

动态调整直方图剪枝 PDA 声控 拨号器的应用与实现

· 论文 ·

陈德锋¹, 郑方¹, 吴文虎¹, 刘建¹, 邓菁¹, 宋战江², 周迅溢²

(1. 清华大学 计算机系 智能技术与系统国家重点实验室 语音技术中心, 北京 100084;
2. 北京得意音通技术有限责任公司, 北京 100085)

【摘要】 以使用嵌入式操作系统 Pocket PC 的个人数字助理(PDA)为实验平台研究了基于非特定人语音命令识别的可定制声控拨号器。针对 PDA 存储空间和运算能力的限制,在保证性能的前提下从严格控制搜索空间和提高解码速度出发,提出了结合搜索路径分数差值实时调整剪枝宽度的动态调整直方图剪枝策略,提出了利用速查表加速似然计算的方法,并在通过实验验证后采用较少维数的特征,结合扩展声韵母进行声学建模等措施,有效地解决了上述问题。在实际 PDA 设备上实验表明,在词表大小为 200 个人名时,识别正确率达 98.70%,而识别速度比采用标准算法的参考系统提高了约 80 倍,同时节省了约 30%搜索存储空间。

【关键词】 语音识别; 声控拨号; 个人数字助理(PDA); 动态调整直方图剪枝

【中图分类号】 TN912.34 **【文献标识码】** A

Application and Implementation of Dynamically-Adjustable Histogram Pruning for PDA Voice Dialing

CHEN De-feng¹, ZHENG Fang¹, WU Wen-hu¹, LIU Jian¹, DENG Jing¹, SONG Zhan-jiang², ZHOU Xun-yi²

(1. Center for Speech Technology, State Key Laboratory of Intelligent Technology and System,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;
2. Beijing d-Ear Technologies Co. Ltd., Beijing 100085, China)

【Abstract】 Memory and speed are two demanding factors that must be faced when applying voice dialing speech recognition system to Pocket PC. We proposed a novel decoding method, dynamic histogram pruning adjusted by the difference scores of token paths, to precisely control the decoding search space and improve decoding efficiency. Besides, a new acoustic modeling method based on Extended Initial/Final(XIF) with less feature dimension is proven suitable for embedded speech recognition. By using the above methods developed, we implemented a speaker-independent and user definable voice dialing speech recognition system with good performance on a real PDA device. In 200-word-sized vocabulary, it obtained the accuracy of 98.70% and better recognition speed with 30% decoding space saving in comparison to the baseline system.

【Key words】 speech recognition; voice dialing; user-definable vocabulary; dynamically-adjustable histogram pruning

1 引言

语音和语言作为人类最直接、最自然和最主要的交流方式,在无线通信和个人数字助理 PDA(Personal Digital Assistant)迅猛发展的今天,其相应的嵌入式语音识别技术逐渐成为新的研究热点。国内外许多公司及研究机构,像 IBM, Microsoft, 得意音通, 言丰科技, 清华大学, 中科院自动化所以及中科院声学所等,都在开发自己的嵌入式语音识别引擎。

PC 上中高性能的非特定人大词汇量连续语音识别在实验室已经取得令人满意的效果,但在 PDA 这样的嵌入式移动设备上,由于以下原因还无法发挥其性能:(1)运算速度不高。处理器一般运行的是 MIPS 指令,处理能力小,只相当于 PC486 水平;(2)存储空间不大。存储卡一般只有 16 M 或 32 M,与目前 PC 的内存大小相差若干个数量级;(3)声音信道不同。内置传声器信噪比较低,声音信道不同于普通的 PC 或固定电话信道,而目前用以训练声学模型的数据库大多不

是 PDA 设备环境下的。

因此,在 PDA 上进行嵌入式语音命令识别应用,要求算法在保证识别效果的前提下,减少时空复杂度,以适应嵌入式平台存储资源少、实时性要求高的特点^[1]。目前主要从以下两个层面进行:

(1) 针对 DSP 专用处理器芯片的应用。主要是基于芯片处理器加快语音识别的速度,比如清华大学语音技术中心基于 ADSP 芯片的语音命令开发^[2]、清华大学电子系的定点 DSP 芯片片上系统 SOC 实现^[3,4]等,都是把语音识别算法固化到芯片上,在早期硬件资源不够发达的情况下,被较多地采用。不足之处是:这样的系统往往都是特定人识别系统,需要在线训练模型,运算精度不高,可支持的词表规模不大,变化不灵活,识别效果不够好。

(2) 基于算法层面的应用。出发点是减少模型的规模和搜索空间,提高解码的速度;同时也可降低运算精度,增加软件层面的定点计算,进一步提高运算速度^[5,6]。随着嵌入式设备硬件资源的不断发展,计算和存储能力的提高,这种方式的应用越来越多。它既可以在线训练模型,实现特定人语音识别,也可以事先载入训练好的模型,实现非特定人语音识别,具有识别精度高、应用灵活的特点。不足之处是:声学模型需要在规模和精度之间折衷,还需要严格控制识别过程中搜索空间的消耗以及提高搜索速度。

笔者以嵌入式操作系统 Pocket PC 为实验平台,研究基于非特定人语音识别的声控拨号系统(声控拨号器)。针对 PDA 上的存储空间和运算速度,在不影响识别率的基础上,降低系统的时空复杂度,采取了以下措施:

(1) 研究了正确候选搜索路径在 Viterbi 束搜索中的排名顺序与输入语音帧数的动态变化关系,研究了 Viterbi 束搜索中排序后相邻排名的搜索路径分数差值随时间变化的关系,进而提出了一种结合搜索路径分数差值动态调整直方图剪枝的搜索策略;

(2) 研究了搜索解码过程中的似然分重复计算问题,提出在 Viterbi 解码中利用速查表加速计算方法;

(3) 研究了在不特别降低识别性能的情况下如何选用较少的特征维数进行声学建模,结合扩展声韵母作为识别基元,提出了一种适合嵌入式设备上声控拨号器的声学建模方法。

经过实验研究,笔者将上述方法综合应用于“得意声控拨号器”上,改进了系统的性能。

2 动态调整直方图剪枝的搜索策略

一般来讲,语音识别中搜索策略主要需要解决以下 3 个问题^[7]:(1) 准确性:有效地利用各种知识源,使识别结果尽可能的准确;(2) 低消耗:尽量少地占用系统资源,包括需要的内存空间,硬盘空间等;(3) 高效率:尽量快地得出识别结果,理想情况是语音流一经输入立刻得到识别的文本结果。

在计算机资源有限的情况下,必须做到这三者的平衡,而后两个则是嵌入式系统着力要解决的问题。大量的实验研究表明,在语音识别系统中,解码的计算量占整个语音识别计算量的 80% 以上^[8],故提高搜索解码的速度和效率对于整个语音识别系统性能的提高具有决定性的意义。一般来说,基于 HMM 的搜索,采用时间同步宽度优先的 Viterbi 方法^[7]会占用很大的搜索空间,搜索计算量也很大,显然不适合嵌入式设备,尤其当词汇量很大时,问题更为突出。最好的策略是,采用剪枝技术^[9]将计算量集中在最有可能的一些搜索路径上,但是,不恰当的剪枝也有可能把最优的路径过早地剪除。故选择一种好的剪枝策略显得非常重要。

2.1 传统剪枝策略

传统的剪枝算法可分为基于分数阈值剪枝和基于排名阈值剪枝。前者称为束剪枝(Beam Pruning),后者称为直方图剪枝(Histogram Pruning)^[10]。

基于分数阈值剪枝算法在每一时刻只保留似然概率分数最大的前若干条搜索路径,其中剪枝分数阈值称为束宽度。通常情况下,束宽度在剪枝过程中固定不变,这也称为固定束宽度剪枝。固定束宽度剪枝方法在阈值较小时可大幅度减少搜索空间,但当大多数路径的分数都非常接近时,则会导致搜索路径数急剧增加。故该方法对于搜索存储空间的控制并不够有效,因单纯从分数阈值上无法预知可保留的搜索路径个数。

基于排名的直方图剪枝可以有效地避免类似情况,直方图剪枝在任何情况下都只保留固定个数的、分数排名靠前的候选者,可以防止分数阈值剪枝中出现的搜索路径空间的无法预知性。其中,该固定排名个数称为直方图剪枝的剪枝宽度。对于在嵌入式系统上的声控拨号器来说,采用基于排名的直方图剪枝方法可能是比较好的选择。但是,固定的剪枝宽度不能动态适应实际的情况,反而不利于在正确率和速度之间找到合适的平衡。要想获得高正确率,需要增大剪枝宽度,这势必会导致计算量很大,因此需要进行改进。

2.2 时间动态调整剪枝策略

一般说来,当剪枝宽度太小时,直方图剪枝的识别正确率比较低;当加大剪枝宽度后,直方图剪枝的识别正确率随着提高。但是当剪枝宽度增加到一定程度后,识别率会呈现饱和,其上限就是标准 Viterbi 剪枝方法的识别率。因此,识别结果的好坏和剪枝的宽度大小有很大的关系。笔者对 200 个中文人名(150 个三字词+50 个两字词)构成的词表进行了标准 Viterbi 方法的性能测试,统计得到如图 1 所示的含有正确候选的路径的排名趋势图。

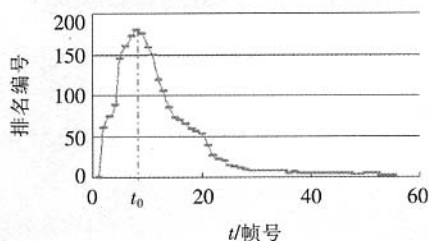


图 1 搜索路径中含有正确候选的路径的排名趋势图

从图 1 可以看出,在标准 Viterbi 解码过程中含正确候选的路径排名随时间(语音帧号)的变化统计上有如下规律:

(1)在初始阶段,其排名随着语音帧数的增加迅速后移(排名编号变大),后移的程度也并不固定,但平均来讲不会超过词表的大小。这是因为开始时的语音还不足以体现词的差异,故而正确候选的路径似然分数排名并不一定比其它的高。

(2)在中间阶段,其排名随着语音帧号的增加开始逐渐前移(排名编号下降)。这是因为足够长的语音可以很好地区分不同的词,输入语音与正确候选的声学模型匹配程度越来越高。

(3)在后面阶段,其排名随着语音帧号的增加一直稳定的排在前面,而且与之相近的几个候选路径也一直稳定排在前列。原先排名落后很多的路径很少能在这个阶段突然排在前列,最终成为识别结果的总是在很少的前几名路径之中。

从这些实验结果可看出:(1)剪枝宽度不必固定,可随时间减小;(2)要保证正确候选的路径在开始阶段留在剪枝宽度内,初始剪枝宽度要足够大,但是不必超过词表大小;(3)排名曲线峰值点的出现帧号(t_0)(图 1),即初始阶段的持续时间,实验上发现与词表平均词长(Len ,字数)以及输入语音总长度(T ,帧数)粗略地存

在如下对应关系

$$t_0 = 1.5 \times T / (Len + 2) \quad (1)$$

基于以上事实,笔者提出一种随时间动态变化剪枝宽度的直方图剪枝方法,称之为动态调整直方图剪枝。剪枝宽度在初始阶段要足够大,其大小与词表的大小有一定的函数关系,然后剪枝的路径宽度随着输入语音帧号(时间)的增加而动态减小。该动态变化过程可视为一个非线性时变系统,为简化起见,笔者利用分段线性的方法来近似描述图 1 所示的趋势。某个 t (单位是帧号)时刻的排名剪枝宽度

$$W_d(t) = g(t) \quad (2)$$

其中 $g(t)$ 是一个输入语音帧号 t 的分段线性函数,如下定义

$$g(t) = \begin{cases} g(t_0) & 0 \leq t < t_0 \\ g(t_0) - a_1 \times (t - t_0) & t_0 \leq t < t_1 \\ \dots & \dots \\ g(t_{m-1}) - a_m \times (t - t_{m-1}) & t_{m-1} \leq t < t_m \end{cases} \quad (3)$$

这里是用 m 段来近似逼近图 1 中的下降趋势,显然函数的形状是由常数 $\{a_1, \dots, a_m\}$ 和 $\{t_1, \dots, t_m\}$ 所决定的。调整这组常数可以调整剪枝宽度 $W_d(t)$ 下降速度,在实际中,这组常数通常是按照前松后紧的原则确定的。可见,通过分段线性函数 $g(t)$ 对剪枝宽度的控制,减少了峰值点(t_0)后的路径打分计算,可以达到大幅度减少解码中的打分次数的目的。

2.3 结合搜索路径分数差值动态调整剪枝策略

为了对剪枝宽度进行更精确的动态调整,笔者继续研究剪枝过程中搜索路径似然分数的情况。实验发现,在剪枝初始阶段,有时排名相差很多的路径之间的似然分数相差很小,而它们在搜索后期很可能上升为前几名,所以应该尽可能保证这些路径参与后续的路径扩展打分运算。这在当词表中有比较多的相同或相近发音时尤其需要。而在搜索后面阶段,排名相差不多的路径之间的似然分数相差已经很大,排名名次很难再上升从而成为识别结果。基于此,笔者将搜索路径的似然分数情况反映到直方图剪枝的剪枝宽度调整中。

为了更精确地定义 t 时刻的 $W_d(t)$,笔者引入 t 时刻路径的似然分数差值(即与第一名的分数差值) $\Delta Score(t)$:如果似然分数差值很大,说明排名比较靠后的路径成为最终结果的可能性非常小,则可以进一步加快剪枝宽度的收缩;反之,说明排名比较靠后的路径和排名靠前的路径成为最终结果的可能性非常接

近,则减缓剪枝宽度的收缩。时刻 t 的 $W_d(t)$ 由

$$W_d(t) = W_d(t-1) + f(\Delta Score(t)) \quad (4)$$

计算得到。其中函数 $f(\cdot)$ 是似然分数差值 $\Delta Score(t)$ 的调整函数: 如果该帧剪枝宽度处的路径似然分数差值大于某个分数差值阈值, 则将 $W_d(t)$ 缩小直到路径似然分数差值接近某个分数差值阈值; 如果该帧剪枝宽度处的路径似然分数差值小于某个分数差值阈值, 则将 $W_d(t)$ 扩大直到路径似然分数差值接近某个分数差值阈值。这里的路径分数差值阈值可以是动态的, 也可以是固定的, 实验发现动态的效果会更好, 即在初始阶段该分数差值阈值较小, 在后续阶段较大。

综合以上两种方式的动态调整, 最终在某个时刻 t 的 $W_d(t)$ 可表示为

$$W_d(t) = W_d(t-1) + g(t) + f^1 \Delta Score(t) \quad (5)$$

2.4 搜索策略比较实验

为了检验上述动态调整直方图剪枝方法(DHP)的性能, 笔者将其与标准 Viterbi 解码算法(VTB)、固定束宽度算法(FHP)进行了实验比较, 其中固定束剪枝(FHP)的剪枝宽度为 50。其中词表大小为 200, 平均词长 Len 为 2.7 字, 设语音长度为 T 帧, 则由式(1)可知 DHP 的峰值点 $t_0 = 0.32T$, 函数 $g(t)$ 中 $m=4$, 常数组 $\{a_1, \dots, a_m\}$ 和 $\{t_1, \dots, t_m\}$ 分别为 $\{100/T, 400/T, 200/T, 120/T\}$ 和 $\{0.4T, 0.5T, 0.75T, T\}$, 而相应的调整函数 $f(\cdot)$ 的动态阈值为 $\{80.00, 90.00, 100.00, 120.00\}$, 在 Dopod 686 (Intel StrongARM, 206 MHz, 32 MB RAM) 上得到实验结果如图 2 所示。

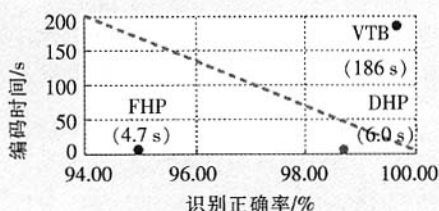


图 2 不同搜索策略的性能比较

可见, 在剪枝解码过程中, 影响识别正确率的最重要的因素是, 在初始阶段设置合适的剪枝宽度以便把正确候选包含进去; 而影响速度的最重要因素是, 在后续阶段将剪枝宽度有效地降下来。从实验结果来看, FHP 方法虽然在节省运算上很有效, 但是因为剪枝宽度固定不变而且在初始阶段太小, 导致本来能成为正确结果的路径过早被剪枝掉。当然, 如果笔者在实验中

选取固定剪枝宽度大, 则可能的结果是正确率提高而解码速度降低。

3 优化搜索实现中的似然计算速度

搜索剪枝在实际实现中通常可在两个层面进行: 状态级剪枝和模型级剪枝^[11]。在状态级上的束搜索剪枝的时候, 对所有状态进行判断是否保留; 在模型级的束搜索剪枝时, 则是对所有的基元模型进行分数判断是否保留, 若决定被剪掉, 那么该模型里的所有状态都将被剪掉。模型级的剪枝在存储上以及实现上都比较高效而且计算量更节省, 故更适用于嵌入式系统上的识别搜索。

为了进一步提高搜索效率, 减少搜索空间的存储, 笔者除了采用树型词典表示^[9]外, 还提出对搜索中的大量状态似然打分计算利用速查表进行优化。实验研究发现, 在搜索解码过程中, 计算每帧输入语音与所有候选语音模型每个状态的匹配得分过程包含了大量的(相同模型节点或状态节点的)重复计算, 而一般而言, 在搜索过程中, 状态似然分数的计算量占整个搜索过程计算量的 80% 以上^[12], 因此状态似然分数打分的优化对整个搜索解码的速度优化是另外一个重要的因素。针对以上这些情况, 笔者提出采用数据共享技术的思路, 以表格形式(即速查表)动态存储词表中节点模型的状态与每帧输入语音的匹配得分, 在计算每帧输入语音与所有候选语音模型每个状态的匹配得分时, 如果该状态在速查表中已经存在(即前面已打过分), 则直接得到似然得分; 否则, 计算该状态似然得分, 并记录在速查表中。通过这种速查表的方法, 可节省搜索中的大量重复似然分数计算, 从而大大提高了搜索解码的速度, 而不影响正确率。

在采用动态调整直方图剪枝的搜索策略的前提下, 笔者对是否应用速查表共享技术在不同词表大小的情况进行了对比实验, 实验结果见图 3, 其中百分数表示采用速查表共享技术后节省计算量的程度。

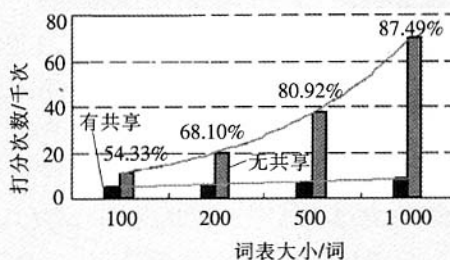


图 3 数据共享技术对搜索的优化实验

从以上实验可以发现,如果没有采用速查表共享技术时,打分次数随着词表大小以二次多项式增大;而采用速查表动态共享技术后,当词表大小较小时,打分次数以低斜率增长,但当词表大小继续增大后,则保持在一个比较稳定的范围。因为当词表涵盖了所有的模型状态后,搜索解码过程中速查表动态共享技术能够保证每个模型状态每一帧只计算一次,那么总的打分次数就只和模型状态总数和输入语音长度有关。

4 适用于嵌入式声控拨号的声学建模

声学模型的建模对于语音识别系统的性能也具有决定性作用。目前在非特定人语音命令识别系统中普遍采用两种声学建模方法:一是整词建模、整词匹配的方法。其优点是计算量小识别率高,缺点是词表必须固定不能及时变化。这是早期应用采取的方法;二是基于词树搜索^[13]和声韵母建模^[14]的方法。其优点是建模精度高,适用于较大规模的词表,而且使用时无需训练,更改词表方便,利于实际应用,缺点是算法复杂度高。

在实际应用中,为了能够实现无需在线训练、词表更改灵活性、可定制词表等要求,必须采用比词更小的基元来建模,笔者从整词建模方法和基于声韵母建模、词树搜索的方法中获得启发,选择扩展声韵母^[14]作为模型基元,离线事先训练好基元模型,在系统中应用的时候,根据词表中词条的声韵母序列拼接成整词模型,可以方便地定制各种词表。

在特征方面, Mel 倒谱系数 MFCC^[16] 抗噪性能较好,适应了嵌入式设备使用环境的多样和噪声。考虑到语音频带中低频携带噪音成分较多和高频成分携带了基音谐波特性,对于非特定人系统的语音特征影响不大。为了适应嵌入式设备上存储资源有限的特点,笔者在提取特征时从滤波器中过滤掉语音频带的低频和低频成分,并减少特征维数,忽略二阶差分系数,最后经过 CMN(Cepstral Mean Normalization)去噪处理,得到“{9 维 MFCC+1 维能量}+一阶差分”^[16]一共 20 维特征参数。笔者用该 20 维特征与 42 维特征 ({13 维 MFCC+1 维能量}+一阶差分+二阶差分) 和 28 维特征 ({13 维 MFCC+1 维能量}+一阶差分) 针对声控拨号器进行了实验比较,同样的训练数据和测试数据,得到如图 4 所示的实验结果。

实验证明,在语音命令识别中,采用这样的 20 维特征相对于 42 维特征的识别率降低是可以接受的,而且,这样做不但能够减小模型的规模(因为语音特征向

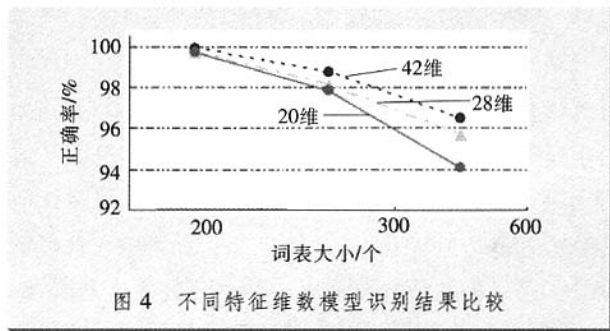


图 4 不同特征维数模型识别结果比较

量的维数小),也可以减少似然分计算时间,从而提高系统的速度,更具现实意义。

5 声控拨号器的实现及实验分析

基于以上基于搜索策略中的各种因素的分析声学建模分析,笔者设计并实现了一种在 PDA 上实用的汉语声控拨号系统——“得意声控拨号器”。它以 PDA 的联系人姓名为词表,自动提取联系人姓名信息构建词表,用户通过语音直接进行联系人快速拨号,非常方便快捷。同时用户可以非常方便地定制声控拨号词表:修改联系人姓名的读音,比如用昵称来代替姓名原来的读音等。而且用户也可定制声控拨号系统的识别参数,该声控拨号系统是个非特定人语音命令识别系统,无须在线训练,无须预先录制声控标签,而且识别率高、识别速度快、模型小(仅有 290 KB),非常适合嵌入式设备上的应用。

为了测试该声控拨号系统的性能,笔者也进行了系统性能测试。实验中,采用标准 Viterbi 解码方法,没有采用数据共享技术的系统作为参考系统(Baseline);与之进行比较的是综合采用了结合搜索路径分数差值动态调整直方图剪枝的搜索策略、动态数据共享策略的系统(d-EarVD)。

测试词表为 200 个随机的中文人名,其中两字词 50 个,三字词 150 个。测试的模型是利用 HTK^[17]工具从大量电话语音库训练得到,测试数据是在 Dopod 掌上电脑(采用 Pocket PC 2002 操作系统和 StrongARM 芯片)上通过 8 K/16 bit 采样实地得到的,由 6 个人(4 男 2 女),每人说一遍词表中的人名,一共 1 174 句,每个声音录制长度约为 3 s。实验结果如表 1 所示,其中搜索空间是指搜索解码过程所占用的存储空间。

从实验结果可以看出,综合采用适用于嵌入式系统的语音识别策略后,改进系统比基线系统在正确率方面只有轻微的下降,相对只有 0.96%,这在实际的嵌

入式系统应用中是可以接受的。而在速度和搜索的存储空间上有很大的改进,速度加快了约 80 倍,搜索存储空间减少了约 30%,提高了系统的实用性。而采用文中改进系统的“得意声控拨号器”成功推向市场也从实践上证明了改进系统的可用性和采用的各种策略的有效性。

表 1 改进系统与基线系统比较实验

系统	正确率/%	识别时间/s	搜索空间/KB
Baseline	99.66	186.0	190
d-EarVD	98.70	2.3	120

6 总结

提出结合搜索路径分数差值动态调整直方图剪枝的搜索策略和基于速查表的动态数据共享技术进行搜索优化,并基于声韵母建模和整词识别的非特定人语音命令识别方法,采用一套模型规模小精度高的特征,实现了一种适用于嵌入式系统的声控拨号系统——“得意声控拨号器”,在 PDA(Pocket PC)上实现,在词表大小为 200 个人名时,识别正确率达到 98.70%。与基于标准的 Viterbi 搜索的基线系统相比,识别速度提高约 80 倍,同时节省约 30%搜索存储空间。

未来除了使声控拨号在嵌入式环境下的更加快速和节省资源,还需要从以下方面进行改进:

(1)克服环境噪声的影响。PDA 作为一种移动通信工具,任何使用环境都有可能出现,尤其是噪声环境,需要增强系统的抗噪性能。

(2)克服用户口音的影响。让更多的用户能够随心所欲地使用声控拨号。

随着移动设备的技术迅速发展,语音识别的技术难点也在不断突破,相信以上困难也将会逐步被解决。

参考文献

[1] 方敏,浦剑涛,等. 嵌入式语音识别系统的研究和实现,第七届全国人机语音通讯会议,2003. 109-112.
 [2] 郑方,胡起秀,等. 介绍一种傻瓜式声控电话机,第四届全国人机语音通讯会议,1996. 165-168.
 [3] 朱纯益,陆建华,刘润生. 基于 DSP 的声控电子记事本的设计与实现. 电子技术应用,2002(9):71-73.
 [4] 荆嘉敏,刘加,刘润生. 基于 HMM 的语音识别技术在嵌入式系统中的应用. 电子技术应用,2003,29(10):12-14.
 [5] 邓勇刚,徐波,黄泰翼. Palm PC 语音识别算法及实现. 计算机研究与发展,2000,37(8):937-941.
 [6] 谢凌云,杜利民,刘斌. 嵌入式语音识别系统的快速高斯

计算实现. 计算机工程与应用,2004(23):30-31,70.
 [7] 张国亮. 大词表连续语音识别中搜索算法的研究与实现(硕士学位论文). 清华大学计算机系,2001.
 [8] 张国亮,郑方. 语音识别中基于两层词法树的跨词搜索算法. 清华大学学报(自然科学版),2003,43(7):981-984.
 [9] Suontausta J., Fast decoding in large vocabulary name dialing, ICASSP, 2000.
 [10] 张东滨,杜利民. 语音识别的自适应束剪枝方法. 电声技术,2004(8):41-45.
 [11] Janne Suontausta, Fast Decoding Techniques for Practical Real-time Speech Recognition System, IEEE Workshop ASRU99,Keystone,Clorado, 1999.
 [12] Imre Kiss, Marcel Vasilache, Low Complexity Technique for Embedded ASR System, ICSLP, Denver, Colorado USA, 2002.
 [13] 张国亮,郑方. 基于两层词法树的大词表连续语音识别搜索算法. 第六届全国人机语音通讯学术会议,2001. 239-242.
 [14] 李峰,浦剑涛,等. 基于声韵母建模基元拼接和整词识别的非特定人孤立词语音识别系统的研究. 第七届全国人机语音通讯会议,2003. 87-90.
 [15] 李净,徐明星. 汉语连续语音识别中声学模型基元比较. 音节、音素、声韵母,第六届全国人机语音通讯会议,2001. 267-271.
 [16] Fang Zheng and Guoliang Zhang. Integrating the energy information into MFCC, International Conference on Spoken Language Processing (ICSLP' 00), pp. I-389~292, Oct. 16-20, Beijing.
 [17] Yong, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. and Woodland, P., The HTK Book (for HTK Version 2.2), Cambridge University(1999).

作者简介

陈德锋,硕士,主要研究方向为嵌入式语音识别研究与应用。
 郑方,副教授,博士,清华大学语音技术中心实验室主任,研究方向为语音识别,声纹识别,自然语言理解。
 吴文虎,教授,博导,研究方向为语音识别。
 刘建,博士,主要研究方向为关键词检出,声学特征等。
 邓菁,博士,主要研究方向为声纹识别。
 宋战江,博士,得意音通技术有限责任公司技术总监,研究方向为语音识别产品研发。
 周迅溢,硕士,得意音通技术有限责任公司,研究方向为语音识别产品嵌入式研发。

[收稿日期] 2005-07-23