西安工程科技学院学报
Journal of Xi'an University of Engineering Science and Technology
第21卷第5期(总87期)　　2007年10月　　Vol. 21, No. 5(Sum No. 87)

文章编号:1671-850X(2007)05-0848-04

VQ声纹识别算法和实验

李爱平,党幼云

(西安工程大学 电子信息学院,陕西 西安 710048)

摘要:采用线性预测倒谱系数(linear prediction cepstrum coefficient,LPCC)作为语音的特征参数,矢量量化(vector quantity, VQ)方法进行模式匹配,探讨声纹识别以实现身份认证,并对此识别方法进行了相关的实验.通过验证,这种方法可以区分不同的说话人,并且在做说话人辨认实验时可达到较高的识别率.

关键词:矢量量化;线性预测倒谱系数;说话人确认;说话人辨认

中图分类号:TN 912.34　　**文献标识码**:A

0 引　言

身份认证在人类的社会生活中自古有之.传统的身份认证方法(如使用身份证、护照、钥匙、智能卡、密码、口令等)存在携带不便,易伪造、遗失,因使用过多、使用不当而损坏或不可读、密码易被破解等诸多问题,安全性、可靠性差.每个人所固有的生物特征,具有与其他人不同的惟一性和在一定时期内不变的稳定性,而且不会丢失,不易伪造和假冒,所以被认为是终极的身份认证媒介[1].

在生物识别领域中,声纹识别也称为说话人识别,以其独特的方便性、经济性和准确性等优势受到世人瞩目,并日益成为人们日常生活和工作中重要且普遍的安全认证方式.说话人识别是一种根据说话人语音波形中反映说话人生理和行为特征的语音参数自动识别说话人身份的技术[2].说话人识别可以看作是语音识别的一种,是指通过对说话人语音信号的特征分析与参数提取,从而能对说话人身份进行辨认和确认.它与语音识别的不同之处在于:前者力求挖掘出包含在语音信号中的说话人的个性因素,强调的是不同人之间的个性差异;而后者则是为了提取语音信号中包含的词语的共性信息,尽量把不同说话人的差别归一化[3].

目前,说话人识别的研究重点在对各种反映说话人特征的声学参数的线性或非线性处理以及新的说话人识别模式匹配方法上,如动态时间归整(dynamic time warping, DTW)、矢量量化(vector quantity, VQ)、隐马尔可夫模型(hidden markov models, HMM)、人工神经网络(artificial neural networks,ANN)以及这些方法的组合技术等.基于HMM的方法需要较多的模型训练数据、较长的训练及识别时间,而且还需要较大的内存空间,而基于VQ算法的自动说话人识别系统便于硬件实现[4].本文采用矢量量化方法(VQ),这种匹配方法不需要对时间进行对齐,简化了系统的复杂度,判断速度快,识别精度高.

1 线性预测倒谱系数(LPCC)的提取及矢量量化

1.1 LPCC的提取

倒谱特征是用于说话人个性特征表征和说话人识别最有效的特征之一.语音信号是声道频率特性和

收稿日期:2007-05-29

通讯作者:党幼云(1962-),女,陕西省澄城县人,西安工程大学教授. E-mail:xk_dyy@tom.com

激励信号源两者共同作用的结果. 而说话人的个性特征很大程度上取决于说话人的发音声道，即声道频谱特性，因此需要将这两者进行分离. LPCC 参数可由线性预测系数(Linear Prediction Coefficients, LPC)递推得到，与直接计算倒谱系数相比，LPCC 的计算量要小的多. 其递推公式为

$$c_n = a_n + \sum_{i=1}^{n-1}(i/n)c_i a_{n-i}, 1 \leqslant n \leqslant p, \tag{1}$$

$$c_n = \sum_{i=1}^{n-1}(i/n)c_i a_{n-i}, n > p. \tag{2}$$

式中　$a_1, a_2, \cdots, a_p$ 为 p 阶 LPC 特征向量. 当 LPCC 的阶数不超过 LPC 阶数 p 的时候，用(1)式进行计算；如果 LPCC 阶数大于 p，则用(2)式进行计算，此时实际上是一种外推.

线性预测倒谱系数比较彻底的去掉了语音产生过程中的激励信息，主要反映声道特性，而且只需 10 余个倒谱系数就能较好地描述语音的共振峰特性，计算量小；其缺点是对辅音的描述力差，抗噪性能也较弱.

1.2 矢量量化

矢量量化是一种极其重要的信号压缩方法，广泛地应用于语音信号压缩领域. 基本思想是将若干个标量数据组构成一个矢量，然后在矢量空间给以整体量化，从而压缩数据而不损失多少信息.

在基于 VQ 的说话人识别系统中，矢量量化起着双重作用. 在训练阶段，把每一说话者所提取的特征参量进行分类，产生不同码字所组成的码本. 在识别阶段，先从待识别的语音中提取特征矢量序列 X，然后计算待识别人语音的特征矢量 X 与集合内每个人的码本的距离，距离测度使用绝对值平均误差. 根据计算得出的距离值最小的说话人作为结果输出. 在说话人确认中还需要设置一个阈值，用于判断是否为说话人本人，这个阈值可以根据实际的结果进行调整.

2 实验及结果分析

2.1 语音库的建立

实验的录音数据使用 PC 机声卡通过音频采样级别 12kHz、采样精度为 16bit、单声道的 A/D 变换转化成数字信号存储. 共有说话人 13 人，其中男性 12 人，女性 1 人，每人分别读不同的文字 2 遍，一遍取语音长度为 18s 左右，作为注册语音；另一遍取语音长度为 6s 左右，作为识别语音.

2.2 系统的实现

系统的实现主要包括语音信号的预处理、特征提取、训练和识别过程.

预处理包括分帧、加窗和端点检测. 为了进行短时分析，必须对信号进行分帧处理，本系统采用一帧帧长为 256 点，帧移为 100. 为了使帧与帧之间平滑过渡，保持连续性，用可移动的有限长窗口进行加权的方法来实现. 窗函数的选择对于短时分析参数的特性影响很大. 本系统采用海明窗.

端点检测的目的就是从连续的声音中间检测出每一段语音的起始点和终止点. 准确的检测语音开始需要用短时能量和过零率配合来检测.

特征提取时先求出 10 阶线性预测分析系数，再递推出 15 阶的倒谱系数.

训练的过程即码本形成的过程. 对输入语音所形成的所有原始特征矢量，使用 LBG 算法形成码本并存储.

用于测试的语音数据同样要经过预处理过程得到原始特征矢量，然后与在训练过程中得到的模型模板加以比较，并根据一定的相似性准则进行判定.

2.3 判决距离测度的估计

从前面论述可知，VQ 识别对说话人的判决是根据量化平均误差来进行的. 希望一个人的码本模板对同一个人的语音参数的量化误差小，而对于不同人的量化误差大，只有这样才能够正确地工作[5]. 首先把一个人对“只能留住”的一遍发音中提取 15 阶的 LPCC 参量序列作为训练集合，通过 LBG 算法得到代表这个人的码本，然后分别将同一个人对语音“只能留住”的另一次发音和另外一个人对语音“喜欢随遇而安”的发音进行 VQ 匹配，比较他们各自的均方量化误差. 同时，改变码本容量，观察随之发生的变化，得到如图 1 所示的曲线.

在图 1 中，曲线 c 对应本段训练语音，曲线 b 对应同一个人的另一次发音，曲线 a 对应另外一个人的

发音.从图1中可以看到,量化均方误差的确可以区分不同的人,而且当码本容量大于一定的数值之后,曲线 c 训练集的量化误差逐渐趋于零.

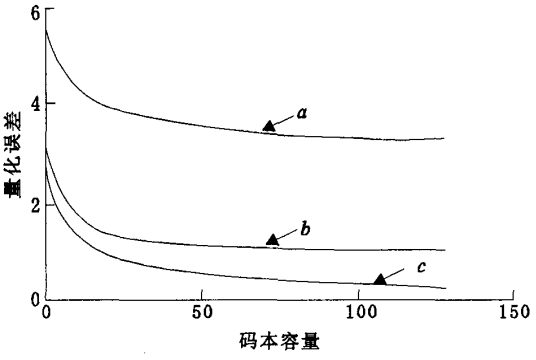

图1 语音信号的LPCC系数VQ量化误差曲线

2.4 说话人辨认实验

说话人辨认用以判断某段语音是若干人中哪一个所说,是多选一的问题.识别时,把所提取的参数与训练过程中的每一个人参考模型加以比较,返回距离最近即绝对误差最小的码本所对应的说话人作为系统的识别结果.表1为码本容量为64的各说话人与其他说话人码本之间的距离.其中横向对应训练语音长度为18s的说话人,纵向为与其相对应的识别语音长度为6s的说话人.

表1 码本容量为64时,各说话人之间的距离

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.8349	1.9251	2.2141	1.885	1.8995	2.0814	2.1801	2.0537	1.8939	2.1101	2.0147	1.8509	2.1566
2	2.0245	1.8368	2.1826	2.0056	1.9907	2.2029	2.1215	1.9995	1.9799	2.1367	2.0088	1.9962	2.1778
3	2.1082	2.2911	2.2639	2.17	2.0285	2.0652	1.9902	2.12	2.2781	2.1759	2.1207	2.0477	2.1269
4	1.9414	2.0113	1.9704	1.7851	1.8568	2.0618	2.0647	1.9666	1.912	1.9855	1.9065	1.9571	2.1393
5	2.0683	2.1813	2.1896	1.9681	1.7893	2.0627	2.0813	2.0524	1.8867	2.1131	1.9862	2.0026	2.106
6	1.9795	2.1321	2.0924	1.9142	1.8358	1.6736	2.0382	1.768	1.9486	2.0077	1.9006	1.8989	1.7686
7	2.31	2.5685	2.8109	2.3067	2.2649	2.4178	1.9769	2.6617	2.4723	2.5262	2.2123	2.293	2.5225
8	2.2328	2.236	2.2776	2.2288	2.2676	2.4627	2.5034	1.9401	2.2859	2.1983	2.4096	2.207	2.2803
9	2.0085	2.0899	2.1205	2.1252	1.9613	2.1686	2.2002	1.8976	1.8312	2.0645	1.9772	2.0358	1.9854
10	2.3043	2.3287	2.1386	2.1896	2.2206	2.1604	2.2722	2.0873	2.3527	1.9496	2.2466	2.1011	2.2704
11	2.0076	2.072	2.1113	1.9322	1.9273	2.0189	1.9768	2.0436	1.8928	2.1284	1.7897	1.9603	2.0855
12	1.8651	1.909	1.8792	1.8558	1.9283	1.857	2.0696	1.8596	1.9649	1.8719	1.9968	1.6903	1.9607
13	2.2536	2.5434	2.497	2.4029	2.2691	2.3908	2.5178	2.2618	2.4026	2.4375	2.375	2.3059	1.9707

由上述讨论可知,表1中对角线上的数值应为各行的最小数值.从表1可知,13人中有1人被误识(表中方框所示),其他说话人均能被正确识别.

为了比较不同码本容量对识别率的影响,将上述13人的语音分别进行码本容量为8,16,32和64的VQ量化,相应的识别结果如表2所示.从表2可以看出,随着码本容量的增加,系统的识别率也逐渐增加.当码本容量等于64时,系统的识别率达到92.31%.在码本容量为128时,识别率可达到100%.可见矢量量化的码本容量越大,其包含的码字矢量个数也越多,对说话人特征分布的描述就越细致,训练时产生的码本性能就越好,但这样会增加训练时需要的计算量和存储量.因此在保证识别率、降低码本的计算量和存储量的情况下,合理选择VQ的码本容量对系统的实现十分重要.

表2 测试码本大小对于系统识别能力的影响

码本容量	8	16	32	64	128
测试人数	13	13	13	13	13
正确数	8	11	12	12	13
识别率/%	61.54	84.61	92.31	92.31	100

2.5 说话人确认实验

说话人确认是确认一个人的身份,只涉及一个特定的参考模型和待识别模式之间的比较,系统只做出"是"或"不是"的二元判决,识别时是将从输入语音中导出的特征参数与其声音为某人的参考量相比较,如果二者的距离小于规定的阈值,则予以确认,否则予以拒绝.若以说话人1为参考模板,说话者自身和其他12位冒认者分别与其进行匹配,结果可以从说话人辨认实验看出,匹配距离为:1.834 9, 2.024 5,

2.108 2，1.941 4，2.068 3，1.979 5，2.31，2.232 8，2.008 5，2.304 3，2.007 6，1.865 1 和 2.253 6.

从结果可以看出其他 12 位冒认者的匹配距离均大于与其自身进行匹配的距离，因此阈值的设定可依据参考模板的主人对于自身模板匹配距离的最大值以及其他冒认者对参考模板的最小值匹配距离来确定，中间可适当加些裕量以提高其对未来的鲁棒性.

3　结束语

采用 LPCC 作为特征参量的基于矢量量化的说话人识别系统可以达到一个较高的识别率，VQ 算法可以识别出不同的说话人，但在做说话人确认实验时，其中阈值的设定是跟据大量实验得到的经验值，因而在一定程度上会影响其识别性能. 通常，在不同的应用环境中判决阈值是不一样的，所以，如何在说话人识别中确定最佳判决阈值还需要做进一步的深入研究.

参考文献：

[1] 卢管明，李海波，刘莉. 生物特征识别综述[J]. 南京邮电大学学报，2007(1)：81-89.
[2] RABINER L R，JUANG B H. Fundamentals of speech recognition [M]. 北京：清华大学出版社，1999.
[3] 马莉，党幼云. 特定人孤立词语音识别系统的仿真与分析[J]. 西安工程科技学院学报，2007(3)：371-373.
[4] 赵力. 语音信号处理[M]. 北京：机械工业出版社，2003.
[5] 江太辉. 基于 VQ 的说话人识别算法与实验[J]. 计算机工程与应用，2004(9)：77-99.

Algorithm and experiment of speaker recognition system based on VQ

LI Ai-ping，DANG You-yun

(School of Electronics and Information，Xi'an Polytechnic University，Xi'an 710048，China)

Abstract：Taking LPCC as the voice feature parameters，the VQ as the model matching method，the correlative experiment is carried on. Experimental result proves that this method can distinguish different speakers，and has a high precision of recognition to speaker identification.

Key words：VQ；LPCC；speaker verification；speaker identification

编辑、校对：武　晖

全国期刊出版形式规范检查结果公布
西安工程大学主办的两种期刊为首批合格期刊

国家新闻出版总署自 2007 年 7 月 1 日启动全国期刊出版形式规范检查。检查组将按期缴送样刊的 7 300 多种期刊全部初检、复检完毕，并经各省(市、自治区)新闻出版局报刊处认真核实，确定首批 3 305 种为合格期刊。检查中，刊物被分为 A、B、C 三类，A 类为合格期刊，B 类为待合格期刊(有几项指标空缺或不规范)，C 类为严重不合格期刊。A 类期刊名单已公布于《中国新闻出版报》2007 年 12 月 11～12 日第 3 版。经核查，西安工程大学主办的《西安工程科技学院学报》及《纺织高校基础科学学报》均被列入规范检查合格名单中，成为首批合格期刊。

据统计，全国 8 500 余种期刊，首批合格的占 38.4%(未按时缴送样刊的被判为不合格)；陕西省 164 种科技期刊中，首届合格 81 种，合格率达到 49.4%，高于全国水平 30%。

西安工程大学学报编辑部

2007-12-13