

用于客观音质评价的扩展巴克谱失真测度

杨波

(成都市成华区人民法院 成都 610051)

【摘要】扩展巴克谱失真测度应用同时性掩蔽效应,可以充分反映听觉系统的感知特性,在巴克谱失真距离的计算中排除被噪声掩蔽门限掩蔽的临界频带。研究中对噪声掩蔽门限进行了尺度伸缩调整,扩展巴克谱失真测度和巴克谱测度对比实验分析也印证了所提出的测度的良好性能。对汉语通信系统进行的客观音质评价,不仅具有较好的普适性,而且主客观评价结果之间具有更高的相关系数,相关度达到0.969。

关键词 语音系统; 客观音质评价; 同时性掩蔽; 巴克谱距离; 相关系数
中图分类号 TP912.3 文献标识码 A

Extensive Bark Spectral Distortion Measurement for Objective Speech Quality Assessment

YANG Bo

(Chengdu Municipal Chenghua District People's Court, Chengdu 610051)

Abstract The performance of the Extensive Bark Spectral Distortion EBSD measure is examined for speech data with coding distortions by varying the scaling the noise masking threshold and omitting the spreading function in noise masking threshold calculation. Some comparative experiment results of the EBSD and Bark Spectral Distortion(BSD) show that the EBSD has better performance than BSD in Chinese communication channels, and the correlation coefficient reaches 0.969.

Key words speech system; objective speech quality evaluation; simultaneous masking; Bark spectral distance; correlation coefficient

主观语音质量评价符合人耳对音质的实际感觉,但是测量相当麻烦,不够灵活,容易受条件限制和人的主观因素影响等。因此,寻找一个比较准确地反映统计平均意义上的语音质量客观评价方法成了人们十分向往的目标。在众多的客观音质评价的方法中,基于听觉模型的巴克谱失真(Bark Spectral Distortion ,BSD)测度具有较好的客观评价性能^[1],获得了较高的主客观相关性^[2],其合理之处在于应用了人耳响度识别的对数特性,但是这种算法还没有考虑人耳听觉系统更精确的响应机制。近年来随着对人耳听觉机理研究的深入,出现了更全面更精确的听觉模型,即听觉响应系统所具有的同时性掩蔽效应^[3-4]。

本文研究了一种能较好地体现人的主观听觉感知特性,具有高的主、客观相关系数的评价测度,主要是利用噪声信号对纯音信号的同时掩蔽特性,在BSD测度中排除被掩蔽的失真距离,并将这种扩展的巴克谱失真测度(Extended Bark Spectral Distortion ,EBSD)应用于多种特定干扰失真语音信号的客观音质评价中。因此,深入研究噪声掩蔽语音的这种掩蔽听觉模型,及其在客观音质评价中的应用,不仅在军事上具有现实意义,而且在军、民语音通信装备的研制、内外场试验进行音质客观评价,以及在语音增强、语音自动识别之类的特征提取的研究中,都会得到广泛应用^[5]。

1 EBSD测度算法

EBSD测度主要分为3步:(1) 听觉响度差计算;(2) 噪声掩蔽门限(Noise Masking Threshold ,NMT)估计;

收稿日期:2004-02-19

基金项目:总参通信部预研基金资助项目

作者简介:杨波(1967-),男,工程师,主要从事语音编码器质量评价等方面的研究。

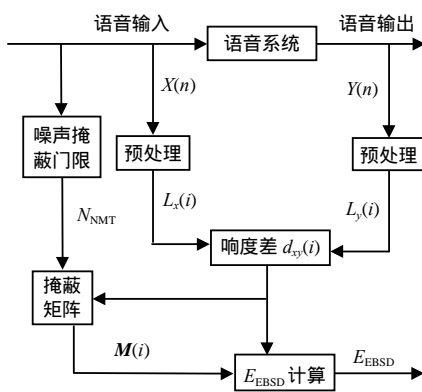


图1 EBSD测度工作框图

(3) EBSD计算, 工作原理框图如图1所示。预处理不同于文献[4], 在信号频谱滤波采样中应用扩展函数计算巴克谱系数失真距离, 它充分地反映了人耳听觉系统的感知特性。图1中 $L_x(i)$ 与 $L_y(i)$ 分别为输入信号和输出信号第 i 帧的响度, $d_{xy}(i)$ 为 $L_x(i)$ 和 $L_y(i)$ 之间的响度差, N_{NMT} 是由输入信号计算得来的噪声掩蔽门限值。

为了确定某一临界带是否可以感知, 定义一个掩蔽矩阵符号 $M(i, j)$, j 表示第 i 帧的第 j 个Bark带。当失真可以被感知, $M(i, j)$ 为1, 反之 $M(i, j)$ 为0, 它也是一个只包含0和1元素的矩阵。而 $M(i, j)$ 的取值由 $d_{xy}(i)$ 与 N_{NMT} 比较得来:

$$\begin{cases} M(i, j) = 1 & d_{xy}(i, j) > N_{NMT}(i, j) \\ M(i, j) = 0 & d_{xy}(i, j) \leq N_{NMT}(i, j) \end{cases} \quad (1)$$

下面把所有听觉能感知的响度差求和得到一帧信号的失真距离, 即计算巴克谱失真时不考虑被掩蔽的距离:

$$E_{EBSD}(i) = \sum_{j=1}^{15} d_{xy}^2(i, j) M(i, j) \quad (2)$$

式中 $E_{EBSD}(i)$ 为第 i 帧的失真距离, 在300 ~ 3400 Hz语音频带取15 Bark。最后整个测试信号总失真距离为:

$$E_{EBSD} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{15} d_{xy}^2(i, j) M(i, j) \quad (3)$$

式中 N 为总帧数。

2 NMT的计算

确定输入语音的噪声掩蔽门限值 N_{NMT} 是计算 E_{EBSD} 测度的关键步骤, 符合人耳听觉掩蔽特性的NMT由信号临界频带分析、对巴克谱应用谱扩展函数和听阈比较3步来完成^[6]。

2.1 临界频带分析

信号作 FFT 后得到功率谱 $P(k)$, 求得每个 Bark 带在一帧信号中的能量分布, 得到临界带能量:

$$B_{ij} = \sum_{k=b_{jl}}^{b_{jh}} P_i(k) \quad (4)$$

式中 b_{jl} 是第 j 个 Bark 带所对应谱线的下限序号; b_{jh} 是对应谱线的上限序号。

2.2 扩展函数的应用

扩展函数用来描述各个临界带信号对其它临界带信号的掩蔽特性。实际上扩展函数表现为滤波器组, 这个滤波器组与Mel谱三角形滤波器组形状十分相似。特征是某一频率信号对于高于它的频率信号的掩蔽效应比较明显, 反之掩蔽效应很弱, 故下向谱扩展可以忽略。对 B_{ij} 用扩展函数进行滤波卷积:

$$C(j) = \sum_{j'=1}^{17} S(j, j') B(j') \quad (5)$$

式中 $S(j, j')$ 为扩展函数, j 表示被掩蔽的临界带指数, j' 表示掩蔽信号的临界带指数, 临界频带间距小于25, 即 $|j' - j| \leq 25$ 。然后将扩展巴克谱 $C(j)$ 在语音频带内应用等响度曲线预加重得到信号的响度级(方), 最后将响度级转换为语音信号的响度(宋)^[11]。

2.3 计算NMT

语音信号的音调特性决定了掩蔽程度的大小。同时性掩蔽效应包括纯音掩蔽噪声和噪声掩蔽纯音两种情况, 因此各部分的贡献大小与信号的音调特性(似噪性)有关。本文采用了谱平坦度量度(Spectral Flatness Measure, SFM)法来判定信号的音调性。根据音调因子 α 的值可以确定一个对 $C(j)$ 的修正值^[6]:

$$O_j = \alpha(14.5 + j) + 5.5(1 - \alpha) \quad (6)$$

最后得到 Bark 带噪声掩蔽门限:

$$N_{NMT_j} = 10^{\lg C(j) - 0.1 O_j} \quad (7)$$

3 实验结果

实验从反映汉语基本统计特性的军用标准 SJ26771-96“军用通信设备音质 MOS 评价法”的语音材料表中选用了两个测试语句表,原始语音去掉无声段由计算机仿真得到各种受扰的失真语音,每句语音经过主观评价得到平均意见得分(MOS 分)。采用的被测语音干扰条件为具有若干个不同性质及干扰强度的 8 Kb/s CS-ACELP(G.729)受扰语音条件。按最小二乘法准则进行二次多项式拟合,求出主客观评价的相关系数 ρ [2]。

N_{NMT} 的推导是假定语音信号都是纯音信号,而实际的语音信号为复音信号,所以 N_{NMT} 并不完全准确地符合失真距离的估计。在这里定义一个尺度因子 β 对 N_{NMT} 进行尺度伸缩,从而来分析 E_{EBSD} 测度性能,式(1)改为:

$$\begin{cases} M(i, j) = 1 & d_{xy}(i, j) > \beta N_{\text{NMT}}(i, j) \\ M(i, j) = 0 & d_{xy}(i, j) \leq \beta N_{\text{NMT}}(i, j) \end{cases} \quad (8)$$

式中 $0 < \beta < 1$ 。表 1 所示是 G.729 失真条件时对不同尺度因子 β 得到的 E_{EBSD} 测度性能与 BSD 测度的比较,可见在 β 等于 0.4 时混合声的主客观相关性最好。

表 1 两种测度的相关系数比较

	E_{EBSD}					BSD
	$\beta=0.2$	0.4	0.6	0.8	1.0	
女	0.964	0.965	0.958	0.948	0.935	0.940
男	0.943	0.947	0.949	0.950	0.946	0.902
混合	0.965	0.969	0.966	0.961	0.952	0.933

4 结束语

利用同时性掩蔽效应对巴克谱失真测度进行改进,即使在 BSD 测度用于 G.729 语音编码器经过不同干扰失真都已达到主客观评价高度相关的情况下,对同样的实验条件 E_{EBSD} 得到的结果还是较大地提高了主客观评价的相关系数,这表明 EBSD 测度具有一定的应用价值。

另外也注意到,目前的掩蔽模型主要是由欧美等少数发达国家所确定,而这些模型是根据模型制定国的人群所做的实验统计得到,由于各国人群的生理参数有一定的差别,因此现有模型未必很适合中国使用,有必要开展适合中国人的掩蔽模型研究及其应用。

本文研究工作得到中国电子科技集团公司第三十研究所张知易研究员给予的指导,在此表示感谢!

参 考 文 献

- [1] Wang S, Sekey A, Gersho A. An objective measure for predicting subjective quality of speech coders[J]. IEEE Journal on Selected Areas in Communication, 1992, 10(5): 819-829.
- [2] 黄惠明, 王 瑛, 赵思伟, 等. 语音系统客观音质评价研究[J]. 电子学报, 2000, 28(4): 112-114.
- [3] 杨 波, 傅汝林, 张知易. 利用前向掩蔽效应改进巴克谱失真测度[J]. 通信技术, 2003, 5: 5-6.
- [4] 杨 波, 傅汝林, 张知易. 一种改进的客观音质评价方法[J]. 成都理工大学学报, 2004, 1: 108-110.
- [5] Yang W, Dixon M, Yantorno R. A modified bark spectral distortion measure which uses noise masking threshold[C]// IEEE Speech Coding Workshop, Pocono Manor USA, 1997:55-56.
- [6] Johnston J D. Transform coding of audio signals using perceptual noise criteria[J]. IEEE Journal on Selected Areas in Communication, 1988, 6(2): 314-323.

编 辑 孙晓丹