编钟非谐加法合成：参数辨识、瞬态建模与声学分析

16钟位、三力度分层与36分量模型

摘要

编钟音色包含非整数倍频率关系、近频分量拍动及随时间变化的衰减结构。仅以少量谐波或统一包络描述，难以兼顾敲击起音与后续共鸣。本研究以16个钟位、轻中重三个力度的48条独立立体声实录为基础，采用36个阻尼振荡分量描述持续共鸣，并分别设置短时起音修正和有界余韵修正。本次采样更新同时核对文件字节与解码后的PCM数据：41条录音保持一致，沿用其已验收参数；7条录音发生变化，重新辨识对应力度层。发声阶段仅读取数值参数，不以钟声采样或录音残差作为激励。通过分量拆分、时频分析、实录对照及跨实现数值比对，考察各部分对合成结果的作用。参数辨识与实现一致性分别评价；同次敲击的重建能力不等同于对独立敲击的预测能力，听觉相似度仍需通过重复采样和受控听辨进一步评价。

关键词： 编钟；非谐加法合成；参数辨识；瞬态；近频拍动；声学可视化

---

1. 研究对象与数据范围

研究对象为16个钟位，每个钟位具有轻、中、重三个力度的独立实录，共48条录音，对应48个参数层。录音采样率均为48 kHz，采用双声道记录。南吕轻力度现有独立录音，按与其他力度层相同的流程进行参数辨识和重建评价，不再由中力度参数推导。

本次更新分别从文件字节与解码后的PCM数据两方面核对录音。41条录音在两种核对中均保持一致，其参数沿用已验收结果；另有7条新录音，涉及蕤宾轻、中、重力度，南吕轻力度，以及半黄钟轻、中、重力度。对应录音标识分别为08_soft、08_middle、08_heavy、11_soft、14_soft、14_middle及14_heavy。更新范围据此限定为这7个力度层的重新辨识。沿用参数并不免除新版整体检查：参数装载、输出增益及分析通路仍需在完整48层范围内核验。

每个力度层包含36个持续分量。单个分量由频率、幅度衰减时间T60、左右声道幅度及左右声道初相位组成，共6个参数；每层因此具有216个持续段参数。完整起音修正另含432个系数。“36分量”表示采用的模型规模，并不意味着钟体仅有36个物理振型。

实录同时包含钟体响应、敲击条件、房间反射、话筒特性与背景噪声。由实录辨识的参数可能吸收其中多种影响，故本文将其界定为**录音驱动的非谐加法模型**，不将参数拟合结果等同于完整的钟体力学辨识。

“参数合成”与“采样播放”在此有明确区别：前者按频率、幅度、相位及包络重新计算波形；实录对照则使用独立的音频播放通路。录音用于辨识参数，不参与合成通路的发声。48条独立实录指48条具有独立文件标识的录音，不表示统计独立性，也不表示每一钟位、每一力度均已具备独立于拟合记录的测试样本。

2. 持续共鸣与近频拍动

2.1 阻尼振荡分量

令采样率为Fs = 48,000 Hz，离散采样点n对应时间t = n/Fs，单位为秒。第m个分量在声道c中的持续波形为：

$$b_{m,c}(t) = A_{m,c} e^{-\alpha_m t} \cos(2\pi f_m t + \phi_{m,c})$$

$$\alpha_m = \frac{\ln(1000)}{T_{60,m}}, \qquad a(t) = 1 - e^{-t/0.0015}$$

其中，f为中心频率，A为数字幅度系数，φ为初相位，α为正衰减率。左右声道共用f和T60，分别保留A与φ。a(t)为时间常数1.5 ms的开声门，在持续分量与起音修正相加后统一作用，以减小起点不连续。

T60定义为单个基础分量的幅度降至初值1/1000，即下降60 dB所需的时间，不等于录音时长或房间混响时间。幅度系数A采用数字幅度，尚未与声压建立校准关系。若T60超过有效观测区间，超出的衰减段属于模型外推。

加法模型不要求各中心频率为某一基频的整数倍，因此能够保留钟类声音的非谐结构。阻尼振荡分量也可从并联模态或状态空间的角度解释；在缺少钟体测振资料时，不能据此将每个拟合分量归属于特定物理振型。[1,4]

2.2 近频分量的具体实例

倍夷则重力度属于本次录音及参数均沿用的力度层。其前两个分量按频率排列如下，作为近频叠加的参数实例。

参数	分量1	分量2
频率 / Hz	439.230324	443.176232
T60 / s	20.203470	20.361750
左声道幅度系数	0.005701	0.008307
右声道幅度系数	0.003106	0.011182
左声道初相位 / rad	−1.947051	1.353620
右声道初相位 / rad	−1.722021	1.365212

两频率相差约3.946 Hz，对应拍动周期约0.253 s。叠加后的强弱起伏取决于频差，也受初相位、相对幅度及各自衰减的影响。该现象来自近频分量的干涉，与对单个振荡器施加任意颤音调制具有不同的参数含义。[1]

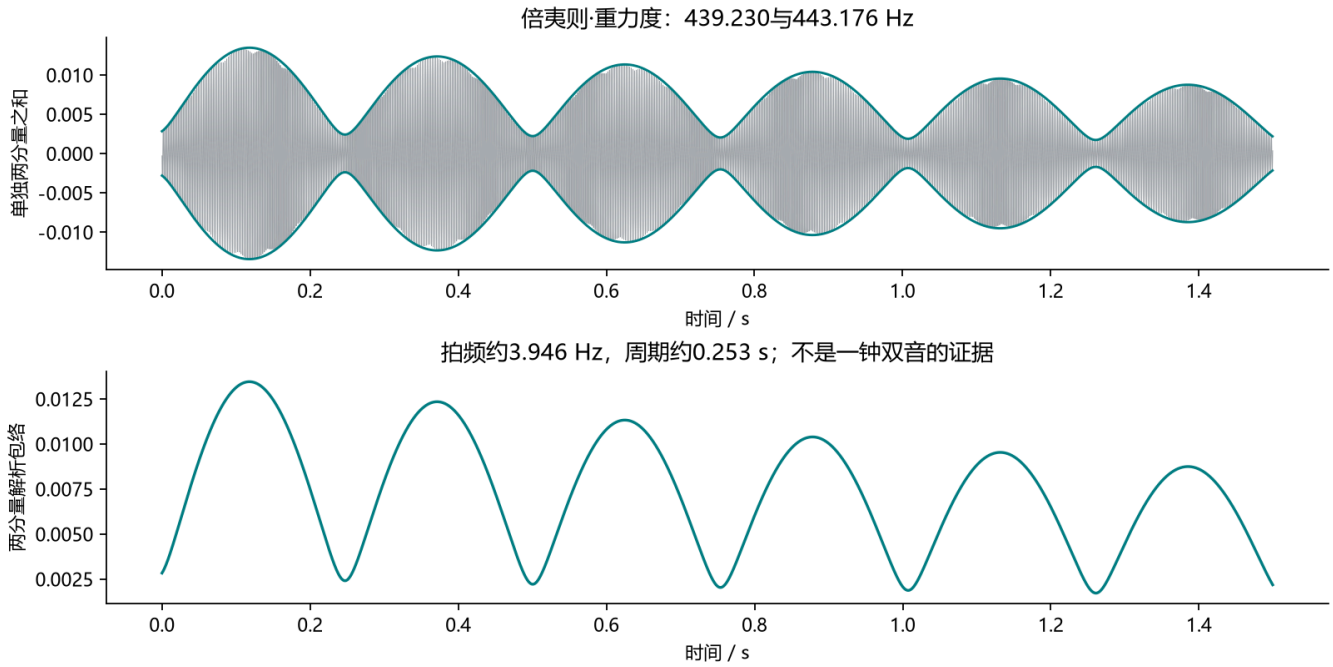


图1 倍夷则重力度层的两个近频分量。包络反映所选分量的叠加关系，不代表完整钟声的总包络。近频拍动也不构成“一钟双音”的证据；不同击点对应的双音机制不在本文建模范围内。

最低拟合频率、谱峰位置与听觉音高并非同一概念。界面以经能量阈值筛选后的最低持续分量为参考，再按A4 = 440 Hz的十二平均律换算邻近音名，作为检索标识；筛选方法见第7节。模型仍保留实录辨识的频率，不据此判定历史律制。“Kangxi14”作为来源标识，也不构成年代、历史定音标准或具体律制的声学证据。

---

3. 参数辨识方法

3.1 单次敲击分段

分析取每条录音的前8 s，使用5 ms的RMS帧检测起音，并保留阈值之前的一帧。若后段出现明显的再次敲击，则截短分析区间，避免将多次激发共同拟合为一次自然衰减。

分段起点作为相位估计及波形比较的时间基准，不视为槌头接触时刻的精密测量。合成与实录比较时沿用相同起点；否则，即使音色相近，数毫秒的时间偏移也可能产生较大的逐点误差。

41个未变更力度层保留其已验收的辨识结果；7个新录音层重新完成分段、候选频率提取、持续段估计及短时修正。参数复用依据为录音数据一致性，不以钟位名称相同作为替代依据。

3.2 候选频率提取

候选谱峰分别从起点后25 ms、长度750 ms的片段，以及起点后150 ms、长度2.8 s的片段提取。分析采用Hann窗，在高于150 Hz、低于9500 Hz的范围内搜索谱峰，以5 dB显著度及高于相对峰值−58 dB的幅度阈值筛选，再合并近邻候选，构建36个载频位置。

7个新录音层中的多数可直接提取36个候选；半黄钟轻力度（14_soft）在上述阈值下仅检出12个清晰共振。为保持统一的36分量结构，该层缺少的24个候选频率中心由同一钟位的重力度录音补充，再以轻力度录音本身估计幅度、初相位及衰减参数。这属于同钟多力度的候选频率先验，不意味着36个共振均在轻力度录音中被独立观测，也不能将各力度层的频率依据视为完全独立。该先验只用于参数辨识，不将重力度的录音波形混入轻力度合成输出。

FFT用于估计频谱结构，发声阶段仍由振荡分量计算波形。零填充细化的是频谱取样网格，不会延长观测时间，也不改变观测窗决定的本征分辨能力。[3]

3.3 频率与衰减的联合估计

相邻候选频率相差不超过28 Hz时归为一组，通过复数异频分析、滤波与降采样，在同一频带内联合估计频率和衰减率。拟合通常从起点后100 ms开始，终点不超过7 s，并受该次敲击的有效长度约束。频率调整范围为候选值附近±0.9 Hz，衰减率限制为正值。

7个新录音层沿用复数异频分析与变量投影方法：将频率、衰减率作为非线性参数，在每次非线性参数更新时求解对应的线性复幅度，以减少需要联合搜索的非线性未知量。非线性参数先在组内能量较强的声道估计，再以共同频率和衰减率分别求取左右声道的复幅度。该处理保留双声道之间的幅度及相位差，避免直接平均声道造成的抵消。

这一过程属于正弦分析与重合成的参数估计框架。[2] 本文采用固定中心频率及指数衰减，不包含完整的逐帧频率轨迹跟踪。重新辨识改变的是录音对应的参数值，持续共鸣、起音与余韵的理论结构保持一致。

3.4 持续段幅相校准

在既有频率、T60和余韵形状固定的条件下，对7个重新辨识层进一步求取幅度与相位的联合修正。系数求解从300 ms开始，终点不超过6.8 s；末1 s不参与本次系数求解，其余观测按50 ms块划分训练与选参区间。各分量通过有界正交系数调整复幅度，左右声道分别估计；更新持续段后重新求取短时起音系数。

完整修正可能改善整体误差，却损害局部起音或尾音。因此，将原参数与候选参数转换为复状态，沿两者之间选取小步长，再检验多分辨率谱幅、持续段波形及各局部窗口。复状态插值同时处理持续分量与起音系数，并不在发声时混入录音。当前仅南吕轻力度、半黄钟中力度及重力度通过筛选，其余层保持既有参数。

筛选要求谱收敛误差相对下降至少0.5%，300 ms后有效区间的波形误差下降至少0.1 dB；起音、过渡、主体、后段及条件保留区间的误差增量均不得超过0.05 dB，谱对数幅度误差也使用0.05 dB容差。这些门槛是本实现的工程判据，不是听觉可辨阈值。末段参与步长选择，且频率与衰减此前已由同次录音估计，故全部结果仍属条件性重建检查，不称为独立测试。

4. 起音的短时幅相修正

持续段参数由起音后的信号估计，反向延伸至激发时刻，未必能重现敲击初段的幅度和相位关系。单纯降低总增益会同时削弱后续共鸣，因此将起音误差单独建模。

在每个中心频率上设置三个短时指数包络，时间常数为6、24、75 ms。令θ = 2πft + φ，声道c中第m个分量的修正项为：

$$q_{m,c}(t) = A_{m,c} w(t) \sum_{k=1}^{3} e^{-t/\tau_k} [C_{m,c,k} \cos\theta_{m,c}(t) + S_{m,c,k} \sin\theta_{m,c}(t)]$$

同频率的余弦与正弦构成正交基，系数C和S共同调整短时幅度与相位，并允许正、负取值。36个分量、2个声道、3个时间常数及2种正交基，对应每层432个起音系数。所选时间常数是该模型的设计参数，并非适用于全部编钟的固定声学常数。瞬态中的幅度与相位关系需同时考虑。[5]

修正门w(t)在前150 ms内为1，在150—300 ms内平滑下降，300 ms后为0：

$$u = \mathrm{clip}\left(\frac{t - 0.15}{0.15}, 0, 1\right), \qquad w(t) = 1 - 3u^2 + 2u^3$$

短时修正具有独立的指数包络，不再乘基础分量的长衰减或余韵乘数。持续共鸣与短时修正先相加，再共同乘以开声门。改变这一运算次序会改变结果，不能仅凭相同的频率表认定实现等价。

7个新录音层的起音系数采用列能量归一化的岭回归求解，以5 ms时间块划分训练、选参和条件性比较区间；时间块用于分隔同次敲击中的观测点，不表示独立录音。列能量归一化用于减小基函数能量差异对系数估计的影响，岭正则用于抑制病态或近相关基函数引起的系数放大。这些措施约束的是参数估计过程，不增加新的发声分量，也不改变上述短时幅相模型。具体系数、约束及逐层误差应以对应版本的参数与分析记录为准。

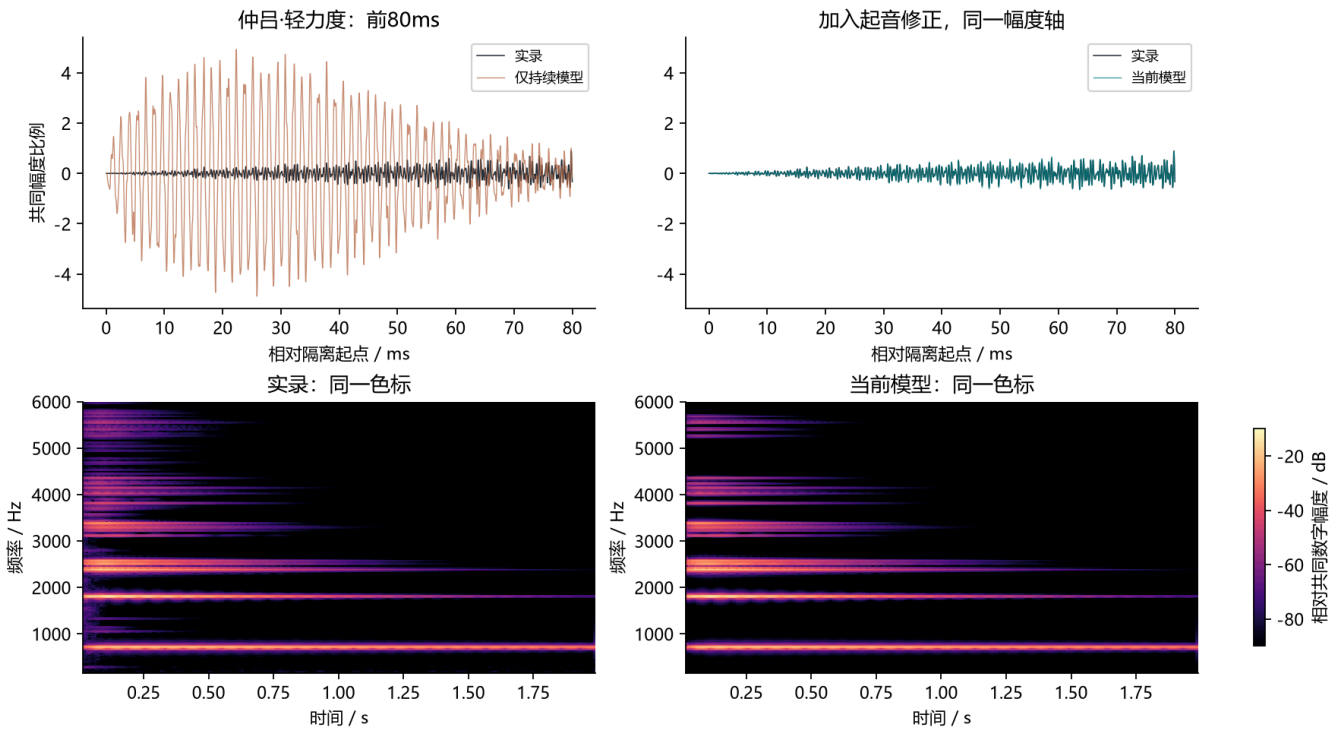


图2 仲吕轻力度的同次敲击对照。该层属于本次沿用参数的范围；持续段模型与完整模型采用相同的波形幅度轴和时频色标。图示结果限于该实例，不能替代7个新录音层的评价、全库统计或受控听辨。

短时包络会造成频谱展宽。因此，36个中心频率并不意味着合成声仅包含36根无限窄的谱线。该修正仍由参数计算，不叠加槌击采样，也不使用录音残差作为发声激励。

---

5. 余韵的有界包络修正

部分实录的尾音不能由单一指数衰减充分描述。模型因此保留有界余韵修正，并通过门控拟合及评价决定各力度层是否采用该项；未采用时保持基础衰减结构。7个新录音层按各自有效尾段重新评价，不能沿用旧录音的适用层清单。

第m个持续分量在300 ms后乘以：

$$h_m(t) = 1 + c_m(1 - e^{-\max(0,\,t-0.3)/\tau_{\text{layer}}})^2, \qquad |c_m| \le 0.75$$

c控制中后段包络的相对增减，τ_layer从0.25、0.75、2 s中选择，每层共用一个时间常数；左右声道共用各分量的c，以限制模型自由度。300 ms之前h为1，因而不改变已建模的起音部分。

该修正保留中心频率、初相位、基础T60及远期指数衰减斜率，但改变中间时段的包络形状。修正后的完整包络不保证处处单调，其实际下降60 dB所需时间也不宜继续直接解释为基础T60。

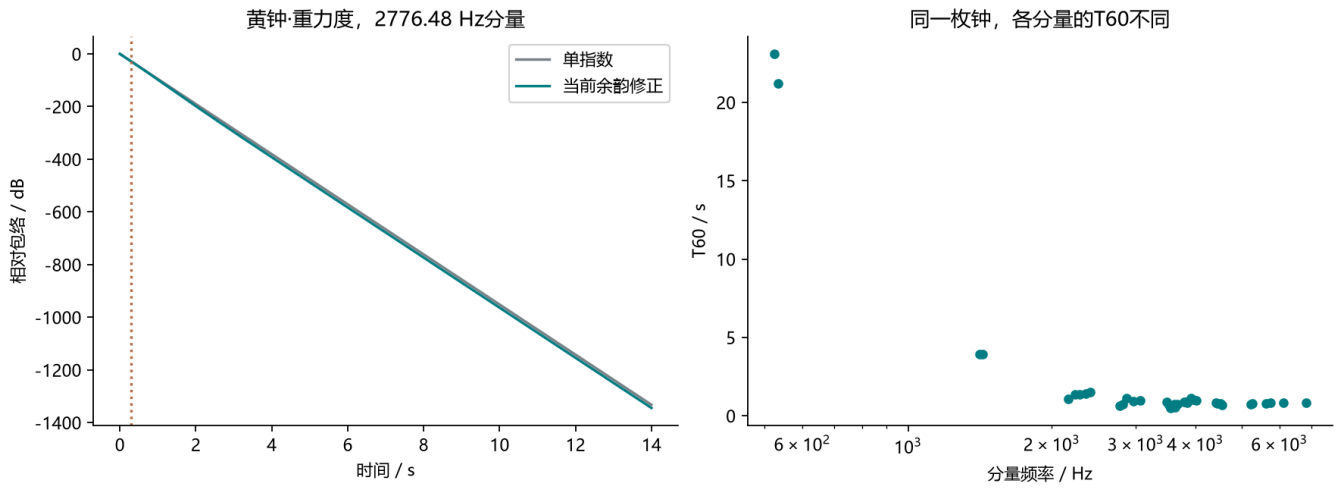


图3 黄钟重力度的基础衰减与余韵修正。该层属于本次沿用参数的范围。T60描述基础分量；余韵乘数描述有限时间内的包络调整，二者不具有相同的物理测量含义。

余韵系数拟合保留同次敲击的最后1 s，不将其用于该项系数求解。然而，频率、T60及候选方案选择仍依赖该次录音，故末段比较属于条件性的同次重建检查，不能视为对独立录音的预测验证。门控或保留尾段也不能消除同次录音内进行参数选择所带来的评价局限。

---

6. 输出结构与离散实现

6.1 完整输出

第l个力度层、声道c的输出为：

$$y_c(t) = G_l V(v) F_D(t) a(t) \sum_{m=1}^{36} [b_{m,c}(t) h_m(t) + q_{m,c}(t)]$$

G_l为层增益，V(v)为演奏力度增益，F_D(t)为结束前80 ms的淡出。未采用余韵修正时，h = 1。分析试听固定使用该层参考力度，即V = 1。

初始音色库以首500 ms的双声道RMS为依据，将轻、中、重的标定目标设为−31、−24、−18 dBFS，再结合峰值约束共同衰减。此次更新保留41个复用层的增益，7个重新辨识层则以其对应原输出的首500 ms RMS为目标，并逐层施加−6 dBFS峰值约束，以维持前后版本的电平关系。标定目标不等于所有音色的最终实测值；实际RMS及峰值均从输出重新测得。

三个力度层分别对应MIDI velocity 1—48、49—95、96—127；层内另作连续增益控制。MIDI键48—63用于寻址16个钟位，并不将同一钟声按半音移调。钟声以激发后的自由衰减为主体，不设置恒定的持续电平；开声门、短时修正及各分量衰减共同决定起音和延续过程。

6.2 复极点递推

阻尼振荡可改写为复数状态，在音频块之间保留相位与幅度：

$$p_m = e^{-\alpha_m/F_s + j2\pi f_m/F_s}, \qquad z_{m,c}[0] = A_{m,c}e^{j\phi_{m,c}}$$

$$z_{m,c}[n+1] = p_m z_{m,c}[n], \qquad b_{m,c}[n] = \Re\{z_{m,c}[n]\}$$

每一步递推同时推进相位并缩小幅度。状态跨块保留，可避免每个音频块重新起振造成的边界不连续。短时修正也可按相应衰减率递推，将初值写成A(C−jS)exp(jφ)，其实部对应A[C cosθ + S sinθ]。

数值实现利用SciPy的lfilter接口传递复数状态。[6] 递推形式与直接计算采用相同参数及输出公式，区别在于计算组织与内存使用方式，不构成新的声学模型。

---

7. 声学分析与可视化

分析界面将合成参数与信号测量分别呈现，避免将参数示意误读为声学测量。各图的含义如下。

目录中的“最低显著持续分量”按参数计算：各分量的左右声道幅度平方取均值，乘以指数衰减包络平方在前2 s内的积分，作为相位平均的持续能量代理；保留不低于最大值−40 dB的分量，选取其中最低频率。共同增益及相位平均的共同系数不影响相对筛选。此量不含实际相位、开声门、起音与余韵修正，也不含分量之间的干涉项，因而不等于完整波形的实测能量。阈值仅用于避免极弱拟合分量主导目录标识，不据此认定听觉基频或物理振型。

分析视图	表示内容	解释范围
合成波形	双声道数字幅度随时间变化	可观察起音与叠加结构，不直接表示声压
分量频率	中心频率及相对幅度系数	属于参数视图，不是完整信号的FFT频谱
持续包络	各基础分量的指数幅度衰减	不包含总包络中的拍动及短时修正
时频结构	合成缓冲的短时傅里叶分析	描述频谱随时间的变化
模态分布	分量频率、T60与相对幅度	属于拟合参数分布，不是钟体振型测量
起始相位	所选分量的左右声道初相位与相对幅度	不等同于整段声场或双耳定位结果
幅度谱对照	合成与实录的时间平均谱形	各自峰值归一化，不能比较绝对响度

时频分析取2 s、48 kHz的合成缓冲，采用2048点周期Hann窗，帧移512点，仅使用完整帧。频率间隔为23.4375 Hz，完整分析范围为0—24 kHz，图像显示150—10,000 Hz的对数频率区间。

FFT幅度除以窗函数之和；正频率幅度乘2，直流与奈奎斯特频点除外。随后取左右声道幅度平方的平均。该结果是幅度平方谱，不是按频率带宽归一化的功率谱密度。图像在各显示分箱内取最大值，不能据此读取超出分析窗分辨能力的精细结构。

功率谱质心使用0—24 kHz全部频点的时间平均功率作为权重。声道相关采用未作时延补偿的Pearson相关系数。前80 ms能量比例以同一2 s缓冲的双声道能量为分母。监听增益不参与上述计算；显示的dBFS也不能换算为未经校准的声压级。

对照试听先对齐分段起点，并按首500 ms RMS匹配电平，再对合成与实录施加共同的峰值保护衰减。等RMS不等于等响度。2 s预览结束前施加20 ms淡出，仅用于避免截断；完整模型的衰减参数保持不变。

三维钟架承担钟位选择与声音参数的对应展示。16个区域为人工划分，模型未拆分为独立钟体，也不呈现测得的振型。钟位标识与音色名称对应，不作为模型、实物及采样身份完全一致的证明。

---

8. 验证协议与误差评价

8.1 实现一致性

实现一致性检查包含54组：48个力度层各比较0.45 s输出，另设6组2 s的分量及修正组合。采用更新后的参数，在相同采样率、时间起点和层增益下，比较Python直接求值与浏览器双精度实现。54组均满足最大绝对样本差小于10^{-8}的门槛，实测最大差为1.59×10^{-15}。该数值是数字波形的计算差异，不是模型与实录之间的误差。

复极点递推还应以不同块长检查跨块状态连续性，比较完整输出与块边界处的误差。此项与直接求值和浏览器实现的比对分别记录，不能以其中一项替代另一项。

上述结果由更新后的参数重新计算，不以旧版结果代替。实现一致性只说明所检计算通路的数值关系，不能换算为音色真实性或主观质量评分。

8.2 全部48层的实录重建评价

逐层检查覆盖全部48个力度层，并区分41个复用层与7个重新辨识层。每层核对录音对应关系、36个持续分量、参数有限性与约束、输出峰值及有效比较区间；南吕轻力度按独立实录层纳入检查，不再作为推导层留空。

实录比较分别考察0—80 ms、80—300 ms及300—1500 ms三个区间的波形误差，并使用512、2048、8192点窗长比较多分辨率谱幅。短窗侧重瞬态定位，长窗提供较细的频率采样。不同窗长的测量结果分别解释，不将多项指标任意合并为单一“真实度”。

波形误差受时间对齐、相位及电平影响；谱幅误差不保留全部相位信息，二者可能给出不同判断。误差计算仅使用实录的有效片段，不将补零部分纳入比较。有效数据不足以覆盖某一评价窗口时，应注明窗口缩短或指标缺失，不能以零值替代。

全部48层均已生成双声道合成文件，并通过完整读回、参数有限性、力度层顺序及无数字削波检查。7个重新辨识层的音头修正均通过既定筛选，其中4层采用余韵修正，另3层保留单指数持续包络。南吕轻力度的多分辨率谱收敛误差从持续模型的0.2967降至含音头修正模型的0.0821，持续段幅相校准后为0.0807；半黄钟中、重力度的对应值分别由0.0863降至0.0847、由0.1266降至0.1234。这些指标采用前2 s观测；结果均针对同次敲击，不能据此推断独立敲击下的相似度。逐层数据记录全部48层的有效区间和误差，不将旧版统计值用于本次实录。

图2和图3分别说明起音及余韵修正的作用，结论限于相应音色和敲击实例。对其余钟位及力度的解释应结合逐层数据，不以单幅示例推断全库效果。

---

9. 方法比较与研究限制

非谐加法模型以并联阻尼振荡分量描述共鸣；数字波导通常通过延迟线、滤波及反馈表示传播和反射。[4] 两种结构之间存在数学联系，但复极点递推不因此等同于数字波导。

吴致暐的编钟合成研究采用非谐数字波导，并从录音取得约50—100 ms的逆滤波残差作为激励。[7] 本文不使用录音残差激励，而通过持续分量及短时幅相系数计算起音，发声路径与该方法不同。

48条独立实录完善了16钟位、三力度的数据覆盖，但每层参数辨识与当前重建评价仍针对同次敲击。不同力度各有一条录音，不等于同一条件下已有多次重复测量；更不能据此判断模型对不同敲击位置、力度波动、话筒位置及房间条件的适用性。进一步评价需要各钟位、各力度的多次独立采样，预先划分拟合与测试记录，并在固定电平处理条件下随机呈现、隐藏版本标签，分别评价起音、共鸣、拍动及尾音。

36分量规模、有限的起音基函数及有界余韵修正均为建模取舍，不能保证涵盖全部噪声成分、时变频率或复杂激发过程。分块选参、正则化及参数边界有助于约束估计，但不构成最优解、完美重建或听觉不可区分的证明。

本研究尚未形成受控听辨结果。双声道参数描述的是记录条件下的数字信号关系，不替代房间声场测量；声学重建结果也不构成生理或治疗效果的证据。

---

10. 结论

16钟位的非谐加法模型以36个持续分量为基础，将起音幅相修正和中后段包络修正分开处理。各部分具有独立的参数定义、作用区间和运算位置，能够在不播放钟声采样的条件下重新计算声音，并通过分量拆分和声学视图分析其构成。

数据范围为48条独立实录。41条录音经文件字节与解码PCM核对保持一致，沿用其已验收参数；7条新增或变化录音重新辨识，南吕轻力度由对应录音建立。更新保持既定理论模型；54组跨实现比对及48层离线输出检查通过。其结论分别限于计算一致性与同次实录的参数化重建，不替代实际演奏链路及受控听辨评价。

现阶段方法的解释范围仍为给定记录条件下的参数化重建。独立敲击验证、受控听辨及有校准依据的声学测量，仍是评价音色相似度与适用范围的必要环节。

参考文献

1. Karjalainen, M., Välimäki, V., & Esquef, P. A. A. (2002). Efficient modeling and synthesis of bell-like sounds. *Proceedings of DAFx-02*, 181–186. 全文。
2. McAulay, R. J., & Quatieri, T. F. (1986). Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(4), 744–754. DOI: 10.1109/TASSP.1986.1164910. 全文。
3. Smith, J. O. III. *Spectral Audio Signal Processing*. Zero-Padding for Interpolating Spectral Peaks。
4. Smith, J. O. III. *Physical Audio Signal Processing*. State Space to Modal Synthesis；Digital Waveguide Modeling Elements。
5. Röbel, A. (2003). A new approach to transient processing in the phase vocoder. *Proceedings of DAFx-03*. 全文。
6. SciPy. *scipy.signal.lfilter*. 接口与差分方程。
7. 吳致暐（2012）。《以非諧音數位波導為基礎之中國編鐘聲音合成模型》。國立交通大學碩士論文。全文。