arXiv:2301.12354v1 [cs.SD] 29 Jan 2023

由音乐音频承载的艺术曲线隐写术

Christopher J Tralie，Ursinus学院，ctralie@alumni.princeton.edu

2023年1月31日

摘要

在这项工作中，我们创作了勾勒图像和三维形状的艺术性闭合环曲线，并将它们隐藏在音乐音频中，作为一种隐写术。我们使用旅行商艺术创作勾勒图像轮廓的艺术性平面环路，并使用三角网格上的哈密顿回路创作填充三维曲面的艺术性空间环路。我们的嵌入方案旨在有损压缩后忠实保留这些环路的几何形状，同时让音频听众无法察觉它们的存在。为此，我们将曲线的每个维度隐藏在不同的频率中，并扰动该频率的幅度滑动窗口和，使其尽可能匹配相应维度的目标曲线，同时将尺度信息隐藏在该频率的相位中。在此过程中，我们利用曲线的几何特性，帮助更有效地隐藏和恢复曲线。我们的方案简单高效，使用非负最小二乘框架即可完成编码，解码则十分简单。我们在大型图像和音频数据集上对该技术进行了定量验证，并展示了众包听音测试的结果，验证了隐藏的信息确实不会造成干扰。

1 引言/背景

隐写术是将一个数据流“明目张胆地”隐藏在另一个称为“载体”的数据流中的过程。在本文中，我们关注使用音频作为载体来存储图像，而这一方向在学术文献中相对鲜有研究（第 1.1 节）。与任何隐写技术一样，人们可以将其作为一种通用的隐蔽通信手段；艺术家也可以用它来隐藏图像，并为自己的曲目添加水印。一个例子是，Aphex Twin 曾将一张人脸图像嵌入其歌曲“Windowlicker”的声谱图中 [30]，，不过这种技术会产生非常响亮、不和谐且刺耳的声音。相比之下，我们希望实现一种无法被察觉的图像嵌入方式。然而，与 Aphex Twin 一样，我们也希望图像能够经受有损音频压缩，从而让人们可以通过社交媒体传递图像，这使问题变得更加具有挑战性。

为了让我们的音频隐写术系统尽可能稳健且易于使用，我们制定了以下设计目标：

1. 隐藏数据在听觉上应当不可察觉 2. 隐藏数据在有损压缩后应当得到忠实保留 3. 隐藏数据应当能够抵抗帧错位，或者应当能够在没有任何先验信息的情况下恢复帧对齐 4. 检索隐藏数据时不应需要元数据；仅凭（有损压缩后的）音频即可完成 5. 应当能够从部分音频片段中部分恢复数据；也就是说，无需等待整个数据流即可恢复信号

目标 1 和目标 2 彼此存在冲突，同时满足这两个目标是本工作的最大挑战。我们在第 2.1 节中制定了一个目标函数，以权衡这两个目标。

我们不直接将图像存储在音频中，而是将问题限定为隐藏描绘图像的艺术曲线（第 1.2 节）或三维表面形状（第 1.3 节）。我们将曲线的每个维度隐藏在不同的频率中，并使用频率幅值的滑动窗口求和在时间上对其进行平滑处理，从而使其更加抗压缩，正如第 3 节所示。此外，由于我们知道曲线在时间上的相邻采样点之间只会发生小幅移动，因此可以利用这一事实从帧错位中恢复，以满足目标 3，具体解释见第 2.4 节。最终，重建曲线所需的全部隐藏信息都存储在音频中各频率的幅度和相位中，因此无需元数据（目标 4）。最后，恢复单个曲线采样点所需的信息在时间上是局部化的，因此可以从部分音频中对曲线进行部分解码（目标 5）。

1.1 音频隐写术的相关工作

在介绍构造艺术曲线的技术（第 1.2 节、第 1.3 节），并最终将其隐藏在音频中（第 2 节）之前，我们先简要回顾音频隐写术领域的相关工作，以便将我们的工作置于背景之中（更多信息请参阅这些综述文章 [12, 14]）。

与大多数其他隐写载体一样，最简单的音频隐写术技术依赖于改变未压缩编码的最低有效位（例如 [11]）。其他早期研究则利用了人类听觉系统的“频率掩蔽”特性。二进制数据通过控制以这种方式被掩蔽的两个频率的相对振幅来存储 [17, 18]。

除了振幅扰动之外，也可以保持振幅不变而改变相位 [36, 37]； [29] 技术在这方面尤为

成功：它通过操纵全通滤波器的极点来编码二进制数据。还可以通过添加和操纵信号的回声来隐藏二进制数据 [21], ；这种方法比其他技术更能抵抗有损压缩，因为它类似于自然环境的脉冲响应；不过，比特率相当低（每秒几十比特）。对于压缩音频，另一种方法是利用 MP3 文件本身的属性来隐藏信息 [2,32]。还有一些技术将无线传输方案应用于音频领域，例如开关键控（OOK）[28]（低比特率为 8 bps）和正交频分复用（OFDM）[15]（较高比特率接近 1 kbps）。

音频隐写术文献中隐藏的大多数消息都是二进制格式，但近期有少数研究侧重于隐藏图像像素，利用深度学习协同训练“隐藏网络”和“揭示网络”，分别将图像编码到音频流中，再解码并将其叠加到载体中，同时尽可能提高解码后图像和编码后音频的质量 [10,13,16,34]。我们不仅也尝试隐藏图像，而且持有相似的观点：与其试图精确提取二进制序列，我们可以容忍重建结果中逐渐累积的噪声。基于深度学习的方法面临的挑战包括需要使用大量样例进行训练、难以纳入能够模拟有损音频压缩影响的损失项，以及难以训练出能够抵抗帧错位的模型（目标 3）。在我们的工作中，我们通过采用一种基于频率隐藏的简单模型，避开了这些挑战。

1.2 旅行商艺术

为了设计能够代替图像的二维曲线，我们采用旅行商问题（TSP）艺术 [4,6,23],，这是一种自动化艺术线描流程，通过计算一条简单闭环 来近似原始图像。这里的“简单”并不意味着缺乏复杂性，而是指曲线不与自身相交（例如，参见 Jordan 曲线定理 [5]）。构造此类曲线时，我们首先按照“点画图案”在平面上布置一组点，以近似表示图像的亮度（例如，较暗区域的点更密集），然后通过近似的旅行商问题（TSP）巡回路径将这些点连接成一个回路。图 1 展示了一个示例。我们采用 TSP 艺术技术 [23],，并作了若干修改，详见下文。

TSP 艺术的第一步是生成点画图案 X ，以尽可能逼近灰度图像 G。遵循 [23], 的作者，我们采用 Secord 的 Voronoi 点画技术 [33], ：该技术根据某些密度权重 W 对点进行初始采样，然后反复将点移向其 Voronoi 区域的加权质心（这是 Lloyd 算法的一种实例），使点分布得更加均匀。Secord [33] 将像素 W_{ij} 处的权重设为与其亮度成反比，并在亮度高于某个亮度阈值 b 时将权重设为 0。这会使更多的点集中在较暗的区域；不过，该算法可能无法在较亮区域之间的重要边缘处采样到任何点（ [27] 的作者也观察到了这一现象）。为缓解这一问题，

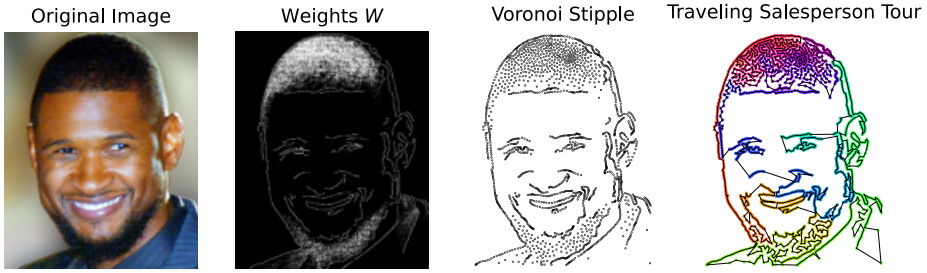
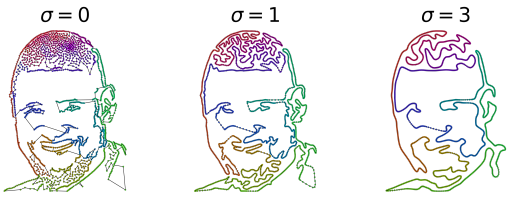


图 1：我们改进后的 TSP 艺术创作流程。巡回路径上的颜色表示闭合回路的相位。

图 2：由于 mp3 压缩会给我们的嵌入曲线引入噪声，我们在 $\sigma = 1$ 处使用一次曲率缩短流迭代对其进行预平滑处理，从而平滑曲线且不会引入任何交叉。



我们对原始图像应用 Canny 边缘检测器 [8]，并将边缘上所有像素的权重设为 1，使最终权重既能增加较暗区域中的采样点，也能增加各种边缘上的采样点。这一改进对线描图尤其有帮助。

确定点描后，TSP 艺术的下一步是用一条闭合回路“连接各点”，并恰好访问每个点描点一次，这条回路称为“巡回路径”。一种常见的目标函数用于衡量巡回路径是否避免了过多的“跳跃”，即总行进距离，也就是所有边长之和；达到最优值的巡回路径称为旅行商问题（*TSP*）巡回路径。由于旅行商问题是 NP 困难的，[23] 的作者使用 Concorde TSP 求解器 [1] 求近似解。我们采用一种更简单的方法：先通过对点描点的最小生成树进行深度优先遍历，构造 TSP 的一个 2-近似解；这本身已经是最优巡回路径的 2-近似解。接着，我们通过一系列 2-opt 松弛 [22]迭代改进这条巡回路径；也就是说，若对于某个 $i > 1$, $i < j < N$ 这 4 个点 X_i，X_{i+1}，X_j和 X_{j+1} 之间的距离满足

$$d(X_i, X_j) + d(X_{i+1}, X_{j+1}) < d(X_i, X_{i+1}) + d(X_j, X_{j+1}) \tag{1}$$

这样便可以执行交换，得到一条距离更短的新巡回路径。这相当于将巡回路径中索引从 $i+1$ 到 j（含首尾）的部分反转。只要这种交换仍然可行，我们就重复执行此步骤。尽管这并不能保证得到最优的 TSP 巡回路径，但确实能够生成简洁且具有美感的巡回路径；也就是说，所有交叉都被消除了。

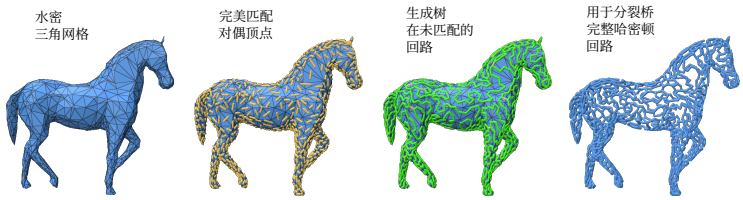


图 3：Gopi and Epstein 的算法在普林斯顿网格分割基准中的一个马模型上的示例 [19]， [9]。

1.2.1 曲率缩短流

由于 MP3 压缩会在我们的嵌入曲线中引入噪声，因此我们会在嵌入之前对其进行平滑处理，以改善视觉质量。为此，我们采用 Mokhtarian 和 Mackworth [31] 描述的曲率缩短流的数值版本，该方法适用于分段线性曲线（例如我们的 TSP 巡回路径）。具体而言，该技术通过将每条曲线的坐标与平滑后的高斯函数及其导数进行卷积来实现数值处理。

为了通过一步曲率缩短流近似地平滑曲线，Mokhtarian 和 Mackworth [31] 证明，只需先按弧长重新参数化，然后将曲线与高斯核[1]卷积一次即可。根据 Gage-Hamilton-Grayson 定理，经历曲率缩短流的简单曲线仍保持简单，并最终变为凸曲线；在反复应用该流的过程中，曲线会不断收缩，直至缩成一个点。图 2 展示了对不同 σ 取值应用一次曲率缩短流的示例。

1.3 水密三角网格上的哈密顿回路

除了平面图像点描图案上的二维旅行商问题巡回路线外，我们还采用 Gopi and Epstein 的方法，创建填充三维形状表面的艺术三维空间环路 [19]。他们指出，水密三角形网格（即没有边界的网格）的对偶图存在完美匹配。换言之，可以将三角形划分为若干对相邻三角形（见图 3 第二列）。实际应用中，我们使用 Blossom-5 算法 [24] 来求解对偶图的完美匹配。随后，在未配对的三角形之间添加边，会形成一组不连通回路（见图 3 第三列）；再通过一组连接配对三角形的生成边，即可将这些回路连接起来。最后，将生成边拆分为连接不连通回路的桥接边，把它们合并成一条覆盖整个三角网格表面的大回路。

[1]使用高斯函数进行卷积的妙处在于，γ 甚至不需要可微，因此这种方法适用于我们的分段线性 TSP 巡回路径。

2 音频中的曲线嵌入

现在，我们将介绍一种用于在音频中隐藏艺术空间曲线的新算法。在详细说明之前，我们先定义定量测量指标，用于衡量与原始音频的匹配程度（目标 1）以及隐藏曲线的几何质量（目标 2）。设原始载体音频为 x ，隐写音频为 y，两者各包含 N 个采样点。于是，我们将隐写信噪比定义为以分贝（dB）为单位的：

$$\mathrm{snr}(y|x) = 10\log_{10}\left(\sum_{j=1}^{N} x_j^2\right) - 10\log_{10}\left(\sum_{j=1}^{N}(y_j - x_j)^2\right) \tag{2}$$

对于质量的几何测量，设 X_i 是由 M 个 $\mathbb{R}^d$ 中的 Y_i 目标曲线点组成的序列， M 是由 $\mathbb{R}^d$中的个重构点组成的序列。然后，我们将平均几何失真简单定义为这些点之间欧氏距离的平均值：

$$\mathrm{distortion}(Y|X) = \frac{1}{M}\sum_{j=1}^{M} ||X_j - Y_j||_2 \tag{3}$$

2.1 最小二乘问题的表述

现在，我们对一个目标函数进行表述，使其在公式 2 和公式 3 之间进行权衡。设 $T = (T_1, T_2, \ldots\ldots, T_N)$ 是要隐藏的目标曲线点序列，其中每个点为 $T_i \in \mathbb{R}^d$；设 $T_{i,\ m}$ 表示 T_i 的 m^{th} 坐标；再设 x 是一组将作为载体的音频样本。目标是扰动 x 的样本，使得 x 的某个函数尽可能接近 T 。我们选择的函数基于幅度短时傅里叶变换（STFT）的“时间正则化”版本，称为滑动窗口和 STFT（SWS-STFT）。首先，我们根据选定的窗口长度 w 计算一个不重叠的 STFTS ，其总帧数为 N ：

$$S_{k,j} = \sum_{n=0}^{w} x_{jw+n}\left(e^{-i2\pi kn/w}\right) = M_{k,j}\left(e^{iP_{k,j}}\right) \tag{4}$$

我们将 S 其分解为幅度和相位分量 M 和 P。接下来，我们选择一组 d 频率 k_i， $i \in \{1, 2, \ldots, d\}$其中每个频率都将隐藏 T的一个不同维度。给定第二个窗口长度ℓ后，我们定义如下滑动窗口和函数，并将其应用于我们希望扰动的 S 幅度的每一行 k_i ，以获得 SWS-STFT

$$\mathrm{SWS}^{\ell}(M)_{k_i,j} = \sum_{n=0}^{\ell-1} M_{k_i,j+n} \tag{5}$$

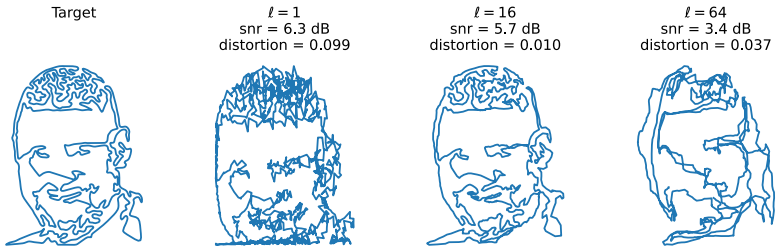


图 4：在 $\lambda = 0.1$固定时改变 ℓ，使用最低的两个非直流频率进行承载。公式 6 中，SW-STFT 的 ℓ 越大，曲线越平滑，也越有可能经受压缩；但 ℓ 过大可能导致过度平滑。

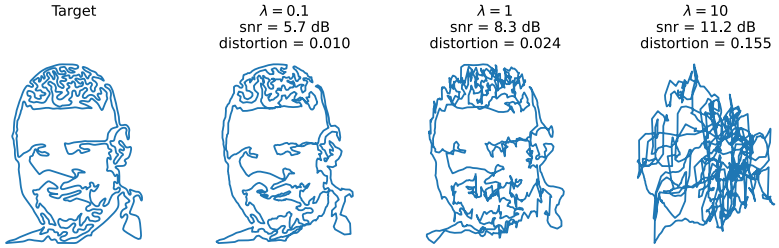


图 5：在 $\ell = 16$固定时改变 λ。公式 6 中，较小的 λ 可提高几何保真度（目标 2），但会降低音频质量（目标 1）；音频质量以信噪比衡量。

其作用是平滑幅度频谱图中嘈杂的行 k_i，使 SW-STFT 中的各行与目标曲线中更平滑的过渡相匹配。SW-STFT 的每一行包含 $N - \ell + 1$ 个样本。暂时假设 T 恰好包含这么多样本；对于 length(T) $> N - \ell + 1$ 的情况，我们将在第 2.2 节讨论。随后，我们寻求幅度 M 的一个扰动版本 $\hat{}$，使得每个坐标 i 都隐藏在 M 的单个频率 $\hat{}$ 索引 k_i 中。为此，我们一次针对一个坐标维度 $i = 1, \ldots d$ 最小化以下目标函数：

$$f(\hat{M}_{k_i}) = \sum_{j=1}^{N-\ell+1} \left(\left(\sum_{n=0}^{\ell-1} \hat{M}_{k_i,j+n} \right) - T_{i,j} \right)^2 + \lambda \sum_{j=1}^{N} \left(M_{k_i,j} - \hat{M}_{k_i,j} \right)^2 \qquad (6)$$

$\hat{}$ 受 $M_{k_i,j} \geq 0$ 约束。换言之，根据 λ，我们希望扰动信号的幅度 SW-STFT 尽可能匹配目标坐标（最小化公式 3），同时尽可能保留原始音频（最小化$\hat{}$公式 9）。更大的 λ 意味着信号 M 会更贴合原始音频，但代价是曲线中的噪声会更多。图 5 展示了一个示例。

ˆ ˆ 求解 $M_{k_i,j}$ 后，我们将 M_{k_i} 的所有行替换为 M_{k_i}，并使用原始相位 P 对 Me^{iP} 执行逆短时傅里叶变换（第 2.3 节将解释何时也需要修改相位）。然后，我们以压缩格式保存生成的音频，并通过加载该音频、提取幅度 SW-STFT 中相应的分量来恢复隐藏信号。

2.1.1 计算复杂度

将公式 6 最小化可以表述为一个稀疏非负线性最小二乘问题，并且存在大量算法（例如 [7]），可通过反复对线性系统及其伴随算子作用于解的迭代估计值进行求值，高效地求解此类系统。此外，尽管公式 6 表明，在每次迭代中计算目标函数的时间复杂度为 $O(N\ell)$，但我们仅使用 $O(N)$ 次运算来实现线性算子及其伴随算子，且与 ℓ 无关，这是借鉴了计算机视觉中的一维“积分图”技巧 [25]。例如，令 $C_{k_i,j} = \sum_{i=0}^{j} \hat{}M_{k,i}$，以及 $C_{k_i,j<1} = 0$。那么公式 6 可以改写为

$$f_i(\hat{M}) = \sum_{j=1}^{N-\ell+1} \left((C_{k_i,j+\ell-1} - C_{k_i,j-1}) - T_{i,j}\right)^2 + \lambda \sum_{j=1}^{N} \left(M_{k_i,j} - \hat{M}_{k_i,j}\right)^2 \quad (7)$$

还值得注意的是，公式 6 中更高的 λ 会导致隐含线性系统中矩阵的条件数 [2] 更低，从而使公式 6 的最小化速度更快。但如图 5 所示，使用较小的 λ 值可能仍然值得。在实践中，我们发现这表现为两种情况之间的差异：一种是对 30 秒音频进行编码，$\lambda = 0.1$ 需要几秒钟；另一种是在中央处理器上进行编码，$\lambda = 10$ 只需不到一秒。

2.1.2 非重叠窗的选择

尽管其他一些研究使用带有重叠窗口的变换 [16, 37]，但独立改变不同的短时傅里叶变换频点会由于重叠窗口之间存在差异，导致得到的短时傅里叶变换不对应任何真实信号。像 Griffin-Lim 这样的算法 [20] 可以恢复出其声谱图与扰动后的短时傅里叶变换之间的局部最小距离的信号，但我们发现，这一步会在目标音频和载体音频中引入不可接受程度的失真。即使对于实值变换（例如离散余弦变换，该变换被 [16]使用）也会出现这一问题。与 [36], 类似，我们通过使用非重叠窗来规避这一问题。

除窗效应外，非重叠窗口的主要缺点是会导致短时傅里叶变换帧数量减少。例如，对于一个时长为 30 秒、采样率为 44100 Hz、窗口长度为 1024 的音频片段，我们最多只能得到约 1292 帧，这相当于使用帧移为 512 的重叠短时傅里叶变换时所得帧数的一半。不过，简单进行一个思想实验便可看出，这仍然是一个

[2]矩阵的条件数定义为最大奇异值与最小奇异值之比，较低的条件数在数值上更为理想。

合理的数据速率。假设我们使用滑动窗口长度 $\ell = 16$，总共得到 1277 个载波样本。再假设，我们对二维平面曲线的嵌入精度大致相当于更传统的二进制编码方案中每个坐标使用 8 比特量化的精度。那么，30 秒内传输的比特总数为 $1277 \times 8 \times 2$，即约为 681 比特/秒。对于三维曲线，这一数值将升至约 1022 比特/秒。这些数值与近期其他旨在抵抗噪声的技术相当（例如，[15] 的 OFDM 技术可达到 900 比特/秒）。此外，我们所对应的“比特错误”是加性坐标噪声，并且隐藏信号会随着噪声增加而连续退化，而不是在比特错误过高时突然进入失效状态。

2.2 平移、缩放和重新参数化目标，以获得更好的拟合

本节将说明如何调整目标曲线，使其更好地匹配给定的 SW-STFT，从而无需大幅扰动 STFT 幅度即可使 SW-STFT 匹配目标曲线，进而在几何形状质量相同的情况下，使隐藏曲线的嵌入更难被察觉。

2.2.1 垂直平移／缩放

为了使隐藏信号难以察觉，一个关键步骤是平移并重新缩放隐藏曲线的目标坐标，使其匹配 SW-STFT 分量的动态范围。我们首先为每个分量选取一个尺度比 a_i，该比值等于标准差 $\mathrm{stdev}_j\left(\mathrm{SWS}^{\ell}(M)_{k_i,\,j}\right)/\mathrm{stdev}_j(T_{i,\,j})$。然后，令 N 为两个信号的长度，计算垂直平移量 b_i，其定义为

$$b_i = \frac{1}{N}\sum_{j=1}^{N} \mathrm{SWS}^{\ell}(M)_{k_i,j} - a_i T_{i,j} \tag{8}$$

^ ^ 缩放后的目标坐标 T_i 定义为 $T_{i,\,j} = a_i T_{i,\,j} + b_i$。遗憾的是，我们会丢失分量之间的相对尺度信息，但我们将在第 2.3 节说明如何隐藏并恢复这些信息

2.2.2 维特比目标重新参数化

除了在垂直方向上平移和缩放目标坐标外，我们可能还需要在时间维度上对其重新参数化，因为一般而言，SWS-STFT 序列的样本数 ^ 与平移后的目标 T 并不相同。此外，目标曲线是循环的，因此起点可以任意选择；沿曲线向左或向右遍历，或以恒定速度遍历，也都无关紧要。这使我们在扰动信号之前，就有很大的自由度来选择目标的重新参数化方式，以便使其与信号最佳匹配。设 N_M 为 SW-STFT 中的帧数，N_T 为目标的样本数，并假设 $N_T > N_M$（我们总可以对目标曲线重新采样，使这一条件成立）。

算法 1维特比目标重参数化

过程 ViterbiTargetReparam($\hat{T}$, SWS$^{\ell}(M)$, K)
  $N_T \leftarrow$len($\hat{T}$) ▷ 目标点数量
  $N_M \leftarrow$ len(SWS$^{\ell}(M)$) ▷ SW-STFT 帧数
  $S[i > 1, j] \leftarrow 0$ ▷ $N_T \times N_M$ 累积代价矩阵
  $S[1, j] \leftarrow \sum_{i=1}^{d}(\hat{T}_{i,1} - \text{SWS}^{\ell}(M)_{k_i,j})^2$
  $B[i, j] \leftarrow 0$ ▷ $N_T \times N_M$ 指向最佳前序状态的回溯指针
  对于 $j = 2:\ N_M$ **执行**
    对于 $t = 1:\ N_T$ **执行**
      $S[t, j] =$ 最小值$_{k=(t-K) \text{ 模 } N_T}^{k=(t-1) \text{ 模 } N_T}\ S[k, j-1]$ ▷ 查找最优前序状态
      $S[t, j] \leftarrow S[t, j] + \sum_{i=1}^{d}(\hat{T}_{i,\ t} - \text{SWS}^{\ell}(M)_{k_i,\ j})^2$ ▷ 加到匹配代价上
      $B[t, j] =$ 使其最小的参数$_{k=(t-K) \text{ 模 } N_T}^{k=(t-1) \text{ 模 } N_T}\ S[k, j-1]$ ▷ 保存对最优状态的引用前一状态
    结束循环
  结束循环
  回溯 B 以获得最优序列 Θ
返回 Θ
结束过程

设 $\Theta = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_{N_M}\}$ 为目标中的新索引，并设 $K > 0$ 为正整数，使得 $1 \leq ((\theta_i - \theta_{i-1}) \bmod N_T) \leq K$；也就是说， K 是目标重新参数化在相邻时间步之间最多可跳过的样本数。对于一个 d 维的平移目标 T，我们要寻找使下列目标函数 $\hat{}$ 最小的 Θ

$$g(\Theta = \{\theta_1, \theta_2, \ldots, \theta_{N_M}\}) = \sum_{j=1}^{N_M} \sum_{i=1}^{d} \left(\text{SWS}^{\ell}(M)_{k_i,j} - \hat{T}_{i,j} \right)^2 \tag{9}$$

对于固定的 K，我们使用维特比算法在 MNK 时间内获得使 O 中公式 9 最小化的 Θ。算法 1 给出了更多详细信息。

在实践中，我们从 $K = 1$ 开始重新运行算法，并反复递增 K，直到最优的 Θ 至少在目标曲线上完整遍历一圈。由于顺时针遍历或逆时针遍历目标曲线是任意的，因此我们随后再次针对序列的逆序版本运行这一过程，并保留使公式 9 最小的结果。作为经验法则，我们发现，当目标曲线的样本数约为 SWS-STFT帧数的 1.5-2 倍时，能够为维特比算法提供足够的调整空间。在第 3 节的实验中，我们将为 1200-1300 个 SWS-STFT帧生成包含 2000 个样本的 TSP 和哈密顿序列。图 6 展示了在这些条件下运行该算法的一个示例，对象是图 4 第 3 列中的 Usher 示例。

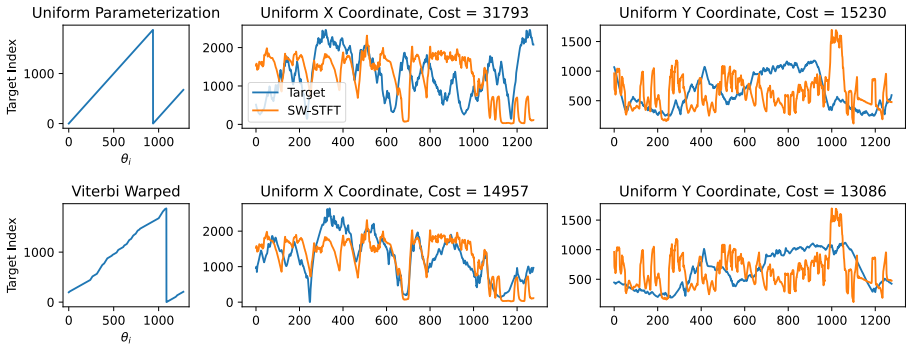


图6：循环移动目标环并以非均匀速度遍历它，可以描绘出相同的形状，同时比从目标上的任意位置开始进行均匀参数化更好地匹配给定的 SW-STFT。

2.3 在相位中存储分量尺度

由于我们有意重新缩放目标曲线的维度，使其匹配每个 SW-STFT 分量的动态范围，因此我们会失去维度之间的长宽比。但由于我们只扰动了频率索引 k_i 的幅度分量，因此仍然可以通过扰动相位来存储额外信息。为此，我们采用 [36] 的作者提出的技术，将每个维度的相对尺度存储在相位中。我们将相同的尺度存储在每个 STFT 帧的相位中，并在解码时将相位的中位数作为该尺度。

2.4 恢复帧对齐

到目前为止所描述的方法适用于与每个窗口对齐的音频，但如果待解码的音频存在错位，则需要开展额外工作来实现目标3。为此，我们利用这样一个事实：隐藏曲线在相邻采样点之间仅发生轻微移动；TSP巡回路径被定义为长度最小，而网格上的哈密顿回路中的相邻采样点，在原始网格上只会在相邻三角形之间移动。如果嵌入是帧对齐的，则曲线长度应达到最小。相反，如果嵌入未进行帧对齐，曲线会变得嘈杂，并且更有可能在采样点之间快速跳动。因此，我们可以在从0到STFT窗口长度的所有可能移位中，选择使长度最小的对齐方式。图7展示了一个示例。我们将在第3节中对此进行实证评估。

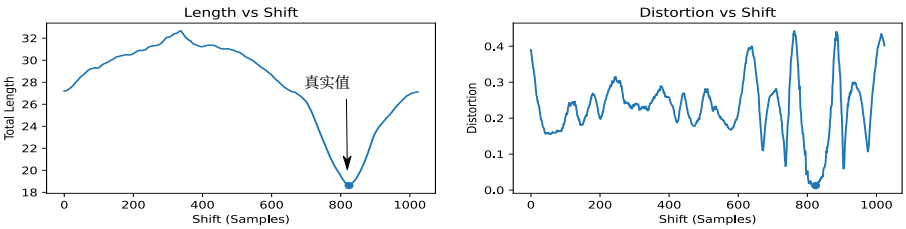


图7：使曲线长度最小的偏移，最有可能就是将音频重新对齐到帧窗口所需的偏移。在此示例中，嵌入使用最低的两个非直流频率，且 $\lambda = 0.1$，$\ell = 16$，全局最小值与真实值完全一致。

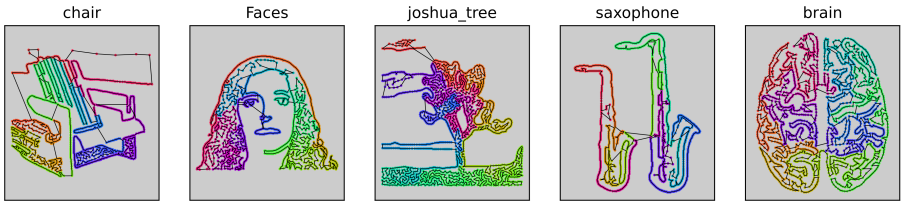


图8：Caltech-101 数据集 [26]上的 TSP 艺术示例。

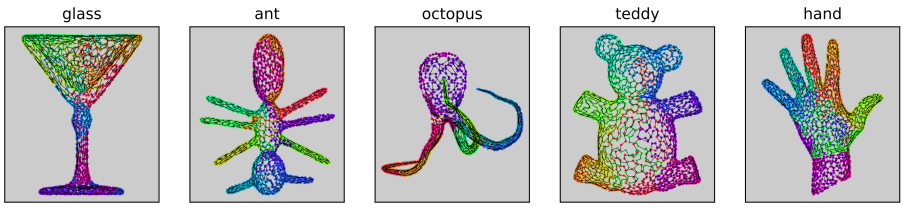


图9：普林斯顿网格分割基准 [9]中的三角形网格上的哈密顿回路示例。

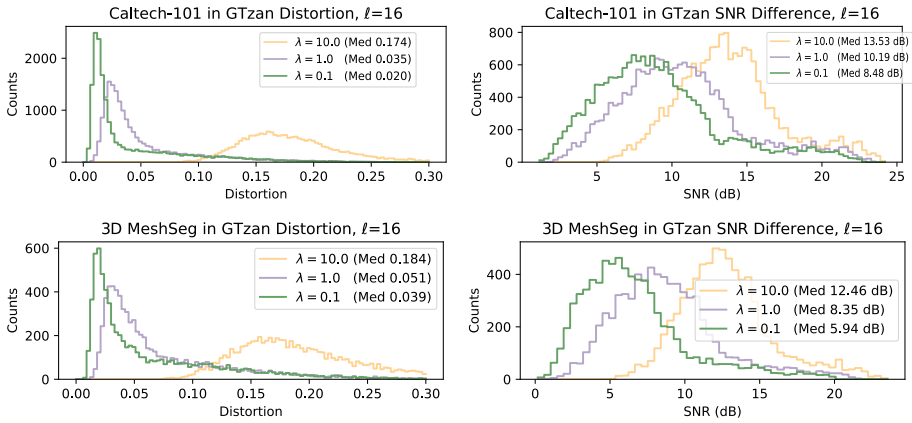


图10：将曲线嵌入 Tzanetakis 音乐流派数据集中的音频片段后得到的结果，$\ell = 16$固定。正如公式 6 所预期的，随着 λ 增加，失真和信噪比都会上升。

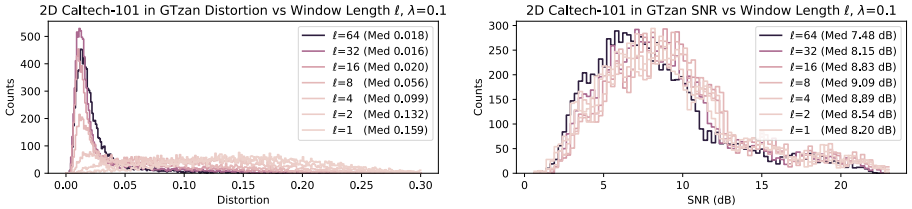


图11：将曲线嵌入 Tzanetakis 音乐流派数据集中的音频片段，并在 ℓ 变化窗口长度， λ固定。适中的窗口长度在信噪比和失真方面都是最佳选择。我们建议使用 $\ell = 16$。（三维情况类似；请参阅补充材料）

3 实验

现在，我们对系统的性能进行定量评估。为生成大量曲线，我们在 Caltech-101 数据集约 10,000 张图像上生成 TSP 艺术 [26]（例如图8），并在普林斯顿网格分割数据库中的 380 个无泄漏三角网格上生成哈密顿回路 [9]（例如图9）。随后，我们使用 Tzanetakis 音乐流派数据集 [35] 中的 1000 段 30 秒音频片段作为载体音频。对于 Caltech-101 数据库，我们将图像分成若干组，每组 10 张，并分别编码到一个载体音频中，同时改变 λ 和 ℓ。类似地，对于网格分割数据集，我们将每条哈密顿路径隐藏在 Tzanetakis 数据集的三个不同音频片段中。所有情况下，我们都使用窗口长度为 1024、采样率为 44100 Hz 的 STFT，并以 64 kbps 的码率通过有损 MP3 压缩对音频进行编码。

总体而言，我们发现三维嵌入的失真略高、信噪比较低

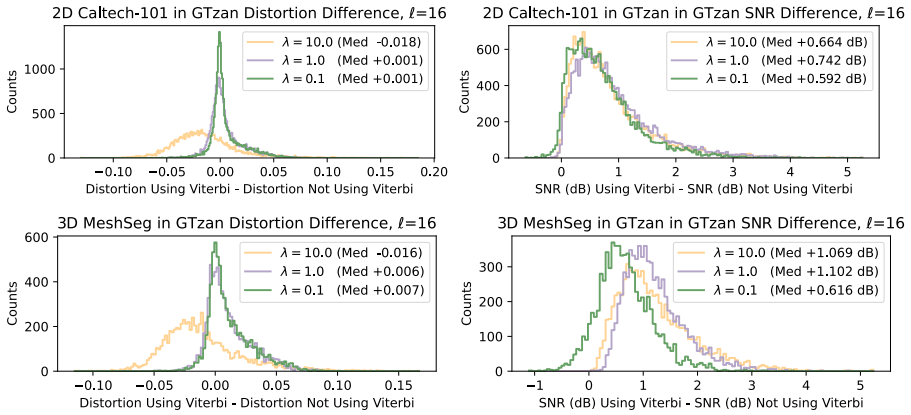


图12：使用维特比对齐对目标进行预扭曲（第2.2节）总体上改善了所得的失真和信噪比。

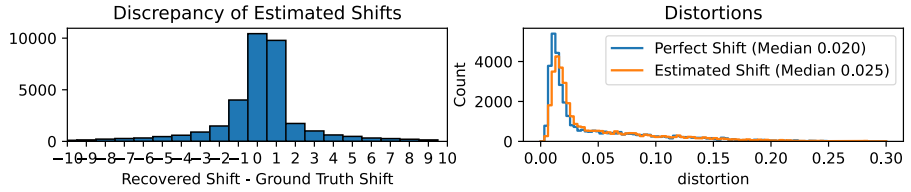


图13：通过最小化长度来估计帧对齐（第2.4节），其效果几乎与完全掌握对齐信息一样好。在 Caltech-101 的所有图像上使用$\lambda = 0.1$，$\ell = 10$ 时，对于 STFT 窗口长度为 1024 的情况，93%的偏移量与真实值相差不超过10，这几乎不会影响曲线的几何形状。

相比二维嵌入，这一结果是合理的，因为在三维中还需要隐藏一个额外坐标。正如图10所示，随着 λ 增加，信噪比和失真都会提高，这符合预期。此外，如图11所示，增加窗口长度会对几何失真产生积极影响，而适中的窗口长度则能带来最佳信噪比。在图12中，我们还看到第2.2节中的维特比对齐对信噪比有积极影响。尽管 $\approx$+1dB看起来可能并不显著，但它会对音频产生巨大影响，尤其是在安静区域。最后，我们在 Caltech-101 数据集上进行了实验：为每个嵌入随机选择4个帧偏移，并在图13中看到，第2.4节的算法能够很好地恢复对齐关系。

3.0.1 主观听音实验

尽管上述实验结果令人鼓舞，但信噪比可能会产生误导；由于心理声学现象，人耳更容易听到的频率实际上可能具有更高的信噪比。

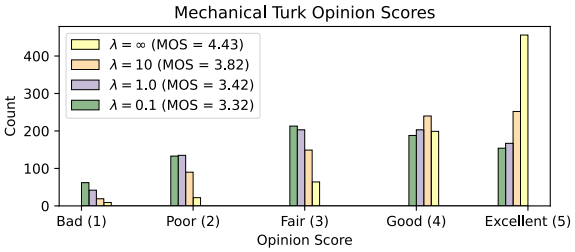


图14：Amazon Mechanical Turk 上的听音实验结果。较低的 λ 会导致较低的平均意见分，这正如预期；不过，程度尚不足以令人无法忍受。

为了解决这一问题，我们在 Amazon Mechanical Turk 上开展了一项众包听音实验，将 Caltech-101 数据集中的随机图像嵌入到每段 Tzanetakis 音频片段中。我们将音频片段分为 4 组：不嵌入（对照组），以及 $\lambda = 0.1$,1,10。我们请听众根据 [3] 的 5 级损伤等级量表对噪声质量进行评分（5：无法察觉；4：可察觉但不令人烦恼；3：略微令人烦恼；2：令人烦恼；1：非常令人烦恼）。本次实验共有 46 位不同的 Turker 用户，其中 21 位至少参与了 40 次评分。图14展示了结果。平均意见分（MOS）与 λ相关，但 $\lambda = 0.1$ 和 $\lambda = 1$之间差异很小，这表明鉴于前者的几何失真较低，可将其作为经验法则。

4 讨论 / 补充材料

我们提出了一种基于模型的方案，用于在音频中隐藏艺术曲线。经定量测量和人工评估验证，这些曲线在有损压缩后仍能保留，同时也较难被察觉。我们将曲线的维度隐藏在经过时间正则化的短时傅里叶变换频率幅度中，不过，任何正交分解的系数都可以用于此目的（我们也实现了小波，但发现其在听觉上更容易被察觉）。我们的方案极为简单，无需训练。解码几乎是即时完成的，因为只需计算少数几个频率的短时傅里叶变换。

为了展示我们的处理流程，我们使用 WebGL 和 JavaScript 创建了一个交互式查看器，可加载并解码任意 MP3 文件。查看器会播放解码得到的曲线，并使其与音乐同步。我们提供了多种预先计算的二维和三维示例，以展示我们的能力。要查看补充材料、用于编码和解码的源代码以及在线示例，请访问 https://github.com/ctralie/AudioCurveSteganography

参考文献

[1] Applegate, D.：《Concorde：用于求解旅行商问题的代码》。http://www. math. princeton. edu/tsp/concorde. html（1986 年）

[2] Atoum, M.S.、Ibrahimn, S.、Sulong, G.、Zeki, A.、Abubakar, A.：探索 MP3 音频隐写术面临的挑战。载于：2013 年高级计算机科学应用与技术国际会议。第 156–161 页。IEEE（2013 年）[3] Bassia, P.、Pitas, I.、Nikolaidis, N.：时域中的鲁棒音频水印。IEEE 多媒体汇刊 **3**（2），232–241（2001 年）[4] Bosch, R.：连接点：TSP 艺术的方方面面。载于：布里奇斯吕伐登会议：数学、音乐、艺术、建筑与文化。第 235–242 页（2008 年）[5] Bosch, R.：作为乔丹曲线的 Jordan。《献给 Gardner 的数学魔法》第 175 页（2009 年）[6] Bosch, R.、Herman, A.：通过旅行商问题绘制连续线条。《运筹学快报》**32**（4），302–303（2004 年）[7] Branch, M.A.、Coleman, T.F.、Li, Y.：用于大规模有界约束最小化问题的子空间法、内点法和共轭梯度法。SIAM《科学计算杂志》**21**（1），1–23（1999 年）[8] Canny, J.：一种边缘检测的计算方法。IEEE《模式分析与机器智能汇刊》第 679–698 页（1986 年）[9] Chen, X.、Golovinskiy, A.、Funkhouser, T.：三维网格分割基准。ACM《图形学汇刊》（SIGGRAPH 会议论文）**28**（3）（2009 年 8 月）[10] Cui, W.、Liu, S.、Jiang, F.、Liu, Y.、Zhao, D.：用于图像嵌入音频隐写术的多阶段残差隐藏[11] Cvejic, N.、Seppanen, T.：一种用于高容量音频隐写术的小波域最低有效位嵌入算法。载于：2002 年 IEEE 第十届数字信号处理研讨会及第二届信号处理教育研讨会论文集。第 53–55 页。IEEE（2002 年）[12] Djebbar, F.、Ayad, B.、Meraim, K.A.、Hamam, H.：数字音频隐写术技术的比较研究。EURASIP《音频、语音与音乐处理期刊》**2012**（1），25（2012 年）[13] Dom`enech Abell´o, T.：将图像隐藏在其口述叙事中。硕士学位论文，加泰罗尼亚理工大学（2022 年）[14] Dutta, H.、Das, R.K.、Nandi, S.、Prasanna, S.R.M.：数字音频隐写术概述。IETE《技术评论》**37**（6），632–650（2020 年）[15] Eichelberger, M.、Tanner, S.、Voirol, G.、Wattenhofer, R.：接收隐藏在音乐中的数据。载于：第二十届移动计算系统与应用国际研讨会论文集。第 33–38 页。ACM（2019 年）

[16] Geleta, M., Punti, C., McGuinness, K., Pons, J., Canton, C., Giro-i Nieto, X.: PixInWav：用于在音频中隐藏像素的残差隐写[17] Gopalan, K.：通过频谱修改实现统一音频与图像隐写。载于：2009年IEEE工业技术国际会议。第1–5页。IEEE（2009年）[18] Gopalan, K., Wenndt, S.：通过不可察觉音调插入实现隐蔽数据传输的音频隐写术。载于：IASTED通信系统与应用国际会议（CSA 2004）论文集，加拿大班夫（2004年）[19] Gopi, M., Eppstien, D.：具有任意拓扑的流形的单条带三角剖分。载于：《计算机图形学论坛》。第23卷，第371–379页。Wiley Online Library（2004年）[20] Griffin, D., Lim, J.：基于修正短时傅里叶变换的信号估计。IEEE声学、语音与信号处理汇刊**32**（2），236–243（1984年）[21] Gruhl, D., Lu, A., Bender, W.：回声隐藏。载于：信息隐藏国际研讨会。第295–315页。Springer（1996年）[22] Johnson, D.S., McGeoch, L.A.：旅行商问题：局部优化中的一个案例研究。《组合优化中的局部搜索》**1**（1），215–310（1997年）[23] Kaplan, C.S., Bosch, R.：TSP 艺术。载于：Renaissance Banff：数学、音乐、艺术与文化。第301–308页（2005年）[24] Kolmogorov, V.：Blossom V：最小成本完美匹配算法的新实现。《数学规划计算》**1**（1），43–67（2009年）[25] Lewis, J.：快速模板匹配，Vision Interface 95。加拿大图像处理与模式识别学会，第15–19页（1995年）[26] Li, Andreeto, Ranzato, Perona：Caltech 101数据集（2022年4月）。https://doi.org/10.22002/D1.20086[27] Li, H., Mould, D.：通过基于优先级的误差扩散实现结构保持点描。载于：Graphics Interface 2011论文集。第127–134页（2011年）[28]Madhavapeddy, A., Scott, D., Tse, A., Sharp, R.：音频网络：被遗忘的无线技术。《IEEE普适计算》**4**（3），55–60（2005年）[29] Malik, H.M.A., Ansari, R., Khokhar, A.A.：使用全通滤波器实现音频中的鲁棒数据隐藏。IEEE音频、语音与语言处理汇刊**15**（4），1296–1304（2007年）

[30] Mathews, P.D.：以自己的形象创作音乐：Aphex Twin 的脸。《 Nebula》**1**（1），65–73（2004年）[31] Mokhtarian, F., Mackworth, A.K.：面向平面曲线的多尺度曲率形状表示理论。IEEE模式分析与机器智能汇刊**14**（8），789–805（1992年）[32] Qiao, M., Sung, A.H., Liu, Q.：MP3stego 的隐写分析。载于：2009年国际联合神经网络会议。第2566–2571页。IEEE（2009年）[33] Secord, A.：加权沃罗诺伊点彩。载于：第2届国际非真实感动画与渲染研讨会论文集。第37–43页（2002年）[34] Takahashi, N., Singh, M.K., Mitsufuji, Y.：源混合与分离鲁棒音频隐写[35] Tzanetakis, G., Cook, P.：音频信号的音乐流派分类。IEEE语音与音频处理汇刊**10**（5），293–302（2002年）[36] Xiaoxiao Dong, Bocko, M., Ignjatovic, Z.：通过音频信号相位操纵实现数据隐藏。载于：2004年IEEE声学、语音与信号处理国际会议。第5卷，第V–377–80页。IEEE（2004年）[37] Yun, H.S., Cho, K., Kim, N.S.：基于调制复数重叠变换的声学数据传输。IEEE信号处理快报**17**（1），67–70（2009年）