GY

中华人民共和国广播电视和网络视听行业标准

GY/T 349—2021

感知音频质量的客观测量方法

Method for objective measurements of perceived audio quality

（ITU-R BS.1387-1,MOD）

2021-03-29 发布　　2021-03-29 实施

国家广播电视总局　　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件使用重新起草法修改采用 ITU-R BS.1387-1《感知音频质量的客观测量方法》。

本文件与 ITU-R BS.1387-1 相比，在结构上有较多的调整，附录 A 中列出了本文件与 ITU-R BS.1387-1 章条编号变化对照一览表。

本文件与 ITU-R BS.1387-1 的技术性差异及其原因如下：

——为符合 GB/T 1.1—2020 的要求，增加了第 1 章“范围”、第 2 章“规范性引用文件”、第 3 章“术语、定义和缩略语”。

本文件对以下内容进行了编辑性修改：

——删除了附件 1“概述”中过去相关研究情况的叙述内容；

——删除了附件 1 主观领域中的对主观评价的叙述内容；

——删除了附件 2 第 7 章中的关于测试条目从数据库 3 中选择的描述性内容；

——删除了附件 1 的附录 3 中关于 PAQM 的部分论述性语句；

——删除了附件 1 的附录 1 中的版权部分的描述；

——删除了附件 1 的附录 4 中的介绍与历史部分的描述；

——删除了附件 2 的附录 1 验证过程；

——删除了附件 2 的附录 2 参考数据库描述。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国广播电影电视标准化技术委员会（SAC/TC 239）归口。

本文件起草单位：国家广播电视总局广播电视规划院。

本文件主要起草人：覃毅力、邓向冬、韦安明、董文辉、郑涛、汪芮、王倩男。

引　言

考虑到：

a) 对采用低比特率编码算法，以及采用模拟或数字信号处理的系统，传统的客观测量方法（如信噪比和失真的测量）不适用于感知音频质量的测量；
b) 低比特率编码算法已得到迅速应用；
c) 并非所有符合某种规范或标准的系统/设备都可以保证达到规范或标准所规定的最高质量；
d) 通常的主观评价方法不适用于音频质量的连续监测，例如在系统运行的情况下；
e) 在整个测量领域中，感知音频质量的客观测量方法将补充或替代传统的客观测量方法；
f) 感知音频质量的客观测量方法可以有效地对主观评价方法进行补充；
g) 对一些应用，需要可实时测量的方法。

建议对于本文件所列的应用，使用本文件规定的方法进行感知音频质量的客观测量。

感知音频质量的客观测量方法

1 范围

本文件规定了感知音频质量的客观测量方法。

本文件适用于在电视节目或广播节目的收录、分配、传送和监测等环节，也适用于编解码器等音频处理设备的研究、开发、测试和维护。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GY/T 298—2016 音频系统小损伤主观评价方法（ITU-R BS.1116-3,MOD）

ITU-R BS.1284-1：2003 声音质量主观评价通用方法（General methods for the subjective assessment of sound quality）

3 术语、定义和缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

绝对误差值 absolute error score；AES

反映SDG置信区间大小与SDG和ODG之差的关联平均值，其计算公式见公式（1）。

$$AES = 2\sqrt{\frac{\sum((ODG - SDG)/CI)^2}{N}} \quad \cdots\cdots（1）$$

式中：

CI——置信区间大小，若CI＜0.25，则CI＝0.25；

N ——被评价音频素材的数量。

3.1.2

基本音频质量 basic audio quality

一个通用主观属性，该属性包含了任意及所有可检测到的参考信号及其处理版本之间的差异。

3.1.3

编码余量 coding margin

一个质量参数，表示编码损伤从不可感知到可感知的阈值余量。

3.1.4

模型输出变量 model output variables；MOV

感知测量方法的中间输出值。这些变量以基本心理声学研究为基础，用于进一步描述编码损伤特性。

3.1.5

主观差异等级　subjective difference grade；SDG

根据GY/T 298—2016开展的音频主观评价，采用5级损伤标度对隐藏参考和被测信号的基本音频质量进行打分得到相应的评分等级，由被测信号评分等级减去隐藏参考信号评分等级所得的差值，见公式（2）。[1]

$$SDG = G_{UDT} - G_{Ref} \quad \cdots\cdots (2)$$

式中：

G_{UDT}——被测信号评分等级；

G_{Ref}——隐藏参考信号评分等级。

3.1.6

客观差异等级　objective difference grade；ODG

感知测量方法的主要输出参数，相应于主观差异等级，为通用基本音频质量的测量参数。[2]

3.1.7

离线测量　off-line measurement

一种测量程序，其测量过程不会影响正在进行节目传输的系统。

3.1.8

在线测量　on-line measurement

一种测量程序，测试过程需依赖于正在进行传输的系统或是节目传输的一部分。

3.2　缩略语

下列缩略语适用于本文件。

ADB　平均失真块（Average Distorted Block）

ASD　听觉频谱差异（Auditory Spectral Difference）

BAQ　基本音频质量（Basic Audio Quality）

CI　置信区间（Confidence Interval）

DC　直流（Direct Current）

DFT　离散傅里叶变换（Discrete Fourier Transform）

DIX　干扰指数（Disturbance Index）

EHS　谐波失真结构（Error Harmonic Structure）

ERB　等效矩形带宽（Equivalent Rectangular Bandwidth）

FFT　快速傅里叶变换（Fast Fourier Transform）

FIR　有限脉冲响应（Finite Impulse Response）

IIR　无限脉冲响应（Infinite Impulse Response）

ITU　国际电信联盟（International Telecommunication Union）

ISO　国际标准化组织（International Standards Organization）

JNLD　临界可察觉电平差（Just Noticeable Level Difference）

MFPD　最大过滤检测概率（Maximum Filtered Probability of Detection）

NL　噪音响度（Noise Loudness）

NMR　噪声掩蔽比（Noise-To-Mask Ratio）

PAQM　感知音频质量测量（Perceptual Audio Quality Measure）

PERCEVAL　感知评价（Perceptual Evaluation）

1）　理想情况下，SDG数值范围为0～-4。如果参考信号没有被正确识别，则数值为正数。
2）　ODG数值范围为0～-4。

POM 感知客观测量（Perceptual Objective Measure）
Ref 参考信号（Reference Signal）
ROEX ROEX函数（Rounded Exponential）
ROV 输出值比率（Rate of Output Values）
SCM 主观编码余量（Subjective Coding Margin）
SPL 声压级（Sound Pressure Level）
Win 窗口平均值（Windowed Average）

4 概述

在数字广播电视系统中，音频质量是一个非常关键的因素。判定音频质量的主要方法包括音频质量主观评价和客观测量。由于音频主观评价既费时又昂贵，而传统音频客观指标如信噪比或总谐波失真与感知音频质量没有可靠的关联性，因此需提出一种客观测量方法用于音频质量测量。

本文件所规定的感知音频质量客观测量方法是在对已有测量方法如干扰指数（DIX）、噪声掩蔽比（NMR）、感知音频质量测量（PAQM）、感知评价（PERCEVAL）、感知客观测量（POM）以及工具箱法（Toolbox Approach）进行研究的基础上形成的，输出可靠有用的信息，用于多种应用场景。通过对上述六种方法的性能进行研究，提取其中最有用的工具，并将这些工具融合形成一个新的测量方法，即本标准规定的测量方法。本文件规定的测量方法已经在许多测试场所经过了仔细验证，且已证明能够为许多应用生成既可靠又有用的信息。不过本文件中的客观测量方法无法取代正式听音测试。

附录B规定了客观感知音频质量的测量方法的原则和特点。

5 应用

感知音频质量客观测量的基本示意图见图1。

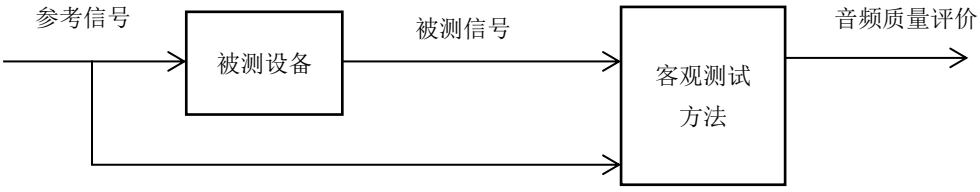


图1 客观测量的基本示意图

本文件规定的测量方法适用于大部分模拟或数字音频信号处理设备，可着重用于音频编解码方面的应用。

该测量方法适用于实时在线测量的应用场景，也适用于非实时离线测量的应用场景。在实时在线测量时，被测设备适宜的最大延时宜小于等于200ms，最大不应大于1s。

本文件规定的测量方法可用于以下八类应用场景，应与表1相符合。

表1 应用范围

序号	应用名称	简介	版本
1	系统/设备的评价	对音频处理设备（多数情况指编解码器）的不同实现方案进行评价	基础/高级
2	感知质量的排序	针对某个设备或线路在投入运行前的快速测量过程	基础

表1（续）

序号	应用名称	简介	版本
3	在线监测	对工作中的音频传输进行连续监测	基础
4	设备或连接状态	对某个设备或某个线路进行详细分析	高级
5	编解码器识别	识别特定编解码器的类型或实现方案	高级
6	编解码器开发	对编解码器性能特性进行尽可能地分析	基础/高级
7	网络规划	对特定条件下的传输网络在性能和成本方面进行优化	基础/高级
8	主观评价辅助	作为筛选听音测试中关键素材的工具	基础/高级

八类应用场景详细的说明见附录C。

6 版本

考虑到不同的经济成本和性能要求，本文件规定的客观测量方法提供了两个版本。基础版本适用于低成本实时实现方案，高级版本侧重于最高的准确度。由于高级版本增加了额外准确度，它的复杂度比基础版本增加了约四倍。

每种应用所适用的版本应符合附录C的要求。

7 主观领域

主观评价与客观测量之间需要相互补充，示意图见图2。通常的音频主观评价，例如基于GY/T 298—2016的评价，是经过精心设计的，用以得出尽可能准确表征音频质量的可靠评价结果。不过主观评价的结果也不一定能完全反映出真实的感觉。客观测量方法可通过音频质量主观评价进行验证。

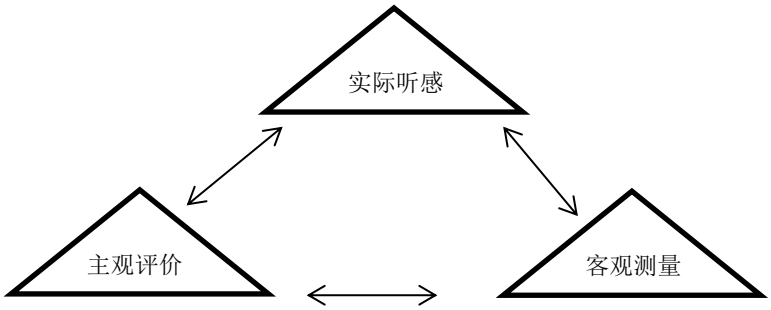


图2 验证示意图

本文件中的测量方法主要关注那些在主观领域中可采用GY/T 298—2016进行评价的应用。GY/T 298—2016中测量方法的基本原则可以简要描述为：听音者在A、B、C三个音源中切换并评价，其中音源A为已知的参考信号，音源B和C为隐藏的参考信号和被测信号的随机排列。

按照连续5级损伤等级，听音者通过对比B与A，C与A，对B和C的损伤进行评价。B和C中的其中一个为隐藏源，难以将其与A区分开，另一个则可能会反映出一些损伤。参考源和另一个音源之间的任何感知上的差异均应视为损伤。通常来说，只考虑“基本音频质量”这一属性，它是一个总体属性，涵盖了参考信号与被测信号之间可感知到的所有差异。

损伤等级标度采用ITU-R BS.1284-1：2003中给定的连续且带锚点的ITU-R 5级损伤等级标度，应与图3相符合。

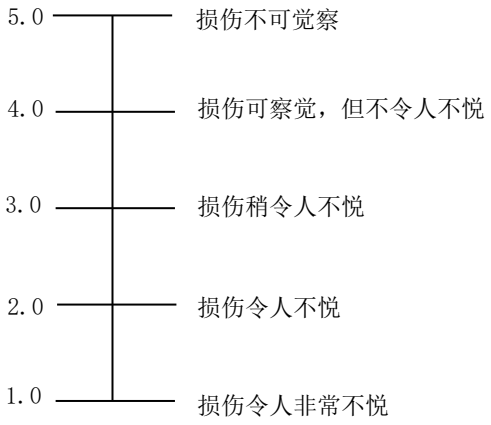


图3　ITU 五级损伤标度

主观评价结果的分析通常以主观差异等级（SDG）为基础。

SDG值的理想范围应是0～-4。0表示损伤不可察觉，-4表示损伤令人非常不悦。

8　分辨率和精准度

客观差异等级（ODG）是客观测量方法的输出变量，相当于主观领域中的SDG。ODG的精度精确到小数点后一位。当任意两个ODG之差超过10%时则表明差异显著，在测试过程需要注意，避免出现这类情况。

鉴于缺少独立的参数对客观测量方法的准确度进行完整描述，因此在验证过程中需要考察多个参数。性能参数一是SDG与ODG之间的关联性。客观测量方法的性能可能随着引入损伤的类型和程度等参数变化而变化。性能参数二是异常值的数量。异常值是指测量出来不符合预定容差的值。根据用户要求，评分等级表靠上部分即高质量音频，测量方法的准确性应最高，评分等级表中下部分即中等及较差质量音频，测量的准确度可以稍降低。关联性可较好地评价客观测量方法的准确性，但还需考察异常值；从异常值的角度来看，即便测量方法具有相当高的关联性，测量方法仍然有可能隐藏无法接受的特性。性能参数三是绝对误差值，它反映了SDG置信区间的大小与SDG和ODG之差的关联平均值。

9　要求及限制

整个测量期间，应将被测设备的信号和参考信号的时间准确度校正到24个采样值内。本文件中不涉及同步机制，不同测量方法的实现方案可有不同的同步机制。

10　模型的描述

10.1　概述

10.1.1　客观测量方法概述

本文件规定的感知音频质量的客观测量方法包括一个耳朵周边模型、多个中间处理环节（即激励模式的预处理）、基于心理声学的MOV计算方法和将MOV集合映射到代表被测信号基本质量的映射算法，应

与图4相符合。耳朵周边模型有两种模型，一种以FFT为基础（简称FFT耳朵模型），一种以滤波器组为基础（简称滤波器组耳朵模型）。除了计算误差信号时有所不同（仅使用FFT耳朵模型部分），其他情况下，两种耳朵周边模型的总体结构一样。

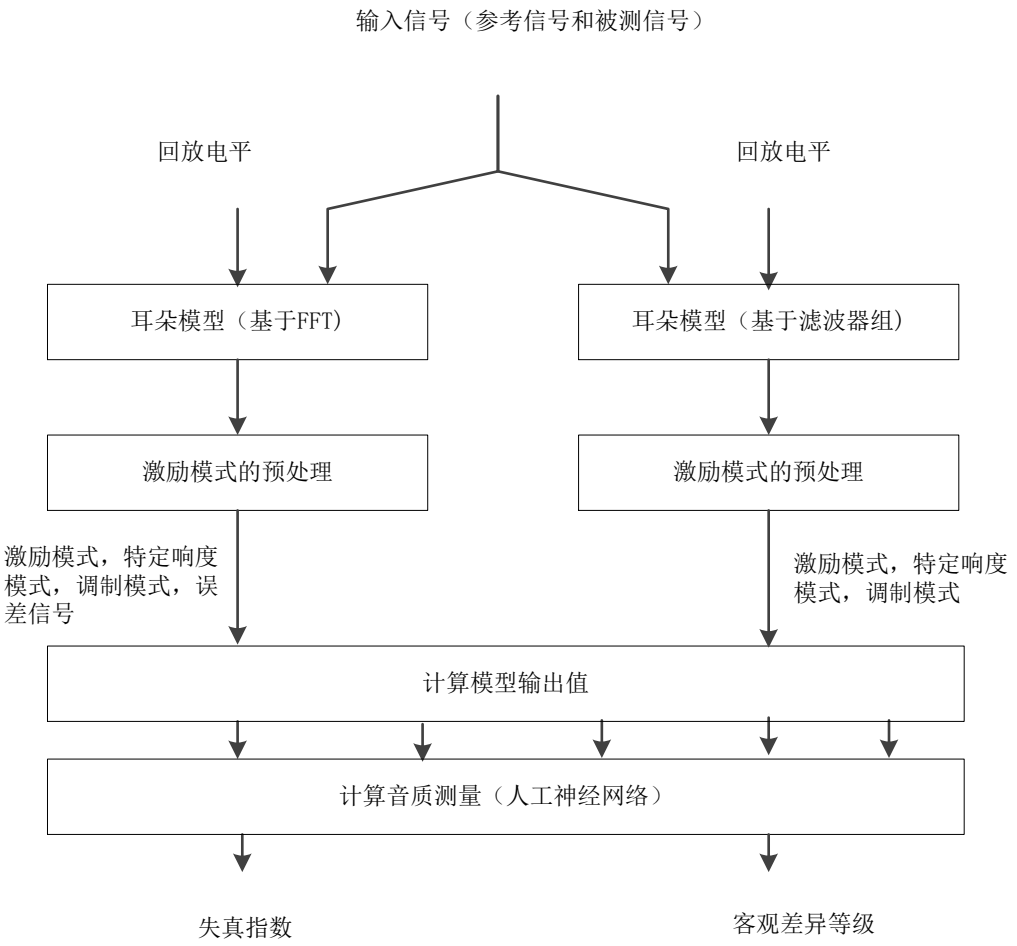


图4　测试方案的常用模块结构图

用于计算模型输出变量值（MOV）的输入包括：

——用于测试和参考信号的激励模式；

——用于测试和参考信号的频谱自适应的激励模式；

——用于测试和参考信号的指定响度模式；

——用于测试和参考信号的调制模式。

误差信号，即测试信号和参考信号间差异的频谱（仅适用于FFT耳朵模型）。

如果没有其他说明，立体声信号左右声道的所有计算都独立执行，左右声道采用的方式一样。

本文件给出了两种实现方式，即基础版本和高级版本。

在所有给出的公式中，“Ref”表示所有根据参考信号计算得到的模式，“Test”表示所有根据被测信号计算得到的模式，“k”表示离散频率变量（如频率频带），“n”表示离散时间变量（如帧计数器或样本计数器）。如果k和n的值没有明确定义，计算时就会计算所有可能的k和n值。其他缩写在其出现的地方会有说明。

在MOV中，后缀“A”表示滤波器组耳朵模型计算出来的变量，“B”表示FFT耳朵模型计算出来的变量。各个MOV应符合附录D的要求，模型补充说明应符合附录E的要求。

10.1.2 基础版本

基础版本只包含FFT耳朵模型计算得到的MOV值，不包括滤波器组耳朵模型计算得到的MOV值。基础版本采用11个MOV值，预测感知音频基本质量。

10.1.3 高级版本

高级版本包含以滤波器组耳朵模型计算得到的MOV值以及FFT耳朵模型计算得到的MOV值。频谱适应激励模式和调制模式仅用于以滤波器组为基础的模型计算。高级版本采用5个MOV值来预测感知音频基本质量。

10.2 耳朵周边模型

10.2.1 FFT 耳朵模型

10.2.1.1 FFT 耳朵模型概述

耳朵周边模型和模型中基于 FFT 处理的激励模式的预处理应符合图 5 的要求。

FFT 耳朵模型的输入为 48kHz 采样、时间对齐的参考和测试信号，输入信号被分割成长度为 0.042s 的帧，帧间重叠率为 50%。使用 Hann 窗口和短期 FFT，将每个帧转换到频域，并对输入信号进行定标，调整到回放电平。为模仿外耳和中耳的频率响应，需对频谱系数进行加权。通过将加权频谱系数组合对应到临界频带，实现了信号到音高标度的转换。通过增加频率偏移，模拟听觉系统中的内部噪声。采用电平扩展函数，模拟频率域中的频谱听觉滤波器。时域分布则负责前向掩蔽效应。

所得的激励模式用于计算指定响度模式和掩蔽模式。最终的时域分布之前的模式(“未抹除的调制模式”)用于计算调制模式。

为模仿误差信号，外耳和中耳滤波器输出的参考信号和测试信号模式将被组合起来，并分组对应到临界频带，从而映射至音高标度。

这些输出与激励信号均用于计算MOV值。

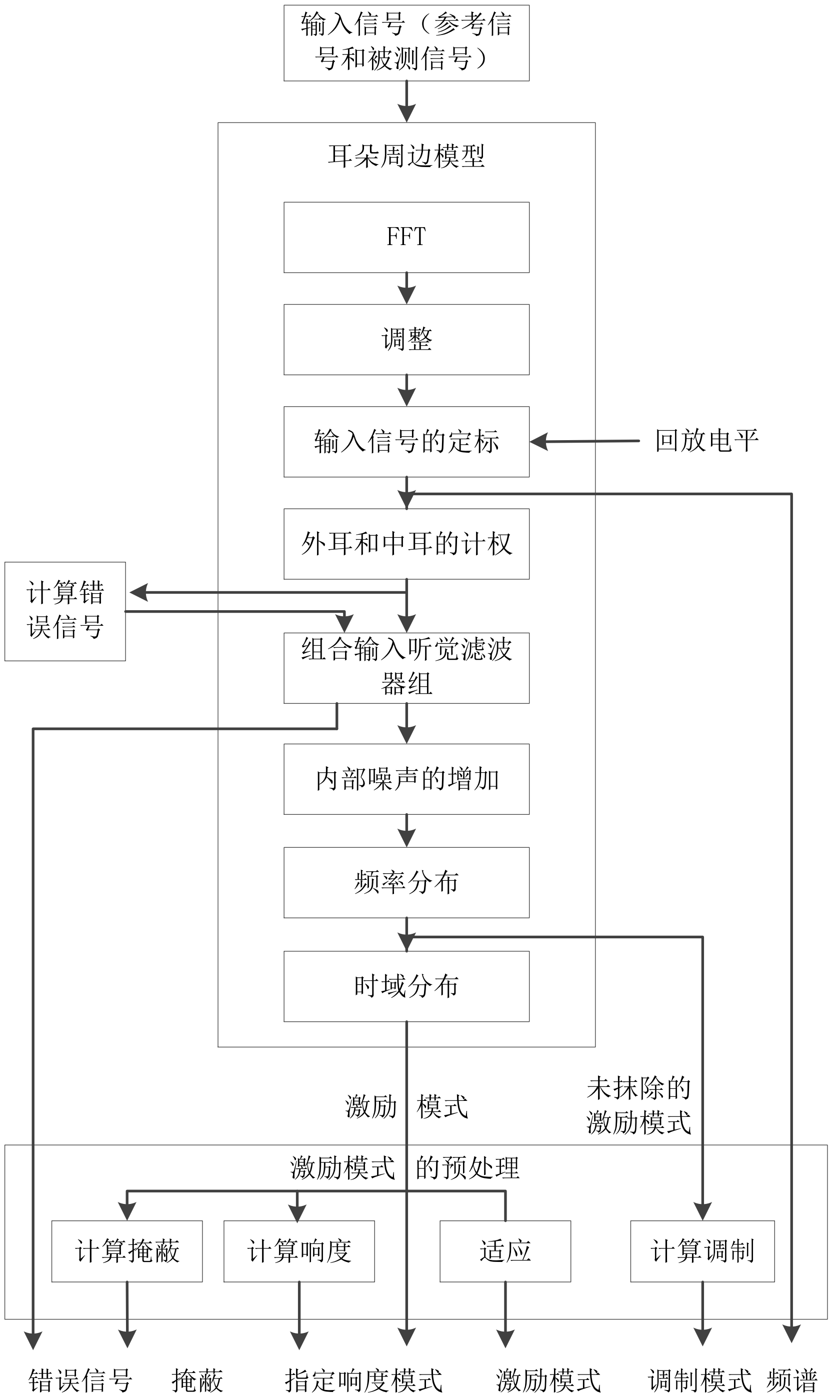


图5　耳朵周边模型和模型中基于FFT处理的激励模式的预处理

10.2.1.2　时间处理

FFT耳朵模型的输入，测试和参考信号划分成具有2048个取样点的帧，相邻的帧与帧之间具有1024个取样点的重叠，见公式（3）。

$$t_n[k_t,n] = t[1024\ n + k_t] \quad \cdots\cdots (3)$$

式中：

n——时间帧数量，取值为0,1,2⋯；

k_t ——帧内的时间计数器，取值为0...2047。

10.2.1.3 FFT

通过使用Hann窗口实现从时间域到频率域的映射，见公式（4）和公式（5）。

$$h_w[k] = \frac{1}{2}\sqrt{\frac{8}{3}}\left[1 - cos\left(2\pi\frac{k}{N-1}\right)\right] \quad \middle| \quad N = 2048 \qquad \cdots\cdots (4)$$

$$t_w[k_t,n] = h_w[k_t] \times t_n[k_t,n] \qquad \cdots\cdots (5)$$

然后，采用短期傅里叶变换，见公式（6）。

$$F_f[k_f,n] = \frac{1}{2048}\sum_{k_t=0}^{2047} t_w[k_t,n]\, e^{-j\frac{2\pi}{2048}k_f k_t} \qquad \cdots\cdots (6)$$

FFT的比例因子可根据一个满刻度正弦波的设定声压级Lp计算得到，见公式（7）和公式（8）。

$$fac = \frac{10^{\frac{Lp}{20}}}{Norm} \qquad \cdots\cdots (7)$$

$$F[k_f,n] = fac \times F_f[k_f,n] \qquad \cdots\cdots (8)$$

其中，归一化因数Norm的计算过程为：把一个1019.5Hz、0dB的满刻度正弦波作为输入信号，计算10帧以上的频谱系数最大绝对值。

如果声压级未知，建议Lp设置为92dB $_{SPL}$。

10.2.1.4 外耳和中耳

外耳和中耳的频率响应可由一个频率加权函数进行表示，见公式（9）。

$$W[k] = -0.6 \times 3.64 \times \left(\frac{f[k]}{1000}\right)^{-0.8} + 6.5 \times e^{-0.6\cdot\left(\frac{f[k]}{1000}-3.3\right)^2} - 10^{-3} \times \left(\frac{f[k]}{1000}\right)^{3.6} \qquad \cdots\cdots (9)$$

其中：

$$f[k] = 23.4375k \qquad \cdots\cdots (10)$$

表示在k行的频率表现。

FFT输出见公式（11）。

$$F_e[k_f,n] = \left|F[k_f,n]\right| \times 10^{\frac{W[k_f]}{20}} \qquad \cdots\cdots (11)$$

Fe[kf]为“外耳加权FFT输出”。

10.2.1.5 分组到临界频带

听觉音高标度可通过Schroeder等人提出的近似法进行计算，见公式（12）。

$$\mathrm{z} = 7\mathrm{arsinh}\left(\frac{f}{650}\right) \qquad \cdots\cdots (12)$$

音高z的单位是巴克（Bark）。

滤波器的频率边界范围为80Hz～18000Hz。对于基础版本而言，滤波器频带的宽度和间距对应的分辨率为0.25Bark，对于高级版本而言，对应的分辨率为0.5Bark。

可推断出基础版本的频带数量为109，应与表2相符合；高级版本的频带数量为55，应与表3相符合。

表2　用于基础版本的 FFT 耳朵模型的频带

组别（k）	低频（$f_l[k]$） Hz	中心频率（$f_C[k]$） Hz	高频（$f_u[k]$） Hz	频率带宽（$f_W[k]$） Hz
0	80	91.708	103.445	23.445
1	103.445	115.216	127.023	23.577
2	127.023	138.87	150.762	23.739
3	150.762	162.702	174.694	23.932
4	174.694	186.742	198.849	24.155
5	198.849	211.019	223.257	24.408
6	223.257	235.566	247.95	24.693
7	247.95	260.413	272.959	25.009
8	272.959	285.593	298.317	25.358
9	298.317	311.136	324.055	25.738
10	324.055	337.077	350.207	26.151
11	350.207	363.448	376.805	26.598
12	376.805	390.282	403.884	27.079
13	403.884	417.614	431.478	27.594
14	431.478	445.479	459.622	28.145
15	459.622	473.912	488.353	28.731
16	488.353	502.95	517.707	29.354
17	517.707	532.629	547.721	30.014
18	547.721	562.988	578.434	30.713
19	578.434	594.065	609.885	31.451
20	609.885	625.899	642.114	32.229
21	642.114	658.533	675.161	33.048
22	675.161	692.006	709.071	33.909
23	709.071	726.362	743.884	34.814
24	743.884	761.644	779.647	35.763
25	779.647	797.898	816.404	36.757
26	816.404	835.17	854.203	37.799
27	854.203	873.508	893.091	38.888
28	893.091	912.959	933.119	40.028
29	933.119	953.576	974.336	41.218
30	974.336	995.408	1016.797	42.461
31	1016.797	1038.511	1060.555	43.758
32	1060.555	1082.938	1105.666	45.111
33	1105.666	1128.746	1152.187	46.521
34	1152.187	1175.995	1200.178	47.991

表 2（续）

组别（k）	低频（f_l[k]） Hz	中心频率（f_C[k]） Hz	高频（f_U[k]） Hz	频率带宽（f_W[k]） Hz
35	1200.178	1224.744	1249.7	49.522
36	1249.7	1275.055	1300.816	51.116
37	1300.816	1326.992	1353.592	52.776
38	1353.592	1380.623	1408.094	54.502
39	1408.094	1436.014	1464.392	56.298
40	1464.392	1493.237	1522.559	58.167
41	1522.559	1552.366	1582.668	60.109
42	1582.668	1613.474	1644.795	62.128
43	1644.795	1676.641	1709.021	64.226
44	1709.021	1741.946	1775.427	66.406
45	1775.427	1809.474	1844.098	68.671
46	1844.098	1879.31	1915.121	71.023
47	1915.121	1951.543	1988.587	73.466
48	1988.587	2026.266	2064.59	76.003
49	2064.59	2103.573	2143.227	78.637
50	2143.227	2183.564	2224.597	81.371
51	2224.597	2266.34	2308.806	84.208
52	2308.806	2352.008	2395.959	87.154
53	2395.959	2440.675	2486.169	90.21
54	2486.169	2532.456	2579.551	93.382
55	2579.551	2627.468	2676.223	96.672
56	2676.223	2725.832	2776.309	100.086
57	2776.309	2827.672	2879.937	103.627
58	2879.937	2933.12	2987.238	107.302
59	2987.238	3042.309	3098.35	111.112
60	3098.35	3155.379	3213.415	115.065
61	3213.415	3272.475	3332.579	119.164
62	3332.579	3393.745	3455.993	123.415
63	3455.993	3519.344	3583.817	127.823
64	3583.817	3649.432	3716.212	132.395
65	3716.212	3784.176	3853.348	137.136
66	3853.348	3923.748	3995.399	142.051
67	3995.399	4068.324	4142.547	147.148
68	4142.547	4218.09	4294.979	152.432
69	4294.979	4373.237	4452.89	157.911

表 2（续）

组别（k）	低频（f_l[k]） Hz	中心频率（f_C[k]） Hz	高频（f_u[k]） Hz	频率带宽（f_W[k]） Hz
70	4452.89	4533.963	4616.482	163.592
71	4616.482	4700.473	4785.962	169.48
72	4785.962	4872.978	4961.548	175.585
73	4961.548	5051.7	5143.463	181.915
74	5143.463	5236.866	5331.939	188.476
75	5331.939	5428.712	5527.217	195.278
76	5527.217	5627.484	5729.545	202.329
77	5729.545	5833.434	5939.183	209.637
78	5939.183	6046.825	6156.396	217.214
79	6156.396	6267.931	6381.463	225.067
80	6381.463	6497.031	6614.671	233.208
81	6614.671	6734.42	6856.316	241.646
82	6856.316	6980.399	7106.708	250.392
83	7106.708	7235.284	7366.166	259.458
84	7366.166	7499.397	7635.02	268.854
85	7635.02	7773.077	7913.614	278.594
86	7913.614	8056.673	8202.302	288.688
87	8202.302	8350.547	8501.454	299.152
88	8501.454	8655.072	8811.45	309.996
89	8811.45	8970.639	9132.688	321.237
90	9132.688	9297.648	9465.574	332.887
91	9465.574	9636.52	9810.536	344.962
92	9810.536	9987.683	10168.013	357.477
93	10168.013	10351.586	10538.46	370.447
94	10538.46	10728.695	10922.351	383.891
95	10922.351	11119.49	11320.175	397.824
96	11320.175	11524.47	11732.438	412.264
97	11732.438	11944.149	12159.67	427.231
98	12159.67	12379.066	12602.412	442.742
99	12602.412	12829.775	13061.229	458.817
100	13061.229	13296.85	13536.71	475.48
101	13536.71	13780.887	14029.458	492.748
102	14029.458	14282.503	14540.103	510.645
103	14540.103	14802.338	15069.295	529.192
104	15069.295	15341.057	15617.71	548.415

表 2（续）

组别（k）	低频（f_l[k]） Hz	中心频率（f_C[k]） Hz	高频（f_u[k]） Hz	频率带宽（f_W[k]） Hz
105	15617.71	15899.345	16186.049	568.339
106	16186.049	16477.914	16775.035	588.986
107	16775.035	17077.504	17385.42	610.385
108	17385.42	17690.045	18000	614.58

表3　用于高级版本的 FTT 耳朵模型的频带

组别（k）	低频（f_l[k]） Hz	中心频率（f_C[k]） Hz	高频（f_u[k]） Hz	频率带宽（f_W[k]） Hz
0	80	103.445	127.023	47.023
1	127.023	150.762	174.694	47.671
2	174.694	198.849	223.257	48.563
3	223.257	247.95	272.959	49.702
4	272.959	298.317	324.055	51.096
5	324.055	350.207	376.805	52.75
6	376.805	403.884	431.478	54.673
7	431.478	459.622	488.353	56.875
8	488.353	517.707	547.721	59.368
9	547.721	578.434	609.885	62.164
10	609.885	642.114	675.161	65.277
11	675.161	709.071	743.884	68.723
12	743.884	779.647	816.404	72.52
13	816.404	854.203	893.091	76.687
14	893.091	933.119	974.336	81.245
15	974.336	1016.797	1060.555	86.219
16	1060.555	1105.666	1152.187	91.632
17	1152.187	1200.178	1249.7	97.513
18	1249.7	1300.816	1353.592	103.892
19	1353.592	1408.094	1464.392	110.801
20	1464.392	1522.559	1582.668	118.275
21	1582.668	1644.795	1709.021	126.354
22	1709.021	1775.427	1844.098	135.077
23	1844.098	1915.121	1988.587	144.489
24	1988.587	2064.59	2143.227	154.64
25	2143.227	2224.597	2308.806	165.579
26	2308.806	2395.959	2486.169	177.364
27	2486.169	2579.551	2676.223	190.054
28	2676.223	2776.309	2879.937	203.713

表 3（续）

组别（k）	低频（f_l[k]） Hz	中心频率（f_c[k]） Hz	高频（f_u[k]） Hz	频率带宽（f_w[k]） Hz
29	2879.937	2987.238	3098.35	218.414
30	3098.35	3213.415	3332.579	234.229
31	3332.579	3455.993	3583.817	251.238
32	3583.817	3716.212	3853.348	269.531
33	3853.348	3995.399	4142.547	289.199
34	4142.547	4294.979	4452.89	310.343
35	4452.89	4616.482	4785.962	333.072
36	4785.962	4961.548	5143.463	357.5
37	5143.463	5331.939	5527.217	383.754
38	5527.217	5729.545	5939.183	411.966
39	5939.183	6156.396	6381.463	442.281
40	6381.463	6614.671	6856.316	474.853
41	6856.316	7106.708	7366.166	509.85
42	7366.166	7635.02	7913.614	547.448
43	7913.614	8202.302	8501.454	587.84
44	8501.454	8811.45	9132.688	631.233
45	9132.688	9465.574	9810.536	677.849
46	9810.536	10168.013	10538.46	727.924
47	10538.46	10922.351	11320.175	781.715
48	11320.175	11732.438	12159.67	839.495
49	12159.67	12602.412	13061.229	901.56
50	13061.229	13536.71	14029.458	968.229
51	14029.458	14540.103	15069.295	1039.837
52	15069.295	15617.71	16186.049	1116.754
53	16186.049	16775.035	17385.42	1199.371
54	17385.42	17690.045	18000	614.58

频率到音高的映射可采用后续介绍的算法完成，其中$Fsp[k_f]$，可以是“外耳加权FFT输出”的能量表示，见公式（13）。

$$Fsp[k_f,n] = \left|F_e[k_f,n]\right|^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots（13）$$

$Fsp[k_f]$也可以是误差信号的能量表示，见公式（14）。

$$Fsp[k_f,n] = \left|F_{noise}[k_f,n]\right|^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots（14）$$

误差信号的计算见10.3.5。

该处理阶段的输出为频率组的能量P_e[k,n]。

伪码：

/* inputs */

Fsp[ ]：输入能量
/* outputs */
Pe[ ]：映射到音高的能量
/* intermediate values */
I ：频率组索引
k ：\fft行的索引
Z ：频率组的数目
    基础版本是109
    高级版本是55
fl[] ：频率组的低频
fu[] ：频率组的高频
Fres ：频率分辨率的常数

```
  Fres = 48000/2048;
  for(i=0; i<Z; i++ )
  {
   Pe[i]=0;
   for(k=0;k<1024;k++)
   {
     /* line inside frequency group */
     if( (( k-0.5)*Fres >= fl[i])  && ((k+0.5)*Fres <= fu[i]))
     {
       Pe[i] += Fsp[k];
     }
     /* frequency group inside*/
     else if( (( k-0.5)*Fres < fl[i])  && ((k+0.5)*Fres > fu[i]))
     {
       Pe[i] += Fsp[k]*(fu[i]-fl[i])/Fres;
     }
     /* left border */
     else if( ((k-0.5)*Fres < fl[i]) && ((k+0.5)*Fres > fl[i]))
     {
      Pe[i] += Fsp[k]*( (k+0.5)*Fres - fl[i])/Fres;
     }
     /* right border
     else if( ((k-0.5)*Fres < fu[i]) && ((k+0.5)*Fres > fu[i]);
     {
      Pe[i] += Fsp[k]*(fu[i]- (k-0.5)*Fres)/Fres;
     }
     /* line outside frequency group */
     else
     {
       Pe[i] += 0;
```

```
        }
      }

       /* limit result */
        Pe[i]=max(Pe[i],0.000000000001);
    }
```

10.2.1.6 添加内部噪声

对每个频率组的能量增加一个频率偏移P_{Thres}，见公式（15）和公式（16）。

$$P_{Thres}[k] = 10^{0.4\times0.364\times\left(\frac{f_c[k]}{1000}\right)^{-0.8}} \quad \cdots\cdots(15)$$

$$P_p[k,n] = P_e[k,n] + P_{Thres}[k] \quad \cdots\cdots(16)$$

该处理阶段的输出$P_p[k,n]$代表“音高模式”。

10.2.1.7 分布

采用电平扩展函数，在频率上抹掉音高模式$P_p[k,n]$。扩展函数为双侧指数函数。低斜率固定为27dB/Bark，而高斜率则取决于频率和能量。

斜率按照公式（17）、公式（18）和公式（19）进行计算。

$$S_u\left[k,\ L[k,\ n]\right] = -24 - \frac{230}{f_c[k]} + 0.2\mathrm{L}[k,\ n] \quad \cdots\cdots(17)$$

$$S_l\left[k,\ L[k,n]\right] = 27 \quad \cdots\cdots(18)$$

式中：

$$L[k,n] = 10\lg\left(P_p[k,n]\right) \quad \cdots\cdots(19)$$

每个频率组k的分布是独立的，见公式（20）。

$$E_2[k,n] = \frac{1}{Norm_{SP}[k]}\left(\sum_{j=0}^{Z-1} E_{line}[j,k,n]^{0.4}\right)^{\frac{1}{0.4}} \quad \cdots\cdots(20)$$

其中E_{line}通过公式（21）得到：

$$E_{line}[j,k,n] = \begin{cases} \dfrac{10^{\frac{L[j,n]}{10}}\cdot 10^{\frac{-res\cdot(j-k)\cdot s_l[j,L[j,n]]}{10}}}{\sum_{\mu=0}^{j-1}10^{\frac{-res\cdot(j-\mu)\cdot s_l[j,L[j,n]]}{10}}+\sum_{\mu=j}^{Z-1}10^{\frac{res\cdot(\mu-j)\cdot s_u[j,L[j,n]]}{10}}} & \text{当 } k < j \text{ 时} \\ \dfrac{10^{\frac{L[j,n]}{10}}\cdot 10^{\frac{res\cdot(k-j)\cdot s_u[j,L[j,n]]}{10}}}{\sum_{\mu=0}^{j-1}10^{\frac{-res\cdot(j-\mu)\cdot s_l[j,L[j,n]]}{10}}+\sum_{\mu=j}^{Z-1}10^{\frac{res\cdot(\mu-j)\cdot s_u[j,L[j,n]]}{10}}} & \text{当 } k \geq j \text{ 时} \end{cases} \quad \cdots\cdots(21)$$

$Norm_{SP}[k]$按照公式（22）和公式（23）进行计算。

$$Norm_{SP}[k] = \left(\sum_{j=0}^{Z-1}\tilde{E}_{line}[j,k]^{0.4}\right)^{\frac{1}{0.4}} \quad \cdots\cdots(22)$$

$$\tilde{E}_{line}[j,k] = \begin{cases} \dfrac{10^{\frac{-res\cdot(j-k)\cdot s_l[j,0]}{10}}}{\sum_{\mu=0}^{j-1}10^{\frac{-res\cdot(j-\mu)\cdot s_l[j,0]}{10}}+\sum_{\mu=j}^{Z-1}10^{\frac{res\cdot(\mu-j)\cdot s_u[j,0]}{10}}} & \text{当 } k < j \text{ 时} \\ \dfrac{10^{\frac{res\cdot(k-j)\cdot s_u[j,0]}{10}}}{\sum_{\mu=0}^{j-1}10^{\frac{-res\cdot(j-\mu)\cdot s_l[j,0]}{10}}+\sum_{\mu=j}^{Z-1}10^{\frac{res\cdot(\mu-j)\cdot s_u[j,0]}{10}}} & \text{当 } k \geq j \text{ 时} \end{cases} \quad \cdots\cdots(23)$$

res表示音高的分辨率，基础版本为0.25Bark，高级版本为0.5Bark。

该处理阶段的模式E_2[k,n]将用于调制模式的计算，称为“未被抹除的激励模式”。

10.2.1.8 时域分布

为了模拟前向掩蔽，每个频率组的能量随时间用一阶低通滤波器进行分布处理。时间常数取决于每个频率组的中心频率（见公式(12)以及表2），并按照公式（24）进行计算。

$$\tau = \tau_{min} + \frac{100}{f_c[k]}(\tau_{100} - \tau_{min}) \quad \cdots\cdots (24)$$

式中：

τ_{100}——取值为0.030s；

τ_{min}——取值为0.008s。

一阶低通滤波器根据公式（25）和公式（26）进行计算。

$$E_f[k,n] = a \times E_f[k,n-1] + (1-a) \times E_2[k,n] \quad \cdots\cdots (25)$$

$$E[k,n] = \max\left(E_f(k,n), E_2(k,n)\right) \quad \cdots\cdots (26)$$

其中a可利用上述时间常数，通过公式（27）进行计算。

$$a = e^{-\frac{4}{187.5}\cdot\frac{1}{\tau}} \quad \cdots\cdots (27)$$

n表示实际的帧，k表示频率组，E_f[k,0]=0

在该处理阶段，模式E[k,n]对应激励模式。

10.2.1.9 掩蔽阈值

掩蔽是一种效应，指的是当一个虽然微弱但明显可感知的信号在相对较强的信号出现时，会感知不到该信号的一种现象。该阈值可采用加权函数m[k]对激励模式进行加权计算得到，见公式（28）和公式（29）。

$$m[k] = \begin{cases} 3.0 & k \cdot res \leq 12 \\ 0.25k \times res & k \cdot res > 12 \end{cases} \quad \cdots\cdots (28)$$

$$M[k,n] = \frac{E[k,n]}{10^{\frac{m[k]}{10}}} \quad \cdots\cdots (29)$$

在该处理阶段，模式M[k,n]代表掩蔽模式。

10.2.2 滤波器组耳朵模型

10.2.2.1 概述

耳朵周边模型和模型中基于滤波器组部分的激励模式的预处理应与图6相符合。

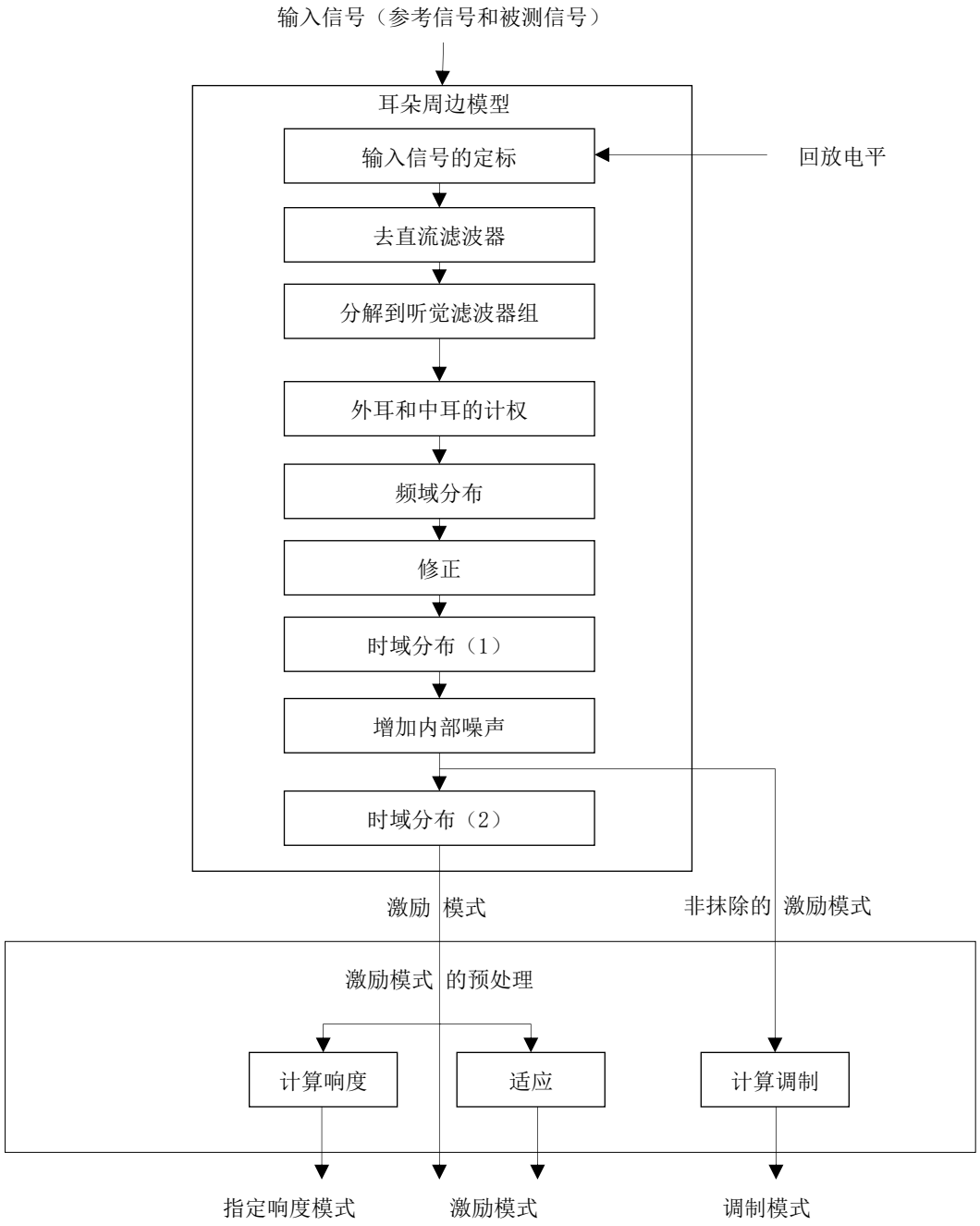


图6　耳朵周边模型和模型中基于滤波器组部分的激励模式的预处理

滤波器组耳朵模型的输入包括被测信号和参考信号，首先将其电平调整到设定的回放电平，然后经过一个高通滤波器，消除其DC成分和次声成分。此后，信号经过线性相位滤波器分解成带通信号，其中线性相位滤波器在可感知音高标度上均匀分布。为了模拟外耳和中耳的频谱特性，采用频率加权函数对带通信号进行频率加权。采用电平分布函数输出的频域卷积对听觉滤波器的频谱分辨率级进行模拟。

通过使用带通信号的Hibert变换（整流）计算出信号包络。为了模拟后向掩蔽，在处理中采用一个具有窗口函数的时域卷积。随后，增加频偏，以模拟听觉系统的内部噪音并模仿静音阈值。最后，使用指数分布函数再一次进行时域卷积，以实现前向掩蔽。

现阶段获得的激励模式可用于计算指定的响度模式，最后一次时域分布处理前的模式（未被抹除模式）用于计算调制模式。这些模式与激励模式一起组成了模型值计算的基础。为了区分被测设备的稳态频率响应的影响与其他失真，被测信号和参考信号的激励模式在频谱上要彼此适应。调制模式和指定响度模式可通过对自适应激励模式和非自适应激励模式进行计算得到。

10.2.2.2 二次采样

在滤波器组输出处，对信号进行采样因子为32的下采样；经过第一次时域分布处理后，对信号进行采样因子为6的下采样，应与图7相符合。

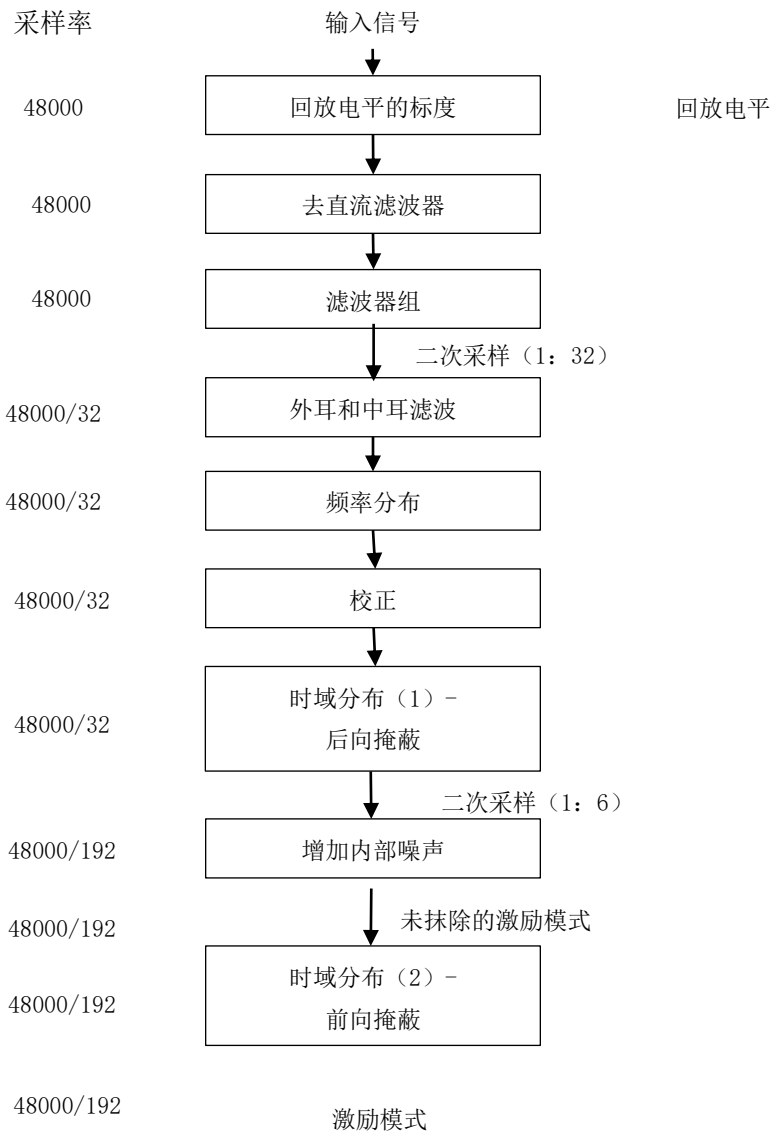


图7 以滤波器组为基础的耳朵周边模型的二次采样

10.2.2.3 回放电平的设置

输入的标度因数fac可通过对满刻度输入信号的设定回放电平进行计算得到，见公式（30）。

$$fac = \frac{10^{Lmax/20}}{32767} \quad \cdots\cdots (30)$$

在未知准确的回放电平的情况下，推荐L_{max}为92dB$_{SPL}$。

10.2.2.4 去直流滤波器

由于滤波器组对输入信号中的次声波非常敏感，因此需对输入信号应用一个去直流滤波器。本文件采用了截止频率为20Hz的四阶Butterworth高通滤波器，该滤波器由两个二阶IIR滤波器串联实现，见公式（31）。

$$y_n = x_n - 2x_{n-1} + x_{n-2} + b_1 y_{n-1} + b_2 y_{n-2} \quad \cdots\cdots (31)$$

其中第一个滤波器的因数为：

$$b_{1,2} = 1.99517, -0.995174$$

第二个滤波器的因数为：

$$b_{1,2} = 1.99799, -0.997998$$

10.2.2.5 滤波器组

对被测信号和参考信号的每个声道，滤波器组均由 40 个滤波器对组成。这些滤波器在听觉音高标度上均匀分布且具有恒定的绝对带宽。每对滤波器由两个滤波器组成，这两个滤波器具有相同的频率响应，但在相位响应上差 90° 。因此，第二个滤波器的输出代表第一个滤波器输出的 Hilbert 变形（若假设第一个滤波器代表的复数信号的实部，则第二个滤波器的输出可代表其虚部）。他们的脉冲响应包络是一个 $\cos^2$ 形状。滤波器定义见公式（29），滤波器的中心频率、脉冲响应长度和额外延时应与表 4 相符合（其中，k 表示滤波器，n 表示时间采样，T 表示两个采样间的间隔：T=1/48000）。把 $h_{re}(k,n)$ 和 $h_{im}(k,n)$ 值当作因数，可以把它们看作 FIR 滤波器。当输入信号受时间限制时，滤波器输出仍可以通过非常快的递归算法计算得到，见公式（32）。

$$\left.\begin{aligned} h_{re}(k,n) &= \frac{4}{N[k]} \cdot \sin^2\left(\pi \cdot \frac{n}{N[k]}\right) \cdot \cos\left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2}\right) \cdot T\right) \\ h_{im}(k,n) &= \frac{4}{N[k]} \cdot \sin^2\left(\pi \cdot \frac{n}{N[k]}\right) \cdot \sin\left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2}\right) \cdot T\right) \end{aligned}\right| 0 \le n < N[k]$$

$$h_{re}(k,n) = h_{im}(k,n) = 0 \quad \left| \begin{array}{l} n < 0 \\ n \ge N[k] \end{array} \right. \quad \cdots\cdots (32)$$

表4 滤波器的中心频率、脉冲响应长度和额外延时

滤波器指数（k）	中心频率（f_c[k]） Hz	脉冲响应长度/样本（N[k]）	额外延时样本（D[k]）
0	50.00	1456	1
1	116.19	1438	10
2	183.57	1406	26

表 4（续）

滤波器指数（k）	中心频率（f_C[k]） Hz	脉冲响应长度/样本（N[k]）	额外延时样本（D[k]）
3	252.82	1362	48
4	324.64	1308	75
5	399.79	1244	107
6	479.01	1176	141
7	563.11	1104	177
8	652.97	1030	214
9	749.48	956	251
10	853.65	884	287
11	966.52	814	322
12	1089.25	748	355
13	1223.10	686	386
14	1369.43	626	416
15	1529.73	570	444
16	1705.64	520	469
17	1898.95	472	493
18	2111.64	430	514
19	2345.88	390	534
20	2604.05	354	552
21	2888.79	320	569
22	3203.01	290	584
23	3549.90	262	598
24	3933.02	238	610
25	4356.27	214	622
26	4823.97	194	632
27	5340.88	176	641
28	5912.30	158	650
29	6544.03	144	657
30	7242.54	130	664
31	8014.95	118	670
32	8869.13	106	676
33	9813.82	96	681
34	10858.63	86	686
35	12014.24	78	690

表 4（续）

滤波器指数（k）	中心频率（f_C[k]） Hz	脉冲响应长度/样本（N[k]）	额外延时样本（D[k]）
36	13292.44	70	694
37	14706.26	64	697
38	16270.13	58	700
39	18000.02	52	703

中心频率范围为50Hz～18000Hz。听域音高标度依据Schroeder等人提出的近似算法进行计算得到，见公式（33）。

$$z = 7\ \mathrm{arsinh}\left(\frac{f}{650}\right) \quad \cdots\cdots（33）$$

为了使所有的滤波器具有相等的延时，对每个滤波器的输入延迟D个采样。其中，D表示信号自身脉冲响应长度与具有最长脉冲响应的滤波器的脉冲响应的长度之差的一半，见公式（34）；在实现时，不应增加样本的额外延时，但用于符合性测试的参考实现方案应包含该额外延时。

$$D[k] = 1 + \frac{1}{2}(N[0] - N[k]) \quad \cdots\cdots（34）$$

对滤波器的输出进行下采样因子为32的二次采样，也就是说，每个输出值需要对每一个滤波器的32个输入样本进行计算得到[3)]。

10.2.2.6 外耳和中耳滤波

外耳和中耳的频率响应通过频率加权函数进行模拟得到。该函数用于滤波器输出，见公式（35）。

$$W[k] = -0.6 \times 3.64 \times \left(\frac{f_c[k]}{1000}\right)^{-0.8} + 6.5e^{-0.6\cdot\left(\frac{f_c[k]}{1000}-3.3\right)^2} - 10^{-3} \times \left(\frac{f_c[k]}{1000}\right)^{3.6} \quad \cdots\cdots（35）$$

伪码：

```
/* inputs */
out_re,out_im :滤波器组输出(实部和虚部)
W  :加权函数(见公式(32))
/* outputs */
out_re,out_im :滤波器组输出
/* intermediate values */
K :滤波器指数
Wt :权重因数
/* 外耳和中二滤波器 */
for(k=0..39)
{
    Wt= pow(10,W[k]/20)
    out_re[k]*= Wt;
```

3) 事实上，较高频率频带中的滤波器的包络并不完全满足采样定理。虽然混叠现象仅在非常特殊的情况下发生（如高频部分被高于 1.5kHz 的频率调制时），且与该现象相关的问题也没有出现在已知的数据库中，但必须强调的是，混叠问题是有可能出现的，特别是测试信号为人工合成信号时。

```
    out_im[k]*= Wt;
}
```

10.2.2.7 频域分布

采用电平分布函数，将滤波器组的输出值对应分布到频率上。分布函数是一个双侧指数函数。低斜率恒定为31dB/Bark，高斜率则在-24dB/Bark到-4dB/Bark之间变换。

高斜率s[k]，可根据公式（36）计算得到。

$$s[k] = \min\left(-4,\ -24 - \frac{230}{f_c[k]} + 0.2\ L[k]\right) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots（36）$$

对于每个滤波器声道而言，电平L[k]是独立计算的。首先求取滤波器输出值的平方绝对值，然后将其转换到分贝（dB）标度。中心频率可从表4中获得。斜率的线性表示在时间上可由一个时间常数为100ms的一阶低通滤波器进行平滑处理。

需要分别计算表示信号实部的滤波器和表示信号虚部的滤波器的频率分布。首先计算在高斜率下的频率分布（与电平有关），之后采用一阶IIR滤波器算法计算低斜率下的频率分布。

伪码：

```
/* inputs */
out_re,out_im :滤波器组输出（实部和虚部）
z[ ] :滤波器组中心频率的临界频带速率，单位Bark(见表4和公式(30))
/* outputs */
A_re,A_im :输出模式
/* intermediate values */
j,k :滤波器标识
a,b :时间平滑因数
dist :计算串扰的常数
L[ ] :每个滤波器输出处的电平
s[ ] :上分布的本地斜率
d1,d2 :缓冲
/* static */ 来自前帧的值被预留；在测量开始时，该值初始化为0
c1, cu[ ] :信号分布部分
```

10.2.2.8 修正

滤波器输出处的能量等于滤波器信号实部和信号虚部的平方和，见公式（37）。

$$E_0[k,n] = A_{re}[k,n]^2 + A_{im}[k,n]^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots（37）$$

10.2.2.9～10.2.2.11中的计算都是基于这些能量进行的。

10.2.2.9 时域分布 - 后向掩蔽

为了模拟后向掩蔽，滤波器输出处的能量随时间采用FIR滤波器进行分布处理。该FIR滤波器具有12抽头的$\cos^2$形脉冲响应（相当于滤波器组输入采样速率下的384个样本的滤波器响应）。在时间分布处理后，对输出进行6倍的下采样。所得结果乘以校正系数cal_1=0.9761，得到相对给定回放电平的适合输出电平，见式（38）。

$$E_1[k,n] = \frac{0.9761}{6} \sum_{i=0}^{11} E_0[k,6n-i] \times cos^2\left(\pi \frac{(i-5)}{12}\right) \quad \cdots\cdots\cdots（38）$$

10.2.2.10 添加内部噪声

在第一次时域分布处理之后，对每个滤波器声道的能量增加频率偏移E_{Thres}，见公式（39）。

$$E_{Thres}[k] = 10^{0.4\times0.364\left(\frac{f_c[k]}{1000}\right)^{-0.8}} \quad \cdots\cdots（39）$$

该处理阶段的模式$E_2[k,n]$见公式（40），将用于调制模式的计算，并被称为“未被抹除的激励模式”。

$$E_2[k,n] = E_1[k,n] + E_{Thres}[k,n] \quad \cdots\cdots（40）$$

10.2.2.11 时域分布－前向掩蔽

为了模拟前向掩蔽，每个滤波器声道中的能量随时间用一阶低通滤波器进行处理抹除。时间常数取决于每个滤波器的中心频率（见表4）并按照公式（41）计算。

$$\tau = \tau_{min} + \frac{100}{f_c[k]} \cdot (\tau_{100} - \tau_{min}) \quad \cdots\cdots（41）$$

式中：

τ_{100}——取值为$0.020s$；

τ_{min}——取值为$0.004s$。

一阶低通滤波器按照公式（42）进行计算。

$$E[k,n] = a \times E[k,n-1] + (1-a) \times E_2[k,n] \quad \cdots\cdots（42）$$

其中时间常数$a = e^{-\frac{192}{48\,000\cdot\tau}}$可根据公式（43）进行计算得到。

$$a = e^{-\frac{192}{48000\tau}} \quad \cdots\cdots（43）$$

该处理阶段的模式E[k, n]被称为激励模式。

10.3 激励模式的预处理

10.3.1 激励模式的预处理概述

本条中的大多数算法均适用于滤波器组耳朵模型和FFT耳朵模型。由于这两种耳朵模型的二次采样因子和频带数量不同，与该因子相关的常数可采用与耳朵模型相关的变量步长（StepSize）和Z进行描述。对于FFT耳朵模型，步长值取1024，z的值可以取55（高级模式）或者109（基础模式）。对于滤波器组耳朵模型，步长值取192，z的值取40。若无其他说明，所有变量和递归滤波器的初始值均为0。

10.3.2 电平和模式调整

10.3.2.1 电平调整和模式调整概述

为了补偿测试信号和参考信号之间的电平差异和线性失真，测试信号和参考信号的电平均调整到两个信号的平均电平。

第一步，每个滤波器声道中的能量采用一阶低通滤波器进行平滑处理。时间常数取决于滤波器组的中心频率，计算见公式（44）。

$$\tau = \tau_{min} + \frac{100}{f_c[k]} \cdot (\tau_{100} - \tau_{min}) \quad \cdots\cdots（44）$$

式中：

τ_{100} ——取值为$0.050s$；

τ_{min}——取值为$0.008s$。

一阶低通滤波器的计算见公式（45）和公式（46）。

$$P_{Ref}[k,n] = a \cdot P_{Ref}[k,n-1] + (1-a) \cdot E_{Ref}[k,n] \quad \cdots\cdots (45)$$

$$P_{Test}[k,n] = a \cdot P_{Test}[k,n-1] + (1-a) \cdot E_{Test}[k,n] \quad \cdots\cdots (46)$$

其中，E_{Test}和E_{Ref}是相互调整后的激励模式，而时间常数a可根据公式（47）进行计算得到。

$$a = e^{-\frac{StepSize}{48\,000 \cdot \tau}} \quad \cdots\cdots (47)$$

10.3.2.2 电平调整

可通过公式（48）对低通输入模式P_{Test}和P_{Ref}进行计算得到瞬时校正系数LevCorr。

$$LevCorr[n] = \left(\frac{\sum_{k=0}^{Z-1}\sqrt{P_{Test}[k,n]\cdot P_{Ref}[k,n]}}{\sum_{k=0}^{Z-1}P_{Test}[k,n]}\right)^2 \quad \cdots\cdots (48)$$

如果校正系数大于1，参考信号应除以校正系数，否则测试信号应乘以该校正系数，见公式（49）和公式（50）。

$$E_{L,Ref}[k,n] = E_{Ref}[k,n]/LevCorr[n] \,|LevCorr[n] > 1 \cdots\cdots (49)$$

$$E_{L,Test}[k,n] = E_{Test}[k,n] \cdot LevCorr[n] \,|LevCorr[n] \leq 1 \cdots\cdots (50)$$

10.3.2.3 模式调整

每个声道的校正系数可通过比较测试信号和参考信号滤波器输出的时间包络得到，见公式（51）。

$$R[k,n] = \frac{\sum_{i=0}^{n} a[k]^i \cdot E_{L,Test}[k,n-i] \cdot E_{L,Ref}[k,n-i]}{\sum_{i=0}^{n} a[k]^i \cdot E_{L,Ref}[k,n-i] \cdot E_{L,Ref}[k,n-i]} \quad \cdots\cdots (51)$$

值a根据公式（41）给出的时间常数按照公式（44）计算得出。如果R[k, n]大于1，测试信号的校正系数设置为$R[k,n]^{-1}$，参考信号的校正系数设置为1。相反，如果R[k, n]不大于1，则参考信号的校正系数设置为R[k, n]，测试信号的校正系数设置为1，见公式（52）。

$$R_{Test}[k,n] = \frac{1}{R[k,n]}, \quad R_{Ref}[k,n] = 1 \Big| R[k,n] \geq 1$$

$$R_{Test}[k,n] = 1, \quad R_{Ref}[k,n] = R[k,n] \Big| R[k,n] < 1 \quad \cdots\cdots (52)$$

如果公式(48)的分母为0(则R[k, n]将不存在)且分子大于0，设置R_{Test}[k, n]为0并设置R_{Ref}[k, n]为1。当公式（48）中的分子为0，R_{Test}[k, n]和R_{Ref}[k, n]的比率可从后续频带计算得到。如果后续频带不存在（如k=0），则R_{Test}[k, n]和R_{Ref}[k, n]比率设置为1。

校正系数在M个滤波器声道上进行平均，并通过公式（44）和公式（47）中给出的相同时间常数随时间进行平滑处理，见公式（53）。对滤波器组耳朵模型，频率窗口M的宽度是3，对FFT耳朵模型，频率窗口M的宽度是4（高级模式）或8（基础模式）。

$$PattCorr_{Test}[k,n] = a \cdot PattCorr_{Test}[k,n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Test}[k+i,n]$$

$$PattCorr_{Ref}[k,n] = a \cdot PattCorr_{Ref}[k,n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Ref}[k+i,n] \quad \cdots (53)$$

$$\left|\begin{array}{ll} M_1 = M_2 = \dfrac{M-1}{2} & |M \text{ 奇} \\ M_1 = \dfrac{M}{2}-1,\ M_2 = \dfrac{M}{2} & |M \text{ 偶} \end{array}\right. \quad \left|\begin{array}{ll} M_1 = M_2 = \dfrac{M-1}{2} & |M \text{ odd} \\ M_1 = \dfrac{M}{2}-1, M_2 = \dfrac{M}{2} & |M \text{ even} \end{array}\right.$$

在频率标度的边界，即频率窗口将超过滤波器组范围的地方，频率窗口的宽度相应降低，见公式(54)。

$$\mathrm{M_1 = \min(M_1, k)},\ \mathrm{M_2 = \min(M_2, z-k-1)},\ \mathrm{M = M_1 + M_2 + 1} \quad \cdots\cdots (54)$$

为了获得频谱自适应模式，电平自适应输入模式采用对应的校正系数进行加权，见公式（55）和公式（56）。

$$E_{P,Ref}[k,n] = E_{L,Ref}[k,n] \cdot PattCorr_{Ref}[k,n] \quad \cdots\cdots (55)$$

$$E_{P,Test}[k,n] = E_{L,Test}[k,n] \cdot PattCorr_{Test}[k,n] \quad \cdots\cdots (56)$$

10.3.3 调制

通过对未抹除的激励模式E_2[k, n]取0.3次幂，并通过公式（57）和公式（58），计算得到简化的响度。该值和其时域得到的绝对值随时间进行抹除。

$$\bar{E}_{der}[k,n] = a \cdot \bar{E}_{der}[k,n-1] + (1-a) \cdot \frac{48000}{StepSize} \cdot |E_2[k,n]^{0.3} - E_2[k,n-1]^{0.3}| \cdots\cdots (57)$$

$$\bar{E}[k,n] = a \cdot \bar{E}[k,n-1] + (1-a) \cdot E_2[k,n]^{0.3} \quad \cdots\cdots (58)$$

值a根据公式（59）给出的时间常数按照公式（47）进行计算。

$$\tau = \tau_0 + \frac{100\ \mathrm{Hz}}{f_c} \cdot (\tau_{100} - \tau_0) \left|\begin{array}{l} \tau_{100} = 0.050\ s \\ \tau_0 = 0.008\ s \end{array}\right. \quad \cdots\cdots (59)$$

根据$\bar{E}_{der}$和$\bar{E}$,按照公式（60）计算每个滤波器输出包络的调制测量值。

$$Mod[k,n] = \frac{\bar{E}_{der}[k,n]}{1 + \bar{E}[k,n]/0.3} \quad \cdots\cdots (60)$$

值$\bar{E}$在下文中还将用于调制差异的计算。

10.3.4 响度

被测信号和参考信号的指定响度模式通过公式（61）进行计算。

$$N[k,n] = const \cdot \left(\frac{1}{s[k]} \cdot \frac{E_{Thres}[k]}{10^4}\right)^{0.23} \cdot \left[\left(1 - s[k] + \frac{s[k] \cdot E[k,n]}{E_{Thres}[k]}\right)^{0.23} - 1\right] \quad \cdots\cdots (61)$$

该公式与Zwicker及Feldtkeller在1967年提出的定义一致。被测信号和参考信号的整体响度是所有滤波器声道中所有大于0的指定响度的总和，见公式（62）。

$$N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} \max(N[k,n], 0) \quad \cdots\cdots (62)$$

为计算1kHz 40dB$_{SPL}$正弦波1宋时的整体响度，FFT耳朵模型的比例常数const为1.07664，滤波器组耳朵模型的比例常数const为1.26539。阈值s和阈值点的激励E_{Thres}分别根据公式（63）和公式（64）进行计算。

$$E_{Three}[k] = 10^{0.364 \cdot \left(\frac{f}{1000}\right)^{-0.8}} \quad \cdots\cdots (63)$$

和

$$s[k] = 10^{\frac{1}{10}\left(-2-2.05\cdot atn\left(\frac{f}{4000}\right)-0.75\cdot atn\left(\left(\frac{f}{1600}\right)^2\right)\right)} \quad \cdots\cdots\cdots\cdots\cdots\cdots (64)$$

注：由于存在不同的耳朵周边模型，此处所指的响度与ISO 532(声学 - 计算响度电平的方法1975)中定义的不一致。

10.3.5 误差信号的计算

误差信号只在FFT耳朵模式中计算，在频域里计算经外耳和中耳过滤后的参考信号和测试信号的幅度频谱之间的差异得到该值，见公式（65），其中参考信号和测试信号的外耳加权FFT输出见10.2.1.4。

$$F_{noise}[k_f,n] = \left|\left|F_{eref}[k_f,n]\right| - \left|F_{etest}[k_f,n]\right|\right| \quad \cdots\cdots\cdots\cdots\cdots\cdots (65)$$

采用10.2.1.5描述的算法，将F_{noise}映射到音高域。

该算法的输出$P_{noise}[n,k]$，被称为噪声模式。

10.4 模型输出变量的计算

10.4.1 概述

用于预测基本音频质量的MOV值应与表5相符合。

表5 用于预测基本音频质量的 MOV 值

MOV	在FFT耳朵模型中计算	在滤波器组耳朵模型中计算	适用的版本	
			基本版本	高级版本
$WinModDiff1_B$	是	否	是	否
$AvgModDiff1_B$	是	否	是	否
$AvgModDiff2_B$	是	否	是	否
$RmsModDiff_A$	否	是	否	是
$RmsNoiseLoud_B$	是	否	是	否
$RmsNoiseLoudAsym_A$	否	是	否	是
$AvgLinDist_A$	否	是	否	是
$BandwidthRef_B$	是	否	是	否
$BandwidthTest_B$	是	否	是	否
总NMR_B	是	否	是	否
$RelDistFrames_B$	是	否	是	否
Segmental NMR_B	是	否	否	是
$MFPD_B$	是	否	是	否
ADB_B	是	否	是	否
EHS_B	是	否	是	是

10.4.2 调制差异

10.4.2.1 概述

被测信号和参考信号的时域包络调制间的差异，通过计算每个滤波器声道的本地调制差异后再进行计算得到，见公式（66）。其中，可根据公式（60）计算出参考信号R_{test}的Mod_{test}和Mod_{Ref}。

$$ModDiff[k,n] = w \cdot \frac{\left|Mod_{test}[k,n] - Mod_{Ref}[k,n]\right|}{offset + Mod_{Ref}[k,n]}$$
$$\begin{cases} w = 1.0 & \left|Mod_{test}[k,n]\right| > Mod_{Ref}[k,n] \\ w = negWt & \left|Mod_{test}[k,n]\right| < Mod_{Ref}[k,n] \end{cases} \quad \cdots\cdots\cdots\cdots\cdots\cdots (66)$$

瞬时调制差异可通过计算所有滤波器所有声道的本地调制差异的平均值得到，见公式（67）。

$$ModDiff[n] = \frac{100}{Z} \sum_{k=0}^{Z-1} ModDiff[k,n] \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (67)$$

静音阈值在计算时考虑了电平相关的加权系数，见公式（68）。该电平相关的加权系数是根据公式（55）给出的参考信号的修正激励模式以及内部噪声函数计算得到，滤波器组耳朵模型的内部噪声定义见公式（39），FFT耳朵模型的内部噪声定义见公式（15）。

$$TempWt[n] = \sum_{k=0}^{Z-1} \frac{\overline{E}_{ref}[k,n]}{\overline{E}_{ref}[k,n] + levWt \cdot E_{Thres}[k]} \text{ ^ } 0.3 \quad \cdots\cdots\cdots\cdots\cdots\cdots (68)$$

采用加权系数TempWt[n]的瞬时调制差异ModDiff[n]的时域平均值将在10.5.2中描述。常数negWt、offset和levWt的值应与表6相符合。

表6　估计总体调制差异的模型输出变量

MOV(Xxx=Win/Avg/Rms)	negWt	offset	levWt
$XxxModDiff1_B$	1	1	100
$XxxModDiff2_B$	0.1	0.01	100
$XxxModDiff_A$	1	1	1

10.4.2.2　$RmsModDiff_A$

模型输出变量$RmsModDiff_A$调制差异的方均值。该差异从滤波器组耳朵模型中计算得到。瞬时平均值参考10.5.2，常数应与表6相符合。

10.4.2.3　$WinModDiff1_B$

模型输出变量$WinModDiff1_B$调制差异的窗口平均值。调制差异根据FFT耳朵模型进行计算得到。时域平均法参考10.5.2.4，常数应与表6相符合。公式（65）给出时间加权系数不能应用于该MOV。

10.4.2.4　$AvgModDiff1_B$和$AvgModDiff2_B$

模型输出变量$AvgModDiff1_B$和$AvgModDiff2_B$是依据FFT耳朵模型计算得到的调制差异的线性平均值。$AvgModDiff1_B$和$AvgModDiff2_B$之间的差异是因为采用了不同常数。时间平均法见10.5.2.2，常数应与表6相符合。

10.4.3　噪声响度

10.4.3.1　概述

模型输出变量估计了当掩蔽参考信号出现情况下所增加失真的部分响度。如果没有掩蔽信号出现，噪声响度的计算公式（见公式（69））用于计算噪声的指定响度；如果与掩蔽信号相比噪声非常小的情况下，噪声响度的计算公式用于计算噪声掩蔽比。

$$NL[k,n] = \left(\frac{1}{s_{test}} \times \frac{E_{Thres}}{E_0}\right)^{0.23} \times \left[\left(1 + \frac{max(s_{test} \times E_{test} - s_{ref} \times E_{ref}, 0)}{E_{Thres} + s_{ref} \times E_{ref} \times \beta}\right)^{0.23} - 1\right] \quad \cdots\cdots (69)$$

其中E_0恒等于1，E_{Tres}指内部噪音函数$E_{Tres}[k]$，定义见公式（36），s可通过公式（70）进行计算。

$$s = ThresFac_0 \times Mod[k,n] + S_0 \quad \cdots\cdots (70)$$

如果没有特殊说明，频谱自适应激励模式（见10.3.2）作为输入，即$E_{Test} = E_{P,\ Test[k,n]}$和$E_{ref} = E_{p,Ref[k,n]}$。系数β决定了掩蔽的数量，可通过公式（71）计算得到。

$$\beta = exp\left(-\alpha \times \frac{E_{test} - E_{ref}}{E_{ref}}\right) \quad \cdots\cdots (71)$$

瞬时噪声响度值需测试信号和参考信号的左右任一声道的噪声响度超过N_{Thres}=0.1sone，50ms后开始计算，见10.5.2.5.2。

在谱平均中，瞬时值由每个临界频带滤波器组的数量进行归一化，而不是由滤波器组的总数进行归一化，也就是，谱平均的结果乘以值为24的系数。

如果瞬时噪声响度低于阈值NL_{min}，设置其为0。估算整体噪声响度的MOV值应与表7相符合。

表7　估算整体噪声响度的 MOV

MOV(Xxx=Win/Avg/Rms)	α	$ThresFac_0$	S_0	NL_{min}
$XxxMissingComponents_B$	1.5	0.15	1	0
$XxxNoiseLoud_B$	1.5	0.15	0.5	0
$XxxMissingComponents_A$	1.5	0.15	1	0
$XxxNoiseLoud_A$	2.5	0.3	1	0.1
$XxxLinDist_A$	1.5	0.15	1	0

10.4.3.2　$RmsNoiseLoud_A$

$RmsNoiseLoud_A$指滤波器组耳朵模型计算得到的噪声响度方均值。时间平均法见10.5.2.3，常数见表7。

10.4.3.3　$RmsMissingComponents_A$

$RmsMissingComponents_A$指滤波器组耳朵模型得到的噪声响度方均值。为了生成在测试信号里丢失的部分（与参考信号相比），可互换测试和参考信号与频谱自适应激励模式一起进行计算。时间平均值见10.5.2.3，常数应与表7相符合。

10.4.3.4　$RmsNoiseLoudAsym_A$

$RmsNoiseLoudAsym_A$是丢失信号组成部份（见10.4.3.3）的响度和噪声（见10.4.3.2）响度的方均值的加权总和，这两个响度都是根据滤波器组耳朵模型进行计算，见公式（72）。

$$RmsNoiseLoudAsym = RmsNoiseLoud + 0.5RmsMissingComponents \quad \cdots\cdots (72)$$

10.4.3.5 $AvgLinDist_A$

$AvgLinDist_A$为被测信号和参考信号频谱自适应过程中丢失的信号组成部份的响度。它以参考信号的频谱自适应模式用作参考，把参考信号未被调整的激励用作测试信号。该MOV的计算基于滤波器组耳朵模型。时间平均法参考10.5.2.2，常数应与表7相符合。

10.4.3.6 $RmsNoiseLoud_B$

$RmsNoiseLoud_B$是FFT耳朵模型计算得到的噪声响度方均值。时间平均法见10.5.2.3，常数应与表7相符合。

10.4.4 带宽

10.4.4.1 概述

MOV估计了FFT行内被测信号和参考信号的平均带宽。

对于每帧而言，本地带宽B_{wRef}[n]和B_{wTest}[n]按照10.4.4.2进行计算。

10.4.4.2 伪码

```
/* inputs */
FLevRef[], FlevelTest[] :FFT输出电平，dB
/* outputs */
BwRef, BwTest :输出模式
/* intermediate values */
K  :FFT行指数
ZeroThreshold :带宽阈值

ZeroThreshold = FLevelTst(921);
BwRef = BwTst = 0.0;
for(k=921;k<1024;k++)
    {
      ZeroThreshold=max(ZeroThreshold,FLevelTst(k));
    }
for (k = 920; k>=0; k--)
  {
    if (FLevelRef[k] >= 10.0+ZeroThreshold)
       {
         BwRef = k+1;
         break;
       }
  }
for (k = BwRef-1; k>=0; k--)
 {
   if(FLeveltest[k] >= 5.0+ZeroThreshold)
       {
       BwTest=k+1;
```

```
        break;
        }
}
```

10.4.4.3 BandwidthRef$_B$和BandwidthTest$_B$

BandwidthRef$_B$指BwRef的线性平均值，BandwidthTest$_B$指BwTest的线性平均值。平均计算时，只考虑BwRef 大于346的帧。测试条目的开始和结束部分能量较小的帧会被忽略。时间平均法见10.5.2.2。

10.4.5 噪声掩蔽比(NMR)

10.4.5.1 概述

下列模型的值可通过对噪音和掩蔽的值进行计算得到。
当前帧n的本地NMR计算见公式（73）。

$$NMR_{local}[n] = 10\,lg \quad \frac{1}{Z}\sum_{k=0}^{Z-1}\frac{P_{noise}[k,n]}{M[k,n]} \quad \cdots\cdots(73)$$

10.4.5.2 总 NMR$_B$

模型输出变量总NMR$_B$即噪声掩蔽比的线性平均值可通过公式（74）计算得到：

$$NMR_{tot} = 10lg\frac{1}{N}\sum_{n}\left(\frac{1}{Z}\sum_{k=0}^{Z-1}\frac{P_{noise}[k,n]}{M[k,n]}\right) \quad \cdots\cdots(74)$$

在条目开始和结束处的低能量帧会被忽略（见10.5.2.5.4）。

10.4.5.3 Segmental NMR$_B$

模型输出值Segmental NMR$_B$指本地NMR线性平均值。时间平均法见10.5.2.2。
在条目开始和结束处的低能量帧会被忽略（见10.5.2.5.4）。

10.4.6 相对波动 FRames$_B$(Relative Disturbed FRames$_B$)

模型输出值相对波动FRames$_B$（即RelDistFrames$_B$)表示该条目中相对帧总数具备公式（75）条件的帧的数量。

$$\max_{\forall k}\left(10\,lg\left(\frac{P_{noise}[k,n]}{M[k,n]}\right)\right) \geq 1.5\ \text{dB} \quad k \in [0,Z-1] \quad \cdots\cdots(75)$$

在条目开始和结束处的低能量帧会被忽略（见10.5.2.5.4）。

10.4.7 检测概率

10.4.7.1 概述

本条定义的MOV以$\tilde{E}[k,n]$(k频带,n帧)为基础进行计算，$\tilde{E}[k,n]$是激励模式E[k,n]的对数表示形式，见公式（76）。

$$\tilde{E}[k,n] = 10\ lg(E[k,n]) \quad \cdots\cdots(76)$$

对于每帧n：

每个声道c（c的值指左声道和右声道）单独执行步骤公式（77）～公式（91）。参考信号的对数激励模式为$\tilde{E}_{ref}[k,n]$，被测信号的对数激励模式为$\tilde{E}_{test}[k,n]$。

对于每个频带k。

——计算非对称平均激励，见公式（77）。

$$L[k,n] = 0.3\ \max(\tilde{E}_{ref}[k,n],\ \tilde{E}_{test}[k,n]) + 0.7\ \tilde{E}_{test}[k,n] \quad \cdots\cdots (77)$$

——计算有效检测步长 s。公式（78）为临界可感知的电平差异测量方法的近似法。

如果L[k, n]＞0，则

$$\mathrm{s[k,n]} = 5.95072\left(\frac{6.39468}{\mathrm{L\,[\,k,\ n]}}\right)^{1.71332} + 9.01033\times10^{-11}\mathrm{L[k,n]}^4 + 5.05622\times10^{-6}\mathrm{L[k,n]}^3 - 0.00102438\mathrm{L[k,n]}^2 + 0.0550197\mathrm{L[k,n]} - 0.198719 \quad \cdots\cdots (78)$$

否则

$$\mathrm{s[k,n]} = 1.0\times10^{30} \quad \cdots\cdots (79)$$

——计算签名误差 e，见公式（80）。

$$e[k,n] = \tilde{E}_{ref}[k,n] - \tilde{E}_{test}[k,n] \quad \cdots\cdots (80)$$

——如果$\tilde{E}_{ref}[k,n] > \tilde{E}_{test}[k,n]$，则斜率 b 的陡度设置为 4.0，否则设置为 6.0。该过程模拟了如下效应：相对于参考信号，增加被测信号的信号能量比降低被测信号的能量时效果要显著。

——计算标度系数 a，见公式（81）。

$$a[k,n] = \frac{10^{\frac{lg(lg(2.0))}{b}}}{s[k,n]} \quad \cdots\cdots (81)$$

——计算检测概率，见公式（82）。公式（81）设置了标度系数 a，如果 e[k, n]等于 s[k, n]，$p_c[k,n]$变为 0.5。

$$p_c[k,n] = 1 - 10^{(-a[k,n]\cdot e[k,n])^b)} \quad \cdots\cdots (82)$$

——计算大于阈值的所有步进数，见公式（81）。

$$q_c[k,n] = \frac{|INT(e[k,n])|}{s[k,n]} \quad \cdots\cdots (83)$$

——双声道检测概率见公式（84）。

$$p_{bin}[k,n] = max(p_{left}[k,n], p_{right}[k,n]) \quad \cdots\cdots (84)$$

——双声道的大于阈值的步进数见公式（85）。

$$q_{bin}[k,n] = max(q_{left}[k,n], q_{right}[k,n]) \quad \cdots\cdots (85)$$

——帧 n 声道 c 的总检测概率见公式（86）。

$$P_c[n] = 1 - \prod_{\forall k}\ (1 - p_c[k,n]) \quad \cdots\cdots (86)$$

其中c可以是左声道，也可以是右声道，或是双声道。帧n声道c中大于阈值的步进数，见公式（87）。

$$Q_c[n] = \sum_{\forall k}\ q_c[k,n] \quad \cdots\cdots (87)$$

10.4.7.2 最大过滤检测概率(MFPD$_B$)

每个声道c的平滑检测概率的计算见公式（88）。

$$\tilde{P}_c[n] = (1 - c_0) \times P_c[n] + c_0 \times \tilde{P}_c[n - 1] \quad \cdots\cdots (88)$$

其中，$P_c[-1]=0$，常数c_0取决于步长（StepSize），见公式（89）。

$$c_0 = 0.9^{StepSize/1024} \quad \cdots\cdots (89)$$

c_0把灵敏度降低至非常小的失真。

最大过滤检测概率(MFPD)的计算见公式（90）。

$$PM_c[n] = \max\left(PM_c[n-1] \cdot c_1, \tilde{P}_c[n]\right) \quad \cdots\cdots (90)$$

其中$PM_C[-1]=0$，常数c_1取决于步长，见公式（91）。

$$c1 = 0.99^{StepSize/1024} \quad \cdots\cdots (91)$$

根据遗忘原理，c_1模拟了音频片段在起始处的失真比在结束处的失真感觉要轻的现象。注意：该常数对模拟被测者不能选择素材中的一小段进行测评的听音测试非常有用。使用GY/T 298—2016推荐听音测试得到的数据校正现有模型，c_1应为1.0。

MFPD是最后一帧的$PM_{bin}[n]$值。

10.4.7.3 平均失真块(ADB_B)

双声道$P_{bin}[n]$的检测概率高于0.5的有效帧数记为$n_{distorted}$。

对于所有有效帧，大于双声道$Q_{bin}[n]$阈值的总步进数可采用公式（92）进行计算。

$$Q_{sum} = \sum_{\forall n} Q_{bin}[n] \quad \cdots\cdots (92)$$

平均失真块ADB的失真采用以下公式计算：

——如果 $n_{distorted}$ 等于 0，则 ADB=0(听不见失真)；

——如果 $n_{distorted}$＞0 且 Q_{sum}＞0，则 $ADB=lg((Q_{sum})/n_{distorted})$；

——如果 $n_{distorted}$＞0 且 Q_{sum}=0，则 ADB=-0.5。

注：在此处块相当于帧。

10.4.8 误差的谐波结构

10.4.8.1 概述

含有较强谐波的参考信号（如低音单簧管、大键琴）具有规则排列的波谷和波峰的频谱特性。在某些情况下，误差信号可能会继承这样的结构。例如，对混杂在这类信号中的噪声，当该噪声处于信号较低的频谱波谷位置时，该噪声将未被掩蔽。此时所得出的误差频谱在结构上与原始频谱相似，但会为了与波谷位置相对应而存在频率偏移。该结构可能会导致声调失真，突显误差。

误差指参考和被测信号的频谱记录中的差异，每个误差都经外耳和中耳的频率响应进行加权(见公式(7))。在这里，未采用心理声学模型的激励模式，因为从非线性频率到音高巴克的转换不会抹除谐波结构。

10.4.8.2 EHS_B

谐波结构幅度可通过识别和测量自相关函数频谱中的最大峰值得到。按照公式（93），每个关联性可由两个向量夹角的余弦值计算得到，其中，$\vec{F}_0$是误差向量，$\vec{F}_t$是延迟t时刻的误差向量。关联性长度等于最大延迟（见下例，关联性长度为256）。

$$C = \frac{\vec{F}_0 \cdot \vec{F}_t}{|\vec{F}_0| \cdot |\vec{F}_t|} \quad \cdots\cdots (93)$$

用于获取自相关函数的最大延迟为小于18kHz对应的FFT频率组成数量一半的2的最大次幂。

例如，在采样率为48kHz且FFT窗口为2048个采样的情况下，18kHz对应的FFT组成为(18/24)×1024=768，因此，最大延迟可能是384。实际的延时可能是256，即为小于384的2的最大幂次方。相关函数的第一个值可通过对齐Ft[0]与F_0[0]获得，最后一个值可通过对齐Ft[0]与F_0[255]对齐获得。

所得关联向量采用归一化Hann窗口进行取值，通过减去平均值以消除DC成分，然后采用FFT计算功率谱。频谱中第一个波谷后的最大峰值为自相关函数中的主频率。帧间最大峰值的平均值乘以1000.0就是谐波失真结构（EHS）变量。

10.5 平均法

10.5.1 频谱平均法

10.5.1.1 概述

如果MOV描述（见10.4）中没有特殊说明，10.5.1.2的计算用于频带本地值的平均计算。

10.5.1.2 线性平均

线性平均值的计算见公式（94）。

$$AvgS = \frac{1}{Z} \times \sum_{k=0}^{Z-1} S[k] \quad \cdots\cdots (94)$$

其中，S表示MOV，Z表示频带的数量。

10.5.2 时域平均

10.5.2.1 概述

如果MOV描述（见10.4）中没有特殊说明，10.5.2.2～10.5.2.5的计算用于瞬时值的平均计算。时间加权系数为W，Z指频带的数量。

10.5.2.2 线性平均

线性平均值（带前缀Avg）的算法见公式（95）。

$$AvgX = \frac{1}{N} \times \sum_{n=0}^{N-1} X[n] \quad \cdots\cdots (95)$$

其中，X表示MOV，N表示计算X的瞬时值所用的时域采样数。

在采用加权的情况下（见10.4.2），线性平均值的算法见公式（96）。

$$AvgX = \frac{\sum_{n=0}^{N-1} W[n] \cdot X[n]}{\sum_{n=0}^{N-1} W[n]} \quad \cdots\cdots (96)$$

10.5.2.3 方均值

方均值（带前缀Rms）的算法见公式（97）。

$$RmsX = \sqrt{\frac{1}{N} \times \sum_{n=0}^{N-1} X[n]^2} \quad \cdots\cdots (97)$$

其中，X表示MOV，N表示计算X的瞬时值所用的时域采样数。

在采用加权的情况下，方均值的算法见公式（98）。

$$RmsX = \sqrt{Z} \times \sqrt{\frac{\sum_{n=0}^{N-1} W[n]^2 \cdot X[n]^2}{\sum_{n=0}^{N-1} W[n]^2}} \quad \cdots\cdots (98)$$

10.5.2.4 窗口化的平均值

窗口化平均值（带前缀 Win）的算法见公式（99）。

$$WinX = \sqrt{\frac{1}{N-L+1} \times \sum_{n=L-1}^{N-1} \left(\frac{1}{L} \times \sum_{i=0}^{L-1} \sqrt{X[n-i]}\right)^4} \quad \cdots\cdots\cdots\cdots\cdots\cdots (99)$$

其中，X表示MOV，N表示计算X的瞬时值所用的时域采样数，L表示时域采样中滑动时间窗口的长度。窗口长度约为100ms，也就是，FFT耳朵模型的L等于4，滤波器组耳朵模型的L等于25。

10.5.2.5 帧的选择

10.5.2.5.1 延迟的平均算法

计算MOV时，在进行时域平均计算过程中应剔除测量开始后的初始0.5s的值。延迟的平均法用于的MOV包括WinModDiff1、AvgModDiff1、AvgModDiff2、RmsNoiseLoudness、RmsNoiseLoudAsym、RmsModDiff、AvgLinDist。

10.5.2.5.2 响度阈值

计算MOV时，当任一对应音频声道（测试信号和参考信号）的整体响度达到N_{Thres} 50ms后才开始计算，这之前所测得的所有瞬时值都不计入时间平均值中。响度阈值仅适用于10.4.3描述的MOV。

10.5.2.5.3 能量阈值

在单声道，或在参考和测试信号的左或右信道中，若某个具有2048个采样的帧的最近一半的能量小于8000[4]，则忽略该帧。帧与帧之间有50%的重叠，仅对含有新数据的半帧进行评价。本条的应用排除了对很小能量的帧的处理。

本条仅适用于10.4.7描述的MOV值。

10.5.2.5.4 数据边界

相对于正式参考文件，如果处理后的文件在参考文件的头或者尾含有噪音，那么相关误差可能会很大，因为参考电平接近-∞。若该误差被认定是损伤，则通过应用数据边界拒绝准则可以忽略该误差。

当首次打开文件时，要识别参考信号中真实数据的开始和结尾位置。将原始数据的开始作为起点，对音频中的一个声道从头到尾进行扫描，直至5个连续采样的绝对值总和超过200的地方作为真实数据的开始位置，同理，也可以将原始数据的结尾作为起点从尾向前进行扫描。完全处于该范围以外的帧将会被忽略。

本条适用于所有MOV的计算。

10.5.3 音频声道上的平均

若无特殊说明，立体声信号的MOV等于时域平均后的左右声道MOV的线性平均。

10.6 感知基本音频质量的估算

10.6.1 概述

感知基本音频质量的估算，主要是采用具有隐藏层的人工神经网络将多个MOV映射成一个数值的方法。

10.6.2 人工神经网络

4) 该数值指 16bit 有符号整数格式表示的输入数据，范围为－32768～32767，一般用于 CD。

神经网络的激活函数是一个非对称sigmoid函数，见公式（100）。

$$\mathrm{sig}(x) = \frac{1}{1 + e^{-x}} \quad \cdots\cdots (100)$$

网络在隐藏层中需要I个输入和J个节点。输入比例系数$a_{min}[i]$和$a_{max}[i]$、输入加权$w_x[i]$、输出加权$w_y[j]$和输出标度系数b_{min}和b_{max}确定了映射关系，将输入映射到失真指数（DI），见公式（101）。

$$DI = w_y[J] + \sum_{j=0}^{J-1}\left(w_y[j] \cdot \mathrm{sig}\left(w_x[I,j] + \sum_{i=0}^{I-1} w_x[i,j] \cdot \frac{x[i] - a_{min}[i]}{a_{max}[i] - a_{min}[i]}\right)\right) \quad \cdots\cdots (101)$$

失真指数DI与感知基本音频质量（ODG）直接相关。DI和ODG之间的关系可用公式（102）表示。

$$ODG = b_{min} + (b_{max} - b_{min}) \cdot \mathrm{sig}(DI) \quad \cdots\cdots (102)$$

10.6.3 基础版本

基础版本仅采用FFT耳朵模型。它采用以下MOV值：$BandwidthRef_B$、$BandwidthTest_B$、总NMR_B、$WinModDiff1_B$、ADB_B、EHS_B、$AvgModDiff1_B$、$AvgModDiff2_B$、$RmsNoiseLoud_B$、$MFPD_B$和$RelDistFrames_B$。通过采用10.6.2中描述的神经网络可将这11个MOV映射成质量指数。该质量指数在隐藏层中具有3个节点。映射的参数应与表8～表12相符合。

表8 基础版本中使用的 MOV

MOV	目的
$WinModDiff1_B$	调制中的变化（与粗糙度相关）
$AvgModDiff1_B$	
$AvgModDiff2_B$	
$RmsNoiseLoud_B$	失真的响度
$BandwidthRef_B$	线性失真（频率响应等）
$BandwidthTest_B$	
$RelDistFrames_B$	可感知失真的频率
总NMR_B	噪声掩蔽比
$MFPD_B$	检测概率
ADB_B	
EHS_B	误差的谐波结构

表9 适用于基础版本输入的比例系数

index(i)	MOV(x[i])	$a_{min}[i]$	$a_{max}[i]$
0	$BandwidthRef_B$	393.916656	921
1	$BandwidthTest_B$	361.965332	881.131226
2	总 NMR_B	－24.045116	16.212030
3	$WinModDiff1_B$	1.110661	107.137772
4	ADB_B	－0.206623	2.886017
5	EHS_B	0.074318	13.933351
6	$AvgModDiff1_B$	1.113683	63.257874
7	$AvgModDiff2_B$	0.950345	1145.018555
8	$RmsNoiseLoud_B$	0.029985	14.819740

表9（续）

index(i)	MOV(x[i])	$a_{min}[i]$	$a_{max}[i]$
9	$MFPD_B$	0.000101	1
10	$RelDistFrames_B$	0	1

表10　适用于基础版本输入节点的加权系数

index(i)	MOV(x[i])	节点1(w_X[i,0])	节点2(w_X[i,1])	节点3(w_X[i,2])
0	$BandwidthRef_B$	-0.502657	0.436333	1.219602
1	$BandwidthTest_B$	4.307481	3.246017	1.123743
2	总NMR_B	4.984241	-2.211189	-0.192096
3	$WinModDiff1_B$	0.051056	-1.762424	4.331315
4	ADB_B	2.321580	1.789971	-0.754560
5	EHS_B	-5.303901	-3.452257	-10.814982
6	$AvgModDiff1_B$	2.730991	-6.111805	1.519223
7	$AvgModDiff2_B$	0.624950	-1.331523	-5.955151
8	$RmsNoiseLoud_B$	3.102889	0.871260	-5.922878
9	$MFPD_B$	-1.051468	-0.939882	-0.142913
10	$RelDistFrames_B$	-1.804679	-0.503610	-0.620456
11	偏差	-2.518254	0.654841	-2.207228

表11　适用于基础版本输出节点的加权系数

节点1(w_y[0])	节点2(w_y[1])	节点3(w_y[2])	偏差(w_y[3])
-3.817048	4.107138	4.629582	-0.307594

表12　适用于基础版本输出节点的比例系数

比例系数	b_{min}	b_{max}
系数值	-3.98	0.22

10.6.4　高级版本

高级版本既采用滤波器组耳朵模型，也采用FFT耳朵模型。它采用了5个MOV，包括：$RmsModDiff_A$、$RmsNoiseLoudAsym_A$、$AvgLinDist_A$、Segmental NMR_B 和EHS_B。通过10.6.2中描述的神经网络，可以将这些MOV映射成一个质量指数。该质量指数在隐藏层里具有5个节点。映射的参数应与表13～表17相符合。

表13　高级版本中使用的 MOV

MOV	目的
$RmsNoiseLoudAsym_A$	失真的响度
$RmsModDiff_A$	调制中的变化（与粗糙度相关）
$AvgLinDist_A$	线性失真（频率响应）
Segmental NMR_B	噪声掩蔽比
EHS_B	误差的谐波结构

表14　适用于高级版本输入节点的比例系数

index(i)	MOV(x[i])	$a_{min}[i]$	$a_{max}[i]$
0	$RmsModDiff_A$	13.298751	2166.5
1	$RmsNoiseLoudAsym_A$	0.041073	13.24326
2	Segmental NMR_B	-25.018791	13.46708
3	EHS_B	0.061560	10.226771
4	$AvgLinDist_A$	0.024523	14.224874

表15　用于高级版本输入的加权系数

index(i)	MOV(x[i])	节点1 ($w_X[i,0]$)	节点2 ($w_X[i,1]$)	节点3 ($w_X[i,2]$)	节点4 ($w_X[i,3]$)	节点5 ($w_4[i,4]$)
0	$RmsModDiff_A$	21.211773	-39.913052	-1.382553	-14.545348	-0.320899
1	$RmsNoiseLoudAsym_A$	-8.981803	19.956049	0.935389	-1.686586	-3.238586
2	Segmental NMR_B	1.633830	-2.877505	-7.442935	5.606502	-1.783120
3	EHS_B	6.103821	19.587435	-0.240284	1.088213	-0.511314
4	$AvgLinDist_A$	11.556344	3.892028	9.720441	-3.287205	-11.031250
5	偏差	1.330890	2.686103	2.096598	-1.327851	3.087055

表16　用于高级版本的输出节点的加权

节点1 ($w_X[i,0]$)	节点2 ($w_X[i,1]$)	节点3 ($w_X[i,2]$)	节点4 ($w_X[i,3]$)	节点5 ($w_4[i,4]$)	偏差 ($w_y[5]$)
-4.696996	-3.289959	7.004782	6.651897	4.009144	-1.360308

表17　用于高级版本输出节点的比例系数

比例系数	b_{min}	b_{max}
系数值	-3.98	0.22

10.7　实现方案的一致性

10.7.1　概述

本条给出一组测试序列，以验证测量方法的正确实现方案。

10.7.2 测试素材

测试序列为16段，其MOV和DI值的范围较大。

10.7.3 一致性测试的设置

测试序列由ITU-R提供，均为WAV文件（Microsoft RIFF格式），采样格式均为48kHz、16bit PCM。ITU提供的测试和参考信号均已经进行时间对齐和电平调整，因此不需要增加额外的增益或延时。测量算法的听音电平应调整到92dB SPL。

10.7.4 可接受的容许空间

为了符合本文件的规定，所有测试条目的DI值应与表18和表19的值一致，容差应小于±0.02[5)]。如果某个实现方案得到的值超出该容差范围，则该方案不符合本文件。

10.7.5 测试项目

测试序列的DI值和ODG值应与表18和表19相符合，其中表18的DI值为基础版本的DI值，表19为高级版本的DI值。其中，被测条目文件的文件名用“cod”表示，参考测试条目文件的文件名用“ref”表示，例如bcodtri.wav为被测序列，对应的参考序列名称为breftri.wav。

表18 用于基础版本的测试素材以及其 DI 值

测试序列	DI	ODG
acodsna.wav	1.304	-0.676
bcodtri.wav	1.949	-0.304
ccodsax.wav	0.048	-1.829
ecodsmg.wav	1.731	-0.412
fcodsb1.wav	0.677	-1.195
fcodtr1.wav	1.419	-0.598
fcodtr2.wav	-0.045	-1.927
fcodtr3.wav	-0.715	-2.601
gcodcla.wav	1.781	-0.386
icodsna.wav	-3.029	-3.786
kcodsme.wav	3.093	0.038
lcodhrp.wav	1.041	-0.876
lcodpip.wav	1.973	-0.293
mcodcla.wav	-0.436	-2.331
ncodsfe.wav	3.135	0.045
scodclv.wav	1.689	-0.435

5) 为达到该精度，需采用 IEEE 浮点算法。

表19　用于高级版本的测试序列以及其 DI 值

测试序列	DI	ODG
acodsna.wav	1.632	-0.467
bcodtri.wav	2.000	-0.281
ccodsax.wav	0.567	-1.300
ecodsmg.wav	1.594	-0.489
fcodsbl.wav	1.039	-0.877
fcodtrl.wav	1.555	-0.512
fcodtr2.wav	0.162	-1.711
fcodtr3.wav	-0.783	-2.662
gcodcla.wav	1.457	-0.573
icodsna.wav	-2.510	-3.664
kcodsme.wav	2.765	-0.029
lcodhrp.wav	1.538	-0.523
lcodpip.wav	2.149	-0.219
mcodcla.wav	0.430	-1.435
ncodsfe.wav	3.163	0.050
scodclv.wav	1.972	-0.293

附 录 A
（资料性）
本文件与 ITU-R BS.1387-1 相比的结构变化情况

本文件与 ITU-R BS.1387-1 相比在结构上有较多的调整，具体章条编号对照情况见表 A.1。

表 A.1 本文件与 ITU-R BS.1387-1 的章条编号对照情况

本文件章条编号	ITU-R BS.1387-1 章条编号
1	—
2	—
3.1	Glossary
3.2	Abbreviations
4	附件 1 中 1
5	附件 1 中 2
6	附件 1 中 3
7	附件 1 中 4
8	附件 1 中 5
9	附件 1 中 6
10	附件 2
10.1	附件 2 中 1
10.2	附件 2 中 2
10.3	附件 2 中 3
10.4	附件 2 中 4
10.5	附件 2 中 5
10.6	附件 2 中 6
10.7	附件 2 中 7
附录 A	—
附录 B	附件 1 附录 4
附录 C	附件 1 附录 1
附录 D	附件 1 附录 2
附录 E	附件 1 附录 3
—	附件 2 附录 1
—	附件 2 附录 2

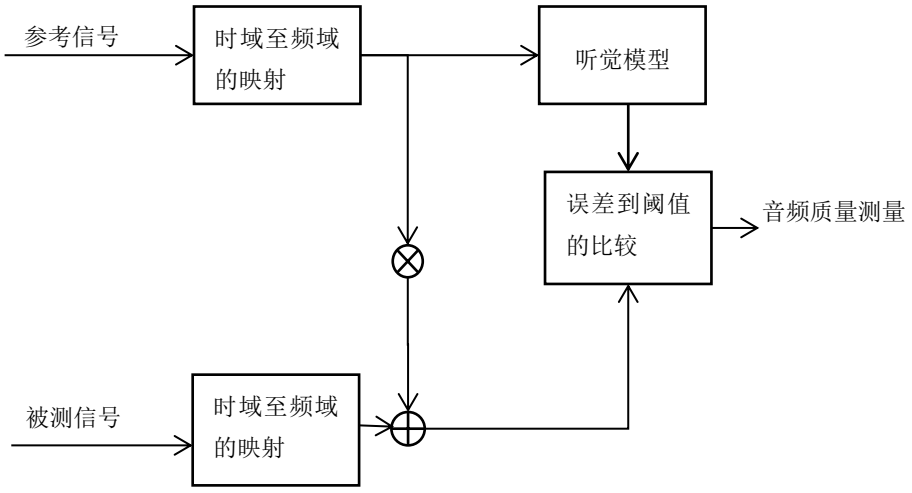
附 录 B
（规范性）
感知音频质量的客观测量方法的原则和特点

B.1 感知音频质量客观测量方法的通用结构

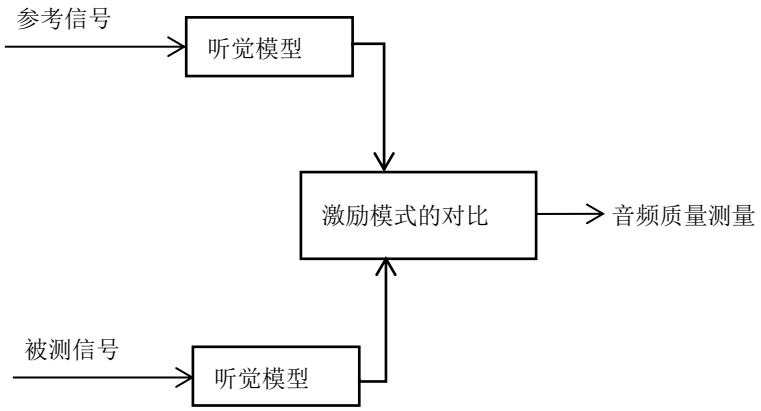
所有感知音频质量的客观测量方法均包括两组输入信号：一组为参考信号，另一组为被测信号。若参考信号不能传输给测量设备且信号是通用的信号时，参考信号可以存储在测量设备中，成为内部参考信号。输入信号必须时序一致。

可以通过两种不同的方法把心理声学引入测量方法。第一种方法与音频编码的结构非常相似：用参考信号估算实际掩蔽阈值，再将被测信号和参考信号的差异与该掩蔽阈值进行比较，这就是“掩蔽阈值概念”，用于噪音响度和噪声掩蔽比中。输入信号间的差异可以在时域中进行计算，或作为短时能量频谱进行计算其差异。后者对时间对准误差提供了更高的稳健性，但降低了时间分辨率。由于时域中的差异通常对相位失真非常敏感，因此不再使用。第二种方法与人类听觉系统中的生理学过程相近：对参考信号和被测信号的内部表征进行计算。内部表征指人类大脑对信号比较后得到的有用信息的评价，该方法称之为内部表征比较法，并用于 ASD。

两种方法的示意图见 B.1。



a） 掩蔽阈值法



b） 内部表征比较法

图 B.1 在感知音频质量的客观测量中用于不同方法的生理听觉概念

B.2 心理声学及认知基础

B.2.1 心理声学认知基础的概述

本条对人类听觉系统的属性进行讨论。这些属性在对音频信号的感知质量进行评价的过程中具有突出的作用。本条着重阐述这些属性是如何进行建模的。

B.2.2 外耳及中耳传输特性

通常而言，声音信号必须经过外耳及中耳，到达内耳，在内耳进行声音检测和分析处理。外耳及中耳的职能就如同一个滤波器组对输入信号进行处理。听觉神经中出现的噪声与血液流动产生的噪声一起叠加到输入信号上，噪声幅度在低频比较大。外耳及中耳传输功能与内部噪声限制了对较小音频信号的识别能力，对听力的绝对阈值产生了巨大的影响。

B.2.3 感知频率标度

人类耳朵中的声压受体是毛细胞。它们位于内耳，具体来说是在耳蜗里。在耳蜗里，实现了频率到位置的转变。最大激励位置取决于输入信号的频率。在耳蜗指定位置上的每个毛细胞对频率标度上的重叠范围负责。音高的感知与毛细胞的恒定距离相关。

采用心理声学实验不同，频率到音高的转换函数也不同：

Zwicker 和 Feldtkeller 在 1967 年将频率划分为 24 个非重叠带，也就是所谓的临界频带，单位为赫兹（Hz），应与表 B.1 相符合。表 B.1 包含了频带的上限截止频率。该表还包括对应的 Bark 标度，具体为：1Bark 对应 100Hz，24Bark 对应 15500Hz。

表 B.1 Zwicker 定义的临界频带

临界频带	上限截止频率 Hz	临界频带	上限截止频率 Hz	临界频带	上限截止频率 Hz
1	100	9	1080	17	3700
2	200	10	1270	18	4400
3	300	11	1480	19	5300
4	400	12	1720	20	6400

表 B.1（续）

临界频带	上限截止频率 Hz	临界频带	上限截止频率 Hz	临界频带	上限截止频率 Hz
5	510	13	2000	21	7700
6	630	14	2320	22	9500
7	770	15	2700	23	12000
8	920	16	3150	24	15500

在 Cohen 和 Fielder 在 1992 年发表的论文中提到，采用巴克（Bark）标度得到的感知音频客观测量结果最优。

B.2.4 激励

每个毛细胞对一个范围的频率做出反应，类似于滤波器特性。滤波器的斜率最好用 B.2.3 定义的感知频率标度进行表示，采用该标度表示的滤波器形状几乎与中心频率无关。激励的下行斜率（约 27dB/Bark）与输入信号电平 L 无关，上行斜率与输入信号电平有关，低电平的斜率比高电平的斜率大，上行斜率的范围为（－5～－30）dB/Bark，应符合图 B.2。这个陡峭特性是由于两种不同毛细胞的反馈机制造成的，且需要一定的时间来解决。因此，从信号开始几毫秒之后的静态信号中可能获得最佳听觉频率分辨率。含有多种组成的信号的激励模式需要非线性相加。

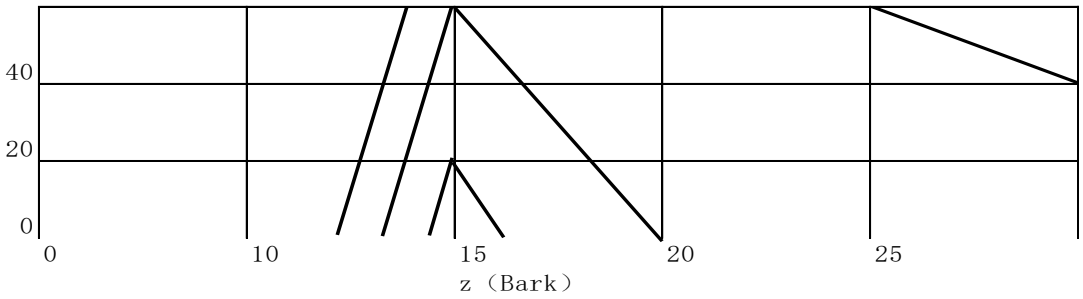


图 B.2 激励的电平依存关系

在接触信号以后，毛细胞及神经处理需要一些时间来恢复，直到敏感度完全恢复为止。恢复过程所需时间取决于信号的电平及持续时间，可持续数百毫秒。信号从毛细胞传输到大脑的时间，高电平信号比低电平信号快。因此，如果开始的信号声音较大，就能够掩蔽随之而来的柔和信号。

另一个模拟激励的方法以 Moore 在 1986 年提出的 ERB 标度为基础。这个方法使用了 Moore 在 1986 年提出的 ROEX 滤波器。在感知音频质量的客观测量相关文献中，以 Zwicker 和 Feldtkeller 在 1967 年提出的算法以及 Terhardt 在 1979 年提出的算法为基础的模型得出的结果较好。

B.2.5 检测

不同音频信号的激励被传到人类大脑。大脑根据信息呈现的细节程度和持续时间进行划分，有 3 个不同的存储区：长期记忆，短期记忆和超短期记忆。在听音测试中，超短期记忆发挥着显著的作用。如果听众或评价员听到音频的持续时间小于 5s～8s，那么信号的绝大多数细节被保留了。在 GY/T 298

—2016 规定的程序考虑到了这个因素，因此，GY/T 298—2016 规定的程序中，允许评价员选择音频中非常短的一部分进行仔细听。在检测阈值处，检测概率是 50%，在阈值周围检测概率在 0～100%平滑分布。

临界可察觉电平差(JNLD)是电平差的检测阈值。JNLD 受输入电平的影响，对声音较小的信号，检测需要大差异（电平：20dBSPL，JNLD：0.75dB)；对声音较大的信号，对较小差异的敏感度要高一些（电平：80dBSPL，JNLD：0.2dB)。数据是根据调幅实验获得的，检测概率的原理掩蔽见图 B.3。

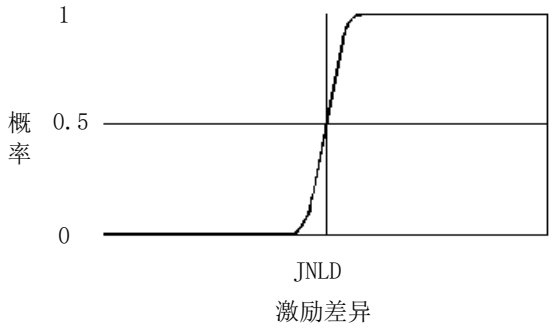


图 B.3　检测概率的原理掩蔽

B.2.6　掩蔽

单独播放时清晰可听的信号,在与其他信号同时播放时则可能完全听不见,该效应称之为掩蔽效应，掩蔽其他信号的信号称为掩蔽信号，被掩蔽的信号称为被掩蔽信号。以下两种情况必须要区分开：

——同时掩蔽：

这种情况下，掩蔽信号和被掩蔽信号同时出现且处于准稳态。如果掩蔽信号具有一个离散带宽，对于低于或高于掩蔽信号的频率，其听力阈值都会提高。掩蔽的数量取决于掩蔽信号和被掩蔽信号的结构。在类噪声信号掩蔽音调信号的情况下，掩蔽数量几乎与频率无关。如果被掩蔽信号的声压电平比掩蔽信号的电平低约5dB，那么被掩蔽信号就听不见了。如果情况对调，也就是音调信号掩蔽类噪声信号，掩蔽数量取决于掩蔽信号的频率。这可以通过公式$\left(15.5 + \frac{z}{Bark}\right)$ dB进行估算，其中z表示掩蔽信号的临界频带率。另外，在高信号电平处，非线性效应降低了掩蔽信号的掩蔽阈值。类似效应也出现在“音调掩蔽音调”的情况下。多信号的掩蔽阈值是非线性叠加的。总之，所得的掩蔽阈值比每个独立信号生成的掩蔽阈值要高。

——时间性掩蔽：

这种情况下，掩蔽信号和被掩蔽信号在不同的时间出现。在掩蔽信号衰减后不久的掩蔽阈值更接近于该掩蔽信号的同时掩蔽阈值，而不是绝对阈值。根据掩蔽信号的持续时间，阈值的衰减时间可以在5ms（掩蔽信号：持续0.05ms的高斯脉冲）到150ms之间（掩蔽信号：持续1s的粉红噪声）。正好处于强信号之前的弱信号将会被掩蔽，这种后向掩蔽效应的持续时间约为5ms。如果被掩蔽信号刚超过阈值，但在掩蔽信号出现之前它还未被感知到，它可能会被看做是掩蔽信号的变化。对不同的视听者，后向掩蔽的效果差异很大。

B.2.7　响度和部分掩蔽

音频信号的感知响度取决于它们的频率、持续时间和声压级。由于自动掩蔽，复杂信号的响度比其

组成部分各自响度相加的总和要低。在音频质量测量中，参考信号中所增加的多余失真的响度，也就是噪音响度，将随参考信号产生的部分掩蔽而降低。

B.2.8 尖锐度

尖锐度是感觉的一个基本值，与音质相关。如果声音主要包含高频组成部分，那么这个声音听起来就非常尖锐。例如，在高频处的正弦音或者带限噪声，或截止频率大于3kHz的高通噪声，都可以称之为尖锐。而音频信号的详细频率结构对尖锐度的影响不大。Von Bismarck在1974年做过与尖锐度相关的基础研究。

Aures在1984年做过另一个与尖锐度相关的研究。与Bismarck定义的加权函数相比，其研究结果给出了一个略加修正的加权函数。很低和很高的临界频带比率，对尖锐度属性贡献较小，但在临界频带比率在4Bark～20Bark间对尖锐度属性贡献较大。另外，这些研究表明，对于声压级剧烈变化和具有强高频内容的音频信号，其尖锐度不能仅由整体响度决定，还需要考虑一个加权函数，该加权函数取决于整体响度。

B.2.9 认知处理

感知音频质量受认知效应的强烈影响。这可以通过一个简单的实验进行阐述。

具有明显可闻背景噪声的参考信号由不能传输那些背景噪声的音频设备进行处理。因为噪音是不受欢迎的失真，在听音测试中，参考信号的评分等级将比被处理后的信号差。另一方面，如果参考信号最重要的部分是柔和的背景噪声，那么以相同方式处理后的信号的得分将更低。

要将所有认知效应列举出来并不是本文件的范畴，但本文件列举了以下示例。

示例1：

线性与非线性失真的分离：

线性失真没有非线性失真那样讨厌。采用输出信号的自适应逆滤波可以轻松地将线性失真从非线性失真中分离出来。本文中所指定的方法是从非线性失真中把线性失真分离出来。

示例2：

听觉场景分析：

由Bregman在1990年提出的听觉场景分析是一个认知过程，该过程允许听音者把不同的听觉事件分开，并将它们分成不同的目标组。Beerends和Stemerdink在1994年介绍了一个实际的方法，该方法对于量化听觉场景分析效应非常有用。如果时间频率组成部分没有进行编码，剩余信号仍然可以形成一个连贯的听觉场景，此时若引入一个新的不相干的时间频率组成部分，就会形成两个不同的感知。因为分成了两种不同的感知，比起新引进的失真在响度上引起的偏差来说，这种失真更加令人讨厌。这会导致感知干扰之间的不对称，主要是未对时间频率组成部分进行编码的部分和新引入的时间频率组成部分二者造成的。

示例3：

信息掩蔽：

信息掩蔽可以通过类熵频谱时间复杂性量度进行模型化。在主观评价前，这个效应很可能与评价人员的培训次数相关。该效应首个模型由Beerends等人在1996年提出，利用该模型对一个超过100ms的时间窗口的本地复杂性进行了评价。如果本地复杂性较高，时间窗口内的失真比本地复杂性低的情况更难听见。Leek and Watson在1984年提出，经过培训可以把掩蔽阈值降低几十个分贝（dB）。

示例4：

频谱时间加权：

音频信号中的频域和时域中携带了更多信息，因此可能比其他信号重要。研究发现频率和时域加权对语音编解码器的质量判定具有非常重要的作用。Beerends和Stemerdink在1994年提出，在语音中，频域-时域组成部分如共振峰，明显地比其他组成部分携带了更多信息。在音乐中，信号中的所有频域-时域组成部分即使无声也可能携带了信息。

附 录 C
（规范性）
应用

C.1 概述

本附录规定了感知音频质量客观测量方法所适用的主要应用及其技术要求。

部分应用需要实时的客观测量方法，部分应用则不需要实时测量。对于实时测量的方法，建议测量设备的最大延时应大于等于200ms，不允许超过1s。

此外，在线测量与离线测量存在不同的区别。在离线测量中，测量程序可以完全访问设备或连接，而在线测量则意味着程序正在运行，不能被测量所中断。

C.2 主要应用

C.2.1 实现方案的评价

广播公司、网络运营商或其他厂商在选购某款设备或者进行设备验收测试时，应要对不同的设备实现方案进行评价，特别是音频编解码器。

这类应用要求高准确性，特别是评价小损伤以及正确地对不同实现方案进行等级划分时。关注输出变量方面，对用户而言，简单的输出如ODG就足够了；但音频编解码器的开发者则关注模型输出变量，他们可以通过适当的模型输出变量集进行更加深入的分析。

对该应用，两个模型版本均适用，推荐采用高级版本。

C.2.2 感知质量的排序

这是一个快速程序，在设备或线路投入应用前进行。其目的是进行功能性检查和质量。测量设备将由操作人员控制。任何类型的失真都可能会出现。

对于这类应用，应进行实时测量，一般需要采用测试信号或预设音频信号。应提供并展现ODG值，且1s内至少应刷新两次；如果采用特殊测试信号，则在测试信号结束后直接给出ODG值。

该应用采用基础版本就足够了。

C.2.3 在线监测

这是一个在线连续过程，发生在音频在线传输的过程中。音频节目绝不能被测量程序打断。因此，测量的时候应采用节目信号本身或预先设定的音频片段作为测试信号。预先设定的音频片段可以是火车鸣笛声或叮当声。测量设备由操作人员控制。

对该应用，应进行实时测量。应提供并展现ODG值，1s内至少应刷新两次或者在预设信号结束后直接显示ODG。不需要展现MOV值。

该应用采用基础版本就足够了。

C.2.4 设备或连接状态

为确认音频连接或设备的功能性，需要不时地进行大量的质量检查。与在线监测和感知质量的排序应用不同，该应用需要对许多技术参数进行检查。

除了ODG值外，测量系统还应显示所有MOV值，通过这些参数，对设备或连接状态给感知音频质量造成的影响进行详细描述。该应用不需要实时测量。

该应用推荐使用高级版本。

C.2.5 编码器识别

为了识别解码器（不同算法或相同算法的不同实现方案），测量系统应具备存储、恢复以及比较各个模式特征的功能。各模式之间的相似性可以认为是不同编解码器实现方案的相似性的测量结果。因此该程序可用于识别特殊编解码器的实现方案与类型。

测量系统应尽可能多地记录与模式相关的信息。仅有ODG值无法提供足够的信息。

该应用推荐采用基础版本，无需实时测量。

注：推荐方法的实践经验极少。此外，用于测量模式之间的相似性的方法还没有确定。

C.2.6 编解码器开发

对这类应用，测量方法应尽可能准确且详细地描述被测设备的性能特性，特别是小损伤系统。

连续监测需要实时处理，但这并不是高级版本的必备功能；而微小劣化和详细信息则需要高级版本的支持。测量系统输出的显示速率与计算速率应相等，支持在4s内直接访问历史输出值。

对该类应用，建议使用高级版本。但对实时测量，采用基础版本就能满足要求。在测量过程中，应实现实时分析、非实时分析以及逐帧分析；可通过峰值等方式，显示任意严重的失真。理想的情况下，要求可以访问所有的MOV值。

C.2.7 网络规划

网络规划需要在规划过程中针对不同节点的预期质量进行评价。可以采用网络仿真软件，模拟不同音频处理阶段，以检查不同的配置，从而达到优化音频质量的目的。在后期，实际的音频处理组件可以按照所选参数进行配置并测试。

网络规划由系统工程师完成。工程师应遍历对音频质量产生影响的网络特性的详细信息。根据网络的具体应用，以适用MOV集为基础，对可能的网络配置进行等级划分。因此，该类应用不仅需要ODG，还需要MOV值。该类应用不需要实时评价。

针对这类应用，两个版本均适用，但推荐采用高级版本。

C.2.8 主观评价的辅助

客观测量方法可用于筛选主观听音评价的关键音频素材。所有的MOV值可用于关键材料的分类。

该类应用要求准确性最高，推荐采用高级版本。然而，为了减少用于选择关键材料的时间，也可进行实时测量。

C.2.9 应用类别的总结

上述主要应用对客观测量方法的要求应符合表C.1。

表 C.1　测量方法的要求

序号	应用	类别	实时	最小ROV[a] Hz	是否在线	模型版本
1	实现方案的评价	诊断	否	—	否	均可
2	感知质量的排序	操作	是/否	2	否	基础版本
3	在线监测	操作	是	2	是	基础版本
4	设备或连接状态	诊断	是/否	—	是/否	高级版本
5	编解码器识别	诊断	否	—	否	均可
6	编解码器开发	开发	是/否	—	否	均可
7	网络规划	开发	是/否	—	否	均可
8	主观评价辅助	开发	是/否	—	否	高级版本
[a] 输出值速率（单位为赫兹（Hz））。						

C.3　测试信号

C.3.1　概述

测试信号分为自然信号与合成信号两类。本文件中所列举的自然测试信号由音频质量主观评价中已使用的关键音频序列组成，来自于ITU-R以及其他组织。在传送点和测量点应获取该信号。因此，测量设备应具备存储功能。

合成信号是一个数学上的定义，根据控制方式可有多种多样。这些信号可以在传送点或测量点生成，测量设备不需要具备额外的存储功能。由于这类信号的本质，从主观上对其进行分级极其困难。因此，测量方法还未实现这些信号的主观验证。

C.3.2　自然测试信号的选择

在验证过程中用到的测试信号子集应符合表C.2。经过验证，确定了本文件的有效性。表中还给出了这些信号在低比特率编码中典型的失真类型。

表 C.2　测试信号的子集列表

序号	音频序列	文件名	备注
1	响板	cas	注1
2	单簧管	cla	注2
3	音棒	clv	注1
4	长笛	flu	注2
5	钟琴	glo	注1，注2，注5
6	大键琴	hrp	注1，注2，注4
7	定音鼓	ket	注1
8	马林巴琴	mar	注1
9	钢琴-舒伯特	pia	注2
10	定调管	pip	注4
11	Ry Cooder	ryc	注2，注4
12	萨克斯管	sax	注2

表C.2（续）

序号	音频序列	文件名	备注
13	风笛	sb1	注2，注4，注5
14	英文女声演说	sfe	注3
15	英文男生演说	sme	注3
16	德文男生演说	smg	注3
17	小军鼓	sna	注1
18	Mozart女高音	sop	注4
19	铃鼓	tam	注1
20	小号	tpt	注2
21	三角铁	tri	注1，注2，注5
22	大号	tub	注2
23	Susanne Vega	veg	注3，注4
24	木琴	xyl	注1，注2
注1：瞬态：前回声敏感，时域中的噪音模糊。 **注2**：音调结构：噪声敏感，粗糙。 **注3**：自然语音（音调部分和攻击的关键组合）：失真敏感，攻击模糊。 **注4**：复合声：强调被测设备。 **注5**：高带宽：强调被测设备，高频损失，被调节目的高频噪音。			

C.3.3 持续时间

自然测试信号的持续时间应与其用于听音评价的时间一致。持续时间通常为10s～20s。测试信号的关键部分（揭露大多数损伤）极有可能只占据持续时间中非常小的部分。

合成测试信号的持续时间应足够长，以给被测编解码器足够的压力。被测编解码器可能含有用于编码音频信号的缓冲区。考虑到这些缓冲区大小以及测量方法中的时间常数，一个序列中的每个测试条目的持续时间应大于500ms。持续时间可以规定这么短，是因为这些信号不会用于主观听音评价。

C.4 同步

测量过程中，被测信号和参考信号应时间同步。该原则适用于自然信号和合成信号。

C.5 版权问题

表C.2给出的测试信号的免费使用，仅限用于测试用途以及本文件规定的客观测量方法。

这些信号主要来自EBU(EBU SQAM 光盘)，需获得版权许可。

附 录 D
（规范性）
输出变量

D.1 简介

本文件规定的客观测量方法可对音频质量进行测量，并给出对应的感知音频质量值。测量方法将听觉系统的基本性质模型化，多个中间过程模拟了生理和心理声学效应。

中间过程的输出值可用于描述损伤的特性。这些参数称为模型输出变量（MOV）。在测量模型的最终阶段，MOV值组合生成一个输出值，该输出值与主观评价的结果相互对应。

D.2 模型输出变量

用于计算客观差异等级的MOV的描述见表D.1。下标A表示该值来自滤波器组模型，下标B表示该值来自FFT模型。客观差异等级可以仅从FFT部分进行预测（基础版本），也可以对FFT与滤波器组的组合进行预测（高级版本）。需要随时间取平均值。

表 D.1 模型输出变量的描述

模型输出变量	描述
$WinModDiff_B$	参考信号与被测信号间的调制（封装）中窗口平均差异
$AvgModDiff1_B$	平均调制差异
$AvgModDiff2_B$	平均调制差异，侧重于引入的调制和参考信号具有少量或无调制的地方的调制变化
$RmsModDiff_A$	调制差异的Rms值
$RmsMissingComponents_A$	丢失频率组件的噪音响度的Rms值（用于$RmsNoiseLoudAsym_A$）
$RmsNoiseLoud_B$	平局噪音响度的Rms值，侧重于引入的组件
$RmsNoiseLoudAsym_A$	$RmsNoiseLoud_A$+0.5$RmsMissingComponents_A$
$AvgLinDist_A$	平均线性失真的一种量度
$BandwidthRef_B$	参考信号的带宽
$BandwidthTest_B$	被测设备的输出信号的带宽
$TotNMR_B$	平均总噪声掩蔽比的对数
$RelDistFrames_B$	至少一个频带含有重大噪音组成部分的帧的相对分数
$AvgSegmNMR_B$	分段噪声掩蔽比的平均对数
$MFPD_B$	低通滤波后的最大检测概率
ADB_B	平均失真块（等同于帧），做为总失真与严重失真帧的数量的比值的对数
EHS_B	随时间而产生的谐波结构误差

D.3 基本音频质量

主观听音测试中最常见的参数是基本音频质量（BAQ）。BAQ可以用主观差异等级进行表示。在主观听音测试中，将被测信号的评价等级减去参考信号的评价等级得到SDG。主观差异等级通常是负数。本文件对应的输出参数称为ODG，可通过大量可靠的测试序列作为基础，把MOV映射到ODG。

ODG 是一个客观测量的参数，与主观感知质量对应。听音评价中，评价者的任务是对测试素材的 BAQ

进行评价，ODG 也是基本音频质量的一个量度。

D.4 编码余量

在未来可能比较有价值的另一个参数是编码余量。编码余量是描述不可感知损伤的一种方式。测试人员通过将损伤放大至可感知的范围测得主观编码余量（SCM），即编码余量描述了损伤的可闻阈值余量。

为了确定该阈值，在听音测试时，需要对损伤进行放大或者衰减。差异法是较适合的一种方法，通过将时间同步后的原始信号与编码信号的差异信号放大并叠加到原始信号。此时，可闻阈值的检测最好与一个强制选择方法一起执行。每个评价者通过发达或衰减损伤后，得到可闻阈值，通过对这些可闻阈值进行平均从而得到主观编码余量。编码余量为负数表示听得见损伤，若余量为正数则表示听不见损伤。与基本音频质量不同，编码余量表示了损伤从听不见变为听得见却不恼人的一个量度。Feiten在1997年发表的文章中描述了主观编码余量测量方法的定义和检验。

客观编码余量同样可以根据模型输出变量计算得到。但只有极少数用于主观编码余量的测试条目被评价过。本文件不提供根据模型输出值计算客观编码余量的方法。

D.5 用户要求

根据应用的不同，用户对于测量方法的输出变量的要求也不同。对某些应用，如附录 C 中介绍的应用 2、应用 3，测量也是操作程序的一部分。在这些情况下，考虑到用户可能对测量技术没有深入了解，测量的输出需要易读易懂。测量方法仅输出一个与感知音频质量相对应的值为最佳。

上述方式同样适用于其他应用，如应用 1 和 4。但对 5～8 这一类应用，建议包含更加复杂的输出值，有利于那些掌握测量方法中相关原理知识的用户进行深入分析。

附 录 E
（规范性）
模型补充说明

E.1 概述

根据 GY/T 298—2016，在听音测试中，通过对音频测试条目进行评价得到对应的 SDG 值，众多评价员的平均 SDG 值代表了该条目的主观质量。测试条目可能含有不同类型的音频失真，音频质量应对整个测试条目进行整体的考察。因此以物理测量为基础预测 SDG 需要一个准确的外耳周听觉系统模型。该系统应与音频质量评判的认知一致。

推荐的客观测量模型，将参考信号和被测信号进行相互比较，输出一系列模型输出变量。通过优化技术，将这些模型变量映射成客观差异等级。通过该优化技术，在一个足够大的数据集上使得 ODG 与其对应的平均 SDG 之间的均方差最小。

这里介绍了两个不同的模型：其一为基于 DFT 版本，可以用于实时监测；另一个版本为同时基于滤波器组和 DFT，该版本可提供更加准确的结果。基于 DFT 的版本称为基础版本，而后者则称为高级版本。

基础版本和高级版本的处理过程应符合图 E.1。

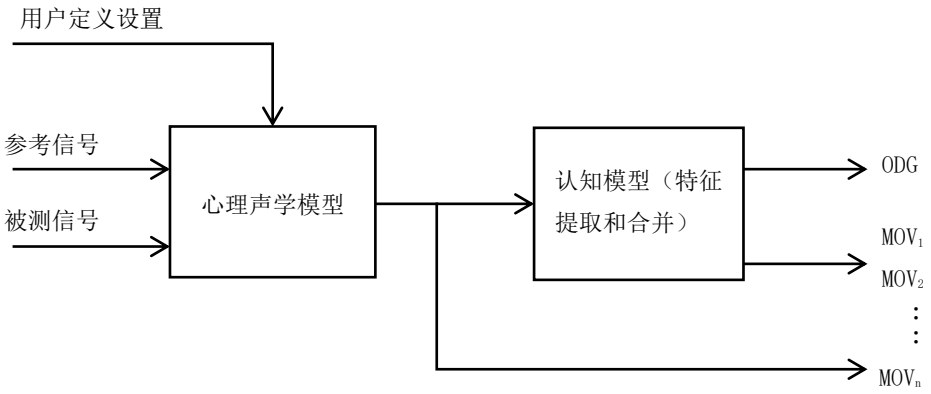


图 E.1 模型的处理过程

E.2 音频处理

与主观评价相同，测试信号的质量通过与参考信号进行比较得到。首先将参考信号和被测信号（单声道或立体声信号）转换成为心理声学的表示形式，再对这些表现形式进行比较，从而获得客观差异等级。处理过程见图 E.1。

E.3 用户定义设置

测量方法需要设定听音测试声压。现实中，该值可以是一个 1019.5Hz 的全幅正弦波形成的声压级（dB SPL），若确切的声压未知，则建议听音声压级设定为 92dB SPL。

E.4 心理声学模型

心理声学模型将时域信号的连续帧转换为基底膜表现方式。该过程结合使用 DFT 和滤波器组。DFT

将数据转换到频率域，并将数据从频率标度映射至音高标度，即心理声学的等效频率。模型的滤波器组部分，则通过带通滤波器的带宽和空间直接频率映射音高。

本文件中采用了两种不同的方法实现同时掩蔽。一些模型输出变量通过采用掩蔽阈值概念计算得到，而另一些变量则通过对内部表征进行比较而获得。第一方式通过使用心理及生理掩蔽函数，直接计算掩蔽阈值。模型输出变量以物理误差信号到该掩蔽阈值之间的距离为基础进行计算得到。通过对内部表征进行比较，将被测信号和参考信号的能量转换到比邻的音高域，以获得激励模式。模型输出变量以这些激励模式的比较为基础。非同步掩蔽通过在时间上抹除信号表征而实现。

绝对阈值由两个部分模拟计算得到，一部分通过应用频率加权函数进行模拟，另一部分通过对激励模式应用频率偏移实现。该阈值是最小可听声压的近似值。

心理声学模型的主要输出是激励模式和掩蔽阈值，均是时间和频率的函数。不同模型输出可用于后续处理。

E.5 认知模型

认知模型对心理声学模型生成的帧序列上的信息进行压缩。质量测量最重要的信息源是参考信号和被测信号在频域和音高域之间的差异。在频域，测试两者间频谱带宽以及谐波结构误差。而在音高域，误差测量主要是对激励包络调制和激励幅度进行测量计算。

所得的参数经过了加权，因此针对特殊音频失真，这些参数的最终计算结果即客观等级差异与主观差异等级十分接近。基础版本采用 11 个参数用于计算生成客观差异等级，而高级版本仅采用了 5 个参数。优化方案采用反向传播神经网络学习算法（见 10.6）。

参 考 文 献

[1] AURAS, W. [September, 1984] Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse. Dissertation an der Fakultät für Elektrotechnik der Technischen Universität München, Federal Republic of Germany.

[2] BEERENDS, J. G. and STEMERDINK, J. A. [December, 1992] A perceptual audio quality measure based on a psychoacoustic sound representation. J. Audio Eng. Soc., Vol. 40, p. 963-978.

[3] BEERENDS, J. G. and STEMERDINK, J. A. [February, 1994] Modeling a cognitive aspect in the measurement of the quality of music codecs. Contribution to the 96th AES Convention, preprint 3800. Amsterdam, Netherlands.

[4] BEERENDS, J. G. and STEMERDINK, J. A. [March, 1994] A perceptual speech quality measure based on a psycho-acoustic sound representation. J. Audio Eng. Soc., Vol. 42, p. 115-123.

[5] BEERENDS, J. G., van den BRINK, W. A. C. and RODGER, B. [May, 1996] The role of informational masking and perceptual streaming in the measurement of music codec quality. Contribution to the 100th AES Convention, preprint 4176. Copenhagen, Denmark.

[6] BRANDENBURG, K. [1987] Evaluation of quality for audio encoding at low bit rates. Contribution to the 82nd AES Convention, preprint 2433. London, United Kingdom.

[7] BREGMAN, A. S. [1990] Auditory Scene Analysis: The Perceptual Organisation of Sound. MIT Press, Cambridge MA, United States of America.

[8] COHEN, E. A. and FIELDER, L. D. [May, 1992] Determining noise criteria for recording environments. J. Audio Eng. Soc., Vol. 40, p. 384-402.

[9] COLOMES, C., LEVER, M., RAULT, J. B. and DEHERY, Y. F. [April, 1995] A perceptual model applied to audio bit-rate reduction. J.Audio Eng. Soc., Vol. 43, p. 233-240.

[10] FEITEN, B. [March, 1997] Measuring the Coding Margin of Perceptual Codecs with the Difference Signal. 102nd AES-Convention, preprint 4417. Munich, Federal Republic of Germany.

[11] GRUSEC, T., THIBAULT, L. and SOULODRE, G. [September, 1997] EIA/NRSC DAR systems subjective tests. Part 1: Audio codec quality. IEEE Transactions on Broadcasting, Vol. 43, 3.

[12] KARJALAINEN, J. [March, 1985] A new auditory model for the evaluation of sound quality of audio system. Proceedings of the ICASSP, p. 608-611. Tampa, Florida, United States of America.

[13] LEEK, M. R. and WATSON, C. S. [1984] Learning to detect auditory pattern components. J. Acoust. Soc. Am., Vol. 76, p. 1037-1044.

[14] MEARES, D. J. and KIM, S. W. [July, 1995] “NBC time/frequency module subjective tests: overall results”, ISO/IEC JTC 1/SC 29/WG 11 N0973 MPEG95/208.

[15] MOORE, B. C. [1986] Frequency Selectivity in Hearing. Academic Press, London, United Kingdom.

[16] MOORE, B. C. [1989] An introduction to the psychology of hearing. Academic Press, London, United Kingdom.

[17] PAILLARD, B., MABILLEAU, P., MORISETTE, S. and SOUMAGNE, J. [1992] Perceval: Perceptual evaluation of the quality of audio signals. J. Audio Eng. Soc., Vol. 40, p. 21-31.

[18] SCHROEDER, M. R., ATAL, B. S. and HALL, J. L. [December 1979] Optimizing digital speech coders by exploiting masking properties of the human ear. J. Acoust. Soc. Am., Vol. 66, p. 1647-1652.

[19] SOULODRE, G., GRUSEC, T., LAVOIE, M. and THIBAULT, L. [March 1998] Subjective evaluation of state-of-the-art 2-channel audio codecs. J. Audio Eng. Society.

[20] SPORER, T. [October 1997] Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio. 103rd AES-Convention, preprint 4512. New York, United States of America.

[21] TERHARDT, E. [1979] Calculating Virtual Pitch, Hearing Research. Vol. 1, p. 155-182.

[22] THIEDE, T. and KABOT, E. [1996] A New Perceptual Quality Measure for Bit Rate Reduced Audio. Contribution to the 100th AES Convention, preprint 4280. Copenhagen, Denmark.

[23] TREURNIET, W. C. [1996] Simulation of individual listeners with an auditory model. Proceedings of the Audio Engineering Society, Reprint Number 4154. Copenhagen, Denmark.

[24] von BISMARCK, G. [1974] Sharpness as an attribute of the timbre of steady sounds. Acustica, 30, p. 159-172.

[25] ZWICKER, E. and FASTL, H. [1990] Psycho-acoustics, Facts and Models. Berlin; Heidelberg: Springer Verlag, Federal Republic of Germany.

[26] ZWICKER, E. and FELDTKELLER, R. [1967] Das Ohr als Nachrichtenempfänger. Stuttgart: Hirzel Verlag, Federal Republic of Germany.