

文章编号:1672-058X(2013)03-0062-05

自适应谱增强在 MBE 语音编码中的应用研究*

黄姗姗, 许钢**, 李远军

(安徽工程大学 安徽省电气传动与控制重点实验室, 安徽 芜湖 241000)

摘要:通过对多带激励(multi-band excitation, MBE)语音编码的研究,将自适应谱增强技术应用在多带激励语音编码中,提升了合成语音的清晰度,改善了合成语音的质量;仿真实验表明:自适应谱增强技术使合成语音的基频和二次谐波的谱幅度明显提升,有效地增强了合成语音的清晰度。

关键词:MBE; 自适应谱增强; 语音编码; 语音解码

中图分类号:TN912.3

文献标志码:A

在低速率语音编码中,合成语音与原始语音的匹配是一个重要的问题^[1]。在解码端对语音信号进行一些处理,来增强合成语音质量的方法有很多,比如将得到的合成语音通过感知加权滤波器,或者自适应后置滤波器,或者自适应谱增强滤波器等^[2]。感知加权滤波器是利用人耳的掩蔽效应设计得来的,依据最小均方预测误差(MPSE)搜索时,谱噪声的不同分量对应的加权值不同,来调节谱噪声的形状,但感知加权滤波器无法减小总的噪声能量,谱峰处合成语音与原始语音的误差反而会变大;自适应后置滤波器也是利用人耳的听觉特征,它和感知加权滤波器的区别是直接对输出信号起作用,将谱噪声的形状进一步修改,抑制对于听觉较敏感的噪声,但是自适应后置滤波器的设计过程比较复杂。在此采用自适应谱增强滤波器^[3-5],这种自适应谱增强技术广泛应用于基于共振峰谱包络合成的中低速率语音编码中,在多带激励语音解码端,将合成语音通过自适应谱增强滤波器,使合成语音与原始语音在共振峰范围内有更好的匹配,实现原理相对较简单,算法复杂程度低,是加强低速率语音编码合成语音质量的实用技术^[6,7]。

1 语音编码

改进的多带激励(MBE)语音编码的原理框图如图 1 所示。对输入信号 $s(n)$ 加窗(图中用的是三角窗)分帧,加窗之后的语音信号用 $s_w(n)$ 表示,进行线性预测分析得到线性预测系数 a_i , 然后对 $s_w(n)$ 进行 FFT,基音估计得到基音频率 w_0 ,由 w_0 得到各次谐波,并划分出频带,得到每个子频带的清浊音标志 V_k 和谐波幅度 A_m 。

1.1 基音估计

基音周期的估计分为初始估计和精细估计两步。对语音信号进行基音初始估计时,计算出与各个候选基音周期对应的误差函数值后,选择使误差函数最小的周期作为初始基音估计值。误差函数 $E(P)$ ^[8] 的定义如下:

收稿日期:2012-09-25;修回日期:2013-01-10.

* 基金项目:安徽高校自然科学基金重点研究项目(KJ2011A035).

作者简介:黄姗姗(1988-),女,安徽淮南人,硕士研究生,从事信号处理的研究.

** 通讯作者:许钢(1972-),男,安徽芜湖人,副教授,从事数字信号处理和 SOPC(可编程片上系统)的研究.

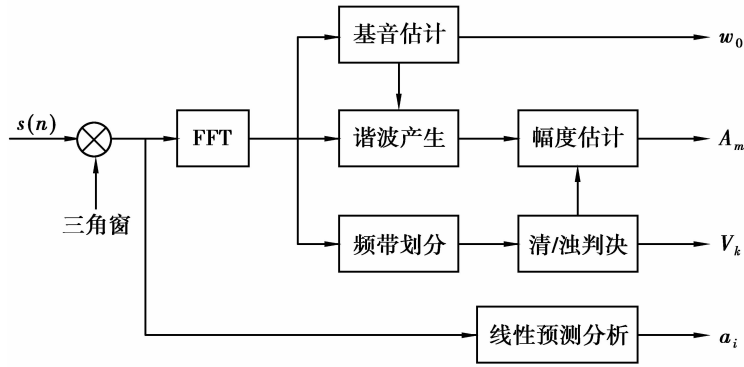


图 1 改进的 MBE 语音编码器原理框图

$$E(P) = \frac{\sum_{n=-j}^j s_{lp}^2(n)w_l^2(n) - P \sum_{k=-\lfloor \frac{m}{P} \rfloor}^{\lfloor \frac{m}{P} \rfloor} \phi(kP)}{[1 - P \sum_{n=-j}^j w_l^4(n)][\sum_{n=-j}^j s_{lp}^2(n)w_l^2(n)]}$$

自相关函数^[8]定义:

$$\phi(t) = \sum_{n=-j}^j s_{lp}(n)w_l^2(n)s_{lp}(n+t)w_l^2(n+t)$$

其中, $j=170, m=114$ 。

在进行基音初始估计后要进行基音平滑, 来保证基音周期的连续性。这里采用基音跟踪算法进行平滑, 该算法分为两个部分: 后向基音跟踪和前向基音跟踪。对每个当前帧进行后向基音跟踪得到后向估计 P_B , 它保持当前帧与过去帧之间基音周期的连续性; 对当前帧进行前向基音跟踪得到前向估计 P_F , 它保持当前帧与将来帧之间基音周期的连续性。然后根据判决规则比较 P_B 和 P_F , 选择一个作为平滑后当前帧基音周期的初始估计 P_I 。

在得到基音周期的初始估计 P_I 之后, 产生这 10 个: $P_I - 9/8, P_I - 7/8, \dots, P_I + 7/8, P_I + 9/8$ 基音精细估计的候选值 P_0 。因为精估是在频域中进行的, 所以要将时域周期值转换为相应的基音频率 $w_0 = 2\pi/P_0$ 。然后对这 10 个点进行细搜索, 确定更为精细的基音周期 w_0 。

1.2 V/U 判决

V/U 的判决可以依据每个谐波处加窗语音谱与合成谱拟合的程度来判断。如果拟合误差小于一定的门限值, 就判别该谐波频带是浊音带, 否则为清音带, 采用 DFT 表达时, 归一化误差^[8]:

$$\xi_m = \frac{\sum_{l=a_m}^{b_m} |S_W(l) - S_{W_r}(l, w_0)|^2}{\sum_{l=a_m}^{b_m} |S_W(l)|^2}$$

1.3 谱幅度估计

对于浊音子频带和清音子频带, 谐波幅度的估计方法是不同的。浊音子频带, $A_m(w_0)$ 能很准确地表达谱包络。所以, 可以利用落在浊音子频带的合成谐波幅度作为第 m 个频带的谐波幅度估计^[8], 即:

$$A_m = |A_m(w_0)|$$

而清音谐波幅度^[8]则由每个清音子频带的语音信号的 RMS(均方根)来表达, 即:

$$A_m = \left[\frac{1}{\sum_{n=-110}^{110} w_R(n)} \right] \left[\frac{\sum_{l=a_m}^{b_m} |S_W(l)|^2}{b_m - a_m} \right]^{1/2}$$

1.4 线性预测分析(LPA)

使用 200 个样点的哈明窗对输入信号加权,再进行 10 阶的 LPA,当前帧的最后一个样点作为窗的中心。第一步用传统的 Levinson-Durbin 算法求出线性预测系数 $a_i (i=1, 2, \dots, 10)$;第二步对 a_i 作 15 Hz 的带宽扩展,即 $a_i = 0.994^i \times a_i (i=1, 2, \dots, 10)$ 。10 阶线性预测滤波器的传递函数^[1]:

$$A(z) = 1 - \sum_{i=1}^{10} a_i z^{-i}$$

2 语音解码

改进的多带激励(MBE)语音解码的原理框图如图 2 所示。MBE 语音合成的基本思想是将语音信号分解成浊音语音 $s_v(n)$ 与清音语音 $s_u(n)$ 两个分量信号,由基音频率 w_0 ,清浊音标志 V_k 和谐波幅度 A_m 这 3 个参数分别合成这两个分量信号,将两个合成的分量信号相加得到完整的重建语音信号 $s_r(n)$,最后将重建语音通过自适应谱增强滤波器后得到最终的输出语音。

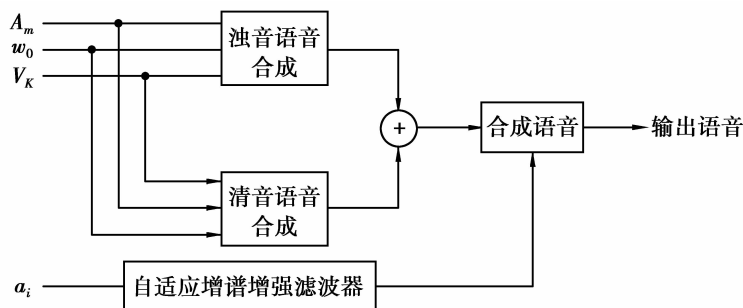


图 2 MBE 语音解码器原理框图

3 自适应谱增强技术

不论合成的或自然的语音,基音脉冲间典型的共振响应一般不会全部衰减,人在说话时声道处的共振峰带宽在一个基音周期内可能会发生变化,致使共振峰之间合成语音谱的波谷没有原始语音谱的波谷尖锐,合成语音峰-峰之间的谷底比自然语音小,使合成语音听起来闷闷的。为了使合成语音波形和原始语音波形在共振峰所在的那些频带匹配得更好,在多带激励语音编码(MBE)算法中引入了自适应谱增强滤波。自适应谱增强是通过将激励信号经过自适应谱增强滤波器来实现的。通过提升激励谱的共振峰谱幅度,增加所有短时谱在共振峰处的信噪比。

自适应谱增强滤波器是根据线性预测滤波器的极点设计的,用它来突出合成语音的共振峰结构同时改善合成波形和输入带通波形在共振峰范围的匹配情况,使合成语音更自然。自适应谱增强滤波器是一个零极点滤波器,其阶数为 10 阶,再级联一个一阶的 FIR 滤波器,经过 15 Hz 带宽扩展的 LPA 系数作为滤波器系数,其传递函数 $H(z)$ ^[1]:

$$H(z) = \frac{A(\alpha z^{-1})}{A(\beta z^{-1})} (1 + \mu z^{-1})$$

其中, $A(z)$ 为线性预测滤波器的传递函数, $\alpha = 0.5p, \beta = 0.8p$ 。这里取 $p = 1, \mu = 0$, 得到一段女声语音和一段男声语音自适应谱增强效果分别如图 3、图 4 所示,虚线是未经过自适应谱增强滤波器的合成语音的频谱曲线,实线是经过自适应谱增强滤波器的合成语音的频谱曲线,由图 3、图 4 可以看出,经过自适应谱增强滤波器的合成语音的基频和二次谐波的幅度有明显增加。

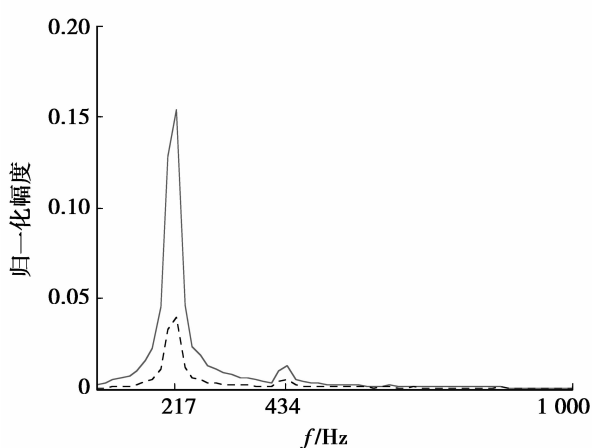


图3 一段女声语音自适应谱增强效果图

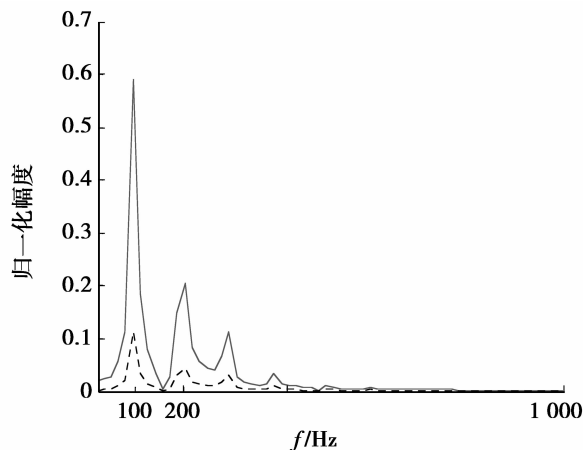


图4 一段男声语音自适应谱增强效果图

自适应谱增强滤波器就怎样匹配共振峰提供了简单有效的办法。此极/零点滤波器可以减少共振峰频率间的量化噪声。自适应滤波器的极点是由线性预测滤波器的扩展得到的,其中 $\beta = 0.8$ 。由于这个全极点滤波器使谱的摆动增大,低通滤波因此受到影响,使用 $\alpha = 0.5$ 的全零点滤波器降低整个滤波器的摆动,同时不减小共振峰增强。在自适应谱增强滤波器中每个 z^{-1} 项用和 βz^{-1} 来代替,使合成语音比较好地匹配共振峰区自然语音的带通特性,可以增加合成语音的主观感知质量。

4 实验结果及分析

下面运用传统算法和改进算法分别对一段女声和男声“多带激励”的语音信号进行语音编解码,在MATLAB中仿真实现,结果如图5和图6所示。

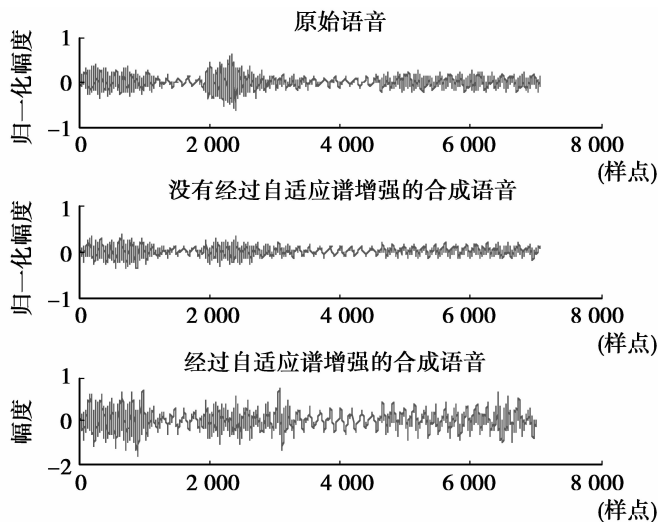


图5 女声原始语音波形和合成语音波形

对于图5、图6,第一个波形是原始语音波形,第二个波形是没有加入自适应谱增强的合成语音,第三个波形是加入了自适应谱增强的合成语音。从图5、图6的波形图上看,没有经过自适应谱增强滤波器的语音波形看起来和原始语音波形更加接近,但是听起来却恰恰相反,经过了自适应谱增强滤波器的合成语音听起来比不经过的更清晰、更自然。

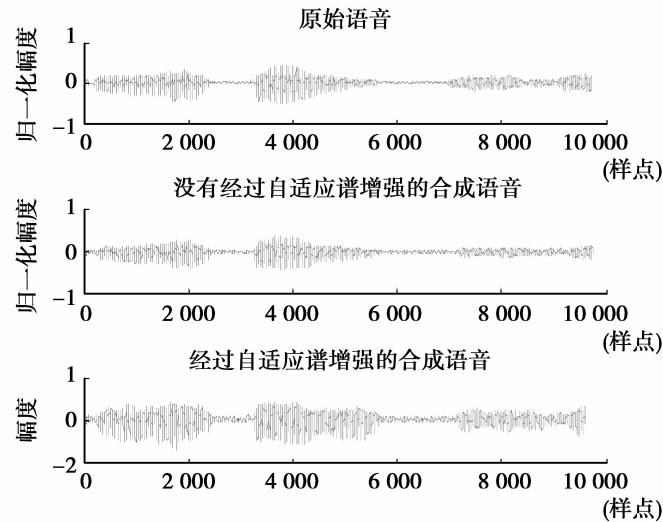


图 6 男声原始语音波形和合成语音波形

参考文献:

- [1] 王炳锡. 语音编码[M]. 西安:西安电子科技大学出版社,2002
- [2] 王洪,唐凯. 低速率语音编码[M]. 北京:国防工业出版社,2006
- [3] 赖长庆. 混合激励线性预测声码器算法的研究[D]. 成都:电子科技大学,2003
- [4] 石亚楠. 基于 MELP 的低速率语音编码算法研究[D]. 西安:西安电子科技大学,2010
- [5] 谭治国. MELP 2.4K 语音编解码算法在 DSP 上的实现[D]. 成都:电子科技大学,2004
- [6] 铁莹. 低速率语音编码的研究[D]. 北京:北京邮电大学,2011
- [7] 黄慧猛. 基于数字对讲机的 MELP 声码器算法研究与实现[D]. 北京:北京化工大学,2011
- [8] 李昌立,吴善培. 数字语音—语音编码实用教程[M]. 北京:人民邮电出版社,2004

Application of Self-adaptive Spectral Enhancement to Speech Coding Based on MBE

HUANG Shan-shan, XU Gang, LI Yuan-jun

(Anhui Key Laboratory of Electric Drive and Control, Anhui Polytechnic University,
Anhui Wuhu 241000, China)

Abstract: By studying the speech coding of multi-band excitation (MBE), self-adaptive spectral enhancement is applied to MBE speech coding, the definition of synthetic speech is upgraded and the quality of synthetic speech is improved. The result of simulation experiment shows that the amplitudes of fundamental frequency and second harmonic of synthetic speech are both increased obviously by using self-adaptive spectral enhancement. So, the definition of synthetic speech is enhanced effectively.

Key words: MBE; self-adaptive spectral enhancement; speech coding; speech decoding