文章编号：2095-4980(2013)01-0136-06

一种融合 IB 准则特征的说话人分段聚类方法

张　力，张连海，许友亮

(信息工程大学 信息工程学院，河南 郑州 450002)

摘　要：针对说话人分段与聚类算法中先验知识不足的问题，利用基于信息瓶颈(IB)准则和基于隐马尔科夫模型(HMM)/高斯混合模型(GMM)方法间的互补性，提出了一种基于特征层融合的说话人分段与聚类算法。该算法将基于 IB 准则算法的输出结果进行对数变换和降维处理；然后利用变换后的特征与传统梅尔频率倒谱系数(MFCC)特征分别训练说话人 GMM 模型，并在得分域对说话人类别的得分进行加权融合；根据融合的得分，进行基于 HMM/GMM 模型的说话人分段与聚类。实验表明，融合后的特征可以为系统提供更多的先验信息，比传统方法的误配率降低了 1.2%。

关键词：信息瓶颈准则；说话人分段聚类；HMM/GMM 模型；系统融合

中图分类号：TN912.3　　**文献标识码：**A

A speaker segmentation and clustering method combined with IB features

ZHANG Li，ZHANG Lian-hai，XU You-liang

(School of Information Engineering，Information Engineering University，Zhengzhou Henan 450002，China)

Abstract：The performance of the speaker segmentation and clustering system usually degrades because of lacking prior information about the speakers. To solve the problem, a novel approach that combines the algorithms based on Information Bottleneck(IB) principle and Hidden Markov Model(HMM)/Gaussian Mixture Model(GMM) is proposed by using the complementarity of these two algorithms. After logarithmic transform and Principal Component Analysis to reduce dimension, the output of the IB algorithm is then used to train the speaker GMM model. Along with the speaker GMM model trained by the traditional Mel Frequency Cepstral Coefficient(MFCC) feature, the scores between different speaker clusters are computed respectively and then combined using linear weighted sum method. Lastly, the HMM/GMM based speaker segmentation and clustering is performed with the combined score. Experiments show that the IB features provide more prior information for the system and the speaker match error rate is reduced by 1.2% compared to that in traditional method.

Key words：Information Bottleneck principle；speaker segmentation and clustering；Hidden Markov Model/Gaussian Mixture Model；system combination

说话人分段聚类主要解决一段语音中“谁在什么时候说话”的问题，即确定语音段中说话人的个数并将语音段与说话人一一对应。说话人分段聚类应用广泛：可用于大词汇量连续语音识别中的说话人自适应和作为基于说话人的索引和检索的前端处理等。近年来，系统融合方法逐渐成为提高说话人分段聚类的热门方法之一。传统的融合方法大都采用“级联”的方式，即将 2 种不同的算法或模块串行使用，将前一步骤的分段结果作为后一步骤的初始化结果，然后更新说话人数量以及说话人的边界[1]。一些学者尝试在说话人分段的输出结果上进行融合，2005 年，Tranter S E 等[2]提出利用投票机制将 2 个系统的输出进行融合，性能比原系统分别提高了 1.64%和 2.56%。2006 年，Mergnier S 等[3]将不同系统产生的说话人标签在帧级进行融合，然后再进行重分段去除多余说话人并更新说话人变换点。2010 年，Bozonnet 等人[4]将自下而上聚类系统和自上而下聚类系统进行融合，较原系统有较大改进。目前说话人分段聚类采用的最主流的方法是基于 HMM/GMM 模型的方法，该方法在历年的美国国家标准与技术研究所(National Institute of Standards and Technology，NIST)评测中都取得了较好性能。2009 年，Deepu

收稿日期：2012-04-07；**修回日期：**2012-05-31
基金项目：国家自然科学基金资助项目(61175017)

Vijayasenan[5]提出了一种基于 IB 准则的说话人分段聚类方法，该方法在性能方面与基于 HMM/GMM 模型的方法相当，但在运算速度上有大幅提升。基于 HMM/GMM 模型的方法与基于 IB 准则的方法都是自下而上的说话人聚类方法，但由于出发点和算法不同，结果存在着一定的互补性，本文将基于 IB 准则的说话人分段聚类的输出作为基于 HMM/GMM 方法的输入特征，与传统的 MFCC 特征相结合，以提高说话人分段聚类系统的性能。

1　基于 HMM/GMM 模型的说话人分段聚类系统

基于 HMM/GMM 模型的说话人分段聚类方法采用自下而上的聚类结构，HMM/GMM 模型的结构如图 1 所示。对 HMM 模型进行初始化时，模型的状态数等于初始说话人类别的个数，每一个状态代表一个说话人类别，其概率分布函数可由该状态对应的 GMM 模型表示。模型中的各状态可认为是由一系列的子状态组成，同一个状态的各子状态共用同一个 GMM 模型。通过设置这些子状态链的长短，可调整各状态的最短驻留时间，一般情况下最短驻留时间设为 2.5 s。

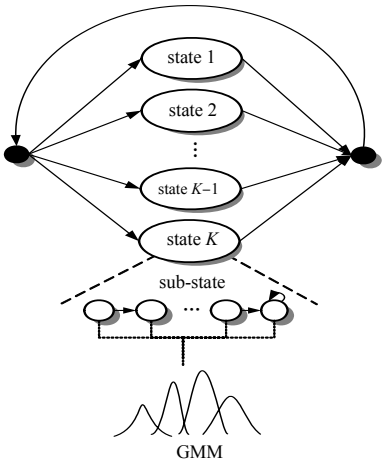

Fig.1 Structure of HMM/GMM model
图 1 HMM/GMM 模型结构示意图

聚类初始化时，初始聚类的个数即初始说话人个数 K 一般大于实际说话人数量，将语音数据切分为 K 个等长的数据段，并利用最大期望(Expectation Maximization，EM)算法相应的 GMM 模型。由于初始说话人个数大于实际说话人数，存在一个说话人的语音被分到多个说话人类别中的情况，所以，需要根据优化的贝叶斯信息准则(Bayesian Information Criterion，BIC)[6]对这些初始说话人类别进行聚类，将相同说话人的语音聚到同一个说话人类别中。聚类的过程如图 2 所示，计算各说话人类别之间的 ΔBIC 值，若存在 $\Delta BIC \geqslant 0$，则将其中最大值对应的 2 个说话人类别进行合并，得到新的说话人类别。ΔBIC 的计算式为：

$$\Delta BIC = BIC(C_{i\cup j}) - \left(BIC(C_i) + BIC(C_j)\right) = \log L(\boldsymbol{X}_{i\cup j}, C_{i\cup j}) - \left(\log L(\boldsymbol{X}_i, C_i) + \log L(\boldsymbol{X}_j, C_j)\right) \tag{1}$$

式中：C_i 和 C_j 表示说话人类别 i 和 j；$C_{i\cup j}$ 为合并后的说话人类别；$\boldsymbol{X}_i$ 和 $\boldsymbol{X}_j$ 表示说话人类别 i 和 j 中的观测向量；$\boldsymbol{X}_{i\cup j}$ 为合并后的说话人类别 $C_{i\cup j}$ 的观测向量。当所有的 ΔBIC 的值都小于零时，聚类停止。

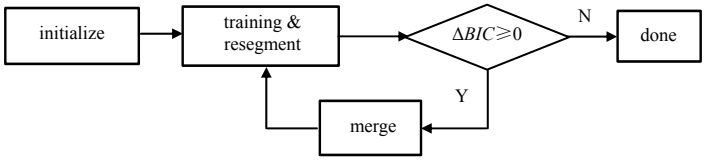

Fig.2 Flow of speaker segmentation and clustering based on HMM/GMM
图 2 基于 HMM/GMM 模型方法聚类流程图

2　基于 IB 准则的说话人聚类系统

2.1 IB 准则

IB 准则由 Tishby[7]提出后，在图像和文本聚类中应用广泛。IB 准则的主要思想是：给定数据集合 X，聚类集合 C 是 X 的压缩表达形式，引入相关变量 Y，在将 X 转化为 C 的过程中，X 中关于相关变量 Y 的信息将传递给 C。IB 准则既要求对数据 X 进行最大化的压缩，即 C 用最少的编码长度表示，同时又要使 C 中保留尽可能多的关于相关变量 Y 的信息，即最大化式(2)[5]：

$$\mathcal{F} = I(Y,C) - \frac{1}{\beta} I(C,X) \tag{2}$$

式中 β 为拉格朗日乘子，用于调整 C 中保留关于 Y 的信息量 $I(Y,C)$ 和 C 中包含关于原始数据 X 的信息量 $I(C,X)$ 之间的比例。压缩后的聚类集合 C 中包含的原始数据 X 的信息可用它们之间的互信息表示：

$$I(C,X) = \sum_{x\in X, c\in C} p(x)p(c|x)\log\frac{p(c|x)}{p(c)} \tag{3}$$

同理，压缩后的聚类 C 中保留的关于相对变量 Y 之间的信息可表示为：

$$I(Y,C) = \sum_{y\in Y, c\in C} p(c)p(y|c)\log\frac{p(y|c)}{p(y)} \tag{4}$$

式中：

$$p(c|x) = \frac{p(c)}{Z(\beta,x)}\exp\left(-\beta D_{KL}\left[p(y|x)\,\|\,p(y|c)\right]\right) \tag{5}$$

$$p(y|c)=\sum_x p(y|x)p(c|x)\frac{p(x)}{p(c)} \tag{6}$$

$$p(c)=\sum_x p(c|x)p(x) \tag{7}$$

式(5)~式(7)中：$Z(\beta,x)$ 为归一化因子；$D_{KL}[p(y|x)\|p(y|c)]$ 表示 $p(y|x)$ 和 $p(y|c)$ 之间的 Kullback-Leibler 距离；$p(x)$, $p(c)$ 分别表示数据 x 和聚类 c 的概率；$p(y|x)$ 表示已知数据 x 时相对变量 y 的概率；$p(y|c)$ 表示已知聚类 c 时相对变量 y 的概率；$p(c|x)$ 表示已知数据 x 其属于聚类 c 的概率。

2.2 基于 IB 准则的说话人分段聚类

在说话人分段聚类系统中，若 $S=\{s_1,s_2,\cdots,s_T\}$ 为语音数据的特征序列，将 S 切分为 K 个等长的语音数据段 $x_i=\{s_{(i-1)\times L+1},s_{(i-1)\times L+2},\cdots,s_{i\times L}\}$，$L$ 为等长数据段中包含的特征帧数。X 为等长语音数据段 $\{x_1,x_2,\cdots,x_K\}$ 的集合，C 为说话人聚类的集合，相关变量 Y 是由全部检测语音数据训练而成 GMM 模型混元的集合。利用 IB 准则进行聚类时，采用硬判决的方式判断语音数据段属于哪个说话人类别。聚类初始化阶段，X 被划分为 K 个初始类，即每一个等长的连续语音特征序列都认为是一个说话人类别，这些类在聚类合并的过程中，为了使得目标方程(2)最大，必须使得每次合并聚类导致$\mathcal{F}$的减少量最小。

在聚类过程中，若合并类 c_i 和 c_j，数据进一步被压缩，会造成目标方程(2)的值减少，减少量 $\Delta\mathcal{F}$ 可表示为：

$$\Delta\mathcal{F}=(p(c_i)+p(c_j))\bar{d}_{ij} \tag{8}$$

式中

$$\bar{d}_{ij}=JS\left[p(y|c_i),p(y|c_j)\right]-\frac{1}{\beta}JS\left[p(x|c_i),p(x|c_j)\right] \tag{9}$$

JS 表示 2 个分布之间的 Jensen-Shannon 距离，可以表示为：

$$JS\left[p(y|c_i),p(y|c_j)\right]=\pi_i D_{KL}\left[p(y|c_i)\|q_Y(y)\right]+\pi_j D_{KL}\left[p(y|c_j)\|q_Y(y)\right] \tag{10}$$

式中：

$$q_Y(y)=\pi_i p(y|c_i)+\pi_j p(y|c_j) \tag{11}$$

$$\begin{cases}\pi_i=p(c_i)/\left(p(c_i)+p(c_j)\right)\\ \pi_j=p(c_j)/\left(p(c_i)+p(c_j)\right)\end{cases} \tag{12}$$

当类 c_i 和 c_j 进行合并后，新的聚类 c_r 的一些特性可以由式(13)和式(14)给出，

$$p(c_r)=p(c_i)+p(c_j) \tag{13}$$

$$p(y|c_r)=\frac{p(y|c_i)p(c_i)+p(y|c_j)p(c_j)}{p(c_i)+p(c_j)} \tag{14}$$

在进行说话人聚类的过程中，每次迭代时，合并使 $\Delta\mathcal{F}$ 最小的 2 个类 c_i 和 c_j，直到保留的说话人类的个数达到规定的说话人个数。由于在说话人分段与聚类任务中，实际说话人的个数往往是未知的，可以利用归一化互信息(Normalized Mutual Information，NMI)判断实际说话人的个数，即设定归一化互信息的门限值，当归一化互信息小于门限值时，则停止聚类。

3 基于 IB 特征的融合系统

基于 IB 准则的说话人分段聚类方法采用的是一种硬判决方式，所有的语音帧只对应唯一的说话人。该方法只根据全部语音数据建立一个背景模型，并不对所有的说话人类别分别建立 GMM 模型，在保证说话人类别保留最多的关于背景模型混元信息的前提下，选择 JS 距离最小的 2 个说话人类别进行合并。系统最终输出的信息包括语音帧隶属于哪个说话人类别的信息以及说话人类别关于背景模型混元的信息。而基于 HMM/GMM 模型的方法，需要对每个说话人类别建立 GMM 模型，并利用优化的 BIC 准则对说话人类别进行聚类。可以看出，2 种聚类方法在说话人类别的建模、类间的距离度量以及输出结果所包含的信息上存在很大的差异，所以这 2 种方法存在着一定的互补性。

在基于 IB 准则的说话人分段与聚类方法中，给定语音数据的特征序列 $S=\{s_1,s_2,\cdots,s_T\}$，对于其中 t 时刻的特征向量 $\boldsymbol{s}_t\in x_j$，x_j 为第 j 个等长数据段，对应唯一的说话人类别 c_i，即 $p(c_i|x_j)=1$。对于任一个等长数据段 x_j，

其对于相关变量 Y 的概率 $p(Y|x_j)$ 是可求的。那么，每一个说话人类别 c_i 关于相关变量 Y 的概率 $p(Y|c_i)$ 可通过对 $\{p(Y|x_j)|x_j\in c_i\}$ 求均值得到。本文定义 $\alpha_t=\{c_i|\boldsymbol{s}_t\in x_j, x_j\in c_i\}$，将数据帧和其最终对应的说话人类别相关联，对于任意 α_t， $p(Y|\alpha_t)$ 都可以求得。

基于 IB 准则的说话人分段聚类方法最终输出的信息可以通过矩阵 $\boldsymbol{A}$ 来表示：

$$\boldsymbol{A}=[p(Y|\alpha_1),p(Y|\alpha_2),\cdots,p(Y|\alpha_T)],\ t=1,2,\cdots,T \tag{15}$$

$\boldsymbol{A}$ 是一个$|Y|\times T$ 的矩阵，其中，$|Y|$为由测试语音训练而成的背景模型的混元数，T 为语音数据的总帧数。可以看出，基于 IB 准则方法的输出矩阵 $\boldsymbol{A}$ 中包含有多重信息量：首先，从矩阵 $\boldsymbol{A}$ 中可以得出 2 个不同的特征向量 $\boldsymbol{s}_{t1}$ 和 $\boldsymbol{s}_{t2}$ 是否隶属于同一个说话人类别(若相同，则 $\alpha_{t1}=\alpha_{t2}$，且 $p(Y|\alpha_{t1})=p(Y|\alpha_{t2})$)；其次，可以得出各个说话人类别对于相关变量的概率分布 $p(Y|\alpha_t)$(不同的说话人类别的 $p(Y|\alpha_t)$ 也不相同)。矩阵 $\boldsymbol{A}$ 中不仅包含了说话人个数及各帧隶属于哪个说话人的信息，还通过不同的关于相关变量的概率分布，体现出各个说话人类别之间的差异。

矩阵 $\boldsymbol{A}$ 中，由于$|Y|$的数值往往较大，且 $p(Y|\alpha_t)$之间的数值差异较小，存在着较多的冗余信息。为了提高运算速度和检测性能，首先对矩阵中的各分量取对数，增加各数据间的差异性，然后利用主成分分析(Principal Component Analysis，PCA)对矩阵进行降维处理，去除冗余信息，减少运算量。

降维后得到的矩阵 $\boldsymbol{A}'$ 即特征矩阵，本文采用文献[8]中融合特征的方法，利用输入的 MFCC 和基于 IB 准则的特征流，分别建立对应的说话人 GMM 模型。对于输入特征向量 $\boldsymbol{s}_t$，其对应于说话人类别 c_i 的概率 $b_{c_i}(\boldsymbol{s}_t)$ 可由式(16)给出：

$$\log b_{c_i}(\boldsymbol{s}_t)=\log\sum_r \omega_{c_i}^r N\ (s_t,\mu_{c_i}^r,\Sigma_{c_i}^r) \tag{16}$$

式中 $\omega_{c_i}^r,\mu_{c_i}^r,\Sigma_{c_i}^r$ 分别为说话人类别 c_i 对应的 GMM 模型的权重、均值和方差。不同特征流产生不同概率后，对这些概率进行线性加权，以获得融合后的概率分布。线性加权由式(17)给出：

$$b_{c_i}(\boldsymbol{s}_t)=P_{\text{mfcc}}\log b_{c_i}^{\text{mfcc}}(s_t^{\text{mfcc}})+P_{\text{IB}}\log b_{c_i}^{\text{IB}}(s_t^{\text{IB}}) \tag{17}$$

式中：s_t^{mfcc} 和 s_t^{IB} 分别为 MFCC 和基于 IB 准则的特征序列；P_{mfcc} 和 P_{IB} 为它们的权重，且有 $P_{\text{mfcc}}+P_{\text{IB}}=1$。在实验部分，将详细介绍如何选定权重值。此时，式(1)可以写为：

$$\begin{aligned}\Delta BIC=P_{\text{mfcc}}\Delta BIC_{\text{mfcc}}+P_{\text{IB}}\Delta BIC_{\text{IB}}=P_{\text{mfcc}}\left[\sum_{s_t^{\text{mfcc}}\in c_{i\cup j}}\log b_{c_{i\cup j}}^{\text{mfcc}}(s_t^{\text{mfcc}})-\sum_{s_t^{\text{mfcc}}\in c_i}\log b_{c_i}^{\text{mfcc}}(s_t^{\text{mfcc}})-\sum_{s_t^{\text{mfcc}}\in c_j}\log b_{c_j}^{\text{mfcc}}(s_t^{\text{mfcc}})\right]+\\ P_{\text{IB}}\left[\sum_{s_t^{\text{IB}}\in c_{i\cup j}}\log b_{c_{i\cup j}}^{\text{IB}}(s_t^{\text{IB}})-\sum_{s_t^{\text{IB}}\in c_i}\log b_{c_i}^{\text{IB}}(s_t^{\text{IB}})-\sum_{s_t^{\text{IB}}\in c_j}\log b_{c_j}^{\text{IB}}(s_t^{\text{IB}})\right]\end{aligned} \tag{18}$$

特征融合的流程图如图 3 所示，当得到融合的 ΔBIC 值后，按照基于 HMM/GMM 模型的方法进行说话人聚类合并，并进行维特比重分段，重新划分说话人类别的边界，然后再根据新的说话人类别按照不同的特征流训练新的 GMM 模型，直到 ΔBIC 值全部小于零，停止聚类。

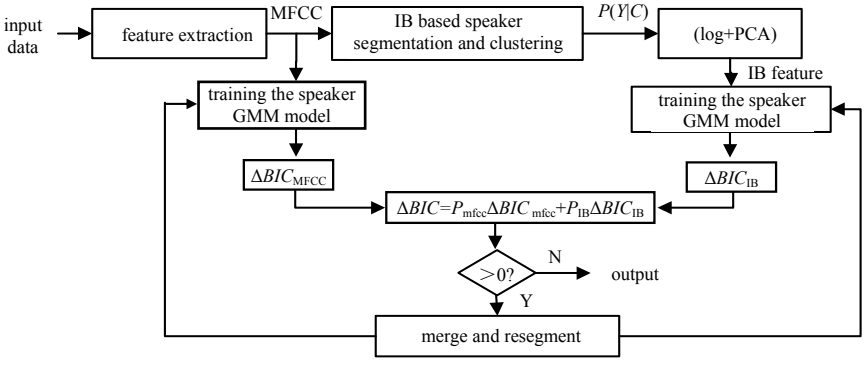

Fig.3 Feature fusion process
图 3 特征融合流程图

4　实验

4.1 实验设置

为测试算法的有效性，本文随机选取了 24 段美国之音(Voice of America，VOA)网络新闻广播数据并进行人工标注，每段音频数据长度约为 5 min。取其中的 12 段音频数据作为实验的开发集，另外 12 段作为实验的测试集。本实验采用的基线系统为基于 HMM/GMM 模型的说话人分类聚类系统，该系统提取的特征为在窗长为 32 ms，帧移为 10 ms 下提取的 19 维的 MFCC 参数。在评价指标方面，为了方便比较算法的性能，本文忽略由语音/非语音检测[9]产生的虚警率和误警率，只采用说话人误配率作为测试的评价指标。说话人误配率 *Spkr* 的计算公式为：

$$Spkr = \frac{T_{\text{err}}}{T_{\text{tot}}} \times 100\% \tag{19}$$

式中：T_{err} 为正确的说话人被判定为其他说话人的语音数据的时间长度；T_{tot} 为语音数据的时间总长度。

利用 MFCC 特征和基于 IB 准则的特征训练的 GMM 模型的混元数均为 5 个，在计算基于 IB 准则的特征过程中，按照文献[10]中的方法对参数进行设置，拉格朗日乘子 β 取 0.1，NMI 门限值设为 0.3。

4.2 参数 P_{IB} 的调试

利用开发集数据，取 PCA 的主分量个数 M 为 5，在不同的 P_{IB} 值上对系统的说话人误配率进行比较，选择说话人误配率最低值对应的 P_{IB} 值作为该参数最终值。实验结果如图 4 所示。

由图 4 可知，当 P_{IB} 取 0 时，即只采用 MFCC 特征作输入特征进行说话人分段与聚类，说话人误配率为 13.9%；当 P_{IB} 取 1 时，即只采用基于 IB 准则的特征输入进行说话人分段与聚类，说话人误配率为 28.3%。同时可以看出，当 P_{IB} 为 0.1 时，说话人误配率得到最小值 12.5%，比只采用 MFCC 特征的方法在说话人误配率上降低 1.4%。随着 P_{IB} 值的不断增加，说话人误配率也随之增加，这是因为基于 IB 准则的特征本身维数较低，信息量有限，只能够作为 MFCC 特征的一个补充，提供关于各说话人的一些先验信息。要进行准确的说话人分段聚类，主要依靠维数较高的 MFCC 特征参数包含的关于说话人的声学上的信息。所以，随着 P_{IB} 增加，MFCC 特征在特征中所占的比重下降，从而导致检测性能的降低。

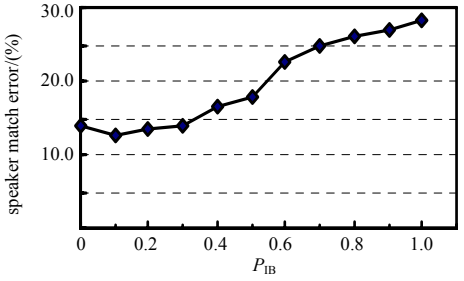

Fig.4 Experiment results with different P_{IB}
图 4 不同 P_{IB} 值下的实验结果

4.3 PCA 中主成分个数 M 的调试

利用开发集数据，按照调试 P_{IB} 参数的方法，将 P_{IB} 的值固定为 0.1。在 PCA 中，前 k 个主成分有多大的综合能力，一般用这 k 个主成分的方差和在全部方差中所占比重来表示，这个比重称为累积贡献率。通过计算，当 k=2 时，累积贡献率的值已经超过 80%，能够包含原变量中绝大部分的信息量。本实验中 PCA 的主分量的个数 M 从 1 到 5 之间取值，通过比较不同主分量个数下系统的说话人误配率，选取说话人误配率最低值对应的 M 作为系统最终的取值。图 5 给出了实验结果，从图中可以看出，当 M=2 时，系统的说话人误配率取到最低值 11.4%。当 M=1 时，对特征矩阵进行 PCA 处理后，只存在一个主分量，不能够完全包含原特征矩阵中所含的有效信息，导致说话人误配率较高。而当 $M \geqslant 3$ 时，特征矩阵 $\boldsymbol{A}'$ 可能包含一些冗余信息，导致系统的性能下降。所以本文选取 M=2 作为 PCA 处理后的主分量个数。

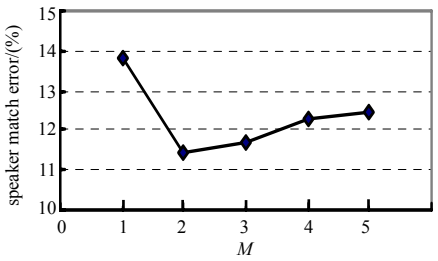

Fig.5 Experiment results with different M
图 5 不同 M 值下的实验结果

4.4 对比测试结果

表 1 给出了在测试集下基线系统和本文提出的融合 IB 特征方法的检测结果，可以看出，本文提出的融合 IB 特征方法比基于 HMM/GMM 的方法的说话人误配率降低了 1.2%，性能相对提高了 9.2%。图 6 给出了 12 段测试音频数据在基于 HMM/GMM 模型的方法和在本文提出的融合 IB 特征方法下的检测数据结果，可以看出，采用融合 IB 准则的方法，除了在数据 2 和 5 时的说话人误配率测试结果比基于 GMM/HMM 的方法略高外，在其他数据下的测试结果性能较基于 HMM/GMM 的方法有所提升。这充分证明，基于 IB 准则的特征，为说话人聚类提供了可用信息，使得说话人分段聚类时，说话人类别的纯度更高，各类中被错分出去的说话人数据减少，使得说话人误配率降低。

表 1 不同方法性能对比
Table1 Performance comparison of different methods

method	speaker match error/(%)
HMM/GMM based method	13.1
proposed method	11.9

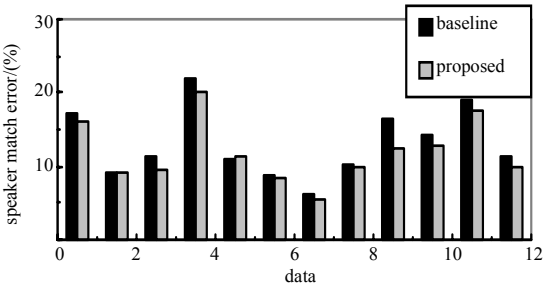

Fig.6 Comparison of the results with different methods
图 6 对比测试结果示意图

5　结论

本文提出了一种融合基于 IB 准则特征的说话人分段与聚类算法，并与传统的基于 MFCC 特征方法进行了比较。实验证明，基于 IB 准则的特征能够与传统的 MFCC 特征形成互补，提供更多的有效信息，有助于提高说话人聚类的纯度，有效降低说话人误配率。

参考文献:

[1] Vijayasenan D,Valente F,Bourlard H. Combination of agglomerative and sequential clustering for speaker diarization[C]// Proceedings of ICASSP 2008.Las Vegas,NV:[s.n.],2008:4361-4364.

[2] Tranter S E. Two-way cluster voting to improve speaker diarisation performance[C]// Proceedings of ICASSP'05. Philadelphia, Pennsylvania,USA:[s.n.], 2005:753-756.

[3] Mergnier S,Moraru D,Fredouille C,et al. Step-by-step and integrated approaches in broadcast news speaker diarization[J]. Computer Speech and Language, 2006,20(2-3):303-330.

[4] Bozonnet S,Evans N W D,Anguera X,et al. System output combination for improved speaker[C]// 11th Annual Conference of the International Speech Communication Association. Makuhari,Japan:[s.n.], 2010:2642-2645.

[5] Vijayasenan D,Valente F,Bourlard H. An Information Theoretic Approach to Speaker Diarization of Meeting Data[J]. IEEE Transactions on Audio, Speech and Language Processing, 2009,17(7):1382-1393.

[6] Anguera X. Robust speaker diarization for meetings[D]. Barcelona:Universitat Politecnica De Catalunya, 2006.

[7] Tishby N,Pereira F,Bialek W. The information bottleneck method[C]// Proceedings of the 37th Annual Allerton Conference on Communication,Control and Computing. Monticello,IL,USA:[s.n.], 1999:368-377.

[8] Pardo J,Anguera X,Wooters C. Speaker diarization for multiple-distant-microphone meetings using several sources of information[J]. IEEE Transactions on Computers, 2007,56(9):1212-1224.

[9] 贾兰兰. 一种快速稳健的语音 / 音乐分类方法[J]. 信息与电子工程,2008,6(4):281-283,288. (JA Lanlan. A Fast and Robust Speech/Music Discrimination Approach[J].Information and Electronic Engineering, 2008,6(4): 281-283,288.)

[10] Vijayasenan D,Valente F,Bourlard H. An Information Theoretic Combination of MFCC and TDOA Features for Speaker Diarization[J]. IEEE Transactions on Audio,Speech and Language Processing, 2011,19(2):431-438.

作者简介:

张　力(1985-)，男，江西省九江市人，在读硕士研究生，主要研究方向为语音信号处理、说话人分段与聚类.email:zl_261200@sina.com.

张连海(1971-)，男，郑州市人，副教授，主要研究方向为语音信号处理、语音编码.

许友亮(1985-)，男，乌鲁木齐市人，在读硕士研究生，主要研究方向为连续语音识别.