中山大学学报（自然科学版）
第30卷 第1期 ACTA SCIENTIARUM NATURALIUM Vol.30 No.1
1991年 UNIVERSITATIS SUNYATSENI 1991

基于矢量量化的实时发音人确认系统研究

陈志成 陈云凤
（无线电电子学系）

摘 要

采用矢量量化的方法，在IBM-PC/XT机上应用高速DSP系统ATD320构成一个与文本有关的实时发音人确认系统。系统可达到96%的确认率而仅有6.8%的错误接受率，系统的确认时间约为0.5s。介绍了一种确认门限预置的方法和在DSP器件TMS32020上加速码书训练的方法，并提出改善确认系统性能的几点设想。

关键词 发音人识别，矢量量化

1 前 言

矢量量化(Vector Quantization)是七十年代末和八十年代初发展起来的一种高效的数据压缩技术[1～3]，其原理与香农(Shannon)在率失真理论中提出的信源块编码（Source Block Coding）相同：将信源发出的信号抽样后，不是逐个进行量化（标量量化），而是将k个($k>2$)抽样值形成k维空间的一个矢量，以后将此矢量一次量化。因为矢量量化有效地应用了矢量中各分量的四种相互关联的性质（线性依赖性、非线性依赖性、概率函数的形状及矢量维数）来消除信源的冗余度，所以它可以进行高效的数据压缩。矢量量化是率——失真理论的一种具体的实现方法，性能大大优于标量量化。

与传统的发音人识别方法不同，本文采用矢量量化的方法来实现发音人确认[4,5]。为了把矢量量化技术应用到发音人识别系统中，将每个发音人作为一个信息源，用矢量量化的方法模拟信息源的特征，因而每个特征码书所用的训练序列必须是基于发音人而不是基于发音内容的。因此每个基于发音人的码书可反映一个人的发音特征。建立好特征码书之后，就可以用该码书对测试者的输入矢量进行编码量化，最后得到一个平均量化误差（平均失真）。将此失真与预先设定的判别门限相比较，如果平均失真小于门限值，则测试者被接受；如果大于门限值，则测试者被拒绝。

2 发音人确认系统构成

本文研究的发音人确认系统是在IBM-PC/XT计算机上应用高速数字处理系统ATD-320构成。ATD320系统的核心是新一代数字信号处理器TMS32020，其运算速度可达500万次/s。ATD-320系统作为IBM-PC/XT的子系统，不但有自己的数据和程序存储器，而且有一个全局存储器可与IBM-PC/XT共享，它既能由主机管理控制，又可以

本文1989年11月14日收到

脱离主机控制独立进行数据处理，与主机构成一个主从机高速数字处理系统。ATD-320系统的详细介绍可参阅文献[6]。

系统采用自制的语音信号采样板，语音信号经过放大之后，进行0～4000Hz低通滤波，再用12位A/D转换器得到离散的语音信号。语音采样数据存贮于ATD-320系统的全局存贮器里，再由IBM-PC/XT机控制ATD-320系统对数据进行处理。发音人确认系统的数据处理框图如图1所示。

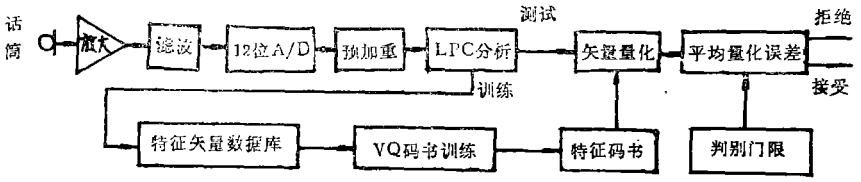

图1 发音人确认系统的数据处理过程

Fig. 1 Data processing of the speaker verification system

由于IBM-PC/XT与ATD320构成的主从系统两者可以同时工作，因此对语音数据的处理采用类似SYSTOLIC ARRAY的处理形式：由IBM-PC/XT完成模数转换以及帧能量计算及预加重处理，每当主机完成一帧的数据采样，即通知ATD-320系统对该帧语音数据进行线性预测分析。由于ATD-320系统能在16ms以内完成一帧（256点）的LPC分析，因此语音采样与处理可以同时进行。采样工作完成之后，系统的工作仅剩下端点确定和量化编码，可以达到实时性的要求。

3 语音信号采样及线性预测分析

采用12位A/D进行语音数据采样，采样频率为10kHz。采样前先根据背景噪声确定参考能量门限，在采样结束后用文献[7]的方法通过比较信号的能量进行端点确定。设背景能量为E_0采样时当帧能量$E>3E_0$时作为信号的起点，当$E<1.5E_0$后几帧作为信号的结束。采样结束后再根据帧能量向前确定准确的端点：$E>1.5E_0$时作为起点；$E<2E_0$时作为信号的终点。为了加快处理速度，采样程序在读入采样数据时马上进行能量计算及预加重处理，预加重系数为15/16，预加重后的语音数据存在ATD-320系统的全局存贮器里。

用格型法直接求部分相关系数(PARCOR系数)[8]，格型公式为

$$K_i=2\sum_{m=0}^{N-1}\left[e^{(i-1)}(m)b^{(i-1)}(m-1)\right]\Big/\left\{\sum_{m=1}^{N-1}\left[e^{(i-1)}(m)\right]^2+\sum_{m=0}^{N-1}\left[b^{(i-1)}(m-1)\right]^2\right\} \tag{1}$$

其中$e^{(i)}(m)$为前向预测误差序列

$$e^{(i)}(m)=e^{(i-1)}(m)-K_i^{(i-1)}b(m-1) \tag{2}$$

$b^{(i)}(m)$为后向预测误差序列

$$b^{(i)}(m)=b^{(i-1)}(m-1)-K_i^{(i-1)}e(m-1) \tag{3}$$

零阶预测器等效于完全不用预测器，因此有

$$e^{(0)}(m)=b^{(0)}(m)=s(m) \tag{4}$$

如果用(1)式求K_i，则有$-1\leqslant K_i\leqslant+1$，计算结果归一化，便于DSP系统进行快速处理。LPC分析由ATD-320数字处理系统完成。LPC分析帧长$N=256$，阶数为12，步长为128。TMS32020的片内B_0，B_1块作误差$e(m)$和$b(m)$寄存器，数据处理采用16位定点运算，每帧LPC分析时间约为16ms。每帧的$K_i(i=1,2,\cdots12)$参数作为一个矢量。

4　特征码书的建立

4.1　训练序列的建立

用格型法对采样进行12阶的LPC分析，每帧LPC分析得到一个12维的矢量。发音人重复某个固定发音直到数据库矢量数超过2500个为止。数据库大小约64KB。系统采用汉语“零”的发音作为识别发音，数据库的建立约需重复发音70次。

4.2　用TMS32020实现的矢量量化算法

本文采用全搜索分裂式LBG算法[6]进行码书训练，训练程序的步骤如下：

①预置1位码书的两个初始码字；

②对每个训练矢量用全搜索方法找出与之最相近的码字，并累加于相应单元；

③求出新的聚类中心并以之作为新的码字；

④重复步骤(2)(3)直到平均失真降至某个预定值为止；

⑤如果码书位数达到要求，则停止训练，输出最后码书。

⑥将码书中每个码字乘以分裂因子得到两个新的码字，从而得到下一位初始码书。

⑦转到步骤(2)。

整个处理过程在数字处理器件TMS32020上完成，其中全局存贮器存放训练矢量数据库，ATD-320系统的DRAM存放训练码书和矢量和的值。TMS32020的片内RAM中B_2块作中间变量存贮，B_0，B_1块作失真计算用。平均失真值计算如下

$$D=(1/P)\sum_{i=1}^{p}(K_{mi}-K_i)^2 \tag{5}$$

展开(5)式，可得

$$D=(1/P)\left[\sum_{i=1}^{p}K_{mi}^2+\sum_{i=1}^{p}K_i^2-2\sum_{i=1}^{p}K_{mi}\cdot K_i\right] \tag{6}$$

对于全搜索算法，上式仅需计算$\sum_{i=1}^{p}K_{mi}\cdot K_i$，前两项均可查表得到。对于12维LPC矢量，仅增加15%的存贮量而训练速度可以提高50%。训练码书的结构除矢量数据外，还包含每个矢量的平方和$\sum_{i=1}^{p}K_{mi}^2$。因此对于一个P维码字，需要$P+2$个单元来存贮，这样也使编码时的速度加快。经过上述处理，码书的训练时间大大缩短了。一般码书训练在普通IBM-PC/XT机上需要数小时才能完成，本系统在TMS32020上采用上述方法仅用数分钟就可以完成码书的训练，表1给出了系统训练码书的迭代次数和训练时间，

其中收敛门限为0.0001，训练矢量数$N=2500$。

表1 TMS32020码书训练时间比较

Tab. 1 Comparison of codebook training time on TMS32020

码书位数	1	2	3	4	5	6	7	8
迭代次数(次)	9	9	13	19	18	21	15	14
时间(前)(s)	2	5	12	30	60	144	225	378
时间(后)(s)	2	4	8	20	38	84	142	242

5 发音人确认系统及其性能分析

5.1 发音人确认系统

经过数据库采集和特征码书训练，得到能反映发音人固有特征的码书，该码书是构成发音人确认系统的基础。本文采用的码书大小为$N=64$和$N=256$码字的码书作试验。系统包括背景噪声采样、连续语音数据采样、端点确定、发音中间停顿处理、LPC分析、矢量编码失真计算及最后判决等处理过程。

确认判决门限的设定是系统设计的重要步骤，门限大小的设定直接影响确认系统的性能和错误接受率(FA)和错误拒绝率(FR)的分布。为了确定每个特征码书相应的确认判决门限，采用统计确认错误率曲线的方法来确定系统的判决门限。用汉语"零"的发音作确认发音，注册者的发音经训练后得到两个特征码书BOOKO6和BOOKO8，其位数分别为6位和8位。本文对以这两个码书为基础的确认系统进行研究。先对注册者和冒充者各进行反复多次测试，统计两者的量化失真分布，从而得到各自的确认失真值统计曲线如图2和图3所示。由图可见，注册者的失真值分布趋于集中在某一个失真值两边，而冒认者的失真值分布较为分散，要从统计分布曲线直接得到门限值是不容易的。为此，本文根据得到的失真分布曲线，通过统计不同确认门限值所对应的确认错误率，分别得到特征码书BOOKO6和BOOKO8的确认错误率曲线如图4和5所示。由图可见，对于注册者，FR曲线单调下降；对于冒认者，FA曲线单调上升。因此，如果$\sqrt{FA*FR}>0$，则两曲线必有一交点，在交点处有$FA=FR$。取交点处对应的门限值作为确认门限，则这时系统具有最优的识别率。对于每个特征码书，都要设定其相应的失真门限并作为码书内容的一部分，因此对于多个注册人的发音人确认系统，要进行多次的确认门限确定。显然，这种门限确认方法具有直观、准确等优点。对每个码书用注册者的发音"零"进行150次测试，再用冒认者的发音"零"进行150次测试，从而得到各

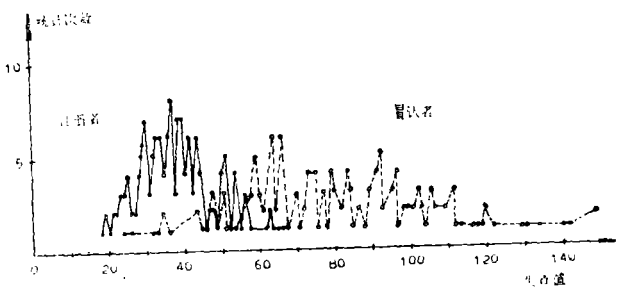

图2 确认失真值统计曲线（码书位数R=6，测试次数N=150）

Fig. 2 Statistical curve of verification distortion

自的确认量化失真值分布。根据统计分布得出系统的确认率曲线，从而得出相应特征码书BOOKO6的确认门限为51而BOOKO8的确认门限为14。

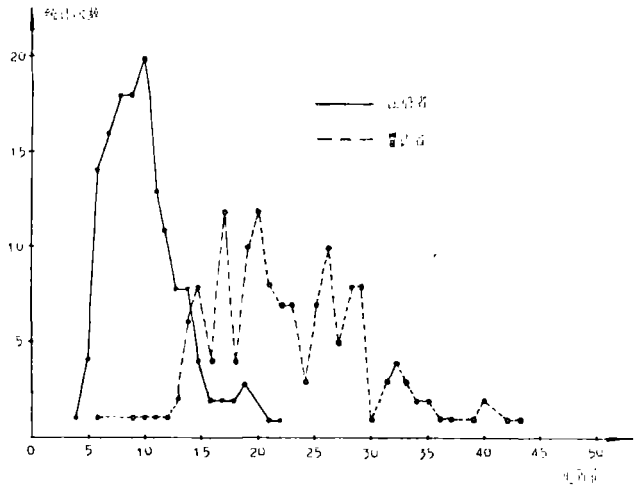

图3　确认失真值统计曲线（码书位数R = 8，测试次数N = 150）
Fig. 3　Statistical curve of verification distortion

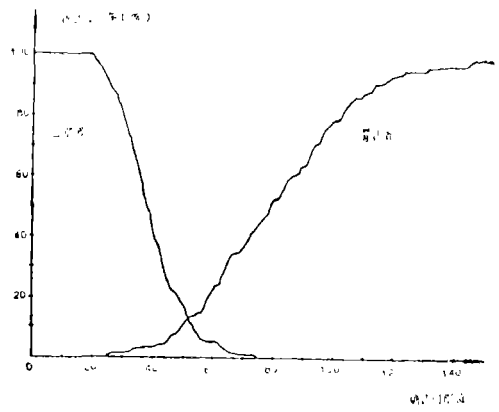

图4　确认错误率曲线（码书位数R = 6，测试次数N = 150）
Fig. 4　Verification error rate curve

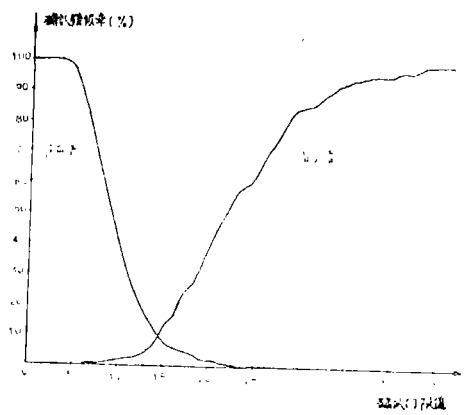

图5　确认错误率曲线（码书位数R = 8，测试次数N = 150）
Fig. 5　Verification error rate curve

5.2　系统性能分析

为了简便起见，将用特征码书BOOKO6和BOOKO8构成的发音人确认系统分别称为SV6系统和SV8系统。首先是注册人（作者）本身对SV6和SV8系统发音确认测试，总共发音测试各50次，其结果如表2所示。

为了测定系统的错误接受率（FA），作者选取了两组冒认者对系统进行测试，每组5个人，其中一组与注册者同性，另一组与注册者异性。每人测试50次来确认系统的错误接受率。测试结果如表3所示。对于SV6系统，总的平均错误接受率为8.2%；而SV8系统总的平均错误接受率为5.6%。对于与注册者同性的冒认者，其FA比与注册

人异性的FA高出约2.5%，而SV8系统的FA比SV6系统的FA低2.6%。对于SV6系统，其错误确认率为$\sqrt{FA*FR}=7\%$，对于SV8系统，其错误确认率为$\sqrt{FA*FR}=4.7\%$。

表2 发音人确认性能测试

Tab. 2 Performance test of the speaker verification system

性能指标	门限值	测式次数	拒绝次数	接纳次数	FR	确认率
SV6系统	51	50	3	47	6%	94%
SV8系统	14	50	2	48	4%	96%

表3 确认系统错误接受率测试（$N=50$）

Tab. 3 False acceptance rate (FA) test of the verification system

冒认者		男性						女性					
		YMS	LZL	MIS	LWX	CYW	平均值	SYC	XH	XT	LC	WG	平均值
SV6	FA(次)	4	6	5	6	3	4.8	4	3	3	4	3	3.4
系统	FA(%)	8%	12%	10%	12%	6%	9.6%	8%	6%	6%	8%	6%	6.8%
SV8	FA(次)	2	4	3	5	3	3.4	3	3	1	2	2	2.2
系统	FA(%)	4%	8%	6%	10%	6%	6.8%	6%	6%	2%	4%	4%	4.4%

6 小结与讨论

本文采用以矢量量化为基础的发音人识别方法，在IBM-PC/XT机上构成一个发音人确认系统，并且利用新一代的DSP器件TMS32020达到实时处理的要求。根据初步的测试结果，系统仅以一个汉语单字“零”的发音就可以达到超过90%的确认率，表明这种确认方法是可行的。它可以不作任何时间弯曲处理，节省处理时间，其次，它也不需要象统计技术那样需要估算所选参数的基本概率密度，第三，由于矢量量化能够高效地进行数据压缩并具有良好性能，系统可以达到较高的识别精度和较好的适应性。另外，这种方法在确认精度和速度上还有很大的潜力。系统性能改善可从以下两方面考虑：①进一步提高确认率，降低确认错误率。②改善FA与FR的分布，使系统在较小的FA时具有较高的确认率。现分别讨论如下：

（1）本文仅以单字“零”作为确认发音，就取得了较高的确认率。实验证明选择不同的发音对系统的确认率仅有轻微的影响，所以单字“零”发音的测试结果对本文的确认系统是有普遍意义的。就本文的确认系统来说，影响系统确认率的主要因素是发音的长度和码书训练序列的长度。因此，选用词组或句子作为确认发音（如用一数字串作为确认发音）将会提高系统的确认率。

（2）由实验结果可见，增加码书位数可提高系统的确认率。但实验结果显示码书大小增加4倍而确认率仅增加2%。显然，增加码书位数来提高确认率并非有效的途径。

（3）提高确认率的另一个办法是利用多个不同的发音进行发音人确认。每个发音对应一个特征码书，系统利用这个特征码书组进行确认。作者相信采用较小位数的多个特征码书的确认性能将比采用单码书高位数的确认性能有较大的改善，其代价是使特征码

书训练和确认程序复杂程度提高。

（4）由于本文仅采用LPC预测系数作特征参数，单一的声道参数往往不能达到很好的拒绝模仿性能，而在实用系统中这又是至关重要的。实际上，如何在使$FA=0$的情况下尽量提高确认率也是实际应用上急需解决的问题之一。使用混合特征参数是解决问题的较好方法。如果本文的确认系统在利用LPC参数作特征参量的基础上再加入波形特征参数（如基音、短时谱等）进行混合参数确认，将会改善系统的确认性能。

参考文献

〔1〕 Buzo A et al., *IEEE Trans.*, **ASSP-28** (1980), 5, 562

〔2〕 Linde Y et al., *IEEE Trans.*, **ASSP-28** (1980), 1

〔3〕 胡征等编著，矢量量化原理与应用，西安电子科技大学出版社，1988

〔4〕 Burton D K, *IEEE Trans.*, **ASSP-35** (1985), 2, 133

〔5〕 Soong F K et al., *Proc. of ICASSP*, 1985, 387

〔6〕 数字信号处理器TMS32020用户手册，1986

〔7〕 Rabiner L R et al., *Bell Syst. Tech. J.*, 54 (1975), 297

〔8〕 Rabiner L R et al., *Digital processing of speech signals*. 1978

〔9〕 Juang B H et al., *IEEE Trans.*, **ASSP-30** (1982), 2

Real-Time Text-Dependent Speaker Verification Based upon Vector Quantization

Chen Zhicheng•　　*Chen Yunfeng*

Abstract

Based on high speed DSP TMS32020, this paper describes a real-time text-dependent speaker verification system on a IBM personal computer. A new method for speaker verification based upon vector quantization (VQ) is used. In this approach, a speaker-based VQ codebook is designed to characterize a particular speaker saying a particular uttrance. Later, the same uttrance is spoken by a speaker to be verified and linear prediction analysis is performed to obtain PARCOR coefficient vectors. The verification decision is made by comparing these vectors' average quantization distortion with a prespecified codebook-specific threshold. In the preliminary tests, 96 percent verification accuracy with only 6.8 percent false acceptance rate is achieved. With the application of DSP TMS32020, only a half second verification time is needed. In this paper, an approach to specify the verification threshold is described and a method to speed up the VQ codebook training on TMS32020 is proposed. Finally, some comments on this verification system are made and several ways to improved the performance of this verification system are also discussed.

Keywords speaker recognition, vector quantization

• Department of Radio and Electronics