

实时的 Gammatone 听感知滤波器组的 FPGA 实现

贾 瑞¹, 李冬梅²

(1 清华大学 微电子学研究所, 北京 100084; 2 清华大学 电子工程系, 北京 100084)

摘 要: Gammatone 听感知滤波器组能大幅度的提高数字助听器、语音增强和语音识别等语音处理系统的性能, 但因其庞大复杂的计算量以及不可实时的综合方式限制了其实际应用. 设计了一种高效的运算单元结构和复用架构, 提出了一种通过延时补偿的方法简化其综合的方式, 并在 FPGA 上实现了 128 通道的便于综合的可实时处理语音信号的 Gammatone 滤波器组. 通过测试验证, 该设计延时为 20 ms, 满足实时性要求, 能很好地重构语音信号, 并提高语音增强系统的性能.

关键词: Gammatone 滤波器组; 语音增强系统; 延时补偿; 实时性; FPGA

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-7180(2015)01-0035-05

Real-time Implementation of Auditory
Gammatone Filter Bank with FPGAJIA Rui¹, LI Dong-mei²

(1 Institute of Microelectronics, Tsinghua University, Beijing 100084, China;

2 Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Gammatone Filter Bank (GTFB) is perfectly suitable for speech processing systems like digital hearing aids, speech enhancement and speech recognition system. However, the high computational complexity and non-real-time characteristic in the hardware synthesis restrict its practical application. In this paper, an efficient core computation unit structure is proposed. Besides, a low complexity synthesis algorithm is proposed and the real-time characteristic is achieved by compensating the filter delay. The measurement results verified show that a 128-band high-efficient real-time reconfigurable Gammatone Filter Bank is realized in FPGA. The design has only 20 ms delay and improve the dual-microphone CASA speech enhancement system performance.

Key words: Gammatone Filter Bank; speech enhancement system; delay compensating; real-time; FPGA

1 引言

语音处理系统近年来得到了广泛应用和迅速发展, 如数字助听器、语音增强系统和语音识别系统等, 人们对其性能的要求也越来越高. 语音处理系统的核心模块主要由前端子带滤波器组和后端分通道语音处理算法决定. 在同样的分通道语音处理算法的情况下, 子带滤波器组的分频方式和计算效率对整个系统的性能有着至关重要的影响. Gammatone 滤波器组 (Gammatone Filter Bank, GTFB) 是

一种模拟人耳听觉特性的滤波器组, 其中心频率的分布以及每个子带的滤波特性都与入耳基底膜的特性对应^[1]. 无论数字助听器、语音增强系统还是语音识别系统, 在应用 GTFB 作为子带滤波器组后, 其性能都有较大提升^[2-3]. 但在实际应用中, 很多语音处理系统, 尤其是数字助听器和语音增强系统都要求具有很强的实时性, 而在以往的研究中, 由于 GTFB 庞大的计算量以及特殊的综合方式, 传统的 PC 机以及 DSP 无法将其实时地实现, 多是以非实时的方式进行仿真研究.

收稿日期: 2014-03-20; 修回日期: 2014-05-17

基金项目: 国家高技术研究发展计划 (2012AA012301); 国家自然科学基金项目 (61171001)

本文针对 16 kHz 采样率和 16 bit 精度的语音信号,设计一种复用的 128 通道 GTFB 硬件实现结构.提出一种简单的近似相位补偿的综合方式,最终在 Xilinx 公司 Virtex5 XC5VFX110T 的 FPGA 上实现实时的可综合 GTFB,并在双通道 CASA(计算听觉场景分析)语音增强系统中进行测试验证.

2 GTFB 的 FPGA 设计与实现

Gammatone 滤波器(Gammatone Filter,GF)是 Johannesma 在 1972 年提出的^[1],能够很好地模拟人耳基底膜尖锐的滤波特性^[4],是一个定义在时域的带通滤波器,其表达式为

$$g(t, f_c) = A t^{n-1} e^{-2\pi b t} \cos(2\pi f_c t + \varphi) \quad (1)$$

式中, A 为增益系数,一般可取 1; n 为滤波器阶数, n 为 4 时能很好地模拟人耳听觉特性; b 为等效带宽系数 $b = \text{ERB}(f_c)$, $\text{ERB}(f_c) = 24.7 + 0.108 f_c$, φ 为初始相位,一般设为 0; f_c 为中心频率.

GTFB 则是由一系列中心频率不同的 GF 组合形成.其中心频率先在等价矩形带宽(equivalent rectangular bandwidth, ERB) 尺度上做均等划分,然后映射到线性尺度,可确定各个 GF 的中心频率.线性尺度上的频率 f_{LIN} 与 ERB 尺度上频率 f_{ERB} 映射关系为^[5]: $f_{\text{ERB}} = 21.4 \lg(4.37 \times 10^{-3} f_{\text{LIN}} + 1)$. 16 通道的 GTFB 的频响特性如图 1 所示,各通道中心频率呈非线性分布.

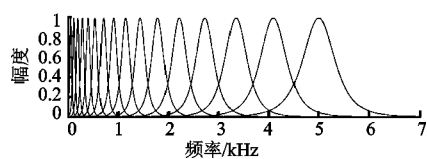


图 1 16 通道的 GTFB 频响特性

本文所设计的 GTFB 为 128 通道,各通道的增益为 1,中心频率分布范围为 80~5 000 Hz,能够覆盖语音信号的绝大部分频率,并且满足语音信号处理系统的需求.

2.1 GTFB 运算单元设计

数字滤波器可以由 IIR 和 FIR 滤波器等方式实现. FIR 滤波器具有严格的线性相频特性,同时由于其非递归结构,能够保证系统稳定.而 IIR 滤波器能用较少的阶数实现更为精确的频响特性. GF 具有精确的数学表达式,可推导出 IIR 的实现方式.同时,心理声学研究表明,人耳听觉对于声音信号的相位信息并不敏感,对滤波器的线性特性要求不高^[6],加之 GF 本身在其通带内就具有近似的线性相位,

完全可以满足系统对相频特性的要求.另外, FIR 需要至少 400 阶以上才能较好地实现 GF 的频响特性,这样在 16 kHz 的典型采样率下,就有至少 25 ms 的延时,而 IIR 只需要 8 阶,考虑到实现过程中加上的流水线延时也只有 12 个样点即 0.75 ms 的延时,实时性能远好于 FIR. 基于以上考虑,本文采用 IIR 滤波器实现 GTFB.

为得到 GF 的传递函数,对式(1)做拉普拉斯变换可得

$$G(s) = \frac{6[(s+B)^4 - 6(s+B)^2\omega^2 + \omega^4]}{[(s+B)^2 + \omega^2]^4} \\ = G_1(s) \cdot G_2(s) \cdot G_3(s) \cdot G_4(s) \quad (2)$$

其中,

$$B = 2\pi b, \omega = 2\pi f_c,$$

而

$$G_n(s) = \frac{A_n(s+z_n)}{s^2 + \frac{\omega_p}{Q}s + \omega_p^2} \quad (n = 1, 2, 3, 4) \quad (3)$$

$$z_n = -(\pm \omega - B \pm \sqrt{2}\omega)$$

$$Q = \frac{1}{2} \sqrt{1 + \left(\frac{f_c}{b}\right)^2}, \omega_p = \sqrt{\omega^2 + B^2},$$

即 $G(s)$ 可由四个级联的二阶滤波器 $G_n(s)$ 组成,对式(3)利用冲激响应不变法进行 Z 变换可得

$$G_n(z) = \frac{[(z_n - \alpha)e^{-aT} \sin(\beta T) - e^{-aT} \cos(\beta T)]z^{-1} + 1}{A_n \frac{\beta}{e^{-2aT}z^{-2} - 2e^{-aT} \cos(\beta T)z^{-1} + 1}} \quad (4)$$

式中, $\alpha = \frac{\omega_p}{2Q}$, $\beta = -\sqrt{\omega_p^2 - \frac{\omega_p^2}{4Q^2}}$, A_n 为增益调节系数.将各级滤波器增益归一化.将式(4)进一步简化表示为

$$G_n(z) = A_n \frac{a_n z^{-1} + 1}{b_2 z^{-2} + b_1 z^{-1} + 1} = A_n \cdot H_n(z) \quad (5)$$

其中,

$$a_n = \frac{(z_n - \alpha)e^{-aT} \sin(\beta T)}{\beta} - e^{-aT} \cos(\beta T),$$

$$b_2 = e^{-2aT}, b_1 = -2e^{-aT} \cos(\beta T).$$

式中, $H_n(z)$ 对应的时域迭代关系为 $y(k) = x(k) + a_n x(k-1) - b_1 y(k-1) - b_2 y(k-2)$ 推广到四级二阶级联的整个通道,实现的框图如图 2 所示.

2.2 GTFB 运算单元及控制系统的 FPGA 实现

滤波器组的实现通常分为并联型、串联型以及折叠型. 并联型实现由于采用并行运算的方式,其计算速率最高,每一个通道独立使用一个计算单元,因

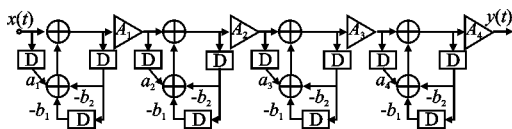


图 2 GTFB 单个通道的结构框图

此满足实时计算所需的时钟频率最低,但是消耗的硬件资源也最多。串联型实现指所有 128 个通道都复用了 1 个计算单元,硬件资源节省了 128 倍,但与此对应的串联实现需要较高的时钟频率。而折叠型的实现是指所有的运算都只用一个乘法器和一个累加器,最大限度地节约了硬件资源,但是需要最高时钟频率来满足计算的实时性。对于 16 kHz 的语音采样率以及 128 通道的 GTFB,需要 320 MHz 的时钟才能满足实时性,系统实现复杂度高。同时采用折叠型的实现难以保留图 2 中的级联式结构。取消级联结构后会增大滤波器系数定点化带来的量化噪声,降低计算精度。对于 IIR 滤波器系统,可能会使在单位圆内部的极点在量化后转移到单位圆外,从而增加系统不稳定的风险。

本文结合以上三种结构提出了一种级联式的折叠结构实现 GTFB。保留了图 2 中的级联结构,同时对每一级的结构做折叠变换,只用一个乘法器和一个累加器实现,从而达到了在保证一定精度的情况下硬件消耗最小化。表 1 所示为四种实现方式所需的资源和时钟频率。图 3 中描述的是用折叠的方式实现单级运算单元的结构。

表 1 四种实现方案资源占用及时钟频率对比

	并联	串联	折叠	本文
乘法器	1 536	12	1	4
加法器	1 536	12	1	4
时钟频率/kHz	16	11	320	80

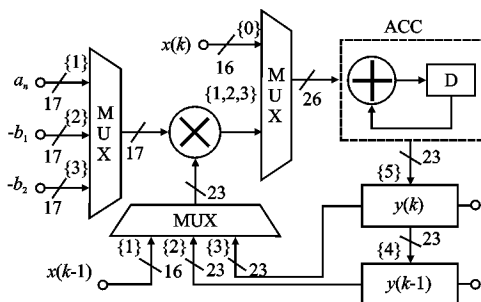


图 3 GTFB 单级运算单元结构

128 通道的 GTFB 通过复用上述基本运算单元来实现。128 个通道的系数各不相同,设计中将各通道系数存储在片内 ROM 中,计算过程中产生的中

间数据存储在片内 RAM 中,以各个通道的序号作为存储器地址。当检测到有输入信号来临时,核心控制单元在主时钟作用下顺序输出地址信息和片选信号,使存储器输出相应的系数和中间数据。同时核心控制单元输出每个地址期间,输出对应的四个基本控制状态,使运算单元在等待新系数和中间数据、读取新系数和中间数据、运算产生新数据、输出并存储新数据四个状态之间转换,从而完成一个通道数据计算。在核心控制单元顺序输出所有的地址信息之后,运算单元也完成了对 128 个通道数据的计算,此时运算单元和核心控制单元停止工作以节省功耗,等待下一个输入信号的到来。

2.3 可实时的综合方案设计

完整的滤波器组包括分析滤波器组和综合滤波器组。在 ERB 尺度下划分中心频率的 GTFB 有以下性质:

$$\sum_{m=1}^M |G_m(e^{j\omega})|^2 = \sum_{m=1}^M G_m(e^{j\omega}) G_m(e^{-j\omega}) \approx c \quad (6)$$

式中, M 为滤波器组的通道个数, $G_m(e^{j\omega})$ 为 m 通道滤波器的频响特性, c 是常实数^[7]。因此当语音信号 $x(k)$ 通过 Gammatone 滤波器 $G_m(z)$ 得到 $y_m(k)$ 后,再通过一个系统函数为 $G_m(z^{-1})$ 的滤波器得到 $s_m(k)$,相加即可近似重构 $x(k)$ 。但 $G_m(z^{-1})$ 是非因果系统,因此在 GTFB 的非实时应用研究中,采用将 $y_m(k)$ 在时域上反转,再将 $y_m(-k)$ 通过 $G_m(z)$ 得到 $s_m(-k)$,反转之后得到 $s_m(k)$,以此等效 $G_m(z^{-1})$ 系统。如图 4 所示。由于需要将整个语音信号在时域反转,因此无法实时处理语音。

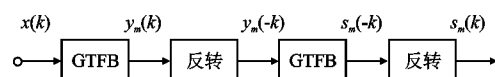


图 4 GTFB 非实时综合的流程

通过对 GTFB 频率特性的研究,本文提出了一种通过延时补偿近似实时重构输入信号的方法。类似于式(6),GTFB 也满足如下性质:

$$\sum_{m=1}^M |G_m(e^{j\omega})| \approx \lambda \quad (7)$$

式中, λ 是一个常实数。因此,当各个通道的 GTFB 具有相似的相频特性时,就使得

$$\left| \sum_{m=1}^M G_m(e^{j\omega}) \right| \approx \lambda' \quad (8)$$

研究发现 GF 的相频特性具有近似线性的特点,其群延迟可近似看作一个常数。不同通道的群延迟不同,如图 5 所示,通道 1 和通道 2 的群延迟分别为 τ_1 和 τ_2 ,因此在时域上表现为两个通道输出信号包络

的延时也分别为 τ_1 和 τ_2 。

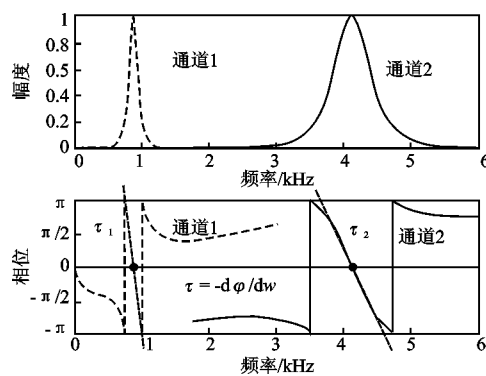


图 5 GTFB 不同通道的相位特性

由于各个通道的延时不同,若将各个通道的输出直接相加作为综合方式,会使综合后整体幅频特性产生较大畸变,如图 6(a)所示.由各个通道不同的群延迟可知,各通道的输出在时域上的不同延时,根据这些不同的延时,通过对各个通道增加相应数目的寄存器,即在时域上对各个通道进行延时补偿,使各通道的延时相同.经延时补偿后,直接对各通道输出相加完成综合,能使整体的幅频特性趋于平坦,即在所需频带内具有全通特性,如图 6(b)所示,从而达到重构输入语音信号的目的.

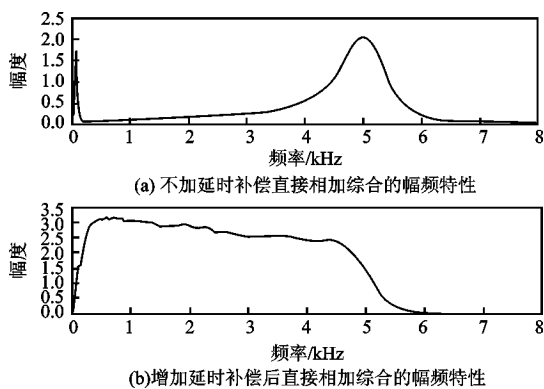


图 6 增加延时补偿前后 GTFB 整体幅频特性的对比

对于 128 通道的 GTFB 需要补偿的最大延时为 17.6 ms,加上分析滤波器组的理论延时 0.75 ms,最终本设计的理论延时为 18.35 ms.通过对所实现的系统进行测试,输入的延时在 20 ms 左右,能够很好地满足系统对实时性处理的要求.

3 系统验证及结果分析

本系统经过在 Matlab 中进行浮点数仿真以及定点化验证后,采用 Xilinx Virtex5 XC5VFX110T 的 FPGA 实现,整个系统所占用资源如表 2 所示.

表 2 GTFB 实现占用 FPGA 资源情况

	使用数	使用率/%
Slice Registers	5 776 个	8
Slice LUTs	5 184 个	7
Occupied Slices	2 465 个	14
Block RAM/FIFO	1 114 kB	20
DSP48Es	8	12

为了验证本文 FPGA 实现的可用性,从 Timit 标准语音库中随机抽取了包括中英语音在内的 5 组语音信号和多种单频信号分别作为输入进行测试,并通过逻辑分析仪测得输出数据,经 Matlab 进行分析验证.

3.1 关于系统实时性的验证测试

通过对输出信号与输入信号波形进行对比,测量系统对于输入信号的延时全部在 20 ms 左右,部分测试结果如表 3 所示.

表 3 系统延时测量结果

语音	sx198	sx144	si1295	si486	1 kHz
延时/ms	20.3	19.1	19.4	20.8	18.7

根据对人耳听觉特性的相关研究,20ms 左右的延时不会影响人耳的听觉感受.测试结果表明,该设计具有较小的延时、良好的实时性,能满足数字助听器、语音增强系统对于分子带滤波器实时性的要求.

3.2 关于语音信号重构性的验证测试

PESQ (Perceptual evaluation of speech quality) 主观语音质量评估是 ITU-T P. 862 建议书提供的客观 MOS 值(衡量通信系统语音质量的重要指标)评价方法^[8].能够用于评价未经处理的语音信号和经过语音处理系统后的语音信号的质量评估. PESQ 的分数在 4.5 到 1.0 之间,4.5 表示没有失真,1.0 表示失真情况最为严重.

测试采用上述 Timit 标准语音库中的五组语音,通过 GTFB 处理后,将各通道信号相加综合后与原始语音进行 PESQ 对比评测,测试结果如表 4 所示.

表 4 输出信号与输入信号 PESQ 评分

语音	sx198	sx144	si1295	si486	Chinese
PESQ	4.318	4.283	4.287	4.336	4.304

测试结果的 PESQ 评分大都在 4.3 左右,表明该 GTFB 设计能对输入语音信号实现高质量重构.

3.3 应用于双麦克风 CASA 语音增强系统的效果

双麦克风 CASA 语音增强系统有主从两个麦克风

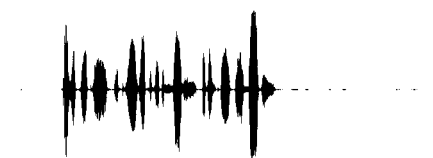
风采集语音信号,经分子带滤波器后,采用 CASA 算法处理,从而提高主麦克风语音的信噪比,实现语音增强.图7表示为主麦克风的原始语音波形以及应用不同的分子带滤波器处理后的波形.图7(c)和图7(d)分别表示 PC 机仿真的非实时 GTFB 和本文 FPGA 实现的实时 GTFB 应用于语音增强系统的结果,与图7(b)中应用 WOLA(一种常用的分子带滤波器组)滤波器组的语音增强系统相比,降噪效果更明显,处理后的语音更清晰.经测量,图7(b)中将主麦克风的信噪比提升 18.31 dB,图7(c)中能提升 29.3 dB,而图7(d)中能提升 27.5 dB,图7(a)中的降噪明显优于图7(b)中的降噪,且与图7(c)中的降噪相当.另外,从主观上去评价处理后语音的可懂度方面,应用 FPGA 实现的 GTFB 系统效果优于 WOLA 滤波器组,与非实时仿真的 Gammatone 滤波器效果类似.



(a)主麦克风原始语音波形



(b)应用 WOLA 滤波器组处理的语音波形



(c)应用非实时的 Gammatone 滤波器组处理的语音波形



(d)应用本文的 Gammatone 滤波器组处理的语音波形

图7 原始语音和处理后的语音波形对比

4 结束语

本文通过对 Gammatone 滤波器组频响特性的研究,完成了基于 FPGA 的系统实现,并对其运算结构进行优化和分析,提出一种级联的折叠结构,同

时针对 Gammatone 滤波器组复杂且不可实时的综合方式^[9].采用了一种通过延时补偿的方法简化综合方式,使延时减小到约 20 ms,重构信号的 PESQ 评分为 4.3 左右.该滤波器在语音增强系统当中的测试表明,对其性能有明显提升.本文的设计解决了 Gammatone 滤波器组的实时性实现,使 Gammatone 滤波器组可以真正应用于实时系统.

参考文献:

- [1] Johannesma P I M. The pre-response stimulus ensemble of neurons in the cochlear nucleus[C]// IPO Symposium on Hearing Theory. IPO, Eindhoven, Netherlands, 1972: 58-69.
- [2] 董保帅. 近距离双麦克风语音增强算法研究[D]. 北京: 清华大学电子系, 2012.
- [3] Qi J, Wang D, Jiang Y, et al. Auditory features based on gammatone filters for robust speech recognition[C]// Circuits and Systems (ISCAS), 2013 IEEE International Symposium on. China: Beijing, IEEE, 2013: 305-308.
- [4] 陈世雄, 宫琴, 金慧君. 用 Gammatone 滤波器组仿真人耳基底膜的特性[J]. 清华大学学报: 自然科学版, 2008, 48(6): 1044-1048.
- [5] 俞一彪. 基于互信息理论的说话人识别研究[D]. 上海: 上海大学, 2004.
- [6] Kuo Y T, Lin T J, Li Y T, et al. Design and implementation of low-power ANSI S1. 11 filter bank for digital hearing aids[J]. Circuits and Systems I: Regular Papers, IEEE Transactions on, 2010, 57(7): 1684-1696.
- [7] Lin L, Holmes W H, Ambikairajah E. Auditory filter bank inversion[C]// Circuits and Systems, ISCAS 2001. The 2001 IEEE International Symposium on. Australia: Sydney, IEEE, 2001(2): 537-540.
- [8] 凌渝. 基于 PESQ 算法的语音质量客观评价方法研究[J]. 警察技术, 2011(5): 011.
- [9] 于增辉, 黑勇, 陈黎明, 等. 多通道数字助听器算法及低功耗 VLSI 设计[J]. 微电子学与计算机, 2012, 29(4): 14-18.

作者简介:

贾瑞 男, (1988-), 硕士研究生. 研究方向为语音信号处理及数字电路设计. E-mail: rui.jia_thu@gmail.com.

李冬梅 女, (1966-), 副教授, 博士生导师. 研究方向为模拟及数模混合集成电路设计、SOC 集成电路设计、语音信号处理及芯片设计.