

基于自适应 Mel 滤波器组的 MFCC 特征提取的 SOC 设计

陶佰睿^{1,2}, 郭 琴¹, 苗凤娟^{1,2}, 李青龙¹

(1. 齐齐哈尔大学 通信与电子工程学院, 黑龙江 齐齐哈尔 161006; 2. 中国科学院 上海技术物理研究所, 上海 200083)

摘 要: 说话人声纹身份认证技术中的关键是特征参数的准确性和模式识别的速率. 为此, 对识别对象的性别予以区分, 并进行参数可自适应调整的 Mel 滤波器组设计, 即通过 Quartus II 平台在 Altera 的 DE2 系列型号为 EP2C35F672C6 的开发板上完成高效率说话人声纹特征提取的 SOC(片上系统)设计. 设计具体步骤如下: 首先, 设计截止频率为 400 Hz 和 200 Hz 的低通滤波器以完成男女生基音频率的检测; 然后, 依据计算出的每一帧语音频谱的频率范围确定 Mel 滤波器组的最高频率并完成参数设计; 最后, 在 Quartus II 平台上完成 Verilog-HDL 代码设计, 并封装为 IP 核完成 SOC 设计以及编译、仿真和下载验证. 结果表明, Mel 滤波器组利用率的提高有利于提高特征参数的准确性和识别速度.

关键词: 声纹身份认证; 自适应梅尔滤波器组; 性别识别; 片上系统

中图分类号: TP391.42 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2016.03.003

0 引言

互联网的普及促进了我国信息产业的快速发展, 人们通过互联网进行日常生活和生产变得越来越便利, 但互联网本身的安全问题也引起了人们对用户个人信息安全的关注. 声纹识别相对其他生物识别技术具有语音获取便捷、易于被人们接受和准确度高优点, 而声纹特征提取的准确度和提取速度是声纹识别技术推广应用的关键^[1-3].

文献[4]针对说话人识别中梅尔频率滤波器组结构分布不均匀的问题, 在低频区域密集而在中、高频分布稀疏, 提出适用于说话人识别的改进 Mel 滤波器与 MidMel 滤波器组相结合得到两种混合特征参数. 文献[5]提出对传统的提取 Mel 频率倒谱系数(MFCC)过程中计算 FFT 这一步骤进行频谱重构, 对频谱进行噪声补偿重建, 使之具有很好的抗噪性. 文献[6]利用人类听觉模型提出一种语音感知哈希算法, 该算法对倒谱系数(MFCC)算法中每帧的滤波器数量进行控制, 得到每帧语音的自适应梅尔频率倒谱系数(MFCC). 传统提取语音特征参数中的 Mel 滤波器组

的频率范围是不可变的, 这样就会导致 Mel 滤波器组的利用率降低^[7-9]. 而笔者提出的自适应 Mel 滤波器组能够对每帧的滤波器组的频率范围进行调整.

首先, 依据男女生基音频率的不同, 设计 2 个截止频率不同的低通滤波器对识别人的性别进行区分. 其次, 考虑到每一帧的语音频谱的频率范围的不同, 为提高 Mel 滤波器组的利用率, 加快识别速度, 笔者设计的 Mel 滤波器组参数可自适应调整, 主要依据每一帧语音频谱的频率范围确定 Mel 滤波器组的最高频率来达到自适应的目的. 最后, 在 Quartus II 平台上完成 Verilog 代码编写、Modelsim 时序仿真以及板级的下载验证, 封装 IP 核完成 SOC 设计.

1 总体设计

男女生生理结构的不同决定了声音听觉特性的不同, 男生基音频率在 100 ~ 200 Hz 之间, 女生基音频率在 200 ~ 350 Hz 之间. 因此笔者设计截止频率为 400 Hz 和 200 Hz 的低通滤波器来对男女生进行区别.

收稿日期: 2015-10-04; **修订日期:** 2015-12-01

基金项目: 国家自然科学基金资助项目(61204127); 黑龙江省自然科学基金资助项目(F201438); 黑龙江省教育厅科学技术研究项目(12541899, 12531774); 黑龙江省教育科学“十二五”规划 2014 年度课题(GBC1214089); 黑龙江省高校科技成果产业化前期研发培育资助项目(1254CGZH04)

作者简介: 陶佰睿(1972—), 男, 吉林洮南人, 齐齐哈尔大学教授, 博士, 主要从事集成电路设计和声纹识别技术研究, E-mail: Tbr_sir@163.com.

同一个说话人的每一帧的语音频谱是有所区别的,传统的 Mel 滤波器组的频带范围为 0 ~ 4 000 Hz,并不是针对每一帧的语音频谱设计的,这样会造成传统 Mel 滤波器组中的个别滤波器并没有被使用.因此,笔者将依据每一帧语音频谱的

频率范围确定 Mel 滤波器组的最高频率,不但可使 Mel 滤波器组的参数进行自适应调整,而且提高了 Mel 滤波器组的利用率,对声纹识别率的提高有很大作用.总设计流程图如图 1 所示.

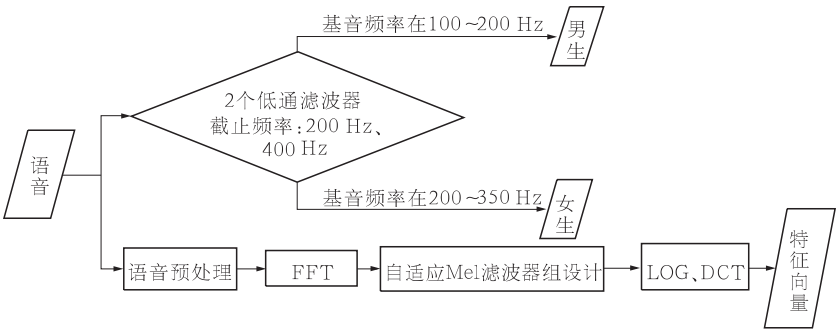


图 1 总体设计流程图

Fig.1 Flow chart of overall design

2 低通滤波器设计

2.1 参数设计

在 Matlab R2014a 的 FDATool 界面中设置响应类型为低通滤波器,使用汉宁窗进行设计,采样频率设置为 8 000 Hz,截止频率分别设置为 400 Hz 和 200 Hz,得到时域图分别如图 2 和图 3 所示,分别得到 2 组滤波器系数,都为 12 阶,见表 1 所示.

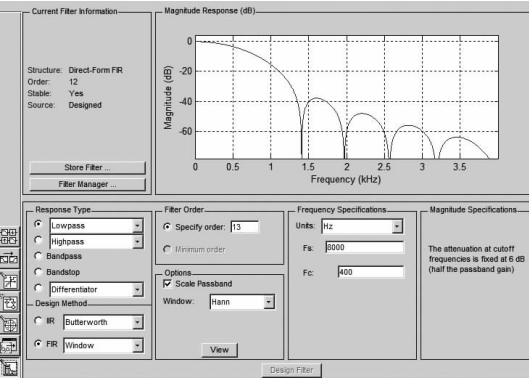


图 2 截止频率为 400 Hz

Fig.2 Cutoff frequency of 400 Hz

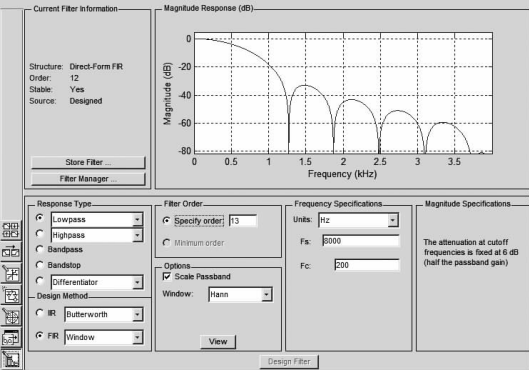


图 3 截止频率为 200 Hz

Fig.3 Cutoff frequency of 200 Hz

表 1 2 组滤波器系数

Tab.1 Filter coefficients of two sets					
序号	第 1 组	第 2 组	序号	第 1 组	第 2 组
1	0.005	0.007	7	0.165	0.154
2	0.025	0.031	8	0.141	0.136
3	0.059	0.065	9	0.102	0.103
4	0.102	0.103	10	0.059	0.065
5	0.141	0.141	11	0.025	0.031
6	0.165	0.154	12	0.005	0.007

2.2 Verilog HDL 代码设计

由低通滤波器的传递函数可以得到当滤波器的阶数为 12 时的表达式,可通过式(1)得出:

$$y(n)=c_0 \cdot x(n)+c_1 \cdot x(n-1)+\cdots+c_{11} \cdot x(n-11). \quad (1)$$

式中: c_0 等为滤波器的抽头系数; n 为时刻; x 为序列; y 为滤波器输出.

由表达式(1)可得出,滤波器的输出是由序列 x 中的 12 个序列点分别与对应的 12 个系数进行相乘再累加的结果.因此编写 Verilog 代码,包括乘法模块、加法模块、控制模块.综合后的 RTL 视图如图 4 所示.图中端口 A 中存放的是序列 x ,端口 B 中存放的是 12 组系数.

3 自适应 Mel 滤波器组设计

3.1 Mel 滤波器组的频带设计

采样率设置为 8 kHz,量化位数设置为 16 bit,由于语音的短时平稳性,帧长设置为 20 ms.对每一帧语音信号进行 FFT 变换,可以得到每一帧语音序列的频谱,不同帧的语音频谱应该是不相同的.因此,计算每一帧语音频谱作为 Mel 滤波器组的最高频率值.设计流程如图 5 所示.

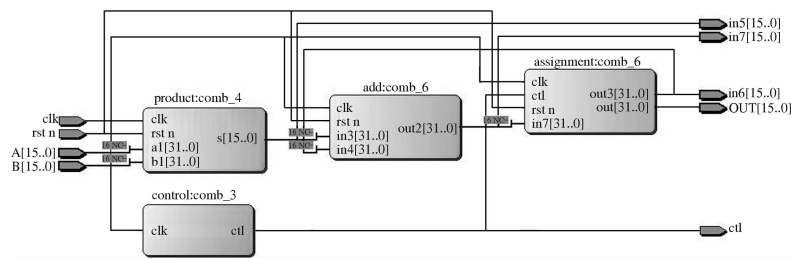


图 4 FIR 的 RTL 视图

Fig. 4 RTL view of FIR

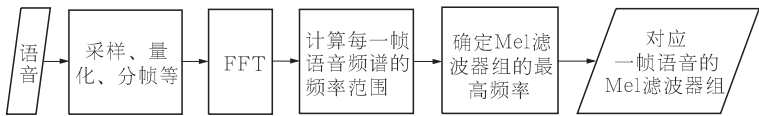


图 5 自适应 Mel 滤波器组的设计流程

Fig. 5 Design flow of adaptive Mel filter banks

3.2 Verilog HDL 代码设计

MFCC 是基于人耳的听觉特性的一种参数,可通过式(2)得出.根据 Zwicker 的研究可知,临界频率带宽随着频率的变化而变化,并与 Mel 频率的增长一致,在 1 000 Hz 以下大致呈线性分布,带宽为 100 Hz 左右;在 1 000 Hz 以上呈对数增长.

Mel(f) = 2 595log₁₀(1 + f/700). (2)

式中:f为线性频率,Hz;Mel(f)为 Mel 频率,mel.

Mel 滤波器组的频率响应函数表达式可通过式(3)得出:

W_l(k) = { k - o(l) / (c(l) - o(l)), o(l) ≤ k ≤ c(l); h(l) - k / (h(l) - c(l)), c(l) ≤ k ≤ h(l). (3)

式中:o(l)为第 l 个三角形滤波器的上限截止 Mel 频率,mel;c(l)为第 l 个三角形滤波器的中心 Mel 频率,mel;h(l)为第 l 个三角形滤波器的下限截止 Mel 频率,mel;k 为在 o(l)和 c(l)之间的整数.

由表达式(3)可看出,每个三角形滤波器都是对称的,选择右边边的表达式,绘制成图形如图 6 所示.

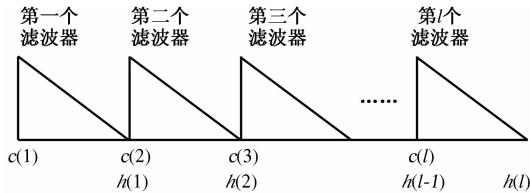


图 6 三角形滤波器组波形

Fig. 6 Waveforms of triangle filters

采样率 f_s 设置为 8 000 Hz,滤波器个数 p 为 18,FFT 变换点数 n 为 160,f 为一帧语音频谱的频率范围,f 决定了 Mel 滤波器组的最高频率,Mel 滤波器组的输出用 W1、W2、⋯、Wl 等表示,inc 表

示下限截止 Mel 频率与中心 Mel 频率的间隔.设计流程如图 7 所示.

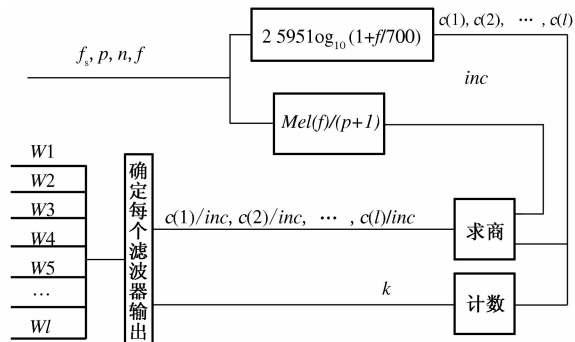


图 7 Mel 滤波器组的设计流程图

Fig. 7 Design flow of Mel filter banks

4 编译仿真与下载验证

4.1 Modelsim 时序仿真

编写 testbench 文件并输入激励源,调用 Modelsim 进行时序仿真.当每帧语音的频谱范围小于传统 Mel 滤波器组的梅尔频率 2 146 mel 时,在相同的时间间隔内自适应滤波器组中的最后一个滤波器实现输出的时间会远小于传统三角形滤波器组中的最后一个滤波器实现输出的时间.随机选取第 10 帧语音序列,得到自适应 Mel 滤波器组和传统 Mel 滤波器组输出分别见图 8 和图 9.时间间隔为 2 μs 内的自适应 Mel 滤波器组的最后一个滤波器在接近 960 ns 时输出为 0,而传统 Mel 滤波器组在接近 1.8 μs 后才输出为 0.可以看出,传统三角形滤波器组中的个别滤波器是不需要被使用的,而自适应三角形滤波器组可以提高滤波器的利用率.

4.2 MATLAB 性能仿真

为验证特征参数的准确性,利用 Matlab 中自带的神经网络工具箱,设计一个根据 10 个人的不同语音帧的特征参数来确定出其中的一个人的神

神经网络系统. 神经网络的输入数据为特征参数, 本文中包括 2 类, 传统 Mel 滤波器组提取的特征参数和自适应 Mel 滤波器组提取的特征参数, 目标

输出为 0 或 1, 0 代表不是该人, 1 代表是此人. 最终得到的误差率如图 10 和 11 所示. 图中, 误差 = 预测值 - 实际值.

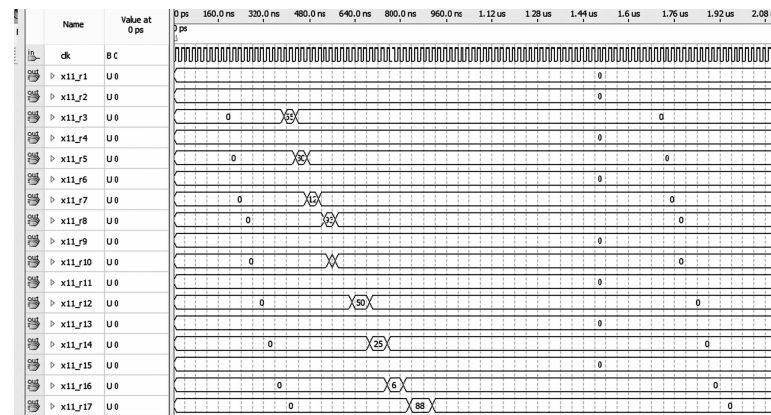


图 8 对应第 10 帧语音的自适应 Mel 滤波器组的 Modelsim 仿真

Fig.8 Modelsim simulation of adapted Mel filter banks corresponding to the first 10 frame voice

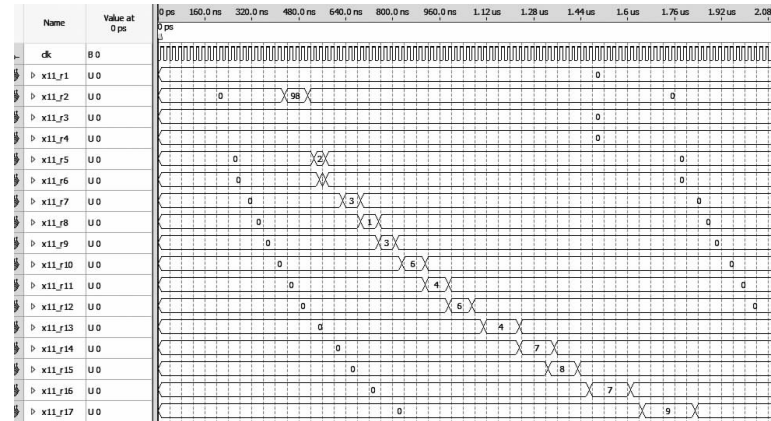


图 9 对应第 10 帧语音的传统 Mel 滤波器组的 Modelsim 仿真

Fig.9 Modelsim simulation of traditional Mel filters corresponding to the first 10 frame voice

由图 10 和图 11 可知, 利用自适应 Mel 滤波器组提取出的特征参数相对传统 Mel 滤波器组进行识别时, 误差率更小, 准确性得到了提高.

4.3 编译

在 QUARTUS II 平台中编写 Verilog 代码, 在时序仿真、综合之后进行编译, 消耗的硬件资源如图 12 所示.

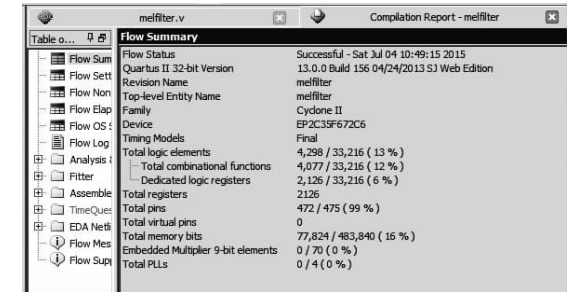


图 12 编译报告

Fig.12 Compiled report

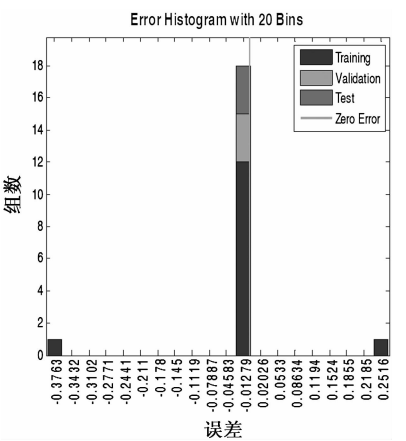


图 10 自适应滤波器的误差

Fig.10 Error of adapted filters

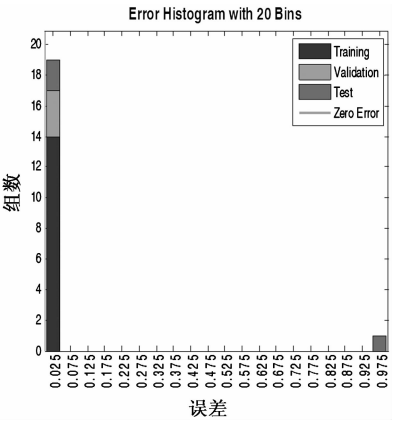


图 11 传统滤波器的误差

Fig.11 Error of traditional filters

4.4 生成 IP 核与下载验证

利用 SOPC Builder 的自定义 IP 技术, 在 QUARTUS II 平台添加 VERILOG 文件、CPU、时钟、定时器等组件, 生成 IP 核, 如图 13 所示.



图 13 SOPC Builder 组件

Fig.13 Component of SOPC Builder

5 结束语

根据每一帧语音频谱的频率范围来确定每一

个 Mel 滤波器组的频带范围,可以有效地提高 Mel 滤波器组的利用率,避免了有些频段的滤波器未被使用情况的发生,有助于提高特征参数的准确性以及识别率,最后通过了硬件验证. 因此,利用 IP 核技术完成 SOC 设计,可方便模块调用,以及减少硬件资源的使用,提高设计的开发速度.

参考文献:

[1] HIRAYAMA N, YOSHINO K, ITOYAMA K, et al. Automatic speech recognition for mixed dialect utterances by mixing dialect language models [J]. IEEE ACM Transactions on audio speech and language processing, 2015, 23(2):373-382.

[2] 陈晨, 胡贝贝. 基于语音识别技术的机载短波应急通信[J]. 电讯技术, 2014(01):6-11.

[3] DUFOUR R, Estève Y, Deléglise P. Characterizing and detecting spontaneous speech: application to speaker role recognition[J]. Speech communication, 2014, 56(1):1-18.

[4] 黄锐, 陆安江, 张正平. 一种改进型的 MEL 滤波器

混合特征参数提取方法研究[J]. 通信技术, 2014 (12):1388-1391.

[5] 胡政权, 曾毓敏, 宗原, 等. 说话人识别中 MFCC 参数提取的改进[J]. 计算机工程与应用, 2014(7): 217-220.

[6] 黄羿博, 张秋余, 袁占亭, 等. 融合 MFCC 和 LPCC 的语音感知哈希算法[J]. 华中科技大学学报(自然科学版), 2015, 43(2):124-128.

[7] SHARMA R P, FAROOQ O, KHAN I. Wavelet based sub-band parameters for classification of unaspirated hindi stop consonants in initial position of cv syllables [J]. International journal of speech technology, 2013, 16(3):323-332.

[8] ZENG F Z, ZHOU H. Speaker recognition based on a novel hybrid algorithm [J]. Procedia engineering, 2013, 61:220-226.

[9] MAHMOOD A, ALSULAIMAN M, MUHAMMAD G. Automatic speaker recognition using multi-directional local features (mdlf) [J]. Arabian journal for science and engineering, 2014, 39(5):3799-3811.

Design of MFCC Feature Extraction Based on Adaptive Mel Filter Banks for SOC Application

TAO Bairui^{1,2}, GUO Qin¹, MIAO Fengjuan^{1,2}, LI Qinglong¹

(1. Computing Center, Qiqihar University, Qiqihar 161006, China; 2. National Laboratory for Infrared Physics, Shanghai Institute of Technical Physics, Shanghai 200083, China)

Abstract: The accuracy of characteristic parameter and pattern recognition rate among speaker voiceprint authentication technologies are important. In this paper, adaptive Mel filter banks are designed after the recognition of the gender, and the SOC (system-on-chip) design of high efficiency speaker voiceprint feature extraction is completed on the EP2C35F672C6 development board of Altera's DE2 series. First of all, two low-pass filters cutoff frequency of 200 Hz and 400 Hz are designed to complete the pitch frequency detection of male and female students. Then, the parameters of Mel filter banks are calculated by the highest frequency deduced from the frequency range of speech spectrum. Then, Verilog-HDL code encapsulated as IP core for SOC design, compilation, simulation, and download authentication are finished on the Quartus II platform. The results show that adaptive Mel filter banks can improve both the accuracy of characteristic parameters and the speed of recognition.

Key words: voiceprint authentication; adaptive mel filter banks; gender recognition; SOC