

# 基于汉语语音素材的响度算法误差研究

李德民, 祝培生\*, 路晓东, 陶皖琪, 周凯宇, 任锡洁

(大连理工大学 建筑与艺术学院, 辽宁 大连 116024)

**摘要:** 对两类在国际上得到广泛接受的响度模型即 Moore 和 Zwicker 时变响度模型及 ITU (国际电信联盟)标准推荐响度模型进行了对比研究. 通过耳机回放的主观听音实验评估了这两类响度模型在计算汉语语音素材总响度时的有效性. 模型的评价采用了 4 种不同的统计量, 对模型的绝对精度和相对精度进行了度量. 研究发现, ITU-R BS. 1770-4 推荐响度模型相对简单, 更易于对语音素材的总响度进行度量; 在计算汉语语音素材的总响度时, ITU 标准推荐响度模型比 Moore 和 Zwicker 时变响度模型的精度高, 提高了 0.5~2 dB.

**关键词:** 响度模型; 汉语语音素材; 总响度

**中图分类号:** TU112

**文献标识码:** A

**doi:** 10.7511/dllgxb201903008

## 0 引言

在声学中, 响度 (loudness) 是声压的主观感知. 声音的物理属性与感知响度的关系受物理、生理和心理因素共同影响. 越来越多的研究表明, 传统的 A 计权声级评价方法不能够准确反映人对响度的真实感知<sup>[1-3]</sup>. 由于语音素材和回放条件的差异, 使用传统的声压级度量方法预测响度的效果较差, 听音人会感觉到忽高忽低、不舒适的响度变化. 主观听音实验是声环境研究中最重要方法之一, 由于响度的偏差会对听觉感知产生很大的影响, 精准地控制和测量响度对实验的成功至关重要.

目前国际上广泛接受的响度模型有两类, 一类是基于 Zwicker<sup>[4]</sup> 和 Moore<sup>[5]</sup> 的时变响度模型, 另一类是基于 ITU-R BS. 1770-4 标准<sup>[6]</sup> 推荐的响度算法. 前者多应用于心理声学的研究, 后者多应用于数字化音频的等响度标准化处理. 自 1933 年 Fletcher 等<sup>[7]</sup> 提出了响度概念以来, 响度一直是心理学领域研究的热点. 其中 Zwicker 和 Moore 的时变响度模型都比较成功, 成为 2017 年的国际标准<sup>[4-5]</sup>. 有研究表明, Moore 等<sup>[8]</sup> 提出的以等矩形带宽 (equivalent rectangular bandwidth,

ERB) 来近似临界频带, 使得 Moore 响度具有高于 Zwicker 响度的分辨率; 近年来, 国际电信联盟 (ITU) 及欧洲广播联盟 (EBU) 相继提出了测量响度的新方法, 该方法广泛应用于电声领域的等响度控制<sup>[6]</sup>, 并成为了 ITU-R BS. 1770-4 标准. 美国、澳大利亚、日本等国和欧盟的标准都是基于 ITU 标准制定的 (如美国的 ATSC A/85、欧洲的 EBU R128、澳大利亚的 OP 59、日本的 TR-B32). 目前国内还没有对应于 ITU-R BS. 1770-4 的标准, 仅有局限于电视、音频等领域的少量研究. 文献<sup>[9-10]</sup> 对 ITU-R BS. 1770 中的核心算法进行了解读, 但近些年新标准又进行了部分修订<sup>[11]</sup>. 文献<sup>[12]</sup> 简要地介绍了 ITU-R BS. 1770 中的响度参数, 但缺少进一步更系统的研究.

Skovenborg 等<sup>[13]</sup> 对比了 12 种不同响度模型在估计语音和音乐素材响度方面的差别并提出了一种新的响度算法, 但该算法与 ITU-R BS. 1770-4 推荐的算法相比, 精度没有显著提高, 其实验细节也有改进的空间. Zorilă 等<sup>[14]</sup> 使用耳机回放实验的方法研究了 Moore 时变响度模型在等响度修正句子方面的有效性, 证明了即使 RMS 声级一样的句子信号, 频谱动态范围不同响度也不同. 这些研究结果也表明, 响度的评价是非常复杂的.

Moore 和 Zwicker 的时变响度模型, 算法原理相似, 都是基于心理声学的研究; ITU-R BS. 1770-4 推荐的响度模型, 基于绝对、相对阈限的频率计权方法, 似乎只是一种纯粹的、简单的算法. 这两类模型原理完全不同, 本文首先对上述两类响度模型的原理给予概述, 然后根据汉语语言特点设计基于耳机回放的响度匹配实验, 最后对两类响度模型的误差进行评估.

## 1 响度计算模型

### 1.1 Moore 和 Zwicker 时变响度模型

Moore 和 Zwicker 时变响度模型是建立在对人耳听觉特性的深入理解上, 考虑了临界频率、频域掩蔽、时间效应等影响因素. Moore 时变响度模型是对 Zwicker 时变响度模型进行改进后得到的, 基于解析式的计算方法使得 Moore 时变响度模型具有更高的计算精度. Zwicker 和 Moore 时变响度模型基于短时傅里叶变换, 善于表现响度随时间变化的细节.

#### 1.1.1 Zwicker 时变响度模型及其计算步骤

利用短时傅里叶变换, 取样本帧长为 24 ms, 帧移为 2 ms. 计算过程为: ①将以相同时间窗截取的语音信号段分别通过 1/3 倍频程滤波器组, 计算功率谱与密度; ②将声音频率划分为 24 个特征频带, 单位是 Bark, 用来计算频域的掩蔽; ③根据特征频带带宽对低频频带进行合并; ④模拟外耳中耳传递函数并确定声场类型(自由场或混响场); ⑤计算各临界频带特征响度; ⑥通过积分特征响度曲线, 得到瞬时响度; ⑦根据瞬时响度计算短时响度  $Z_{stl}$  (Zwicker short-term loudness) 和长时响度  $Z_{ltl}$  (Zwicker long-term loudness)<sup>[4]</sup>.

1.1.2 Moore 时变响度模型及其计算步骤 为了模拟人耳对低频的不敏感特性, Moore 时变响度模型计算激励曲线是基于多分辨率的傅里叶变换频谱分析. 对于高频使用较短的窗长(2 ms), 对于低频使用较长的窗长(64 ms). Moore 时变响度模型以 1 ms 的帧移计算, 以等矩形带宽近似临界频带. 激励模式的计算基于人耳基底膜的频率响应(roex 听觉滤波器). 将激励曲线转化为特征响度后, 特征响度曲线下的面积就是一段时间间隔内的瞬时响度; 根据瞬时响度计算短时响度 (Moore short-term loudness) 和长时响度 (Moore long-term loudness)<sup>[5-7]</sup>.

Moore 时变响应模型模拟了 3 种不同的情

况, 建立了 3 个采用 4 097 个点的 FIR 滤波器来模拟外耳中耳传递函数<sup>[5]</sup>. 本文使用耳机回放进行主观响度听音测试, 因此响度模型适用于扩散声场、中耳传递函数条件<sup>[15]</sup>.

1.1.3 总响度表示方法 Moore 和 Zwicker 时变响度模型可以根据声音信号的长度表现响度随时间变化的细节. 由于听众能容易地判断时变声音的整体响度, 为了表示一段声音信号的总响度, 不同的指标被开发出来, 用单值表示整体的声音响度. 研究者<sup>[7-8, 15]</sup>认为, 随着时间的推移, 整体响度要高于响度的平均值. 在 ISO 532-1<sup>[4]</sup> 和 DIN 45631/A1<sup>[16]</sup> 中建议使用百分数响度. 但有研究<sup>[17-18]</sup>建议使用均值响度, 认为均值响度比使用百分数响度在评价时变声音总响度方面有更好的表现. 也有研究<sup>[14]</sup>建议使用峰值响度, 认为响度的峰值更能代表时变信号的总响度.

本研究根据查阅的文献和标准的建议, 对于 Moore 时变响度模型使用峰值响度  $M_{peak}$  (Moore peak loudness) 和均值响度  $M_{mean}$  (Moore mean loudness) 表示一段声音的总响度, 对于 Zwicker 时变响度模型使用百分数响度  $Z_{95\%}$  (Zwicker 95th percentile loudness) 和均值响度  $Z_{mean}$  (Zwicker mean loudness).

### 1.2 ITU-R BS. 1770-4 推荐响度模型

ITU-R BS. 1770-4 推荐响度模型是基于绝对阈限和相对阈限的 K 计权方法, 适用于多种语音素材. 设置阈限的目的是保证参与响度计算的部分为有效部分, 与上述基于心理声学的时变响度模型不同的是, 它似乎只是一种纯粹的计算方法. 该模型多用于数字化音频的响度标准化处理, 在计算多种音频信号总响度方面得到了广泛应用.

ITU-R BS. 1770-4 推荐响度模型计算过程为: ①将声音信号通过 K 计权滤波器, 该滤波器被设计为两级子滤波器. 第一级滤波器的增益(图 1)从 1 kHz 到 3 kHz 内逐渐提升至 4 dB, 3 kHz 以上保持 4 dB 的增益不变, 以模拟人头部的声学效应. 第二级为修正低频 B 曲线 (RLB-revised low frequency B curve) 滤波器, 该滤波曲线(图 2)主要模拟人耳对低频声音的不敏感性. ②计算每个声道信号的均方值. ③根据测量间隔分帧并计算每一帧内的响度(瞬时响度). ④将低于阈限的瞬时响度去除, 计算总响度. 将阈限分成绝对阈限和相对阈限, 绝对阈限设置为 -70 LKFS

(loudness units, K-weighted, relative to full scale, LKFS 这一单位等价于 dB), 将低于绝对阈限的部分去除; 相对阈限设置为  $-10$  dB, 即在计算绝对阈限的基础上, 再将低于当前响度  $10$  LKFS 的部分去除. 设置绝对阈限和相对阈限的目的是将声音中静音部分和本底噪声去除, 以保证参与总响度计算的部分为有效响度.

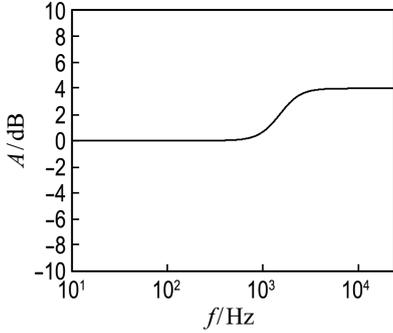


图1 考虑了头部声学效应的第一级滤波器  
Fig. 1 The first stage filter used to account for the acoustic effects of the head

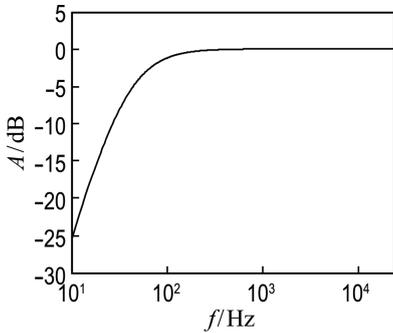


图2 第二级计权曲线  
Fig. 2 The second stage weighting curve

滤波后信号的均方值  $z_i$  由下式计算:

$$z_i = \frac{1}{T} \int_0^T y_i^2(t) dt \quad (1)$$

式中:  $y_i$  是输入信号(由上述两级滤波器滤波后).  $i \in I, I = \{L, R, C, L_s, R_s\}$ , 是输入通道的集合,  $L, R, C, L_s, R_s$  表示左、右、中、左后、右后.  $T$  是时间间隔.

K 计权滤波后测量间隔  $T$  内的响度  $L(K)$  被定义为

$$L(K) = -0.691 + 10 \lg \sum_i G_i z_i \quad (2)$$

其中  $G_i$  是不同声道的权重系数, 前置声道  $G_i$  等于 1, 后置声道  $G_i$  等于 1.41, 相当于增加了 1.5 dB, 表示人耳对后置声道的声音更敏感.

为了计算有效响度, 需要将信号进行分帧, 逐一计算每帧的响度(瞬时响度). 然后设置绝对阈限和相对阈限, 将小于阈限部分的值(瞬时响度)去除, 不参与计算. 设置帧长  $T_g = 400$  ms, 帧移  $s = \frac{1}{4} T_g$ ,  $z_{ij}$  按下式计算:

$$z_{ij} = \frac{1}{T_g} \int_{T_g j s}^{T_g(j+1)s} y_i^2(t) dt \quad (3)$$

其中  $j \in \left\{0, 1, 2, \dots, \frac{T - T_g}{T_g s}\right\}$ .

瞬时响度  $l_j$  按下式计算:

$$l_j = -0.691 + 10 \lg \sum_i G_i z_{ij} \quad (4)$$

K 计权滤波后, 去除静音和本底噪声部分的有效响度  $L_{eq}(K)$  按下式计算:

$$L_{eq}(K) = -0.691 + 10 \lg \sum_i G_i \left( \frac{1}{J_g} \sum_1^{J_g} z_{ij} \right) \quad (5)$$

其中  $J_g$  是瞬时响度中大于绝对阈限( $-70$  LKFS)和相对阈限( $-10$  dB)的数量.

### 1.3 与其他滤波曲线的比较

A 计权(A-weighted)声级评价标准在国际标准中是最常用的一种声级评价方法<sup>[19]</sup>, 反映了人耳感知的相对响度和对低频的不敏感性. A 计权是基于 40 方 Fletcher-Munson 等响曲线<sup>[19]</sup>, 因此 A 计权声压级适合安静(严格讲应该是 40 方以下的响度级)的声环境.

Robinson 于 2001 年提出了响度修正的 RG (Reply Gain) 算法. 该算法基于 80 方 Fletcher-Munson 等响曲线<sup>[19]</sup>, 取得了较好的响度预测效果, 被运用于数字化音频响度的标准化处理. 本文为了与 A 计权滤波器进行对比, 使用 RG 算法中的 80 方等响曲线滤波器, 借鉴 RG 算法的优点, 希望能够在算法上得到改进.

在图 3 中将本文使用的计权方法进行了对比. 从图中可以看出 RG 算法的计权曲线比上文提到的 K 计权和 A 计权曲线都要复杂. 各滤波曲线都表达了人耳对低频的不敏感, RG 算法的滤波曲线和 K 计权滤波曲线都表达了 3 kHz 处外耳道对声音的共振作用.

本文使用 4 种不同的滤波曲线(图 3), 根据上文中的响度算法原理, 编写了 Matlab 程序. 根据声音信号长度将音节信号设置帧长 4 ms, 帧移 1 ms, 句子信号设置帧长 400 ms, 帧移 100 ms.

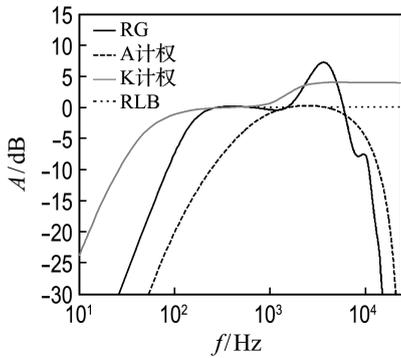


图 3 不同计权曲线的比较

Fig. 3 Comparison of different weighting curves

### 1.4 多段声音信号的等响度修正方法

使用下面的方法将多段声音信号进行等响度修正. 首先将音节或句子等语音信号经过图 3 中的滤波曲线进行滤波(A、K、RG、RLB 滤波曲线). 按式(5)计算信号间的有效响度  $L_{eq}(N)$ ,  $N$  表示所使用的不同滤波曲线. 总响度修正量  $l(N)$  按下式计算:

$$l(N) = L_m(N) - L_{eq}(N) \quad (6)$$

其中  $L_m(N)$  是信号间的总响度平均值.

接下来计算各信号的响度修正系数:

$$h(N) = 10^{\frac{l(N)}{20}} \quad (7)$$

将信号依次使用式(8)的方法进行修正, 得到修正后的信号:

$$Y(N) = yh(N) \quad (8)$$

其中  $y$  是输入信号;  $Y(N)$  是使用不同的方法, 将总响度进行修正后的输出信号.

## 2 基于耳机回放的响度听音实验

进行了基于耳机回放的响度匹配实验, 统计不同算法的响度预测误差, 得到以 dB 为单位的 4 种统计量, 以对比分析不同响度模型的适用性. 为了避免耳机、耳郭耦合(耳道传输函数  $H_pTF$ )的影响, 将测试信号进行了  $H_pTF$  的逆滤波均衡<sup>[20]</sup>.

### 2.1 耳机回放实验的有效性

在过去的几十年里, 耳机和扬声器之间的响度感知差异常常让声学家们感到困惑: 耳机的声压级往往要提高多达 10 dB 才能达到与扬声器相同的响度. 这种所谓的“缺失的 6 dB”问题引发了许多猜测, 很多人甚至对耳朵是纯声压探测器的假设提出了质疑<sup>[21-22]</sup>.

但有研究表明<sup>[23-24]</sup>, “缺失的 6 dB”来自于各种假象以及被忽略的细节, “6 dB”并没有真正缺

失. 最小可听阈的差异是由于封闭式耳机容易引起堵耳效应, 在耳道内放大的生理噪声掩蔽了低频, 使耳机听音的最小可听阈提高<sup>[25]</sup>.

耳机与扬声器听音在远高于阈值时的响度差异似乎更难以解释, 生理噪声的掩蔽此时已经不起作用. Rudmose<sup>[23]</sup>的研究认为其主要是由一种心理声学现象造成的, 即在近处的声源需要发出更大的声压级才能达到远处声源相同的响度. 用 Rudmose 的话来说, 远处的声源有“更大的声学大小”, 因此显得更响. 实验对象一旦认识和观察到了这种现象, 可以通过训练来消除.

因此, 在不考虑心理声学现象影响的情况下, 人的耳朵可以看成是单纯的声压探测器<sup>[23]</sup>. 本文使用耳机回放的方法进行响度匹配实验避免了头转动方向(头相关传递函数)、房间混响、心理声学现象等不确定因素的影响, 相比于声场聆听具有更高的实验精度.

### 2.2 使用的语音信号

由于汉语中音节是常听到的最自然的语音单位, 本文根据汉语语音特点, 采用了两种语音信号进行实验, 一种是单音音节, 一种是句子.

单音音节使用的是 GB/T 15508—1995<sup>[25]</sup>规定的语音平衡音节表, 由一位男播音员在消声室中以大约每秒 4 个音节的语速进行录制的, 共 80 个. 音节的长度在 150~400 ms, 单声道, 采样率是 48 000 Hz, 位数为 32 位.

句子使用的是 GSBM. 6001—1989<sup>[26]</sup>规定的标准测试语音信号, 共 80 个句子, 单声道, 采样率是 44 100 Hz, 位数为 16 位.

### 2.3 受试者

一共 10 位听力正常的汉语为母语受试者参与了本文的实验, 年龄 18~25 岁, 听力测试的阈值均小于 20 dBHL. 正式实验之前先进行预实验, 确保受试者熟悉实验的全部流程.

### 2.4 实验设备

录音系统: 传声器 B&K 4189 (供电系统为 B&K 1704)、声卡 B&K ZE-0948、录音软件 Audition 3.0.

耳机回放系统: 耳机 Sennheiser HD 650 (开放式)、声卡 B&K ZE-0948、耳机功放 Rane-HC4s、回放软件使用在 Matlab 平台编程实现的 GUI 程序.

### 2.5 响度匹配实验

2.5.1 实验方法及其步骤 主观响度匹配实

验<sup>[27]</sup>共有 10 名受试者参与. 为此专门编写了 Matlab GUI 程序, 受试者在计算机前操作鼠标和键盘完成实验.

实验流程如图 4 所示, 实验前将等响度均衡后测试信号的声压级调整为很宽的动态范围(测试信号与参考信号的声压级差在 -15~20 dB 任意取值). 程序将 8 段测试信号使用耳机

(Sennheiser HD 650 开放式) 传送给受试者. 受试者分别将测试信号的响度与参考信号的响度进行比较, 然后通过程序调整测试信号的声压级(调整步长 0.25 dB), 直到与参考信号的主观响度相匹配. 最后统计受试者的调整声压级  $S_i$  (以 dB 为单位), 即主观感知响度. 参考信号的声压级固定, 耳机回放的声压级约为 60 dB(A).

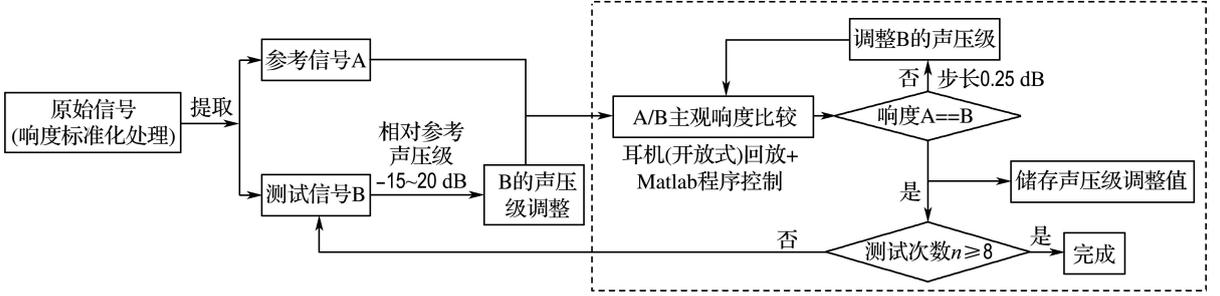


图 4 响度匹配实验程序框图

Fig. 4 A block diagram of loudness-matching experiments

一共对 160 个汉语音节和句子信号(80 个音节信号和 80 个句子信号)进行了响度匹配实验. 每次实验每位受试者需要进行 8 组听音(音节和句子×4 种方法), 也就是需要进行 8 次成对比较的响度匹配实验. 每次实验用时大约 30 min, 分 4 d 完成.

2.5.2 误差统计量 响度估算误差  $E_i$  按下式计算:

$$E_i = M_i - S_i \quad (9)$$

其中  $M_i$  是测试信号与参考信号的相对声压级差(也是算法预测响度),  $S_i$  是受试者的声压级调整值(也是主观感知响度).

基于响度估算误差  $E_i$ , 下面 4 种统计量被计算出来, 用来评估不同算法预测响度的能力.

平均绝对误差(average absolute error)是误差绝对值的平均, 表示估算误差的集中趋势, 按下式计算:

$$E_{aac} = \frac{1}{n} \sum_{i=1}^n |E_i| \quad (10)$$

绝对标准偏差(absolute standard deviation)是绝对误差的标准偏差, 表示估计误差的离散情况, 使用下式表示:

$$E_{asd} = \sqrt{\frac{1}{n} \sum_{i=1}^n (|E_i| - E_{aac})^2} \quad (11)$$

均方根误差(root mean square error)是误差

的均方根值, 表示误差估计的有效值, 使用下式表示:

$$E_{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^n E_i^2} \quad (12)$$

95% 绝对误差(95th percentile absolute error), 表示有 95% 的绝对误差小于  $E_{p95ae}$ .

2.5.3 实验结果及分析 图 5 和 6 给出了算法预测响度和主观感知响度之间的线性关系, 图中直线斜率是 1, 截距是 0, 散点与直线越接近表明算法预测响度与主观感知响度越接近, 也就表明算法的响度预测精度越高. 其中圆圈代表音节材料响度匹配结果, 加号代表句子材料响度匹配结果.

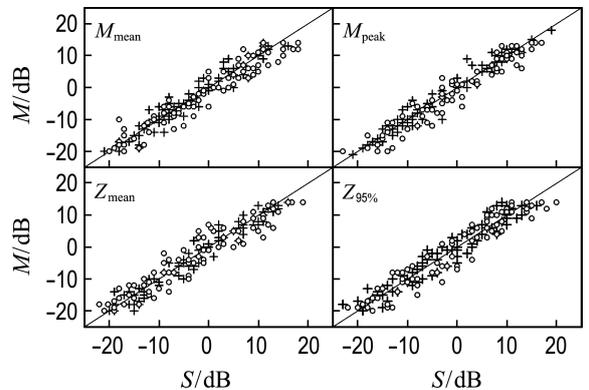


图 5 时变响度模型

Fig. 5 Time-varying loudness models

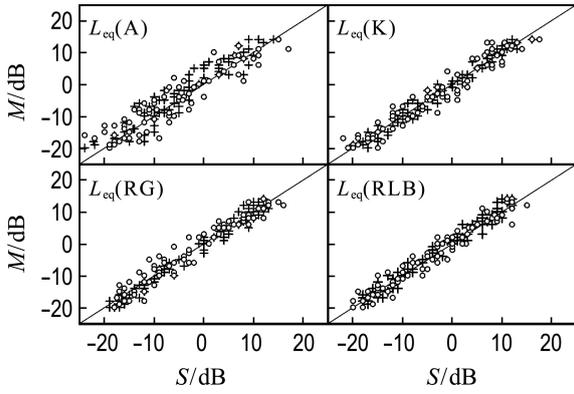


图 6 ITU-R BS. 1770-4 推荐的响度模型

Fig. 6 Recommended loudness model of ITU-R BS. 1770-4

表 1 显示了不同算法的确定系数  $R^2$  和相关系数  $r$ , 确定系数  $R^2$  测度了回归直线对观测数据的拟合程度. 若所有观测点都落在直线上, 残差平方和等于零, 确定系数  $R^2$  等于 1, 表示拟合是完全的. 如果  $y$  的变化与  $x$  无关,  $x$  完全无助于解释  $y$  的变化, 则确定系数等于 0. 相关系数  $r$  是确定系数的平方根. 从表中可以看出  $L_{eq}(RLB)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RG)$  的确定系数  $R^2$  比  $Z_{mean}$ 、 $Z_{95\%}$ 、 $M_{mean}$ 、 $M_{peak}$  的要高,  $L_{eq}(A)$  的确定系数  $R^2$  最小. 本文实验得到的相关系数  $r$  与文献[6]中实验得到的相关系数非常接近, 差值小于 0.01.

表 2 列出了不同算法的误差统计量并与文献[13]中的部分数据(括号中数据)对比. 由于使用

的听音信号不同(本文使用的是语音信号, 而文献[13]使用的是语音和音乐信号), 实验数据存在 1~2 dB 的差异. 从表中可以看出  $L_{eq}(RG)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RLB)$  方法在预测汉语语音素材响度方面表现很好, 对于句子响度的预测, 均方根误差  $E_{rmse}$  约为 1.8 dB, 平均绝对误差  $E_{aac}$  约为 1.6 dB, 绝对标准偏差  $E_{asd}$  约为 1.4 dB, 95% 绝对误差  $E_{p95ae}$  约为 2.5 dB, 响度的误差在主观上基本感觉不到. 对于音节响度的预测, 均方根误差  $E_{rmse}$  约为 2.7 dB, 平均绝对误差  $E_{aac}$  约为 2.1 dB, 绝对标准偏差  $E_{asd}$  约为 1.6 dB, 95% 绝对误差  $E_{p95ae}$  约为 4.5 dB, 响度的误差在主观上很难察觉到.

表 1 确定系数  $R^2$  与相关系数  $r$

Tab. 1 Determination coefficient  $R^2$  and correlation coefficient  $r$

	$R^2$		$r$	
	音节	句子	音节	句子
$L_{eq}(A)$	0.819	0.877	0.905	0.936
$L_{eq}(RG)$	0.932	0.960	0.965	0.980
$L_{eq}(K)$	0.942	0.967	0.971	0.983
$L_{eq}(RLB)$	0.941	0.920	0.970	0.959
$M_{mean}$	0.897	0.912	0.947	0.955
$M_{peak}$	0.916	0.924	0.957	0.961
$Z_{mean}$	0.881	0.910	0.939	0.954
$Z_{95\%}$	0.907	0.930	0.953	0.964

表 2 8 种不同响度量方法的误差估计

Tab. 2 Deviation meters for eight loudness performance metrics

响度	$E_{aac}$			$E_{asd}$			$E_{rmse}$			$E_{p95ae}$			dB
	音节	句子	平均	音节	句子	平均	音节	句子	平均	音节	句子	平均	
$L_{eq}(A)$	3.02	2.53	2.78(1.85)	2.43	1.80	2.12	3.89	3.12	3.51(2.29)	6.50	5.25	5.88(3.83)	
$L_{eq}(RG)$	2.14	1.30	1.72	1.64	1.35	1.50	2.69	1.43	2.06	4.50	2.00	3.25	
$L_{eq}(K)$	1.84	1.60	1.72	1.57	1.27	1.42	2.42	1.50	1.96	4.25	2.50	3.38	
$L_{eq}(RLB)$	1.78	1.35	1.57(0.855)	1.54	1.09	1.32	2.36	1.78	2.07(1.08)	4.00	2.50	3.25(2.17)	
$M_{mean}$	2.59	1.41	2.00	1.89	1.12	1.51	3.20	2.10	2.65	6.00	4.00	5.00	
$M_{peak}$	2.32	1.78	2.05	1.79	1.19	1.49	2.90	2.07	2.49	5.75	4.25	5.00	
$Z_{mean}$	2.78	1.48	2.13	2.02	1.31	1.67	3.43	2.12	2.78	6.00	4.50	5.25	
$Z_{95\%}$	2.81	2.05	2.43(1.68)	1.87	1.19	1.53	3.37	1.72	2.55(2.08)	5.50	4.00	4.75(3.60)	

从表 2 可以看出以  $Z_{mean}$ 、 $Z_{95\%}$ 、 $M_{mean}$ 、 $M_{peak}$  4 种方法为代表的时变响度模型, 对于句子的响度预测, 均方根误差  $E_{rmse}$  约为 2.1 dB, 平均绝对误差  $E_{aac}$  约为 2.0 dB, 绝对标准偏差  $E_{asd}$  约为

1.3 dB, 95% 绝对误差  $E_{p95ae}$  约为 4.5 dB, 响度的误差在主观上很难感觉到. 对于音节响度的预测, 均方根误差  $E_{rmse}$  约为 3.4 dB, 平均绝对误差  $E_{aac}$  约为 2.8 dB, 绝对标准偏差  $E_{asd}$  约为 2.0 dB, 95%

绝对误差  $E_{p95ae}$  约为 6.0 dB, 响度的误差较大, 在主观上能较为明显地察觉到. 将上述 4 种度量方法与  $L_{eq}(RG)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RLB)$  方法进行比较: 综合 4 种误差统计量的实验结果, 在进行汉语语音素材总响度计算时, 使用 ITU 标准推荐响度模型将会比 Moore 和 Zwicker 时变响度模型的计算精度提高 0.5~2 dB.

由表 2 可以看出,  $L_{eq}(A)$  方法的均方根误差  $E_{rmse}$  约为 3.5 dB, 平均绝对误差  $E_{aac}$  约为 3 dB, 95% 绝对误差  $E_{p95ae}$  约为 6 dB, 响度的误差在主观上能够较为明显地感知. 使用  $L_{eq}(A)$  进行等响度修正, 误差的最大范围可能在 5~7 dB. 将  $L_{eq}(RG)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RLB)$  3 种方法与  $L_{eq}(A)$  方法进行比较: 不同的计权方法可能会对总响度的计算带来较大影响, 使用 ITU-R BS. 1770-4 推荐的计权方法可能比传统的 A 计权方法的计算精度提高接近 1 倍.

本文使用方差分析的方法, 比较了几种方法在评价汉语音节和句子的主观响度方面是否存在显著性差异. 检验结果  $p < 0.001$ . 经过两两比较发现, ITU-R BS. 1770-4 推荐的响度模型与 Moore 和 Zwicker 时变响度模型在预测汉语语音素材响度方面存在差异, 而两类模型中不同算法之间没有显著性差异.

图 7 是不同响度量度方法的绝对误差分布图, 从图中可以看出 ITU-R BS. 1770-4 推荐的响度模型与 Moore 和 Zwicker 时变响度模型的误

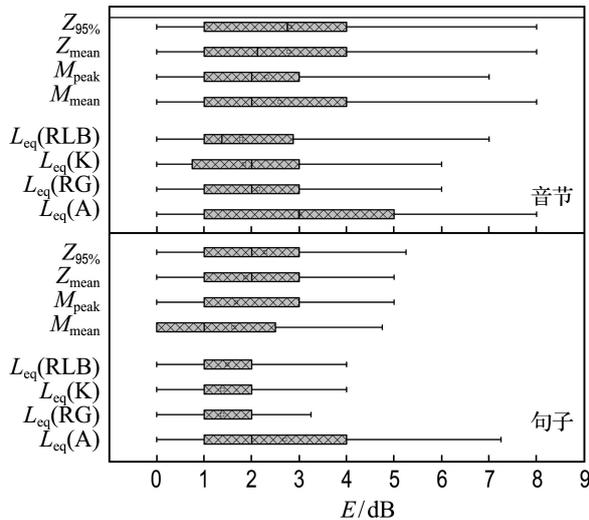


图 7 8 种不同的响度量度方法的绝对误差

Fig. 7 Absolute error of eight loudness performance metrics

差分布情况. 对比图 3 可知, RLB 和 K 计权滤波器之间的差异很小, 实验也证明了响度预测精度没有太大的差异. 对比  $L_{eq}(RG)$  和  $L_{eq}(A)$  的数据结果可知, 虽然 A 计权曲线与 RG 曲线在低频和中高频都有较大差别, 但是对响度预测精度起作用的主要是低频部分.

从表 2 和图 7 中可以明显看出, 音节似乎比句子的响度更难以准确度量, 可能主要有两方面的原因: 一方面是因为音节比句子更难以感知响度的差异, 因为音节信号的长度短, 感知响度的时间短; 另一方面是因为音节信号的频谱动态范围变化更大, 增加了算法预测的难度.

将不同的模型根据计算精度进行分级,  $L_{eq}(RLB)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RG)$  3 种方法的计算精度最高;  $Z_{mean}$ 、 $Z_{95\%}$ 、 $M_{mean}$ 、 $M_{peak}$  之间并没有显著的差异, 计算精度次之;  $L_{eq}(A)$  方法的计算精度最低.

### 3 结 语

本文研究了两类在国际上得到广泛认可的响度模型, 对比了它们在计算汉语语音素材总响度方面的有效性和误差.

经过研究发现, Moore 和 Zwicker 时变响度模型算法较复杂, 但在预测汉语语音素材总响度方面的可靠性并不是最高的, 原因可能是 Moore 和 Zwicker 时变响度模型使用合成信号和稳态信号来开发和验证, 在计算非稳态信号的响度时, 引入短时傅里叶变换, 总响度使用均值响度、峰值响度或者百分数响度来表示, 因此该算法更善于表现响度随时间变化的细节, 而 ITU-R BS. 1770-4 推荐的响度模型的一个关键优点是它的简单性, 允许以非常低的成本实现, 且更易于对语音的总响度进行度量.

$L_{eq}(RG)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RLB)$  方法在预测汉语语音素材的响度方面表现较好, 响度计算误差的有效值  $E_{rmse}$  均小于 2 dB, 误差平均值  $E_{aac}$  均小于 1.7 dB, 绝对标准偏差  $E_{asdl}$  均小于 1.5 dB, 95% 绝对误差  $E_{p95ae}$  接近于 3 dB, 响度误差在主观上基本感觉不到. 对于音节和句子,  $L_{eq}(A)$  方法的  $E_{rmse}$  和  $E_{aac}$  都接近 3 dB,  $E_{p95ae}$  分别为 6.50 和 5.25 dB, 响度的误差在主观上能够较为明显地感知, 使用  $L_{eq}(A)$  进行等响度修正, 误差的最大范围在 5~7 dB.

虽然  $L_{eq}(RG)$ 、 $L_{eq}(K)$ 、 $L_{eq}(RLB)$  3 种方法都取得了较好的响度预测效果,但  $L_{eq}(RLB)$  方法未能表达 3 kHz 处外耳道对声音的共振作用,因此  $L_{eq}(K)$  和  $L_{eq}(RG)$  方法对含有高频成分更多的音频素材(例如音乐)可能会有更好的表现.综合 4 种统计量的实验结果,在进行汉语语音素材总响度计算时,使用 ITU 标准推荐响度模型将会比 Moore 和 Zwicker 时变响度模型的计算精度提高 0.5~2 dB.不同的计权方法可能会对总响度的计算带来较大影响,使用 ITU-R BS. 1770-4 推荐的计权方法可能比传统的 A 计权方法的计算精度提高接近 1 倍.

## 参考文献:

- [1] 徐中明,周小林,张芳,等. Moore 响度在车内噪声分析中的应用 [J]. 振动与冲击, 2013, **32**(1): 169-173.  
XU Zhongming, ZHOU Xiaolin, ZHANG Fang, *et al.* Application of Moore loudness in interior noise analysis [J]. **Journal of Vibration and Shock**, 2013, **32**(1):169-173. (in Chinese)
- [2] CHARBONNEAU J, NOVAK C, GASPARD R, *et al.* A-weighting the equal loudness contours [J]. **Journal of the Acoustical Society of America**, 2012, **131**(4):3502.
- [3] 毛东兴. 响度感知特征研究进展 [J]. 声学技术, 2009, **28**(6):693-696.  
MAO Dongxing. Recent progress in hearing perception of loudness [J]. **Technical Acoustics**, 2009, **28**(6):693-696. (in Chinese)
- [4] ISO/TC 43 Acoustics. Acoustics — Methods for Calculating Loudness — Part 1: Zwicker Method: ISO 532-1: 2017 [S]. Geneva: ISO, 2017.
- [5] ISO/TC 43 Acoustics. Acoustics — Methods for Calculating Loudness — Part 2: Moore-Glasberg Method: ISO 532-2: 2017 [S]. Geneva: ISO, 2017.
- [6] International Telecommunication Union. Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level: ITU-R BS. 1770-4 [S]. Geneva: ITU, 2017.
- [7] FLETCHER H, MUNSON W A. Loudness, its definition, measurement and calculation [J]. **Journal of the Acoustical Society of America**, 1933, **5**(2):377-430.
- [8] MOORE B C, GLASBERG B R, VARATHANATHAN A, *et al.* A loudness model for time-varying sounds incorporating binaural inhibition [J]. **Trends in Hearing**, 2016, **20**:1-16.
- [9] 向海燕. 数字高清时代的电视节目响度 [C] // 2009 四川电视节广播电视技术研讨会论文集. 西安: 中国电影电视技术学会, 2009:304-313.  
XIANG Haiyan. The loudness of television programs in the digital HD era [C] // **Paper Collection of Television Technology Seminar on 2009 SCTVF**. Xi'an: China Society of Motion Picture and Television Engineers, 2009:304-313. (in Chinese)
- [10] 陈章虹. 音频响度测试与控制系统的研究与实现 [D]. 长沙: 中南大学, 2014.  
CHEN Zhanghong. Research and implementation of audio loudness test and control system [D]. Changsha: Central South University, 2014. (in Chinese)
- [11] 应俊. 浅谈响度标准概念及相关软件的使用 [J]. 电视工程, 2017(1):34-37.  
YING Jun. A brief discussion on the concept of loudness standard and the application of relevant software [J]. **Television Engineering**, 2017(1):34-37. (in Chinese)
- [12] 盛轶骏. 电视节目响度差异问题的探讨 [J]. 世界广播电视, 2011(1):58-60.  
SHENG Yijun. Discussion on the difference in loudness of TV programs [J]. **International Broadcast Information**, 2011(1): 58-60. (in Chinese)
- [13] SKOVENBORG E, NIELSEN S H. Evaluation of different loudness models with music and speech material [C] // **117th Audio Engineering Society Convention**. New York: AES, 2004:6234.
- [14] ZORILÁ T-C, STYLIANOU Y, FLANAGAN S, *et al.* Effectiveness of a loudness model for time-varying sounds in equating the loudness of sentences subjected to different forms of signal processing [J]. **Journal of the Acoustical Society of America**, 2016, **140**(1):402-408.
- [15] MOORE B C. Development and current status of the "Cambridge" loudness models [J]. **Trends in Hearing**, 2014, **18**:320-325.
- [16] Deutsches Institut für Normung. Calculation of Loudness Level and Loudness from the Sound Spectrum — Zwicker Method — Amendment 1: Calculation of the Loudness of Time-Variant Sound: DIN 45631/A1 [S]. Berlin: DIN, 2008.

- [17] SCHLITTENLACHER J, HASHIMOTO T, KUWANO S, *et al.* Overall judgment of loudness of time-varying sounds [J]. **Journal of the Acoustical Society of America**, 2017, **142**(4):1841-1847.
- [18] NAMBA S, KATO T, KUWANO S. Evaluation of loudness level of time-varying sounds [C] // **40th International Congress and Exposition on Noise Control Engineering 2011, INTER-NOISE 2011**. Osaka: INCEJ and ASJ, 2011:3077-3083.
- [19] ISO/TC 43 Acoustics. Acoustics — Normal Equal-Loudness-Level Contours; ISO 226; 2003 [S]. Geneva: ISO, 2003.
- [20] SCHÄRER Z, LINDAU A. Evaluation of equalization methods for binaural signals [C] // **126th Audio Engineering Society Convention**. New York: AES, 2009:15-31.
- [21] ANDERSON C M B, WHITTLE L S. Physiological noise and the missing 6 dB [J]. **Acustica**, 1971, **24**(5):261-272.
- [22] SIVIAN L J, WHITE S D. On minimum audible sound fields [J]. **Journal of the Acoustical Society of America**, 1933, **4**(4):288.
- [23] RUDMOSE W. The case of the missing 6 dB [J]. **Journal of the Acoustical Society of America**, 1982, **71**(3):650-659.
- [24] KILLION M C. Revised estimate of minimum audible pressure: where is the "missing 6 dB"? [J]. **Journal of the Acoustical Society of America**, 1978, **63**(5):1501-1508.
- [25] 国家标准化管理委员会. 声学语言清晰度测试方法: GB/T 15508—1995 [S]. 北京: 中国标准出版社, 1995.  
National Standardization Administration. Acoustic — Speech Articulation Testing Method; GB/T 15508-1995 [S]. Beijing: Standards Press of China, 1995. (in Chinese)
- [26] 国家标准化管理委员会. 电声产品声音质量主观评价用节目标样说明书: GSBM. 6001—1989 [S]. 北京: 中国标准出版社, 1989.  
National Standardization Administration. Specification for the Subjective Evaluation of Sound Quality of Electroacoustic Products; GSBM. 6001-1989 [S]. Beijing: Standards Press of China, 1989. (in Chinese)
- [27] SOULODRE G A. Evaluation of objective loudness meters [C] // **116th Audio Engineering Society Convention**. New York: AES, 2004:6161.

## Study of deviations of different loudness algorithms based on Chinese speech materials

LI Demin, ZHU Peisheng\*, LU Xiaodong, TAO Wanqi, ZHOU Kaiyu, REN Xijie

( School of Architecture and Fine Art, Dalian University of Technology, Dalian 116024, China )

**Abstract:** A comparative study has been completed between two types of internationally recognized loudness models: Moore's and Zwicker's time-varying loudness models and loudness models recommended by ITU (International Telecommunication Union). To test the availability of these loudness models in terms of calculating the total loudness of Chinese speech materials, a loudness listening experiment is realized, and acoustic stimuli are presented to subjects through headphones. Four different statistical measures are employed in the evaluation of the models, and the absolute accuracy and relative accuracy of the models are measured. It is found that the loudness model recommended by ITU-R BS. 1770-4 is relatively simple and easier to measure the total loudness of speech materials. When calculating the total loudness of Chinese speech materials, the loudness models recommended by the ITU are more accurate than Moore's and Zwicker's time-varying loudness models, and the improved accuracy is 0.5-2 dB.

**Key words:** loudness model; Chinese speech materials; total loudness