

# 正弦参数自适应联合量化

王 嵩

(北京石油化工学院信息工程学院,北京 102617)

**摘要:** 为实现音频信号的低速率编码,提出一种正弦参数量化的新方案.该方案利用人耳掩蔽效应和实时参数统计特性,基于高速率原理建立并显式求解率失真优化问题,得到具有解析形式的参数量化器,实现了在正弦分量间和正弦参数间动态分配编码比特.本量化方案无迭代运算过程,适合实时低速率音频编码应用.与经典方法比较,平均比特速率每正弦减少约17%.当速率大于15bit/正弦,重建信号的感觉失真小于球形量化方案.

**关键词:** 高速率原理; 正弦编码; 匹配追踪; 自适应联合量化

**中图分类号:** TN911.21      **文献标识码:** A      **文章编号:** 0372-2112 (2014)04-0798-06

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2014.04.027

## Adaptive Joint Parameter Quantization of Sinusoidal Parameters

WANG Song

(School of Information Engineering, Beijing Institute of Petro-Chemical Technology, Beijing 102617, China)

**Abstract:** In order to achieve low bit rate coding of audio signals, this paper presents a new scheme for quantization of sinusoidal parameters. By using the masking effect of the human auditory system and real-time parameter statistical properties, the scheme based on high-rate theory to establish a rate-distortion optimization problem that is solved explicitly, derives the parameter quantizers which have analytical expressions, and dynamically implements the coding rate distribution between sinusoidal components and sinusoidal parameters. The quantization scheme without iterative calculation processes, is suitable for real-time low-rate audio coding applications. Compared with the classical method, the average bit rate of the proposed scheme is reduced 17% per sinusoid. When the rate is larger than 15 bit per sinusoid, the perceptual distortion of the reconstructed signals is smaller than the spherical quantization scheme's perceptual distortion of the reconstructed signals.

**Key words:** high-rate theory; sinusoidal coding; matching pursuit; adaptive joint quantization

## 1 引言

音频信号的正弦模型包括幅度、频率和相位等参数,量化这些参数的传统方法为启发式方法,其特点是分配给各参数的量化比特数要由先验的心理声学知识来确定<sup>[1,2]</sup>.启发式方法具有确定的量化器表达式,实现简单,常用于有较高实时性要求的中等质量音频编码应用.而有更高质量要求的编码应用,往往需要率失真控制来迭代调整比特分配,以使量化误差达到最小.二者比较,启发式方法量化误差较大,而基于迭代的率失真控制量化方法实现复杂,难以满足实时性要求.

文献[3~5]的正弦参数联合量化方法较好解决了这些问题.在给定速率约束的条件下,该方法建立并求解一个感觉加权的均方量化误差最小化问题,得到具有解析形式的量化器,实现了以表达式而非经验值确定分

给各参数的比特数,保证了合成信号的质量.同时,在编码端不设解码器,去掉了运算复杂的解码/编码迭代优化过程.

正弦参数的联合量化基于高速率原理(high-rate theory)<sup>[6]</sup>.该原理假定可分配的比特数足够多,因而量化空间由大量量化胞腔覆盖,在量化胞腔内变量的概率密度函数约等于常数.高速率原理以点密度函数描述量化器,对于标量量化器,量化点密度等于量化步长的倒数.尽管联合量化使用高速率假设,但在低速率下,量化器仍有较好性能<sup>[7]</sup>.

联合量化源于循环对称高斯变量的无限制极点量化方法(Unrestricted Polar Quantization, UPQ)<sup>[8,9]</sup>.基于UPQ量化,文献[5]提出用于正弦参数的熵约束无限制极点量化(Entropy-Constrained Unrestricted Polar Quantization, ECUPQ)方案,但是该方案并未考虑频率参数的量

化问题。

文献[3,4]扩展了 ECUPQ 算法,实现了频率量化,称正弦参数(包括幅度、频率、相位)的联合量化为熵约束无限制球形量化(Entropy-Constrained Unrestricted Spherical Quantization, ECUSQ)。参数的 ECUSQ 表达式除与参数自身取值有关外,还与其它参数的取值有关,这是“球形量化”或者“联合量化”名称的由来。在文献[3]的基础之上文献[4,10]将心理声学效应用于量化器设计,使重建信号具有更好的主观质量。文献[10]还将多帧联合优化引入到 ECUSQ 算法中,进一步提升了算法性能。

文献[3~5,11]运用了输入信号的统计知识来推导量化器,这些由先验估计得到的知识在量化/解量化过程中保持不变。受运算复杂度及编码速率约束,正弦参数估计通常以迭代方式进行(例如匹配追踪算法)。随迭代的进行,残差能量和正弦分量幅度均呈逐次衰减趋势,而分量频率则逐次向高频端演变。可见,各次估计的参数的统计分布是有差异性的,因此,在量化过程中实时估计、应用参数的各次统计知识则更为合理。基于这一思路,本文提出适应参数统计特性变化的联合量化方法,该方法也基于高速率原理,但采取逐次更新参数统计知识的新策略,以优化量化过程。实验证实,新的自适应量化方法能更好地平坦量化器信噪比,当正弦幅度减小时,延缓信噪比降低速度,改善小幅度参数的量化分辨率,显著提升合成信号质量。

## 2 自适应联合量化原理

### 2.1 联合量化原理分析

正弦参数的联合量化利用人耳感觉特性,并通过求解一个熵约束的感觉加权量化误差最小化问题导出最优量化器。

设一帧音频信号为  $x(n)$ ,  $n = n_0, \dots, n_0 + N - 1$ , 以正弦分量的合成对其建模,

$$x(n) = \sum_{l=1}^L a_l \cos(\omega_l n + \varphi_l) \quad (1)$$

这里,  $L$  表示正弦总数,  $\{a_l, \omega_l, \varphi_l\}_{l=1}^L$  表示分量的幅度、频率和相位参数。运用感觉失真测度定义<sup>[13]</sup>, 第  $l$  条正弦分量的量化失真为:

$$D_l = \frac{1}{2\pi} \int_0^{2\pi} \mu_l(\omega) |E_l(\omega)|^2 d\omega \quad (2)$$

其中,  $\mu_l(\omega)$  表示感觉加权, 为原信号在第  $l$  条正弦频率处的掩蔽阈的倒数<sup>[13]</sup>;  $E_l(\omega)$  为第  $l$  条正弦量化误差的加窗傅立叶变换:

$$E_l(\omega) = \sum_{n=n_0}^{n_0+N-1} w(n) (a_l \cos(\omega_l n + \varphi_l) - \hat{a}_l \cos(\hat{\omega}_l n + \hat{\varphi}_l)) e^{-j\omega n} \quad (3)$$

其中,  $w(n)$  为窗函数,  $\{\hat{a}_l, \hat{\omega}_l, \hat{\varphi}_l\}$  为量化后的正弦参

数。假设  $N$  足够大,  $E_l(\omega)$  的主瓣因而足够窄, 大部分误差能量集中于主瓣。由于主瓣带宽远小于听觉的临界带宽, 所以主瓣内掩蔽阈约为常数,

$$D_l \approx \frac{\mu_l(\hat{\omega}_l)}{2\pi} \int_0^{2\pi} |E_l(\omega)|^2 d\omega \quad (4)$$

参考文献[10]的方法, 设常数  $\|\mathbf{w}\|^2 = \sum_{n=n_0}^{n_0+N-1} w(n)^2$ , 并运用 Parseval 定理, 上式可写成

$$D_l \approx \frac{\mu_l(\hat{\omega}_l)}{2} (\|\mathbf{w}\|^2 (a_l^2 + \hat{a}_l^2) - 2a_l \hat{a}_l \cdot \sum_{n=n_0}^{n_0+N-1} w^2(n) \cos((\omega_l - \hat{\omega}_l)n + \varphi_l - \hat{\varphi}_l)) \quad (5)$$

为导出参数量化器的点密度函数, 运用高速率假设: 在量化间隔内, 正弦参数服从均匀分布, 因此, 幅度、频率和相位的最优化点正好位于量化间隔中心, 量化步长  $\Delta\{a_l, \omega_l, \varphi_l\}$  分别由参数的点密度的倒数代替; 并且, 设幅度、频率和相位的联合概率密度为  $f(a, \omega, \varphi)$ , 常数  $\sigma^2 = \left(\frac{1}{\|\mathbf{w}\|^2}\right) \sum_{n=n_0}^{n_0+N-1} w^2(n) n^2$ 。运用泰勒展式展开式(5)中正弦函数项并省略高阶项。于是,  $D_l$  的期望

$$E[D_l] = \frac{\|\mathbf{w}\|^2 \mu_l(\hat{\omega}_l)}{24} \cdot \iiint f(a, \omega, \varphi) (g_{a_l}^{-2} + a^2 (\sigma^2 g_{\omega_l}^{-2} + g_{\varphi_l}^{-2})) da d\omega d\varphi \quad (6)$$

式中,  $E[\cdot]$  表示数学期望算子,  $g_{\{a_l, \omega_l, \varphi_l\}}$  分别表示  $\{a_l, \omega_l, \varphi_l\}$  的量化点密度函数。为了简化表达式, 式(6)省略了点密度函数的自变量, 所以需注意  $g_{\{a_l, \omega_l, \varphi_l\}}$  均为幅度、频率和相位的函数。

由于信号的总量化误差是各正弦分量量化误差的累加, 于是  $L$  条正弦的平均感觉量化失真  $D$  等于:

$$D = \frac{1}{L} \sum_{l=1}^L E[D_l] \quad (7)$$

综合以上, 各参数的优化点密度函数可由如下熵约束的量化失真最小化问题来求解:

$$\begin{aligned} \min_{g_{a_l}, g_{\omega_l}, g_{\varphi_l}} D \\ \text{s.t. } H \leq H_T, \quad l = 1, \dots, L \end{aligned} \quad (8)$$

其中,  $H_T$  为每条正弦的目标熵,  $H$  为实际分配给每条正弦的平均熵。在高速率假设下,

$$H = \frac{1}{L} \sum_{l=1}^L \iiint f(a, \omega, \varphi) \cdot (\log_2 g_{a_l} + \log_2 g_{\omega_l} + \log_2 g_{\varphi_l}) da d\omega d\varphi \quad (9)$$

运用 Lagrangian 乘子法, 将式(8)转化成无约束问题, 即  $\min J$ ,  $J = D + \lambda H$ 。这里,  $\lambda$  为 Lagrangian 乘子。由该无约束问题建立各参数的 Euler-Lagrange 方程<sup>[14]</sup>, 并求解,

即得各参数的点密度函数:

$$g_{a_l} = \left( \frac{\mu_l 2^{\frac{(2H-4b)}{3}}}{\sigma^{\frac{2}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}}, \quad (10a)$$

$$g_{\omega_l} = \left( \frac{\mu_l a_l^2 2^{\frac{(2H-4b)}{3}}}{\sigma^{\frac{-4}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}}, \quad (10b)$$

$$g_{\varphi_l} = \left( \frac{\mu_l a_l^2 2^{\frac{(2H-4b)}{3}}}{\sigma^{\frac{2}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}}, \quad (10c)$$

式中,  $\mu_l = \mu_l(\hat{\omega}_l)$ ,  $\mu_g = \left( \prod_{l=1}^L \mu_l \right)^{\frac{1}{L}}$ , 为感觉权的几何平均值,  $b = \int f_A(a) \log_2(a) da$ , 为对数幅度  $a$  的集合平均值,  $f_A(a)$  为幅度的概率密度函数. 在文献[5, 10, 11]中,  $b$  为先验估计值, 存放于编码端和解码端, 并在正弦参数的量化和解量化过程中保持不变.

## 2.2 统计参数估计与自适应联合量化

正弦音频编码器常用迭代方式估计参数, 各次估计的正弦分量幅度, 即模型/信号的匹配系数, 将随迭代次数的递增而呈逐次减小的趋势, 因而与幅度对应的对数幅度  $a$  的集合平均值  $b$  也将表现出与前者一致的变化趋势. 在参数的估计过程中同步地估计、更新  $b$  值则更为合理. 所以, 量化比特分配就不仅受感觉权影响, 而且还受每次估计出的正弦幅度的统计知识影响. 运用实时更新的统计知识, 就能进一步优化文献[5, 10, 11]所提出的量化方案. 下面, 重写  $b$  为  $b_l = \int f_A(a_l) \log_2(a_l) da_l$ , 表示第  $l$  次估计出的幅度  $a$  的对数估计值的统计平均, 并运用匹配追踪原理<sup>[16]</sup>, 给出  $b_l$  的估计算法, 以及参数的自适应联合量化方案.

第  $l$  次迭代, 最大化原子  $a_l$  与残差  $r_l$  的内积模值是匹配追踪选择原子的准则<sup>[15]</sup>. 由正交性和原子具有单位范数约束, 残差  $r_{l+1}$  的能量为<sup>[16]</sup>:

$$\|r_{l+1}\|^2 = \|r_l\|^2 - \underbrace{|\langle a_l, r_l \rangle|^2}_{a_l^2} \quad (11)$$

式中,  $\langle \cdot, \cdot \rangle$  表示内积算子,  $a_l$  为内积的模. 显然,  $a_l$  也是第  $l$  条正弦的幅度值. 由式(11)可见, 随着迭代次数增加, 残差能量呈递减趋势. 引入变量

$$c_l = \frac{\|r_{l+1}\|}{\|r_l\|}, \quad c_l \in (0, 1) \quad (12)$$

综合式(11)、(12), 有  $a_l^2 = \|r_l\|^2(1 - c_l^2)$ . 递归地应用式(12), 并初始化  $\|r_1\| = \|x\|$ , 得到

$$a_l = \|x\| c_1 \cdots c_{l-1} (1 - c_l^2)^{\frac{1}{2}} \quad (13)$$

式(13)显示  $a_l$  和  $a_{l-1}$  具有如下关系:

$$\frac{a_l}{a_{l-1}} = c_{l-1} \left( \frac{1 - c_l^2}{1 - c_{l-1}^2} \right)^{\frac{1}{2}} \approx c_{l-1} \quad (14)$$

在  $(l-1)$  次迭代, 如果已知  $c_{l-1}$ , 运用式(14)就能估计出第  $l$  条正弦幅度的平均值,  $\eta_{a_l} = a_{l-1} c_{l-1}$ . 参考文献[17]的方法, 第  $l$  条正弦的对数幅度的均值  $b_l$  可按式计算:

$$b_l = E(\log_2(a_l)) \approx \log_2(\eta_{a_l}) + \log_2''(\eta_{a_l}) \frac{\sigma_{a_l}^2}{2} \quad (15)$$

其中,  $\log_2''$  表示二阶微分算子,  $\sigma_{a_l}^2$  为  $a_l$  的方差, 由下式估计<sup>[17]</sup>:

$$\sigma_{a_l}^2 \approx \frac{1}{l-1} \sum_{j=1}^l (a_{l-j} - \eta_{a_{l-j}})^2 \quad (16)$$

设置  $c_0 = 1$ ,  $a_0$  为一帧中最大正弦幅度值. 将式(15)代入式(10), 即得自适应联合量化点密度:

$$g_{a_l} = w_l \left( \frac{2^{\frac{2H}{3}}}{\sigma^{\frac{2}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}} \quad (17a)$$

$$g_{\omega_l} = w_l a_l \sigma \left( \frac{2^{\frac{2H}{3}}}{\sigma^{\frac{-2}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}} \quad (17b)$$

$$g_{\varphi_l} = w_l a_l \left( \frac{2^{\frac{2H}{3}}}{\sigma^{\frac{2}{3}} \mu_g^{\frac{1}{3}}} \right)^{\frac{1}{2}} \quad (17c)$$

不难看出, 式(17)也为均匀标量量化器. 由于存在  $g_{\omega_l} > g_{\varphi_l} > g_{a_l}$  关系, 故分配给频率参数的比特最多, 相位次

之, 而幅度最少. 式中  $w_l = (\mu_l 2^{\frac{-4b}{3}})^{\frac{1}{2}}$ , 称感觉——参数权. 感觉——参数权同时考虑了人耳的听觉特性和参数的统计特性, 它伴随正弦抽取的进行而实时更新. 从式(15)还可以看出, 对于较小的  $a_l$ , 其对数的均值估计也较小, 若  $\mu_l$  不变, 观察  $w_l$  的表达式, 较小  $a_l$  的感觉——参数权将较大.

## 3 自适应联合量化的比特分配方案

将音频信号分成  $K$  帧, 帧结构表示如下:

$$\left\{ \begin{array}{cccccc} \text{NL} & \text{GMI} & \text{MI} \dots & \text{AI} \dots & \text{FI} \dots & \text{PI} \dots \\ \left\{ \begin{array}{cccccc} 6 & 2 & 4 & 5 & 11 & 6 \end{array} \right\}_k \end{array} \right., \quad k = 1, 2, \dots, K \quad (18)$$

在式(18)中,  $\{ \cdot \}$  中第一行表示量化参数名称, 第二行表示分配给各参数的比特数, 即平均比特数. 其中, NL: 正弦条数, 分配 7bit, 表示每帧信号最多抽取  $2^7 = 128$  条正弦. GMI: 感觉权的几何平均值, 2bit, 对数域均匀中升量化. MI: 感觉——参数权, 对数域 4bit 中升量化, 其后省略号表示连续有  $2^{\text{MI}}$  个 MI(下同); AI: 幅度量化索引, 平均 5bit 中升量化; FI: 频率量化索引, 平均 11bit 中升量化; PI: 相位量化索引, 平均 6bit 中平量化.

由式(17), 分配给第  $l$  条正弦的比特数

$$B_l = \frac{\left( \log_2 g_{a_l} + \int f_A(a_l) \log_2 g_{\omega_l} da_l + \int f_A(a_l) \log_2 g_{\varphi_l} da_l \right) \cdot T_l R_l}{\sum_{l=1}^{2^{\text{NL}}} (\log_2 g_{a_l} + \int f_A(a_l) \log_2 g_{\omega_l} da_l + \int f_A(a_l) \log_2 g_{\varphi_l} da_l)} \quad (19)$$

其中,  $T_f = \frac{1}{f_s}$  为帧周期,  $R_l$  为目标速率(已扣除非正弦参数所占用的部分). 显然, 严格按照上式来分配比特, 需要每正弦至少分配 5bit 来表达比特分配的边信息, 这对于低速率音频编码而言是一项不小的开销. 为避免额外开销, 本文采用以下方案.

首先计算分配给幅度参数的总比特数  $B_a$ :

$$B_a = \frac{AI \cdot T_f R_l}{AI + FI + BI} = \frac{5 \cdot T_f R_l}{5 + 11 + 6} \quad (20)$$

然后量化感觉——参数权与感觉权的几何平均值. 按式(17a)计算幅度参数的点密度, 进而计算分配给第  $l$  条正弦幅度的比特数  $B_{a_l}$ :

$$B_{a_l} = \frac{\log_2 g_{a_l}}{\sum_l (\log_2 g_{a_l})} \cdot B_a \quad (21)$$

量化各幅度参数, 最后将量化值代入式(17b)、(17c)计算各分量的频率和相位点密度, 从而得到分配给频率和相位参数的比特数:

$$B_{\omega_l} = \frac{\log_2 g_{\omega_l}}{\sum_l (\log_2 g_{\omega_l} + \log_2 g_{\varphi_l})} \cdot (T_f R_l - B_a) \quad (22)$$

$$B_{\varphi_l} = \frac{\log_2 g_{\varphi_l}}{\sum_l (\log_2 g_{\omega_l} + \log_2 g_{\varphi_l})} \cdot (T_f R_l - B_a) \quad (23)$$

## 4 自适应联合量化性能

本文以主观质量对比和客观的量化误差对比来评价自适应联合量化的性能. 实验使用选自 SQAM<sup>[18]</sup> 的单声道 44.1kHz 采样音频测试样段. 下表列出了样段的名称、类型和长度:

表 1 音频测试样段

名称	类型	长度(s)
响板与吉它	合奏	6
钟琴	独奏	8
ABBA	流行乐	10
木琴	独奏	9
低音吉它	独奏	12
大提琴	独奏	6
钢琴	独奏	10

实验确定每帧信号含 1024 个样点, 加汉宁窗, 且前、后帧叠接半帧. 为准确比较各方案的量化误差, 实验将模型误差(原信号与模型之差)加入到合成信号中, 以去掉其影响. 实验采用文献[19]算法作正弦抽取及参数估计.

实验一, 自适应联合量化(Adaptive Joint Quantization, AJQ)方案与传统量化(Conventional Quantization, CQ)方案<sup>[2]</sup>的主观对比听音测试. 年龄在 20~37 岁之间的 5 位男性和 4 位女性参与了测试(不包括本文作者). 在正

式测试之前每位参试者均要接受听音训练, 以熟悉测试样段, 增加测试结果的可靠性(训练时不告诉听者所听的是何种样段). 测试使用头戴式耳机, 在一间半消声室(本底噪声 < 18dB(A))完成.

实验同时提供给听者三种版本的样段: 各种目标速率的 AJQ 版本(速率范围从 16bit/正弦到 20bit/正弦; 感觉——参数权边信息分配 3bit, 以保证重建信号具有较好的感觉质量), CQ 版本(CQ 方案分配给幅度, 频率和相位参数的比特数分别为 6、11 和 6, 量化效果接近透明音质), 以及原始信号样段.

CQ 方案由以下各式给出.

正弦幅度量化器:

$$I_a = \max\left(\min\left(\left\lfloor \frac{20\log_{10}(a)}{\Delta_a} + 2^6 \right\rfloor, 2^6 - 1\right), 0\right) \quad (24)$$

式中,  $a$  为正弦幅度,  $\Delta_a$  表示幅度的量化步长, 其值为  $\frac{|a_{\text{dB, min}} - a_{\text{dB, max}}|}{2^6} = 1.58\text{dB/step}$ ,  $\lfloor \cdot \rfloor$  表示向下取整,  $\min(A, B)$ 、 $\max(A, B)$ , 分别表示  $A$ 、 $B$  中的较小者、较大者.

频率量化器:

$$I_f = \max(\min(\lfloor R \times 2^{11} \rfloor, 2^{11} - 1), 0) \quad (25)$$

式中,  $R = \frac{\text{ERB}_n(f) - \text{ERB}_n(f_{\min})}{\text{ERB}_n(f_{\max}) - \text{ERB}_n(f_{\min})}$ ,  $\text{ERB}_n(\cdot)$  表示频率/等效矩形带速率变换, 其计算公式为<sup>[12]</sup>:

$$\text{ERB}_n = 21.4\log_{10}(0.00437f + 1) \quad (26)$$

相位量化器:

$$I_p = \max\left(\min\left(\left\lfloor \left(\frac{\varphi}{2\pi} + 0.5\right) \times 2^6 \right\rfloor, 2^6 - 1\right), 0\right) \quad (27)$$

式中,  $\varphi$  表示相位参数.

测试要求听者对样段从 5(最好)到 1(最差)打分. 图 1 是所有听者对所有样段给出的平均意见分(MOS). 图 1 中, 误差条上的圆点表示中位数, 即第二四分位数, 顶端和底部的横线分别表示得分的上四分位和下四分位; AJQ20-AJQ16 分别表示 20bit/正弦到 16bit/正弦的 AJQ 量化方案. 从图中看到, AJQ 方案的性能转折点在 18bit 处, 当比特数高于 18, 量化器具有很好的主观质量, 与 CQ 方案相当, 而当比特数低于 18 后, 主观质量即开始迅速下降. 在相同的主观质量下, AJQ20-AJQ19 的性能好于 CQ 方案, 前者相对后者有  $4 = (6 + 11 + 6) - 19\text{bit}$  的改善.

实验二, 比较 AJQ 量化方案和球形量化方案(SQ)<sup>[11]</sup>的总体感觉失真. 实验定义总体感觉失真等于一帧信号的各量化正弦分量所产生的感觉失真的累加, 其值按式(2)计算. 在 SQ 方案中, 分配 2bit 用于量化感觉权边信息.

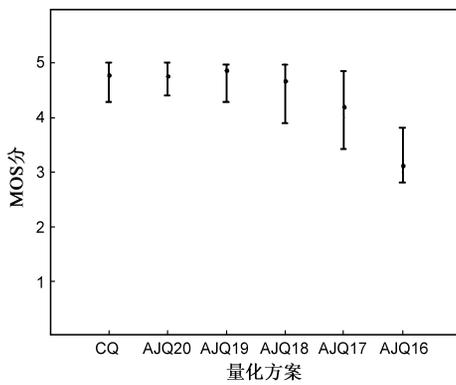


图1 所有听者和所有测试样段的总体平均主观听音测试结果

图2表示随机选择的一帧大提琴信号在6种速率下量化的率失真曲线.从图中看到,当速率大于15bit/正弦时,AJQ方案的感觉失真小于SQ方案的感觉失真,并且伴随比特速率逐渐变大,AJQ对失真的改善越明显;当速率小于15bit/正弦时,AJQ的失真较大.AJQ方案分配3bit编码感觉——参数权边信息,导致在小于15bit/正弦的速率下分配给正弦参数的比特数占总比特数的比例显著减少,是AJQ失真较大的原因.如果小于该速率,AJQ和SQ方案的感觉失真均>50dB,二者的感觉质量均不能被主观接受,所以在小于此速率下比较其量化性能已无实际意义.

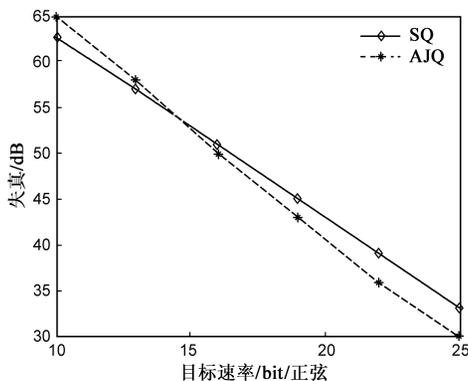


图2 SQ和AJQ量化方案的率失真曲线

表2、表3分别为SQ和AJQ方案在22bit、16bit每正弦速率下量化1帧信号的总体感觉失真对比.

表2 22bit/正弦速率下的总体感觉失真对比

	SQ	AJQ
响板与吉它	43.1dB	42.3dB
钟琴	41.4dB	39.7dB
ABBA	43.7dB	41.2dB
木琴	41.0dB	38.8dB
低音吉它	36.3dB	33.5dB
大提琴	39.2dB	36.0dB
钢琴	40.4dB	37.6dB

表3 16bit/正弦速率下的总体感觉失真对比

	SQ	AJQ
响板与吉它	59.7dB	59.1dB
钟琴	56.8dB	56.2dB
ABBA	58.7dB	58.2dB
木琴	52.3dB	50.8dB
低音吉它	48.4dB	47.5dB
大提琴	50.8dB	50.0dB
钢琴	51.7dB	50.2dB

从表2、表3看出,在两种速率下,所有测试信号的AJQ失真均小于SQ失真.

## 5 总结

本文提出一种正弦参数的自适应联合量化器方案.量化器的优化点密度函数由感觉权和实时参数的统计知识以及熵约束共同决定.按感知重要性和参数统计知识的不同,各正弦分量参数的点密度函数是可变的,以实现比特的优化分配.主观听音测试表明,本方案较传统量化方案有明显的性能改善,在同等的主观质量下,平均比特速率每正弦减少约  $17\% = \frac{4}{(6+11+6)}$ .当速率大于15bit/正弦,本量化方案的感觉失真小于球形量化方案的感觉失真.最近,文献[20, 21]实现了将联合量化方法应用到指数阻尼正弦模型中,和SQ方案一样,这些方法也使用信号参数的先验统计知识.受此启发,后续工作将对非稳态正弦模型(如调幅正弦模型、调频正弦模型)参数的自适应联合量化方法展开研究.

## 参考文献

- [1] Christensen M G, van de Par S. Efficient parametric coding of transients[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1340 - 1351.
- [2] Purnhagen H. Very low bit rate parametric audio coding[D]. Hannover, Germany: PhD dissertation, Hannover University, 2008.
- [3] Korten P E L, Jensen J, Heusdens R. High rate spherical quantization of sinusoidal parameters[A]. 12th European Signal Processing Conference[C]. Vienna, Austria: EURASIP, 2004. 1757 - 1760.
- [4] Vafin R, Prakash D, Kleijn W B. On frequency quantization in sinusoidal audio coding[J]. IEEE Signal Processing Letters, 2005, 12(3): 210 - 213.
- [5] Vafin R, Kleijn W B. Entropy-constrained polar quantization and its application to audio coding[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(2): 220 - 232.

- [6] Gray R M. Source Coding Theory[M]. Norwell, MA: Kluwer, 1990.
- [7] Gray R M, Neuhoff D L. Quantization[J]. IEEE Transactions on Information Theory, 1998, 44(6): 2325 – 2383.
- [8] Wilson G. Magnitude/phase quantization of independent Gaussian variates[J]. IEEE Transactions on Communications, 1980, 28(11): 1924 – 1929.
- [9] Swaszek P F, Ku T W. Asymptotic performance of unrestricted polar quantizer[J]. IEEE Transactions on Information Theory, 1986, 32(2): 330 – 333.
- [10] Korten P, Jensen J, Heusdens R. High-resolution spherical quantization of sinusoidal parameters[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(3): 966 – 981.
- [11] Korten P, Jensen J, Heusdens R. High resolution spherical quantization of sinusoidal parameters using a perceptual distortion measure [A]. International Conference on Acoustics, Speech, and Signal Processing[C]. Philadelphia, USA: IEEE, 2005. 177 – 180.
- [12] Moore B C J. An Introduction to the Psychology of Hearing (fourth edition)[M]. London: Academic Press, 1997.
- [13] van de Par S, Kohlrausch A, et al. A perceptual model for sinusoidal audio coding based on spectral integration [J]. EURASIP Journal on Applied Signal Processing, Special Issue on Anthropomorphic Signal Processing, 2005, (9): 1292 – 1304.
- [14] Komzsis L. Applied Calculus of Variations for Engineers [M]. Boca Raton: CRC Press, 2009.
- [15] Christensen R A, van de Par S. Rate-distortion efficient amplitude modulated sinusoidal audio coding [A]. 38th Asilomar Conference on Signals, Systems and Computers[C]. USA: IEEE, 2004. 2280 – 2284.
- [16] Goodwin M. Adaptive signal models: theory, algorithms, and audio applications[D]. Berkeley, USA: PhD dissertation of University of California, Berkeley, 1997.
- [17] Papoulis A, Pillai S U. Probability, Random Variables and Stochastic Processes[M]. New York: McGraw-Hill, 2002.
- [18] European Broadcasting Union. Sound Quality Assessment Material Recordings for Subjective Tests [DB]. Geneva: EBU, 1988.
- [19] 王嵩, 鲍长春. 基于组合正弦的音频信号建模[J]. 信号处理, 2009, 25(8A): 135 – 138.  
Wang S, Bao C C. An audio signal modeling based on composite sinusoids [J]. Signal Processing, 2009, 25(8A): 135 – 138. (in Chinese)
- [20] Derrien O, Badeau R, Richard G. Entropy-constrained quantization of exponentially damped sinusoids parameters [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. Prague, Czech Republic: IEEE, 2011. 4064 – 4067.
- [21] Derrien O, Badeau R, Richard G. Parametric audio coding with exponentially damped sinusoids [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(7): 1489 – 1501.

#### 作者简介



王 嵩 男, 1972 年 10 月出生, 四川南部县人, 获北京工业大学博士学位, 北京石油化工有限公司讲师, 研究方向为信号的参数建模方法、低比特率音频编码。

E-mail: s. wang@bupt.edu.cn