

# 一种语音变速不变调处理算法的FPGA实现

The FPGA Implementation of Algorithm for Voice Speed Changing

西安交通大学微电子研究所 宋红花 陈贵灿 (西安 710049)

**摘要:** 文章介绍了一种对语音进行变速不变调处理的系统。该系统与ADPCM编解码技术相结合,能更精确地检测和分割“音元”,并通过音元的复制和抽取实现了对播放语速的控制,达到了变速不变调的目的。该系统用FPGA实现,结果表明:采用改进后的音元处理算法,可以大大减少语音处理中所引入的噪声;该算法与ADPCM相结合实现的语音变速系统,具有速度快、占用资源少、芯片面积小和成本低等特点。

**关键词:** 音元,语速控制,自适应差分脉冲编码调制

## 1 引言

语音的变速处理是目前十分流行的一项语音处理技术,即通过改变单位时间内输出的语音信息量,达到改变语速的要求。人们有时希望能加快或减慢语音的播放速度,尤其是减慢语速,对于外语学习等方面有很大的帮助。

在改变语音播放速度方面,传统的技术通常是通过改变放音机的走带速度来实现的。但这种方法会导致放音磁头输出信号频率变低,在改变语速的同时使音调、音色发生变化,甚至无法识别的语音内容。为了达到变速但不变调的效果,可以采用已经成熟的语音信号处理方法,如利用自相关函数、短时平均幅度差函数<sup>[1]</sup>、倒谱法、小波变换等构成的基音检测法,以及基于参数编码的LPC算法等<sup>[2,3]</sup>。但这些方法在单芯片硬件设计和实现语音变速不变调系统时比较繁琐。

本文根据音元概念<sup>[4]</sup>准确地识别和分割音元,并通过对音元的复制和抽取,实现了与ADPCM编解码相结合的语音变速不变调的算法设计和FPGA实现。文献[4]中所描述的算法,只是将语音信号看作短时的、近乎理想的阻尼振荡波来处理,尤其是音元的峰值检测算法中,该文献假设语音信号两个相邻零点之间只有一个极值,从而把零点后出现的第一个极值作为峰值。而实际上,语音信号是在不断波动的,两个相邻零点之间总是存在多个极值点,因此采用该文献的算法,就必定会使阻尼波的各个峰值检测不准确,从而导致较大的音元判断和语音分割误差,最终在语音信号中引入很大的噪声。本文在分析实际语音信号波形特征的基础上,总结出其波动规律,对峰值检测算法进行了改进,增加多个极值比较的过程,提高了峰值检测的准确度,大大地减小了噪声,使语音效果得到了明显的改善。

在应用方面,文献[5]采用音元的抽取和复制同时实现语音的压缩/解压缩功能与变速功能,这样的处理不仅使系统的速度慢,压缩比小,占用的资源多,而且不利于系统的单芯片实现。本文将基于音元的语音处理方法与ADPCM编解码相结合。在录音模式下,在对原始的PCM信号进行ADPCM编码的同时,对其进行峰值检测和音元判断,从而找出并存储语音分割点的相对地址;在放音模式下,在对ADPCM码进行解码的同时,将数据存储器地址与语音分割点的相对地址进行比较,从而只通过地址的重复和删减便可实现语音的变速要求。这样的处理,其优点在于:(1)利用了ADPCM编解码压缩比高,复原性好的特点。(2)与ADPCM编码相结合,音元检测法只需对原始的PCM信号进行实时处理,通过简单的地址操作,便可实现语音的变速不变调处理。这种处理,算法简单直观,速度快,占用资源少,更适合于集成电路的实现。

## 2 音元检测与变速系统

### 2.1 音元概念及音元检测的实现

音元可以被认为是组成语音信号的基本单位,任何语音信号都是由音元构成的。从物理上讲,任何的发声行为都有振动过程,音频信号就是发声系统机械振动的电表现,其波形与机械振动波相对应。所以在很小的时间间隔内,它可以被看作是由微小的阻尼振动波所组成的。这样微小的、具有实际含义的一段阻尼振动波就被称为音元<sup>[4]</sup>。根据阻尼振动理论,阻尼振动波是逐渐衰减的,所以在同一音元中,各极值总体上也是衰减的,可以用阻尼振动包络线方程来描述。

众所周知,人的语音信号是一种典型的非平稳信号。但是由于语音的形成过程是与发音器官的运动密切相关的,因此语音信号常常可假定为短时平稳的,即在10~20ms这样的时间内,其频率特性和

某些物理特征常量是不变的。因此我们可以定义包含了声音的最基本的信息的音元的长度一般为0.1~20ms。

结合音元概念和语音信号的特点,我们可以在准确划分各个音元的基础上,有规律地重复或抽取单个或多个音元的内容来实现语速的变化,而不改变原有语音信号的频谱特性,这就是我们采用的变速不变调方法的理论基础<sup>[4]</sup>。

以音元为单位进行语音分割时,通常以幅值为零的某个特殊点作为分割点。这样进行变速处理,在插入或删除了音元后仍可以保证音频波形之间能光滑地连接,不会使声音产生突变。

音元检测的实现是在频率为8kHz的采样时钟控制下,通过顺次比较每个输入的PCM编码的值来进行峰值检测,并得到最大极值点,进而寻找出最大极值点前的第一个零点位置,作为新的音元的开始和前一个音元的结束。峰值检测通常采用单边极值比较中的正极大值比较法来进行<sup>[4]</sup>。

由于实际的语音信号是不断波动的,因此两相邻零点之间会存在多个极值。但是在文献[4]中,在顺次比较相邻两个零点之间的输入PCM编码时,只要判断当前的PCM码值小于前一个PCM编码,便认为前一个PCM值为峰值,没有考虑此极值之后(在同极性方向上)将会出现更大的极值点的情况,从而把相邻两个零点之间的第一个波动对应的极值作为峰值,造成了峰值检测中很大的误差。图1就是本文在考虑了以下列举的非理想情况,对峰值检测算法进行了改进之后,所得到的峰值检测状态机。这里列举两种常见的非理想情况。

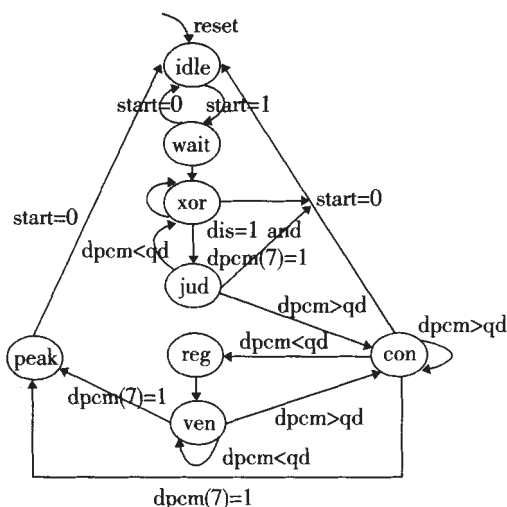


图1 峰值检测状态机

(1) 对于实际语音中出现的(1,-1, 0,-1, 0,-1)情

况,常用的峰值检测会出现错误。我们通过重复的异或检测来修正——即在达到jud状态时,如果dpcm小于qd,并不立即认为qd为峰值,而是回到xor状态重新进行异或检测。

(2) 实际语音中,在两个零点之间存在波动(如0, 1, 2, 1, 3, 4, 2, 2, 0, 7, 8, 9, 4, 3, 2, 1),对于这样的采样值,用通常的波形检测会得到第三个pcm值2为极值,从而发生错误。对此,我们规定:在相邻两个零点之间只有一个最大的正峰值。我们除了进行重复的异或检测,找到相邻两零点之间的所有极值外,还设计了一个峰值寄存器,在reg状态对其赋值。这样就可以在多个极值中准确地找出最大的一个作为峰值。

如图1的状态机中,dpcm、qd和dis分别表示输入语音信号、其延迟后的信号和相邻两个PCM码的最高位异或信号。在各状态中,xor状态表示开始进行正的峰值检测,peak表示已经找到了一个最大极值。采用此算法,可以大大地提高峰值检测以及语音分割的准确度,从而减少了噪声的引入。

## 2.2 语音变速的实现

基于以上的音元检测,我们可以在音元(或音元的整数倍)之间插入或抽取若干个类似相邻音元,达到语音变慢或变快的效果,实现变速不变调的目的。

因为人的耳朵如同眼睛有视觉残留一样也有听觉残留现象,因此允许在音元分割点之后插入一段空白信号。插入的信息单元可以是如下几种<sup>[4]</sup>:(1)空白信号;(2)插入点前一小段信号;(3)经衰减的插入点前一小段信号。经过比较和验证,插入空白信号后的效果较其它两种方法略差,我们采用第二种方法取得了满意的效果。

对一段语音信号以音元作为基本分割单位进行分段,分成的每一小段可包含有1~10个音元。我们的实验证明,1~2个为最佳。本文以两个音元( $S=2$ )为一小段进行分割。图2(a)为 $S=2$ 时的语音分割状态机。其中,st2, st3, st16和st19表示进行峰值检测,st12, st13, st15和st21都表示完成了一个音元的检测,需要记录前一邻近零点地址,不同之处在于,它们分别对应实际语音四种不同波形下检测到音元的情况。检测到每个音元后即进入st17,判断音元计数是否等于2,如果不是,则继续判断下一音元;如果是,则此时音元的开始端便是语音的分割点,从而保证了分割后的每段语音包括两个音元。在st18状态下,同时生成分割点地址标志信号并存

储该分割点地址。这里我们设计分割点地址为前一地址的相对值,这样可以降低地址存储器的容量,从而提高速度。图 2(b)为相应的输出仿真波形。其中 mark 信号表示包含两个音元的语音小段分割点相对地址值。

考虑到在实际的语音信号中,由于噪声等因素,有时会出现同一音元中极值并非严格逐渐递减的情况。图 2 中,我们采用的方法是,如果后一极值大于前一极值的某一预定量时,才确认后一极值

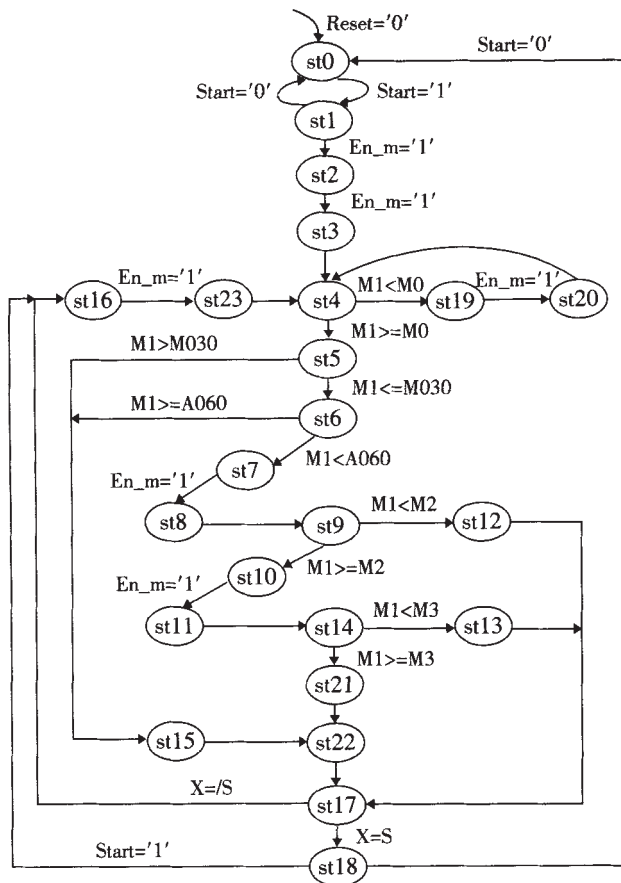


图 2 (a) 语音分割点状态机

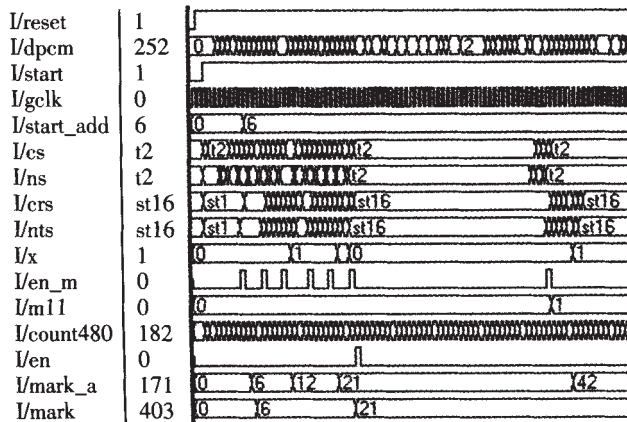


图 2 (b) 输出分割点相对地址波形图

为新音元的最大极值,否则,仍认为是同一音元,继续比较后一极值。我们采用的预定量为前一极值的 30%。采用这种方法可消除噪声等因素对音元分割的影响。图 2(a)中, st3 和 st17 分别进行 A0 的 60% 和 M0 的 130% 计算,以便在 st4 进行比较。

### 2.3 自适应差分脉冲编码调制(ADPCM)

ADPCM (adaptive difference pulse code modulation) 是一种已经非常成熟的语音信号编码技术,它综合了 APCM 的自适应特性和 DPCM 系统的差分特性,是一种性能比较好的波形编码。它的核心思想是<sup>[3]</sup>:①利用自适应的思想改变量化阶的大小,即使用小的量化阶(step-size)去编码小的差值,使用大的量化阶去编码大的差值;②使用过去的样本值估算下一个输入样本的预测值,使实际样本值和预测值之间的差值总是最小。本文采用 ADPCM 编解码算法对 PCM 编码进行 4 倍压缩的编码转换,编码结果存入编码数据存储器。

### 2.4 系统整体结构

基于音元的语音分割与 ADPCM 组成的语音变速不变调处理系统的整体框图如图 3 所示。

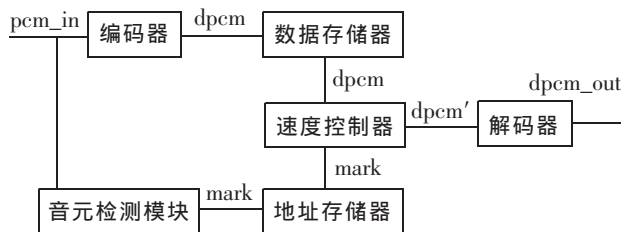
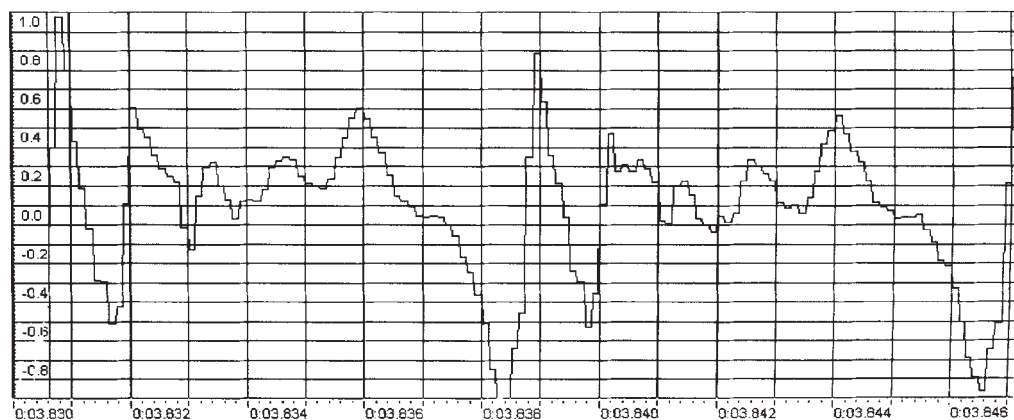


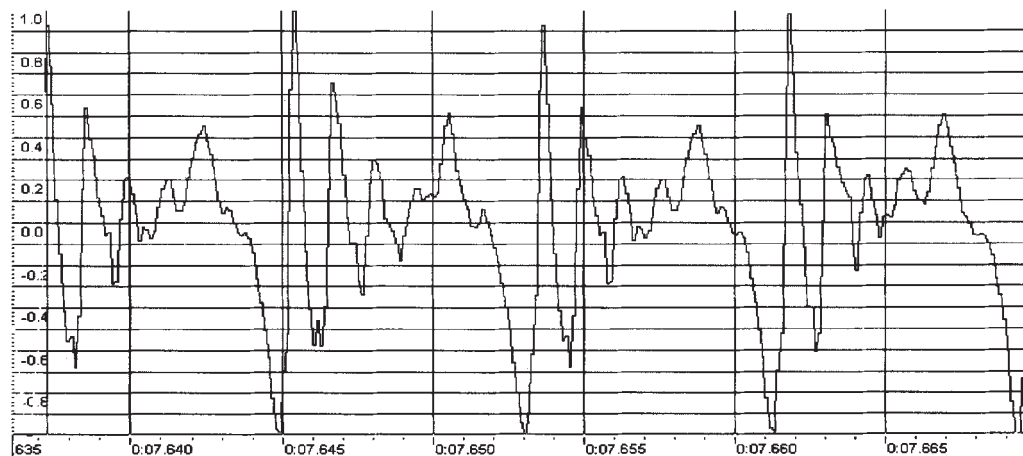
图 3 语音变速系统框图

在录音模式下,语音峰值检测和分割点地址存储均与 ADPCM 编码同时进行,64kb/s 的 PCM 码同时进入编码器和音元检测模块,编码后生成 16kb/s 的 ADPCM 码将存入数据存储器,而音元检测模块产生的分割点地址将存入地址存储器。

在变速放音模式下,音元的复制和抽取亦是与 ADPCM 解码同时进行,并且只通过地址操作来实现变速,通过速度控制器产生的经过音元复制或抽取的新的 ADPCM 码送入解码器,解码后生成的数据流便是经过变速后的语音波形。下面以 0.5 倍速放音为例。在从 ADPCM 编码存储器中读取数据进行解码的同时,将这些原始 ADPCM 数据的值顺序存入另一中间数据存储器,当所读取的 ADPCM 编码存储器的地址与地址存储器中的分割点地址相等时,便将中间数据存储器的数据重复读取和解码一遍,实现音元的复制。由于音元分割地址为相对



(a) 原速放音时的语音波形



(b) 0.5 倍速放音波形(插入音元)

图4 语音变速波形图比较

地址,因此每次音元复制结束后,地址和中间数据存储单元清零。这样重复上述步骤进行每一小段语音的复制,便可实现不改变原始语音的频率,达到慢速放音的目的。

### 3 实验结果与电路实现

本文的设计采用 VHDL 语言,并用 Modelsim 进行了功能验证,后用 Maxplus II 进行了时序验证。图4为本设计的验证结果,它是将仿真得到的二进制语音数据经 Matlab 转换为波形文件格式后,用 Goldwave 软件播放得到的。图4(a)为录入的一段正常速度语音波形,图4(b)为插入前一音元后的信号,将其变慢一倍后的波形。从图中可以看出,后者除了时间上变为前者的两倍外,语音波形基本保持一致,并无明显的噪声引入。实验结果表明,采用本文的处理方法可以在编解码的同时实现语音播放的变速处理。包括 ADPCM 编解码与变速处理的整个系统以 FPGA 实现,共用了 12900 个门。

### 4 结束语

本文改进了音元检测的算法,实现了采用 ADPCM 编解码的语音信号的变速处理。在采样频率为 8KHz,工作模式为录音/放音模式下,针对不同变速要求进行了逻辑功能仿真、时序仿真以及算法综合和 FPGA 实现。结果表明,该系统达到了良好的变速不变调的功能。

### 参考文献

- [1] Baba H, Onishi N, Sakashita Y, Tokizawa H, Tanaka H. Development of a Voice Speed Converting System LSI, Consumer Electronics. Proceedings of International Conference on, Jun 1995.
- [2] 曹志刚, 钱亚生. 现代通信原理. 清华大学出版社, 1992.
- [3] 易克初, 田斌, 付强. 语音信号处理. 国防工业出版社, 2000.
- [4] 苏勇. 音频信号保真变速处理方法. 国际专利 PCT(申请号 PCT/CN96/00074).

(下转第 62 页)



- 2001.9
- [5] 黄文梅等,系统仿真分析与设计,国防科技大学出版社,2001
- [6] 林闯,计算机网络和计算机系统的性能评价,清华大学出版社,2001
- [7] 雷擎,王行刚,应用网络仿真技术进行网络性能评价,计算机应用,2001.12
- [8] 雷擎,王行刚,网络设计与研究集成环境的设计与实现,计算机研究与发展,2002.2
- [9] 冯径,马小骏,面向对象的网络需求分析工具研究与实现,小型微型计算机系统,2000.7
- [10] 顾伯萱,冯径,企业集成网络性能评价方法的研究,小型微型计算机系统,1998.12
- [11] 邹玲,石坚,贺聿志,网络设计与仿真方法的研究,计算机工程与应用,2000.6
- [12] IMiguel Angel Olabe, Juan Carlos Olabe, Telecommunication Network Design Using Modeling and Simulation, IEEE Transaction on education, vol 41, no. 1, february 1998
- [13] S.C.Lin, N. McKeown, "A Simulation Study of IP Switching," Proceedings of Sigcomm' 97, France, 1997.9
- [14] OPNet, <http://www.opnet.com>
- [15] NS-2, <http://www-mash.cs.berkeley.edu/ns/>
- [16] Scheduling Algorithms for Input-Queued Cell Switches, Nick McKeown Ph.D. dissertation, Univ. California, Berkeley. CA. May 1995.
- [17] Nick McKeown, Fast Switched Backplane for a Gigabit Switched Router, <http://www.cisco.com>
- [18] M. Karol; M. Hluchyj; S. Morgan, "Input versus output queueing on a space-division switch," IEEE Transactions on Communications, vol. 35, pp. 1347-1356, Dec 1987.
- [19] N. McKeown, et al. "The Tiny Tera: A small high-bandwidth packet switch core," IEEE Micro, Jan-Feb 1997.
- [20] 江勇,吴建平,徐恪,高性能交换体系结构及其调度算法分析,电子学报,2001年11月
- [21] S.Y.Wang and H.T. Kung, A Simple Methodology for Constructing Extensible and High-Fidelity TCP/IP Network Simulators, IEEE INFOCOM' 99, March 21-25, 1999
- [22] S.Y. Wang, C.L. Chou, C.H. Huang, T.H. Yan, C.C. Hwang, Z.M. Yang, C.C. Chiou, and C.C. Lin, The Design and Implementation of the NCTUns 1.0 Network Simulator, <http://NSL.csie.nctu.edu.tw/nctuns.html>
- [23] Harvard TCP/IP network simulator 1.0, <http://www.eecs.harvard.edu/networking/simulator.html>
- [24] Douglas E.Comer,用TCP/IP进行网际互联 第一卷,第二卷,电子工业出版社,2001
- [25] Internet骨干路由器及发展中的Internet设计, <http://www.juniper.net>
- [26] Internet路由器体系结构的发展及评估大型路由器设计时需考虑的基本可扩展性和性能问题, <http://www.cisco.com>
- [27] 徐恪、熊勇强、吴建平,宽带IP路由器体系结构分析,软件学报,2000年2月
- [28] 梁阿磊,彭路,胡越明,路由器体系结构及其发展,小型微型计算机系统,2001年第22卷第8期
- [29] M160 Internet骨干路由器体系结构, <http://www.juniper.net>
- [30] Cisco 12000千兆比特交换路由器, <http://www.cisco.com>
- XIE Yong-liang, WANG Xing-gang (ICT, Chinese Academy of Sciences, Beijing, 100080)

**Abstract:** Network simulation is a important method for network system design. Because current network simulation tools can't satisfy the requirement of network system design and performance evaluation, We have developed a new kind of network simulation tools NDS-2 oriented to network system design and performance evaluation, which consists of 3 parts: protocol simulation, equipment simulation and application simulation. In this paper, we will present the details of the system design and implementation. Concurrently, we will point out the next research point.

**Keywords:** network simulation, system design, performance evaluation, packet switch architecture

解永良 男, (1974-), 博士研究生, 研究方向为信息网络系统设计, 网络模拟。

王行刚 男, (1937年-), 研究员, 博士生导师, 研究方向为信息网络系统设计。

(上接第 82 页)

- [5] 苏勇, 沈雪良. 一种音频信号的压缩/解压缩方法 (申请号 971064024).

SONG Hong-hua, CHEN Gui-can (Institute of Microelectronics of Xi'an Jiaotong University, Xi'an 710049)

**Abstract:** This paper presents a system of voice speed process-

ing. It incorporates ADPCM codec and allows us to change the voice speed by copying and cutting the voice elements without altering the voice pitch. We complete the algorithm design and FPGA implementation. It is proved that our improvement for the voice speed changing method based on voice elements is valid and available.

**Key words:** Voice element, Voice speech changing, ADPCM