

一种基于 SOLA 的数字音频 变调算法及其在 TMS320C6713 上的实现

蔡杰, 叶梧, 冯穗力

(华南理工大学 电子与信息学院, 广东 广州 510640)

摘要: 数字变调是数字音效处理的一种重要应用。描述了一种基于时域的内插和抽取技术实现音频变调的算法, 利用 SOLA-FS 算法保证播放时间不变。根据实际应用的要求, 对算法的实时性进行了改进, 并在 TMS320C6713 上进行了硬件实现。实现结果证明该算法效果良好。

关键词: 变调 SOLA-FS 数字信号处理 数字音效处理

数字音频变调是许多商用数字设备的一项重要功能, 如语音邮件、多媒体音频信号处理、采样合成器、调和器、声码器等。其中最为广泛的应用是卡拉 OK 中的数字变调装置。目前可查阅的资料中, 实现数字变调的方法均存在各种各样的问题。而一些成熟的算法, 由于商业保密的原因, 并没有向外界公开。本文采用一种时域上的线性内插及抽取的方法实现音频变调, 再结合时域 SOLA-FS 算法, 解决了因为变调后而导致播放时间改变的问题; 讨论了该算法在 TI 公司的 TMS320C6713 DSP 上的实时实现。

1 音频变调的基本原理

目前国际上普遍采用的音阶(音程)划分定律法是 12 平均律。12 平均律是将一个纯八度音分为 12 个均等的部分(即 12 个半音)。根据 12 平均律, 总音列中各音级的物理振动频率相差 $2^{\frac{1}{12}}$ 倍。一个纯八度频率相差两倍。如果将声音各频率成分的振动频率升高 $2^{\frac{1}{12}}$ 倍时, 就相当于音调升高了一个半音; 如果将声音各频率成分的振动频率降低 $2^{\frac{1}{12}}$ 倍时, 相当于音调降低了一个半音。

假设原音频信号的频率为 f , 变调后的频率为 f' , 则它们应满足式(1)的关系:

$$f' = f \times 2^{\frac{i}{12}}, i = \pm 1, \pm 2, \pm 3, \dots \quad (1)$$

当 $i > 0$ 时为升调, 反之则为降调。

现在大多数音调变换设备都使用数字存储器作为记录载体, 只要使音频数据的读入速度不同于输出速度, 就可以实现音调的变换。输出速率大于读入速率则为升调, 输出速率小于读入速率则为降调。大部分的卡拉 OK 机就是根据这一个原理进行变调的。例如 YAMAHA 公司生产的 YSS216B 芯片, 三菱公司的 M65831P 芯片都是用

于变调的硬件芯片。但是用硬件实现变调存在灵活性差等缺点。随着 DSP 技术的发展, DSP 芯片的功能越来越强大, 速度也越来越快, 因此采用软件实现音频变调具有一定的优点。

2 几种变调方法的比较

(1) 改变 WAV 文件的头文件信息, 以改变播放的采样率^[1]。这种方法实际上和硬件实现变调的方法相类似。采样率提高, 则 WAV 文件的播放速度也加快, 播放时间缩短, 产生升调的效果; 采样率降低, 则 WAV 文件的播放速度降低, 播放时间拖长, 产生降调的效果。为了保持播放时间不变, 需要把音频数据分成若干帧。对于升调, 则要把每帧最后一部分复制并补到该帧之后; 对于降调, 则把每帧数据最后部分删除。这个方法的不足之处在于: 升调时每帧之间是连续的, 但在帧内由于补上了一段数据, 在帧内会出现一处不连续。降调时帧内的数据是连续的, 但每帧之间由于有部分数据被删除也会导致不连续。因此, 声音的音质会受到较大的影响, 即使采样频率在 48kHz 时, 也会出现沙沙声和喀嚓声。

(2) 在频域上用线性内插的方法来实现声调的变化^[2]。这种方法的缺点在于: 内插会引入不需要的频率, 特别是在某些能量大的频点。假设要两倍频, 将会引入一些能量为原频点能量一半的频率分量, 这些频率分量会大大影响音频的音质, 而且与原来有用的频率靠得很近, 如果要进行滤波也比较麻烦。另外, 这种方法需要进行傅立叶变换及其反变换, 运算量较大。

(3) 频域搬移法^[3]。与第二种方法相比, 该方法不会引入新的频率分量。但是, 其实现的变调效果会带有金属声。原因在于频率的搬移并不是频率的线性变化。

3 变调算法研究与功能实现

3.1 变调

首先实现音频的变调, 即改变音频的频率。改变频

率可以通过改变采样率来实现。改变采样率的方法有很多,可以通过在相邻两个数据点之间插入新的数据点来达到增采样的目的,也可以通过每隔一定时间抽取一些点来达到降采样的目的。从频谱上来看,增采样可以实现频谱的压缩,从而起到降调的作用,而降采样则有升调的作用。当变调后的频率 f 与原频率 f_0 之比非整数时,可以将内插与抽取结合起来实现。

假设 $\alpha = \frac{f}{f_0} = \frac{M}{L}$, 其中 $M、L$ 均为正整数。要实现任意有理数倍频率的变调,可采用以下公式:

$$X_1(n) = X(\lfloor \frac{n}{L} \rfloor) + \frac{\text{mod}(n, L)}{L} \times \{X(\lfloor \frac{n}{L} \rfloor + 1) - X(\lfloor \frac{n}{L} \rfloor)\} \quad n=0, 1, 2, \dots, (N-1) \times L + 1$$

$$Y(n) = X_1(M \times n) \quad n=0, 1, 2, \dots, N \times L / M \quad (2)$$

其中, N 是帧长, $\lfloor \cdot \rfloor$ 表示取整运算。当 $M > L$ 时实现升调, $M < L$ 时实现降调。如果要进行实时实现的话,帧长 N 不宜取得太大。

3.2 采用 SOLA-FS 算法保持播放时间不变

由于按上述方法进行变调会造成音频信号播放时间缩短(升调)或延长(降调),因此本文采用 SOLA-FS (Synchronized Over Lap-and-Add-Fixed Synthesis) 算法来保持播放时间不变。该算法可以在不改变音调的情况下,对音频数据的长度进行改变。与传统的 SOLA 算法相比, SOLA-FS 算法的输入和输出信号均采用固定的位移间隔,能使计算相关性和交叉渐变过程的运算量大幅降低,从而便于硬件的实时实现。该算法的主要参数如下: W —窗长度; S_a —分析位移; S_s —合成位移; K_{\max} —寻找相关性的最大允许位移; W_{ov} —交叉渐变区域长度。数据压缩时,各参数关系如图 1 所示。数据扩展时,各参数关系如图 2 所示。

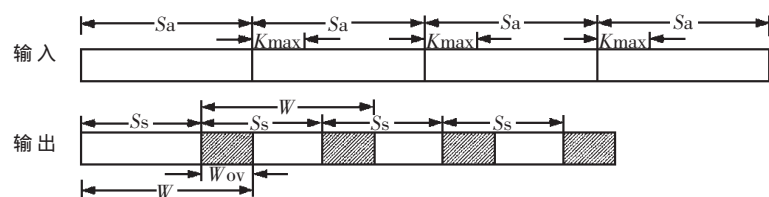


图 1 数据压缩时, SOLA-FS 各参数关系

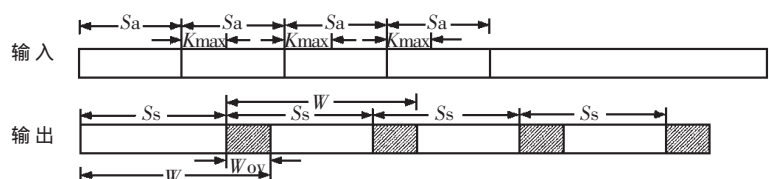


图 2 数据扩展时, SOLA-FS 各参数关系

从图 1、图 2 中可以很容易地证明,如果输入的音频信号长度足够长,则新旧序列的长度之比为 $\frac{S_s}{S_a}$ 。具体算法如下:

设 $X[n]$ 为输入音频信号采样序列, $X_m[n]$ 为第 m 个

窗口的音频信号。则:

$$X_m[n] = \begin{cases} X[m \times S_a + k_m + n] & n=0, 1, \dots, W-1 \\ 0 & \text{否则} \end{cases} \quad (3)$$

设 $Y[n]$ 为输出音频信号序列,它与输入音频信号序列 $X[n]$ 之间的关系为:

$$Y[m \times S_s + n] = \begin{cases} (1 - n/W_{ov}) \times Y[m \times S_s + n] + n/W_{ov} \times X_m[n] & n=0, 1, \dots, W_{ov}-1 \\ X_m[n] & n=W_{ov}, \dots, W-1 \end{cases} \quad (4)$$

(3) 式中, k_m 是最优匹配位移,即最优匹配点与第 m 个窗口之间的距离。解决最优匹配的判决标准, SoLA-FS 算法采用互相关系数 $R_{xy}^m[k]$ 来表示匹配程度。

$$k_m = \max_{0 \leq k \leq K_{\max}} R_{xy}^m[k] \quad (5)$$

$$R_{xy}^m[k] = \frac{r_{xy}^m[k]}{\sqrt{r_{xx}^m[k] r_{yy}^m[k]}} \quad (6)$$

$$r_{xy}^m[k] = \sum_{n=0}^{W_{ov}-1} X[m \times S_a + k + n] \times Y[m \times S_s + n] \quad (7)$$

$$r_{xx}^m[k] = \sum_{n=0}^{W_{ov}-1} X^2[m \times S_a + k + n] \quad (8)$$

$$r_{yy}^m[k] = \sum_{n=0}^{W_{ov}-1} Y^2[m \times S_s + n] \quad (9)$$

3.3 实时改进

由于互相关系数的计算比较复杂,因此在硬件的实时处理中,很有必要采用一种低复杂度的算法来进行匹配性的判决。本文采用平均绝对误差函数(MAE)来简化这一判决过程。

$$MAE(k) = \frac{1}{W_{ov}} \sum_{n=0}^{W_{ov}-1} |X[m \times S_s + n] - X[m \times S_a + k]| \quad (10)$$

$$k_m = \min_{0 \leq k \leq K_{\max}} MAE(k) \quad (11)$$

经改进后,假设交叉渐变区域长度 $W_{ov} = a$, 则计算一次匹配程度。MAE 算法需要进行 a 次减法运算、 $a-1$ 次加法运算、 a 次取绝对值运算和一次乘法运算,而互相关系数法则需要 $3a$ 次乘法运算、 $3(a-1)$ 次加法运算、1 次开根和 1 次除法运算。本文在 VC 上对一段时长为 35 秒的音乐进行非实时处理,互相关系数法耗时大约为 20 秒,而 MAE 耗时仅为 3 秒。可见,采用 MAE 算法后运算量大幅降低。

图 3 为互相关系数法和 MAE 判决方法对同一段音乐判决结果的比较。实验中取 $K_{\max} = 2000$ 。实验证明, MAE 算法得出的结果与互相关系数法得出的结果基本一致。

4 实验仿真及分析

图 4 是一段音乐分别升、降两个半音后的频谱图。从图中可以看出,升调后的频谱是原频谱的拉伸,降调是对原频谱的压缩。以源音乐频谱图 700Hz 附近的一个

谱峰为例,根据式(1)可知,升两个半音后的频率应该为升调前频率的 1.12 倍,即 $700 \times 1.12 = 784\text{Hz}$ 。因此可以在升两个半音后频谱图中 780Hz 附近找到一个谱峰与源音乐频谱图中的那个谱峰相对应。同理可得,对于降调,降两个半音后的频率应该为降调前频率的 0.893,即 $700 \times 0.893 = 625.1\text{Hz}$ 。因此可以在降两个半音后频谱图中的 620Hz 附近找到一个谱峰与源音乐频谱图中的那个谱峰相对应。

仿真表明,本文提出的方法是行之有效的,较好地实现了变调功能。

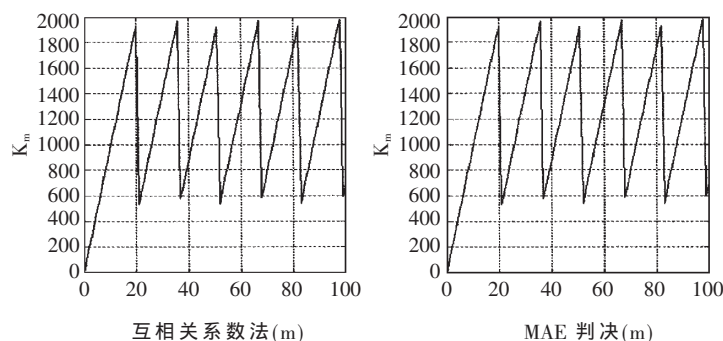
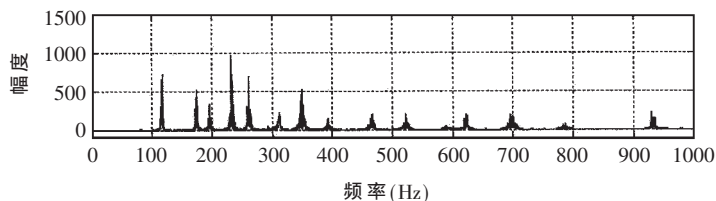
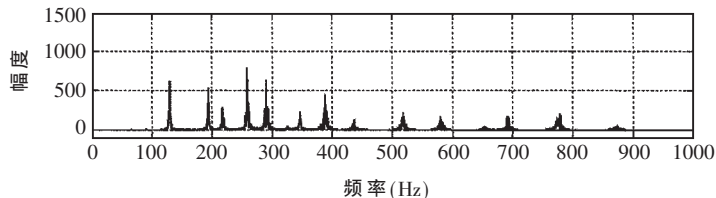


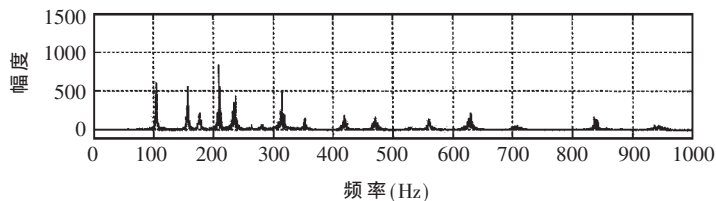
图3 互相关系数法和 MAE 判决结果比较



(a) 源音乐频谱图



(b) 升两个半音后频谱图



(c) 降两个半音后频谱图

图4 频谱仿真效果

5 TMS320C6713 上的实时实现

TMS320C6713 是 TI 公司生产的一款高性能 DSP。该 DSP 片内集成了多通道缓冲串口 McBSP、PC、EDMA 等模块,为数字音频信号处理提供了很大的便利。而且该芯片 1 600MIPS 的运算速度、256KB 二级缓存可以满足

大多数音频处理的要求,特别适用于数字调音台、数字功放等高端音频设备。其硬件实现框图如图 5 所示。

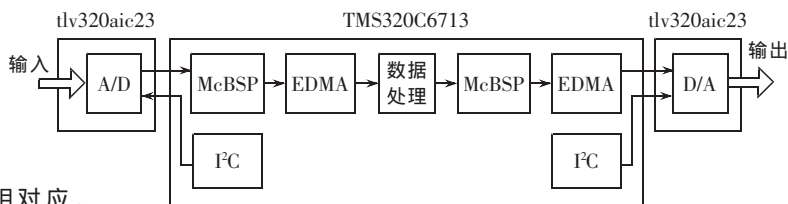


图5 硬件框图

在实验中,A/D 和 D/A 的采样频率为 48kHz,每个采样数据为 16bit,DSP 工作频率为 200MHz(即 1600MIPS)。在实际应用中,往往需要进行连续平滑的变调处理,即从一个音调变换到另一个音调不能让人感觉有严重的噪声或明显的不连续感。要解决这一问题,就要保证每次处理前后输入输出的数据量要相同。在数据缓冲方面,本文采用乒乓缓存的机制,选取合适的帧长,以保证旧数据被新数据覆盖前已经处理完毕。取每一帧数据长度为 len ,先进行变调处理,处理后数据长度变为 $\frac{len}{\alpha}$ 。为了保证长度不

变,再进行 SOLA-FS 算法处理,取 $S_s = len, S_a = \frac{len}{\alpha}$,

则每次处理后输出数据长度也为 len 。设乒乓缓存的长度均为 len ,tlv320aic23 将模拟信号转化为数字信号后,通过 McBSP 传送到 DSP,DSP 利用 EDMA 通道写入乒乓缓存。当其中一个缓存放满后,EDMA 产生一个软中断,DSP 就开始调用变调处理程序。与此同时,后续的音频信号继续通过 McBSP 和 EDMA 写入另一个缓存。

实验结果证明,在双声道立体声、48kHz 采样频率、16bit 数据的实验条件下,用 TMS320C6713 实现本文提出的算法处理 1 秒的音频数据量的时间不超过 100 毫秒,完全满足实时处理要求。

本文提出的数字音频变调方法,通过使用数字信号处理中的内插、抽取技术和 SOLA-FS 算法,结合交叉渐变进行过渡平滑。本方法相对于频域法,运算量小,易于在硬件上实现,既适用于音乐信号,也适用于语音信号,而且变调效果良好。

参考文献

- 1 狄红卫,陈龙云,张永林.数字音乐变调方法的研究与实现.计算机工程与应用,1998;(3):28~30
- 2 李力利.数字音频变调算法的研究.电声技术,2003;(6):18~20
- 3 周婷.非线性编辑中的数字音频处理——用软件实现音频特技.电子科技大学学位论文,2001
- 4 舒非,李桓,李玉柏.语音变调算法的实时改进及其在 DSP 平台上的软件实现.电声技术,2004;(9)
- 5 HEJNA D J, MUSICUS B R, CROWE A S. Method for time-

(接上页)

scale modification of signals. United States Patent 5,175,769, 1992

6 LIN G J, CHEN S G, WU T. High quality and low complexity

pitch modification of acoustic signals. Acoustics, Speech, and Signal Processing, 1995. ICASSP-95, 1995 International Conference, 1995; (5): 2987~2990

(收稿日期: 2006-07-06)

* 基金项目: 湖南省自然科学基金项目(编号: 05JJ30123);
湖南省教育厅科学研究项目(编号: 05C246)