

基于改进谐波恢复算法的语音增强方法

吕言国, 崔慧娟

(清华大学电子工程系信息科学与技术国家实验室, 北京 100084)

摘要: 当信噪比较低时, 语音信号的高次谐波部分会完全淹没在噪声中。针对该情况, 提出一种基于改进谐波恢复算法的语音增强方法。对经过 MMSE-LSA 算法语音增强处理后的时域输出语音信号进行非线性处理, 得到准周期冲激信号, 并将其与原增强信号相乘, 突出语音的谐波分量。实验结果表明, 改进算法较好地解决了低信噪比时谐波失真的问题, 相比传统谐波恢复算法能更好地改善语音高次谐波的质量。

关键词: 语音增强; 谐波恢复; 非线性; 低信噪比; 冲激函数; 低通滤波

Speech Enhancement Method Based on Improved Harmonic Retrieval Algorithm

LV Yan-guo, CUI Hui-juan

(National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

【Abstract】 The higher harmonic part of the speech signal is completely overwhelmed by the noise when Signal to Noise Ratio(SNR) is low. In response to this situation, a new algorithm is proposed. A quasi-periodic impulse signal is calculated based on the distorted signal processed by MMSE-LSA algorithm using a nonlinearity to regenerate harmonics. This artificial signal is then used to multiply the original enhanced signal, making the voice of the harmonic components to be prominent. Experimental results show that the algorithm is a good solution to the problem of Harmonic distortion better and is better than the traditional algorithm when improving the voice quality of high harmonics in low SNR.

【Key words】 speech enhancement; harmonic retrieval; nonlinearity; low Signal to Noise Ratio(SNR); impulse function; Low-pass Filtering(LPF)

DOI: 10.3969/j.issn.1000-3428.2012.04.080

1 概述

由于噪声的存在会明显降低语音识别和语音编码系统的性能, 因此语音增强越来越广泛地应用于上述系统中。针对语音增强的算法有很多种, 比较经典的维纳滤波法和谱减法分别由 Lim J S 和 Oppenheim A V^[1]于 1978 年及 Boll S^[2]于 1979 年提出, 20 世纪 90 年代以来, 基于隐马尔可夫模型的方法得到了重点研究。基于高斯模型^[3]的方法也是很重要的方法。其中, 基于短时谱估计的算法应用最普遍, 基于 MMSE 及 MMSE-LSA^[4]估计算法的增强效果最好。但是, 上述各种算法对低信噪比时高次谐波的恢复效果不好。本文算法在 MMSE-LSA 算法的基础上提出, 可有效地恢复低信噪比下的高次谐波。

2 基于 MMSE-LSA 算法的谐波恢复

2.1 MMSE-LSA 算法分析

设带噪语音的时域表示:

$$m(i) = s(i) + n(i), 0 \leq i \leq N-1 \quad (1)$$

其中, $m(i)$ 是带噪语音; $s(i)$ 是纯净语音; $n(i)$ 是噪声; N 为帧长。其频域表示为:

$$M(k) = S(k) + N(k) \quad (2)$$

其中, $M(k) = C(k)\exp(j\theta(k))$; $S(k) = A(k)\exp(j\alpha(k))$; $N(k)$ 分别为 $m(i)$ 、 $s(i)$ 和 $n(i)$ 的傅里叶变换系数。

语音增强的目的是基于带噪语音估计纯净语音。基于短时谱估计的语音增强算法可写成如下统一形式:

$$|\hat{S}(k)| = G(k)|M(k)| \quad (3)$$

基于 MMSE-LSA 的短时谱估计语音增强算法增益^[4]为:

$$G_{LSA}(k) = \frac{\xi_k}{1 + \xi_k} \exp\left(\frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right) \quad (4)$$

其中, ξ_k 和 v_k 可分别由先后验信噪比求得。基于短时谱估计的语音增强算法原理如图 1 所示。

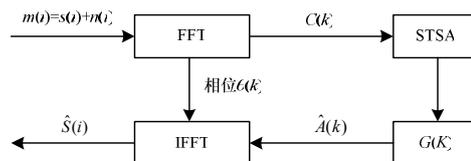


图 1 基于短时谱估计的语音增强方法原理

2.2 谐波恢复算法

图 2 为信号与白噪声混合后的频谱图, 信噪比为 0。当信噪比较低时, 带噪语音高频部分的频谱类似于白噪声的频谱, 此时语音的谐波特性已经消失。

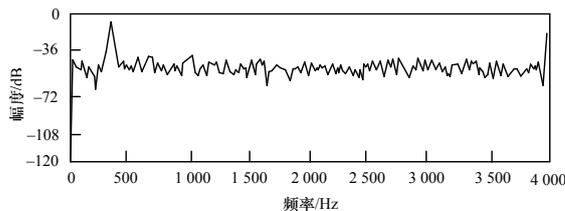


图 2 信噪比为 0 时的带噪语音频谱

基金项目: 国家自然科学基金资助项目(60572081)

作者简介: 吕言国(1980—), 男, 硕士研究生, 主研方向: 语音增强; 崔慧娟, 教授

收稿日期: 2011-05-09 **E-mail:** lvyanguo2000@163.com

为恢复语音的谐波特性,引入谐波恢复算法,将经过基于MMSE-LSA算法的语音增强处理后的时域输出语音 $\hat{s}(t)$ 进行如下非线性处理,得到 $s_{\text{harmonic}}(t)$ ^[5]:

$$s_{\text{harmonic}}(t) = \hat{s}(t) \cdot p(\hat{s}(t)) \quad (5)$$

其中:

$$p(\hat{s}(t)) = \begin{cases} 1 & \hat{s}(t) > 0 \\ 0 & \hat{s}(t) < 0 \end{cases} \quad (6)$$

设语音的周期为 T ,函数 $p(\hat{s}(t))$ 也以 T 为周期。分析可知,因为函数 $p(\hat{s}(t))$ 的周期性使信号的高次谐波都得到一定的加强。但是,函数 $p(\hat{s}(t))$ 的傅里叶变换是 $Sa(\frac{m}{T})$ 函数,谐波幅度收敛速率为 $1/m$,各次谐波得到不同程度的增强,且频率越高增强效果越不明显。

3 谐波恢复算法的改进

3.1 算法原理分析

为解决上述问题,把周期方波函数 $p(\hat{s}(t))$ 替换为周期冲激函数 $p_{\delta}(\hat{s}(t))$:

$$p_{\delta}(\hat{s}(t)) = \begin{cases} 1 & \hat{s}(t) = (\hat{s}(t))_{\text{middle}}, i = 0, 1, \dots \\ 0 & \text{else} \end{cases} \quad (7)$$

其中, $(\hat{s}(t))_{\text{middle}}$ 为 $t \in (T_i, T_{i+1}), i = 0, 1, \dots$ 时,信号 $\hat{s}(t)$ 的第 i 个周期的正半周中间的信号值(如图3中的虚线所示)。如图3所示,通过式(7)的非线性处理可得到语音信号的周期 T ,同时也得到了以 T 为周期的周期冲激函数。

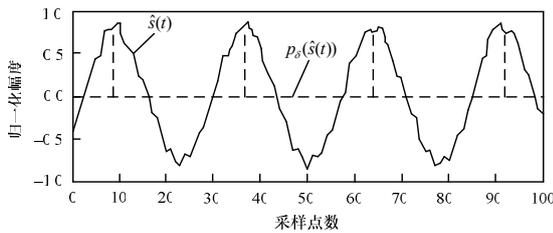


图3 时域信号非线性变换示意图

令:

$$s_{\text{harmonic}}(t) = \hat{s}(t) \cdot p_{\delta}(\hat{s}(t)) \quad (8)$$

得到 $s_{\text{harmonic}}(t)$ 的傅里叶变换:

$$FT(s_{\text{harmonic}}(t)) = FT(\hat{s}(t)) \times FT(p_{\delta}(\hat{s}(t))) = \frac{1}{T} \sum_{m=-\infty}^{m=\infty} F_{\hat{s}}(f - \frac{m}{T}) \quad (9)$$

从式(9)可以看出,信号的各次谐波都得到相同的增强。由于式(9)的卷积结果会产生频谱混叠现象,因此,若想取得较好的处理结果,在进行该运算前应先对信号 $\hat{s}(t)$ 进行低通滤波。

3.2 算法在语音增强中的应用

谐波恢复算法一方面增强了各次谐波,而另一方面由于上述信号及冲激函数并非严格意义的周期信号,因此,运算后也引入了其他频率成分。另外,由式(9)可知,信号 $s_{\text{harmonic}}(t)$ 的傅里叶变换会产生频谱混叠。从图2可以看出,带噪语音的低频部分在低信噪比(Signal Noise Ratio, SNR)的情况下保留了较好的纯净语音的特征。鉴于以上原因,在语音增强中采用如下的谐波恢复算法:

- (1)对信号 $\hat{s}(t)$ 进行截止频率为800 Hz的低通滤波。
- (2)根据滤波后的信号得出函数 $p_{\delta}(\hat{s}(t))$ 。
- (3)将步骤(1)和步骤(2)得出的2个函数相乘,得到 $s_{\text{harmonic}}(t)$ 。
- (4)对函数 $s_{\text{harmonic}}(t)$ 进行傅里叶变换,得到相应的频谱函数 $\hat{S}_{\text{harmonic}}(n, k)$ 。

(5)计算语音频谱函数如下:

$$\hat{S}_{\text{harmonic}}(n, k) = \begin{cases} \hat{S}(n, k) & k < K0 \\ \alpha \cdot \hat{S}(n, k) + (1 - \alpha) \cdot \hat{S}_{\text{harmonic}}(n, k) & k > K0 \end{cases} \quad (10)$$

其中, $\hat{S}(n, k)$ 是函数 $\hat{s}(t)$ 的频谱函数; α 是平滑系数。

为避免谐波恢复算法对信号低频部分的干扰,同时又能较好地恢复高次谐波,采用分段处理的方法对不同频段进行不同的处理,式(10)是分为2段计算语音频谱的公式。由于一般情况下频率越高,带噪语音的谐波失真越严重,为更好地恢复所有的谐波分量,可采用分多段处理的方法。段数越多,取得的增强效果也就越好。但是,每段的平滑系数的确定是一个难点。

3.3 改进算法

众所周知,语音可以分为清音和浊音。浊音呈现明显的周期性,清音则不同,其特征非常类似于白噪声。本文算法基于信号具有周期性,也就是算法主要是针对浊音信号的。在使用谐波恢复算法前,应该先进行清浊音判决,清音则跳过本算法,当判断为浊音时则采用本文算法。

清浊音的判决可采用以下方法得到:本文在3.1节讨论了函数 $p_{\delta}(\hat{s}(t))$ 的求解,在此基础上,判断每一帧的 $p_{\delta}(\hat{s}(t))$ 是否具有近似的周期性,有即是浊音,否则为清音。

4 实验结果与分析

图4(c)是原算法的语谱图,图4(d)是经过本文谐波恢复算法后的语音语谱图,可以看到,本文算法在高次谐波的恢复上明显优于原算法。

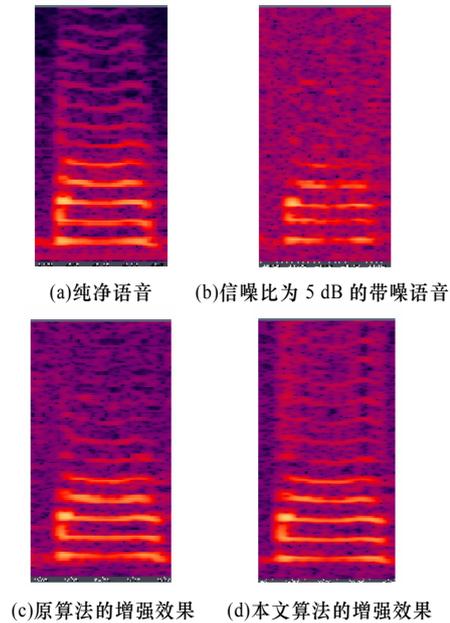


图4 应用算法后的语音谱图比较

语音增强最终目的是提高语音的质量,提高信号信噪比只是他的一个客观指标,有时候信噪比的提高并不能改善语音质量。在语音增强的过程中,最重要的是提高人耳对增强后语音的听觉感受,最大程度地提高语音的清晰度和可懂度。为实现这一目的,需要有比较准确合理的主观评价标准来衡量语音。语音质量的主观评测方法有多种,常用的有平均意见得分(Mean Opinion Score, MOS)、诊断押韵测试(DRT)、判断满意度(DAM)。下面实验结果采用MOS,表1是MMSE-LSA算法以及增加了本文算法的MMSE-LSA算法

(下转第250页)