第3章

数字音频基础

本章学习目标

- 了解声音的物理特性。
- 熟练掌握声音信号的数字化原理。
- 熟练掌握典型声音构件的使用方法。

本章首先向读者介绍声音的物理特性、声音的听觉原理、声音信号如何转换为数字信号、数字音频文件基础知识、简单的音乐制作系统以及一些常用的声音构件加工软件的使用方法。本章的内容是在数字音频领域以及在影视声音制作中，需要了解的一些知识。

3.1 声音的物理特性

生活中人们接触到的声音各种各样，千差万别。其中有许多声音让人们感到悦耳、动听。例如，音叉发出的声音、人歌唱的声音、各种乐器的演奏声等，它们都是物体做规则振动时发出的声音。也有嘈杂刺耳、令人厌烦的噪声，如刹车时的摩擦声、打磨工件的声音、电钻的声音。

随着人们对声音研究的不断深入和声音应用领域的广泛拓展，人们越来越多地研究声音的记录及应用。

声音的记录是为了对其进行回放，或者在回放之前还要用一些手段对其进行加工处理。可以将声音的制作工作看做是声音信号的产生、传播、记录、加工和还原的过程。这一复杂的过程包括了声学、电学、磁学、光学、机械学以及物理学中的许多基础知识，尤其是某些物理知识直接影响人们对声音的认识。因此，要想全面了解数字媒体中有关声音的相关问题，必须首先了解声音的物理特性。

声音从物理学上讲是一种声波，声音在空气中振动时，都会使周围的空气分子被挤到一起，然后又撞开，再挤到一起，再撞开，随着振动造成疏密变化，形成疏密波。它是由一些空气分子的振荡导致了气压的轻微变化，而这个变化被我们的耳朵的鼓膜捕获并向大脑传递。这个过程可以归纳为以下 4 个要素。

第一，声源，必须得有一个可以产生声能的声源，如乐器演奏、音叉、琴弦、扬声器喇叭等。

第二，介质，声音必须通过一定的媒介才可传播，如气体、液体或者固体。

第三，接收者，如可以拾取声能的话筒或者人类的听觉器官——耳朵。获取震动，使耳朵鼓膜产生运动，并可以在耳蜗里转换为神经信号。

第四，大脑，将这些神经信号解释为特定的声音。前三点也是物质世界中声音存在的三大基本条件。

正弦波是声音最基本的元素，是声波的可视化的替代物，它展示沿水平轴经历时间内垂直轴上的压力值。水平线表示周围的正常空气压力，即大气压强，直线上面的值形成波峰，它代表声波中压缩的正压区域。水平直线以下的值形成波谷，它代表低于正常压力的负压区域，这些压力随时间的改变就生成了声音的图形表示，称为声波，如图 3-1 所示。图中 A 处代表水平线，B 处代表低压区，C 处代表高压区。

声音信号基于时间线的 360°运动称为一个周期，相位以度数来表示，用来描述在一个周期内的特定的点，如图 3-2 所示。

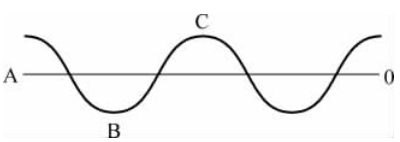

图 3-1　正弦波

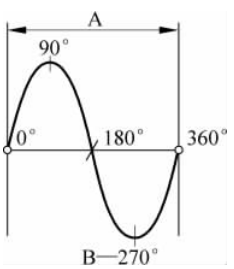

图 3-2　相位

通过对波形的观察，可以直观地感受到声音随时间轴的能量变化，在数字音频工作站环境下，通常将波形放大至单个周期，在对声音进行实际编辑时，往往要在 0°、180°、360°，也就是零交叉点上进行剪辑，零交叉点是指波形与穿越水平线的地方，在此编辑可以避免出现声音的跳点。当制作循环段落的声音时，第一区块的结尾必须连接随后区块的开始处，两者都必须在零交叉点位置进行修剪。通常，循环音乐用于扩展音乐、氛围声轨、持续不变的效果层等。

当两个单一频率波相遇时，同相组合时，它们共同产生一个能量更大的信号。当两个相同的波形相位偏移组合时，两者逐渐产生相长和相消变化，并生成一个新的能量较小的波，这一现象来自于相位消减或抵消，如图 3-3 与图 3-4 所示。

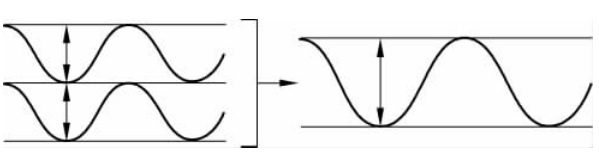

图 3-3　同频同相波相互加强

现实生活中各种声音的声波要复杂得多，一般声波是由各种不同频率的许多简谐振动所组成的，波形的相位不是正好的同相位或者相差 180°。通常把最低的频率称为基音，比

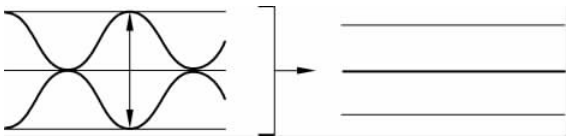

图 3-4　同频反相波相互抵消

基音高的各频率称为泛音。如果各次泛音的频率是基音频率的整数倍，那么这种泛音称为谐音。两个简单波组合后，与原始波形相比变得更加复杂，形成复杂波，如图 3-5 所示。

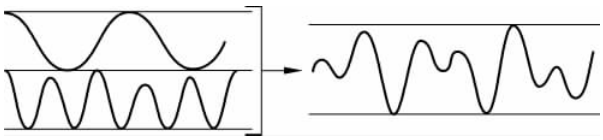

图 3-5　两个简单波组合创造一个复杂波

（1）频率。声音是由物体振动产生的，声音的传播是指振动的传递。物体每秒钟振动的次数，称为频率，单位为赫兹（Hz），即每秒一个周期等于 1Hz，振动越快，频率越高。高频率就意味着高音。人类的频率阈值为 20Hz～20kHz。

（2）波长。对一个声波的整个周期的水平衡量，波长与频率成反比，不同长度的波与其固定的听觉跨度相互影响，低频波的波长很长，从而影响听众对声源方向的判断。在声音的实际应用中，常常使用低频音效来增加听众的害怕反应，一般用于战争或惊悚的场景。听众无法确定声源或声音的方向时，他们就更加具有恐惧感。

（3）振幅。用于描述声音信号当前的能量数或电压，用来表示声音的大小。振幅越大，声波就越强，对于听众的听觉器官的声压越大，声音就越大，听起来越响，描述声压的单位为 dB SPL，dB 指分贝，SPL 指声压级。听众对于声压级的主观感受就是音量。正常人的听觉所能感受到的最小声音级别称为听觉阈值，定为 0 分贝作为参考值，最低的可感知声音与最大的可感知声音之间的范围被称为动态范围。

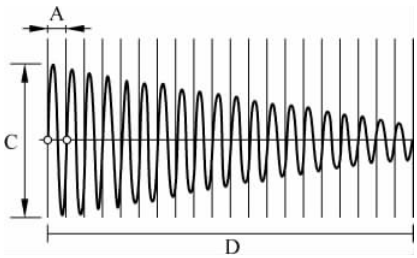

图 3-6　波长、振幅、频率图形描述

图 3-6 表示一个频率为 20Hz 的波形，其中 A 处代表波长，C 处代表振幅，D 处代表一秒钟。

3.2　声音听觉原理

3.2.1　声波与听觉

声以波的形式传播，称为声波，声波借助各种介质向四面八方传播。声波是一种纵波，是弹性介质中传播的压力振动。但在固体中传播时，也可以同时有纵波及横波。

听觉是声波作用于听觉器官，使其感受细胞兴奋并引起听神经的冲动发放传入信息，经各级听觉中枢分析后引起的感觉。这个生理学过程即是听觉。如果听众没有关注某个声

音，那么其所包含的信息不会得到处理或者说是没有意义的。

外界声波通过介质传到外耳道，再传到鼓膜。鼓膜振动，通过听小骨传到内耳，刺激耳蜗内的纤毛细胞而产生神经冲动。神经冲动沿着听神经传到大脑皮层的听觉中枢，形成听觉。

(1) 声波可以营造出空间感。当声能从其声源开始传播，声波的能量在传播过程中有3种类型：直达声、早期反射和后期反射、混响和回声，这三者有助于定义存在声音对象的空间。直达声是从声源直接到达听众的耳朵的声音。早期反射声是从声源经过一次或两次反射后到达耳朵的声音。后期反射声是从声源反射两次以上到达耳朵的声音。声波在室内空间中传播时，会被墙壁、地板等障碍物反射，每反射一次都会被障碍物吸收一部分。这样，当声源停止发声后，声波会在空间中经历多次反射和吸收后才消失，所以声源停止发声后听众还会感觉声音持续一段时间，这种现象就是混响。通常会使用混响的技法来营造空间感。混响是声音的一种存在延续，是声音的拉长；而回声则是离散的重复，是声音的再现。两者的存在使声音有了空间感。在不同的环境中，声音到达耳朵的强度、时间和音色是不同的，根据这些属性听众可以判断出声源的方向和位置。在现实世界中，当声波传播或通过障碍时，高频率的声音容易被吸收。因此，空间越大，高频损失的可能性越大。

(2) 声波的节奏与速度。节奏是可辨的声音与无声之间的组合模式，其循环的快慢称为速度。速度可以恒定、加快或减慢，节奏会影响观众对于画面速度的感知。影视作品中的对白、音效、音乐等都涉及节奏和速度的变化。

(3) 声波能体现出运动感。在影视作品中为了表现出速度感，通常使用多普勒效应的原理，也就是让声源与听众以一定的速度做相对运动，听众所接收到的声音的频率就会变化，从而引起音量与音调的明显变化。当声源与听众做相向运动相互靠近时，听众接收到的声音频率就会升高，当声源与听众做反向运动相互远离时，听众接收到的声音频率就会变低。一个典型的例子就是，火车鸣笛从身边飞驰而过时，音调升高，火车远离时，音调又降低。声源的移动速度越快，多普勒效应就越明显，表现的影片中的人物、动物、物体或镜头的移动速度就越快。

3.2.2 声音三要素

由于人耳听觉系统非常复杂，迄今为止人类对它的生理结构和听觉特性还不能从生理解剖角度完全解释清楚。因此，对人耳听觉特性的研究目前仅限于在心理声学和语言声学。

人耳对不同强度、不同频率声音的听觉范围称为声域。在人耳的声域范围内，声音听觉心理的主观感受主要有响度、音调、音色等特征和掩蔽效应、高频定位等特性。其中响度、音高、音色可以在主观上用来描述具有振幅、频率和相位3个物理量的任何复杂的声音，故又称为声音“三要素”，有的声音听起来音调高，有的声音听起来音调低，声音为什么会有音调高低的不同呢？这主要取决于声音的三要素。而在多种音源场合，人耳掩蔽效应等特性更重要，它是心理声学的基础。

1. 响度

响度又称为声强或音量，它表示的是声音能量的强弱程度，主要取决于声波振幅的大小。声音的响度一般用声压(达因/平方厘米)或声强(瓦特/平方厘米)来计量，声压的单位为帕(Pa)，它与基准声压比值的对数值称为声压级，单位是分贝(dB)。对于响度的心理感

受，一般用单位宋(Sone)来度量，并定义 1kHz、40dB 的纯音的响度为 1 宋。响度的相对量称为响度级，它表示的是某响度与基准响度比值的对数值，单位为方(phon)，即当人耳感到某声音与 1kHz 单一频率的纯音同样响时，该声音声压级的分贝数即为其响度级。可见，无论在客观和主观上，这两个单位的概念是完全不同的，除 1kHz 纯音外，声压级的值一般不等于响度级的值。

响度是听觉的基础。正常人听觉的强度范围为 0～140dB。固然，超出人耳的可听频率范围的声音，即使响度再大，人耳也听不出来。但在人耳的可听频域内，若声音弱到或强到一定程度，人耳同样是听不到的。当声音减弱到人耳刚刚可以听见时，此时的声音强度称为“听阈”。一般以 1kHz 纯音为准进行测量，人耳刚能听到的声压为 0dB，通常大于 0.3dB 即有感受、声强为 10～16W/cm^2 时的响度级定为 0 方。而当声音增强到使人耳感到疼痛时，这个阈值称为“痛阈”。仍以 1kHz 纯音为准来进行测量，使人耳感到疼痛时的声压级约达到 140dB。

实验表明，听阈和痛阈是随声压和频率变化的。听阈和痛阈随频率变化的等响度曲线之间的区域就是人耳的听觉范围。通常认为，对于 1kHz 纯音，0～20dB 为宁静声，30～40dB 为微弱声，50～70dB 为正常声，80～100dB 为响音声，110～130dB 为极响声。而对于 1kHz 以外的可听声，在同一级等响度曲线上有无数个等效的声压频率值，例如，200Hz 的 30dB 的声音和 1kHz 的 10dB 的声音在人耳听起来具有相同的响度，这就是所谓的等响。小于 0dB 听阈和大于 140dB 痛阈时为不可听声，即使是人耳最敏感频率范围的声音，人耳也觉察不到。人耳对不同频率的声音听阈和痛阈不一样，灵敏度也不一样。人耳的痛阈受频率的影响不大，而听阈随频率变化相当剧烈。人耳对 3～5kHz 声音最敏感，幅度很小的声音信号都能被人耳听到，而在低频区和高频区人耳对声音的灵敏度要低得多。响度级较小时，高、低频声音灵敏度降低较明显，而低频段比高频段灵敏度降低更加剧烈，一般应特别重视加强低频音量。通常 200Hz～3kHz 语音声压级以 60～70dB 为宜，频率范围较宽的音乐声压以 80～90dB 最佳。

2. 音调

音调也称为音高，表示人耳对声音调子高低的主观感受。客观上音高大小主要取决于声波基频的高低，频率高则音调高，反之则低，单位用赫兹(Hz)表示。主观感觉的音高单位是“美”，通常定义响度为 40 方的 1kHz 纯音的音高为 1000 美。赫兹与“美”同样是表示音高的两个不同概念而又有联系的单位。

人耳对响度的感觉有一个从听阈到痛阈的范围。人耳对频率的感觉同样有一个从最低可听频率 20Hz 到最高可听频率 20kHz 的范围。响度的测量是以 1kHz 纯音为基准，同样，音高的测量是以 40dB 声强的纯音为基准。实验证明，音高与频率之间的变化并非线性关系，除了频率之外，音高还与声音的响度及波形有关。音高的变化与两个频率相对变化的对数成正比。不管原来频率多少，只要两个 40dB 的纯音频率都增加一个倍频程，人耳感受到的音高变化则相同。在音乐声学中，音高的连续变化称为滑音，一个倍频程相当于乐音提高了一个八度音阶。根据人耳对音高的实际感受，人的语音频率范围可放宽到 80Hz～12kHz，乐音较宽，效果音则更宽。

3. 音色

音色又称为音品，由声音波形的谐波频谱和包络决定。声音波形的基频所产生的听得

最清楚的音称为基音，各次谐波的微小振动所产生的声音称泛音。单一频率的音称为纯音，具有谐波的音称为复音。每个基音都有固有的频率和不同响度的泛音，借此可以区别其他具有相同响度和音调的声音。声音波形各次谐波的比例和随时间的衰减大小决定了各种声源的音色特征，其包络是每个周期波峰间的连线，包络的陡缓影响声音强度的瞬态特性。声音的音色色彩纷呈，变化万千，高保真(Hi-Fi)音响的目标就是要尽可能准确地传输、还原重建原始声场的一切特征，使人们真实地感受到诸如声源定位感、空间包围感、层次厚度感等各种临场听感的立体环绕声效果。

另外，表征声音的其他物理特性还有音值，又称为音长，是由振动持续时间的长短决定的。持续的时间长，音则长，反之则短。从以上主观描述声音的 3 个主要特征看，人耳的听觉特性并非完全线性。声音传到人的耳朵内部经过处理后，除了基音外，还会产生各种谐音及它们的和音及差音，并不是所有这些成分都能被感觉。人耳对声音具有接收、选择、分析、判断响度、音高和音品的功能。例如，人耳对高频声音信号只能感受到对声音定位有决定性影响的时域波形的包络(特别是变化快的包络在内耳的延时)，而感觉不出单个周期的波形和判断不出频率非常接近的高频信号的方向；以及对声音幅度分辨率低、对相位失真不敏感等。这些涉及心理声学和生理声学方面的复杂问题。

3.2.3 人耳的掩蔽效应

一个较弱的声音(被掩蔽音)的听觉感受被另一个较强的声音(掩蔽音)影响的现象称为人耳的掩蔽效应。被掩蔽音单独存在时的听阈分贝值，或者说在安静环境中能被人耳听到的纯音的最小值称为绝对听阈。实验表明，3～5kHz 绝对听阈值最小，即人耳对它的微弱声音最敏感。而在低频和高频区绝对听阈值要大得多。在 800～1500Hz 范围内听阈随频率变化最不显著，即在这个范围内语言可储度最高。在掩蔽情况下，提高被掩蔽弱音的强度，使人耳能够听见时的听阈称为掩蔽听阈，被掩蔽弱音必须提高的分贝值称为掩蔽量。

1. 掩蔽效应

已有实验表明，纯音对纯音、噪声对纯音的掩蔽效应具体表现为以下几个方面。

1) 纯音间的掩蔽

对处于中等强度时的纯音最有效的掩蔽是出现在它的频率附近。低频的纯音可以有效地掩蔽高频的纯音，而反过来则作用很小。

2) 噪声对纯音的掩蔽噪声是由多种纯音组成的，具有无限宽的频谱

若掩蔽声为宽带噪声，被掩蔽声为纯音，则它产生的掩蔽门限在低频段一般高于噪声功率谱密度 17dB，且较平坦，超过 500Hz 时大约每 10 倍频程增大 10dB。若掩蔽声为窄带噪声，被掩蔽声为纯音，则情况较复杂。其中位于被掩蔽音附近的由纯音分量组成的窄带噪声即临界频带的掩蔽作用最明显。所谓临界频带，是指当某个纯音被以它为中心频率，且具有一定带宽的连续噪声所掩蔽时，如果该纯音刚好能被听到时的功率等于这一频带内噪声的功率，那么这一带宽称为临界频带宽度。临界频带的单位为巴克(Bark)，1Bark 等于一个临界频带宽度。频率小于 500Hz 时，1Bark 约等于 freq/100。频率大于 500Hz 时，1Bark 约等于 9+4log(freq/1000)，即约为某个纯音中心频率的 20%。通常认为，20Hz～16kHz 范围内有 24 个子临界频带。而当某个纯音位于掩蔽声的临界频带之外时，掩蔽效应仍然存在。

2. 掩蔽类型

1）频域掩蔽

所谓频域掩蔽，是指掩蔽声与被掩蔽声同时作用时发生掩蔽效应，又称为同时掩蔽。这时，掩蔽声在掩蔽效应发生期间一直起作用，是一种较强的掩蔽效应。通常，频域中的一个强音会掩蔽与之同时发声的附近的弱音，弱音离强音越近，一般越容易被掩蔽；反之，离强音较远的弱音不容易被掩蔽。例如，一个 1000Hz 的音比另一个 900Hz 的音高 18dB，则 900Hz 的音将被 1000Hz 的音掩蔽。而若 1000Hz 的音比离它较远的另一个 1800Hz 的音高 18dB，则这两个音将同时被人耳听到。若要让 1800Hz 的音听不到，则 1000Hz 的音要比 1800Hz 的音高 45dB。一般来说，低频的音容易掩蔽高频的音，在距离强音较远处，绝对听阈比该强音所引起的掩蔽阈值高，此时噪声的掩蔽阈值应取绝对听阈。

2）时域掩蔽

所谓时域掩蔽，是指掩蔽效应发生在掩蔽声与被掩蔽声不同时出现时，又称为异时掩蔽。异时掩蔽又分为导前掩蔽和滞后掩蔽。若掩蔽声音出现之前的一段时间内发生掩蔽效应，则称为导前掩蔽，否则称为滞后掩蔽。产生时域掩蔽的主要原因是人的大脑处理信息需要花费一定的时间，异时掩蔽也随着时间的推移很快会衰减，是一种弱掩蔽效应。一般情况下，导前掩蔽只有 3～20ms，而滞后掩蔽却可以持续 50～100ms。

3.2.4 声音立体感与立体声

所谓声音立体感，就像自己处在声音的发生现场，在现场左耳和右耳所听到的声音是不一样的，如左边的琴声大一些，而右边鼓声大一些，就像身临其境。人们听声音时，可以分辨出声音是由哪个方向传来的，从而大致确定声源的位置。人们之所以能分辨声音的方向，是由于人们有两只耳朵的缘故。例如，在人们的右前方有一个声源，那么，由于右耳离声源较近，声音就首先传到右耳，然后才传到左耳，并且右耳听到的声音比左耳听到的声音稍强些。如果声源发出的声音频率很高，传向左耳的声音有一部分会被人头反射回去，因而左耳就不容易听到这个声音。两只耳朵对声音的感觉的这种微小差别，传到大脑神经中，就使人们能够判断声音是来自右前方。就像立体电影，左眼和右眼所看到的东西，也是有微小变化的。

立体声是指具有立体感的声音，它是一个几何概念，指在三维空间中占有位置的事物。因为声源有确定的空间位置，声音有确定的方向来源，人们的听觉有辨别声源方位的能力。特别是有多个声源同时发声时，人们可以凭听觉感知各个声源在空间的位置分布状况。从这个意义上讲，自然界所发出的一切声音都是立体声，如雷声、火车声、枪炮声、风声、雨声等。

当人们直接听到这些立体空间中的声音时，除了能感受到声音的响度、音调和音色外，还能感受到它们的方位和层次。这种人们直接听到的具有方位层次等空间分布特性的声音，称为自然界中的立体声。立体声源于双声道的原理，立体声和双声道不是一个概念，但是它们是因果关系。

众所周知，立体声能给人真实的声音环境感。不论是纯声音节目，还是影视作品，从两个音箱或者更多个音箱中传出的立体声响效果，必然比孤单的简陋扬声器震撼得多。必须弄清楚什么是立体声，如果仅仅把声音信号用导线分接到两个或者更多个音箱上播放，并不是真正的立体声。

现实环境中的声音，从各个方向到达听者的声波，其强度和成分，在绝大多数情况下是不相同的。例如，有人在听者的右边发出一个声音，那么对于听者来说，从左、右两边来的声波情况肯定是不一样的。

正因为这样的方向差异性，即使听者闭眼不看，也可以判断出声音来自什么方向。所以说，要想营造出一个完整的声音环境假象，必须尽量在每个方向上都忠实地还原该方向上的声波振动状况。

因此，如果仅是在各个音箱上播放完全相同的信号，就不能复原真实的声场，而其仅有的实际效果只不过是让声源显得更庞大一些。这就是假立体声，只有立体声设备的外观形式，却没有立体声回放的正确信号。例如，两位歌手在剧场舞台两侧，相距甚远，同时演唱，如果只用录音笔录一条信号，回来播放时将其导线分配到环绕听者的多个音箱上，听者也只能感到两个歌手的声音并在一处，而没有离得很远，这就导致声音环境的还原在这里不真实，即使录音的音质再好也无济于事。

正确的做法是采用两个或多个话筒进行环绕式录音进行现场录制，这样就得到了不止一条声音信号，再正确地将这组声音同步送到各个音箱上播放。

一般地说，每个音箱只接收一条信号，称为一个声道。所以，两个音箱的设备可以称为双声道立体声设备，多个音箱环绕听者的就可以称为多声道环绕声设备，这里添加“设备”两字，是为了强调设备仅是实现立体声的条件之一，更重要的是拥有立体声信号，当然还包括信号与设备的正确连接。

当今我国电影的声音已经采用环绕立体声格式了，而电视节目正处于从单声道到多声道发送的过渡极端，广播除了少数频率的专门立体声节目之外，还都是单声道。

在这里还要了解一个声像的概念。声像又称为虚声源或感觉声源。当人们在听音环境良好的音乐厅欣赏音乐时，毫无疑问地能体会出一种身心的愉悦。这除了作品本身的感染力和演奏者高度技艺以及听者的艺术素养以外，也与人耳的听觉能体现出音乐的现场感和包围感不无关系。精于乐感的行家即使不看舞台也能细微地分辨出，小提琴在左前方、鼓在左后方、钢琴在右前方、大提琴在右后方、长笛在中前方而黑管在中后方等声部的空间位置。利用一个完善的立体声记录和重放系统，当人们再度聆听时，仍然可以分辨出上述的各乐器的位置。这种在听音者听感中所展现的各声部空间位置，并由此而形成的声画面，通常称为声像。

用两个或两个以上的音箱进行立体声放音时，听音者对声音位置的感觉印象，有时也称这种感觉印象为幻象，声音图像的空间分布由人的双耳效应决定。立体声放音正是以声像的形式，再现原来声音的空间分布，从而使人们产生一种幻觉，诱发立体感觉。利用立体声技术还可以人为地改变原来的声像位置，通常称为声像移动。

双声道的“声像(Pan)”则是理解立体声的首要概念。在双声道播放的声音中，如果哪个成分给人的主观听感是来源于标有 Pan 的位置，则该位置就是这个声像，如图 3-7 所示。在含有多个声源的声音中，每个声源的 Pan 都可能各不相同。当然，每个声源的 Pan 也可以随时间变化，如电影里的踱步声，从一边逐渐移向另一边，也有可能需要让某些声源的 Pan 位置不明确，这就要求音频制作人

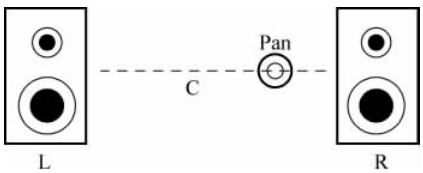

图 3-7　声像

员在制作时进行一些处理。

3.3 声音信号的数字化

爱迪生发明的留声机用机械的方法实现了对声波的记录和重放，电声录音技术把声波信号转化为电流电压的变化，并在磁带上通过对信号的控制实现了对声音的编辑重塑，而进入到数字时代，电信号的变化进一步被转换为二进制数字信号进行描述和处理。

我们已经身处在一个以“数字”为标志的时代中，数字技术极大地推动了艺术作品的生产和传播，通过各种媒介，大众有幸能轻易地领略到数量空前巨大的艺术作品。不但艺术作品的观赏变得十分容易，甚至其创作方式也发生了巨大的变革，只要给个人计算机配上相应的外设和编写相应的软件，它就可以使我们成为很多领域的“专家”，通过人对计算机下达各种命令，也就是说把人脑的思维传达给它，其余的工作都是由计算机所做的事情了。数字艺术家创作正在变得数据化，每个人都可以把自己的艺术灵感和创意较为容易地表现出来。

平时在读书看报的时候，一边用眼睛浏览报纸上的文字，一边用大脑不断地解释文字的含义。但是，如果手上拿的书报上印的是甲骨文，相信除了少数专家，一般人即使看到了文字也很难明白其含义，文字可以说是一种“人能够阅读的数据”，并且这种能力也是可以训练的。当然，人们感知数据的形式不是单一化的，例如，可以利用录音带、电影胶片、录像带记录声音和视频的模拟信息。在这种情况下，所用的载体不再是纸张，而是胶片、磁带。对这种形式的数据，人们不能直接阅读，必须借助录音机、放映机、放像机等这些播放设备才能看到，从而知道上面信息的含义。而让计算机理解和处理信息的基础，就是必须将所有的信息转换成计算机可以识别的信息。计算机可以识别的信息只有二进制码，也就是“0”和“1”两种状态，所以无论是文字、图像，还是声音、视频，对计算机而言，数字化以后都是“0”和“1”的排列组合，这样计算机就可以对它们进行处理、存储和传输了。

3.3.1 模拟信号与数字信号

1. 模拟信号

模拟信号是在时间和数值上都是连续变化的一类信号。模拟信号利用对象的一些物理属性来表达、传递信息。例如，非液体气压表利用指针螺旋位置来表达压强信息。在电学中，电压是模拟信号最普遍的物理参数，除此之外，频率、电流和电荷也可以被用来表达模拟信号。

任何的信息都可以用模拟信号来表达。这里的信号常常指物理现象中被测量的对变化有响应的表达，如声音、光、温度、位移、压强，这些物理量可以使用传感器测量。模拟信号中，不同的时间点位置的信号值可以是连续变化的。而对于数字信号，不同时间点的信号值总是处于预先设定的离散点，因此如果物理量的真实值不能在这些预设值中被找到，那么这时数字信号就与真实值存在一定的偏差。

在电子技术中，典型的模拟信号主要有工频信号、射频信号、视频信号等。

模拟是一组随时间改变的数据，如某地方的温度变化，汽车在行驶过程中的速度，或电路中某节点的电压幅度等。有些模拟信号可以用数学函数来表示，其中时间是自变量而信号本身则作为因变量。离散时间信号是模拟信号的采样结果，离散信号的取值只在某些固

定的时间点有意义，其他地方没有定义，而不像模拟信号那样在时间轴上具有连续不断的取值。

若离散时间信号在各个采样点上的取值只是原来模拟信号取值的一个近似，那么我们就可以用有限字长来表示所有的采样点取值，这样的离散时间信号成为数字信号。将一组精确测量的数值用有限字长的数值来表示的过程称为量化。从概念上讲，数字信号是量化的离散时间信号，而离散时间信号则是已经采样的模拟信号。

模拟信号的主要优点是其精确的分辨率，在理想情况下，它具有无穷大的分辨率；与数字信号相比，模拟信号的信息密度更高；由于不存在量化误差，它可以对自然界物理量的真实值进行尽可能逼近的描述。

模拟信号的另一个优点是当达到相同的效果，模拟信号处理比数字信号处理更简单。模拟信号的处理可以直接通过模拟电路组件实现，如运算放大器等，而数字信号处理往往涉及复杂的算法，甚至需要专门的数字信号处理器。

模拟信号的主要缺点是它总是受到杂音的影响。信号被多次复制，或进行长距离传输之后，这些随机噪声的影响可能会变得十分显著。在电学里，使用接地屏蔽、线路良好接触、使用同轴电缆或双绞线，可以在一定程度上缓解这些负面效应。

噪声效应会使信号产生有损。有损后的模拟信号几乎不可能再次被还原，因为对所需信号的放大会同时对噪声信号进行放大。如果噪声频率与所需信号的频率差距较大，可以通过引入电子滤波器，过滤掉特定频率的噪声，但是这一方案只能尽可能地降低噪声的影响。因此，在噪声的作用下，虽然模拟信号理论上具有无穷分辨率，但并不一定比数字信号更加精确。

尽管数字信号处理算法相对复杂，但是现有的数字信号处理器可以快速地完成这一任务。另外，随着计算机系统的逐渐普及，使得数字信号的传播和处理都变得更加方便。例如，照相机等设备都逐渐实现数字化，尽管它们最初必须以模拟信号的形式接收真实物理量的信息，最后都会通过模拟/数字转换器转换为数字信号，以方便计算机进行处理，或通过互联网进行传输。

2. 数字信号

数字信号是在时间和数值上都是不连续变化的一类信号。数字信号是离散时间信号的数字化表示，通常可由模拟信号获得。二进制码就是一种数字信号。二进制码受噪声的影响小，易于由数字电路进行处理，所以得到了广泛的应用。

随着电子技术的飞速发展，数字信号的应用也日益广泛。很多现代的媒体处理工具，尤其是需要和计算机相连的仪器都从原来的模拟信号表示方式改为使用数字信号表示方式。常见的例子包括手机、视频或音频播放器和数码相机等。

一般情况下，数字信号是以二进制数来表示的，因此信号的量化精度一般以比特(bit)来衡量。数字信号有诸多的特点。

1）抗干扰能力强、无噪声积累

在模拟通信中，为了提高信噪比，需要在信号传输过程中及时对衰减的传输信号进行放大，信号在传输过程中不可避免地叠加上的噪声也被同时放大。随着传输距离的增加，噪声累积越来越多，以致使传输质量严重恶化。

对于数字通信，由于数字信号的幅值为有限个离散值，在传输过程中虽然也受到噪声的

干扰，但当信噪比恶化到一定程度时，即在适当的距离采用判决再生的方法，再生成没有噪声干扰的和原发送端一样的数字信号，所以可实现长距离高质量的传输。

2）便于加密处理

信息传输的安全性和保密性越来越重要，数字通信的加密处理比模拟通信容易得多，以话音信号为例，经过数字变换后的信号可用简单的数字逻辑运算进行加密、解密处理。

3）便于存储、处理和交换

数字通信的信号形式和计算机所用信号一致，都是二进制代码，因此便于与计算机联网，也便于用计算机对数字信号进行存储、处理和交换，可使通信网的管理、维护实现自动化、智能化。

4）设备便于集成化、微型化

数字通信采用时分多路复用，不需要体积较大的滤波器。设备中大部分电路是数字电路，可用大规模和超大规模集成电路实现，因此体积小、功耗低。

5）便于构成综合数字网和综合业务数字网

采用数字传输方式，可以通过程控数字交换设备进行数字交换，以实现传输和交换的综合。另外，电话业务和各种非电话业务都可以实现数字化，构成综合业务数字网。

6）占用信道频带较宽

一路模拟电话的频带为 4kHz 带宽，一路数字电话约占 64kHz，这是模拟通信目前仍有生命力的主要原因。随着宽频带信道（光缆、数字微波）的大量利用以及数字信号处理技术的发展，数字电话的带宽问题已不是主要问题了，可以将一路数字电话的数码率由 64kbps 压缩到 32kbps 甚至更低的数码率。

以上介绍可知，数字通信具有很多优点，所以各国都在积极发展数字通信。近年来，我国数字通信得到迅速发展，正朝着高速化、智能化、宽带化和综合化方向迈进。

模拟信号和数字信号之间可以相互转换：模拟信号一般通过 PCM 脉码调制方法量化为数字信号，即让模拟信号的不同幅度分别对应不同的二进制值，例如采用 8 位编码可将模拟信号量化为 $2^8=256$ 个量级，实际应用中常采取 24 位或 30 位编码。数字信号一般通过对载波进行移相的方法转换为模拟信号。计算机、计算机局域网与城域网中均使用二进制数字信号，目前在计算机广域网中实际传送的则既有二进制数字信号，也有由数字信号转换而得的模拟信号。但是更具应用发展前景的是数字信号。

3.3.2 声音信号数字化简介

声音或者图像，需要记录和传送，必须通过工具转化为声音信号。

“模拟”一词本意为“类似”、“相似”，在工程技术领域中，它有“连续”的含义；而“数字”则具有“不连续”、“离散”的含义。一般说到模拟量都是指“连续量”，而数字量则是指“离散量”。

不论是声音还是图像，要想记录和传送，都要通过声电转换器（话筒）或者广电转换器（摄像机），将其转变成相应的电信号，所得到的电信号在时间和幅度上都是连续变化的，因此称为模拟信号。模拟音频信号在传输过程中容易受到干扰，干扰杂音不能与有用信号区别开来，容易损害原信号，记录和复制时也不可避免地会造成信噪比下降，每转录一次，就会增加一定的噪声，这是不可避免的，随着转录次数的增多，噪声不断增加，使信噪比下降

严重。

为了克服模拟信号的缺点，就需要有高精度、高性能的装置以及高水平的维护保养技术。把模拟信号数字化是克服这些缺点的最有效办法。数字音频信号是指把声音信号数字化，将本来是模拟量的声音信号变换成离散的二进制码进行记录，这种二进制码仅由“1”和“0”组成，重放时只需识别出“1”和“0”就可以了，因而抗杂音抗干扰的能力特别强。

从模拟信号到数字信号这个转化过程通常称为模/数转换(A/D)，要经过采样、量化和编码的 3 个过程。

1. 采样

话音信号是模拟信号，它不仅在幅度取值上是连续的，而且在时间上也是连续的。要使话音信号数字化并实现时分多路复用，首先要在时间上对话音信号进行离散化处理，这一过程称为采样。所谓采样，就是每隔一定的时间间隔，抽取话音信号的一个瞬时幅度值，也就是采样值，采样后所得出的一系列在时间上离散的采样值称为样值序列。这样文字叙述采样有些抽象，下面形象地举例说明。例如，电影，大家都知道原始电影是用胶片制作的，是一张张连续的胶片组成的，但是人眼看到的画面是连续的，其实这个连续的画面是由每秒 24 幅静止的画面顺次播放的，即它是一系列时间上连续的画面组成的。在拍摄时，摄影机以每秒钟拍摄 24 幅画面的方法记录下真实的活动场面。在放映时，人们的眼睛就从这一系列单幅静止的画面的连续放映中重构出运动的画面。

电影的例子说明，通过一定的手段对信息进行采样加以记录，在一定情况下并不能影响信息的还原。然而采样涉及一个名词就是采样周期，就是间隔多长时间，对被采样的对象进行一次采样。通常以采样频率来衡量，采样频率是指一秒钟内采样的次数，采样频率与采样周期互为倒数，如果一秒钟采样 50 次，则称采样频率为 50Hz。单位时间内抽取多少个样本值，是以赫兹(Hz)为单位的，从这个意义上讲，电影的图像的采样频率就是 24Hz。不同采样率的应用级频率范围如表 3-1 所示。

表 3-1 不同采样率的应用级频率范围

采样率	应　　用	频率范围
11 025Hz	低端多媒体	0～5512Hz
22 050Hz	FM 收音机(高端多媒体)	0～11 025Hz
32 000Hz	优于 FM 收音机(标准的广播率)	0～16 000Hz
44 100Hz	CD	0～22 050Hz
48 000Hz	标准 DVD	0～24 000Hz
96 000Hz	蓝光 DVD	0～48 000Hz

采样频率越高，则采样后的采样点排列越密集，得到这种脉冲序列值越接近原信号，或者说明脉冲序列表示的原信号失真越小，它可恢复的音频信号分量越丰富，其声音保真度越好，如图 3-8 所示。这样看来，采样频率越高声音越接近原始信号。但采样频率的提高，将对同一长度的信号增加采样点，进而增加了数字处理的运算量。但是，如果采样频率太低就会产生信息丢失，在恢复信号时会产生失真。例如，电影画面如果每秒钟只保留一幅画面，就会影响到画面播放的流畅性。为了既减少采样点，减小数据量，又能使采样后的信号不失真，因此，选择合适的采样频率非常关键，瑞典科学家奈奎斯特和美国数学家香农经过研究

得出了重要的采样定理。

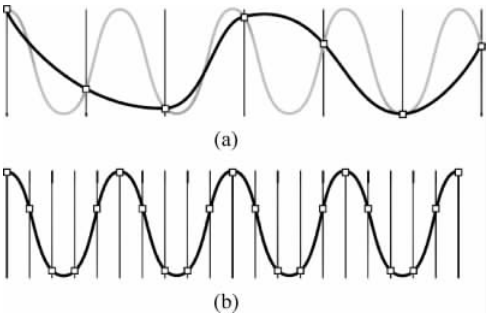

图 3-8　两个采样频率

采样定理说明了一个频带限制在 $0\sim f_m$ 之间的低通模拟信号能被一个离散取样序列所替代而不会有任何信息的丢失，还描述了如何从取样中重建原始的连续信号，并且进一步指出：采样频率 f_s 必须至少是信号最高频率 f_m 的两倍，即 $f_s\geqslant 2f_m$，才能不失真地从采样值恢复出原始信号。采样频率也称为奈奎斯特频率。

采样定理明确地说明了一个给定宽的波形应该以多高的频率来进行取样。例如，制造厂商选择 44.1kHz 作为 CD 唱片的采样频率，可记录的最高音频为 22kHz，人耳听觉范围的频率带宽为 20Hz～20kHz，这正是人耳能分辨的最高音频再加上一定的保护频带，这样可获得音质与原始声音几乎毫无差别的听觉效果，这也就是人们常说的超级高保真音质的重要参数，这是综合考虑了唱盘播放时间的实用性和媒介造价等因素制定的。除此之外，声音的数字化参数中常见的采样频率还有 48kHz 和 96kHz。

例如，一路电话信号的频带为 300～3400Hz，$f_m=3400$Hz，则采样频率 $f_s\geqslant 2\times 3400=6800$Hz。如按 6800Hz 的采样频率对 300～3400Hz 的电话信号采样，则采样后的样值序列可不失真地还原成原来的话音信号，话音信号的采样频率通常取 8000Hz。对于 PAL 制电视信号，视频带宽为 6MHz，按照 CCIR601 建议，亮度信号的采样频率为 13.5MHz，色度信号为 6.75MHz。

采样后的样值序列在时间上是离散的，可进行时分多路复用，也可将各个采样值经过量化、编码变换成二进制数字信号。

2. 量化

采样是对时间的计算，而量化是对采样量的计算。在音频系统中，则是对采样时刻的音频信号数值的计量。采样和量化是将模拟信号数字化的基础，它们分别决定了系统的带宽和分辨率两个特性参数。

采样把模拟信号变成了时间上离散的脉冲信号，但脉冲的幅度仍然是模拟的，还必须进行离散化处理，才能最终用数码来表示。这就要对幅值进行舍零取整的处理，这个过程称为量化。

量化的过程就是通过采集波形各个点上的能量进行振幅数字化，并为每一个点分配一定的电压值。量化的单位为比特，比特深度表示了音频的精度，它影响着波形的最终形状和质量。增加采样率，声波波形的所有部分就更加聚合到最近的值。量化生成声音像素，随着量化精度的提高，解析度也相应提高，信号就更趋向于原始声音。量化比特深度如表 3-2 所示。

表 3-2 量化比特深度

位 深 度	应 用	量化级别	动态范围
8bit	电话、网络	256	48dB
16bit	CD	65 536	96dB
24bit	DVD	16 777 216	144dB
32bit	最优质	4 294 967 296	192dB

打个简单的比方，量化的概念有点像人们买衣服，虽然每个成年人的身材千差万别，但是被“量化”为 XS、S、M、L、XL、XXL、XXXL 等几个有限的尺寸，这样对制造商来说，工作可以大大简化，如果针对每个人不同身材生产，那就是个性化定制，显然“数据量”会非常大。所以，有时可能并不很关注每一个规格具体是多少，而是更关注有多少个规格供我们选择。

量化有两种方式，分别是均匀量化和非均匀量化。量化方式中，取整时只舍不入，即 0～1V 间的所有输入电压都输出 0V，1～2V 间所有输入电压都输出 1V 等。采用这种量化方式，输入电压总是大于输出电压，因此产生的量化误差总是正的，最大量化误差等于两个相邻量化级的间隔 Δ。量化方式在取整时有舍有入，即 0～0.5V 间的输入电压都输出 0V，0.5～1.5V 间的输出电压都输出 1V 等。采用这种量化方式量化误差有正有负，量化误差的绝对值最大为 Δ/2。因此，采用有舍有入法进行量化，误差较小。

上面所述的采用均匀间隔量化级进行量化的方法称为均匀量化或线性量化，这种量化方式会造成大信号时信噪比有余而小信号时信噪比不足的缺点。

如果使小信号时量化级间宽度小些，而大信号时量化级间宽度大些，就可以使小信号时和大信号时的信噪比趋于一致。这种非均匀量化级的安排称为非均匀量化或非线性量化。数字电视信号大多采用非均匀量化方式，这是由于模拟视频信号要经过校正，而校正类似于非线性量化特性，可减轻小信号时误差的影响。

实际信号可以看成量化输出信号与量化误差之和，因此只用量化输出信号来代替原信号就会有失真。一般来说，可以把量化误差的幅度概率分布看成在 $-\Delta/2$～$+\Delta/2$ 之间的均匀分布。可以证明，量化失真率，即与最小量化间隔的平方成正比。最小量化间隔越小，失真就越小。最小量化间隔越小，用来表示一定幅度的模拟信号时所需要的量化级数就越多，因此处理和传输就越复杂。所以，量化既要尽量减少量化级数，又要使量化失真看不出来。一般都用一个二进制数来表示某一量化级数，经过传输在接收端再按照这个二进制数来恢复原信号的幅值。所谓量化比特数，是指要区分所有量化级所需几位二进制数。例如，有 8 个量化级，那么可用 3 位二进制数来区分，因此称 8 个量化级的量化为 3 比特量化。8 比特量化则是指共有 256 个量化级的量化。

在实际应用中，量化级总是有限的，因此模拟信号转换为数字信号后，很明显，在这个“归属”的过程中有一定的误差，也就是说会损失一定的信息，这个误差称为量化噪声。这也正如人们生活当中有很多的时候在允许的范围内进行四舍五入，舍去无关紧要的细枝末节，当然这样做的结果与真实值相比是有误差的。如果量化等级的数目增多，意味着分出的等级数会增多，那么引进的量化噪声就会减小。

量化误差与噪声是有本质的区别。因为任一时刻的量化误差是可以从输入信号求出，而噪声与信号之间就没有这种关系。可以证明，量化误差是高阶非线性失真的产物。但量

化失真在信号中的表现类似于噪声，也有很宽的频谱，所以也被称为量化噪声并用信噪比来衡量。

对于音频信号的非均匀量化也是采用压缩、扩张的方法，即在发送端对输入的信号进行压缩处理再均匀量化，在接收端再进行相应的扩张处理。

目前国际上普遍采用容易实现的 A 律 13 折线压扩特性和 μ 律 15 折线压扩特性。我国规定采用 A 律 13 折线压扩特性。

采用 13 折线压扩特性后小信号时量化信噪比的改善量可达 24dB，而这是靠牺牲大信号量化信噪比换来的，会亏损 12dB。

3. 编码

采样、量化后的信号还不是数字信号，需要把它转换成相应的数制代码，将量化后的幅值转变成相应的数制代码的过程称为编码。最简单的编码方式是二进制编码。使用二进制编码就要确定代码的位数，单位是比特。具体来说，就是用比特二进制码来表示已经量化了的样值，每个二进制数对应一个量化值，然后把它们排列，得到由二值脉冲组成的数字信息流。编码过程在接收端，可以按所收到的信息重新组成原来的样值，再经过低通滤波器恢复原信号。用这样方式组成的脉冲串的频率等于抽样频率与量化比特数的积，称为所传输数字信号的数码率。显然，抽样频率越高，量化比特数越大，数码率就越高，数据量就越大，所需要的传输带宽就越宽。编码可以用不同的方法进行，脉冲编码调制是把模拟信号转换成数字信号的基本方法。

编码可以使用各种码，除了上述的自然二进制码，还有其他形式的二进制码，如格雷码和折叠二进制码等。这 3 种码各有优缺点。

(1) 自然二进制码和二进制数一一对应，简单易行，它是权重码，每一位都有确定的大小，从最高位到最低位依次为 2 的 N 次幂排列，可以直接进行大小比较和算术运算。自然二进制码可以直接由数/模转换器转换成模拟信号，但在某些情况，例如从十进制的 3 转换为 4 时二进制码的每一位都要变，使数字电路产生很大的尖峰电流脉冲。

(2) 格雷码则没有这一缺点，它在相邻电平间转换时，只有一位发生变化，格雷码不是权重码，每一位码没有确定的大小，不能直接进行比较大小和算术运算，也不能直接转换成模拟信号，要经过一次码变换，变成自然二进制码。

(3) 折叠二进制码沿中心电平上下对称，适用于表示正负对称的双极性信号。它的最高位用来区分信号幅值的正负。折叠码的抗误码能力强。

经过采样、量化和编码的过程，先将时间连续的模拟信号转换成时间离散、幅度连续的采样信号，再将这些信号转换成时间离散、幅度离散的数字信号，最后将量化后的信号编码形成一个二进制码输出。计算机对这些二进制数据以音频文件的形式进行存储、编辑和处理，并且可以将其还原成原始的波形进行播放。这个过程称为解码，即数/模变换。

3.3.3 声音质量与数据率

所谓声音的质量，是指经传输、处理后音频信号的保真度。

目前，业界公认的声音质量标准分为四级，即数字激光唱盘 CD-DA 质量，其信号带宽为 10Hz～20kHz；调频广播 FM 质量，其信号带宽为 20Hz～15kHz；调幅广播 AM 质量，其信号带宽为 50Hz～7kHz；电话的话音质量，其信号带宽为 200～3400Hz。可见，数字激

光唱盘的声音质量最高，电话的话音质量最低。除了频率范围外，人们往往还用其他方法和指标来进一步描述不同用途的音质标准。

对模拟音频来说，再现声音的频率成分越多，失真与干扰越小，声音保真度越高，音质也越好。如在通信科学中，声音质量的等级除了用音频信号的频率范围外，还用失真度、信噪比等指标来衡量。对数字音频来说，再现声音频率的成分越多，误码率越小，音质越好。通常用数据率来衡量，采样频率越高，量化比特数越大，声道数越多，存储容量越大，当然保真度就高，音质就好。

声音的类别特点不同，音质要求也不一样。例如，语音音质保真度主要体现在清晰、不失真、再现平面声像。乐音的保真度要求较高，营造空间声像主要体现在用多声道模拟立体环绕声，或虚拟双声道 3D 环绕声等方法，再现原来声源的一切声像。

音频信号的用途不同，采用压缩的质量标准也不一样。例如，电话质量的音频信号采用 ITU-TG·711 标准，8kHz 取样，8bit 量化，码率 64Kbps。AM 广播采用 ITU-TG·722 标准，16kHz 取样，14bit 量化，码率 224Kbps。高保真立体声音频压缩标准由 ISO 和 ITU-T 联合制订，CD11172-3MPEG 音频标准为 48kHz、44.1kHz、32kHz 取样，每声道数码率 32～448Kbps，适合 CD-DA 光盘用。

对声音质量要求过高，则设备复杂；反之，则不能满足应用。一般以音质够用，又不浪费存储空间为原则。

大家知道，计算机发出的信号都是数字形式的。比特是计算机中数据量的单位，也是信息论中使用的信息量的单位。英文 bit 来源于 binary digit，意思是“二进制数字”，因此一个比特就是二进制数字中的一个 1 或 0。网络技术中的速率是指连接在计算机网络上的主机在数字信道上传送数据的速率，它也称为数据率或比特率。速率是计算机网络中最重要的一个性能指标。

数字音频文件存储在计算机中要占据一定的空间，然而不同的采样频率、量化深度和录制时间生成的音频文件大小也不相同。例如，用标准的 CD 音质 44.1kHz、16bit 来进行两个声道立体声采样、录制 3 分钟的音频，那么该未经压缩的声音数据文件的大小为：1 秒内采样 44.1×1000 次，每次的数据量是 16×2＝32(bit)，那么 3 分钟的总共数据量是 44 100×32×60×3＝254 016 000(bit)，换算成计算机中的常用单位(Byte，B)，总数据量是 254 016 000/8/(1024×1024)＝30.28MB。同样，可以计算出衡量音频质量的指标，音频流码率，即每秒钟音频的二进制数据量。上述例题的数据率是 176.4Kbps。

一个汉字在计算机里占用两个字节，那么 176.4KB 的空间可以存储 176.4KB/2＝88 200 个汉字。也就是说 1 秒的数字音频数据量与近 9 万个汉字的数据量相当。由此可见，数字音频文件的数据量是十分庞大的。

如果要衡量一个数字音频音质好坏，通常需要参考以下指标。

(1) 采样频率：即采样点之间的时间间隔。采样间隔时间越短，音质越好。

(2) 量化深度：指单位电压值和电流值之间的可分等级数。可分等级越多，音质越好。

(3) 音频流码率：数字化后，单位时间内音频数据的比特容量。数据率越大，音质越好。它是数字音乐压缩效率的参考性指标，表示记录音频数据每秒钟所需要的平均比特值。CD 中的数字音乐比特率为 1411.2Kbps，也就是记录 1 秒钟的 CD 音乐，需要 1411.2×1024 比特的数据，近乎于 CD 音质的 MP3 数字音乐需要的比特率为 112～128Kbps。

以上3个方面的指标中，前两个指标也是绝对指标，而音频流码率是一个相对指标，可以间接考察音频的质量。同样一个声音，在保证声音效果基本相同的情况下，用不同的编码方式进行编码，在存储容量上就会有大小的差别，当然在播放时需要的解码时间也不一样。

因此，音频流码率是一个间接的指标，当两个数字音频文件用同样的编码方式时，可以用它来衡量它们之间的音质好坏，但对于不同编码方式的数字音频文件，就不一定适用。

3.4 数字音频文件

计算机的产生为音频处理带来了极大的方便，将模拟音频转换为数字音频进行处理，不但可以高保真，还可以应用各种数字音效，丰富音乐的创作手段。

数字音频是一种利用数字化手段对声音进行录制、存放、编辑、压缩或播放的技术，它是随着数字信号处理技术、计算机技术、多媒体技术的发展而形成的一种全新的声音处理手段。数字音频的主要应用领域是音乐后期制作和录音。

计算机数据的存储是以0和1的形式存取的，那么数字音频就是首先将音频文件转化，接着再将这些电平信号转化成二进制数据保存，播放的时候就把这些数据转换为模拟的电平信号再送到喇叭播出，数字声音和一般磁带、广播、电视中的声音就存储播放方式而言有着本质区别。相比而言，它具有存储方便、存储成本低廉、存储和传输的过程中没有声音的失真、编辑和处理非常方便等特点。

3.4.1 数字音频文件大小的计算

模拟音频经过采样、量化、编码的过程转变为数字音频，为了获得高质量的数字音频，就必须提高采样的分辨率和频率，以采集更多的信号样本。然而采样数越多音频数据量也就越大。音频数字化的过程中其数据传输率与在计算机中的实时传输有直接关系，因此，数据传输率是计算机处理时要掌握的一个基本技术参数。对于无压缩的数字音频来说，数据传输率按如下公式计算：

数据传输率＝采样频率×量化位数×声道数

其中数据传输率以比特每秒(bps)为单位；采样频率以赫兹(Hz)为单位；量化位数以比特(b)为单位，单声道取值1，立体声取值2。

如果采用无压缩编码形式，音频数字化所需的存储空间按如下公式计算：

音频数据量＝数据传输率×持续时间/8

其中数据量以字节(Byte)为单位，8比特位一个字节，所以在公式中要除以8得到以字节为单位的结果。数据传输率以比特每秒(bsp)为单位，持续时间以秒为单位。以CD音质为例，采用44.1kHz，16位量化位数，采用立体声标准，其音频数据传输率为：

数据传输率＝44.1kHz×16×2＝1411.2Kbps

那么，一分钟的数字音频数据量为：1411.2Kbps×60s/8＝10 584 000B≈10.09MB。

由此可见，数字音频的数据量比较大，对于计算机存储和数据传输都会带来不便。因此，在实际应用中，要根据音源的质量和实际需要灵活运用。或者采用通常的办法，在编码过程中利用各种编码算法对其进行压缩，以得到较好的音质和较小的文件量。

3.4.2 数字音频文件的常用压缩方法

现实世界中的声音非常复杂，人们通常采用的是脉冲代码调制编码，即 PCM 编码。PCM 通过采样、量化、编码 3 个步骤将连续变化的模拟信号转换为数字编码。

相对于声音的模拟信号，音频编码最多只能做到无限接近，任何数字音频编码方案都是有损的，因为无法完全还原。在计算机应用中，能够达到最高保真水平的就是 PCM 编码，CD、DVD 以及常见的 WAV 文件中均有所应用。因此，PCM 被认为是无损编码，因为 PCM 代表了数字音频中最佳的保真水准，但是 PCM 也只能做到最大程度的无限接近。PCM 编码的最大的优点就是音质好，最大的缺点就是体积大。人们通常把 MP3 列入有损音频编码范畴，是相对 PCM 编码的。强调编码的相对性的有损和无损，是为了说明要做到真正的无损是困难的。数字音频都要经过压缩的过程，常用压缩率来描述数字声音的压缩效率，也就是音乐文件压缩前和压缩后大小的比值。

要减小音频文件量，降低磁盘占用空间，只有两种方法，降低对音频的采样率或者压缩音频。降低指标是不可取的，因此专家们研发了各种压缩方案。因为音频的用途不一样，各种音频压缩编码所达到的音质和压缩比也不一样，要根据需要来确定。

以下列举数字音频文件的常用压缩方法。

1. WAV

WAV 是微软公司开发的一种声音文件格式，它符合 RIFF(Resource Interchange File Format)文件规范，用于保存 Windows 平台的音频信息资源，被 Windows 平台及其应用程序所支持。WAV 对音频流的编码没有硬性规定，除了 PCM 之外，还有几乎所有支持 ACM 规范的编码都可以为 WAV 的音频流进行编码。支持多种音频位数、采样频率和声道，标准格式的 WAV 文件和 CD 格式一样，也是 44.1kHz 的采样频率、速率 88Kbps、16 位量化位数，因此 WAV 格式的声音文件质量和 CD 相差无几，也是目前 PC 上广为流行的声音文件格式，几乎所有的音频编辑软件都兼容 WAV 格式。

另外，苹果公司开发的 AIFF(Audio Interchange File Format)格式和为 UNIX 系统开发的 AU 格式，它们都和 WAV 非常相像，在大多数的音频编辑软件中也都支持这几种常见的音乐格式。

PCM 编码的 WAV 文件是音质最好的格式，Windows 平台下，所有音频软件都能够提供对它的支持。在开发多媒体软件时，往往大量采用 WAV，用做事件声效和背景音乐。PCM 编码的 WAV 可以达到相同采样率和采样大小条件下的最好音质，因此也被大量用于音频编辑、非线性编辑等领域。

2. MP3

MP3 格式诞生于 20 世纪 80 年代的德国，所谓的 MP3，是指 MPEG 标准中的音频部分，也就是 MPEG 音频层。根据压缩质量和编码处理的不同分为三层，分别对应“*.mp1”、“*.mp2”、“*.mp3”这 3 种声音文件。MPEG 音频文件的压缩是一种有损压缩，MPEG3 音频编码具有 10∶1～12∶1 的高压缩率，同时基本保持低音频部分不失真，但是牺牲了声音文件中 12kHz 到 16kHz 高音频这部分的质量来换取文件的尺寸，相同长度的音乐文件，用 *.mp3 格式来储存，一般只有 *.wav 文件的 1/10，而音质要次于 CD 格式或 WAV 格式的声音文件。由于其文件尺寸小，音质好，因此在它问世之初还没有什么别的音频格式可以

与之匹敌，因而为 *.mp3 格式的发展提供了良好的条件。直到现在，这种格式还是风靡一时，作为主流音频格式的地位难以被撼动。

MP3 格式压缩音乐的采样频率有很多种，可以用 64Kbps 或更低的采样频率以节省存储空间，也可以用 320Kbps 的标准达到极高的音质。

MP3 具有不错的压缩比，使用 LAME 编码的中高码率的 MP3，听感上已经非常接近源 WAV 文件。使用合适的参数，LAME 编码的 MP3 很适合于音乐欣赏。几乎所有著名的音频编辑软件也提供了对 MP3 的支持，可以将 MP3 像 WAV 一样使用，但由于 MP3 编码是有损的，因此多次编辑后，音质会急剧下降，MP3 长远的历史和不错的音质，使之成为应用最广的有损编码之一，网络上可以找到大量的 MP3 资源，MP3 是被多个播放媒介支持的最好的编码之一。MP3 也并非完美，在较低码率下表现不好。MP3 也具有流媒体的基本特征，可以做到在线播放。

3. VQF

.vqf 是雅马哈公司格式，它的核心是减少数据流量但保持音质的方法来达到更高的压缩比，可以说技术上也是很先进的，但是由于没有做很好的维护，这种格式没有得到推广。.vqf 可以用雅马哈的播放器播放。同时雅马哈也提供从 *.wav 文件转换到 *.vqf 文件的软件。

4. WMA

WMA（Windows Media Audio）格式来自于微软，在低码率下，有着好过 MP3 的音质表现，它和日本 Yamaha 公司开发的 VQF 格式一样，是以减少数据流量但保持音质的方法来达到比 MP3 压缩率更高的目的，WMA 的压缩率一般都可以达到 1∶18 左右，WMA 的另一个优点是内容提供商可以通过 DRM（Digital Rights Management）方案如 Windows Media Rights Manager 7 加入防复制保护。这种内置了版权保护技术可以限制播放时间和播放次数，甚至播放的机器等，这可以有效打击盗版，另外 WMA 还支持音频流技术，适合在网络上在线播放，作为微软抢占网络音乐的开路先锋，可以说是技术领先、风头强劲，更方便的是不用像 MP3 那样需要安装额外的播放器，而 Windows 操作系统和 Windows Media Player 可以直接播放 WMA 音乐，还可以直接把 CD 光盘转换为 WMA 声音格式，在 Windows XP 操作系统中，WMA 是默认的编码格式。WMA 这种格式在录制时可以对音质进行调节。同一格式，音质可以与 CD 媲美，压缩率较高的可用于网络广播。

5. OGG

OGG 格式完全开源，完全免费，和 MP3 有相似之处的新格式。OGGVorbis 也是对音频进行有损压缩编码，但通过使用更加先进的声学模型去减少损失，因此相同码率编码的 OGGVorbis 比 MP3 音质更好一些，文件也更小一些。另外，MP3 格式是受专利保护的。发布或者销售 MP3 编码器、MP3 解码器、MP3 格式音乐作品，都需要付专利使用费。而 OGGVorbis 就完全没有这个问题。目前，OGGVorbis 虽然还不普及，但在音乐软件、游戏音效、便携播放器、网络浏览器上都得到广泛支持，OGG 具有流媒体的基本特征。和 MP3 一样，OGGVorbis 是一种灵活开放的音频编码，能够在编码方案已经固定下来后还能对音质进行明显的调节和新算法的改良。因此，它的声音质量将会越来越好，和 MP3 相似，OGGVorbis 更像一个音频编码框架，可以不断导入新技术逐步完善。和 MP3 一样，OGG 也支持 VBR。

OGG 是一种非常有潜力的编码，在各种码率下都有比较惊人的表现，尤其中低码率下。OGG 除了音质好之外，还是一个完全免费的编码。OGG 有着非常出色的算法，可以用更小的码率达到更好的音质，128Kbps 的 OGG 比 192Kbps 甚至更高码率的 MP3 还要出色。OGG 的高音具有一定的金属味道，因此在编码一些高频要求高的乐器独奏时，OGG 的这个缺陷会暴露出来。

6. FLAC

FLAC 即是 Free Lossless Audio Codec 的缩写，为无损音频压缩编码。FLAC 是一套著名的自由音频压缩编码，其特点是无损压缩。不同于其他有损压缩编码如 MP3 与 AAC，它不会破坏任何原有的音频资讯，所以可以还原音乐光盘音质。现在它已被很多软件及硬件音频产品所支持。FLAC 是免费的并且支持大多数的操作系统。并且 FLAC 提供了在开发工具 Autotools、MSVC、Watcom C、ProjectBuilder 上的 Build 系统。

7. APE

APE 是目前流行的数字音乐文件格式之一。与 MP3 这类有损压缩方式不同，APE 是一种无损压缩音频技术，也就是说当用户将从音频 CD 上读取的音频数据文件压缩成 APE 格式后，还可以再将 APE 格式的文件还原，而还原后的音频文件与压缩前的一模一样，没有任何损失。APE 的文件大小大概为 CD 的一半，但是随着宽带的普及，APE 格式受到了许多音乐爱好者的喜爱，特别是对于希望通过网络传输音频 CD 的朋友来说，APE 可以帮助他们节约大量的资源。作为数字音乐文件格式的标准，WAV 格式容量过大，因而使用起来很不方便。因此，一般情况下把它压缩为 MP3 或 WMA 格式。压缩方法有无损压缩、有损压缩以及混成压缩。MPEG、JPEG 就属于混成压缩，如果把压缩的数据还原回去，数据其实是不一样的。当然，人耳是无法分辨的。因此，如果把 MP3、OGG 格式从压缩的状态还原回去的话，就会产生损失。然而 APE 压缩格式即使还原，也能毫无损失地保留原有音质。所以，APE 可以在不损失高音质的情况下，进行压缩和还原。当然，目前只能把音乐 CD 中的曲目和未压缩的 WAV 文件转换成 APE 格式，MP3 文件还无法转换为 APE 格式。事实上 APE 的压缩率并不高，虽然音质保持得很好，但是压缩后的容量也没小多少。一个 34MB 的 WAV 文件，压缩为 APE 格式后，仍有 17MB 左右。

APE 的本质，其实是一种无损压缩音频格式。庞大的 WAV 音频文件可以通过 Monkey's Audio 这个软件压缩为 APE。很多时候它被用做网络音频文件传输，因为被压缩后的 APE 文件容量要比 WAV 源文件小一半多，可以节约传输所用的时间。更重要的是，通过 Monkey's Audio 解压缩还原以后得到的 WAV 文件可以做到与压缩前的源文件完全一致。所以 APE 被誉为无损音频压缩格式，Monkey's Audio 被誉为无损音频压缩软件。与采用 WinZip 或者 WinRAR 这类专业数据压缩软件来压缩音频文件不同，压缩之后的 APE 音频文件是可以直接被播放的。

3.4.3 常用的数字音频文件格式

不同的编码方式对应计算机中不同的文件格式，反映在计算机中就是文件的后缀名不同。数字音频的常见格式有以下几种。

1. WAV

WAV 格式是微软公司开发的一种声音文件格式，也称为波形声音文件，是最早的数字

音频格式，被 Windows 平台及其应用程序广泛支持。WAV 格式支持许多压缩算法，支持多种音频位数、采样频率和声道；采用 44.1kHz 的采样频率，16 位量化位数，因此 WAV 的音质与 CD 相差无几，但 WAV 格式对存储空间需求太大，不便于交流和传播。

2. MIDI

MIDI 是 Musical Instrument Digital Interface，乐器数字接口的缩写，是数字音乐/电子合成乐器的统一国际标准。它定义了计算机音乐程序、数字合成器及其他电子设备交换音乐信号的方式，规定了不同厂家的电子乐器与计算机连接的电缆和硬件及设备间数据传输的协议，可以模拟多种乐器的声音。MIDI 文件就是 MIDI 格式的文件，在 MIDI 文件存储的是一些指令。把这些指令发送给声卡，由声卡按照指令将声音合成出来。MID 文件格式由 MIDI 继承而来，MID 文件并不是一段录制好的声音，而是记录声音的信息，然后再告诉声卡如何再现音乐的一组指令。这样一个 MIDI 文件每存储一分钟的音乐只用 5～10KB。今天，MID 文件主要用于原始乐器作品、流行歌曲的业余表演、游戏音轨以及电子贺卡等。*.mid 文件重放的效果完全依赖声卡的档次。*.mid 格式的最大用途在计算机作曲领域。*.mid 文件可以用作曲软件编写，也可以通过声卡的 MIDI 接口把外接音序器演奏的乐曲输入计算机中，制成 *.mid 文件。

3. CDA

大家都很熟悉 CD 这种音乐格式了，扩展名为.cda，其采样频率为 44.1kHz，16 位量化位数。CD 存储采用了音轨的形式，又称为“红皮书”格式，记录的是波形流，是一种近似无损的格式。

4. MP3

其全称是 MPEG-1 Audio Layer 3，它在 1992 年合并至 MPEG 规范中。MP3 能够以高音质、低采样率对数字音频文件进行压缩。也就是说，文件数据量较大的音频文件能够在音质丢失很小的情况下把文件压缩得更小。

5. MP3 Pro

该格式是由瑞典 Coding 科技公司开发的，包含了两大技术：一是来自于 Coding 科技公司所特有的解码技术；二是由 MP3 的专利持有者法国汤姆森多媒体公司和德国 Fraunhofer 集成电路协会共同研究的一项译码技术。MP3 Pro 可以在基本不改变文件大小的情况下改善原先的 MP3 音乐音质，它能够在用较低的比特率压缩音频文件的情况下，最大限度地保持压缩前的音质。

6. WMA

WMA 是微软针对互联网音频、视频领域的格式。WMA 格式是以减少数据流量但保持音质的方法来达到更高的压缩率为目的，其压缩率一般可以达到 1∶18。此外，WMA 还可以通过 DRM 方案加入防止复制保护，或者加入限制播放时间和播放次数，甚至是播放机器的限制，可有效地打击盗版。

7. MP4

该格式采用的是美国电报电话公司(AT&T)所研发的以知觉编码为关键技术的 A2B 音乐压缩技术，也可以称为 AAC 技术，由美国网络技术公司及 RIAA 联合公布的一种新的音乐格式。MP4 在文件中采用了保护版权的编码技术，只有特定的用户才可以播放，这样有效地保证了音乐版权的合法性。另外，MP4 的压缩比例达到了 1∶15，体积较 MP3 更小，

但音质却没有下降。

8. SACD

SACD(Super Audio CD)是由 Sony 公司正式发布的。它的采样率为 CD 格式的 64 倍，即 2.8224MHz。SACD 重放频率宽达 100kHz，为 CD 格式的 5 倍，24 位量化位数，远远超过 CD，声音的细节表现更为丰富、清晰。

9. QuickTime

QuickTime 是苹果公司于 1991 年推出的一种数字流媒体，它面向视频编辑、Web 网站创建和媒体技术平台，QuickTime 几乎支持所有主流的个人计算机平台，可以通过互联网提供实时的数字化信息流、工作流与文件回放功能。

10. VQF

VQF 格式是由 Yamaha 和 NTT 共同开发的一种音频压缩技术，它的压缩率能够达到 1∶18，因此相同情况下压缩后 VQF 的文件体积比 MP3 小 30%～50%，更便于网上传播，同时音质极佳，接近 CD 音质(16 位 44.1kHz 立体声)。但 VQF 未公开技术标准，至今未能流行开来。

11. DVD Audio

DVD Audio 是新一代数字音频格式，与 DVD Video 尺寸及容量相同，为音乐格式的 DVD 光碟，采样频率为 48kHz/96kHz/192kHz 和 44.1kHz/88.2kHz/176.4kHz 可供选择，量化位数可以为 16 位、20 位或 24 位，它们之间可自由地进行组合。低采样率的 192kHz、176.4kHz 虽然是二声道重播专用，但它最多可收录到六声道。而以二声道 192kHz/24b 或六声道 96kHz/24b 收录声音，可容纳 74min 以上的录音，动态范围达 144dB，整体效果出类拔萃。

12. MD

MD(Mini Disc)由 Sony 公司出品。MD 之所以能在一张小小的盘中存储 60～80min 采用 44.1kHz 采样的立体声音乐，就是因为使用了 ATRAC 算法，也就是自适应声学转换编码压缩音源。这是一套基于心理声学原理的音响译码系统，它可以把 CD 唱片的音频压缩到原来数据量的大约 1/5 而声音质量没有明显损失。ATRAC 利用人耳听觉的心理声学特性以及人耳对信号幅度、频率、时间的有限分辨能力，编码时将人耳感觉不到的成分不编码、不传送，这样就可以相应减少某些数据量的存储，从而既保证音质又达到缩小体积的目的。

13. RA

RA(RealAudio)是由 Real Networks 公司推出的一种文件格式，其最大的特点就是可以实时传输音频信息，尤其是在网速较慢的情况下，仍然可以较为流畅地传达数据，因此 RealAudio 主要适用于网络上的在线播放。现在 RealAudio 文件格式主要有 RA、RM、RMX 3 种，这些文件的共同之处在于随着网络带宽的不同而改变声音的质量，在保证大多数人听到流畅声音的前提下，令带宽较宽敞的听众获得较好的音质。

14. Liquid Audio

Liquid Audio 是一家提供付费下载的音乐网站。它通过在音乐中采用自己独有的音频编码格式来提供对音乐的版权保护。Liquid Audio 的音频格式就是所谓的 LQT。如果想在 PC 中播放这种格式的音乐，就必须使用 Liquid Player 和 Real Jukebox 中的一种播放

器。这些文件也不能够转换成 MP3 和 WAV 格式，这使得采用这种格式的音频文件无法被共享和刻录到 CD 中。如果非要把 Liquid Audio 文件刻录到 CD 中，就必须使用支持这种格式的刻录软件和 CD 刻录机。

15. Audible

Audible 拥有 4 种不同的格式，即 Audible1、Audible2、Audible3、Audible4。Audible. com 网站主要是在互联网上贩卖有声书籍，并对他们所销售的商品、文件通过 4 种 Audible. com 专用音频格式中的一种提供保护。每一种格式主要考虑音频源以及所使用的收听设备。格式 1、格式 2 和格式 3 采用不同级别的语音压缩，而格式 4 采用更低的采样率与 MP3 相同的解码方式，所得到的语音更清楚，而且可以更有效地从网上下载。Audible 所采用的是他们自己的桌面播放工具，这就是 Audible Manager，使用这种播放器就可以播放存在 PC 或者是传输到便携式播放器上的 Audible 格式文件。

16. VOC

在 DOS 程序和游戏中常会遇到这种文件，它是随声霸卡一起产生的数字声音文件，与 WAV 文件的结构相似，可以通过一些工具软件方便地相互转换。

17. AU

在 Internet 上的多媒体声音主要使用该种文件格式。AU 文件是 UNIX 操作系统下的数字声音文件，由于早期 Internet 上的 Web 服务器主要是基于 UNIX 操作系统的，因此这种文件成为网络上唯一使用的标准声音文件。

18. AIFF

AIFF(. aif)是 Apple 公司开发的声音文件格式，被 Macintosh 平台和应用程序所支持。

19. Amiga

Amiga 声音(. svx)格式是 Commodore 公司开发的声音文件格式，被 Amiga 平台和应用程序所支持，不支持压缩。

20. MAC

MAC 声音(. snd)是 Apple 公司开发的声音文件格式，被 Macintosh 平台和多种 Macintosh 应用程序所支持。

21. S48

S48(Stereo，48kHz)采用 MPEG-1 layer 1、MPEG-1 layer 2(简称 MP1、MP2)声音压缩格式，由于易于编辑、剪切，因此在广播电台应用较为广泛。

22. AAC

AAC 实际上是高级音频编码的缩写。AAC 是由 Franuhofer IIS-A、杜比和 AT&T 共同开发的一种音频格式，它是 MPEG-2 规范的一部分。AAC 所采用的运算法则与 MP3 的运输法则有所不同，AAC 通过结合其他的功能来提高编码效率。AAC 的音频算法在压缩能力上远远超过了以前的一些压缩算法。它还同时支持多达 48 个音轨、15 个低频音轨、更多种采样率和比特率、多种语言的兼容能力、更高的解码效率。总之，AAC 可以在比 MP3 文件缩小 30%的前提下提供更好的音质。

从另外一种角度而言，一种数字音频格式就对应这一数字音频的编码方式，在播放不同的音频文件格式时可能需要不同的播放器，技术上一般称为解码器或者音频解码算法。

数字音频给人类的生活带来了前所未有的变化，它以音质优秀、传播无损耗、可进行多

种编辑和转换而成为主流，并且应用于各个方面，如音响设备、IP 电话、卫星电话、数字卫星电视及专业录音、制作等。展望未来，数字音频将会应用于更多的领域，而且会拥有更清晰、更真实的音质，更小巧的体积，更方便的传输和转换功能。

3.5 MIDI 与 MIDI 音乐制作系统

MIDI 是由世界上主要电子乐器制造厂商建立起来的一个通信标准，并于 1988 年正式提交给 MIDI 制造商协会，便成为数字音乐的一个国际标准。

MIDI 标准规定了电子乐器与计算机连接的电缆硬件以及电子乐器之间、乐器与计算机之间传递数据的通信协议等规范。MIDI 标准使计算机音乐程序、不同厂家生产的电子合成乐器以及其他电子设备互相发送和接收音乐数据。MIDI 文件记录的是一系列指令而不是数字化后的波形数据，所以它占用存储空间比 WAV 要小很多。

3.5.1 MIDI 简介

通俗地讲，MIDI 可以理解为电脑音乐的统称，它包括协议、硬件设备等所有的相关技术。可以为不同乐器创建数字声音，可以模拟小提琴、小号、二胡、钢琴等常见的乐器。MIDI 音乐在计算机作曲领域的使用最为广泛。它可以由作曲软件编写，也可以通过声卡的 MIDI 接口把外接音序器演奏的音乐输入计算机中生成 MIDI 文件。

与数字音频文件有所不同，一个 MIDI 文件可能每分钟只有 10KB 大小，因此每兆磁盘空间可以存储上百分钟的 MIDI 音乐。MIDI 是一种典型的参数声音文件。

参数声音记录的方法，与采样记录不同。它完全不采用“按照特定的间隔记录介质分子的振动状态”的方法，而是记录声音在某些“关键点”上的“关键信息”。MID 文件并不是一段录制好的声音，而是记录声音的信息，然后再告诉计算机上的声卡如何再现声音的一组指令。当回放时，参照这些信息，进行提示和指导，使回放的声音尽量接近声音的原来面貌。其实这种方法更接近人类所使用的乐谱记录音乐的方式。

与 MIDI 音乐的记录方式作对比，乐谱本质上也是记录音乐的一些关键信息，而不是记录每个时刻声音振动的具体情况。对照着乐谱演奏，从乐谱上可以直接读出每个音多高，歌词是什么，速度有多快。乐谱里面的“音高”、“拍子”、“歌词”、“表情符号”等，实际上就是关键信息，它为正确的再现声音，提供了依据和参考。经过分析可知，乐谱只能采用参数方式，不能采取采样记录。

MIDI 文件就如同一张数字化的乐谱，使用计算机 CPU 强大的运算能力，在音序器软件中处理这些数字化乐谱上的数据，将这些指令发送到计算机音频卡上的合成器芯片中，根据乐谱上记录的各种参数，通过分析处理，选用相应的乐器演奏相应的音符，再按照数字化乐谱中规定的演奏节奏、演奏强度等进行处理。经由合成器芯片解释的 MIDI 指令符号被翻译形成波形，然后通过声音发生器送往扬声器进行播放。

参数记录的方式也具有明显的局限性。当人们拿着乐谱唱歌时，由于演唱者个体嗓音的不同，演唱过程中对音乐的理解和投入的感情是不同的，因此乐谱上音乐的原貌不会被绝对精准的再现出来，它只能尽最大可能保留音乐的原貌。同样，用参数方法记录下的声音文件，由于关键信息不可能完全代表声音的原貌，因此回放的效果可能因回放设备的不同，而

与该声音的本来面目有所不同。参数记录的这些缺点，是由其原理决定的，虽然能采取措施在一定程度上进行弥补，却不能完全消除。

不过，参数记录方法与采样记录方法相对比有一个明显的优势，这就是数据存储效率极高。MIDI 文件记录的是乐器指令，它与记录文字的 Word 文档的记录方式差不多，所需的存储空间很小，通常几十 KB 的文件可以记录长达数分钟的音乐。而波形文件，其文件的数据量是由音频的采样频率、量化位数、声道数以及压缩比等参数的具体值决定的，一般数据量都很大。MIDI 文件由于抓住了关键信息，无须对声音做每秒成千上万次的采样逐点记录，所以必然极大地减少了文件的数据量。

在具体实现技术上，如果通过播放设备的电路来回放参数声音文件，就不能使用普通的乐谱，而是要用专门的数据字节来代表乐谱上的各个参数。例如，用一组数据代表一个音，在这组数据中，分别用不同位置的数字代表此音的音高、音色、开始时间、持续长度等信息，电路即可依据这些数据在正确的时间，调用正确的音色种类，用正确的方式发出这个音。很多组这样的数据就构成了一个“参数音乐文件”。MIDI 音乐文件就是运用这种基本方法才诞生的。

3.5.2 MIDI 音乐合成法

MIDI 音乐文件与传统波形文件是不相同的，要掌握 MIDI 音乐的合成方法，就必须知道 MIDI 音乐文件与波形文件之间的区别，就要知道其记录声音的原理。

MIDI 音乐文件与波形文件的不同点表现为记录原理不同、声音来源不同、文件容量不同、适用范围不同。

(1) 记录原理不同。波形文件通过对声波进行采样、量化、编码的过程，得到一批量化的数值，再对离散的采样数值进行编码形成二进制数字存储起来，从而形成数字化的音频信号数据，这些数据再经过还原，形成波形后送到扬声器进行播放。MIDI 文件不记录声音的采样数据，它只记录音符的键、音量、力度和速度等信息，记录的是描述演奏过程的指令，这些指令发送到合成器合成声波进行播放。

(2) 声音来源不同。波形文件可以获取广泛的音源，包括各种乐器的声音、各种语音、各种音响效果。而 MIDI 文件通过接收电子乐器演奏的指令数据，并由音序器记录，它更多的是记录音乐文件，不能记录人讲话的声音。

(3) 文件容量不同。使用参数记录方式记录的 MIDI 文件和使用采样记录方式记录的波形文件，类似于计算机图形学中使用参数记录方式的矢量图和使用采样记录方式的位图。由于记录方式的不同，矢量图的文件量明显小于位图的文件量。MIDI 文件比波形文件所需占用的存储空间要小得多。

(4) 适用范围不同。MIDI 文件记录的是乐谱指令，对其编辑修改非常方便，用户可以通过音序器任意改变 MIDI 文件的音色、节奏，还可以更换不同的乐器，通常用在影片配乐的前期创作工作中使用。波形文件音源广泛，音质好，但数据量大，在影片配乐中常用于后期混音工作中。

当作曲家创作一段音乐并将它交给演奏人员时，实际上是传送该音乐作品的表演数据。用来记载这些为再现该乐曲所需要的指令的载体是纸张。但是很显然，即使将乐谱贴近耳

朵也不会听到任何音乐，这是因为乐谱实际上并不等于音乐，这不过是演奏音乐的一组指令。乐谱中的指令指示演奏家要演奏哪些音符、该持续多长时间、应该使用多大的力度以及其他各种各样重要的信息。

MIDI 音乐文件就相当于一份电子化的乐谱。MIDI 文件采用参数记录的方法记录着声音，它主要用于记录音乐，难以记录语音，因为乐器声音的关键信息相对较少，声波随时间的变化总体来说相对简单。而人类的语言声音，无论是频率状况还是随时间的变化状况，都难以使用很少的几个参数来描述。MIDI 采用参数记录的方法，不会像 PCM 那样详细地记录每个声音每个时刻的细节，但是它能记录的信息种类，仍然比乐谱的记录更加复杂和全面。除了音高、发音的开始时刻、发音的结束时刻、力度、音量、音色、节奏等，还有许多传统乐理体系中从未出现过，专属于计算机音乐技术的概念。例如，通道、音色库号、弯音程度、调制程度等。但不论是不是传统的，这些概念对于 MIDI 设备来说，都只是一个个的“关键信息”，或称为“关键事件”，简称“事件”。参数音乐是众多事件的有机组合，而不是对众多采样值的机械排列。

MIDI 文件是由一系列 MIDI 指令和消息组成的文件，都采用二进制编码的形式进行传输，以 8 位(1 字节)为一组基本数据。因为一个字节能够表示的不同信息只有 255 种，有时不够使用，所以每个 MIDI“事件”可能由 1～3 个字节组成。MIDI 文件中每个信息由若干字节组成，这些信息分为两类，包括状态信息和数据信息两部分内容。状态信息标明该事件的性质，如演奏一个音符或某一个声音音高的升降。数据信息标明该事件在该性质下发生的程度或细节，描述演奏者演奏了哪个音符或这个音符升降的程度。

如果直接阅读这些 MIDI 指令代码，是很不直观的。所以 MIDI 制作软件就起到了一个“翻译”作用，它以人们熟悉的乐谱、单词和十进制数值为载体，供人制作和编辑想要的音乐，并将这些音乐内容转换成 MIDI 代码，形成音序文件。

在明白了声音记录的方法之后，再探讨一下音频的相互转换的问题，也就是参数数字声音和采样数字声音的相互转换，通俗地讲，就是“MIDI 和音频的相互转换”。这个问题的答案很特别，两者之间转换的方向不同，其难度也有很大的不同。

把 MIDI 转换成音频，其现实性和可行性，都很容易理解。把 MIDI 内录成音频文件或者使用像 Audition 等融音序处理和音频处理于一身的专业级软件，通过运算直接把音序计算成音频数据，可以直接输出。或者一边播放 MIDI 文件，一边打开音频编辑软件对其进行声音录制，就可以将 MIDI 文件转换成波形文件。

把音频转换成 MIDI，却没有那么简单。就相当于使用一些语音识别软件所实现的功能，把一段讲话变成文档中的文字。一般要依靠专门的制作员手工完成，即制作依靠乐理知识和听辨能力，自行分析出音乐中的各个关键信息，并依照这些信息另行制作 MIDI 文件。但是不同的乐曲有不同的乐器参与，混合后会产生不同的效果，单从波形分析，很难进行区分，也就达不到理想的效果。

两类乐曲的转换，有非常大的难度，这与两类声音的记录原理在本质上的差异有很大的关系。把参数记录的音乐，在播放过程中随时采样，对计算机来说很容易，但计算机难以使用固定的方法，从千变万化的声音波形采样中分析出足够的“关键信息”，当然也就无法将其总结为音乐参数了。

3.5.3 计算机合成 MIDI 音乐

MIDI 是音乐和计算机结合的产物，它是用于在音乐合成器、电子乐器、计算机之间交换音乐信息的一种标准协议。MIDI 产生声音的方法与声音波形采样输入的方法有很大的不同。它不是将模拟信号进行数字编码，而是把 MIDI 音乐设备上产生的每个动作记录下来。例如，在电子键盘上演奏，MIDI 文件记录的不是实际乐器发出的声音，而是记录弹奏时弹的音符、节奏信息以及各种表情控制信息，包括按键的速度、按键的力度、有无颤音以及音色的变化等。然后对记录下来的信息进行编辑修改，可以对一些音乐段落作升调或降调处理，插入或删除一段音乐，或者改变一段音符的位置。经过编辑修改的弹奏信息随时都可以发送给音源，由音源发出相应的音色声音。

利用 MIDI 制作软件可以直接在计算机上创作、演奏 MIDI 音乐，也可以通过计算机声卡上的 MIDI 接口，从带 MIDI 输入的乐器中采集音乐，形成 MIDI 文件，或用连接在计算机上的 MIDI 键盘直接创作音乐，形成 MIDI 文件。MIDI 作曲系统的核心部分是一个被称为音序器的软件。这个软件即可以装到个人计算机里。它是提供 MIDI 数据录制、编辑和回放的专业软件。其工作原理类似虚拟指挥家在指挥软件音源进行演奏，所有的音频软件都内置 MIDI 音序功能模块，如 Adobe Audition、Cakewalk、Cubase、Pro Tools 等音频制作软件。

音序器实际上是一个音乐词处理器，应用它可以记录、播放和编辑各种不同 MIDI 乐器演奏出的乐曲。音序器并不真正的记录声音，它只记录和播放 MIDI 信息，这些信息就像印在纸上的乐谱一样，它本身不能直接产生音乐，MIDI 本身也不能产生音乐，但是它包含有如何产生音乐所需的所有指令，如用什么乐器、演奏什么音符、演奏的速度快慢、演奏的力度多强等。

音序器的使用过程完全与专业录音棚里多轨录音机一样，可以把许多独立的声音记录在音序器里，其区别仅仅是音序器只记录演奏时的 MIDI 数据，而不记录声音。它可以一轨一轨地进行录制，也可以一轨一轨地进行修改，当用户弹键盘音乐时，音序器记录下从键盘来的 MIDI 数据。每一轨的音乐信息依次添加，属于不同声部的演奏信息可以被分别记录在不同的 MIDI 通道中。一旦把所需要的数据存储下来以后，可以播放刚作好的曲子。如果觉得这一声部的曲子不错，可以把别的声部加上去，新加上去的声部播放时完全与第一道同步。也就是说，音序器可以将所有 MIDI 通道中的演奏信息同时演奏。这样，一个人就可以完成整个乐队的多声部的演奏和录音任务。

计算机合成 MIDI 音乐的优势：作为单独设备的音序器，音轨数相对少一些，大概 8～16 轨，而作为计算机软件的音序器几乎多达 50 000 个音符，64～200 轨以上；音序器与磁带不同，它只受到硬件有效的 RAM(随机存储器)和存储容量的限制，所以作曲、配器根本用不着担心录制时间的限制，不用考虑像使用磁带录制声音时的录制时间问题。

MIDI 技术的一大优点就是它传输和存储在计算机里的数据量相当小，一个包含有一分钟立体声的数字音频文件需要约 10 兆字节的存储空间。然而，一分钟的 MIDI 音乐文件只有 2KB。这也意味着，在乐器与计算机之间的传输数据是很低的，也就是说即使最低档的计算机也能运行和记录 MIDI 文件。

通过使用 MIDI 音序器可以大大降低作曲和配器成本，根本用不着庞大的乐队来演奏。

音乐编导在家里就可把曲子创作好，配上乐器，再也用不着大乐队在录音棚里逐个声部去录制了。只需要用录音棚里的计算机或键盘，把存储在键盘里的 MIDI 音序器的各个声部的全部信息输入到录音机上即可。

MIDI 程序的设计目标就是要将所要演奏的音乐或音乐曲目，按其进行的节奏、速度、技术措施等要求，转换成 MIDI 控制语言，以便在这些 MIDI 指令的控制之下，各种音源在适当的时间点上，以指定的音色、时值、强度等、演奏出需要的音响。在录音系统中，还要控制记录下这些音响。MIDI 所适应的范围只是电声乐曲或模拟其他乐器的乐曲。

MIDI 技术的产生与应用，大大降低了乐曲的创作成本，节省了大量乐队演奏员的各项开支，缩短了在录音棚的工作时间，提高了工作效率。一整部乐曲的作曲、配器、录音等工作，只需要一位音乐编导、一位录音师即可将乐器作曲、配器、演奏，录音工作全部完成。

综上所述，在计算机中合成 MIDI 音乐的流程可以总结为：通过计算机音频卡上的 MIDI 接口，MIDI 电子乐器与计算机相连接；计算机通过音序器软件采集 MIDI 电子乐器、计算机键盘或鼠标发出的一系列指令，这一系列的指令可以存储为一个“. mid”的文件；然后通过音序器软件对 MIDI 文件进行编辑、修改；至此已完成 MIDI 乐曲的创作。在播放的时候，这些 MIDI 指令被送到音频卡上的合成器芯片处，由合成器芯片将 MIDI 指令符号进行解释并产生波形，然后通过声音发生器送到扬声器处播放出来。从输入音符到最后播放出声音，全程都是在计算机内部完成的。

3.6 声音构件的加工与制作

3.6.1 Windows 录音机

Windows 录音机是一套 Windows 操作系统自带的声音处理应用程序，可以从麦克风录制声音，并兼容大部分声卡。可以录制、混合、播放和编辑声音。可以向文件中添加声音、删除部分声音文件、更改回放速度、更改回放音量、更改声音的存储文件类型、添加回音等。可以使用不同的算法压缩声音。录音可以保存为 WMA 及 WAV 格式，一次可录 60s 甚至更长，音质可达 CD 标准。使用 Windows 录音机录音的方法如下。

1. 连接麦克风

麦克风的种类很多，人们常见的就是话筒和头戴式麦克风，一般在计算机上录音使用头戴式麦克风较为方便一些。当然，如果要录制出较好的效果，就必须使用专业的录音麦克风或者高档的头戴式麦克风。

2. 设置录音参数

特别注意，这项设置中，不同的声卡则设置有细微的差别。双击位于任务栏的声音控制图标，弹出“音量控制”窗口，选择“选项”→“属性”菜单项，在弹出的窗口中的“调整音量区”选择“录音”，然后在“显示下列音量控制”中选择“线性输入”，用于外部声音音频电流的输入。如果是使用麦克风录音，则必须选中 Microphone 选项。可以选择多个音量控制项。单击“确定”按钮后，音量控制窗口就出现各种录音方式的音量控制栏。这时用户就可以选择要使用的某种录音方式，然后再调节音量控制栏中的音量，用户可以根据自己的输入设备调节录音音量。另外，也可以调节左右声道的音量比例。

3. 开启 Windows 中的录音机

打开“开始”菜单，选择“程序”→“附件”→“娱乐”→“录音机”选项，打开 Windows 中自带的录音机程序。如果没有这一项，可以通过控制面板中的“添加”→“删除程序”来安装录音程序。

4. 设置 WAV 录音文件的格式

在录音程序“文件”菜单中选择“属性”选项，进行录音文件的格式设置。先在“录音位置”栏中选择“录音格式”，再单击“开始转换”按钮。在弹出窗口中的“选择声音”栏中选择“C 质量”即可。如果有特殊需要，可以按自己的要求选择其他的格式。为了避免将 WAV 格式压缩转换为 MP3 出现问题，尽量选择 16 位声音格式。

5. 设置录音质量

在录音程序“编辑”菜单中选择“音频属性”选项，然后在“录音”栏中选择高级属性，最后在弹出的窗口中调节“采样率转换质量”，一般情况下都可以选择“一般”选项，如果录制高质量的声音需要调节到“最佳”选项。

6. 开始录音

单击录音机程序界面中的“录音”按钮，然后打开录音机、收音机、随身听，或者对着麦克风讲话，录音程序即开始录制。用它来录制歌曲等意义不大，只能用来录制自己的一些短小的嘱咐话语，对声音文件的编辑和处理功能都比较简单。

3.6.2 GoldWave

GoldWave 是一个集声音编辑、播放、录制和转换的音频工具，体积小巧，功能强大。可以支持多种音频文件的编辑，包括 WAV、OGG、VOC、IFF、AIF、AFC、AU、SND、MP3、MAT、DWD、SMP、VOX、SDS、MOV 等音频文件格式，用户也可以从 CD、VCD、DVD 或其他视频文件中提取声音。用户使用 GoldWave 不仅可以对声音任意剪裁拼接，还可以对声音素材施加多普勒、回声、混响、降噪、变调等效果，把声音素材处理成想要的效果。

GoldWave 具有人性化的界面，可以定制操作界面，操作简便。在软件中可以同时打开和编辑多个文件。GoldWave 允许使用很多种声音效果，如倒转、回音、摇动、边缘、动态和时间限制、增强、扭曲等。软件具有精密的过滤器，如降噪器和突变过滤器等，可以修复声音文件。

GoldWave 是标准的绿色软件，不需要安装且体积小巧，将压缩包的几个文件释放到硬盘下的任意目录里，直接单击 GoldWave.exe 就开始运行了。

1. 导入素材

选择“文件”菜单的“打开”命令，指定一个将要进行编辑的文件，然后按 Enter 键。整个主界面从上到下被分为三大部分，最上面是菜单命令和快捷工具栏，中间是波形显示，下面是文件属性。用户的主要操作集中在占屏幕比例最大的波形显示区域内，如果是立体声文件则分为上下两个声道，可以分别或统一对它们进行操作，如图 3-9 所示。

2. 选择声波文件

要对文件进行各种音频处理之前，必须从整段音频中选择出需要的音频素材进行编辑。GoldWave 的选择方法很简单，利用鼠标的左右键配合进行，在合适的位置上单击鼠标左键可以确定选择部分的起始点，在结束位置上单击鼠标右键选择“设置结束标记”命令就可以

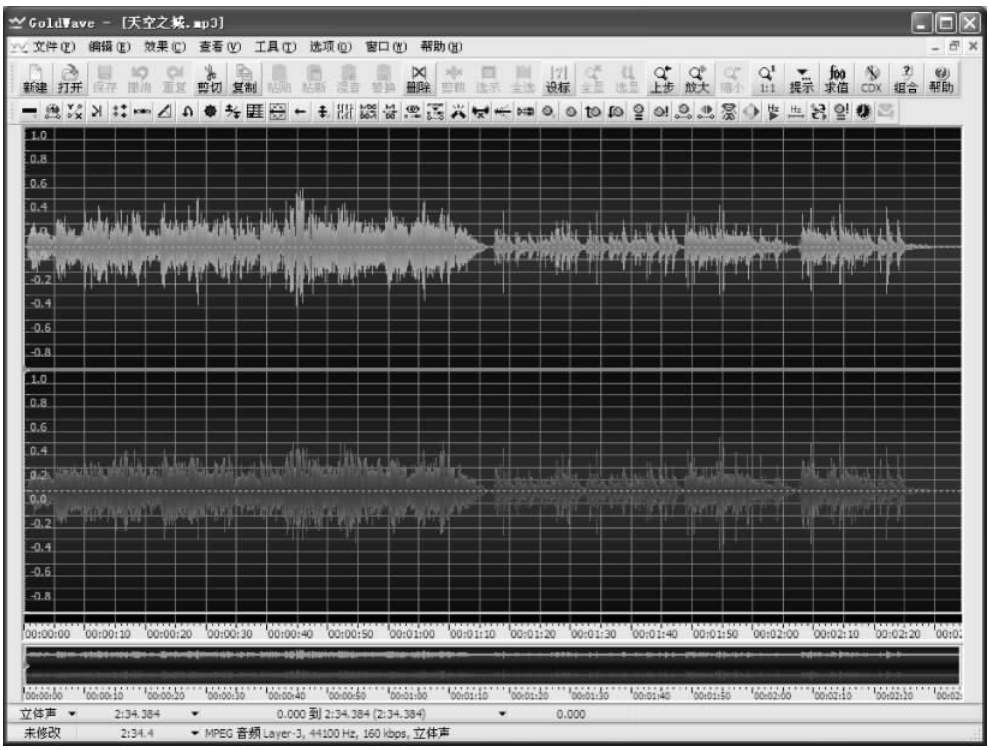

图 3-9　导入素材

确定选择部分的终止点，被选中的音频就将以高亮度显示，后续的所有操作都只会对这个高亮度区域施加作用，其他的阴影部分不会受到影响，如图 3-10 所示。

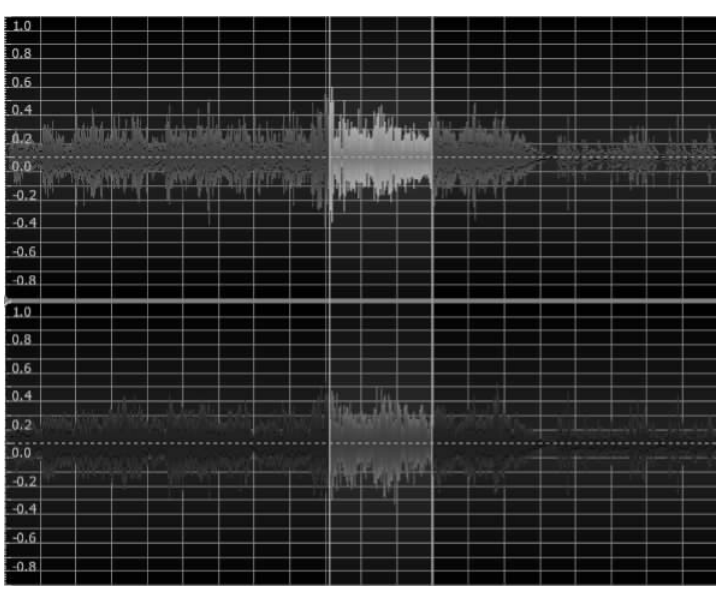

图 3-10　选择波形

3. 编辑操作

GoldWave 对音频的编辑与 Windows 其他应用软件一样，可以选择“编辑”菜单下的“剪切”、“复制”、“粘贴”、“删除”等基础操作命令实现相应的操作。GoldWave 的这些常用操作命令实现起来十分容易，要进行一段音频的“剪切”操作，首先要对剪切的部分进行选择，然后选择“编辑”→“剪切”命令，这段高亮度的选择部分就消失了，只剩下其他未被选择的阴影部分。然后使用“查看”命令，并重新指定指针的位置到将要粘贴的地方，用“编辑”→“粘贴”命令就能将刚才剪掉的部分还原出来。同理，可以使用“复制”、“删除”命令进行复制、删除操作。如果在删除或其他操作中出现了失误，用 Ctrl＋Z 组合键就能够进行命令撤销。

4. 时间标尺及状态显示

在波形显示区域的下方有一个指示音频文件时间长度的标尺，它以秒为单位，清晰地显示出任何位置的时间情况，通过标尺用户可以掌握音频处理时间、音频编辑的长短等，因此，在实际操作中要利用好参照标尺，辅助声音的编辑。打开一个音频文件之后，立即会在标尺下方显示出音频文件的格式以及它的时间长短，这就给制作者提供了准确的时间量化参数，根据这个时间长短来计划进行各种音频处理，往往会减少很多不必要的操作过程。有的音频文件太长，一个屏幕不能显示完毕，一种方法是用横向的滚动条进行拖放显示，另一种方法是改变显示的比例。在 GoldWave 中，用户通过滚动鼠标滚轮改变标尺的显示比例，或者使用“查看”菜单下的“放大”、“缩小”命令来完成，更方便的方法是使用 Shift＋↑组合键放大和用 Shift＋↓组合键缩小。如果想更详细地观测波形振幅的变化，那么就可以加大纵向的显示比例，方法同横向一样，使用“查看”菜单下的“垂直放大”、“垂直缩小”命令来完成，或者使用 Ctrl＋↑组合键放大，使用 Ctrl＋↓组合键缩小，这时会看到出现纵向滚动条，拖动它可以在纵向进行细致的查看，如图 3-11 所示。

图 3-11　时间标尺及状态显示

5. 声道选择

对于立体声音频文件来说，在 GoldWave 中的显示是以平行的水平形式分别进行的，上方表示左声道，下方表示右声道。有时在编辑中只想对其中一个声道进行处理，另一个声道要保持原样不变化，通过执行“编辑”→“声道”→“左声道”命令，可以选择立体声音频的左声道，后续所做的操作只会影响左声道的音频，另一个声道仍然以深色显示不受到任何影响，如图 3-12 所示。

6. 插入空白区域

在指定的位置插入一定时间的空白音频也是音频编辑中常用的一项处理方法，执行“编辑”→“插入静音”命令，在弹出的对话框中输入插入的时间，然后单击“确定”按钮，就可以在指针停留的位置插入一段空白的音频区域，如图 3-13 所示。

7. 制作典型音频效果

在 GoldWave 的“效果”菜单中提供了多种常用音频特效的命令，如滤波器、回声、偏移、

混响等，每一种特效都是日常音频制作领域使用最为广泛的效果，掌握它们的使用方法能够更方便地在动画制作、音效合成方面进行操作，得到令人满意的效果，如图 3-14 所示。

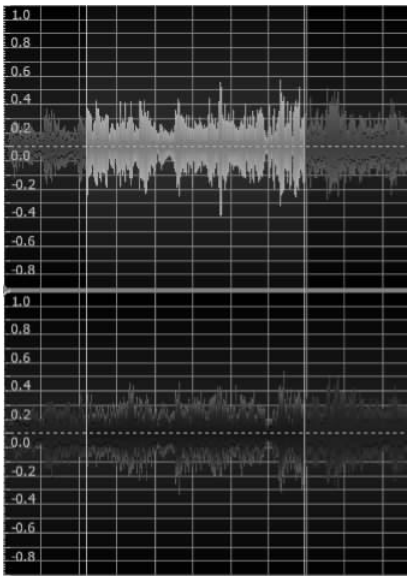

图 3-12　选择左声道

图 3-13　设置插入静音的时间

图 3-14　音频效果菜单

1）回声

当声音投射到距离声源有一段距离的大面积上时，声能的一部分被吸收，而另一部分声能要反射回来，如果听者听到由声源直接发来的声和由反射回来的声的时间间隔超过十分之一秒，它就能分辨出两个声音，这种反射回来的声就是回声。GoldWave 中的回声效果制作方法是执行“效果”→“回声”命令，在弹出的对话框中输入延迟时间、音量大小和反馈数值。延迟时间值越大，声音持续时间越长，回声反复的次数越多，效果就越明显。而音量控制是指返回声音的音量大小，这个值要控制得小一些，使回声的效果更接近真实。选中“产生尾声”复选框之后，能够使声音听上去更润泽、更具空间感。或者在“预置”下拉列表中选择回声的预设值，也可以达到很好的回声效果，如图 3-15 所示。

图 3-15　设置回声效果参数

2）压缩器/扩展器效果

在录制歌时，由于唱歌时气息、力度的掌握等原因，往往录制出来的效果有不令人满意的地方。有的语句用力过大，发音过强，甚至造成过载失真；有的语句却过于轻柔，造成信号微弱。如果对这些录音后的音频数据使用压缩器/扩展器效果就会在很大程度上减少这种情况的发生。压缩效果可以把高音压下来，把低音提上去，对声音的力度起到均衡的作用。在 GoldWave 中，用户可以执行“效果”→“压缩器/扩展器”命令，执行后弹出对话框。对相应的参数进行设置，在它的参数中最重要的是阈值的确定，它的取值就是压缩开始的临界点，高于这个值的部分就被以此值的比率进行压缩。选中对话框中的“使用平滑器”复选框，该项表示声音的润泽程度，在压缩过程中选择平滑器，可以获得良好的声音效果。也可以直接使用该对话框中的预置值，如图 3-16 所示。

3）镶边效果

使用镶边效果能在原来音色的基础上给声音再加上一道独特的“边缘”，使其听上去更有趣、更具变化性，以满足影视声音中一些特殊的用途。执行“效果”→“镶边器”命令，设置镶边的作用效果主要由镶边和频率两项参数决定，改变它们的取值就可以得到很多意想不到的奇特效果。也可以在对话框中选择镶边器的预置值，获得更多的镶边效果，如图 3-17 所示。

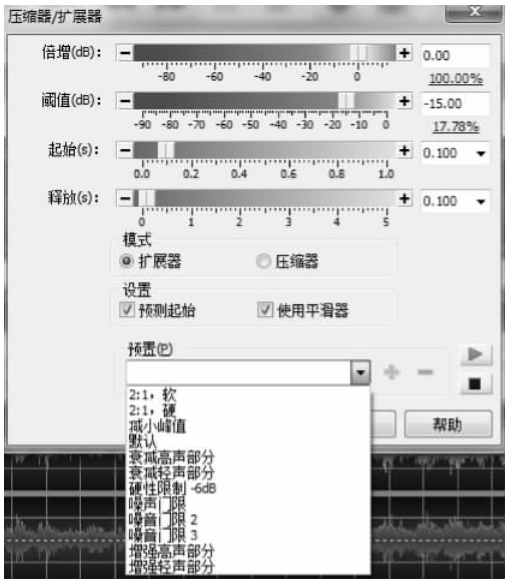

图 3-16　设置压缩器/扩展器效果参数

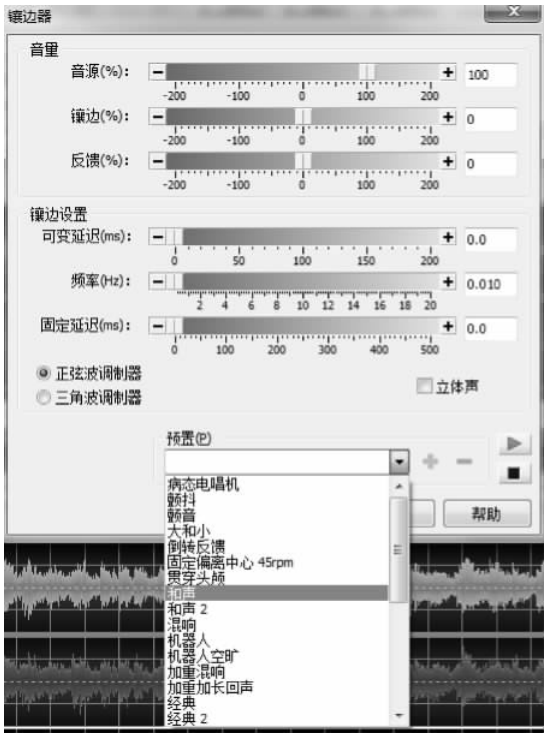

图 3-17　设置镶边效果参数

4）均衡器

均衡调节也是音频编辑中一项十分重要的处理方法，它能够合理改善音频文件的频率结构，达到理想的声音效果。执行“效果”→“均衡器”命令，弹出“均衡器”对话框后，最简单

快捷的调节方法就是直接拖动代表不同频段的数字标识到一个指定的大小位置，声音每一段的增益不能过大，以免造成声音的过载失真，如图 3-18 所示。

5）音量效果

GoldWave 的“音量”效果子菜单中包含了自动增益、更改音量、淡出、淡入、匹配音量、最佳化音量、外形音量等命令，满足制作者各种调节音量的需求。改变音量大小命令是直接以百分比的形式对音量进行提升或降低的，其取值不宜过大。最佳化音量命令既不会使声音过载，又在最大范围内提升了音量。它是 GoldWave 中很实用的一个命令，一般在歌曲刻录 CD 之前都要做一次音量最佳化的处理。如果想对不同位置的音频进行不同的音量变化就必须使用“外形音量”命令，打开“外形音量”对话框后直接用鼠标添加、调整音量点的位置，即可根据需要调节声音的音量，如图 3-19 所示。

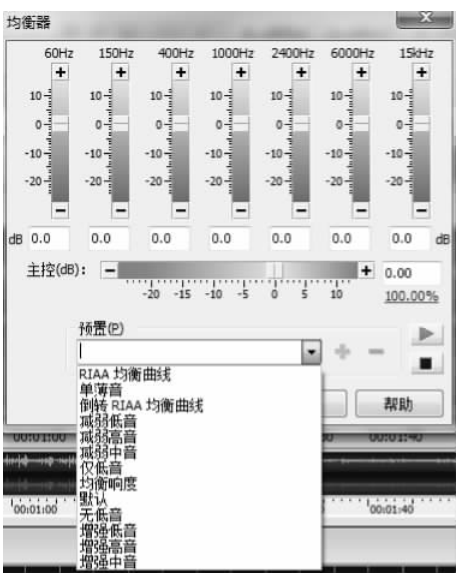

图 3-18　设置均衡器参数

图 3-19　音量效果菜单

6）声像效果

声像效果是指控制左右声道的声音位置并进行变化，达到声像编辑的目的。GoldWave 的声像效果中，交换声道位置和声像包络线最为常用。交换声道位置就是将左右声道的数据互换。而声像包络线与音量包络线非常类似，能够更灵活地控制不同地方的不同声像变化。可以选择“声像”对话框中的预置值，进行声像的设置，如右声道到左声道、整个左声道等，如图 3-20 所示。

当然除了上面介绍的几种效果之外，GoldWave 还可以制作反向、静音、倒转、偏移等效果，它们的使用方法都非常简单，反复实验和使用就可以找到用户满意的声音效果。

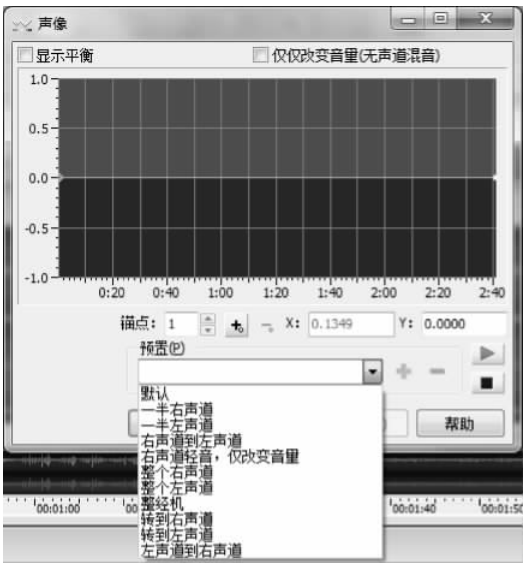

图 3-20　设置声像效果参数

8. 其他实用功能

GoldWave 除了提供丰富的音频效果制作命令外，还可以进行 CD 抓音轨、批量格式转换、多种媒体格式支持等非常实用的功能。批处理命令可以把一组声音文件转换为不同的格式类型。在对话框中添加要转换的多个文件，并选择转换后的格式和路径，然后按下“开始”键。该功能可以根据需要转换声音的音质，在批处理命令选项中提供了很多声音转换的预置值，如转换立体声为单声道，转换 8 位声音到 16 位声音，或者是文件类型支持的任意属性的组合。还可以把原有的声音文件压缩为 MP3 的格式，在保持出色的声音质量的前提下使声音文件的存储容量缩小为原有尺寸的 1/10 左右。批处理命令还可以为一组声音添加特效、添加组合、添加编辑。批处理命令如图 3-21 和图 3-22 所示。

图 3-21　批处理添加效果窗口

使用 GoldWave 的“工具”→“CD 读取器”命令，可以将 CD 音乐复制为一个声音文件。为了缩小尺寸，也可以把 CD 音乐直接提取出来并存储为 MP3 等压缩格式。

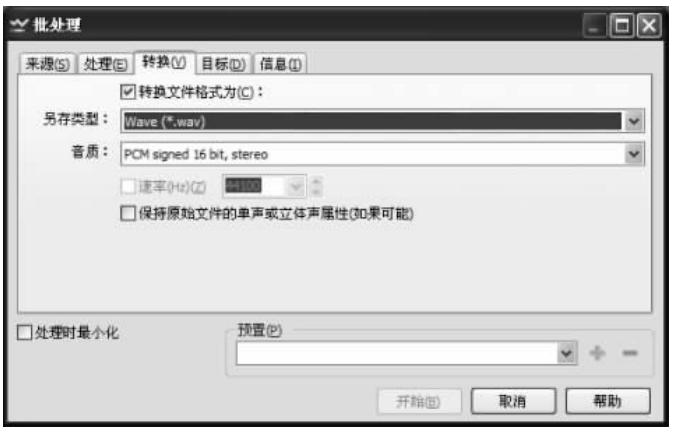

图 3-22　批处理转换窗口

9. 支持多种媒体格式

在 GoldWave 的“打开声音文件”对话框中就可以发现，除了支持基础的 WAV 格式外，它还可以直接编辑 MP3 格式、AIFF 格式甚至是视频 MPG 格式的音频文件。这样用户就不用在各种格式之间来回转换了，方便了用户的操作。GoldWave 的多种媒体格式支持带给了用户更高的工作效率，如图 3-23 所示。

图 3-23　选择存储格式

3.6.3　Cool Edit

Cool Edit 是一个集录音、混音、编辑于一体的多轨数字音频编辑软件。用户使用 Cool Edit 可以对音频添加多种特效，如放大、降低噪声、压缩、扩展、回声、失真、延迟等效果。还可以同时处理多个音频文件，轻松地在几个文件中进行剪切、复制、粘贴、合并、重叠声音操作。使用它可以生成噪声、低音、静音、电话信号等声音。该软件还包含有 CD 播放器。另外，它还可以在 AIF、AU、MP3、RAW PCM、SAM、VOC、VOX、WAV 等文件格式之间进行转换，并且能够保存为 RealAudio 格式。

它可以在普通声卡上同时处理多达 128 轨的音频信号，具有极高的编辑精度，极其丰富的音频处理效果，还支持视频文件的回放和混缩，并能进行实时预览和多轨音频的混缩合成，是音频处理的优秀软件。

使用 Cool Edit 编辑音乐就像音响工程师一样。用户可以使用它记录自己的音乐、声音等，可以编辑它，可以让它与其他的声音或音乐进行混缩，然后添加反转、合唱团效果，还有回响等增强效果。制作完成的音乐可以保存成常用的各种音乐格式，可以在软件内部非常方便地把歌曲烧录成 CD，也可以存储为便于网络传输的音频格式，放置在网页上或使用电子邮件进行传递。

在 Cool Edit 软件中可以先后完成声音采集、声音编辑、添加音效、混缩、输出等工作。下面以制作一段有背景音乐的诗歌朗诵为例，介绍 Cool Edit 的使用。首先录制朗诵诗歌的声音，然后对录制的声音进行降噪、声音标准化等编辑工作，再添加音效，导入背景音乐与录音进行合成，最后进行作品的输出。

1. 声音采集

首先，在计算机的声卡上连接好话筒设备。执行“文件”→“新建”命令，弹出“新建波形”对话框，选择适当的采样率、录音声道和采样精度。在这里选择 CD 音质，即 44100Hz，立体声，16 位的采样精度。单击“确定”按钮，就会新建一个波形文件，如图 3-24 所示。

单击 Cool Edit 主窗口左下部的红色“录音”按钮，开始录音，如图 3-25 所示。

图 3-24　“新建波形”窗口

图 3-25　音频控制按钮

对准话筒开始朗诵诗歌，完成录音后，单击 Cool Edit 主窗口左下部的“停止”按钮结束录音。此时，在 Cool Edit 窗口中将出现刚录制的诗歌的波形图。单击左下部的“播放”按钮，可以播放它。如果波形图是一条直线或者波形不够明显，放音时将没有声音或声音很小，那么要检查音源选择是否正确、录音电平是否设置值太低。

在录制完诗歌音频后，先将这段音频文件保存下来，以便将来制作不同的效果处理。选择软件界面左上角的“切换为波形编辑界面”按钮，或者在左侧“文件”窗口中双击录制的音频文件切换至录制音频音轨，选择菜单栏中的“文件”→“另存为”选项，在弹出的对话框中选择合适的文件目录及文件名，保存类型选择无损压缩的 WAV 文件。

2. 声音编辑

首先对录制的声音进行降噪的处理，降噪的方法有很多如采样、滤波等，其中效果最好的是采样降噪法。把刚才录制的诗歌声音波形进行放大，将噪声区内波形最平稳且最长的一段选中，一般选择没有音乐信号的间隔处。然后选择菜单栏中的“效果”→“噪声消除”→“降噪器”命令，弹出“降噪器”对话框，单击“噪声采样”按钮，对噪声进行分析，然后单击对话框中的“确定”按钮，再在波形图中双击，选中整个波形，再次选择“效果”→“噪声消除”→“降噪器”命令，然后单击“确定”按钮，就可以对整个声音波形进行降噪操作。这样就可以达到良好的降噪效果，如图 3-26 所示。

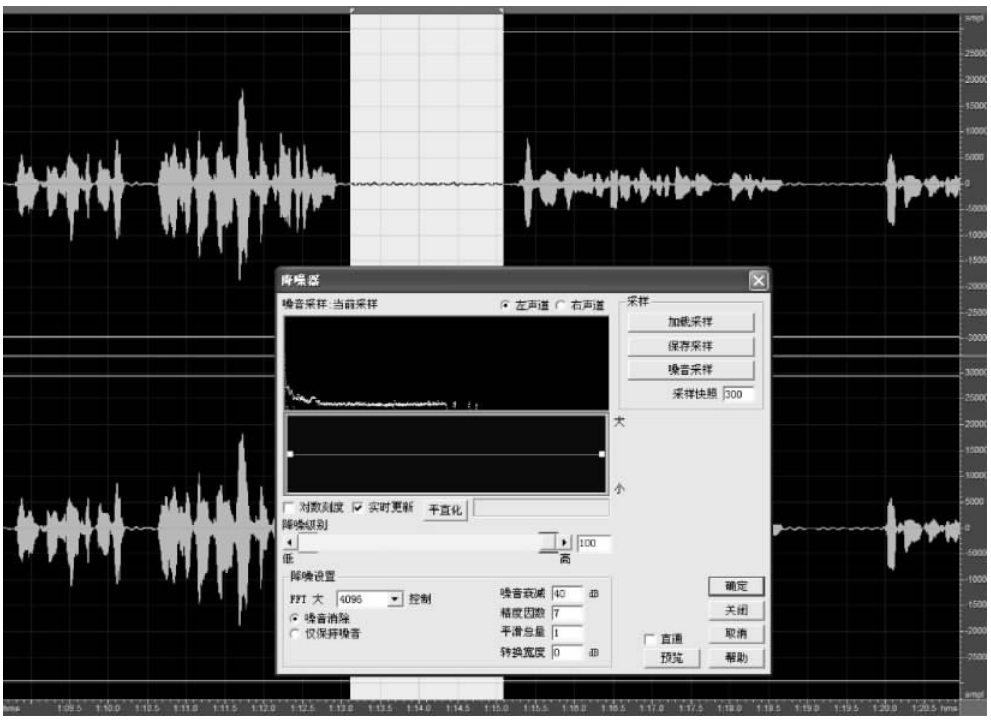

图 3-26　噪声采样窗口

3. 多轨窗口编辑

为录制的诗歌加入配乐，并在多轨窗口中进行编辑。单击左上角的“切换”按钮回到单轨编辑窗口。选择菜单“文件”→“打开波形”命令，在对话框中选择一首合适的音乐，支持 WAV、MP3 甚至 CD 音轨等很多格式，在选定音乐之前，单击“播放”按钮可以试听音乐。也可以提取视频文件中的配音，支持 AVI、MPEG 等格式。在打开的音乐波形上单击鼠标右键，执行“插入到多轨中”命令，将它插入多轨窗口。插入到多轨的波形将被自动放置在空缺的最上面一个轨道。

单击软件界面左上角的“单轨/多轨切换”按钮，打开 Cool Edit 的多轨编辑视窗。在软

件界面左侧的“文件”窗口中，右键单击前面录制的声音文件，执行“插入到多轨中”命令，那么声音文件就会插入多轨视窗的第一轨。再选择音乐“天空之城.mp3”放置在“文件”窗口中，作为诗歌的配乐，右键单击音乐素材“天空之城.mp3”，执行“插入到多轨中”命令，那么音乐素材文件就会插入到多轨视窗的第二轨。在多轨视窗中完成声音导入之后，单击“播放”按钮，听一下初步的合成效果，如图 3-27 所示。

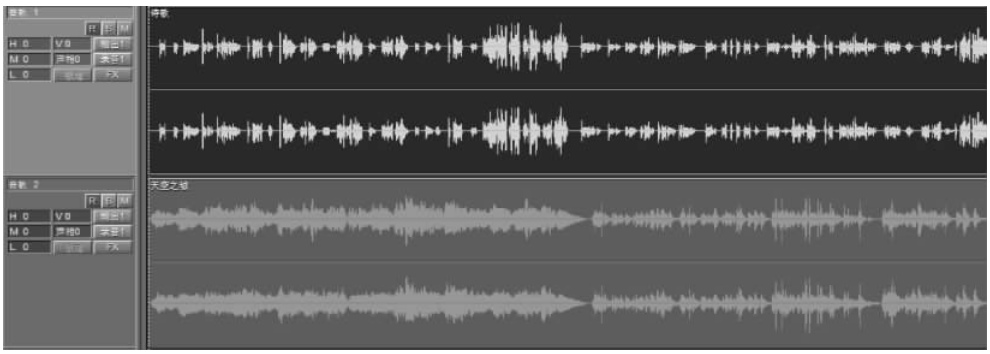

图 3-27　多轨编辑窗口

Cool Edit 的多轨窗口有一条黄色的垂直亮线，播放音频时，随着它的移动，作用于经过的所有轨道。在波形上按住鼠标右键不放向两边拖动轨道上的波形，可以改变波形在轨道上的位置。也可以上下拖动，移至其他轨道。轨道的左侧按钮中，有 3 个带有 R、S、M 标识的按钮，分别代表录音状态、独奏和静音，可以按照需要选用与取消对此轨道的作用。3 个按钮左侧还有音量与声像的选取项，可以单击鼠标右键打开控制推杆，调整该轨道的音量或者声像值。在此保持声像不变，降低背景音乐的音量，以突出朗诵的诗歌声音。在第二条轨道上调整背景音乐的音量为“－15dB”。多轨窗口中控制声像和音量，如图 3-28 所示。

(a) 声像　　(b) 音量

图 3-28　声像和音量控制

4. 添加音效

在 Cool Edit 中可以随意添加音效。用户可以方便地制作出各种专业、迷人的声音效果。例如，可以产生音乐大厅的环境混响效果，可以根据录音电平动态调整输出电平，能够

在不影响声音质量的情况下，改变乐曲音调或节拍等。通过添加音效调整录制的声音，使其更加悦耳。

1）音量量化

录制好的声音波形可能过小或者过大。过大就会造成波形上下两边特别整齐，这表明已经大于0dB，形成了“消峰”失真，虽然有工具可以修整这种“消峰”的现象，但也要尽量地避免。使用音量控制效果器，先选取波形，在选择菜单“效果”→“波形振幅”→“渐变”命令，打开“波形振幅”对话框。左上角有两个标签，分别是“恒量改变”与“淡入/出”两个选项卡，使用右侧的预置窗口的现成效果完成操作。在之前录制的诗歌音频中，整体音量偏小，要全选音频波形，调整整个声波的音量，选取右侧的一个预置效果，在此选择“6dB Boost”选项，也就是提升音量。可以单击右下侧的“预览”按钮监听效果，满意后，单击“确定”按钮。淡入/淡出的效果与之相似，选取开头或结尾的约5秒以下的波形片段，再在“预置”效果中选择Fade In和Fade Out选项，为录制的声音设定淡入/淡出效果。试听效果满意后，单击“确定”按钮，如图3-29所示。

图3-29　设置波形振幅参数

2）音调调整

可以对录制的声音做一些润色。例如，如果声音的音调较低沉，可以提升高音使它变得更清晰，如果声调偏高，可以将它调整的柔和悦耳。选取波形局部或全部，选择菜单“效果”→“滤波器”→“图形均衡器”命令，就会打开“图形均衡器”对话框。对话框中有3个标签，分别是10段均衡、20段均衡、30段均衡，任选其一，做适当的调整，单击“预览”按钮试听效果，一边试听一边调整。满意后，单击“确定”按钮，如图3-30所示。

3）美化声音

经过上一步的调整，声音已经很清晰了，可是声音的音色不够生动。在菜单“效果”→“常用效果器”命令中，分别有合唱、延迟、动态延迟、混响、房间混响、回声等。可选中一部分波形打开相应对话框后，按下“预览”按钮试听效果，可以边试听边调整。采用“预置”选项窗口中提供的现成效果，可以直观地调整声音效果。用效果器为自己的声音加上恰当的诸如回声、混响等效果，使制作的声音不显得太干，变得更加圆润，如图3-31所示。

图 3-30　设置图形均衡器参数

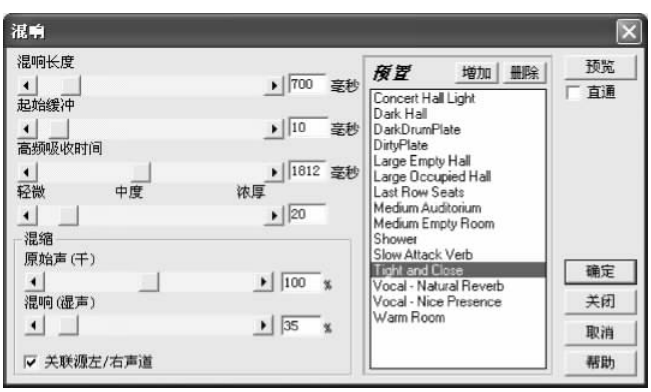

图 3-31　设置混响参数

5. 多轨混音

此时，在多轨窗口中，应该有两轨波形，第一轨是录制的诗歌波形，第二轨是“天空之城.mp3”配乐。下面要调整两条声音轨道的音量比例。使它们之间声音主体突出，音乐不喧宾夺主。开始时应该先让音乐响起，以营造出气氛，确定作品悠扬的基调。然后让朗诵诗歌的声波淡入，诗歌的声波出现前，让音乐渐弱至某一个恒定的音量，等到诗歌朗诵完毕时，音乐的音量回到正常。依次选取“查看”→“音量包络”和“查看”→“声像包络”命令。在每条音轨的上部将出现亮绿色的音量控制包络线和位于中间的亮蓝色的声像控制包络线。用鼠标可对各个轨道声波的局部或者全部，进行音量与声像的控制。单击第二条轨道任意的空白处选中此轨，在声波起始和结束前的音量控制线上，单击鼠标加入控制点，上下拖动控制点，以减弱音量，音量百分比同时会显示出来。依次类推，在需要恢复音量的部位，做类似操作，使音量复原。也可以对声像做相应的调整，使被修改的部位听起来声音是来自不同的方位。控制点一般由两个组成，一个是执行开始，一个是执行结束。按键盘上 HOME 键，使

播放头回到左侧起点位置，再按“播放”键试听两轨合成效果，确定没有问题之后，选择“文件”→“另存为”命令保存工程文件，文件名及格式为“配乐诗朗诵.ses”，如图 3-32 所示。

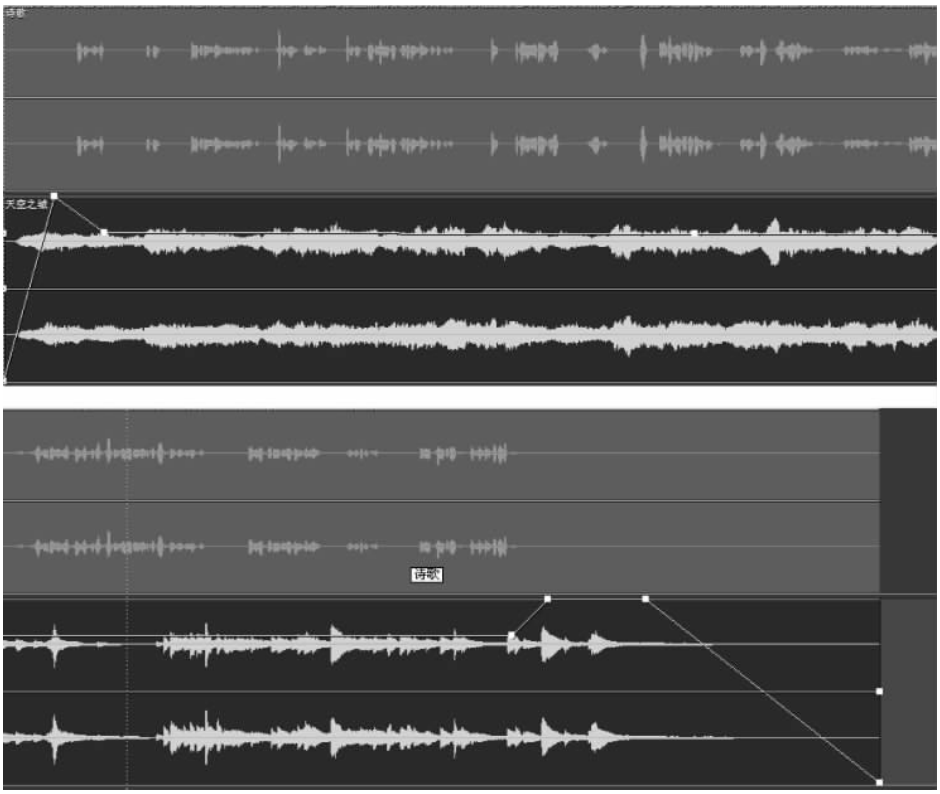

图 3-32　编辑配乐音量包络

6. 合成输出

全部调整完毕以后，进行最后一次试听，因为这时还是多轨的 WAV 格式，不便于保存与传输，要把它变成所需要的音频格式。执行“文件”→“混缩另存为”命令，Cool Edit 会把若干轨道变为只具有左右声道，同时包含之前经过编辑操作的所有声音波形的一个文件。可以选择保存多种音频格式，如 MP3，ASF，AIF 等，要混缩为 WAV 格式时，为保证通用性，一般选用 Windows PCM 的 WAV 格式。

Cool Edit 是一款操作简单且易学的音频编辑软件，但是要真正做出优质的音响效果还需要在应用中不断地积累经验。

3.7　本章小结

本章讲解了声音创作的传统理论和数字化的制作平台，数字音频技术可以把声音信号数字化，并在数字状态下进行记录，并进一步加工处理，而且还可以在数字化的平台上制作强大的声音特效。这样利用声音制作软件的强大功能在小工作室就可以完成影视声音的制作。本章还以实例的形式重点讲解了 GoldWave、Cool Edit 等主流音频制作软件。

习题 3

1. 简述声音的三要素。
2. 简述立体声的概念。
3. 简述数字音频信号的特点。
4. 简述 MIDI 音乐的概念。
5. 试着使用 Cool Edit 软件对一段会议发言录音进行降噪处理。