第 5 章

CHAPTER 5

音频处理系统

在过去的几十年里，数字音频处理的技术、系统和设备已经发生了重大的改变。发展到现在，基于软件的系统几乎将硬件系统推离出了音频处理链的过程。现在大多数处理过程可以在通用计算机或平板电脑上完成，而各种麦克风、扬声器与计算机的接口仍然需要专用的硬件解决方案。近来，数字信号处理技术改进了麦克风和扬声器，使得低功率的设备可以通过专用硬件，例如数字信号处理器(DSP)、可编程门阵列(FPGA)来实现。而大型调音台仍然受益于使用 FPGA 进行接口和内部多信号处理内核的硬件集成方法。音频网络将向着更高集成度和更小体积硬件设备方向发展。

具有扩展接口的新型 DSP 在计算机庞大且繁多的不同应用领域中得到应用。它们为模数转换器(ADC)/数模转换器 (DAC)、局域网 (LAN)和无线局域网(WLAN)提供了多种接口，使其更方便用于具有低功耗和低延迟的音频设计。数字信号处理器设备及应用如表 5.1 所示。基于不同操作系统的计算机技术的进步，出现了新的基于软件的音频系统。利用中央处理单元(CPU)和图形处理单元(GPU)的音频处理显著地加速实现了音频计算和录制。从手机到平板电脑到通用计算机都需要使用专门的音频处理任务单元，当然，软件程序员的熟练技能仍然是很重要的。

表 5.1　DSP(AD/TI/NXP/STM/Cadence/Xilinx)设备及应用

LAN/WLAN 片上系统	浮点 DSP	定点 DSP	现场可编程门阵列
树莓派	PC/CPU/GPU	Sigmastudio	Xilinx
流媒体	DAWs	耳机	麦克风
网络	音响合成器	扬声器	混音台
无线扬声器	录音	合成器效果	
系统	混频		
	母版制作		
块延迟	块延迟	1～2 样本	0 样本

5.1　数字信号处理器

5.1.1　定点 DSP

AD 转换器的离散时间和离散幅度输出通常以 2 的补码格式表示，这些数字序列的处理是通过定点或浮点运算进行的。处理后信号的输出再次采用 2 的补码格式，并馈送到 DA 转换器。有符号小数表示(2 的补码)是定点数表示算法的常用方法。为了处理生成和模运算，要使用无符号整数。图 5.1 给出了一个典型的定点 DSP 的示意图。主要构建模块是程序控制器、带有乘法累加器(MAC)的算术

逻辑单元(ALU)、程序和数据存储器以及与外部存储器和外围设备的接口。所有模块通过内部总线系统相互连接，内部总线系统具有独立的指令和数据总线。数据总线本身可以由多个并行总线组成，例如，可以将乘法指令的两个操作数并行传输到 MAC。内部存储器由指令、数据随机存取存储器(RAM)和附加的只读存储器(ROM)组成。该内部存储器允许快速执行内部指令和数据传输。为了增加存储空间，地址/控制和数据总线连接到外部存储器，如可擦除可编程(EP)ROM、ROM 和 RAM。外部总线系统与内部总线体系结构的连接对外部指令的高效执行以及对外部数据的处理有很大的影响。为了连接串行操作的 AD/DA 转换器，多个 DSP 都提供了具有高传输速率的特殊串行接口。此外，一些处理器支持直接连接到 RS232 接口。微处理器的控制可以通过 8 位字长的主机接口来实现。

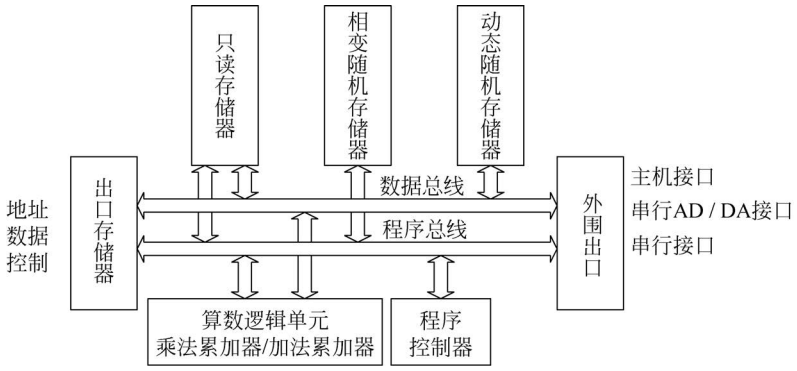

图 5.1　定点 DSP 原理图

定点 DSP 在字长和周期时间方面的介绍可以从相应制造商的网站上得到。如果量化影响了所应用算法的稳定性和数值精度，算法的精度基本上可以提高一倍。通过与组合乘法和累加命令的处理时间(以处理器周期为单位)相关的周期时间可以了解特定处理器类型的计算能力。周期时间直接由最大时钟频率产生。指令处理时间主要取决于内部指令和数据结构，以及处理器的外部存储器连接。在复杂算法和海量数据处理中，对外部指令和数据存储器的快速访问具有特殊的意义。串行数据连接与 AD/DA 转换器的连接通过特殊的主机接口进行控制，因此可以避免复杂的接口电路。对于独立的解决方案，也可以从简单的外部 EPROM 加载程序。

信号处理算法需要以下软件命令。

(1) MAC(乘法和累加)——组合乘法和加法命令。

(2) 将两个用于乘法的操作数同时传送到 MAC(并行移动)。

(3) 位反寻址[用于快速傅里叶变换(FFT)]。

(4) 模寻址(用于加窗和过滤)。

不同的信号处理器实现 FFT 有不同的处理时间，具有改进架构的最新信号处理器具有更短的处理时间。对于不同的处理器，组合乘法和累加命令(应用如加窗、滤波)的指令周期大致相等，但必须考虑外部操作数的处理周期。

5.1.2　浮点 DSP

典型浮点 DSP 的框图如图 5.2 所示，不同架构的主要特点是双端口原理(德州仪器)和外部哈佛架构(模拟设备)。浮点 DSP 内部有多个总线系统来加速数据传输到处理单元，片上直接存储器(DMA)访问控制器和缓存存储器支持更高的数据传输速率。

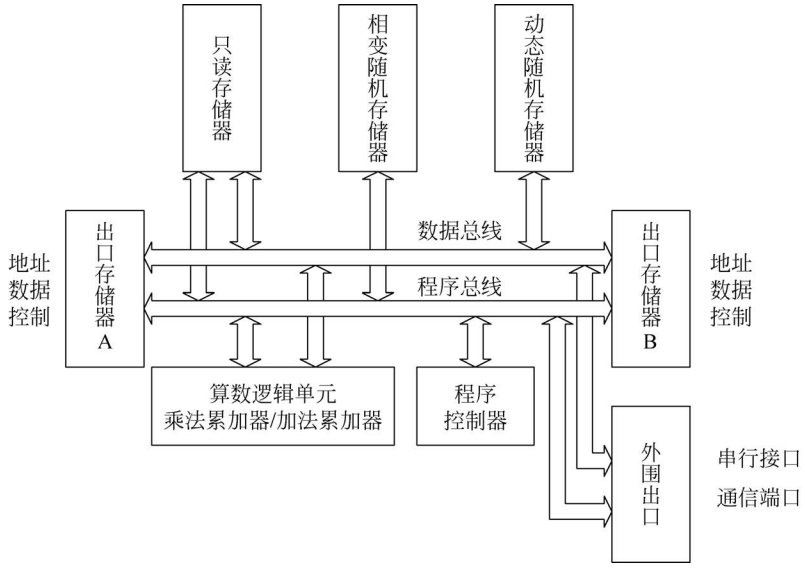

图 5.2　典型浮点 DSP 的框图

5.2　数字音频接口

为传输数字音频信号，AES(音频工程学会)和 EBU(欧洲广播联盟)已经建立了两个传输标准。这些标准用于双通道传输和多达 64 个音频的多通道传输，称为多通道音频数字接口(MADI)。

5.2.1　双通道 AES/EBU 接口

双通道 AES/EBU(也称 AES3)接口定义了专业和消费者模式。两种模式的外框是相同的，如图 5.3 所示。对于一个采样周期，定义一帧，它由两个子帧组成，分别用于前导码为 X 的通道 1 和前导码为 Y 的通道 2。总共 192 帧形成一个块。块开始的特征是一个特殊的前序 Z。子帧的位分配由 32 位组成，如图 5.4 所示，前导码由 4 位(0～3 位)和多达 24 位(4～27 位)的音频数据组成，子帧的最后四位表示有效性(数据字或错误的标志，V)、用户状态(可用位，U)、信道状态(192bits/块＝24 字节编码的信道状态信息，C)和奇偶校验(偶数奇偶校验，P)。串行数据位的传输是用双相位标记码进行的。这是通过时钟(双比特率)和串行数据位之间的异或关系来完成的(见图 5.5)。在接收端，通过检测前导码(X＝11100010，Y＝11100100，Z＝11101000)是否违反编码规则实现时钟检索(图 5.6)。信道状态信息的

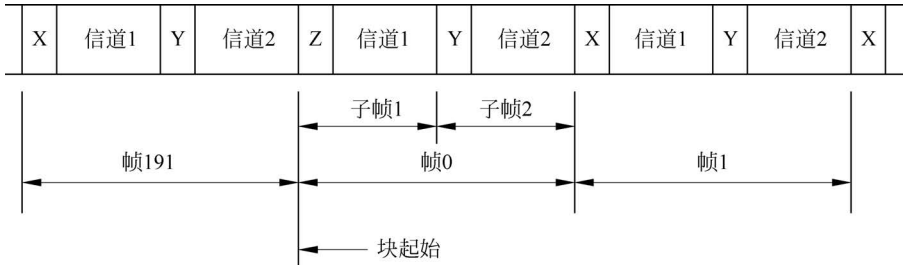

图 5.3　双通道格式

24字节的含义总结在表5.2中。该信道状态信息的前三个重要字节的精确位分配如图5.7所示。在字节0的各个字段中，除了专业/消费者模式和数据/音频特性，还指定了预加重和采样率（参见表5.3和表5.4）。字节1确定信道模式（表5.5）。消费者格式（通常被标记为SPDIF，索尼/飞利浦数字接口格式）在信道状态信息的定义和输入/输出的技术规格方面不同于专业格式。信道信息前四位的位分配如图5.8所示。对于消费级应用，使用带有RCA连接器的同轴电缆或带有TOSLINK连接器的光纤电缆。对于专业用途，使用带XLR连接器和对称输入/输出（专业格式）的屏蔽两线引线。专业接口电气规范如表5.6所示。

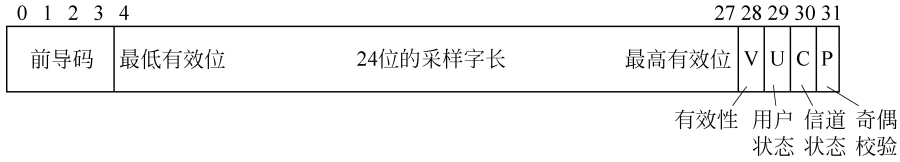

图5.4 双信道格式（子帧）

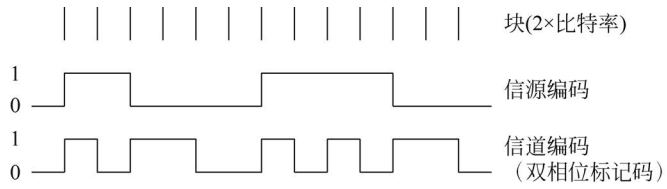

图5.5 信道编码

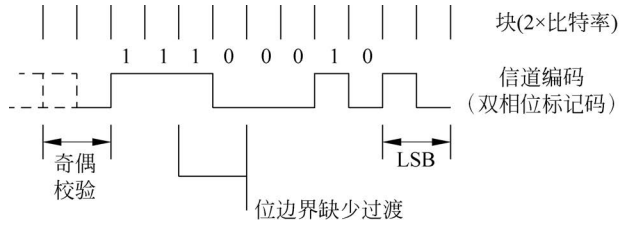

图5.6 前导码X

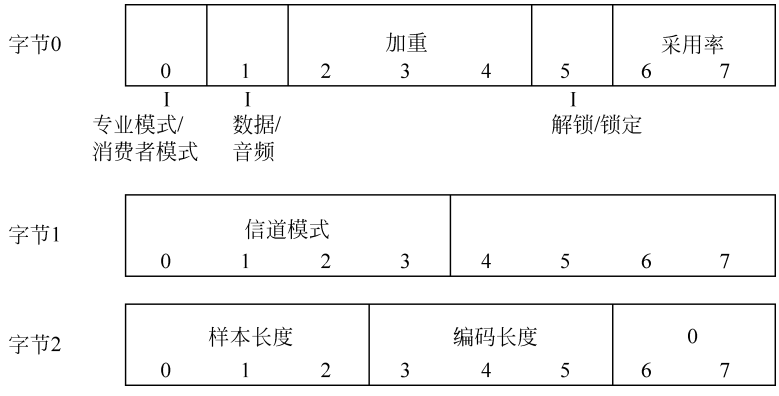

图5.7 字节0～2表示的信道状态信息

表 5.2　信道状态字节

字　节	说　明	字　节	说　明
0	加重,采样率	6～9	4 字节 ASCII 源
1	信道选用	10～13	4 字节的 ASCII 目的地
2	样本长度	14～17	4 字节的本地地址
3	字节 1 的向量	18～21	时间码
4	参考位	22	标志
5	保留	23	循环冗余校验码

表 5.3　加重

字　节	说　明	字　节	说　明
0	无指示,启用覆盖	6	50/15 μs 加重
4	无指示,覆盖禁用	7	CCITT J.17 加重

表 5.4　采用率

字　节	说　明	字　节	说　明
0	无指示(默认 48kHz)	2	44.1kHz
1	48kHz	3	32kHz

表 5.5　信道模式

字　节	说　明	字　节	说　明
0	无指示(默认 2 通道)	3	初级/中级(A=初级,B=中级)
1	双通道	4	立体声(A=左,B=右)
2	单声道	7	指向字节 3 的向量

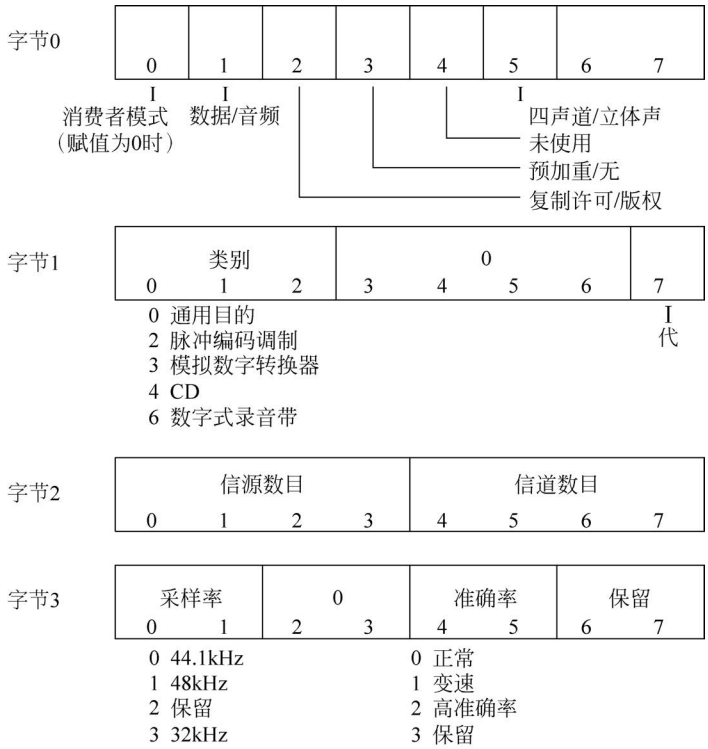

图 5.8　字节 0～3(消费者模式)

表 5.6 电气专业接口规范

输出阻抗	信号幅度	抖　动
110Ω	2～7V	最大 20ns
输入阻抗	信号幅度	连　接
110Ω	最小 200mV	平衡

5.2.2 多通道音频数字接口(MADI)

为了连接不同位置的音频处理系统，可使用多通道音频数字接口(MADI)，如图 5.9 所示。模拟/数字 I/O 系统由 AD/DA 转换器、AES/EBU 接口(AES)和抽样速率转换器(SRC)组成，通过双向 MADI 链路连接到数字分发系统。实际的音频信号处理是在特殊的 DSP 系统中进行的，这些系统通过 MADI 链路连接到数字分发系统。MADI 格式被标准化为 AES10，由双通道 AES/EBU 格式派生，允许在一个采样周期内传输 64 个数字单声道(见图 5.10)。MADI 帧由 64 个 AES/EBU 子帧组成。每个通道都有一个包含图 5.10 所示信息的前导信号。位 0 负责标识第一个 MADI 通道(MADI 信道 0)。表 5.7 显示了采样率和相应的数据传输率。当采样率为 48kHz 时，要求最大数据速率为 98.304Mb/s。通过 75Ω 同轴电缆 BNC 连接器(50m，信号振幅 0.3～0.6V)或 SC 连接器光纤线(长达 2km)传输。

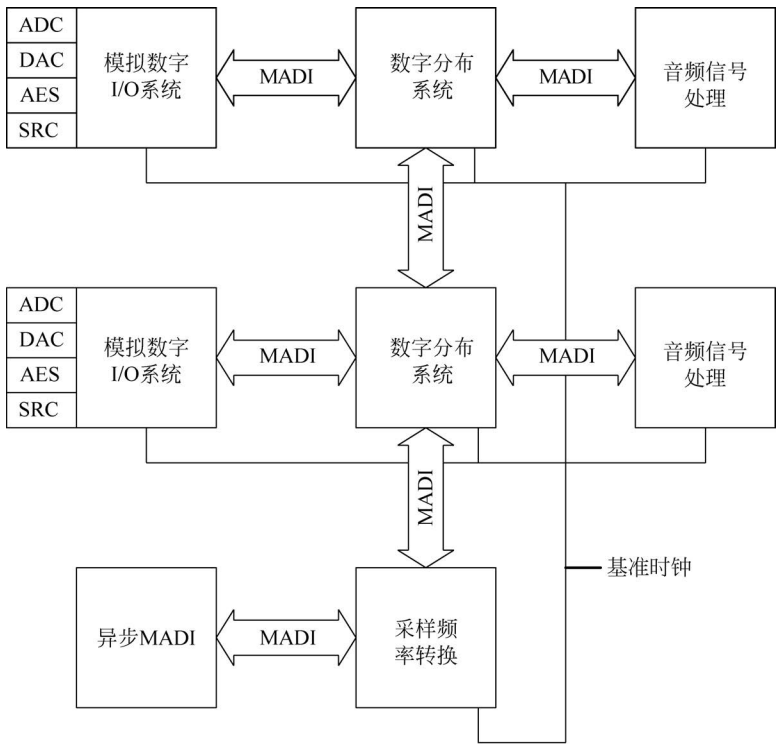

图 5.9 多通道音频数字接口(MADI)的系统连接

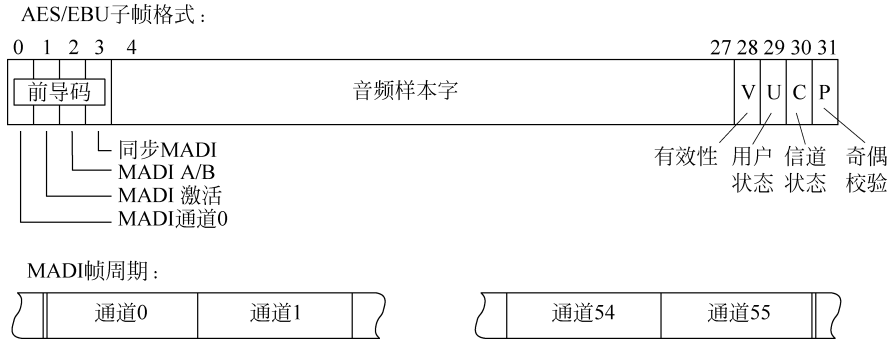

图 5.10　多通道音频数字接口(MADI)的帧格式

表 5.7　MADI 规格

采样率	32～48kHz±12.5%	数据传输率	100Mb/s
传输速率	125Mb/s	最大数据传输速率	98.304Mbit/s(64 channels at 48kHz)

5.2.3　HDMI 中的音频

高清多媒体接口(HDMI)是一种专有的音频/视频接口标准。数字音频数据与视频和辅助数据一起发送到 TMDS(最小化差分信号)数据通道，如图 5.11 所示。图 5.12 显示了通道的物理引脚，可以看到每个 TMDS 通道分别被屏蔽，并作为差分信号在双绞线上传输。音频数据可以包含任意压缩的、非压缩的、脉冲编码调制(PCM)，可以为单通道或多通道格式，最多 8 通道(自 HDMI 2.0 以来，达到 32 通道)，并在数据岛周期内封装的视频消隐间隔期间传输。为了确保发送端只发送音频或视频数据，接收端有能力接收，在连接建立过程中，显示数据通道(DDC)中有一个基于 I^2C 的信息交换。其标准称为扩展显示识别数据(EDID)，由视频电子标准协会(VESA)发布。它包含制造商 ID 和型号名称以及支持的视频格式的信息，如分辨率、纵横比、位深度和刷新率。EDID 的音频描述符包含音频格式、通道数量、采样率和位深度等信息。

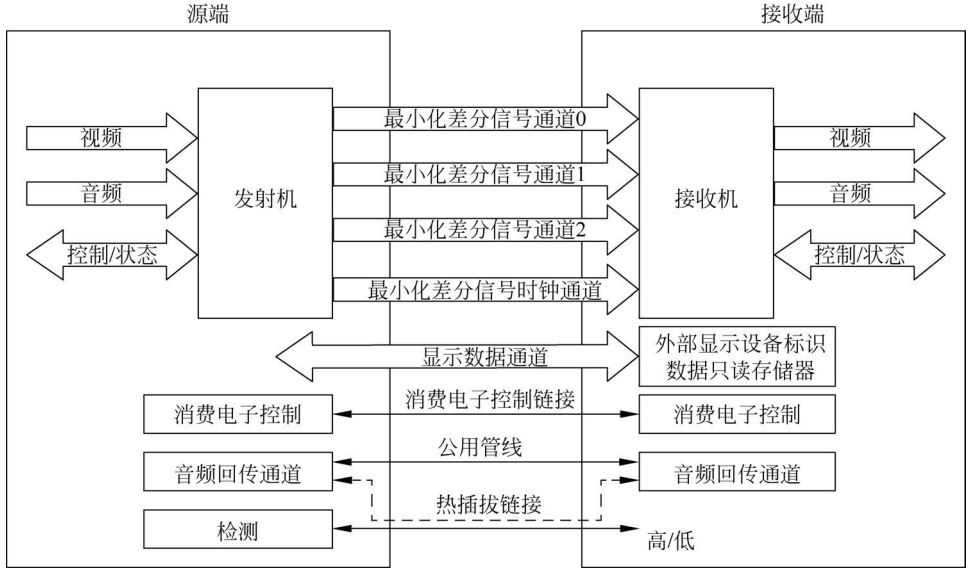

图 5.11　高清视频信息的标准格式框图

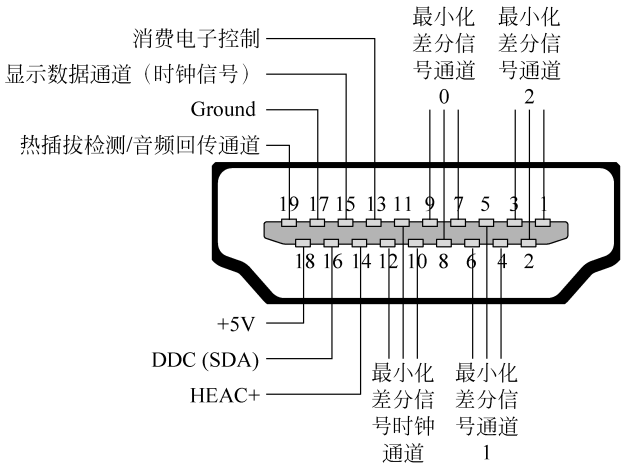

图 5.12　HDMI 1.4 中 HDMI 引脚排列

除了从发射机到接收机的音频传输，HDMI 还指定了从接收机到发射机的音频传输，称为音频返回通道（ARC），这是 HDMI 1.4a 的一个功能，也是自 HDMI 2.1 以来的一个更新版本，称为 eARC（增强 ARC）。例如，电视是信号的来源，把一个 AV 接收器连接到电视作为显示输出，那么接收电视节目时，电视节目的音频可以通过同一 HDMI 线发送回 AV 接收器。音频回传通道（ARC）和 HDMI 以太网音频返回通道（HEAC）通过以太网连接传输，HEAC 是 HDMI 一个单独的电缆。HEAC 与 HPD（热插拔检测）共用引脚，如图 5.12 所示。图 5.11 中消费电子控制（CEC），它是一种双向连接，可以仅通过遥控器来控制所有连接的 HDMI 设备。

目前，通过 HDMI 传输的最大数据速率是 HDMI 2.1 标准 2 定义的 42.6Gbit/s。

5.2.4　计算机音频接口

有几种计算机音频接口，最常见的是：通用串行总线（USB）（音频数据格式设备类定义）、雷电、蓝牙音频。

特别地，USB 和 Thunderbolt 允许更高的数据速率，数据速率见表 5.8，因此输入和输出信号就需要更多的音频通道数。对比来看，一个普通抽样速率为 48kHz @24 位量化比特对应 1.152Mb/s 比特率。而专业的音频录音在 192kHz@24 位的设置参数下，在纯数据的情况下也只需要 4.608Mb/s。音频数据将以 PCM 值的形式被无压缩的、分成块后进行传输。

表 5.8　USB 和 Thunderbolt 的数据速率

类　型	比 特 率	类　型	比 特 率
USB 2.0	480Mbit/s	Thunderbolt 1.0	10Gbit/s
USB 3.0	5Gbit/s	Thunderbolt 2.0	20Gbit/s
USB 3.1	10Gbit/s	Thunderbolt 3.0	40Gbit/s

蓝牙所提供的无线通信在消费市场被广泛应用，而 USB 和 Thunderbolt 在专业设备方面则处于领先地位。两者在外部音频设备和计算机之间都通过特定的接口电缆提供短距离的连接。对于更远的距离，网络接口将成为主要解决方案。

在软件方面，需要一个设备驱动程序，它创建一个软件接口，并使操作系统（OS）和其他程序可以访

问硬件。对于音频设备，通常该驱动程序与 OS 音频引擎进行对话。然后，音频引擎为应用程序提供标准的 API。表 5.9 中列出了最常见的几种 API。

表 5.9　最常见的几种 API

操作系统	音频引擎
Linux	ALSA（高级 Linux 声音架构）
	OSS(开放声音系统)(过时)
Windows	ASIO(音频流输入/输出)
	WASAPI（Windows 音频会话 API)连接到 WDM(Windows 音频驱动程序)
macOS and iOS	核心音频

5.2.5　音频网络接口

音频网络接口的主要应用可以分为两种。第一种描述基于网络的音频协议，这些协议主要用于局域网中具有专用硬件接口的音频传输。它们通常用于现场活动或录音场景，并且越来越多地取代模拟音频传输。由于低延迟在这种情况下是非常重要的，局域网上的数据速率又不小，因此音频大多以 PCM 值和未压缩的包形式发送。第二种描述更多基于互联网的音频协议，这些协议通常只基于软件，用于通信。它是互联网语音协议（VoIP）的通用术语，通常会与会议软件中的视频结合使用。然而，几乎所有的现场音频传输，如音乐流媒体或在线广播，都是以类似的方式完成的。由于互联网带宽非常昂贵，通常音频会通过编码算法进行压缩。

1. 基于网络的音频

有大量的基于网络的音频协议彼此不兼容。有些是开放标准，有些是专有标准，但它们都共享相同的物理连接——以太网电缆。这些电缆大多由铜制成，但光纤连接也是可能的。协议之间的根本区别在于它们所使用的网络层。这也定义了网络的拓扑结构、与现有以太网的互操作性，并可能影响可靠性和延迟。协议分为第 1 层、第 2 层和第 3 层协议。该分类基于 ISO OSI 模型。

1）第 1 层网络协议

该协议只使用以太网电缆的物理连接，定义电缆引脚分配，而不使用以太网帧结构。因此，只使用支持该协议的特殊硬件之间的点对点连接（第一层网络协议使用硬件之间的点对点连接，但是这个硬件必须是支持该协议的特殊硬件）。这意味着每个连接都需要额外的导线。因为有一个专用网络，所以可以在保证延迟的情况下创建一个可靠的连接。以 AES50 为例，图 5.13 给出了现场演出录制的典型网络，它主要由 the Music Tribe（Midas、Klark Teknik、Behringer)使用。

AES50 使用高达 100m(100Mbit/s)的 CAT5 铜缆，有 48 个输入/48 个输出通道@48kHz/24 位，24 输入/24 输出通道通道@96kHz/24 位，5Mbit/s 辅助数据连接，用于 ADC 的远程增益控制，延迟 $<100\mu s$，使用 AES/EBU 流（增强 AES10/MADI)，用于音频和同步连接的单独信号对。

2）第 2 层网络协议

音频数据被封装在以太网包中，然后可以在以太网网上传输。然而，由于这些包低于互联网协议（IP）级别，因此需要特殊的以太网集线器和交换机来同时处理音频协议和 IP 数据包。以基于第 2 层的最重要的协议 AVB 为例，图 5.14 给出了直播录制的典型网络。

（1）关键方面：定时和同步、流预留协议，可靠性和延迟有保障。

（2）需要特殊硬件来创建网络（AVB 兼容交换机）。

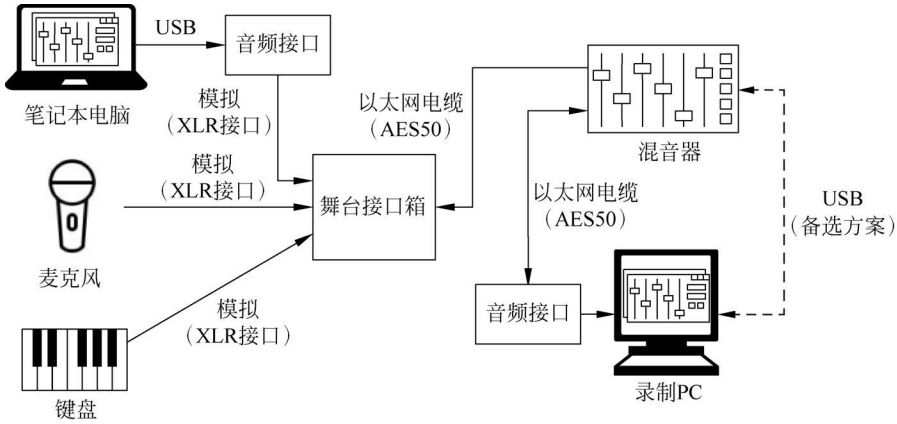

图 5.13　第 1 层网络（AES50 在这种实际场景中取代了传统的多芯线）

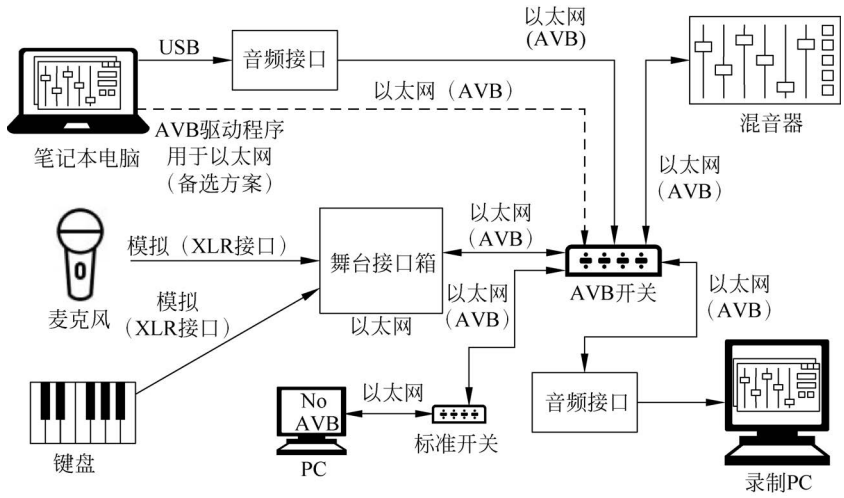

图 5.14　第 2 层网络（PC 无法接收 AVB 流，因为它是通过不支持 AVB 的标准路由器连接的）

（3）如果硬件支持，则可以与同一链路上的 IP 连接共存。

（4）支持高达 196kHz/32 位的音频。

（5）多个信道封装在 AVB 流中。一个流最多可包含 64 个频道。对于两个设备之间的多个流，通道数仅受到硬件限制和链路带宽的限制。

（6）典型接口：128 个输入/128 个输出通道@48kHz/24 位（利用率 14%，1000M 以太网）。

（7）100Mb/s、1Gb/s 或光纤连接。

（8）通过精确时间协议（PTP）实现纳秒精度的同步。

（9）网络中，包含一个必须配置的主时钟，其他设备为从时钟。

（10）网络包含扬声器、监听器和 AVB 功能的交换机。

（11）用于在整个网络路径中预留带宽的多流预留协议（MSRP）。

（12）带宽在传输和分块之前就被保留，因此带宽得到保证。

（13）默认情况下，延迟限制为 2ms，相当于网络中的 7 跳。如果数据包较早到达接收器，则播放将

被延迟。优点是延迟是恒定的。

(14) AVB 在 IEEE 802.1(开放标准,无许可费)中定义。

(15) 远程增益控制可以通过 IP 链路完成,但标准中没有定义,因此供应商之间可能不兼容。

3) 第 3 层网络协议

与其他协议标准不同,第三层协议使用标准和低价的硬件,可以很容易地集成到现有的局域网中。需要注意的是,该协议依赖于只提供尽力交付的 IP 协议,在设计上是不可靠的。为了弥补这一点,需要添加更大的缓冲区,这将导致更高的延迟,如果网络因其他协议而过载,可能会丢失一些包,这将导致音频故障。然而,延迟仍然很小,足以在实时应用程序中使用。图 5.15 显示了这样一个带有 PC 记录站的实时应用程序。下面给出第三层协议的三种不同示例。

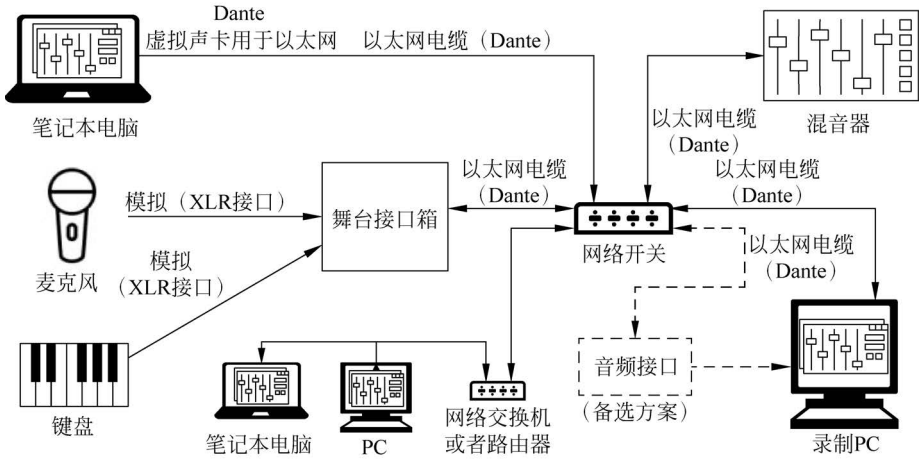

图 5.15　第 3 层网络(网络中的所有设备都可以连接到任何 Dante 设备,而不依赖于使用的网络硬件)

(1) Dante。

① 使用 UDP(用户数据报协议)数据包。

② 支持高达 196kHz/32 位的音频。

③ 使用 Dante Controller 软件可以轻松设置。

④ PTP 用于时钟同步。

⑤ 可以进行单播和多播连接。

⑥ 能够发送和接收 AES67 流的特定 Dante 设备。

(2) Ravenna。

① 基于 RTP(实时传输协议)。

② 使用 16 位或 24 位(L16/L24)的线性 PCM。

③ 使用 PTP 进行同步。

④ 使用 DiffServ。

⑤ Ravenna 与 AES67 兼容。

(3) AES67。

① 设计用于创建各种 3 层网络协议之间的互操作性,如 RAVENNA、Dante 等。

② 支持 44.1kHz、48kHz 和 96kHz 的采样率。

③ 使用 16 位或 24 位(L16/L24)线性 PCM。

在介绍了各种类型的网络音频协议之后，必须指出的是，所有这些协议都只指定了音频的传递。ADC/DAC 增益控制等控制消息不是标准协议的一部分。为了控制设备，最常用的是专有应用程序或软件。它们很可能是基于 TCP/IP 的，并且在不同供应商之间不兼容。当然还有其他标准来实现互操作性，如 OSC、MIDI 或 OCA（开放控制体系结构/AES70），但通常它们需要某种配置来通过软件控制内部的一切，并且通常不提供完整的功能集。

2. 基于互联网的音频

在本地网络中，数据通常以未压缩的 PCM 发送，而在互联网通信中，大多数音频数据都需要经过编码。编码和解码降低了每个流所需的数据速率，但需要花费时间（<40ms），且计算成本很高。在本地网络上，数据速率通常不是限制因素，因此并不通过降低数据速率来减少延迟。在互联网上的音频传输，低的数据速率是有益的，延迟不是那么关键。有专门设计用于使用 VoIP 和电信软件传输的特殊编解码器。

有很多 VoIP 和电信软件可以使用。基于计算机的 IP 音频的基本结构如图 5.16 所示，图 5.16 显示了从模拟音频源（如麦克风）到扬声器的单向传输示例。经过 ADC 转换器，音频由操作系统音频引擎处理后发送到 VoIP 应用程序。在应用中，根据使用情况，可以执行许多音频处理，例如噪声消除，并且可以应用诸如 EQ 或噪声门之类的滤波器。之后，音频被分组、编码，然后通过网络发送。在接收器应用程序中，编码后的包被排序和解码，然后返回到操作系统音频引擎，并在连接到 DAC 的扬声器上播放。

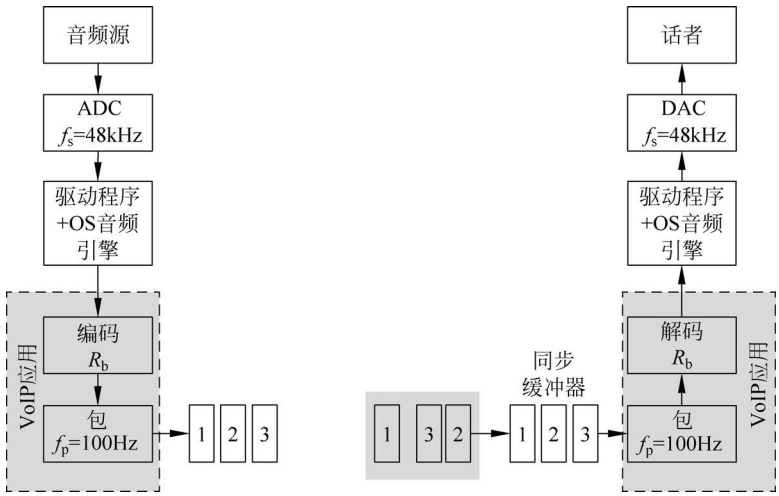

图 5.16　IP 音频

5.3　双通道系统

大多数双通道立体声处理系统可以通过一个 DSP 实现，该 DSP 通过串行接口连接到 ADC 和 DAC，以及一些用于参数调整的控制逻辑单元，如图 5.17 所示。对于更复杂的功能（如 USB、HDMI、LAN、WLAN），则由片上系统（SoC）设备提供 DSP 计算和扩展通信接口。

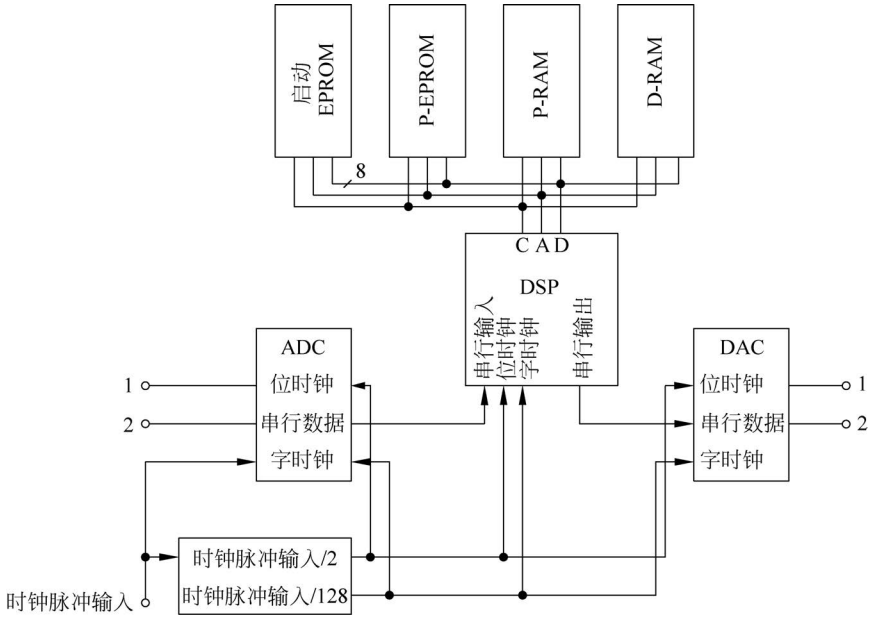

图 5.17　双通道 AD/DA 转换器的双通道 DSP 系统(C=控制,A=地址,D=数据)

5.4　多通道系统

多通道系统可以使用多核 CPU、GPU 和 FPGA(现场可编程门阵列)对多个音频通道进行并行 DSP 处理。一个具有 N 个输入和 K 个输出通道的混合控制台如图 5.18 所示,可以通过单个 FPGA 和适应于 ADC/DAC/AES-BU/MADI 设备的音频接口以及 XGbit/s USB 或以太网等标准计算机接口实现。

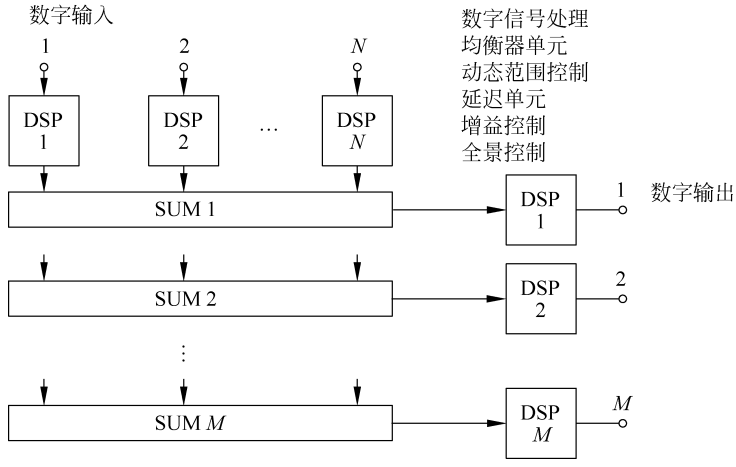

图 5.18　多通道系统的混合控制台