DOI: 10.13382/j.jemi.B2307091

基于 DTW-GMM 的光纤传感系统声纹识别方法*

杨佳沛[1] 王 宇[1,2,3] 彭广建[1] 白 清[1,2,3] 刘 昕[1,2,3] 靳宝全[1,2]

(1. 太原理工大学电子信息与光学工程学院 太原 030024;2. 太原理工大学新型传感器与智能控制教育部重点实验室 太原 030024;3. 山西省交通科技研发有限公司 太原 030024)

摘 要:为了满足易燃易爆环境的声纹识别需求,设计了直线型萨格奈克干涉光纤声音传感系统,利用维纳滤波算法对语音数据进行了降噪,通过三电平削波法获取了基音周期特征,采用动态时间规整算法筛选了说话人样本,并提取了梅尔频率倒谱系数特征,运用高斯混合模型-期望最大化算法开展了声纹识别实验研究,同时探究了光纤声音传感系统的频率响应特性与声纹特征,研究了采集语音幅值对声纹识别结果的影响。实验结果表明,系统可实现 300~3 500 Hz 频率段的声音信号感知,声音幅值从 0.9 V 降至 0.15 V 时最大与次大对数似然值之差由 35.5 降至 10.9,识别结果从成功变为失败。重复性实验表明,在 10 km 的传感光纤上,距声源 2 m 位置处,传感系统可对 400 段时长为 3~5 s 之间的文本无关语音段实现准确检测,且综合识别准确率为 94.75%。本系统有望为易燃易爆环境中的设备故障、应急救援、渗漏监测等领域提供声纹识别的解决方案。

关键词: 光纤传感;萨格奈克干涉;声纹识别;高斯混合模型

中图分类号: TH741 **文献标识码**: A **国家标准学科分类代码**: 510.5020

Voiceprint recognition method of optical fiber sensing system based on DTW-GMM

Yang Jiapei[1] Wang Yu[1,2,3] Peng Guangjian[1] Bai Qing[1,2,3] Liu Xin[1,2,3] Jin Baoquan[1,2]

(1. College of Electronic Information and Optical Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 2. Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan 030024, China; 3. Shanxi Transportation Technology Research & Development Company Limited, Taiyuan 030024, China)

Abstract: In order to meet the demand of voiceprint recognition in flammable and explosive environment. A linear Sagnac interference optical fiber acoustic sensor system has been designed. Speech data was denoised using the Wiener filtering algorithm, and pitch features were extracted through three-level clipping. Speaker samples were screened using dynamic time warping, and Mel-frequency cepstral coefficients were extracted as features. Voiceprint recognition experiments were conducted utilizing the Gaussian mixture model-expectation maximization algorithm, concurrently investigating the frequency response characteristics of the optical fiber acoustic sensor system and their relationship with voiceprint features. The influence of the amplitude of acquired speech on voiceprint recognition outcomes was studied. Experimental results demonstrate that the system can realize the sound signal perception in the frequency range of 300~3 500 Hz. When the sound amplitude decreases from 0.9 to 0.15 V, the difference between the maximum and second-largest log-likelihood values drops from 35.5 to 10.9, the recognition result changed from success to failure. Repetition experiments show that, at a distance of 2 meters from the sound source along a 10-kilometer sensing fiber, the system accurately recognizes 400 speech segments of 3 to 5 seconds duration, unrelated to any specific text, achieving an overall identification accuracy rate of 94.75%. This system holds promise as a solution for voiceprint recognition in applications such as equipment fault diagnosis and emergency response within flammable and explosive environments.

收稿日期:2023-11-29 Received Date: 2023-11-29

*基金项目:山西省重点研发计划项目(202102130501021)、山西省水利科学技术研究与推广项目(2024GM18)、中央引导地方科技发展资金项目(YDZJSX20231B004)、山西省科技创新团队项目(201805D131003)资助

Keywords: optical fiber sensing; Sagnac interference; voiceprint recognition; Gaussian mixture model

0　引　言

光纤传感技术作为新型传感器,因其高灵敏度、抗电磁干扰、抗腐蚀性、本质安全等特点,已被作为井下应急救援通信[1]、光纤水听器[2]、地震声波监测仪[3]等声学监测设备在多种领域应用。

2023 年,贵州大学的左一武等[4]提出并验证了一种基于大角度倾斜光纤光栅包层模的低频声传感方案,有望应用于低频声探测领域。2021 年,长春理工大学的余双勇等[5]设计出特种光纤高感知结构,使光纤声波传感器具有更好的采集质量。2021 年,中国科学院上海光学精密机械研究所的王照勇等[6]将分布式光纤声波传感器用于新兴地震波检测。2018 年,太原理工大学的宋晓达等[7]设计出一种本质安全无源检测的矿用双通道光纤拾音系统。2021 年,南昌航空大学尹玺等[8]设计出一种基于法布里-珀罗微腔结构的光纤声传感系统,整个系统结构简单,有望在声音检测领域应用。2021 年,上海工程技术大学的吴虎等[9]利用端点检测与信号重组的方法进行光纤声音信号识别,可以有效识别出施工声、触摸声、噪声、汽车声。近几年,国内科研人员为光纤传感技术的发展做出了贡献,为声音检测领域提供了新方法和新思路。

声纹特征和虹膜、视网膜、指纹、人脸等常见生物特征类似,也是一种可代表一个人独一无二的标识[10],所以声纹识别技术也叫作说话人识别技术。相较于其他生物特征,声纹特征具有高采样灵活度、高识别准确率、低采样成本、可远程识别等独特优势[11]。2000 年,麻省理工学院的 Reynolds 等[12]提出高斯混合通用背景模型并将其应用于声纹识别,结果表明系统的验证性能和效率得到显著提升。2011 年,加拿大魁北克大学蒙特利尔分校的 Dehak 等[13]在传统的高斯混合模型的方法上提出了 i-vectors 模型,相较于联合因子分析法准确率提升 4%。2018 年,约翰霍普金斯大学的 Snyder 等[14]提出了基于深度神经网络的说话人识别模型 x-vectors,与 i-vectors 模型相比,具有更好地大规模训练数据集,性能更卓越。而随着声纹识别技术的逐步发展,其已经被广泛应用于机械故障分析[15-16]、居民住宅和办公场所的声音锁[17]、人机交互身份确认[18]、刑事侦查[19]等领域。然而在煤矿井下、木制文物建筑等易燃易爆场合,声纹识别技术应用的需求难以解决。

本文运用直线型萨格奈克(Sagnac)光纤传感器采集声音信号,利用维纳滤波算法降噪,提取基音周期和梅尔频率倒谱系数(MFCC)特征,利用动态时间规整算法(dynamic time warping,DTW)比较基音周期,利用高斯混合模型-期望最大化算法(GMM-EM)算法确定 MFCC 特征归属说话人。系统有望应用于易燃易爆环境中的设备故障、应急救援等声纹检测及识别领域。

1　理论分析

1.1　萨格奈克干涉光纤传感原理

如图 1 所示为传统的直线型萨格奈克干涉系统结构,传感系统主要包括激光器、光电探测器、2×2 耦合器、2×1 耦合器、传感光纤、延迟光纤、上位机、采集卡、法拉第旋转镜。图示的光路系统中有两条光程相等的干涉光路:

1)2×2 耦合器—1—延迟光纤—3—2×1 耦合器—传感光纤—法拉第旋转镜反射—传感光纤—2×1 耦合器—4—2—2×2 耦合器

2) 2×2 耦合器—2—4—2×1 耦合器—传感光纤—法拉第旋转镜反射—传感光纤—2×1 耦合器—3—延迟光纤—1—2×2 耦合器

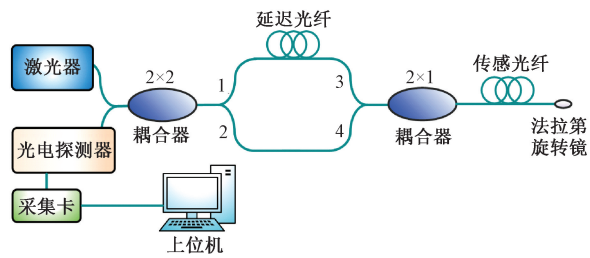

图 1　直线型萨格奈克干涉结构

Fig. 1　Linear Sagnac interference structure diagram

两路光程相等的光路在 2×2 耦合器处发生干涉,通过光电探测器将光强信号输出为电信号至上位机。

声音信号经空气作用在传感光纤上形成振动信号。振动信号在光纤内由于泊松效应、光弹效应、应变效应引起振动作用处的激光相位变化。相位变化量可以表示为:

$$\Delta y(t) = y_2(t) - y_3(t) = 4\sum_{s=1}^{N}\left\{E_s \sin\left(\omega_s \frac{L_d n}{2c}\right)\cos\left(\omega_s \frac{L_2 n}{c}\right)\cdot \left[\cos\omega_s\left(t - \frac{(L_d + 2L)n}{c}\right) + \varphi_s\right]\right\} \tag{1}$$

式中:t 为时间变量,E_s 表示相移幅值,ω_s 是正弦波的频率,φ_s 是正弦波相位初始值,N 表示各频率分量个数,L_d 是延迟光纤的长度,L_2 为声音作用光纤处距离传感光纤末端的距离,传感光纤的长度为 L,n 表示单模光纤折射

率,c 为光速。单个频率分量对应的相位差幅值为:

$$Z = 4E_s \sin\left(\omega_s \frac{L_d n}{2c}\right) \cos\left(\omega_s \frac{L_2 n}{c}\right) \tag{2}$$

可以看出相位差幅值与声波作用引发的激光相移幅值成正比,进而引起干涉光强的变化,之后光电放大探测器将光强信号转为电信号,再输出至上位机进行数据采集与处理。

1.2　维纳滤波

维纳滤波过程中,混有噪声的语音信号可以表示为 $x(n)=s(n)+d(n)$,$s(n)$ 为离散期望信号,$d(n)$为离散噪声信号,n 为自然数。

假设维纳滤波器的单位脉冲响应为 $h(n)$,其输出表示为:

$$\hat{s}(n) = x(n) \cdot h(n) = \sum_{m=-\infty}^{+\infty} x(n-m)h(m) \tag{3}$$

式中:$x(n)$表示输入的混有噪声的信号,$\hat{s}(n)$ 为期望信号的估计。构造维纳滤波器的过程就是使其输出与期望信号最为接近,$s(n)$ 与 $\hat{s}(n)$ 之间的均方误差可以用式(4)来表示:

$$\varepsilon = E[\{s(n) - \hat{s}(n)\}^2] = E\left[\left\{s(n) - \sum_{m=-\infty}^{+\infty} x(n-m)h(m)\right\}^2\right] \tag{4}$$

$h(m)$的数据长度表示为 l,令 ε 对 $h(m)$每一个元素的偏导数等于0:

$$\frac{\partial \varepsilon}{\partial h(i)} = E[-2\{s(n) - \hat{s}(n)\}x(n-i)] = 0 \tag{5}$$

根据式(1)、(3)可得到维纳-霍夫方程:

$$E[s(n)x(n-i)] = \sum_{m=-\infty}^{+\infty} h(m)E\{(n-m)x(n-i)\} \tag{6}$$

式中:$i=1,2,\cdots,l$,共有 l 个方程,对求解得到的 $h(1)$,$h(2)$,$\cdots$,$h(l)$做反 z 变换,就得到维纳滤波器 $H(z)$。

1.3　基音周期的提取

人类通过声带振动发出的声音信号分为清音和浊音,其中浊音具有明显的周期信号特征,浊音的振动频率也就是基音频率,相应的周期就是基音周期。基音周期是人类声纹的重要特征,也是描述语音信号激励源的重要参数。基音的频率变化范围从老年男性的 50 Hz 到成年女性的 500 Hz,频程大导致了基音检测提取的困难,至今没有找到普适通用的检测方法。本文运用三电平削波互相关函数法提取基音周期。

三电平中心削波法的输入输出函数为:

$$d'_i(n) = C'[x_i(n)] = \begin{cases} 1, x_i(n) > C_L \\ 0, |x_i(n)| \leqslant C_L \\ -1, x_i(n) < -C_L \end{cases} \tag{7}$$

首先对语音数据进行分帧,式(7)中 $x_i(n)$表示数据分帧后的第 i 帧离散数据,C_L 表示削波电平值,$d'_i(n)$表示第 i 帧的削波器的输出。一般来说 C_L 的取值方法为找出每一帧前 100 个样本点中最大幅值和后 100 个样本点中最大幅值,将两者较小的那个值乘以系数 0.68。式(8)为中心削波函数的输出 $d_i(n)$:

$$d_i(n) = C(x_i(n)) = \begin{cases} x_i(n) - C_L, x_i(n) > C_L \\ 0, x_i(n) \leqslant C_L \\ x_i(n) + C_L, x_i(n) < -C_L \end{cases} \tag{8}$$

然后求 $d(n)$ 和 $d'(n)$ 的互相关值:

$$S(k) = \sum_{n=1}^{N} d(n) \cdot d'(n+k) \tag{9}$$

式中:$k=0,1,2,\cdots,N/2$,当 $S_{\max} < 0.25S(0)$ 时,这一帧就为清音,基音周期值为0。而 $S_{\max} >= 0.25S(0)$ 时基音周期取最大值 $S_{\max}$ 的位置 k 值,公式表示为 $P = \text{argmax}S(k)$。

1.4　梅尔频率倒谱系数特征提取

梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC),被广泛应用于说话人识别领域。MFCC 特征提取流程如下:

1)对离散语音信号 $X(n)$进行降噪处理。

2)对降噪后的信号分帧加窗,目的是消除各帧左右两端信号不连续问题。所用窗函数为汉明窗,定义如下:

$$W(n) = \begin{cases} 0.54 - 0.46\cos\left(\dfrac{2\pi n}{N-1}\right), 0 \leqslant n \leqslant N-1 \\ 0, \text{其他} \end{cases} \tag{10}$$

3)时域信号 $X(n)$经快速傅里叶变换转化为频谱 $X(k)$。变换公式为:

$$X(k) = \sum_{n=0}^{N-1} X(n) e^{-j2\pi nk/N} (0 \leqslant n, k \leqslant N-1) \tag{11}$$

4)将频谱 $X(k)$通过梅尔滤波器组转换为 Mel 频谱,三角滤波器组的频率响应函数为:

$$H_m(k) = \begin{cases} 0 & (k < f(m-1)) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & (f(m-1) \leqslant k \leqslant f(m)) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & (f(m) < k \leqslant f(m+1)) \\ 0 & (k > f(m+1)) \end{cases} \tag{12}$$

式中:m 指第 m 个梅尔滤波器,$f(m)$为滤波器的中心频率,k 为输入信号的频谱分量。本文实验所用到的三角滤波器组如图2所示,每个不同的颜色表示不同的三角滤波器,每个滤波器的中心频率值如式(15)所示,在中心频率点响应值为1,在两边响应值逐渐衰减至0,因为梅尔频率反应的是人耳对频率感知的灵敏程度,而人耳

对低频段更灵敏,所以滤波器在低频段密集在高频段稀疏,本文系统的采样频率为 16 000 Hz,所以根据奈奎斯特定律最大信号频率为 8 000 Hz,梅尔滤波器组的个数选取通常值 24。

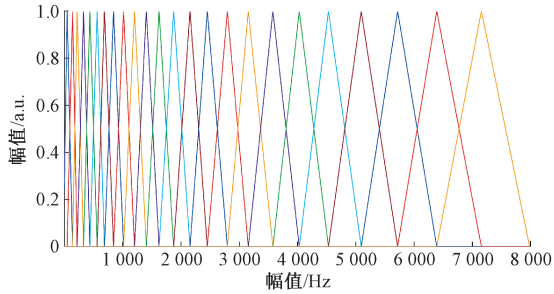

图 2　梅尔滤波器组

Fig. 2　Mel filter bank

梅尔频率与真实频率的关系如式(13)所示:

$$F_{\mathrm{mel}}(f)=2\ 595\times\lg\left(1+\frac{f}{700}\right)$$
$$f=700\left(10^{\frac{F_{\mathrm{mel}}(f)}{2\ 595}}-1\right) \tag{13}$$

式中:$F_{\mathrm{mel}}(f)$ 表示梅尔频率,f 表示真实频率。

所以图 2 中梅尔滤波器组与 X 轴的除 0 外的 25 个交点真实频率为:

$$f_i(i)=F_{\mathrm{mel}}^{-1}\left(F_{\mathrm{mel}}(f_1)+i\frac{F_{\mathrm{mel}}(f_{\mathrm{h}})-F_{\mathrm{mel}}(f_1)}{M+1}\right) \tag{14}$$

其中,最低频率 f_1 为 0,最高频率 f_{h} 为 8 000 Hz,M 表示滤波器组个数为 24,i 表示第几个交点。梅尔滤波器组的中心频率为:

$$f(m)=\frac{f_i(m-1)+f_i(m+1)}{2} \tag{15}$$

5)对梅尔频谱取对数能量,进而转化为对数频谱 $S(m)$,转化公式如下:

$$S(m)=\ln\left(\sum_{k=0}^{N-1}\mid X(k)\mid^2H_m(k)\right),0\leqslant m\leqslant M \tag{16}$$

式中:M 为梅尔滤波器组最大数量。

6)$S(m)$ 经过离散余弦变化,得到梅尔频率倒谱系数参数 $C(n)$:

$$C(n)=\sum_{m=1}^{M-1}S(m)\cos\left(\frac{\pi n(m+1/2)}{M}\right),0\leqslant m<M \tag{17}$$

1.5　动态时间规整算法

DTW 算法用来求解不相等序列特征之间的距离,来衡量两段序列的相似性,可以用于语速不同、时间不同的语音识别。DTW 算法计算过程如下:

1)特征序列 $A=\{a_1,a_2,a_3,\cdots,a_n\}$ 与 $B=\{b_1,b_2,b_3,\cdots,b_n\}$ 之间的距离公式为:

$$\mathrm{dis}(a,b)=(a-b)^2 \tag{18}$$

2)定义累积距离矩阵为 P,用式(19)求取矩阵中的每个元素 $P(x_i,y_j)$ 元素值。

$$D[i,j]=\mathrm{dis}(a_i,b_j)+\min\{D[i-1,j],D[i,j-1],D[i-1,j-1]\} \tag{19}$$

3)在特征序列的所有归一化路径中找到累积距离最小的路径,并计算其最小归一化距离。计算公式为:

$$L=\min[\mathrm{mean}(\sum D[i,j])] \tag{20}$$

求解过程需要满足的条件为:

1)边界条件,特征序列 A 和 B 的首元素和尾元素分别对齐;2)随着步长的增加 A 和 B 序列元素必须单调变化;3)连续性条件,X 和 Y 中所有元素都不可缺少,并且只能与其相邻的点对齐[20]。

1.6　高斯混合模型-期望最大化算法

高斯混合模型表示多个高斯概率密度函数的统计模型,由 M 个多维高斯分布函数加权求和得到,即:

$$d(\boldsymbol{x}_i)=\sum_{i=1}^{M}\beta_i d_i(\boldsymbol{x}_i) \tag{21}$$

其中,M 为高斯混合模型分量的个数,$\boldsymbol{x}_i$ 为特征矢量,$d_i(\boldsymbol{x}_i)$ 的加权系数表示为 β_i,$d_i(x_i)$ 为 S 维高斯分布函数。$d_i(\boldsymbol{x}_i)$ 和 β_i 满足如下两个公式:

$$d_i(\boldsymbol{x}_i)=\frac{1}{(2\pi)^{\frac{S}{2}}\mid\boldsymbol{\Sigma}_i\mid^{\frac{1}{2}}}\exp\left\{-\frac{(\boldsymbol{x}_i-\mu_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_i-\mu_i)}{2}\right\}\ \sum_{i=1}^{M}\beta_i=1 \tag{22}$$

故 GMM 模型可由如下参数集表示:

$\lambda=\{\beta_i,\mu_i,\Sigma_i(i=1,2,\cdots,M)\}$

协方差矩阵 $\boldsymbol{\Sigma}_i$ 可取为对角阵,即:

$\boldsymbol{\Sigma}_i=diag\{\sigma_{i0}^2,\sigma_{i1}^2,\cdots,\sigma_{iS-1}^2\}$

高斯混合模型第 i 个分量对应的第 k 维特征分量的方差表示为:$\sigma_{ik}^2(k=0,1,\cdots,S-1)$。

声纹识别的前提需要事先训练建立各个说话人的模型。这个过程就是估计 GMM 模型的参数。本文采用期望最大化(expectation-maximization,EM)算法进行最大似然估计,EM 算法估计 GMM 模型参数的步骤如下:

1)确定初始权值、均值、协方差矩阵。通过聚类分析法确定均值矢量,协方差矩阵初始化为单位矩阵,每个矢量初始权值为 1/16。

2)对每个特征向量 x_i 用式(22)计算高斯混合分量 $d_j(x_i)$,特征矢量落入隐状态 j 的概率公式为:

$$d(j\mid x_i,\lambda)=\frac{\beta_j d_j(x_i)}{\sum_{k=1}^{M}\beta_k d_k(x_i)} \tag{23}$$

3)用下面3个公式重估模型参数,其中加权系数重估公式为:

$$\beta_j^{(k+1)} = \frac{1}{L}\sum_{i=1}^{L} d^{(k)}(j \mid x_i, \lambda) \tag{24}$$

其中,L表示参与二次迭代的高斯混合模型分量。

均值矢量重估公式为:

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^{L} d^{(k)}(j \mid x_i, \lambda) x_i}{\sum_{i=1}^{L} d^{(k)}(j \mid x_i, \lambda)} \tag{25}$$

方差重估公式为:

$$\sigma_{jk}^2 = \frac{\sum_{i=1}^{L} d^{(k)}(j \mid x_i, \lambda)(x_{ik} - \mu_{jk}^{(k+1)})^2}{\sum_{i=1}^{L} d^{(k)}(j \mid x_i, \lambda)} \tag{26}$$

4)用式(23)更新$d(j \mid x_i, \lambda)$,用下式更新似然函数的期望:

$$F(\lambda, \lambda^E) = \sum_{j=1}^{M}\sum_{i=1}^{L} \log(\beta_j) d(j \mid x_i, \lambda^E) + \sum_{j=1}^{M}\sum_{i=1}^{L} \log(d_j(x_i \mid \lambda_j)) d(j \mid x_i, \lambda^E) \tag{27}$$

5)若当前$F(\lambda, \lambda^E)$与上一次迭代值的差小于设定的收敛域,则迭代结束并且当前的参数估计值即为模型参设,否则跳转步骤2)重复。

1.7　说话人判定

假设待识别的说话人语音特征矢量集为$\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$,说话人为候选人中第$n$个人的后验概率为:

$$P(\lambda_n \mid \boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \lambda_n) P(\lambda_n)}{p(\boldsymbol{X})} = \frac{p(\boldsymbol{X} \mid \lambda_n) P(\lambda_n)}{\sum_{m=1}^{N} p(\boldsymbol{X} \mid \lambda_m) P(\lambda_m)} \tag{28}$$

其中,$p(X)$是特征矢量集$\boldsymbol{X}$的概率密度,对于每个说话人都相等,$p(\boldsymbol{X} \mid \lambda_n)$为第$n$个人条件下,特征矢量$\boldsymbol{X}$的概率密度,$P(\lambda_n)$为第$n$个人的先验概率,可假设两个待识别人说话的概率相同都为$1/n$,所以判定结果等价于:

$$n^* = \arg\max_{1 \leqslant n \leqslant N} [p(\boldsymbol{X} \mid \lambda_n)] \tag{29}$$

为了简化计算一般采用对数似然函数:

$$L(X \mid \lambda_n) = \ln p(\boldsymbol{X} \mid \lambda_n) \tag{30}$$

所以判定公式为:

$$n^* = \arg\max_{1 \leqslant n \leqslant N} [L(\boldsymbol{X} \mid \lambda_n)] \tag{31}$$

最大对数似然值对应的说话人即为声纹识别确定的说话人。

2　实验方案设计与分析

按照图1所示光路系统搭建图3所示的实验系统,加入隔离器的目的是避免散射光和反射光对激光器造成影响。其中,光源为1 550 nm宽带激光器,延迟光纤为3 000 m,传感光纤为10 km,探头为200 m裸纤内置法拉第旋转镜。3×3耦合器的两路光作为光电探测器的输入端,输出连续电信号。说话人距探头2 m范围内说话,声音作用于传感光纤后引起激光信号变化,光信号经探头内法拉第旋转镜反射,在3×3耦合器处发生干涉,干涉光信号经过光电探测模块转为连续电信号,采集卡采样频率为16 kHz,采集卡将连续电信号转变为离散数字电信号输入至上位机,上位机进行信号存储与处理。

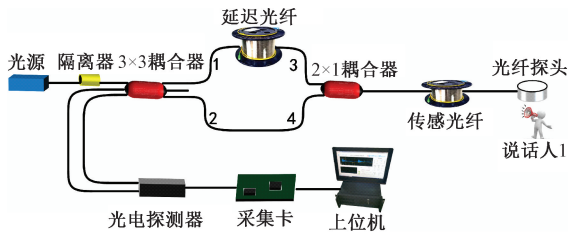

图3　实验系统示意图

Fig. 3　Experimental system diagram

距光纤探头2 m范围内使用扬声器播放频率为1 kHz输出声压为94 dB的正弦信号音频,之后保持其他变量不变通过更换光纤探头长度然后观察采集信号的变化。图4为光纤探头长度为50、100、150、200 m时采集信号的波形变化,可以看出除了光纤探头长度为50 m时信号严重失真,采集信号波形幅值随着光纤探头长度增加而增加,分别为158.4、190.4、261.5、380.3 mV。由于系统为了面向实际工程应用,将裸纤封装在直径大约10 cm的圆柱体机械结构中,在有限的空间内光纤探头的最大长度为200 m,这是本文光纤探头长度选取为200 m的原因。同时图3中传感光纤长度10 km是为了配合煤矿井下巷道长度而进行的选取。

图5所示为声纹识别流程图,声纹识别的前提条件需要一段说话人的语音作为训练语音并且存入数据库中,说话人的其他语音为识别语音。本文的说话人识别就是建立说话人识别数据库,之后判断识别语音属于数据库中的哪个说话人。训练语音通过维纳滤波降噪,然后提取特征参数基音周期(pitch)和MFCC,其中MFCC特征参数通过EM算法得到高斯混合模型参数:加权系数(prob),均值(mean),协方差对角矩阵(cov),然后将说话人姓名连同pitch、mean、cov、prob存入数据库中。识别语音也通过降噪之后获得pitch以及MFCC,第1阶段使用DTW算法计算该识别语音pitch与数据库中所有样本pitch间距离,筛选出距离相对小的35%的样本。第2阶段,计算识别语音MFCC特征矢量在筛选出的样本GMM模型下的似然概率,得出似然概率最大的样本就是说话人。

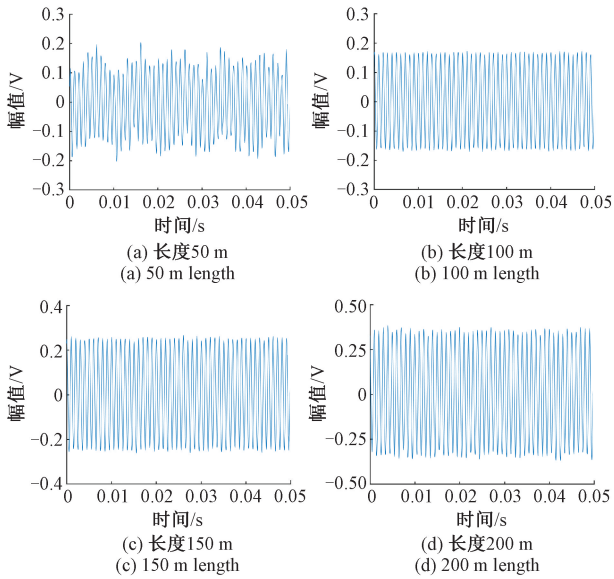

图 4　光纤探头长度影响

Fig. 4　Influence of fiber probe length

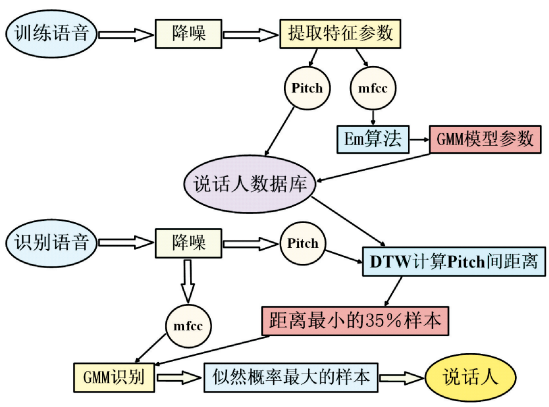

图 5　声纹识别流程

Fig. 5　Voiceprint recognition process

本文声纹识别方法选用 pitch 特征筛选说话人样本的原因在于当数据库注册说话人数量达到 100 人以上时，声纹识别第 2 阶段识别过程较第 1 阶段识别过程花费几乎多 1 倍的时间。而大多数情况下相同人之间 pitch 距离比不同人之间的 pitch 距离小至少 0.005。在 21 组识别错误的语音中错误原因为第 1 阶段筛选 35%样本时错误将正确说话人筛掉的有 4 组，并且将这 4 组语音跳过第 1 阶段筛选直接进行第 2 阶段识别时也依旧识别错误。所以利用 pitch 特征筛选 35%说话人样本节省识别时间的同时并没有让系统性能减少。DTW 算法是较欧式距离更准确的衡量时间序列相似性的方法，pitch 距离的比较筛选实际上是衡量识别语音 pitch 特征与数据库中注册 pitch 特征的相似性问题[21]，这是本文选用 DTW 算法的原因。

高斯混合模型以及对应的期望最大化算法是本文声纹识别的核心算法，其中高斯混合模型被广泛应用于信息识别领域，而期望最大化算法是一种有效求解高斯混合模型参数最常用的算法，这一算法在技术上是成熟的，输出结果稳定可预期。MFCC 特征维数为 16 维的原因在于 GMM 模型高斯分量个数 M 是根据数据的时间长度来确定的，一般数据长度小于 30 s 取 $M=16$，大于 30 s 取 $M=32$[22]。

3　实验结果分析

3.1　系统频率响应特性分析

由于本文主要运用光纤声音传感系统采集人声，国际上指定的数字电话机的通信标准为 300～3 400 Hz，而人讲话的频率主要集中在 1～3 kHz 之间。

本文对于光纤声音传感系统各频率段的响应度的测量如图 6 所示，使用扬声器在光纤探头 2 m 范围内播放单频声音信号，分别测量了 300、800、1 000、1 500、2 000、2 500、3 000、3 500Hz 频率响应波形，其幅值分别为 45.9、120.5、368.7、987.5、336.8、239.5、254.0、285.6 mV。光纤声音传感系统的各频率的响应度可以表示为传感器输出电压值与声源声压值的比值，图 6 中各信号强度分别为 78、91、94、102、97、95、97、98 dB，计算结果为：288.9、169.8、368.7、392.2、237.9、212.9、179.4、179.8 mV/Pa。

直线型萨格奈克声音传感系统对于不同频率段的响应度存在差异，这种差异还体现在采集到的语音信号 MFCC 特征与纯净语音相比不同。如图 7 所示为纯净语音的 MFCC 特征，如图 8 所示为系统采集语音信号的 MFCC 特征。两者对比可以发现，纯净语音的 MFCC 特征表现为第 1 维 MFCC 值过大，系统采集语音的 MFCC 特征表现为前 5 维 MFCC 值都过大。图 2 梅尔滤波器组的个数为 24 可知 MFCC 特征维度最大为 24 维，而 MFCC 维数越低相应的数值越大，同时受系统频率响应特征的影响也越大，所以为了解决系统频率响应度不同带来的特征差异问题，本文声纹识别使用的 16 维 MFCC 特征维度为第 9 维到第 24 维，同时数据库注册语音选用声音传感系统采集的语音。

3.2　系统声纹识别结果分析

假设数据库中注册有 4 位测试人员时，4 位测试人员的注册语音 pitch 与 MFCC 如图 9 和 10 所示。根据图 5 声纹识别流程，将 4 位测试人员的一段注册语音的 pitch 以及用 MFCC 特征获得的 mean、cov、prob 存入数据库中并且将其分别命名为测试人员 1、测试人员 2、测试人员 3、测试人员 4。

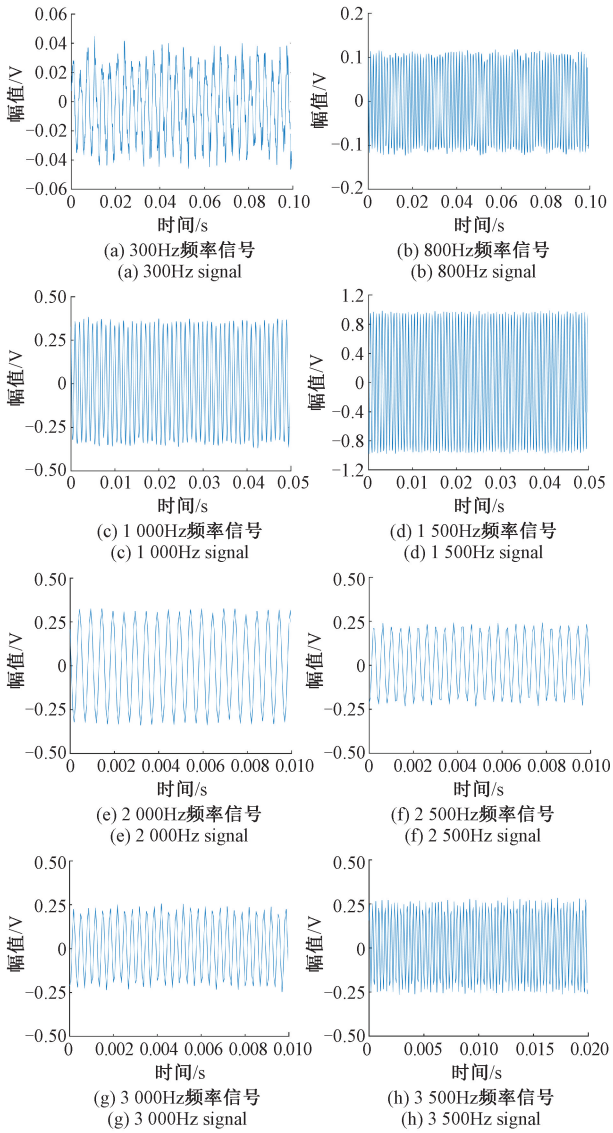

图 6　各频率段采集信号波形

Fig. 6　Waveforms of different frequency band acquisition

在距光纤探头 2 m 范围内利用扬声器播放音频信号，扬声器的输出分贝为 60~75 dB，在正常人说话分贝范围内。然后通过光纤声传感器采集声音信号，再通过维纳滤波提高信号信噪比。声纹识别系统将维纳滤波算法作为一种语音预处理方法，注册语音以及识别语音使用的都是预处理之后的语音。

识别语音 1 通过维纳滤波降噪前后的波形如图 11 所示，降噪前信号幅值取最大幅值的绝对值，为 0.836 6，噪声信号取无人说话段信号最大幅值的绝对值，为 0.045 47，降噪前 SNR 为 25.30 dB。同理，降噪后 SNR 为 35.65 dB。维纳滤波使信号 SNR 提升了 10.35 dB。提取降噪后识别语音 1 的基音周期与 MFCC 特征，如图 12 所示。

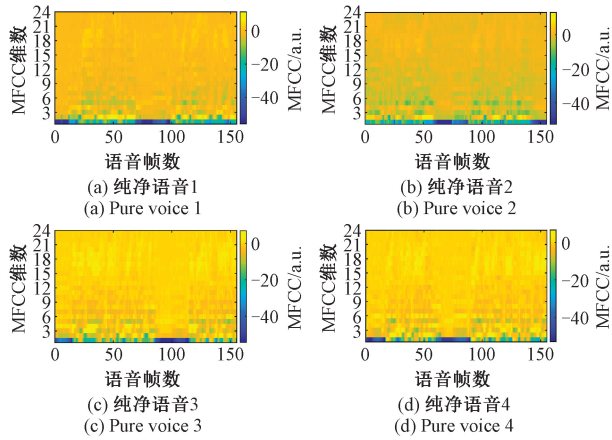

图 7　纯净语音的 MFCC 特征

Fig. 7　MFCC features of pure speech

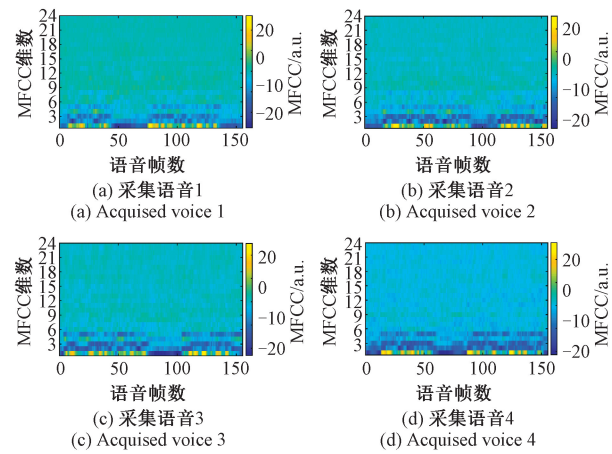

图 8　系统采集信号 MFCC 特征

Fig. 8　MFCC characteristics of system acquisition signal

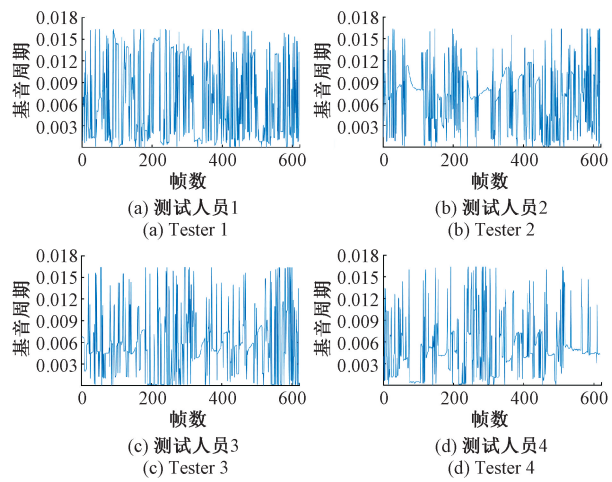

图 9　数据库中 4 位注册人的 pitch 特征

Fig. 9　Pitch characteristics of 4 registrants in the database

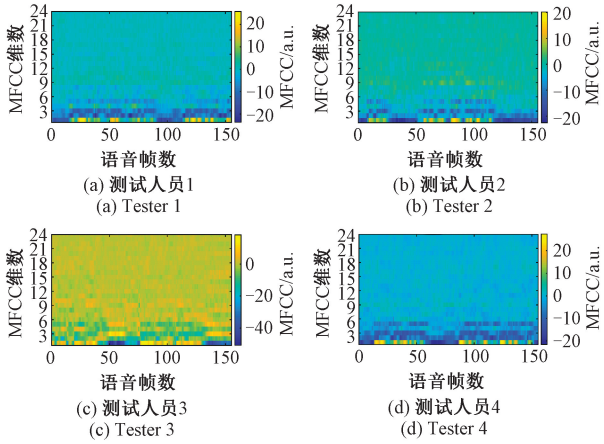

图 10　数据库中 4 位注册人的 MFCC 特征

Fig. 10　MFCC characteristics of 4 registrants in the database

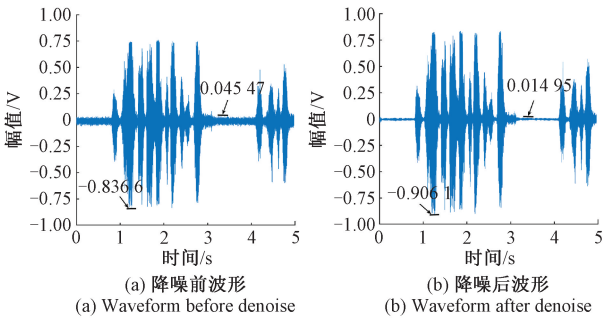

图 11　识别语音 1 的维纳滤波前后波形

Fig. 11　Waveforms of speech 1 to be recognized before and after wiener filtering noise reduction

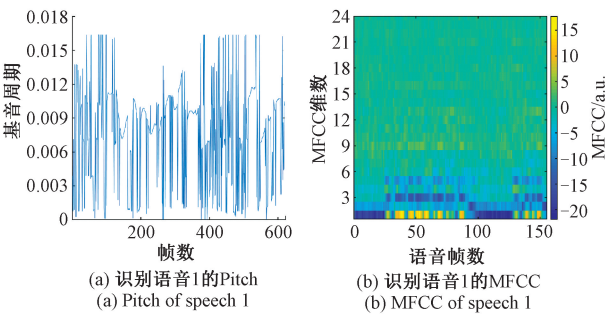

图 12　降噪后识别语音 1 的 pitch 与 MFCC 特征

Fig. 12　Pitch and MFCC features of speech 1 to be recognized after noise reduction

识别第 1 阶段通过 DTW 算法计算识别语音 1 的 pitch 与图 9 所示数据库中 4 位注册人的 pitch 特征之间的距离，测试人员 1、测试人员 2、测试人员 3、测试人员 4 的距离分别为 0. 173 6、0. 148 3、0. 160 0、0. 162 5，通过筛选距离小的两个数据库说话人测试人员 2、测试人员 3 进入识别第 2 阶段。

先经过第 1 阶段 pitch 周期筛选后，根据式(30)，利用识别语音 1 的第 9 维到第 24 维 MFCC 特征作为矢量特征分别计算在测试人员 2 的 GMM 模型与测试人员 3 的 GMM 模型中的对数似然值，分别为 −1 491. 1 和 −1 662. 9，测试人员 2 对应的对数似然值更大，所以最终确定说话人为测试人员 2。

在声纹识别中 A 的语音被错误识别为 B 的情况主要有两种，第 1 种为当声纹识别第 1 阶段筛选数据库注册说话人时将正确说话人错误筛掉；第 2 种为第 2 阶段计算识别语音 MFCC 在筛选后的数据库各注册说话人 GMM 模型下的对数似然值时，对数似然值最大的并不是说话人的情况。

图 13 为识别语音 2 预处理后的波形图，识别语音 2 来自与说话人测试人员 3。第 1 阶段计算其与数据库中测试人员 1、测试人员 2、测试人员 3、测试人员 4 的 pitch 距离分别为 0. 171 8、0. 149 3、0. 155 8、0. 148 0。而筛选进入第 2 阶段的为测试人员 2 和测试人员 4，然而识别语音 2 实际来源于测试人员 3，所以识别错误。

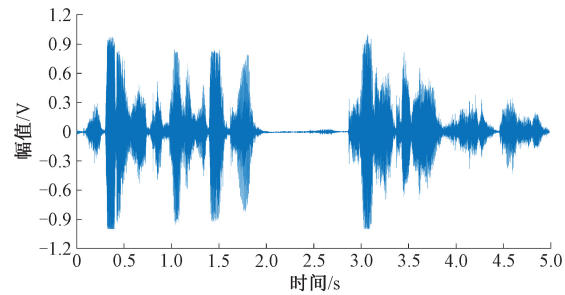

图 13　识别语音 2 的波形

Fig. 13　Waveform of speech 2 to be recognized

图 14 为识别语音 3 预处理后的波形图，识别语音 3 实际来源于测试人员 2，第 1 阶段分别计算识别语音 pitch 与数据库中注册说话人测试人员 1、测试人员 2、测试人员 3、测试人员 4pitch 之间的距离为 0. 171 8、0. 149 3、0. 155 8、0. 148 0。筛选进入第 2 阶段的为测试人员 2、测试人员 4。

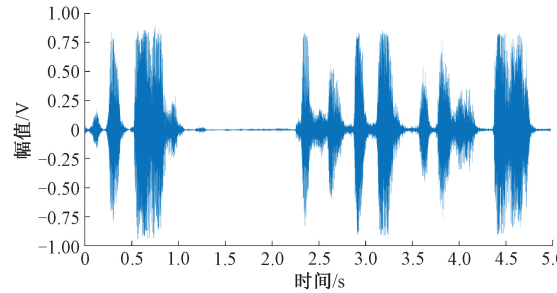

图 14　识别语音 3 的波形

Fig. 14　Waveform of speech 3 to be recognized

识别第 2 阶段分别计算识别语音 3 的 MFCC 特征在测试人员 2 的 GMM 模型与测试人员 4 的 GMM 模型中的对数似然值,分别为-1 817.9 和-1 550.3。对数似然值大的说话人为测试人员 4,而实际上识别语音 3 来源于测试人员 2,所以识别错误。

陌生人语音识别一般指的是非数据库中说话人语音的识别。本文的声纹识别过程需要经历两个阶段,第 1 阶段 pitch 距离的筛选并无法区分陌生人。而对于陌生人语音识别来说,在第 2 阶段最大对数似然值对应的说话人是经过整个声纹识别筛选之后数据库中最接近的说话人。

通过对于声纹识别第 2 阶段最大的对数似然值的观察来发现规律,将识别语音最大对数似然值除以 MFCC 特征帧数定义为每帧最大对数似然值,图 15 为 24 组实验每帧最大对数似然值,其中前 12 组实验为非数据库中说话人语音,从图中可以看出非数据库中说话人语音的每帧最大对数似然值都小于-10.5,后 12 组实验为可以正确识别的数据库中说话人语音,其每帧最大对数似然值都大于-10.5。所以在误差允许的范围内当识别语音第 2 阶段每帧最大对数似然值小于-10.5 时,将识别语音定义为非数据库中说话人,也就是陌生人。

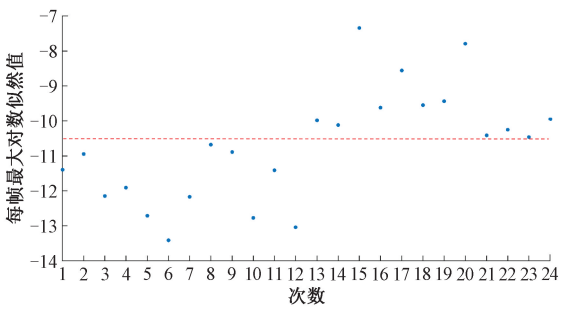

图 15 24 组实验每帧最大对数似然值

Fig. 15 Maximum log-likelihood value per frame in 24 experiments

为了验证系统声纹识别综合准确率,进行了 400 段时长为 3~5 s 之间的文本无关语音段的识别,最终有 379 段语音正确识别出说话人,所以综合识别准确率为 94.75%。本文的 21 组识别错误的语音中,其中 A 的声音被错误识别为 B 的有 13 组,概率为 3.25%;被错误识别为陌生人的有 8 组,概率为 2%。

3.3 声音幅值对识别结果的影响分析

直线型萨格奈克声音传感系统采集声音信号的幅值对声纹识别成功率有影响,这种影响主要因为系统拥有固有噪声,如图 11 所示系统采集信号无人说话段的幅值为 0.045 47。图 16 为一段识别语音随着幅值的降低,其在识别阶段最大似然值与次大似然值之差的变化过程。在图 16 中 P1 点位置幅值为 0.403 5 V 时,次大似然值对应的说话人改变;在 P2 点位置幅值为 0.151 5 V 时,最大似然值对应的说话人改变,也就是识别结果改变。

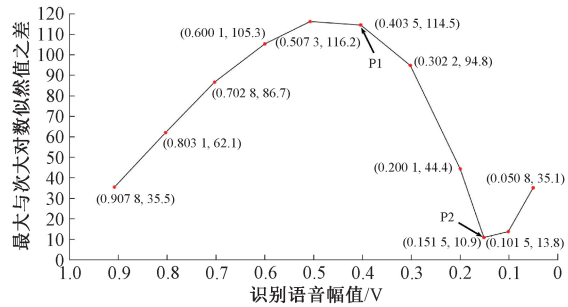

图 16 识别语音幅值对识别结果的影响

Fig. 16 Influence of the amplitude of speech to be recognized on the recognition result

4 结 论

本文介绍了直线型萨格奈克干涉光纤传感系统的工作原理以及 DTW-GMM 声纹识别原理。构建了直线型萨格奈克干涉光纤声音传感系统,利用维纳滤波算法将数据信噪比提升 10.35 dB,通过三电平削波法获取采集语音的基音周期特征,利用动态时间规整算法筛选 35%说话人样本,进一步提取 MFCC 特征,然后运用 GMM-EM 算法最终确定说话人。并进行了 400 段语音的说话人识别成功率实验。结果表明,传感光纤 10 km,距光纤探头 2 m 内,在语音段时长范围在 3~5 s,传感系统的语音综合识别成功率为 94.75%,有望应用于易燃易爆环境中的设备故障、应急救援、渗漏监测等声纹检测及识别领域。

参考文献

[1] 康志坚,张红娟,高妍,等. 煤矿光纤传感应急通信系统设计[J]. 工矿自动化,2020,46(11):72-76,82.
KANG ZH J, ZHANG H J, GAO Y, et al. Design of emergency communication system for optical fiber sensing in coal mine[J]. Industrial and Mine Automation, 2019, 46(11):72-76,82.

[2] 畅楠琪,黄晓砥,王海斌. 基于 EKF 参数估计的光纤水听器 PGC 解调方法研究[J]. 中国激光, 2022, 49(17):118-129.
CHANG N Q, HUANG X J, WANG H B. Research on PGC demodulation method of fiber optic hydrophone based on EKF parameter estimation[J]. Chinese Journal of Lasers, 2022, 49(17):118-129.

[3] 寇华东,王伟君,闫坤,等. 分布式光纤声波传感背景噪声近地表成像:在北京房山的应用[J]. 地震,2023,43(3):50-65.

KOU H D, WANG W J, YAN K, et al. Near-surface imaging of background noise by distributed optical fiber acoustic sensing: application in Fangshan, Beijing[J]. Earthquake, 2019, 43(3):50-65.

[4] 左一武,田晶,杨清,等. 一种基于大角度倾斜光纤光栅包层模的低频声传感方案[J]. 物理学报,2023,72(12):176-182.
ZUO Y W, TIAN J, YANG Q, et al. A low frequency acoustic sensing scheme based on large-angle inclined fiber grating cladding mode[J]. Acta Physica Sinica, 2023, 72(12):176-182.

[5] 余双勇,衣文索,陈昊玥,等. 分布式声传感型特种光纤结构设计[J]. 仪器仪表学报,2021,42(3):59-69.
YU SH Y, YI W S, CHEN H Y, et al. Structure design of special fiber based on distributed sound sensing[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 59-69.

[6] 王照勇,卢斌,叶蕾,等. 分布式光纤声波传感及其地震波检测应用[J]. 激光与光电子学进展,2021,58(13):83-94.
WANG ZH Y, LU B, YE L, et al. Distributed optical fiber acoustic wave sensing and its application in seismic wave detection[J]. Laser and Optoelectronics Progress, 2019, 58(13):83-94.

[7] 宋晓达,王宇,张建国,等. 矿用双通道光纤拾音系统设计[J]. 电子测量与仪器学报,2018,32(4):43-50.
SONG X D, WANG Y, ZHANG J G, et al. Design of dual-channel optical fiber pickup system for mining[J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(4):43-50.

[8] 尹玺,万生鹏,熊新中,等. 基于法布里-珀罗微腔结构的光纤声传感系统研究[J]. 激光与光电子学进展,2021, 58(3):180-186.
YIN X, WAN SH P, XIONG X ZH, et al. Research on fiber optic acoustic sensing system based on Fabry-Perot microcavity structure [J]. Progress in Laser and Optoelectronics, 2021, 58 (3): 180-186.

[9] 吴虎,孔勇,王振伟,等. 基于端点检测与信号重组的光纤分布式传感信号识别[J]. 光子学报, 2021, 50(11):123-130.
WU H, KONG Y, WANG ZH W, et al. Signal recognition of fiber distributed sensing based on endpoint detection and signal recombination[J]. Acta Photonica Sinica, 2021, 50(11):123-130.

[10] 郑方,李蓝天,张慧,等. 声纹识别技术及其应用现状[J]. 信息安全研究,2016,2(1):44-57.
ZHENG F, LI L T, ZHANG H, et al. Voiceprint recognition technology and its application status [J]. Information Security Research, 2016, 2(1):44-57.

[11] 刘晓晨,潘孝勤,曹金璇,等. 声纹识别和语音识别技术在公安领域的应用[J]. 网络安全技术与应用,2021,244(4):153-155.
LIU X CH, PAN X Q, CAO J X, et al. Application of voice print recognition and speech recognition technology in the field of public security [J]. Network Security Technology and Application, 2021, 244(4):153-155.

[12] REYNOLDS D A, QUATIERI T F, DUNN R B. Speaker verification using adapted Gaussian mixture models[J]. Digital Signal Processing, 2000, 10(1-3):19-41.

[13] DEHAK N, KENNY P, DEHAK R, et al. Front-end factor analysis for speaker verification[J]. IEEE Trans on Audio Speech and Language Processing, 2011, 19(4):788-798.

[14] SNYDER D, GARCIA-ROMERO D, SELL G, et al. X-vectors: Robust DNN embeddings for speaker recognition[C]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[15] 王丰华,王邵菁,陈颂,等. 基于改进 MFCC 和 VQ 的变压器声纹识别模型[J]. 中国电机工程学报,2017,37(5):1535-1543.
WANG F H, WANG SH J, CHEN S, et al. Transformer voiceprint recognition model based on Improved MFCC and VQ[J]. Proceedings of the CSEE, 2017, 37(5): 1535-1543.

[16] 崔佳嘉,马宏忠. 基于改进 MFCC 和 3D-CNN 的变压器铁心松动故障声纹识别模型[J]. 电机与控制学报,2022,26(12):150-160.
CUI J J, MA H ZH. Voiceprint recognition model for transformer core Loose fault based on improved MFCC and 3D-CNN[J]. Electric Machines and Control, 2022, 26(12):150-160.

[17] 王涛,王国中,朱林林. 一种基于声纹识别的智能门锁系统设计与实现[J]. 电子测量技术,2019,42(3):107-111.
WANG T, WANG G ZH, ZHU L L. Design and implementation of Intelligent door lock system based on voiceprint recognition [J]. Electronic Measurement Technology, 2019, 42(3):107-111.

[18] 苏学军,谢存祥,于文龙. 基于 STM32 的语音声纹识别系统设计[J]. 电子测量技术,2020,43(24):1-5.
SU X J, XIE C X, YU W L. Design of voiceprint recognition system based on STM32 [J]. Electronic Measurement Technology, 2019, 43(24):1-5.

[19] 李昊霖,吴苑菲. 基于声纹识别技术精准打击电信网络诈骗犯罪研究[J]. 网络安全技术与应用,2023,269(5):141-144.

LI H L, WU Y F. Research on accurate suppression of telecom network fraud based on voiceprint recognition technology [J]. Network Security Technology and Application, 2023, 269(5):141-144.

[20] 侯德华,张庆,李忠玉,等. 基于 DTW 算法的复合改性沥青相容性评价研究[J]. 化工新型材料, 2023, 51(S1):191-196.
HOU D H, ZHANG Q, LI ZH Y, et al. Compatibility evaluation of composite modified asphalt based on DTW algorithm[J]. New Chemical Materials, 2023, 51(S1): 191-196.

[21] 谭珊,赵仲勇,杨建,等. 基于动态时间规整的变压器绕组变形故障诊断方法研究[J]. 高压电器, 2024, 60(2):108-118.
TAN SH, ZHAO ZH Y, YANG J, et al. Research on fault diagnosis method of transformer winding deformation based on dynamic time warping [J]. High Voltage Electrical Apparatus, 2024, 60(2):108-118.

[22] 刘永红. 说话人识别系统的研究[D]. 成都:西南交通大学, 2003.
LIU Y H. Research on speaker recognition system[D]. Chengdu:Southwest Jiaotong University, 2003.

作者简介

杨佳沛, 2020 年于太原理工大学获得学士学位,现为太原理工大学硕士研究生,主要研究方向为光纤传感器。
E-mail: 1256972787@qq.com

Yang Jiapei received his B. Sc. degree in 2020 from Taiyuan University of Technology. Now he is a M. Sc. candidate in Taiyuan University of Technology. His main research interest includes fiber optic sensors.

王宇(通信作者), 2014 年于法国塞吉-蓬图瓦兹大学获得博士学位,现为太原理工大学副教授,主要研究方向为光纤传感器。
E-mail: wangyu@tyut.edu.cn

Wang Yu (Corresponding author) received his Ph. D. degree in 2014 from Cergy-Pontoise University. Now he is an associate professor in Taiyuan University of Technology. His main research interest includes fiber optic sensors.