

基于 DSP 的语音处理方法研究与实现

Study and Realization Based on DSP for Speech Processing

刘纪红^{1,2} 徐心和³ 倪 艳³

(东北大学信息科学与工程学院 沈阳110004)¹

(北京大学视觉与听觉信息处理国家重点实验室 北京100871)²

(东北大学信息科学与工程学院 沈阳110004)³

Abstract Digital Signal Processor is developed rapidly in recent years. In this paper, authors design and realize a software system for speech processing based on TMS320C5402 DSK which is developed by Texas Instrument Corp. This system could finish the A/D, filter and spectrum analysis for the speech signal. Authors also analyze the related arithmetic using the theory of DSP.

Keywords DSP, Speech processing

1 引言

数字信号处理技术已经广泛地应用于数字通信、雷达、遥感、声纳、语音合成、图像处理、测量与控制、高清晰度电视、数字音响、多媒体技术、地球物理学、生物医学工程、振动工程,以及机器人等各个领域。随着科学技术的发展,其研究领域和应用领域还将不断地发展和扩大。

数字信号处理(DSP)芯片,是近几年发展起来的适于高速运算的集成电路。基于它开发的应用系统,由于具有高速、便携、可靠、软件可升级等突出特点,而成为近几年的研究热点。

数字信号处理技术之所以发展得这样快,应用得这样广,是与它的突出优点分不开的。归纳起来,它有以下4个方面的优点:(1)精度高;(2)灵活性大;(3)可靠性高;(4)时分复用。

数字信号处理技术的实现方法,可以分为3类:(1)软件实现;(2)硬件实现法;(3)软、硬件结合实现法。本论文用到的是软、硬件结合实现法,即利用数字信号处理器(DSP 芯片),通过配置的硬件和编

程,实现所要求的数字信号处理任务。

目前,基于 DSP 芯片的语音和图像处理软、硬件开发是众多工程人员的研究焦点。随着计算机性能价格比的不断提高以及有关数字处理方法的发展,数字化的语音和图像处理技术在科学研究上、工业生产上或管理部门中都得到越来越多的应用。由于语音和图像处理的数据量大,数据处理相关性高,并且具有严格的帧、场时间限制,因此如何针对语音和图像处理的特点对 DSP 进行优化编程,充分发挥其性能就成为提高整个系统性能的关键。本文在语音处理方面做了一些工作。

TMS320系列是美国 Texas Instrument 公司推出的 DSP 芯片,该系列包括16位定点运算和32位浮点运算两大类芯片。它们非常适合应用在高速控制和矩阵数学运算中。该系列内部采用高度并行处理设计,具有丰富灵活的指令集、运算速度快、性能价格比高等特点。从1982年该公司推出第一代 DSP 芯片——TMS320C10以来,已发展出5个大类:C1X, C2X, C3X, C4X, C5X。C5X 是一种定点运算的 DSP 芯片类型,很适合于语音和图像处理。它有很高的运

- 2 李英. 基于 Agent 的预测与交通控制系统研究:[博士学位论文]. 天津:天津大学,2000
- 3 徐晋晖. Agent 模型与联盟机制研究:[博士学位论文]. 北京:清华大学,2000
- 4 UNY CAO Y. Cooperative Mobile Robotics: Antecedents and Directions. Autonomous Robotics, 1997, 4
- 5 Parker. Heterogeneous Multi-Robot Cooperation. www.ai.mit.edu/publications/pubs/Db/search
- 6 Shoham Y. Agent-Oriented Programming. Artificial Intelligence, 1993, 51~92
- 7 Wooldridge M. This is MYWORLD: The Logic of an Agent-Oriented DAI Testbed. In: M. J. Wooldridge, N. R. Jennings, eds. Intelligent Agents, ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Proceedings. Berlin: Springer, 1995, 160~178
- 8 Alonso E. How individuals negotiate societies. In: ICMAS98.

- 1998, 18~25
- 9 Wooldridge M, Jennings N R. Formalizing the cooperative problem solving process. Readings in Agents, 1996, 430~440
- 10 Dunin-Keplicz B, et al. Compositional Formal Specification of Multi-agent Systems [J]. Lecture Notes in AI, 1994
- 11 薛宏涛,等. 多智能体系统体系结构及协调机制研究综述. 机器人, 2001, 1
- 12 周洪玉,等. 基于 Agent 的多机器人动态协作机制. 机器人技术与应用, 2001, 6
- 13 www.robot.sjtu.edu.cn/robocup
- 14 Wooldridge M, Jennings N R. Intelligent Agents: Theory and Practice. The Knowledge Engineering Review, 1995, 10 (2): 115~152
- 15 李春,等. 协作多机器人系统的模块化设计与实现. 机器人, 1999, 3

算速度和大的存储空间、很高的性能价格比、广泛的应用领域和成熟的开发技术。本论文工作采用 TMS320C5402 DSP 芯片。

2 研究平台和系统功能

本论文工作所用的硬件系统是 TI 公司的 TMS320C5402 DSK(DSP Starter Kit)板,软件平台

是 TMS320C5402 DSK 板的集成开发环境 CCS (Code Composer Studio)和 matlab,所应用的 PC 操作系统是 windows 98.编程语言是 C、汇编语言。

Code Composer Studio 扩展了一套具有调试及实时分析能力的基本代码生成工具。利用 Code Composer Studio 进行 DSP 软件开发的过程如图1所示。

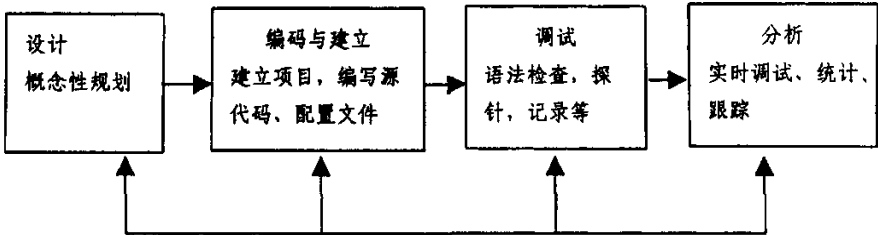


图1 利用 CCS 进行 DSP 软件开发的过程

3 系统功能的设计实现

本论文是在 TMS320C5402 DSK 上实现语音信号的采集、滤波和频谱显示。所实现的语音信号处理系统的功能框图如图2所示。下面逐一介绍。

3.1 语音处理功能的总体设计



图2 语音信号处理系统的功能框图

A/D 和 D/A 部分：
由于 TMS320C5402 DSK 板自带 TCL30244 A/D 和 D/A 芯片,我们对其进行初始化编程,并对其进行调用就可以完成这部分工作。语音信号可以两种方式进入系统,一种是直接通过麦克(MIC)实

时录入,另外一种是通过音频线传入声音文件。在最后的实现中,我们选定后者,因为这种方法可以提供比较稳定的语音信号文件,以便我们后来的实验分析。

FIR 功能部分：

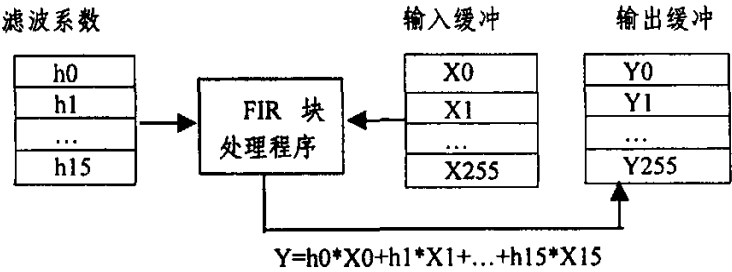


图3 16点 FIR 块滤波处理函数的数据处理流程

16点 FIR(有限长冲击响应滤波器)滤波处理函数的数据处理流程如图3所示。为了运算方便,已将输入数据 X(n)进行了倒序处理。
滤波部分的工作主要是两个方面:一是软件方面的 FIR 算法如何实现,另一是硬件方面对输入输出数据流的时序控制。我们所用的软件开发环境对 DSP 有一些特殊指令,其中就有我们要用到的 FIR 指令。在软件编程实现中,我们充分利用了这一优化指令来实现滤波功能。在配合使用这个函数的过程

中我们用到了 TMS320C5402 DSP 芯片的两个硬件性质:DMA(直接存取存储器)和 McBSP(多通道缓冲串口),通过它们来控制输入输出数据流。
输出部分功能:由于语音信号具有时变性,且每次的平均速度或过程中的瞬间速度也各有所不同,在语音信号处理中,若采用时域特性来识别,可能会带来较大的困难。所以我们在实验中,通过频谱图像显示形式,将处理前后的数据做 FFT 变换并显示出来。实践证明这样做,不仅可以使实验结果很直接,

对于我们做理论分析和实验结果分析也是很有帮助的。

3.2 语音采集和滤波器的理论设计

语音处理系统首先要将模拟的语音信号采样量化为数字量,而要得到不失真的语音信号,必须要根据采样定理,选择合理的采样频率。

1. 采样定理

我们从音频线上采集来的数字信号,对其进行滤波,频域处理,最后得到我们要的特征信息。为了后面数据的不失真性,信号在采样时就应该注意到采样频率的问题。在后面频谱显示部分也涉及到采样频率。

采样定理要求:一个频谱受限的信号 $f(t)$, 如

果频谱范围为: $-f_c \sim +f_c$, 则要使信号 $f(t)$ 不失真, 采样频率 f_s 必须大于等于 $2f_c$ 。参看图4, 在数字域下只有满足 $\omega_s \geq 2\omega_m$ 的条件, $F_s(\omega)$ 才不会产生频谱的混叠。这样, 抽样信号 $f_s(t)$ 保留了原连续信号 $f(t)$ 的全部信息, 完全可以由 $f_s(t)$ 恢复出 $f(t)$ 。图4中, 示出了当抽样频率 $\omega_s > 2\omega_c$ (不混叠时) 及 $\omega_s < 2\omega_c$ (混叠时) 两种情况下冲击抽样信号的频谱。

为了较准确地反映原语音信息, 由于人的语音信号主要集中在 $300\text{Hz} \sim 3000\text{Hz}$, 我们用 8kHz 作为采样频率这样是可以满足奈奎斯特采样定律的。这一点将在我们后面对频域的处理和显示中有所体现。

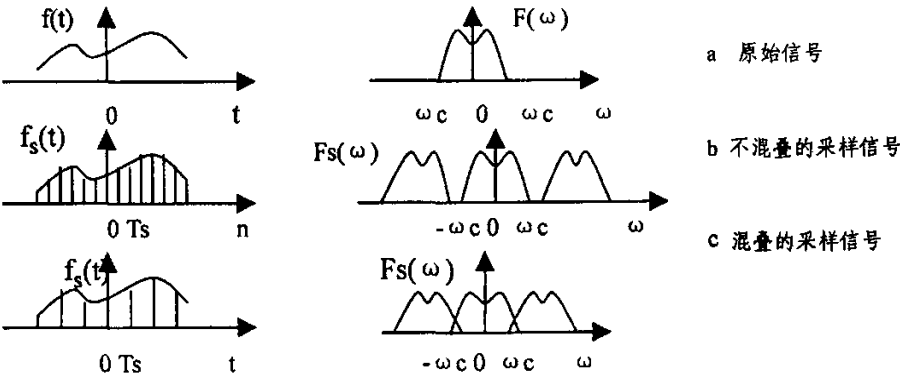


图4 不同的采样频率的效果图

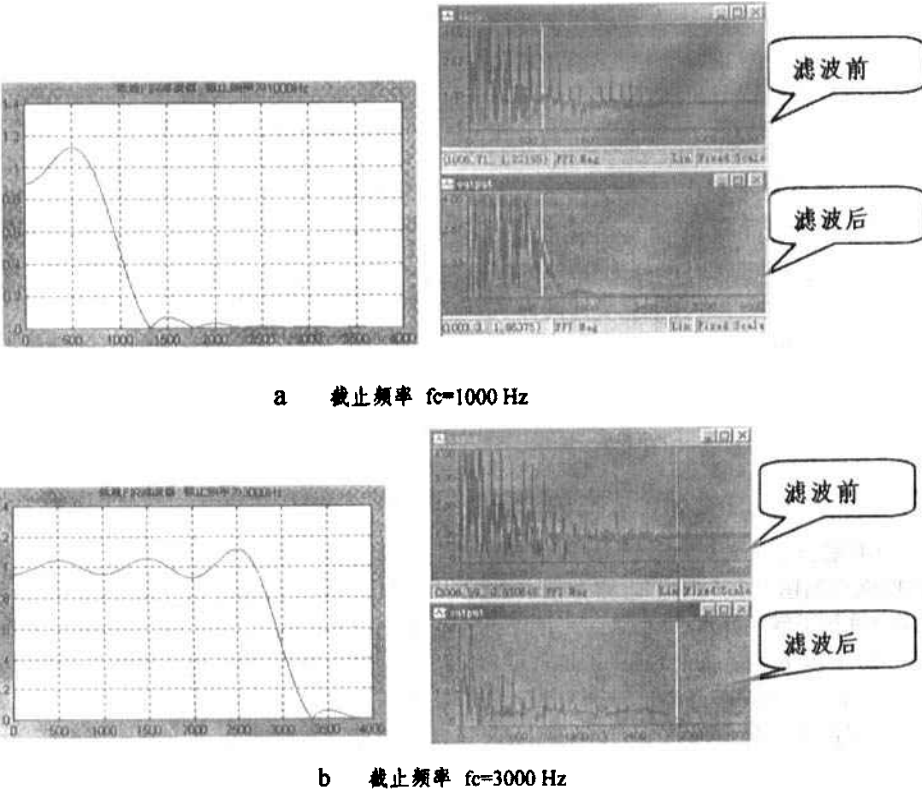


图5 不同的截止频率所对应的不同的滤波器特性及其处理效果图

2. FIR 理论设计

在本论文中,我们采用有限长单位冲击响应(FIR)滤波器,并使用窗函数设计法。经过理论推导,可得滤波器系数如下:

$$h(n) = \begin{cases} \frac{\omega_c}{\pi} \frac{\sin[\omega_c(n - \frac{N-1}{2})]}{\omega_c(n - \frac{N-1}{2})}, & 0 \leq n \leq N-1 \\ 0, & n \text{ 为其它值} \end{cases}$$

3.3 实验结果

我们设定的三个截止频率分别是1000 Hz、2000 Hz 和3000 Hz。在设计中对语音信号的采样频率设定为8k Hz,就可以定出滤波器系数序列的具体值了。将不同的滤波器系数序列带入编好的处理程序当中,在图5中,我们就可以看到不同的截止频率所对应的不同的滤波器特性及其处理效果了。

图5中,频率波形显示的滤波截止频率分别是1000Hz 和3000Hz 左右(黑线所示),这与理论上所设定的低通滤波器正好是符合的。在实验进行的现场,我们听到了滤波前、后语音的差异,滤波后已经

没有了滤波前所夹杂的高频干扰吱吱声。 $f_c = 1000$ Hz 时,语音的效果有断续现象, $f_c = 3000$ Hz 时语音的效果很好。这是因为人类语音的频率范围是300-3000Hz,当设定截止频率为3000Hz,不仅达到了很好的滤波效果,同时对音质的损伤相对1000Hz 要小很多,在图5所示的300-3000Hz 范围内,我们可以看到颇有规律的语音频谱信号。因此,在系统要求不是很苛刻的情况下,我们可以初步设定人类语音信号低通截止频率为3000Hz。在某些情况下,可以采用带通滤波器来得到更好的效果。

小结 本论文工作可以作为用 DSP 系统进行语音处理和分析的研究、试验平台,系统具有便于升级、高速的特点,并具有较高的应用价值。

参考文献

- 1 王念旭,等编著. DSP 基础与应用系统设计. 北京航空航天大学出版社,2001
- 2 程佩青,等编著. 数字信号处理教程. 清华大学出版社,1995
- 3 L. R. 拉宾纳, R. W. 谢弗. 语音信号数字处理. 科学出版社,1983
- 4 雷传华. 连接数字语音识别系统的 DSP 实时实现. 上海交通大学学报,1999,33(12)

(上接第148页)

1. 求测度 $\mu(x, y)$ 。

首先取一系列邻域半径分别为:1, $\sqrt{2}$, 2, $\sqrt{5}$, $2\sqrt{2}$, 3, $\sqrt{10}$, 然后由以上定义的测度公式,计算出相应的测度。

2. 求 Hölder 指数 $\alpha(x, y)$ 及广义维数 $D(x, y)$ 。分别根据定义(1)、(2),用最小二乘法拟合由步骤(1)所得的数据点对,由拟合直线的斜率即可求出每点的 $\alpha(x, y)$ 及 $D(x, y)$ 。

3. 求谱 $f(\alpha(x, y))$ 。利用步骤(2)所得数据,根据公式(3)计算出每一点谱 $f(\alpha(x, y))$ 。

4. 提取边缘。当 $f(\alpha(x, y)) \approx 1$ 时,标记点 (x, y) 为边缘点。

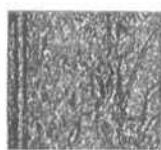
四、试验及分析

图(a)为256×256的 Lenna 原图,图(b)为用分

形方法^[1]检测出的边缘图,图(c)为采用本文的方法检测出的边缘图。由此看出,分形方法对噪音很敏感,检测出的伪边缘及噪音较多,而本文的方法具有较强的抗噪性,且能检测出较丰富的边缘。



图(a)



图(b)



图(c)

参考文献

- 1 巫兆聪,方圣辉. 基于分形理论的 SAR 图像边缘检测. 2000, 25(1): 334~336
- 2 Levy-Vehel J, Mignot P. Multifractal segmentation of images. Fractals, 1994, 2(3): 371~377
- 3 李后强,汪富泉. 分形理论及其在分子科学中的应用. 科学出版社,1997