

文章编号: 1002- 8684(2006) 09- 0007- 03

一种改进的时域音频变调方法及其软件实现

·实用技术·

蔡 杰

(华南理工大学 电子与信息学院, 广东 广州 510640)

【摘 要】变调是音效处理的一种, 其方法有很多种。采用了一种基于改进的时域音频变调方法对音频实现变调处理, 能在变调的同时保证播放时间不变, 通过交叉渐变的方法提高音质, 并在计算机上用高级语言进行了仿真实验。

【关键词】变调; 数字信号处理; 数字音频处理

【中图分类号】TN912

【文献标识码】B

The Realization of an Improved Method of Pitch Shifting

CAI Jie

(Electronic Engineering Department, South China University of Technology, Guangzhou 510640, China)

【Abstract】An improved method of pitch shifting based on time domain is proposed. Using this method, the pitch shifting is realized. In the meanwhile, the length of the signal is kept as that of the original signal. Finally, cross-fade is used to improve the quality of the acoustics signals.

【Key words】pitch shifting; digital signal process; digital audio process

1 引言

改变声音信号的音调是许多商用设备的一项功能, 如语音邮件、多媒体音频信号处理、采样合成器、调音器和声码器等, 其中运用最广泛的应属卡拉 OK 机了。由于每个人的声调都不同, 人们在唱歌的时候都希望伴奏音乐的音调适合自己的嗓音。所以在很多卡拉 OK 机中设计了专门的硬件来完成这一功能。由于商业保密的需要, 很多成熟的算法并没有对外公开, 而能查阅到的方法也存在着各种各样的问题。笔者采用一种线性内插及抽取的方法实现音频变调, 再结合经典的复制和删除的保长方法, 保持音频信号播放时间不变, 并使用音频交叉渐变的方法保证音质, 实验证明效果良好。

2 音频变调基本原理

目前, 世界上普遍采用的定律法是 12 平均律。12 平均律是将一个纯八度分为 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 2 倍。如果将声音各频率成分的振动频率升高 $2^{\frac{1}{12}}$ 倍, 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, 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_i , 则它们应满足以下关系^[1]

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

当 $i > 0$ 时为升调, 反之为降调。

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。只要使音频数据存入存储器的速度不同于从存储器中输出的速度, 就可以实现音调的变换。输出速率大于存入速率则为升调, 输出速率小于存入速率则为降调。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^[2]。但是用硬件实现变调存在灵活性差的问题。目前, 一般计算机声卡都没有专门硬件变调功能。因此, 这使得用软件实现音频变调就具备一定的优点。

3 常用的变调方法及其比较

一种实现变调的方法^[3]是改变 WAV 文件的头文件信息, 改变播放的采样率。这实际上与硬件实现变调

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 48 kHz 时,也会出现沙沙声和喀嚓声。

另一种方法^[4]是在频域上用线性内插的方法来实现频率的提高与降低,从而实现声调的变化。这个方法的缺点在于:内插的方法会引入不需要的频率。特别是在某些能量大的频点,假设要升 2 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。而且这些频率分量与原来有用的频率贴得很近,若要进行滤波也较麻烦。另外,这种方法需要进行傅里叶变换及其反变换,运算量较大。

还有一种方法^[5]是频域搬移法。与上一种方法相比,该方法不会引入新的频率分量。但是,该方法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。

4 变调实现

4.1 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。改变采样率的方法很多,可以通过在相邻两个数据点之间内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距离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,从而有降调的作用,而降采样则有升调的作用^[6]。当变调后的频率 f 与原频率 f_0 之比为非整数时,可以将内插与抽取结合起来实现。

假设变调因子 $\alpha = \frac{f}{f_0} = \frac{M}{L}$, 其中 M, L 均为正整数。

要实现任意有理数倍频率的变调,采用以下公式

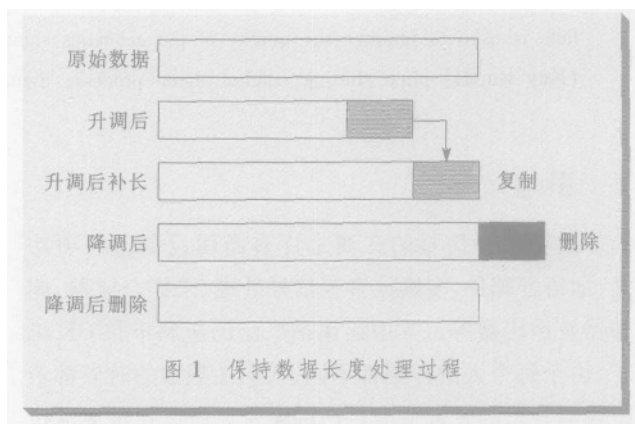
$$X_1(n) = X\left(\left\lfloor \frac{n}{L} \right\rfloor\right) + \frac{\text{mod}(n, L)}{L} \times \left\{ X\left(\left\lfloor \frac{n}{L} \right\rfloor + 1\right) - X\left(\left\lfloor \frac{n}{L} \right\rfloor\right) \right\}, n=0, 1, 2, \dots, (N-1) \times L + 1$$

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

其中, N 是帧长, $\lfloor \cdot \rfloor$ 表示取整运算, mod 为取模运算。当 $M > L$ 时实现升调, $M < L$ 时实现降调。若要进行实时实现, 帧长 N 不宜取得太长。

4.2 保持播放时间不变

变调后, 为了保持播放时间不变, 需要对变调后的数据进行适当的处理。以升调为例, 升调会使原数据变短。先将经变调处理后的数据分成一个个帧, 在每一帧后面补上该帧最后一部分数据, 补数据的具体长度应为帧长的 $\left(1 - \frac{M}{L}\right)$ 倍。帧长的选择必须适当, 如果选得太短, 会导致变调无效果; 选得太长, 可以明显地感觉到声音多次被重复或打断。经实验证明, 帧长取 10 000 点左右比较适合。对于采样率为 48 kHz 的音频信号, 10 000 点相当于 20 ms 的数据量, 人耳一般分辨不出, 符合人耳对声音的分辨特性。数据保持长度不变的处理过程如图 1 所示。



4.3 过渡带平滑

为了克服复制或删除数据而导致的噼啪声, 笔者采用交叉渐变的方法, 用于平滑过渡部分。公式如下

$$CF(m) = X_1(m) \times \left(1 - \frac{m}{\text{len}}\right) + X_2(m) \times \frac{m}{\text{len}}, \quad 0 \leq m \leq \text{len} - 1 \quad (3)$$

式中: len 是交叉渐变区域的长度。

经交叉渐变处理后, 噼啪声等噪声明显减少, 音质得到了提高。交叉渐变区域的长度的选取也非常讲究。过短的话, 平滑的效果不明显; 过长的话, 会导致时域波形图出现凹谷, 让人感觉声音时大时小, 影响音质。

5 实验仿真及分析

图 2 是对一段音频信号分别升、降两个半音后的

软件仿真效果图。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。原信号是 300 Hz、500 Hz 和 1 000 Hz 这 3 个正弦信号的叠加。根据式(1)可知,升两个半音后的音频信号各频率应该为升调前频率的 1.12 倍。因为 $300 \times 1.12 = 336$, $500 \times 1.12 = 560$, $1\,000 \times 1.12 = 1\,120$, 所以可以在升两个半音后信号频域图中的 330 Hz、560 Hz 及 1 100 Hz 这 3 个频点附近找到与原信号频域图中相对应的 3 个谱峰。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。因为 $300 \times 0.893 = 267.9$, $500 \times 0.893 = 446.5$, $1\,000 \times 0.893 = 893$, 所以可以在降两个半音后信号频域图中的 270 Hz、450 Hz 及 890 Hz 3 个频点附近找到与原信号频域图中相对应的 3 个谱峰。

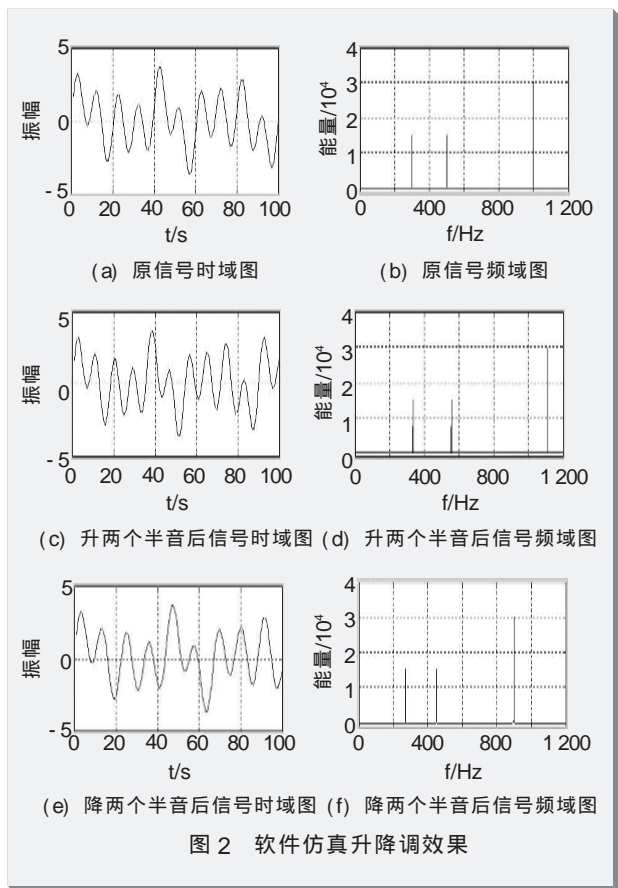


图2 软件仿真升降调效果

实验表明,在时域上,音频信号经过变调及交叉渐变处理后基本没有明显不连续点。在频域上,也没有引入新的频率分量,频谱清晰,无混浊。综上所述,文中所使用的改进方法是行之有效的,变调功能基本实现。

6 结束语

笔者提出了改进的音频变调方法,通过使用数字

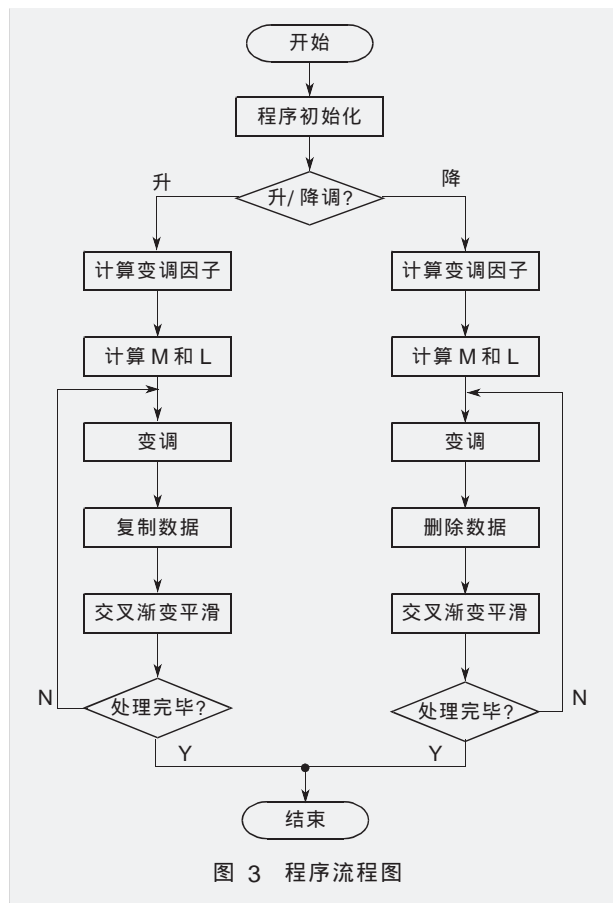


图3 程序流程图

信号处理中的内插和抽取技术以及常用的保长方法,结合交叉渐变进行过渡平滑。该方法相对于频域法,运算量小,易于在计算机上用软件实时实现,且既适用于音乐信号,也适用于语音信号,在对音质要求不算高的情况下是一个不错的实现方法。

参考文献

- [1] 韩宪柱. 数字音频技术与应用[M]. 北京:中国广播电视出版社,2003.
- [2] 刘宪坤. 数字音响技术[M]. 北京:人民邮电出版社,1993.
- [3] 狄红卫,陈龙云,张永林. 数字音乐变调方法的研究与实现[J]. 计算机工程与应用,1998(3):28-30.
- [4] 李力利. 数字音频变调算法的研究[J]. 电声技术,2003(6):18-20.
- [5] 周婷. 非线性编辑中的数字音频处理——用软件实现音频特技[D]. 成都:电子科技大学学位论文,2001.
- [6] 金连文,韦岗. 数字信号处理简明教程[M]. 北京:清华大学出版社,2003.

作者简介

蔡杰,硕士研究生,主要研究方向为通信与信息系统及数字信号处理。

[责任编辑] 闫雯雯

[收稿日期] 2006-07-10