利用 DSP 实现的实际环境下语音识别方法 [1]

肖圣兵 赵 力* 刘海滨* 吴镇扬*

(苏州大学电子工程系 苏州 215006)
*(东南大学无线电工程系 南京 210096)

摘 要 该文提出了一种在实际环境下利用 DSP 实现的语音识别方案，通过户外实际环境的语音识别实验，这种方法的有效性得到了验证.

关键词 语音识别， DSP, 噪声
中图号 TP391.42, TN912.3

1 引 言

如同手机语音拨号识别装置一样，用一片数字信号处理器 (DSP) 实现的，在户外公共场所操作的语音识别装置，必须具有很强的抗环境噪声的能力以及实时的高识别率的性能，而且最好是非特定人的识别系统. 为此，本文利用对于传统的谱相减 (SS: Spectrum Subtraction) 降噪声技术 [1] 的修改以及概率尺度的动态规划 (DP: Dynamic Programming) 算法 [2] ，提出了一种在实环境下利用 DSP 实现的非特定人语音识别方案，并且实验证明了该方案的有效性.

2 具有输入幅值谱自适应的 SS 方法

利用 SS 法进行降噪处理仍然是当今主要的降低环境噪声的方法. 设对于第 t 帧幅值谱的第 i 元素，噪声下的语音功率是 $|y_i(t)|^2$ ，估计的噪声功率是 $|\overline{n}_i|^2$ ，除噪后语音的功率是 $|\overline{s}_i(t)|^2$ ，则传统的 SS 法如下.

$$|\overline{s}_i(t)|^2 = \begin{cases} |y_i(t)|^2 - \alpha|\overline{n}_i|^2, & |y_i(t)|^2 > \alpha|\overline{n}_i|^2 \\ 0, & \text{其它} \end{cases} \tag{1}$$

式中， α 为权值. 由于传统的 SS 方法考虑噪声为平稳噪声，所以对于整个语音段，噪声功率取相同的值. 同时对于整个语音段， α 一般也取相同的值. 而实际环境下的噪声，例如展览会中的展示隔间内的噪声是非平稳噪声，所以用相同的噪声功率值是不确切的. 同样，传统的 SS 方法用相同的权值 α ，有可能发生减除过度或过少的问题，使得有的区段要么噪声消除不够，要么减除过多产生 $|\overline{s}_i(t)|^2$ 失真. 为此，本文对传统的 SS 方法进行了如下修改. 首先，对于噪声功率估计，采用 (2) 式:

$$|n_i(t)|^2 = (1-\beta)|n_i(t-1)|^2 + \beta|x_i(t)|^2, \qquad 0 < \beta < 1 \tag{2}$$

在当前区段用语音以外的输入帧 $|x_i(t)|^2$ ，对噪声功率进行逐次更新. 其次，让 α 和输入语音功率相适应，使权值 α 按 (3) 式:

$$\alpha_i(t) = \begin{cases} C_1, & |y_i(t)|^2 < \theta_1 \\ [(C_2 - C_1)/(\theta_2 - \theta_1)]|y_i(t)|^2 + C_1, & \theta_1 < |y_i(t)|^2 < \theta_2 \\ C_2 & |y_i(t)|^2 > \theta_2 \end{cases} \tag{3}$$

随输入语音功率谱值改变，以避免产生减除过多或过少的问题. 式中 θ_1 和 θ_2 为门限制阈值，C_1 和 C_2 为常数， $(\theta_2 > \theta_1, C_2 > C_1)$.

[1] 2001-12-17 收到， 2002-08-08 改回

3 概率尺度的 DP 识别方法

用一片 DSP 实现的语音识别装置，为了节约它的存储和运算成本，一般采用矢量量化 (VQ) 方法或者 DP 方法进行识别，因为对于小词汇量单词或词组识别系统来讲，VQ 和 DP 方法足以满足识别性能的要求。但是，传统的 VQ 和 DP 方法一般只能适用于特定人的语音识别系统。为了使 DSP 语音识别装置也能适用于非特定人的语音识别，本文提出了利用概率尺度的 DP 进行识别的方法 [2]。例如对于如图 1 所示的非对称型 DP 路径 [3]，具有概率尺度的 DP 方法的递推公式可以用 (4) 式：

$$G(i,j)=\max\begin{cases}G(i-2,j-1)+\lg p(\boldsymbol{y}_{i-1}|j)+\lg p(\boldsymbol{y}_i|j)+\lg p_{\mathrm{PS1}}(j)\\ G(i-1,j-1)+\lg p(\boldsymbol{y}_i|j)+\lg p_{\mathrm{PS2}}(j)\\ G(i-1,j-2)+\lg p(\boldsymbol{y}_i|j)+\lg p_{\mathrm{PS3}}(j)\end{cases}\tag{4}$$

表示。这里 $\lg p_{\mathrm{PS1}}(j)$，$\lg p_{\mathrm{PS2}}(j)$，$\lg p_{\mathrm{PS3}}(j)$ 分别表示 3 条路径 PS1，PS2，PS3，即 $q((i-2,j-1)\to(i,j))$，$q((i-1,j-1)\to(i,j))$，$q((i-1,j-2)\to(i,j))$ 的状态转移概率。

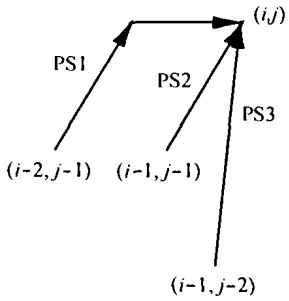

图 1 非对称型 DP 路径

3.1 条件概率 $p(\boldsymbol{y}_i|j)$ 的作成 假定在状态 j 观测到的 $\boldsymbol{y}_i$ 是符合 $(\boldsymbol{\mu}_j,\Sigma_j)$ 的高斯分布，则条件概率 $p(\boldsymbol{y}_i|j)$ 由下式给定。

$$p(\boldsymbol{y}_i|j)=(2\pi)^{-p/2}|\Sigma_j|^{-1/2}\times\exp\{-1/2(\boldsymbol{y}_i-\boldsymbol{\mu}_j)^t\Sigma_j^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_j)\}\tag{5}$$

为了求出各个时刻的均值和方差，首先选择一个学习样本序列作为核心样本，然后输入一个同类的学习数据和核心样本进行 DP 匹配寻找最佳路径函数 F，这时各个时刻的均值和方差可以通过最佳路径函数 F 找出和核心样本对应时刻的输入帧矢量进行计算和更新，如此重复直到同类的学习数据用完为止，渐近地求出各个时刻的均值和方差。在学习数据较少时，可以利用分段区间数据计算均值和方差，尤其是方差。

3.2 状态转移概率的作成 各个学习数据和核心样本进行 DP 匹配时，记下各时刻选择的路径情况 (如图 1 所示的 3 个路径之一)，学习完毕后，假定在时刻 j，3 个路径被选择的总数分别是 PS1(j), PS2(j), PS3(j)，则此时的 3 个状态转移概率可由 (6) 式给定：

$$\left.\begin{aligned}p_{\mathrm{PS1}}(j)&=\mathrm{PS1}(j)/\{\mathrm{PS1}(j)+\mathrm{PS2}(j)+\mathrm{PS3}(j)\}\\ p_{\mathrm{PS2}}(j)&=\mathrm{PS2}(j)/\{\mathrm{PS1}(j)+\mathrm{PS2}(j)+\mathrm{PS3}(j)\}\\ p_{\mathrm{PS3}}(j)&=\mathrm{PS3}(j)/\{\mathrm{PS1}(j)+\mathrm{PS2}(j)+\mathrm{PS3}(j)\}\end{aligned}\right\}\tag{6}$$

3.3 *识别方法*　识别时，对于输入语音信号序列利用 (4) 式和各个模型进行 DP 匹配，给出最大得分的模型所对应的类别即为识别结果.

4 识别试验和结果

识别装置结构如图 2 所示. 采用 TMS320C54x 系列 16 位定点 DSP[4−7]，该片内部附有 64kROM 和 64kRAM，经过测算本系统全部程序占有 31k 存储容量.

识别实验由 2 部分组成. 在第 1 个实验中，为了评价上述概率尺度的 DP 识别方法，我们进行了非特定人语音识别实验. 我们采用 35 个 4 位数汉语连续语音数字进行了识别试验. 邀请 20 名男性每个人对 35 个 4 位数字各发音 4 遍，其中 12 个人的发音作为训练用数据，另 8 个人的发音作为识别用数据. 试验结果表明 8 人的平均识别率是 96.9%. 第 2 个实验是户外实际场所语音识别实验. 我们选择 50 个人名，由 1 人对 50 个人名各发音 3 遍，其中 2 遍发音作为训练用数据，另 1 遍发音作为识别用数据. 实验是在市民广场、交通路口和学校食堂 3 种不同的环境下进行，结果是这 3 种环境下的识别概率分别是 98%、92%、96%，达到了较高的识别精度.

5 结　论

本文介绍了利用改进的 SS 降噪声技术和概率尺度的 DP 算法实现的单片 DSP 实际环境下语音识别方案. 实验证明利用该方法实现的语音识别装置，可以在户外不同场合的实际环境下，达到高识别率的性能，并且适宜于非特定人的语音识别，因此，是一种有效的汉语语音识别方法.

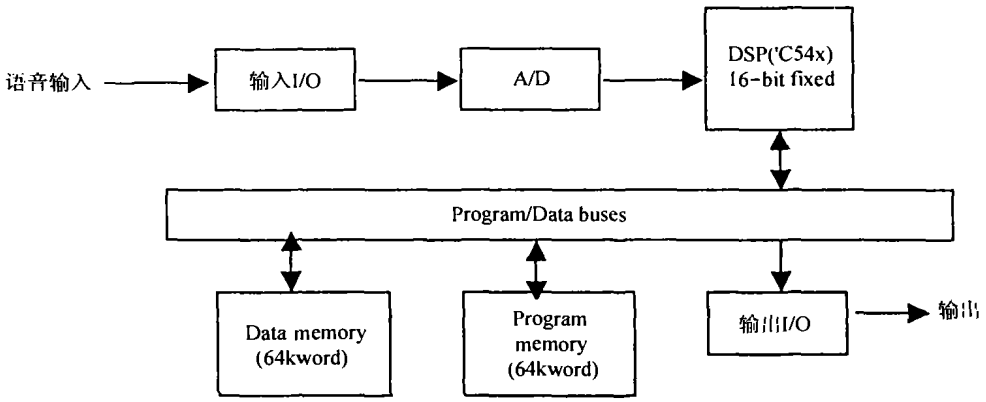

图 2　识别装置框图

参　考　文　献

[1] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. on Acoust., Speech & Signal Processing, 1979, 27(2), 113–120.

[2] L. Zhao, H. Suzuki, S. Nakagawa, A comparison study of probability functions in HMMs through spoken digit recognition, IEICE Trans. on Info. And Syst., 1995, E78-D(6), 669–675.

[3] 新美康永，音声认识，东京，日本共立出版社，1987, 334–345.

[4] TMS320C54x DSP Reference Set, VOLUME1: CPU and Peripherals, U.S.A.: Texas Instruments, 1994.

[5] TMS320C54x DSP Reference Set, VOLUME2: Mnemonic Instruction, U.S.A.: Texas Instruments, 1994.

[6] TMS320C54x DSP Reference Set, VOLUME3: Algebraic Instruction, U.S.A.: Texas Instruments, 1994.
[7] TMS320C54x DSP Reference Set, VOLUME4: Application Guide, U.S.A.: Texas Instruments, 1994.

THE REALIZATION OF SPEECH RECOGNITION BY DSP

Xiao Shengbing Zhao Li* Liu Haibin* Wu Zhenyang*

(*Dept. of Electronic Eng., Suzhou University, Suzhou* 215006, *China*)

*(*Dept. of Radio Eng., Southeast University, Nanjing* 210096, *China*)

Abstract This paper presents the method for recognizing Chinese speech by utilizing DSP. Through experiment of Chinese speech recognition in the noise environment of outdoor, the effectiveness of the method is confirmed.

Key words Speech recognition, DSP, Noise

肖圣兵: 男, 1964 年生, 副教授, 研究方向为电子信息工程.
赵 力: 男, 1958 年生, 副教授, 研究方向为信号处理.
刘海滨: 男, 1974 年生, 在读博士, 研究方向为语音信号处理, 语音识别等.
吴镇扬: 男, 1949 年生, 教授、博士生导师, 研究方向为信号处理.