

文章编号:1001-9081(2006)05-1232-02

基于线性预测和最大似然的基音检测算法

李 晋, 王 玲

(湖南师范大学 工学院, 湖南 长沙 410081)

(jin_li81@163.com)

摘 要:根据语音信号产生机理,结合常用的线性预测和最大似然法,提出了一种有效的基音检测算法。该算法采用频域分块估计候选基音周期的范围,提高了算法的计算速度。仿真实验表明,该算法与传统方法相比其基音检测结果有明显的改善,克服了随机错误及倍频、半频错误,并且在低信噪比下鲁棒性较好。

关键词:线性预测编码;最大似然;基音检测

中图分类号: TP37 **文献标识码:** A

Pitch detection algorithm based on LPC and maximum likelihood

LI Jin, WANG Ling

(Institute of Engineering, Hunan Normal University, Changsha Hunan 410081, China)

Abstract: According to the mechanism of speech signal, an effective pitch detection algorithm by combined liner predictive coding with maximum likelihood was proposed. The algorithm adopts frequency partition to estimate the range of pitch period, improving the computational speed. Simulation experiments show that the pitch extracting results of the algorithm are superior to that of the conventional algorithms, overcoming the random errors and multiple or half frequency errors, especially robust at low SNR.

Key words: LPC(liner predictive coding); maximum likelihood; pitch detection

0 引言

基音周期是描述语音信号激励源的重要特征之一。由于语音信号可视为一个动态非平稳随机过程,这使得语音波形变化十分复杂。声带振动的频率不仅受到背景噪声,个人声带的长短、厚薄、韧性和发音习惯的影响,还与发音者的性别、年龄、发音时的力度及情感等有关^[1]。至今还没有一种通用的方法能准确可靠地估计出任何人在任何情况下的基音周期。

目前,基音检测方法主要是基于传统语音模型,可分为时域法、频域法和时频域混合法。时域法最具代表性的是短时平均幅度差函数(AMDF)法^[2]和短时自相关(ACF)法^[3]。一般AMDF法在语音信号的幅度或频率变化比较快时,基音检测精度明显下降。ACF法通常会导致“倍频”或“半频”错误。频域法中的倒谱(CEP)法^[4]由于引入了对数运算,致使进行数字处理时增加了计算量。之后,在倒谱法的基础上提出了一种改进的功率二次谱(Power Spectrum Reprocessing)方法^[5],但因除零频外的最大谱峰的纵坐标值在过渡音中将变得不清晰,导致横坐标估值不准确。时频域结合的小波变换(Wavelet)法^[6]则易受噪声的干扰,随着信噪比的下降,基音检测的误差越来越大。针对这些问题,本文提出了一种基于线性预测和最大似然的基音检测算法,在噪声环境下测试,该算法能有效检测到基音周期,减小了提取误差,鲁棒性较好。

1 语音信号的线性预测(LPC)

线性预测是表示语音信号波形的重要方法之一,其分析思想是用过去若干个取样值和激励信号的线性组合来表示语

音信号现在或未来的取样值^[7]:

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (1)$$

式中: a_k 为线性预测系数; p 为预测阶数; $u(n)$ 为语音激励信号; G 为幅度因子。

p 阶线性预测器的输出为:

$$\hat{s}(n) = - \sum_{k=1}^p a_k s(n-k) \quad (2)$$

预测误差为:

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^p a_k s(n-k) \quad (3)$$

根据最小均方误差(LMS)准则,按照式(4)使预测误差功率 ε 最小,可以决定唯一的一组线性预测系数 $a_k(k=1,2,\dots,p)$ 。

$$\varepsilon = E\{|e(n)|^2\} = \min \quad (4)$$

对语音信号产生机理进行分析可知,语音信号可以看成是一个线性时变系统在准周期脉冲序列或随机噪声激励下所产生的输出,可用图1所示的模型表示。

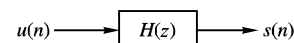


图1 语音信号产生模型

系统函数 $H(z)$ 是一个全极点模型,对于语音信号,确定了各个线性预测系数 a_k 后,可得到其频率响应的估值,即LPC谱:

$$H(e^{-j\omega}) = \frac{G}{1 + \sum_{k=1}^p a_k e^{-j\omega k}} \quad (5)$$

收稿日期:2005-11-25;修订日期:2006-02-22

作者简介:李晋(1981-),男,湖南衡东人,硕士研究生,主要研究方向:语音识别、信号处理;王玲(1962-),女,湖南长沙人,教授,主要研究方向:多媒体信息处理、模式识别。

由式(1)和式(3)可知,如果经 p 阶线性预测器产生的输出与实际语音信号越接近,则 $e(n)$ 就越接近激励信号源,而不包含声道响应信息。因此,LPC相当于一线性滤波器,对语音信号 $s(n)$ 建立一个全极点自回归(AR)模型,使得该模型产生的信号 $\hat{s}(n)$ 和实际语音信号 $s(n)$ 的谱包络相匹配。

2 基于线性预测和最大似然的基音检测算法

本文所提算法的基音周期检测框图见图2。首先经过前置低通滤波器滤掉高频信息和噪声信号,来提高清、浊音判决和基音周期检测的准确性。然后,利用线性预测的方法得到声道冲激响应。最后,通过候选声门激励与声道冲激响应的卷积来重构语音信号的短时频谱,根据最大似然法确定基音周期。

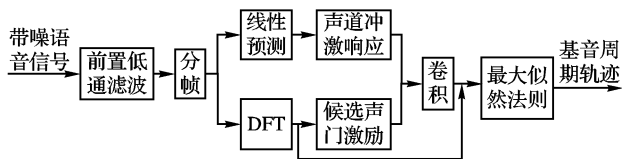


图2 基于LPC和最大似然的基音周期检测框图

2.1 预处理

语音信号中包含丰富的谐波分量,其中多次谐波成分甚至比基波分量还要强,而且声道共振峰有时会严重影响音源的谐波结构,要直接提取出仅与声门振动有关的音源信息变得相对困难。因此,本文采用一个截止频率为800Hz的5阶低通滤波器对去均值的语音信号进行处理,去除高频噪声和语音信号第三、第四个高频共振峰的影响,同时达到不削弱基音频率信息的目的。

2.2 清、浊音判断

对语音信号进行分帧并采用短时平均能量作为判断清浊音的标志:设定一个阈值,当语音帧的短时平均能量大于阈值时,该帧语音判为浊音,对其进行 $p(p=10\sim15)$ 阶线性预测,得到声道冲激响应,并做出LPC谱 $H(e^{-j\omega})$;否则判为清音,基音频率为0。

2.3 频域分块估计候选基音周期范围

由于声门的开闭使得语音信号浊音段呈现准周期性变化,浊音信号的短时频谱具有谐波结构,其谐波峰值包含基音周期的信息。又因基音周期随时间连续变化,其动态范围相对较大,一般在1.25ms~20ms之间变化。为避免候选基音周期数目过多,而使算法计算效率下降,本文提出把搜索范围分块为三个区域,2ms~4ms为儿童和中年女性,3.5ms~10ms为年轻人的一般情况,8ms~20ms为老人,搜索原始语音信号的短时频谱 $S(e^{-j\omega})$ 对应每个区域中频率范围的第一最大峰,并把相应的峰值记录为谐波频率 f_k 。然后,将谐波频率 f_k 转换到时域,得到谐波信号的时域周期 x_k 。最后,将 $\{x_k-1, x_k, x_k+1\}$ 作为基音周期的候选值 t_k ,并进行周期延拓得到声门激励 $u(n)$,加汉明窗进行频谱分析得到 $U(e^{-j\omega})$ 。

2.4 最大似然法确定基音周期

根据候选声门激励 $u(n)$ 与声道冲激响应 $h(n)$ 的卷积来重构语音信号的短时频谱。由时域卷积和频域乘积的关系可得:

$$u(n) * h(n) = \text{IDFT}[U(e^{-j\omega}) \cdot H(e^{-j\omega})] \quad (6)$$

重构的语音信号的短时频谱为:

$$S'(e^{-j\omega}) = U(e^{-j\omega}) \cdot H(e^{-j\omega}) \quad (7)$$

将 $S(e^{-j\omega})$ 与 $S'(e^{-j\omega})$ 归一化后进行相似度比较,若均方误差最小则重构的短时频谱所对应的候选值即为基音周期:

$$\varepsilon(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S(e^{-j\omega}) - S'(e^{-j\omega})|^2 d\omega \quad (8)$$

3 计算机模拟

为了验证本文所提算法的有效性,用Matlab进行仿真,实验所用的原始语音(如图3)是男声普通话“今天偷了菜,明天就想偷米、偷钱”的时域波形,16kHz采样,16bits量化,数据长度共62976个采样点,LPC分析为15阶,帧长取512个采样点(32ms),帧移为256点。这样,即使声带振动频率低至100Hz时,在一帧信号中也能振荡3周左右。

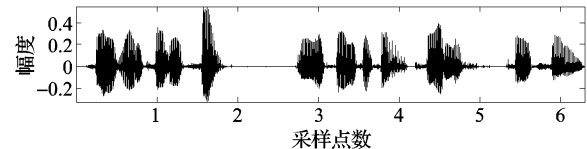


图3 原始语音波形

对语音样本加入均匀白噪声,使其信噪比 $\text{SNR}=3\text{dB}$,波形如图4所示。

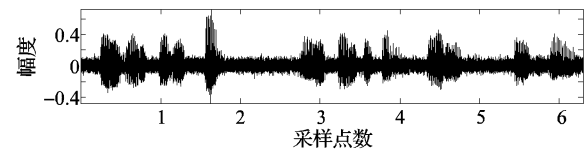


图4 SNR=3dB下语音波形

分别在信噪比 $\text{SNR}=3\text{dB}$ 下用倒谱法、自相关法和本文所提算法进行仿真得到基音周期轨迹曲线如图5~图7。三图比较可知,倒谱法效果最差,ACF法在非语音段会产生识别误差,影响了基音提取精度,而基于LPC和最大似然的检测方法在非语音段几乎不会产生干扰,且有效地减少了基音检测中的随机错误和倍频、半频错误,基音轮廓清晰、平滑,无大的跳变,清音和浊音段区别明显,符合语音信号慢时变性的一般规律,其检测性能明显优于ACF法。由此可见,本文所提算法有较强的抗噪性。

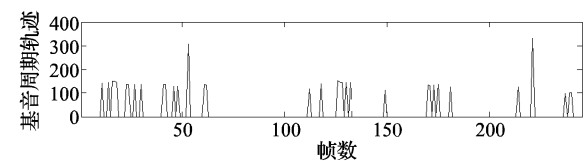


图5 倒谱法

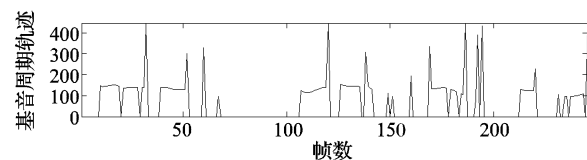


图6 ACF法

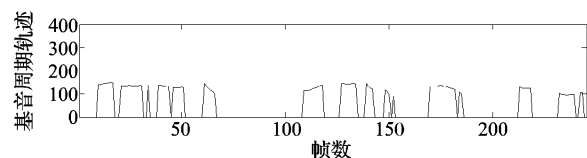


图7 基于LPC和最大似然的检测方法

4 结语

本文提出了一种基于线性预测和最大似然的基音检测算法,该算法从语音信号产生机理出发,采用频域分块估计候选基音周期的范围,结合线性预测和最大似然法来确定基音周

(下转第1236页)

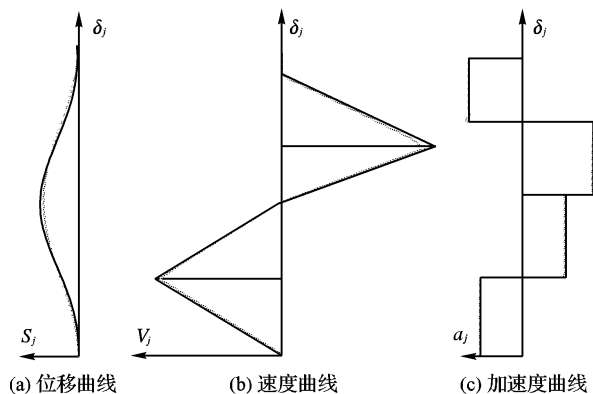


图4 位移曲线、速度曲线和加速度曲线

表1 理论曲线方程系数

	b_0	b_1	b_2
0 ~ 90°	0	0	4.051 796 5
90° ~ 180°	-20	25.464 812 4	-4.052 847 48
180° ~ 250°	-46.122 448 9	42.094 893 97	-6.699 616 11
250° ~ 330°	170.156 25	-59.078 612 34	5.129 384 9
330° ~ 360°	0	0	0

设实际运动规律用二次多项式表示:

$$S_{\text{实}} = b'_0 + b'_1\beta + b'_2\beta^2 \quad (4)$$

根据从动件位移与转角分别对应的值求解出 b'_0 、 b'_1 、 b'_2 , 见表2。

表2 反求曲线方程系数

	b'_0	b'_1	b'_2
0 ~ 90°	-0.001 3	0.002 527	4.051 894
90° ~ 180°	-20.007 267 3	25.472 541 79	-4.055 012 379
180° ~ 250°	-46.218 373 5	42.140 798 11	-6.705 098 768
250° ~ 330°	169.974 168 8	-58.790 155 78	5.096 289 46
330° ~ 360°	0.021 582 765	-0.007 514 586	0.000 652 408

对比式(3)令:

$$\begin{cases} b'_0 = b_0 + \Delta b_0 \\ b'_1 = b_1 + \Delta b_1 \\ b'_2 = b_2 + \Delta b_2 \end{cases} \quad (5)$$

由式(3)~(5)得到误差:

$$\Delta s = \Delta b_0 + \Delta b_1\beta + \Delta b_2\beta^2 \quad (6)$$

表3 反求误差分析

	Δb_0	Δb_1	Δb_2	Δs_{max}
0 ~ 90°	-0.001	0.002	0	0.001 3
90° ~ 180°	-0.007	0.007	0.003	0.036
180° ~ 250°	-0.09	0.045	-0.05	-0.057 7
250° ~ 330°	-0.182	0.28	-0.03	0.047 1
330° ~ 360°	0.021	-0.007	0.000 6	0.004

4.2 从动件为余弦加速度运动规律反求

从动件余弦加速度运动的运动规律为:

$$\text{推程: } S = \frac{h \left(1 - \cos \left(\frac{\pi * \delta}{\delta_0} \right) \right)}{2} \quad (7)$$

$$\text{回程: } S = \frac{h \left(1 + \cos \left(\frac{\pi * \delta}{\delta_0} \right) \right)}{2} \quad (8)$$

根据得到的离散点反求出 h , 将 h 代入方程可以得到余弦加速度运动方程, 误差结果如表4所示。

表4 反求误差分析

	h	Δs_{max}	Δs_{min}
推程	20.000 5	0.000 5	$3.4 * 10^{-7}$
回程	19.997 78	0.002 23	$1.5 * 10^{-6}$
标准	20	0	0

4.3 从动件为摆动运动规律反求

如果凸轮机构的从动件是摆动件, 则只需将位移 S , 速度 V 和加速度 a 替换为角位置 φ 角速度 ω 和角加速度 α , 行程 h 以摆杆的最大摆角 ψ 替换即可。摆角误差为:

$$\Delta\phi \approx \Delta S/l, \quad l \text{ 为摆杆的长度} \quad (9)$$

5 结语

本文提出了利用 CATIA 根据平面盘形凸轮机构的数字模型反求凸轮机构从动件运动规律的方法, 并开发了反求从动件的运动规律算法及程序, 通过实例证明, 提出的方法可以有效地对各种平面凸轮机构进行运动规律反求, 且反求结果的精确度较高。这为重新设计出工作性能达到甚至优于原设计指标凸轮机构的反求设计提供了比较可靠的数据, 且在对进口设备中的凸轮机构进行仿制和反求设计提供了较好的指导作用。

参考文献:

- [1] 盛选禹. CATIA 三维机械设计实例[M]. 北京: 机械工业出版社, 2003.
- [2] 刘影, 杭九全, 万耀青. 反求工程与现代设计[J]. 机械设计, 1998, (12): 1-4.
- [3] 赵韩, 丁爵曾, 梁锦华. 凸轮机构设计[M]. 北京: 高等教育出版社, 1993.
- [4] 谢进, 万朝燕, 杜立杰. 机械原理[M]. 北京: 高等教育出版社, 2004. 98-105.
- [5] Ranghavyacharyulu E. Nonlinear vibration analysis of cam-follower systems with pneumatic coupling[A]. Proc of the 6th world Congress on Theory of Machines and Mechanisms[C], 1983. 1213-1216.
- [6] 刘炳文, 许蔓舒. Visual Basic 程序设计教程[M]. 北京: 清华大学出版社, 2000.

(上接第1233页)

期。通过计算机模拟表明: 与传统方法相比, 该算法的抗噪性能好, 在低信噪比下仍能准确获取基音周期的轨迹曲线, 克服了随机错误和倍频、半频错误, 提高了基音估计精度, 而且算法的复杂度低, 效率高。

参考文献:

- [1] 冯康, 时慧琨. 语音信号基音检测的现状与展望[J]. 微机发展, 2004, 14(3): 95-101.
- [2] ROSS MJ, SHAFFER HL, COHEN A. Average magnitude difference function pitch extractor[J]. IEEE Transactions on ASSP, 1974, 22(5): 353-362.

- [3] 马祥杰, 孟相如, 张百生. 一种自相关基音检测算法[J]. 通信技术, 2003, (5): 9-11.
- [4] NOLL AM. Cepstrum pitch determination[J]. Journal of the Acoustical Society of America, 1967, 41(2): 293-309.
- [5] 张天骐, 张站, 权进国, 等. 语音信号基音检测的二次谱方法[J]. 计算机应用, 2005, 25(4): 934-936.
- [6] 樊桂花, 唐斌. 基于小波变换的语音基音周期鲁棒性检测[J]. 信号处理, 2004, 20(4): 365-368.
- [7] 柏静, 韦岗. 一种基于线性预测与自相关函数法的语音基音周期检测新算法[J]. 电声技术, 2005, (8): 43-46.