ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 31,No. 2,Mar. 2016,pp. 362－369
DOI:10.16337/j.1004-9037.2016.02.017

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

基于因子分析和特征映射的耳语说话人识别

张庆芳　赵鹤鸣　龚呈卉

(苏州大学电子信息学院,苏州,215006)

摘　要：为了解决耳语音识别系统中训练语音和测试耳语音来自不同发音模式的失配问题,本文提出一种基于联合因子分析(Joint factor analysis,JFA)与特征映射(Feature mapping,FM)的失配信息补偿算法。该算法首先用联合因子分析法计算说话人发音模式信息,并对发音模式因子和发音模式空间参数进行优化,接着对语音参数用发音模式信息进行特征映射后再进行训练和识别,以减少发音模式对系统的影响。实验结果表明,基于因子分析和特征映射的方法可以有效地提取训练语音中的说话人信息,提高耳语识别系统的识别率。

关键词：耳语说话人识别;联合因子分析;特征映射;正常音;耳语音

中图分类号：TP391　　**文献标志码**:A

Whispered Speaker Identification Based on Factor Analysis and Feature Mapping

Zhang Qingfang, Zhao Heming, Gong Chenghui

(School of Electronics and Information,Soochow University, Suzhou, 215006, China)

Abstract: Aiming at the mismatch between training speech and test speech from different speaking manners, a kind of feature processing algorithm is proposed based on joint factor analysis and feature mapping. The speaking mode information is extracted by joint factor analysis algorithm, then the speaking mode factor and space are optimized. Before training and test,the feature is mapped by speaking mode information to reduce the speaking mode effects. The experimental results show that the proposed algorithm can effectively extract the speaker information of the training speech, and improve the recognition rate of whispered speaker recognition system.

Key words: whisper speaker recognition; joint factor analysis; feature mapping;normal speech; whispered speech

引　　言

耳语说话人识别是通过对待识别说话人的耳语音分析来识别说话人的身份,它在安全场所的身份识别、犯罪鉴定等方面有着重要的意义,因此耳语说话人识别及其相关研究越来越受到重视[1-5]。由于耳语音受发音情绪、发音环境等影响较大,在实际应用时,很难获取与测试语音相同背景的足够的训练用耳语音,于是出现训练语音和测试语音失配问题,即训练用语音来自与测试耳语音不同的信道,或与

基金项目:国家自然科学基金(61071215,61372146)资助项目;苏州市科技计划(SYG201033)资助项目。
收稿日期:2014-12-28;**修订日期**:2015-05-16

测试语音发音时说话人的情绪不同,或与测试语音的发音模式不同,这些问题使得耳语识别系统的识别率严重下降[6-8]。常见的发音模式有耳语、低声说、正常说、大声说及大声喊,本文主要研究耳语音和正常音这两种来自不同发音模式的语音在耳语说话识别系统中的情况,即训练语音为正常音,而测试语音为耳语音,目前此方面的研究较少。文献[9]采用线性频率倒谱系数(Linear frequency cepstral coefficients,LFCC)参数,利用特征映射的方法提高了不同发音模式下失配系统的识别率。文献[10]提取了语音的瞬时频率参数,将它应用到失配系统中。这些方法都采用了基于通用背景模型(Universal background models,UBM)的特征映射这一补偿手段,在补偿过程中将说话人和说话模式等信息混合在一起补偿,在映射过程中无法独立提取独立的说话人发音模式信息。联合因子分析方法能根据训练语音的数据来提取出说话人信息和信道信息等,但是联合因子分析方法需要充分的训练数据。

本文尝试将联合因子分析和特征映射相结合,首先用联合因子分析法从训练用的耳语音和正常音中提取出说话人的发音模式信息,然后用提取出的发音模式信息进行特征映射。结果表明使用本文方法系统的平均正确识别率高于基于 UBM 的特征映射法。

1 联合因子分析

联合因子分析[11,12]是模型域的一种信道补偿技术,它在不匹配信道下的耳语音说话人识别中已经起到了很好的作用。联合因子分析法[13]认为说话人的每段话对应的超向量含有说话人超向量和信道超向量这两部分,可表示为

$$\boldsymbol{M}_h(s)=\boldsymbol{m}+\boldsymbol{M}(s)+\boldsymbol{M}(c) \tag{1}$$

式中:s 为某个说话人;h 为说话人的某段话;$\boldsymbol{m}$ 为与说话人无关且与信道无关的超向量;$\boldsymbol{m}+\boldsymbol{M}(s)$ 为说话人超向量,它描述的为受说话人情绪、说话方式等影响的部分;$\boldsymbol{M}(c)$ 为信道超向量,它描述为受信道影响部分。联合因子分析法通过估计并消除信道因子实现模型补偿,提高不同信道下的说话人识别系统的识别率。

在耳语说话人识别系统中,当训练语音来自与测试耳语音不同的说话模式时,系统性能下降[8],显然说话模式是影响说话人识别系统的一个重要因素。可以将语音信号中说话人信息分为与发音模式无关的说话人个性信息和说话人发音模式信息两部分组成,则说话人的每段话对应的超向量可表示为

$$\boldsymbol{M}_h(s)=\boldsymbol{m}+\boldsymbol{M}_g(s)+\boldsymbol{M}_f(s)+\boldsymbol{M}(c) \tag{2}$$

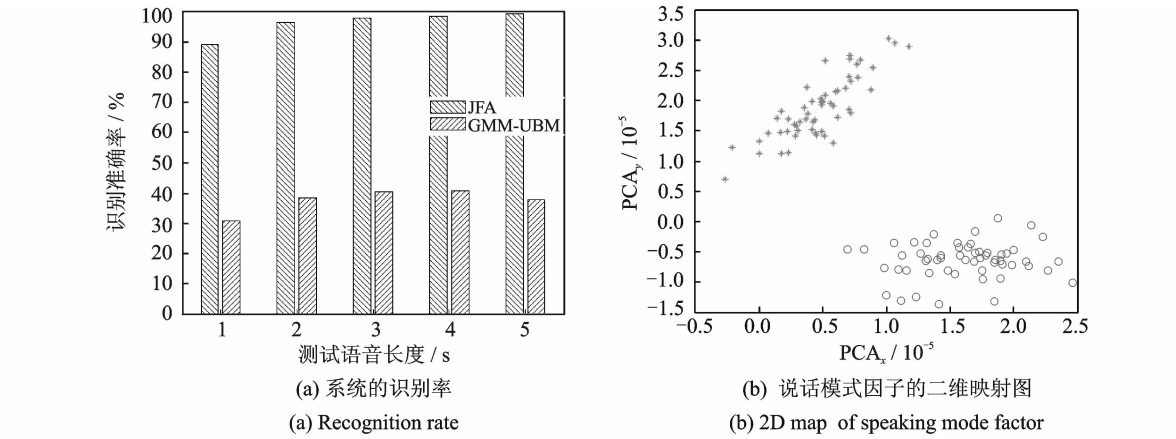
式中:$\boldsymbol{m}+\boldsymbol{M}_g(s)$ 为说话人超向量,它描述说话人个性特征与说话模式无关部分;$\boldsymbol{M}_f(s)$ 为说话模式超向量,它描述受说话模式影响部分。

本文研究训练语音和测试耳语音来自相同信道的情况,则不存在信道失配的问题,即不需要进行信道补偿,失配来源于说话人发音模式的不同,需要补偿的是说话人的发音模式信息,所以式(2)中说话人的每段话对应的超向量可表示为

$$\boldsymbol{M}_h(s)=\boldsymbol{m}+\boldsymbol{z}*\boldsymbol{g}(s)+\boldsymbol{v}*\boldsymbol{f}(s) \tag{3}$$

式中:$\boldsymbol{z}$ 为说话人空间;$\boldsymbol{g}(s)$ 为说话人因子;$\boldsymbol{z}*\boldsymbol{g}(s)$ 为说话人个性特征与发音模式无关的部分;$\boldsymbol{v}$ 为说话人发音模式空间;$\boldsymbol{f}(s)$ 为说话人发音模式因子;$\boldsymbol{v}*\boldsymbol{f}(s)$ 为说话人发音模式超向量,它描述受说话人的发音模式影响部分,由于训练语音均来自同一信道,所以与信道相关的信息包含在超向量 $\boldsymbol{m}$ 中。本文的目标是保留说话人之间差异,抑制说话人发音模式的差异。

当有足够的训练语音数据时,即联合因子分析(Joint factor analysis,JFA)模型训练时,每个说话人有足够的正常音和耳语训练语音,如果在说话人模型训练时仅采用正常语音,测试时采用耳语音,如图1(a)所示,JFA 系统的平均识别率为 96.25%,高斯混合通用背景模型(Gaussian mixture model-UBM,GMM-UBM)系统的识别率为 37.68%,此时联合因子分析系统的识别率较高,说明联合因子分析同样适用于发音模式不同的失配系统中。本文对说话人因子向量进行主成分分析,得到二维向量(PCAx,



(a) 系统的识别率 (b) 说话模式因子的二维映射图

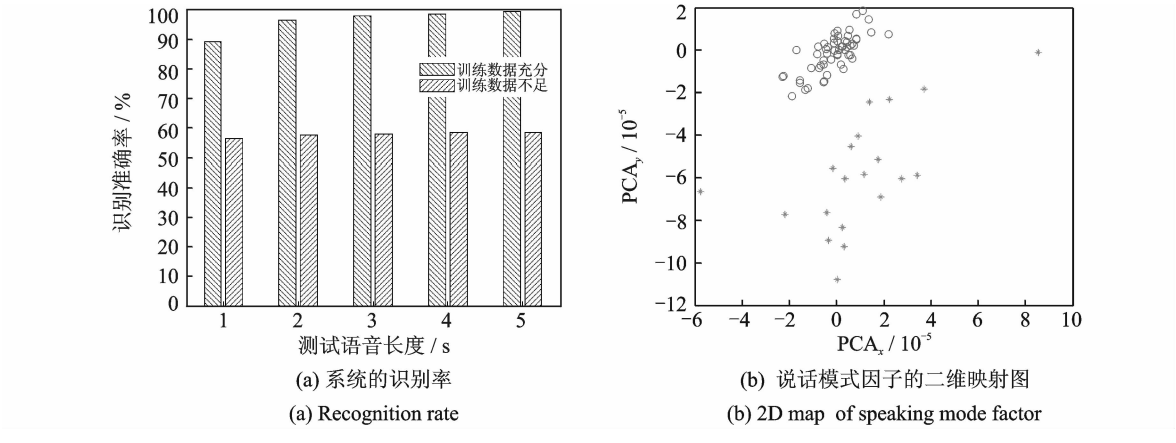
(a) Recognition rate (b) 2D map of speaking mode factor

图 1 训练数据充足情况下系统的数据

Fig. 1 System performance with sufficient training data

PCAy),图 1(b)为训练数据充足情况下说话人因子向量的二维映射图。

但是当没有足够的训练语音数据时,即在 JFA 模型训练时,没有足够的耳语训练语音,仅有一半说话人有耳语训练语音,如图 2(a)所示,JFA 识别系统识别率大大下降,平均识别率仅为 57.62%,正确的识别主要来自于具有充分耳语训练语音的一半测试者。由此可知当训练数据不足时,说话人的空间估计受到很大的影响,使得系统识别率下降,这说明仅仅单独使用 JFA 模型不能很好地解决训练数据不充分情况的问题。图 2(b)为训练数据不足情况下说话人因子向量的二维映射图,从图 2(b)中可以发现,说话模式因子空间的区分度依旧很好,这主要是因为 JFA 模型训练时具有完整的说话人发音模式数据,所以提取出的发音模式因子能较好地反映训练数据的发音模式信息,因此本文将联合因子分析中的发音模式空间和发音模式因子用于特征映射。



(a) 系统的识别率 (b) 说话模式因子的二维映射图

(a) Recognition rate (b) 2D map of speaking mode factor

图 2 训练数据不足情况下的系统的数据

Fig. 2 System performance with insufficient training data

2 基于因子分析的特征映射法

对于不同信道下的正常音说话人识别系统,文献[14]将信道因子用于特征补偿,较好地提高了识别系统的识别率。考虑到失配系统中说话人发音模式因子的估计较好,本文将发音模式因子用于失配信

息补偿,流程图如图 3 所示。

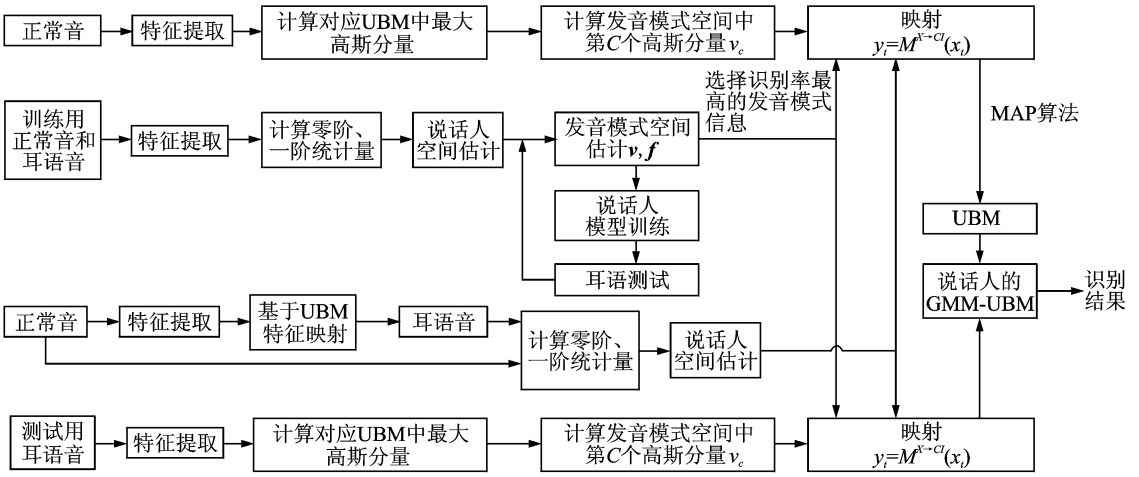


图 3 基于联合因子和特征映射的系统工作流程

Fig. 3 Workflow of system based on JFA and FM

2.1 背景模型的训练

通用背景模型 UBM$\{\boldsymbol{\mu},\Sigma,\boldsymbol{w}\}$ [15] 与说话人和说话人发音模式均无关,其中 $\boldsymbol{\mu}$,Σ,$\boldsymbol{w}$ 分别为 UBM 的均值、方差和权重矩阵。一般 UBM 由所有发音模式的语音数据一起训练得到,但是在失配模式下,耳语音数据相对较少,耳语音和正常语音数据量不均会使得 UBM 质心偏移,因此本文首先分别用 EM 算法训练正常语音的 $\mathrm{UBM_N}$ 和耳语音的 $\mathrm{UBM_W}$,然后将二者合并为一个与发音模式无关的 UBM。其模型训练过程如图 4 所示。

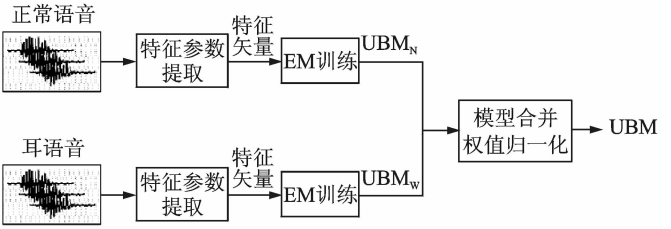


图 4 UBM 模型的训练过程

Fig. 4 Training process of UBM model

2.2 JFA 模型

2.2.1 基本算法

(1)统计量的计算。在训练说话人空间、说话人发音模式空间,以及说话人训练和其测试的时候都需要用到统计量,在本系统中,由于只更新模型的均值参数,所以只估计了零阶和一阶统计量。假定说话人 s 以及多段语句 $h=1,\cdots,H(s)$,令

$$N_{h,c}(s)=\sum_t \gamma_t(c) \tag{4}$$

式中:$\gamma_t(c)$ 为第 t 帧对应于 UBM 的后验概率(即高斯占有率);h 为说话人 s 的第几段话;c 为某个高斯元。可得

$$N_c(s)=\sum_h N_{h,c}(s) \tag{5}$$

令 $\boldsymbol{N}(s)$ 表示为 $CF\times CF$ 的对角矩阵,每个小矩阵块为 $N_c(s)\boldsymbol{I}(c=1,\cdots,C)$,$\boldsymbol{I}$ 为 $F\times F$ 的单位阵。然后再计算其一阶统计量

$$\boldsymbol{F}_{hc}(s)=\sum_t \gamma_t(c)\,(\boldsymbol{Y}_t-\boldsymbol{m}_c) \tag{6}$$

$$\boldsymbol{F}_c(s)=\sum_h \boldsymbol{F}_{hc}(s) \tag{7}$$

式中：$\boldsymbol{m}_c$ 为第 c 个 UBM 高斯元的均值向量；$\boldsymbol{Y}_t$ 为第 t 帧特征向量。将 $\boldsymbol{F}_c(s)$ $(c=1,\cdots,C)$ 向量串接形成 $\boldsymbol{F}(s)$，维数为 $CF\times 1$。

(2) 说话人空间的估计。说话人空间的估计采用的方法是类似于 EM 步骤的算法，分为 E 步和 M 步，具体为：

步骤 E 设定说话人 s 有 $H(s)$ 段话组成，先假定说话人模型空间是随机产生，然后分别求出说话人因子的一阶和二阶统计量。令

$$\boldsymbol{l}(s)=\boldsymbol{I}+\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{N}(s)\boldsymbol{z} \tag{8}$$

则

$$E(\boldsymbol{g}(s))=\boldsymbol{l}^{-1}(s)\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(s) \tag{9}$$

$$E(\boldsymbol{g}(s)\boldsymbol{g}(s)^{\mathrm{T}})=E(\boldsymbol{g}(s))E(\boldsymbol{g}(s)^{\mathrm{T}})+\boldsymbol{l}^{-1}(s) \tag{10}$$

根据文献[16]可知，当因子的先验分布为高斯分布时，其相应的后验分布仍为高斯分布，通过式(9)可以求出说话人因子，在这个步骤中，通过对说话人 s 的所有语音段进行求和而得到，显然说话模式的影响就相应地平均化，从而可以认为只有说话人信息存在。

步骤 M 通过上面的统计量来估计说话人空间，令

$$\boldsymbol{\Phi}_c=\sum_s N_c(s)E[\boldsymbol{g}(s)\boldsymbol{g}^{\mathrm{T}}(s)] \quad c=1,\cdots,C \tag{11}$$

$$\boldsymbol{\Theta}=\sum_s \boldsymbol{F}(s)E[\boldsymbol{g}^{\mathrm{T}}(s)] \tag{12}$$

式中：$\boldsymbol{\Phi}_c$ 和 $\boldsymbol{\Theta}$ 的维数分别为 $R_s\times R_s$ 和 $CF\times R_s$，对于每个高斯元 c，通过如下的矩阵方程更新 $\boldsymbol{z}$ 的第 i 行

$$\boldsymbol{v}_i\boldsymbol{\Phi}_c=\boldsymbol{\Theta}_i \quad i=(c-1)\times F+1,\cdots,(c-1)\times F+F \tag{13}$$

(3)发音模式空间的估计。估计说话人发音模式空间的方法和说话人空间的估计相似，但在估计说话人发音模式空间时，一阶统计量的计算以 $\boldsymbol{m}=\boldsymbol{z}*E(\boldsymbol{g}(s))$ 为基准，而不以 $\boldsymbol{m}$ 为中心，所以在说话人发音模式空间进行估计时，要先计算说话人空间。

2.2.2 发音模式因子优化

在进行 JFA 模型训练过程中，由于说话人空间和发音模式空间都随机产生，所以每次训练后得到的 JFA 参数不同，使得系统的识别率也不同。本文随机产生 5 次训练初始值，并用训练后得到的 JFA 参数进行识别，图 5 所示为训练数据充分时不同 JFA 参数下的系统识别率，平均识别率均在 95% 以上，最高为 96.67%，最低为 95.65%，为了获得更好的 JFA 参数，本文在训练发音模式空间时，基于识别率对 JFA 参数进行优化，具体步骤如下：(1)初始化发音模式矩阵 $\boldsymbol{v}$，计算初始化发音模式因子 $f(s)$；(2)训练说话人模型；(3)进行说话人测试，并计算平均识别率。重复步骤(1～3)共 10 次，选择平均识别率最高的一组 JFA 参数作为最后训练结果。

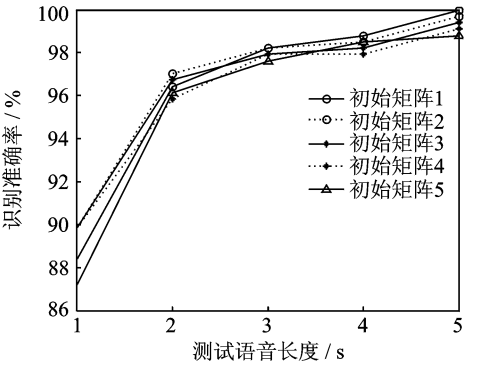


图 5 不同初始参数下 JFA 的系统识别率
Fig. 5 Recognition rates of JFA system with different initial parameters

2.3 语音的特征映射

文献[10]将传统的基于 UBM 特征映射方法[17]应用到不同发音模式下的说话人识别系统，提高了识别率，证明了特征映射同样适用于不同发音方式下的说话人识别系统。特征映射流程如下：

(1)对于某说话人 s 的某帧特征参数 $\boldsymbol{x}_t$，首先计算出其对应通用背景模型 UBM 中的最大高斯分量 $c=\underset{1\leqslant j\leqslant M}{\arg\max}\{W_j N(\boldsymbol{x}_t\,|\,\mu_j^X,\Sigma_j^X)\}$；

(2)根据最大高斯分量的序号 c,找出对应的说话人模式空间 v 中的的第 c 个高斯分量 $\boldsymbol{v}_c$;

(3)对于特征矢量 $\boldsymbol{x}_t$,通过式(14)映射到与发音模式无关的特征矢量 $\boldsymbol{y}_t$。

$$\boldsymbol{y}_t = \boldsymbol{M}^{X \to Cl}(\boldsymbol{x}_t) = \boldsymbol{x}_t - \boldsymbol{v}_c * \boldsymbol{f}_h(s) \tag{14}$$

2.4 说话人模型的训练和测试

本文采用正常语音的特征矢量 $\boldsymbol{x}_t^N$ 作为说话人模型的训练数据,在说话人训练过程中,通过式(14)对 $\boldsymbol{x}_t^N$ 进行特征映射,将训练语音中的与发音模式相关的成分去除,得到与发音模式无关的特征矢量 $\boldsymbol{x}_t$,接着用 MAP 算法对背景模型 UBM$\{\boldsymbol{\mu}, \Sigma, \boldsymbol{w}\}$ 作自适应训练,获得每个人的 GMM-UBM 模型。

在说话人模型测试时,测试语音全部采用耳语音,对测试用的耳语音同样进行特征映射,来减少训练语音和测试语音中发音模式的差异,接着再进行识别。整个训练和测试流程如图 4 所示。这种方法与传统的特征映射相比,它的优势在于根据联合因子分析的原理,发音模式空间的估计是在去除说话人信息的基础上进行估计,因此更好地描述了发音模式信息,相应的发音模式因子更加精确,而基于 UBM 的特征映射就没有考虑到说话人,只是单纯地做自适应。

3 实验与结果分析

3.1 数据描述

语音数据均来自于某大学电子信息学院的耳语音数据库,本文选用在安静环境下使用手持式传声器录制的正常音和耳语音。测试者共 56 人,全部为男性,耳语音的语音长度为 70～90 s,正常音的语音长度约为 20 s,语音均采用 PCM 编码,采样频率为 8 kHz,语音处理均在 Matlab 平台上实现。

3.2 实验系统

(1) 预处理和特征提取。首先采用预加重来提升语音信号高频部分的能量,高通滤波器的形式为 $H(z)=1-\mu z^{-1}$。对耳语音进行端点检测时,只切除语句的开头、结尾和大段的空白部分。同时还对语音进行去直流和归一化处理。对每帧语音提取 12 维 MFCC 及其相应的一阶差分系数,总共 24 维。

(2) 模型训练。选用每人后 30 s 的耳语音作为耳语测试语音,将剩下的耳语音作为耳语训练语音,每人的正常音(20 s)全部作为训练语音。由于数据库中正常音的数据量偏少,而耳语音数据量偏多的特点,所以本文采用分开训练 UBM 的方法,具体是分别采用 56 人的正常音(每人 20 s)和 56 人的耳语音(每人切分 30 s 后所剩下的耳语音)训练 64 元的 $\mathrm{UBM_N}$ 和 64 元的 $\mathrm{UBM_W}$,然后将二者合并成一个总的 128 元的 UBM 模型。JFA 模型训练过程中,采用前 28 人的正常音和耳语音的语音库来进行训练计算发音模式空间,在说话人模型训练时仅采用后 28 人正常音进行训练,其中说话人空间计算中后 28 人耳语信道的数据对正常音通过基于 UBM 的特征映射获得。

(3) 实验设置。实验比较了不同系统的识别率。在本文中,训练数据不足的情况是所有 56 人的正常音和前 28 人的耳语音用来进行训练,而将后 28 人耳语音用来进行测试,测试时长分为 1,2,3,4,5 s。

3.3 实验数据分析

图 6 给出在训练数据充足情况下,基于 UBM 的特征映射法和基于联合因子法的系统识别率比较。在基于 UBM 的特征映射系统中,训练时采用后 28 人的正常音和耳语训练语音计算映射参数正常音模型 λ_{N}、耳语音模型 λ_{W},将每个说话人的正常音映射为与发音模式无关的语音,并用 MAP 算法自适应出此说话人的 UBM-GMM 模型。测试时先将测试耳语音映射为与发音模式无关的语音,然后再进行识别。联合因子法在训练 JFA 模型时采用后 28 人的正常音和耳语训练音,在说话人模型训练时仅采用耳语音(取后 30 s 语音)。从识别结果可以看出,联合因子法明显优于基于 UBM 的特征映射法,这主

要是因为联合因子分析法的补偿是对说话模式信息进行单独补偿,而基于 UBM 的特征映射法将语音中的所有信息混合在一起补偿。

图 7 给出在训练数据不充分下,即训练语音和测试语音为发音模式失配情况时,分别采用基于 UBM 的特征映射法和本文提出的因子映射法的系统识别率比较。基于 UBM 的特征映射系统在 1～5 s 测试时间长度时,平均识别率为 46.31%,本文所提出的因子映射法的平均识别率为54.76%,识别率提升 8.45%。从测试结果可发现,当训练语音中缺少充分的耳语音时,仅采用基于 UBM 的特征映射法识别率较低,这主要因为基于 UBM 的特征映射法在对特征进行映射时将说话人的发音信息和其他个性信息一起进行映射,而因子映射法在映射时只是降低语音之间的发音模式差异,较好地保护了说话人与发音模式无关的个性特征,有助于系统的性能提升,同时说明正常音作为一种相对容易获取的语音,虽然和耳语音不属于同一说话方式,但是由于其中含有的说话人信息,所以仍然可以用于耳语识别系统性能的提升。

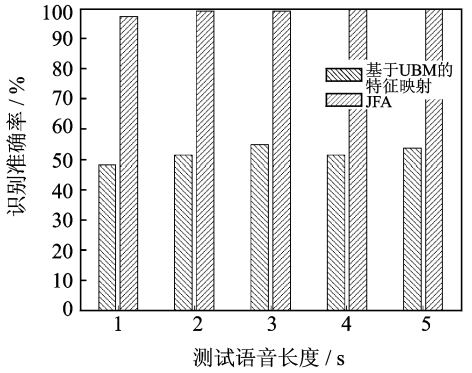


图 6 训练数据充足情况下不同方法的识别率

Fig. 6 Comparison of recognition rates with sufficient training data

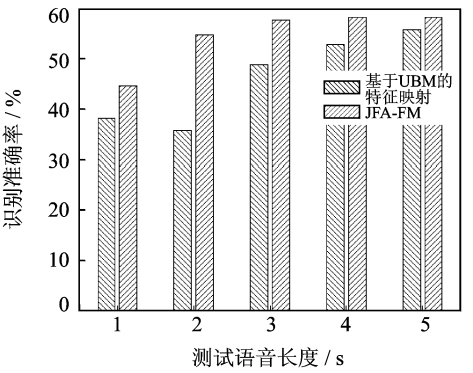


图 7 训练数据不充分情况下不同方法的识别率

Fig. 7 Comparison of recognition rates with insufficient training data

4 结束语

本文对不同发音模式下耳语说话人识别系统在不同发音模式下进行研究,将常用于信道补偿的联合因子分析算法应用于不同发音模式的耳语说话人识别系统。当训练数据充足时,单独使用联合因子分析算法可以直接获得较好的系统识别率,而且相比于特征映射,此方法无需进行信道识别。但当训练数据缺少时,仅使用因子分析算法,系统的性能大大下降,针对这个问题,本文将特征映射与因子分析算法相结合,通过因子分析算法提取发音模式信息,并利用它进行特征映射,以减少发音模式对说话人识别系统的影响。通过实验发现,采用因子映射法系统的识别率获得了明显的提高。但是本文仅对正常音和耳语音这两种来自不同发音模式的语音进行研究,实际上常见的发音模式还有低声说、大声说和大声喊等,后续将收集这些不同发音模式语音,进一步验证本文方法的有效性。

参考文献:

[1] Jin Q, Jou S C S, Schultz T. Whispering speaker identification[C]// 2007 IEEE International Conference on Multimedia and Expo. Beijing, China: IEEE, 2007: 1027-1030.

[2] Lin Wei, Yang Lili, Xu Boling. A new frequency scale of Chinese whispered speech in the application of speaker identification [J]. Progress in Natural Science, 2006, 16(10): 1072-1078.

[3] Jawarkar N P, Holambe R S, Basu T K. Speaker identification using whispered speech[C]// 2013 International Conference on Communication Systems and Network Technologies (CSNT). Washington, DC, USA: [s. n.], 2013: 778-781.

[4] Gong Chenghui, Zhao Heming, Tao Zhi. Speaker identification of whispered speech with perceptible mood[J]. Journal of Mul-

timedia,2014,9(4):553-561.
[5] Fan X, Hansen J H L. Speaker identification for whispered speech based on frequency warping and score competition[C]// 9th Annual Conference of the International Speech Communication Association. Brisbane, Australia: INTERSPEECH, 2008:1313-1316.
[6] Wang Yanlei, Zhao Heming, Gu Xiaojiang. A study on speaker and session variability in speaker recognition of Chinese whispered speech[C]// 2010 the 2nd International Conference on Industrial Mechatronics and Automation (ICIMA). Wuhan, China: [s. n.],2010: 292-295.
[7] 顾晓江, 赵鹤鸣, 吕岗. 模型与特征混合补偿法及其在耳语说话人识别中的应用[J]. 声学学报, 2012, 37(2): 198-203.
Gu Xiaojiang, Zhao Heming, Lü Gang. An application in whispered speaker identification using feature and model hybrid compensation method[J]. Acta Acustica, 2012, 37(2): 198-203.
[8] Zhang C, Hansen J H L. Analysis and classification of speech mode: Whispered through shouted[C]// INTERSPEECH 2007. Antwerp, Belgium: ISCA, 2007: 2289-2292.
[9] Fan X, Hansen J H L. Speaker identification with whispered speech based on modified LFCC parameters and feature mapping[C]// Acoustics, Speech and Signal Processing, 2009. Taipei, China: IEEE, 2009: 4553-4556.
[10] 王敏,赵鹤鸣,张庆芳. 基于瞬时频率估计和特征映射的汉语耳语音话者识别[J]. 数据采集与处理,2011,26(6):686-690.
Wang Min, Zhao Heming, Zhang Qingfang. Speaker identification with Chinese whispered speech based on instantaneous frequency estimation and feature mapping[J]. Journal of Data Acquisition and Processing, 2011, 26(6): 686-690.
[11] Kenny P. Joint factor analysis of speaker and session variability: Theory and algorithms[R]. CRIM 08-13, Montreal: CRIM, 2005.
[12] Dehak N, Kenny P J, Dehak R, et al. Front-end factor analysis for speaker verification[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2011, 19(4): 788-798.
[13] Kenny P, Ouellet P, Dehak N, et al. A study of interspeaker variability in speaker verification[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2008, 16(5): 980-988.
[14] 李轶杰,郭武,戴礼荣. 话者识别的信道补偿[J]. 小型微型计算机系统,2008,29(12):2344-2347.
Li Yijie, Guo Wu, Dai Lirong. Model session variability in speaker verification[J]. Journal of Chinese Computer Systems, 2008, 29(12): 2344-2347.
[15] Reynolds D, Thomas A, Quatieri F, et al. Speaker verification using adapted Gaussian mixture models[J]. Digital Signal Processing, 2000, 7(1): 19-41.
[16] Kenny P. Eigenvoice modeling with sparse training data[J]. IEEE Trans, Audio, Speech and Language Processing, 2005, 13(3): 345-354.
[17] Reynolds D A. Channel robust speaker verification via feature mapping[C]// ICASSP'03. Hong Kong, China: IEEE, 2003.

作者简介:

张庆芳(1981-),女,博士研究生,研究方向:语音信号处理,E-mail: qfzclear@aliyun.com。

赵鹤鸣(1957-),男,教授,博士生导师,研究方向:语音信号处理、人工神经网络理论与应用。

龚呈卉(1981-),女,博士研究生,研究方向:语音信号处理。