arXiv:2601.06662v1 [eess.AS] 10 Jan 2026

基于反卷积的频率箱特定衰减脉冲响应去混响滤波器

Stefan Ciba
https://www.researchgate.net/profile/Stefan-Ciba

摘要

本文介绍了一种稳健的单通道逆滤波器，用于对非理想录音进行去混响，并通过真实音频进行了验证。所开发的方法侧重于计算和修改离散脉冲响应，以滤除已知数字单通道录音设置及房间特性（如早期反射和混响）带来的影响。其目标是实现更干燥、更清晰的信号重建，理想情况下得到直达路径信号。时域脉冲响应由倒谱域计算得出，并通过频谱中针对每个频率格的指数衰减进行渐隐处理。衰减率通过盲估计获得，即针对每个频率格，估算录制输出信号与测试信号之间的混响时间比。修改后的脉冲响应通过反卷积对录制的音频信号进行滤波。盲估计广为人知，其突出的优势是对噪声和非理想因素具有较强的鲁棒性。将盲估计用于脉冲响应修改，以通过反卷积滤除特性，在需要调整特定频率格的混响及其他特性的场景中具有优势，例如音频制作、声学仿真或虚拟现实应用。该滤波器可应对非理想录音条件，并可作为单个音频通道处理的初始步骤之一。某些实时应用可受益于此方法；但对于录制期间脉冲响应发生变化的情况，某些应用可能需要实时或空间自适应，而手动重新校准则不可行。直达路径信号的估计是许多应用的关键。

关键词 去混响 · 频谱脉冲响应修改 · 逆滤波方法

1 引言

自20世纪60年代和20世纪70年代数字技术问世以来，数字音频信号处理领域取得了显著发展，使人们能够以前所未有的精度和灵活性处理声音 [1, 2]。
基于与脉冲响应卷积，对房间和录音设置的声学特性进行建模和修改，是数字音频处理中的一项关键技术。采用线性时不变 (LTI)系统方法，将音频信号与测量或合成的脉冲响应进行卷积，从而能够模拟混响、滤除早期反射，并补偿录音链引入的音色失真 [2, 3]。

传统的混响与去混响方法通常依赖于参数模型，或对脉冲响应 [3, 4]进行稀疏表示。然而，通过稀疏且经过频谱修正的脉冲响应进行卷积，并通过指数衰减函数明确纳入 $T60$ ，为单通道去混响提供了一种直接且灵活的方法。该方法基于既定的信号处理原理，能够有针对性地抑制早期反射、混响以及不理想的录音特性。这种方法不同于密切相关的最新研究成果，例如“时频域语音去混响” [5], 、“使用IIR结构近似真实脉冲响应” [6] 以及“离散时间小波变换域中的系统表示与辨识” [7]。

2 信号处理

2.1 用于系统辨识的周期正弦扫频

为表征一个系统，可以使用周期性线性正弦扫频（啁啾信号）作为测试信号 [8]。正向啁啾信号的频率在时长 T内从 f_0 线性增加至 f_1：

$$\acute{x}(t) = \sin\left(2\pi\left[f_0 t + \frac{f_1 - f_0}{2T}t^2\right]\right), \quad t \in [0, T]. \tag{1}$$

信号以 $f_s = 48$ kHz 的采样率进行采样，因此包含 $N = Tf_s$ 个采样点，计数范围为 $n = 0$，1，2，$...N-1$。通过使用 $t_n = {}^{n}{}_{模}{}_{f_s}{}^{N}$进行索引，正啁啾信号的表达式变为离散且周期性的：

$$\acute{x}_n = \sin\left(2\pi\left(f_0 t_n + \frac{f_1 - f_0}{2T}t_n^2\right)\right). \tag{2}$$

通过数组平铺，啁啾向量还可以扩展至 P 个周期：

$$x_n = \bigoplus_{p=0}^{P-1} \acute{x}_{n-pN}. \tag{3}$$

2.2 脉冲响应估计

给定离散时间序列 x_n 及其记录信号 y_n，两者均包含 N 个采样点，它们之间的关系可由一个具有 $y_n = x_n * h_n$ [9] 的离散线性时不变系统描述。其中，脉冲响应 h_n 描述了信号链，其中可能包含令人不悦的房间特性，例如造成早期反射、混响、共振，以及传感器和扬声器的非线性。
通过对相关信号进行反卷积，可以去除脉冲响应中的相关性，以进行系统辨识：

$$h_n = y_n * x_n^{-1}. \tag{4}$$

所选用的反卷积方法是在倒谱域中对测试信号和录制信号逐帧相减，然后反变换回时域。随后将脉冲响应合并并归一化，以得到一个时域表示。

将离散时间信号分割为 N_{frames}帧，帧长为 $N_{\text{DFT}} = 5 \cdot f_s$，因此根据采样率 f_s选择了相应的离散傅里叶变换大小，使脉冲响应 的持续时间符合常见的混响和延迟时间。对于每个帧索引 $\eta = 0$，1，2，..... $N_{\text{frames}} - 1$，帧边界以跳步长度 $N_{hop} = \frac{N_{\text{DFT}}}{2}$沿信号移动，从而形成 $o = 50\%$ 的重叠。频率格按索引$\mu = 0$，1，2，..... $N_{\text{DFT}-1}$计数。同样，离散时间步长由索引 ν表示，用于统计一帧中的样本数。子带可表示为角频率 $\Omega = \frac{2\pi}{N_{\text{DFT}}}$乘以索引 μ。离散傅里叶变换频谱采用矩形窗 $\omega_\nu^{(\text{rect})}$逐帧处理。

复数加窗短时频谱为：

$$\underline{Y}_{\mu,\eta} = \sum_{\nu=0}^{N_{\text{DFT}}-1} y_{\eta N_{hop}-\nu} \cdot \omega_\nu^{(\text{rect})} e^{-j\Omega\nu}. \tag{5}$$

原始信号的短时频谱为 $X_{\mu,\eta}$。记录信号的短时频谱为 $Y_{\mu,\eta}$。为保证数值稳定性，通过$|X_{\mu,\eta}|_\varepsilon = \max(|X_{\mu,\eta}|, \varepsilon)$进行正则化。
实倒谱通过如下变换获得：

$$C_{\mu,\eta}^{(x)} = \Re\left\{\frac{1}{N_{\text{DFT}}}\sum_{\mu=0}^{N_{\text{DFT}}-1} \ln|\underline{X}_{\mu,\eta}|_\varepsilon e^{j\Omega\mu}\right\}. \tag{6}$$

在倒谱域 反卷积通过减法来完成，以获得脉冲响应：

$$C_{\mu,\eta}^{(h)} = C_{\mu,\eta}^{(y)} - C_{\mu,\eta}^{(x)}. \tag{7}$$

脉冲响应反向变换回时域，反之亦然。虽然虚部应为零，但为了避免数值误差，只使用实部：

$$h_{\nu,\eta} = \Re\left\{ \frac{1}{N_{\text{DFT}}} \sum_{\mu=0}^{N_{\text{DFT}}-1} \overbrace{e^{\left(\sum_{\mu=0}^{N_{\text{DFT}}-1} C_{\mu,\eta}^{(h)} e^{-j\Omega\mu}\right)}}^{H_{\mu,\eta}} e^{j\Omega\mu} \right\}. \tag{8}$$

脉冲响应通过求平均进行累积：

$$\overline{h}_\nu = \frac{1}{N_{\text{frames}}} \sum_{\eta=0}^{N_{\text{frames}}-1} h_{\nu,\eta}. \tag{9}$$

为确保脉冲响应的尺度一致，最大绝对值为 1，需要对其进行归一化：

$$h_\nu = \frac{\overline{h}_\nu}{\max_\nu |\overline{h}_\nu|}. \tag{10}$$

这些步骤得到一个归一化的单一脉冲响应，可作为良好的整体表示。

2.3 频谱脉冲响应修改

时域脉冲响应（例如公式 (10) 中的脉冲响应）经短时傅里叶变换处理后，根据盲 $T60$ 估计对频率分箱进行衰减。

2.3.1 各频率格混响时间的盲估计

ISO 3382 标准 [10] 正式将 $T60$ 定义为：空间声能衰减 60 dB 所需的时间。尽管已知脉冲响应，本研究仍采用 $T60$的盲估计来塑造每个频率格的脉冲响应，以针对特定的声学和系统伪影进行反卷积滤波；文献中尚未见到这种方法。此外，$T60$ 混响时间是传统上用于测量、表征和塑造室内声学的量，通常作为全局参数或频带平均参数使用 [11]。Ratnam 等人提出了一种用于盲单通道 $T60$ 估计[12]的最大似然（ML）算法。为对脉冲响应进行频率格特定的修改，需通过以下步骤获得衰减数组。离散信号 x_n 的功率谱密度（PSD）是其短时傅里叶变换幅度的平方：

$$S_{\mu,\eta} = |\underline{X}_{\mu,\eta}|^2. \tag{11}$$

每个频率格的能量为：

$$E_{\mu,\eta} = \frac{\sum_{\eta'=\eta}^{N_{\text{frames}}-1} S_{\mu,\eta'}}{\sum_{\mu'=0}^{N_{\text{DFT}}-1} S_{\mu,\eta'}}. \tag{12}$$

$T60$ 每个频率格的时间定义为累积能量降至某个阈值（例如 0.001，对应幅度 -60 dB）以下所需的时间：

$$\eta_\mu^{(\text{T60})} = \min\left\{\eta : E_{\mu,\eta} < 0.001\right\}. \tag{13}$$

重叠因子为：

$$o = 1 - \frac{N_{hop}}{N_{\text{DFT}}}, \tag{14}$$

每个频率格的 $T60$ 估计值（以秒为单位）为：

$$T60_\mu = \frac{o \cdot \eta_\mu^{(\text{T60})}}{f_s}. \tag{15}$$

2.3.2 衰减函数的应用

原始测试信号 x_n 与其录制信号 y_n的 $T60$之比 ρ，用于对每个频率格的指数衰减进行加权： $\rho_\mu^{(\text{T60's})} = \frac{T60_\mu}{(y_n)\ T60_\mu\ (x_n)}$。

对于逐帧时间长度 $\tau = \eta \cdot \frac{N_{hop}}{f_s}$，指数衰减函数为：

$$D_{\mu,\eta} = e^{\left(-\frac{\tau}{\rho_\mu^{(\text{T60's})}}\right)}. \tag{16}$$

逐帧衰减修正后的脉冲响应为：

$$h'_{\nu,\eta} = \frac{1}{N_{\text{DFT}}} \sum_{\mu=0}^{N_{\text{DFT}}-1} \underline{H}_{\mu,\eta} \cdot D_{\mu,\eta}\, e^{-j\Omega\mu}. \tag{17}$$

通过 $\omega_\nu^{(\text{hann})}$ [2] 应用合成窗（汉宁窗），此外还必须处理重叠，这通过恒定重叠相加法（COLA）完成：

$$h'_n = \sum_{\substack{\eta \\ 0 \le n-\eta N_{hop} < N_{\text{DFT}}}} \omega_{n-\eta N_{hop}}^{(\text{hann})} \cdot h'_{n-\eta N_{hop},\eta}. \tag{18}$$

2.4 滤波器组的应用

可以对修正后的脉冲响应 $h''_n = h'_n \cdot e^{-dk\cdot n}$ 应用额外的整体指数衰减 dk，以避免回声。最后，通过反卷积，将修正后的脉冲响应作为频谱滤波器组$H''_{\mu,\ \eta}$ 应用于任意录制信号的频谱$\underline{Z}_{\mu,\ \eta}$：

$$\underline{Z}''_{\mu,\eta} = \frac{\underline{Z}_{\mu,\eta}}{\underline{H}''_{\mu,\eta}}. \tag{19}$$

短时傅里叶变换和逆短时傅里叶变换均使用汉宁窗。逆短时傅里叶变换则使用 COLA 方法。帧大小设为与脉冲响应 相同的采样点数；如有必要，则通过严格避免削波来归一化增益，以获得所需的滤波信号 $\hat{}z''_n = z''_n / max(|z''_n|)$。

3 测试设置

使用二合一工作站 （HP ZBook Studio x360 G5 (6TW61EA#ABD)）的内置麦克风，在客厅斜屋顶区域的一个角落（面积约为 1 Ar）录制了正弦扫频和语音音频。正弦扫频声音由二合一工作站的内置扬声器播放。内置降噪功能已关闭。目标是改善语音音质，使说话者坐在二合一工作站前方时，声音干净清晰。另一个假设是，劣化的语音可懂度可以得到改善。

同样，正弦扫频和鼓组律动也在同一位置录制。正弦扫频通过鼓组旁的一台 Alesis Active M1 MKII (8N) 近场监听音箱播放。尽管录音条件不佳，目标仍是获得干净、清晰、具有录音棚级原始录音质量的录音。

4 目标

若干客观测量指标证明了该方法的适用性。其中一些指标用于比较系统脉冲响应与滤波后的脉冲响应，另一些则用于比较滤波信号与未滤波信号。

通过将原始信号 z_n 和滤波后的音频信号z''_n 纳入考量，计算了<主要 id='1'>对数信号功率衰减</主要>（LPA），以表示滤波器的效率：

$$LPA = 10log_{10}\left(\frac{\overline{(z''_n - \overline{z''_n})^2}}{\overline{(z_n - \overline{z_n})^2}}\right). \tag{20}$$

<主要 id='1'>短时客观可懂度测量 </主要>（STOI）用于测量语音可懂度 [13]。它仅适用于语音音频。<主要 id='4'>语音质量感知评估 </主要>（PESQ）是一种对人类听觉感知进行建模的算法，可捕捉失真、背景噪声、削波和延迟等影响。它测量<主要 id='6'>平均意见分、客观收听质量</主要>（MOS-LQO），其本质上反映了音频通道的清晰度，对于非语音音频通道同样是必要的。MOS-LQO 越高，音频通道的质量越好。它广泛用于电信、网络电话和音频处理中，以测量语音增强或劣化 [14]。

Zwicker 的时变响度模型可估算人类听者感知到的声音强度或响度 [15]。

基于 DIN 45692 标准（Zwicker 模型）的尖锐度量化了声音听起来有多刺耳或尖厉，取决于 3 kHz 以上频率的突出程度 [16]。

粗糙度衡量声音听起来有多刺耳或“颤动”，与人耳最容易察觉的频段内快速、可听见的振幅波动有关（20–300 Hz 调制）[17]。

清晰度比 D50 是 50 毫秒内早期能量与总能量的比值 [18]。它用于表示脉冲响应中的混响和后期反射有多明显。

测量感知音质的算法可分别通过模块化声音质量指标工具箱（MoSQITo） [19] [20], 、PESQ [21]和 STOI [22],使用。此外，其他测量指标通过专有软件进行评估，即 Cubase 5.0.1 build 147（Steinberg Media Technologies GmbH） [23]。

该 m 量可在结果部分的表格中找到 1.

5 结果

该滤波器产生了几种明确的效果，这些效果显示在客观结果表 1 中。由于信号中的混响经过滤波，信号振幅和功率都会降低。LPA 测量在一张图中展示了这一效果，并为其命名。音频中最显著的估计音高有所升高，这是因为滤波后的房间特性使较低频率逐渐衰减。滤波器会降低 $T60$ ，这正如预期，至少对于那些出现混响的频率而言是如此。响度可能令人不适，滤波后有所降低，但尖锐度则有所增加。粗糙度几乎未受滤波器影响，只能观察到轻微下降。D50 显示，50 毫秒后的反射已消失。

语音音频录制装置原始和修改后的脉冲响应显示在数值图 2 中。原始未修改的脉冲响应中可以看到缓慢衰减、早期反射，以及一些导致晚期反射的不理想特性。滤波器根据数值图 1 中的 $T60$ 衰减矩阵，针对不同频率分箱有选择地减弱这些不理想特性。由于语音录音的脉冲响应与鼓录音不同，滤波结果也大不相同。STOI 测量不受滤波器影响：滤波不会带来语音增强或语音劣化；也不会因滤波而向语音样本添加或移除噪声。出乎意料但原因显而易见的是，语音音频中的 $T60$ 在滤波后降至零。

6 结论

该滤波器有意采用激进的设计，能够滤除混响，产生干净清晰的声音，并最大限度地减少房间和录音设置特征，使直达路径信号占据主导地位。不过，由于该滤波器确实达到了预期效果，为了获得更悦耳的声音特征，输出可能还需要进一步处理；但显然，它为后续音频处理提供了一个极佳的起点。

7 补充材料

所研究方法的 Python 代码和音频样本可在 GitHub 上找到：https://github.com/fanci90/Digital-Deconvolution-Audio-Filter-repo.git。

Tab表 1：客观测量指标。带星号（名称*）的统计数据取自 Cubase 5.0.1 build 147 .

	鼓录音 劣化（z_n）	鼓录音 滤波（z_n''）	语音 劣化（z_n）	语音 滤波后的（z_n''）
最小样本值*	-1	-0.919	-0.045	-0.045
最大样本值*	+1	+0.948	0.046	0.043
峰值振幅*	0dB	-0.46dB	-26.75dB	-27.43dB
直流偏移*	-∞dB	-∞dB	-85.30dB	-86.44dB
估计音高*	2248.0Hz / C#6	9649.6Hz / D8	1445.2Hz / F#5	1530.5Hz / G5
最小均方根功率*	-∞dB	-∞dB	-61.03dB	-63.28dB
最大均方根功率*	-1.84dB	-12.54dB	-32.82dB	-32.61dB
平均值*	-14.38dB	-28.84dB	-41.34dB	-41.27dB
LPA	-16.29dB		$7.63 \cdot 10^{-5}$dB	
$T60$	5.475ms	0.186ms	0.1816ms	0ms
STOI	N/A	N/A	0.9978	0.9978
Δ STOI	N/A		0	
MOS-LQO	4.4202	4.5314	4.6308	4.6337
Δ MOS-LQO	0.1112		$2.9 \cdot 10^{-3}$	
响度	0.90320	0.30145	0.14764	0.14393
粗糙度	0.23022	0.22169	0.21995	0.21946
尖锐度	3.26095	3.95266	3.09846	3.16039
D50	42。9%	100。0%	71.7%	100。0%

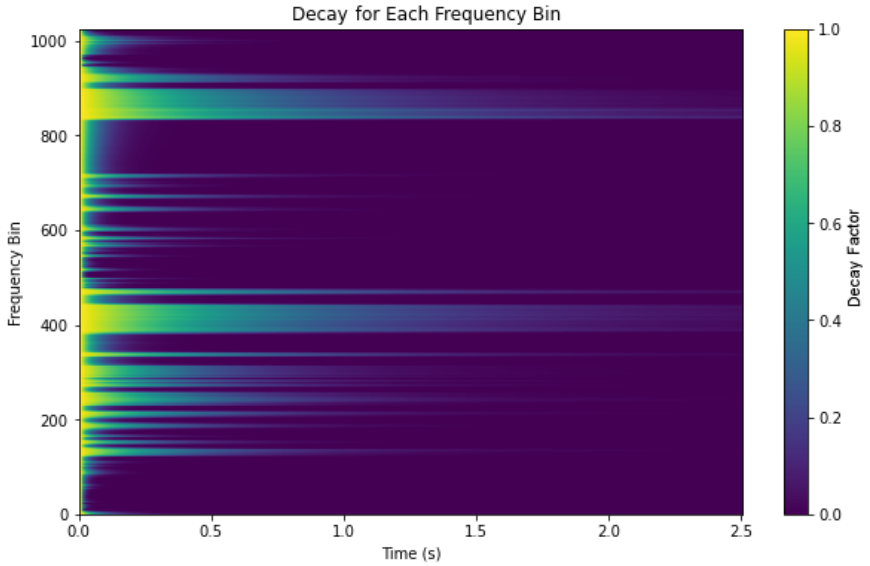


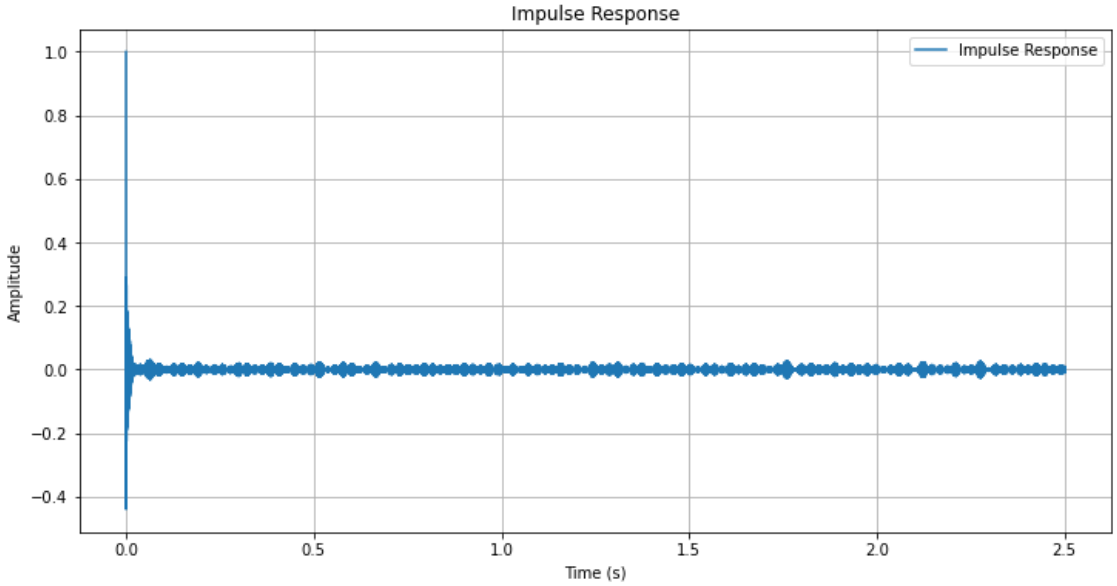
图1： $T60$ 衰减矩阵。

致谢

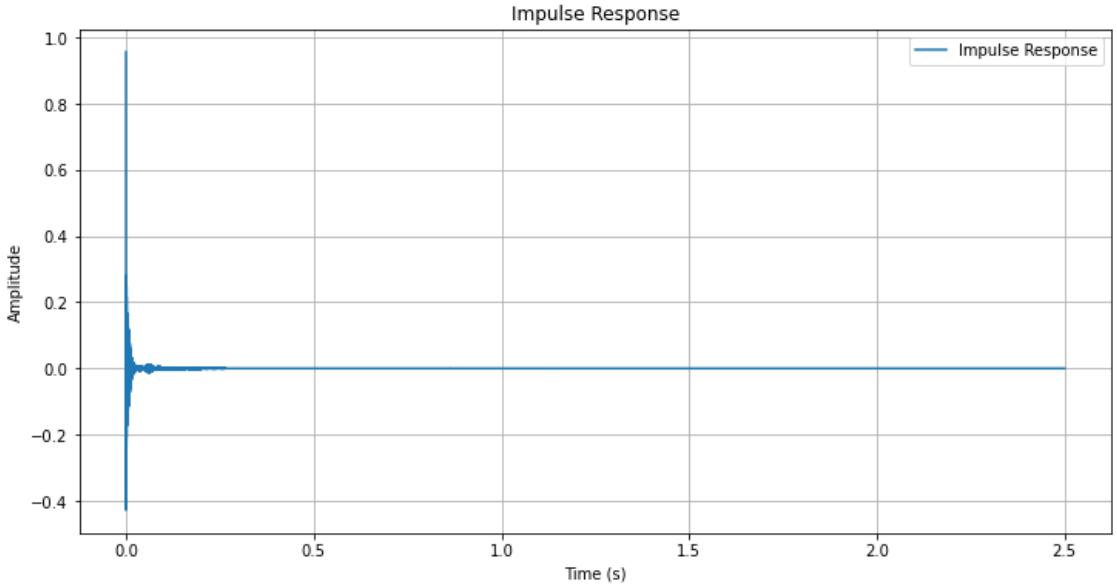
作者谨此感谢来自柏林工业大学音频通信组的 Simon Ciba 硕士（其代表作包括《WhisPER——一种用于进行听音测试的新工具》），感谢他认真审阅本文并提出宝贵意见。

参考文献

[1] A. V. Oppenheim，R. W. Schafer，离散时间信号处理，普伦蒂斯霍尔，1989。ISBN：0-13-216292-X



(a) 原始。



(b) 修改后。

图 2： 脉冲响应。

[2] Julius O. Smith III，频谱音频信号处理，W3K Publishing，2011。在线获取：https://ccrma.stanford.edu/~jos/sasp/。[3] J. A. Moorer，关于混响这回事，《计算机音乐期刊》，第3卷，第2期，第13–28页，1979。doi:10.2307/3680287。[4] E. A. P. Habets，《使用频谱增强进行单麦克风和多麦克风语音去混响》，博士论文，埃因霍温理工大学，荷兰，2007。https://research.tue.nl/files/1972985/200710970.pdf。[5] A. Oyzerman，I. Cohen，《时频域语音去混响》，硕士论文，以色列理工学院，2012。https://israelcohen.com/wp-content/uploads/2018/05/AnnaOyzerman_MSc_2012.pdf。

[6] A. Primavera 等人，《使用IIR结构近似真实脉冲响应》，EUSIPCO会议论文集，2011。https://www.eurasip.org/Proceedings/Eusipco/Eusipco2011/papers/1569422051.pdf。[7] I. Cohen，《离散时间小波变换域中的系统表示与辨识》，IEEE信号处理汇刊，2007。https://israelcohen.com/wp-content/uploads/2018/05/ASM2007.pdf。[8] A. Farina，《使用扫频正弦技术同时测量脉冲响应和失真》，*AES*第*108*届会议，2000。https://www.aes.org/e-lib/browse.cfm?elib=10211。[9] N. Fliege，系统理论。B. G. Teubner Stuttgart，1991（信息技术）。ISBN：3-519-06140-6[10] ISO 3382-1:2009，声学——室内声学参数的测量——第*1*部分：演出空间。https://www.iso.org/standard/40979.html（该标准可付费购买，但其方法在开放文献中被广泛引用）。[11] H. Kuttruff，室内声学（第6版），CRC Press，2016。（提供开放获取章节）ISBN：978-1498740436[12] Ratnam 等人，混响时间盲估计，《美国声学学会期刊》，114(5)，第2877–2892页，2003。https://www.ee.columbia.edu/~dpwe/papers/Ratnam03-reverb.pdf[13] Taal, Cees H.、Hendriks, Richard C.、Heusdens, Richard 和 Jensen, Jesper，《用于时频加权含噪语音的短时客观可懂度测量》，2010年IEEE声学、语音与信号处理国际会议。第4214–4217页，DOI：10.1109/ICASSP.2010.5495701。[14] （国际电信联盟），I. T. U.：建议书 P.862：语音质量感知评估（PESQ）：用于评估窄带电话网络和语音编解码器端到端语音质量的客观方法。2001。[15] ISO 532-2:2017(E)，声学——响度计算方法——第*2*部分：*Moore-Glasberg*方法，国际标准化组织，日内瓦，2017。https://cdn.standards.iteh.ai/samples/63078/a42f5c5bd7014d5f99cc3a7d4fa0c0c9/ISO-532-2-2017.pdf。[16] HEAD Acoustics，《心理声学分析 I：响度与尖锐度计算》，应用说明，201x。https://cdn.head-acoustics.com/fileadmin/data/global/Application-Notes/SVP/Psychoacoustic-Analyses-I_e.pdf。[17] H. Fastl，《声音质量评价的心理声学基础》，慕尼黑工业大学，2006。https://mediatum.ub.tum.de/doc/1138486/729841.pdf。[18] T. Houtgast 和 H. J. M. Steeneken，《室内声学中的MTF概念及其在语音可懂度估计中的应用综述》，美国声学学会期刊，第77卷，第3期，第1069–1077页，1985。[19] Green Forge Coop. MOSQITO（版本1.1.1）。https://doi.org/10.5281/zenodo.10629475[20] Eomys Engineering，《模块化声音质量指标工具箱》（MoSQITo），2025。https://github.com/Eomys/MoSQITo[21] C. Ludlow，python-pesq：语音质量感知评估（PESQ），2025。https://github.com/ludlows/python-pesq.git[22] M. Pariente，pystoi：短时客观可懂度（STOI），2025。https://github.com/mpariente/pystoi[23] Steinberg Media Technologies GmbH，Cubase 5.0.1 build 147 [计算机软件]，2009。https://o.steinberg.net/index.php?id=1782&L=1