arXiv:2305.04531v1 [cs.SD] 8 May 2023

一种分析音频设备采样抖动的方法

Makoto Takeuchi[a] 和 Haruo Saito
Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, 153-8902, JAPAN

本文提出了一种分析音频设备采样抖动的方法。该方法基于时域分析，利用录制的正弦波中过零点的时间波动来表征抖动。当同一播放信号同时馈入两台音频录音机时，该方法能够分别评估音频播放器和音频录音机中的抖动。实验使用了最高采样率为每秒 192 000 个采样点的市售便携式设备。结果表明，该方法能够识别音频播放器中数十皮秒量级的抖动。此外，该方法还能利用音频设备的左右声道，将抖动与相位无关噪声分离。因此，该方法可用于评估音频设备、信号发生器和时钟源的性能。

[ ]

I. 引言

音频设备中的采样抖动是指采样时刻相对于理想时序的误差，即

$$t[i] = i - ( )$$

1 f^{-1} f_S

S，其中，表示数模转换器（DAC）和模数转换器（ADC）的采样频率。采样抖动 $j_S(t)$ 会导致采样时刻变为 $t[i] = ($ $i-1)f^{-1}$ S $+ j_S(t[i])$。因此，采样抖动会影响音频设备的性能。传统上，采样抖动通常在频域中进行分析[1–4]。在这种方法中，首先使用音频播放器播放频率为 $f_P/4$ 的正弦波，其中 f_P 是音频播放器的采样频率，然后对其进行录制，接着使用旁瓣电平较低的窗函数（如 Blackman-Harris 窗）分析其频率响应。除了频域分析（FDA）之外，研究人员还开展了采样抖动的时域分析（TDA）[5]。通过希尔伯特变换，可以获得真实的抖动波形 $j_S(t)$。时域分析的优势在于，可以从录制波形中分别提取抖动和幅度调制（AM），而这是频域分析无法实现的。

本研究提出了一种高效且强大的方法，用于表征音频设备中的采样抖动。该方法包含两个关键要素。第一个是改进的TDA，称为过零点分析（ZCA）。应用该方法时，先分析录制波形的过零点（ZCP），然后计算理想正弦波的过零点。录制波形与理想正弦波之间过零点的时间差包含抖动信息。我们将这些时间差称为“过零点波动（ZCF）”。ZCA使我们能够即使输入信号同时包含抖动和幅度调制，也能从录制波形中提取抖动信息。第二个关键要素是使用两台音频录音机同时录制同一播放信号，从而生成两个独立的波形。我们将此称为“双录音机设置（DRS）”。由于ZCA能够保留绝对时间信息，因此我们可以精确比较并计算所生成的两个波形中ZCF的正相关和负相关。根据概率加法规则，可以分别评估播放器和录音机的采样抖动。

需要指出的是，所提出的方法既不需要与录音机内部时钟同步的可选输出时钟信号，也不需要精度高于内部时钟的外部时钟发生器。因此，该方法可以使用低成本录音机实现，并且适用于终端用户。所提出的方法还可应用于高频相位噪声和抖动测量。通过将两台录音机替换为两台具有更高采样率的数字示波器，我们可以使用该方法对各种能够输出高频正弦波的信号发生器和时钟发生器进行性能评估。

该方法与麦克风的互易校准有些相似，这种校准方法自20世纪40年代以来便已得到应用[6,7]；然而，它不需要双向使用设备。DRS与用于相位噪声测量的交叉频谱法（CSM）的配置类似[8]。在CSM中，通过使用两台乐器重复进行FDA，可将乐器的影响降低至$1/\sqrt{m_{CS}}$，其中 m_{CS} 为测量次数。所提出的方法即所谓的CSM的TDA版本。通过使用DRS，可以消除乐器的影响。

在本研究中，我们重点关注音频设备的性能；然而，这项研究也为该领域作出了贡献。

[a]takeuchi@phys.c.u-tokyo.ac.jp

人类可听度[9–11]。此前的抖动研究[9] 表明，音乐信号中随机抖动的感知检测阈值较高，但当时并未控制原始抖动，即添加额外抖动之前已存在的抖动。研究人员可以使用本文提出的方法，快速筛选出抖动最小的乐器。借助该方法，可以进一步研究：使用抖动水平较低的音频播放器，并采用叠加了轻微人工抖动的音乐信号对受试者进行充分训练后，检测阈值会降低多少。此外，在所有可听度研究中，所提出的方法还有助于诊断播放器是否运行正常。

本研究其余部分安排如下。第II 节介绍所提出方法的基本原理。第III 节介绍实验流程。第IV 节呈现我们的结果。在第V节中，我们从多个角度展开讨论；第VI节则总结全文并展望未来。我们进行了数值计算，以确认所提出方法的准确性，相关细节见附录。

II. 原理

A. 播放信号和录制波形中的噪声分类

纯正弦波 $F_{\mathrm{pure}}(t)$ 表示如下：

$$F_{\mathrm{pure}}(t) = A_0 \cos(\omega t + \theta_0), \tag{1}$$

其中，t、ω、θ_0和 A_0 分别是波的时间、角频率、初始相位和幅度。当播放一个预期会重现纯正弦波的数字音频文件时，产生的播放信号并非纯正弦波。它包含 (i) 抖动、(ii) 幅度调制和 (iii) 与相位无关的 (PI) 噪声。这里，我们分别解释这三种噪声模式。(i) 抖动是每个点处播放时序的偏差。存在抖动时，t 会被 $t + j(t)$替代，其中 $j(t)$是播放器的抖动。因此，纯正弦波 $F_{\mathrm{pure}}(t)$会变为

$$F_1(t) = A_0 \cos\{\omega[t + j(t)] + \theta_0\} \tag{2}$$
$$= F_{\mathrm{pure}}(t) + n_{\mathrm{jitter}}(t). \tag{3}$$

当 $\omega j(t) \ll 1$时，抖动导致的噪声 $n_{\mathrm{jitter}}(t)$ 表示如下：

$$n_{\mathrm{jitter}}(t) \approx -\omega j(t) A_0 \sin(\omega t + \theta_0). \tag{4}$$

这表示当信号接近零时，抖动会很明显，即 $F_1(t) \approx 0$，而当 $j(t)$ 较大时，抖动更容易测量。抖动的概念示意图见 图 1(a)。(ii) 幅度调制是波的振幅随时间发生的变化。若存在幅度调制，则 A_0 会被 $A_0 + A_{\mathrm{M}}(t)$ 取代，其中 $A_{\mathrm{M}}(t)$ 是表示时刻 t的幅度调制的连续函数。纯正弦波 $F_{\mathrm{pure}}(t)$ 将变为

$$F_2(t) = [A_0 + A_{\mathrm{M}}(t)] \cos(\omega t + \theta_0) \tag{5}$$
$$= F_{\mathrm{pure}}(t) + n_{\mathrm{AM}}(t). \tag{6}$$

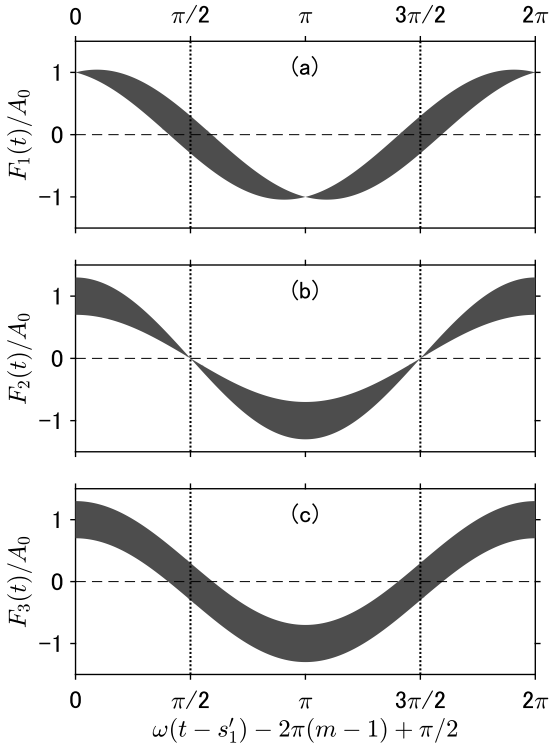


图1。（a）由抖动、（b）幅度调制和（c）PI噪声调制的正弦信号示意图。水平轴表示纯正弦波的相位。时间 $t = s_1'$ 是纯正弦波的第一个过零时间，m 是周期数。填充带表示重复过程中波动的范围。

幅度调制 $n_{\mathrm{AM}}(t)$ 引起的噪声表示如下：

$$n_{\mathrm{AM}}(t) = A_{\mathrm{M}}(t) \cos(\omega t + \theta_0). \tag{7}$$

因此，与抖动的变化趋势相反，当 $\cos(\omega t + \theta_0)$ 增大时，幅度调制也会增大。幅度调制的概念示意图如图 1(b) 所示。(iii) 实际波形不仅包含抖动和幅度调制，还包含 PI 噪声，它代表所有未归类为抖动或幅度调制的其他类型的噪声。如果存在 PI 噪声，则纯正弦波 $F_{\mathrm{pure}}(t)$ 将变为

$$F_3(t) = F_{\mathrm{pure}}(t) + n_{\mathrm{PI}}(t), \tag{8}$$

其中 $n_{\mathrm{PI}}(t)$ 表示相位间隔噪声。相位间隔噪声的概念示意图如图 1(c) 所示。考虑这三种噪声模式后，总噪声 $n_{\mathrm{total}}(t)$ 和实际播放信号 $c(t)$ 可表示为

$$n_{\mathrm{total}}(t) = n_{\mathrm{jitter}}(t) + n_{\mathrm{AM}}(t) + n_{\mathrm{PI}}(t), \tag{9}$$
$$c(t) = F_{\mathrm{pure}}(t) + n_{\mathrm{total}}(t). \tag{10}$$

在实际的数字音频播放器中，$n_{\mathrm{jitter}}(t)$ 主要来自内部时钟模块，而 $n_{\mathrm{AM}}(t)$ 来自 DAC单元。DAC 的输出放大器会对 $n_{\mathrm{PI}}(t)$ 产生影响。

在所提出的方法中，我们关注非纯正弦波的过零点。$F_{\mathrm{pure}}(t)$ 的过零点表示为 s'_k，满足 $F_{\mathrm{pure}}(s'_k) = 0$；因此，它们可表示为

$$s'_k = \frac{1}{\omega}\left(k\pi - \frac{\pi}{2} - \theta_0\right), \tag{11}$$

其中，k 是过零点的索引。在 $t = s'_k$ 处，播放信号 $c(t)$ 不等于零，其值为

$$c(s'_k) = (-1)^k \omega A_0 j(s'_k) + n_{\mathrm{PI}}(s'_k), \tag{12}$$

这可由式（10）推导得出。

上述噪声分类同样适用于录音过程。尽管数据数组是一组离散变量，而非连续函数，但下文仍将其称为“波形”。录制纯正弦播放信号时，录制波形并不等于纯正弦波的采样数据点，因为其中包含（i）抖动、（ii）幅度调制和（iii）PI噪声。在实际的数字音频录音机中，抖动主要源自内部时钟模块，而幅度调制则源自模数转换器单元。模数转换器的驱动放大器会产生 PI 噪声。

正如本研究后文所示，对于抖动值较小（通常低于 100 ps）的新型音频设备，PI 噪声会变得与抖动相当。利用播放器和录音机的左右声道，可以分离抖动和 PI 噪声。

B. 数字音频播放器的建模

在此，我们展示上一节中引入的抖动、幅度调制噪声和相位间隔噪声如何在实际回放过程中实现。因此，我们引入了一个数字音频播放器模型。单声道数字音频播放器的示意图如图2（a）所示。参数固定为与第III 节所述实验条件相同的值，也可以根据实验条件任意选择。假设播放器中的数模转换器是理想且无噪声的，其以 N_{P} 位精度执行转换，采样率为f_{P}。在本研究中，N_{P} 和 f_{P} 分别设置为24位和48千赫。

播放波形表示为 $v[i]$，其中i 为自然数，$v[i]$ 为 N_{P} 位有符号整数。$v[i]$ 的长度为 N_{total}。本研究中使用的波形 $v[i]$ 如图 2(b)所示。如图所示，它可以分为五个部分，标记为 (i)、(ii)、(iii)、(iv) 和 (v)。为提高图中的可见性，已减少点的数量。横轴表示索引 i。(i) 静音部分，此时播放波形在$1 \leqslant i \leqslant i_{\mathrm{main}} - N_{\mathrm{F}} - 1$内为 $v[i] = 0$。如下所示，i_{main} 是主体部分的第一个索引，N_{F} 是淡变部分的长度。我们设定 $i_{\mathrm{main}} = 480\ 000$ 和 $N_{\mathrm{F}} = 240\ 000$。因此，静音部分的时间长度为 $(i_{\mathrm{main}} - N_{\mathrm{F}})f_{\mathrm{P}}^{-1} \approx 5$ 秒。(ii) 淡入部分：播放

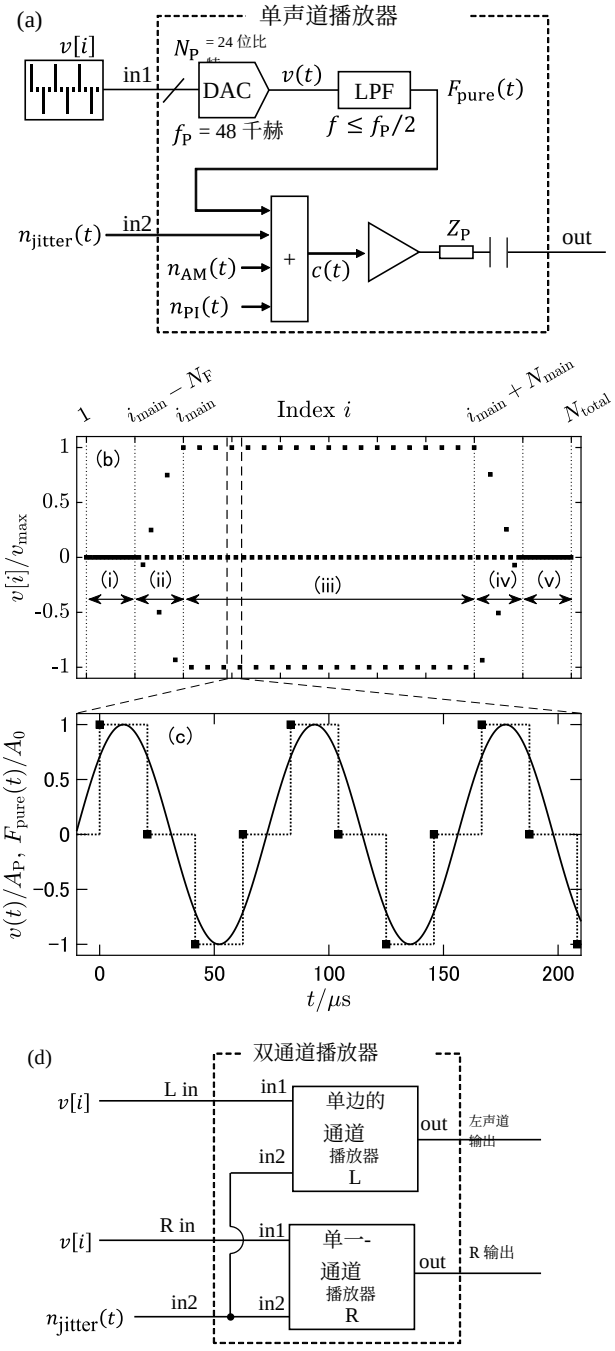


图 2. (a) 单声道数字音频播放器的模型图。(b) 播放波形；为提高可见度，已减少点的数量。(c) 数模转换器输出 $v(t)$（虚线）、低通滤波器后的信号 $F_{\mathrm{pure}}(t)$（实线）与播放波形 v_i（黑色方块）之间的关系。(d) 展示双通道数字音频播放器的模型图。

波形的形式为

$$v[i] = \left\{ v_{\min} + \left[1 + \cos\left(\pi \frac{i - i_{\text{main}}}{N_{\text{F}}} \right) \right] \times \frac{v_{\max} - v_{\min}}{2} \right\} \cos\left(2\pi \frac{\text{mod}[i - i_{\text{main}}, 4]}{4} \right) \tag{13}$$

波形位于 $i_{\text{main}} - N_{\text{F}} \leqslant i \leqslant i_{\text{main}} - 1$区域内，其中 $v_{\max}$:=$2^{N_{\text{P}}-1} - 1 = 8\ 388\ 607$ 是 N_{P}位有符号整数的最大值。淡入部分的初始振幅 $v_{\min}$ 设为 256。淡入部分的时间长度为 $N_{\text{F}} f^{-1}{}_{\text{P}} \approx 5$ 秒。(iii) 主体部分：播放波形是 ($v_{\max}, 0, -v_{\max}, 0$) 重复 $i_{\text{main}} \leqslant i \leqslant i_{\text{main}} + N_{\text{main}} - 1$ 次，即主体部分中的播放波形表示如下：

$$v[i] = v_{\max} \cos\left(2\pi \frac{\text{mod}[i - i_{\text{main}}, 4]}{4} \right). \tag{14}$$

在本研究中，我们设定 $N_{\text{main}} = 1\ 440\ 000$。因此，播放开始后 $(i_{\text{main}}) f^{-1}{}_{\text{P}} \approx 10$ 秒时，波形的主体部分开始，并持续 $(N_{\text{main}}) f^{-1}{}_{\text{P}}$ ≈30 秒。(iv) 淡出部分：该部分在主体部分之后开始，长度与淡入部分相同。淡出部分的波形是淡入部分的反向序列，如式（13）所示。(v) 第二静音部分：该部分在淡出部分之后开始，长度与第一静音部分相同。因此，播放信号的总长度，即其五个部分长度之和，为 $N_{\text{total}} f^{-1}{}_{\text{P}} = (2i_{\text{main}} + N_{\text{main}}) f^{-1}{}_{\text{P}} \approx 50$ 秒。

数模转换器的输出电压表示为 $v(t)$，是一个方波。对于 $t_{\text{P}} + (i-1) f^{-1}{}_{\text{P}} \leqslant t < t_{\text{P}} + i f^{-1}{}_{\text{P}}$，$v(t)$ 表示为 $v(t) = A_{\text{P}} v[i]/v_{\max}$。播放开始的时间为 t_{P}。在本节中，我们设定$t_{\text{P}} = -(i_{\text{main}} + N_{\text{main}}/6) f^{-1}{}_{\text{P}} \approx -15$ 秒。

数模转换器（DAC）之后连接一个低通滤波器（LPF）。假设 LPF 的截止频率为 $f_{\text{P}}/2$。由于方波 $v(t)$经过 LPF 平滑处理，LPF 后的输出电压变为式（1）所表示的纯正弦波。由于播放波形的主体部分由式（14）表示，因此纯正弦波的频率变为 $f_{\text{C}} := \omega/2\pi = f_{\text{P}}/4$。$v(t)$ 与 $F_{\text{pure}}(t)$ 之间的关系如图 2(c) 所示。虚线和实线分别表示 $v(t)$ 和 $F_{\text{pure}}(t)$。为了与 $v(t)$ 进行比较，播放波形 $v[i]$ 在 $t = t_{\text{P}} + (i-1) f^{-1}{}_{\text{P}}$ 和$v(t) = A_{\text{P}} v[i]/v_{\max}$ 所对应的位置以黑色方块表示。如上所述，$v(t)$ 中的 $3f_{\text{C}}$、$5f_{\text{C}}$、$\cdots$ 分量被 LPF 完全衰减，仅保留频率为 f_{C} 的纯正弦波。

播放信号 $c(t)$ 是通过将$n_{\text{jitter}}(t)$、$n_{\text{AM}}(t)$、$n_{\text{PI}}(t)$ 加到 $F_{\text{pure}}(t)$ 上得到的，如方程（9）和（10）所示。在实际的音频播放器中，抖动主要由 f_{P}的波动引起。然而，在该模型中，数模转换器和低通滤波器均为理想状态，抖动噪声是在低通滤波器之后加入的。输出由带有直流（DC）隔直电容的缓冲放大器产生。

双通道数字音频播放器的模型图见图 2(d)。它由两个单通道数字音频播放器组成，并假设它们的抖动输入等电位。该模型图用于第IIF节。

C. 数字音频录音机建模

为了表示实际录音过程中出现的抖动、幅度调制和PI噪声，我们在图3(a)中给出了单通道数字音频录音机的模型图。与

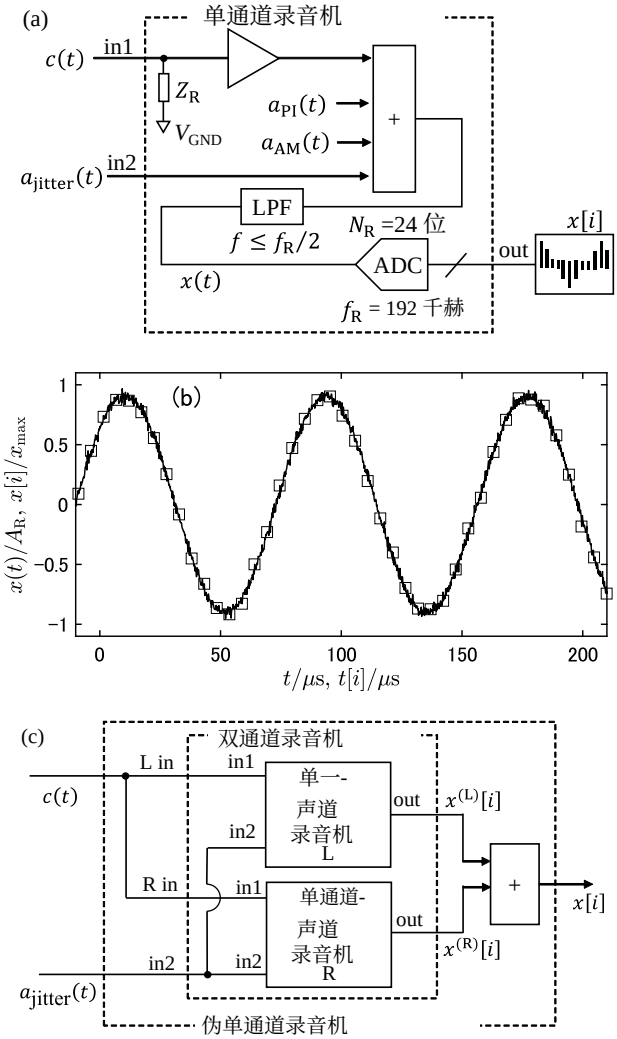


图3. (a) 单通道数字音频录音机的模型图。(b) 低通滤波器（LPF）之前的信号 $x(t)$（实线）与录制波形 x_i（白色方块）之间的关系。(c) 双通道数字音频录音机的模型图。图中展示了将双通道录音机用作单通道录音机的方法。

上一节中，参数固定为与第III节所述实验条件相同。播放信号 $c(t)$被馈送至缓冲放大器

具有输入阻抗 Z_{R}。随后，加入抖动 $a_{\mathrm{jitter}}(t)$、幅度调制 $a_{\mathrm{AM}}(t)$ 和 PI噪声 $a_{\mathrm{PI}}(t)$，并通过低通滤波器衰减高频分量。假设低通滤波器的截止频率为 $f_{\mathrm{R}}/2$。ADC 前的电压信号表示为 $x(t)$。它们之间的关系与方程 (9) 和 (10) 类似。因此，这种关系可以表示如下：

$$a_{\mathrm{total}}(t) = a_{\mathrm{jitter}}(t) + a_{\mathrm{AM}}(t) + a_{\mathrm{PI}}(t), \tag{15}$$

$$x(t) = \mathcal{LF}\{c(t) + a_{\mathrm{total}}(t)\}, \tag{16}$$

其中，$\mathcal{LF}\{\ \}$ 表示低频分量。

假设录音机中的模数转换器是理想且无噪声的，它以 N_{R} 位精度和 f_{R}的采样率进行转换。录制波形表示为 $x[i]$。第 i个值表示为 $x[i]$ =floor$[x_{\max}\{x(t[i])/A_{\mathrm{R}}\}]$，其中 A_{R} 是一个具有电压量纲的常数，而 $x_{\max}$: $= 2^{N_{\mathrm{R}}-1} - 1$ 是 N_{R} 位有符号整数的最大值。本研究中，N_{R} 和 f_{R} 分别设为24位和192千赫。模数转换的时序记为 $t[i]$。值 $t[i]$ 表示为

$$t[i] = t_{\mathrm{R}} + (i-1)f_{\mathrm{R}}^{-1}, \tag{17}$$

其中，t_{R} 表示录音开始的时间。我们假设模数转换器在播放开始前便已开始工作，并且 $t[i]$ 的长度足以录下完整的播放信号。当模数转换器之前的电压信号为 $x(t[i])$ $= A_{\mathrm{R}}$ 时，录制波形变为 $x[i] = x_{\max}$。在真实的音频录音机中，抖动源于 f_{R}的波动；然而，在此模型中，低通滤波器和模数转换器均为理想的，且抖动是在低通滤波器之前加入的。

.....之间的关系 $x(t)$ 与 $x[i]$ 如图 3(b) 所示。实线表示 $x(t)$。录制波形 $x[i]$ 在 $t = t[i]$ 和$x(t)$ 处以白色方块表示 $= x[i]$。横轴范围与图 2(c) 相同。图中实线的波动显示了人为引入的随机噪声。由于 ADC 的采样率与频率 $F_{\mathrm{pure}}(t)$ 之比为 $f_{\mathrm{R}}/f_{\mathrm{C}} = 16$，每个波长包含 16 个采样点[12]。

我们在图 3(c) 中展示了双通道数字音频录音机的模型图。该录音机由两台单通道数字音频录音机组成；假设单通道录音机的抖动输入等电位。为获得第IV A、IV B和IV C节所述的实验结果，我们将双通道录音机作为单通道录音机使用：连接两个模拟输入，并对其波形取平均值 $x^{(L)}[i]$ 和 $x^{(R)}[i]$。我们将这种设置称为“伪单通道录音机”。作为例外，我们分别分析波形 $x^{(L)}[i]$ 和 $x^{(R)}[i]$ ，以估算数字音频录音机的抖动。详情请参见第V D节。

D. ZCA

在 ZCA 中，我们首先寻找录音机中的电压信号，即 $x(t)$，穿过 t 轴时的时刻，同时 $0 \leqslant t \leqslant T$。为此，我们根据采样数据 $x[i]$ 重建连续函数 $x'(t)$，使其满足对于 $0 \leqslant t \leqslant T$，$x'(t) \approx x(t)$。重建过程包括三个步骤。(i) 为避免采样数据的边界效应，将窗函数 $w(t[i])$ 乘以 $x[i]$ ，得到 $x[i]w(t[i])$，其中 $w(t)$ 为 Blackman 型，其表达式如下：

$$w(t) = \begin{cases} 0.42 + 0.5\cos(\pi f_{\mathrm{R}} t/N) + 0.08\cos(2\pi f_{\mathrm{R}} t/N) & \\ \qquad (-N f_{\mathrm{R}}^{-1} \leqslant t < 0) & \\ 1 \qquad (0 \leqslant t \leqslant T) & \\ w(T-t) \qquad (T < t \leqslant T + N f_{\mathrm{R}}^{-1}). & \end{cases} \tag{18}$$

我们将 $N = 48\ 000$和 $T = 4Nf^{-1}{}_{\mathrm{R}} \approx 1$s 设定为...。因此，数据长度 $w(t[i])$ 变为 $6N$，而 $w(t)$ 的定义域变为 -0.25 s $\leqslant t \leqslant 1.25$ s。于是，数据长度 $x[i]w(t[i])$ 变为 $6N$。(ii) 乘以窗函数后，使用快速傅里叶变换（FFT）方法，按照 $N_{\mathrm{over}} = 64$ 的过采样因子对数据点进行插值。结果，数据点数量增加至 $6N_{\mathrm{over}}N$。根据所需精度调整 N_{over} 的值。通过进行数值模拟[13]，我们确认 $N_{\mathrm{over}} = 64$ 足够大。得益于 MATLAB[14]，中使用的 FFTW 库，计算时间几乎可以忽略不计。我们还对数据应用了带宽限制，以消除直流分量。(iii) 用直线连接插值点。经过这三个步骤，便可从离散数据 $\{x[i_{\mathrm{main}} - N], \cdots, x[i_{\mathrm{main}} + 5N]\}$ 得到连续函数 $x'(t)$。

图 4(a) 展示了获取 $x'(t)$ 的过程。横轴的范围与图 3(b) 相同。白色方块表示记录的波形 $x[i]$，黑色圆点表示通过 FFT 方法插值得到的点。实线表示 $x'(t)$。图 4(b) 展示了图 4(a) 中第一和第二个过零点附近的放大视图。为了便于查看，图 4(a)(b) 中的 FFT 插值采用了过采样因子 $N_{\mathrm{over}} = 4$ 。可以看到， $N_{\mathrm{over}} - 1$ 点插值在两个记录的数据点之间。如图 4(b) 所示，过零时刻标记为 $t = s_1, s_2, \cdots, s_M$，其中$M$ 表示当 $0 \leqslant t \leqslant T$ 时过零点的数量。由于抖动，得到的序列 $s_1, s_2, \cdots, s_M$并非等间隔。

其次，我们确定等间距的点 s'_k，这些点在 ZCA 的公式（11）中引入。为此，使用最小二乘法对 s_k 进行直线拟合。概念示意图见 图 5。拟合函数如下：

$$s'(k) = \frac{k-1}{2f'_{\mathrm{C}}} + s'_1, \tag{19}$$

其中 f'_{C} 是录音机测得的播放信号频率。中的偏差 s_k 相对于直线小于 100 ps；为便于观察，图中将其放大显示在 图 5 中。根据式（19）的拟合函数，

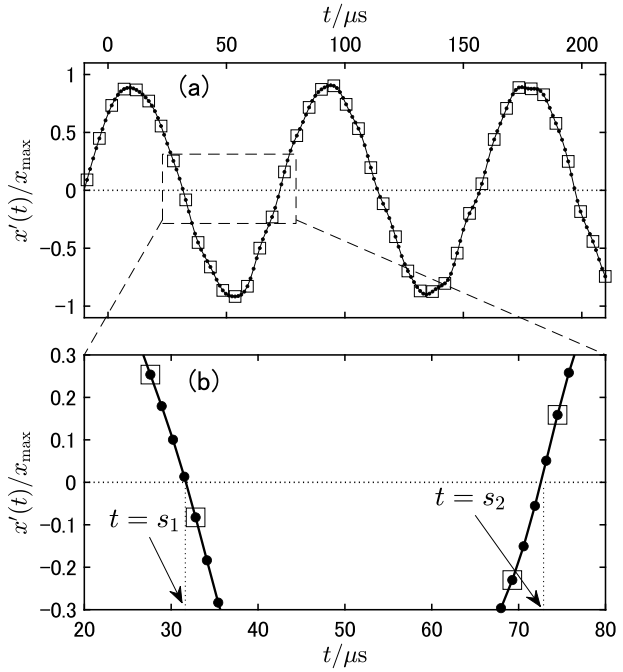


图 4. (a) 录制波形 x_i（白色方块）、通过 FFT 插值得到的点（黑色圆圈）以及连续函数 $x'(t)$（实线）。为便于观察，FFT 插值采用的过采样因子为 $N_{\text{over}} = 4$，而非 $N_{\text{over}} = 64$。(b) 20 µs $\leqslant t \leqslant$ 80 µs 的放大图。

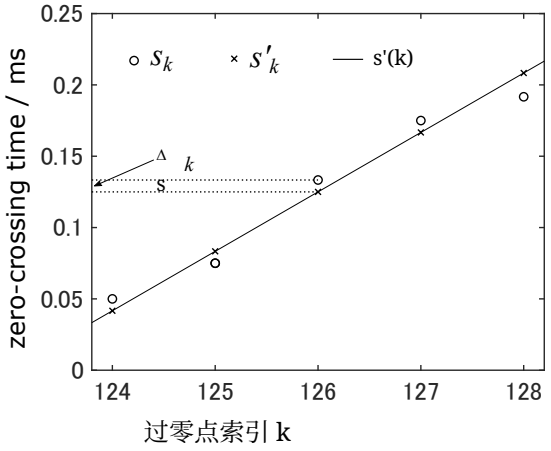


图 5. 过零时间（ms）与过零索引 k 的关系。s_k 表示录制数据的过零时间，s'_k 表示对应纯正弦波的过零时间。相邻的时间 s'_k 等间隔，而 s_k 则不等间隔。为便于观察，s_k 与s'_k 之间的差异经过了极度放大。

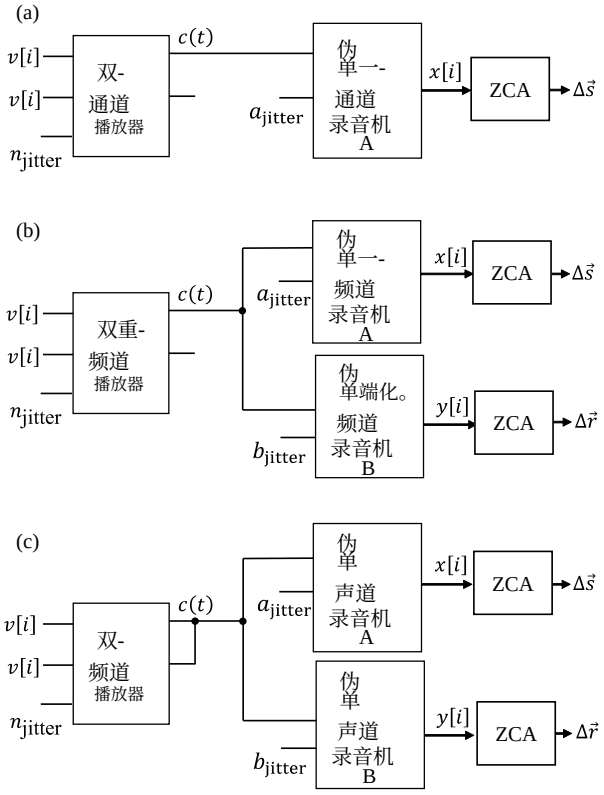


图6。(a) SRS。(b) DRS。(c) 用于将抖动与相位间隔噪声分离的设置。

获得第 k 个等距点 $s'_k := s'(k)$。频率 f'_C 是 $0 \leqslant t \leqslant T$ 期间的平均频率。因此，该分析对频率为 $f \geqslant 1/T$ 的短期漂移敏感，但对长期漂移不敏感。

在 ZCA 中，我们最终得到 ZCF，它给出 s_k 和 s'_k 之间的差值，表达式如下：

$$\Delta s_k = s'_k - s_k. \tag{20}$$

当 PI 噪声 $n_{\text{PI}}(t)$ 和录音机噪声 $a_{\text{total}}(t)$ 均可忽略时，ZCF Δs_k 会转变为抖动 $j(s'_k)$。由于重建函数 $x'(t)$ 每个周期会两次穿过 t 轴，因此可以按照$f_Z = 2f_C$的重复频率获得 ZCF 值。换言之，由 Δs_k 重建的 $j(t)$ 的带宽变为 $f \leqslant f_Z/2 = f_C$。这是预期的结果，因为抖动类似于频率调制，而在频率调制中，不可能传输高于载波频率的频率。如果只观察上升沿或下降沿的过零点，则 $j(t)$ 的带宽会受到限制，变为 $f \leqslant f_C/2$，这不足以完美地重建 $j(t)$。

E. 单录音机设置（SRS）和 DRS

首先，我们考虑一台播放器与一台录音机相连的情况[图 6(a)]。我们将这种情况称为

“单录音机设置（SRS）”。在 SRS 中，播放器噪声和录音机噪声均包含在 ZCF 中，并表示为 Δs_k。播放器噪声、录音机噪声与 ZCF 之间的关系如下所示：

$$(-1)^k \omega A_0 \Delta s_k = n_{\text{jitter}}(s'_k) + n_{\text{PI}}(s'_k) + a_{\text{jitter}}(s'_k) + a_{\text{PI}}(s'_k). \quad (21)$$

下文中，$\mathbb{V}\{\,\}$ 表示数据的方差。根据公式（21），ZCF 的方差 $\mathbb{V}\{\Delta s_k\}$ 变为

$$\mathbb{V}\{\Delta s_k\} = (\sigma_{n1})^2 + (\sigma_{a1})^2, \quad (22)$$

其中，σ_{n1} 和 σ_{a1} 分别是播放器和录音机的 ZCF 均方根（RMS）。更具体地说，σ_{n1} 和 σ_{a1} 可表示如下：

$$(\sigma_{n1})^2 = \mathbb{V}\{j(s'_k)\} + \frac{\mathbb{V}\{n_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}, \quad (23)$$

$$(\sigma_{a1})^2 = \frac{\mathbb{V}\{a_{\text{jitter}}(s'_k)\} + \mathbb{V}\{a_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}. \quad (24)$$

方程（22）左侧可使用实验数据获得，其结果将在第 IVA 节中描述。

<escape>其次，我们考虑一台播放器与两台录音机相连接的情况 [图 6(b)]；这种配置称为“DRS”。在 DRS 中，每个 ZCF 都包含播放器噪声和录音机噪声，分别表示为 Δs_k 和 Δr_k。对于 DRS，公式（21）仍然成立。与公式（21）一样，Δr_k 可表示为</escape>

$$(-1)^k \omega A_0 \Delta r_k = n_{\text{jitter}}(r'_k) + n_{\text{PI}}(r'_k) + b_{\text{jitter}}(r'_k) + b_{\text{PI}}(r'_k), \quad (25)$$

其中 r'_k 是利用录音机B采集的波形（表示为$y[i]$）通过 ZCA 获得的等间隔时间，$b_{\text{jitter}}(t)$ 是录音机B的抖动，而 $b_{\text{PI}}(t)$ 是录音机B的 PI噪声。由于两台录音机同时对同一台播放器输出的播放信号进行采样，因此在实际情况下，我们可以认为 $s'_k = r'_k$。请注意，s'_k 和 r'_k 具有相同的索引 k。因此，r'_k 式（25）中的 s'_k 可以替换为 s'_k，进而得到以下结果：

$$(-1)^k \omega A_0 \Delta r_k = n_{\text{jitter}}(s'_k) + n_{\text{PI}}(s'_k) + b_{\text{jitter}}(s'_k) + b_{\text{PI}}(s'_k). \quad (26)$$

利用方程 (21) 和 (26)，可得以下四个方程：

$$\mathbb{V}\{\Delta s_k\} = (\sigma_{n2})^2 + (\sigma_{a2})^2, \quad (27)$$

$$\mathbb{V}\{\Delta r_k\} = (\sigma_{n2})^2 + (\sigma_{b2})^2, \quad (28)$$

$$\mathbb{V}\{\Delta s_k - \Delta r_k\} = (\sigma_{a2})^2 + (\sigma_{b2})^2, \quad (29)$$

$$\mathbb{V}\{\Delta s_k + \Delta r_k\} = 4(\sigma_{n2})^2 + (\sigma_{a2})^2 + (\sigma_{b2})^2, \quad (30)$$

其中，σ_{n2}、σ_{a2}和 σ_{b2} 分别是播放器、录音机A和录音机B的ZCF均方根。需要注意的是，播放器的影响被抵消了在 $\Delta s_k - \Delta r_k$中。此外，播放器在 $\Delta s_k + \Delta r_k$中产生了双重贡献。这是因为 s'_k 和 r'_k 对于两台乐器都是相同的，尽管 Δs_k 和 Δr_k是在不同的乐器上测量的。

这些方程左侧的值可利用实验数据得到。利用方程 (27)–(29)，我们可以分别评估播放器中的噪声 (σ_{n2}) 和录音机中的噪声 (σ_{a2} 和 σ_{b2})。方程 (30) 可用于验证计算结果。实验结果见第 IVB小节。

这些噪声，σ_{n2}、σ_{a2} 和 σ_{b2}，可表示为

$$(\sigma_{n2})^2 = \mathbb{V}\{j(s'_k)\} + \frac{\mathbb{V}\{n_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}, \quad (31)$$

$$(\sigma_{a2})^2 = \frac{\mathbb{V}\{a_{\text{jitter}}(s'_k)\} + \mathbb{V}\{a_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}, \quad (32)$$

$$(\sigma_{b2})^2 = \frac{\mathbb{V}\{b_{\text{jitter}}(s'_k)\} + \mathbb{V}\{b_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}. \quad (33)$$

F. 将抖动与PI噪声分离

最后，我们考虑 图 6(c) 所示的情况。该设置使我们能够将播放器的抖动与 PI 噪声分离。此设置不同于图 6(b)，后者将播放器的左、右声道信号捆绑在一起。播放器的 ZCF 均方根（表示为 σ_{n3}）可像在 DRS 中那样获得。

如图2(d)所示，双通道播放器由两个抖动相当的单通道播放器组成。此外，两个单通道播放器的PI噪声彼此独立。因此，在该模型的假设下，σ_{n3} 中的PI噪声降低至

$$(\sigma_{n3})^2 = \mathbb{V}\{j(s'_k)\} + \frac{1}{2}\frac{\mathbb{V}\{n_{\text{PI}}(s'_k)\}}{(\omega A_0)^2}. \quad (34)$$

利用方程（31）和（34），可得 $\mathbb{V}\{j\,(s'_k)\}$。因此，可以通过测量 σ_{n2} 和 σ_{n3}来确定抖动的均方根值。该关系表示为

$$\text{dev}\{j(s'_k)\} = \sqrt{2(\sigma_{n3})^2 - (\sigma_{n2})^2}, \quad (35)$$

其中，dev$\{\,\}$ 表示偏差，即dev$\{\,\} := \sqrt{\mathbb{V}\{\,\}}$.

III. 实验流程

A. 音频播放器和录音机

在实验中，我们使用了三台相同的便携式音频设备（DR-100MKIII；TASCAM，日本）。这些设备具有多项优点：不受交流电源质量的影响，从而确保测量结果具有可重复性和独立性；价格低廉且易于获得。

三台设备中的一台（编号1）用作播放器，其余两台（编号2和编号3）用作录音机。图7（a）和（b）分别对应于第IIE小节所述的 SRS 和 DRS。图7（c）

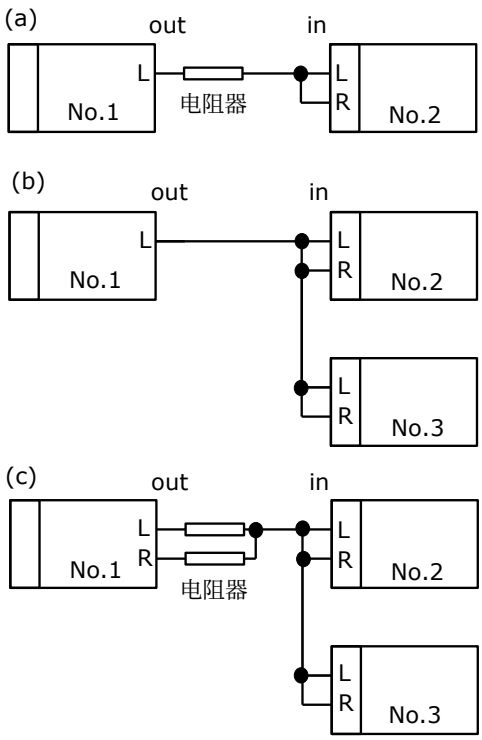


图7。（a）SRS；设备1号的左声道输出同时馈送至设备2号的左、右声道。接入了一个200 Ω的匹配电阻。（b）DRS；设备1号的左声道输出同时馈送至设备2号和设备3号的左、右声道。（c）输出为设备1号左右声道的总和，并同时馈送至设备2号和设备3号的左、右声道。

展示了分离抖动与 PI 噪声所需的设置，如第IIF小节所述。

录音机的设备设置汇总于表I中。三种设置的录音电平为

表I。录音机的设备设置。

文件格式	WAV24
采样率 192千赫兹	
文件类型	立体声
XRI	OFF
双重录音	OFF
信号源	外部线路
A/D滤波器	FIR1
双ADC	ON
低切	OFF
录音电平 +3分贝	

通过插入匹配电阻器进行调整，使其相等，

如图7（a）（c）所示。这对于防止录音中的电平变化影响相位间隔噪声是必要的。

B. 如何同步不同的波形

播放文件的详细信息见第 II B 小节。主体部分的长度为 $N_{\mathrm{main}} f^{-1}{}_{\mathrm{P}} \approx 30$ 秒。淡入和淡出部分的长度均为 $N_{\mathrm{F}} f^{-1}{}_{\mathrm{P}} \approx 5$ 秒。因此，播放信号中有 $(N_{\mathrm{main}} + 2N_{\mathrm{F}})/4 = 480\,000$ 个正弦波周期，录制波形也是如此。在我们的改进的TDA中，分析程序计算来自两台录音机的两个正弦波中的周期数，并分配一个共同的过零索引。该特性在第 IVB 小节中具有重要意义。

IV. 测量结果

A. SRS

使用 SRS [图7(a)]，我们播放并录制了 $f_{\mathrm{C}} = 12$ 千赫的正弦波。为了消除播放器中非时钟源产生的低频噪声，对录制波形进行了有限带宽范围内的处理，范围为 $f_{\mathrm{C}} - B_{\mathrm{w}} \leqslant f \leqslant f_{\mathrm{C}} + B_{\mathrm{w}}$。因此，我们分析了 $1/T \leqslant f \leqslant B_{\mathrm{w}}$带宽内的抖动。在此分析中，$B_{\mathrm{W}}$ 设置为 6 千赫。随后，使用 MATLAB 代码获得 ZCF。图 8(a) 显示了获得的 ZCF，Δs_1、Δs_2、$\cdots$和Δs_M。图 8(b) 显示了获得的 ZCF 的分布，其形状类似于高斯曲线。如公式（22）所示，该 ZCF 包含播放器和录音机产生的抖动及 PI 噪声的影响。ZCF 的均方根为 $\{(\sigma_{n1})^2 + (\sigma_{a1})^2\}^{1/2} = 55.3$ 皮秒。

B. DRS

使用 DRS [图7(b)]，我们分别从Δs_k 和 Δr_k 2号和3号设备中获得了 ZCFΔs_k 和 Δr_k 。图9 展示了 $\Delta s_k - \Delta r_k$和 $\Delta s_k + \Delta r_k$的分布，其中前者明显比后者窄。这是因为播放器的影响在 $\Delta s_k - \Delta r_k$中被抵消，而在 $\Delta s_k + \Delta r_k$中则产生双重贡献。根据实验数据，我们可以确定

$$E_1 = \mathrm{dev}\{\Delta s_k\}, \tag{36}$$
$$E_2 = \mathrm{dev}\{\Delta r_k\}, \tag{37}$$
$$E_3 = \mathrm{dev}\{\Delta s_k - \Delta r_k\}, \tag{38}$$
$$E_4 = \mathrm{dev}\{\Delta s_k + \Delta r_k\}, \tag{39}$$

其中，E_1、E_2、E_3和 E_4 分别是根据 $\{\Delta s_1, \cdots, \Delta s_M\}$、$\{\Delta r_1, \cdots, \Delta r_M\}$、$\{\Delta s_1 - \Delta r_1, \cdots, \Delta s_M - \Delta r_M\}$和 $\{\Delta s_1 + \Delta r_1, \cdots, \Delta s_M + \Delta r_M\}$计算得出的标准差。$E_1$、$E_2$、$E_3$和 E_4的值分别为 $E_1 = 56.0$ ps、$E_2 = 56.1$ ps、$E_3 = 50.6$ ps和 $E_4 = 100.0$ ps。利用公式（27）、（28）和（29），可得以下均方根值

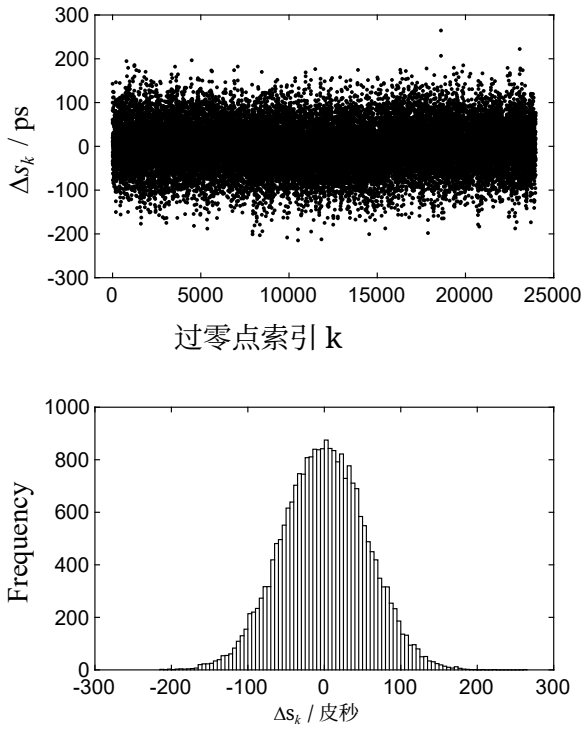


图8. (a) 1 s 内获得的 ZCF。共有 $M = 24\ 000$个过零点。(b) 获得的 ZCF 直方图。

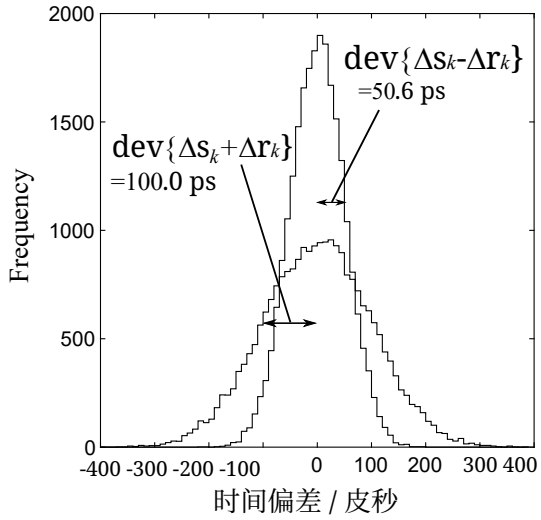


图9。 $\Delta s_k - \Delta r_k$ 和 $\Delta s_k + \Delta r_k$的直方图。前者和后者的均方根值分别为50.6 ps和100.0 ps。

播放器的 ZCF：

$$\sigma_{n2} = 43.1 \text{ ps}, \tag{40}$$

以及录音机的均方根值

$$\sigma_{a2} = 35.7 \text{ ps}, \tag{41}$$
$$\sigma_{b2} = 35.9 \text{ ps}. \tag{42}$$

这些数值满足方程（30），表明两台录音机之间不存在相关性。

C. 播放器的抖动和 PI 噪声

在本小节中，我们分别测量播放器的抖动和 PI 噪声。我们使用图7(c) 所示的设置测量了 ZCF。采用与上一小节相同的方法，我们确定了播放器 ZCF 的均方根值，σ_{n3}。所得数值为

$$\sigma_{n3} = 33.5 \text{ ps}. \tag{43}$$

该值小于上一小节计算出的 $\sigma_{n2} = 43.1$ ps。如方程（31）–（34）所示，这一差异可解释为平均处理导致 PI 噪声降低。由此，我们得到

$$\text{dev}\{j(s'_k)\} = 19.7 \text{ ps}, \tag{44}$$
$$\frac{\text{dev}\{n_{\text{PI}}(s'_k)\}}{\omega A_0} = 38.4 \text{ ps}. \tag{45}$$

这些结果表明，在评估抖动值较小（通常低于 100 ps）的新型音频设备的抖动时，必须考虑 PI 噪声。正如第IVA节所述，抖动是在 1 Hz $\leqslant f \leqslant$ 6 kHz 带宽内计算的。与此同时，PI 噪声是在 6 kHz $\leqslant f \leqslant$ 18 kHz 带宽内分析的。

V. 讨论

A. 抖动检测限

使用所提出的方法，测量抖动的精度高于使用现有方法时的精度[1–5]。这部分归功于近年来模数转换器性能的提升。我们测量了采样率为192千赫、位深为24位的波形，而先前的研究使用的是采样率为44.1或48千赫、位深为16位的数模转换器[5]。所提出方法的检测限受量化噪声影响，量化噪声表示为 j_{LSB}，可通过求解以下方程得出：

$$x_{\max}\frac{A_0\omega j_{\text{LSB}}}{A_{\text{R}}} = 1. \tag{46}$$

结果变为 $j_{\text{LSB}} \approx 1.76$ ps。

σ_{n2} 的平均值和 $\text{dev}\{j(s'_k)\}$ 的平均值按如下方式获得：

$$\sigma_{n2} = 42.34(14) \text{ ps}, \tag{47}$$
$$\text{dev}\{j(s'_k)\} = 21.1(6) \text{ ps}, \tag{48}$$

使用不同十个时间域的 σ_{n2} 和 $\text{dev}\{j(s'_k)\}$ 值。括号 ( ) 中的数字表示均值的标准差。因此，检测限

介于 j_{LSB} 和 21.1 ps 之间，预计小于 10 ps。此外，检测限取决于所使用的录音机；使用性能更高的录音机可以进行更精确的测量。

B. 回放噪声的相位依赖性

现在我们来考察总回放噪声的相位依赖性，即 $n_{\mathrm{total}}(t)$。人们可能会认为，相位依赖性分析能够分离抖动、幅度调制和 PI噪声；遗憾的是，如下所示，这种方法并不可行。

首先，我们证明 $n_{\mathrm{total}}(t)$ 的相位依赖性总能用两个参数 A 和 B 表示。为此，我们将时间 t 表示为相位 θ 和周期数 m 的函数，如下所示：

$$t(\theta,m)=\frac{\theta-\theta_0+2\pi(m-1)}{\omega}, \tag{49}$$

其中，将 m 的最大值（记作 $m_{\max}$）设为

$$m_{\max}=\mathrm{floor}\left(\omega T/2\pi\right), \tag{50}$$

并将 θ 的取值范围限制为

$$0\leqslant\theta<2\pi. \tag{51}$$

值得注意的是，$t(\pi/2,m)=s'_{2m-1}$，而 $t(3\pi/2,m)=s'_{2m}$。为简便起见，以下将 $n_{\mathrm{total}}(t(\theta,m))$ 的表达式替换为 $n_{\mathrm{total}}(\theta,m)$。同样的规则也适用于 $j(t)$、$A_{\mathrm{M}}(t)$ 和 $n_{\mathrm{PI}}(t)$。因此，回放噪声表示为

$$\begin{aligned}&n_{\mathrm{total}}(\theta,m)\\&=A_0\omega j(\theta,m)\sin\theta+A_{\mathrm{M}}(\theta,m)\cos\theta+n_{\mathrm{PI}}(\theta,m).\end{aligned} \tag{52}$$

下文中，$\mathbb{V}\{n_{\mathrm{total}}(\theta,m)\}$ 表示根据 $n_{\mathrm{total}}(\theta,2)$、$n_{\mathrm{total}}(\theta,3)$ $\cdots$ 以及$n_{\mathrm{total}}(\theta,m_{\max}-1)$计算得到的方差。由式（52）可得

$$\begin{aligned}\mathbb{V}\{n_{\mathrm{total}}(\theta,m)\}&=\frac{1-\cos(2\theta)}{2}(\omega A_0)^2\mathbb{V}\{j(\theta,m)\}\\&\quad+\frac{1+\cos(2\theta)}{2}\mathbb{V}\{A_{\mathrm{M}}(\theta,m)\}+\mathbb{V}\{n_{\mathrm{PI}}(\theta,m)\} \qquad (53)\\&=A\cos(2\theta)+B, \qquad (54)\end{aligned}$$

其中，设

$$A:=\frac{\mathbb{V}\{A_{\mathrm{M}}(\theta,m)\}-(\omega A_0)^2\mathbb{V}\{j(\theta,m)\}}{2}, \tag{55}$$

$$B:=\mathbb{V}\{n_{\mathrm{PI}}(\theta,m)\}+\frac{(\omega A_0)^2\mathbb{V}\{j(\theta,m)\}+\mathbb{V}\{A_{\mathrm{M}}(\theta,m)\}}{2}. \tag{56}$$

我们假设 $\mathbb{V}\{j(\theta,m)\}$、$\mathbb{V}\{A_{\mathrm{M}}(\theta,m)\}$ 和 $\mathbb{V}\{n_{\mathrm{PI}}(\theta,m)\}$ 不依赖于 θ。因此，只要上述假设成立，$\mathbb{V}\{n_{\mathrm{total}}(\theta,m)\}$ 的相位依赖性始终可以用两个参数 A 和 B 表示。

因此，可以确认以下行为：(i) 当抖动和幅度调制不可忽略时，偏移量 B 不等于 PI噪声；(ii) 当抖动和幅度调制相当时，幅度 A 消失；(iii) 当 PI噪声可以忽略时，可以通过计算$B\pm A$来获得抖动和幅度调制；(iv) 当 PI噪声不可忽略时，无法从 A 和 B中获得抖动、幅度调制和 PI噪声。行为 (iv) 表明，需要进一步考虑如何将 $\mathbb{V}\{j(\theta,m)\}$ 与 B区分开来。为此设计的流程将在第 II F 节和第 IVC节中说明。

C. 与 CSM 的比较

如引言中所述，DRS 与 CSM 的设置类似[8]。CSM 可视为 FDA 和 DRS 的组合，而所提出的方法是 ZCA 和 DRS 的组合。对于 CSM，通过平均处理可降低来自两台乐器的噪声，交叉频谱可达到被测设备的功率谱。

基于 CSM 的商用产品旨在评估频率高于 1 兆赫的时钟发生器[15][16]。这主要是因为在音频频段内进行频率转换在技术上具有挑战性。因此，迄今为止尚未使用 CSM 对音频信号进行评估。所提出的方法是 ZCA 和 DRS 的组合，能够处理音频信号，作为 CSM 的替代方案似乎是可行的。

D. 录音机的抖动和相位间隔噪声

在本小节中，我们分别求得录音机的抖动和PI噪声。如图3(c)所示，双通道录音机由两台具有共模抖动的单通道录音机组成。左、右输入端的PI噪声相互独立，分别表示为$a_{\mathrm{PI,\ L}}(t)$和$a_{\mathrm{PI,\ R}}(t)$。录制波形L和R的ZCF分别表示为 $\Delta s_k^{(\mathrm{L})}$ 和 $\Delta s_k^{(\mathrm{R})}$。与方程(27)–(33)类似，我们得到以下方程。

$$\begin{aligned}(E_5)^2&=\mathbb{V}\{\Delta s_k^{(\mathrm{L})}\}\\&=(\sigma_{n2})^2+\frac{\mathbb{V}\{a_{\mathrm{jitter}}(s'_k)\}}{(\omega V_0)^2}+\frac{\mathbb{V}\{a_{\mathrm{PI,L}}(s'_k)\}}{(\omega V_0)^2}, \qquad (57)\\(E_6)^2&=\mathbb{V}\{\Delta s_k^{(\mathrm{R})}\}\\&=(\sigma_{n2})^2+\frac{\mathbb{V}\{a_{\mathrm{jitter}}(s'_k)\}}{(\omega V_0)^2}+\frac{\mathbb{V}\{a_{\mathrm{PI,R}}(s'_k)\}}{(\omega V_0)^2}, \qquad (58)\\(E_7)^2&=\mathbb{V}\{\Delta s_k^{(\mathrm{L})}-\Delta s_k^{(\mathrm{R})}\}\\&=\frac{\mathbb{V}\{a_{\mathrm{PI,L}}(s'_k)\}+\mathbb{V}\{a_{\mathrm{PI,R}}(s'_k)\}}{(\omega V_0)^2}, \qquad (59)\\(E_8)^2&=\mathbb{V}\{\Delta s_k^{(\mathrm{L})}+\Delta s_k^{(\mathrm{R})}\}\\&=4(\sigma_{n2})^2+4\frac{\mathbb{V}\{a_{\mathrm{jitter}}(s'_k)\}}{(\omega V_0)^2}\\&\quad+\frac{\mathbb{V}\{a_{\mathrm{PI,L}}(s'_k)\}+\mathbb{V}\{a_{\mathrm{PI,R}}(s'_k)\}}{(\omega V_0)^2}, \qquad (60)\end{aligned}$$

其中， E_5、 E_6、 E_7和 E_8 分别是根据 $\{\Delta s_1^{(L)}, \cdots, \Delta s_M^{(L)}\}$、 $\{\Delta s_1^{(R)}, \cdots, \Delta s_M^{(R)}\}$、 $\{\Delta s_1^{(L)} - \Delta s_1^{(R)}, \cdots, \Delta s_M^{(L)} - \Delta s_M^{(R)}\}$和 $\{\Delta s_1^{(L)} + \Delta s_1^{(R)}$、 $\cdots$、 $\Delta s_M^{(L)} + \Delta s_M^{(R)}\}$计算得到的标准差。使用本文生成的实验数据， E_5、 E_6、 E_7和 E_8 分别为63.7、63.1、61.9和110.6 ps。因此，我们得到

$$\frac{\mathrm{dev}\{a_{\mathrm{PI,L}}(s'_k)\}}{\omega V_0} = 44.3 \text{ ps}, \tag{61}$$

$$\frac{\mathrm{dev}\{a_{\mathrm{PI,R}}(s'_k)\}}{\omega V_0} = 43.3 \text{ ps}, \tag{62}$$

$$(\sigma_{n2})^2 + \frac{\mathbb{V}\{a_{\mathrm{jitter}}(s'_k)\}}{(\omega V_0)^2} = (45.9\mathrm{ps})^2. \tag{63}$$

在 $\sigma_{n2} = 43.1$ 方程（40）中取 ps，可得

$$\frac{\mathrm{dev}\{a_{\mathrm{jitter}}(s'_k)\}}{\omega V_0} = 15.7 \text{ ps}. \tag{64}$$

本实验使用的播放器和录音机是同一款产品；因此，我们预计播放器和录音机的抖动相近。方程（44）和（64）的结果支持这一预期。

VI. 总结与展望

本文提出了一种高效且强大的方法，可用于高精度抖动测量。该方法基于两个关键要素：ZCA 和 DRS。通过分析录制波形（$x[i]$ 和 $y[i]$），ZCA 使我们能够确定录音机中电压信号（s_k 和 r_k）的过零时间，以及纯正弦波（s'_k 和 r'_k）的过零时间。它们各自的差值“ZCF（Δs_k 和 Δr_k）”包含播放器噪声（σ_{n2}）和录音设备噪声（σ_{a2} 和 σ_{b2}）的信息。若使用 DRS 测量 ZCF，则可以通过计算 ZCF（$\mathbb{V}\{\Delta s_k + \Delta r_k\}$ 和 $\mathbb{V}\{\Delta s_k - \Delta r_k\}$）之间的正相关和负相关，从 ZCF 中消除录音机噪声。因此，可以独立确定播放器噪声。播放器噪声（σ_{n2}）由抖动（$\mathrm{dev}\{j(s'_k)\}$）和 PI 噪声（$\mathrm{dev}\{a_{\mathrm{PI}}(s'_k)\}$）构成。要将两者分离，需要进行一些考量。一种可行的做法是，在左声道和右声道输出合并时测量播放器噪声（σ_{n3}）。

我们使用商用音频设备演示了所提出的方法。抖动和PI噪声的均方根值分别确定为$\mathrm{dev}\{j(s'_k)\} \approx 20$ ps和$\mathrm{dev}\{n_{\mathrm{PI}}(s'_k)\}/(\omega A_0) \approx 40$ ps。这些结果表明，所提出的方法能够评估小于PI噪声的抖动值。所提出方法的高精度意味着它将成为未来开发超高性能设备的有力手段。利用此类设备，可以对音乐等现实生活中的声音进行更明确、更定量的研究。这将为未来的研究奠定基础。

附录：数值模拟

如IID节所述，我们开发了一个用 MATLAB 代码编写的 ZCA 程序。为了确认该程序能够准确推导出 ZCF，我们准备了多个录制波形模拟样本，并在纯正弦波中加入人工抖动、幅度调制、PI噪声和录音设备噪声。这些模拟波形表示如下：

$$x[i] = \mathrm{floor}\left[x_{\max}\frac{A_0\cos(\omega t[i]+\theta_0) - A_0\omega j(t[i])\sin(\omega t[i]+\theta_0)}{A_{\mathrm{R}}} + \frac{A_{\mathrm{M}}(t[i])\cos(\omega t[i]+\theta_0) + n_{\mathrm{PI}}(t[i]) + a_{\mathrm{total}}(t[i])}{A_{\mathrm{R}}}\right], \tag{A1}$$

其中， $j(t[i])$、 $A_{\mathrm{M}}(t[i])$、$n_{\mathrm{PI}}(t[i])$和 $a_{\mathrm{total}}(t[i])$的幅度和带宽均任意选取。纯正弦波 $f_{\mathrm{C}} = \omega/2\pi$ 的频率在 $f_{\mathrm{P}}/4 - 120$ Hz与 $\leqslant f_{\mathrm{C}} \leqslant f_{\mathrm{P}}/4 + 120$ Hz之间变化，其中 $f_{\mathrm{P}} = 48$ 千赫为播放设备的采样频率，以此验证程序的准确性。 $t[i+1]$ 与 $t[i]$ 之间的时间间隔设为 $f^{-1}_{\mathrm{R}} \approx 5.2083$ μs，其中 $f_{\mathrm{R}} = 192$ 千赫为录音机的采样率。下面给出在以下三种条件下获得的结果：(i) $\mathrm{dev}\{j(t[i])\} \neq 0$ 和 $A_{\mathrm{M}}(t) = n_{\mathrm{PI}}(t) = a_{\mathrm{total}}(t) = 0$，(ii) $\mathrm{dev}\{A_{\mathrm{M}}(t[i])\} \neq 0$ 和 $j(t) = n_{\mathrm{PI}}(t) = a_{\mathrm{total}}(t) = 0$，以及 (iii) $\mathrm{dev}\{n_{\mathrm{PI}}(t[i])\} \neq 0$ 和 $j(t) = A_{\mathrm{M}}(t) = a_{\mathrm{total}}(t) = 0$。使用这些模拟波形，我们确认了程序的结果准确无误。

1. 模拟波形的生成

首先，我们生成不受带宽限制的人工抖动、幅度调制和相位间隔噪声，分别表示为$j_{\mathrm{FBW}}(t[i])$、$A_{\mathrm{M,\ FBW}}(t[i])$和$n_{\mathrm{PI,\ FBW}}(t[i])$。其表达式如下：

$$j_{\mathrm{FBW}}(t[i]) = J \cdot \mathrm{randn}[i], \tag{A2}$$

$$A_{\mathrm{M,FBW}}(t[i]) = A_0\omega J \cdot \mathrm{randn}[i], \tag{A3}$$

$$n_{\mathrm{PI,FBW}}(t[i]) = A_0\omega J \cdot \mathrm{randn}[i], \tag{A4}$$

其中，randn[ ] 是 MATLAB 函数，用于生成标准差为 1 的正态分布随机数。载波频率设为 $\omega/2\pi$ =11.884 877 千赫。载波幅度设为 A_0 =$0.9A_{\mathrm{R}} = 0.9$ FS，其中 FS 表示录音设备输入端的满量程。用J表示的抖动幅度设为 $J = 160$ ps。因此，人工抖动的偏差为 $\mathrm{dev}\{j_{\mathrm{FBW}}(t[i])\} = 160$ ps。根据采样定理， $j_{\mathrm{FBW}}(t[i])$的带宽为 $f_{\mathrm{R}}/2 = 96$ 千赫。

其次，我们对 $j_{\mathrm{FBW}}(t[i])$,$A_{\mathrm{M,FBW}}(t[i])$ 和 $n_{\mathrm{PI,FBW}}(t[i])$ 施加带宽限制，得到人工抖动 $j(t[i])$、幅度调制 $A_{\mathrm{M}}(t[i])$ 和 PI噪声 $n_{\mathrm{PI}}(t[i])$。所施加的带宽限制包括低通滤波器（ LPF ）和

带通滤波器（BPF）类型，表示为

$$j(t[i]) = \mathcal{LF}\{j_{\mathrm{FBW}}(t[i])\} \tag{A5}$$
$$A_{\mathrm{M}}(t[i]) = \mathcal{LF}\{A_{\mathrm{M,FBW}}(t[i])\} \tag{A6}$$
$$n_{\mathrm{PI}}(t[i]) = \mathcal{BPF}\{n_{\mathrm{PI,FBW}}(t[i])\} \tag{A7}$$

其中，$\mathcal{LF}\{\ \}$ 表示 0 6$f\ \leqslant\ B_{\mathrm{W}}$ 的频率分量，$\mathcal{BPF}\{\ \}$ 表示 $f_{\mathrm{C}} - B_{\mathrm{W}} \leqslant f \leqslant f_{\mathrm{C}} + B_{\mathrm{W}}$ 的频率分量。我们将低通滤波器的截止频率和带通滤波器的半带宽设为 $B_{\mathrm{W}} = 6$ 千赫，约为$f_{\mathrm{C}}/2$。由于带宽限制，人工抖动的偏差减小至 dev

$$\{j\ t[i]\ \} = (\ )\ (\ )$$

$J\{B_{\mathrm{W}}/\ f_{\mathrm{R}}/2\ \}^{1/2} = 40$ ps。同样，幅度调制和 PI噪声的偏差分别变为 $(\omega A_0)^{-1}$dev$\{A_{\mathrm{M}}(t[i])\}$ ≈40 ps 和 $(\omega A_0)^{-1}$dev$\{n_{\mathrm{PI}}(t[i])\} \approx 40\sqrt{2}$ ps。

最后，我们将公式（A1）中的参数分别设为 $x_{\max} = 2^{23} - 1$和 $\theta_0 = 0$。因此，创建了以下模拟波形：（i）纯正弦波与抖动，（ii）纯正弦波与幅度调制，以及（iii）纯正弦波与 PI噪声。这些模拟波形以 24 位 WAV 文件格式保存，然后加载。

2. 虚拟波形的 FDA

我们展示了模拟波形 (i)、(ii) 和 (iii) 的频谱，分别见 图 10(a)、(b) 和 (c)。横轴表示频率 f，纵轴表示以满刻度分贝每赫兹表示的功率谱密度 (PSD)$^{-1}$，其中 FS 表示满量程。

使用 (i) 进行 FDA 的结果如 图 10(a) 所示。在频率 $f\ \approx\ 12$千赫处可见一个峰值。其高度和宽度分别为 -5 满刻度分贝每赫兹$^{-1}$ 和 1 赫兹。该峰值对应于纯正弦波的功率谱密度；峰值面积约为 ($10^{-5/10}$ FS^{2}赫兹$^{-1}$) (1 赫兹) $= 10^{-5/10}$ FS^2。录音机吸收的载波相对功率表示如下

$$P_{\mathrm{C}}/\mathrm{FS}^2 = \frac{\langle\{F_{\mathrm{pure}}(t)\}^2\rangle}{A_{\mathrm{R}}^2} = \frac{(A_0/A_{\mathrm{R}})^2}{2}, \tag{A8}$$

其中，符号 $\langle\ \rangle$ 表示时间平均值。因此，P_{C} 计算为 $P_{\mathrm{C}} =$ (0.9 FS)$^2/2\ \approx 10^{-4/10}$ FS^2。该值对应于峰值面积。白噪声底位于约 -200满刻度分贝 Hz^{-1}的水平，对应于量化噪声；噪声底的面积计算为 ($10^{-200/10}$ FS^2 Hz^{-1})($f_{\mathrm{R}}/2$) $\approx 10^{-150/10}$ 2

FS。$N_{\mathrm{R}} = 24$ 位模数转换器的量化噪声理论信噪比计算为 $6.02N_{\mathrm{R}} + 1.76$ 分贝 =146 dBc[17]。因此，量化噪声的相对功率为 $10^{-146/10}\{(1\ \mathrm{FS})^2/2\} = 10^{-149/10}$ FS^2。该值接近图 10(a)中白噪声底的面积。在 6 千赫处可见矩形频谱 6 $f \leqslant 18$ 千赫。其顶部水平为 -155 满刻度分贝 Hz^{-1}。该矩形频谱对应于人工抖动的功率谱密度；矩形频谱的面积计算为 ($10^{-155/10}$FS^2 Hz$^{-1} - 10^{-195/10}$FS^2 Hz^{-1})(18 千赫 –

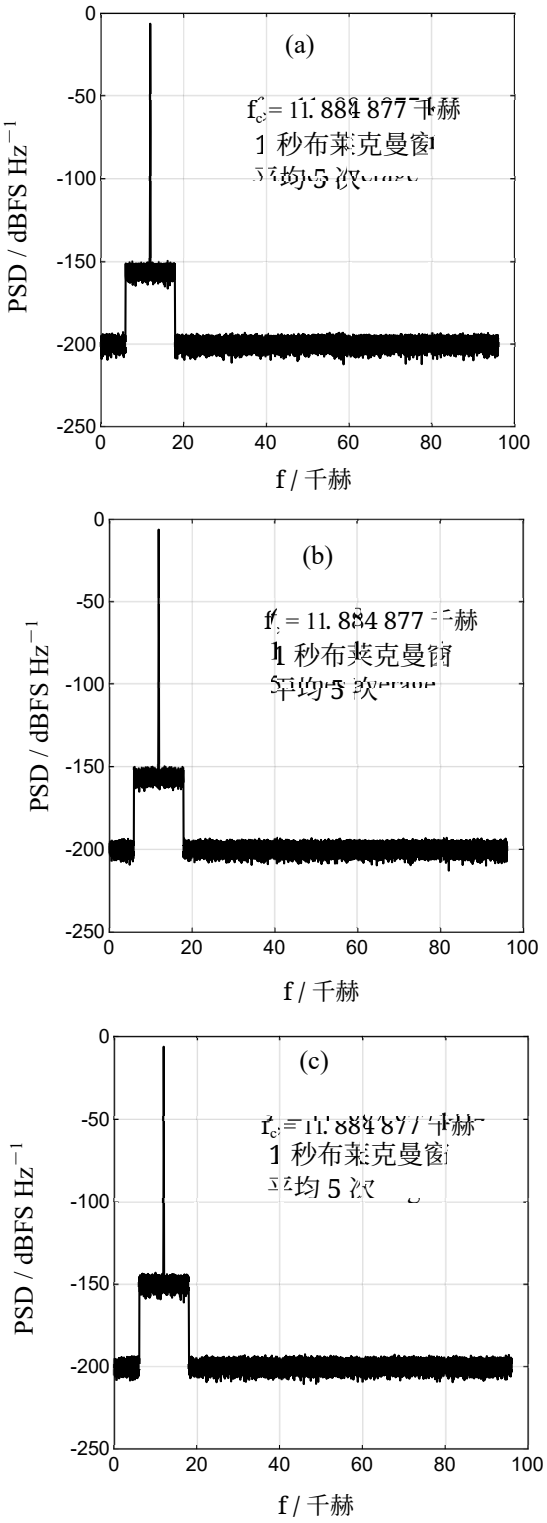


图 10。FDA 的结果。(a) (i) 的功率谱密度。(b) (ii) 的功率谱密度。(c) (iii) 的功率谱密度。

6 千赫) $= 10^{-114/10}\,\mathrm{FS}^2$。根据公式（A1），抖动的相对功率可写为

$$P_{\mathrm{jitter}}/\mathrm{FS}^2 = \frac{\langle\{n_{\mathrm{jitter}}(t)\}^2\rangle}{A_{\mathrm{R}}^2} = \frac{(\omega A_0/A_R)^2\mathbb{V}\{j(t[i])\}}{2}. \tag{A9}$$

因此，P_{jitter} 计算为 $P_{\mathrm{jitter}} \approx 10^{-114/10}\ \mathrm{FS}^2$。该值接近图 10（a）中矩形频谱的面积。

采用（ii）进行 FDA 得到的结果如图 10（b）所示。这与图 10（a）相似，可以采用类似的方式进行解释。在 $f \approx 12$ 千赫处观察到一个峰值，这对应于纯正弦波的功率谱密度。白噪声底出现在 -195 满刻度分贝 Hz^{-1} 的电平处，这对应于量化噪声。在 6 千赫处观察到一个矩形频谱，范围为 $\leqslant f \leqslant 18$ 千赫。顶部电平为 -155 满刻度分贝 Hz^{-1}。该矩形频谱对应于人工幅度调制的功率谱密度；矩形频谱的面积计算为 $10^{-114/10}\ \mathrm{FS}^2$。根据公式（A1），幅度调制的相对功率可写为

$$P_{\mathrm{AM}}/\mathrm{FS}^2 = \frac{\langle\{n_{\mathrm{AM}}(t)\}^2\rangle}{A_{\mathrm{R}}^2} = \frac{\mathbb{V}\{A_{\mathrm{M}}(t[i])\}}{2A_{\mathrm{R}}^2}. \tag{A10}$$

因此，P_{AM} 计算为 $P_{\mathrm{AM}} \approx 10^{-114/10}\ \mathrm{FS}^2$。该值接近图 10（b）中所示矩形频谱的面积。

可以看出，（i）和（ii）的功率谱密度无法区分。这是由公式（A5）和（A6）的条件造成的，也是预期结果。我们选择这些条件，是因为这些结果清楚地显示了 FDA 的缺点；抖动和幅度调制无法与 FDA 区分。正如我们在引言部分所述，TDA，例如希尔伯特变换分析[5]（HTA）和 ZCA，使我们能够消除幅度调制。

FDA 在（iii）条件下的结果如图 10（c）所示。与图 10（a）的区别在于，矩形频谱的顶端电平变为 $-150\ \mathrm{dBFSHz}^{-1}$。该角谱对应于人工相位间隔噪声的功率谱密度；矩形频谱的面积计算为 $10^{-109/10}\ \mathrm{FS}^2$。根据公式（A1），相位间隔噪声的相对功率表示为

$$P_{\mathrm{PI}}/\mathrm{FS}^2 = \frac{\langle\{n_{\mathrm{PI}}(t)\}^2\rangle}{A_{\mathrm{R}}^2} = \frac{\mathbb{V}\{n_{\mathrm{PI}}(t[i])\}}{A_{\mathrm{R}}^2}. \tag{A11}$$

因此，P_{PI} 计算为 $P_{\mathrm{PI}} \approx 10^{-108/10}\ \mathrm{FS}^2$。该值接近图 10（c）中所示矩形频谱的面积。这表明 FDA 能够测量总噪声 $(P_{\mathrm{jitter}} + P_{\mathrm{AM}} + P_{\mathrm{PI}})/P_{\mathrm{C}}$。然而，由于 P_{AM} 未被消除，相位噪声，即 $(P_{\mathrm{jitter}} + P_{\mathrm{PI}})/P_{\mathrm{C}}$会被 FDA 高估。因此，绝对测量采样抖动需要采用时域分析。

3. 虚拟波形的 HTA

在本小节中，我们将 HTA[5] 视为一种现有的时域分析技术。虚拟波形(i)使用 HTA 进行分析。结果如图 11(a) 所示。水平轴表示时间 t。从虚拟波形 (i) 获得的点 $j_{\mathrm{H}}[i]$ 绘制在 $t = (i-1)f^{-1}{}_{\mathrm{R}}$ 处，并以点标记表示。人工抖动 $j(t)$ 以实线绘出以供比较。如图所示，HTA 成功使我们能够以 f_{R} 的采样率对 $j(t)$ 进行采样。

虚拟波形（ii）的 HTA 结果如图 11（b）所示。得到的抖动 $j_{\mathrm{H}}[i]$ 绘制在 $t = (i-1)f^{-1}{}_{\mathrm{R}}$ 处，并以圆点标记表示。作为比较，人工调幅 $A_{\mathrm{M}}(t)$ 用虚线表示。可以看出，从（ii）得到的 HTA 数据点几乎为零，且不受幅度调制影响。这是时域分析相较于 FDA 的优势：可以观察到抖动，同时消除幅度调制。

使用模拟波形 (iii) 得到的 HTA 结果如图 11 $-j_{\mathrm{H}}[i]$ $\omega t +$ (c)所示。θ_0 $t = i-1\ f^{-1}$ 上变频后的抖动正弦波 sin( ) 在 ( )R 处以点标记绘出。作为对比，人工相位间隔噪声 $n_{\mathrm{PI}}(t)$以点划线表示。根据公式 (4)，必须对 $j_{\mathrm{H}}[i]$ 进行上变频。由图可见，HTA 无法对 $n_{\mathrm{PI}}(t)$进行采样。使用 HTA 时，相位噪声 $(P_{\mathrm{jitter}}+P_{\mathrm{PI}})/P_{\mathrm{C}}$ 被低估。因此，对于采样抖动的绝对测量，必须采用优于 HTA 的改进的 TDA。

4. 模拟波形的 ZCA

在本小节中，我们将 ZCA 视为一种改进的 TDA。使用开发的 ZCA程序对模拟波形 (i) 进行分析。结果如图 12(a)所示。横轴表示时间 t。得到的 ZCF Δs_k 在 $t = s'_k$ 处以叉形标记绘出。作为对比，$j(t)$以实线表示。如 图 12(a)所示，ZCA 成功地使我们能够在理想时刻对 $j(t)$进行采样，该时刻为 $t = s'_k$。与 图 11(a)相比，显著不同之处在于，数据点出现在$t = s'_k$，即载波波形的过零时刻；如第二节 D 中所述，ZCA 的采样率为 $2f_{\mathrm{C}}$。该采样率足以根据采样定理从 $j(t)$重构 Δs_k。

使用虚拟波形 (ii) 得到的 ZCA 结果如图 12(b) 所示。得到的 ZCF Δs_k 绘制在 $t = s'_k$ 处，并以叉形标记表示。作为比较，$A_{\mathrm{M}}(t)$以虚线绘出。可以看出，ZCF 接近于零，且不受幅度调制的影响。根据公式（21），这是自然的；可以观察到抖动和 PI噪声，而幅度调制则被消除。从 (i) 和 (ii) 的 ZCA 结果来看，我们可以说，我们开发的 ZCA 程序在 10 ps 量级范围内具有足够的精度。

使用虚拟波形 (iii) 得到的 ZCA 结果如图 12(c) 所示。得到的 ZCF $(-1)^k\Delta s_k$ 绘制在 $t = s'_k$ 处，并以叉形标记表示。根据公式（21），需要使用因子 $(-1)^k$。作为比较，$n_{\mathrm{PI}}(t)$以点线绘出。可以看出，ZCA 成功地使我们能够对 $n_{\mathrm{PI}}(t)$进行采样。这表明，使用 ZCA 可以恰当地估算相位噪声 $(P_{\mathrm{jitter}} + P_{\mathrm{PI}})/P_{\mathrm{C}}$。

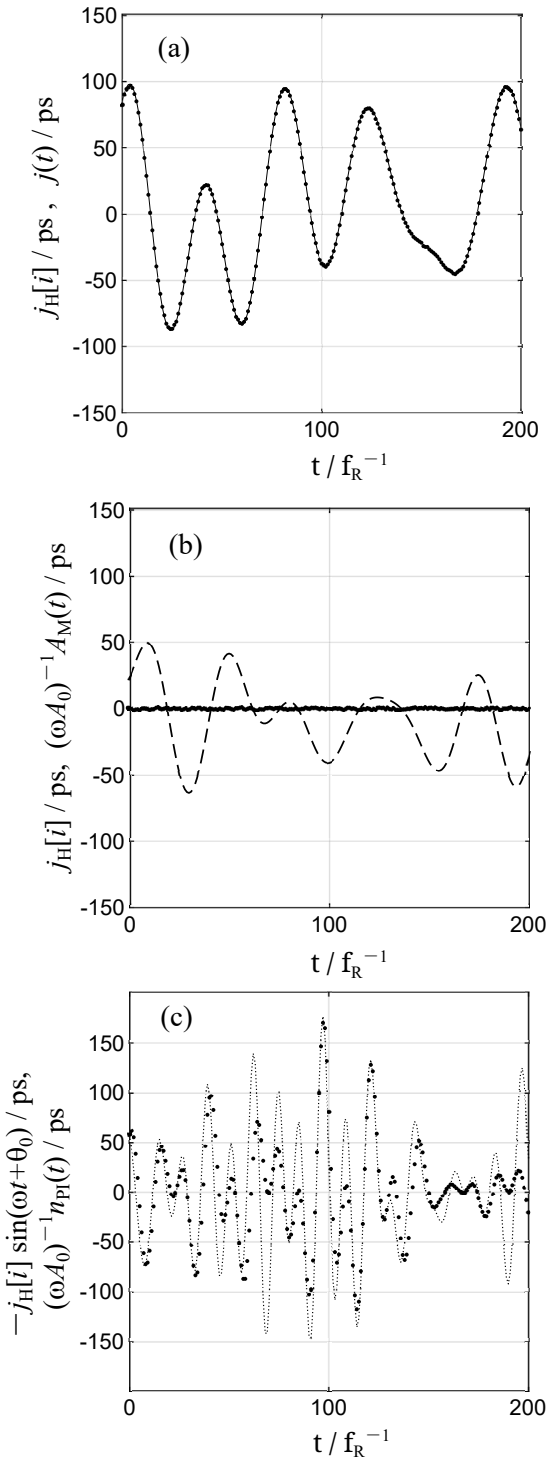


图11. HTA 的结果。(a) 圆点表示从 (i) 中获得的 $j_H[i]$。实线表示 $j(t)$。(b) 圆点表示从 (ii) 中获得的 $j_H[i]$ 。虚线表示 $A_M(t)$。(c) 圆点表示从 (iii) 中获得的 $j_H[i]$ 。点线表示 $n_{PI}(t)$。

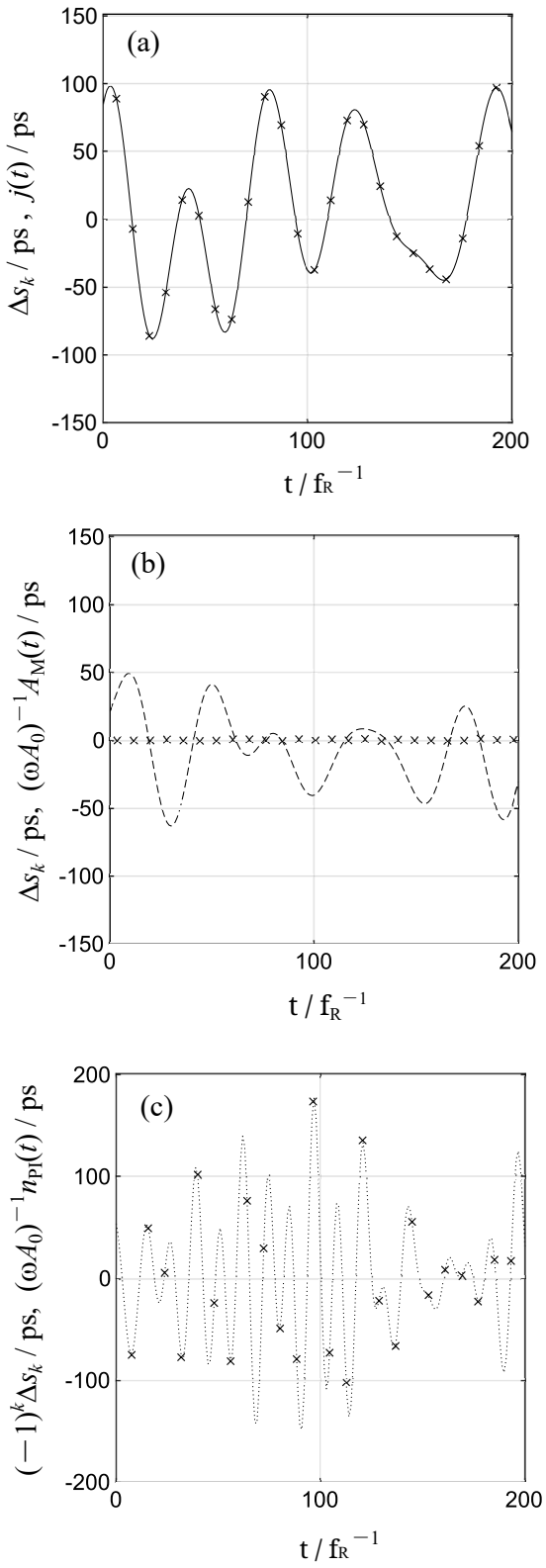


图12。（a）从（i）得到的 ZCF Δs_k （叉号）。实线表示 $j(t)$ 。（b）从（ii）获得的 ZCF Δs_k （叉号）。虚线表示 $A_M(t)$ 。（c）从（iii）取得的 ZCF Δs_k （叉号）。点线表示 $n_{PI}(t)$ 。

[1]J. Dunn，《数字音频设备中的抖动：规格与评估》，载于 *Proceedings of the 93rd AES Convention*，美国加利福尼亚州旧金山（1992年10月1日至4日），第3361页。[2]J. Dunn 和 I. Dennis，《数字音频系统中抖动相关问题的诊断与解决》，载于 *Proceedings of the96th AES Convention*，荷兰阿姆斯特丹（1994年2月26日至3月1日），第3868页。[3]J. Dunn，《数字音频系统中抖动相关问题的诊断与解决》，载于 *Proceedings of the AES 9th UK Conference: Managing the Bit Budget (MBB)*，英国伦敦（1994年5月16日至17日），第148–166页。[4]J. Dunn，《抖动理论》，Audio Precision 技术说明 TN-23，Audio Precision，俄勒冈州比弗顿（2000年）。[5]A. Nishimura 和 N. Koizumi，《使用解析信号测量模数转换器和数模转换器中的采样抖动》，《声学科学与技术》31(2)，172–180（2010年）。[6]W. R. MacLean，《无需基准标准的声音绝对测量》，《美国声学学会杂志》12，140–146（1940年）。[7]S. Barrera-Figueroa，《频率高达150千赫时测量麦克风的自由场互易校准》，《美国声学学会杂志》144(4)，2575–2583（2018年）。[8] E. Rubiola 和 F. Vernotte，《互谱实验法》（2010年） muarXiv:physics.ins-det/1003.0113。[9]K. Ashihara、S. Kiryu、N. Koizumi、A. Nishimura、J. Ohga、M. Sawaguchi 和 S. Yoshikawa，《离散信号的检测阈值——抖动导致的数字音频失真，《声学科学与技术》26(1)，50–54（2005年）。[10]V. R. Melchior，“高分辨率音频：历史与展望”，《音频工程学会杂志》67(5)，246–257（2019年）。[11]H. Nittono，“高分辨率音频的高频声音成分无法在听觉感觉记忆中被检测到”，《科学报告》10，21740（2020年）。[12]播放器产生频率为 $f_C = f_P/4$ 的正弦波；然而，录音机测得的频率并不完全是 $f_P/4$。[13]数值模拟结果见附录。[14]M. Frigo 和 S. G. Johnson，“FFTW（3.3.10 版） [计算机程序],”（2005年） muhttp://www.fftw.org （最后查看于2022年4月27日）。[15]G. Feldhaus 和 A. Roth，“采用互相关的 1 兆赫至 50 吉赫直接下变频相位噪声分析仪”，2016年欧洲频率与时间论坛（EFTF）1–4（2016年），doi：mu10.1109/EFTF.2016.7477759。[16]G. Feldhaus、G. Roesel、A. Roth 和 J. Wolle，“相位噪声的测量不确定度分析与溯源性”，应用说明第1EF95号，Rohde & Schwarz，德国慕尼黑（2016年）。[17]W. Kester，“揭开这个著名公式的神秘面纱：‘信噪比 $= 6.02N + 1.76$ 分贝’，以及为何你应该关注它”，Analog Devices MT-001 教程，Analog Devices（2009年）。