比

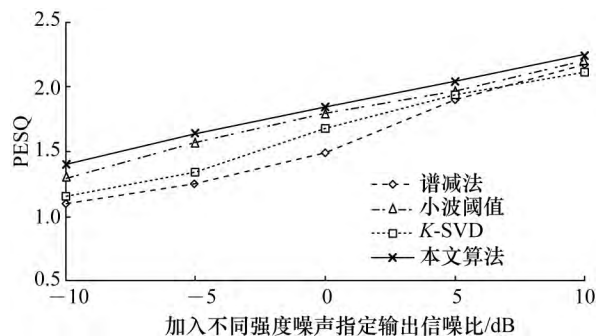


图 7 各算法不同信噪比下增强处理后的 PESQ

为了进一步验证本文算法的降噪性能, 采用汉语清晰度诊断押韵测试 (diagnostic rhyme test method of Chinese articulation, DRTC)^[14] 对降噪后的语音进行了主观评测。实验采用在加入不同强度的白噪声, 并指定输出带噪语音的信噪比为 -5 、 0 、 5 dB 下的 10 对韵字语音, 听音人为 5 男 5 女听力正常的研究生。DRT 实验结果如表 1 所示, 可以看出, 本文提出的算法总体上优于传统的谱减法、小波阈值去噪法和 K-SVD 算法。

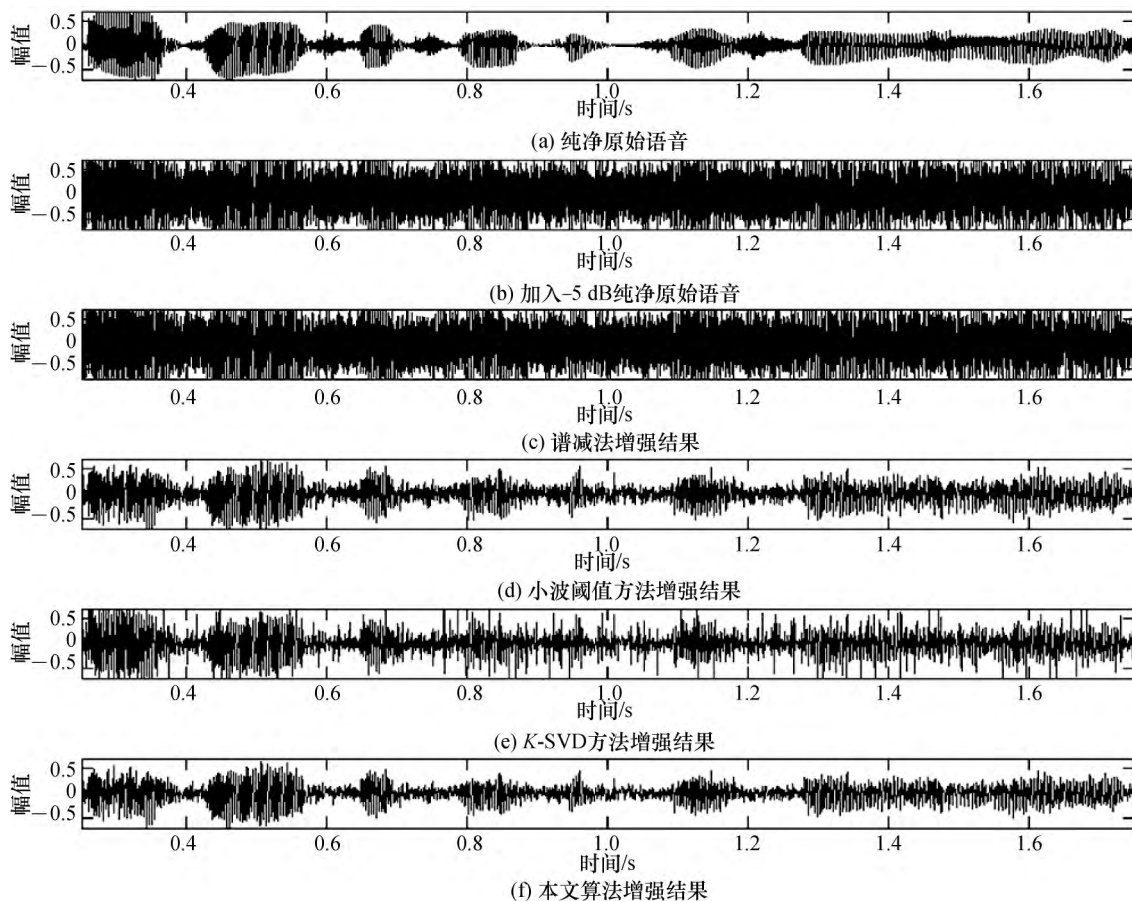


图 8 加入白噪声 -5 dB 时 4 种算法增强后的波形

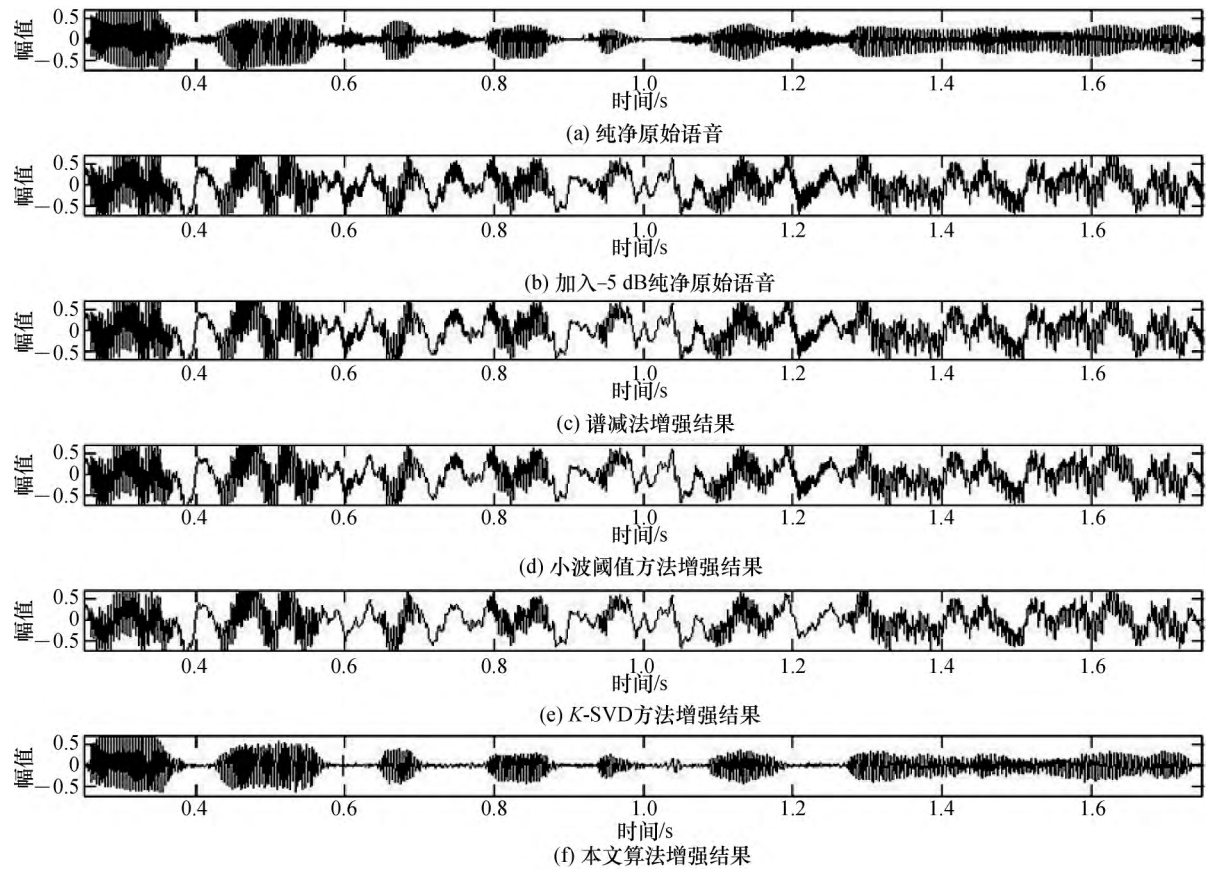


图 9 加入色噪声-5 dB 时 4 种算法增强后的波形

表 1 加入不同噪声时 4 种算法的 DRT 评分

SNR/dB	DRT 评分			
	谱减法	小波阈值	K-SVD	本文算法
-5	50%	64%	54%	70%
0	62%	70%	66%	86%
5	76%	76%	82%	92%

5 结 论

本文提出了一种 EEMD 与 K-SVD 字典训练结合的语音增强算法。首先对带噪声语音进行 EEMD 分解,得到的各 IMF 分量分别与纯净语音和噪声做互相关分析,对疑似噪声 IMF 分量做自相关分析,从而去除噪声 IMF 分量。同时,将过渡 IMF 分量再次进行 EEMD 分解,去除噪声成分。将去噪后的过渡 IMF 分量与剩余 IMF 分量组合成新的带噪声语音,并在采用纯净语音训练得到的过完备字典上进行稀疏表示,通过稀疏系数重构得到增强之后的语音信号。不同信噪比情况下各算法对带噪声语音的去噪对比实验结果表明,本文提出的算法优于其他几种算法。

参考文献 (References)

[1] Chen J, Benesty J, Huang Y, et al. New insights into the noise reduction Wiener filter [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, **14**(4): 1218-1234.

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

[3] Donoho D L. De-noising by soft-thresholding [J]. *IEEE Transactions on Information Theory*, 1995, **41**(3): 613-627.

[4] 王文波, 张晓东, 汪祥莉. 基于主成分分析的经验模态分解消噪方法 [J]. *电子学报*, 2013, **41**(7): 1425-1430. WANG Wenbo, ZHANG Xiaodong, WANG Xiangli. Empirical mode decomposition de-noising method based on principal component analysis [J]. *Chinese Journal of Electronics*, 2013, **41**(7): 1425-1430. (in Chinese)

[5] 石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展 [J]. *电子学报*, 2009, **37**(5): 1070-1081. SHI Guangming, LIU Danhua, GAO Dahua, et al. Advances in theory and application of compressed sensing [J]. *Chinese Journal of Electronics*, 2009, **37**(5): 1070-1081. (in Chinese)

[6] Donoho D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289-1306.

- [7] Michal A, Michael E, Alfred B. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing*, 2006, **54**(11): 4311-4322.
- [8] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[C]// Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. London, UK: Royal Society, 1998: 903-995.
- [9] Wu Z, Huang N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method [J]. *Advances in Adaptive Data Analysis*, 2009, **1**(1): 1-41.
- [10] Karahanoglu N B, Erdogan H. A* Orthogonal matching pursuit: Best-first search for compressed sensing signal recovery [J]. *Digital Signal Processing*, 2012, **22**(4): 555-568.
- [11] Engan K, Aase S O, Husøy J H. Multi-frame compression: Theory and design [J]. *Signal Processing*, 2000, **80**(10): 2121-2140.
- [12] 李月, 彭蛟龙, 马海涛, 等. 过渡内蕴模态函数对经验模态分解去噪结果的影响研究及改进算法 [J]. *地球物理学报*, 2013, **56**(2): 626-634.
- LI Yue, PENG Jiaolong, MA Haitao, et al. Study of the influence of transition IMF on EMD de-noising and improved algorithm [J]. *Chinese Journal of Geophysics*, 2013, **56**(2): 626-634. (in Chinese)
- [13] Donoho D L, Elad M, Temlyakov V N. Stable recovery of sparse overcomplete representations in the presence of noise [J]. *IEEE Transactions on Information Theory*, 2006, **52**(1): 6-18.
- [14] Li Z, Tan E C, McLoughlin I, et al. Proposal of standards for intelligibility tests of Chinese speech [J]. *IEEE Proceedings-Vision, Image and Signal Processing*, 2000, **147**(3): 254-260.

(上接第 285 页)

- [7] Havenith G, Richards M, Wang X, et al. Apparent latent heat of evaporation from clothing: Attenuation and "heat pipe" effects [J]. *Journal of Applied Physiology*, 2008, **104**(1): 142-149.
- [8] Oliveira A, Gaspar A, Quintela D. Measurements of clothing insulation with a thermal manikin operating under the thermal comfort regulation mode: Comparative analysis of the calculation methods [J]. *European Journal of Applied Physiology*, 2008, **104**(4): 679-688.
- [9] Celcar D, Meinander H, Gersak J. Heat and moisture transmission properties of clothing systems evaluated by using a sweating thermal manikin under different environmental conditions [J]. *International Journal of Clothing Science and Technology*, 2008, **20**(4): 240-252.
- [10] ISO 9920. Ergonomics of the thermal environment- Estimation of thermal insulation and water vapour resistance of a clothing ensemble [S]. Geneva: International Organization for Standardization, 2009.
- [11] ASTM F 1291. Standard test method for measuring the thermal insulation of clothing using a heated manikin [S]. West Conshohocken: American Society for Testing and Materials, 2010.
- [12] Howie R. Assessment of the scientific validity of ISO 7933/EN 12515 [C]// 1st European Conference on Protective Clothing. Stockholm, Sweden: European Society on Protective Clothing, 2000: 163-166.
- [13] ASTM F 2370. Standard test method for measuring the evaporative resistance of clothing using a sweating manikin [S]. West Conshohocken: American Society for Testing and Materials, 2010.
- [14] Wang F M, Gao C, Kuklane K, et al. A study on evaporative resistances of two skins designed for thermal manikin tore under different environmental conditions [J]. *Journal of Fiber Bioengineering and Informatics*, 2009, **1**(4): 211-215.
- [15] Song G, Paskaluk S, Sati R, Crown E, et al. Thermal protective performance of protective clothing used for low radiant heat protection [J]. *Textile Research Journal*, 2011, **81**(3): 311-323.