

文章编号:1004-9037(2014)02-0304-05

一种改进的基音周期提取算法

赵 祎 张 盛 林孝康

(清华大学深圳研究生院深圳市信息科学与技术重点实验室,深圳,518055)

摘要:基音周期的提取在语音信号处理领域有着广泛的应用。受基于归一化自相关函数基音周期提取算法和多带激励(Multi-band excitation, MBE)声码器中基音检测算法的启发,本文提出了一种改进的基音周期提取算法。该算法主要由预处理、时域基音粗估、基音平滑、时变滤波搜索、小数基音周期估计等5个部分组成。实验表明:该算法能达到更高的搜索准确度,得到更加平滑的基音周期曲线;与传统自相关检测算法相比,该算法有很好的抗噪性。

关键词:基音周期;归一化自相关;基音平滑;时变滤波

中图分类号:TN912.3

文献标志码:A

Improved Algorithm for Pitch Period Detection

Zhao Yi, Zhang Sheng, Lin Xiaokang

(Shenzhen Key Laboratory of Information Science and Technology, Graduate School at Shenzhen,
Tsinghua University, Shenzhen, 518055, China)

Abstract: The extraction of pitch period has a wide range of applications in the field of speech signal processing. Inspired by traditional autocorrelation algorithm and the pitch detection method that used in multi-band excitation (MBE) vocoder, an improved algorithm is put forward for pitch period extraction. This algorithm has five parts: pre-process, pitch rough estimation in time domain, pitch smoothing, search with time variable filter, and decimal pitch estimation. Experimental results show that this new algorithm can achieve higher accuracy and compared with traditional autocorrelation algorithm, this approach has a better noise immunity.

Key words: pitch period; normalized autocorrelation; pitch smoothing; time variable filter

引 言

语音信号的大部分能量集中在浊音信号中,浊音信号呈现准周期性。称浊音的周期为基音周期,在时域波形里反映为峰值点和峰值点之间的距离。基音周期的提取在语音合成、语音识别、语音编码等领域有着广泛的应用。由于语音信号变化的复杂性,迄今为止,并未找到一个完善的适用于不同环境 and 要求,不同讲话者的可靠准确的检测基音周期的办法^[1]。

现有的基音周期的检测算法大致可以分为三

类^[2]:波形估计法、相关处理法以及变换法。波形估计法^[3]直接通过语音时域上的波形分析周期峰值,算法简单,硬件实现容易,但语音信号幅度或频率快速变化时,精度下降很明显。相关处理法^[4]通过比较原始语音和原始语音移位后信号间的相似性来确定基音周期。由于它硬件实现较为简单且具有很强的抗波形相位失真能力,在语音信号处理中广泛使用。变换法^[5]通常是将语音信号变换到频域和倒谱域来估计,倒谱参数是一种重要的说话人个性特征参数,主要反映声道响应的特性^[6],效果较好但是算法复杂。

受基于归一化自相关函数基音检测算法^[7]和

多带激励(Multi-band excitation, MBE)声码器中基音检测算法^[8]的启发,本文提出了一种改进的基音周期检测算法,通过计算归一化自相关函数搜索每一语音帧的备选基音周期,利用维特比算法搜索最佳基音转移路径,平滑基音曲线,在时域得到的粗略基音周期的基础上,利用时变滤波器进一步从频域准确化基音周期,在时域利用样条插值搜索小数基音周期。实验表明:该算法能在较低的复杂度下达到较高的搜索准确度,得到平滑的基音周期曲线。与传统的基于归一化自相关函数的基音检测算法相比,该算法有很好的抗噪性。

1 背景算法

1.1 基于归一化自相关函数的基音检测算法

令语音信号表示为 $s(n)$, 语音信号的归一化自相关函数的表示为

$$\rho(\tau) = \frac{\sum_{n=0}^{N-1} s(n)s(n-\tau)}{\sqrt{\sum_{n=0}^{N-1} s^2(n) \sum_{n=0}^{N-1} s^2(n-\tau)}} \quad (1)$$

式中: τ 为移位距离, N 为分析帧长, $\rho(\tau)$ 为归一化自相关函数(NCCF)。根据 Cauchy 不等式, 可以证明当且仅当移位后的语音信号与原始语音信号完全重合时, $\rho(\tau)$ 达到最大, 最大值为 1。

$$\left| \sum_{n=0}^{N-1} s(n)s(n-\tau) \right| \leq \sqrt{\sum_{n=0}^{N-1} s^2(n) \sum_{n=0}^{N-1} s^2(n-\tau)} \quad (2)$$

式(2)表明一段周期信号的归一化自相关函数在其周期点会出现较大的峰值。

自相关方法是目前使用最广泛的基音检测方法, 此算法最大的问题在于, 由于一般的语音浊音段时域波形不是很理想的准周期信号, 它的自相关函数也不会只有一个很明显的最大峰值, 会出现很多峰值点, 寻找的基音周期是其中一个峰值点的位置。如何排除倍频和半频等干扰的影响, 搜索到准确的基音周期是一个棘手的问题。常见的方法是在预处理的时候利用低通滤波或者数值滤波消除谐波或者共振峰的影响, 而这两种线性时不变的滤波方法往往会给基频的提取带来一定的误差。

1.2 MBE 声码器中基音检测算法

MBE 模型中整个基音检测算法的基本思想是通过比较重见谱和原始谱来确定基音周期。可分

为 4 个步骤: 高、低通滤波, 基音粗估, 基音平滑以及基音细搜索。该算法是在频域进行基音检测, 算法准确度高但是计算复杂度也很高。

1.3 语音信号频谱

图 1 为语音信号经过短时傅里叶变换后得到的三维频谱图。图 2 将三维转化成了二维, 图中黑线浓度越高, 代表该时刻该频率点处频谱能量越大。语音信号的频谱是随着时间而变化的, 若用普通的线性滤波器, 是无法准确分离出基频的。因此本文中的算法没有在一开始就进行滤波, 而是在时域粗略估计得到基音周期后利用时变滤波器准确的分离出基波。

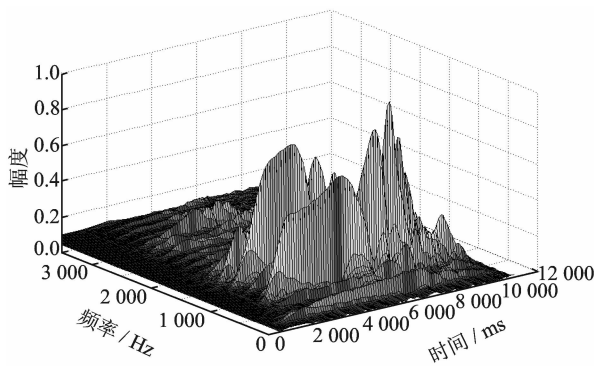


图 1 语音信号短时傅里叶变换三维频谱图

Fig. 1 Example of three-dimensional spectrum for short-time Fourier transform of speech signal

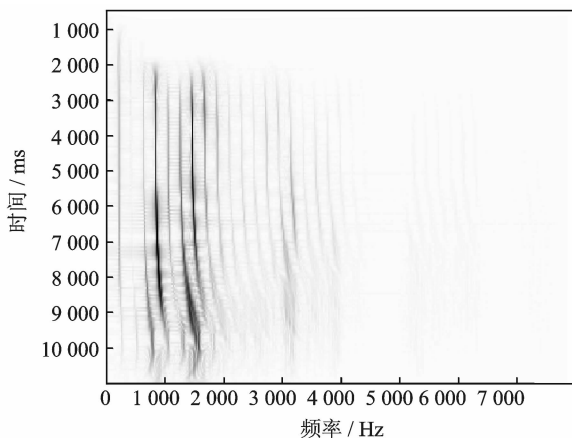


图 2 语音信号短时傅里叶变化二维频谱图

Fig. 2 Example of two-dimensional spectrum for short-time Fourier transform of speech signal

2 算法实现

本文中的算法主要分为 5 个部分: 去均值, 时域基音周期粗估计, 基音平滑, 时变滤波搜索以及

小数基音周期搜索。算法流程图如图 3 所示。

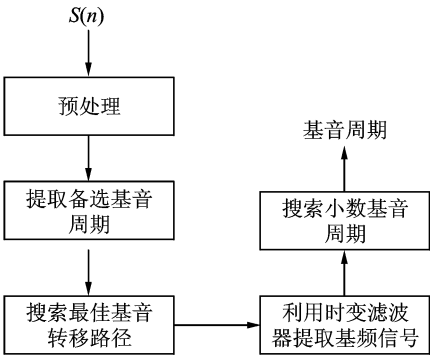


图 3 算法流程图
Fig. 3 Flow chart of algorithm

2.1 去均值

当语音信号在分析窗里有非零均值或有非常低的低频噪声出现时,归一化自相关函数在所要求的所有延迟上都产生高的相关^[8]。鉴于此,在计算 $\rho(\tau)$ 时首先减掉均值,减去均值的信号为

$$s'(n) = s(n) - \mu \tag{3}$$

其中

$$\mu = \frac{1}{N} \sum_{n=0}^{N-1} s(n) \tag{4}$$

2.2 时域基音周期粗估计

首先归一化自相关函数的计算方法如式(1)所示。为了能准确地提取基音周期,自相关函数截取的点数,即位移应该在两个周期以上。由自相关函数法计算出的峰值基本上能反映实际的基音周期。

由于语音信号呈现准周期性,在与该帧波形具有较强相似性的地方, $R(\tau)$ 会出现峰值,然而这些峰值中究竟哪个才是准确的周期位置,还需进一步判断。因为这些峰值可能对应基音周期的整数倍或分数倍,也可能是清浊音过渡阶段的干扰所造成。

为防止峰值点过多影响计算效率,可设立一个门限 V_{th} ,当且仅当峰值的大小超过该门限时,记录其位置。由于对应准确基音周期的自相关函数值较大,并且归一化自相关函数最大值为 1,所以,可以忽略掉那些大小不超过 0.5 的峰值点;同时,该门限大小应当与该帧归一化自相关函数的峰值的具体大小有关。假设某帧语音信号出现峰值点的位置为 τ_1, τ_2, τ_3 ,对应的峰值大小为 $R(\tau_1), R(\tau_2), R(\tau_3)$,门限 V_{th} 应当不小于前 n 个峰值的均值, n

的具体取值可根据精度和运算量大小的要求来定,如式(5)所示

$$V_{th} = \max\{0.5, \frac{1}{n} \sum_{i=1}^n R(\tau_i)\} \tag{5}$$

记录所有大于门限的峰值点,作为该帧的备选基音周期。基音提取的目标转为从这些峰值中选出最合理的一个。如果所有的峰值大小都低于 0.5,可以认为该帧周期性很弱,有可能是清音帧,备选基音周期标记为空,最佳基音周期直接设为 0。

2.3 基音平滑

借鉴文献[9]中动态规划的思想,这里采用维特比算法搜索最佳基音周期。维特比算法的关键就是确定转移路径的权值,通过局部最优来达到全局最优。

如何确定相邻帧不同基音之间的转移概率?假设当前帧第 i 个备选基音周期的归一化自相关函数的值为 $R(\tau_i)$,前一帧第 j 个备选基音周期的归一化自相关函数的值为 $R(\tau_j)$ 。由于语音信号的平稳性,基音周期曲线也应该是平滑的^[10]。相邻帧之间的基音周期应当是连续的。由于基音周期一般不会发生突变,在度量转移概率时,周期的值越接近,转移的概率就越大。另外转移概率应当与前一帧基音值本身大小有关^[11],因此可定义这两个备选基音周期之间的路径转移权值为

$$W_{ij} = \frac{abs(R_1(\tau_i) - R_2(\tau_j))}{R_1(\tau_i)} \tag{6}$$

每次搜索需要前溯几帧?一般来说,每次规划参与的帧的数目应当综合考虑计算复杂度和精确度的要求。不宜过多,过多有可能导致一味的追求平滑而忽略了基音偶尔有可能存在的跳变。也不宜过少,过少会导致平滑效果不理想。一般来说,以 4~6 帧为宜。

备选基音状态为空的帧怎么处理?由于基音曲线的平滑仅是针对连续浊音帧来讲^[12],遇到备选基音状态为空的帧,直接将该帧最佳基音周期设为 0。将其之后的第一个不为空的帧作为起始帧处理。

2.4 时变滤波搜索

根据时域搜索得到的初步的基音周期,设计时变滤波器,追踪基频的变化轨迹。

短时傅里叶变换可以展示信号瞬时频谱随时间变化的关系。设某帧语音时域搜索得到的基音周期为 T_0 ,对该帧语音做短时傅里叶变换,短时傅

里叶变换以 T_0 为长度,如式(7)所示,第 i 点的短时傅里叶变换为

$$SIFT[m,k]=\sum_{n=i-\lfloor\frac{T_0}{2}\rfloor}^{i+\lfloor\frac{T_0}{2}\rfloor-1}s[n]g[n-m]\exp\left(-j\frac{2\pi kn}{T_0}\right)\quad (7)$$

式中: $m\geqslant0;k<T_0$ 。

其中 $g[n]$ 时域表达式如式(8)所示,对于不同语音帧, $g[n]$ 是随着基音周期的变化而变化的

$$g(n)=u(n)-u(n-T_0)\quad (8)$$

根据式(7)得到的语音信号的频谱设计时变滤波器^[13],追踪基频的轨迹。对残留的频谱做傅里叶反变换得到时域信号 $s'(n)$,计算 $s'(n)$ 的归一化自相关函数,记录峰值位置作为较为准确的基音周期。

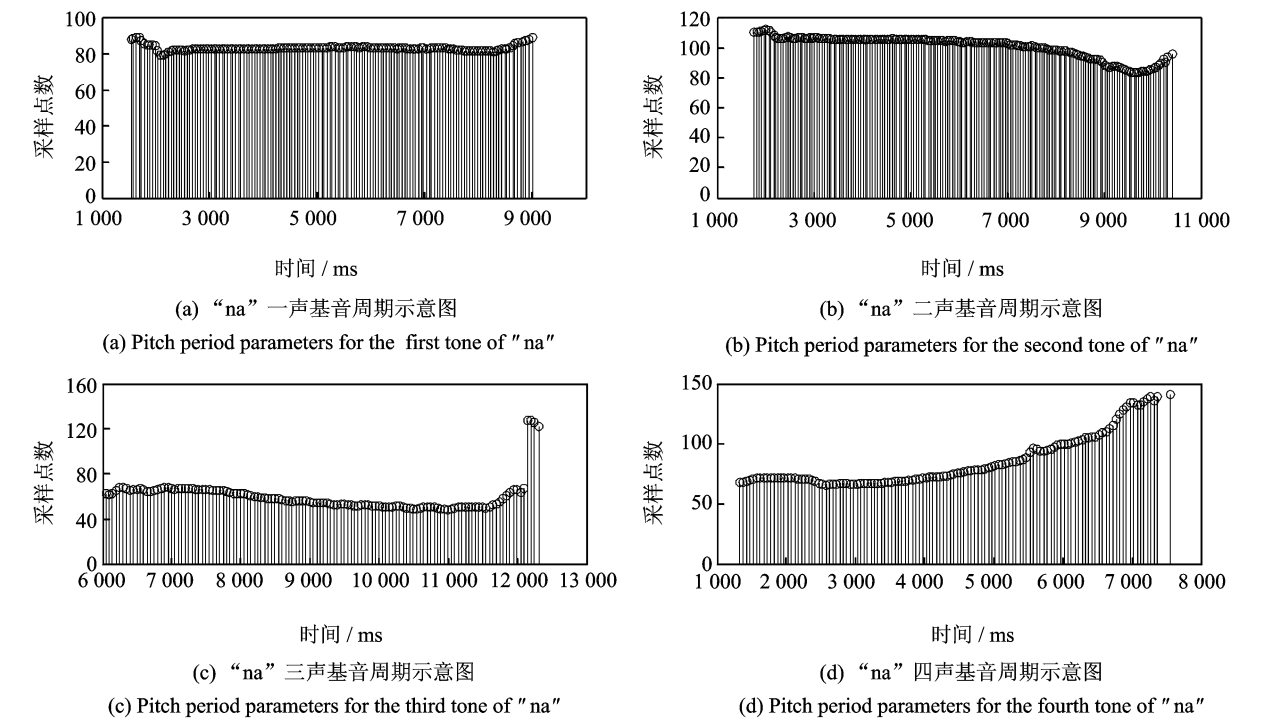


图 4 汉语普通话“na”4 个声调基音周期图
Fig. 4 The four tones of "na" in Mandarin

3.2 准确性和抗干扰性分析

在语音信号波形上叠加不同功率的白噪声,测试该算法提取效果。首先分别对 50 个纯净的语音文件用本文中提到的算法和传统短时自相关法检测其基音周期,各自得到相应的两个标称值,然后混入不同功率的白噪声,分别再用这两种方法检测。混入噪声后检测到的值若与原信号检测值相比误差小于 80% 即判为正确检测。实验结果表明,当信噪比在 20 dB 以上时,两种算法抗干扰性几乎没有什么差别。而当信噪比低于 15 dB 时,本

2.5 小数基音周期搜索

在搜索得到整数基音周期以后,为了提高精确度,还需要确定小数基音周期。假设搜索到的整数基音周期为 T ,可以在 $[T-1:T+1]$ 的范围内进行样条插值,然后卷积,依旧是搜索出现最大峰值点的位置作为最佳小数基音周期。时域样条插值相当于提高了时间域的分辨率,而频域仅仅是拓展了频谱周期,不改变频谱形状,也不改变谐波位置。

3 实验结论

3.1 实验结果示例

以“na”这个音的 4 个声调为例,对该算法进行测试,测试结果列在图 4 中。从图中可以看出,该算法能得到较为平滑的基音周期曲线。

文中的算法要明显优于普通自相关法(见表 1)。
表 1 本文算法与普通自相关法抗噪性对比
Table 1 Noise resistance comparison between proposed algorithm and traditional autocorrelation algorithm

SNR/dB	准确率/%	
	本文算法	普通自相关法
20	97.13	96.92
15	96.66	92.71
10	95.42	89.56
5	95.15	86.97

4 结束语

本算法抗干扰性、准确度明显高于传统的自相关法,但是本算法所占存储空间要多,复杂度也较传统的自相关法高。在本算法中,采用了维特比算法进行最佳路径搜索,有效地平滑了基音周期;同时利用时变滤波器过滤频谱,再通过滤波后的频谱对应的时域信号搜索最佳基音周期,有效提高了搜索的准确度。利用时域插值搜索小数基音周期,在较小的计算量的前提下有效提高了计算结果的精度。另外,该算法中时变滤波的思想可应用于语音信号谐波的提取中。

参考文献:

- [1] Zeng L, Chen L, Xiao Q. Pitch period estimation base on voiced degree weighted sub-frame octave region dynamic programming[C]//Wireless Communications and Signal Processing (WCSP), 2010 International Conference on. [S.l.]: IEEE, 2010:1-5.
- [2] Rabiner L, Cheng M, Rosenberg A, et al. A comparative performance study of several pitch detection algorithms[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1976, 24(5): 399-418.
- [3] Yang Y, Zhang H, Guo X. A pitch tracking method mixing ACF & AMDF algorithms based on correlations [C]//Image Analysis and Signal Processing (IASP), 2011 International Conference on. [S.l.]: IEEE, 2011:553-556.
- [4] So Y J, Jia J, Cai L H. Analysis and improvement of auto-correlation pitch extraction algorithm based on candidate set[C]//Recent Advances in Computer Science and Information Engineering. Berlin, Heidelberg: Springer. 2012:697-702.
- [5] Bahja F, Elhaj I, Hassan E, et al. On the use of wavelets and cepstrum excitation for pitch determination in real-time [C]//Multimedia Computing and Systems (ICMCS), 2012 International Conference on. [S.l.]: IEEE, 2012:150-153.
- [6] 丁琦,平西建. 针对语音变换的语音篡改检测[J]. 数据采集与处理, 2012, 27(1): 57-62.
Ding Qi, Ping Xijian. Speech tampering detection for voice transformation[J]. Journal of Data Acquisition and Processing, 2012, 27(1): 57-62.
- [7] 鲍长春,樊昌信. 基于归一化自相关函数的基音检测算法[J]. 通信学报, 1998, 19(10): 27-31.
Bao Changchun, Fan Changxin. Pitch detection algorithm based on normalized cross-correlation function [J]. Journal of China Institute of Communications, 1998, 19(10): 27-31.
- [8] Hu J, Xu S, Chen J. A modified pitch detection algorithm[J]. Communications Letters, IEEE, 2001, 5(2): 64-66.
- [9] Kawahara H, Masuda-Katsuse I, de Cheveigné A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds[J]. Speech communication, 1999, 27(3): 187-207.
- [10] Zhao X, O'Shaughnessy D, Minh-Quang N. A processing method for pitch smoothing based on autocorrelation and cepstral F0 detection approaches [C]//Signals, Systems and Electronics, 2007. ISSSE'07. International Symposium on. [S.l.]: IEEE, 2007: 59-62.
- [11] 朱小燕,王昱,刘俊. 汉语声调识别中的基音平滑新方法[J]. 中文信息学报, 2001, 15(2): 45-50.
Zhu Xiaoyan, Wang Yu, Liu Jun. An approach of fundamental frequencies smoothing for chinese tone recognition [J]. Journal of Chinese Information, 2001, 15(2): 45-50.
- [12] 计哲,李晔,崔慧娟,等. SELP 声码器基音周期参量化合成改进算法[J]. 高技术通讯, 2010, 20(1): 45-48.
Ji Zhe, Li Ye, Cui Huijuan, et al. An improved algorithm for pitch quantization in SELP vocoders[J]. Chinese High Technology Letters, 2010, 20(1): 45-48.
- [13] Benisty H, Avargel Y, Cohen I. Adaptive system identification using time-varying Fourier transform [C]//Applications of Digital Information and Web Technologies. [S.l.]: IEEE, 2009: 652-657.

作者简介:赵玮(1991-),女,硕士研究生,研究方向:低速率语音编码技术的硬件实现, E-mail: zhaoyi11@mails. tsinghua. edu. cn; 张盛(1975-),男,副教授,研究方向:数字对讲机 SOC 系统; 林孝康(1947-),男,教授,研究方向:移动通信。

