

一种新的修正 Mel 频率映射的应力影响下变异语音识别方法^{*}

A New Method of Modified Mel Frequency for Recognition of Stressed Speech Under G-force

毕继武 韩纪庆

(哈尔滨工业大学计算机科学与工程系 哈尔滨150001)

Abstract Based on the spectrograms analysis and the individual frequency bands of speech under G-force, in this paper, a new Mel frequency scale is proposed, and the related MFCC (Mel Frequency Cepstrum Coefficient) is adopted as the features for recognition of stressed speech under G-force. It is shown from the experiments that the proposed method is better than other methods of Mel-based features for stressed speech recognition.

Keywords Stressed speech recognition, G-force, Mel frequency, Mel frequency cepstrum coefficient

1 引言

话者在心理紧张和情绪变化时语音会发生变异,如在身体不适(如感冒)时、恐惧、愤怒、环境噪声影响下,以及生理情况受到影响(如加速度变化)时,其语音都会发生改变。发音变异可以引起各语音参数的不同变化,以致常规语音识别系统的识别率大大下降。从八十年代末开始,就有人研究顽健(robust)语音识别问题,针对噪声下的研究工作已经取得了非常显著成效,但对应力影响下变异语音识别的研究则很有限。应力影响下变异语音识别是一个难度较大的问题,因为应力影响同人的发音过程具有直接的物理关系,其所引起的发音变异有别于单一的因心理的、感知的或生理的因素所引起的变异。

为了提高变异情况下语音识别算法的性能,人们提出了许多方法。这些方法可归纳为三类^[1]:1)顽健特征方法,2)变异补偿方法,3)模型调整或训练方法。为了克服变异语音数据有限这一不利因素,本文将研究的重点集中在采用顽健特征的方法上。

Mel 频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)^[2]是语音识别中采用的主要特征参数,它是根据人耳的听觉特性和掩蔽效应原理^[3],在有关的心理实验基础上发现,人耳对不同频段信号的敏感程度不同,因而采用了一种 Mel 频率划分方法,并在此基础上,经过离散余弦变换而求出的倒谱系数。在噪声环境下 MFCC 表现出很强的顽健性(robustness)。通过对应力影响下变异语音的分析,我们发现,其共振峰的位置和带宽都发生了改变。所以当语音信号发生变异时,一种可能的改进语音识别性能的方式,就是改变 Mel 频率映射函数,从而使 MFCC 参数特征获得更好的顽健性。本文正是根据这样的想法,设计了相应的 Mel 频率映射函数来进行变异语音识别。

2 应力影响下变异语音库^[4]

应力影响下的变异语音数据是在航空模拟飞行器中,实验者通过头戴麦克风将由加速度变化引起的 G-Force 下的变异语音录制到录音机,然后再利用声卡输入到计算机中,以 PCM 格式存储在硬盘上。由于条件限制,分别采集了在3g、

4g、5g、6g 四种超重情况时的变异语音数据。参加数据采集的话者是空军第四研究院的两名江苏籍战士,他们的年龄分别为18岁和21岁。

在采集应力下的变异语音时,由于受到实际条件与人体承受能力所限,因此选用了具有代表性的0—9十个数字组成的小词表。3g、4g、5g、6g 四种情况下每个词读3遍,这样每个话者共12遍变异语音样本数据。对应的正常情况(1g)下的语音数据选用了0—9十个数字及十二个单字词,十四个双字词,四个三字词,每个词11遍作为正常语音样本数据。

3 应力影响下变异语音分析

3.1 语音谱分析

我们先对典型变异语音的语谱图进行了分析。图1中分别给出了话者1在1~6g 下“1”的语谱图^[5]。从图中可以看出,在超重情况下,语音高频部分能量明显增强,共振峰位置向高频部分移动,第一共振峰不再明显。第二共振峰位置上移,带宽加宽。因此,使用传统的基于 Mel 频率映射的 MFCC 特征将不适合这种变异情况下的语音识别。我们模仿 Mel 频带划分的方法,通过分频带实验来考核各个频带的识别性能。然后根据实验结果,对性能优异的频带进行加强,从而提高语音识别方法的性能。

3.2 分频带分析

图2给出了正常语音训练/识别、正常语音训练/变异语音识别、变异语音训练/识别三种情况下,根据分频带测试的结果^[6]。可以看出,正常语音训练/识别,其识别率最大值出现在500Hz 处,在500~1750Hz 这段频带曲线变化比较平稳,可以说这些频带的特征对于正常语音识别的影响较大,因此传统的 Mel 频带在这部分划分得较细。正常语音训练/变异识别时,其识别率的最大值出现在1500Hz,识别率较大的频带集中在1000~1500Hz,部分数据在1750点处也表现出较高的识别率。变异语音训练/识别时,因第二共振峰位置向高频方向移动,最大值出现在1750Hz 处,同样识别性能较好的频带也相对后移,主要集中在1000~2000Hz 频段。

基于上面的分析结果,我们着重研究了对 Mel 频带重新划分来提高变异语音识别性能的方法。

^{*}国家自然科学基金资助项目(项目批准号:60085001),教育部留学回国人员科研启动基金资助项目。毕继武 硕士研究生,主要研究方向为语音识别。韩纪庆 教授,博士。

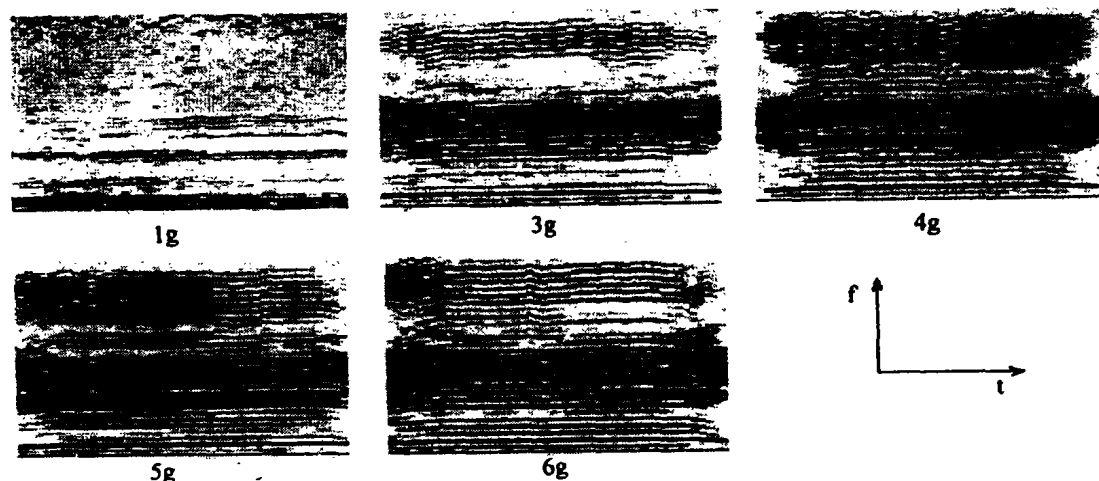


图1 G-Force 变异情况下话者1发音“l”的语谱图

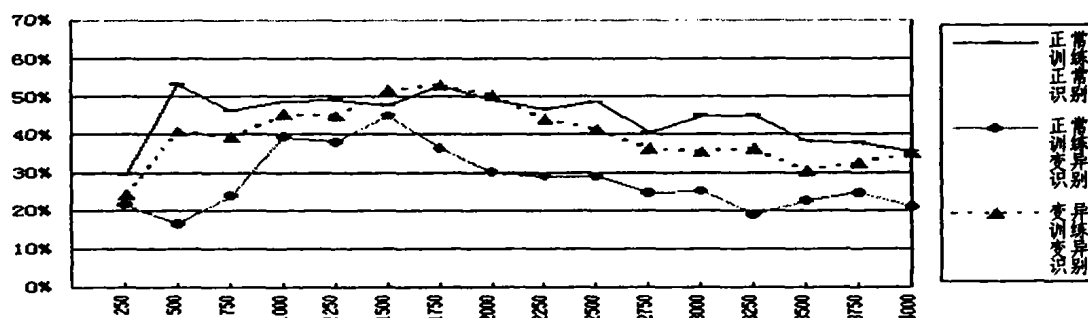


图2 分频带测试结果

4 修正 Mel 频率映射的方法

4.1 Mel 频率倒谱系数

Mel 频率倒谱系数,作为一种具有顽健性,且较能忠实地反映语音特性的参数,在语音识别中得到了广泛的应用。它是将信号的频率轴通过映射函数变换为 Mel 频率刻度,再变换到倒谱域而导出的倒谱系数。其中 Mel 频率的划分是引入人耳的一些听觉特性而建立的。由于人耳对于声波频率高低

的感觉与实际频率的高低不成线性关系,而是近似为对数关系,因此在进行声学测量时,频率刻度常常取为非线性刻度,其映射函数形式为:

$$mel_scale = 2595 \times \log\left(1 + \frac{f}{700}\right) \quad f=0, \dots, 4000Hz \quad (1)$$

其中 f 为频率,它将 0~4000Hz 频率映射为 16 个等分的 Mel 频带。

图3给出了 Mel 频率映射函数曲线,可以看出在 1000Hz 以下频率轴划分较细。

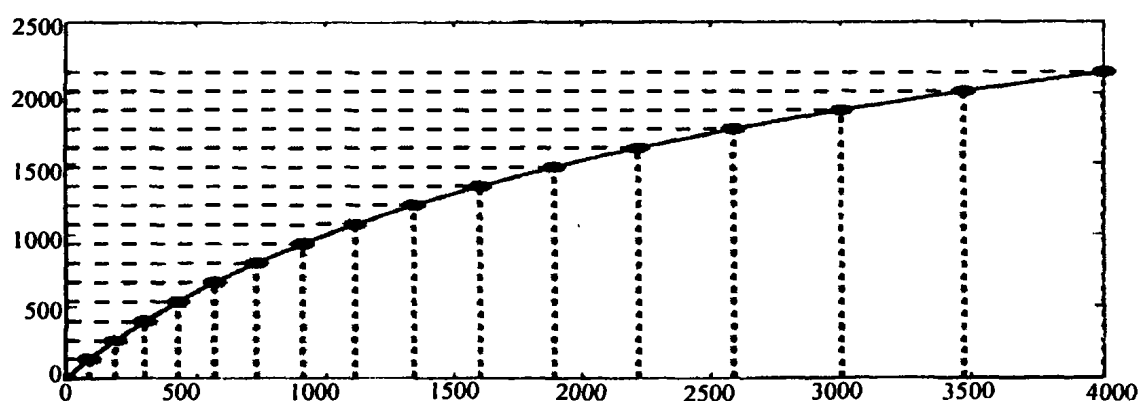


图3 传统 Mel 刻度映射函数曲线

4.2 修正 Mel 频率映射的方法

通过前面的分析,我们期望在正常语音识别中突出 500~1750Hz 频带的特性,这可以通过改变曲线的形式来完成。

前面的(1)式可以写成通用的形式:

$$mel_scale = C \times \log\left(1 + \frac{f}{k}\right) \quad f=0, \dots, 4000Hz \quad (2)$$

保持曲线两端点位置不变,代入端点值(4000, 2146),可

以将(2)式写成如下形式:

$$2146 = C \times \log(1 + \frac{4000}{k}) \quad (3)$$

映射函数的曲线形式会因 k 值的调整而改变,各个频带的划分也会因曲线形状的变化而移动。 k 值减小,曲线上移,1000Hz 以下频带划分更细,不符合要求。 k 值增大,曲线下移,这样频带划分在 0~2000Hz 区间分布的点较为平均。但 k 不能无限地增大,否则会过分趋近 $y = (2146/4000) \times f$ 这一线性函数,这是不符合人耳的听觉特性的。实验表明当 k 取 1000 时,对正常语音的识别效果较好,同时对变异语音的识别效果也有所提高。这样就可获得新的频率变换形式:

$$M_{\text{mel-scale}} = 3070 \times \log(1 + \frac{f}{1000}) \quad f=0, \dots, 4000\text{Hz} \quad (4)$$

然而这种变换不能很好地突出变异语音识别中性能较好的 1000~2000Hz 频带的特性。为此,我们借鉴 Bou—Ghazale^[2] 将映射函数划分为两段进行描述的思想,对该函数的前半段采用指数形式,后一段保持对数形式,这样就可以突出中

频段的作用。因为性能较好的频带都在 2000Hz 以下,所以我们就将两部分的分界点设在 2000Hz。

用如下的公式来分别表示指数对数形式:

$$y_1 = C_1 \times (10^{f/k_1} - 1) \quad 0 \leq f \leq 2000\text{Hz} \quad (5)$$

$$y_2 = C_2 \times \log(1 + f/k_2) \quad 2000 < f \leq 4000\text{Hz} \quad (6)$$

将式(6)表示成式(4)的形式。令 $f = 2000\text{Hz}$, 式(4)、(5)相等,根据经验将 C_1 取值 700。可以求得如下的分段映射函数形式:

$$M_{\text{Expolog}} = \begin{cases} 700 \times (10^{f/1000} - 1) & 0 \leq f \leq 2000\text{Hz} \\ 3070 \times \log(1 + \frac{f}{1000}) & 2000 < f \leq 4000\text{Hz} \end{cases} \quad (7)$$

图 4 给出了新的映射函数曲线,可以看出,它对 1000~2000Hz 之间中频带的划分比较细,因此计算出的 MFCC 特征参数也能够充分体现出这一频段特征的作用,从而提高变异语音识别的性能。

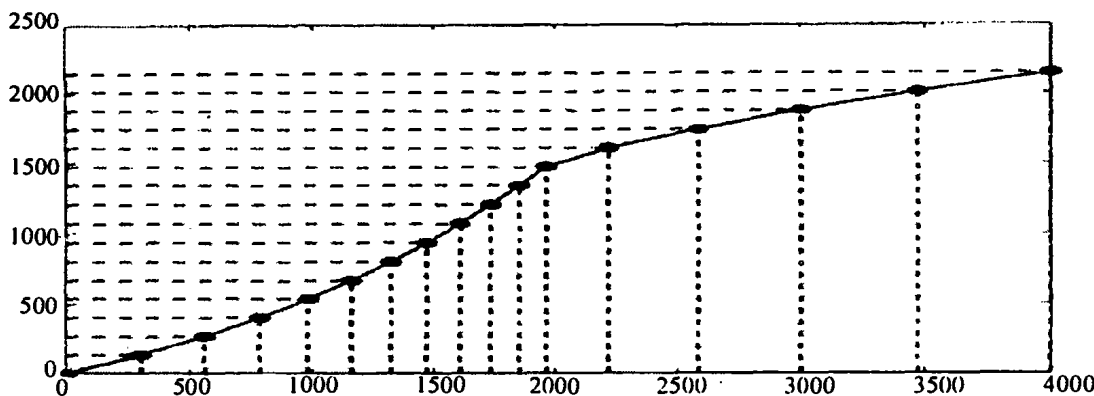


图 4 修正的分段 Mel 映射函数曲线

5 实验情况

我们用实验考核了用 Mel 频率映射的方法对变异语音识别性能的影响。实验中,比较了分别使用传统的 Mel 频率划分、M-Mel 频率划分、Expolog 频率划分,以及本文研究的 M-Expolog 频率划分时的性能。采用的 MFCC 特征参数,分别是在上述各个基础上通过对数运算和离散余弦变换而求出的。实验中语音信号以 16kHz 的采样率量化为 16bit 的数字量,每帧 256 点,帧移 128 点。窗函数采用汉明窗。每一实验的帧特征中包括能量、12 阶 MFCC,以及它们的一阶、二阶差分,共 39 维特征。识别器选用的是基于 HMM 的识别器。表 1 给出了两个话者采用上述几种特征时不同的实验结果。

从表 1 的实验结果可以看出,应用 M-Mel 映射函数提取的特征,因为平均地突出了 2000Hz 以下频带的特征,所以其结果比应用传统的 Mel 刻度映射函数识别效果都要好一些,但是提高幅度不是很大。在所有的特征中,识别效果最好的还是修正的 M-Expolog,因为它充分考虑了 1000~2000Hz 频带的作用,满足变异语音识别中突出这部分频带的要求,所以它对变异语音的识别效果都要明显好于前面各种频率划分方法。在正常语音训练/变异识别时,两话者识别率平均提高了 10.4%;在变异语音训练/测试时,两话者识别率平均提高了 9.2%,其中话者 1 在使用 M-Expolog 频率划分时,提高效果非常明显,比应用传统的 Mel 频率划分时,这两项实验的识别率分别提高了 14.2%、11.7%。

表 1 两个话者采用不同 Mel 频率映射时的识别结果

频率映射形式	正常训练/识别			正常训练/变异识别			变异训练/识别		
	话者 1	话者 2	平均	话者 1	话者 2	平均	话者 1	话者 2	平均
Mel 刻度	89.1%	82.4%	85.8%	50.0%	34.2%	42.1%	75.0%	77.5%	76.3%
M-Mel 刻度	91.8%	85.1%	88.5%	54.2%	36.8%	45.5%	79.0%	80.0%	79.5%
Expolog 刻度	88.1%	83.6%	85.9%	58.3%	39.2%	48.8%	82.5%	83.3%	82.9%
M-Expolog 刻度	91.8%	85.5%	88.7%	64.2%	40.8%	52.5%	86.7%	84.2%	85.5%

结论 本文在对应力下变异语音的语谱图和分频带进行分析的基础上,重新调整了 Mel 刻度映射函数形式,设计了在变异语音识别中顽健性能较好的修正 M-Expolog 映射方法,通过实验比较了这种方法与先前方法在语音识别中的性

能。实验结果表明,采用修正的 M-Expolog 方法,在变异语音识别中有较好的顽健性,较传统的 Mel 映射方法识别率有明显提高。

致谢 感谢吕成国、马永林硕士提供的变异语音分析实

非线性判别函数的基于 MDL 标准的勒让德多项式构造法^{*}

A Constructing Method of Non-linear Discriminating Function Based on MDL Criterion Using Legendre Polynomials

翟正利 孙忠林 吴哲辉

(山东科技大学信息科学与工程学院 泰安271019)

Abstract In Model Recognition, although a non-linear discriminating function may be superior to linear or quadratic classifiers, it is difficult to construct. This paper proposes a constructing method using Legendre polynomials. The selection of an optimal set of Legendre is determined by the MDL criterion. Results using many data show the effectiveness of this method.

Keywords Non-linear discriminating function, Legendre polynomials, MDL

1. 引言

在模式识别中,分类决策问题是最基础也是最重要的内容,所谓分类决策就是根据被识别对象特征的观察值将其分到某个类别中去,其基本作法是在样本训练集基础上确定某个判决规则(即样本特征空间的一个函数,也称为判别函数或分类器),使按这种判决规则对被识别对象进行分类所造成的错误率最小或引起的损失最小。

当样本分布是正态分布时,从最小错误率来说二次分类器是最优分类器。此外,作为一种特殊情况,如果这些分布有相同的协方差矩阵,那么线性分类器就是最优的^[1]。然而这些假定在很多实际问题中并不成立,因此线性或二次分类器对大量的问题是不能胜任的。另一方面,k-近邻或分段线性分类器不假设样本分布的统计模型并且决策规则是基于训练样本的,然而,在这两种方法中,决策规则通常太复杂以至于对未知样本无法工作。总之,线性和二次分类器显得太简单,k-近邻和分段线性分类器显得太复杂。

尽管非线性判别函数已被证明是理论上比较有效的,然而如何构造非线性判别函数却比较困难。本文基于最小描述长度(简记为 MDL)标准和勒让德多项式的正交性质给出了构造非线性判别函数的一种较为有效的方法。对许多数据的试验结果表明了该方法的有效性。

2. 非线性判别函数

首先考虑两类问题,即已知样本类别共两类(记为 ω_1 和 ω_2),要将被识别对象归为其中之一。判别函数分两个阶段来构造:①定义域的变换;②用多项式构造之。主要思想是:先把特征向量的定义域 x 变换为 y ,此变换为非线性变换,即 y 是 x 的一个非线性函数;然后构造 y 的一个线性函数作为判别

函数。例如,已经知道“排斥或”问题能通过定义域的变换来解决。在二维空间中,当两个样本(0,0)和(1,1)来自 ω_1 ,另外两个样本(1,0)和(0,1)来自 ω_2 时,这些样本不能被一个线性分类器判别。然而,我们可通过定义域的变换

$$y_1 = x_1, y_2 = x_2, y_3 = x_1 x_2$$

构造一个线性分类函数

$$f(y) = y_1 + y_2 - 2y_3 - 1/3$$

来对其分类。

2.1 定义域的变换

设一个从 R^n 到 R^m 的定义域变换:

$$x = (x_1, x_2, \dots, x_n)^T \rightarrow y(x) = (y_0(x), y_1(x), \dots, y_m(x))^T$$

其中, $y_0(x)$ 是一个常量, $y_i(x)$ 是 x 的变换函数($i=1, \dots, m$)。

此时,我们用勒让德多项式来表示 $y_i(x)$, r -次勒让德多项式为:

$$P_r(x) = \frac{1}{2^r r!} \frac{d^r (x^2 - 1)^r}{dx^r},$$

在 $[-1, 1]^n$ 上的规范化勒让德多项式为:

$$Q_r(x) = \frac{P_r(x)}{\|P_r(x)\|},$$

其中, $\|P_r(x)\|^2 = \int_{-1}^1 P_r^2(x) dx$ 。

由勒让德多项式的正交性质知,

$$\{Q_1(x), Q_2(x), \dots, Q_r(x), \dots\}$$

构成了最简明的标准正交基。现定义 $y_i(x)$ ($i=0, 1, 2, \dots, m$) 如下:

$$y_0(x) = 1/\sqrt{2}$$

$$y_i(x) = Q_{r'_1}(x_1) Q_{r'_2}(x_2) \dots Q_{r'_n}(x_n)$$

$y_i(x)$ 由序列 $r' = (r'_1, r'_2, \dots, r'_n)$ 指定, $r'_j \in \{0, 1, 2, \dots\}$, y_i

(x) 的次数定义为 $\deg(y_i(x)) = \sum_{j=1}^n r'_j$, 则 $\{y_i(x)\}$ 构成了 $[-1$

^{*} 国家自然科学基金资助课题(60173053)。翟正利 讲师,硕士研究生,主要研究方向为 Petri 网理论与应用。孙忠林 副教授,主要研究方向为图像识别与数据库理论。吴哲辉 教授,博士生导师,主要研究方向为 Petri 网理论与应用、算法设计与分析等。

验结果。

参考文献

- 1 Bou-Ghazale S, Hansen J H L. A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress. IEEE Trans. Speech and Audio Processing, 2000, 8(4): 429~442
- 2 Picone J W. Signal Modeling Techniques in Speech Recognition.

Proceedings of IEEE, 1993, 81(9): 1215~1247

- 3 易克初,田斌,付强. 语音信号处理. 国防工业出版社,2000
- 4 吕成国,周健,诸光. 高噪声有变异语音库的建立. 见:第五届全国人机语音通讯学术会议论文集. 1998. 377~381
- 5 吕成国,王承发,张磊,韩纪庆. G-stress 和 Lombard 影响下变异语音谱图. 高技术通讯,2000(增刊):377~381
- 6 马永林,韩纪庆,张磊,王承发. 基于修正 Mel 频率映射的应力影响下语音识别方法. 计算机工程与应用,2002(录用待发表)