

文章编号:1001-9081(2009)02-0389-03

混合窗函数和子带频谱质心在 MFCC 特征提取过程中的应用

赵欢¹, 张林¹, 陈珍文²

(1. 湖南大学 计算机与通信学院, 长沙 410082; 2. 北京邮电大学 计算机科学与技术学院, 北京 100876)

(CZW198224@163.com)

摘要:为改善低信噪比环境下语音的质量,在传统 MFCC 特征提取的基础上,提出了两种提高识别系统鲁棒性的方法。一种方法利用混合窗函数对旁瓣的抑制来提高系统的鲁棒性;另一种方法是基于频谱峰值位置受背景噪声影响相对较小,将子带幅度信息和 Mel 子带频谱质心(MSSC)相结合。实验表明混合窗函数和子带频谱质心(MSSC)以及它们相结合的系统与使用传统 MFCC 的基准系统相比,在低信噪比的平稳噪声环境下系统的鲁棒性得到了一定的提高。

关键词:语音识别;Mel 倒谱系数;低信噪比;子带频谱质心

中图分类号:TP391.4 **文献标志码:**A

Using mixed window function and subband spectrum centroid in MFCC feature extraction process

ZHAO Huan¹, ZHANG Lin¹, CHEN Zhen-wen²

(1. College of Computer and Communication, Hunan University, Changsha Hunan 410082, China;

2. School of Computer Science and Technology, Beijing University of Post and Telecommunications, Beijing 100876, China)

Abstract: In order to improve the quality of speech in low SNR, two methods were proposed to improve the robustness of the system in this paper based on the traditional MFCC feature extraction. One is to use the side lobe suppression of mixed window function to improve the robustness of system; the other is to incorporate subband amplitude information with Mel-subband spectrum centroid(MSSC) because spectral peak position remains practically unaffected in the presence of background noise. Experimental results show that mixed window function and MSSC and their combination system could improve the robustness of system compared to the benchmark system based on traditional MFCC in the low SNR of stationary noises.

Key words: speech recognition; MFCC; low signal to noise ratio; subband spectrum centroid

0 引言

如何在噪声环境中特别是低信噪比的情况下保持语音识别系统的性能仍然是一个难题。目前所采用的提高噪声环境下语音鲁棒性的方法通常可以分为三类:鲁棒特征提取、语音增强和模型补偿。前两个方法主要针对特征提取过程研究去除噪声干扰的方法,主要有:短时修正相关法(Short-term Modified Coherence, SMC)^[1], PLP 特征^[2], RAS-MFCC^[3], 倒谱均值减^[4], 谱减法^[5], 维纳滤波法^[6], 等等。模型补偿主要是通过模型补偿的方法减少训练环境与测试环境的不匹配。

鲁棒语音特征通过寻找对噪声影响不敏感的语音特征来提高系统的性能,不需要估计环境噪声的特征,因此在实际中得到广泛应用。Mel 倒谱系数(Mel-Frequency Cepstrum Coefficients, MFCC)在一定程度上模拟了人耳某些听觉机理,在有信道噪声和频谱失真的情况下具有较好的稳健性,大量研究表明,噪声环境下基于 MFCC 特征提取的系统可以得到较高的识别率。

近年来子带语音识别得到了广泛研究^[7-8], Paliwal 的研究表明子带频谱质心(Subband Spectrum Centroid, SSC)非常接近频谱中的峰值位置^[9],而频谱峰值位置信息可以提高语音识别系统的性能。由于频谱峰值位置受背景噪声的影响相

对较小,因此基于 SSC 的语音前端处理能够提高语音识别系统的鲁棒性。近几年来,许多研究者把 SSC 作为 MFCC 的附加特征或作为基于 SSC 的新特征矢量^[9-11],一定程度上提高了语音识别系统的性能。但是这些方法中,频谱峰值位置信息与幅度信息的利用是相互独立的。

本文提出了一种新的方法,将混合窗函数和子带频谱质心(MSSC)与基于 MFCC 的语音前端处理相结合,实验表明,该方法与使用传统 MFCC 的识别系统相比,提高了语音识别系统的鲁棒性。

1 MFCC 前端处理的改进

1.1 基于混合窗函数的语音前端处理

语音信号是一种非平稳的时变信号,其产生过程与发声器官的运动紧密相关。通过对发声机理的研究了解到,发声器官的状态变化速度较声音振动的速度要缓慢得多,因此语音信号是短时平稳的,每个短时平稳的语音段称为一个语音帧。通常我们采用一个长度有限的窗函数来截取语音信号形成语音帧。理想窗函数的频率响应要求主瓣无限狭窄且没有旁瓣,但这种窗函数在实际工程中无法实现。在标准的 MFCC 提取过程中,采用的是汉明(Hamming)窗函数。

一个 N 点的汉明(Hamming)窗函数定义如式(1):

收稿日期:2008-09-02;修回日期:2008-10-18。

基金项目:湖南省科技计划项目(05FJ3046);湖南省财政厅项目(湘财教指[2006]52号)。

作者简介:赵欢(1967-),女,湖南长沙人,教授,博士,主要研究方向:嵌入式系统设计与仿真、语音信息处理;张林(1980-),女,湖南株洲人,硕士研究生,主要研究方向:语音信号处理;陈珍文(1982-),男,苗族,湖南湘西人,硕士研究生,主要研究方向:语音信号处理、数据挖掘。

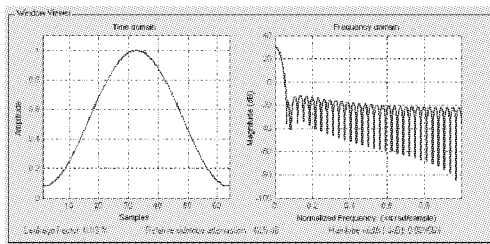
$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N-1}\right), 0 \leq n \leq N \quad (1)$$

考虑在窗函数的主瓣宽度不变的情况下,而使得旁瓣能够更好的抑制,本文选用一种混合窗函数来代替汉明窗函数,其定义如式(2):

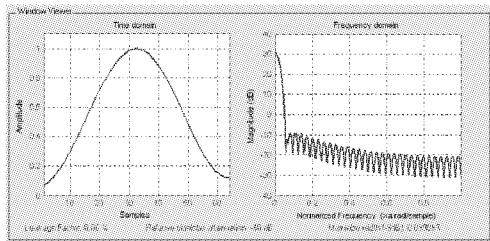
$$w(n) = \begin{cases} 0.42 - 0.36\cos\left[\frac{2\pi(n-1)}{N-1}\right] + 0.22\sin\left[\frac{\pi(n-1)}{N-1}\right], & 1 \leq n \leq N/2 \\ 0.56 - 0.44\cos\left[\frac{2\pi(n-1)}{N-1}\right], & \frac{N+1}{2} \leq n \leq N \end{cases} \quad (2)$$

公式中所有的系数是根据实验得出的经验值。图 1 是汉明窗与此混合窗在时域与频域特性的比较。可见,混合窗主瓣频带和汉明窗的一样宽,但其对旁瓣的抑制比汉明窗强。

加窗后的语音帧经过快速傅立叶变换(FFT)得到其频谱,Mel 频率用以模拟耳蜗(cochlea)的频率响应,语音频谱的幅度或能量通过 Mel 域滤波器组得到 Mel 域滤波器组幅度或能量。



(a) 汉明窗函数



(b) 混合窗函数

图 1 汉明窗与混合窗在时域与频域的特性比较

1.2 基于子带频谱质心(MSSC)的 Mel 滤波器组处理

假设频段 $[0, F_s/2]$ 被分成 M 个子带,其中 F_s 是语音信号的采样频率。对第 m 个子带,假设它的最低和最高频率边界分别为 l_m 和 h_m ,设子带滤波器为 $w_m(f)$,频率 f 处的能量为 $P(f)$ 。根据 Paliwal 的研究^[9],第 m 个子带的质心由式(3)计算得到:

$$C_m = \frac{\int_{l_m}^{h_m} f w_m(f) P^\gamma(f) df}{\int_{l_m}^{h_m} w_m(f) P^\gamma(f) df} \quad (3)$$

其中 γ 是一个经验值。根据 Bojana 等人的实验表明^[11],当 γ 为 1 时,系统可以取得较好的性能,因此,我们在实验中采用同样的数值。

为了将子带频谱质心信息和传统的基于 MFCC 的前端处理结合起来,在本文中, M 为 Mel 滤波器组的个数, f 采用的是 Mel 频率, $w_m(f)$ 为传统 MFCC 提取方法中使用的三角滤波器,而 $P(f)$ 为 $w_m(f)$ 与 f 的乘积, l_m 和 h_m 分别为第 m 个三角滤波器的下限和上限频率,根据我们实验的实际情况,我们将上述的公式变形为式(4):

$$C_m = \frac{\sum_{f=l_m}^{h_m} f w_m^2(f) |X(f)|}{\sum_{f=l_m}^{h_m} w_m^2(f) |X(f)|} \quad (4)$$

其中 $|X(f)|$ 为频率 f 的语音信号的幅度谱,而 $w_m(f)$ 由式(5)计算得到。

$$w_m(f) = \begin{cases} \frac{f-l_m}{o_m-l_m}, & l_m \leq f \leq o_m \\ \frac{h_m-f}{h_m-o_m}, & o_m < f \leq h_m \end{cases} \quad (5)$$

其中, o_m 为第 m 个滤波器的中心频率。通过式(4)(5),可以得到 Mel 子带频谱质心序列 $\{C_m, 1 \leq m \leq M\}$ 。新的 Mel 滤波器组输出由式(6)计算得到:

$$fbank'(m) = \frac{fbank(m)(C_m - o_m)}{h_m - l_m} \quad (6)$$

其中 $fbank(m)$ 是第 m 个滤波器的初始输出,而 $fbank'(m)$ 为新的输出。

$\{fbank'(m), 1 \leq m \leq M\}$ 经过对数非线性变换、DCT 和倒谱系数提升后,得到基于 MSSC 的倒谱系数,其提取过程如图 2 所示。

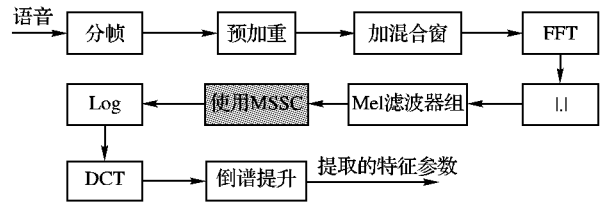


图 2 基于 MSSC 的语音特征参数提取示意图

2 实验

实验使用的语音数据包括在安静环境下录制的 4 个男性和 6 个女性对长度不大于 8 的连续汉语数字串 0~9 的发音。语音信号的采样率为 16 kHz, 16 bit 量化精度。语音样本采用 HTK 工具中的工具 HSGen 生成符合任务语法的 250 个样本^[12]。语音数据分为训练集和测试集两部分。每个说话人有 25 个语音串,其中 20 个作为训练集,5 个作为测试集。对测试集中每个语音文件加入高斯白噪声,其信噪比从 5 db 到 -10 db,间隔为 5 db。

实验中采用的语音帧长为 25 ms,帧移为 10 ms。对每 1 帧语音数据,其预加重系数为 0.97。经过混合窗后,使用 512 点的 FFT 来计算每帧语音的频谱。采用 26 维的 Mel 滤波器,计算前 13 维 ($C_0, \dots, C_{11}, C_{12}$) 的倒谱系数,以及它们的一阶二阶动态特征,其中 C_0 为能量分量,最后得到 39 维的语音特征序列。语音的模型采用从左到右无跳转的连续 HMM。HMM 有 5 个状态,每个状态下观察值的概率密度函数采用单数据流单高斯概率密度函数,转移矩阵为对角矩阵。HMM 模型的训练和识别都采用英国剑桥大学用 C 语言开发的开放源代码的语音识别工具 HTK3.3^[12]。

为评测不同的系统的鲁棒性,实验中使用相同的干净语音对模型进行训练,采用不同信噪比下的语音分别进行测试。表 1 和图 3 为不同信噪比的语音, MFCC 特征、利用混合窗替代汉明窗的特征、加入 MSSC 的特征以及混合窗 + MSSC 的特征的识别率。由实验结果可以看出,低信噪比下,使用混合窗函数的系统与传统基于 MFCC 的系统相比,平均识别率提高了 1.57%。使用 MSSC 的系统与传统基于 MFCC 的系统相

比,平均识别率提高了 19.14%。使用混合窗 + MSSC 的系统与传统基于 MFCC 的系统相比,平均识别率提高了 17.13%。

表1 不同信噪比的测试语音在不同系统中的识别率 %

系统	信噪比/dB			
	-10	-5	0	5
MFCC	9.79	9.79	9.79	23.08
混合窗函数	9.79	9.79	10.14	25.87
MSSC	27.97	31.82	31.47	37.76
混合窗函数 + MSSC	22.73	27.62	31.82	38.81

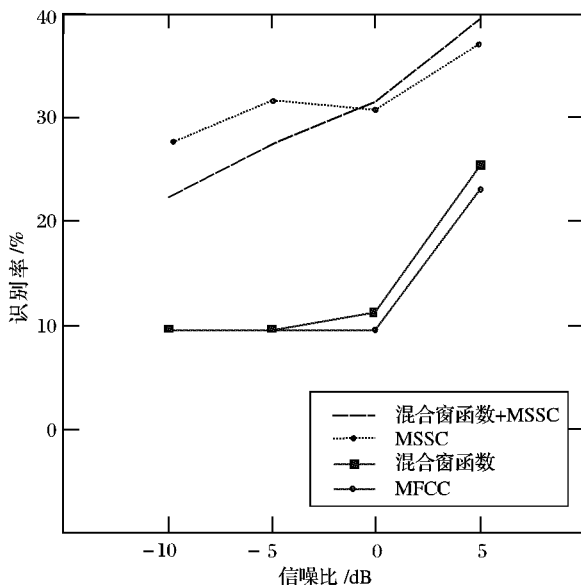


图3 不同信噪比的测试语音在不同系统中的识别率

3 结语

本文提出两种低信噪比下鲁棒语音识别方法:一种是将传统 MFCC 特征提取中的汉明窗函数用一个混合窗函数代替,另一种是将 Mel 子带频谱质心(MSSC)和基于 MFCC 的语音前端处理相结合。实验表明,混合窗函数和 MSSC 在低信噪比下都可以提高语音识别系统的鲁棒性。鲁棒特征提取的目的是为了提取不受或受噪声干扰较小的特征参数,但是也导致了无法利用特定噪声的先验知识,在以后的实验中,可以考虑在特征提取的过程中融合语音增强的某些方法。本文在实验中仅使用了高斯白噪声,在其他噪声环境下,使用这两种方法的识别系统的鲁棒性还需进一步的实验验证。

(上接第 359 页)

情况下, $|T| \ll 2^{32}$, 因而 α 值几乎为 0, 误报数也为 0。

4 结语

本文通过增加 group indication 域把路由器分成不同的组,这样在重构攻击路径时,大大减少了所需要的组合次数,改进了 FMS 算法高计算开销和高误报率的缺点,为更快、更精确地追踪到攻击者打下了基础。

参考文献:

- [1] SAVAGE S, WETHERALL D, KARLIN A, et al. Practical network support for ip traceback[EB/OL]. [2008-06-5]. <http://www.ece.cmu.edu/~adrian/630-f04/readings/SWKA.pdf>.
- [2] SONG D X, PERRIG A. Advanced and authenticated marking schemes for IP traceback[EB/OL]. [2008-06-5]. <http://www>.

参考文献:

- [1] MANSOUR D, JUANG B H. The short-term modified coherence representation and its application for noisy speech recognition[C]// IEEE Transactions on Acoust Speech Signal Process. New York: IEEE, 1989: 95-804.
- [2] HERMANSKY H, HANSON B, WAKITA H. Perceptually based linear predictive analysis of speech[C]// IEEE International Conference on ICASSP 85. New York: IEEE, 1985: 509-512.
- [3] YOU K H, WANG H C. Robust features for noisy speech recognition based on temporal trajectory filtering of short-Time autocorrelation sequences[J]. Speech Communication, 1999, 28(1): 13-24.
- [4] VIKKI O, BYE D, LAURILA K. A recursive feature vector normalization approach for robust speech recognition in noise[EB/OL]. [2008-06-5]. <http://www.karilaurila.com/Publications/13.doc>.
- [5] HIRSCH H G, EHRLICHER C. Noise estimation techniques for robust speech recognition[C]// Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1995: 153-156.
- [6] 刘波, 李锦宇, 戴礼荣, 等. 语音识别中的两级 MEL 域滤波器组维纳滤波方法[J]. 信号处理, 2004, 20(3): 133-137.
- [7] SANGITA R, SHARMA. Multi stream approach to robust speech recognition[D]. Oregon: Oregon Graduate Institute of Science and Technology, 1999.
- [8] SANGITA TIBREWALA, HVNEK HERMANSKY. Subband based recognition of noisy speech[C]// Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1997: 1255-1258.
- [9] PALIWAL K K. Spectral subband centroid as features for speech recognition[C]// Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1998: 617-288.
- [10] TSUGE S, FUKADA T, SINGER H. Speaker normalized spectral subband parameters for noise robust speech recognition[C]// Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1999: 285-288.
- [11] BOJANA G, PALIWAL K K. Robust feature extraction using subband spectral centroid histograms[C]// Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 2001: 85-88.
- [12] YOUNG S. The htk book(for htk version 3.3)[K]. Entropic Cambridge Research Laboratory, 2005.

cs.berkeley.edu/~dawnsong/papers/iptrace.pdf.

- [3] CARTER R L, CROVELLA M E. Dynamic server selection using dynamic path characterization in wide-area networks[C]// Proceedings of the 1997 IEEE INFOCOM Conference. Washington, DC: IEEE Computer Society, 1997: 1014.
- [4] WOLFGANG T, ROTHERMEL K. Dynamic distance maps of the internet[C]// IEEE Infocom. Israel: IEEE, 2000, 1: 275-284.
- [5] Cooperative association for internet data analysis[EB/OL]. [2008-06-06]. <http://www.caida.org/Tools/Skitter/Summary/>.
- [6] STOICA I, ZHANG HUI. Providing guaranteed services without per flow management[EB/OL]. [2008-06-5]. <http://www.cs.columbia.edu/~zwb/my/oral/qos/sigcomm99/sz.ps>.
- [7] 周曜, 徐长江, 徐佳, 等. 基于随机包标记方案的 IP 追踪性能分析. 计算机科学[J], 2007, 34(12): 78-81.