文章编号:1002-8684(2009)09-0061-04

基于 MFCC 参数的说话人特征提取算法的改进

·论文·

张　晶，范　明，冯文全，董金明
（北京航空航天大学 电子信息工程学院，北京 100191）

【摘　要】在说话人识别系统中，特征参数的提取对语音训练和识别有着重要的影响。对于特征参数提取模块，提出了一种新的特征参数提取算法 MFCC_E(Efficient MFCC)。相对于标准算法 MFCC_S(Standard MFCC)，MFCC_E 在特征提取模块部分减少了 53%的计算量。最终实验结果说明 MFCC_E 的识别率为 90.3%，仅比标准 MFCC 算法 92.0%的识别率降低 1.7%。因为 MFCC_E 算法的这种特点，使其能够更有效的适用于硬件实现。

【关键词】特征提取；MFCC_S；MFCC_E
【中图分类号】TP311　　【文献标识码】A

An Efficient Speaker Feature Extraction Method Based on MFCC

ZHANG Jing，FAN Ming，FENG Wen-quan，DONG Jin-ming
(School of Electronic and Information Engineering，Beihang University，Beijing 100191，China)

【Abstract】 Feature extraction is a significant module for speech training and recognition in speech recognition system. A new algorithm of feature extraction MFCC_E(Efficient MFCC) is introduced. Compared to the standard algorithm MFCC_S (Standard MFCC)， the new algorithm reduces the computation power by 53%. The simulation results indicate MFCC_E has a recognition accuracy of 90.3%， and there is only an 1.7% reduction compared to MFCC_S which has 92.0% recognition accuracy. The new algorithm is acceptable for hardware implement for its advantage.

【Key words】feature extraction；MFCC_S；MFCC_E

1　引言

真正意义上的“自动”说话人识别的研究始于 20 世纪 60 年代，此后 40 多年间人们提出了多种语音参数模型，其中 Mel 频率倒谱系数(Mel-Frequency Cepstrum Coeffients，MFCC)应用最为广泛，尤其是在如何提高其识别率方面，人们对 MFCC 参数进行了很多的研究[1-2]。然而这些算法都需要经过大量的计算，这不仅提高了成本，更为重要的是降低了其硬件实现的可行性。笔者通过对标准 MFCC 算法的研究[3-4]，提出一种新算法使其更加适用于硬件实现。

2　标准 MFCC 参数提取

2.1　Mel 频率倒谱系数(MFCC)[5]

人耳对不同频率的语音具有不同的感知能力，实验发现，在 1 000 Hz 以下，感知能力与频率成线性关系，而在 1 000 Hz 以上，感知能力则与频率成对数关系。因此人们提出了 Mel 频率的概念，其意义为：1 *Mel* 为 1 000 Hz 的音调感知程度的 1/1 000。频率 f 与 *Mel* 频率之间的转换公式为

$$Mel(f)=2\,595\lg(1+f/700) \tag{1}$$

式中，f 为频率，单位：Hz。

2.2　标准 MFCC_S 的提取过程

MFCC 即为基于上述 Mel 频率的概念而提出的，其提取及计算过程如图 1 所示。

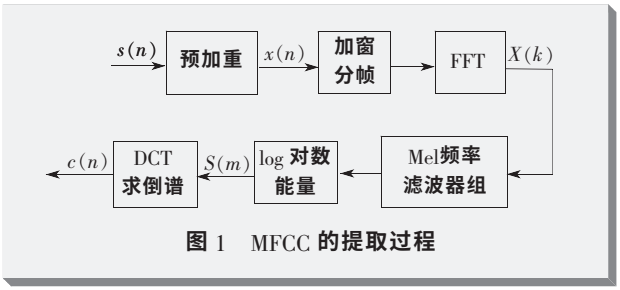

图 1　MFCC 的提取过程

提取及计算过程为：

(1) 原始语音信号 $f(n)$ 经过预加重、分帧、加窗等模块的处理，得到每个语音帧的时域信号 $x(n)$。在预加重模块中，笔者采用数字滤波器 $H(z)=1-\mu z^{-1}$，μ 值取 0.97，即 $f'_n=f_n-\mu f'_{n-1}$。加窗及分帧模块如图 2 所示，其中窗长 20 ms(320 点)，窗移 10 ms。

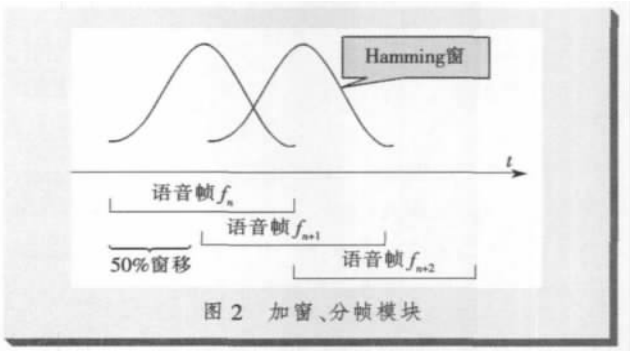

图 2 加窗、分帧模块

(2) 将时域信号 $x(n)$ 后补若干个 0 以形成长为 N(本文 N 取 512)的序列，然后经过 FFT 模块后得到线性频谱 $X(k)$，

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi nk/N},\quad 0\leqslant n,k\leqslant N-1 \tag{2}$$

(3) 将上述线性频谱 $X(k)$ 通过 Mel 频率滤波器组模块得到 Mel 频谱，并通过对数能量的处理得到对数频谱 $S(m)$[5]。

$$S(m)=\ln\left[\sum_{k=0}^{N-1}|X(k)|^2H_m(k)\right],\quad 0\leqslant m<M \tag{3}$$

其中 Mel 频率滤波器组为在语音的频率谱范围内设置若干个带通滤波器 $H_m(k)$，每个滤波器具有三角形滤波特性。图 3 为计算 Mel 倒谱的三角滤波器组。

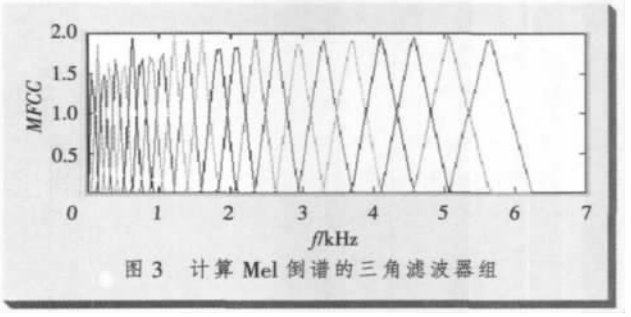

图 3 计算 Mel 倒谱的三角滤波器组

(4) 将上述对数频谱 $S(m)$ 经过离散余弦变换(DCT)变换到倒频谱域，可得到 Mel 频率倒谱系数(MFCC_S 参数)

$$c(n)=\sum_{m=0}^{M-1}S(m)\cos\left[\frac{\pi n(m+1/2)}{M}\right],\ 0\leqslant m<M \tag{4}$$

2.3 差分 MFCC_C 参数的提取

在谱失真测度定义中通常不用 0 阶倒谱系数，MFCC_S 分析的滤波器组数取 33，系数选取了前面 16 个(C_0~C_{15})。从图 3 中可看出第一维 MFCC_S 系数的能量很大且含有直流信息，而低阶分量较高阶分量更容易受加性噪声的干扰。故在系统中将 C_0 称为能量系数，不作为倒谱系数的一员。MFCC 参数如图 4 所示。

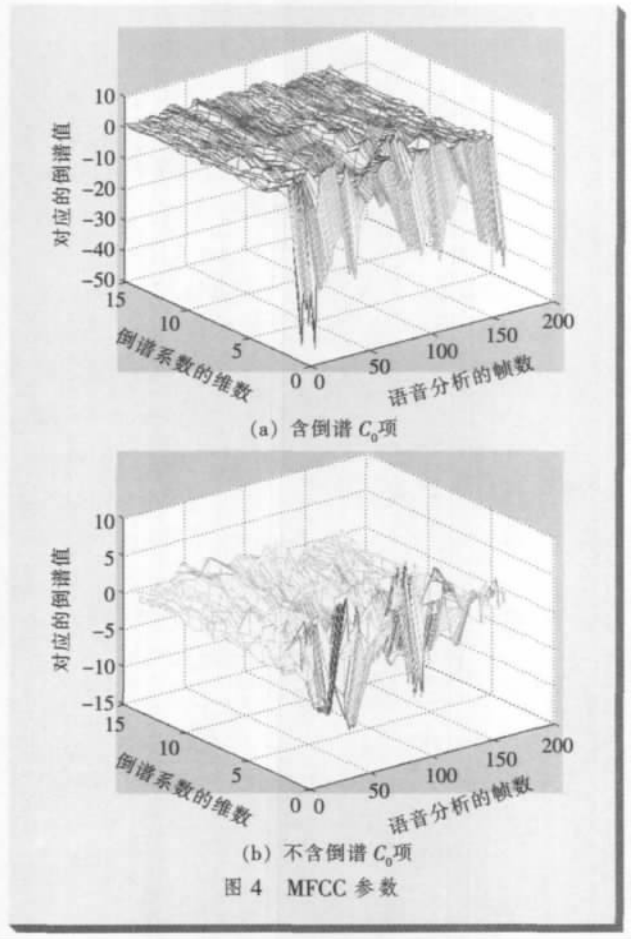

图 4 MFCC 参数

二次特征提取是对原始特征向量序列进行再分析。通过对特征向量运用加权、差分、筛选等方法，进一步剥离出隐藏在语音背后的说话人特征。特征差分用于获取语音特征向量的连续动态变化轨迹，其研究对象是一段语音的特征向量序列。差分公式为

$$\Delta C_l(m)=\sum_{k=-2}^{2}kC_{l-k}(m),\quad 1\leqslant m\leqslant p \tag{5}$$

式中，l 与 $l-k$ 表示第 l 与 $l-k$ 帧，m 表示第 m 维。

由于不同人之间说话的差别，利用单一参数很难达到可靠的性能要求，为了更有效地表征说话人特征，采用几个特征参数的组合式用来提高实际系统的性能。当各组合参数间相关性不大时，会有很好的效果。笔者使用参数 MFCC_S 和 △MFCC 相结合的方法，形成新的特征参数 MFCC_C。这种混合特征识别方法能

使辨认系统的误识率有明显的降低。其提取过程如图5所示。

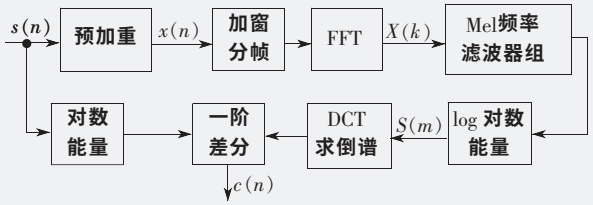

图5 差分参数MFCC_C提取过程

在MFCC_C的提取过程中，预加重，分帧加窗，FFT，Mel滤波器组，DCT等模块与MFCC_S相同。将语音信号的对数能量 $E=\lg\sum_{n=1}^{320}f_n^2$ 代替 C_0，与 ΔC_l 构成12维差分参数。这样MFCC_C就由24维特征参数构成。

3 MFCC_E参数提取

如表1所示，在MFCC_C的提取过程中，每帧语音帧在加窗模块中消耗20×16=320的乘法运算，FFT模块需要256×lb512=2 304的乘法运算，在Mel滤波器组模块中需要消耗256的乘法运算，在DCT模块中消耗24×12=288的乘法运算，所以在MFCC_C的提取过程中总共需要3 168的乘法运算，可看出对于一帧语音帧来说需要消耗大量的乘法计算。

表1 不同特征参数计算量对比

特征参数	计算量				
	加窗	FFT	Mel	DCT	总和
MFCC_C	320	2 304	256	288	3 168
MFCC_E	160	1 024	0	288	1 472

笔者提出一种新的特征参数MFCC_E的提取方法。相对于MFCC_C，新算法降低了53%的乘法计算量。MFCC_E的提取过程如图6所示，虚线部分是与MFCC_C参数提取过程不同的模块。

(1) 在预加重模块中对 μ 值进行了修正，用31/32取代0.97，即

$$f'_n=f_n-\mu f_{n-1}=f_n-\frac{31}{32}f_{n-1} \tag{6}$$

将式(6)改写为

$$f'_n=f_n-(f_{n-1}-\frac{1}{32}f_{n-1}) \tag{7}$$

由式(7)只需将 f_{n-1} 向右移5 bit就可实现预加重模块。这样就可用简单的加法和移位运算来代替复杂乘法运算节省开销，便于硬件实现。而这一修正不会对识别率造成严重的影响，如表2所示。

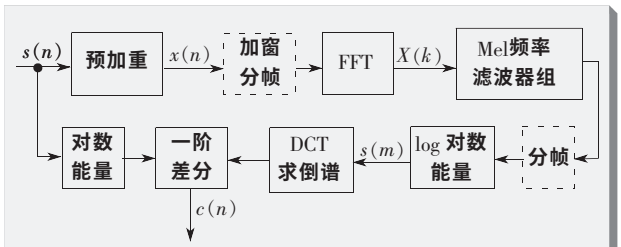

图6 MFCC_E提取过程

表2 实验采用的主要参数及识别率对比

特征参数	μ	窗长	FFT点数	Mel滤波器	识别率/%		
					3 s	6 s	9 s
MFCC_S	0.97	320	512	三角滤波器	84.2	86.5	87.6
MFCC_C	0.97	320	512	三角滤波器	86.4	89.8	91.3
MFCC_C	31/32	320	512	三角滤波器	86.8	90.3	92.0
MFCC_E	31/32	160	256	三角滤波器	85.7	88.6	90.3
MFCC_E	31/32	160	256	矩形滤波器	85.6	88.2	89.8

(2) 如图2所示，在MFCC_C参数的提取过程中，加窗的过程也完成了分帧模块。在MFCC_E参数提取过程中，笔者将分帧模块放到Mel滤波器组模块之后。如图7所示，将加窗后的语音帧 s_n 称之为“子帧”，一个子帧包含10 ms语音信号(160点)，而且相邻2个子帧帧移为0。在语音帧长度上，MFCC_C一帧的长度 f_n 等于MFCC_E中相邻2个子帧 s_n，s_{n+1} 之和。

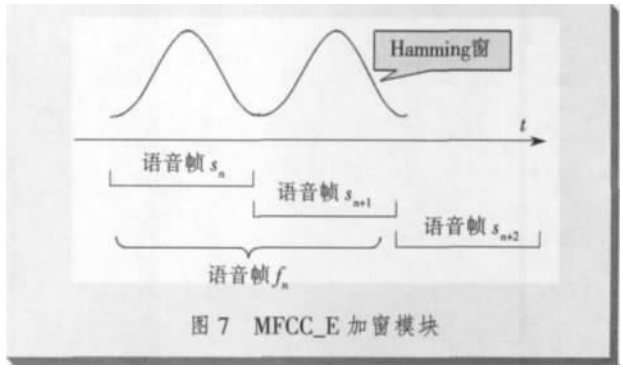

图7 MFCC_E加窗模块

(3) 加窗模块后，将对语音帧进行FFT运算，这与MFCC_C相同，然而因为一个子帧只包含160点语音信号，所以在FFT模块中只需进行256点的FFT运算，其消耗的计算量为128×lb 256=1 024。

(4) Mel滤波器应尽可能地压缩频谱的动态范围，而且要在整个频谱范围内平滑频谱的频率响应。矩形滤波器满足上述要求，如图8所示，在Mel滤波器组模块中，用矩形滤波器代替三角滤波器，其频率范围与三角滤波器相同。因为矩形滤波器的输出只有“0”，“1”这2个值，所以只需对经FFT模块后的频谱值 $X(k)$ 进行“加”或“不加”的操作。这样就把式(3)中复杂的乘法运

算变成了简单的加法运算，便可得到对数频谱 $S(m)$。对于一帧有 256 点的子帧来说，采用三角滤波器需要 256 次的乘法运算，现在只需 256 次加法运算。

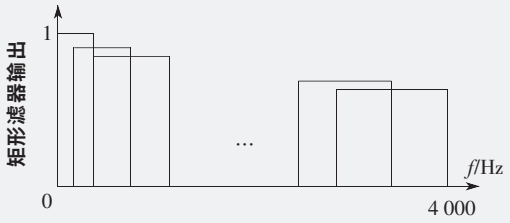

图 8　MFCC_E Mel 滤波器组模块

（5）　分帧模块如图 9 所示，f_n，f_{n+1} 为 MFCC_S，MFCC_C 中的一帧语音帧，其长度为 20 ms（320 点），帧移为 10 ms（160 点）。s_n，s_{n+1}，s_{n+2} 为 MFCC_E 中的子帧，每帧长度为 10 ms（160 点）。s_n，s_{n+1} 经由 FFT，Mel 滤波器组模块后的输出结果为 $S_{n,k}$，$S_{n+1,k}$，相加得到对数能量谱 $F_{n,k}$，同理得到 $F_{n+1,k}$。由图 9 可知，$F_{n,k}$，$F_{n+1,k}$ 为 f_n 和 f_{n+1} 的对数能量谱。由于将分帧模块放到 Mel 滤波器组模块之后，如表 1 所示，这可减少近 50%的计算量。

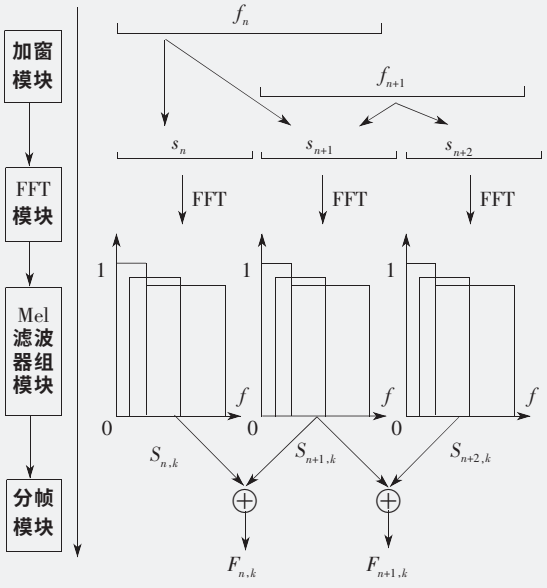

图 9　MFCC_E 分帧模块

（6）　DCT 模块及差分模块与 MFCC_C 的提取过程相同，将得到 24 个特征参数。新的特征参数提取算法将乘法计算量由 3 168 减小到 1 472。如表 2 所示，在 MFCC_E 的提取过程中每帧语音帧在加窗模块中消耗10×16=160 的乘法运算，FFT 模块需要 128×lb 256=1 024 的乘法运算，在 DCT 模块中消耗 24×12=288 的乘法运算。

4　实验结果分析

实验采用自己录制的语音，录音人数为 20 人，每人录制 6 次。前 5 次作为模型训练，最后一次作为测试语音。笔者采用与文本无关的高斯混合模型（GMM）为语音的声学模型，并以最大后验概率（MAP）算法进行模型的训练并实现最终的模型匹配。实验中采样率为 16 kHz，采用 Hamming 窗对语音信号进行加窗、分帧，帧长为 20 ms，帧移为 10 ms。表 2 为实验中所用到的主要参数及实验结果。

如图 10 所示，3 种参数中 MFCC_C 的识别率最高，而 MFCC_E 在降低了 53%计算量的情况下其识别率仅比 MFCC_C 低 1.7%左右。由结果可看出含有 项的特征参数识别系统性能下降，这主要是因为含有直流信息而低阶 MFCC 分量较高阶分量更容易受加性噪声的干扰，所以 MFCC_S 的识别率最低。如图 11~12 所示，正如在第 3 节中介绍的一样，在 MFCC_E 中对 μ 值和 Mel 滤波组的改进不会对识别率造成严重影响。

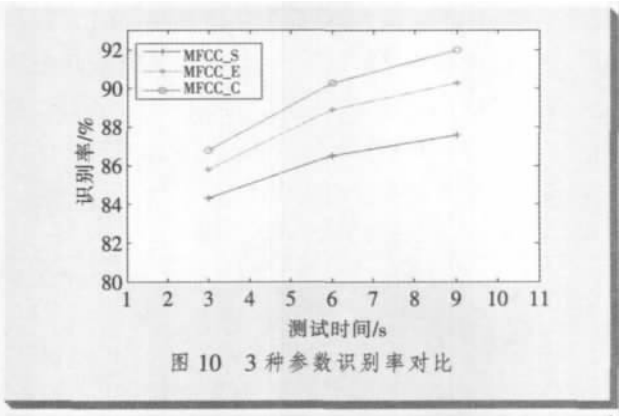

图 10　3 种参数识别率对比

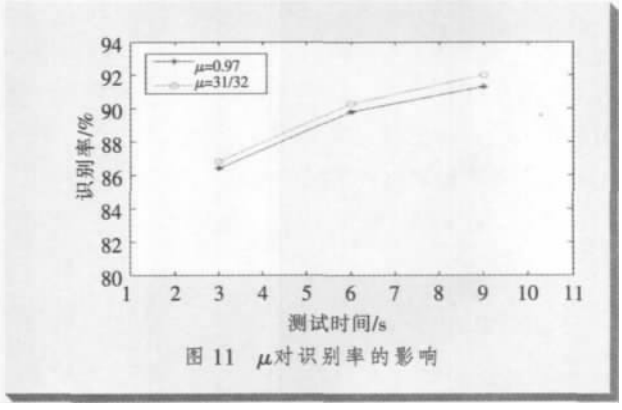

图 11　μ 对识别率的影响

5　结语

笔者介绍了一种新的特征参数 MFCC_E 的提取方法，新算法将乘法计算量由 3 168 次减小到 1 472 次，而识别率仅降低 1.7%。对标准算法进行多次改进，

（下转第 69 页）

白噪声，受测试者为听力正常的4男4女。在SNR_{in}=-5，0，5，10 dB这4种情况下，分别对硬、软阈值算法和笔者算法增强后的语音进行试听。试听结果为：受测试者均认为笔者算法增强后的语音可懂度和清晰度明显高于硬、软阈值算法，去噪后语音清晰，接近原始语音，听觉舒适且没有疲劳感。由此说明笔者方法是有效的。

6 结语

采用小波包变换可细致地将语音的高频与低频部分进行分解，从而避免了采用小波变换不能对语音高频部分细分的不足[7]；更进一步，笔者没有采用传统的阈值函数处理小波包系数，而是构造了一种新的阈值函数，这个函数既兼顾了硬、软阈值函数的优点，同时又在一定程度上弥补了这2种方法的缺陷。通过仿真实验明显可以看出，笔者方法的去噪效果优于传统的硬、软阈值方法，说明改进阈值函数可行且非常有效。

参考文献

[1] DONOHO D L，JOHNSTONE J M. Ideal spatial adaptation by wavelet shrinkage [J]. Biometrika，1994，81(3)：425-455.

[2] DONOHO D L. De-noising by soft-thresholding[J]. IEEE Trans. on Information Theroy，1995，41(3)：613-627.

[3] 杨永明，路陈红.小波包分析在一维及二维信号去噪中的应用[J].西安建筑科技大学学报：自然科学版，2004，36(3)：364-367.

[4] WALDEN A T，MCCOY E，PERCIVAL D B. The variance of multi-taper spectrum estination for real Gaussian processes [J]. IEEE Trans. on Signal Processing，1994，42(2)：479-482.

[5] 崔锦泰.小波分析导论[M].西安：西安交通大学出版社，1995.

[6] CHANG S，KWON Y，YANG S，et al. Speech enhancement for non-stationary noise environment by adaptive wavelet packet[C]//Proceedings of IEEE International Conference on Acoustics，Speech，and Signal Processing.[S.l.]：IEEE Press，2002：561-564.

[7] 周静，陈允平，周策，等.小波系数软硬阈值折中方法在故障定位消噪中的应用[J].电力系统自动化，2005，29(1)：65-68.

作者简介

邓玉娟，助教，主要研究方向为智能控制与信号处理、小波分析理论及其应用等。

[责任编辑] 侯莉　　[收稿日期] 2009-06-21

（上接第64页）

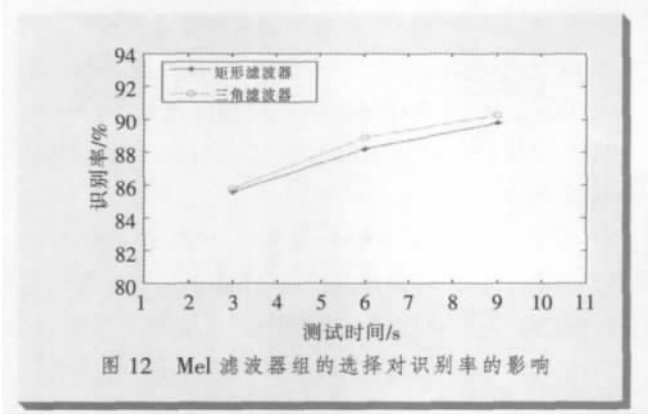

图12 Mel滤波器组的选择对识别率的影响

旨在于使其更加适于硬件，并且已经在FPGA上实现了新算法的说话人识别系统。笔者希望新算法能够给说话人识别系统的硬件实现带来高效、高速、廉价的应用前景。

参考文献

[1] 杨行峻，迟惠生.语音信号数字处理[M].北京：电子工业出版社，1995.

[2] PHADKE S，LIMAYE R，VERMA S，et al. On design and implementation of an embedded automatic speech recognition system [C]// Proceedings of 17th International Conference on VLSI design. Mumbai：[s.n.]，2004：127-132.

[3] HATCH A，PESKIN B，STOLCKE A. Improved phonetic speaker recognition using lattice decoding[C]//Proceedings of International Conference on Acoustics，Speech and Signal Processing. Philadelphia：[s.n.]，2005，1：169-172.

[4] HIRSCH Hans-gunter，PEARCE D. The AURORA experimental framework for performance evaluation of speech recognition systems under noisy conditions [C]// Proceedings of ISCA ITRW ASR2000. Paris：[s.n.]，2000，9：18-20.

[5] 蔡莲红，黄德智，蔡锐.现代语音技术基础与应用[M].北京：清华大学出版社，2003.

作者简介

冯文全，副院长，副教授，主要研究方向为微波技术和集成电路设计；

董金明，教授，主要研究方向为微波技术；

张晶，硕士研究生，主要研究方向为信号处理与语音识别。

[责任编辑] 史丽丽　　[收稿日期] 2009-05-24