

噪声自适应的多数据流复合子带语音识别方法

张军 韦岗

(华南理工大学电信学院 广州 510640)

摘要 首先针对现有丢失数据语音识别技术中的边缘化(marginalisation)技术在特征运用上的局限,提出了一种倒谱特征分量的可靠性估计方法,将边缘化技术推广到常用的倒谱语音识别系统中;然后利用基于全带和子带倒谱特征的边缘化识别器在不同噪声中的互补性能,提出了一种噪声自适应的多数据流复合子带语音识别方法。实验结果表明,所提识别方法可以自适应地选出全带和子带数据流中受噪声影响较小者并以此为主要依据进行识别,有效地提高了识别系统在多变噪声环境中的鲁棒性。

关键词 语音识别, 丢失数据, 边缘化, 多数据流, 复合子带

中图分类号: TP391.42

文献标识码: A

文章编号: 1009-5896(2006)07-1183-05

Noise Adaptive Multi-stream Hybrid Sub-band Approach for Robust Speech Recognition

Zhang Jun Wei Gang

(College of Electronic & Information Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract This paper first proposes a new method for evaluating the reliability of cepstral components and extends the marginalisation technique to cepstral recognizers. Then a noise adaptive multi-stream hybrid sub-band approach is proposed for robust speech recognition by making use of the complementary performances between full-band and sub-band cepstral marginalisation recognizers in different noises. Experimental results show that the proposed approach can turn to the less distorted data stream automatically and improve the robustness of the speech recognizer in various noisy environments effectively.

Key words Speech recognition, Missing data, Marginalisation, Multi-stream, Hybrid sub-band

1 引言

尽管经过多年的努力,有噪环境下的语音识别仍然是语音识别研究中的一个难题,并日益成为语音识别技术进入实用化的主要障碍之一。在过去的研究中,众多研究者提出了大量提高语音识别系统在噪声中鲁棒性的方法,虽然其中的一些(如模型补偿等)在一定条件下可以取得较好的效果,但在面对现实生活中复杂多变的环境噪声时,现有的技术还显得非常不足。

丢失数据(Missing data)技术^[1-3]是近年来提出的一类提高语音识别系统识别过程鲁棒性的新技术。它利用语音特征中的冗余信息,将语音特征分量按其失真程度分为可靠分量和不可靠分量,识别时采用忽略不可靠分量对输出概率的贡献(称为边缘化, Marginalisation),或估计恢复不可靠分量(称为数据重建, Data imputation)的方法,来有效地提高识别系统的抗噪声能力。由于这类技术不需要预先知道环境噪声的模型,也不需要重新训练或修改语音模型,因此与模型补偿等技术相比更能适应多变的噪声环境,所需的运算量也更少。

由于丢失数据技术的一个基本假设是观察矢量中的一部分数据是可靠的,因此一般认为,对语音频谱的任何去相关变换都会将频谱局部的失真扩散到整个特征矢量,不利于丢失数据技术的运用^[4]。基于这种思想,目前文献中的丢失数据技术基本上只应用于滤波器组特征,极大地限制了其使用的范围。虽然数据重建技术允许对恢复后的频谱特征进行进一步的变换,转化为其他特征进行识别,但这样的处理需要额外的基于频谱特征的语音模型,并且运算量也较大。

倒谱是目前语音识别中最常用的一类语音特征。针对现有丢失数据技术在特征运用上的局限,本文首先对边缘化技术直接应用于倒谱特征的可行性进行了探讨,并提出了一种倒谱特征分量的可靠性估计方法,将边缘化技术推广到倒谱语音识别系统中。在此基础上,本文对基于全带倒谱特征和子带倒谱特征的边缘化技术在不同噪声环境中的性能进行了观察研究,并利用它们之间的互补性提出了一种噪声自适应的多数据流复合子带语音识别方法。实验结果表明,本文提出的识别方法可以自适应地选出全带和子带数据流中受噪声影响较小者并以此为主要依据进行识别,进一步提高了识别系统在多变噪声环境中的鲁棒性。

2 基于倒谱特征的边缘化技术

在边缘化技术的模型中,观察矢量的输出概率采用可靠

2004-11-05 收到, 2005-07-22 改回
国家自然科学基金(60502041)和广东省自然科学基金博士启动基金(65300146)资助课题

分量的边缘概率来计算,即对于具有混合高斯概率密度函数和对角方差矩阵的HMM语音识别系统,观察矢量输出概率为^[1]

$$p(\mathbf{o} | C_k) = \sum_{m=1}^{M_k} w_{mk} \prod_{i, \text{reliable}} N(o_i; \mu_{imk}, \sigma_{imk}^2) \quad (1)$$

其中 C_k 为 HMM 模型中第 k 个状态, \mathbf{o} 为观察矢量, $N(o; \mu, \sigma^2)$ 为一维高斯概率密度函数, w_m 为高斯分量的权重。

由式(1)可以看到,边缘化技术在理论上并不需要局限于滤波器组特征,任何特征只要在噪声的影响下仍能包含一定数量的可靠分量,边缘化技术都应能有效地提高系统的识别率,其效果将取决于所采用特征及其分量受噪声影响的情况。为此,本文对将频谱失真扩散到整个特征矢量的特征(如全带倒谱特征等)和将频谱失真限制在一定范围内的特征(如子带倒谱特征、滤波器组特征等)在不同噪声环境中的失真进行了观察研究,实验中发现:(1)前者尽管对窄带噪声性能不理想,但对于一些常见的全带噪声,其总的失真有时反而会比后者更小。这个现象在一些文献的实验结果中也有反映^[5-6];(2)前者在受到全带噪声影响时,通常也会出现大部分失真集中在某些特征分量中的现象。图1中为英文数字“ZERO”在10dB的白噪声和F16噪声中12维全带MFCC特征各分量的平均归一化均方误差(Normalized mean square error, NMSE)^[7]。图中可以清楚看到倒谱特征分量失真间的差别。这些观察结果说明,对于全带噪声,对倒谱特征直接应用边缘化技术应同样能有效地提高识别系统的识别率,并且某些情况下其性能还可能更优。

为了能有效地将边缘化技术应用到倒谱识别系统中,本文针对加性噪声的情况提出了一种倒谱特征的不可靠分量性判决准则,其基本原理如下所述。

令 \mathbf{c}_x , \mathbf{c}_y 和 \mathbf{s}_x , \mathbf{s}_y 分别表示倒谱域和 LOG 频谱域的干净语音信号和带噪语音信号, $\mathbf{c}_n = \mathbf{c}_y - \mathbf{c}_x$, $\mathbf{s}_n = \mathbf{s}_y - \mathbf{s}_x$ 表示倒谱域和 LOG 频谱域的噪声,则 \mathbf{c}_y 可以表示为

$$\mathbf{c}_y = \mathbf{c}_x + \mathbf{c}_n = \mathbf{T}\mathbf{s}_x + \mathbf{T}\mathbf{s}_n = \mathbf{T}\mathbf{s}_y \quad (2)$$

其中 $\mathbf{T} = \begin{bmatrix} t_{11}, \dots, t_{1J} \\ \vdots & \ddots & \vdots \\ t_{I1}, \dots, t_{IJ} \end{bmatrix}$ 为 LOG 频谱域到倒谱域的线性变换。根据式(2),当某倒谱分量的噪声幅度 $|c_{n,i}|$ 超出一定阈值时,则可以合理地认为该分量为不可靠特征分量。通常情况下, \mathbf{s}_x 可以由增强算法(如谱相减等)估计得到,因此 \mathbf{c}_n 可以用下式进行估计:

$$\hat{\mathbf{c}}_n = \mathbf{c}_y - \hat{\mathbf{c}}_x = \mathbf{T}\mathbf{s}_y - \mathbf{T}\hat{\mathbf{s}}_x = -\mathbf{T}\mathbf{s}_g \quad (3)$$

其中 $\hat{\mathbf{c}}_n$, $\hat{\mathbf{c}}_x$, $\hat{\mathbf{s}}_x$ 分别表示 \mathbf{c}_n , \mathbf{c}_x , \mathbf{s}_x 的估计值, $\mathbf{s}_g = \hat{\mathbf{s}}_x - \mathbf{s}_y = (\log g_1, \log g_2, \dots, \log g_J)^T$, g_j 为增强算法在线性频谱域的增益。根据式(2)和式(3)可得倒谱特征的一种不可靠分量判决准则为

$$|c_{y,i}| \leq \alpha_1 |\hat{c}_{n,i}| \quad (4)$$

其中 $\hat{c}_{n,i}$ 为倒谱域中噪声第 i 个分量的估计值, α_1 为调节门限的参数。

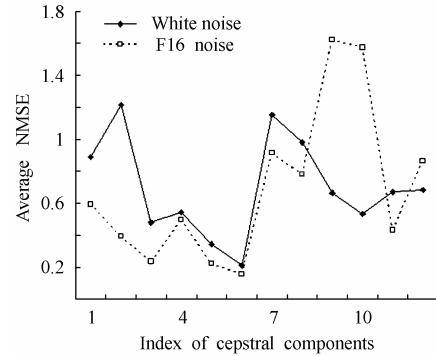


图1 英文字母“ZERO”在10dB白噪声和F16噪声中MFCC特征各分量的平均NMSE

Fig.1 Average NMSE of “ZERO” in 10dB white noise and F16 noise

在实际应用中,由于增强算法在很多情况下(特别在干净语音的频谱被噪声完全掩盖时)并不能准确地估计出 \mathbf{s}_x , 因此单独使用准则式(4)具有一定的局限性。为了进一步提高不可靠分量判决的准确性,本文提出了另一种基于噪声掩盖原理的倒谱分量可靠性估计方法。

对于 \mathbf{s}_y 中的每个分量,可以近似地认为干净信号或噪声中的一种占主要地位,其贡献掩盖了另外一种信号的贡献,因此可以近似地看成是干净信号或噪声。令

$$s'_{x,j} = \begin{cases} s_{y,j}, & s_{y,j} > \log(\beta \bar{N}_j) \\ 0, & \text{其它} \end{cases} \quad (5)$$

$$s'_{n,j} = \begin{cases} s_{y,j}, & s_{y,j} \leq \log(\beta \bar{N}_j) \\ 0, & \text{其它} \end{cases} \quad (6)$$

其中 \bar{N} 为噪声频谱的估计值。则 s'_x 与 s'_n 可以近似地认为是 LOG 频谱域的干净信号和噪声,因此 \mathbf{c}_y 可以表示为

$$\mathbf{c}_y = \mathbf{T}\mathbf{s}_y = \mathbf{T}\mathbf{s}'_x + \mathbf{T}\mathbf{s}'_n \quad (7)$$

根据式(7),每个带噪语音信号的倒谱分量 $c_{y,i}$ 都可以看作由多个 LOG 域的干净语音信号 $s'_{x,j}$ 和噪声 $s'_{n,j}$ 加权叠加而成,因此当参与叠加的噪声项 $t_{ij}s'_{n,j}$ 的幅度和或能量和大于一定阈值时,可以认为该分量为不可靠特征分量,由此可得倒谱特征的另一种不可靠分量判决准则为

$$\sum_j |t_{ij}s_{y,j}| \leq \alpha_2 \sum_j |t_{ij}s'_{n,j}| \quad (8)$$

准则式(4)和式(8)相互独立,并且能够相互补充,因此将两种准则结合起来使用应能取得更好的效果。本文采用“或”运算将两个准则结合起来,即当下式成立时,

$$|c_{y,i}| \leq \alpha_1 |\hat{c}_{n,i}| \quad \text{或} \quad \sum_j |t_{ij}s_{y,j}| \leq \alpha_2 \sum_j |t_{ij}s'_{n,j}| \quad (9)$$

判决 $c_{y,i}$ 为不可靠分量。

在实际应用中,差分特征经常与静态特征结合使用,以提高识别系统的识别率和鲁棒性。一阶差分特征一般可采用多项式 $\Delta \mathbf{c} = \mathbf{G} \sum_{l=1}^L l \mathbf{c}(l)$ 来计算,其中 l 为帧序号,

$G = 1 / \sum_{l=-L}^L l^2$ 。定义倒谱差分特征的噪声为 $\Delta \mathbf{c}_n = \Delta \mathbf{c}_y - \Delta \mathbf{c}_x$,

则由式(3)可得

$$\Delta \hat{\mathbf{c}}_n = \Delta \mathbf{c}_y - \Delta \hat{\mathbf{c}}_x = G \sum_{l=-L}^L \hat{l} \hat{\mathbf{c}}_n(l) \quad (10)$$

$\Delta \hat{\mathbf{c}}_n$ 和 $\Delta \hat{\mathbf{c}}_x$ 分别为 $\Delta \mathbf{c}_n$ 和 $\Delta \mathbf{c}_x$ 的估计值。由式(10)可得到差分特征的不可靠分量判决准则为

$$|\Delta c_{y,i}| \leq \alpha'_1 |\Delta \hat{c}_{n,i}| \quad (11)$$

同时, 根据式(7)有

$$\Delta \mathbf{c}_y = G \sum_{l=-L}^L l T \mathbf{s}_y(l) = G \sum_{l=-L}^L l T \mathbf{s}'_x(l) + G \sum_{l=-L}^L l T \mathbf{s}'_n(l) \quad (12)$$

由式(12)可得到另一差分特征的不可靠分量判决准则为

$$\sum_{l=-L}^L \sum_j |l t_{ij} s_{y,j}(l)| \leq \alpha'_2 \sum_{l=-L}^L \sum_j |l t_{ij} s'_{n,j}(l)| \quad (13)$$

与静态特征相类似, 本文将式(11)和式(13)结合起来一起使用, 即当

$$|\Delta c_{y,i}| \leq \alpha'_1 |\Delta \hat{c}_{n,i}| \text{ 或 } \sum_{l=-L}^L \sum_j |l t_{ij} s_{y,j}(l)| \leq \alpha'_2 \sum_{l=-L}^L \sum_j |l t_{ij} s'_{n,j}(l)| \quad (14)$$

时, 判决 $\Delta c_{y,i}$ 为不可靠分量。

3 噪声自适应的多数据流复合子带语音识别方法

实验中发现, 尽管边缘化技术可以有效地提高倒谱特征在全带噪声中的识别率, 但由于倒谱中的去相关变换将局部的频谱失真扩散到整个特征, 因此其对窄带噪声的效果依旧并不理想。此外, 实验中还发现, 对于不同的全带噪声, 基于将频谱失真扩散到整个特征矢量的特征(如全带倒谱特征等)和基于将频谱失真限制在一定范围内的特征(如滤波器组特征、子带倒谱特征等)的边缘化技术在性能上有时也会有较大的差别。因此如果能将这些对不同噪声性能各异的识别器合理地结合起来, 并根据环境噪声的不同自适应地选择性能较好的识别器进行识别, 则有可能进一步提高识别器在多变噪声环境中的综合性能。基于这个思想, 本文提出了一种噪声自适应的多数据流复合子带语音识别方法, 其基本原理如图2所示。

输入语音信号经过增强等前处理后, 分别提取特性各异

的特征作为不同的数据流。考虑到子带倒谱特征与滤波器组特征相比同样具有将局部的频谱失真限制在某些特征分量范围内的功能, 并且特征维数和分量间相关性更小, 与全带倒谱特征的互补性也更好^[6,8], 因此文中的数据流分别采用了全带倒谱特征和子带倒谱特征。为了避免复杂的子带权重调节, 子带特征由每个频带提取出来的语音特征采用特征组合(Feature combination)^[5]的方式构成。

对全带和子带数据流, 分别采用准则式(9)和式(14)检测出特征中的不可靠分量, 并采用边缘化技术计算每个观察矢量的输出概率。不同数据流的输出概率在 HMM 状态层采用几何加权平均(Geometric weighted average)的方式结合起来, 即对于具有混合高斯概率密度函数和对角方差矩阵的 HMM, 观察矢量的输出概率采用下式计算:

$$p(\mathbf{o} | C_k) = \prod_{s=1}^S [p_s(\mathbf{o}_s | C_{ks})]^{\gamma_s} \\ = \prod_{s=1}^S \left[\sum_{m=1}^{M_{sk}} w_{msk} \prod_{i, \text{reliable}} N(o_{is}; \mu_{imsk}, \sigma_{imsk}^2) \right]^{\gamma_s} \quad (15)$$

其中 $p_s(\mathbf{o}_s | C_{ks})$ 为数据流 s 的输出概率, S 为数据流个数, M_{sk} 为状态 k 中数据流 s 的混合高斯概率密度函数个数, γ_s 为数据流权重。文中数据流权重采用固定权重的形式, 取值原则为可保证对干净语音具有较好的识别效果。

在式(15)中, 如果设 A 为数据流 s 中任一观察矢量 \mathbf{o}_s 所有分量的集合, 则当该观察矢量未受噪声影响时, 数据流 s 的对数加权概率应为 $\gamma_s \log(p_s^A(\mathbf{o}_s | C_{ks}))$, 其中 $p_s^A(\mathbf{o}_s | C_{ks})$ 为可靠分量集 A 时数据流 s 的输出概率。现设 B, C 为 A 的子集且有 $A \supseteq B \supseteq C \supseteq \Phi$, 则当可靠分量集分别为 B 和 C 时, 数据流 s 的对数加权概率为

$$\gamma_s \log(p_s^B(\mathbf{o}_s | C_{ks})) = \gamma_s \frac{\log(p_s^B(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \log(p_s^A(\mathbf{o}_s | C_{ks})) \\ = \gamma_s^B \log(p_s^A(\mathbf{o}_s | C_{ks})) \quad (16)$$

$$\gamma_s \log(p_s^C(\mathbf{o}_s | C_{ks})) = \gamma_s \frac{\log(p_s^C(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \log(p_s^A(\mathbf{o}_s | C_{ks})) \\ = \gamma_s^C \log(p_s^A(\mathbf{o}_s | C_{ks})) \quad (17)$$

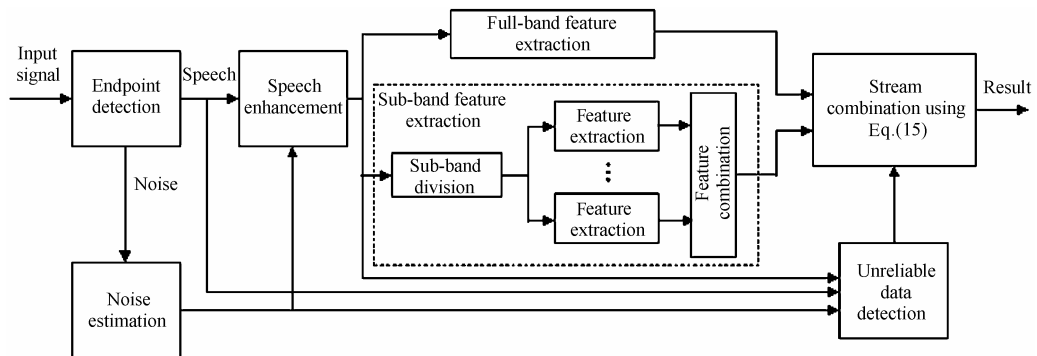


图2 本文所提的识别方法基本原理框图
Fig.2 Block diagram of the proposed approach

其中 $\gamma_s^B = \frac{\log(p_s^B(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \gamma_s$, $\gamma_s^C = \frac{\log(p_s^C(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \gamma_s$, 可

以看作是数据流在可靠分量集为 B 和 C 时的等效权重, $p_s^B(\mathbf{o}_s | C_{ks})$ 和 $p_s^C(\mathbf{o}_s | C_{ks})$ 分别为可靠分量集取 B 和 C 时数据流 s 的输出概率。根据边缘化技术的概率计算方法, 易得

$$0 \leq p_s^A(\mathbf{o}_s | C_{ks}) \leq p_s^B(\mathbf{o}_s | C_{ks}) \leq p_s^C(\mathbf{o}_s | C_{ks}) \leq 1 \quad (18)$$

因此有

$$1 \geq \frac{\log(p_s^B(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \geq \frac{\log(p_s^C(\mathbf{o}_s | C_{ks}))}{\log(p_s^A(\mathbf{o}_s | C_{ks}))} \geq 0 \quad (19)$$

即

$$\gamma_s \geq \gamma_s^B \geq \gamma_s^C \geq 0 \quad (20)$$

由以上分析可以看到, 如果以干净语音的情况为基准, 则观察矢量中可靠分量数目的减少等效于数据流权重的降低。由于可靠分量的数目通常随观察矢量受噪声影响程度的增大而减少, 因此式(15)相当于引进了一种自适应的数据流权重调节机制, 即当某一数据流受噪声影响小时, 其等效权重较高, 而当该数据流受噪声影响大时, 其等效权重较低。这使得基于式(15)的识别器在不同数据流受噪声影响程度差异较大时, 能通过各数据流等效权重的差异自动选择出失真较小的数据流并以此为主要依据进行识别; 而当各个数据流受噪声影响差不多, 其等效权重的比例接近干净语音的情况时, 又可利用多数据流中可靠分量数多于单数据流的优势, 在所采用的特征之间具有较好互补性的情况下取得比单数据流识别器更好的识别效果。因此可以预期, 本文所提识别方法能有效地提高识别系统在多变噪声环境中的性能。

4 实验结果

实验中使用的干净语音数据来自 TI46 孤立词语音数据库, 包括 8 位男性对 10 个英文数字 0-9 的发音。每人对每个字母发音 26 次, 其中的 10 次用于语音模型的训练, 16 次用于测试。噪声数据采用了 NoiseX92 中 4 种特性不同的噪声, 分别为白噪声, F16 噪声, PINK 噪声和 FACTORY 噪声。带噪语音由干净语音与噪声在时域中按不同的信噪比叠加得到。

干净语音的模型采用从左到右无跳转的连续 HMM。HMM 具有 5 个状态, 每个状态下观察矢量的概率密度函数采用了 3 高斯的混合高斯概率密度函数, 每个高斯分量采用对角的方差矩阵, 数据流采用等重加权。HMM 模型的训练和识别均采用 HTK3.1^[9] 工具包来完成。

实验中使用了滤波器组、全带 MFCC 和子带 MFCC 3 种语音特征, 其中滤波器组特征由 MEL 频率上均匀分布的 24 个三角滤波器的 LOG 能量输出构成; 全带 MFCC 特征为滤波器组特征 DCT 变换的前 12 阶系数(不含 0 阶 DCT 系数); 子带特征的选取以与全带特征在安静环境和噪声中均具有较好的互补性为原则, 文中参考了文献[8]的子带特征提取方法, 经验地选取前 14 维滤波器组特征(相当于 0Hz - 1552Hz)

和后 10 维滤波器组特征(相当于 1449Hz - 6250Hz)中提取的 7 维和 5 维 MFCC 系数共同组成 12 维子带 MFCC 特征, 实验表明这样的设计在本文的识别任务中具有较好的性能。识别过程中同时使用了以上特征的一阶差分。

为了消除端点检测的不准确对所提识别方法性能检验的影响, 实验中语音的起止点采用先验的方法来确定。带噪语音在噪声段的频谱平均值用于估计噪声的频谱特性。实验中所有的边缘化技术和所提识别方法均采用常规线性谱相减法对输入语音进行增强。

首先检验边缘化技术应用于倒谱识别系统时的性能。表 1 中分别给出了全带 MFCC、常规线性谱相减以及基于倒谱特征的边缘化技术的识别率, 不可靠分量准则采用式(9)和式(14), $\alpha_1(\alpha_1')$ 和 β 为 1.0, $\alpha_2(\alpha_2')$ 为 2.0。为了进一步验证本文所提不可靠分量检测准则的有效性, 表中还给出了单独采用基于增强算法的式(4), 式(8)(称为准则 1)和基于噪声掩盖原理的式(11), 式(13)(称为准则 2)时边缘化技术的识别率。

由表 1 可以看到, 无论采用何种准则, 边缘化技术都可以有效地提高倒谱识别系统在 4 种噪声中的识别率。这说明了本文所提的不可靠分量判决方法的有效性, 同时也验证了本文对边缘化技术在倒谱识别系统中效果的设想。此外, 比较使用式(3), 式(7)和式(11), 式(13)与本文采用的式(8), 式(14)的效果可以看到, 将基于增强算法的准则与基于噪声掩盖原理的准则结合起来使用可以进一步提高系统的识别率。

为了检验本文所提的多数据流复合子带识别方法的效果, 实验中对所提识别方法与基于滤波器组特征、全带 MFCC 特征和子带 MFCC 特征的边缘化技术的性能进行了比较, 实验结果如表 2 所示。其中在基于滤波器组特征的边缘化技术中, 可靠分量的检测采用 SNR 准则^[1], 当 SNR 大于 0dB 时判决为可靠分量。

由表 2 可以看到, 随着噪声的不同, 基于滤波器组特征、全带 MFCC 特征和子带 MFCC 特征的边缘化技术在性能上出现了较大的起伏, 这说明当使用环境经常发生变化时, 单独使用任何一种识别器都难以取得很好的效果。本文所提识别方法在全带和子带单数据流识别器性能相差较大时, 识别率与其中性能最优者几乎保持一致, 当单数据流识别器识别率相近时, 其性能优于单数据流识别器, 有效地提高了识别器在不同的噪声环境中的综合性能。

5 结束语

本文首先对边缘化技术直接应用于倒谱特征的可行性进行了探讨, 并提出了一种倒谱特征分量的可靠性估计方法, 将边缘化技术推广到更常用的倒谱语音识别系统中。在此基础上, 本文利用基于全带倒谱特征和子带倒谱特征的边缘化识别器在不同噪声中的互补性能, 提出了一种噪声自适应的多数据流复合子带语音识别方法。实验结果验证了本文所提识别方法的有效性。本文在实验中仅采用了全带 MFCC 和子带 MFCC 两种数据流, 如果在实际应用中能合理地设计

表 1 边缘化技术应用于倒谱识别系统时的识别率(单位：%)

Tab.1 Recognition rates of the marginalized cepstral recognizer

SNR(dB)	白噪声					F16 噪声				
	20	15	10	5	0	20	15	10	5	0
MFCC	91.9	67.8	33.7	17.4	10.8	99.5	96.5	69.5	30.1	11.3
谱相减	98.7	90.5	70.6	38.6	20.6	99.1	96.8	84.1	61.9	19.5
边缘化技术(准则 1)	99.0	94.4	81.2	53.1	33.3	99.7	98.7	92.6	78.6	45.7
边缘化技术(准则 2)	99.4	93.4	77.0	45.4	18.5	99.8	98.8	89.3	67.2	20.6
边缘化技术(本文所提准则)	99.5	94.9	82.8	57.6	34.1	99.8	98.9	94.4	82.0	57.3
SNR(dB)	PINK 噪声					FACTORY 噪声				
	20	15	10	5	0	20	15	10	5	0
MFCC	99.2	94.8	61.0	27.0	10.6	99.7	97.7	77.3	32.8	14.9
谱相减	99.1	95.2	79.8	49.6	16.9	99.1	94.5	79.4	50.4	14.1
边缘化技术(准则 1)	99.8	98.9	94.8	79.6	38.2	99.8	98.8	95.9	80.4	41.2
边缘化技术(准则 2)	99.8	97.2	84.4	57.3	20.3	99.5	96.6	83.2	52.0	15.1
边缘化技术(本文所提准则)	99.8	99.1	95.3	82.0	45.2	99.8	99.1	96.6	85.4	44.1

表 2 本文所提的多数据流复合子带识别方法的识别率(单位：%)

Tab.2 Recognition rates of the proposed multi-stream hybrid sub-band approach

SNR(dB)	白噪声					F16 噪声				
	20	15	10	5	0	20	15	10	5	0
基于滤波器组的边缘化技术	99.4	97.7	91.3	68.3	34.2	99.6	98.1	91.0	61.4	34.5
基于全带 MFCC 的边缘化技术	99.5	94.9	82.8	57.6	34.1	99.8	98.9	94.4	82.0	57.3
基于子带 MFCC 的边缘化技术	99.4	97.7	90.0	70.9	28.2	99.6	98.1	91.8	76.8	51.7
多数据流复合子带识别方法	99.6	97.3	90.0	70.4	34.1	99.9	98.6	94.8	81.9	58.0
SNR(dB)	PINK 噪声					FACTORY 噪声				
	20	15	10	5	0	20	15	10	5	0
基于滤波器组的边缘化技术	99.8	99.1	95.4	62.0	30.3	98.6	92.0	59.2	29.4	12.6
基于全带 MFCC 的边缘化技术	99.8	99.1	95.3	82.0	45.2	99.8	99.1	96.6	85.4	44.1
基于子带 MFCC 的边缘化技术	99.5	98.4	95.9	82.7	51.3	99.8	99.1	96.8	84.8	44.3
多数据流复合子带识别方法	99.8	99.1	97.3	86.6	50.2	99.8	99.2	97.5	89.3	44.6

更多互补性能好的数据流，则识别系统的性能可望能得到进一步的提高。

参 考 文 献

[1] Cooke M, Green P, Josifovski L, Vizinho A. Robust automatic speech recognition with missing and unreliable acoustic data. *Speech Communication*, 2001, 34(3): 267 – 285.

[2] 罗宇, 杜利民. 基于隐马尔可夫模型局部最优状态路径的数据重建算法. *电子与信息学报*, 2004, 26(5): 722 – 726.

[3] PalomÃ¡kim K J, Brown G J, Wang D L. A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation. *Speech Communication*, 2004, 43(4): 361 – 378

[4] Veth J, Cranen B, Boves L. Acoustic backing-off as an implementation of missing feature theory. *Speech Communication*, 2001, 34(3): 247 – 256.

[5] Okawa S, Bocchieri E, Potamianos A. Multi-band speech recognition in noisy environments. *IEEE International*

Conference on Acoustics, Speech, and Signal Processing, Seattle, Washington USA, May 12-15, 1998: 641 – 644.

[6] Hariharan R, Kiss I, Viikki O. Noise robust speech parameterization using multiresolution feature extraction. *IEEE Trans. on Speech and Audio Processing*, 2001, 9(8): 856 – 865.

[7] Huerta J, Stern R. Speech recognition from GSM codec parameters. *International Conference on Spoken Language Processing*, Sydney, Australia, November 30-December 4, 1998, 4: 1463 – 1466.

[8] 蒋文建, 韦岗. 噪声下差分复合子带语音识别方法. *通信学报*, 2002, 23(1): 18 – 24.

[9] Young S, Kershaw D, Odell J, et al.. *The HTK Book* (for HTK version 3.1). Cambridge, UK: Cambridge University Tech Services Ltd, 2001.

张 军： 男，1975 年生，讲师，研究方向为信号处理、语音处理。
 韦 岗： 男，1963 年生，教授，博士生导师，研究方向为通信理论、信号处理。