语音质量客观评价的一步策略

付 强[1],易克初[1],田 斌[1],张知易[2]
(1. 西安电子科技大学 ISN 国家重点实验室,陕西西安 710071;2. 西南通信研究所,四川成都 610041)

摘 要: 本文提出了一种基于一步策略的语音质量客观评价新方法. 它利用前向神经网络的多维非线性映射原理,将传统的包含平均失真测度计算和平均失真测度到 MOS 的映射这两个步骤合为一步来实现. 其特点是可以充分地反映人耳听觉系统的感知特性,且计算简便. 在统计学意义上它还是 MOS 的一致估计. 主客观评价结果的相关性实验表明,一步法的相关系数可达到 0.95 左右,明显优于两步法.

关键词: 语音质量客观评价; 神经网络
中图分类号: TN912.3 **文献标识码**: A **文章编号**: 0372-2112 (2001) 07-0885-03

One-Step Strategy of Speech Quality Objective Assessment

FU Qiang[1], YI Ke-chu, TIAN Bin[1], ZHANG Zhi-yi[2]
(1. *National Key Laboratory of ISN, Xidian University, Xi'an, Shanxi* 710071, *China*;
2. *South-West Communication Institute, Chengdu, Sichuan* 610041, *China*)

Abstract: This paper proposes a novel method for objective assessment of speech quality based on one-step strategy. Using the theory of multi-dimension non-linear mapping of the feedfoward neutral network, the method combines the traditional two steps of average distortion computing and mapping from the average distortion value to the Mean Opinion Score (MOS) into one step. It can embody adequately the perception properties of the human auditory system with simple computing and also is a unanimous estimate for MOS in statistical sense. Experimental results show that the correlation coefficient between the subjective test score and objective MOS estimate of one-step method can reach up to about 0.95, which is obviously better than that of the conventional two-step methods.

Key words: speech quality objective assessment; neural network

1 引言

语音质量的客观评价一般都是建立在度量待测系统的输入语音(原始语音)和输出语音(失真语音)之间某种差异的基础上,采用两步法实现[1~3]:第一步计算原始语音和失真语音特征之间的平均失真值;第二步由在统计回归分析基础上构造的主客观评价相关模型,找出对应的主观评价得分,如平均意见得分(MOS)等. 两步法的主要缺陷在于其性能依赖于语音特征的选择和失真测度的定义,然而对听觉系统对语音的感知进行数学描述是困难的. 同时,回归分析需要精确指定待分析变量之间的函数形式,这往往也是较为困难的.

本文所提出的语音质量客观评价方法的一步策略,将语音特征的平方误差空间到一维失真测度的非线性映射,和失真测度到主观得分估计的非线性参数回归分析过程,整合到一个前向神经网络当中实现.

2 语音质量客观评价的一步策略

2.1 一步法基本原理

一步法用一个具有 M 个输入节点,1 个输出节点的前向神经网络来实现. 该网络的第 i 个输入为 $e_i = [(x_{1i} - x'_{1i})^2, (x_{2i} - x'_{2i})^2, \cdots, (x_{Mi} - x'_{Mi})^2]$,即原始语音特征与失真特征的平方差值矢量. 训练时取每一个 e_i 所对应的 MOS 为其期望输出,设某种失真条件的第 i 个输入矢量 e_i 相应的网络输出值为 m_i,这种失真条件的 MOS 估值就是该失真条件下所有 m_i 的统计平均.

2.2 多层感知器语音质量评价系统

用一个 $M \times K \times 1$ 的三层感知器来构成客观评价系统,其表达式为

$$m_i = 5 - y_i = \sum_{j=1}^{N}\left(w_j \cdot f\left(\sum_{m=1}^{M}\left(w_{mj}[x_{mi} - x'_{mi}]^2 \right) \right) \right) = W_2 \cdot [f(W_1 \cdot e_i)]^T \quad (1)$$

式中,$M = 14$ 为 MFCC 矢量的维数,N 是隐节点的个数. $W_1 = [w_{mj}]^T$, $W_2 = [w_j]^T$, $f(x) = 1/1 + e^{-x/\beta}$,$\beta$ 用于控制曲线的陡度. 系统的结构图如图 1 所示.

2.3 径向基函数网络语音质量客观评价系统

RBFN 需三层节点:输入层、隐层和输出层,每个隐节点

收稿日期:2000-04-24;修回日期:2000-12-08
基金项目:国家自然科学基金(No. 69872027)

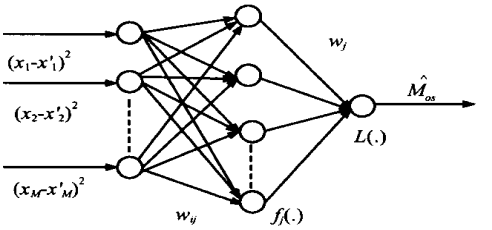

图 1 多层感知器语音质量客观评价系统

对接受的输入模式估值一核函数,估值结果作为隐节点的输出,网络的输出是所有隐节点的核函数输出的加权和. 隐节点的核函数一般取为径向对称的高斯函数,即

$$f_j(e_i)=\exp\left[-\frac{1}{2}(e_i-c_j)^T\Sigma_j^{-1}(e_i-c_j)\right],\ j=1,2,\cdots,N \tag{2}$$

这里 c_j 和 Σ_j^{-1} 分别为第 j 个基函数的中心矢量和协方差矩阵的逆,N 为隐节点的个数. 由于 x_i, x'_i 均为 MFCC 参数,各维分量相互独立,所以 e_i 各维分量也相互独立,故 Σ_j 采用对角阵,为简化问题取 $\sigma_j^2=\sigma_{j0}^2=\cdots=\sigma_{jm}^2=\sigma_{jM}^2$,对每一个 e_i,系统的输出为

$$m_i=5-y_i=\sum_{j=1}^{N}w_jf_j(e_i) \tag{3}$$

RBFN 参数的训练时间是多项式级的,比相当规模的 BP 网要快 2 - 3 个数量级[4].

3 一步策略的统计性能分析

一步法采用的统计平均值为 I 个相互独立的随机变量 $m_i(i=1,2,\cdots,I)$ 的算术平均,即

$$M_{os}=\frac{1}{I}\sum_{i=1}^{I}m_i=\frac{1}{I}\sum_{j=1}^{J}\left(\sum_{i=1}^{I_j}m_i\right)\approx\frac{1}{J}\sum_{j}^{J}\left(\frac{1}{I_j}\sum_{i=1}^{I_j}m_i\right)=\frac{1}{J}\sum_{j}^{J}M_{os}^j \tag{4}$$

进一步分解,其首先是对较小的随机变量集合$\{m_i\mid i=1,2,\cdots,I_j\}$ 做算术平均,结果为 M_{os}^j,再对 J 个 M_{os}^j 取算术平均. 这里设 I_j 大致相等,于是 $I\approx J\cdot I_j$. 在 I 是个大数时,I_j 也应是个大数,根据中心极限定理[5],随机变量 M_{os}^j, $j=1,2,\cdots,J$ 都将趋向于一维高斯分布. 当分组适当时,这 J 个高斯变量的均值 μ_i 和方差 σ_j^2 可做到近似地相同. 而统计量 M_{os} 仍为高斯随机变量,其均值与原均值 μ_i 相等,而方差仅为原方差的 $1/J$[5],即 $\sigma_{MOS}^2=\sigma_j^2/J$,即 M_{os} 是 MOS 的无偏一致估计[5].

传统的两步法的统计平均是在失真值一级进行的,由于其趋势接近于倒数规律[1~3],为简化讨论,令

$$M'_{os}=1\Big/\left(\sum_{i=1}^{I}d_i/I\right) \tag{5}$$

既然平均失真的倒数是 MOS 估值,那么每个失真值的倒数也应该是它相应的 MOS 值,

$$m'_i=\frac{1}{d_i} \tag{6}$$

于是,统计量 M'_{os} 可表示为

$$M'_{os}=1\Big/\left(\frac{1}{I}\sum_{i=1}^{I}\frac{1}{m'_i}\right)=I\Big/\sum_{i=1}^{I}\frac{1}{m'_i} \tag{7}$$

上式是一种调和平均,所以它不是 MOS 的一致估计.

4 实验结果

本文按照汉语的统计特性经过声学平衡建立原始语音库,失真语音库则包括了原始语音通过 ADPCM、CVSD、G.729 系列、G.728、IMBE 和 LPC-10 等在内的多种语音编码器以及它们的若干级联形式,以及这些编码器受到不同信噪比(SNR)等级的加性和乘性信道噪声干扰时的情况,共计 28 种失真条件,分布均匀. 为评估各种方法的主客观测试结果的吻合性,在 $M_{os}\sim M_{os}$ 平面做二次拟合曲线,相关系数定义为

$$r=\sqrt{\sum_{k=1}^{K}(M_{os}(k)-m_y)\Big/\sum_{k=1}^{K}(M_{os}(k)-m_y)} \tag{8}$$

式中 $M_{os}(k)$ 为第 k 个失真条件对应的 MOS 估值,m_y 为所有 K 种失真条件下的 MOS 估值的平均. 标准偏差为

$$s=\sqrt{\frac{1}{K}\sum_{k=1}^{K}(M_{os}(k)-M_{os}(k))^2} \tag{9}$$

图 2～4 分别给出了一步法和两步法在 28 种失真条件下所得到主客观测试结果的二次拟合曲线,表 1 列出了它们的相关系数及标准偏差.

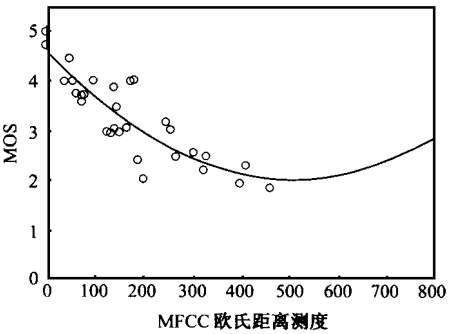

图 2 两步法主客观测试结果拟合曲线

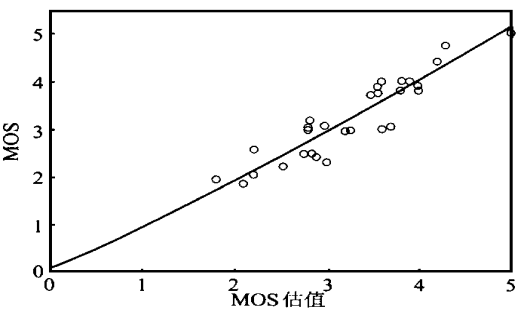

图 3 MLP 系统主客观测试结果拟合曲线

表 1 一步法和两步法的主客观相关系数及标准偏差

	训练集		测试集	
	r	s	r	s
两步法	0.89	0.40	0.87	0.43
MLP 系统	0.97	0.21	0.93	0.28
RBFN 系统	0.98	0.16	0.96	0.19

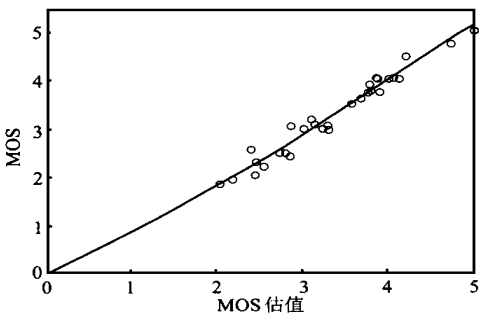

图 4　RBFN 系统主客观测试结果拟合曲线

5　结论

基于一步策略的语音质量客观评价系统实现了从多维特征平方差值矢量空间向一维主观测度的直接映射,避免了对人耳听觉系统感知行为或生理结构进行复杂的数学描述,且计算简便.而且它还将主客观测试值的统计回归分析过程纳入到神经网络当中,也避免了由于模型假定不当所产生的误差,使得评价结果更加可靠.统计学分析表明,对于某一种失真条件而言,一步评价系统的输出是 MOS 的无偏一致估计,其统计特性明显优于两步法.

参考文献:

[1] John R Deller, John G Proakis, John H L. Hansen Discrete-Time Processing of Speech signals [M]. Macmillan Publishing Company, New York 100022, 1993.

[2] Gray R M, Buzo A, JR Gray A H, Matsuyama Y. Distortion Measures for Speech Processing [J]. IEEE Trans. on ASSP, August 1980, ASSP-28(4).

[3] Hermansky H. Perceptual linear predictive (PLP) analysis of speech [J]. J. Acoust. Soc. Am. 1990, 87(4): 1738 - 1752.

[4] Moody J, Darken C J. Fast learning in networks of locally tuned processing units [J]. Neural Computation, 1991, 3: 246 - 257.

[5] A Papoulis. Probability, Random Variables, and Stochastic Processes (Second Edition) [M]. McGraw-Hill, Inc. 1984.

作者简介:

付　强　男. 1972 年生于陕西西安. 分别于 1994 年 7 月、1997 年 4 月获得工学学士、硕士学位. 现为西安电子科技大学综合业务网国家重点实验室博士生,学科信号与信息处理. 目前主要研究领域为语音和通信信号处理等. 已发表相关学术论文 10 余篇.

易克初　男. 1943 年生于湖南连源. 1967 年毕业于华中理工大学无线电技术专业,后来分别在中国科技大学和西安电子科技大学获硕士和博士学位,并在英美做高访各一年,现为西安电子科技大学综合业务网国家重点实验室副主任,博士生导师. 目前感兴趣的领域有语音处理,通信信号处理,卫星通信,互联网应用.

田　斌　男. 1970 年生于山东荷泽. 分别于 1992 年 7 月,1995 年 4 月获得西安电子科技大学计算机系学士、硕士学位. 2000 年 3 月获得西安电子科技大学通信与信息系统博士学位,现任西安电子科技大学综合业务网国家重点实验室副教授. 目前感兴趣的研究领域为语音编码、通信信号处理、高速网络模型、IP 网络服务质量及电子商务,发表学术论文 10 余篇.