一种新的用于语音主观质量评价的谱失真参数 [1]

杨 震 毕厚杰

(南京邮电学院信息工程系 南京 210003)

摘 要 该文分析和讨论了各种语音主观质量评价的客观方法，提出了一种考虑了人耳的屏蔽效应，且正比于人耳听觉的 Bark 域谱失真参数 PBSD(Perception-based Bark Spectral Distortion)，用来映射语音的主观 MOS 分值. 实验表明，基于该参数及其主客观映射关系，所得到的各类语音编译码系统的主观 MOS 预测分，平均和最大预测偏差均较小. 论文最后利用 PBSD 参数代替 MSE 参数，设计的语音编码系统，改善了解码语音的主观听觉质量.

关键词 语音信号，MOS 分，预测

中图号 TN912.3

1 引 言

语音信号是目前千家万户进行通信所主要使用的手段. 各种场合各种网络环境所开发和应用的语音编译码标准和通信系统很多，因此评价语音数字化编译码系统和通信系统的质量，一直是备受人们关注的重要课题.

由于在语音数字化压缩及通信领域，语音最终的接受者多数是人，所以目前广泛使用且公认的质量评价标准是主观的评价标准. 语音质量主观评价标准也有很多 [1]，其中最常用的是平均意见分值 (MOS, Mean Opinion Score). MOS 反映的是人对语音听觉质量的整体感觉，一般分 5 个等级，以 5，4，3，2，1 分评定.

MOS 尽管很常用，但测量相当麻烦，因此，寻找一个客观可计算的语音质量评价标准，而又比较准确反映了统计平均意义上，语音质量主观评价的结果，一直是人们十分向往的. 这样的一种标准，具有下列优点：(1) 测试简单，快捷，不需花费大量人力物力. (2) 具有完全可重复性，且不受不同环境和不同对象的影响. (3) 为设计开发各种新的语音编译码和通信系统，提供了质量测试手段，有助于设计者进行比较和选择.

2 语音质量的主观质量客观评价方法

要进行语音质量的主观质量客观评价，原理由图 1 所示：

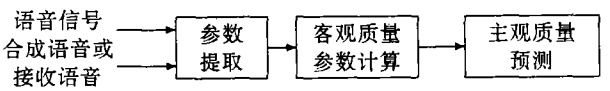

图 1 语音质量的主观质量客观评价系统原理

许多学者已对这个问题进行了研究 [2−5]. 进行客观质量参数分析时，可用参数很多，比如文献 [4] 共测试了 32 种，文献 [5] 试验了 9 种等；总的可分成四大类：(1) 信噪比类. 指语音信号时域波形的匹配，常采用总信噪比和分段平均信噪比. (2) 基于线性预测编码 (LPC) 的测度. 主要是 LPC 分析所得的各种参数，比如线谱对参数、对数面积比等. (3) 谱距离测度. 衡量

[1] 1998-09-25 收到，2000-09-16 定稿
邮电部预研基金支助 (98 部预 5)

语音信号频谱的匹配，比如 I-S(Itakura-Saito) 频谱距离， bark 谱距离 (BSD) 等。 (4) 其它。如相关函数，过零率等。

根据客观质量参数分析结果，预测主观语音质量常用的方法有： (1) 曲线拟合。从某种参数的失真分布到 MOS 的映射，多以回归曲线逼近为主，用得较多，如文献 [1,3,5] 。 (2) 用模糊逻辑法。如文献 [2] 寻找了语音电平、串音及多径效应等与语音质量的模糊隶属函数。 (3) 用神经网络归类。文献 [4] 用 Abductive 网络，也有采用二层 BP 网络的。 (4) 用 VQ(Vector Quantization) 分类。

本文提出一种新的用于语音主观质量评价的谱失真参数 (PBSD, Perception-based Bark Spectral Distortion) ，并基于它来映射语音的主观 MOS 和改善语音编码系统的质量。

3 语音质量评价客观参数的计算

设已有被评价的作过能量归一化预处理的两种语音信号，一是原始未处理的语音 $s(n)$ ，二是处理过的语音 $\hat{s}(n)$ 。 $\hat{s}(n)$ 可以是 $s(n)$ 经编译码系统处理后的恢复语音，也可以是 $s(n)$ 经一通信系统传输后收方接收到的语音。我们试图寻找一种能更好反映人耳的听觉感受特性的特征参数。

由于语音质量的根本评价依据是人的主观听觉，所以，基于人耳感知特性的参数，应是主客观质量映射的最佳分析参数。不过，这一方面必须依赖声学和医学的研究成果，另一方面，人的感知听觉特性的分析相当复杂，因而在话带语音编码领域鲜有人使用。

3.1 Bark 谱及人耳听觉特性

人耳对外界声音信号的听觉感受，主要取决于声音信号的音高 (声音的频率) 、响度 (声音的强弱) 和掩蔽效应 (耳朵对一个声音的听觉感受，受到另一个声音影响的现象) 等因素。人耳对声音的听觉感受特性以 '临界频带'(critical band) 来描述，比普通的赫兹为单位的频率刻度要好，因为在一个临界频带内，很多心理声学特性是一样的，比如掩蔽效应。在人耳听觉范围内临界频带的划分如表 1 所示。

表 1 临界频带的划分 (单位： Hz)[6]

频带数	下界频率	上界频率	频带数	下界频率	上界频率
1	0	100	13	1720	2000
2	100	200	14	2000	2320
3	200	300	15	2320	2700
4	300	400	16	2700	3150
5	400	510	17	3150	3700
6	510	630	18	3700	4400
7	630	770	19	4400	5300
8	770	920	20	5300	6400
9	920	1080	21	6400	7700
10	1080	1270	22	7700	9500
11	1270	1480	23	9500	12000
12	1480	1720	24	12000	15500

临界频带这个参数提出的意义是可将人耳当作一个并联的滤波器组 [7] ，各个滤波器有不同的带宽，分别对听觉作出不同的贡献，因而在研究失真语音的主观听觉质量时，可以将失真在各个临界频带上的分布求出，分别考虑各带的掩蔽效应、听觉响度与频率关系，从而研究各带失真对听觉的影响。

临界频带的单位一般用 bark 来表示， 1 bark 用来指明一个临界频带的频率宽度 [8] ，若记 bark 域的频率变量为 b, 赫兹 (Hertz) 域频率变量为 f ，则有

$$f = 600 \sin h(b/6) \tag{1}$$

各临界频带间信号仍有一定相互影响，这可以用一个扩展函数 (expanding function)$e(b)$ 来计算 [3]，各个临界频带的扩展函数在 bark 域是相同的：

$$e(b) = 10^{0.7-0.75(b-0.215)-1.75}\sqrt{0.196+(b-0.215)^2} \tag{2}$$

知道了各临界频带内信号能量分布，再考虑了各带的屏蔽效应 (许多文献如文献 [3]，未考虑这一效应)，还不意味着它们对听觉感知的贡献大小已知，因为人耳对不同频率区域的信号有不同的灵敏度，图 2 中 Fletcher-Munson 曲线 [8] 反映了人耳的这一特性。

图中最下面一根曲线是人耳的听觉阈值，声强低于它，人耳听不见，而信号频率不同，阈值相差也很大，如 30Hz 处声音信号得比 4kHz 处声音信号强约一百万倍，二者才有相同的响度级。人对声音真正感受的响度以宋 (sone) 为单位，与响度级 phon 的关系为 [6]

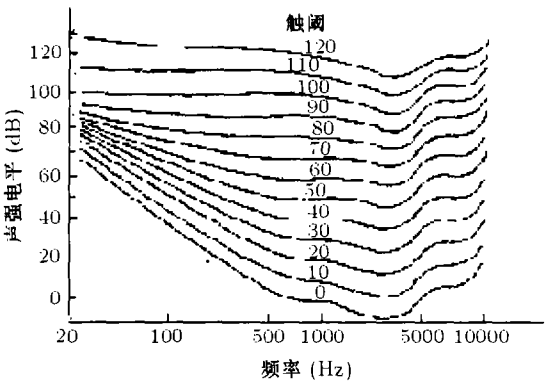

图 2　人耳听觉等响度级 (Fletcher-Munson) 曲线

$$S = 0.063 \times 10^{0.03p} \tag{3}$$

3.2 新的正比于人耳听觉的 bark 域谱失真 PBSD 参数的计算

人耳对声音的相位变化不敏感，所以在评价语音质量时，主要以语音的功率谱为分析对象，探讨各部分谱失真引起的主观感觉。

观察表 1，在 0~4kHz 话带内，有 17 个临界频带，我们需要考察这 17 个 bark 谱频带内的失真对听觉的影响。用于进行主客观质量映射的新的 PBSD 参数的计算分成以下几步：

(1) 将经过能量归一化预处理的 $s_i(n)$, $\hat{s}_i(n)$ 作 FFT 分析得 $S_i(k)$, $\hat{S}_i(k)$；i 表示第 i 帧语音信号。然后求相应的归一化功率谱

$$P_i(k) = |S_i(k)|^2, \qquad \hat{P}_i(k) = |\hat{S}_i(k)|^2 \tag{4}$$

(2) 每个 bark 谱带抽一次样，失真影响统一考虑，求各临界频带内 bark 谱功率：

$$B_{ij} = \sum_{k=b_{jl}}^{b_{jh}} P_i(k), \quad \hat{B}_{ij} = \sum_{k=b_{jl}}^{b_{jh}} \hat{P}_i(k) \tag{5}$$

j 表示第 j 个 bark 临界频率，b_{jl} 是它对应的 f 域下限频率，b_{jh} 是上限频率，它们可以根据表 1 和抽样频率 f_s 求出

$$b_{jl} = N \times \frac{B_{jl}}{f_s}, \quad b_{jh} = N \times \frac{B_{jh}}{f_s} \tag{6}$$

N 是 FFT 分析点数，B_{jl} 是表中第 j 个临界频带的 f 域下界频率，B_{jh} 是上界频率。

(3) 临界频带间相互影响的计算，是 $P(b)$(bark 域的功率谱密度，可根据 (1) 式转换得到) 与谱扩展函数 $e(b)$ 的直接卷积 [3]。由于这两者均是非线性函数，该卷积的计算颇为麻烦，文献 [3] 将这种运算折返到赫兹域，等效成临界频带的加权函数运算，每个加权函数形似衰减指数函

数，这可以预先计算好，所以避免了频谱扩展函数难以计算的问题，此处也借鉴这样的方法，修正后 (4) 式的各临界频带内 bark 谱功率记为： B'_{ij}, $\hat{B}'_{ij}$ 。

(4) 考虑各带的失真

$$E_{ij} = |B'_{ij} - \hat{B}'_{ij}| \tag{7}$$

是否能被掩蔽。这又分成两种可能： (1)E_{ij} 在听觉阈值下； (2)E_{ij} 能被 B'_{ij} 掩蔽。

听觉阈值与掩蔽，是在声强级电平 (图 2 纵坐标) 单位下讨论问题的，因此首先必须将 E_{ij} 和 B'_{ij} 转换成声强。我们设计了如下的转换方法：

根据文献 [7], 声强是在垂直于声传播方向上通过单位面积能量的速率，或者说是单位面积上输送的平均功率。因此假设声音是通过一个阻抗 R 的喇叭发出的，喇叭放在室内四周无障碍物处，不考虑声波反射，人离它 s 米 (实验中就取 1 米)，声音沿半球面各方向均匀纵向传播，则人所在的位置处半球面上的声强 I 乘半球面面积应等于声音的功率 P ， I 正是我们所想求的变量。忽略能量损耗，这个功率 P 应等于喇叭发出的功率，再设喇叭的转换效率为 η ，那么输送给喇叭的功率是 P/η ，等于

$$P/\eta = I \times 2\pi \times S^2/\eta \tag{8}$$

另一方面，设被分析的语音信号是电压信号，那么前面求得的 bark 域功率 E_{ij} 和 B'_{ij} 其实是单位阻抗上，在持续一帧时间内的等效的总能量。根据 Parseval 定理，时域能量等于频域能量，由于将每个临界频带抽样等效成一个音调，那么总能量为 E_{ij} 和 B'_{ij} 的音调对应的正弦波电压幅度 A 为

$$A = 2\sqrt{E_{ij}/m} \quad 或 \quad A = 2\sqrt{B_{ij}/m} \tag{9}$$

m 是每帧中样点数；因此传送给喇叭的功率等于

$$p = E_{ij}/(mR) \quad 或 \quad p = B_{ij}/(mR) \tag{10}$$

将 (8) 和 (10) 式结合，于是可求得声强为

$$I_E = \eta E_{ij}/(2\pi mRS^2), \quad I_B = \eta B'_{ij}/(2\pi mRS^2) \tag{11}$$

式中 I 的单位是 $\mathrm{W/m^2}$ ，对应的以分贝为单位的声强级为

$$\mathrm{IL} = 10\log_{10}(I/I_{\mathrm{ref}}) \tag{12}$$

I_{ref} 是参考声强，等于 $10^{-12}\mathrm{W/m^2}$ 。

根据 (11) 式 I_E 对应的声强级，与图 2 曲线中听觉阈值对比，判断第一种掩蔽是否成立。

掩蔽效应的第二步是考虑 I_E 能否被 I_B 掩盖。这需要计算 I_B 与 I_E 之差的绝对值，再与掩蔽门限比较即可判断；不过，掩蔽有两种，音调掩蔽噪声和噪声掩蔽音调，二者掩蔽门限不同 [6] 。若某临界频带 i 内掩蔽音是音调，其强度为 x_i(dB)，则强度在 $x_i - (14.5 + i)$(dB) 以下的噪声是听不见的；反之，若某临界频带 i 内掩蔽音是类似噪声的信号，强度为 y_i, 则强度在约 $y_i - 5.5$(dB) 以下的失真音是听不见的。

如何才能判断一个临界频带内信号特性是什么呢？下面提出一个判别方法。

用 (5),(6) 式计算 B_{ij} 时，仔细观察会发现，以下新定义的变量：

$$P_i'(k) = P_i(k)/B_{ij}, \quad b_{jl} \le k \le b_{jh} \tag{13}$$

即第 j 个临界频带内的归一化功率谱，含有概率密度函数的特征：每个分量大于等于 0 而小于等于 1，且所有 $P_i'(k)$ 相加等于 1。而信息论中计算信息熵时有一个重要特性：当全部随机变量有相等的概率时给出最大的熵，而某一个变量概率等于 1 时，全部随机变量的熵最小。所以熵值可以作为随机变量概率是否均匀分布 (平坦化) 的一个度量，于是定义：

$$H_{ij} \equiv -\sum_{k=b_{jl}}^{b_{jh}} P_i'(k) \log P_i'(k) \tag{14}$$

为第 i 帧语音的第 j 个临界频带内的信号分布熵，根据它的值与一个门限值比较来判定该带内信号属性 (门限值由实验定)，以决定掩蔽效应的属性。

(5) 将不能被屏蔽的临界频带内失真功率，转换成与人听觉响度成正比的宋 (sone) 变量。(3) 式可以用来计算宋，但必须首先求出 phon，即响度级。一个特定频率和强度的音调，其对应的响度级 phon, 定义成与 1kHz 处音调有同样听觉响度的以分贝为单位的强度：

$$P(\text{phon}) = IL + 30\log_{10} F(f) \tag{15}$$

$F(f)$ 是一个经验公式，可以根据图 2 等响度曲线确定 [3]。

有了各带的响度级和不能被屏蔽的误差响度级，就可用 (3) 式计算它们对应的人耳听觉响度宋，设记为 S_{ij} 和 SE_{ij}。

(6) 计算不能被掩蔽的相对平均谱失真响度 PBSD。

$$\text{PBSD} = \frac{1}{M}\sum_i \sum_{j\in[1,N]} \frac{SE_{ij}}{S_{ij}} \tag{16}$$

N 是临界频带数，$j \in [1,N]$ 表示只计算那些不能被掩蔽的临界频带的谱失真，M 是语音信号的分帧数。

4 主客观参数的映射

研究主客观评价值之间的映射，必须要有一些已知主观评价质量的系统，以便与客观评价质量进行对比。我们选用了：(1) G.711 PCM 系统；(2) G.721ADPCM 系统；(3)G.726 ADPCM 系统；(4)G.723.1MP-MLQ/ACELP 系统；(5) 基本 CELP 系统。

根据资料显示，这几种系统的解码语音质量的平均意见分 (MOS, Mean Opinion Score) 大致为：(1)4.0~4.5(码率 56~64 kbit/s); (2) 4.3(码率 32kbit/s); (3) 2.8~4.0(码率 16~24kbit/s); (4)3.5~3.0(码率 5.3~6.3kbit/s); (5)2.8~3.2(码率 5.6~6.1kbit/s)。

试验是针对 7 个说话人的声音进行的，说话人由 4 男 2 女 1 个儿童构成。而声音的主观质量评价由 10 位人员组成，进行主观质量打分，后面所列 MOS 均是这 10 人的 MOS。实验测得的目前广泛使用的两种失真距离参数：平均分段信噪比 SNRseg-avg 和平均倒谱距离 (CD, Cepstrum Distance)，以及相应系统的主观 MOS 如表 2 所示：

从表 2 中可见，信噪比 SNR 不一定与主观听觉质量成正比，如 ADPCM 系统，在码率为 16kbit/s 时 (第 3, 6, 9)，其量化信噪比与 G.723.1 系统 (第 15, 17, 19) 也不相上下，不过，前者码率是后者的近 3 倍，性能却明显差于后者。以上说明 SNR 与 MOS 分之间，并无一一对应

关系，SNR 因而不是一个好的质量评价标准。而 CD 值与 MOS 关系要明显一些，一般 CD 值小，MOS 高；但基本 CELP 系统与 G.723.1 系统有类似的 CD 值，而后者的 MOS 却明显高于前者，说明 CD 值亦不是一个很理想的、可以用来准确估计主观质量的客观质量评价参数。

针对表 2 中各系统的 CD 值与 MOS 的对应关系，作图后根据形状用抛物曲线拟合法，可得到主客观质量映射关系：

$$\mathrm{MOS} = a + b \times \mathrm{CD} + c \times \mathrm{CD}^2 \tag{17}$$

其中 a 、 b 、 c 是待求参数。采用最小二乘法进行曲线拟合，求解得

$$\mathrm{MOS} = 4.6017 - 2.0208 \times \mathrm{CD} + 0.42725 \times \mathrm{CD}^2 \tag{18}$$

表 2 各种系统的分段平均信噪比、平均倒谱距离性能与相应的 MOS

实验序号	码率 (kb/s)	SNR 值 (dB)	CD 值	MOS	实验序号	码率 (kb/s)	SNR 值 (dB)	CD 值	MOS
1	32	27.1	0.16389	4.2	12	64	35.08	0.0479	4.5
2	24	21.97	0.31949	4.0	13	6.3	7.15	0.58065	3.8
3	16	12.08	0.88326	2.8	14	5.3	6.39	0.62173	3.6
4	32	28.19	0.27812	4.3	15	6.3	10.62	0.51988	3.9
5	24	21.72	0.60166	3.9	16	5.3	9.19	0.55526	3.7
6	16	9.89	1.42158	2.7	17	6.3	11.47	0.57072	3.8
7	32	30.18	0.23477	4.3	18	5.3	9.91	0.62357	3.6
8	24	23.28	0.56061	3.9	19	6.3	11.36	0.55679	3.8
9	16	11.08	1.42559	2.7	20	5.3	9.69	0.63225	3.6
10	32	29.16	0.26395	4.3	21	5.6	10.08	0.57818	3.1
11	56	32.46	0.0782	4.4	22	6.1	11.34	0.55296	3.2

* 说明：实验 1-3 、 13-14 和 21-22 中，语音样本是一句长 14.61s 的声音 (含 3 男 1 女声音)，实验 4-6 和 15-16 中，样本是一句长 21.4s 的女声，实验 7-9 和 17-18 中样本是一句长 22.5s 的男声，实验 10-12 和 19-20 中样本是一句长 25.3s 的童声. 实验 1-10 为 ADPCM 系统，11 和 12 为 PCM 系统，13-20 为 G.723.1 系统，21 和 22 为 CELP 系统，分析帧长 32ms，50% 重叠 。

4.1 CD 值与 MOS 关系的映射

根据这条曲线，对各点的 MOS 估计偏差进行了统计，结果为，平均偏差为 0.185，最大偏差为 0.476 。另外，还针对在曲线拟合中没有使用的新的编码系统，进行了测试，结果为，平均偏差为 0.191, 最大偏差为 0.446；编码系统是 G.727 嵌入式 ADPCM 系统和多类激励线性预测 MELP 系统，两者均是可变速率编码系统，实验中试验了多种码率。

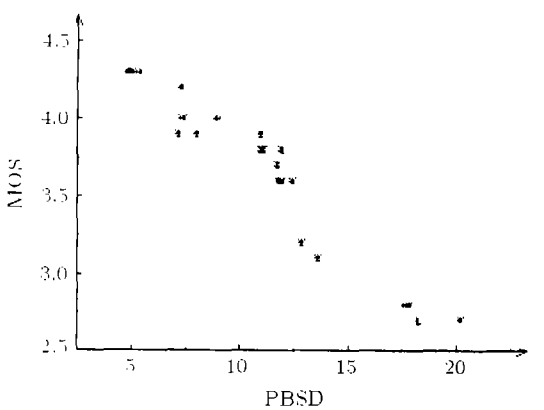

图 3 PBSD 与 MOS 的关系图

4.2 PBSD 与 MOS 的映射

对同样的语音，采用 3.2 节方法计算的 PBSD 参数与 MOS 关系如表 3 和图 3 所示 (分析帧长 32ms 以便作 FFT) 。

表 3 中的 PBSD 值大，MOS 低，未遇到反例，说明 PBSD 值确实较其它参数更能反映语音主观质量的优劣。根据图 3 形状，采用抛物线进行曲线拟合，得到拟合方程为

表 3　各种系统的 PBSD 谱失真值与相应的 MOS 实验

实验序号	PBSD 值	实验序号	PBSD 值	实验序号	PBSD 值
1	7.3155	9	18.149	17	11.0731
2	8.9571	10	4.7735	18	11.8303
3	17.844	11	7.3502	19	11.0002
4	5.3264	12	17.6584	20	11.8932
5	8.0153	13	11.9238	21	13.6069
6	20.190	14	12.4265	22	12.8413
7	4.9983	15	10.9684		
8	7.1896	16	11.7099		

$$\mathrm{MOS} = 4.7015 - 0.07112 \times \mathrm{PBSD} - 1.846 \times 10^{-3} \times \mathrm{PBSD}^2 \tag{19}$$

计算得各系统估计的 MOS 与实测的 MOS 的平均估计偏差为 0.108, 最大估计偏差为 0.283 , 可见明显好于基于 CD 值的估计。同样，也对拟合数据集以外的编码系统进行了测试，平均和最大估计偏差分别为 0.098 和 0.257, 再次证明 (19) 式的估计效果较好。

5 新的基于 PBSD 谱质量评估的语音编码系统

新的语音主观质量的客观计算法，不仅可以评价各类语音系统的质量，还可以取代常规的 MSE 失真准则，用来设计改进现有和将来的语音编码和处理系统，因而具有相当重要的意义。此处以 ADPCM 系统来进行试验。

ADPCM 系统是波形编码，量化编码直接根据 MSE 准则。另外，ADPCM 编码系统中是包含了解码系统的，基于这一点，本文设计了一个新的 ADPCM 系统，引入基于 PBSD 参数的量化编码机制，取代对原系统中误差信号的 MSE 量化编码法，系统框图如图 4 所示。

PBSD 参数的求解，是基于一段一段语音的，所以图中缓存子系统将存储一段长 Nms (实验中取 20ms) 的语音，作为计算 PBSD 参数的输入，输入信号并被开汉明窗。图 4 系统与 G.726 ADPCM 国际标准不同之处，就在于量化编码准则，新系统不是基于每点预测误差波形的量化噪声最小化来进行量化编码的，而是基于一段输入语音与解码语音的 bark 域感觉误差最小化来进行量化级的选取。对表 2 中一句 3 男 1 女声音 (长 14.61s) , 进行编译码处理，根据 (19) 式计算 MOS , 实验结果如表 4 .

可见，本文新方法对语音编码主观质量有明显改进。

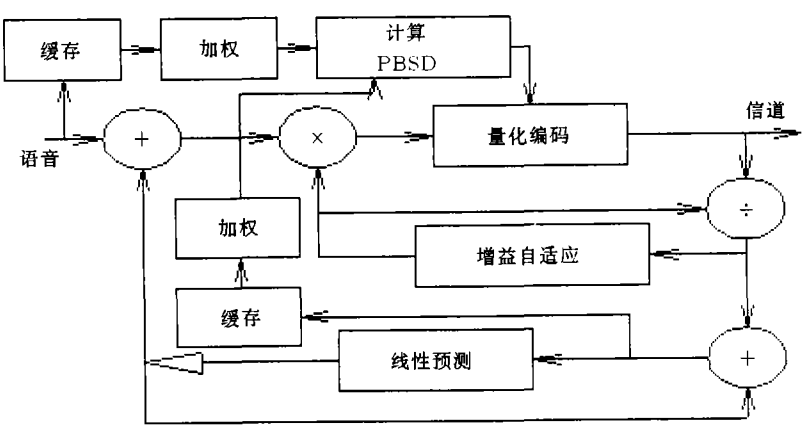

图 4　基于 bark 域感觉误差的 ADPCM 编码系统

表 4 采用不同失真准则的编码系统的 MOS

比特 / 每样值	4	3	2
基于 G.726 ADPCM 的 MOS	4.3154	3.9267	2.8466
基于 PBSD ADPCM 的 MOS	4.5087	4.1495	3.1328

6 结 论

本文提出了一种正比于人耳听觉的 bark 域谱失真参数 PBSD, 用来映射语音的主观 MOS。实验表明，基于该参数所得到的各类语音编译码系统的主观质量预测，平均估计偏差很小，估计精度优于常规的基于分段信噪比和倒谱失真参数所估计的语音主观质量精度。可望根据该参数，修正各类语音编译码系统的设计准则，获得更好的解码语音质量。

参 考 文 献

[1] S. R. Quackenbush, T. P. Barnwell, M. A. Clements, Objective Measures of Speech Quality, New York, U.S.A., Prentice Hall, 1988, 第 2 章.

[2] 丁瑾，钟涛，胡健栋，语音质量的一种新的评价方法，电子学报， 1997, 25(4), 6-9.

[3] Shihua Wang, A. Sekey, A. Gersho, An objective measure for predicting subjective quality of speech coders, IEEE Journal on Selected Areas in Communications, 1992, 10(5), 829-829.

[4] M. M. Meky, Tarek, N, Saadawi, A perceptually-based objective measure for speech coders using abductive network, ICASSP'96, Atlanta, U.S.A., 1996, 479-482.

[5] Nobuhiko Kitawaki, *et al*, Artificial voice signal for objective quality evaluation of speech coding system, ICC'89, Boston, MA, U.S.A., 1989, 373-379.

[6] J. D. Johnston, Transform coding of audio signals using perceptual noise, IEEE on Selected Areas in Communications, 1988, 6(2), 314-323.

[7] L. E. Kinsler *et al*, Fundamental of Acoustics, New York, U.S.A., John Wiley & Sons Inc., 1982, third edition, 246-278.

[8] T. W. Parsons 著，文成义等译，语音处理，西安电子科技大学情报资料室， 1989,3, 49-114.

A NEW PARAMETER OF SPECTRAL DISTORTION FOR PREDICTING SUBJECTIVE QUALITY OF SPEECH

Yang Zhen Bi Houjie

(*Dept. of Info. Eng., Nanjing Inst. of Posts and Telecom., Nanjing* 210003, China)

Abstract This paper analyses various objective measures for the prediction of subjective quality of speech. A new Perception-based Bark Spectral Distortion(PBSD) parameter is presented, which takes the masking property of human ear into consideration, to predict Mean Opinion Score(MOS) of speech quality. Experiments prove that this map from objective measure to subjective MOS based on the calculation of PBSD has rather small prediction error. The PBSD parameter is applied to designing new speech codec in place of MSE parameter and the subjective quality of decoded speeches is improved.

Key words Speech signal, Mean Opinion Score, Prediction

杨 震： 男，1961 年生，博士，副教授，主要研究方向为信号处理，编码， ATM 和 IP 。已发表论文 30 多篇.

毕厚杰： 男， 1932 年生，教授，博士生导师，主要研究方向图像通信，图像处理， IP 网络，已发表论著 7 部，论文 70 多篇.