

基于相似度的高精度基音检测算法

陈雪勤¹, 刘正², 赵鹤鸣¹

(1. 苏州大学电子信息学院, 江苏苏州 215021; 2. 苏州经贸职业技术学院, 江苏苏州 215009)

摘要: 提出了一种具有较高精度且抗噪性能强的基音检测算法。该算法将线性预测残差看作语音源信号的近似, 对其进行频谱分析, 依据残差幅度谱算得基音周期的粗估值。然后回到时域信号, 根据基音周期粗估值设计一长度可调的窗, 通过窗函数在语音段连续取两段语音信号作相似度运算, 可根据最大相似度值计算出准确的基音周期。该方法准确性高, 在噪声环境下也具有较好的效果。

关键词: 基音检测; 相似度; 线性预测残差信号

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-3630(2008)-05-0704-04

A similarity-based high resolution pitch detection algorithm

CHEN Xue-qin¹, LIU Zheng², ZHAO He-ming¹

(1. School of Electronics and Information Engineering, Soochow University, Suzhou 215021, Jiangsu, China; 2. Mechanical and Electrical Engineering Department, Suzhou Institute of Trade and Commerce, Suzhou 215009, Jiangsu, China)

Abstract: A pitch detection algorithm of high resolution and robust anti-noise is proposed in this paper. Firstly, the linear predictive residual is used as the approximate of original speech signal and is transformed by FFT. Secondly pitch is calculated based on the magnitude spectrum of residual and a window, the length of which could be adjusted, is designed. Finally the more exact pitch is obtained by similarity calculation of two successive speech signal segments chosen by the window in each frame signal. This algorithm is of high resolution and effective in low SNR environment.

Key words: pitch detection; similarity; linear predictive residual signal

1 引言

汉语是一种有调语音, 其声调的主要载体为基音频率。基音频率是发浊音时声带振动的频率, 它是语音时域特征里最重要的参数。而基音检测是语音处理中非常重要的问题, 在语音编码、语音合成、说话人识别、情感识别^[1-4]中占据重要地位。在低速率语音编码中, 准确检测语音信号的基音频率非常关键, 它直接影响到整个声码器系统的性能。近年来人们已从时域、频域和时频混合域出发, 针对不同情况提出了多种有效的基音检测算法。时域算法如自相关法、短时平均幅度差法等, 频域算法如倒谱法等, 时频域算法

有小波变换算法等及其对应的改进算法^[5-8]。

前述时域和频域算法在精度和计算量两者之间往往产生矛盾, 并且受共振峰影响都可发生基音误判。时频混合域算法往往在时域进行基音初估, 再在频域进行基音细搜索, 纯净语音下通常具有很好的性能。由于实际应用中语音易受环境噪声干扰, 因此基音检测技术的研究热点和难点都集中于提高算法抗噪性能。由于大多数基音检测算法针对纯净语音提出, 信噪比较低时算法性能均有明显下降。信噪比越低, 基音误判越严重, 以致后处理措施也无能为力。本文提出的基音检测算法计算量小, 精度高, 可适用于低信噪比环境。

2 相似度函数的定义

对于两个有限长度的信号 $x_1(n)$ 和 $x_2(n)$, 用 S 表示它们之间的相似度, S 应满足条件: ① $0 \leq \frac{|(x_1, x_2)|}{\|x_1\|_2 \cdot \|x_2\|_2} \leq 1$; ② 当 $x_1 = x_2$ 时, $S(x_1, x_2) = 1$; ③ $S(x_1, x_2) =$

收稿日期: 2007-10-07; 修回日期: 2008-01-31

基金项目: 国家自然科学基金(60572076), 江苏省高校自然科学基金(05KJB510113)

作者简介: 陈雪勤(1974-), 女, 江苏扬州人, 硕士, 讲师, 研究方向为语音信号处理。

通讯作者: 陈雪勤, E-mail: chenxueqin@suda.edu.cn

$S(x_2, x_1)$ 。依据已上条件,可定义:

$$S(x_1, x_2) = \frac{|(x_1, x_2)|}{\|x_1\|_2 \cdot \|x_2\|_2} \quad (1)$$

式中, $\|x\|_2 = (x, x)^{\frac{1}{2}}$, (x, x) 为 l^2 上的内积, $(x, y) = \sum_{n=0}^{N-1} x(n)y(n)$ 。由 suchy-Schwarz 不等式可以知道 $|(x, y)| \leq \|x\|_2 \cdot \|y\|_2$, ($\forall x, y \in l^2$), 所以 $0 \leq \frac{|(x_1, x_2)|}{\|x_1\|_2 \cdot \|x_2\|_2} \leq 1$, 即可以满足条件①。在 $x_1 = x_2$ 时, $|(x_1, x_2)|$ 可进一步分解为 $|(x_1, x_2)| = \|x_1\|_2^2 = \|x_2\|_2^2 = \|x_1\|_2 \cdot \|x_2\|_2$, 所以 $\frac{|(x_1, x_2)|}{\|x_1\|_2 \cdot \|x_2\|_2} = 1$, 即 $S(x_1, x_2) = 1$, 满足条件②。由于 $x_1, x_2 \in l^2$, 且 x_1, x_2 都是实数, 所以根据内积性质有 $(x_1, x_2) = (x_2, x_1)$, 也就是 $S(x_1, x_2) = S(x_2, x_1)$, 所以满足上述条件③。

3 基音周期检测方法

本文介绍的基音提取算法可分为以下几步:(1)利用 LPC 残差信号的频谱求基音频率的粗估值;(2)通过对多个粗估值的概率统计,排除倍频的出现;(3)在粗估值的基础上利用时域信号的最大相关度,求得最终基音频率。它对基音进行提取的准确度大于 98%。

3.1 基于残差频谱的基频粗估

线性预测分析的基本思想是:由于语音样点之间存在相关性,所以可以用过去的样点值来预测现在或未来的样点值,即一个语音的抽样能够用过去若干个语音抽样或它们的线性组合来逼近。通过使实际语音抽样和线性预测抽样之间的误差在某个准则下达到最小值来决定唯一的一组预测系数。而这组预测系数就反映了语音信号的特性,可以作为语音信号特征参数用于语音识别、语音合成。

$$\hat{s}(n) = -a_1s(n-1) - a_2s(n-2) - \dots - a_p s(n-p) \quad (2)$$

式中, a_i 是对过去时刻的语音抽样 $s(n-i)$ 加权系数, P 是预测器的阶数,一般 $P=10$ 。真实信号与预测信号的差值即为预测残差 $\varepsilon(n)$ 。

$$\varepsilon(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^P a_i s(n-i) \quad (3)$$

从理论上讲,预测残差信号 $\varepsilon(n)$ 中已不包含声道响应信息,但却完整地包含激励信息。如图 1 所示,与原信号频谱相比较,残差信号的频谱消除了声道变化的信息。因而各次谐波的幅度比较均衡,在此基础上进行基音频率的提取可以消除共振峰的影响。

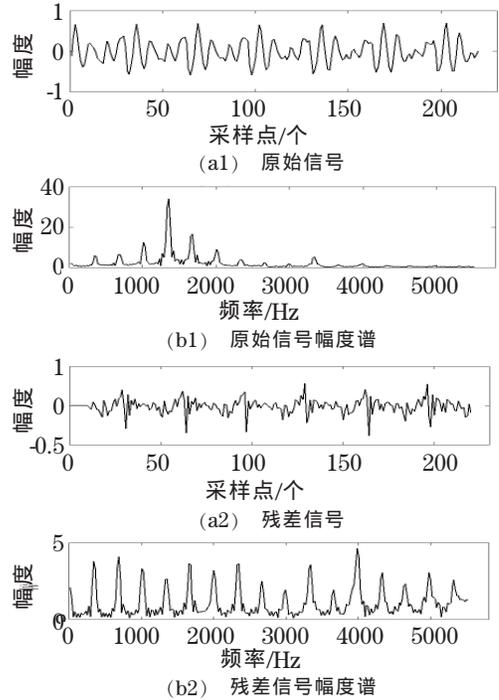


图 1 信号及其残差信号的时频比较

Fig.1 Comparisons of waveform and amplitude spectrum between signal and residual

人的基音频率分布于 50Hz~400Hz, 在采样率为 8000Hz, 512 点 DFT 的情况下, 100 点幅度谱中约可包含 4~25 个峰值, 因此取前 100 点作为分析对象就足够了。为提高准确性,对残差幅度谱作削波(如图 2),对削波后的幅度谱进行峰值提取,在这 100 点中连续取尽可能多的谐波峰值数 N , 其对应的峰值位置以 $P[n], n=1 \sim N$ 表示。相邻峰之间的间距表示为 $Q[n] = Q[n+1] - Q[n], n=1 \sim N-1$ 。由 $Q[n]$ 可计算出各自对应的基频值,如式(4)所示。

$$F[n] = \frac{f_s}{512} \times Q[n], n=1 \sim N-1 \quad (4)$$

观察图 2 可以发现,在提取峰值时有可能取到倍频。那如何才能消除这一现象呢?如前所述,通过多个连续谐波峰值,计算出多个基音频率 $F[n]$,若发现有个别极大值,应将其抛弃,取大概率的基音频率值作为基音频率的粗估值 F ,则基音周期粗估值为 $P = \frac{1}{F}$ 。

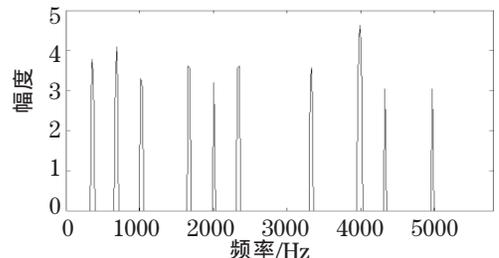


图 2 削波后的残差幅度谱

Fig.2 Amplitude spectrum of residual after clipping

3.2 相似度函数进一步作基音检测

利用 LPC 残差信号的频谱作基频检测,由于中间做了频率变换,受窗长的影响其结果比较粗糙。但是这一基音周期值至少可以给出一个比较可靠的范围。在此基础上对时域信号进一步作最大相似度搜索,可以得到较高精确度的基音周期所对应的采样点数。

这里,定义信号 $x(n)$ 的相似度为:

$$S_m(x_1, x_2) = \frac{|(x_1, x_2)|}{\|x_1\|_2 \cdot \|x_2\|_2} \quad m=1 \sim 31 \quad (5)$$

x_1, x_2 为语音信号 $x(n)$ 中相邻的两段序列, m 为取不同长度 x_1, x_2 时所产生的一组相似度值的序号。在前一节基频初估的基础上, x_1, x_2 长度变化有据可依,有限次实验证明, m 值一般不超过 21, 为保证可靠性,最大值可放宽至 31。结合相似度函数的定义可知,对应于时域语音信号,相似度函数是两段语音的互能量,而 $S_m(x_1, x_2)$ 表征的是语音段的滑动自相关性,如果 $S_m(x_1, x_2)$ 在某处等于 1, 根据随机信号理论,用来计算 $S_m(x_1, x_2)$ 的两个语音段在波形上就相似,另外由于这里参与计算的语音段是相邻的,其幅度基本一致,所以这时可以近似认为短时条件下的语音段具有周期性。实际计算中,一般 $S_m(x_1, x_2)$ 不可能为 1, 但如果语音段为浊音,那么相关系数总会比较接近于 1, 所以可以通过设置门限来判断语音信号在短时条件下是否为浊音,这时最大的 $S_m(x_1, x_2)$ 对应的信号长度就是基音周期。

其中 x_1, x_2 的取值方法如下。设一帧语音 $x(n)$, 根据 3.1 节中残差幅度谱算得基音周期的粗估值为 M (这里的 M 为换算为一个基音周期时域信号所对应的采样点数), 在此基础上,选取一个长度可调的矩形窗 $w(n)$, $n=1 \sim L$, L 为窗长,分别从 $M-D$ 点可调至 $M+D$ 点,共 $2D+1$ 种长度。当窗长为某个 L 值时,令

$$x_1 = x(n) \cdot w(n), n=1 \sim L$$

$$x_2 = x(n) \cdot w(n-L), n=1 \sim L$$

将 x_1, x_2 代入式(5)做相似度计算,可得到一个相似度值,当 L 从 $M-D$ 点调至 $M+D$ 共可获得 $2D+1$ 点相似度值。在窗长恰好与基音周期长度一致时,应该得到一个最大值,此时的长度即为基音周期长度。由此可以算出基音频率。

图 2 中语音信号通过基音粗估法得到 M 等于 32, 在此基础上左右各偏移 8 点,即对不同长度的语音共做了 17 次相似度计算,得到如图 3 所示的相似度曲线,可见在第 10 点(对应 $M=33$)时取得最大相

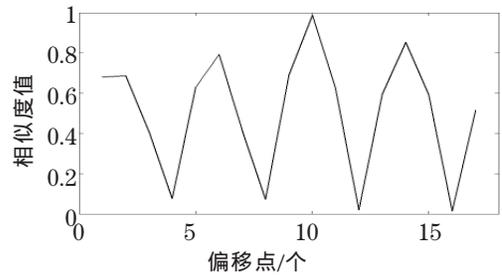


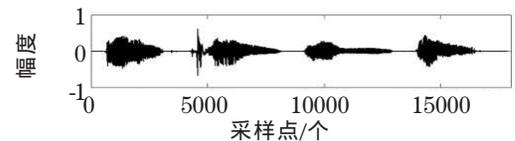
图 3 二段语音的 17 点相似度值

Fig.3 Similarity value of two segments of signals

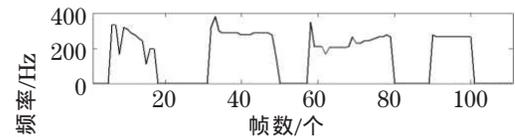
似度。由此可计算出最后基音频率。

4 实验与分析

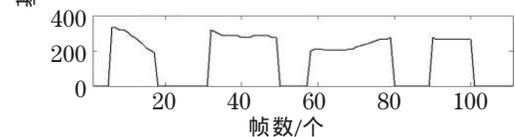
图 4 比较了本文算法和传统自相关法的基频检测效果。分别取纯净语音和信噪比 $SNR=-5\text{dB}$ 时的女声“报刊文摘”为分析对象,通过连续语音的基音检测效果对算法进行评价。由图 4 可见,在纯静的语音条件下,本文算法可准确计算出浊音段的基音



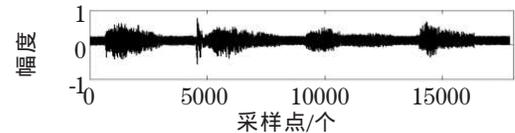
(a1) 纯净语音“报刊文摘”



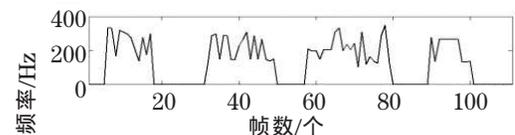
(b1) 传统自相关算法基音轨迹



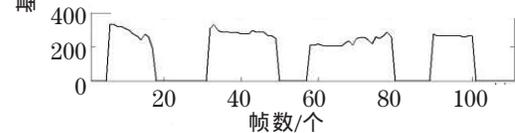
(c1) 本文算法基音轨迹



(a2) 含噪语音“报刊文摘” $SNR=-5\text{dB}$



(b2) 传统自相关算法基音轨迹



(c2) 本文算法基音轨迹

图 4 不同信噪比条件下的两种算法的基频检测效果

Fig.4 Pitch tracks of a speech signal detected with two algorithms for different SNR

表 1 100 个浊音在三种算法下的基频检测误差率比较

Table 1 Error rate comparison of pitch detection of 100 voiced speeches between three algorithms

算法类型	误差率/(%)		
	偏长误差	偏短误差	总误差率
自相关法	4.32	1.38	5.70
AMDF	3.95	1.11	5.06
本文算法	1.05	0.51	1.56

频率,有利于语音的声调检测。而传统自相关法有部分“野点”,主要由倍频或半频产生。但这些野点在连续语音的基音轨迹中可以通过平滑的手段消除。因此在这种情况下,两种算法都可以较好地声调检测提供基频轨迹,但本文算法的准确性更高,在语音合成时会有更好的贡献。在信噪比较低如 $SNR=-5dB$ 时,本文算法依然能够较为准确的提取连续语音的基音频率,而传统自相关所提取的基音轨迹已经不能表现语音的声调信息,即使平滑也无能为力。

采用自相关法、短时平均幅度差法和本文算法对 100 个浊音进行基音频率检测,分别统计各个算法的基频检测的产生误差的概率,总的误差概率由偏长误差概率和偏短误差概率构成。如表 1 所示。

由表 1 可见,本文算法可较大程度降低基频检测的误差率,保证基音频率检测算法的高精度。

5 结 论

本文介绍的算法可获得比传统自相关算法更加精确的基频值,并且可同时实现语音的清浊音判决。这一速度快且精度高的特点对语音声码器合成语音技术颇具意义,可产生更加符合人物个性的语音。同时,由于具备较强的抗噪性、鲁棒性强,可适用于低信噪比的条件下。

参 考 文 献

- [1] McClellan S, Gibson J D, Rutherford B K. Efficient pitch filter encoding for variable rate speech processing[J]. IEEE Transactions on Speech and Audio Processing [1063-6676], 1999, 7(1): 18-29.
- [2] Jong-Soon J, Jeong-Jin K, Myung-Jin B. Pitch alteration technique in speech synthesis system[J]. IEEE Transactions on Consumer Electronics [0098-3063], 2001, 47(1): 163-167.
- [3] BAI Junmei, ZHENG Rong, XU Bo. Robust speaker recognition integrating pitch and Wiener filter[A]. 2004 International Symposium on Chinese Spoken Language Processing[0 7803 8678 7][C]. 2004: 69-72.
- [4] 王治平, 赵力, 邹采荣. 基于基音参数规整及统计分布模型距离的语音情感识别[J]. 声学学报, 2006, 31(1): 28-34.
WANG Zhiping, ZHAO Li, ZOU Cairong. Emotional speech recognition based on modified parameter and distance of statistical model of pitch[J]. Acta Acustica, 2006, 31(1): 28-34.
- [5] 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.
BAI Jing, WEI Gang. A new method for speech signals pitch detection based on LPC and autocorrelation[J]. Audio Engineering, 2005, (8): 43-46.
- [6] 刘建, 郑方, 等. 基于混合幅度差函数的基音提取算法[J]. 电子学报, 2006, 34(10): 1925-1928.
LIU Jian, ZHENG Fang, et al. Combined magnitude difference function based pitch tracking algorithm[J]. Acta Electronica Sinica, 2006, 34(10): 1925-1928.
- [7] ZENG Yumin, TANG Min. Modified cepstrum model based speech pitch detection algorithm[A]. Proceedings of the 16th International Conference on Computer Communication[C]. 2004, 2: 1227-1231.
- [8] 李辉, 戴蓓蓓, 陆伟. 基于前置滤波和小波变换的带噪语音基音周期检测方法[J]. 数据采集与处理, 2005, 20(1): 100-104.
LI Hui, DAI Beiqian, LU Wei. Pitch detection method for noisy speech signals based on pre-filter and wavelet transform[J]. Journal of Data Acquisition&Processing, 2005, 20(1): 100-104.