第14卷第1期　　空军工程大学学报(自然科学版)　　Vol. 14 No. 1
2013年2月　　JOURNAL OF AIR FORCE ENGINEERING UNIVERSITY(NATURAL SCIENCE EDITION)　　Feb. 2013

基于小波变换的说话人识别技术

檀蕊莲[1,2]，柏　鹏[2]，李　哲[2]，姚战宏[2]，栾前进[3]
(1.武警工程大学信息工程系，陕西西安，710086；
2.空军工程大学装备发展与运用研究中心，陕西西安，710051；
3.总参陆航部驻西安地区军代室，陕西西安，710000)

摘要　说话人识别技术是通过判断待识别人语音与预先提取的说话人语音特征是否匹配来鉴别说话人身份的一种生物认证技术，环境噪声是说话人识别技术走向实用化的一个主要障碍。针对噪声环境中说话人识别性能较差的不足，结合小波变换的优点，提出了将小波变换技术与传统的特征参数提取方式相结合的方法。该方法首先对语音信号进行小波分解，在此基础上再对小波系数进行阈值处理，仅保留阈值以上的数据，而后提取相关性不大的传统特征参数进行组合，分别作为说话人识别系统的输入矢量。仿真结果表明：在噪声环境中，说话人识别系统能较好识别出说话人，经过小波变换后再提取特征参数的方法可以得到更高的识别率，大大提高说话人识别系统的识别性能。

关键词　说话人识别；动态时间规整；矢量量化；小波变换；组合特征
DOI　10.3969/j.issn.1009-3516.2013.01.019
中图分类号　TN912.3　　**文献标志码**　A　　**文章编号**　1009-3516(2013)01-0085-05

Research on Speaker Recognition Based on Wavelet Transform

TAN Rui-lian[1,2], BAI Peng[2], LI Zhe[2], YAO Zhan-hong[2], LUAN Qian-jin[3]
(1. Information Engineering Department, Armed Police Engineering University, Xi'an 710086, China; 2. Research Center of Development and Application of Equipment, Air Force Engineering University, Xi'an 710051, China; 3. Military Representative Office of the General Staff of Army aviation in Xi'an, Xi'an 710000, China)

Abstract: Speaker recognition is a kind of biological authentication technology which distinguishes speakers' identity by matching the voice distilled beforehand. However, the noise circumstance is an obstacle disturbing this technology walking up to practicality. Concerning the shortcoming of poor speaker recognition performance in noisy environments and combining the advantages of wavelet transform, a method of combining the wavelet transform technology with the traditional characteristic parameter extraction mode is proposed. In this method, the speech signal is decomposed by the wavelet, and then wavelet coefficients are processed by threshold. Only the data above the threshold are retained. The traditional characteristic parameters of little correlation are extracted to use as the input vector of the speaker recognition system. The simulation results indicate that the use of the method can better identify the speaker. A higher recognition rate can be obtained through the wavelet transform first and then the extraction of characteristic pa-

收稿日期：2012-08-30
基金项目：国家自然科学基金资助项目(61174194)
作者简介：檀蕊莲(1982—)，女，广西钦州人，讲师，博士生，主要从事信号处理、数据链应用与研究.
E-mail：tanruilian_19821225@yahoo.com.cn

rameters. The application of this method greatly improves the performance of the speaker recognition system

Key words: speaker recognition; dynamic time warping; vector quantization; wavelet transform; feature combination

说话人识别,即根据输入语音确定发音者的身份,是利用生物特征进行身份鉴别和认证的方法之一,是一种高效的人机交互,身份识别及信息检索手段,广泛应用于银行、工厂等领域[1]。目前,已有多种快速辨认说话人的算法[2]。实验表明,说话人识别系统在应用环境和训练环境匹配的情况下可以得到令人满意的结果[3],然而将在安静环境下训练的模型应用于实际有背景噪声的环境中,系统的识别性能就会明显下降,环境噪声已经成为说话人识别技术走向实用化的主要障碍之一。寻找更加有效的强抗噪性能的说话人识别特征参数是说话人识别研究的热点。

目前,小波分析主要应用于语音编码、端点检测、基音周期提取等方面,在说话人识别技术中,用小波分析来提取特征参数还处于研究阶段。小波变换是一种具有分辨率可变,实现简单和无平稳要求等优点的时频局部分析方法,能同时在时、频域中对信号进行分析。研究表明,语音信号的有用信息主要集中在低频部分,高频部分蕴含着大部分的噪声[4],小波变换在各频段的恒 Q(品质因数)特性与人耳听觉对信号的加工特点相一致[5]。生理学研究显示,对听觉起关键作用的耳蜗内基底膜,其作用相当于一组建立在薄膜振动基础上的恒 Q 带通滤波器。小波变换的这一良好特性为利用小波变换提取语音特征参数奠定了基础。

研究表明[6-7],直接利用小波系数作为特征参数,其识别率较低,识别性能较差,但具有较好抗噪声性能,因此难点主要在于如何把直接小波系数转化为代表说话人个性特征的参数。本文重点分析说话人特征参数的提取问题,结合小波变换的特点,提出了在噪声环境中更为有效的组合特征参数提取方法,可大大提高系统的识别性能,具有很好的推广性。

1 小波变换原理

如果函数 $\psi(t)\in L^2(R)$,并且满足允许性条件(完全重构条件或恒等分辨条件)[8]:

$$C_\psi=\int_R\frac{|\hat{\psi}(w)|^2}{|w|}\mathrm{d}w<\infty \tag{1}$$

则称 $\psi(t)$ 是一个基本小波或母小波(Mother Wavelet),$\psi(t)$ 定下来后,通过母函数的伸缩(Dilation)和平移(Translation)可得:

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)\quad a,b\in R;a\neq 0 \tag{2}$$

上式称为一个小波序列。式中:a 为伸缩因子,b 为平移因子。

对于任意的函数 $f(t)$ 在 $L^2(R)$ 上的连续小波变换定义为:

$$W_f(a,b)=\langle f,\psi_{a,b}\rangle=|a|^{-1/2}\int_R f(t)\,\overline{\psi\left(\frac{t-b}{a}\right)}\mathrm{d}t \tag{3}$$

其重构公式(逆变换)为:

$$f(t)=\frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{|a|^2}W_f(a,b)\psi\left(\frac{t-b}{a}\right)\mathrm{d}a\mathrm{d}b \tag{4}$$

连续小波变换主要用于理论分析方面,在实际运用中,尤其是在计算机上实现,离散小波变换更适于计算机处理,因此,连续小波必须加以离散化。离散小波定义为:

$$\psi_{j,k}(t)=a_0^{-j/2}\,\psi\left(\frac{t-ka_0^j b_0}{a_0^j}\right)=a_0^{-j/2}\,\psi(a_0^{-j}t-kb_0) \tag{5}$$

离散化小波变换系数可表示为:

$$C_{j,k}=\int_{-\infty}^{\infty}f(t)\psi_{j,k}^*(t)\mathrm{d}t=\langle f,\psi_{j,k}(t)\rangle \tag{6}$$

其重构公式为:

$$f(t)=C\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}C_{j,k}\psi_{j,k}(t) \tag{7}$$

式中 C 是一个与信号无关的常数。

2 基于小波变换的组合特征提取

2.1 思路分析

现有的说话人特征提取算法大多采用基于语音信号短时平稳假设而提取的短时声道参数,如 Mel 频率倒谱系数(MFCC)和线性预测倒谱系数(LPCC)及其差分倒谱系数(ΔMFCC 或 ΔLPCC)等。MFCC 是基于人耳的听觉特性提取的,在噪声环境中也能有较好的表现;LPCC 能够很好地体现每个人特定的声道运动特性;差分倒谱系数在一定程度上反映了说话人的动态特性。然而,文献[1]指

出,语音信号是一种典型的非平稳信号,短时傅里叶变换仅仅是将信号在频域中展开,并不能给出信号在某个时间点上的变化情况,因而容易产生误差和遗漏。因此,基于上述方法提取的语音特征参数仅仅对说话人的静态特征进行了描述,忽略了说话人的动态特征。

小波分析是一种信号的时间-尺度分析方法,它具有多分辨分析的特点,是一种时间窗与频率窗都可改变的时频局部化分析方法。在低频时有较高的频率分辨率和较低的时间分辨率,在高频时有较低的频率分辨率和较高的时间分辨率,因而小波变换更适合处理语音、图像等非平稳信号。文献[8]指出,语音信号经过小波变换后,可以在各尺度空间提供不同频率的语音信号构成信息,从而精确检测到声门闭合产生的语音波形突变点,对说话人的动态特征进行了描述。将小波变换方法与传统特征提取方法结合使用,可以更全面体现说话人的个性特点。

由于低阶倒谱参数分量包含的说话人个性信息不多,而高阶倒谱参数分量幅度很小,因此对系统性能的改善作用有限,为了突出对说话人识别有效分量的作用,需要提升某些阶次的倒谱系数来突出说话人信息,本文对所提取的高阶倒谱系数进行加权,加权系数取升正弦窗函数,即:

$$w_n=\begin{cases}1+h\sin(n\pi/m), & n=1,2,\cdots,L\\ 0, & \text{其他}\end{cases} \tag{8}$$

式中:n 为实际帧长;L 为窗的长度;m 取 $L-1$,h 取 $\omega/2$。

2.2　改进的组合特征参数提取方法

具体步骤如下:

Step1　小波分解:将经过预处理的语音信号,用 MATLAB 算法进行小波变换,求出各尺度上的小波系数,本文选用 db6 小波,分解级数为 3 层。

Step2　阈值处理:该处理主要是针对低频部分的小波系数用门限进行。用该阈值来甄别携带能量较少的系数,即小于或等于该阈值的小波系数将其视为携带能量较少者,实际处理中将这些值作为零来处理,而仅仅保留阈值以上的数据,具体处理如下:

$$x(t)=\begin{cases}\operatorname{sgn}[x(t)](|x(t)|-\delta), & |x(t)|>\delta\\ 0 & |x(t)|\leqslant\delta\end{cases} \quad t=1,2,\cdots,N \tag{9}$$

式中:N 为语音序列的长度;sgn[]为符号函数;δ 为阈值,采用 $\delta=a\sigma$ 来估计,其中 a 为一常数,取 0.3;σ 为小波系数的标准方差,可通过下面的式子来估计:

$$\sigma^2=\frac{1}{N/2}\sum_{i=1}^{N/2}(x_i-\bar{x})^2,\quad i=1,2,\cdots,N/2 \tag{10}$$

式中:$\{x_i,i=1,2,\cdots,N/2\}$为小波系数;$\bar{x}$ 为均值。

Step3　组合特征参数提取:

方案 1:分别对低频部分进行 m 阶的 LPCC 特征参数提取,对高频部分进行 n 阶的 ΔLPCC 特征参数提取。

方案 2:分别对低频部分进行 m 阶的 MFCC 特征参数提取,对高频部分进行 n 阶的 ΔMFCC 特征参数提取。

方案 3:分别对低频部分进行 m 阶的 LPCC+ΔLPCC 特征参数提取,对高频部分进行 n 阶的 MFCC+ΔMFCC 特征参数提取。

方案 4:分别对低频部分进行 m 阶的 MFCC+ΔMFCC 特征参数提取,对高频部分进行 n 阶的 LPCC+ΔLPCC 特征参数提取。

Step4　加权特征组合:将经过处理的特征参数赋予一定的权重,然后排列成特征向量作为识别系统的输入使用。在为每帧数据计算完参数之后分别在每个倒谱系数之前乘以不同的权系数,以提高特征在系统中的抗噪性能。此时的特征参数空间维数为$(m+n)\times 3$。

实验证明,文中 LPCC 和 ΔLPCC 的阶数取 12,MFCC 和 ΔMFCC 的阶数取 16 时识别效果较好。则此时特征参数空间维数为 84。低频部分特征参数的权重取 1,高频部分特征参数的权重取 2。

动态时间规整算法(Dynamic Time Warping,DTW)和矢量量化(Vectov Quantization,VQ)方法是目前较为成熟的说话人识别方法[9],将以上提取的组合特征参数组成特征向量分别供动态时间规整算法(DTW)模型和矢量量化(VQ)模型进行识别。

3　特征参数在噪声环境下的性能分析和比较

为了使实验数据更为可靠,本文建立了 40 个人的语音库,在语音的录制过程中考虑了语速快慢、音量、时间间隔等影响说话人辨认系统性能的因素,实验者皆说普通话,个人略带地方色彩。在相对安静的环境下采集这 40 个说话人的语音,每人朗读一段 20 s 的课文作为系统的训练模板。重新采集这 40 个人的语音,内容与训练模板相同,每采集一次语音就加入一次高斯白噪声,并设置 5 个不同的信噪比作为本实验 5 个测试样本存入电脑文件夹中,采集完毕可方便调用。将测试样本分别输入动态时间规整(DTW)模型和矢量量化(VQ)模型进行识别,实

验数据见表 1 和表 2。

表 1 基于 DTW 模型的特征参数在不同信噪比条件下的识别率

Tab. 1 Recognition rate of feature parameters in different SNR condition base on DTW model %

信噪比/dB	LPCC	MPCC	LPCC+ΔLPCC	MPCC+ΔMFCC	方案 1	方案 2	方案 3	方案 4
−10	15.0	15.0	22.5	25.0	25.0	25.0	22.5	25.0
−5	15.0	17.5	30.0	30.0	40.0	40.5	45.0	45.0
0	22.5	27.5	50.0	52.5	55.0	52.5	62.5	65.0
20	50.0	52.5	70.0	72.5	72.5	72.5	77.5	80.0
50	70.0	75.0	90.0	90.0	90.0	90.0	92.5	92.5
干净	82.5	85.0	95.0	97.5	92.5	92.5	95.0	95.0

表 2 基于 VQ 模型的特征参数在不同信噪比条件下的识别率

Tab. 2 Recognition rate of feature parameters in different SNR condition base on VQ model %

信噪比/dB	LPCC	MPCC	LPCC+ΔLPCC	MPCC+ΔMFCC	方案 1	方案 2	方案 3	方案 4
−10	15.0	17.5	22.5	25.0	27.5	27.5	32.5	35.0
−5	15.0	17.5	30.0	30.0	40.0	40.0	45.0	47.5
0	35.0	30.0	52.5	75	57.5	57.5	62.5	65.0
20	52.5	55.0	75.0	75.0	77.5	77.5	80.0	82.5
50	72.5	77.5	90.0	92.5	90.0	92.5	95.0	95.0
干净	82.5	87.5	97.5	100	95.0	95.0	97.5	97.5

1)从图 1 和图 2 中可以看出,在干净语音条件下,采用传统的组合特征参数的系统识别率略高于本文提出的组合方案,主要原因可能是在阈值处理时对低频部分的小波系数用门限进行了处理。小于或等于阈值的小波系数被其视为携带能量较少者,处理中将这些值作为零来处理,而仅仅保留阈值以上的数据,这样做造成了部分表征说话人个性特征的信息丢失,从而影响了识别率。但从总体来说,识别率较高。

2)在信噪比为 0～50 dB 时,本文提出的特征参数提取方法较传统方法有所提高。方案 1 和方案 2 的识别性能和传统方法的识别性能几乎一样;方案 3 和方案 4 的识别率略高于传统方法,主要是采用了 LPCC+ΔLPCC 和 MFCC+ΔMFCC 进行组合,它们之间的相关性比较小,即有动态特征也有静态特征识别,因此识别效果更好。

3)当信噪比为−10～0 dB 时,本文提出的特征参数提取方法使得系统的识别性能有了明显的提高,充分说明了本文提出的方法的有效性,特别是在小于 0 dB 后,系统的识别率骤然提高,说明该方法在噪声环境下有较好的鲁棒性。

4)方案 4 的识别率比方案 3 的略高,这主要因为小波分解后,低频部分蕴含了丰富的频率特征,而 MFCC 是基于人的听觉特性的提取方法,比 LPCC 的识别性能好,因此在低频采用 MFCC 及其差分,在高频又采用基于人的声道特性的 LPCC 及其差分,能更好地提取动态和静态特征,因此识别率也就有了一定的提高。

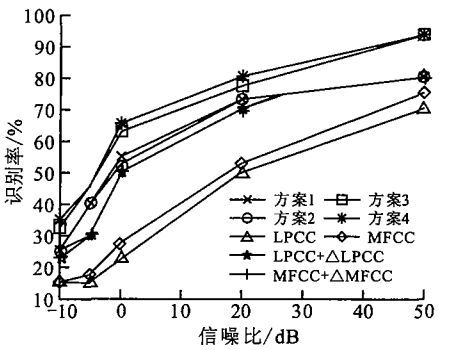

图 1 基于 DTW 模型的特征参数在不同信噪比条件下的识别率

Fig. 1 Recognition rate of feature parameters in different SNR condition base on DTW model

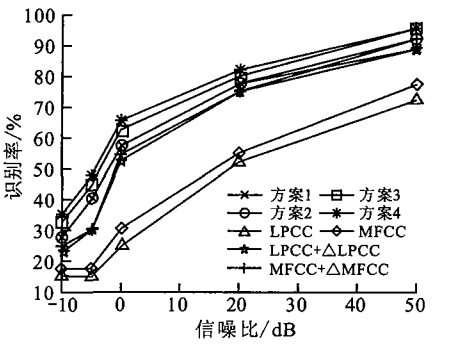

图 2 基于 VQ 模型的特征参数在不同信噪比条件下的识别率

Fig. 2 Recognition rate of feature parameters in different SNR condition base on VQ model

4 结语

在噪声环境下,说话人识别系统的识别性能明

显降低,本文提出的组合特征参数提取方法,结合小波变换去噪处理,能有效改善这一缺点,其中方案 4 的特征组合具有较强的抗噪性能,在较强噪声环境下比传统特征参数更具有鲁棒性,对于自然环境条件下的说话人识别来说也有一定的研究意义。

参考文献(References):

[1] 檀蕊莲. 基于小波变换的说话人识别技术研究[D]. 西安:武警工程大学, 2008.
TAN Ruilian. Research of speaker recognition based on wavelet transform [D]. Xi'an: Armed police engineering university, 2008. (in Chinese)

[2] 王刚,邬晓钧. 基于参考说话人模型和双层结构的说话人辨认[J]. 清华大学学报:自然科学版,2011,51(9):1261-1266.
WANG Gang, WU Xiaojun. Speaker recognition based on reference speaker model and two-tier structure[J]. Journal of tsinghua university: natural science edition, 2011,51(9):1261-1266. (in Chinese)

[3] 王永奇. 语音增强用于抗噪声的汉语说话人识别[J]. 微电子学与计算机,2006,23(2):166-168.
WANG Yongqi. Chinese speech enhancement for anti-noise speaker recognition [J]. Microelectronics and computer, 2006,23(2):166-168. (in Chinese)

[4] Joxicic S T, Saric Z A. Caustic analysis of consonants in whispered speech[J]. Journal of voice,2008,22(3):263-274.

[5] 芮贤义,俞一彪. 噪声环境下说话人识别组合特征提取方法[J]. 信号处理,2006,22(5):673-677.
RUI Xianyi, YU Yibiao. Feature extraction methods of speaker recognition in noisy environment[J]. Signal processing, 2006,22(5):673-677. (in Chinese)

[6] Woo S Ch, Lim Ch P, Osman R. Development of a speaker recognition system using wavelets and artificial neural networks[J]. Processing of 2001 international symposium on intelligent, multimedia, video and speech processing, 2001,2-4:413-416.

[7] Kinney A, Stevens J. Wavelet packet cepstrum [J]. the thirty-sixth asilomar and computers, analysis for speaker recognition,2002,1(3-6):206-209.

[8] 杨福生. 小波变换的工程分析与应用[M]. 北京:科学出版社,2003.
YANG Fusheng. Engineering analysis and application of wavelet transform [M]. Beijing: Science press, 2003. (in Chinese)

[9] 赵力. 语音信号处理[M]. 北京:机械工业出版社, 2005.
ZHAO Li. Speech signal processing [M]. Beijing: Mechanical industry publishing house, 2005. (in Chinese)

本刊相关链接文献:

[1]柳革命,吴姚振. 响度特征量化的改进算法[J]. 空军工程大学学报:自然科学版, 2011,12(4):91-94.

[2]常思江,王中原,韩成辉. 一种基于过程噪声控制的弹道滤波方案[J]. 空军工程大学学报:自然科学版,2011,12(1):51-54.

[3]伍友利,方洋旺,王洪强,等. 乘性/加性噪声 Markov 跳变系统线性均方最优控制[J]. 空军工程大学学报:自然科学版,2009,10(5):32-36.

[4]田野,王永良,张永顺,等. SαS 噪声环境下的宽带信号二维波达方向估计算法[J]. 空军工程大学学报:自然科学版,2009,10(4):71-75.

(编辑:徐楠楠)