# 声音的量子表示：从机械波到量子电路

Paulo V. Itabora´ı[1] 和 Eduardo R. Miranda[1]

[1]普利茅斯大学计算机音乐研究跨学科中心（ICCMR），英国普利茅斯{paulo.itaborai,eduardo.miranda}@plymouth.ac.uk

arXiv:2301.01595v1 [quant-ph] 1 Jan 2023

**摘要**

截至本文撰写时，量子音频仍是一个非常年轻的研究领域，即便在量子信号处理领域也是如此。本章介绍量子音频的最新进展，并讨论音频信号的量子表示方法。目前，没有任何一种量子表示策略宣称最适用于音频应用。每种策略都有其优缺点。可以认为，未来的量子音频表示方案将针对特定应用采用多种策略。作者还讨论了.....。

**注意：**这是章节提交前草稿的未编辑节略版本，标题相同，发表于 E. R. Miranda 编著的《量子计算机音乐：基础、方法与高级概念》（第 223–274 页）。有关应用示例，以及基于量子音频表示的声音合成方法及其在开发新型乐器方面的潜力的讨论，请参阅该书中的版本。

https://link.springer.com/book/10.1007/978-3-031-13909-3

## 1 引言

声音与图像有许多共同之处。从信息处理的角度来看，它们都只是信号。但显然，我们对它们的感知方式不同。

大多数用于声音的信号处理方法同样适用于图像，反之亦然。它们的主要区别在于维度。例如，声音是时域中的一维（1D）信号，而图像是二维（2D）信号。从数学角度来看，信号的维度越高，表示和处理它就越复杂。因此，先了解一维信号的表示和处理方法如何运作，再将其推广到二维（图像）、三维（视频）乃至更高维度，是合乎逻辑的。总的来说，音频和视觉信号处理教材通常会这样引入这一主题；例如， [1] [2]。

因此，从历史角度来看，似乎可以合理地预期，在图像和视频的量子表示及相关算法出现之前，研究领域就已出现声音的量子表示及潜在算法。令人惊讶的是，事实并非如此。人们对开发用于图像处理（主要用于人脸识别及类似应用）的量子算法抱有浓厚兴趣，因此迅速提出了量子图像表示与处理方法，而一维声音却被落下了将近

十年。考虑到语音技术对电子产业的重要性，这一差距令人意外。更何况，人们早已期望利用量子计算进行自然语言处理 [4]。

首批从理论上描述如何在量子处理器上表示图像的论文发表于 2003 [17]和 2005 [16]。相比之下，首批提出声音量子表示方法的论文发表于 2015 [6] 和 2018 [7]；这些方法基于 2011 [5]提出的一种图像表示方法。事实上，迄今设计的大多数量子信号处理算法都面向图像应用。量子声音领域需要迎头赶上。这正是本章的写作动机。

本章结构如下：首先，第3节将简要介绍声音的基本特征与概念，以及后文将用到的模拟和数字音频表示方法。该节总体说明原始声音内容如何从一种媒介转换到另一种媒介。本节介绍的概念将帮助我们凭借更好的直觉迈入量子领域。本节最后初步说明量子信息逻辑将如何应用于音频信号。

引言之后是一个简短章节（3），介绍如何在量子态中编码时间信息。接下来的注释（3.2）指出，文献中用于量子音频的命名法存在一些混淆问题，并提出一种新的命名系统，以统一并澄清用于编码音频信息的不同策略。

接下来的两个章节将深入探讨使用前文提出的命名系统，在量子机器中表示音频的各种定义和方案。例如，第 4 节探讨基于系数的表示，而第 5 节则聚焦于基于状态的表示。

第 6 节总结前文各节介绍的表示方法，并讨论在考虑为先进量子硬件构建和运行量子音频电路时应注意的一些障碍。

第**??** 节介绍一些基本的量子信号处理运算与电路设计，例如量子音频加法和拼接，以及逐样本比较。

最后一节（**??**）简要探讨近期及短期内的一些潜在艺术应用。具体而言，其中有一个案例研究，探讨了波表合成，以及利用基于系数的表示中的可配置参数实现的简单效果器。

# 2 从经典到量子

本节的主要目的是回顾声音在数字计算机及一般电子设备中的表示方式，并为理解如何在量子计算中表示声音奠定基础。

为了将给定的声音存储在量子计算机中（即在本文撰写时），需要先将声音录制为模拟音频，再进行模数转换。随后，需要将数字音频转换为某种量子音频表示（图1）。下面我们将详细介绍这些步骤。

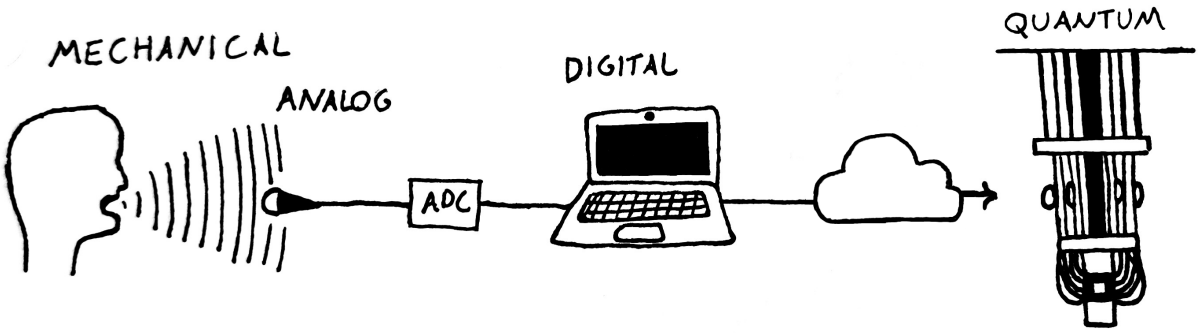


**图1：** 从机械声音到量子声音的信号路径。

## 2.1 从机械到模拟

从最广义上说，波是携带特定介质中扰动信息的物理实体。例如，电磁波携带电磁场中扰动的信息。然而，声音是一种机械波。通过向声学介质（如空气）中的一组耦合振动系统提供能量，便可产生声音。

到达麦克风时，振动的空气压力——也就是声音——会转变为电压。这就是模拟音频。实现这种机械电转换的方法有很多。最早开发出的技术之一，至今仍应用于动圈式麦克风。

动圈式麦克风由三个主要部件组成：一片薄振膜、一块圆柱形磁铁和一个线圈。线圈缠绕在磁铁周围，但不与磁铁接触。它可以沿圆柱体的轴线自由移动。振膜连接在线圈的一端。因此，振动的声音会带动振膜前后运动。结果，线圈会在磁铁中往复运动——或者从线圈的视角来看，磁铁会在线圈中往复运动。简而言之，法拉第定律告诉我们，磁铁在导电线圈中移动时会感应出电流。这种运动会产生可测量的电压。因此，振荡的线圈会在其两端感应出振荡电压。这样，机械声音便转化成了变化的电压信号，按定义而言，这就是模拟音频。

### 2.1.1 音频编码

模拟音频在机械介质与电介质之间建立了直接联系。我们可以说，它代表声音。

模拟音频作为一种传播和处理声音的新技术应运而生，而在它发明之前，这是不可能实现的。例如，歌手可以试着唱得更大声，甚至放声尖叫，试图让远处的听众听见自己，但仍可能无法如愿。或者，这位歌手可以将自己的声音转换成音频，然后通过导线或电磁波传输信号，直到信号到达听众那里。随后，扬声器会将音频重新转换成声音。需要注意的是，传输、接收和转换过程容易受到噪声干扰，也容易出错。

为了在长距离传输（例如无线电传输）过程中减少噪声和错误，人们开发了最初的音频编码方案，以有效表示电音频信息。最常用的三种模拟音频编码方案是调幅（AM）、调频（FM）和调相（PM） [3]。在这些情况下，原始模拟信号通过以下方式进行编码：

对信号参数进行可测量的变化，例如频率、振幅（即功率）或相位。

## 2.2 从模拟到数字

数字计算机无法处理连续的音频流，因为音频流包含所有可能时间点上的电压信息。音频信号中这样的时间点有无穷多个。因此，需要将模拟信息数字化。这通过模数转换器（ADC）设备完成，该设备可将模拟信号转换为数字信号。具体如何实现？如何将连续的时间和振幅信息转换为离散的二进制编码?

本节介绍采样 和量化的概念，它们分别是对时间和振幅进行离散化。

音频采样是对连续信号进行快照并将其（即快照捕获的数值）存储为带有时间索引的样本 的过程（图2）。

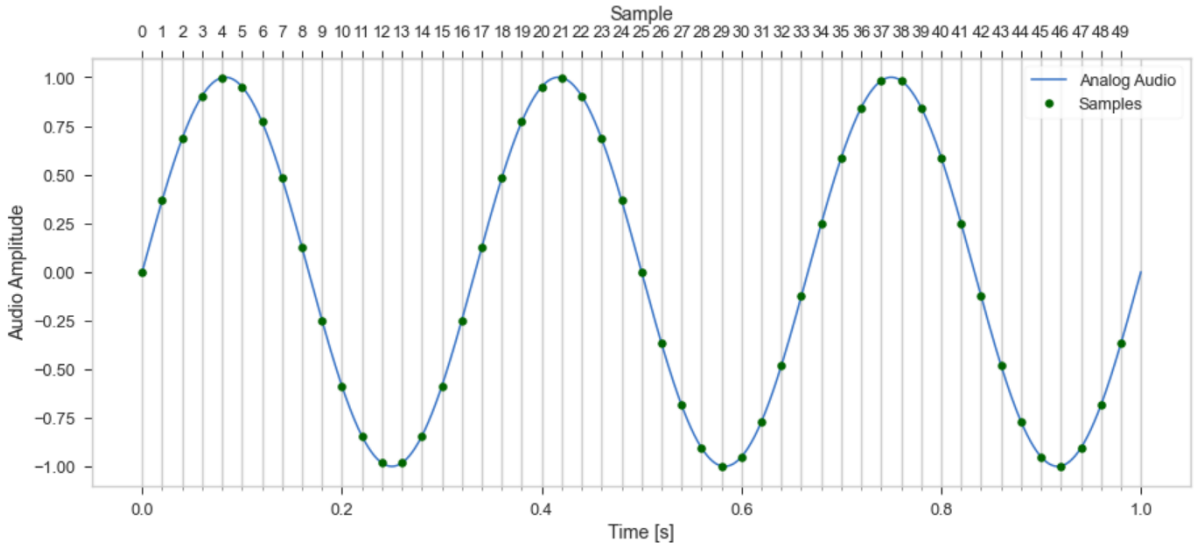


**图2：** 采样后的正弦波。

如图2所示，通常以等间隔的时间间隔对信号进行快照。间隔的速度称为采样率 或采样频率。换句话说，采样率规定了单位时间内采集的样本数量。高品质音频系统使用的采样率为44,100 Hz（即每秒采集44,100个快照）。采样率的作用是将索引 $k$ =

0，1，2，.....对应到各自的时间瞬间 $t_k$。也就是说，采样率 $S_R$ 是负责将快照时刻（以时间单位表示）转换为时间索引（无量纲）的常数（式 1）。该索引 $k$ 可以用二进制形式表示，因此能够存储在经典计算机中。

$$t_k = \frac{k}{S_R} \tag{1}$$

现在，我们来看看量化这一步。请注意，此处的量化概念与量子力学毫无关系。这里，量化是指将连续范围限制在一组预先规定的取值中。

让我们来看看模拟信号的振幅如何量化为二进制数（图3）。其振幅通常表示在−1到1的范围内。

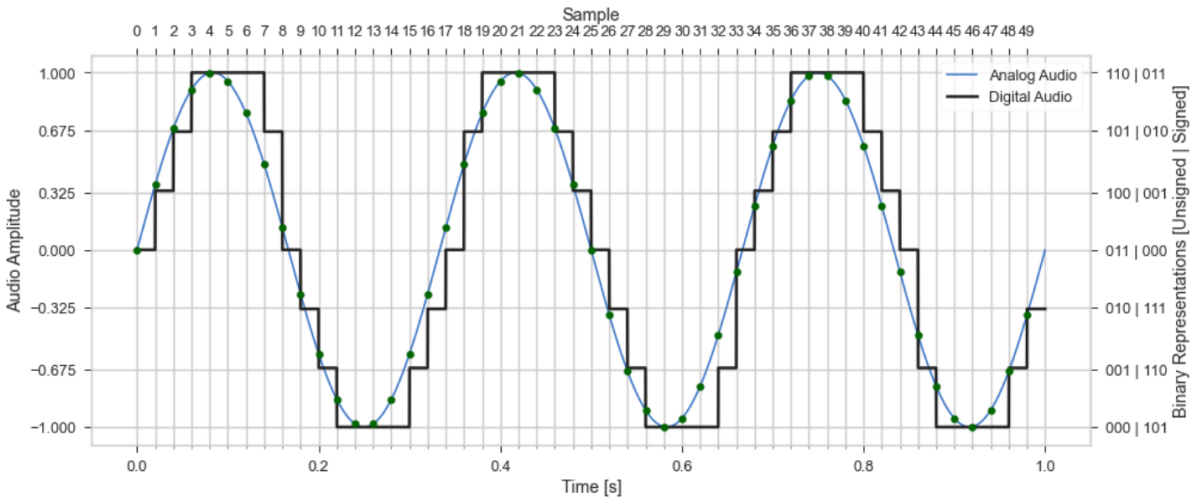


**图3：** 模拟波及其对应的量化。

量化在某些方面与采样类似，但它沿着纵轴进行。主要区别在于，数值可以为负。不过，我们可以重新缩放并平移整个信号，使其完全落在正值域内，而不失一般性。这是音频信号处理中的一种实用操作。

沿某个确定的轴平移信号的方法，是对信号的所有值加上或减去一个确定的量。本例中，我们希望将整个信号上移 1，因此给每个振幅值都加上 1。这样处理后，振幅范围变为 0 到 2，而非 $-1$ 和 1。此时范围已为正值，但我们还可以进一步将信号重新缩放到 0 到 1 的范围；这对后续讨论量子音频很重要。重新缩放通过乘法和除法完成，当然在本例中，我们将振幅除以 2。采用这个单位范围后，我们就能更轻松地进行量化。如果使用 $n$ 位字来存储每个振幅，就会在该范围内划分出 $2^n$ 个细分区间。随后，将电压值四舍五入，使其符合细分方案所允许的最接近的二进制值。

因此，如图3所示，我们通过将音频信号映射到网格上，成功地将其数字化。

### 2.2.1 二进制补码方案

上述量化方案完全有效，且适用于正值范围内的任何类型的信号。然而，我们已经看到，音频信号也具有负值。

当然，这本身并不是问题。上文已说明如何将音频信号移至正值域。不过，这种移位操作往往会使音频信号不适合进行某些特定操作。一个值得注意的例子是振幅乘法。将一个值移至严格的正值域，可能会大幅改变简单乘法的结果（如式 2 所示）。

$$(a_1 + shift)(a_2 + shift) \neq a_1 a_2 + shift \tag{2}$$

没有负数这一点，至少也可能带来一些不便。比如，考虑如何将两个音频信号相加。对数字化音频波进行相加的方法，是逐项求和其振幅。具体来说，设想一个过度简化的数字噪声消除算法。它依赖负数。首先，接收到音频信号中的一些不需要的噪声。然后，生成一份该噪声的副本，并将其极性反转（换句话说，所有振幅都乘以 $-1$）。最后，

将反相信号与原始信号相加。每个振幅都与其自身的相反数相加，也就是说，它们彼此抵消，噪声便消失了。这种现象通常称为相消干涉[1]。即使可以通过复杂的方法，仅用正数实现预期效果，采用较省力的替代方案仍然更可取。

一种可行的解决方案——也是通常用于表示音频信号的方法——是采用对二进制整数的另一种解读方式。这样既能表示负数，也有利于进行二进制加法运算。这种解决方案称为二进制补码编码方案。

理解二进制中表示负数的底层逻辑非常重要，因为这对量子音频表示也很有帮助。

二进制补码编码方案的核心思想是将二进制字分为两组：一半表示负数，另一半表示正数。具体做法是将最高有效位（即最左侧的位），也就是 MSB，作为符号位。如果 MSB 为 0，则该数为正数；否则为负数（式 3）。

$$0010... \longrightarrow \overset{(+\ or -)}{0} \quad \overset{(number)}{010...} \tag{3}$$

然而，单靠符号位还不足以表示一个“理想”的负数。要使其有用，还需要符合二进制加法性质。

让我们考虑一台采用4位架构的计算机，并使用最高有效位来表示后面3位的信号，不应用任何其他逻辑。现在，让我们试着用这种新的解释将2与 $-2$ 相加。考虑式4中简化的二进制加法规则，看看将这些数相加时（式5）会发生什么。

$$\begin{aligned} &0 + 0 = 0 \\ &0 + 1 = 1 + 0 = 1 \\ &1 + 1 = 10 \quad \{\text{The bit to the left is added by } 1\} \end{aligned} \tag{4}$$

$$2 \to 0\,010\,; \quad -2 \to 1\,010 \ \therefore\ 0\,010 + 1\,010 = 1\,100 \to -4\,(?) \tag{5}$$

显然，式 5 所示的做法行不通。不过，这个问题可以解决。我们可以为编码方案再添加一个条件。也就是说： $x + (-x) = 0$。有趣且方便的是，这个条件会彻底改变负数的表示方式，如式 6 所示。现在，我们可以通过式 7 验证，加法能够完美实现。

$$\begin{array}{ccccc} \textit{Regular Integer} & & & \textit{Two's Complement Integer} & \\ \begin{bmatrix} 15 \\ 14 \\ ... \\ 8 \end{bmatrix} & \begin{bmatrix} 1\ 111 \\ 1\ 110 \\ ... \\ 1\ 000 \end{bmatrix} & & \begin{bmatrix} 0\ 111 \\ ... \\ 0\ 001 \\ 0\ 000 \end{bmatrix} & \begin{bmatrix} 7 \\ ... \\ 1 \\ 0 \end{bmatrix} \\ = & & \searrow & = & \\ \begin{bmatrix} 7 \\ ... \\ 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0\ 111 \\ ... \\ 0\ 001 \\ 0\ 000 \end{bmatrix} & & \begin{bmatrix} 1\ 111 \\ 1\ 110 \\ ... \\ 1\ 000 \end{bmatrix} & \begin{bmatrix} -1 \\ -2 \\ ... \\ -8 \end{bmatrix} \end{array} \tag{6}$$

$$2 \to 0\,010\,; \quad -2 \to 1\,110 \ \therefore\ 0\,010 + 1\,110 = 0\,000 \to 0 \tag{7}$$

[1]这也出现在量子理论的语境中，因为量子系统也表现出类波行为，因此具有叠加态和干涉等类波性质。但不要因此产生误解。经典波与量子粒子波在本质上仍然不同，即使它们具有相似的数学性质

### 2.2.2 将数字音频表示为数组

音频数字化后，其二进制信息会存储在一种数据结构中。数字音频最常见的数据结构是数组向量，如图 4 所示。时间信息会成为数组中特定位置的索引，而对应的振幅则成为存储在该位置的变量。这样，只需提供时间索引，就可以检索相应的振幅。

| $a_0$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ |
|---|---|---|---|---|---|---|---|
| $t_0$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ |

**图 4：** 音频数组可视化。

为了高效地将二进制数据结构表示、存储并传输为音频信息流，通常会使用数字音频编码方案。与上文所示的模拟情况类似，数字域提供了一些可控（或可调制）的参数来表示信息。例如，我们可以使用由压电晶体精确控制的脉冲来表示数字信息流。在音乐产业中占据主导地位的表示方式是 PCM 音频，即脉冲编码调制。PCM 凭借可靠的无损存储和传输能力，已成为数字音频的标准。此外，与脉冲幅度调制（PAM）、脉冲位置调制（PPM）和脉冲宽度调制（PWM [3]等其他脉冲表示方式相比，它更不易受到噪声和错误的影响。

需要指出的是，二进制补码数字音频广泛应用于各种场景。然而，现代 32 位和 64 位操作系统，以及 Python 等高级编程语言，通过使用浮点数进一步改进了量化音频振幅的表示。浮点数可以高精度地存储实数（$\in \mathbb{R}$）的良好近似值，并保留足够多的有效小数位。因此，我们不必存储表示量化范围中间值的整数，而可以增加一个预处理层，存储该数值本身的良好近似值，并利用浮点数的灵活性在数字介质中处理信号。不过，这需要占用大量空间和计算能力。正如本章后续内容将会看到的，我们距离实现类似的量子计算机音频方案仍很遥远，但这并非不可想象。

## 2.3 从数字到量子

直观地理解量子音频表示并不难，因为狄拉克符号能够清楚地展示如何将音频信息（时间和振幅）转化为量子态（图 5）。当然，其底层的量子计算范式可能有些反直觉。不过，这种对量子音频组织结构的直观鸟瞰式理解，为我们指明了最终目标，也为我们穿越充满未知的量子算法道路提供了指引。在这条道路上，我们将探讨如何制备和测量这些量子音频表示，以及它们在近期和远期的潜在应用。

需要指出的是，即使在量子信号处理这一领域中，量子音频仍是一个非常年轻的研究方向。因此，许多基本问题仍有待探讨。目前还没有哪种音频表示策略能够被认为最适合音频应用。每种策略都有其特定的优缺点。可以说，量子音频将采用多种表示方式，分别面向特定应用。

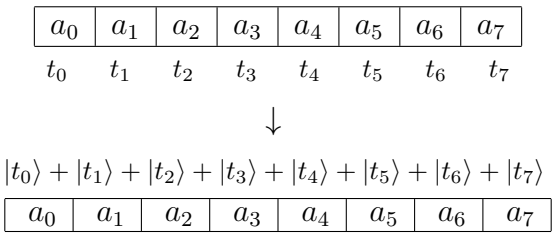


**图 5：** 从数字到量子

# 3 量子音频的制备与检索

量子技术的近期发展前景表明，量子计算机不会在许多方面取代经典计算机。相反，它们将共存，形成混合型经典-量子处理环境。这种情况已经出现在当前运行的量子系统中，其中许多可通过云服务访问，并与经典机器交互。近期的量子音频应用也不会例外。这些应用高度依赖经典计算机，不仅要用它们提供待处理的数字音频文件，还要用它们对量子电路测量结果进行后处理。此类工作有一些广为人知的经典编程语言，例如 Python 和 MatLab。在接下来的章节中，我们将详细考察迄今提出的量子音频表示方式，探讨量子音频如何制备和测量，并展望其可能的应用。首先，让我们介绍时间索引与量子叠加态之间的联系。

## 3.1 编码时间信息

如前所述，量子比特是一种二态量子系统，测量结果可以是 0 或 1。它也可以处于态叠加中，如公式 8 所示，其中 $|\alpha|^2$ 和 $|\beta|^2$ 分别表示测量该状态时结果为 0 或 1 的概率。由于所有可能结果的概率之和必须为 1，因此这意味着 $|\alpha|^2 + |\beta|^2 = 1$。

$$|\Psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{8}$$

对于双量子比特态，考虑所有可能的结果，也可以写出类似的公式，如公式 9 所示。

$$|\Psi\rangle = a\,|00\rangle + b\,|01\rangle + c\,|10\rangle + d\,|11\rangle\,; \qquad |a|^2 + |b|^2 + |c|^2 + |d|^2 = 1 \tag{9}$$

由于字母表中的字母有限，无法表示许多这样的概率幅，因此稍微改变一下记法并使用带下标的同一个字母可能更好。例如，以这种方式写出的三量子比特态如公式 10 所示。

$$\begin{aligned} |\Psi\rangle = {} & a_{(000)}\,|000\rangle + a_{(001)}\,|001\rangle + a_{(010)}\,|010\rangle + a_{(011)}\,|011\rangle \\ & + a_{(100)}\,|100\rangle + a_{(101)}\,|101\rangle + a_{(110)}\,|110\rangle + a_{(111)}\,|111\rangle \end{aligned} \tag{10}$$

此时，我们还应对记号做一处有益的调整，以增强直观性。有时，将右矢中的数字理解为与某个整数相关联的经典二进制比特串，而不是寄存器中各个量子比特的状态序列，会更方便（也符合惯例）。

在此之前，我们必须提醒自己，“右矢”中的这些 0 和 1并不是数字。右矢是一种用于书写向量的简化记法。因此，右矢内部写的任何内容，都只是一个约定俗成的标签，用于指代一个向量：

$|label\rangle$。我们用数字作为标签，以降低这些数学实体的抽象程度，并帮助理解它们在系统中的用途。换句话说，上述解释并没有在数学意义上改变状态，只是改变了记法。清楚地认识到这一点并牢记于心至关重要，这样在接下来介绍各种表示法时，才能避免混淆。

因此，将量子比特状态解释为二进制数后，我们可以像式 11 所示那样写出式 10 中相同的 3 量子比特状态。

$$|\Psi\rangle_{3-qubit} = \frac{1}{\sqrt{8}}\big[\,|0\rangle + |1\rangle + |2\rangle + |3\rangle + |4\rangle + |5\rangle + |6\rangle + |7\rangle\,\big] \tag{11}$$

本文介绍的所有量子音频表示都采用相同的时间信息编码策略。它们使用量子寄存器创建与每个音频采样相关的所有可能时间索引的叠加态。这称为时间寄存器。每个状态都是一个索引，表示时间上的一个位置，类似于经典数组。

与该状态相关的任何信息都会采用不同策略对相应的样本进行编码（式 12）。例如，可以使用概率振幅，或使用另一个耦合的量子寄存器。

$$(Amplitude)\,|t_k\rangle \tag{12}$$

请注意，所有表示都需要长度为 $2^n$ 的信号。对于大小不同的音频，我们可以使用零填充。下面我们将更详细地探讨不同方案如何表示振幅信息。

## 3.2 命名法说明

在继续之前，我们先说明本文用于指称下文将要回顾的各种量子音频表示（QAR）方法的命名法。为求清晰并实现系统化，我们对这些方法最初发表于研究论文时所采用的名称作了少许调整。

根据如何编码音频振幅信息，以及如何通过测量将其取回经典领域，这些表示方法可分为两类。第一类称为“概率型”或“基于系数”的表示方法，因为以经典方式取回信息时具有概率性。第二类包括“确定型”或“基于状态”的方法。

如前所述，量子音频表示方法源自用于表示图像而非声音的方法。研究文献介绍了数字音频的量子表示（QRDA）、量子音频灵活表示（FRQA）和多通道音频的量子表示（QRMA）等方法，这些方法都属于基于状态的方法。不过，也有一些用于图像的基于系数的方法，可以（并将在本章中）轻松改编用于音频表示。

此外，本章旨在面向信号处理领域的研究群体；为此，我们将为已有的表示方法提出一套新的命名体系。我们的目的是在未来将这些表示方法与经典表示方法关联、整合或统一起来。为此，我们将以“调制”一词为核心。这意味着我们将提议重新命名，例如，将量子音频灵活表示（其本身已经存在一些混淆问题

，尤其是与量子图像[2]的关系）改称量子态调制（QSM），因为振幅信息存储在多量子比特态中。这套命名系统也为其他基于系数的音频表示方法铺平了道路，例如量子概率振幅调制（QPAM）。

# 4 基于系数的表示

假设我们有一段数字音频 $A$，包含 $N = 2^n\ n \in \mathbb{Z}^*$ 个样本，每个样本量化为 $[-2^{q-1}, -2^{q-1}+1, ..., 2^{q-1}-1]$。也就是说，有 $2^q$种可能的取值。

我们可以推断，表示这些样本最简单的方式是将它们置于量子态$|A\rangle$ 中：对所有可能的状态 $t$ （编码时间）进行叠加，并按照各自的概率振幅（编码样本值）加权，使其类似于数组，只不过处于量子叠加态中（式 13 和式 14）。

我们可以推断，表示这些样本最简单的方式是，首先对所有可能的状态 $t$（编码时间）进行叠加，从而构造量子态 $|A\rangle$。然后，使用各自的概率振幅（编码样本值）对每个时间状态进行加权。

$$\boxed{a_n}\, t_n \longrightarrow \alpha_i \, |t_i\rangle \tag{13}$$

$$|A\rangle_{(3-qubitArray)} = \alpha_0 \, |0\rangle + \alpha_1 \, |1\rangle + \alpha_2 \, |2\rangle + \alpha_3 \, |3\rangle + \alpha_4 \, |4\rangle + \alpha_5 \, |5\rangle + \alpha_6 \, |6\rangle + \alpha_7 \, |7\rangle \tag{14}$$

## 4.1 量子概率振幅调制：QPAM

更一般地，我们可以表示大小为 $N$ 的任意量子音频，其中使用 $n$（或 $\lceil \log\ N \rceil$）个量子比特，并构建一种编码方案，将音频的每个可能振幅值映射为一个概率振幅（式 15）。

$$|A_{QPAM}\rangle = \sum_{i=0}^{N-1} \alpha_i \, |i\rangle \tag{15}$$

量子计算理论通常会在求和上限中明确标出量子比特的数量，例如（$2^n - 1$）。我们则选择使用 $N = 2^n$ 作为主要记法，因为对于具有音频数字信号处理背景的读者来说，这样更容易理解。

通过公式 15，我们实现了一种简单但仍然非常实用的 QAR。我们将这种方法称为量子概率振幅调制音频表示，或 QPAM。这是一个便于使用的名称，因为振幅信息被编码为每个量子态的概率振幅。

[2]直觉上，人们可能会认为，量子音频灵活表示（FRQA）源自量子图像灵活表示（FRQI）。遗憾的是，事实并非如此：FRQI 是一种概率导向表示，而 FQRA 是一种状态导向表示，源自新型增强数字图像量子表示（NEQR）。将 FQRA 命名为此的原因并不明确，或许是出于历史原因。FRQI 是早期的图像表示方法之一（与 FRQA 类似），也可能是文献中最早得到广泛研究并拥有多种应用的表示方法。

### 4.1.1 将数字振幅映射到 QPAM

现在，让我们来考察如何将数字信号表示转换为 QPAM 表示，反之亦然。

音频振幅 $a_n$ 并不等于测量每个样本态的概率振幅 $\alpha_i$。相反，两者之间存在特定的映射 关系，这使得这种编码方案成为可能。让我们通过图 6 所示的假想音频片段来了解其原理。

在量子力学中，我们可以断言，对于任意量子比特 $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$，测量其处于状态 $|0\rangle$ 的概率为 $|\alpha|^2$，前提是：

- 概率是介于 0 和 1 之间的数
- 所有可能状态的概率之和应等于 1

然而，图6中的数字音频振幅是介于 $-1$ 和 1 之间的数。它们的总和不一定为 1。因此，要从数字转向量子，我们需要采取以下步骤来归一化 这些振幅：

- 步骤 1：将所有振幅加 1 $a_n$
- 步骤 2：将振幅除以 2
- 步骤 3：再除以所有振幅之和
- 步骤 4：对结果取平方根

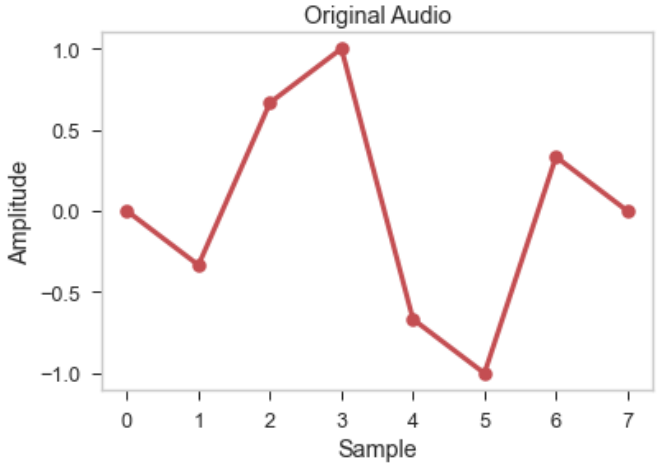


**图6：** 假想音频。

请看图7中向下箭头之前列出的示例振幅。归一化后的值显示在向下箭头之后。首先，我们将每个振幅值加 1，使其移至正值域。此时，振幅范围为 0 到 2。接着，我们对其进行缩放，使其落在 0 到 1 的范围内。然后，将所有值相加，并用每个值除以总和。此时，所有振幅之和应为 1。

最后一步是取这些振幅的平方根，将其转换为概率值（公式 16）。

$$\alpha_i = \frac{1}{\sqrt{g}}\sqrt{\frac{(a_i+1)}{2}}; \qquad g = \sum_k \frac{(a_k+1)}{2} \tag{16}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.0 | −0.3 | 0.7 | 1.0 | −0.7 | −1 | 0.3 | 0.0 |

↓

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.3 | 0.2 | 0.5 | 0.6 | 0.1 | 0.0 | 0.4 | 0.3 |

**图7：** 归一化过程。

### 4.1.2 QPAM 制备

QPAM 表示是最容易准备的。由于信息存储在概率振幅中，我们只需将量子音频初始化为所需的量子叠加态即可。可以利用概率振幅计算一个酉门来实现这一点。另一种方法是构造一个量子电路，将初始化为 $|0000...\rangle$ 状态的一组量子比特映射到任意状态。幸运的是，从实际应用的角度来说，我们不必费心计算这些矩阵和量子电路。如今大多数量子计算编程工具都提供了初始化量子比特状态的命令。因此，只需假设向相应命令提供一组概率振幅，就能初始化任意态叠加（图 8）。因此，要准备 QPAM 量子音频，我们只需将数字采样值转换为概率振幅。

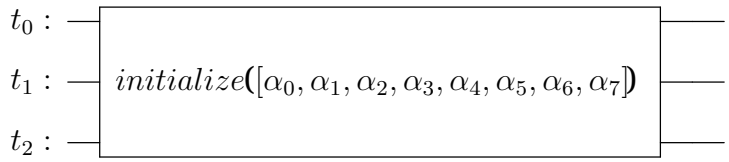


**图8：** 为 QPAM 初始化量子比特状态。

因此，制备任意 $n$ 量子比特状态通常需要 $\mathcal{O}(n)$ 个简单操作，这意味着 QPAM 制备需要 $\mathcal{O}(\lceil \log\ N \rceil)$ 个操作。

### 4.1.3 QPAM 检索

从量子域检索信息，意味着对我们的量子系统进行测量。测量量子系统时，该系统所有可能的状态结果中，只有一个会在测量装置中“显现”。换句话说，一旦测量时间寄存器，叠加态就会被“破坏”，最终只会返回一个可能的时间状态。

图9展示了图6所示音频的一种假设性 QPAM 表示（即振幅直方图）。此外，图10展示了通过一台据称完美的量子机器检索测量结果后重构的相应音频。让我们来详细探讨一下。

概率型表示之所以得名，正是因为它们的检索策略。检索音频需要准备多个完全相同的该音频量子版本，并对所有测量结果进行统计分析。分析会返回测量结果的直方图（例如图 9），我们可以据此估算测得每种状态的概率。考虑到系统最初的制备方式，就会发现直方图本身已经是经过缩放和平移的数字音频。我们只需使用公式 16 的反向版本，将音频缩放回原始范围，该版本针对 $a_i$，见公式 17。

$$a_i = 2g|\alpha_i|^2 - 1 \tag{17}$$

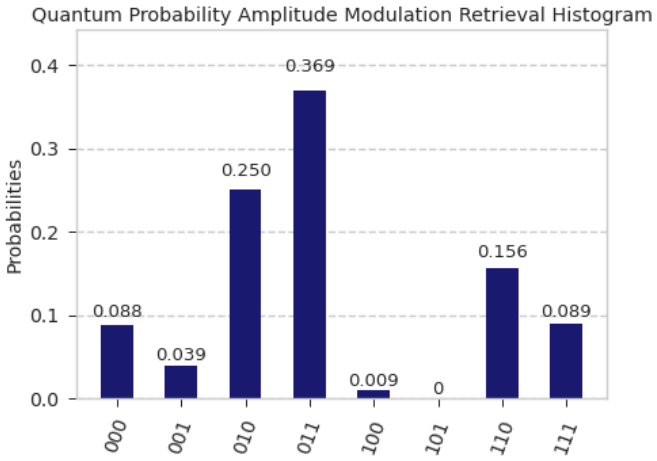


**图9：** 一种假设的 QPAM 表示。

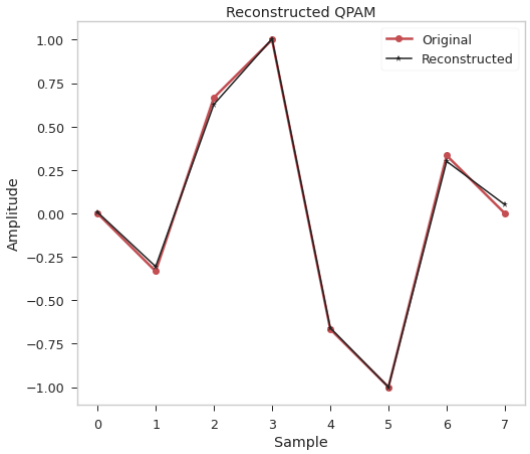


**图10：** 由图9中的 QPAM 重建的音频。

需要注意公式 17 的两个关键方面。首先，求和符号内的和是一个定义好的常数值，记作 $g$。其次，是在 $\alpha_i$ 上加上一个“|”。我们需要谨慎处理 $\alpha_i$，因为严格来说，它是一个复数（$\alpha_i^2 \neq |\alpha_i|^2$）。不过，有价值的信息在于其模平方，这对应于测得 $|i\rangle$ 的概率 $p_i = |\alpha_i|^2$。完成测量后，直方图箱不会返回 $p_i$，而是返回一个近似值 $\tilde{p}_i$。

将公式 17 替换为 $\tilde{p}_i$，即可完成检索数字音频的最后一步后处理，如公式 18 所示。

$$a_i = 2g\tilde{p}_i - 1 \tag{18}$$

### 4.1.4 注意事项

公式 16 中的常数 $g$ 带来了一个重要的注意事项。该公式包含与数字振幅 $a_k$相关的项。具体是什么注意事项?

让我们来考虑准备阶段。常数 $g$ 是分母，这意味着各项 $(a_i+1)\ 2$ 之和不能等于 0。如果 $a_i$ 的所有值都等于 $-1$，就会出现这种情况，而这完全可以是一个有效的音频信号。

从检索的角度来看，还有另一点需要考虑。$g$ 的存在使得必须对原始输入进行归一化，才能检索到 QPAM 表示。

不过，更一般地说，可以将 $g$ 视为任意常数，用来定义信号的整体振幅范围 。

如果没有适当评估， $g$ 的这种选择随意性可能会成为问题。它会不必要地改变还原后音频的振幅范围。我们通过一个思想实验来说明这一点。

考虑一个包含 8 个样本的 QPAM 音频信号。所有样本值均为 0。公式 16 确认了 $g = 4$。接着，假设某个量子算法将此信号转换为图 9 所示的测量波形。信号将使用公式 17 进行重建。具体而言，考虑直方图的峰值。应用公式 17 将得到 $2g|0.369|^2 - 1 = 0.089288$。此外，如果选取一个低概率样本：$2g|0.09|^2 - 1 = -0.999352$。因此，振幅范围会压缩。

解决这一重建问题的一种可能方法是使用输出的归一化结果来计算新的 $g$，如公式 19 所示。

$$g = \sum_k \tilde{p}_k \tag{19}$$

## 4.2 单量子比特概率振幅调制：SQPAM

另一种基于系数的量子音频表示实例，可以从一种称为量子图像灵活表示（Flexible Representation of Quantum Images，简称 FRQI）的图像表示方案中推导出来 [9]。我们提议将这种方法称为 单量子比特概率振幅调制，简称 SQPAM。

SQPAM 的工作方式与 QPAM 类似。不过，它不使用原始概率振幅，而是使用 $\lceil \log\ N \rceil + 1$ 个量子比特。它改进了编码逻辑，使用三角函数将样本编码到另一个专用量子比特的概率振幅中，并将其作为新寄存器 $|\gamma_i\rangle$ 添加。这种编码方案比 QPAM（公式 22）更可靠，也更易于编辑。

$$a_n \longrightarrow |\gamma_i\rangle = \cos\theta_i |0\rangle + \sin\theta_i |1\rangle \tag{20}$$

$$\begin{aligned} |A\rangle_{3-qubit} = \frac{1}{\sqrt{8}}\Big[&(\cos\theta_0 |0\rangle + \sin\theta_0 |1\rangle) \otimes |0\rangle + (\cos\theta_1 |0\rangle + \sin\theta_1 |1\rangle) \otimes |1\rangle + \\ &(\cos\theta_2 |0\rangle + \sin\theta_2 |1\rangle) \otimes |2\rangle + (\cos\theta_3 |0\rangle + \sin\theta_3 |1\rangle) \otimes |3\rangle + \\ &(\cos\theta_4 |0\rangle + \sin\theta_4 |1\rangle) \otimes |4\rangle + (\cos\theta_5 |0\rangle + \sin\theta_5 |1\rangle) \otimes |5\rangle + \\ &(\cos\theta_6 |0\rangle + \sin\theta_6 |1\rangle) \otimes |6\rangle + (\cos\theta_7 |0\rangle + \sin\theta_7 |1\rangle) \otimes |7\rangle\Big] \end{aligned} \tag{21}$$

将其推广到大小为 N 的音频时，有：

$$|A_{SQPAM}\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} (\cos\theta_i |0\rangle + \sin\theta_i |1\rangle) \otimes |i\rangle \tag{22}$$

### 4.2.1 SQPAM 音频振幅的映射

与前一种表示方式类似，需要经过一些预处理步骤，将振幅映射到归一化范围内。三角函数常用于表示概率振幅，因为其绝对值范围为 0 到 1，且余弦函数与正弦函数之间的关系完美满足归一化要求[3]。

[3] 余弦 $(\theta)^2$ + 正弦$(\theta)^2 = 1$

本质上，这描述的是一种简单的变量替换，其中 $\alpha_i = \cos\theta_i$ 和 $\beta_i = \sin\theta_i$。请注意，这里只考虑实数，与 QPAM 类似。

引入额外的量子比特 $|\gamma_i\rangle$ 显著改善了音频的表示方式，包括编码和检索。首先，编码方案不再依赖音频内容的整体大小（叠加态中的状态总数及其相对振幅）。这些内容被局部编码在新量子比特中，因此可以使用旋转矩阵 作为量子门独立操控。这开辟了新的可能性。

将音频振幅 $a_i$ 映射为角度的函数 $\theta_i$ 如公式 23 所示。该映射也可以理解为以下一组指令:

- 步骤 1：将所有振幅 $a_n$加 1。
- 步骤 2：将振幅除以 2。
- 步骤 3：求平方根。
- 步骤 4：对结果求反正弦 （只要在整个实现过程中遵循相同的约定，也可以使用反余弦 ）。

$$\theta_i = \sin^{-1}\left(\sqrt{\frac{a_i + 1}{2}}\right) \tag{23}$$

映射后的角度数组将用于制备 SQPAM 量子音频态。

### 4.2.2 制备

SQPAM 态制备需要 $n$ 个量子比特用于时间编码（与 QPAM 相同），还需要 1 个量子比特用于振幅（$\gamma_i$）。如式 24 所示，它们都可以初始化为 $|0\rangle$ 态。

$$|000...\rangle \equiv |0\rangle \otimes |0\rangle \otimes |0\rangle \otimes ... \equiv |0\rangle^{\otimes n+1} \equiv \underset{(\gamma_i)}{|0\rangle} \otimes |0\rangle^{\otimes n} \tag{24}$$

对时间寄存器中的每个量子比特施加哈达玛门，使其处于均匀叠加态（图11）。由此得到的量子态如式 25 所示。

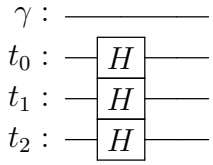


**图11：** 对 3量子比特时间寄存器施加哈达玛门的量子电路片段。

$$|\Psi\rangle = \frac{1}{\sqrt{N}}\Big[\,|0\rangle \otimes |0\rangle \;+\; |0\rangle \otimes |1\rangle \;+\; |0\rangle \otimes |2\rangle \;+ ... +\; |0\rangle \otimes |N-1\rangle\,\Big] \tag{25}$$

此时，我们应该讨论将振幅量子比特转换为所需形式（$|0\rangle \rightarrow |\gamma_i\rangle$）的具体策略。有一个非常适合此任务的量子门，称为 $R_y$。它会绕布洛赫球的 $y$ 轴旋转（式 26[4]）。

$$R_y(2\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{26}$$

[4] $R_y$（$2\theta$）的形式与许多线性代数教材中所见的二维旋转矩阵相同。主要区别在于，在布洛赫球中，角度 theta 的旋转速度是普通欧几里得空间中的两倍。

我们来看看将 $R_y(2\theta)$ 作用于量子态 $|0\rangle$ 时会发生什么。记住，量子态表示向量，因此这将归结为一次矩阵向量乘法（式 27）。

$$R_y(2\theta_i)\,|0\rangle = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix} = \cos\theta_i\,|0\rangle + \sin\theta_i\,|1\rangle = |\gamma_i\rangle \tag{27}$$

最后，需要有一种方法以某种方式遍历时间寄存器的不同索引，从而对它们应用特定的 $R_y(2\theta_i)$。这样一来，角度 $|\gamma_i\rangle$ 就会与各自的状态 $|i\rangle$纠缠在一起。如果这些状态处于叠加态，该如何访问并关联特定状态?

通过使用 $R_y$ 的多重受控版本来控制值设置量子门 $R_y$，即可实现所需的关联。让我们来探究受控门对量子态的影响。

设想一个长度为 1 个量子比特的音频示例。假设按照第一个制备步骤，在时间寄存器中创建了一个叠加态。该量子态可用式 28 描述。

$$|\phi_0\rangle = \frac{1}{\sqrt{2}}\Big[\,|0\rangle\,|0\rangle + |0\rangle\,|1\rangle\,\Big] \tag{28}$$

随后，应用一个受控 $R_y(2\theta)$ 门。通过分析图 12 中的量子电路，可以得出结论：只有当时间量子比特为 1 时，才会应用 $R_y$ 门。式 29 展示了所得的量子态。

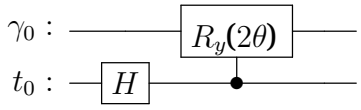


**图12：** 受控 $R_y(2\theta)$的应用

$$|\phi_1\rangle = cR_{y[t_0]}\,|\phi_0\rangle = \frac{1}{\sqrt{2}}\Big[\,|0\rangle\,|0\rangle + R_y(2\theta)\,|0\rangle\,|1\rangle\,\Big] \tag{29}$$

现在，让我们在前后各放置一个 $X$ 门（图13）。会有什么效果? 如果时间量子比特恰好处于 $|0\rangle$ 态，它就会翻转为 $|1\rangle$。因此，它会触发控制条件，从而得到式30。随后，它会翻转回 $|0\rangle$。请注意，否则它不会触发受控量子门。




**图13：** 使用两个 X门反转控制条件

$$|\tilde{\phi_!}\rangle = (I\otimes X)cR_{y[t_0]}(I\otimes X)\,|\phi_0\rangle = \frac{1}{\sqrt{2}}\Big[R_y(2\theta)\,|0\rangle\,|0\rangle + |0\rangle\,|1\rangle\,\Big] \tag{30}$$

$X$ 量子门构建了一种有效切换控制条件的策略：当量子比特处于 $|0\rangle$ 态而非 $|1\rangle$ 时触发。其表示方式是在控制位置标记一个白色圆圈，如图 13 右侧所示。

对于双量子比特系统，可以使用相同的策略来访问特定索引。在这种情况下，需要同时验证两个控制条件。例如，

第三个样本 - $\theta_2$ - 的地址是 $|10\rangle$。因此， $t_0$ （最右侧的位）的控制条件是"$|0\rangle$"，而 $t_1$ 是"$|1\rangle$"。公式 31 表明， $R_y(2\theta_2)$ 仅应用于第三个位置。

通过使用四个 2 控制 $R_y(2\theta_i)$ 门，可以逐一寻址并相应地存储每个振幅（图 14）。

$$
\begin{aligned}
&|\phi_0\rangle = \tfrac{1}{2}\Big[\,|0\rangle\,|00\rangle + |0\rangle\,|01\rangle + |0\rangle\,|10\rangle + |0\rangle\,|11\rangle\,\Big] \therefore \\
&|\phi_{\theta_2}\rangle = (I\otimes X\otimes I)(ccR_{y[t_0t_1]}(\theta_2))(I\otimes X\otimes I)\,|\phi_0\rangle \\
&= \tfrac{1}{2}\Big[\,|0\rangle\,|00\rangle + |0\rangle\,|01\rangle + R_y(\theta_2)\,|0\rangle\,|10\rangle + |0\rangle\,|11\rangle\,\Big]
\end{aligned}
\tag{31}
$$

推广这一策略，我们得到一个用于制备 SQPAM 音频的第二个制备步骤的算法。对于每个时间状态，

- 将该状态标签的二进制版本存储在一个经典变量中。
- 如果字符串中指定位置的位为零，则在相应的量子比特上应用一个 $X$ 门。

- 应用 $R_y(2\theta)$ 门。
- 按照第二条指令，再应用一个 $X$ 门。

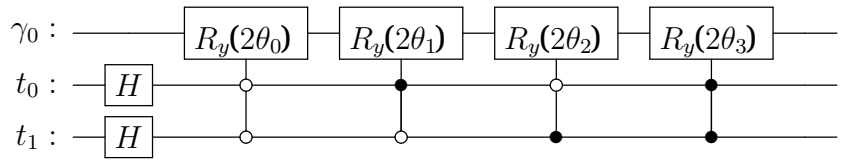


**图14：** 使用 $R_y$ 个量子门，通过值设置操作制备 SQPAM 表示

请注意，这种策略依赖于经典数值。根据框架所使用的编程语言，可以在构建量子电路时轻松实现这一点，只需使用经典的"for循环"即可。

这种准备策略（使用多重控制门）可以称为赋值操作。这种操作也适用于其他表示形式。

此准备过程应用 $\lceil logN\rceil$ 个哈达玛门，以及 $N$ $N$控 $R_y(2\theta)$ 门。由于 $N$控 $R_y(2\theta)$ 具有 $\mathcal{O}(N)$ 个基本操作，这意味着准备 SQPAM 音频需要 $\mathcal{O}\left(N^2\right)$ 个操作。

### 4.2.3 检索

图15展示了 SQPAM 检索直方图的预期结果。每个样本都编码在嵌入 $|\gamma_i\rangle$ 个概率中的互补正弦和余弦对里（式32）。因此，分箱数为16，而不是8。

在方程23中，有一个反正弦函数。因此，选取正弦分量（$p_{\gamma_i}(|1\rangle)$）后，就会呈现音频的轮廓。

然后，只需使用公式 33 即可重建信号[5]。图16展示了使用图15中的直方图和公式 33 进行的信号重建。

$$
p_{\gamma_i}(|0\rangle) = (\cos(\theta_i))^2 \qquad ; \qquad p_{\gamma_i}(|1\rangle) = (\sin(\theta_i))^2 \tag{32}
$$

[5]试想在式33的分子中使用余弦项（$p_{\gamma_i}(|0\rangle)$）。会发生什么？三角函数的互补性会使重建音频极性反向。

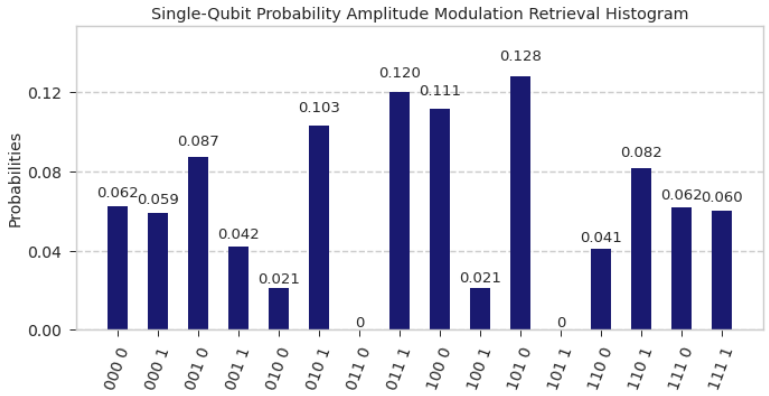


**图15：** 假设的 SQPAM 表示

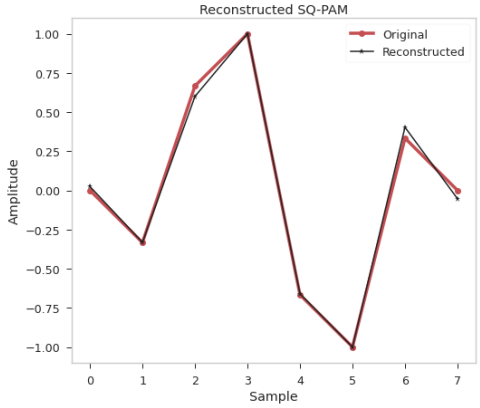


**图16：** 由图15中的 SQPAM 重建的音频

$$a_i = \frac{2p_{\gamma_i}(|1\rangle)}{p_{\gamma_i}(|0\rangle) + p_{\gamma_i}(|1\rangle)} - 1 \tag{33}$$

# 5 面向状态的表示

在上一节中，量子音频表示方案将时间域存储为多量子比特态的叠加态，并将振幅存储为概率振幅。本节的主要区别在于，音频振幅也将存储在多量子比特态中。这个新的量子寄存器与时间寄存器组合后，将产生一组纠缠信息的时间-振幅对。如果测量给定的时间样本 $t_i$，就会得到一个相关状态；测量该状态后，即可获得相应的振幅。

它可以被视为前述表示的一条演化分支（参见量子图像表示时间线）。例如，它们有可能 得到更细致的操控，并用于更广泛的应用。当然，这需要根据具体情况逐一讨论。目前还无法确定哪种表示类型更适用于不同的音频和音乐应用。下一节将对此作进一步讨论。

这些“更精细”的能力所带来的结果是，它们可能会增加表示和处理音频所需的空间。换言之，存储量子音频所需的量子比特数量与已知潜在应用的数量之间存在权衡。

此外，发展以下表示策略的另一个历史动因，与其制备复杂度有关（例如，编码所需的指令数量）。而且，将音频振幅存储为二进制量子态的逻辑，与第 2.2 节中量化后的数字振幅相一致。

## 5.1 QSM 和 uQSM

以下两种量子音频表示源自新型增强图像量子表示（NEQR） [5]。它们是最早提出的音频量子表示，即 QRDA（数字音频量子表示）和 FRQA（量子音频灵活表示）。从记法上看，两者完全相同，但对量子比特状态的解释略有不同。它们分别被解读为无符号整数或有符号整数。

QRDA [6] 是文献中最早提出的量子音频表示。它指明了其制备和检索技术的发展方向，并奠定了面向状态的表示的许多基础。在第 2.2 节中，我们讨论了如何将模拟振幅转换为数字值。样本被近似为量化方案中最接近的二进制整数。本节将使用相同的量化振幅来制备多量子比特态。

QRDA 使用专用的量子寄存器，通过严格的正整数对样本进行编码（这直接继承自 NEQR 图像；NEQR 图像对范围为 0 到 1 的灰度信息进行编码）。而 FRQA，即量子音频灵活表示，则使用二进制补码方案来解释振幅（第 2.2.1 节）。

有符号整数使音频混合变得极为方便，因为只需使用常规的逐位加法算法即可。这表明，相较于 QRDA，FRQA 对状态的解释方式更适用于量子音频。

除了变量类别的解释方式不同，制备和检索的逻辑完全相同。这些相似之处表明，可以将它们的名称统一为我们提出的命名系统：量子态调制，简称 QSM。如果某个特定应用明确要求使用无符号整数（QRDA），我们可以在缩写前加上小写字母“u”（uQSM）。今后还可以将这一术语扩展到其他变量类型，例如定点数（第 5.3 节）或浮点数。

## 5.2 QSM

考虑一个包含 8 个样本的数字音频文件，其振幅值采用 3 位深的二进制补码方案进行量化。假设信号的第七个样本表示为以下整数比特串：”001”。得到这个比特串后，我们通过先将专用振幅寄存器初始化为基态，再翻转最右侧的量子比特，以某种方式制备一个相应的 3 量子比特字符串。随后，这个振幅态可以与时间寄存器的相应状态纠缠，在此例中为 $t_7$，如公式 34 所述。

$$(a_7, t_7) \equiv (001, 110) \longrightarrow |001\rangle\,|110\rangle \equiv |a_7\rangle \otimes |t_7\rangle \tag{34}$$

对所有样本应用这一结构，将得到方程35和36所示的叠加态。更一般地，QSM 可以写成方程37的形式。

$$
\begin{aligned}
\overset{binary\,notation}{|A_{QSM}\rangle} = \frac{1}{\sqrt{8}}\Big[ & |000\rangle \otimes |000\rangle \ + \ |111\rangle \otimes |001\rangle \ + \ |010\rangle \otimes |010\rangle \ + \\
& |011\rangle \otimes |011\rangle \ + \ |110\rangle \otimes |100\rangle \ + \ |101\rangle \otimes |101\rangle \ + \\
& |001\rangle \otimes |110\rangle \ + \ |000\rangle \otimes |111\rangle \Big]
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
\overset{decimal\,notation}{|A_{QRDA}\rangle} = \frac{1}{\sqrt{8}}\Big[ & |0\rangle \otimes |0\rangle \ + \ |-1\rangle \otimes |1\rangle \ + \ |2\rangle \otimes |2\rangle \ + \ |3\rangle \otimes |3\rangle \ + \\
& |-2\rangle \otimes |4\rangle \ + \ |-3\rangle \otimes |5\rangle \ + \ |1\rangle \otimes |6\rangle \ + \ |0\rangle \otimes |7\rangle \Big]
\end{aligned} \tag{36}
$$

$$
|A\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |S_i\rangle \otimes |i\rangle \tag{37}
$$

时间寄存器使用 $n$ 个量子比特（音频长度），而振幅寄存器需要 $q$ 个量子比特（量子比特深度）。

如前所述，这种策略有一些可预见的优势。用于振幅的多量子比特寄存器与数字音频直接相关。这种相似性表明，可以开发一些量子算法来执行与已知数字信号处理相同的操作。在此基础上，开发者可以探索量子媒体是否能够改进、扩展或优化其中的一些操作。这一“直觉”层面促使基于状态的表示相较于探索较少的基于系数的表示迅速流行起来。

### 5.2.1 准备

这里将采用 SQPAM 情况中提到的相同策略：赋值操作。在这种情况下，操作将使用多重受控 CNOT 门来翻转振幅寄存器中的量子比特状态，而不是使用 $R_y(2\theta_i)$。因此，算法将访问特定的振幅状态，并使其与各自的时间状态纠缠。

赋值操作：对于每个时间状态，

- 步骤 1：将时间状态标签的二进制形式存储在一个经典变量中。
- 步骤 2：获取相同索引处量化音频样本的二进制形式
- 步骤 3：如果时间变量某个位置上的位为 0，则在相应的量子比特上应用 X 门。检查所有位。
- 步骤 4：如果音频样本某个位置上的位为 1，则应用多重受控 CNOT 门。检查所有位。
- 步骤 5：重复第三条指令以重置控制条件。

图 17 和图 18 展示了一个音频的制备电路，该音频包含 8 个样本，位深为 3 位

与 SQPAM（$R_y(2\theta_i)$）相比，图 17 中音频的制备电路将执行 $N$ 次赋值操作，需要 $\mathcal{O}(q + \log\ N)$ 条基本指令。总之，QSM 的制备电路速度快了指数倍，只需 $\mathcal{O}(qN \log\ N)$ 次简单操作。

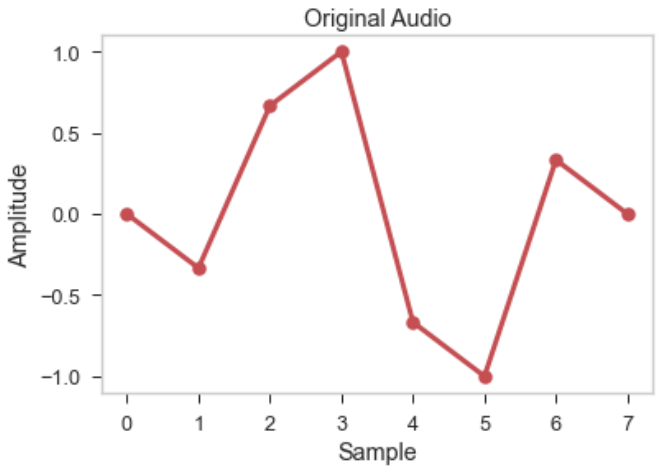


**图17：** 用于制备 QSM 电路的音频示例

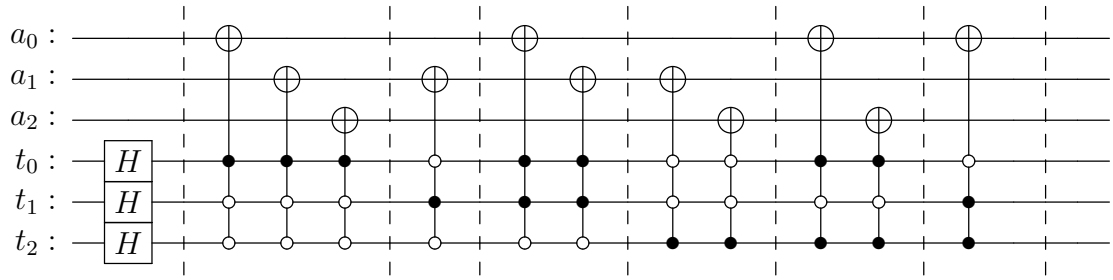


**图18：** 使用值设置操作制备 QSM 表示

### 5.2.2 检索

检索基于状态的量子音频的有趣之处在于，它可以被视为确定型。这保证了数字音频能够完美重建。这是因为样本以二进制状态表示。

在基于系数的版本中，样本是统计量，需要通过大量测量和统计分析来近似。只要测量次数有限，就始终存在固有的统计误差。这是在误差落入某个容差范围内（取决于音乐、心理声学和感知因素）与时间消耗（主要取决于电路的制备和测量次数）之间的权衡 [6]。

对于状态检索，我们只需测量所有有效的时间-振幅对即可还原音频。每个时间状态都与一个振幅态纠缠。因此，从理论上讲，这可以确保测量给定的时间状态时，必然会得到与之相关联的振幅；不存在其他可能的组合态。

既然如此，原则上我们可以对每个时间状态只测量一次，从而检索出所有相应的振幅。遗憾的是，基于处于叠加态的量子时间状态，可能无法设计出一种方法，保证每个样本只测量一次。换句话说，受控测量（仅根据另一个量子态来测量一个量子态）是非幺正指令，无法表示为一组量子门。 [7]。

[6]在一篇关于使用基于系数的图像表示来生成生成式量子图像的论文中，James Wootton [15]表示，他测量了 $n$量子比特图像态 $4^n$ 次，之后才认为该状态足以近似满足其应用需求。要实现可靠检索，[7]这个次数可能要高得多，而且会呈指数级增长。我们可以构建量子电路，利用测量结果来控制量子门（依据延迟测量原理）。但只能使用酉指令。

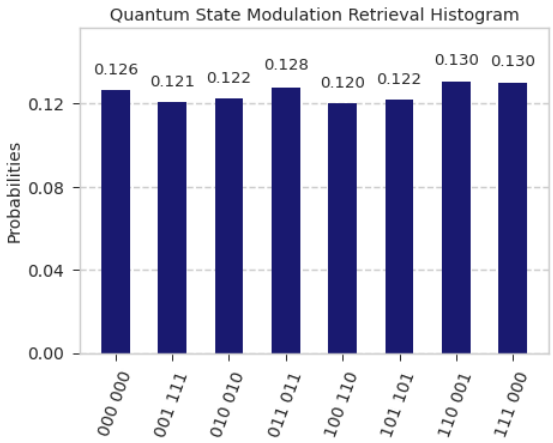


**图19：** 假设的QSM直方图

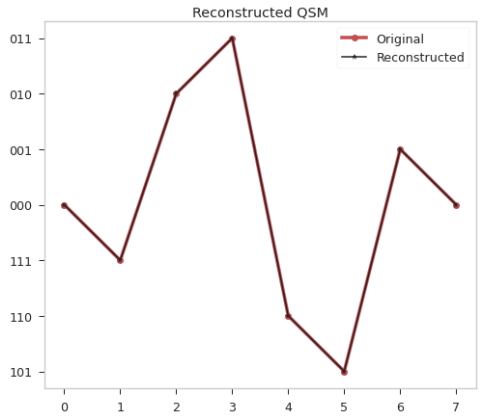


**图20：** 根据图19中的QSM重建的音频

话虽如此，检索 QSM 的策略在许多方面与之前的方法相似：制备多个副本、进行测量，并分析直方图。需要进行足够多次测量，以确保有足够高的概率测得所有可能的状态。随后分析直方图，并将有效的测得状态解析为数字振幅和时间索引。这种分析方法也是在现实中的量子处理器上应对量子错误的良好策略：进行足够多次测量，以区分噪声与信号。在理想情况下，状态调制表示会对可能的状态（包含音频振幅及其时间索引的状态）进行均衡叠加，这意味着测得每种状态的可能性都相同（图 19）。此外，振幅寄存器的测得状态会使用二进制补码方案解码为相应的数字振幅（图 20）。

QSM框架之所以强大，是因为它与数字表示具有相关性。它看起来就像处于叠加态的 PCM 音频。这不禁让人想到比较已知经典数字信号处理算法的计算效率（当然，即使量子并行性有助于降低计算复杂度，我们也需要考虑制备和检索所消耗的时间）。

## 5.3 定点 QSM

制备复杂度以及将量子态解读为整数的方式，使 QSM（FRQA）广为人知。尽管如此，这也引出了一个问题：为什么要止步于整数？我们完全可以将量子比特寄存器解读为定点数（或浮点数），并采用 QSM 的相同态结构。Li 等人正式提出了一种基于 QSM 的定点数表示 [19]，并将其命名为 QRDS（数字信号量子表示）。为便于本文术语表述，这种表示也可以称为定点 QSM 或 fpQSM。

如公式 38 所示，它本质上将用于表示整数的 $q$ 个量子比特分为三部分：一个符号位、 $m$ 个整数位和 $q-m-1$ 个小数位[8]，。

$$|A\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |fixed_i\rangle \otimes |i\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |x_i^m, x_i^{m-1} \cdots x_i^0 \,.\, x_i^{-1} \cdots x_i^{q-m-1}\rangle \otimes |i\rangle \tag{38}$$

## 5.4 多通道音频

[21] 提出的 QRMA（多通道音频量子表示）扩展了量子音频面向状态表示的逻辑。在我们的命名系统中，它称为多通道量子状态调制（MQSM）。它通过在量子电路中引入一个新的量子寄存器，进一步发展了 QSM 的实现逻辑：通道寄存器 （公式 39）。借助这个新寄存器，我们可以寻址立体声音频的左声道（$|0\rangle$）和右声道（$|1\rangle$）等（公式 40）。此外，它还能将多通道音频高效存储在同一个量子态中（或将多个长度相同的单声道音轨存储在一个声音库中）。

$$(a_{n,m}, t_{n,m}, c_m) \longrightarrow |c_j\rangle \otimes |a_{i,j}\rangle \otimes |t_{i,j}\rangle \tag{39}$$

$$|A_{Stereo}\rangle = \frac{1}{\sqrt{2N}} \begin{bmatrix} |L\rangle \otimes \big( |0\rangle|0\rangle + |0\rangle|1\rangle + |0\rangle|2\rangle + |0\rangle|3\rangle + ... \big) \\ |R\rangle \otimes \big( |0\rangle|0\rangle + |0\rangle|1\rangle + |0\rangle|2\rangle + |0\rangle|3\rangle + ... \big) \end{bmatrix} \tag{40}$$

$$|A_{MQSM}\rangle = \frac{1}{\sqrt{NC}} \sum_{i=0}^{N-1} \sum_{j=0}^{C-1} |c_j\rangle \otimes |a_{i,j}\rangle \otimes |t_{i,j}\rangle \tag{41}$$

MQSM 的一般公式如公式 41 所示，其中 $C = 2^c$ 是所用声道的数量。与 $N$ 类似，如果声道数量不是 2 的幂次方，则在制备量子态之前，会将其余可能的声道补零。

相应地，可以在基于系数的表示中添加通道寄存器。正如下一节所示，它可以成为特征提取算法的有力工具。例如，可以按照公式 42 推导出 SQPAM 的多通道版本。

$$|A_{MSQPAM}\rangle = \frac{1}{\sqrt{NC}} \sum_{i=0}^{N-1} \sum_{j=0}^{C-1} |c_j\rangle \otimes (\cos\theta_i |0\rangle + \sin\theta_i |1\rangle) \otimes |i\rangle \tag{42}$$

[8]该 $q-m-1$ 已调整为符合我们的记法。在原始记法中，文本使用的是 $n+1$ 量子比特，而不是 $q$

# 6 总结

量子系统中音频编码的不同策略源自量子图像表示。此外，还提出了量子音频表示术语的变更（未来也可将其扩展至图像及其他信号）。表1对所讨论的关键信息和特征进行了比较。

沿表格向下查看，各种表示法需要越来越多的量子比特空间来存储所有信息。因此，概率型表示能够在相同硬件中存储更大的样本。位深也是如此：由于概率振幅是连续变量，经过检索过程后，它们能够存储的幅度分辨率可达到经典计算机所能提供的水平。相比之下，基于状态的表示使用量子比特二进制字，要达到相当的分辨率则需要多得多的量子空间。此外，位深 $q$ 也会影响量子态调制的制备复杂度。

另一方面，仅考虑音频大小 $N$ 时，制备 SQPAM 态所需的指令数量，对于较大的音频样本来说，明显高于确定型表示法。

| 表示法 | 原始缩写 | 基本结构 | 振幅 Mapping | 所需空间 | 制备复杂度 | 检索 |
|---|---|---|---|---|---|---|
| QPAM | —— | $\alpha_n \lvert t_n\rangle$ | 归一化，严格正向量 | $\lceil$对数 $N\rceil$ | $\mathcal{O}(N)$ | 概率型 |
| SQPAM | —— | $(\cos\theta_n \lvert 0\rangle + \sin\theta_n \lvert 1\rangle) \lvert t_n\rangle$ | 角度向量 | $\lceil\log N\rceil + 1$ | $\mathcal{O}(N^2)$ | 概率型 |
| uQSM | QRDA | $\lvert U_n\rangle \lvert t_n\rangle$ | 无符号的整数向量 | $\lceil$对数 $N\rceil + q$ | $\mathcal{O}(qN\lceil$对数 $N\rceil)$ | 确定型 |
| QSM | FRQA | $\lvert S_n\rangle \lvert t_n\rangle$ | 有符号整数向量-tor | $\lceil$对数 $N\rceil + q$ | $\mathcal{O}(qN\lceil$对数 $N\rceil)$ | 确定型 |
| fpQSM | QRDS | $\lvert Fp_n\rangle \lvert t_n\rangle$ | 定点整数向量 | $\lceil$对数 $N\rceil + q$ | $\mathcal{O}(qN\lceil$对数 $N\rceil)$ | 确定型 |
| MQSM | QRMA | $\lvert c_i\rangle \lvert S_n\rangle \lvert t_n\rangle$ | 多个有符号整-数向量 | $\lceil$对数 $N\rceil + q + \lceil$日志 $C\rceil$ | $\mathcal{O}(CqN\lceil$日志 $N\rceil)$ | 确定型 |

表1：音频量子表示的比较

这些比较对于设想音乐品质音频在量子硬件上的近期应用至关重要。设想一位作曲家出于某种假想的艺术目的，想在量子机器中处理两个不同的音频样本。第一个样本可能来自一种实时音频信号处理语言，其中有一段包含64个样本的数据块，需要在一种全新的量子音频效果上运行。第二个样本存储在计算机上，音频时长约为1.5秒，包含 $2^{16} = 65536$ 个样本，这些样本将被输入一个大型量子声音数据库。两个样本均为CD音质：采样率为44,100 Hz，分辨率为16位。我们将比较三种表示方法：QPAM、SQPAM和QSM。

根据表1，QPAM需要6个量子比特和 $\sim$ 64条指令来存储第一个样本；SQPAM需要7个量子比特和 $\sim$ 4096 条指令。QSM则需要额外分配16个量子比特，以达到所需的幅度分辨率，总计需要22个量子比特和 $\sim$ 6144 条指令。

对于第二段音频，时间寄存器的规模会扩大。QPAM 将使用 16 个量子比特和 $\sim$ 65 千条指令；SQPAM 将使用 17 个量子比特，以及数量惊人的 $\sim$ 4.3 十亿条指令。最后，QSM 将需要 32 个量子比特和 $\sim$ 16 百万条指令。

这一比较的一个显著结果是，SQPAM 可能更适合处理小型信号样本或分段音频块，这对信息检索算法很有用。另一方面，QSM 能够有效处理大型声音。此外，QPAM 在准备阶段所需的资源最少，这对于近期应用非常有用。

## 6.1 在真实硬件上运行

本章着重梳理并构建一条平顺、易于理解的理论过渡路径，将数字音频与量子音频这一新领域联系起来，揭示计算层面的潜在问题，提出更统一的术语，并最终展望潜在的艺术应用。这也是本章着重进行高层次讨论的原因：我们假设量子比特是完美的，且能够被完美操控，这类量子比特称为逻辑量子比特。即便如此，还是有必要谈谈如何在真实硬件上运行这些算法。撰写本文时，公众可以使用的最先进计算机拥有三到十六个噪声相当大的物理量子比特。我们可以纠正潜在错误，以更好地近似逻辑量子比特。为此，可以将多个物理量子比特组合成容错量子比特。这一领域称为量子纠错 [22]。目前最先进的机器及其量子比特尚不具备容错能力。

现实中的量子比特容易受到各种错误和噪声的影响。例如，外部电磁干扰可能会导致其状态自发翻转。它也往往会损失能量，转变到较低能级的状态。此外，随着时间推移，量子系统可能会失去相干性。而且，施加到量子比特上的每个门或指令都有可能失效或不准确，从而引入更多噪声，并可能导致无法纠正的错误。此外，量子比特之间可能以不同方式相互干扰，具体取决于它们在实际量子芯片上的位置。

遗憾的是，制备量子音频表示所需的量子电路深度，对于近期的容错能力不足的机器而言可能过高（如上一节所示）。在容错能力不足的机器中，指令越多（物理量子比特越多），结果的可靠性就越低[9]。这一技术障碍促使量子音频社区转向理论和模拟，而非在最先进技术的真实硬件上进行实验。

作为未来分析和比较的参考，本文强调了所介绍的部分编码方案（QPAM、SQPAM、QSM）在最先进硬件上的表现。实验使用了 IBM Q 硬件 [23]。具体而言，使用了一块名为 *ibmq casablanca* 的 7 量子比特（量子体积为 32）芯片。这些实验（以及以下各节中的结果和模拟）是在 QuTune 项目 [25] 范围内，使用以 Python 和 Qiskit 实现的量子音频模块进行的。QuTune 模块用于将信号映射并生成量子音频电路（见图 21）。随后，将量子电路输入 IBMQ 系统。最后，取回并进行后处理的结果直方图见图 22、24 和 26。重建音频信号见图 23、25 和 27。

QPAM（图 23、22）具有最低的制备复杂度（约 8 条指令和 3 个量子比特）。它似乎保留了原始信息的大部分。

另一方面，SQPAM (24, 25) 引入了赋值操作，成本更高。检索到的音频损失了大部分振幅范围（请注意 y 轴上的数值），但保留了部分原始轮廓。

[9]虽然这一局限可能确实存在，但从艺术创作的角度看，在合理的噪声／劣化美学框架下，它也可以被视为一种优势。

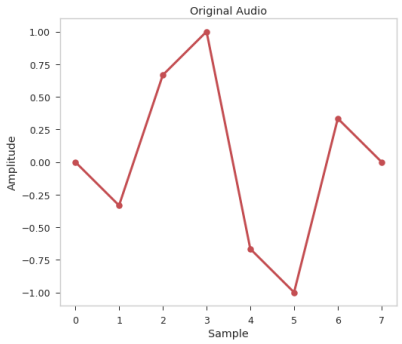


**图21：** 含有8个样本的测试音频。

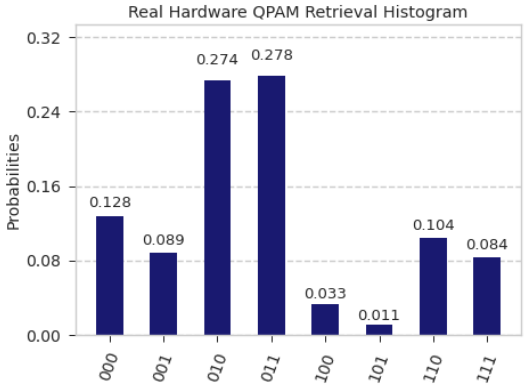


**图22：** 生成的 QPAM 直方图。

相比之下，QSM（图 26）产生了一个结果。令人惊讶的是，它与原始文件的相似性消失了。这可能是由制备过程中大量的多重受控指令以及操作所需的时间造成的。此外，一些量子比特可能已经翻转到较低的能量态。尽管它能够保证精确检索，但重建过程依赖于最有可能被测量到的状态。在这一信息丢失阶段，QSM 音频已无法恢复 (27)[10]。

# 7 总结

本章讨论了迄今提出的各种音频量子表示方法。量子音频仍是一个非常年轻的研究领域。自然，本文介绍的不同方法各有优缺点。在现阶段，我们无法预见本文讨论的方法中是否有任何一种最终会被工业界采用，因为实现这些方法所需的硬件尚不可用。

尽管如此，如上所示，我们已经可以着手开发各种方法，并基于模拟以及对量子处理器未来发展的合理推测，畅想有趣的应用（例如量子电吉他效果踏板）。量子计算有一个重要的注意事项需要解决，那就是存储器问题。从本质上说，量子信息无法像经典处理器处理数字信息那样进行存储和读取。量子随机存取存储器（qRAM）技术的发展将对量子音频产生重大影响。目前已有多种关于开发 qRAM 的方案（例如，[26, 27]），但通用型实用设备仍未见踪影。

[10]不过，在 QSM直方图（26）中，量子比特序列上略微显现出了周期性。这能否用于艺术创作?

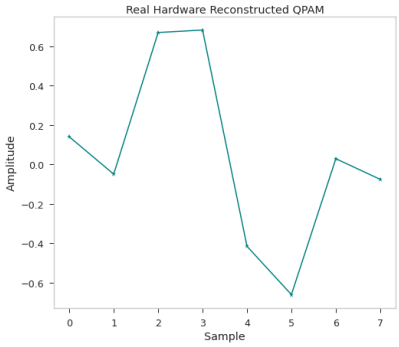


**图23：** 图22中的重建音频。

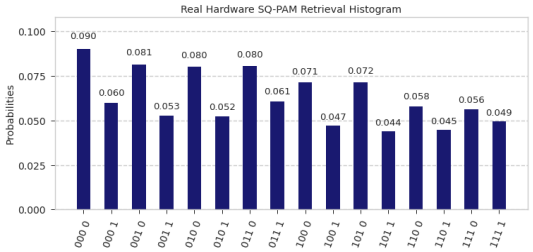


**图24：** 提取的 SQPAM 直方图。

quantumaudio：一个用于在 Qiskit 中实现音频量子表示的 Python 软件包，现已推出：

https://pypi.org/project/quantumaudio/

quantumaudio 由英国普利茅斯大学计算机音乐研究跨学科中心（ICCMR）的量子计算机音乐团队开发和维护。

# 8 致谢

作者感谢普利茅斯大学和 QuTune 项目的支持。该项目由英国计算与模拟量子技术中心资助。

# 参考文献

[1] Rabiner, Lawrence R., and Bernard Gold.数字信号处理理论与应用。恩格尔伍德克利夫斯：Prentice-Hall，1975年。[2] Broughton. S. Allen & Bryan, Kurt. 离散傅里叶分析与小波：信号与图像处理的应用。John Wiley & Sons, Ltd，2008年。[3] Ziemer, Rodger E., and William H. Tranter.通信原理，第*3–4*章。John Wiley & Sons，2014年。[4] K Meichanetzidis, S Gogioso, G De Felice, N Chiappori, A Toumi and B Co- ecke. 近期量子计算机上的量子自然语言处理，2021年。arXiv:2005.04147 [cs.CL]

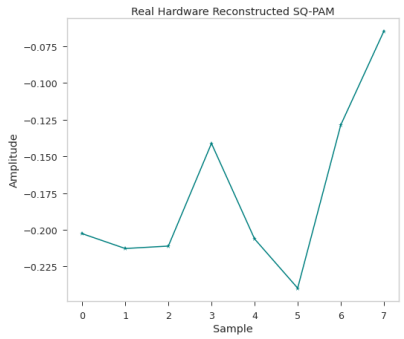


**图25：**来自图24的重建音频。

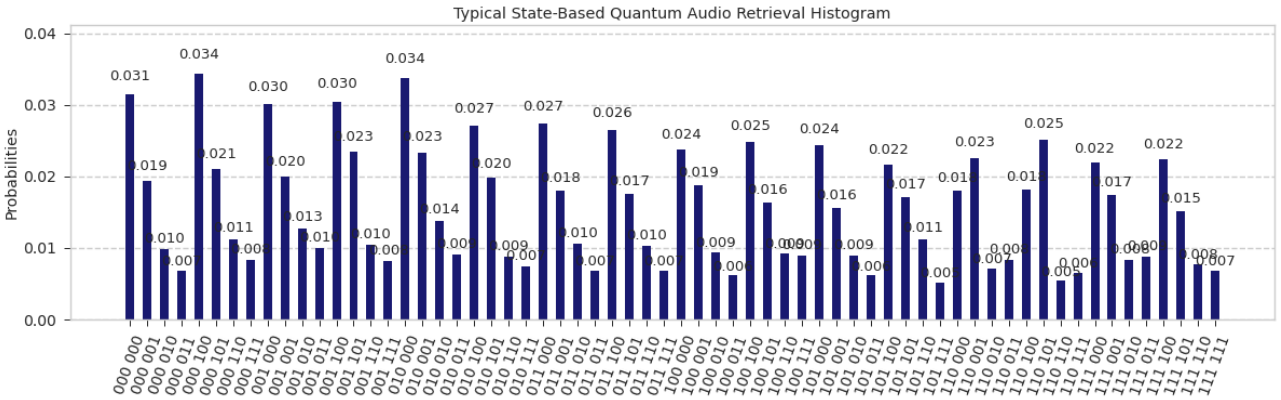


**图26：**检索到的 QSM 直方图。

[5]Yi Zhang, Kai Lu, Yinghui Gao, and Mo Wang. *NEQR*：一种新颖的增强型数字图像量子表示。《量子信息处理》，12(8)：2833–2860，2013年。[6] Wang, Jian. *QRDA*：数字音频的量子表示。《国际理论物理学杂志》，55(3)：1622–1641，2016年。[7] Fei Yan, Abdullah M Iliyasu, Yiming Guo, and Huamin Yang. 量子计算机上音频信号的灵活表示与操作。《理论计算机科学》，752：71–85，2018年。[8] Vlatko Vedral, Adriano Barenco, and Artur Ekert.用于基本算术运算的量子网络。《物理评论A》，54(1)：147，1996年。[9] Fei Yan, Abdullah M Iliyasu, and SalvadorE Venegas-Andraca. 量子图像表示综述。《量子信息处理》，15(1)：1–35，2016年。[10] D Wang, Z-H Liu, W-N Zhu, and SZ Li. 基于多目标扩展广义托福利门的量子比较器设计。《计算机科学》，39(9)：302–306，2012年。[11] Kehan Chen, Fei Yan, Abdullah M Iliyasu, and Jianping Zhao. 探索在量子音频信号上实现隐写协议。《国际理论物理学杂志》，57(2)：476–494，2018年。[12] Kehan Chen, Fei Yan, Abdullah M Iliyasu, and Jianping Zhao. 一种量子音频水印方案。载于2018年第37届中国控制会议（CCC），第3180–3185页。IEEE，2018年。[13] Kehan Chen, Fei Yan, Abdullah M Iliyasu, and Jianping Zhao.基于量子离散余弦变换的双量子音频水印方案。《国际理论物理学杂志》，58(2)：502–521，2019年。

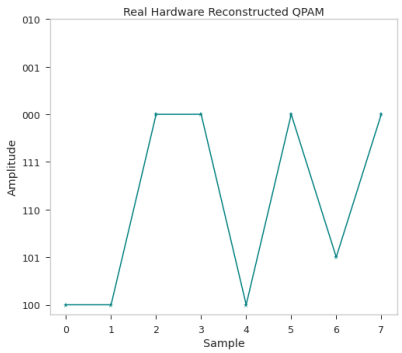


**图27：**根据图26进行的重建尝试。

[14] Michael A Nielsen 和 Isaac Chuang。《量子计算与量子信息》，第*216–221*页。美国物理教师协会，2002年。[15] James R Wootton。使用量子计算进行程序化生成。发表于数字游戏基础国际会议，第1–8页，2020年。[16] Salvador El´ıas Venegas-Andraca。离散量子漫步与量子图像处理。2005年。[17] Salvador E Venegas-Andraca 和 Sougato Bose。使用量子力学存储、处理和检索图像。发表于《量子信息与计算》，第5105卷，第137–147页。国际光学与光子学学会，2003年。[18] S¸ahin，Engin。基于有符号整数的量子傅里叶变换的量子算术运算。《国际量子信息杂志》18(06):2050035，2020年。[19] Li, P., Wang, B., Xiao, H. 等。数字信号的量子表示与基本操作。《国际理论物理学杂志》，57:3242–3270，2018年。[20] Yan, Fei & Iliyasu, Abdullah & Le, Phuc & Sun, Bo & Dong, Fangyan & Hirota, Kaoru。量子计算机上多对图像的并行比较。《国际创新计算与应用杂志》。5. 199-212，2013年。[21] S¸ahin，Engin，和 Ihsan Yilmaz。*QRMA*：多通道音频的量子表示。《量子信息处理》18.7：1-30，2019年。[22] Lidar, Daniel A. 和 Todd A. Brun，编。量子纠错。剑桥大学出版社，2013年。[23] *IBM Quantum*。https://quantum-computing.ibm.com/，最后访问日期：2022年2月1日。[24] Miranda, E. R.。计算机声音设计：合成技术与编程。Focal Press；第2版，2002年。[25]*QuTune* 项目——量子计算机音乐资源。https://iccmr-quantum.github.io/，最后访问日期：2022年2月1日。[26] Chen, K. C., Dai, W., Errando-Herranz, C., Lloyd, S. 和 Englund, D。自旋-光子网络中的可扩展高保真量子随机存取存储器。PRX Quantum，2，030319，2021。[27] Asaka, R., Sakai, K. 和 Yahagi, R.。通过量子漫步实现量子随机存取存储器。《量子科学与技术》，6(3)：035004，2021。