

基于 CMN 和 PMC 算法的语音增强失真补偿方法研究

苗玉杰¹, 刘雪飞¹, 张晓敏²

(1 中国环境管理干部学院, 河北 秦皇岛 066004; 2 秦皇岛职业技术学院, 河北 秦皇岛 066004)

摘 要: 语音增强技术在低信噪比情况下, 由于语音增强带来的失真使得系统的识别性能严重下降. 因此提出一种结合特征空间的倒谱均值归一化算法(CMN)和模型空间的并行模型合并算法(PMC)的语音增强失真补偿技术. 实验结果表明, 该方法有效提高了低信噪比情况下的语音信号识别率.

关键词: 语音增强; 倒谱均值归一化; 并行模型合并

中图分类号: T912

文献标识码: A

文章编号: 1000-7180(2011)06-0160-03

Compensation of Speech Enhancement Distortion with Combination of CMN and PMC

MIAO Yue-jie¹, LIU Xue-fei¹, ZHANG Xiao-min²

(1 Environment Management College of China, Qinhuangdao 066004, China;

2 Vocational Technology College of Qinhuangdao, Qinhuangdao 066004, China)

Abstract: The performance of speech recognition systems degrades severely when speech enhancement technique is applied under low the SNR ratio because of the distortion induced by speech enhancement. This paper presents a Parallel Model Combination (PMC) in model space and Cepstral Mean Normalization(CMN) in characteristic space to compensate the distortion. Results are presented that demonstrate the effective combination of robust methods in different space can significantly improve the recognition rate.

Key words: speech enhancement; Cepstral Mean Normalization(CMN); Parallel Model Combination(PMC)

1 引言

利用语音增强算法在信号空间进行消噪, 是提高语音识别系统噪声鲁棒性的一种方法. 实验表明: 语音增强在中等信噪比范围内, 对系统性能有比较明显的提高, 在低信噪比情况下, 剩余噪声所造成的模型空间的失配仍足以对识别产生很大的影响; 增强对语音造成的损伤则引入了模型和特征之间一种新的失配, 同样造成识别器性能的下降, 这种损伤对应着语音频谱的变形, 包括统计特性的变化和某些鉴别信息的损失^[1]. 语音频谱变形程度会随着信噪比的降低而加剧. 这两个对语音识别不利的因素我们这里统称为语音增强失真, 语音增强失真包括增强引起的语音频谱变形和增强后的剩余噪声, 它们

限制了语音增强对系统抗噪声性能的进一步提高. 目前采用的语音增强失真补偿算法较少, 主要采用单一的倒谱均值归一化算法^[2].

本文根据语音增强失真的等效模型, 借鉴一些对信道噪声和背景噪声比较有效的抗噪声语音识别算法, 运用模型空间的并行模型合并算法(PMC)和特征空间的倒谱均值归一化算法(CMN), 对语音增强失真进行补偿, 从而实现带噪语音信号识别率的提高.

2 倒谱均值归一化算法

倒谱均值归一化(Cepstral Mean Normalization, CMN)算法运算复杂度小, 但对语音识别系统抗噪声性能的提高却十分有效. CMN 算法主要用

收稿日期: 2010-08-15; 修回日期: 2010-11-01

基金项目: 河北省科技厅科学技术研究与发展计划项目(2007-009)

来消除信道卷积噪声在倒谱域造成的偏差.图1给出了倒谱均值归一化算法的实现过程.

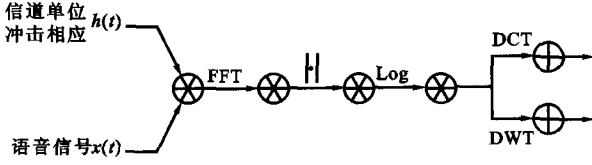


图1 提取 MFCCs 和 MFDWCs

从图1中可以看到,信道噪声和语音信号在时域为卷积的关系,在频谱和功率谱域则为乘性关系,再经过对数变换和 DCT 或 DWT 线性变换,信道噪声对应于倒谱域的加性偏差^[3].

假设语音信号在倒谱域的均值接近于0,则含噪语音倒谱特征均值将主要为信道卷积噪声造成的偏差.因此减去倒谱特征均值的归一化过程将消除信道造成特征空间的失配.语音增强引起的语音失真等效于信道卷积噪声,因此利用 CMN 算法可以有效消除语音失真在特征空间造成的失配.所以采用了 CMN 算法作为一种对语音增强失真进行补偿的方法.

假设 $C(t)$ 是带噪语音的倒谱, m 是倒谱均值,则归一化后的倒谱 $E(t)$ 为

$$E(t) = C(t) - m \quad (1)$$

CMN 算法的关键在于倒谱均值 m 的计算,这里采用式(2)的求解方式.

$$m(t) = \begin{cases} \lambda \cdot \sum_{i=1}^N (C(i)/N + (1-\lambda) \cdot C(t+N)), & t=1 \\ \lambda \cdot m(t-1) + (1-\lambda) \cdot C(t+N), & 2 \leq t \leq T-N \\ m(T-N), & T-N < t \leq T \end{cases} \quad (2)$$

式中, N 是计算倒谱均值 m 的窗宽, λ 是每次计算的更新步长,二者需满足: $1 - \lambda^N = 1/\sqrt{2}$

3 并行模型合并算法

3.1 并行模型合并算法理论框架

并行模型合并 (Parallel model combination, PMC) 算法是一种非常经典的在模型空间展开的抗噪声语音识别算法.图2是并行模型合并算法的原理图.

首先通过事先采集到的噪声数据获取噪声的统计先验,然后在识别过程中对以 MFCC (或 MFDWC) 为特征的语音声学模型进行相应的调整,以使之同对应的含噪语音相匹配^[4].假设噪声和语音相互独立,噪声可以用单高斯分布或多个高斯分布的

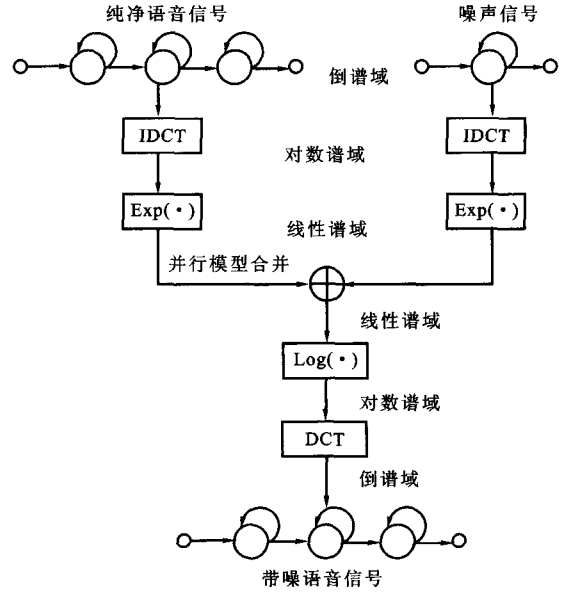


图2 并行模型合并算法

叠加来准确描述其统计规律,且噪声不影响纯净语音识别路径中状态的转移,则 PMC 算法对噪声的补偿就是对 HMM 各状态输出概率密度函数的补偿,反映在对均值和方差的调整上.

由图2可看出,并行模型合并算法是在对数正态假设的基础上,将现实场景的噪声模型与原始的干净语音模型合并,训练噪声模型,得到一组有噪声的语音模型,以获得其统计特性.在模型补偿过程中,纯净语音 HMM 和噪声 HMM 的均值和方差通过逆离散余弦变换从倒谱域转换到对数谱域,然后在对数谱域进行模型调整,即所谓模型合并.合并后的均值和方差反映了对应的含噪语音的均值和方差,最后通过离散余弦变换或小波变换将调整后的均值和方差转换回倒谱域,并用于含噪语音的识别^[5].在对数谱域,可以用高斯分布或者是多个高斯分布的叠加来描述噪声和纯净语音的概率密度函数,对高斯分布的合并过程比较简单.并行模型合并算法 (PMC) 可由式(3)描述:

$$o^{\text{cep}}(\tau) = \text{DCT}\{\lg[g \cdot \exp(\text{IDCT}(s^{\text{cep}}(\tau) + h^{\text{cep}}(\tau))) + \exp(\text{IDCT}(n^{\text{cep}}(\tau)))]\} \quad (3)$$

式中, $o^{\text{cep}}(\tau)$ 、 $s^{\text{cep}}(\tau)$ 、 $n^{\text{cep}}(\tau)$ 及 $h^{\text{cep}}(\tau)$, 分别代表第 τ 帧合成的带噪语音、干净语音、加性噪声以及卷积噪声的倒谱特征参数. cep 表示倒谱域中的参数, lg 表示对数域以及线性域 (频谱域) 中的参数, DCT 与 IDCT 是离散余弦变换及其反变换, g 是一个增益控制因子.利用上式,可以合成出带噪语音模型.处理的参数是基于高斯分布的均值和方差.

运用 PMC 技术进行带噪语音参数估计基于以

下几点假设:

- (1) 语音和噪声相互独立;
- (2) 语音和噪声在时域是相加关系. 此外, 还假定谱估计足够平滑, 以保证语音和噪声在能量上呈现加性;
- (3) 在对数谱域和倒谱域中, 多高斯混合模型包含足够的体现观察矢量的信息;
- (4) 由纯净语音数据生成语音模型的过程不受加性噪声影响.

3.2 对数正态近似算法

对基于 HMM 的语音识别系统而言, 噪声对识别的影响在于其引入了纯净语音模型和含噪语音特征之间的失配, 从而造成识别性能的下降. 从模型角度对噪声的影响进行分析, 为了方便起见, 假设识别系统所使用的特征为对数谱系数, 且语音在对数谱域服从高斯分布. PMC 算法中, 模型合并过程可以采用数值积分法, 对数正态近似法 (Log-Normal Approximation) 和对求和近似法 (Log-Add Approximation). 本文中采用了对数求和近似算法, 其计算量非常小, 而且只对模型均值作调整. 图 3 是对数求和近似算法的原理^[6].

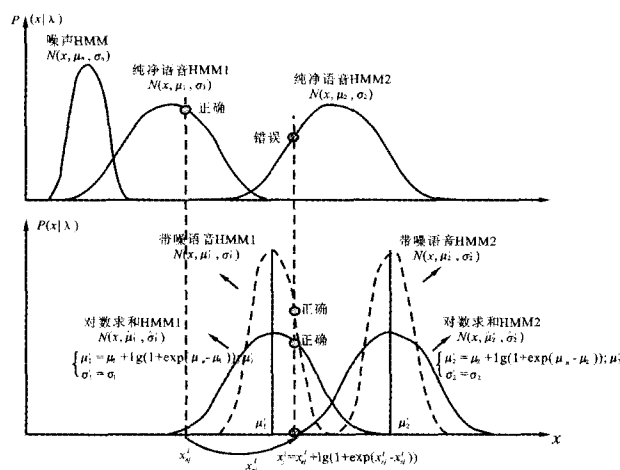


图3 对数相加近似算法

由于并行模型合并是在对数谱域进行的, 所以在对数谱域进行分析, 考虑第 j 维对数谱的情况, 假设该维特征在纯净语音模型 1、纯净语音模型 2 和噪声模型的概率密度分别为 $N(x, \mu_1, \sigma_1)$ 、 $N(x, \mu_2, \sigma_2)$ 和 $N(x, \mu_n, \sigma_n)$, x_{ij}^l 是和模型 1 对应的纯净语音的特征, 下标 j 表示第 j 个 Mel 滤波器的输出, 上标 l 表示对数谱域. 则识别器可以根据纯净语音模型对 x_{ij}^l 做出正确的判断. 但是当出现噪声时, 假设 x_{ij}^l 是噪声特征, 含噪语音的特征则由 x_{ij}^l 变为 x_{ij}' , 和 x_{ij}' 匹

配的模型应该为在此噪声情况下所训练出的含噪语音模型 1 和 2, 此时的概率密度为 $N(x, \mu_1', \sigma_1')$ 和 $N(x, \mu_2', \sigma_2')$. 如果用含噪语音模型来进行识别则系统仍然可以给出正确结果, 但是对于没有采用任何抗噪声技术的系统来说, 必然使用和含噪语音存在失配的纯净语音模型来分类, 因此难免出现将 x_{ij}' 误判为模型 2 的错误, 当噪声强度加大时, 失配程度随之加剧, 系统的识别过程盲目性就会增加, 使得识别率变得很低.

在图 3 中可以观察到加性噪声对模型均值、方差的影响为: 均值变大, 方差变小, 而模型相对距离减小. 这一现象是由 Mel 倒谱频率系数 (或 MFDWCs) 特征提取过程中对数变换的非线性造成的, 其中 x_{ij} 和 x_{nj} 是语音和噪声在第 j 个 Mel 滤波器的合并功率. 由于噪声的加入使得能量加大, 因此含噪语音对数谱均值变大. 对数变换点因能量增大而提升, 而其斜率变小, 假设噪声功率方差非常小, 则含噪语音方差会相应的被压缩. 由于模型总数不变, 因此模型之间的相对距离减小, 这体现了模型之间鉴别能力减弱.

因为语音增强后的剩余噪声具有加性背景噪声的性质, 所以采用 PMC 算法中的 Log-Add 合并过程在模型空间对剩余噪声进行补偿, 以消除其对语音识别的影响.

4 实验结果分析

语音信号的采样率为 1024 kHz, 谱分析采用时长为 32 ms、步长为 10 ms 的 hamming 窗进行, 用 FFT 计算每帧数据的能量谱. 运用模型空间的并行模型合并算法 (PMC) 和特征空间的倒谱均值归一化算法 (CMN) 提取参数, 分别得到 DCT、DWT 方法提取的 MFCCs 和 MFDWCs, 结果如表 1 所示.

表 1 综合运用 CMN 和 PMC 补偿技术的识别结果 %

噪声类型	MFCC			MFDWC		
	-5	0	5	-5	0	5
CMN	24.3	48.0	80.3	25.6	53.2	81.8
PMC	46.0	68.7	82.3	46.3	70.5	86.1
CMN-PMC	64.3	87.1	92.6	68.4	87.5	94.1

由表 1 可以看出, 综合运用倒谱均值归一化方法和并行模型合并算法对于提高噪声环境下语音识别系统的识别率非常有效.

(下转第 167 页)

将来要使用的长度达128比特的IPv6地址,可借鉴二分查找的算法以64、32为分界点构建1棵深度为2的二叉树.在此基础上,再构建步长为 $\{m, n, \dots, n\}$ 多分支Trie树,多分支Trie即为二叉树的叶子结点.这样对于任一IPv6分组,在步长为 $\{16, 4, 4, 4, 4\}$ Trie树结构下,最多情况只需要7次存储器访问次数.

5 结束语

一个好的路由查找算法,应该在查找速度、路由表更新的开销,以及所需存储空间的大小^[6]等方面综合考虑,以取得较佳的综合性能为主要目标.

本文借鉴Trie树的数据结构思想,建立了固定步长的路由表.在此基础上,由添加到路由表中的路由信息依据目的网段和掩码长度,生成了一棵转发域信息Trie树.一方面,由于采用了固定步长的多分支Trie树结构从根本上降低了树的高度,因而就提高了查找速度.另一方面,由于对转发域信息建立了一棵Trie树,在这样的数据结构下,对因路由的变化而引起的结点的更新,只是在找到相应的表项或节点后,修改相应的数据,创建或删除转发域信息Trie树上的结点.因此,更新操作的时间花销仍主要在查找,其算法复杂度与查找算法相当.综上所述,带有转发信息树的多分支Trie结构既保证了查找算法的简单性,又能保证路由信息的易更新性,其综合性能是比较理想的.它是一种具有算法简单、查

找速度快、存储空间利用率较高、便于向IPv6过渡等特点的快速IP路由查找算法.

参考文献:

- [1] 颜顺华,刘宏利.一种改进的端到端混合QoS模型[J].微电子学与计算机,2009,26(5):36-39.
- [2] 刘宝起,王长山,李瑞.基于保证QoS的片上网络路由算法[J].微电子与计算机,2009,26(7):53-56.
- [3] Yeim-Kuan Chang, Yung-Chieh Lin. A fast and memory efficient dynamic IP lookups algorithm based on B-tree [C]// International Conference on Advanced Information Networking and Applications. Bradford: IEEE,2009:278-284.
- [4] Douglas E Comer. Internetworking with TCP/IP Vol I: principle, protocols, and architectures [M]. 4ed. London: Prentice Hall, 2000.
- [5] Waldvogel M, Varghese G, Turner J, et al. Scalable high speed IP routing lookups [J]. Computer Communication Review, 1997,27(4):25-26.
- [6] 谭明锋,高蕾,龚正虎,等. IP路由查找算法研究概述[J]. 计算机工程与科学,2006,28(6):87-91.

作者简介:

邵国良 男,(1985-),硕士研究生.研究方向为无线通信与嵌入式系统.

李广军 男,(1950-),教授,博士生导师.研究面向为通信系统设计、通信专用集成电路设计、无线传感器网络、嵌入式系统设计、EDA/SoC/SOPC设计.

(上接第162页)

5 结束语

虽然语音降噪增强方法的效果很好,但这种方法会带来一定的增强失真,包括增强引起的语音频谱变形和增强后的剩余噪声,它们限制了语音增强对语音识别系统抗噪声性能的进一步提高.针对这些问题,分析了特征空间的倒谱均值归一化方法和模型空间的并行模型合并算法,从识别率上给出了不同补偿技术的补偿效果.

参考文献:

- [1] 王晶,傅丰林,张运伟.语音增强算法综述[J].声学电子工程,2005(1):22-26.
- [2] Ding Qi, Xu Wang, Xu Jinfu, et al. Speech enhancement based on hearing masking properties and subspace [J]. Signal Processing,2008,(1):307-310.

- [3] Ruhi Sarikaya. Robust and efficient technology for speech recognition in noise[J]. Abstracts International, 2009,62(10):4699.
- [4] Oma Y N B, McInnes F R, Jack M A. Improving performance of spectral subtraction in speech recognition using a model for additive noise[J]. Speech and Audio Processing,2008,(11):579-582.
- [5] 王晶,傅丰林,陈建.基于小波变换多阈值语音增强处理研究[J].声学电子工程,2010,(4):32-35.
- [6] 靳雪莲,杨益新.水下环境噪声采集系统的设计与实现[J].微电子学与计算机,2010,27(5):120-123.

作者简介:

苗玉杰 男,(1972-),硕士,讲师.研究方向为信号处理技术、计算机控制工程及算法.

刘雪飞 女,(1972-),硕士,副教授.研究方向为信号处理技术,模式识别.