

基于短时傅里叶变换的音频变速算法及其 DSP 实现

杨洪柏, 陈蕾蕾, 李战伟

(上海开放大学 信息与工程学院, 上海 200433)

摘要: 介绍了一种基于短时傅里叶变换的音频变速方法, 并在 DSP 平台上对其进行优化, 提高了运算速度。仿真测试结果表明, 该方法可达到良好的音频变速效果。

关键词: 音频变速; 短时傅里叶变换; FFT; 瞬时频率; DSP

中图分类号: TN912

文献标识码: A

文章编号: 1674-7720(2013)16-0042-03

Audio time-scaling algorithm based on STFT and its implementation on DSP platform

Yang Hongbai, Chen Leilei, Li Zhanwei

(School of Information and Engineering, Shanghai Open University, Shanghai 200433, China)

Abstract: This paper introduces an audio time-scaling algorithm based on STFT. The algorithm is optimized to be implemented on DSP platform and the computing speed is increased. The test results indicate that the DSP algorithm achieves good results of audio time-scaling.

Key words: audio time-scaling; STFT; FFT; instantaneous frequency; DSP

近年来, 随着数字技术的飞速发展, 带有数字音频播放功能的电子产品越来越多地进入人们的日常生活, 如便携式音乐播放器、数码语音复读机等。在这些电子产品中, 有一种被称为音频变速的功能日益受到重视, 即在不丢失音频内容、不改变音调的前提下使音频变快或变慢, 其本质是对信号在时间尺度上进行压缩或扩展。音频变速是一项很实用的功能, 对于数码复读机而言, 语音变速可以使听者通过操控语音的速度来提高外语学习的效果; 对于音乐播放器而言, 音乐变速可提高音乐欣赏的趣味性。

目前, 对音频变速算法的研究主要局限在语音领域, 它可分为时域方法和频域方法两大类。时域方法仅限于时域处理, 典型算法为基音同步重叠相加 PSOLA (Pitch-Synchronous Overlap and Add) 算法^[1]等。时域方法虽然运算简单, 但往往音质较差。对变速后语音品质的追求导致了频域方法的出现。20 世纪 90 年代, LAROCHE J 等人^[1-2]提出了基于短时傅里叶变换 STFT (Short-Time Fourier Transform) 的语音变速方法, 它可在各种变速比下实现良好的语音变速效果。本文将这种基于 STFT、面向语音的变速算法扩展应用于包括纯音乐和歌曲在内的各种音频信号, 取得了良好的变速效果,

不仅音调保持不变, 而且音色也基本保持不变。本文针对 DSP 平台进一步对该音频算法进行优化, 大大提高了算法计算速度。最后采用 Visual DSP 软件对算法进行了编程和测试。

1 基于 STFT 的音频变速算法

根据信号傅里叶分析理论, 音频信号可以被视为一系列正弦分量的线性叠加。信号频谱决定了音调, 其中关键因素是各正弦分量的瞬时频率。如果计算出正弦分量瞬时频率, 那么根据时间压缩或扩展的要求构造相位连续的短时频谱就可实现音频变速而不改变其音调。为了准确估计音频的瞬时频率, 需要将信号分帧, 并利用前后帧对应正弦分量的相位信息。上述音频变速算法需要借助正反短时傅里叶变换, 实现音频信号的分析与合成。该算法具体步骤如下。

(1) 原始信号分帧。音频一般是时变的非平稳信号, 但可假设其在短时间(例如 10 ms~30 ms)内是平稳的。因此可将音频信号按时间分帧, 以便于频谱分析。设原始音频信号为 $s(m)$, $m=0, 1, 2, \dots$ 。将 $s(m)$ 分为等长的帧(称为分析帧), 相邻帧之间有重叠。帧长为 L 个采样点, 每帧步进长度为 P 个采样点, 那么第 i 帧信号可表达为:

$$x_i(n) = s(iP + n) \quad (i=0, 1, 2, \dots; n=0, 1, 2, \dots, L-1) \quad (1)$$

《微型机与应用》2013 年第 32 卷第 16 期

其中, P 一般取 $[L/8 \ L/4]$ 区间内的整数。

(2) 分析帧信号加窗。为使短时频谱计算更准确, 对分析帧信号进行加窗处理, 窗函数 $h(n)$ 可选海明窗或汉宁窗, 取窗长等于帧长 L 。

$$x_i'(n) = h(n) \cdot x_i(n) \quad (2)$$

(3) 计算分析帧短时频谱。利用离散形式的短时傅里叶变换计算加窗后帧信号 $x_i'(n)$ 的频谱:

$$x_i(k) = \sum_{n=0}^{N-1} x_i'(n) \cdot e^{-j \frac{2\pi}{N} kn} \quad (k=0, 1, 2, \dots, N-1) \quad (3)$$

其中, $N \geq L$ (如果 $N > L$, 则 $x_i'(n)$ 后面添 $N-L$ 个零)。当然, 可采用快速傅里叶变换(FFT)算法进行频谱计算^[3]。

步骤(1)~(3)的计算如图 1 所示。

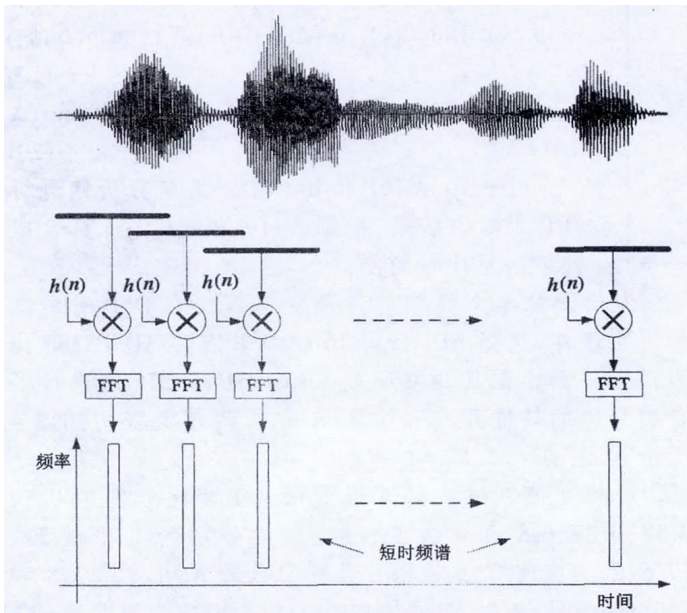


图 1 原始音频信号的分帧、加窗和短时傅里叶变换

(4) 瞬时频率计算。在求得相邻两帧的频谱后, 可以根据频谱的相位信息求得相应时间段内信号的瞬时频率。对于特定的 k , 频谱分量 $X_i(k)$ 的相位角为:

$$\varphi_i(k) = \text{mod} \left(\arctan \left(\frac{\text{Im}(x_i(k))}{\text{Re}(x_i(k))} \right) \cdot 2\pi \right) - \pi \quad (4)$$

其中, $\text{Im}()$ 表示取虚部计算, $\text{Re}()$ 表示取实部计算, $\text{mod}()$ 为取模计算。 $\varphi_i(k)$ 的取值范围为 $[-\pi, \pi)$ 。

设相邻两帧频谱相位角分别为 $\varphi_i(k)$ 和 $\varphi_{i+1}(k)$, 则角度差值为:

$$\delta_i(k) = \varphi_{i+1}(k) - \varphi_i(k) - \frac{2\pi \cdot k \cdot P}{N} \quad (5)$$

由于相邻两帧在时间上相隔 P 个采样周期, 对于任一 k 值, 如果频谱分量 $X_i(k)$ 的数字角频率准确地等于 $\frac{2\pi \cdot k}{N}$, 那么 $\delta_i(k)$ 将等于零。反之, 如果计算所得 $\delta_i(k)$ 不

等于零, 则说明 $X_i(k)$ 的数字角频率与频谱采样频率 $\frac{2\pi \cdot k}{N}$ 之间有偏差。为求此偏差, 首先对 $\delta_i(k)$ 作解卷绕操作:

$$\delta_i'(k) = \text{mod}(\delta_i(k) + \pi, 2\pi) - \pi \quad (6)$$

解卷绕的目的是当 $\delta_i(k)$ 不在 $[-\pi, \pi)$ 范围时, 使其等

价映射至 $[-\pi, \pi)$ 范围内。

对于第 i 帧信号, 各正弦分量的瞬时频率如下, 如图 2 所示。

$$f_i(k) = \frac{2\pi \cdot k}{N} + \frac{\delta_i'(k)}{P} \quad (k=0, 1, \dots, N/2) \quad (7)$$

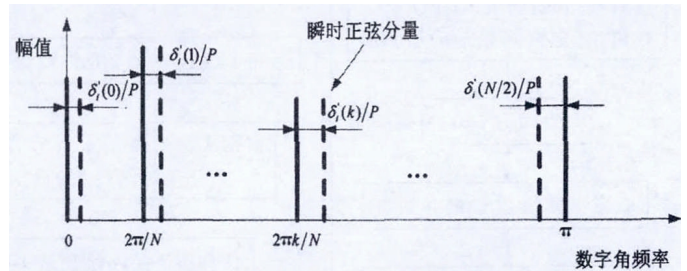


图 2 正弦分量的瞬时频率示意图

(5) 合成帧频谱计算。将音频变速比 r 定义为合成音频速度与原始音频速度的比值。 r 取值范围一般为 $[0.5, 2]$ 。假设 r 在变速过程中为恒定值, 那么合成帧的步进长为:

$$Q = \text{round}(P/r) \quad (8)$$

其中, $\text{round}()$ 为圆整操作。

合成帧与分析帧之间起始点的对应关系如图 3 ($r < 1$) 和图 4 ($r \geq 1$) 所示。可以根据分析帧频谱推算出合成帧频谱。设第 i 合成帧频谱为 $Y_i(k)$, 其初始相位 $\beta_i(k)$ 可以根据分析帧各正弦分量瞬时频率和合成帧步进长进行递推计算:

$$\beta_0(k) = \alpha_0(k) \quad (9)$$

$$\beta_i(k) = \beta_{i-1}(k) + f_{i-1}(k) \cdot Q \quad (i > 0) \quad (10)$$

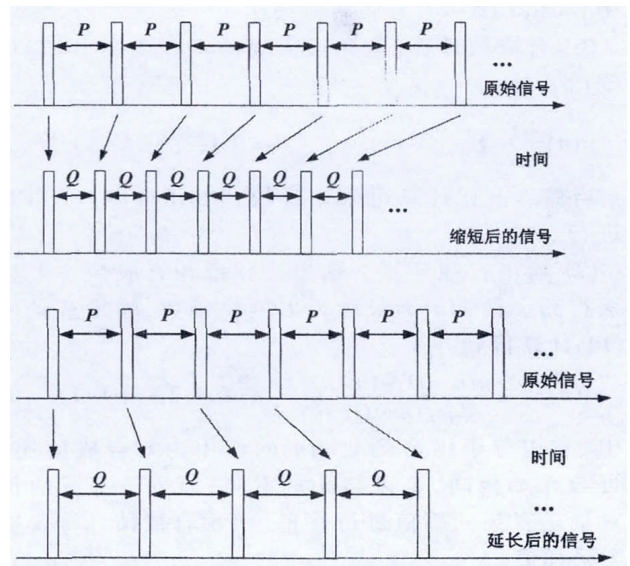


图 3 合成帧与分析帧之间起始点的对应关系 ($r \geq 1$)

$Y_i(k)$ 的幅值可保持与 $X_i(k)$ 相同。这样, 第 i 合成帧的频谱为:

$$Y_i(k) = |X_i(k)| \cdot e^{j \cdot \beta_i(k)} \quad (k=0, 1, \dots, N/2) \quad (11)$$

其中, $|X_i(k)| = \sqrt{(\text{Re}(X_i(k)))^2 + (\text{Im}(X_i(k)))^2}$ 。由于实信号频

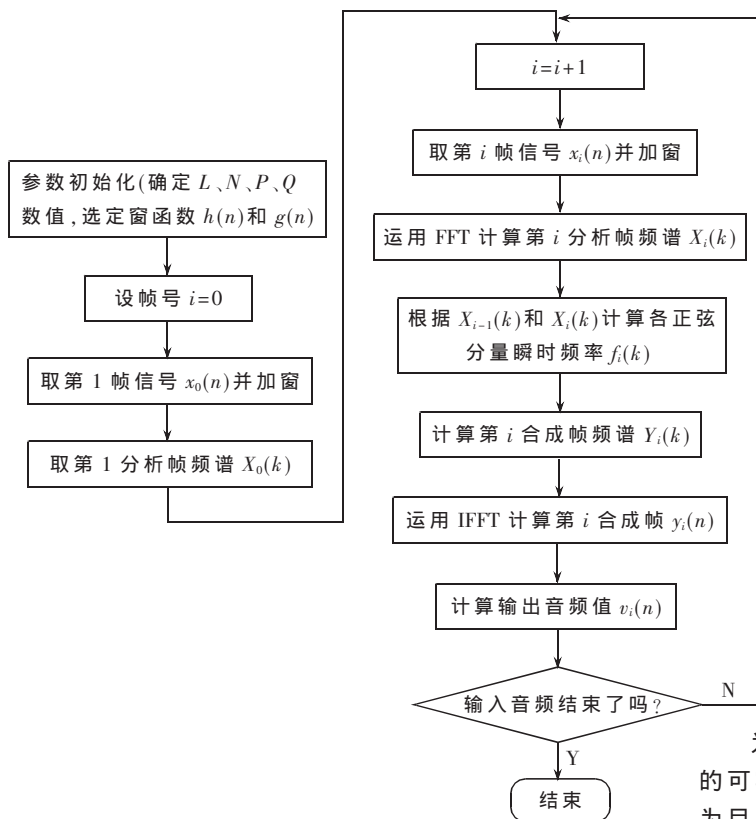


图 4 DSP 程序流程图

谱的共轭对称性,当 $k=N/2+1, \dots, N-1$ 时,可按式(12)计算频谱分量:

$$Y_i(k) = \text{conj}(Y_i(N-k)) \quad (12)$$

其中, $\text{conj}(Y_i(N-k))$ 为取共轭操作。

(6) 合成帧计算。利用反离散傅里叶变换(IDFT)得到合成语音帧:

$$y_i(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y_i(k) \cdot e^{j \frac{N}{2} kn} \quad (n=0, 1, 2, \dots, N-1) \quad (13)$$

当然,上述计算可采用反快速傅里叶变换(IFFT)算法进行。

(7) 输出音频计算。输出音频由各合成帧按重叠相加法得到。设 $g(n)$ 为长度为 L 的窗函数,输出音频可按式(14)计算得到^[1]:

$$v_i(n) = \frac{\sum_i g(n-iQ) y_i(n-iQ)}{\sum_i g^2(n-iQ)} \quad (n=0, 1, 2, \dots, Q-1) \quad (14)$$

其中,右边分子内窗函数 $g(n)$ 的作用为对合成帧加权,它可与 $h(n)$ 相同,也可与 $h(n)$ 不同;而分母部分的作用是对输出音频的幅值进行修正。每次合成帧计算仅输出 Q 个采样值。

2 DSP 平台上算法的优化和实现

在音乐播放器和语音复读机等嵌入式设备中,音频处理任务一般由数字信号处理器 DSP(Digital Signal Processor)来完成。由于 DSP 的运算速度和内存等资源往往是受限制的,在 DSP 平台上实现上述音频变速算法时需要对其进行优化以简化运算。通过分析发现,在式(11)

所表达的合成帧频谱计算中,开方运算是一项耗时的运算。为此,本文对合成帧频谱计算过程进行了改进,提出了一种无开方运算的合成帧频谱计算方法。对每一合成帧计算如下角度差:

$$\Delta_i(k) = \beta_i(k) - \alpha_i(k) \quad (15)$$

那么根据式(11)可得:

$$Y_i(k) = |X_i(k)| \cdot e^{j(\alpha_i(k) - \Delta_i(k))} \quad (16)$$

$$\text{Re}(Y_i(k)) = |X_i(k)| \cos(\alpha_i(k) + \Delta_i(k)) \quad (17)$$

$$\text{Im}(Y_i(k)) = |X_i(k)| \sin(\alpha_i(k) + \Delta_i(k)) \quad (18)$$

根据三角函数和角公式以及 $\text{Re}(X_i(k)) = |X_i(k)| \cos(\alpha_i(k))$ 、 $\text{Im}(X_i(k)) = |X_i(k)| \sin(\alpha_i(k))$,可得:

$$\text{Re}(Y_i(k)) = \text{Re}(X_i(k)) \cos(\Delta_i(k)) - \text{Im}(X_i(k)) \sin(\Delta_i(k)) \quad (19)$$

$$\text{Im}(Y_i(k)) = \text{Im}(X_i(k)) \cos(\Delta_i(k)) + \text{Re}(X_i(k)) \sin(\Delta_i(k)) \quad (20)$$

由于 $\text{Re}(X_i(k))$ 和 $\text{Im}(X_i(k))$ 是对分析帧进行 FFT 计算的结果,根据式(19)和式(20)计算合成帧频谱是比较简便的。

为了评估本文介绍的音频变速算法在 DSP 平台上的可实现性,选择 ADI 公司 16 位 DSP 芯片 ADSP2188 作为目标平台,在其开发环境 VisualDSP++ 3.5 中进行了音频变速的软件开发和仿真测试。程序流程图如图 4 所示。

音频变速软件支持 4 级变慢(变速比分别为 0.5、0.57、0.67、0.8)和 4 级变快(变速比分别为 1.25、1.50、1.75、2)。变速算法中 FFT 运算的点数 N 可支持 128 和 256 两种。设定 $L=N$,并选取 $h(n)$ 和 $g(n)$ 均为海明窗。选取采样率为 22.05 kHz 的单声道音频作为测试音频,音频内容分为纯语音、纯音乐和歌曲(含人声和伴奏)。测试结果表明,在上述变速比和 FFT 点数设定下,音频变速后音调保持不变。当 $0.57 \leq r \leq 1.75$ 时,音质良好;而当 $r=0.5$ 或 $r=2$ 时,音质较差。同是变快或变慢情况下,变速比越接近 1,则音质越好。另一方面,256 点 FFT 运算的音质要好于 128 点 FFT 运算的音质。图 5 显示了一段音频的原始波形以及在变快($r=1.5$)和变慢($r=0.67$)时的波形。根据上述测试,为保证变速音频的品质,一般应选 $0.57 \leq r \leq 1.75$ 。

软件仿真结果表明,ADSP2188 的运算能力满足音频变速程序的要求。选定分析帧步进长 P 在 $[L/8, L/4]$ 区间时,算法运算量最大值小于 50 MIPS。而 ADSP2188 的运算能力可达到 75 MIPS,它可满足音频变速实时计算的要求。程序代码长为 1 K Words,所需数据存储空间为 2.4 K Words,远远小于 ADSP2188 所能提供的内存空间。软件测试还表明,省去开方运算后,最多可使变速算法节省 12 MIPS 的运算量,这是相当可观的。

(下转第 47 页)

《微型机与应用》2013 年第 32 卷第 16 期

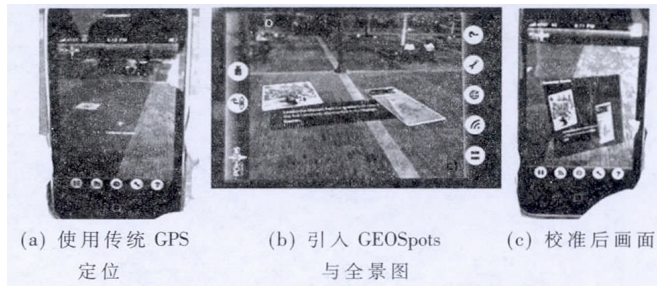


图3 同一 GEOSpot 处叠加在地面上的 AR 内容

两种类型 AR 内容校准前后误差对比如表 1 所示。从表 1 中可以发现,虽然校准后的误差仍比全景图背景误差要大,但是相对于校准前的误差而言则小得多。

表 1 两种类型 AR 内容校准前后误差分析

AR 内容位置分类	校准前误差	全景图误差	校准后误差
叠加在建筑物旁	约 15 m	约 2 m	约 4 m
叠加在地面上	定位错误	与地面偏离约 0°	与地面偏离约 35°

本文进行了 LBS 手机 AR 系统定位精度的研究,提出了提高定位精度的新方法。引入 GEOSpot,并在 GEOSpot 处提供全景图背景来匹配校准 AR 内容的显示,最后以 mixare 开源框架为基础进行实例验证。

虽然目前 LBS 手机 AR 系统受到精度方面的影响,但是随着未来 GPS 定位精度的提高与新的定位方法的出现,实现 LBS 手机 AR 系统的精确定位是很有可能,相信到时候 LBS 手机 AR 应用领域将会大大拓展。

参考文献

- [1] HAKKARAINEN M, WOODWARD C, BILLINGHURST M. Augmented assembly using a mobile phone[C]. IEEE International Symposium on Mixed and Augmented Reality, 2008.
- [2] TAKACS G. Outdoors augmented reality on mobile phone using lox-el-based visual feature organization[R]. MIR, 2008.
- [3] NetExpolorateur. What is Layar[EB/OL]. [2012-08-01]. <http://www.layar.com/what-is-layar>.
- [4] Wikitude. The World's leading Augmented Reality SDK[EB/OL]. [2012-11-20]. <http://www.wikitude.org>.

(收稿日期:2013-04-23)

作者简介:

洪晓燕,女,1988 年生,硕士研究生,主要研究方向:信息集成与交换,智能系统。

纪昌红,男,1989 年生,硕士研究生,主要研究方向:图形图像处理,增强现实。

(上接第 44 页)

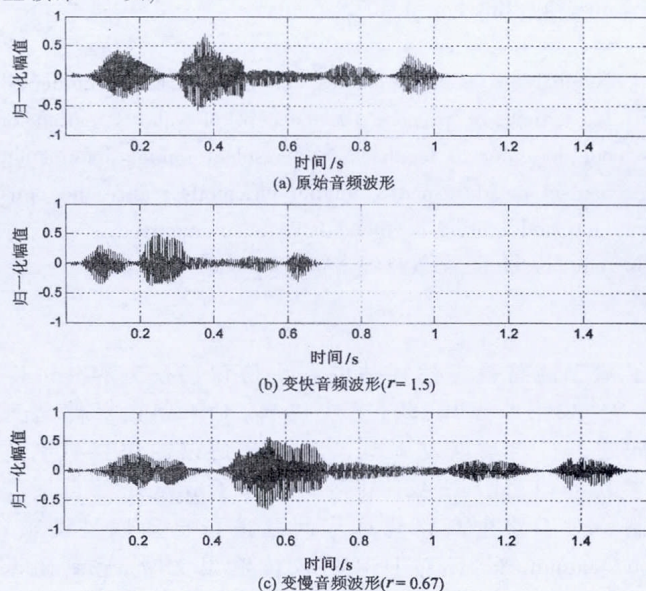


图5 音频变速举例

本文介绍了基于短时傅里叶变换的音频变速算法原理,并针对 DSP 平台对算法进行了优化,减少了音频变速算法计算量。软件仿真测试表明,变速后音频品质良好,并且优化后的算法在 ADSP2188 平台上是可实现的。

参考文献

- [1] MOULINES E, LAROCHE J. Non-parametric techniques for pitch-scale and time-scale modification of speech[J].

Speech Communication, 1995, 16(2): 175-205.

- [2] LAROCHE J, DOLSON M. Improved phase vocoder time-scale modification of audio[J]. IEEE Transactions on Audio and Speech Processing, 1999, 7(3): 1-9.
- [3] 奥本海姆 A V, 谢弗 R W. 离散时间信号处理[M]. 北京: 科学出版社, 2000.

(收稿日期:2013-04-23)

作者简介:

杨洪柏,女,1972 年生,博士,副教授,主要研究方向:机电控制,多媒体技术。

陈蕾蕾,女,1975 年生,博士,讲师,主要研究方向:服务计算,多媒体技术。

欢迎订阅

微型机与应用

半月刊

订阅代号: 82-417

定价: 20 元/本(全年 480 元)