

# 非线性失真多频声音质评价研究

· 论文 ·

岳德刚, 谢志文

(华南理工大学 物理科学与技术学院, 广东 广州 510640)

【摘要】回顾了目前非线性失真客观测量方法, 及其不能够充分反映电声系统非线性失真特性的原因, 并介绍了国内外非线性失真主观音质评价的最新科研成果, 指出利用多频声评价非线性失真将是一个较好的选择。

【关键词】非线性失真; 音质评价; 多频声

【中图分类号】TB53

【文献标识码】A

Sound Quality Evaluation of Nonlinear Distortion with Multitone

YUE De-gang, XIE Zhi-wen

(School of Physics, South China University of Technology, Guangzhou 510640, China)

【Abstract】The present objective measuring methods for nonlinear distortion and the reasons why they cannot sufficiently reflect the characteristics of electro-acoustical systems nonlinearity are analysed. The latest scientific researches of nonlinear distortion about subjective and objective sound quality evaluation around the world are introduced. It is pointed out that multitone will be a good choice to evaluate nonlinear distortion.

【Key words】nonlinear distortion; sound quality evaluation; multitone

## 1 引言

所有的电声换能器件 (例如扬声器和传声器) 以及传输通道都不可避免地带有某种程度的失真, 即接收到或者重发后的信号同原始信号并不是完全一致。失真可分为两种类型: 线性失真和非线性失真, 而非线性失真的存在对电声系统的影响比线性失真更大。如何评价非线性失真的大小, 并据此改善电声产品的性能一直都是人们非常感兴趣的问题。

## 2 非线性失真的测量方法

从广义上来说, 失真可分为两种类型:

(1) 线性失真: 重发后的信号与原始信号相比, 频率成分不变, 但各频率成分的振幅或相位发生变化, 从而产生音质变化或声染色, 称之为线性失真。线性失真可以通过线性滤波器的补偿作用来消除, 并且这种滤波器可以用在电声器件及传输通道之前或之后<sup>[1-2]</sup>。

(2) 非线性失真: 重发后的信号与原始信号相比, 出现了新的频率成分, 称之为非线性失真。这种非线性失真对音质产生的影响是很难用信号处理的方法加以消除的。在扬声器的音质评价中, 非线性失真是影响音质的重要因素。因为人耳对频率的变化非常敏感, 经过训练的人耳可以分辨到 1 Hz 以下。非线性失真还可以

进一步细分: 将一定频谱的乐音馈给扬声器, 如果接收到的声音里增加了新谐波成分称为谐波失真; 如果因其频率分量的相互作用而出现了原来没有的新的频率成分, 则称为互调失真。上述非线性失真的大小通常用正弦信号或者扫频信号来测量。

### 2.1 谐波失真及其测量

谐波失真是由振幅非线性引起的一种失真。根据所用的测试信号不同, 扬声器谐波失真的测试方法有两种: 纯音信号测试法和窄带噪声信号测试法。

(1) 纯音信号测试法: 当扬声器馈给某一频率的纯音信号时, 扬声器输出的声信号中除了原来输入的信号频率 (称为基频) 外, 还出现了二倍于、三倍于……基频的信号, 这些信号称为谐波。这种现象就叫谐波失真。一般用声压谐波系数来描述, 定义为

$$K = \sqrt{\frac{P_2^2 + P_3^2 + \dots + P_n^2}{P_1^2 + P_2^2 + \dots + P_n^2}} \times 100\% \quad (1)$$

式中:  $P_1$  为基频声压的均方根值,  $P_2$  为二次谐波的声压均方根值, 依此类推,  $P_n$  为  $n$  次谐波的声压均方根值。测量方法如图 1 所示, 而总谐波失真系数定义为

$$\gamma = \sqrt{\frac{P_2^2 + P_3^2 + P_4^2 + \dots}{P_1^2}} \times 100\% \quad (2)$$

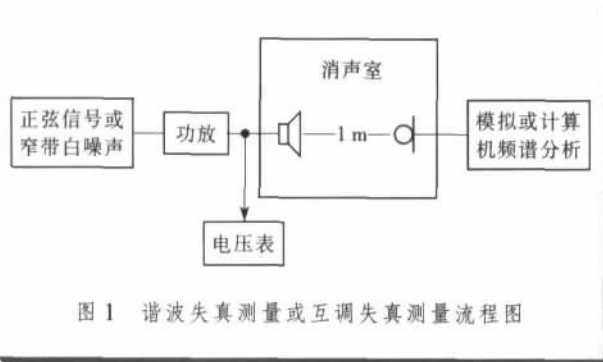


图1 谐波失真测量或互调失真测量流程图

(2) 窄带噪声信号测试法: 测试条件与用纯音信号测试相同, 将扬声器置于消声室中, 馈给扬声器某一1/3倍频程的窄带噪声信号, 其中心频率为 $f$ , 其电压大小由扬声器的额定功率和阻抗决定, 或由离扬声器1 m远处的额定声压级决定。测试原理同样如图1所示。扬声器辐射的声波, 由传声器接收后馈给计算机进行频谱分析, 分别读出中心频率为 $f, 2f, 3f, \dots, nf$ 的1/3倍频程带宽的幅度, 然后按下式计算谐波失真系数。

$$K = \sqrt{\frac{A_2^2 + A_3^2 + \dots + A_n^2}{A_1^2 + A_2^2 + \dots + A_n^2}} \times 100\% \quad (3)$$

式中:  $A_1$ 为基波幅度,  $A_2, \dots, A_n$ 分别为2次, 3次,  $\dots, n$ 次谐波幅度。

用窄带噪声信号测试的结果是某一频带范围内的失真, 而用纯音信号测试的结果则仅为点失真, 为了提高点失真测量的效率, 出现了扫频信号测量非线性失真的方法。

## 2.2 互调失真及其测量

互调失真是指由于系统的非线性使得一个复杂信号中的各分量相互作用而产生出原始信号中没有的新的频率分量。

在测量互调失真时, 一般同时馈给扬声器两个纯音信号, 频率分别为 $f_1$ 和 $f_2$  ( $f_2 > f_1$ ), 则输出信号中含有频率分别为 $f_2 - f_1, f_2 + f_1, f_2 + 2f_1, |f_2 - 2f_1|$ 等的成分, 从而得到互调失真为

$$D = \frac{\sqrt{\sum_i A_{f_2 \pm f_1}^2}}{A_{f_1}} \times 100\% \quad (4)$$

式中:  $A_{f_1}$ 是频率为 $f_1$ 信号的振幅,  $A_{f_2 \pm f_1}$ 是频率为 $f_2 \pm f_1$ 信号的振幅。其测量原理类似谐波失真测量, 同样如图1所示, 但测试信号采用如下形式

$$x(t) = A_{f_1} \sin(2\pi f_1 t) + A_{f_2} \sin(2\pi f_2 t) \quad (5)$$

式中:  $A_{f_1}$ 是频率为 $f_1$ 信号的振幅,  $A_{f_2}$ 是频率为 $f_2$ 信号的振幅。

## 3 非线性失真音质评价的研究现状

扬声器是一类复杂的非线性动态系统, 不同的激励信号通过它们之后所产生的非线性失真的结果有显著的差别, 传统的测量互调失真和总谐波失真 (THD) 的方法是使用正弦信号或扫频信号。然而这两种方法并不能充分反映电声器件的非线性失真信息<sup>[3-4]</sup>。例如, 对应于某个基频的输出信号, 当其输出信号中含有的频率成分分离基频足够近, 且幅度足够小时, 就会被这个基频成分掩蔽, 以至于人耳不能辨别出来; 另一方面, 如果输出信号中的这个频率成分分离基频较远, 就很容易被听到。而且输出信号所产生的多余频率成分还会与基频成分互相作用而产生其他的一些效果 (例如幅度或频率调制), 这种效果能否被人耳感知到取决于输出信号产生的多余频率成分同基频之间的频率分离度和相位差别。而传统意义上的互调失真和谐波失真并没有考虑到这些情况的影响, 所以它们对于电声产品的性能的评价是有待改善的。

Eugene Czerwinski 等学者详细地讨论了谐波失真、互调失真等客观测试信号之所以不能很好地反映电声产品性能的原因。对于一个静态的具有多项式特征的简单非线性系统来说, 激励信号通过该系统导致的互调失真要超过谐波失真。所谓简单的具有静态多项式特征的非线性系统, 指的是对于一个输入信号 $x(t)$ , 该系统的输出 $y(t)$ 由下式决定

$$y(t) = h_1 x(t) + h_2 x^2(t) + h_3 x^3(t) + \dots + h_n x^n(t) \quad (6)$$

式中:  $h_1$ 是输出信号线性部分的系数, 而 $h_2, \dots, h_n$ 决定了高阶非线性系数。当系统的阶数越高 (即 $n$ 越大), 则谐波失真和互调失真的差别就越明显。像扬声器或号角驱动器之类的电声换能器是一类复杂的非线性系统, 它们的谐波失真 (HD) 和互调失真 (IMD) 就更加复杂, 并且和输入信号 $x(t)$ 密切相关, 在互调失真的测量过程中, 不可避免地会掺入谐波失真的影响, 要想使互调失真完全独立于谐波失真之外是很难的。仅仅依据谐波失真根本无法准确评价被测系统的非线性失真程度, 根据人耳的掩蔽效应可知, 高频谐波成分很可能已经被基频所掩蔽, 而在目前的谐波失真或总谐波失真的计算中, 并没有充分考虑到这些情况的影响, 互调失真同样也存在这方面的问题<sup>[9]</sup>。

当掩蔽效应发生时, 对于一个纯音信号来说, 掩蔽曲线会随着其声压级的改变而改变, 如图 2 所示。在图 2 中, 随着基频(掩蔽音)声压级的增大, 掩蔽曲线(a, b, c, d, e)与频率轴围成的图形逐渐由原来的近似对称三角形变为明显的非对称形状, 并且逐渐向高频部分延伸, 从而出现基频声压级较高时, 其所掩蔽的高频成分就越多。图 3 中表示的是含有两个频率成分纯音的掩蔽曲线。非线性失真的最高阶数为 5 阶[见式(5)], 互调失真的差频部分由于处于两个纯音掩蔽曲线之上, 故很容易被人耳感知到, 而其它一些高次谐波成分或互调失真之和却未能被察觉。由图 2、图 3 可看出掩蔽效应的复杂性, 如何在非线性失真的主客观评价中考虑到这些情况, 是一个亟待解决的问题。

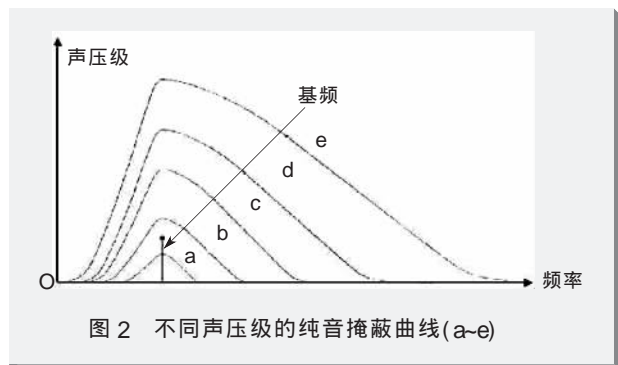


图 2 不同声压级的纯音掩蔽曲线(a-e)

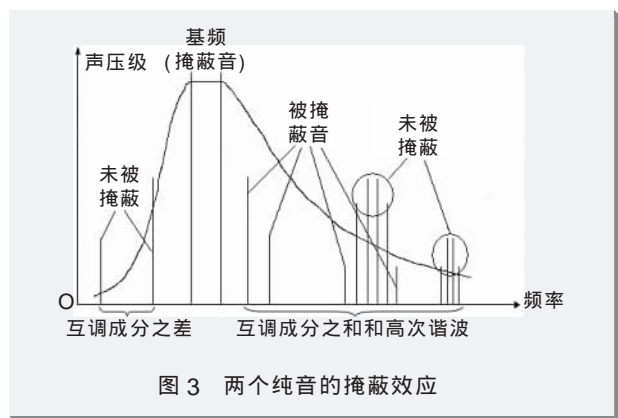


图 3 两个纯音的掩蔽效应

另外一种评价非线性失真的方法是由 Kates 提出的<sup>[6]</sup>, 主要用于助听器的失真评价, 他建议使用通过梳状滤波器的白噪声作为激励信号, 这种信号有点类似于多频声。还有一些学者建议使用输入输出信号的一致性来作为评价非线性失真的方法<sup>[7]</sup>, 所谓一致性就是归一化的输入、输出信号的互相关频谱密度, 它可以使用白噪声、伪随机噪声、最大长度序列(MLS)作为测试信号。然而这种方法存在一定的局限性, 对于时不变系统来说, 主客观评价的一致性符合得相当好, 但在其他方面则不能给出满意的结果。

另外在有关音频信号的编解码、传输、压缩过程中的失真问题的评价方法主要有: ITU-R BS.1387-1 所规定的 PEAQ 标准<sup>[8]</sup>以及 ITU-T P.862 所规定的 PESQ 标准<sup>[9]</sup>。但上述方法仅仅适用于信号本身, 并不适用于评价电声器件, 而且, PESQ 仅仅适用于窄带语言信号测量。

## 4 非线性失真的多频声音质评价

Eugene Czerwinski 等人在文献[3-4]中详细回顾了音频系统的非线性失真测量方法, 着重介绍了利用多频声进行测量的方法, 并将这种方法与其他传统的测量谐波失真、互调失真的方法进行对比。结果证明用多频声测得的结果比用传统的方法包含更多的信息, 也更能反应非线性系统的特性。但是如何利用多频声计算出相应的客观测试参量, 并对电声系统的非线性失真进行评价, 还在进一步探讨当中。

### 4.1 多频声理论基础

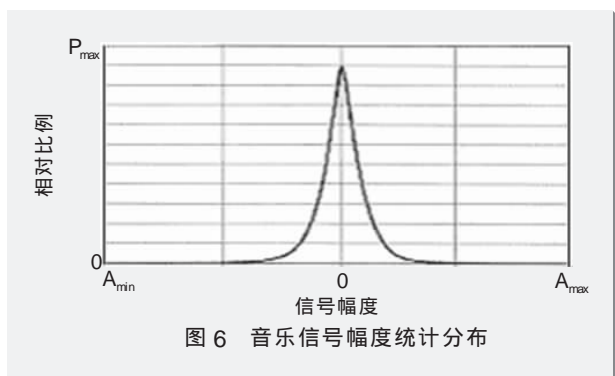
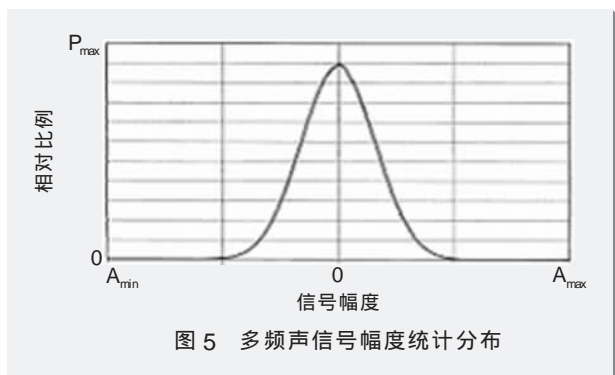
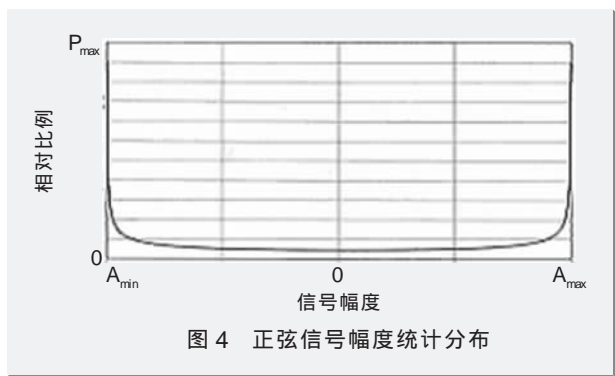
多频声可表示为

$$x(t) = \sum_{i=1}^N A_i \sin(\omega_i t + \varphi_i) \quad (7)$$

式中:  $x(t)$  是由  $N$  个正弦波所组成的多频声信号,  $A_i$  表示第  $i$  个正弦波振幅,  $\omega_i$  表示其角频率,  $\varphi_i$  表示初相位。

在多频声中, 各个不同频率成分可以随着测试目的的不同而采用不同的形式, 例如在频率坐标轴上这些频率成分可以是线性坐标均匀排列, 也可以是按对数坐标均匀排列。比较常用的方法是采用频率成分按对数均匀排列形式, 因为这种形式可以在一定程度上避免输出信号的谐波重叠, 并且信号不具有周期性, 更接近于实际的语音或音乐。由于非线性系统对信号的响应与信号的振幅分布有关, 而正弦波、多频声、音乐信号的振幅的统计分布确实具有不同的特性。因此, 要更好地研究非线性系统在真实工作状态下的表现, 测试信号的振幅分布是一个很重要的参数。图 4、图 5、图 6 分别显示了正弦波、多频声、音乐信号这 3 种信号的振幅统计分布图。其中  $P_{\max}$  表示信号分布最大值,  $A_{\max}$  表示信号幅度最大值。

从这 3 幅图中可以看出: 多频声的幅度分布统计特性同音乐信号的统计特性非常相似, 而正弦信号振幅的统计分布同音乐信号的振幅统计分布大致相反。因此, 从这一点也可看出在客观评价方面用多频声比用正弦波更有优势。



另外一个有关多频声的问题是各个频率成分的相位问题, 它可以通过峰值因数来进行优化。多频声的峰值因数可由下式来计算

$$CF = \frac{|A(t_k)|_{\max}}{\sqrt{\frac{1}{M} \sum_{k=1}^M A^2(t_k)}} \quad t_k \in \{0 \dots t_m\} \quad (8)$$

式中:  $M$  为样本数,  $|A(t_k)|_{\max}$  表示的是在时间间隔  $\Delta t$  内多频声振幅绝对值的最大值, 而分母就是时间间隔  $\Delta t$  内的有效值。

随着多频声峰值因数的减少, 可以在减少信号峰值的情况下增大测量的动态范围, 从而起到一种“平滑”波形的作用, 因而可以相应地增大信号的有效值、提高测量的信噪比。许多科研工作者曾做过这方面的研究工作<sup>[10-11]</sup>。Eugene Czerwinski 等人在文献[4]中提

出了用一种全局优化算法来减少峰值因数的计算方法, 目的是为了寻找在多频声频率按对数分布的情况下的优化算法。

他认为峰值因数的减少是一个相位优化问题, 可以通过调整各个正弦波的相位来达到目的, 即通过式(9)的相位改变来优化峰值因数

$$F(\varphi) = \max \left| \sum_{i=1}^N A_i \sin(\omega_i t + \varphi_i) \right| \quad \min$$

$$t \in \{0 \dots t_m\}, \quad \varphi = \{\varphi_1, \varphi_2, \dots, \varphi_N\} \quad (9)$$

式中:  $F(\varphi)$  为峰值因数因子绝对值。

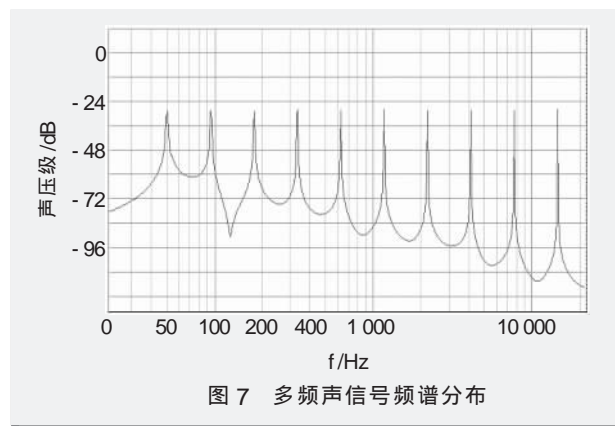
在式(8)中, 由于分母并不随着相位的变化而改变, 所以就可通过使式(9)的  $F(\varphi)$  达到最小来减少峰值因数,  $F(\varphi)$  随着相位的变化有许多局部极小值, 利用全局最优化算法计算出全局最小值。

Chin-Tuan Tan 和 Brian C. J. Moore 在文献[12]里详细讨论了评价非线性失真所需的多频声的最佳频率个数、相位分布及持续时间。他们在这篇文献里指出: 按对数坐标均匀分布含 10 个频率成分的持续时间为 1 s 钟的多频声信号为最佳信号, 其相位及频率如表 1 所示。

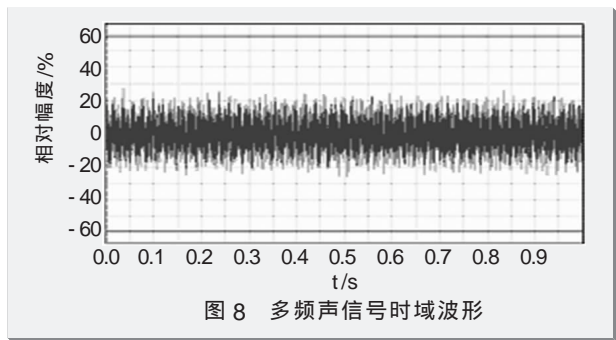
表 1 多频声频率及初相位(频率对数分布, 相位已经过优化)

序号	频率/Hz	相位/rad	序号	频率/Hz	相位/rad
1	50	2.594	6	1 189	0.364
2	94	5.878	7	2 241	2.217
3	178	5.761	8	4 223	5.109
4	335	2.578	9	7 959	0.062
5	631	5.615	10	15 000	0.873

按照上述方法产生的多频声频域及时域波形分别如图 7、图 8 所示。







#### 4.2 多频声的音质评价应用

Chin-Tuan Tan 和 Brian C. J. Moore 在文献[12-13]中利用多频声对非线性失真进行了评价。在文献[12]中, 他们所研究的非线性失真主要有两个方面: 一是人工产生的非线性失真; 另一部分是真实的听筒换能器产生的非线性失真。其所定义的客观评价参量为利用多频声处理而得到的失真分数(即 DS)。主客观实验的评价结果表明, 人工产生的最高相关性可达到-0.98, 而对于真实的听筒换能器其主客观评价的最高相关性仅能达到-0.67。在文献[13]中, 他们对客观参量进行了进一步的改进, 其中最显著的一点就是适当考虑到了人耳听觉特性对音质评价的影响, 得到了新的客观参量, 称之为 Rndin, 并以此来预测非线性失真的大小。结果表明: 人工产生的非线性失真两种评价结果主客观相关性相差不大, 但是对于真实的听筒换能器来说, 两种评价结果后者要明显优于前者。

## 5 总结

如何评价非线性失真的大小, 并利用它来改善电声产品的性能一直都是人们非常感兴趣的问题。现存的用正弦波或扫频信号得到的谐波失真、互调失真不能够很好地反映电声系统的非线性失真特性, 且容易引起较大的偏差。如何改变这一现状, 就需要寻找新的客观测试信号和客观测试参量, 使得主客观评价程度一致性较好, 而用多频声信号来评价电声系统的非线性失真特性是一个较好的选择。由于电声产品性能优劣的最终评价标准是人耳, 因此在客观参量的计算上还要充分考虑到人耳特性的影响, 包括外耳/中耳等响曲线校准、耳蜗的分频作用以及基底膜上的动态压缩和掩蔽效应等。

#### 参考文献

[1] NORCROSS S G, SOULODRE G A, LAVOIE M C. Further investigations of inverse filtering[C]// Proceedings of AES Convention 115. New York: [s.n.], 2003.

- [2] NORCROSS S G, SOULODRE G A, LAVOIE M C. Subjective effects of regularization on inverse filtering[C]// Proceedings of AES Convention 114. Amsterdam [s.n.], 2003.
- [3] CZERWINSKI E, VOISHVILLO A, ALEXANDROV S, et al. Multitone testing of sound system components- some results and conclusions: Part 1: History and Theory[J]. J. Audio Eng. Soc., 2001, 49(11): 1 011- 1 048.
- [4] CZERWINSKI E, VOISHVILLO A, ALEXANDROV S, et al. Multitone testing of sound system components- some results and conclusions: Part 2: Modeling and application [J]. J. Audio Eng. Soc., 2001, 49(12): 1 180- 1 192.
- [5] VOISHVILLO A, TEREKHOV A, TEREKHOV E, et al. Graphing, interpretation and comparison of the results of loudspeaker nonlinearity measurements[J]. J. Audio Eng. Soc., 2004, 52(4): 332- 357.
- [6] KATES J M. A test suit for hearing aid evaluation [J]. J. Rehab. Res. Devel., 1990, 27: 255- 278.
- [7] DYRLUND O. Coherence measurement in hearing instruments, using different broad-band signal[J]. Scandinavian Audiology, 1992, 21: 73- 78.
- [8] International telecommunication union. ITU-R BS.1387.1. Method for objective measurements of perceived audio quality[S].
- [9] International telecommunication union. ITU-T P.862. Perceptual evaluation of speech quality(PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and codecs[S].
- [10] VAN DEN BOS A. A new method for synthesis of low-peak-factor signals[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1987, 35(1): 120- 122.
- [11] FRIESE M. Multitone signals with low crest factor[J]. IEEE Transactions on Communications, 1997, 45(10): 1 338- 1 344.
- [12] TAN Chin-tuan, MOORE B C J. The effect of nonlinear distortion on the perceived quality of music and speech signals[J]. J. Audio Eng. Soc., 2003, 51(11): 1 012- 1 031.
- [13] TAN Chin-tuan, MOORE B C J. Predicting the perceived quality of nonlinearly distorted music and speech signals[J]. J. Audio Eng. Soc., 2004, 52(7/8): 699- 711.

#### 作者简介

岳德刚, 硕士研究生, 主要研究方向为电声技术, 声信号处理; 谢志文, 副教授, 硕士生导师, 主要研究方向为电声技术, 多媒体信号处理等。

[责任编辑] 史丽丽

[收稿日期] 2006-06-17