

◎数据库、信号与信息处理◎

幅值补偿的 AMDF 基音周期检测算法

康广玉¹, 郭世泽², 孙圣和¹KANG Guang-yu¹, GUO Shi-ze², SUN Sheng-he¹

1. 哈尔滨工业大学 自动化测试与控制系, 哈尔滨 150001

2. 总参 54 所, 北京 100001

1. Department of Automatic Test and Control, Harbin Institute of Technology, Harbin 150001, China

2. No.54 Institute from Headquarters of the General Staff, Beijing 100001, China

KANG Guang-yu, GUO Shi-ze, SUN Sheng-he. Amplitude compensation based AMDF pitch detection algorithm. Computer Engineering and Applications, 2009, 45(28): 125-127.

Abstract: Traditional AMDF and its improved algorithm LV-AMDF based pitch detection method easily causes the double or half error. For this problem, this paper analyzes the characters of functions of AMDF and LV-AMDF and their shortages on pitch detection, and then proposes amplitude compensation based AMDF to detect the pitch. This algorithm decreases the occurring of double-frequency and half-frequency errors and increases the detection precision. The experiments show that the proposed algorithm performs better on pitch detection compared with AMDF and LV-AMDF methods.

Key words: pitch detection algorithm; Average Magnitude Difference Function (AMDF); Length-Varied AMDF; amplitude compensation

摘要:传统的基于 AMDF 及其改进算法 LV-AMDF 容易导致检测基音周期时产生加倍、减半等错误。针对该现象, 分析了 AMDF、LV-AMDF 函数的特性及其用于基音检测时存在的不足, 提出了幅值补偿 AMDF 算法来检测基音周期, 降低了基音检测中经常出现的半频、倍频错误, 提高了检测的精度。实验表明该方法基音检测性能优于 AMDF 和 LV-AMDF 的方法。

关键词:基音检测; 平均幅度差函数 (AMDF); 变长 AMDF; 幅值补偿

DOI: 10.3778/j.issn.1002-8331.2009.28.037 文章编号: 1002-8331(2009)28-0125-03 文献标识码: A 中图分类号: TP391

1 引言

语音信号按声带振动与否可分为清音和浊音, 浊音具有准周期性。基音周期作为一个重要的特征参数, 在语音识别、话者识别、语音合成等多个领域有着广泛的应用^[1-3], 因此基音周期的检测方法在语音信号处理领域一直是研究热点。基音周期的范围很宽, 加之容易受到发音音调和情绪等因素的影响, 因此快速有效地检测基音周期是一个十分复杂的问题。针对浊音信号的准周期性, 通常采用短时处理技术, 常用的短时基音周期检测方法如 AMDF 基音检测算法^[1-3]。

AMDF 方法具有运算量小、精度高等优点, 在语音编码中得到广泛应用。实验表明, 当语音信号的变化比较平缓时, 采用 AMDF 进行基音周期估计可以得到较高的估计精度和可靠性, 所需的计算量相对自相关函数法而言也较小。但是当语音信号的幅度或基音周期快速变化时, AMDF 法的基音估计精度会明显下降, 从而直接影响汉语声调识别的准确率^[4]。如经常发生基音估计结果为实际基音周期的 2 倍或 1/2 的情况^[1], 针对该问题, 相关文献提出了不少改进方法^[1-3]。文献[6]提出了 (Length-Varied AMDF) 方法, 对频率快速变化产生的基音检测错误进行了修正。但也存在一些问题, 如基音周期值对于语音帧的前

部分比较吻合, 而后面的偏差就越来越大。对清浊混合帧 (或浊音的起始帧) 的计算结果偏差较大等。

语音信号是一个时变的非稳定信号, 幅值随着时间逐渐衰减。针对此现象, 提出采用幅值补偿 AMDF 算法, 降低由于幅值衰减而造成的基音周期检测错误。实验表明该方法与传统的 AMDF 和 LV-AMDF 方法相比, 基音检测的精度有所提高。

2 算法比较

AMDF 基音检测算法的定义为:

$$D(k) = \sum_{n=0}^{N-k-1} |s_w(n+k) - s_w(n)| \quad (1)$$

$$s_w(n) = s(n) \cdot w(n)$$

其中 $s(n)$ 为语音信号, $w(n)$ 为滑动窗。

基于短时 AMDF 函数的基音周期初步估计为:

$$TP = \arg \min_{k=TP_{\min}}^{TP_{\max}} D(k) \quad (2)$$

其中 TP_{\max} 和 TP_{\min} 分别为预先设定的最大、最小基音周期。

短时 AMDF 函数随着滞后时间的延长, 峰值幅度逐渐下降^[3]。这使得谷值点检测以及谷值点的清晰度检查比较困难, 加

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.60772074)。

作者简介: 康广玉 (1975-), 男, 博士生, 主要研究方向为语音处理、语音合成。

收稿日期: 2008-11-03 修回日期: 2009-01-22

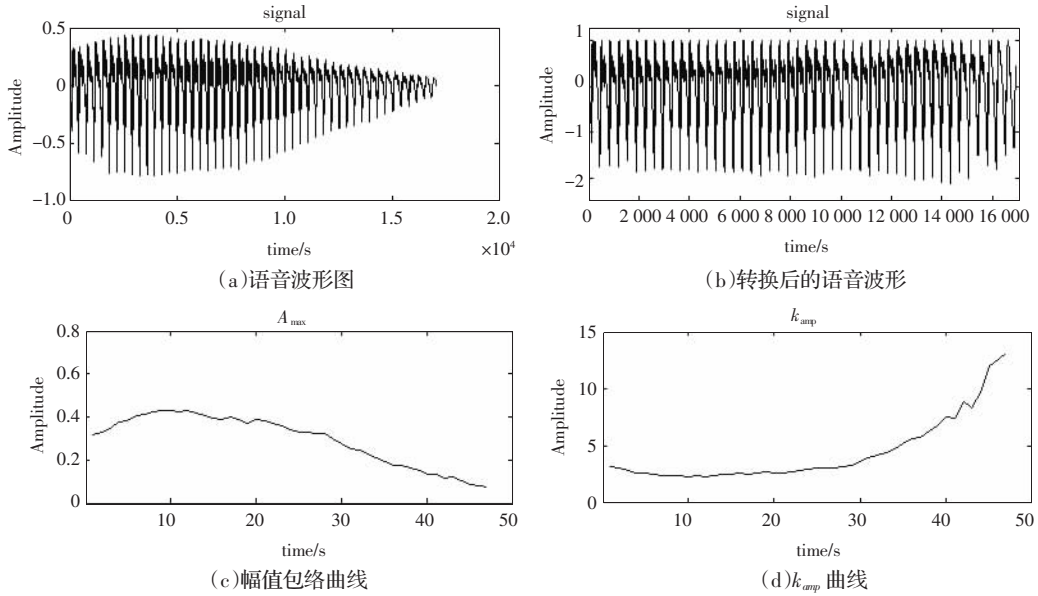


图1 k_{amp} 的计算和幅值转换结果

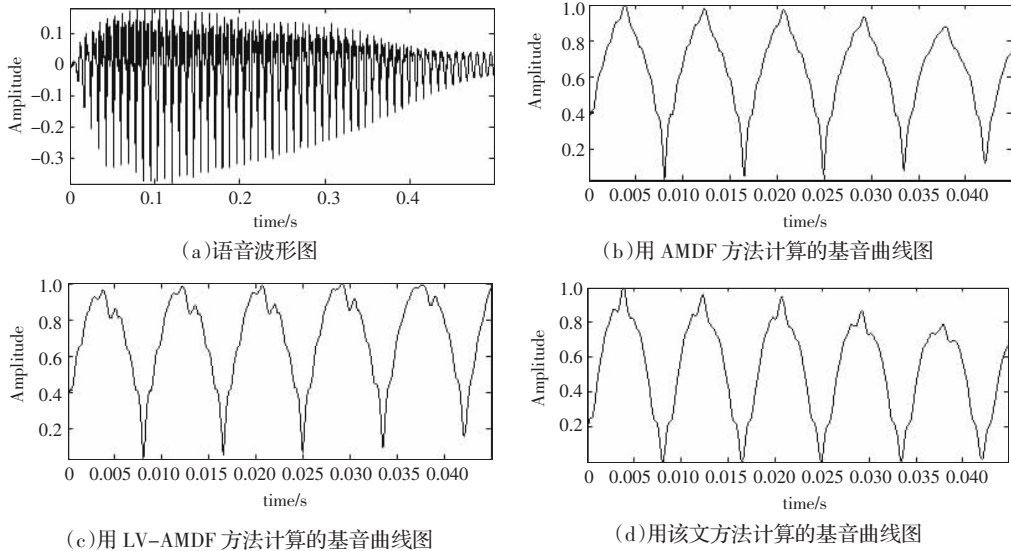


图2 各种算法的 $D(k)$ 波形

倍误判、减半误判的概率比较高。

LV-AMDF 基音检测算法的定义为:

$$D(k) = \frac{\sum_{n=0}^{N-k-1} |s_w(n+k) - s_w(n)|}{\frac{1}{2} \sum_{n=0}^{2k-2} s_w(n)} \quad (3)$$

样本幅度差的计算范围 n 是随延时而变化的, 且延时越长, 计算范围越宽。为了排除窗宽的影响, 式(3)引入了分母项 $\frac{1}{2} \sum_{n=0}^{2k-2} s_w(n)$, 作用是使能量归一化。这种算法可以消除基音周期快速变换造成的错误检测。

该算法具有诸多优点, 但是也存在两个不足的地方:(1) LV-AMDF 受语音帧首部的影响比较大。基音周期的估计值对于语音帧的前部分比较吻合, 而后面的偏差就越来越大。这容易造成以部分帧(前一部分)的估计结果代表整个语音帧的情况, 使估计精度下降。对于以静音开始的语音帧, 还会出现函数分母为零的奇异情况。而且加倍误判的情况比较严重, 对清浊

混合帧(或浊音的起始帧)的估计结果通常还是实际值的好几倍。这种浊音开始部分的估值偏离, 由于缺少先前的参考信息, 更难以纠正^[7]。

幅值补偿 AMDF 算法的定义为:

$$D(k) = \sum_{n=0}^{N-k-1} |s_w(n+k) - k_{amp} s_w(n)| \quad (4)$$

其中 k_{amp} 的计算方法如下: 将语音信号用滑动窗分割, 窗长应大于一个基音周期, 找出窗内语音最大值点 A_{max} , 组成语音信号的幅值包络, 计算其倒数组成 k_{amp} 矩阵, $s_w(n) * k_{amp}$ 使得语音信号由衰减信号变为等幅信号, 如图 1 所示。

各种基音检测算法计算结果如图 2 所示。

3 实验和分析

算法流程如图 3 所示。

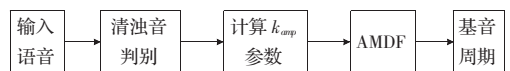


图3 基音检测算法流程图

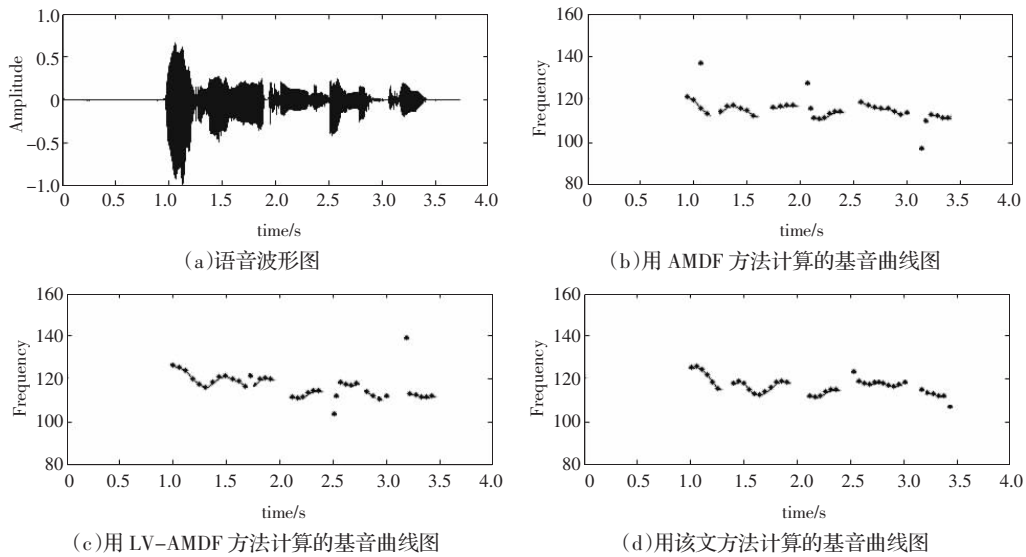


图4 各种算法计算的基音频率曲线图

具体算法如下:

(1)语音信号清浊音判别。

(2)计算 k_{amp} 曲线。

(3)进行幅值补偿 AMDF 计算。因基音周期不会突变,只是在平均基音周期附近移动。所以以后只计算基音周期的 70%~130%这段语音,可减少 40%的计算量。

(4)继续计算直到浊音结束。

实验中所用的说话人语音样本均使用 Sennheiser-e828s 传声器和 Aureon 7.1 Space 专业数字音频卡采集,其采样频率为 PCM 22.05 kHz,16 bit 量化。录音在普通实验室安静环境下进行,该文的实验样本选择了男女各 2 名年龄 22~24 岁身体健康的大学生进行实验,录制了他们在平静时汉语普通话,发音约 1 min,包括词组、句子等。

为了定量描述实验的结果,以手工标记的基音频率为参考标准,选择了两个指标^[5]。

(1)基频错误率

$$e(n) = \frac{e_d(n) - e_p(n)}{e_p(n)} \quad (5)$$

其中 n 表示第 n 个周期, e_d 是用各种基音提取方法得到的结果, e_p 表示基准值(人工检测结果)。当 $50\% \geq |e(n)|$ 或 $|e(n)| \geq 100\%$ 样点时被称为无效检测点。无效基频错误率定义为无效检测帧占总帧数的比例。利用这个指标可以描述算法对语音过渡段和含噪语音的检测能力。

(2)有效基频标准差

$$D = \frac{\sqrt{\sum_{n=1}^N [e_d(n) - e_p(n)]^2}}{N} \quad (6)$$

其中 N 表示总帧数, e_d 是用各种基音提取方法得到的结果, e_p 表示基准值(人工检测结果)。利用这个指标可以描述算法检测结果的稳定性。

实验表明该文方法检测精度高于其他方法(AMDF、LV-AMDF)。同时在不同发音环境和发音方式下,各种方法检测

出的有效基频能够保持基本稳定,而该文方法则更接近于标准值。

表1 无环境噪声基音检测结果

	基频错误率/(%)	有效基频标准差
AMDF	2.6	11.3
LV-AMDF	1.8	10.7
该文方法	0.9	9.6

4 结论

提出了一种幅值补偿 AMDF 方法,通过对语音信号幅值衰减进行补偿,实现基音检测,降低了半频和倍频的错误率,提高了基音检测准确率。同时作为一种检测精度较高的基音检测算法,能够达到实时检测的要求。

参考文献:

- [1] Yang Min, Yang Ying-chun, Wu Zhao-hui. A pitch-based rapid speech segmentation for speaker indexing[C]//Seventh IEEE International Symposium on Multimedia, 2005:6.
- [2] 邵艳秋,韩纪庆,刘挺,等.自然风格言语的汉语句重音自动判别研究[J].声学学报,2006,31(3):203-210.
- [3] Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound[C]//Proc of the Institute of Phonetic Sciences of the University of Amsterdam, 1993, 17:97-110.
- [4] Kobayashi H, Shimamura T. A modified cepstrum method for pitch extraction[C]//The 1998 IEEE Asia-Pacific Conference on Circuits and Systems, 1998:299-302.
- [5] 钱博,李燕萍,唐振民,等.一种基于线性预测残差倒谱的基音检测算法[J].计算机工程与应用,2007,43(32):210-213.
- [6] 顾良,刘润生.高性能汉语语音基音周期估计[J].电子学报,1999,27(1):8-11.
- [7] 张文耀.循环 AMDF 及其语音基音周期估计算法[J].电子学报,2003,31(6):886-890.