

## 电话语音识别中通道补偿方法的性能比较\*

Performance Comparisons of Channel Compensation Methods in Telephone Speech Recognition

韩纪庆 王承发 高文

(哈尔滨工业大学计算机科学与工程系 哈尔滨150001)

**Abstract** A Relative Mel-Frequency Cepstral Coefficients processing method, namely RMFCC, had been proposed for improving the performance of a speech recognition system in our previous work. In this paper, we compare the performance between RMFCC and other channel compensation methods for telephone speech recognition.

**Keywords** Speech recognition, Channel compensation, Robustness

## 1. 引言

语音信号在进入语音识别系统之前,通常受到一些诸如语音产生过程、记录过程以及传输过程中产生的通道畸变的影响。例如,麦克风依赖于其类型和位置的不同能明显地影响语音信号的频谱,电话线网络的频率特性也对语音信号产生了频谱畸变的影响。这些由于传输通道而引入的频谱畸变将直接影响到语音信号的短时频谱分析结果,目前所有的语音识别系统中的参数计算,以及相似度测度都直接或间接地依赖于语音信号的短时频谱分析结果,因而上述的通道畸变将直接影响到语音识别系统的性能。

近年来,基于通道补偿的语音识别方法已经被进行了广泛的研究<sup>[1~4]</sup>。倒谱平均减 CMS(Cepstral Mean Subtraction)<sup>[1]</sup>方法,假设输入语音信号倒谱系数的总体平均为零,通过修正倒谱系数来使得由于通道畸变而引起的训练和测试环境间的不匹配达到最小。如果考虑其简单性,这是一种有效的算法。然而,当通道畸变不能被模型化为线性特性时,该方法的有效性受到了限制。为了处理非线性通道畸变,人们提出了二级 CMS 方法<sup>[2]</sup>,它对非线性通道畸变采用了简单的分段线性处理方法,即先将输入语音信号分为语音和噪声两部分,并分别求出各部分的倒谱系数的平均值,然后每一部分的倒谱系数再分别减去各自的平均值。这种方法被分别有效地用到

了汽车环境下数字识别中的通道补偿<sup>[3]</sup>,以及话者识别中的通道补偿上<sup>[4]</sup>,实验证明其性能优于 CMS 方法,尽管这种二级 CMS 方法,在解决非线性通道畸变的影响上比较有效,但它要求进行语音和噪声的分类,且系统性能依赖于分类结果,而在噪声环境下进行语音和噪声的分类,是一个比较困难的问题,因而这种方法也有其局限性。RASTA(Relative Spectral)处理技术<sup>[5]</sup>通过使用一个低端截止频率非常低的带通滤波器,可以有效地抑制变化缓慢的通道畸变。然而,以往的 RASTA 处理都是被应用到基于感知线性预测 PLP(Perceptual Linear Predictive)分析<sup>[6]</sup>的对数频谱上的,PLP 技术是对人类听觉感知特性的一种工程模拟,但它需要进行如下复杂的运算:

(1) 等响度预加重(Equal-loudness preemphasis):

$$E(\omega) = [(\omega^2 + 56.8 \times 10^6)\omega^4] / [(\omega^2 + 6.3 \times 10^6)^2 \times (\omega^2 + 0.38 \times 10^9)] \quad (1)$$

其中  $\omega$  是频带号,

(2) 强度响度幂律(Intensity-loudness power law):对经等响度预加重后的各频带频谱做一个求3次方根的运算。

(3) 使用自回归模型(Autoregressive model)进行线性预测分析或接着进行倒谱分析。

我们曾在先前的研究中<sup>[7]</sup>,提出了一种相对 Mel 倒谱系数的处理方法 RMFCC(Relative Mel-

\* 本文得到了哈尔滨工业大学回国人员基金的资助。

Frequency Cepstral Coefficients), 用于语音识别中的通道补偿以提高识别系统性能。该方法基于传统的 RASTA-PLP 方法, 它用 Mel 频谱分析代替 PLP 频谱分析以减少计算代价。而后, 基于 Mel 频率对数频谱与其倒谱系数 MFCCs 间的线性关系, 将 RASTA 处理从 Mel 频率对数频谱扩展到 MFCCs 上, 通过使用一个类似于 RASTA 的带通滤波器来进行稳健的 (robust) 语音识别。该方法既改进了识别系统的性能, 又减少了计算代价。本文分析比较了 RMFCC 与其它通道补偿方法用于电话语音识别时的性能改进情况。

## 2. RMFCC 与其它方法的性能比较

我们采用 RMFCC 方法与其它通道补偿方法分别进行了电话语音识别的实验, 所有实验都是采用语音库进行的, 全部语料是从韩国的汉城和大田两个城市的市内电话线网络中, 使用不同类型的话机采集来的 84 个韩国语孤立词的多个话者数据。其中, 训练语音库包含了 22 个男性话者和 18 个女性话者, 测试语音库包含了不同于训练语音库的 22 个男性话者和 18 个女性话者。音库中的男女比例基本上反映了韩国的实际人口分布情况。每个话者分别朗读 93 个句子, 而后经手工分割出各个孤立词。由于舍弃了一些录音较差的发音, 所以整个训练语音库中包含 11381 个发音, 测试语音库中包含了 8036 个发音。

实验中使用一个不特定话者连续密度的 HMM 系统作为基本的识别系统, 以此来比较有关通道补偿方法的有效性。输入的语音信号先以 8kHz 采样率进行数字化, 而后用滤波器  $H(z) = 1 - 0.98z^{-1}$  进行预加重, 每 30ms 使用一个 240 个样本点的汉明窗, 加窗后的信号计算其 256 点的 DFT (离散傅里叶变换), 并求其功率谱。接着进行 Mel 频谱分析, 求得 40 个 Mel 频段上的对数频谱, 然后求其 12 个 MFCCs 系数及其 12 个 MFCCs 系数的一阶差分, 以此作为每帧的特征参数。

本文的实验中, 使用系统误识率来代表识别系统的性能, 而在进行方法间的性能比较时, 使用的是误识率的下降率, 如果用  $\Omega$  来代表误识率的下降率, 则其定义为:

$$\Omega = \frac{ERR_{\text{baseline}} - ERR}{ERR_{\text{baseline}}} \times 100\% \quad (2)$$

其中  $ERR_{\text{baseline}}$  代表被作为比较基准的方法所对应的系统误识率, 而  $ERR$  代表要与基准方法进行比较

的方法所对应的系统误识率。

采用全部语音库, 对所提出的通道补偿方法 RMFCC 与以往的通道补偿方法进行了性能比较实验。表 1 中列出了使用基本识别系统、采用 CMS、二级 CMS、RASTA-PLP ( $\rho = 0.98$  时)、RMFCC 等方法时的误识率, 以及采用上述方法的系统与基本识别系统相比, 误识率的下降情况。在二级 CMS 方法中, 语音和噪声的分类器系数是基于先前的研究<sup>[3]</sup>而选择的。在文[7]中已经讨论了使用线性回归导出的一阶动态倒谱系数 delta-MFCCs 与 RMFCC 有某种内在的联系, 所以, 这里把使用 delta-MFCCs 也作为一种通道补偿方法单独列在表 1 中。这样, 表中的基本系统为没有使用动态特征的“Baseline”系统, 而使用了 delta-MFCCs 的系统为“Baseline- $\Delta C$ ”系统。

表 1 使用不同通道补偿方法时的识别结果

方法	误识率	误识率的下降率
Baseline	11.8%	—
Baseline- $\Delta C$	9.9%	16.1%
CMS	7.8%	33.9%
二级 CMS	7.2%	38.9%
RASTA-PLP	7.1%	39.8%
RMFCC	7.1%	39.8%

从实验的情况看, RMFCC 的系统性能明显优于基本识别系统, 与其相比误识率下降了 39.8%。与使用 Baseline- $\Delta C$  的系统相比, RMFCC 的误识率下降了 28.3%, 而计算代价将有所增加。对一个长度为  $L$  帧的语音, RMFCC 与 Baseline- $\Delta C$  相比增加了  $L$  个乘法和  $L$  个加法。而当用数字信号处理器 DSP (Digital Signal Processor) 实现该算法时, 其所增加的只是  $L$  个乘累加运算。对大多数 DSP 来说, 乘累加运算都是单指令周期完成的, 因此 RMFCC 所增加的并不是很多, 通常都是毫秒级的。

RMFCC 的性能也优于 CMS 和二级 CMS。尽管与二级 CMS 相比, RMFCC 并没有明显改进系统性能, 但它不需要长时估计传输通道的频率特征, 这一点优于 CMS 和二级 CMS; 因为后两者都是后处理算法, 都需要先计算整个发音的倒谱系数平均值以估计传输通道的频率特征, 然后再逐帧减去该平均值, 更主要的在于二级 CMS 需要进行语音和噪声的分类, 而 RMFCC 是一个直接 (straightforward) 的实时算法, 它可以直接使用到任何的测试环境中。

与 RASTA-PLP 相比, RMFCC 在不损失精度

的条件下减少了计算代价。假设 PLP 分析和 Mel 分析在做 Bark 分割和加窗时的运算是相近的,同时 PLP 分析在做完频谱域上的各种听觉感知工程模拟后,不是使用自回归模型做线性预测分析求倒谱系数,而是像 Mel 分析一样,直接进行 DCT(离散余弦变换)变换来求倒谱系数,则 PLP 分析比 Mel 分析也多进行了等响度预加重和强度响度幂律运算。前者对每一帧的每一个频带都要乘以一个如式(1)那样比较复杂的函数;即使事先已经求好该函数在各个频带的值,然后采用查表的方式,对一个发音长度为  $L$  帧的语音,如果使用的频带数为  $B$ ,则需要多进行  $L \times B$  个乘法;后者对每一帧的每一频带都要进行一次求3次方根的运算,所以对  $L$  帧的语音将增加  $L \times B$  个求3次方根的运算。因而,RMFCC 的计算代价明显小于 RASTA-PLP 处理。在一个由 Sun-SPARC 工作站与 S-32C 数字信号处理板构成的主从式系统上,采用 RASTA-PLP 处理,在对全部测试语音库中的数据进行测试时,单一发音的平均处理时间为1秒多,而相应的 RMFCC 的处理时间约为0.6秒。

综上所述,如果同时考虑系统性能和计算代价, RMFCC 方法是上述所有的通道补偿方法中最好的一个。

**结束语** 本文分析比较 RMFCC 与其它通道补偿方法用于电话语音识别时的性能改进情况。可以看出, RMFCC 优于传统的通道补偿方法,它具有不需要长时估计通讯环境的频率特征、计算简单,以及非常适合实时计算等优点,可以很容易地连接到现有大多数语音识别系统上。

**致谢:**感谢韩国科学院系统工程研究所语言理解实验室的 Park Gyu-Bong、Han Munsung、Park Jeongue 等协助完成本实验。

### 参考文献

- 1 Furu S. Cepstral analysis technique for automatic speaker verification. *IEEE Trans. on Acoustics Speech and Signal Processing*, 1981, 29(4): 254~272
- 2 Sankar A, Lee C. Robust Speech Recognition Based on Stochastic Matching. In: *Proc. of 1995 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*. Michigan, USA, Diane Drago. 1995. 121~124
- 3 Gupta S, Soong F, Hammi-Cohen R. High Accuracy Connected Digit Recognition for Mobile Applications. In: *Proc. of 1996 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*. Atlanta, USA, Vijay K. Madisetan. 1996. 57~60
- 4 Reynolds D. The Effects of Handset Variability on Speaker Recognition Performance: Experiments on the Switchboard Corpus. Same to [3]. 113~116
- 5 Hermansky H, Morgan N. RASTA Processing of Speech. *IEEE Trans. on Speech and Audio Processing*, 1994, 2(4): 578~589
- 6 Hermansky H. Perceptual Linear Predictive (PLP) Analysis of Speech. *J. of Acoustical Society of America*, 1990, 87(4): 1738~1752
- 7 韩纪庆, 高文. 基于 MFCCs 滤波的电话语音识别的通道畸变补偿方法. *计算机学报*, 1998(录用待发表)
- 8 韩纪庆, 王承发, 高文. 二阶 CMS 用于电话语音识别的通道补偿. *哈尔滨工业大学学报*, 1999, 1(待发表)

(上接第59页)

算法,运用于 CBR 的改写中,使得具体事例所蕴含的客观规律可以得到很好的体现,这样,既避免了收集专家经验的庞大工作量及进行知识表达的困难,又克服了单纯基于规则进行推理的片面性、静态性。由此,关于新问题的解答,可以更加切合实际情况。

当然,为提高 CBR 的实用性、科学性,还有许多工作要做,比如:影响因素的确定、提取相似事例的方法等。这些都是我们今后要着重研究的。

### 参考文献

- 1 Leak D B, et al. Case-Based CBR: Capturing and

- Reusing Reasoning About Case Adaptation. *Int. J. of Expert Systems*, 1997, 10(2): 197~213
- 2 Gondran M, Minoux M. *Graphs and Algorithms*. U K, 1984
- 3 Watston I. *An introduction to case-based reasoning*. U K, 1995
- 4 Jiao Y, Hudson J A. Fully-Coupled Model for Engineering Systems. *Int. J. Rock Mech. Min. Sci. & Geomech Abstr.*, 1995, 32(5): 491~521
- 5 Cecil I O S. Correlation of bolt-shotcrete support and rock quality parameters in scandinavian tunnels: [Ph. D. Thesis]. *Unt. of Illinois, Urban*, 1970