

三种语音特征信息综合方法的理论分析*

王成友 梁甸农 唐朝京 杨述明

(国防科技大学电子技术系 长沙 410073)

摘要 本文是作者在1997年《声学学报》中文版和英文版发表的三篇文章基础上的进一步研究。这些文章没有对提出的三种语音特征综合方法进行理论分析,本文运用概率论知识,详细地对前面三篇文章提出的三种语音特征信息综合方法进行理论分析。分析结果表明:三种语音特征信息综合方法能够比较好地综合语音特征信息,提高语音识别系统的性能。这一结果与前面三篇文章中实验结果一致。

关键词 语言特征信息综合, 顶尖参数法, 全部参数法, 最可靠参数法, 语音识别
分类号 TN91

The Theory Study of the Three Kinds of Methods for Combining the Information of Speech Features

Wang ChengYou Liang DianNong Tang Zhaojing Yang shuming

(Department of Electronic Technology, NUDT, Changsha, 410073)

Abstract This paper is a continuation of three papers that were published in 《Chinese Journal of Acoustics》in Chinese version and English version. It discusses three kinds of methods for combining the information of speech features that were presented in detail in the previous three papers in accordance with the theory of probability. The result shows the three methods improve the performance in speech recognition system. That is the same with the experiment result.

Key words combination of speech features information, the maximum parameter method, all parameter method, the most reliable parameter method, speech recognition

语音识别技术是学术界一个研究很热的领域。至今的语音识别研究中,已经有相当多的特征被提出和应用,其中一些特征被发现对语音识别输出有着互补性^[1,2,6],如倒谱和 Δ 倒谱。但至今大多数语音识别系统中对特征信息之间互补性的研究不多,主要有两类,一类是通过增大特征空间的维数来综合多种特征信息^[4];另一类是在HMM/VQ识别系统中利用多种码本进行综合^[3]。这些综合没有利用识别器的分辨性能。文献[1,2,6]利用识别器分辨性能,在识别器输出上提出了三种综合方法,并从实验的角度检验了这三种方法的综合性能,但是这三种方法没有进行理论分析。本文运用概率论知识,详细地对文献[1,2,6]提出的三种语音特征信息综合方法进行理论探讨。

1 三种语音特征信息综合方法

在以HMM为模型的语音识别系统中,设系统模板库大小为 N ,被综合的特征数为 M , $\text{recp}(k, W_l)$ 表示对输入语音使用第 k 个特征和 W_l 模板的HMM识别概率,其中 $1 \leq k \leq M, 1 \leq l \leq N$ 。设输入语音为 S_i ,输出语音为 S_o 。 $S_i, S_o \in \{W_1, W_2, \dots, W_N\}$ 。则三种综合方法可以表示如下。

(1) 顶尖参数法

$$\text{令: } \text{Prma}(W_l) = \max_k \text{recp}(k, W_l) \quad (1 \leq l \leq N)$$

$$S_o = f_{\max}(S_i) = \arg \max_{\{W_1, W_2, \dots, W_N\}} \text{Prma}(W_l)$$

(2) 全部参数法

$$S_o = f_{m\max}(S_i) = \arg \max_{\{W_1, W_2, \dots, W_N\}} \left[\frac{1}{M} \sum_{k=1}^M \text{recp}(k, W_l) \right]$$

* 1999年3月8日收稿

第一作者: 王成友, 男, 1966年生, 博士

(3) 最可靠参数法

$$\begin{aligned} \text{令: } \max_l(k) &= \max_{\{W_1, W_2, \dots, W_N\}} \text{recp}(k, W_l) \quad l = k = M \\ W_j &= \arg \max_{\{W_1, W_2, \dots, W_N\}} \text{recp}(k, W_j) \quad l = j = N, 1 = k = M \\ \max_2(k) &= \max_{\{W_1, W_2, \dots, W_N\}, W_l, W_j} \text{recp}(k, W_l) \quad (l = k = M) \\ n &= \arg \max_k \{ \max_l(k) - \max_2(k) \} \\ S_o &= f_{\max}(S_i) = \max_{\{W_1, W_2, \dots, W_N\}} \text{recp}(n, W_l) \end{aligned}$$

2 两个引理和一个定理

为了对三种综合方法进行理论上分析, 先给出两个引理和一个定理。

引理 1 设两个服从正态分布的随机变量 $X_1, X_2, X_1 \sim N(m_1, \sigma_1^2), X_2 \sim N(m_2, \sigma_2^2), X_1$ 与 X_2 相互独立。那么

$$P(X_1 = X_2) = \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} \exp\left(-\frac{(m_1 - m_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right)$$

证明略。

引理 2 设两个服从正态分布的随机变量 $X_1, X_2, X_1 \sim N(m_1, \sigma_1^2), X_2 \sim N(m_2, \sigma_2^2), X_1$ 与 X_2 相互独立。那么

$$P(X_2 > X_1) = \Phi\left(\frac{m_2 - m_1}{\sigma_1^2 + \sigma_2^2}\right) \quad (1)$$

$$\text{其中 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt, x \in R \quad (2)$$

证明略。

设有一个对 N 个孤立词 $\{W_1, W_2, \dots, W_N\}$ 识别的语音处理系统, 当系统输入语音为 W_i , 经系统处理, 输出分别与 $\{W_1, W_2, \dots, W_N\}$ 的似然值为 $\alpha_1, \alpha_2, \dots, \alpha_N$ 。该系统输出正确的概率(一般称识别率)为 P_{W_i} , 则整个识别系统的总体识别率为

$$P_{rec} = \frac{1}{N} \sum_{i=1}^N P_{W_i} \quad (3)$$

设 W_1, W_2, \dots, W_N 之间相互独立, $\alpha_i \sim N(m_{\alpha_i}, \sigma_{\alpha_i}^2), i = 1, 2, \dots, N$ 。有下面定理成立。

定理 1 如果系统以最大似然值方法识别, 则式(3)中词 W_i 的识别率

$$\begin{aligned} P_{W_i} &= \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_1}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_1}^2}\right) \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_2}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_2}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i-1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i-1}}^2}\right) \\ &\quad \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i+1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i+1}}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_N}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_N}^2}\right) \end{aligned} \quad (4)$$

证明

因为 W_1, W_2, \dots, W_N 之间相互独立, 则 $\alpha_1, \alpha_2, \dots, \alpha_N$ 之间相互独立

所以

$$\begin{aligned} P_{W_i} &= P((\alpha > \alpha_1) \& (\alpha > \alpha_2) \& \dots \& (\alpha > \alpha_{i-1}) \& (\alpha > \alpha_{i+1}) \& \dots \& (\alpha > \alpha_N)) \\ &= P(\alpha > \alpha_1) P(\alpha > \alpha_2) \dots P(\alpha > \alpha_{i-1}) P(\alpha > \alpha_{i+1}) \dots P(\alpha > \alpha_N) \end{aligned}$$

根据引理 1

$$P_{W_i} = \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_1}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_1}^2}\right) \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_2}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_2}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i-1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i-1}}^2}\right) \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i+1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i+1}}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_N}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_N}^2}\right)$$

证毕

定理1与(3)式相结合清楚地显示出孤立词语音识别系统用最大似然值方法识别的识别率与似然值的均值与方差的关系。

3 三种语音特征信息综合方法理论分析

现在考虑两种特征识别的情况, 设当输入语音为 W_i 时, 用第一种特征判断, 系统输出与 W_1, W_2, \dots, W_N 的似然值为 $\alpha_1, \alpha_2, \dots, \alpha_N$, 用第二种特征判断, 系统输出与 W_1, W_2, \dots, W_N 的似然值为 $\beta_1, \beta_2, \dots, \beta_N$ 。两种特征相互独立。 $\alpha_i \sim N(m_{\alpha_i}, \sigma_{\alpha_i}^2)$, $\beta_i \sim N(m_{\beta_i}, \sigma_{\beta_i}^2)$, $i = 1, 2, \dots, N$ 。则三种语音特征信息综合方法可作如下分析:

3.1 顶尖参数法

系统总体识别率为

$$P_{rec} = \frac{1}{N} \sum_{i=1}^N P_{W_i}$$

$$P_{W_i} = P((\alpha_1 \alpha_1) \& (\alpha_1 \alpha_2) \dots \& (\alpha_1 \alpha_{i-1}) \& (\alpha_1 \alpha_{i+1}) \dots \& (\alpha_1 \alpha_N) \& (\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_i) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N)) + P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_i) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N) \& (\beta_1 \beta_1) \& (\beta_1 \beta_2) \dots (\beta_1 \beta_{i-1}) \& (\beta_1 \beta_{i+1}) \dots (\beta_1 \beta_N))$$

$$= P((\alpha_1 \alpha_1) \& (\alpha_1 \alpha_2) \dots \& (\alpha_1 \alpha_{i-1}) \& (\alpha_1 \alpha_{i+1}) \dots \& (\alpha_1 \alpha_N)) P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_i) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N)) + P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_i) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) P((\beta_1 \beta_1) \& (\beta_1 \beta_2) \dots (\beta_1 \beta_{i-1}) \& (\beta_1 \beta_{i+1}) \dots (\beta_1 \beta_N))$$

设输入语音为 W_i 时系统单独用第一种特征识别正确率为 P_{α_i} , 用第二种特征识别正确率为 P_{β_i} 。有

$$P_{\alpha_i} = P((\alpha_1 \alpha_1) \& (\alpha_1 \alpha_2) \dots \& (\alpha_1 \alpha_{i-1}) \& (\alpha_1 \alpha_{i+1}) \dots \& (\alpha_1 \alpha_N))$$

$$P_{\beta_i} = P((\beta_1 \beta_1) \& (\beta_1 \beta_2) \dots (\beta_1 \beta_{i-1}) \& (\beta_1 \beta_{i+1}) \dots (\beta_1 \beta_N))$$

$$p_{W_i} = P_{\alpha_i} P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_i) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N)) + P_{\beta_i} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_i) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N))$$

$$= P_{\alpha_i} P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N)) + P_{\beta_i} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) + P_{\alpha_i} P(\alpha_1 = \beta_1) P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N))$$

根据定理1

$$P_{W_i} = P_{\alpha_i} P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N)) + P_{\beta_i} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) + \frac{1}{2\pi(\sigma_{\alpha_i}^2 + \sigma_{\beta_i}^2)} \exp\left(-\frac{(m_{\alpha_i} - m_{\beta_i})^2}{2(\sigma_{\alpha_i}^2 + \sigma_{\beta_i}^2)}\right) P_{\alpha_i} P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N))$$

令

$$P_{\Delta} = \frac{1}{2\pi(\sigma_{\alpha_i}^2 + \sigma_{\beta_i}^2)} \exp\left(-\frac{(m_{\alpha_i} - m_{\beta_i})^2}{2(\sigma_{\alpha_i}^2 + \sigma_{\beta_i}^2)}\right) P_{\alpha_i} P((\alpha_1 \beta_1) (\alpha_1 \beta_2) \dots (\alpha_1 \beta_{i-1}) \& (\alpha_1 \beta_{i+1}) \dots (\alpha_1 \beta_N))$$

$$= P(\alpha = \beta) P_{\alpha} P_{\beta} P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N)) \quad (5)$$

有

$$\begin{aligned} & \min(P_{\alpha} P_{\beta} P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N)) + P_{\Delta}, \\ & P_{\beta} P_{\alpha} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) + P_{\Delta}) \\ & P_{w_i} \max(P_{\alpha} P_{\beta} P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N)) + P_{\Delta}, \\ & P_{\beta} P_{\alpha} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) + P_{\Delta}) \\ & \max(P_{\alpha} + P_{\Delta}, P_{\beta} + P_{\Delta}) \end{aligned} \quad (6)$$

不妨设 $\sigma_{\alpha_k}^2, \sigma_{\beta_k}^2, k = 1, 2, \dots, N$ 为较小的定值。这样可对顶尖参数法粗略分析如下:

(1) 当两种特征识别性能较好, $m_{\beta_i} - m_{\alpha_i}$ 较小时,

$$\begin{cases} P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N)) \\ P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) \end{cases} \text{较大。从图 1 可见, } P(\alpha = \beta) \text{ 较大。} P_{\Delta} \text{ 较大。}$$

根据(5)式, P_{w_i} 集中在一个较大值附近, 综合性能较好。

(2) 当两种特征性能一个较好一个较差, 不妨设第一种特征(对应 α)性能较差, 第二种特征(对应 β)性能较好。 $m_{\beta_i} - m_{\alpha_i}$ 较大时, $P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N))$ 较小, $P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N))$ 较大。由图 1, $P(\alpha = \beta)$ 较小, P_{Δ} 较小。

式(6)变为:

$$\begin{aligned} & P_{\alpha} P_{\beta} P((\alpha \beta_1)(\alpha \beta_2) \dots (\alpha \beta_{i-1}) \& (\alpha \beta_{i+1}) \dots (\alpha \beta_N)) + P_{\Delta} \\ & P_{w_i} P_{\beta} P_{\alpha} P((\beta_1 \alpha_1) \& (\beta_1 \alpha_2) \dots (\beta_1 \alpha_{i-1}) \& (\beta_1 \alpha_{i+1}) \dots (\beta_1 \alpha_N)) + P_{\Delta} \end{aligned}$$

P_{w_i} 分布范围较大, 综合性能有着较大的随机性。

(3) 其它情况, 直接从(6)式观察, 综合性能随机性较强。

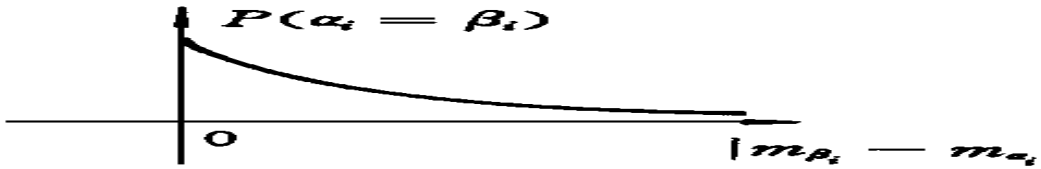


图 1 $P(\alpha = \beta)$ 与 $m_{\beta_i} - m_{\alpha_i}$ 的变化关系图

Fig. 1 The relation of $P(\alpha = \beta)$ and $m_{\beta} - m_{\alpha}$

从上面分析可知, 顶尖参数综合法不能算是一种较好的综合方法。但它在两种特征性能均较好, m_{α_i} 与 m_{β_i} 较接近时, 会表现出一种较强的综合性能。

3.2 全部参数法

设识别系统输入语音为 W_i 时, 全部参数法综合输出与 W_1, W_2, \dots, W_N 似然值为 r_1, r_2, \dots, r_N 。则:

$$r_k = \alpha^k + \beta^k, k = 1, 2, \dots, N。$$

因为 α, β_k 的概率分布是正态分布, r_k 也是正态分布且有:

$$\begin{aligned} m_{r_k} &= E[r_k] = E[\alpha^k + \beta^k] = E[\alpha^k] + E[\beta^k] = m_{\alpha_k} + m_{\beta_k} \\ \sigma_{r_k}^2 &= D[r_k] = D[\alpha^k + \beta^k] = D[\alpha^k] + D[\beta^k] = \sigma_{\alpha_k}^2 + \sigma_{\beta_k}^2 \end{aligned}$$

即

$$r_k \sim N(m_{r_k}, \sigma_{r_k}^2) = N(m_{\alpha_k} + m_{\beta_k}, \sigma_{\alpha_k}^2 + \sigma_{\beta_k}^2)$$

根据定理 1, 输入 W_i 用全部综合法、 α 特征、 β 特征分别识别的正确率

$$P_{r_k} = \Phi\left(\frac{m_{r_i} - m_{r_1}}{\sigma_{r_i}^2 + \sigma_{r_1}^2}\right) \Phi\left(\frac{m_{r_i} - m_{r_2}}{\sigma_{r_i}^2 + \sigma_{r_2}^2}\right) \dots \Phi\left(\frac{m_{r_i} - m_{r_{i-1}}}{\sigma_{r_i}^2 + \sigma_{r_{i-1}}^2}\right)$$

$$\begin{aligned}
 & \Phi\left(\frac{m_{r_i} - m_{r_{i+1}}}{\sigma_{r_i}^2 + \sigma_{r_{i+1}}^2}\right) \dots \Phi\left(\frac{m_{r_i} - m_{r_N}}{\sigma_{r_i}^2 + \sigma_{r_N}^2}\right) \\
 P_{\alpha_i} = & \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_1}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_1}^2}\right) \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_2}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_2}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i-1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i-1}}^2}\right) \\
 & \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_{i+1}}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_{i+1}}^2}\right) \dots \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_N}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_N}^2}\right) \\
 P_{\beta_i} = & \Phi\left(\frac{m_{\beta_i} - m_{\beta_1}}{\sigma_{\beta_i}^2 + \sigma_{\beta_1}^2}\right) \Phi\left(\frac{m_{\beta_i} - m_{\beta_2}}{\sigma_{\beta_i}^2 + \sigma_{\beta_2}^2}\right) \dots \Phi\left(\frac{m_{\beta_i} - m_{\beta_{i-1}}}{\sigma_{\beta_i}^2 + \sigma_{\beta_{i-1}}^2}\right) \\
 & \Phi\left(\frac{m_{\beta_i} - m_{\beta_{i+1}}}{\sigma_{\beta_i}^2 + \sigma_{\beta_{i+1}}^2}\right) \dots \Phi\left(\frac{m_{\beta_i} - m_{\beta_N}}{\sigma_{\beta_i}^2 + \sigma_{\beta_N}^2}\right)
 \end{aligned}$$

函数 $\Phi(x)$ 是一个单调增加函数, 它的图形如图 2。

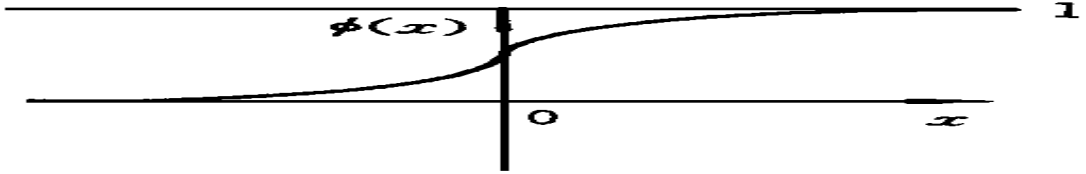


图 2 $\Phi(x)$ 的函数图

Fig. 2 The function of $\Phi(x)$

来看下面式子:

$$\begin{aligned}
 & \frac{m_{r_i} - m_{r_k}}{\sigma_{r_i}^2 + \sigma_{r_k}^2} - \frac{m_{\alpha_i} - m_{\alpha_k}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}, K = 1, 2, \dots, i-1, i+1, \dots, N \\
 = & \frac{(m_{\alpha_i} + m_{\beta_i}) - (m_{\alpha_k} + m_{\beta_k})}{\sigma_{\alpha_i}^2 + \sigma_{\beta_i}^2 + \sigma_{\alpha_k}^2 + \sigma_{\beta_k}^2} - \frac{m_{\alpha_i} - m_{\alpha_k}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}
 \end{aligned}$$

不妨设

$$\sigma_{\alpha_k}^2 = \sigma_{\beta_k}^2, k = 1, 2, \dots, N$$

上式变为

$$\begin{aligned}
 & \frac{(m_{\alpha_i} + m_{\beta_i}) - (m_{\alpha_k} + m_{\beta_k}) - \sqrt{2}(m_{\alpha_i} - m_{\alpha_k})}{\sqrt{2} \sqrt{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}} \\
 & \frac{(m_{\beta_i} - m_{\beta_k}) - 0.4(m_{\alpha_i} - m_{\alpha_k})}{\sqrt{2} \sqrt{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}}
 \end{aligned}$$

同理

$$\begin{aligned}
 & \frac{m_{r_i} - m_{r_k}}{\sigma_{r_i}^2 + \sigma_{r_k}^2} - \frac{m_{\beta_i} - m_{\beta_k}}{\sigma_{\beta_i}^2 + \sigma_{\beta_k}^2} \\
 & \frac{(m_{\alpha_i} - m_{\alpha_k}) - 0.4(m_{\beta_i} - m_{\beta_k})}{\sqrt{2} \sqrt{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}}
 \end{aligned}$$

不妨设

$$P_{\alpha_i} \quad P_{\beta_i}$$

因而只要

$$(m_{\beta_i} - m_{\beta_k}) > 0.4(m_{\alpha_i} - m_{\alpha_k})$$

即使用较差特征系统输出似然值均值之差大于使用较好特征输出似然值均值差的 0.4 倍, 就有

$$\frac{m_{r_i} - m_{r_k}}{\sigma_{r_i}^2 + \sigma_{r_k}^2} > \frac{m_{\alpha_i} - m_{\alpha_k}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2},$$

$$\Phi\left(\frac{m_{r_i} - m_{r_k}}{\sigma_{r_i}^2 + \sigma_{r_k}^2}\right) > \Phi\left(\frac{m_{\alpha_i} - m_{\alpha_k}}{\sigma_{\alpha_i}^2 + \sigma_{\alpha_k}^2}\right),$$

$$k = 1, 2, \dots, i-1, i+1, \dots, N$$

$$P_{r_i} > P_{\alpha_i}, P_{r_i} > \beta_i$$

即 W_i 的识别率将会增大。

对于大多数常被使用性能比较好的语音特征之间这一点容易满足。根据式(3)系统识别正确率为每个词识别正确率之和求平均。因此全部参数法综合性能能够提高识别率而且比较稳定。

3.3 最可靠参数法

令

$$P_{\alpha_i} = P((\alpha_1 \quad \alpha_i) \& (\alpha_1 \quad \alpha_2) \dots \& (\alpha_1 \quad \alpha_{i-1}) \& (\alpha_1 \quad \alpha_{i+1}) \dots \& (\alpha_1 \quad \alpha_{i-1})$$

$$\& (\alpha_1 \quad \alpha_{i+1}) \dots \& (\alpha_1 \quad \alpha_N))$$

$$P_{\beta_i} = P((\beta_{r_1} \quad \beta_i) \& (\beta_1 \quad \beta_2) \dots \& (\beta_{r_1} \quad \beta_{i-1}) \& (\beta_{r_1} \quad \beta_{i+1}) \dots \& (\beta_{r_1} \quad \beta_{i-1})$$

$$\& (\beta_1 \quad \beta_{i+1}) \dots \& (\beta_{r_1} \quad \beta_N))$$

当识别系统输入语音 W 时, 最可靠参数法综合识别正确率为

$$P_{\alpha_i} = \sum_{i_1=0}^N \sum_{j_1=0}^N \sum_{j=0}^N P(\alpha_i - \alpha_{i_1} > \beta_i - \beta_{j_1}) P_{\beta_j} P_{\alpha_{i_1}} P_{\beta_{j_1}}$$

$$+ P_{\beta_i} \sum_{i_1=0}^N \sum_{j_1=0}^N \sum_{j=0}^N P(\beta_i - \beta_{j_1} > \alpha_i - \alpha_{i_1}) P_{\alpha_j} P_{\beta_{i_1}} P_{\alpha_{i_1}}$$

$$+ P_{\alpha_i} P_{\beta_i} \sum_{i_1=0}^N \sum_{j_1=0}^N \sum_{i=0}^N P(\alpha_i - \alpha_{i_1} = \beta_i - \beta_{j_1}) P_{\alpha_{i_1}} P_{\beta_{j_1}} \quad (7)$$

可以推导出

$$(7) \text{ 式 } P_{\alpha_i} + P_{\beta_i} - P_{\alpha_i} P_{\beta_i}$$

$$P_{\alpha_i} \quad P_{\beta_i}$$

不妨设

所以

$$(P_{\alpha_i} + P_{\beta_i} - P_{\alpha_i} P_{\beta_i}) - P_{\alpha_i} = P_{\beta_i} (1 - P_{\alpha_i})$$

这一结果显示, 最可靠参数法提高了每个单词识别正确率, 识别系统总体识别性能得到改善。

4 实验结果

基于 HMM 的语音识别系统, 取量化位数 12 位, 采样率为 16kHz, 预加重传输函数为 $1 - 0.94z^{-1}$, 分析帧长为 10ms, 帧移为 5ms, 线性预测阶数为 16, 让 $\{W_1, W_2, \dots, W_N\} = \{0, 1, \dots, 9\}$ 。采集的语音是由四个男同学和两个女同学发的音, 每个数字发音 10 遍, 共 600 个音。运用其中每人每音 5 遍共 300 个音进行训练, 建立线性预测系数(LPC)、反射系数(GRC)、LPC 倒谱(CEP)和 LPC 谱(LPCC)^[1]四种特征的码本和模板库, 再用 300 个音进行识别。

四种特征分别和综合的识别结果如表 1 和表 2。

表 1 多发音人汉语数字 DHMM/VQ 系统单特征识别率

Table 1 The recognition rate in multiple speakers Chinese digit DHMM/VQ system using a single feature

特征 \ 数码	0	1	2	3	4	5	6	7	8	9	平均
GRC	73.3%	80%	96.7%	100%	86.7%	93.3%	93.3%	96.7%	70%	100%	89%
LPC	80%	100%	70%	83.3%	73.3%	83.3%	86.7%	100%	56.7%	86.7%	82%
CEP	93.3%	90%	73.3%	83.3%	80%	100%	93.3%	100%	93.3%	83.3%	89%
LPCC	53.3%	90%	73.3%	43.3%	63.3%	80%	66.7%	60%	40%	53.3%	62.3%

表2 特征LPC、LPC、LPC、CEP三种方法综合的多发音人汉语数字DHMM/VQ系统识别率

Table 2 The recognition rate in multiple speakers Chinese digit DHMM/VQ system using the combination of Feature GRC, LPC, CEP and LPCC

方法 \ 数码	0	1	2	3	4	5	6	7	8	9	平均
顶尖参数法	96.7%	100%	86.7%	86.7%	90%	100%	93.3%	100%	83.3%	86.7%	92.3%
全部参数法	96.7%	100%	96.7%	96.7%	90%	96.7%	100%	100%	93.3%	86.7%	95.7%
最可靠参数法	96.7%	90%	86.7%	86.7%	86.7%	100%	96.7%	100%	100%	93.3%	94.3%

5 结束语

本文对文献[1, 2, 6]提出的三种语音特征信息综合方法进行了理论分析, 结果表明顶尖参数法综合性能有一定的随机性, 全部参数法和最可靠参数法可以比较好地改善识别系统的识别性能。但是推导过程假设特征之间相互独立, 实际许多语音特征相互不独立; 词汇与参考模式输出似然值的统计分布用正态分布模拟还可以用其它方法模拟更确切。

参考文献

- 1 Wang Chengyou, Tang Shuqi, Liang Diannong. The Methods for Combing the Information of Various Kinds of Features in Speech Recognition. Chinese Journal of Acoustics, 1997, 16(2): 115 ~ 120
- 2 王成友, 汤叔祺, 梁甸农等. 语音识别中多种特征信息综合利用的方法. 声学学报, 1997, 22(2): 111 ~ 115
- 3 Lee K F, Hon H W and Raj R. An Overveiw of the SPHINX Speech Recognition System. IEEE Tran. on ASSP, 1990, 38(1): 35 ~ 45
- 4 Rabiner L R, Wilpon J G and Soong F K. High Performance Connected Digit Recognition, Using Hidden Markov Models. ICASSP-88: 119 ~ 123
- 5 第四届全国人机语音通讯学术会议大会报告, 1996年10月, 北京
- 6 王成友, 汤叔祺, 梁甸农. 噪声对特征综合法语音识别性能的影响. 声学学报, 1997, 22(3): 282 ~ 285
- 7 王成友. 语音特征信息综合方法及汉语连接数字语音识别系统的研究. 国防科技大学博士学位论文, 1997