

# 语音质量客观评价方法研究进展

陈 国, 胡修林, 张蕴玉, 朱耀庭  
(华中科技大学电子与信息工程系, 武汉 430074)

摘 要: 本文回顾了语音质量客观评价方法的研究发展情况, 综述了当前主要语音质量客观评价方法, 介绍了这些评价方法的分类及性能, 并探讨了今后语音质量客观评价方法的发展方向。

关键词: 语音质量; 客观评价

中图分类号: TN91213 文献标识码: A 文章编号: 0372-2112 (2001) 04-0548-05

## Research Advance on Objective Measures of Speech Quality

CHEN Guo, HU Xiulin, ZHANG Yunyu, ZHU Yaoqing

(Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: This paper reviewed the research progress of objective measures of speech quality. First, a survey of present major objective measures of speech quality was described. Secondly, the classification and performance of these objective measures were introduced. At the end, some viewpoints for development direction of objective measures of speech quality were discussed.

Key words: speech quality; objective measures

### 1 引言

语音作为信息传递的重要载体, 与其相关构成的通信、编码、存储和处理等语音系统已成为现代社会信息交流的必要手段, 且已广泛应用于社会各个领域。这些系统的性能好坏成为信息交流是否畅通的重要因素, 而评价这些系统性能优劣的根本标志是在于系统输出语音质量的好坏。因此, 研制灵活、方便、可靠的语音质量评价系统自然成为国内外研究者共同努力的目标。

语音质量包括两方面内容: 清晰度和自然度。前者是衡量语音中字、单词和句的清晰程度, 而后者则是对讲话人的辨识水平。语音质量评价不但与语音学、语言学、信号处理等学科有关, 而且还与心理学、生理学等学科有着密切的联系, 因此语音质量评价是一个极其复杂的问题。语音质量评价从评价主体上讲可分为两大类: 主观评价和客观评价。

主观评价是以人为主体来评价语音的质量。该方式虽较为繁杂, 但由于人是语音的最终接受者, 因此这种评价应是语音质量的真实反映。目前, 国内外使用较多的主观评价方法有: 平均意见分 MOS<sup>[1, 2]</sup>、音韵字可懂度测量 DRT<sup>[3]</sup> 和满意度测量 DAM<sup>[4]</sup> 等。其中, MOS 评分法是一种广为使用的主观评价方法, 它以平均意见分来衡量语音质量, 用五个等级来表示语音的质量等级: 优(5分)、良(4分)、一般(3分)、差(2分)、坏(1分)。显然, 主观评价的优点是符合人对语音质量的感受, 缺点是费时费力费钱, 且灵活性不够, 重复性和稳定性较

差, 受人的主观影响较大等。

为了克服主观评价缺点, 人们不得不寻求一种能够以方便、快捷方式给出语音质量评价值的客观评价方法, 即用机器来自动判别语音质量。不过值得注意的是, 研究语音质量客观评价的目的不是要用客观评价来完全替代主观评价, 而是使客观评价成为一种既方便快捷又能够准确预测出主观评价值的语音质量评价手段。尽管客观评价具有省时省力等优点, 但它还不能反映人对语音质量的全部感受, 而且当前的客观评价方法都是以语音信号的时域、频域及变换域等的特征参量作为评价依据, 没有涉及到语义、语法、语调等这些影响语音质量主观评价的重要因素。

### 2 语音质量客观评价方法的发展回顾

语音质量客观评价从评价结构上可分为基于输入/输出的评价和基于输出的评价, 见图 1。基于输入/输出的评价是以语音系统的输入信号和输出信号之间的误差大小来判别语音质量的好坏, 是一种误差度量; 基于输出的评价是仅根据语音系统的输出信号来进行质量评价。近三十年来, 语音质量客观评价方法研究主要集中在基于输入/输出方式的评价上, 即通过提取两端语音信号的特征参量来建立评价模型。随着信息、通信技术的飞速发展, 这种基于输入/输出的评价方法已满足不了许多领域的实际应用需要, 如在无线移动通信、航空航天以及现代军事等领域, 往往要求客观评价方法具有较高的灵活性、实时性和通用性, 而且在得不到原始输入语音信号

情况下也要能对语音质量进行评价,所以,在九十年代基于输出方式的客观评价方法已开始受到国内外学者的重视。

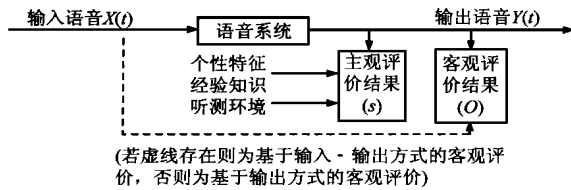


图1 语音质量客观评价的两种结构方式

## 2.1.1 基于输入/输出评价方法的研究发展

最早的客观评价方法可以追溯到四十年代末,由 N. R. French 和 J. C. Steenberg 于 1947 年提出的清晰度指数 AI 方法<sup>[5]</sup>。由于受当时科学技术发展水平限制,语音质量客观评价方法研究并没有得到很大的发展,直到进入七十年代以后,随着通信技术的迅速发展,由于要对语音编码技术、通信设备系统进行选择、性能测试以及标准化等工作,就必须对这些语音系统输出的语音质量进行大规模评测,这样具有方便、快捷、省时、省力等特点的客观评价方法就得到了蓬勃的发展。

在这期间,亚特兰大的乔治亚工学院的 T. P. Barnwell III 和 S. R. Quackenbush 在这一领域所作的研究工作令人瞩目<sup>[6~11]</sup>。T. P. Barnwell III 和 S. R. Quackenbush 对八十年代中期以前的语音质量客观评价研究工作作了系统总结,于 1988 年出版了一本关于语音质量客观评价的专著<sup>[12]</sup>。S. R. Quackenbush 和 T. P. Barnwell 首先建立了语义、语法和音素平衡的未失真语音库,其中所有的未失真语音由四人发音(三男一女),每人 12 个句子,共 48 个未失真句。对这些未失真语音进行 264 种失真处理,得到 11880 个样本的失真语音库。所有的未失真语音和失真语音样本都进行了主观评价,得到主观评价分,再以未失真语音作为参考,对失真语音进行了 600 多种客观评价方法测试,并建立了线性和非线性回归分析,得出与主观评价相关性较高的几种客观评价方法为:谱距离相关度达到 0.180;斜率加权谱距离相关度达到 0.174;对数面积比相关度为 0.162;分段频变信噪比方法相关度达到 0.193(仅对波形编码的失真语音)等<sup>[12]</sup>。他们的工作为以后语音质量客观评价方法的研究奠定了基础。

在八十年代后期提出的语音质量客观评价方法主要有:加拿大 Bell Northern Research 提出的一致函数方法 CHF<sup>[13,14]</sup>, CHF 是一种加权信号失真比测度方法,它通过描述人的听力敏感度、人对噪声的阈值效果以及电话听筒接收的敏感度等来对电话语音质量进行评价,CHF 方法与主观评价的相关度达到 0.1816;日本 NIT 研究人员 N. Kitawaki 等人提出的倒谱距离 CD 方法<sup>[15]</sup>,在当时这是一种与主观评价相关性较好的方法,CD 方法是根据输入和输出语音信号的  $i$  阶 LPC 系数推出各自的倒谱系数  $C_x(i)$  和  $C_y(i)$ ,然后求出它们之间的倒谱距离,其相关度达到为 0.1902;法国 J. Lalou 于 1990 年提出了信息指数  $\Pi$  (Information Index) 方法<sup>[16]</sup>, $\Pi$  方法将语音频谱分为 16 个临界带宽,在每个频带内运用频率加权和听力阈值等计算方法来评价语音质量,其与主观评价的相关度为 0.1688;美国电信科学研究院 ITS 学者 R. Kubichek, E. A. Quincy

等人提出的专家模式识别 EPR 方法<sup>[17]</sup>,EPR 方法运用贝叶斯估计原理以寻找语音信号的特征参数与语音质量之间非线性关系,在 ITS 测试中其相关度达到 0.1886。CHF, CD,  $\Pi$  和 EPR 这四种方法也是原 CCITT 建议方法<sup>[18]</sup>。

进入九十年代后,语音质量客观评价方法研究取得了飞跃的发展,特别是在近几年取得的成果最大,与主观评价的相关度达到了 0.195 左右,同时,基于输出的客观评价方法也得到了有一定发展。

S. Wang 于 1992 年提出的巴克谱失真 BSD 方法对以后的影响较大<sup>[19]</sup>,BSD 方法是以人的听觉能力和听觉心理特点为基础构造出一种听觉转换模型。BSD 在三个层次上模拟了人对语音信号的感知机理:(1)临界频带分析处理;(2)等响曲线预加重;(3)语音强度 $\rightarrow$ 听觉响度变换等。听觉转换模型见图 2。

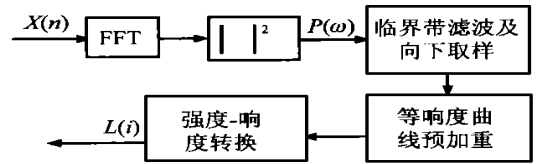


图2 BSD方法的听觉转换模型

输入语音  $x(n)$  和输出语音  $y(n)$  分别按听觉转换模型进行相同的预处理,得到 Bark 谱  $L_x(i)$  和  $L_y(i)$ , 然后进行 BSD 计算。BSD 定义为:  $BSD^{(k)} = \frac{N}{i=1} [L_x^{(k)}(i) - L_y^{(k)}(i)]^2$ 。其中,  $N$  为临界带宽的个数,  $L_x^{(k)}(i)$  为第  $k$  帧输入语音的 Bark 谱,  $L_y^{(k)}(i)$  为第  $k$  帧输出语音的 Bark 谱。继 S. Wang 等之后, R. Yanco, W. Yang 等学者于近几年先后对 BSD 进行了进一步研究,提出了 MBSD 方法。MBSD 与主观评价相关度达到了 0.195 左右<sup>[20~21]</sup>。另外,我国学者在 BSD 基础上也作了许多研究工作<sup>[23~25]</sup>。

ITU-T 于 1996 年公布了对电话频带(300~3400Hz)语音编解码器质量的客观评价方法建议(P. 861 建议):感知语音质量测度 PSQM 方法<sup>[26]</sup>。PSQM 方法与主观评价的相关度可达到 0.194。另外,美国 ITS 的 S. Voran 曾对 ITU-T 的 PSQM 方法的计算作过简化<sup>[27]</sup>,也取得了比较好的效果。但值得注意的是,PSQM 方法只是对高速率编码语音的质量评价效果较好,且要求没有传输比特误码、帧丢失(如在无线移动通信中)或信元丢失(如 ATM 网中)的信道降级等<sup>[26]</sup>。

S. Voran 在总结前人工作基础上,提出了一种度量标准化段 MNB(Measuring Normalizing Block)的语音质量客观评价方法<sup>[28]</sup>,并在 MNB 基础上于 1999 年提出基于 MNB 的听觉距离 AD/MNB(Auditory Distance based on Measuring Normalizing Block)评价方法<sup>[29,30]</sup>。S. Voran 认为人对语音质量优劣好坏的评价过程应包括两个方面,一个是听过程,另一个是判断过程。虽然这两个过程不能被严格区分开,但是人在感觉语音质量时在这两个过程中的行为是有所不同的。象 MBSD、PSQM 等方法都比较侧重于模拟人的听过程(听觉模型),而对判断过程则予以很大的简化。AD/MNB 方法则在考虑听过程的基础上,采用 MNB 方法来模拟人的判断过程,在 MNB 的基础上再求出听觉距离 AD(Auditory Distance),以其作为评价的尺度标准。

由于 AD/MNB 方法综合了听过程和判断过程,因而评价结果与主观评价的相关程度较高。S. Voran 在大量实验基础上提出了两种 MNB 结构: MNB1 型和 MNB2 型,基于这两种结构的 AD 与主观评价结果的相关度都达到 0.95 左右<sup>[30]</sup>。

### 2.1.2 基于输出评价方法的研究发展

以上介绍的客观评价方法都是基于输入/输出的评价,在九十年代中期基于输出的客观评价方法研究已开始有所进展,如 R. Kubichek, J. Liang, C. Jin 等学者提出了基于 PLP 模型的客观评价方法<sup>[31, 32]</sup>。

基于输出的客观评价方法其关键之处在于解决语音多变性问题,不同人具有不同发声特征和基频特性,使得输出语音信号变化极大。在基于输入/输出方式下,输入语音信号可作为已知参考信息,而在输出方式下无参考信息。因此,在基于输出的客观评价中要避免受语音信号中依赖发音人的特征信息的干扰,而必须从输出语音信号中提取出非依赖发音人的特征信息来作为客观评价依据。也就是说,依赖发音人的细节特征信息应从输出语音信号中剔除掉或者对其进行最大地抑制。然而,表达语音信号特征的许多参量都是高度依赖于发音人的,所以从中提取非依赖发音人的特征信息存在着一定难度。不过, H. Hermansky 于 1990 年提出的 PLP (Perceptual Linear Predictive) 模型为解决这一问题提供了一条可尝试的技术路径<sup>[33]</sup>。

PLP 模型的基本思想是模仿人类听觉系统对声音频率和幅度选择接收的特点,把语音频谱在频域范围内进行听觉加权,从而转换成听觉频谱,再求出听觉 AR 模型,然后借用这个听觉模型的感知线性预测系数来描述听觉谱的变化规律。特别值得注意的是, Hermansky 通过大量语音实验发现低阶 PLP 分析能够抑制住语音信号中依赖于发音人的个性信息,当取 5 阶 PLP 分析时这种抑制达到最大<sup>[33]</sup>,恰好这正是基于输出的客观评价方法所需要的。R. Kubichek 等学者就是以低阶 PLP 模型为基础,尝试对没有参照信息的输出语音信号进行质量评价。该方法首先对无失真语音样本进行 PLP 特征提取,并运用矢量量化技术构造出中心参考类集(码本),然后,再以此参考类集作评价基准模板,求出待测试语音与参考类集的最近距离,以其作为该语音质量好坏的评价标准。R. Kubichek 等人分别采用了 PLP、PLP2CD (PLP 倒谱距离) 和 PLP2DCD (PLP 差分倒谱距离) 等参数进行实验测试,与主观评价的相关度分别达到 0.82、0.84 和 0.67 左右<sup>[32]</sup>。

R. Kubichek 等人的研究工作表明这种基于输出的客观评价方法与主观评价具有一定的相关性,但还不成熟,有待于作进一步深入研究。同时也可以看到,虽然这种方式的评价方法是仅以语音系统的输出信号作为评价依据,但是从本质上说这种基于输出方式的客观评价方法还是依赖了一个可参考的输入信息,只不过这个参考信息是经过学习、训练、归纳等过程之后而形成的,是一种广义的输入信息。正如人在进行主观评价时,实质上也是把语音系统的输出信号与人在自身成长过程中不断学习、记忆而得的经验知识作比较。因此,从这个意义上讲,基于输出的客观评价方法实际上是一种广义的基于输入/输出的客观评价方法。

## 3 语音质量客观评价方法的分类

语音质量客观评价研究自七十年代以来迅速发展,国内外学者提出了数以千计的客观评价方法。这些方法从评价结构上可分为基于输入/输出和基于输出这两大类方法,从它们各自使用的主要技术(如谱分析、LPC 分析、听觉模型分析、判断模型分析等)和主要特征参数(时域参数、频域参数、变换域参数等)又可以分为以下六类:

### 3.1 基于 SNR 评价方法<sup>[12]</sup>

信噪比是一种广为应用的简单客观评价方法,高信噪比是高质量语音的必要条件,但不是高质量语音的充分条件。大量的实验表明,单一的 SNR 预测主观评价的能力极差。经过改进的分段信噪比、变频分段信噪比等方法与主观评价的相关度有所提高,但都只是对高速率的波形编码语音而言。

### 3.1.2 基于 LPC 技术评价方法<sup>[12, 15, 34, 35]</sup>

这类方法是以 LPC 分析技术为基础的,把 LPC 系数及其导出参数作为评价的依据参量。由 LPC 导出的方法有: LRC (Linear reflection coefficient)、LIR (Log likelihood ratio)、LSP (Line spectrum pairs)、LAR (Log area ratio)、Itakura、CD<sup>[15]</sup> 等方法以及它们的一些改进方法<sup>[34, 35]</sup>。

### 3.1.3 基于谱距离评价方法<sup>[12, 36, 37]</sup>

基于谱距离的评价方法是以语音信号平滑谱之间的比较为基础的。谱距离评价有很多种,主要有: SD (Spectral Distance)、LSD (Log SD)、FVLISD (Frequency variant linear SD)、FVLOSD (Frequency variant log SD)、WSD (Weighted slope SD)、ILSD (Inverse log SD) 等方法<sup>[36, 37]</sup>。

### 3.1.4 基于听觉模型评价方法<sup>[19~22, 26, 31~33, 38]</sup>

这类评价方法是以人感知语音信号的心理听觉特性为基础。具有代表性的听觉模型方法有 BSD<sup>[19]</sup>、MBS<sup>[20~22]</sup>、PSQM<sup>[26]</sup>、PLP<sup>[31~33]</sup>、MSD (Mel Spectral Distortion)<sup>[38]</sup> 等。

### 3.1.5 基于判断模型的评价方法<sup>[28~30, 39]</sup>

这类评价方法是在选择表达语音质量的特征参量基础上,更主要侧重于模拟人对语音质量的判断过程。如 S. Voran 提出的 AD/MNB 方法<sup>[28~30]</sup>及我国学者丁瑾等提出的模糊决策树方法等<sup>[39]</sup>。

### 3.1.6 其它评价方法<sup>[13, 14, 16, 17]</sup>

主要有一致函数 CHF 法<sup>[13, 14]</sup>、信息指数 II 法<sup>[16]</sup>、专家模式识别 EPR 法<sup>[17]</sup> 等。

## 4 客观评价方法的性能比较及其发展方向

客观评价和主观评价之间的联系常用一种函数映射关系来表示,即  $S_i = T(O_i)$ , 其中,  $T$  为转换函数,  $O_i$  为第  $i$  个样本的客观评价,  $S_i$  为由  $T(\#)$  将  $O_i$  转换成的主观评价,  $T(\#)$  可以是线性或非线性回归关系也可以是多项式拟合关系。由于客观评价实质上是对主观评价的一种预测,因此客观评价方法的性能好坏是与其与实际主观评价的相关性来衡量,即  $S_i$  与  $S_i$  的相关程度,  $S_i$  为实际主观评价。这种相关程度可用它们的相关度  $Q$  和误差范围  $R$  来表示。计算公式见式(1):

$$Q = \frac{\sum_{i=1}^N (S_i - L_s)^2}{\sum_{i=1}^N (S_i - L_s)^2}, R = R_s(1 - Q^2)^2 \quad (1)$$

其中,  $N$  为样本数,  $S_i$  表示实际主观评价值,  $S_i$  表示由客观评价方法预测出的主观评价值,  $L_s$ 、 $R_s$  分别为实际主观评价值的均值和标准偏差。

表 1 分类给出了目前具有代表性的客观评价方法的相关度数值表。由于受测试数据、测试方式等因素的影响,因而在不同文献中使用同样方法却会得到不同结果。这里,本文采取:在同一文献中,对同一评价方法因不同测试数据所得到的不同结果取平均值;对不同文献中的同一评价方法所得到的不同结果取最大值。

从客观评价方法的发展历程以及各种方法的性能来看,语音质量客观评价的研究发展具有以下特点:

(1) 语音质量客观评价研究取得了十分可喜的成绩。基于输入/输出的客观评价方法与主观评价的相关度已达到了 0.95 左右,如 AD/MNB、MBSD 和 PSQM 等方法;基于输出的客观评价方法已开始展开研究,目前相关度已达到 0.84 的水平。

表 1 语音质量客观评价方法性能比较表

语音质量评价方法	Q	语音质量评价方法	Q
基于信噪比评价方法		Auditory freq. weighted log SD	0.78
SNR	0.24#	Inverse log SD	0.75
Segment SNR	0.77#	基于听觉模型评价方法	
Frequency variant seg. SNR	0.93#	MSD	0.86
基于 LPC 分析评价方法		BSD	0.89
Log LPC	0.34	MBSD	0.95
Linear reflection coefficient	0.46	PSQM	0.94
Log likelihood ratio	0.48	PLP	0.82
Line spectrum pairs	0.35	PLP2Cepstral	0.84
Log area ratio	0.62	PLP2Delta Cepstral	0.67
Itakura	0.59	基于判断模型评价方法	
CD	0.90	L(AD)/MNB1	0.95
基于谱距离评价方法		L(AD)/MNB2	0.96
Spectral distance(SD)	0.80	其它评价方法	
Log SD	0.60	II	0.69
Frequency variant Linear SD	0.68	CHF	0.82
Frequency variant log SD	0.70	EPR	0.88
Weighted slope SD	0.74		

(# 只是对波形编码语音的测试结果)

(2) 从语音质量客观评价性能的提高过程来看,客观评价方法研究大致经历了这样几个阶段:时域阶段(如 SNR 等),频域谱分析阶段(如 SD 等),模型参数(LPC 分析)阶段(如 CD 等)、听觉模型阶段(BSD, MBSD, PSQM 等)及听觉模型与判断模型的混合(hybrid)模型阶段(如 AD/MNB 等)。当然,各阶段不是严格分开的,而是互相重叠,特别是新方法往往是建立在传统方法基础上。

(3) 听觉模型在语音质量客观评价研究中占有十分重要的地位。纵观客观评价的发展过程可以清楚看到只要在评价中考虑了人对语音信号的感知特性就会大幅度提高整个评价方法的性能。从近几年的研究成果来看,性能较好的客观评价

方法都以听觉模型为基础,如 BSD、MBSD、PSQM、AD/MNB 等。因此,可以说从人的听觉器官功能和听神经系统的特点来进一步研究客观评价模型是将来的发展趋势之一。

(4) 判断模型的研究已开始得到重视。人对语音质量的评价包含两个过程:听过程和判断过程,因此,有必要构造良好的判断模型,并与听觉模型相结合。虽然目前判断模型的研究已有所进展,如 AD/MNB 方法等,但在这方面的研究工作还有待于进一步深入下去。

(5) 基于输出的客观评价方法研究有所发展,但尚属萌芽阶段。目前这种形式的客观评价方法与主观评价的相关度还较低。尽管该方法具有巨大的发展前景和实际应用价值,但是在理论方法上和使用的技术参数等方面还有待于有突破性的进展。

另外,随着研究的深入,客观评价研究也要求引入新的分析技术,如非线性预测分析,多精度时频分析技术(包括子波变换技术),高阶统计分析技术,模糊处理技术以及人工神经网络技术等等。预计这些技术更能挖掘人的听觉感知机理,更能从类似人的听觉行为来进行语音质量评价,使客观评价过程更接近于主观评价的处理方式。当然,这还与人体生理学、心理学等学科的发展息息相关。

## 5 结束语

本文回顾了语音质量客观评价方法的研究发展情况,介绍了语音质量客观评价方法的分类与性能,同时,探讨了今后语音质量客观评价方法的研究发展方向。随着信息时代的到来,对语音质量客观评价方法研究提出了更加迫切的要求,语音质量客观评价技术将得到进一步发展。

## 参考文献:

- [1] IEEE. IEEE recommended practice for speech quality measurements [J]. IEEE Trans. on Audio and Electroacoust., Sep. 1969: 227- 246.
- [2] N. Kitavaki, M. Honda, K. Itoh. Speech quality assessment methods for speech coding systems [J]. IEEE Communications Magazine, Oct. 1984, 22(10): 26- 33.
- [3] W. D. Voiers. Diagnostic Evaluation of Speech Intelligibility [M]. in M. E. Hawley (Ed.), Benchmark Papers in Acoustics, Stroudsburg, PA: Dowden, Hutchinson and Ross, 1977.
- [4] W. D. Voiers. Diagnostic acceptability measure for speech communication systems [A]. Proc. 1977 IEEE ICASSP [C], 1977: 204- 207.
- [5] N. R. French, J. C. Steinberg. Factors governing the intelligibility of speech sounds [J]. J. Acous. Soc. Amer., Jan. 1947, 19(1): 90- 119.
- [6] T. P. Barnwell III, A. M. Bush. Statistical correlation between objective and subjective measures for speech quality [A]. Proc. 1978 IEEE ICASSP [C], 1978: 586- 590.
- [7] T. P. Barnwell III. Objective measures for speech quality testing [J]. J. Acoust. Soc. Amer., 1979, 66(6): 1658- 1663.
- [8] T. P. Bamwell III. Correlation analysis of subjective and objective measure for speech quality [A]. Proc. 1980 IEEE ICASSP [C], 1980: 706 - 709.
- [9] T. P. Bamwell III. A comparison of parametrically different objective

- speech quality measures using correlation analysis with subjective quality results [A]. Proc. 1980 IEEE ICASSP [C], 1980: 710- 713.
- [10] T. P. Bamwell III, S. R. Quackenbush. An analysis of objectively comparable measures for speech quality testing [A]. Proc. 1982 IEEE ICASSP [C], 1982: 996- 999.
- [11] S. R. Quackenbush, T. P. Bamwell III. The estimation and evaluation of pointwise nonlinearities for improving the performance of objective speech quality measures [A]. Proc. 1983 IEEE ICASSP [C], 1983: 547- 550.
- [12] S. R. Quackenbush, T. P. Bamwell III, M. A. Clements. Objective Measures of Speech Quality [M]. Englewood Cliffs, NJ: Prentice Hall, 1988.
- [13] Bell Northern Research. Objective Evaluation of Nonlinear Distortion Effects on Voice Transmission Quality [M]. Contribution to CCITT, COM XI246E, March 1986.
- [14] Bell Northern Research. Reevaluation of the Objective Method for Measurement of Nonlinear Distortion [M]. Contribution to CCITT, COM XI2172E, June 1987.
- [15] N. Kitawaki, H. Negabuchi, K. Itoh. Objective quality evaluation for low bit rate speech coding systems [J]. IEEE Journal on Sel. Areas in Communications, 1988, 6(2): 242- 248.
- [16] J. Lalou. The information index: An objective measure of speech transmission performance [J]. Ann. Telecommun., 1990, 45(22): 47- 65.
- [17] R. Kubichek, E. A. Quincy, L. L. Kiser. Speech quality assessment using expert pattern recognition techniques [J]. IEEE Pacific Rim Conference on Computers, Communication, and Signal Processing, Jun. 1989: 216- 219.
- [18] R. Kubichek, Atkinson D, A. Webster. Advance in objective voice quality assessment [A]. Proc. of IEEE Global Telecommunication Conference [C], 1991, 3: 1765- 1770.
- [19] S. Wang, A. Sekey, A. Gersho. An objective measure for predicting subjective quality of speech coders [J]. IEEE Journal on Sel. Areas in Communications, June 1992, 10(5): 819- 829.
- [20] W. Yang, M. Dixon, R. Yantorno. A modified bark spectral distortion measure which uses noise masking threshold [A]. in Proc. 1997 IEEE Workshop Speech Coding for Telecommunications [C], 1997: 55- 56.
- [21] W. Yang, M. Benbouchta, R. Yantorno. Performance of the modified bark spectral distortion as an objective speech quality measure [A]. Proc. 1998 IEEE ICASSP [C], 1998: 541- 544.
- [22] W. Yang, R. Yantorno. Improvement of MBSD by scaling noise masking threshold and correlation analysis with MOS difference instead of MOS [A]. Proc. 1999 IEEE ICASSP [C], 1999: 673- 666.
- [23] 陈国, 胡修林等. 一种基于听觉特性的语音失真测度方法 [J]. 声学学报, 2000, 25(5): 463- 467.
- [24] 王瑛, 张知易. 一种基于人耳听觉特性的语音客观测度研究 [J]. 通信技术, 1999, 3: 62- 66.
- [25] 吴淑珍, 赵朝阳. 基于听觉模型的客观音质评价方法研究 [J]. 电子学报, 1999, 27(7): 92- 94.
- [26] ITU-T Rec. P. 861. Objective Quality Measurement of Teleband Speech Codecs [M]. Geneva, 1996.
- [27] S. Voran. A simplified version of the ITU algorithm for objective measurement of speech codec quality [A]. Proc. 1998 IEEE ICASSP [C], 1998: 537- 540.
- [28] S. Voran. Estimation of perceived speech quality using measuring normalizing blocks [A]. Proceedings of the 1997 IEEE Speech Coding Workshop [C], 1997: 83- 84.
- [29] S. Voran. Objective estimation of perceived speech quality part I: Development of the measuring normalizing block technique [J]. IEEE Trans. on Speech and Audio Processing, 1999, 7(4): 371- 382.
- [30] S. Voran. Objective estimation of perceived speech quality part II: Evaluation of the measuring normalizing block technique [J]. IEEE Trans. on Speech and Audio Processing, 1999, 7(4): 383- 390.
- [31] J. Liang, R. Kubichek. Output based objective speech quality [A]. 1994 IEEE 44th Vehicular Technology Conference [C], 1994, 3: 1719- 1723.
- [32] C. Jin, R. Kubichek. Vector quantization techniques for output based objective speech quality [A]. Proc. 1996 IEEE ICASSP [C], 1996: 491- 494.
- [33] H. Hemansky. Perceptual linear predictive PLP analysis of speech [J]. J. Acoust. Soc. Am., 1990, 87(4): 1738- 1752.
- [34] 吴淑珍, L. C. W. Pols. 汉语通信系统客观音质评价方法 [J]. 声学学报, 1998, 23(2): 170- 174.
- [35] 陈国, 胡修林等. 对 Itakura 语音失真测度的改进 [J]. 华中理工大学学报, 1999, 27(10): 101- 103.
- [36] K. H. Lam, O. C. Au. Objective speech measure for Chinese in wireless environment [A]. Proc. 1995 IEEE ICASSP [C], 1995: 277- 280.
- [37] K. H. Lam, O. C. Au. Objective speech quality measure for cellular phone [A]. Proc. 1996 IEEE ICASSP [C], 1996: 487- 490.
- [38] R. Kubichek. Mel cepstral distance measure for objective speech quality assessment [A]. in Proc. IEEE Pacific Rim Conf [C]. Communications, Computers, and Signal Processing, 1993: 125- 128.
- [39] 丁瑾, 钟涛, 胡健栋. 语音质量的一种新的评价方法 [J]. 电子学报. 1997, 25(4): 6- 9.

#### 作者简介:



陈国 1968年生. 1995年在武汉工业大学(现武汉理工大学)自动化系获工学硕士学位, 现为华中科技大学电信系博士生, 目前主要从事数字信号处理、语音质量评价、语音识别、语音及音频编码、多媒体通信系统等方面的研究与开发工作。



胡修林 1945年生. 华中科技大学电信系教授、博士生导师, 长期从事现代通信理论、通信系统、多媒体通信、计算机网络等方面的研究工作, 在国内外发表学术论文60多篇。

张蕴玉 1948年生. 华中科技大学电信系副教授, 主要从事微型计算机应用、信号处理, 信号检测及多媒体通信等方面的研究工作。