

融合统计模型与 EMD 的宽带话音增强方法

周璇, 鲍长春, 夏丙寅

(北京工业大学 电子信息与控制工程学院 语音与音频信号处理研究室, 北京 100124)

摘要: 提出了一种融合统计模型和经验模态分解 (EMD) 的宽带话音增强方法。该方法首先用统计模型增强算法消除含噪语音中的主要噪声成分, 然后用一种基于活动话音检测 (VAD) 的 EMD 增强算法做后处理进一步抑制残留噪声, 从而使以上 2 种方法的优点有效地结合。在 ITU-T G.160 标准下对算法进行了性能测试, 测试结果表明, 与经典的统计模型方法相比, 在不同强度的背景噪声下, 增强话音的信噪比提高都较为明显。同时, 在低信噪比情况下, 该方法能有效抑制增强话音高频部分的音乐噪声, 提高了听觉舒适度。

关键词: 宽带话音增强; 统计模型; EMD; VAD

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-436X(2013)08-0095-07

Combined wideband speech enhancement method based on statistical model and EMD

ZHOU Xuan, BAO Chang-chun, XIA Bing-yin

(Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: A combined wideband speech enhancement method based on statistical model and empirical mode decomposition (EMD) was proposed. First, statistical model was used to eliminate the main noise component in noisy speech. Then, the residual noise was further suppressed by a post-processing module which is a speech enhancement algorithm with voice activity detection (VAD) based on EMD. The advantages of the two methods were combined effectively. The performance of the proposed method was evaluated under the standard of ITU-T G.160. The experimental results indicate that the algorithm is more effective for improving the SNR in the different noise environments than classical statistical model approach. Meanwhile, in low SNR conditions, musical noise is reduced effectively, and the speech sounds more comfortable.

Key words: wideband speech enhancement; statistical model; EMD; VAD

1 引言

便利的话音通信系统已经成为现代人生活和工作中的必需品, 而现实生活里, 传输于通信系统中的话音信号不可避免地要受到各种背景噪声的污染, 导致通信系统不能为接受端提供高质量的话音。因此, 话音增强技术便孕育而生了, 且一直是话音信号处理领域的研究热点。话音增强的目的就是在尽可能不损失话音可懂度的前提下, 有效地抑制含噪语音中的噪声成分, 以改善话音的主

客观质量。

经过近 40 年的发展, 研究者们提出了多种话音增强算法, 其中, 较为经典的有维纳滤波方法^[1]、谱减法^[2]、基于子空间的方法^[3]、基于小波变换的方法^[4]、基于统计模型的方法^[5]及近几年发展较快的基于经验模式分解 (EMD, empirical mode decomposition) 的方法^[6]。以上方法的提出大多是针对窄带语音信号, 即信号带宽限制在 300 Hz ~ 3.4 kHz 之间。然而这种信号与宽带语音信号相比, 其质量已经不能满足人们的需求。在 3G 网络时

收稿日期: 2012-08-27; 修回日期: 2013-04-03

基金项目: 北京市教育委员会科技发展计划重点基金资助项目 (KZ201110005005)

Foundation Item: Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education (KZ201110005005)

代, 宽带语音信号将成为语音通信系统中的主流, 而目前专门针对宽带语音信号增强方法的研究并不是很多。近年, 利用频带扩展技术进行宽带语音增强的方法^[7]被提出, 该算法在低信噪比环境下能够取得较好的增强效果, 但其高频语音具有较明显的失真现象发生。

在传统方法中, 基于统计模型的方法和基于 EMD 的方法均能取得一定的增强效果。其中, 基于统计模型的方法由于较好地利用了语音和噪声信号的先验特性, 因此通常可以获得较好的算法性能。但该方法得到的增强语音在高频部分往往残留较明显的音乐噪声, 影响了主观听觉感受。而基于 EMD 的方法能够有效地将高频部分、语音能量较弱处的噪声从语音信号中分离出来, 但语音与噪声混叠的部分其增强效果不够明显, 导致最终的增强语音听觉感受也不够理想。

为了更好地结合基于统计模型和基于 EMD 的 2 种语音增强方法的优点, 并弥补它们各自的不足, 本文提出了一种基于语音激活检测(VAD, voice activity detection)的 EMD 增强方法, 并将其与统计模型方法进行融合得到了一种整体增强方案。该方案通过了国际电信联盟电信标准化部门(ITU-T)于 2008 年公布的针对移动网络中语音增强设备的评测标准 G.160^[8]的各项相关测试。在与较为经典的统计模型方法的对比中, 测试结果显示, 本文方法去噪性能优于统计模型方法, 且主观听觉质量得到了明显的提升。

2 统计模型与 EMD 简介

2.1 统计模型语音增强算法简介

基于统计模型的语音增强算法的基本原理是: 假设语音信号与噪声的傅里叶变换系数服从一定的统计模型, 在一定的误差准则下可以得到不同的增强语音幅度谱估计器, 目前大多采用贝叶斯法估计增强语音的幅度谱。

根据所选代价函数的不同, 贝叶斯估计器主要有最小均方误差估计器(MMSE, minimum mean-squared error)^[5]、最大后验概率估计器(MAP, maximum a posteriori)^[9]、板仓一斋藤测度估计器(IS, Itakura-Saito)^[10]、COSH 估计器^[10]和加权欧式失真测度估计器(WEDM, weighted euclidean distortion measure)^[10]5 种。

本文采用的是性能相对较好的基于加权欧式失真测度的贝叶斯幅度谱估计器。这是一种引入了感知加权效应的估计器, 其代价函数为

$$d(X_k, \hat{X}_k) = \frac{(X_k - \hat{X}_k)^2}{X_k} \quad (1)$$

其中, \hat{X}_k 是频点 k 处的幅度谱估计值, X_k 是纯净语音的幅度谱。在估计误差相同的情况下, 当 X_k 较小时, 即谱谷处, 代价函数较大; 而当 X_k 较大时, 即谱峰处, 代价函数较小, 这样在一定程度上反映了听觉掩蔽效应的作用。由上述代价函数可以构成贝叶斯风险函数

$$\mathcal{R} = \int_0^{\infty} \left[\frac{(X_k - \hat{X}_k)^2}{X_k} \right] p(X_k | Y_k) dX_k \quad (2)$$

其中, $p(X_k | Y_k)$ 为 X_k 的后验概率密度函数, Y_k 为含噪语音的幅度谱值。令风险函数对 \hat{X}_k 的偏导数为零, 可得纯净语音的估计值 \hat{X}_k 为

$$\hat{X}_k = \frac{1}{\int_0^{\infty} \frac{1}{X_k} p(X_k | Y_k) dX_k} \quad (3)$$

利用高斯统计模型, 可以得到最后的幅度谱估计器为

$$\hat{X}_k = \frac{\sqrt{v_k} \exp(v_k / 2)}{\sqrt{\pi \gamma_k} I_0(v_k / 2)} Y_k \quad (4)$$

其中, $v_k = \xi_k \gamma_k / (1 + \xi_k)$, ξ_k 和 γ_k 分别为含噪语音信号的先验信噪比和后验信噪比, $I_0(\cdot)$ 为零阶修正贝塞尔函数。

与其他估计器相比, WEDM 在听觉感知准则的基础上, 对谱谷处感知较明显的残留噪声分配了较大的失真, 对谱峰处的失真则较小, 这样可以在保证语音质量的前提下对噪声进行有效地抑制, 使增强效果明显提高。

但基于统计模型的语音增强算法往往会在宽带增强语音的高频部分出现较窄带语音更为明显的音乐噪声现象, 而本文提出的融合统计模型与 EMD 的宽带语音增强方法可以有效地解决这一问题。

2.2 EMD 方法简介

EMD 方法^[11]是一种新的非线性、非平稳信号分析处理方法, 该方法完全独立于傅里叶变换方法, 它可以直接将信号分解为一组固有模式函数(IMF, intrinsic mode function)。一个 IMF 分量必须满足以下 2 个条件。

1) 整个信号中, 过零点与极值点的个数相等或最多相差一个。

2) 对于信号的任何一点，由局部极大值获得的上包络线和由局部极小值获得的下包络线的均值为零。

EMD 方法是一种完全基于信号本身的信号分析方法，其具有直观、自适应的特点，算法的实现基于以下假设。

1) 待分解信号至少具有 2 个极值点，一个极大值点和一个极小值点。

2) 特征时间尺度是由 2 个极值点间的时间间隔决定的，而不是过零点。

3) 若信号没有极值点但存在拐点，可以通过对信号求导来获得极值点，再通过求积分获得最终的分解结果。

EMD 方法采用了一种筛选的方式对信号进行逐步分解，对于待分解信号 $x(t)$ 其具体分解步骤如下。

1) 首先找出该信号的局部极值点，然后利用 3 次样条插值将极大值连接为上包络线，将极小值连接为下包络线。

2) 取上下包络线的均值定义为 $m_1(t)$ 。

3) 取 $x(t)$ 和 $m_1(t)$ 的差值定义为 $h_1(t)$ ，即 $h_1(t) = x(t) - m_1(t)$ 。

4) 检查 $h_1(t)$ 是否满足 IMF 的 2 个条件，若满足则定义其为第一个 IMF 分量；若不满足则将 $h_1(t)$ 视为待分解信号，重复上述过程直到找到满足 IMF 条件的 $h_k(t)$ ，定义 $c_1(t) = h_k(t)$ 为原始数据的第一个 IMF 分量。

5) 令 $r_1(t) = x(t) - c_1(t)$ 作为新的待处理信号，重复上述筛分过程，得到各个 IMF 分量，即

$$\begin{aligned} r_1(t) - c_2(t) &= r_2(t) \\ r_2(t) - c_3(t) &= r_3(t) \\ &\vdots \\ r_{n-1}(t) - c_n(t) &= r_n(t) \end{aligned} \quad (5)$$

6) 直到 IMF 分量 $c_n(t)$ 或残余分量 $r_n(t)$ 比预定误差小，或 $r_n(t)$ 变为单调函数时，不再筛选 IMF 分量。

7) 最后，信号 $x(t)$ 分解为

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (6)$$

3 基于 VAD 检测的 EMD 增强方法

由于 EMD 技术能够有效地将不同尺度的分量从原始信号中分离出来，而含噪语音信号中的语音成分和噪声成分大多数分布于不同的尺度中。因此，大多数基于 EMD 的增强方法均采用一种简单阈值

处理方法。该方法能够达到一定的去噪效果，但与传统算法相比其性能较差。本文所提算法首先利用 EMD 技术进行了 VAD 检测，根据检测结果针对不同类型的信号选择更加适合的阈值进行去噪处理。相比一般 EMD 方法，该方法能够极大改善增强语音的质量。

3.1 利用 EMD 分析实际信号

通过对含噪语音中的噪声段信号和语音段信号分别进行 EMD 分解，发现两者 IMF 分量的分布特性有明显的差异。对一帧白噪声信号和一帧浊音段语音信号进行分解得到的 IMF 分量分别如图 1 和图 2 所示。

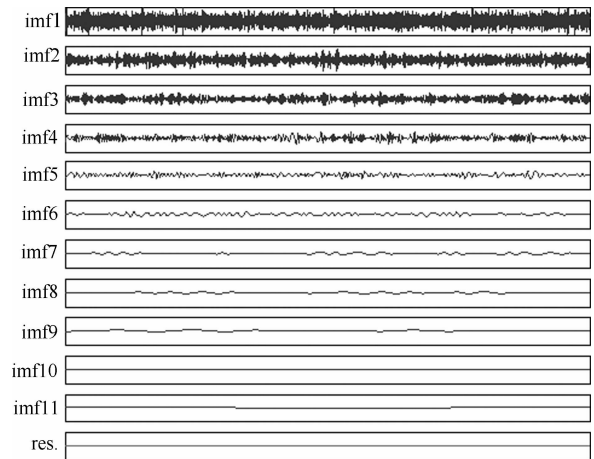


图 1 白噪声信号 EMD 分解结果

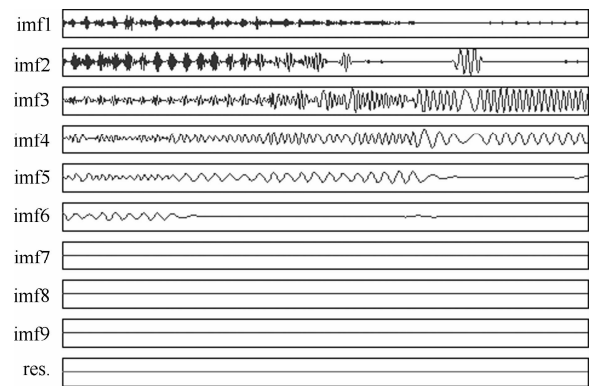


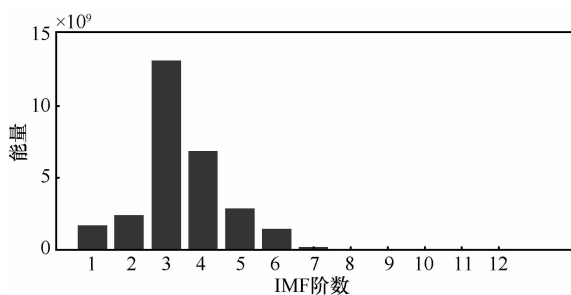
图 2 浊音段语音 EMD 分解结果

从图 1 中可以看出，白噪声经 EMD 分解后其能量主要集中在第一个 IMF 分量中，在第 3 个 IMF 分量之后几乎没有能量存在。而从图 2 中可以看出，纯净语音信号经 EMD 分解之后，其语音成分主要集中在第 2 个 IMF 分量之后的中间尺度的分量中，这是由语音信号的频谱能量主要集中于 300 ~ 3 400 Hz 之间决定的。而小尺度和大尺度的 IMF 分量中能量较小，这也同语音频谱能量的分布相吻合。

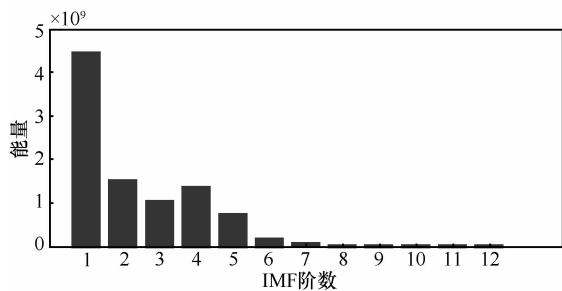
3.2 基于 VAD 检测的 EMD 增强方法

根据上一节的分析可以看出，经 EMD 分解得到的 IMF 分量的分布特性是与当前帧信号的类型密切相关的。本文将一段含噪语音信号分为 3 类：浊音段、清音段和噪声段，将原始含噪语音的构成作为一个参数引入到最终的阈值选取中，这样会更针对性地进行语音去噪。

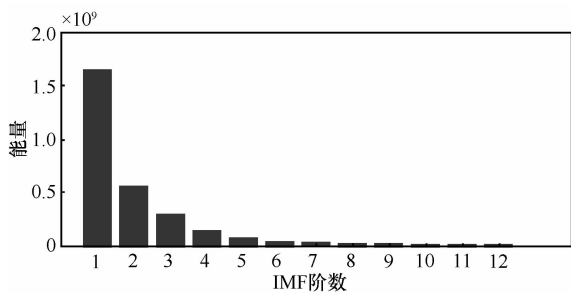
通过研究含噪语音 IMF 分量与噪声 IMF 分量的能量分布特点，本文提出了一种有效的基于 EMD 的 VAD 检测方法。该方法主要是利用 IMF 分量的能量最大值位置来判断当前帧的信号类型的。浊音段语音、清音段语音与噪声段各 IMF 分量的能量分布如图 3 所示。



(a) 浊音段语音



(b) 清音段语音



(c) 噪声段

图 3 IMF 分量能量分布

从图 3 中可以看出，浊音段语音 IMF 分量能量最大的阶数不为 1，而清音段语音和噪声段 IMF 分量能量最大的阶数都为 1。仔细观察可以发现，清音段语音第一个 IMF 分量能量的绝对值要比噪声

段大许多。本文选择了 6 dB、12 dB、18 dB 3 种不同强度白噪声背景下的各 20 句话音进行了 IMF 分量能量分布的统计，统计显示经 EMD 分解后的含噪语音信号的 IMF 分量能量分布均满足以上描述的特性。据此，可以准确地将输入的含噪语音信号分为浊音段语音、清音段语音与噪声段。为了准确性，VAD 算法将输入的长帧信号进行了分帧处理。上述思想的具体实现步骤如下。

- 1) 利用 EMD 算法对输入的长帧信号进行分解，这里取帧长 N 为 4 096。
- 2) 对分解后的 IMF 分量进行分帧，子帧长为 $n=256$ ，共有 16 个子帧。
- 3) 计算各子帧每个 IMF 分量的能量值，即

$$E_i = \sum_{j=1}^n [\text{imf}_i(j)]^2 \tag{7}$$

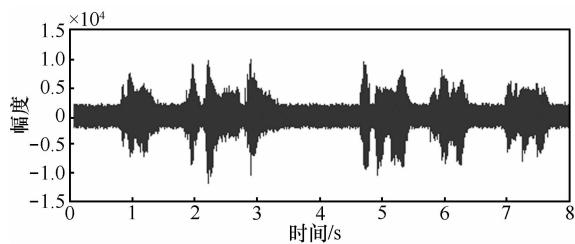
- 4) 确定能量最大值所在的阶数 i ，根据 i 判断是否为浊音。如果 i 不等于 1，则判断为浊音段语音。
- 5) 再根据第一个 IMF 分量的大小判断是清音段语音还是噪声段，计算

$$S = E_1 / E_{1_buff} \tag{8}$$

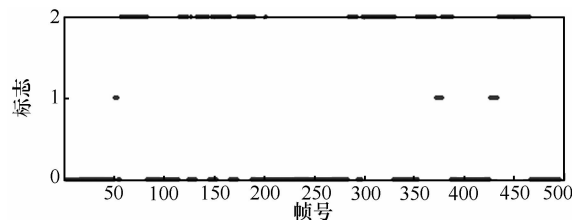
其中， E_{1_buff} 是只在噪声帧进行更新的第一个 IMF 分量能量的历史值。

若 S 大于某一阈值则判断为清音段语音，否则判断为噪声段。

利用上述方法对一段实际含噪语音进行 VAD 检测的实验结果如图 4 所示，0 标志噪声段，1 标志清音段语音，2 标志浊音段语音。



(a) 含噪语音信号



(b) VAD 检测结果

图 4 基于 EMD 的 VAD 检测方法实验结果

从图 4 中可以看出该方法可以准确地将输入含噪话音分为浊音段话音、清音段话音与噪声段 3 种。

根据检测的结果，对不同的信号段，采用不同强度的阈值处理，就可以得到最终的增强结果。

文献[12]提出了一种根据标志的不同和 IMF 阶数的不同自适应变化的阈值选择方法，本文根据该方法得到的阈值进行了相应的阈值处理。下面介绍这种方法的原理和阈值处理过程。

设一段含噪话音信号经 EMD 算法分解为 L 个 IMF 分量， $v(i)$ 为第 i 个 IMF 分量 R_i 的方差。假设被检测到的距离当前帧最近的纯噪声帧的 IMF 分量能量分布为 P_r ，该噪声段 EMD 分解结果中 R_i 在所有 IMF 分量中的能量比重表示为 $P_r(i)$ 。由于含噪话音信号的第一个 IMF 分量方差与实际噪声方差近似成正比关系，因此用被处理帧第一个 IMF 分量的方差进行噪声估计，估计得到

$$\sigma^2 = v(1) / P_r(1) \quad (9)$$

这样就可以设定一个随 IMF 分量阶数和信号种类特性变化的阈值

$$\tau_i = \lambda \sigma P_r(i) \lg\left(\frac{1}{v(i)\sqrt{i}}\right), \quad i=1,2,\dots,L \quad (10)$$

其中， λ 是一个影响阈值大小的参数，它随着 VAD 检测得到的标志变化。若判断为噪声段，则对全部 IMF 分量进行大幅度的衰减，即 λ 取较大的值，一般大于 1；若判断为清音段话音，则对 IMF 分量做适当的衰减，可令 $\lambda=1$ ；而对于浊音段话音，则进行较小幅度的衰减，令 λ 取较小的值，一般小于 1。

确定了每段信号各个阶数的阈值 τ_i 后，可以通过式(11)进行阈值处理，得到经过增强后的 IMF 分量，然后将各个 IMF 分量累加即可得到最终的增强话音为

$$\hat{C}_i(t) = \begin{cases} R_i(t) - \tau_i, & \text{若 } R_i(t) \geq \tau_i \\ 0, & \text{若 } |R_i(t)| < \tau_i \\ R_i(t) + \tau_i, & \text{若 } R_i(t) \leq -\tau_i \end{cases} \quad (11)$$

其中， $\hat{C}_i(t)$ 是经阈值处理后得到的增强后 IMF 分量。

由上，基于 VAD 检测的 EMD 增强算法流程图如图 5 所示。

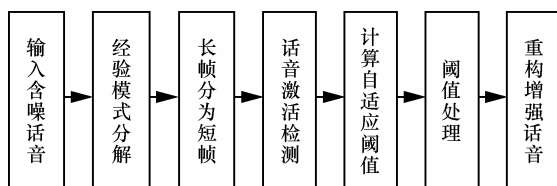


图 5 基于 VAD 检测的 EMD 增强算法流程

4 融合增强算法

由于 EMD 算法能有效地将不同尺度的波形从信号中逐级分离出来，而经统计模型算法增强后，话音中存在的音乐噪声和残留的背景噪声一般是较小尺度的波形，因此，本文考虑利用上节中提出的增强算法进行后处理，以达到去除音乐噪声和残留噪声成分的目的。

首先，利用 EMD 算法对统计模型增强后的话音进行分解，观察其 IMF 分量的特性。一段增强话音被 EMD 分解后的结果如图 6 所示，由于后面的分量能量已经很小，这里只给出了前 6 个 IMF 分量的波形。

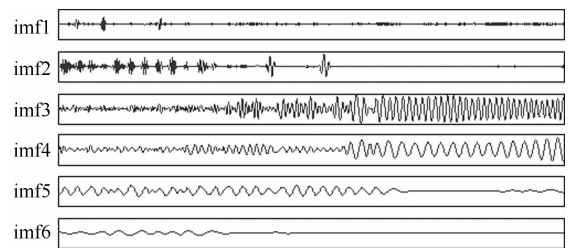


图 6 EMD 算法对统计模型增强后话音的分解结果

从上图 6 可以看出，由于被分解信号已经经过增强处理，分解结果中的小尺度 IMF 分量的能量很低，而在大尺度的 IMF 分量中噪声成分并不明显。因此，只选取前 2 个 IMF 分量进行最后的阈值处理，以达到进一步去除残留噪声和音乐噪声的目的。对于较大尺度的 IMF 分量则不做衰减处理，以免损失过多的话音成分。

由上，EMD 后处理增强的算法流程如图 7 所示。

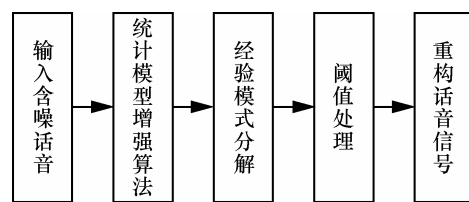


图 7 融合增强算法流程

其中统计模型增强算法采用 WEDM 幅度谱估计器，噪声估计和信噪比估计分别选择最小值控制递归平均 (MCRA, minima-controlled recursive averaging)^[13] 算法和基于判决的信噪比估计方法^[5]。而后处理则采用基于 VAD 检测的 EMD 增强方法。由于本文基于统计模型增强算法选择的帧长为 512 个样点，且进行了叠接处理，而基于 VAD 的 EMD 增强方法选择的帧长是 4 096 个样点。因此，在 2 种算

法结合时，需要首先对统计模型算法后的增强信号缓存 16 帧，然后对这 16 帧的增强信号再进行整体 EMD 分解，做进一步的去噪处理。

5 实验结果及分析

本文中涉及的测试项目主要来自 ITU-T G160 标准。测试中采用的信号均为 16 kHz 采样的宽带信号，序列长度为 8 s。纯净语音信号选自 NTT 标准语音库中的中文子库，信号电平均调整到 -26 dB。高斯白噪声信号由 Cool Edit 软件生成，有色噪声信号从 ITU-T 噪声数据库中选择，包括车内噪声和街道噪声 2 种。

G160 标准是 ITU-T 提出的一项针对移动网络中语音增强设备的国际标准。该标准明确给出了语音增强设备的特征、性能指标及测试方法。性能测试分别在白噪声和有色噪声 2 种环境下进行。其中在有色噪声环境下的测试项目中加入了 G160 中没有提到的语音质量提高测试(PESQI, perceptual evaluation of speech quality improvement)^[14]。具体的测试项目主要包括如下几项。

1) 白噪声背景下

白噪声环境下的性能测试主要包括期望噪声衰减量测试(Q_m)、实际噪声衰减量测试(Q_{n1} , Q_{n2})、信号电平衰减测试(Q_s)和收敛时间测试。测试达标要求如下：

① 实际噪声衰减量必须达到期望噪声衰减量的 ± 3 dB 范围内，即 $(Q_m - 3)$ dB $< Q_n < (Q_m + 3)$ dB；

② 信号电平衰减需满足 -3 dB $< Q_s < 2$ dB。

2) 有色噪声背景下

有色噪声环境下的性能测试主要包括信噪比提高测试(SNRI, signal-to-noise ratio improvement)、整体噪声电平衰减测试(TNLR, total noise level reduction)和 SNRI-to-NPLR difference 测试(DSN)。其中，NPLR(noise power level reduction)表示短暂停噪声帧中语音增强算法的噪声衰减量。DSN 测试用来检测语音增强算法对语音信号功率的影响。测试达标要求如下：

① 信噪比提高需满足 $SNRI \geq 4$ dB；

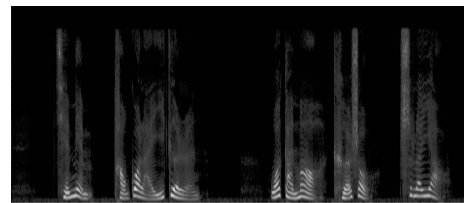
② 整体噪声电平衰减需满足 $TNLR \leq -5$ dB；

③ DSN 测试需满足 -4 dB $\leq DSN \leq 3$ dB。

本文选择性能较好的 WEDM 幅度谱估计器^[10]作为参考算法。下面分别介绍本文所提融合增强方法与参考算法增强效果和各项测试指标的对比。

在白噪声背景下经 2 种增强方法处理后的语音

的语谱图对比如图 8 所示，含噪语音信号的输入 SNR 均为 6 dB。



(a) 纯净语音信号



(b) 含噪语音信号



(c) 参考方法输出语音



(d) 本文方法输出语音

图 8 增强效果对比

从图 8 的仿真结果中不难看出，2 种方法均能有效地消除背景噪声，但是本文所提方法在保持原始语音信号能量不变的前提下，所能提供的噪声衰减量更大。

2 种方法在白噪声环境下和有色噪声环境下各项测试的结果分别如表 1 和表 2 所示。

表 1 白噪声环境下测试结果

测试项目指标		参考方法	本文方法
期望噪声衰减/dB	Q_m	21.49	23.24
实际噪声/dB	Q_{n1}	21.57	23.24
衰减和语音/dB	Q_{n2}	21.62	23.25
电平衰减/dB	Q_s	0.54	0.54
收敛时间/s	T_1	0.70	0.70
	T_2	0.29	0.29
	T_3	0.01	0.01

观察表 1 中数据可以看到, 2 种增强方法的实际噪声衰减量与期望噪声衰减量均非常接近, 满足测试达标要求。但本文所提方法在话音电平衰减与参考方法相同的情况下, 提供了更大的实际噪声衰减量。收敛性测试表明 2 种方法在背景噪声能量发生突变时均能很快地对噪声能量进行跟踪, 收敛性较好。

观察表 2 中数据可以看出, 2 种增强方法均满足 G160 中有色噪声环境下的 3 项测试标准。另外, 本文算法在 DSN 测试结果与参考方法接近, 即话音信号能量损失几乎相同的前提下, 提供了更大的信噪比提高和噪声能量衰减。在 PESQI 测试中, 2 种方法的 PESQI 结果均随着含噪话音信噪比的提高而降低。这是由于信噪比较高时, 噪声对话音的干扰作用相对较小, 含噪话音信号的客观质量本身较大, 因此增强算法对客观话音质量的提高不如在低信噪比环境下明显。在同一信噪比条件下, 本文方法的话音质量提高量均明显高于参考算法。

表 2 有色噪声环境下测试结果

测试项目指标	参考方法	本文方法	
G160	SNRI/dB	12.13	12.93
	TNLR/dB	-19.90	-20.74
	DSN/dB	-1.514	-1.528
PESQI	6 dB	0.291 3	0.324 1
	12 dB	0.250 8	0.301 9
	18 dB	0.211 7	0.258 4

6 结束语

本文首先研究了一种基于 VAD 检测的 EMD 增强方法。然后通过结合统计模型增强方法和基于 EMD 的增强方法各自的优点, 提出了一种融合宽带语音增强方法。该方法首先用统计模型方法去除含噪话音中的主要噪声成分, 然后用 EMD 技术做后处理操作, 以达到去除音乐噪声和残留噪声成分的目的。最终的测试结果表明, 本文所提方法的各项客观测试指标均优于参考方法。另外, 仿真结果表明本文方法能够提供更好的主观听觉感受。

参考文献:

- [1] LIM J S, OPPENHEIM A V. Enhancement and bandwidth compression of noisy speech[J]. Proceedings of the IEEE, 1979, 67(12): 1586-1604.
- [2] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans Acoust Speech Signal Process, 1979, 27(2):

113-120.

- [3] JENSEN S H, HANSEN P C, HANSEN S D, *et al.* Reduction of broadband noise in speech by truncated QSVD[J]. IEEE Trans Speech and Audio Process, 1995, 3(6):439-448.
- [4] DONOHO D L. De-noising by soft-thresholding[J]. IEEE Trans on Information Theory, 1995, 41(3):613-627.
- [5] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. IEEE Trans Acoust Speech Signal Process, 1984, 32(6):1109-1121.
- [6] BOUDRAA O, CEXUS J C, SAIDI Z. EMD-based signal noise reduction[J]. Int J Sig Process, 2004(1):33-37.
- [7] ESCH T, HEESE F, GEISER B *et al.* Wideband noise suppression supported by artificial bandwidth extension techniques[A]. Proc IEEE Int Conf on Acoustics, Speech, and Signal Processing[C]. Dallas, USA, 2010. 4790-4793.
- [8] ITU-T G160. Voice Enhancement Devices[S]. 2008.
- [9] LOTTER T, VARY P. Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model[J]. EURASIP J Appl Signal Process, 2005(7):1110-1126.
- [10] LOIZOU P. Speech enhancement based on perceptually motivated Bayesian estimators of the speech magnitude spectrum[J]. IEEE Trans Speech Audio Process, 2005, 13(5):857-869.
- [11] HUANG N E, SHEN Z, LONG S R, *et al.* The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis[A]. Proc Royal Soc London[C]. London, UK, 1998. 903-995.
- [12] LI X Y, ZOU X J, ZHANG R B, *et al.* Method of speech enhancement based on Hilbert-huang transform[A]. Proceedings of the 7th World Congress on Intelligent Control and Automation[C]. Chongqing, China, 2008. 8414-8418.
- [13] COHEN I, BERDUGO B. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. IEEE Signal Process Lett, 2002, 9(1):12-15.
- [14] ITU-T P862. Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs[S]. 2001.

作者简介:



周璇 (1986-), 男, 山东济南人, 北京工业大学硕士生, 主要研究方向为语音增强。



鲍长春 (1965-), 男, 蒙古族, 内蒙古赤峰人, 博士, 北京工业大学教授、博士生导师, 主要研究方向为语音与音频信号处理。

夏丙寅 (1986-), 男, 北京人, 北京工业大学博士生, 主要研究方向为语音增强。