

语音信号的一种加权的鲁棒特征参数 W_RAS_MFCC

方杰 李英 陶泯

(江南大学通信与控制工程学院 江苏 无锡 214122)

摘要: 本文讨论了一种加权的带通滤波器分析法(WFBA)来加强语音信号相对自相关序列(RAS)美尔倒谱系数的性能,它通过简单而直接的手法使人耳敏感的对数带通滤波器(LFBEs)频谱峰值更加突出,而使相应不敏感的谷值处更加减小,实验表明,在噪声干扰或者有信道失真情况下,语音的相对自相关序列的加权的美尔特征差数(W_RAS_MFCC)比 W_MFCC 以及其它的谱的提升技术具有更强的鲁棒性。

关键词: 相对自相关序列; 加权带通滤波器分析法; 倒谱提升

The Weighted Robust Speech Features- W_RAS_MFCC

FANG Jie LI Ying TAO Min

(School of Communication and Control Engineering, South Yangtze University, Wuxi, 214036, China)

Abstract: In this paper, we discuss the use of weighted filter bank analysis (WFBA) to increase the discriminating ability of mel frequency cepstral coefficients of the relative autocorrelation sequences (RAS_MFCCs). The WFBA emphasizes the peak structure of log filter bank energies (LFBEs) obtained from filter bank analysis while attenuating components with lower energy in a simple, direct way. Experiments show that the weighted mel frequency cepstral coefficients of relative autocorrelation sequences (W_RAS_MFCCs) are more robust than the weighted frequency Cepstral coefficients (W_MFCCs) and other widely used cepstral liftering methods.

Keywords: relative autocorrelation sequences; weighted filter bank analysis; cepstral liftering

引言

目前,带通滤波器分析技术已经广泛应用于各种语音处理技术中,这种技术的主要特点是跟据人耳对不同频率的敏感特性不同,将人耳模拟成带通滤波器组,如 MEL 频率尺度滤波器组等。这种带通滤波器的输出实际上可以看成一种短时的频谱包络,在不同的说话人、背景噪声、以及信道干扰等情况下,这种频谱包络容易产生波动,这就影响了谱分析可靠性。为了减轻这些外在因素的影响,国内外近年来提出了一些谱的提升技术如:正弦提升、线性提升、统计提升、指数提升^[1]等,这些提升技术虽然有一定成效,但还不能达到所期望效果,在此基础上,Wei-Wen Hung, Hasiao-Chuan Wang 等人提出了加权带通滤波器分析技术(WFBA)^[2]较上述方法效果更好,本文对 WFBA 形式稍作改进,并根据语音信号自相关算法能够消除噪声等特点,以语音信号的相对自相关序列(RAS)为对象,提出了一种加权的特征参数 W_RAS_MFCC,实验表明,比 W_MFCC 等特征参数鲁棒性要好。

1 RAS_MFCC

与普通实际频率倒谱分析不同, MFCC 的分析着眼于人耳的听觉特性,人耳听到的声音高低与频率并不成正比关系,而用 Mel 频率尺度则更符合人耳的听觉特性。所谓的 MEL 频率尺度,它的值大体上对应于实际频率的对数分布关系。Mel 频率与实际频率具体关系可用式表示^[3,4]:

$$Mel(f) = 2595 \lg(1 + f/700) \quad (1)$$

这里,实际频率 f 的单位是 Hz。临界频率随着频率的变化而变化,并与 Mel 频率的增长一致,在 1000Hz 以下大致呈线性分布,带宽为 100Hz 左右;在 1000Hz 以上呈对数增长。类似于零界频带划分,可以将语音频率划分成一系列的三角形滤波器序列,即 Mel 滤波器组。取每个三角形滤波器频率带宽内所有信号幅度加权和作为某个带通滤波器的输出,然后对所有滤波器输出做对数运算,再进一步做离散余弦变换(DCT)即得到 MFCC。MFCC 参数的计算过程的具体表达式如下:

$$e(i) = \sum_{k=1}^N |X(k)|^2 * H_i(k) \quad (2)$$

经余弦变换得到 MFCC 系数

$$c(n) = \sum_{i=0}^{M-1} e(i) \cos(\pi n(i-0.5)/M)$$

$$0 \leq n \leq M$$

传统的方法求出倒谱系数之后, 通常还要为这 M 系数乘以不同的权系数进行倒谱系数提升。常用的是正弦提升:

$$w_i = 1 + \frac{D}{2} \sin\left(\frac{\pi i}{D}\right) \quad (3) \quad \text{其中: } D \text{ 为倒谱的阶数。}$$

美谱特征描述人耳对频率高低的非线性感知特性, 具有一定的噪声鲁棒性, 实验表明在大多数情况下, 它的性能优于传统的 LPCC 系数。所以在当今的语音识别主流算法:

“倒谱+HMM”中, 我们更倾向于用 MFCC 系数。但是无论是 LPCC 系数还是 MFCC 系数作为语音特征参数都只能在高信噪比下获得较高的识别率, 在低信噪比的情况下识别率很低, 甚至根本无法识别, 因此如何提取抗噪的语音特征参数一直以来是国内外研究的热点问题, 提出了诸多的方法如减谱, 卡尔曼滤波, 基于子带滤波等语音识别方法。语音自相关性算法能够有效消除噪声, 语音信号的单边自相关序列 (OSA) 再经过帧间滤波可得语音信号的相对自自相关序列 (RAS)^[3], 如下式:

$$Ras(m, k) = \frac{1}{T_L} \sum_{t=-L}^L t * R(m+t, k) \quad (4)$$

其中: $0 \leq m \leq M-1, 0 \leq k \leq N-1$, M, N 分别为语音信号的帧数和帧长。 k 为自相关序号, $t=-L:L$, L 为计算差分帧的范围, 一般取 $2 \sim 3$, $T_L = \sum_{t=-L}^L t^2$ 。求得

语音信号的单边自相关序列 (RAS) 后, 再把它作为语音信号的短时域特征, 求取它的美谱特征参数和差分特征参数 RAS_MFCC 和 Δ RAS_MFCC 作为语音特征, 能够有效消除卷积噪声^[4], 在低信噪比下, 有很高的识

2 W_RAS_MFCC

通过上述的描述, 我们知道求出美谱系数后, 退化的语音信号总会产生谱的波动和识别率的下降。虽然前面的倒谱提升和频率提升方法被广泛的应用于加强语音特征, 仍然需要一种更好的解决办法, 接下来介绍一种加权的带通滤波器分析方法。众所周知, 在滤波器通带比在阻带具有更好的抗噪效果, 因此我们的目的是更好的加强对数滤波器的峰值能量, 来增强美谱系数对环境干扰的鲁棒性, 下面的方法是通过带通滤波器对数能量谱乘以一系列的权系数来达到这一目的的, 又称加权带通滤波器分析 (weighted filter bank analysis WFBA) 实验表明: 该方法比传统的方法有更好的抗噪声和信道失真功能。算法如下:

$$c_m(WFBA) = \sum_{i=1}^Q w(i) \cdot \log[e(i)] \times \cos\left[m \cdot \left(\frac{2 \cdot i - 1}{2}\right) \cdot \frac{\pi}{Q}\right] \quad (5)$$

这里, 我们给出权函数的两种形式:

$$\text{类型 1: } w(i) = \frac{\beta_i}{\sum_{j=1}^Q \beta_j} + 1 \quad \text{其中}$$

$$\beta_i = \sum_{r=1}^Q \left(\frac{\log[e(i) + 1.0]}{\log[e(r) + 1.0]} \right)^{1/F-1} \quad (6)$$

类型 2:

$$w(i) = \frac{\log[e(i) + 1.0]}{\sum_{j=1}^Q \log[e(j) + 1.0]} + 1 \quad (7)$$

对于第一种类型的权函数, 一个模糊集函数用来决定权系数, 通过调整模糊元素 F , 我们可以得到一系列的权系数。当 F 为无穷大时这个时候权系数为 $1/Q$, 在第二种类型, 权系数就是每个波段的对数能量所占的比例。它有计算量小, 不需要计算模糊元素等优点。每个权系数加 1 是为了防止权值过小而产生算法溢出, 第一种我们称之为 FWFBA, 第二种我们称之为 DWFBA。

本文以语音信号的相对自相关序列作为研究对象, 给出了一种加权的语音信号相对自相关美谱特征的提取方法, 通过理论分析和实验验证, 比前面的方法尤其是在对退化语音取得更好识别效果, 该算法的具体流程如下:

- 1) 预加重 $x_n' = x_n - kx_{n-1}$, 其中 k 为预加重系数, 一般取 0.95;
- 2) 加窗 (hamming), 分帧, 帧长为 N ;
- 3) 求出语音信号单边自相关序列;
- 4) 对语音信号的帧求导, 求出单边自相关序列;
- 5) 对单边自相关序列进行 DFT 变换, 利用式 (2) 求出 $e(i)$;
- 6) 根据式 (6) 或 (7) 求出权系数 $w(i)$;
- 7) 再利用 (5) 式求出 RAS_MFCC;
- 8) 再对求得的美谱系数进行一次均值规整, 并求其帧的一阶导数^[6];
- 9) 最后以 $\{W_RAS_MFCC, W_ \Delta RAS_MFCC\}$ 为识别特征参数;

3 实验

1 端点检测

端点检测用语音信号的相对自相关序列短时平均幅度进行检测, 下图是语音信号“2”在 0dB 下语音信号时域波形图以及其相对自相关序列的短时平均幅度, 在同样

的信噪比下语音信号的短时平均幅度接近 0.5, 而相对自相关序列的短时平均幅度只有 0.08, 所以利用语音的这

一特性进行端点检测可行的下表是短时平均幅度发与一般方法进行端点检测的比较

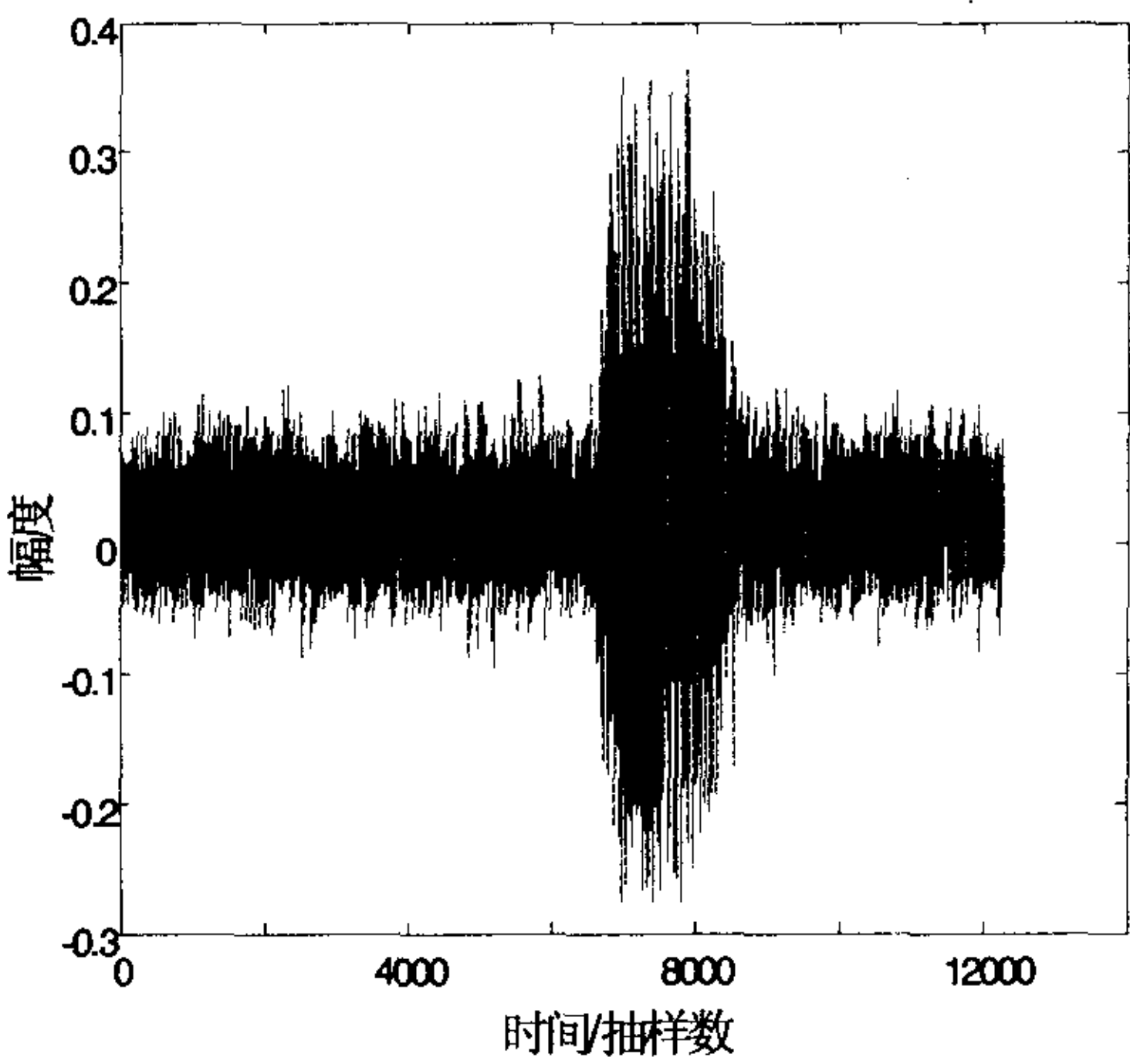


图 1 语音信号“2”在 0dB 下波形图

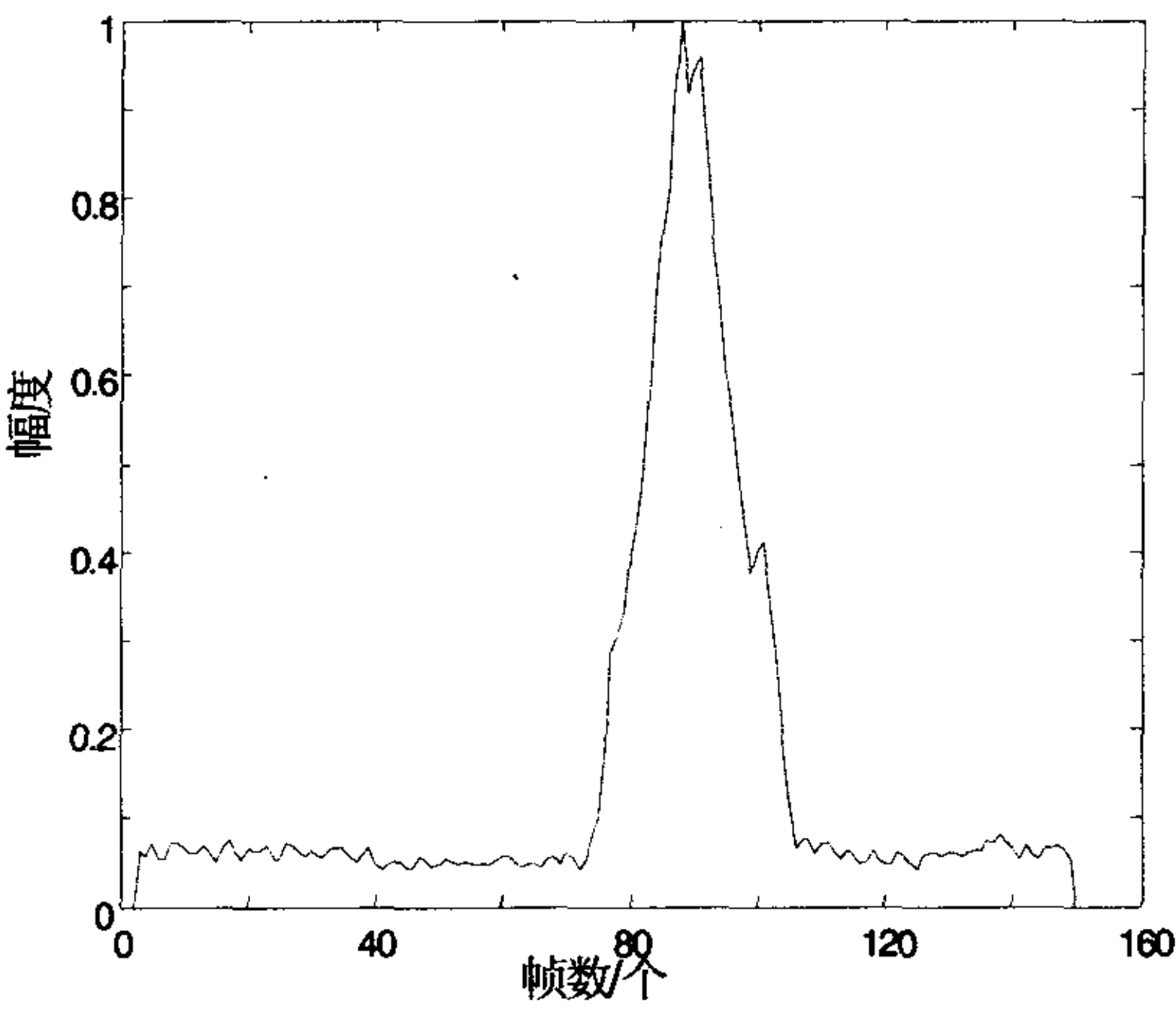


图 2 语音信号“2”的相对自相关序列短时平均幅度

表 1 双门限比较法和本文方法端点检测正确率比较 (正确率)

SNR/dB	>35	25	20	15	10	5dB	0	-5
双门限比较法	1.00	1.00	0.965	0.850	0.810	0.430	0.10	0
本文方法	1.00	1.00	1.00	0.990	0.985	0.985	0.979	0.940

2 识别实验

本实验采用“美尔倒谱+HMM”识别系统, 分别采集 3 男 2 女语音样本, 分别是阿拉伯数字 0~9, 和 20 个汉语短语 (2~4 字), 如北京, 回家, 上学等, 每个词发 10 遍, 采样频率为 8000 Hz, 其中 70%用于训练样本, 另外 30%用于测试。另外取这些词的电话录音各 3 遍, 识别的特征参数及一阶差分特征参数共 24 维, 分别采用 MFCC+

Δ , MFCC+ Δ (SL), W_MFCC+ Δ , RAS_MFCC+ Δ (SL), W_RAS_MFCC+ Δ , 其中 Δ 代表相应特征参数的一阶差分, SL 为正弦提升, 实际中不同信噪比下 FWFBA 要获得最高的识别率 F 取值是不确定的, 2.0 左右普遍效果较好^[2], 故本实验 F 取 2.0, 端点检测用语音信号的相对自相关序列短时平均幅度进行检测, 实验结果见下表:

表 2 语音 MFCC 特征参数及改进形式在不同的信噪比和电话录音下的识别率

SNR/dB	>35	30	20	15	10	5	0	-5	电话录音
MFCC+ Δ	0.976	0.946	0.910	0.854	0.765	0.345	0.111	0	0.448
MFCC+ Δ (SL)	0.992	0.986	0.923	0.878	0.778	0.440	0.126	0	0.451
W_MFCC+ Δ (FWFBA $F=2.0$)	0.982	0.965	0.920	0.897	0.810	0.562	0.163	0.03	0.488
W_MFCC+ Δ (DWFBA)	0.978	0.978	0.921	0.887	0.815	0.556	0.148	0.03	0.484

表 3 语音 RAS_MFCC 特征参数及改进形式在不同的信噪比和电话录音下的识别率

SNR(dB)	>35	30	20	15	10	5	0	-5	电话录音
RAS_MFCC+△	0.974	0.976	0.950	0.904	0.860	0.646	0.311	0.233	0.645
RAS_MFCC+△(SL)	0.994	0.976	0.963	0.918	0.876	0.670	0.326	0.251	0.667
W_RAS_MFCC+△ (FWFBA F=2.0)	0.989	0.978	0.980	0.937	0.907	0.712	0.353	0.283	0.718
W_RAS_MFCC+△ (DWFBA)	0.995	0.978	0.971	0.940	0.901	0.706	0.358	0.270	0.716

表 2 比较了 MFCC 特征参数在不经提升、经过正弦提升和本文方法的比较,可见,在不同的信噪比及有信道失真情况下,以 W_MFCC 特征参数的识别率要高于常用的正弦提升,从表 3 可以看出在低信噪比及有信道失真情况下,以 W_RAS_MFCC 为特征参数的识别率要明显高于 W_MFCC,也要优于经正弦提升后的特征参数 RAS_MFCC,所以通过上述实验可知,W_RAS_MFCC 是语音信号的一种实际可行的鲁棒特征参数。

4 结论

本文讨论了通过对带通滤波器加权分析来提高对数滤波器峰值能量,并结合语音信号的自相关算法能够减小噪声影响,提出了语音信号的一种鲁棒性很强的特征参数 W_RAS_MFCC,实验证明其性能要优于 MFCC 及其改进的特征参数。

参考文献

[1]K.K. Paliwal .On the Use of Filter-bank Energies As Features For Robust Speech Rocognition [C].IEEE.ISSPA. Brisbane, Australia,1999,pp 641~644

[2] Wei-Wen Hung, Member, Hsiao-Chuan Wang. On the Use of Weighted Filter Bank Analysis for the Derivation of Robust MFCCs[J]. IEEE.signal processing letters,2001,8 (1),pp:70~

[3]王定让, 柴配琪. 语音倒谱特征研究[J]. 计算机工程, 2003, 29 (13), 31~33

[4]赵力. 语音信号处理 [M]. 北京:机械工业出版社, 2003

[5] 徐金甫,韦岗. 基于单边自相关序列的语音特征及其在带噪语音识别中的应用 [J]. 计算机工程 , 2000, 26(5), 57~59

[6] 徐文盛,戴蓓倩,方绍武,李辉. 基于连续HMM孤立语音鲁棒识别方法 [J]. 电路与系统学报, 1999, 4 (4), 19~23

作者简介

方杰 (1978-), 男, 安徽寿县人, 汉族, 江南大学通信与控制工程学院 2003 级研究生, 专业“控制理论与控制工程”, 研究方向: 语音信号处理。

联系电话 0510—5911509 13003306424

电子信箱: jackiefang@126.com

李英 (1955-), 女, 浙江金华市人, 汉族, 江南大学通信与控制工程学院副教授 硕士生导师。研究方向: 数字信号处理。0510-5705497 13023325403

陶泯(1978-), 男, 安徽马鞍山人, 汉族, 江南大学通信与控制工程学院 2004 级研究生, 专业“检测与自动化仪表”, 研究方向: 语音信号处理。