维吾尔语音节语音识别与识别基元的研究*)

王昆仑
(新疆师范大学数理信息学院计算机科学与技术系　乌鲁木齐830054)

A Study of Uighur Syllable Speech Recognition and the Base Element of the Recognition

WANG Kun-Lun
(Department of Computer Science & Technology, Xinjiang Normal University, Wulumuqi 830054)

Abstract This paper solves the recognition problems of the Vowel Consonant and efficiency syllable of Uighur speech, which is based on the Center-Distance Continuance Probability Model and Uighur speech database. And it analyzes the result of recognition, brings forward the rationality that Uighur syllable is used as the base element of recognition and the difficulties we find in this study. This paper also gives the method of how to solve these problems.

Keywords Speech recognition, Center-distance continuance probability model, Uighur syllable, Base element of recognition

1 引言

现代维吾尔语(以下简称维语)是维吾尔族人民的主要交际工具,是我国新疆维吾尔自治区的法定工作用语之一,也是新疆其它少数民族共同的交际用语之一。

维语属阿尔泰语系,突厥语族。维语语音有元音8个、辅音24个。由辅音和元音构成维语语音音节,每个音节必须且只能有一个元音,单元音可构成音节。维语句子由词构成。句子中有意群重音和句重音。部分音节在语流中产生语流音变现象,常见的有同化、弱化、脱落以及元音和谐等现象。

维语的发音规律和语音现象有很鲜明的特点,其元音、辅音以及语音结构的最小单位音节的识别对维语语音识别和选取最佳识别基元有重要意义。本文在CDCPM和维吾尔语语音数据库基础上,对维吾尔语元音、辅音和音节进行了识别并对识别结果进行了分析。提出了以维语音节为识别基元的合理性以及要解决的问题和方法。

2 识别模型

使用CHMM[1,2](Continuous Hidden Markov Model)进行语音识别时,主要是求解下列3方面问题:

(1)模型训练问题。对于输出序列 $O=\{O_1,O_2,\cdots,O_T\}$,计算模型参数 $\Lambda=\{\alpha,A,B\}$,使 $P\{O|\Lambda\}$达到最大。

(2)计算 $P\{O|\Lambda\}$问题。已知系统的3项特征参数 α、A、B,估计模型产生输出序列 $O=\{O_1,O_2,\cdots,O_T\}$的概率。

(3)对于模型输出序列 $O=\{O_1,O_2,\cdots,O_T\}$,估计一个对此输出序列最可能经历的状态系列$\{S_1,S_2,\cdots,S_t\}$。

中心距离连续概率模型(CDCPM, Center-Distance Continous Probability Model)[3]是基于CHMM的。CDCPM根据对转移概率在语音识别中的作用和地位的认识,简化了HMM模型,去掉了状态转移概率矩阵。把输出概率(密度)矩阵中各状态的概率密度函数用一个一维的符合CDN(Center-Distance Normal)的分布概率密度函数来描述。CDCPM的混合CDN分布描述为:

$$b_n(O_t)=\sum_{m=1}^{M} g_{nm} N_{CD}(O_t;m_{xnm},m_{ynm}) \tag{1}$$

其中:

$$N_{CD}(x;m_x,m_y)=\frac{2}{\pi m_y}\exp\{-y^2(x,m_x)/\pi m_y^2\} \tag{2}$$

$$m_{xnm}=E[O_t^{(nm)}] \tag{3}$$

$$m_{ynm}=E[y(O_t^{(nm)},m_{xnm})] \tag{4}$$

m 为CDN的混合数,$E[\cdot]$表示求数学期望,$y(\cdot,\cdot)$表示两个向量间的距离。$O_t^{(nm)}$ 表示属于第 n 状态第 m 混合的任一向量。g_{nm}是混合增益。

对一个HMM模型 Λ,用Viterbi算法[2,4]可以得出一个最大似然状态序列:

$$S^{(ML)}=\arg\max_s(P_d\{O|\Lambda,S\}-P\{S|\Lambda\})=\arg\max_s(a_{i_1}\prod_{t=2}^{T}a_{i_{t-1}i_t})\cdot(\sum_{t=1}^{T}b_{i_t}(O_t))$$

以此时的观察序列的输出概率作为其与模型的匹配得分,即

$$P_d\{O|\Lambda,S\}-P\{S|\Lambda\}=(a_{i_1}\prod_{t=2}^{T}a_{i_{t-1}i_t})\cdot(\prod_{t=1}^{T}b_{i_t}(O_t))$$

而CDCPM去掉对匹配影响相对较小的A,取

$$\mathrm{Score}\{O|\Lambda,S^{(ML)}\}=P_d\{O|\Lambda,S^{(ML)}\}=\prod_{t=1}^{T}b_{s_t^{(ML)}}(O_t)$$

作为观察序列的输出概率与CDCPM模型的匹配得分。在Viterbi算法中,令 $a_{ii}=a_{i,i+1}=1/2$。

语音分析窗帧宽取30ms,帧偏移为15ms。每帧语音经Hamming窗加权。对每帧加权语音用Durbin递推算法[5]进行线性预测编码(LPC),分析得到阶数为20的LPC系数。由LPC系数计算得到20阶倒谱特征系数 $C_n(m)$,$m=1\sim Q$(Q为倒谱特征系数维数)。

3 语音数据库

维语语音数据库[6]分为5个部分:音节语音数据库DB Ⅰ、

*)本研究得到国家自然科学基金(项目编号:69562001)和新疆维吾尔自治区"九五"重点攻关科研项目基金(项目编号:G9532603)资助。王昆仑副教授,研究方向为语音识别、中文信息处理和数据压缩。

词语语音数据库DBⅠ、语句语音数据库DBⅡ、数字语音数据库DBⅣ和常用符号语音数据库DBⅤ。数据库语音采样频率22050Hz，采样精度为16bit。

在语音数据库DBⅠ中，我们选择了维语的有效音节2160个，包括8个单音素音节、318个二音素音节、1765个三音素音节、69个四音素音节。这些音节中，有一部分为成词音节(音节本身构成一个词)，其余非成词音节仅作为构词的部件。常用单词多是由两个或三个音节构成。两个音节及以上的词中，最后一个音节为词重读音节。另外，有24个辅音单音素语音可以用做语音分析、语音识别研究之用。发音人发音可以略带方言。要求口齿清楚，发音清晰即可。语料的不同发音人，重复发音次数35遍。

4 音节识别及其分析

CDCPM模型参数包括：模型的状态数N、混合密度数M、特征矢量维数D、混合成分CDN分布参数μ_{nm}、混合增益g_{nm}。($1\leqslant n\leqslant N, 1\leqslant m\leqslant M$)。据实验，特征参数线性预测编码(LPC)阶数取20，倒谱特征系数(CEP)阶数取20。语音数据库是DBⅠ。

4.1 实验一：单音素音节模型的状态数N和混合密度数M实验

实验中，单音素音节有8个元音、24个辅音。模型数为32。测试语音随机取男、女各一人，测试语音数为64个。实验结果见表1。实验看出CDCPM混合密度数M取10、模型的状态数N取21时识别效果较好。

表1 状态数N和混合密度数M实验(%)

N	M	一选	二选	三选	四选	五选
16	8	81.25	87.50	95.31	96.88	98.44
16	10	90.63	95.31	96.88	98.44	98.44
16	12	89.06	93.75	95.31	95.31	98.44
20	10	84.38	89.06	95.31	96.88	96.88
21	10	92.19	98.44	98.44	100	100
22	10	85.94	87.50	93.75	96.88	96.88

4.2 实验二：元音、辅音识别实验

我们对元音和辅音分别进行了识别。元音模型数8个、辅音模型数24个，模型状态数N取21，模型混合密度数M取10。测试语音随机取男、女各一人。元音测试语音数为16个，辅音测试语音数为48个。测试结果见表2。

可以看出，元音一选、二选正识率比辅音高4个百分点。经分析，元音误识出现1次，而辅音误识则多达5次。误识的音及其原因见表3。其中："i"和"e"都是前元音，发"i"音舌位高，发"e"音舌位半高。是一男性发音者的发音不平稳，带明显的重读造成的。经对辅音发音者波形和试听分析，除了"m"音是因为发音者发音不准确造成误识，"m"、"ng"和"n"都是鼻音，"m"是双唇鼻音，"ng"是舌根鼻音，而"n"是舌尖鼻音。其他都是语音的浊音部分很弱或者时长很短时出现误识。

表2 元音、辅音识别实验(%)

语音	一选	二选	三选	四选	五选
元音	93.75	100	100	100	100
辅音	89.58	95.83	100	100	100

表3 误识原因分析

待识语音	误识为	误识原因
"i"	"e"	发音不平稳
"t"	"k"	发音不准确
"g"	"k,t"	发音不准确
"°l"	"l"	擦音部分时长短
"j"	"z"	音强很弱
"m"	"ng,n"	发音不准确

"t"和"k"都是清塞音，声带不震动，呼出气流发"t"时较强，发"k"时强。"g"是浊塞音，该发音者发音时，声带震动不够，接近清塞音。"°l"是浊擦音，发音者由于气息不足，未完全发出浊擦音。"j"也是舌尖浊擦音，发音者将浊音发为清音。

4.3 实验三：二音素音节、三音素音节、四音素音节识别实验

我们选取了二音素音节、三音素音节、四音素音节各32个，分别进行了识别实验。识别结果见表4。模型状态数N仍取10，模型混合密度数M取10，测试语音随机取男、女各一人，测试语音数各为64个。

表4 二音素、三音素、四音素音节识别实验(%)

语音	一选	二选	三选	四选	五选
二音素	85.94	87.50	89.06	89.06	89.06
三音素	82.81	85.94	87.50	90.36	92.19
四音素	81.25	85.94	90.63	92.19	93.75

通过表4我们看到，二音素音节、三音素音节、四音素音节识别正识率在81%～85%之间。分析误识原因，有以下几个：

(1)辅音发音不充分，特别是在尾部的辅音和发音速度较快的发音者的发音，出现减音现象。发塞音和擦音时舌位不到位。随着音节中音素的增加，辅音发音不充分的现象更突出，造成误识率逐步升高。在自然语音流中将更明显。

(2)音节发音的重音问题。重音发音时在音强、音高和音长三方面都有所增加，其中音强和音长是维语重音的基本特征，在音节发音时，一般的发音者按照非重读发音，少数发音者读做重读音节。据分析，发音者在对二音素音节、三音素音节、四音素音节的重读处理时，重读音节的元音的时长比非重读音节的元音的时长长一倍或者更长一点。误识出现较多的是读做重读音节的语音。

(3)发音者音高的变化也影响识别率。经分析，发音者音高无变化或者变化较小的语音识别率较高，而音高有显著上仰或者下降的语音出现误识情况。

(4)非成词音节的识别出现误识。产生这种现象的原因是读非成词音节时发音者读音的随意性较大。

在音节中，元音弱化表现不明显，但辅音弱化的情况就比较明显。辅音的弱化在文字中不能反映出来，但由于发音者发辅音时呼出不带噪音的气流，浊辅音就变成了清辅音，有些清塞音、清塞擦音变成了擦音。通过上述实验，这些维语中特殊的发音现象对识别结果都产生了影响。

5 识别基元

由此可见，维语语音识别基元如果以音素或者音节为单位，由于其发音规律的复杂性，对识别影响较大的有：①元音和辅音的弱化现象。在音节识别中，元音的弱化现象较少，而辅音的弱化问题已经反映了出来。在连续语音流中，这种情况

进一步严重时，将出现元音的脱落、辅音的脱落甚至音节的脱落。②辅音的同化问题。这主要表现在辅音的顺同化（前一个辅音制约后一个辅音的发音）和逆同化（后一个辅音影响前一个辅音的发音）等，这些现象有普遍性。所以，若识别基元以音素为单位，由于前述原因，不但其数量较大，而且其描述和定量都难于实现。若以音节为单位，则要解决两个问题，一是发音的规范化问题或者元音、辅音弱化的识别处理问题；二是由于维语音节是由一个元音与几个辅音或者一个元音单独组成，辅音之间可以连续拼读。一个词由多个音节按音节结构（常用的9种结构）组成，其发音是连续的。所以以音节为识别基元，音节与音节的划分的准确性与正识率成正比。

从维语语音学规律来看，以音节为识别基元并结合音节类型（开音节和闭音节）以及音节结构类型（9种）信息进行规范处理，识别正识率会有所提高。相关问题另文阐述。

结束语 在CDCPM下对元音和辅音的识别中可以看出维语的元音识别率很高，而辅音的识别率较低。二音素音节、三音素音节、四音素音节识别正识率随音素的增加而降低。而最佳识别基元的确定是语音识别必须解决的问题。本文对维语音节的识别，可以看出识别效果与发音规律有密切关系，以音节作为识别基元是可行的。结合维语语音的有限组合规律[7]，是一条可行的研究思路。

参考文献

1 Rabiner L R, Schafer R W. Digital Proce- ssing of Speech Signals. USA: Prentice Hall, Inc., 1978

2 杨行峻，迟惠生. 语音信号数字处理. 北京：电子工业出版社，1995

3 郑方，吴文虎，方棣棠. CDCPM及其在语音识别中的应用. 软件学报，1996，7：69

4 Forney G D. The Viterbi Algorithm. Proc. IEEE, 1973, 61: 268～278

5 Gold D, Rader C M. Digital Proce-ssing of Signals. New York: McGraw-Hill, 1969. 246

6 王昆仑，樊志锦，吐尔洪江，等. 维吾尔语综合语音数据库系统. 见：哈尔滨工业大学第五届全国人机语音通讯学术会议论文集. NCMMSC-96, 1996. 366

7 方晓华. 现代维语教程（上册，语音篇）. 乌鲁木齐，新疆师范大学，1987

（上接第181页）

是：

$\{d_6, d_7, d_0, d_1, d_2, d_3, d_5, d_4\}$；

若 $F(M')<0$，则在搜索区域内的先后搜索方向序列顺序是：

$\{d_7, d_6, d_0, d_1, d_2, d_3, d_5, d_4\}$。

同理，可讨论其他段的优先搜索方向序列。对任一基圆弧，通过以上的方法确定优先搜索后继像素的方向序列表，对在圆弧上概率大的方向优先搜索，减少搜索次数，提高算法的效率。

3.5 整体圆或圆弧的产生

对每一基圆弧，在它的搜索区域内按其所确定方向序列表中的顺序进行搜索。若某基圆弧在识别其所确定的圆或圆弧时搜索到已经搜索过的点，仍继续搜索下去（因为这一点可能是交点），对连续已经搜索过的点进行统计，直到这些点数和大于给定的阈值为止，或到图形的边界或出现像素值为0的像素为止。经上述处理后，就可得到图像中的完整圆或圆弧。对已识别的部分进行标记，以后不再重复进行处理。由于单一图段可能产生多个基圆弧，对产生的所有基圆弧中没有识别的进行处理，识别出所有可能的圆或圆弧，再处理下一个单一图段，这就大大减少了一次的处理量。由于对已识别的部分只是进行标记，没有进行删除，保留图像原有的信息。因此，对图像中的交点不需用阈值等方法进行判断，可直接对圆或圆弧进行整体识别。

对没有标记的像素进行上述处理，直到图像中的所有的像素均被处理，识别出所有的整体圆或圆弧。

结束语 通过获取工程图中的基圆弧，确定由该圆弧所在的圆的跟踪方向和范围，一次对一条圆或圆弧进行整体识别，避免了分段后再合并所需进行的重复跟踪、共圆判断和连接，简化了矢量化过程，可使圆或圆弧识别的速度和精度有较大的提高。通过对扫描图（图6）中的工程图用本文提出的方法进行识别，其识别过程简洁、快速、准确，得到了比较满意的效果（图7）。通过进一步的研究，提高识别范围，就可以达到对各种工程图进行识别。

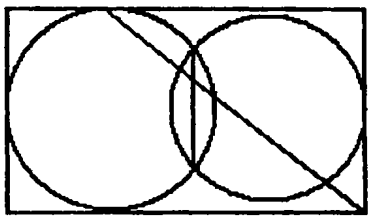

图6 工程图

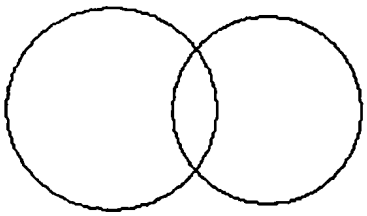

图7 识别图

参考文献

1 谭建荣，彭群生. 基于图形约束的工程图像直线整体识别方法. 计算机学报，1994，17(8)：561～569

2 Jasson D T, Vosseoel A M. Adaptive vectorization of line drawing images. Computer Vision and Image Understanding, 1997, 65(1): 38～56

3 Nagasamy V, Langrana N A. Engineering drawing processing and vectorization system. Computer Vision, Graphics and Image Processing, 1990, 65: 379～397

4 宋晓宇，王永会. 工程图自动矢量化算法的设计与实现. 中国图像图形学报，2000，5(1)：66～69

5 Dori D. Vector-based arc segmentation in the machine drawing understanding system environment. IEEE Transactions on Pattern and Machine Intelligence, 1995, 17(11): 1057～1068