第3章 数字音频处理技术

声音是人类认识自然和进行交流的主要媒体形式，通常的声音主要是指语音、自然声和音乐。如何将声音数字化转换成数字音频，更加方便地进行传输、存储和处理成为多媒体研究的一个重要领域。数字音频信号的处理主要表现在数据采样和编辑加工两个方面。其中，数据采样的作用是把自然声转换成计算机能够处理的数字音频信号；对数字音频信号的编辑加工则主要表现在剪辑、合成、静音、增加混响及调整频率等方面。

3.1 数字音频概述

声音是指通过一定介质（如空气和水等）传播的一种连续波，其本质是机械振动或气流扰动引起周围弹性媒质发生波动的现象，它是一个随着时间连续变化的模拟信号，在物理学中称为声波。声波具有普通波所具有的特性，即反射（Reflection）、折射（Refraction）和衍射（Diffraction），它有如下几个重要指标，如图 3-1 所示。

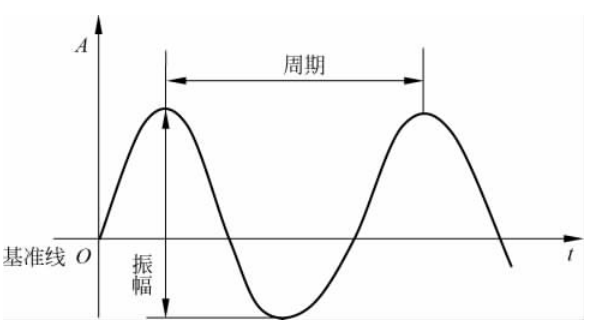


图 3-1　声音关键指标

（1）基准线：提供模拟信号的基准点。

（2）振幅（Amplitude）：波的高低幅度，表示声音的强弱。

（3）周期（Period）：两个相邻波之间的时间长度。

（4）频率（Frequency）：每秒钟振动的次数，以 Hz 为单位。

通常人耳听力频率范围是 20～20 000Hz，如果物体振动频率低于 20Hz 或高于 20 000Hz 人耳就听不到了，高于 20 000Hz 频率的称为超声波，而低于 20Hz 频率的称为次声波。

3.1.1 声音的基本特点

1. 声音传播方向

声音依靠介质的振动进行传播。声源实际上是一个振动源，它使周围的介质（空气、液体、

固体)产生振动,并以波的形式进行传播,人耳如果感觉到这种传播过来的振动,反映到大脑,就意味着听到了声音。

声音以振动波的形式从声源向四周传播,人类在辨别声源位置时,首先对声音到达左、右两耳的微小时间差和强度差异进行辨别,然后经过大脑综合分析而判断出声音来自何方。从声源直接到达人类听觉器官的声音叫作"直达声",直达声的方向辨别最容易。现实生活中,森林、建筑、各种地貌和景物存在于人们周围,声音从声源发出后,需要经过多次反射才能被人们听到,这就是"反射声"。就理论而言,反射声会影响方向的准确辨别。但实际中,反射声不会使人丧失方向感,起关键作用的是大脑的综合分析能力。经过大脑的分析,不仅可以辨别声音的来源,还能丰富声音的层次,感觉声音的厚度和空间效果。

2. 声音的三要素

声音的三要素为音调、音色和音强。就听觉特性而言,这三者决定了声音的质量。

(1) 音调:代表声音的高低。音调与频率有关,频率越高,音调越高,反之亦然。当人们提高唱盘的转速时,声音频率提高,音调也提高。当使用音频处理软件对声音进行处理时,频率的改变可造成音调的改变。

(2) 音强:代表声音的强度,也称为"响度","音量"是指音强。音强与声波的振幅成正比,振幅越大,强度越大。CD 音乐盘、MP3 音乐及其他形式的声音强度是一定的,可以通过播放设备的音量控制改变聆听的响度,也可使用音频处理软件改变声源的音强。定量描述声音强弱的方式有多种,声压和声压级就是其中的两种形式。声压是指在声场中某处由声波引起的压强的变化值,用 P 表示,单位是"帕斯卡(Pa)"。声压越大,声音也就越大。由于人耳对声音强弱的感觉并不与声压的大小成线性关系,而是大体上与声压有效值的对数成正比,因此为了适应人类听觉的这一特性,通常对声压的有效值取对数,用其对数值来表示声音的强弱即声压级,用 SPL 表示,单位为分贝(dB),表达式如下:

$$\mathrm{SPL} = 20\lg \frac{P_{\mathrm{rms}}}{P_{\mathrm{ref}}}$$

式中 20 为参考常量,P_{rms} 是计量点的声压有效值,P_{ref} 是人为定义零声压级的参考声压值,国际协议规定 $P_{\mathrm{ref}} = 2\times10^{-5}$(帕),这是大多数具有正常听力的年轻人刚刚能察觉到的 1kHz 单一频率信号(称为简谐音)存在时的声压值。

(3) 音色:它与声波的形状有关,是由混入基音的泛音决定的。通常声音分为纯音和复音两种类型。纯音是指振幅和周期均为常数的声音,一般只会出现在专用的电子设备中。复音是具有不同频率和振幅的混合音,大自然中的声音大部分是复音。复音中最低的频率称为基频,即"基音",它是声音的基调。其他频率复音称为"谐音",也叫泛音。复音中的基频和谐音决定了复音的音质和音色。各种声源都有自己独特的音色,如各种乐器、不同的人和各种生物等,即使在同一音高和同一声音强度的情况下,人们也能根据音色辨别声源种类。

3. 声音的频谱与质量

声音的频谱有线性频谱和连续频谱之分。线性频谱是具有周期性的单一频率声波;连续频谱是具有非周期性的、带有一定频带的所有频率分量的声波。纯粹的单一频率的声波只能在专门的设备中创造出来,声音效果单调而乏味。自然界中的声音几乎全部属于非周期性声波,这种声波具有广泛的频率分量,听起来声音饱满、音色多样且具有生气。

声音的质量简称"音质",音质的好坏与音色和频率范围有关。悦耳的音色、宽广的频率范围(带宽),能够获得非常好的音质。

4. 声音的连续时基性

声音在时间轴上是连续信号,具有连续性和过程性,属于连续时基性媒体形式。构成声音的数据前后之间具有强烈的相关性。除此之外,声音还具有实时性,这对处理声音的硬件和软件提出了很高的要求。

3.1.2 模拟录音

目前记录声音主要有两种技术,即模拟录音技术与数字化音频技术。

模拟磁性录音技术在数字化音频技术以前已使用多年,这一技术被广泛地用于采集和播放各种各样的声音,如音乐、配音及特殊的声音效果,至今在某些领域还被广泛应用。模拟磁性录音过程就是声→电→磁的转换过程。以录音机为例,其工作过程如图 3-2 所示。

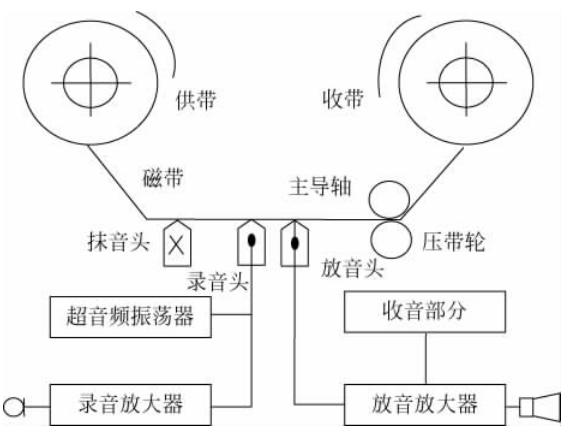


图 3-2 录音机工作过程

这种模拟录音方式是直接记录音频信号的波形,重放时用唱针扫描唱片槽纹或者用放音磁头拾取信号。模拟磁性录音性能受电磁性能的影响较大。磁带的频率特性微小的变化就会对音质产生影响。目前模拟录音动态范围可达 80dB。若进一步提高录音、放音的音质,需借助于数字音频技术。

3.1.3 音频数字化

将时间上连续的模拟音频(自然声或其他种类的声音)转换成时间上不连续的数字音频的过程称为音频的数字化。只有将模拟音频转换为标准数字音频信号,计算机才能进行处理。音频的数字化过程包括采样、量化和编码三大步骤。音频的数字化过程所用到的主要硬件设备便是模拟/数字转换器(Analog to Digital Converter,ADC)。

音频数字化与音频磁记录对于声源产生模拟电信号的捕获方式相同,所不同的是对这种捕获后的电信号的处理方式。音频数字化处理中,并不是利用磁头及磁头线圈进行相关的处理,而是利用硬件按照固定的时间间隔截取该音频电信号的振幅值,振幅值采用若干位二进制数表示,从而将模拟声音信号变成数字音频信号,这样就将连续变化的振动波的模拟声音信号转化为阶跃变化的离散的数字音频信号,如图 3-3 所示。

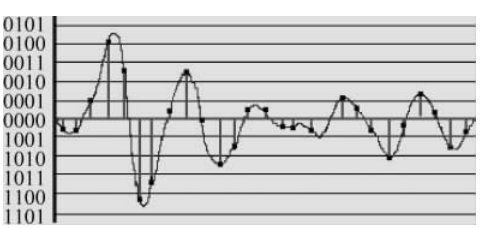


图 3-3 采样过程

截取模拟声音信号振幅值的过程称为“采样”，得到的振幅值称为“采样值”。采样值用二进制数的形式表示，称为“量化编码”。具体的实现过程如下：

1. 采样

根据傅里叶定理，只要在连续的信号量上等间隔地取足够多的“点”，就能逼真地模拟出原来的连续量，这个取点的过程称为“采样”。每秒钟所抽取的模拟音频幅度的样本次数称为采样频率，单位为 Hz(赫)，通常使用 kHz(千赫)，即 1kHz＝1000Hz。采样频率的高低决定了声音失真程度的大小，采样精度越高(“取点”越多)，数字声音越逼真，音质就越好。当然，采集的样本数量越多，数字化声音的数据量也越大。如果为了减少数据量而过分降低采样频率，音频信号增加了失真，音质就会变得很差。采样过程如图 3-4 所示。

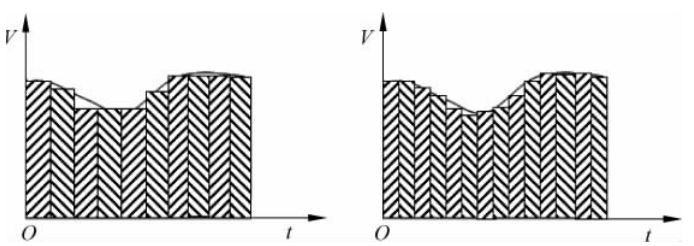


图 3-4　采样过程

采样频率的高低是根据奈奎斯特理论(Nyquist Theory)和音频信号本身的最高频率决定的。奈奎斯特理论指出，采样频率不应低于音频信号最高频率的两倍，这样就能把以数字表达的音频还原成原来的音频，这叫作无损数字化(Lossless Digitization)。采样定律用公式表示为：

$$f_s \geqslant 2f \quad \text{或者} \quad T_s \leqslant T/2$$

其中 f 为被采样信号的最高频率。可以这样来理解奈奎斯特理论，例如音频信号可以看成由许多个正弦波组成的，则振幅为 A、频率为 f 的正弦波至少需要两个采样样本表示，因此音频数据的采样频率 $f_{采样}$ 与声音还原频率 $f_{还原}$ 的关系就可以表示如下：

$$f_{采样} = 2 \cdot f_{还原}$$

目前经常用到的采样频率有 11.025kHz、22.05kHz、44.1kHz、48kHz 等。例如，人耳的可听频率范围为 20Hz～20kHz，根据奈奎斯特采样定理，为保证声音不失真，采样频率至少应保证不低于 40kHz。此外，由于每个人听力范围不同，20Hz～20kHz 只是一个参考范围，因此还要留有一定余地，所以 CD 音频通常采用 44.1kHz 的采样频率，这样的采样频率可以保证即使是采样 22.05kHz 的超声波也不会产生失真。

2. 量化

如图 3-5 所示，把整个声波振幅划分成有限个小等份，每一个小等份赋予一个相同的值，则每个采样点可用振幅的等份数量来描述精度。这些等份值在计算机中用若干位二进制数来表示，这一过程称为量化。从图 3-5 中可以看出采样后的离散值用二进制表示要损失一些精度，量化级别越多，损失越少，音质就越好，声音就越清晰。量化级别是用量化位数表示每个采样点能够表示的数据范围，常用的有 8 位、12 位、16 位、24 位、32 位甚至是 64 位等。要注意的是，8 位(1 个字节)不是说把纵坐标分成 8 份，而是分成 2^8＝256 份。同理，16 位是把纵坐标分成 2^{16}＝65 536 份。通常 16 位的量化级别足以表示从人耳刚能听到的最细微的声音到无法忍受的巨大的噪音这样的声音范围了。无论量化精度有多高，量化过程必然会产生一定的噪音，这个称为量化噪音。但只要选择适当的量化精度，量化噪音就可以控制在人耳感觉不出来

的范围内。

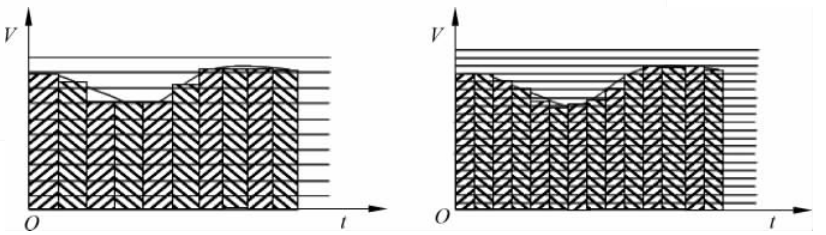


图 3-5 量化过程

采样精度的另一种表示方法是信号噪声比,简称为信噪比(Signal-To-Noise,SNR)。狭义来讲,是指放大器的输出信号的电压与同时输出的噪声电压的比,常常用分贝数表示。一般来说,信噪比越大,说明混在信号里的噪声越小,声音回放的音质量越高,否则相反。信噪比一般不应该低于 70dB,高保真音箱的信噪比应达到 110dB 以上。用下式计算:

$$\mathrm{SNR} = 10\lg[(V_{\mathrm{signal}})^2/(V_{\mathrm{noise}})^2] = 20\lg(V_{\mathrm{signal}}/V_{\mathrm{noise}})$$

其中,V_{signal}表示信号电压,V_{noise}表示噪音电压,SNR 的单位为分贝(dB)。

例如,假设 $V_{\mathrm{noise}}=1$,采样精度为 1 位表示,则 $V_{\mathrm{signal}}=21$,SNR=20lg2≈6dB。

再如,假设 $V_{\mathrm{noise}}=1$,采样精度为 16 位表示,则 $V_{\mathrm{signal}}=2^{16}$,SNR=20×16lg2≈96dB。

3. 编码

采样与量化后的二进制音频数据需要按一定的规则进行组织,以便于计算机进行处理,这就是编码。最简单的编码方案是直接使用二进制的补码表示,也称为脉冲编码调制(Pulse Code Modulation,PCM),它属于非压缩编码。在多媒体计算机中用这种编码方法存储的未压缩的音频数据文件大小可用下面公式来计算:

文件存储量(B)=时间(s)×采样频率(Hz)×采样精度(b)×声道数/8

4. 声道数

声道数是声音通道的个数,指一次采样的声音波形个数。单声道一次采样一个声音波形,双声道(立体声)一次采样两个声音波形,双声道比单声道多一倍的数据量。

3.2 声音的输出与识别

随着计算机科学技术的发展,人们已不再满足于仅仅通过键盘和显示器与计算机交互信息。让计算机能听懂人说的话,或者用语音控制各种自动化系统,即用人类最直接、最方便的交换信息形式——语言来与计算机进行通信,这是人类一直以来的梦想。针对此,诞生了一门新的学科——计算机语音学(Computer Phonetics)。人们对于计算机语音学的研究主要包括以下几个方面:语音编码(Speech Coding)、语音合成(Speech Synthesis)、语音识别(Speech Recognition)、语种识别(Language Identification)、说话人识别(Speaker Recognition)或说话人确认(Speaker Verification)等。

3.2.1 语音输出

实现计算机语音输出通常有两种方法:一是录音/重放,二是文字→语音的转换。若采用第一种方法,首先要把模拟语音信号转换成数字序列,编码后暂存于存储设备中(录音)。需要

时再经解码,重建声音信号(重放)。录音/重放可获得高音质声音,并能保留特定人或乐器的音色,但所需的存储容量随发音时间线性增长。第二种方法是基于声音合成技术的一种声音产生技术,它可用于语音合成和音乐合成。文字→语言转换是语音合成技术的延伸,它能把计算机内的文本转换成连续自然的语声流。若采用这种方法输出语音,应预先建立语音参数数据库和发音规则库等。需要输出语音时系统按需求先合成出语音单元,再按话音学规则或语言学规则连接成自然的话流。文字→语言转换的参数库不随发音时间增长而加大,但规则库却随语音质量的要求而增大。基于语音合成技术的方法众多,根据不同分类标准有不同的合成方法,从研究技术上来分,有发音参数合成、声道模型参数合成和波形编辑合成;从合成策略上来分,有频谱逼近和波形逼近方法。

3.2.2 语音识别

语音识别技术,也称为自动语音识别(Automatic Speech Recognition,ASR),其目的是要将人类语音中的词汇内容转换为计算机可读取的输入,如按键、二进制编码或者字符序列等。这与说话人识别及说话人确认不同,因为后者并不关心语音中所包含的词汇内容,它只对发出语音的说话人进行识别或确认。一般来说,语音识别技术包括语音拨号、语音导航、室内设备控制、语音文档检索和简单的听写数据录入等。在实际应用中,它往往与机器翻译和语音合成技术等自然语言处理技术相结合,以解决具体需求,如语音到语音的翻译等。语音识别技术所涉及的领域广泛,不同领域上的研究成果都对语音识别的发展做出了贡献,主要包括信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等。语音识别技术与语音合成技术结合可以使人们甩掉键盘,通过语音命令进行操作,现在已成为一个非常具有竞争性的新型高技术产业。

1. 语音识别系统的分类

语音识别系统可以根据对输入语音的限制加以分类。如果从说话者与识别系统的相关性考虑,可以将识别系统分为三类:

(1) 特定人语音识别系统:仅考虑对于专人的话音进行识别。

(2) 非特定人语音系统:识别的语音与人无关,通常要用大量不同人的语音数据库对识别系统进行学习。

(3) 多人的识别系统:通常能识别一组人的语音,或者称为特定组语音识别系统,该系统仅要求对要识别的那组人的语音进行训练。

如果从说话的方式考虑,也可以将识别系统分为三类:

(1) 孤立词语音识别系统:要求输入每个词后要停顿。

(2) 连接词语音识别系统:要求对每个词都清楚发音,一些连音现象开始出现。

(3) 连续语音识别系统:连续语音输入是指进行自然流利的连续语音输入时,大量连音和变音就会出现。

如果从识别系统的词汇量大小考虑,也可以将识别系统分为三类:

(1) 小词汇量语音识别系统:通常包括几十个词的语音识别系统。

(2) 中等词汇量的语音识别系统:通常包括几百个到上千个词的识别系统。

(3) 大词汇量语音识别系统:通常包括几千到几万个词的语音识别系统。

随着计算机与数字信号处理器运算能力及识别系统精度的提高,识别系统根据词汇量大小进行分类也会不断发生变化。

2. 语音识别的几种基本方法

一般来说,语音识别的方法有三种:基于声道模型和语音知识的方法、模板匹配的方法及利用人工神经网络的方法。

1) 基于声道模型和语音知识的方法

该方法起步较早,在语音识别技术提出的开始就有了这方面的研究。通常认为常用语言中有有限个不同的语音基元,而且可以通过其语音信号的频域或时域特性来区分。该方法分为两步实现:

(1) 分段和标号。把语音信号按时间分成离散的段,每段对应一个或几个语音基元的声学特性,然后根据相应声学特性对每个分段给出相近的语音标号。

(2) 建立词序列。根据第一步所得语音标号序列得到一个语音基元网格,从词典得到有效的词序列,也可结合句子的文法和语义同时进行。

2) 模板匹配的方法

模板匹配的方法发展比较成熟,目前已达到了实用阶段。在模板匹配方法中要经过 4 个步骤:特征提取、模板训练、模板分类和判决。常用的技术有三种:动态时间规整(Dynamic Time Warping,DTW)、隐马尔可夫(Hidden Markov Model,HMM)理论、矢量量化(Vector Quantization,VQ)技术。

(1) 动态时间规整。

语音信号的端点检测是语音识别中一个非常重要的步骤,所谓端点检测就是正确地标注出语音信号中各种段落(如音素、音节、词素)的始点和终点的位置,从语音信号中排除无声段。在早期,进行端点检测的主要依据是能量、振幅和过零率,但效果往往不明显。20 世纪 60 年代,Itakura 提出了动态时间规整算法。算法的思想就是把未知量均匀地伸长或缩短,直到与参考模式的长度一致。在这一过程中,未知单词的时间轴要不均匀地扭曲或弯折,以使其特征与模型特征对正。

(2) 隐马尔可夫法。

隐马尔可夫法是 20 世纪 70 年代引入语音识别理论领域的,它的出现使得自然语音识别系统取得了实质性的突破。HMM 方法现已成为语音识别的主流技术,目前大多数大词汇量、连续语音的非特定人语音识别系统都是基于 HMM 模型的。HMM 是对语音信号的时间序列结构建立统计模型,将之看作一个数学上的双重随机过程:一个是用具有有限状态数的 Markov 链来模拟语音信号统计特性变化的隐含的随机过程;另一个是与 Markov 链的每一个状态相关联的观测序列的随机过程。前者通过后者表现出来,但前者的具体参数是不可测的。人的言语过程实际上就是一个双重随机过程,语音信号本身是一个可观测的时变序列,是由大脑根据语法知识和言语需要(不可观测的状态)发出的音素的参数流。可见,HMM 合理地模仿了这一过程,很好地描述了语音信号的整体非平稳性和局部平稳性,是一种较为理想的语音模型。

(3) 矢量量化。

矢量量化是一种重要的信号压缩方法。与 HMM 相比,矢量量化主要适用于小词汇量和孤立词的语音识别中。其过程是将语音信号波形的 K 个样点的每一帧,或有 K 个参数的每一参数帧构成 K 维空间中的一个矢量,然后对矢量进行量化。量化时,将 K 维无限空间划分为 M 个区域边界,然后将输入矢量与这些边界进行比较,并被量化为“距离”最小的区域边界的中心矢量值。矢量量化器的设计就是从大量信号样本中训练出好的码书,从实际效果出发

寻找到好的失真测度定义公式，设计出最佳的矢量量化系统，用最少的搜索和计算失真的运算量实现最大可能的平均信噪比。

在实际的应用过程中，人们还研究了多种降低复杂度的方法，这些方法大致可以分为两类：无记忆的矢量量化和有记忆的矢量量化。无记忆的矢量量化包括树形搜索的矢量量化和多级矢量量化。

3）人工神经网络的方法

利用人工神经网络的方法是20世纪80年代末期提出的一种新的语音识别方法。人工神经网络(Artificial Neural Network，ANN)本质上是一个自适应非线性动力学系统，模拟了人类神经活动的原理，具有自适应性、并行性、鲁棒性、容错性和学习特性，其强大的分类能力和输入输出映射能力在语音识别中都很有吸引力。由于ANN不能很好地描述语音信号的时间动态特性，因此常把ANN与传统识别方法结合，分别利用各自优点来进行语音识别。

3. 语音识别系统的结构

语音识别是研究如何利用计算机从人的语音信号中提取有用的信息，并确定其语言含义。其基本原理就是将输入的语音经过处理后，将其和语音模型库进行比较，从而得到识别结果，如图3-6所示。其中语音采集设备是指话筒和电话等将语音输入设备；数字化预处理则包括A/D变换、过滤和预处理等过程；参数分析是指提取语音特征参数，利用这些参数与模型库中的参数进行匹配，从而产生识别结果的过程；语音识别是最终将识别结果输出到应用程序中的过程；模型库是提高语音识别率的关键。

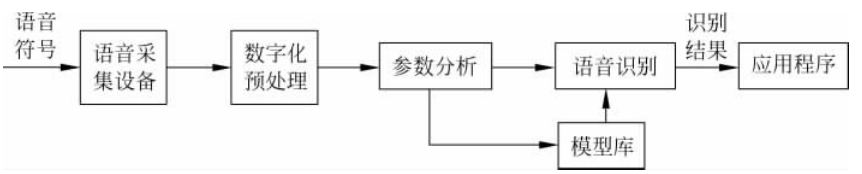


图3-6　语音识别原理

不同的语音识别系统，虽然具体实现细节有所不同，但所采用的基本技术相似，一个典型的语音识别系统的实现过程如图3-7所示。

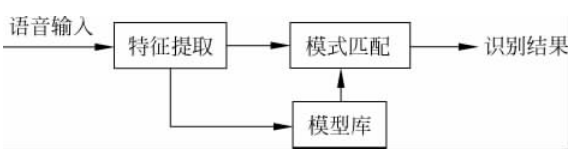


图3-7　语音识别系统的实现过程

完整的基于统计的语音识别系统可大致分为三部分：

(1) 语音信号预处理与特征提取。

选择识别单元是语音识别研究的第一步。语音识别单元有单词(句)、音节和音素三种，具体选择哪一种由具体的研究任务决定。

① 单词(句)。单词(句)单元广泛应用于中小词汇语音识别系统，但不适合大词汇系统，原因在于模型库太庞大，训练模型任务繁重，模型匹配算法复杂，难以满足实时性要求。

② 音节。音节单元多见于汉语语音识别，主要因为汉语是单音节结构的语言，而英语是多音节，并且汉语虽然有大约1300个音节，但若不考虑声调，约有408个无调音节，数量相对

较少。因此,对于中、大词汇量汉语语音识别系统来说,以音节为识别单元基本上是可行的。

③ 音素。音素单元以前多见于英语语音识别的研究中,但目前中、大词汇量汉语语音识别系统也被越来越多地采用。原因在于汉语音节仅由声母(包括零声母有 22 个)和韵母(共有 28 个)构成,且声母和韵母的声学特性相差很大。实际应用中常把声母依后续韵母的不同而构成细化声母,这样虽然增加了模型数目,但提高了易混淆音节的区分能力。由于协同发音的影响,音素单元不稳定,因此如何获得稳定的音素单元还有待研究。

语音识别的一个根本问题是合理地选用特征。特征参数提取的目的是对语音信号进行分析处理,去掉与语音识别无关的冗余信息,获得影响语音识别的重要信息,同时对语音信号进行压缩。在实际应用中,语音信号的压缩率介于 10～100 之间。语音信号包含了大量各种不同的信息,提取哪些信息,用哪种方式提取,需要综合考虑各方面的因素,如成本、性能、响应时间及计算量等。非特定人语音识别系统一般侧重提取反映语义的特征参数,尽量去除说话人的个人信息;而特定人语音识别系统则希望在提取反映语义的特征参数的同时,尽量也包含说话人的个人信息。

线性预测(Linear Prediction,LP)分析技术是目前应用广泛的特征参数提取技术,许多成功的应用系统都采用基于 LP 技术提取的倒谱参数。但线性预测模型是纯数学模型,没有考虑人类听觉系统对语音的处理特点。

Mel 参数和基于感知线性预测(Perceptual Linear Predictire,PLP)分析提取的感知线性预测倒谱在一定程度上模拟了人耳对语音的处理特点,应用了人耳听觉感知方面的一些研究成果。实验证明,采用这种技术,语音识别系统的性能有一定提高。从目前使用的情况来看,梅尔刻度式倒频谱参数已逐渐取代原本常用的线性预测编码导出的倒频谱参数,原因是它考虑了人类发声与接收声音的特性,具有更好的鲁棒性(Robustness)。也有研究者尝试把小波分析技术应用于特征提取,但目前性能难以与上述技术相比,有待进一步研究。

(2) 声学模型与模式匹配。

声学模型通常是将获取的语音特征使用训练算法进行训练后产生。在识别时将输入的语音特征同声学模型(模式)进行匹配与比较,得到最佳的识别结果。声学模型是识别系统的底层模型,并且是语音识别系统中最关键的一部分。声学模型的目的是提供一种有效的方法计算语音的特征矢量序列和每个发音模板之间的距离。声学模型的设计和语言发音特点密切相关。声学模型单元大小(字发音模型、半音节模型或音素模型)对语音训练数据量大小、系统识别率及灵活性有较大的影响。必须根据不同语言的特点和识别系统词汇量的大小决定识别单元的大小。

以汉语为例,汉语按音素的发音特征分为辅音、单元音、复元音和复鼻尾音 4 种;按音节结构分为声母和韵母,并且由音素构成声母或韵母。有时,将含有声调的韵母称为调母。由单个调母或由声母与调母拼接成为音节。汉语的一个音节就是汉语一个字的音,即音节字。由音节字构成词,最后再由词构成句子。

汉语声母共有 22 个,其中包括零声母、韵母共有 38 个。按音素分类,汉语辅音共有 22 个,单元音 13 个,复元音 13 个,复鼻尾音 16 个。

目前常用的声学模型基元为声韵母、音节和词,根据实现目的不同来选取不同的基元。汉语加上语气词共有 412 个音节,包括轻音字,共有 1282 个有调音节字,所以当在小词汇表孤立词语音识别时常选用词作为基元,在大词汇表语音识别时常采用音节或声韵母建模,而在连续语音识别时,由于协同发音的影响,常采用声韵母建模。

基于统计的语音识别模型常用的就是 HMM 模型 λ(N,M,π,A,B),涉及 HMM 模型的相关理论包括模型的结构选取、模型的初始化、模型参数的重估及相应的识别算法等。

(3) 语言模型与语言处理。

语言模型包括由识别语音命令构成的语法网络或由统计方法构成的语言模型,语言处理可以进行语法和语义分析。

语言模型对中、大词汇量的语音识别系统特别重要。当分类发生错误时可以根据语言学模型、语法结构和语义学进行判断纠正,特别是一些同音字必须通过上下文结构才能确定词义。语言学理论包括语义结构、语法规则和语言的数学描述模型等。目前比较成功的语言模型通常是采用统计语法的语言模型与基于规则语法结构命令语言模型。语法结构可以限定不同词之间的相互连接关系,减少了识别系统的搜索空间,这有利于提高系统的识别。

声学模型是识别系统的底层模型,并且是语音识别系统中最关键的一部分。声学模型的目的是提供一种有效的方法计算语音的特征矢量序列和每个发音模板之间的距离。声学模型的设计和语言发音特点密切相关。声学模型单元大小(字发音模型、半音节模型或音素模型)对语音训练数据量大小、系统识别率及灵活性有较大的影响,必须根据不同语言的特点、识别系统词汇量的大小决定识别单元的大小。

3.2.3 语音合成

语音合成又称为文语转换(Text to Speech)技术,已有多年发展历史,是将任意文字信息实时转化为标准流畅的语音朗读出来,它涉及声学、语言学、数字信号处理、计算机科学等多个学科技术,是中文信息处理领域的一项前沿技术。按照研究技术可分为发音参数合成、声道模型参数合成和波形编辑合成,从合成策略上讲可分为频谱逼近和波形逼近。

1. 发音器官参数语音合成

这种方法对人的发声过程进行直接模拟,它定义了唇、舌及声带的相关参数,如唇开口度、舌高度、舌位置和声带张力等。由这些发音参数估计声道截面积函数,进而计算声波。但由于人发音生理过程的复杂性,理论计算与物理模拟之间的差异,合成语音的质量暂时还不理想。

2. 声道模型参数语音合成

这种方法基于声道截面积函数或声道谐振特性合成语音,如共振峰合成器、LPC 合成器。国内外也有不少采用这种技术的语音合成系统。这类合成器的比特率低,音质适中。为改善音质,发展了混合编码技术,主要手段是改善激励,如码本激励、多脉冲激励、长时预测规则码激励等,这样比特率有所增大,同时音质得到提高。作为压缩编码算法,参数合成广泛用于通信系统和多媒体应用系统中。

3. 波形编辑语音合成技术

20 世纪 80 年代末,E. Moulines 和 F. Charpentier 提出了基于时域波形修改的语音合成技术,在 PSOLA(Pitch Synchronous Overlap Add)方法的推动下,此技术得到很大的发展与广泛的应用。波形编辑语音合成技术是直接把语音波形数据库中的波形级联起来,输出连续语流。这种语音合成技术用原始语音波形替代参数,而且这些语音波形取自自然语音的词或句子,它隐含了声调、重音及发音速度的影响,合成的语音清晰自然,其质量普遍高于参数合成。

PSOLA 就是基音同步叠加,它把基音周期的完整性作为保证波形及频谱平滑连续的基本前提。该算法按以下三步实施:对原始波形进行分析,产生非参数的中间表示;对中间表示进行修改;将修改过的中间表示重新合成为语音信号。由于修改的参数不同,又分为 TD-

PSOLA、FD-PSOLA 和 LP-PSOLA。

文语转换系统实际上可以看作是一个人工智能系统。为了合成出高质量的语言,除了依赖于各种规则,包括语义学规则、词汇规则和语音学规则外,还必须对文字的内容有很好的理解,这也涉及自然语言理解的问题。具体的 TTS 的基本结构如图 3-8 所示。

图 3-8 语音合成过程

(1) 语言学处理

对输入的文本进行语言学分析,主要模拟人对自然语言的理解过程——文本规整、断句、词的切分、语法分析和语义分析,使计算机对输入的文本能完全理解,并给出后两部分所需要的各种发音提示。

(2) 韵律处理

为合成语音规划出音段特征,如音高、音长和音强等,使合成语音能正确表达语意,听起来更加自然,提高语音合成系统所输出的语音质量,主要从清晰度、自然度和连贯性等方面进行主观评价。

(3) 声学处理

根据前几部分处理结果的要求,把处理好的文本所对应的单字或短语从语音合成库中提取,把语言学描述转化成语言波形,输出语音,即语音合成。

3.3 音频素材的获取与存储

3.3.1 录音

常用的录音方式有两种,即模拟录音和数字录音。

1. 模拟录音

模拟录音主要是通过录音机等录音设备对声音进行磁记录,将其保存在磁带等磁介质上,然后再通过放音机等播放设备将记录的磁信号还原成音频信号。历史上很多珍贵的影音资料都是利用这种方式记录并保存的。模拟录音是在数字录音技术之前的主要录音手段。对模拟录音文件直接进行相应的编辑、修复较为困难,合成效果有限。随着数字化技术的发展,模拟录音已逐步被数字录音所取代。

2. 数字录音

数字录音通常是利用专业的声音编辑软件(如 GoldWave 和 Adobe Audition)及相关的数字音频录制设备进行录制,能直接得到数字化的音频文件,可根据需要直接在计算机上进行相应编辑、修改及音效合成等操作,极大地方便了操作。

3.3.2 网络及素材库

通过网络下载和购买素材库也可以获得音频素材。对于一些特殊音频效果,由于条件与环境的限制很难直接录制,可通过网络(如中国原创音乐基地 http://5sing.kugou.com/等)

进行搜索下载，方便快捷。

3.3.3　转换及效果合成

通过音频文件的转换与合成来获取音频，经济实惠，操作方便。

1. 音频文件的转换

（1）模拟录音的转换：利用放音机通过数据线或话筒将音频输入计算机的声卡，然后再利用声卡的采集功能对输入的信号进行录制，保存为数字化音频文件，以供使用。

（2）从 CD 和 VCD/DVD 获取：利用 Windows Media Player、GoldWave 及 Adobe Audition 等软件进行截取。

（3）利用软件合成：利用 GoldWave、Adobe Audition 等专业音频编辑软件可以对几种声音进行合成与音效处理，如回音、变调、慢速与快速等操作。

2. 音乐格式转换

在创作多媒体作品的时候，由于开发平台及系统软件的限制，往往对音频的文件格式有所限定和要求，因此音频文件的格式转换也是音频素材制作的一个重要方面。常见的音频格式转换软件有超级转换秀、GoldWave 等，图 3-9 所示为 GoldWave 进行文件转换的操作界面。

3. 音频文件的播放

不同格式的音乐文件往往需要不同的播放器，目前常见的播放器有 Windows Media Player、千千静听、QQ 影音、Foobar2000 和 Kugou 等。

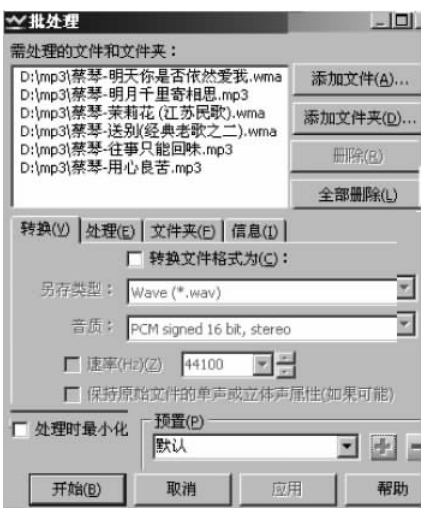


图 3-9　GoldWave 音频文件格式转换界面

3.3.4　数字音频文件

多媒体计算机中的声音可分为三类：波形声音（Wave）、话音（Voice）和音乐（Music）。波形声音实际上包含所有的声音形式，它可以对任何声音进行采样量化，并且恰当地恢复出来，相应的文件格式是 WAV 文件或 VOC 文件。人的说话声虽是一种特殊的媒体，但也是一种波形，所以与波形声音的文件格式相同。音乐是符号化的声音，乐谱可转变为符号媒体形式，对应的文件格式有 MID、MP3、CMF 文件等。

数字音频文件是数字化音频的软载体，目前常用的文件格式有：

1. VOC

VOC 文件是 Creative 公司开发的波形音频文件格式，也是声霸卡（Sound Blaster）所使用的音频文件格式，被 DOS 平台所支持。但是随着 Windows 平台的普及，已逐渐被淘汰，取而代之的是 WAV 格式。VOC 文件由文件头块（Header Block）和音频数据块（Data Block）组成。文件头包含一个标识、版本号和一个指向数据块起始的指针。数据块分成各种类型的子块，如声音数据、静音、标记、ASCII 码文件、重复、重复的结束及终止标记等。

2. WAV

WAV 文件是 Microsoft 公司开发的音频文件格式，它来源于对声音模拟波形的采样。用不同的采样频率对声音的模拟波形进行采样可以得到一系列离散的采样点，以不同的采样位

数(8b 或 16b)把这些采样点的值转换成二进制数,然后存入磁盘,这就产生了声音的 WAV 文件,即波形文件。波形声音是最基本的声音格式,该格式记录声音的波形,因此只要采样频率高、采样字节长、机器速度快,利用该格式记录的声音文件能够和原声基本一致,质量非常高。但其文件尺寸太大,多用于存放简短的声音片段。

3. AIF

AIF/AIFF(Audio Interchange File Format,音频交换文件格式)是 Apple 公司开发的一种声音文件格式,被 Macintosh 平台及其应用程序所支持,SGI 及其他专业音频软件包同样支持 AIFF 格式。AIFF 支持 ACE2、ACE8、MAC3、MAC6 压缩,支持 16 位 44.1kHz 立体声。

4. MIDI

MIDI(Musical Instrument Digital Interface,乐器数字接口)是由世界上主要电子乐器制造厂商共同提出来的一个通信标准,规定了计算机音乐程序、电子合成器和其他电子设备之间交换信息与控制信号的方法。MIDI 文件中包含音符、定时和多达 16 个通道的乐器定义,每个音符包括键、通道号、持续时间、音量和力度等信息,所以 MIDI 文件记录的不是乐曲本身,而是一些描述乐曲演奏过程的指令。因此,MIDI 音频与波形音频完全不同,它不对声波进行采样、量化与编码,而是将电子乐器键盘的演奏信息(包括键名、力度和时间长短等)记录下来,这些消息称为 MIDI 消息,是乐谱的一种数字式描述。对应于一段音乐的 MIDI 文件不记录任何声音消息,而只是包含一系列产生音乐的 MIDI 消息(描述乐曲演奏过程的指令),播放时只需从中读出 MIDI 消息,生成所需的乐器声音波形,经放大处理即可输出。与波形声音相比,由于 MIDI 数据不是声音而是指令,因此它的文件长度非常小。半小时的立体声音乐,如果用波形文件无压缩录制,约需 300MB 的存储空间,而 MIDI 数据大约只需要 200 KB,两者相差 1000 多倍。MIDI 的另一个优点表现在配音方面,由于数据量小,可在多媒体应用中与其他波形声音配合使用,形成伴音的效果,而两个波形文件是不能同时播放的。与波形声音相比,MIDI 声音在编辑修改方面也是十分方便灵活的。例如,可以任意修改曲子的速度和音调,也可改换乐器等。MIDI 的缺陷主要是无法模拟自然界中其他非乐曲类声音,文件的录制比较复杂,需掌握一定的 MIDI 创作及改编作品的专业知识,同时还必须借助于专门的工具如键盘合成器等。

根据 MIDI 的特点,在以下三种情况下比较适合用 MIDI 谱曲。

(1) 长时间播放高质量的音乐。

(2) 从 CD-ROM 或 DVD-ROM 等装载其他数据的同时,以音乐作为背景音响效果。

(3) 用音乐作为背景音响效果,同时播放波形音频或进行文字/语言转换音乐输出。

MIDI 是目前最成熟的音乐格式,实际上已经成为一种行业标准,其科学性、兼容性和复杂程度等各方面远远超过其他标准(除交响乐 CD 和 Unplug CD 外),它的 General MIDI 是最常见的通用标准。作为音乐工业的数据通信标准,MIDI 能指挥各种音乐设备的运转,而且具有统一的标准格式,能够模仿原始乐器的各种演奏技巧,甚至可达到人类无法演奏出的效果。MIDI 文件的扩展名为 MID。

MIDI 设备是处理 MIDI 信息所需要的硬件设备,其基本组成包括:

(1) MIDI 端口。一台 MIDI 设备可以有 1～3 个 MIDI 端口,分别称为 MIDI In、MIDI Out 及 MIDI Thru。

① MIDI In:接收来自其他 MIDI 设备的 MIDI 信息。

② MIDI Out:发送本设备生成的 MIDI 信息到其他设备。

③ MIDI Thru：将从 MIDI In 端口传来的信息转发到相连的另一台 MIDI 设备上。

(2) MIDI 键盘。主要用于 MIDI 乐曲演奏，MIDI 键盘本身并不发音，当作曲人员按下键盘上的按键时就会发出按键信息，产生的也只是 MIDI 音乐消息，经过音序器录制后才生成 MIDI 文件。这些数据可以进一步加工，也可以和其他的 MIDI 数据合并，经过编辑后的 MIDI 文件就可以送合成器播放。

(3) 音序器(Sequencer)。用于记录、编辑和播放 MIDI 的声音文件，音序器既有硬件形式也有软件音序器，目前大多为软件音序器。音序器可捕捉 MIDI 消息，将其存入 MIDI 文件。音序器还可以编辑 MIDI 文件。

(4) 合成器。MIDI 合成器与 WAV 合成器之间没有任何关系，它们是声卡上两个独立的声音合成器单元。MIDI 文件的播放是通过 MIDI 合成器，合成器解释 MIDI 文件中的指令符号，生成所需要的声音波形，经放大后由扬声器输出，声音的效果比较丰富。MIDI 文件也可以不经合成器直接送原 MIDI 设备播放。

目前被广泛采用的 MIDI 合成方式有调频合成(Frequency Modulation，FM)和波形表合成(Wave Table)两种。

(1) 调频合成方式。其原理是根据傅里叶级数而来的，即任何一种波动信号都可被分解为若干个频率不同的正弦波，合成器利用硬件产生的若干个正弦波合成某种乐器的声音。

(2) 波形表合成。其原理是 ROM 中已存储着各种实际乐器的声音采样，合成时以查表方式调用这些样本将其还原回放。它可分为硬波表合成与软波表合成。

① 硬波表合成方式。该合成方式的数字声音样本被保存在 ROM 或 RAM(可动态更换)内。而软波表的数字化样本保存于系统主存中，合成运算靠 CPU 完成，最终的音频合成靠声卡上的 WAV 合成器来完成。

② 软波表合成方式。该合成方式表实际上是针对合成 MIDI 音乐而开发的一套软件，其主要作用是控制 CPU 来完成波表 MIDI 合成器的部分功能。

波表与 FM 的最大区别就在于 FM 通过对简单正弦波的线性控制来模拟音乐乐器、鼓和特殊效果，而波表采用真实的声音样本进行回放，因此采用波表合成的 MIDI 音乐听上去更加接近自然且更具真实感，而 FM 合成的 MIDI 音乐则多带有人工合成的色彩。

5. MP3

MP3(MPEG-1 Layer3)是目前最流行的声音文件格式。MPEG 即动态视频压缩标准，其中的声音部分称为 MPEG-1 音频层，它根据压缩质量和编码复杂度划分为三层，即 Layer1、Layer2、Layer3，分别对应 MP1、MP2、MP3 三种声音文件，并且根据不同的用途，使用不同层次的编码。MPEG 音频编码的层次越高，对应的编码器越复杂，压缩率也越高。MP1 和 MP2 的压缩率分别为 4∶1 和 6∶1～8∶1，而 MP3 的压缩率高达 10∶1～12∶1 或更高。举例来说，一个未经压缩的 50MB 的 WAV 文件压缩成 MP3 文件时可能只有 5MB。不过，MP3 采用的是有损压缩方式，与 CD 相比音质差一些。

由于 MP3 是压缩后产生的文件，因此需要一套 MP3 播放软件进行还原。目前 Windows 自带的媒体播放器和 Winamp 等很多软件都支持这种声音文件格式。为了降低失真度，MP3 采取“感官编码技术”，以极小的声音失真换取了较高的压缩比，这使得 MP3 既能在 Internet 上自由传播，又能被轻易地下载到便携式数字音频设备(如 MP3 随身听)中。这种便携式数字音频设备是基于数字信号处理器(Digital Signal Processing，DSP)的，无须计算机支持便可实现 MP3 文件的存储、解码和播放。MP3 文件的扩展名为.MP3。

6. MP4 音乐

在 MP3 日益成为一种主流的音乐格式之后,现在又出现了 MP4。MP4 并不能望文生义地理解为 MPEG-4 或者 MPEG-1 Layer4 格式。从技术层面讲,MP4 使用的是 MPEG-2AAC 技术,简称 A2B 的技术。它的特点是音质更加完美而压缩比更大(15∶1～20∶1)。MPEG-2AAC 是在采样频率为 8～96kHz 时,可提供 1～48 个声道可选范围的高质量音频编码。AAC(Advanced Audio Coding,先进音频编码)适用于从比特率为 8kb/s 单声道电话语音音质到 160kb/s 多声道超高质量音频信号范围内的编码,并且允许对多媒体进行编码/解码。它增加了诸如对立体声的完美再现、比特流效果音扫描、多媒体控制和降噪等 MP3 没有的特性,使得在音频压缩后仍能完美地再现 CD 的音质。

MP4 真正的含义由来是因为版权问题,对唱片公司来说,MP3 的缺陷就是忽视了著作者和出版者应享有的版权待遇。于是,GMO(Global Music One)公司针对 MP3 提出了基于 AT&T 公司授权的 AAC 改良技术——A2B 的音频压缩方法和应用,并将其命名为 MP4,其用意大概是想表明 MP4 是继 MP3 之后的一种升级换代技术,这正好契合了人们的习惯思维。

A2B 技术主要由三个部分组成:第一,AT&T 的音频压缩技术专利,它可以将 AAC 压缩比提高到 20∶1 而不损失音质;第二,安全数据库,它可以为 A2B 音乐文件创建一个特定的密钥,并将此密钥置于其数据库中,只有 A2B 的播放器才能播放含有这种密钥的音乐;第三,协议认证,这个认证包含了复制许可、允许复制副本数量、歌曲总时间、歌曲可以播放时间及经营销售许可等信息。

7. Real Audio 文件——RA/RM/RAM

Real Audio 文件是由 Real Networks 公司开发的主要适用于网络实时数字音频流技术的文件格式,如今已成为网上在线收听的标准。它将音频文件大大压缩,所以在高保真方面远不如 MP3,不过由于体积小,适合实时收听。与 MP3 相同,它也是为解决网络传输带宽资源而设计的,因此主要追求压缩比和容错性,其次才是音质。

8. CD Audio 音乐

CD Audio 音乐是 CD 唱片采用的格式,又叫“红皮书”格式,是目前音质最好的音频格式,其扩展名为.CDA。在大多数播放软件的“打开文件类型”中都可以看到 *.CDA 格式,这就是 CD Audio 了。CD 音轨可以说是近似无损的,因此它的声音基本上是忠于原声的。CD 光盘可以在 CD 唱机中播放,也能用计算机中的各种播放软件播放。一个 CD 音频文件即一个 *.CDA 文件,这只是一个索引信息,并不真正地包含声音信息,所以不论 CD 音乐的长短,在计算机上看到的“*.CDA 文件”都是 44 字节长。注意,不能直接复制 CD 格式的 *.CDA 文件到硬盘上播放,需要使用像 Exact Audio Copy 这样的抓音轨软件把 CD 格式的文件转换成 WAV 格式的文件。CD Audio 音乐的缺点是无法编辑,文件太大。

9. AAC 文件

AAC(Advanced Audio Coding,高级音频编码)出现于 1997 年,基于 MPEG-2 的音频编码技术,由 Fraunhofer IIS、杜比实验室、AT&T 和 SONY(索尼)等公司共同开发,目的是取代 MP3 格式。2000 年,MPEG-4 标准出现后,AAC 重新集成了其特性,加入了 SBR 技术和 PS 技术,为了区别于传统的 MPEG-2,AAC 又称为 MPEG-4 AAC。

10. 其他音频文件格式

除了上述常见的音频文件格式以外,还有以下几种格式:

- RMI 文件:Microsoft 公司的 MIDI 文件格式,它可以包括图片、标记和文本。

- SND 文件：另一种计算机的波形声音文件格式，Apple 计算机上音频文件的存储格式。
- AU 文件：SUN 和 NEXT 公司的声音文件存储格式，主要用于 UNIX 工作站上。

3.3.5 音质与数据量

本书中所讲的数字音频主要指 WAV 格式的波形音频文件，它是其他格式音频文件转换的基础。数字音频的声音质量好坏取决于采样频率的高低、表示声音的基本数据位数和声道形式。音频文件的数据量由下式算出：

$$v = fbs/8$$

式中 v 代表数据量；f 是采样频率；b 是数据位数；s 是声道数。

例如，CD 质量的参数为：$f=4.1\text{kHz}$，$b=16\text{b}$，$s=2$，则每秒钟的数据量为：

$$v = (44\,100\text{Hz} \times 16\text{b} \times 2) \div 8 = 176\,400\text{B}(\text{约合 } 172\text{KB})$$

如果以 CD 激光盘音质(44 100Hz 的采样频率，16 位，立体声，172KB/s)记录一首 5min(300s)的乐曲，则数据量为：

$$172\text{KB/s} \times 300\text{s} = 51\,600\text{KB}(\text{约合 } 50.39\text{MB})$$

由计算结果可以看出，音频文件的数据量问题不容忽视。为了节省存储空间，通常在保证基本音质的前提下适当降低采样频率。在一般场合，人的语音采用 11.025kHz 的采样频率、8b、单声道已足够；如果是乐曲，22.05kHz 的采样频率、8 位、立体声就已满足要求。

3.4 音频编辑

一般的音频编辑包括录制音频、确定编辑区域、删除片段、设置静音和剪贴片段等相关操作，常用软件有 GoldWave 和 Adobe Audition 等。

3.4.1 GoldWave 软件介绍

GoldWave 是一款功能强大的数字音乐编辑器，它小巧易用，可运行在 Windows XP/7 等环境中。它集声音编辑、播放、录制和转换于一身。可处理的音频文件格式包括 WAV、MP3、OGG、VOC、IFF、AIF、AFC、AU、SND、MAT、DWD、SMP、VOX、SDS、AVI、MOV、APE 等，也可以从 CD、VCD、DVD 及其他视频文件中获取声音。编辑功能主要包括剪辑、合成多个声音素材、制作回声、混响、改变音调和音量、频率均衡控制、音量自由控制及声道编辑等。

本书中使用的是 GoldWave 5.77，把该软件的全部文件复制到硬盘的某个文件夹内，然后在桌面上建立启动文件 GoldWave.exe 的快捷方式。双击桌面上 GoldWave 的快捷方式图标，显示图 3-10 所示的主界面。主界面中的菜单栏用于文件及其他编辑操作；工具栏中的工具按钮用于编辑和产生特效；窗口中的左声道和右声道波形是主要编辑区，坐标轴是时间轴；底部的状态栏提示当前编辑的时间宽度和采样频率等。

3.4.2 简单音频编辑

简单音频编辑包括删除片段、静音处理、剪贴片段、声音反向及生成回声效果等。不论声音素材是单声道还是双声道，编辑操作同样有效。

1. 增减工具

工具栏上的各种工具按钮可以根据需要增减，以方便使用。如果屏幕显示分辨率足够高，

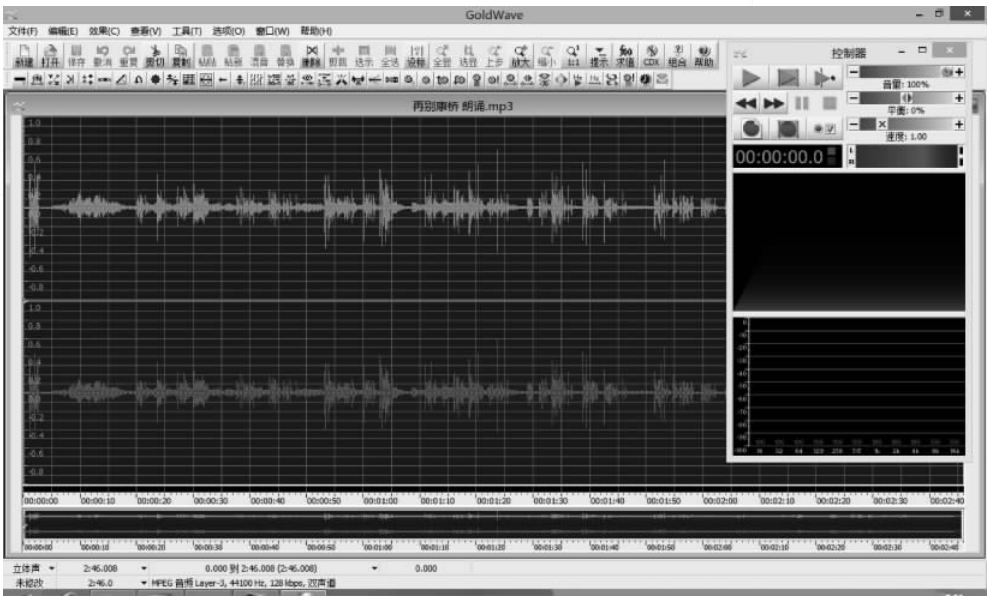


图 3-10　GoldWave 软件的主界面

软件主界面足够大,也可把尽可能多的工具按钮置于工具栏中。

在工具栏上,增减工具按钮的操作步骤如下:

(1) 选择“选项”→“工具栏”命令,出现“工具栏选项”对话框,如图 3-11 所示。

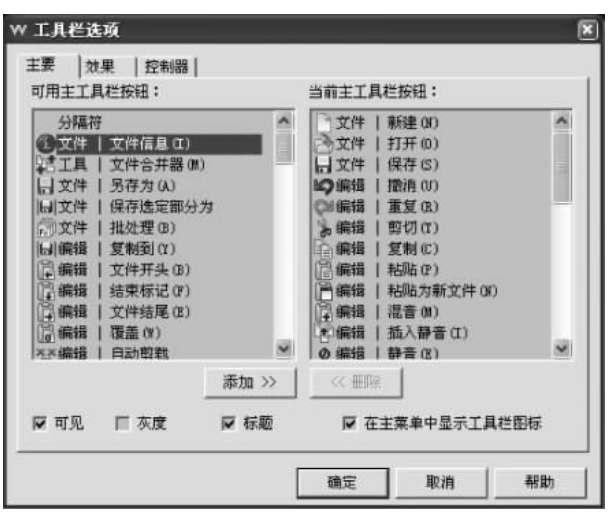


图 3-11　“工具栏选项”对话框

(2) “工具栏选项”对话框中“主要”选项卡右侧的“当前主工具栏按钮”列表框中是当前工具栏中显示的工具按钮。若要增加工具按钮,在左侧“可用主工具栏按钮”列表框中选择一个工具,拖动到右侧“当前主工具栏按钮”列表框中即可。若要减去某个工具按钮,则把右侧“当前主工具栏按钮”列表框中的工具拖动到左侧“可用主工具栏按钮”列表框中即可。

(3) 工具按钮增减完毕,单击“确定”按钮。

2. 打开与关闭声音文件

GoldWave 软件可以直接打开、编辑和保存 MP3 格式、WAV 格式及其他多种格式的声音

文件。

(1) 打开声音文件。单击工具栏中的“打开”按钮，显示“打开声音文件”对话框，查找并选中需要打开的声音文件后，单击“打开”按钮。

GoldWave可同时打开多个声音文件，但受内存容量的限制，打开文件不宜过多。

(2) 关闭声音文件。选择“文件”→“关闭”命令，关闭当前的声音窗口。

3. 删除声音片段

该操作用于取消不需要的部分，如噪声、噼啪声、各种杂音及录制时产生的口误等。删除声音片段的操作步骤如下：

(1) 确定编辑区，在声音文件的音轨上单击设置编辑起点，右击设置编辑终点。

(2) 单击工具栏中的“删除”按钮，编辑区域被删除，其中的声音也一并被删除。

要准确地确认编辑区域，需要仔细聆听，在放大显示状态下反复调整区域。

4. 静音处理

静音处理可以把声音片段处理成一段寂静无声的片段，通常用于去除语音之间的噪声、音乐首尾的噪音和设置两段声音之间的静音间隔等。静音处理的操作步骤如下：

(1) 确定编辑区域。

(2) 单击工具栏中的“静音”按钮，编辑区域变成静音，时间长度不变。

如果默认状态下工具栏中没有静音工具按钮，可通过相应设置添加该工具按钮。

5. 剪贴片段

剪贴片段用于重新组合声音，将某段“剪”下来的声音粘贴到当前声音的其他位置或者粘贴到其他声音素材中。剪贴片段的操作步骤如下：

(1) 确定编辑区域，该区域将是被剪贴的内容。

(2) 单击工具栏中的“复制”按钮，将编辑区域的内容复制到剪贴板中。

(3) 单击任意声音文件波形图的某一位置(该位置是粘贴的起始位置)，单击“粘贴”按钮，剪贴板内的声音被粘贴到波形图中，原有声音被“挤”向后边。

如果希望把剪贴板内容生成新的文件，而单击工具栏中的“粘贴为新文件”按钮，则生成新文件的音轨编辑窗口。这个操作经常用于把某部分从声音素材中分离出来。

6. 恢复操作

一旦发生操作失误，单击工具栏中的“撤销”按钮，可恢复单击撤销工具之前的状态。若单击“重复”按钮，可恢复错误发生之前的状态。

7. 声音反向

声音反向是把声音数据反向排列，形成倒序声音。倒序声音可用于声音的加密传送，对方利用相同的软件，并进行相同的倒序处理才能还原声音。声音反向的操作步骤如下：

(1) 确定编辑区域，把需要进行倒序处理的内容包括在内。

(2) 单击工具栏中的“反向”按钮，编辑区域内的声音变成倒序。

8. 生成回声

制作回声最理想的对象是语音，乐曲和歌曲不宜制作回声，这是由于乐曲和歌曲比较连续，不易听出回声的缘故。产生回声的基本原理如图3-12所示。

在原声1次波上叠加2次波，且2次波比1次波有所延迟，音量小，叠加后的听觉效果就是回声。当然，如果叠加3次波、4次波乃至更多，则产生像左右漂移的效果。具体的操作步

骤如下：

(1) 设置编辑区域，把需要制作回声的部分包括进去。

(2) 单击工具栏中的“回声”按钮，出现图 3-13 所示“回声”对话框。

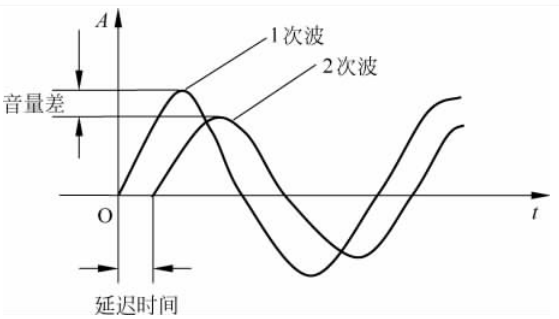


图 3-12　回声的基本原理

图 3-13　“回声”对话框

(3) 在对话框中移动“回声”滑块，确定叠加波形的数量，通常取 2～4。

(4) 移动“延迟”滑块，调整各次波的延迟时间。

(5) 移动“音量”滑块，确定叠加波形的衰减音量。

(6) 若希望回声采用立体声，单击“立体声”选项，使其有效。

(7) 希望回声不绝于耳，单击“产生尾声”选项，使其有效。这是多次波叠加的效果。

(8) 设置完成后单击“确定”按钮。

延迟时间不宜过长，否则声音分离，不像回声。设置完成后，可先单击回声设置对话框中的“试听当前设置”按钮，试听效果满意后单击“确定”按钮结束操作。

3.4.3　高级音频编辑

高级音频编辑包括设置播放控制工具、淡入淡出、混响时间、频率均衡控制、时间调整、响度控制、声道编辑、音频合成及降噪等。

1. 设置播放控制工具

工具栏中的播放器有很多控制工具，如图 3-14(a)所示。这些工具用于监听编辑效果，方便音频编辑。最常用的播放控制工具是两个播放按钮，一个绿色，一个黄色。两个播放按钮的功能可以自行设定。绿色按钮用于聆听编辑区域开始端的声音，黄色按钮用于聆听编辑区域结束端的声音，可方便地确认编辑区域是否准确。具体设置步骤如下：

(1) 选择“选项”→“控制器属性”命令，会出现图 3-14(b)所示“控制器属性”对话框。

(2) 在对话框中的“绿色播放键”选项区域中选择“选定部分”单选按钮，使其有效。

(3) 在“黄色播放键”选项区域中选择“结束部分”单选按钮，使其有效。

绿色和黄色播放键栏目的底部均有“循环”选项，可设定循环播放的次数。

(4) 单击“确定”按钮，结束设置。

2. 淡入淡出

“淡入”和“淡出”是指声音的渐强和渐弱，通常用于产生渐近渐远的听觉效果。两个声音素材交替切换时也经常采用这种处理方式。制作淡入和淡出效果的操作步骤如下：

(1) 确定编辑区域。一般编辑区域总是位于声音素材的开始或末尾。

(a) 播放器中的播放按钮　　(b) “控制器属性”对话框

图 3-14　播放控制工具的定义

(2) 制作淡入效果。单击工具栏中的“淡入”按钮，会出现图 3-15(a)所示的“淡入”对话框。调整滑块可以改变淡入的初始音量，初始音量为 0 时无须拖曳滑块。单击“确定”按钮。

(3) 制作淡出效果。单击“淡出”按钮，显示图 3-15(b)所示的“淡出”对话框。调整滑块可以确定淡出的最终音量，若最终音量为 0 则不动滑块。单击“确定”按钮。

淡入与淡出效果如图 3-15(c)所示。在乐曲的开始和结束阶段有渐进和渐远的听觉感受。

(a) “淡入”对话框　　(b) “淡出”对话框

(c) 波形效果

图 3-15　淡入与淡出效果

3. 混响时间

混响时间的长短是润色音色的技术手段，混响时间稍长，声音显得圆润；混响时间更长一些，声音具有空旷感。

混响原理与回声原理近似，把滞后一小段时间的声音叠加到原声上，叠加的声音音量和延迟时间可调，以产生不同的混响效果。其具体制作步骤如下：

(1) 确定编辑区域。

(2) 单击工具栏中的“混响”按钮，会出现图 3-16 所示的“混响”对话框。

(3) 调整“混响时间”滑块,确定混响时间,单位是秒。混响时间越长,空旷效果越明显。调整“音量”滑块,改变叠加到原声上的声波幅度。调整“延迟深度”滑块,改变延迟时间,从而影响混响总体效果。

4. 频率均衡控制

频率均衡控制是指对低音、中音和高音各个频段进行提升和衰减的控制。该控制使声音的层次和频段分布更为理想,音响效果更好。操作步骤如下:

(1) 确定编辑区域。

(2) 单击工具栏中的“均衡器”按钮,会出现图 3-17 所示的“均衡器”对话框。在对话框中可以看到一个七段均衡器,每个频率段可单独调整。

图 3-16 “混响”对话框

图 3-17 “均衡器”对话框

(3) 移动各个频段的滑块,调整该频段的强弱。

如果各频率段的调整没有固定规则,要根据声音素材的实际情况进行。若乐曲高音不清,中音混浊,则可适当提高 15kHz 和 1000Hz 频段的幅度。

(4) 调整完毕,单击“确定”按钮。

5. 时间调整

制作多媒体产品时,为了和画面同步,需要改变声音的长度。加工音响素材时,也需要精确地控制长度,这就需要进行时间的调整。具体调整步骤如下:

(1) 设定编辑区域。

(2) 单击工具栏中的“时间弯曲”按钮,出现图 3-18(a)所示的“时间弯曲”对话框。在“变化”和“长度”单选按钮两者之间任选一个,改变其数值,即可改变声音的时间长度。聆听效果时会发现音调也随之发生变化。

(3) 若希望改变时间长度时音调不变,在图 3-18(a)所示的对话框中单击 FFT 按钮,显示图 3-18(b)所示的画面。在画面下边改变“FFT 大小”微调框中的数值,数值大,效果好。根据视听效果改变“重叠”下拉到表框中的数值,调整完毕后单击“确定”按钮。

6. 音量自由控制与合成

声音的音量可根据音量曲线自由控制,此举常用于多种声音素材的合成。在一首乐曲中可随意安排某处或多处的音量减小或增加。音量自由控制的典型例子如图 3-19 所示。图 3-19 中背景音乐采用了音量自由控制,在中间某段形成低谷。在曲线低谷时插入语音。待语音结

(a) 直接改变速率　　(b) FFT模式

图 3-18　“时间弯曲”对话框

束后，曲线恢复原有音量值。

1）音量自由控制

音量自由控制的操作步骤如下：

（1）打开语音文件，聆听并记录下该语音的时间长度。

（2）打开背景音乐文件，寻找合适的语音插入点，然后设置编辑区域。该区域应略大于语音文件时间为 2～4s。例如，语音长度为 20s，则编辑区域为 22～24s，如图 3-19 所示。

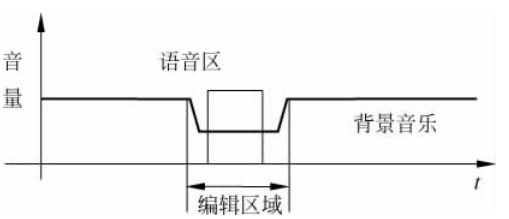


图 3-19　音量自由控制原理

（3）单击工具栏中的“外形音量”按钮，出现“外形音量”对话框。拖动该对话框中的线段形成低谷，与图 3-19 中的背景音乐曲线类似，如图 3-20 所示。

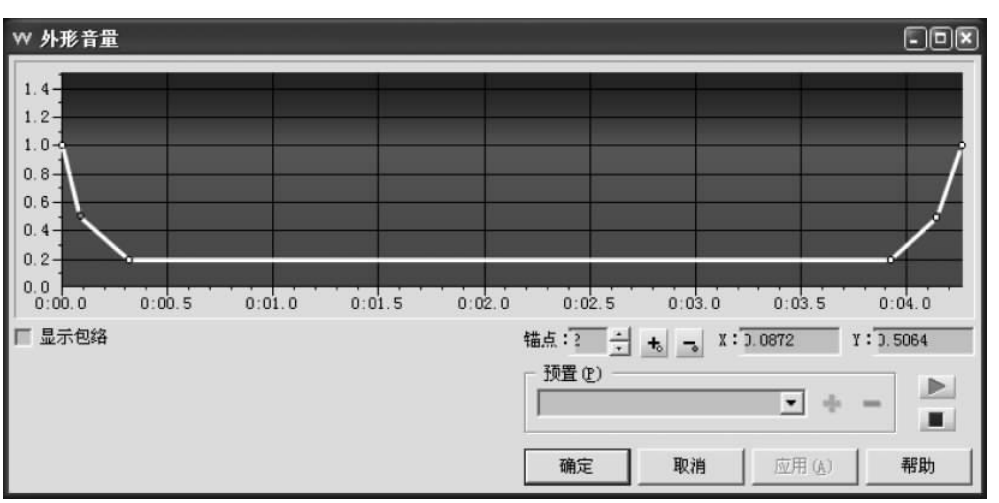


图 3-20　“外形音量”对话框

(4) 单击“确定”按钮。

2) 合成

把语音与背景音乐合成在一起,其位置在背景音乐的低谷处。这种合成手段适用于所有声音素材的合成。合成步骤如下:

(1) 打开参与合成的相关素材,如经过音量自由控制的背景音乐和语音等。

素材窗口多时,可选择“窗口”→“横向平铺”或“窗口”→“纵向平铺”命令,整齐排列各个窗口。整齐排列的画面如图 3-21 所示。

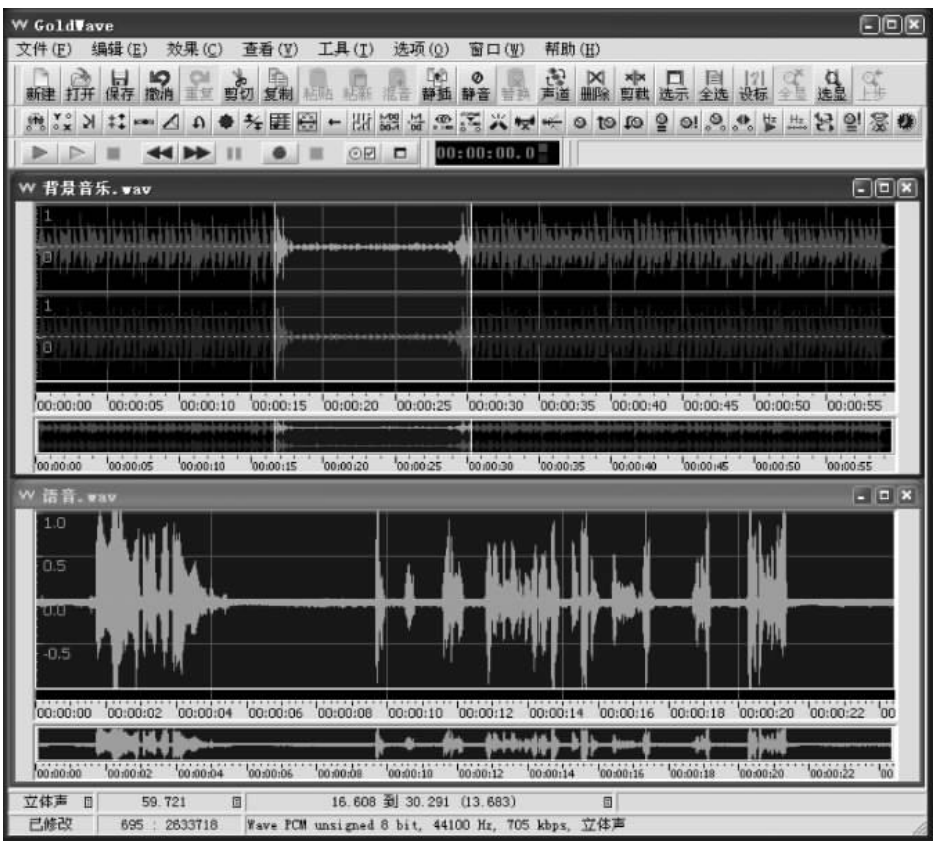


图 3-21 参与合成的素材画面

(2) 单击语音窗口,设置编辑区域,把语音包括在内。

(3) 单击工具栏中的“复制”按钮,将语音复制到剪贴板。

(4) 单击背景音乐窗口,在低谷的开始位置单击,确定合成起点。

(5) 单击“混音”按钮,出现图 3-22 所示的“混音”对话框。在对话框中调整音量滑块,改变将要合成的语音音量。若语音原有音量很小,右移滑块,适当调高音量。

图 3-22 “混音”对话框

(6) 单击“确定”按钮,语音被合成在背景音乐的低谷处。

单声道音频合成到双声道音频中时,自动变成均等的双声道。若双声道音频向单声道合成时,则把两个声道合二为一,变成单声道。

7. 声道编辑

前面介绍的所有编辑手段都是在两个声道间同步进行的。声道编辑可以在两个声道中选择一个进行编辑，把声音素材合成到任意一个声道，制作声像左右漂移效果等功能。

1）选择当前声道

选择当前声道的步骤如下：

（1）右击选择区域，然后选择“声道”按钮，左声道处于当前编辑状态，右声道亮度变暗，处于非编辑状态。

（2）再次右击选择区域，选择“声道”按钮，则右声道成为当前编辑的声道。

（3）再次右击选择区域，然后选择“声道”按钮，恢复到原始的双声道编辑状态。

选择声道后，所有音频编辑手段只对当前声道有效。

2）声道间素材的合成

声道间素材的合成步骤如下：

（1）选择一个声道，设置编辑区域，单击工具栏中的“复制”按钮，把该声道的内容复制到剪贴板中。

（2）切换到另一个声道，重新设定编辑区域。根据需要，单击粘贴按钮或合成按钮，把剪贴板中的内容粘贴（插入效果）或合成到当前声道中。

若使用粘贴功能，由于是插入操作，因而将改变当前声道的时间长度，与另一个声道的同步关系被破坏，应予以充分注意，除非有意制作该效果。

3）制作声像漂移效果

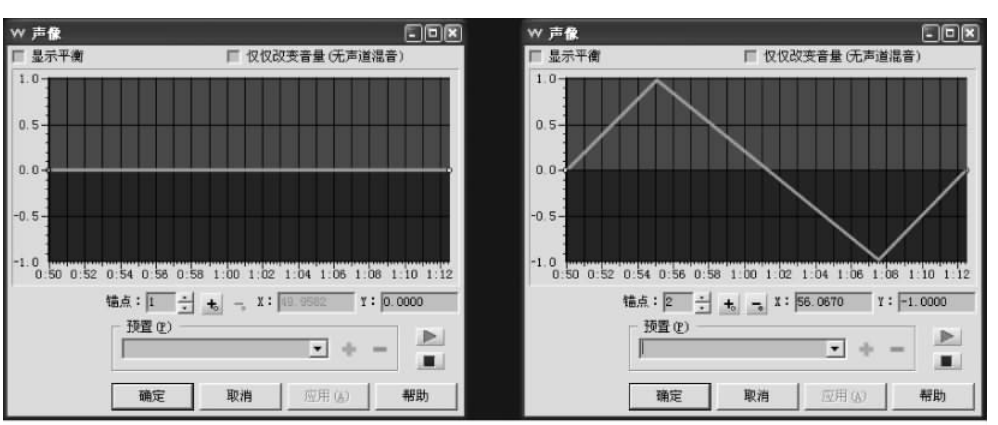
声像漂移是一种听觉感受，声音在左、右声道之间来回漂移，忽左忽右。声像漂移必须在双声道编辑状态下进行，不可只有一个编辑声道。

制作声像漂移效果的操作步骤如下：

（1）设定编辑区域。

（2）单击工具栏中的“声像”按钮，弹出图 3-23 所示的“声像”窗口。该窗口的上半部分是左声道（图中浅色部分），下半部分是右声道（图中深色部分），中间有一条直线。

（3）拖动图 3-23(a)中的直线或上或下移动，如图 3-23(b)所示。该线段表示声音从平衡点到右声道最大值，然后通过平衡点逐渐过渡到左声道为最大值，再回到平衡点。听觉感受是：声音先从中间向右漂移，然后通过中间向左漂移，最后恢复到中间。



(a) 调整前　　(b) 调整后

图 3-23　“声像”窗口

8. 多格式保存

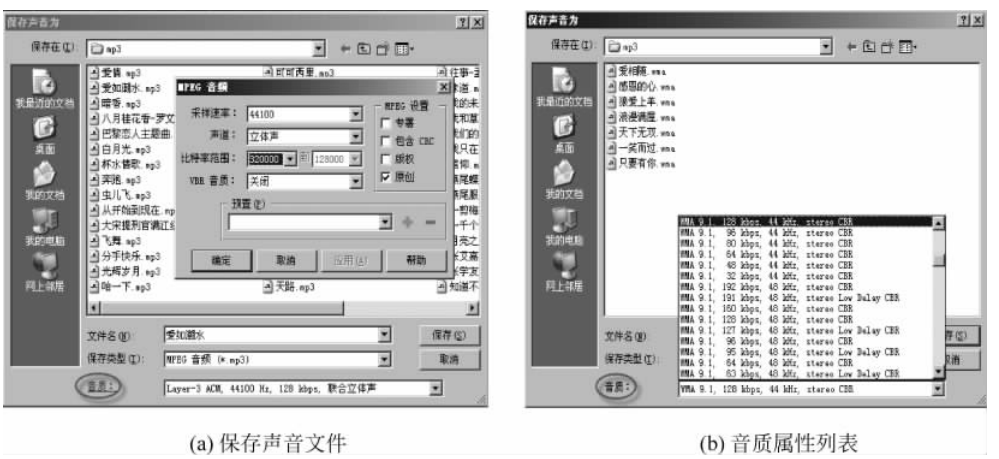
GoldWave 软件可以多种格式保存声音文件,下面以常见的 WAV 格式和 MP3 格式为例进行介绍。

1) 保存 MP3 文件

MP3 声音文件应用广泛,除了用计算机播放以外,使用最广泛的是手机、MP3 和 MP4 播放器。在保存 MP3 声音文件时需要考虑播放设备,从而决定采用何种文件模式。

保存 MP3 文件的操作步骤如下:

(1) 选择"文件"→"另存为"命令,出现图 3-24(a)所示"保存声音为"对话框。



(a) 保存声音文件　　(b) 音质属性列表

图 3-24　保存声音文件界面

(2) 在对话框中指定保存的路径;在"保存类型"下拉列表中选择 MPEG 音频(*.mp3)选项。

(3) 在"音质"下拉列表中选择一种文件模式。

若使用计算机播放 MP3 声音文件,采用"44 100Hz,320kb/s,立体声"模式,该模式数据量大,音质好;若使用 MP3 随身听播放,在"44 100Hz,192kb/s,立体声"到"44 100Hz,96kb/s,立体声"之间选取,kb/s 数值越低,数据量越小,音质越差。

(4) 单击"保存"按钮。

2) 保存 WAV 文件

WAV 声音文件的数据量很大,为了在音质和数据量之间寻求平衡,在保存时要选用不同的文件模式。操作步骤如下:

(1) 选择"文件"→"另存为"命令,出现图 3-24(a)所示的"保存声音为"对话框。

(2) 指定路径,并输入文件名。

(3) 在"音质"下拉列表中选择一种文件模式,如图 3-24(b)中的属性列表。通常在列表中选择"Unsigned 8bit,立体声"模式,若要求音质更好一些,可选择"Unsigned 16bit,立体声"模式。

(4) 单击"确定"按钮。

保存后的文件需要改变采样频率时可使用"录音机"软件。

利用 GoldWave 软件的文件操作,可以方便地实现文件转换。如打开 WAV 格式文件后,保存为 MP3 格式文件,反之亦然。

3.5　综合实例

任务：制作“再别康桥”的配乐朗读。

要求：让人通过听“再别康桥”的配乐朗读体会这篇散文的意境。

制作过程：先录制朗读声音，然后对声音进行降噪、音调调整和幅度调整等修饰处理，最后加入合适的背景音乐进行混音合成，以加强该散文的抒情意境。

3.5.1　录音设置

选择安静的录音场所，关闭计算机中一切可能发声的软件，如 QQ 和旺旺等。

设置录音属性的步骤如下：

(1) 右键单击任务栏右侧的“小喇叭”图标，在弹出的快捷菜单中选择“录音设备”命令，打开“声音”对话框，如图 3-25 所示。

图 3-25　录制声源设置

(2) 如果录制设备中没有“麦克风”，在空白处右键单击，在弹出的快捷菜单中选择“显示禁用的设备”命令；右键单击“麦克风”，在弹出的快捷菜单中选择“启用”命令；再次右键单击“麦克风”，在弹出的快捷菜单中选择“设置为默认设备”命令。

(3) 单击“属性”按钮，打开“麦克风 属性”对话框。选择“级别”选项卡，调整“麦克风”至 70～100 之间，小喇叭处于开启状态。若麦克风音量太小，适当调节“麦克风加强”的数值，如图 3-26 所示。

图 3-26　麦克风属性设置

3.5.2　录音

打开 GoldWave 软件，单击“新建”按钮，在“新建声音”对话框中设置“声道数”为 2，“采样速率”为 44100，如图 3-27 所示。设置完毕，单击“确定”按钮。

图 3-27　新建声音参数设置

单击右侧“控制器”面板上的红色圆形按钮 开始录音。先录取一段长约 5s 的空白声音，然后开始朗读“再别康桥”。录音结束后单击红色方形按钮 。

选择“文件”→“保存”命令，在弹出的“保存声音为”对话框中选择保存的路径和文件格式，如图 3-28 所示。单击“保存”按钮，保存录音文件。

图 3-28　声音保存对话框

3.5.3　消除噪声

在录音的过程中，由于设备和环境的干扰，所录制的声音通常存在噪声，因此在对声音进行编辑之前需要消除噪声。如图 3-29 所示，选中的部分是语音的间隔时间，从波形上可以看出该时间内没有语音，但却有很多不规则的小幅度波形存在，这些波形就是噪声。

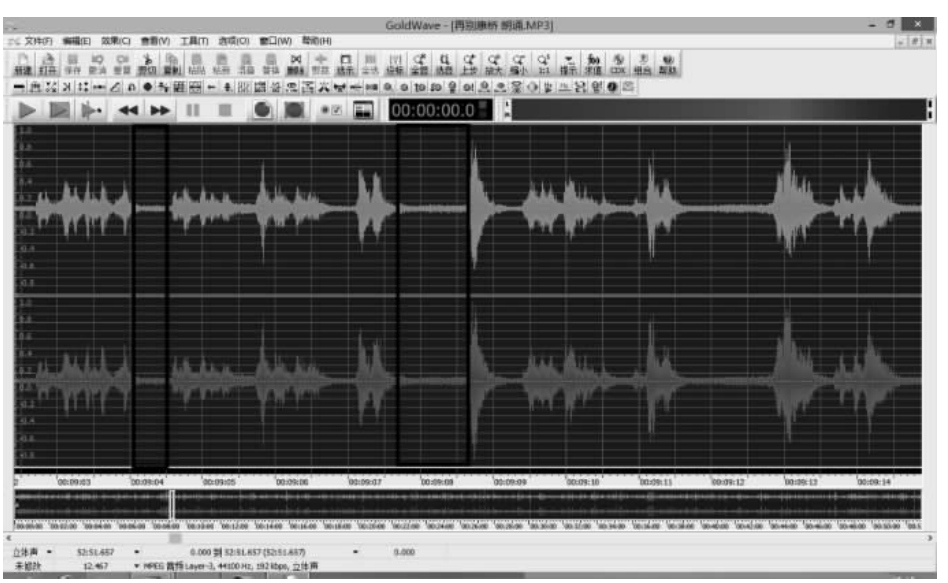


图 3-29　存在噪声的音频波形

消除噪声的步骤如下：

(1) 消除初始噪声。初始噪声是指录音前录制的一段噪音。在波形上右键单击，在弹出的快捷菜单中选择“选择全部”命令，选中整个波形。选择“效果”→“滤波器”→“降噪”命令，打开“降噪”对话框，如图 3-30 所示。在“降噪”对话框的“预置”下拉列表中选择“初始噪音”选项，单击“确定”按钮，可以消除录制波形中所有和初始噪音一样的波形，以达到降噪的效果。

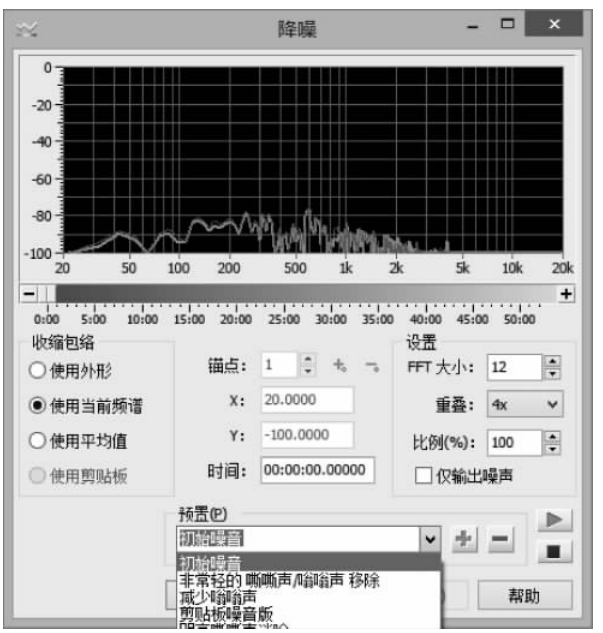


图 3-30 “初始噪音”降噪

(2) 消除录制过程中的噪声。右键单击选中的噪声部分，在弹出的快捷菜单中选择“复制”命令，这一过程叫作取样，可以将噪声选取出来。选择“效果”→“滤波器”→“降噪”命令，打开“降噪”窗口，如图 3-30 所示。在“收缩包络”选择区域中选择“使用剪贴板”单选按钮，在“预置”下拉列表中不要选择，最后单击“确定”按钮完成降噪。

(3) 删除无波形的时间段。查看整个声音波形，右键单击选中的波形中的直线部分，在弹出的快捷菜单中选择“删除”命令以删除直线波形，然后保存文件。

3.5.4 声音的修饰

声音的修饰是指按照作品的要求，对声音的音调和幅度进行调整，并添加淡入淡出和回声等特效的工作过程。本案例中需要降低音调、调整音量大小和进行过渡调整。

(1) 去除爆破音。在对着麦克风讲话时，有些字的声母发音时有突发式冲击波，造成急促的“爆破音”。选择“效果”→“滤波器”→“爆破音/嘀嗒声”命令，打开“爆破音/嘀嗒声”对话框，如图 3-31 所示。按照默认设置，单击“确定”按钮，可以使爆破音大大压缩。

(2) 音调调整。音调是指声音频率的高低。一般来说，女性的声带紧而薄，发出声调高；男性的声带松而厚，发出声调较低。为了使得朗读的声音浑厚沉稳，符合该散文表达的意境，需要将朗读声音音调适当降低。选择“效果”→“音调”命令，打开“音调”对话框，如图 3-32 所示。在“音阶”右侧文本框中输入 90，选中“保持速度”复选框，单击试听按钮 ▶ 试听效果。单击“确定”按钮，结束音调调整。

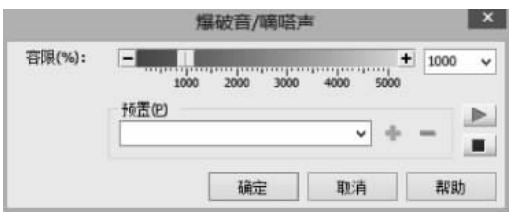


图 3-31　去除爆破音

图 3-32　降低音调

(3) 幅度调整。在录音时,声音波形幅度有一定范围,即最大值有一个限制,如果波幅超出这个极限,将只按最大值记录,那么超出的部分将被截去,这样的波形称为“截顶失真”,这也被称为音量的“过载”,如图 3-33 所示。一般来说,录音过程中声音宜小不宜大,波幅控制在 50%～80%之间。

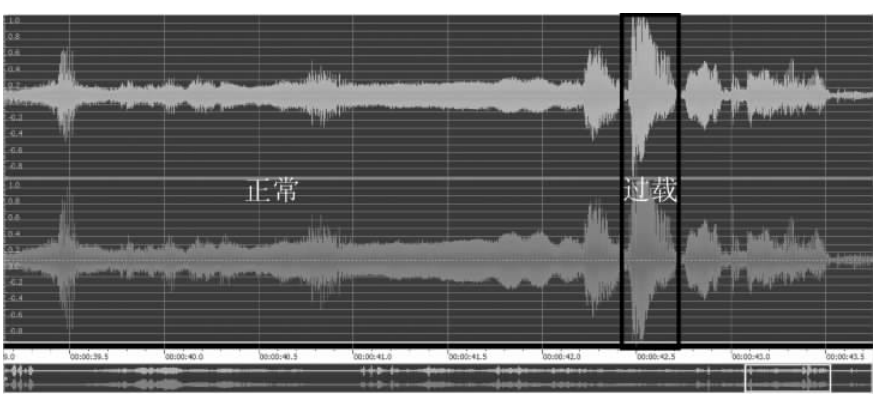


图 3-33　音量过载现象

使用放大工具放大波形并选择声音波形中音量明显偏小或明显过载的声音段,选择“效果”→“音量”→“更改音量”命令,打开“更改音量”对话框,如图 3-34 所示。当拖动滑块时,右边文本框中的数值也随着改变,显示出改变量的分贝数。注意,这里的分贝不是绝对声强,是相对分贝值,即与原声强比例的分贝数。也可以直接在右边文本框中输入数字。按照需要将音量小的波形调大音量,将音量过载的波形调小音量。如果想精细调节,可以单击滑块两端的 − 和 + 按钮,滑块会一点点地移动。

图 3-34　更改音量

(4) 封套调整。在对部分波形进行音量调整之后,调节点处会与未调节部分出现“台阶”,从而产生音量的突变。使用封套调节方法可完成调节处与未调处的平滑过渡。

选择“效果”→“音量”→“外形音量”命令,打开“外形音量”窗口,如图 3-35 所示。选中“显示包络”复选框后,窗口中即出现声音波形包络线,从包络线可以看出波幅大小,此处波形只显示音量大小的绝对值,所以没有负半周。

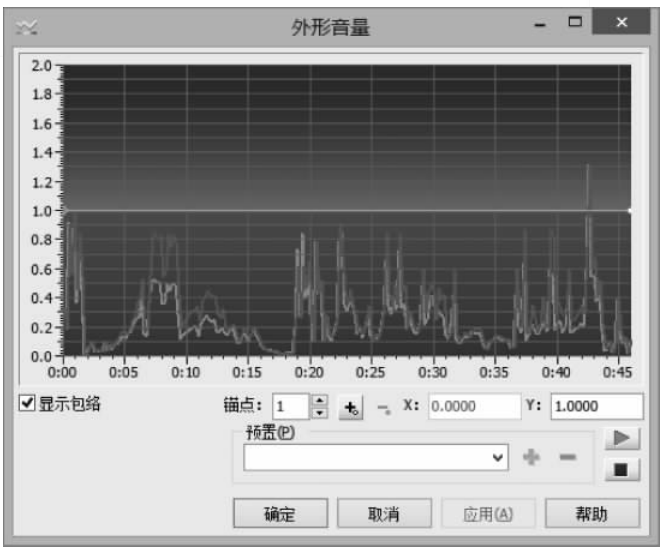


图 3-35 外形音量包络线

声音包络线上面一条横线是“封套线”,线两端各有一个小方块,叫作“节点”,调整线上节点高度可改变相应点的音量大小。在封套线上任意处单击即可在该处添加节点,添加节点后即可改变该点高度来调整相应声段的音量值。例如,在音量大到音量小的过渡处可以把左边音量大的部分压下来,把右边音量小的部分提上去,如图 3-36 所示。

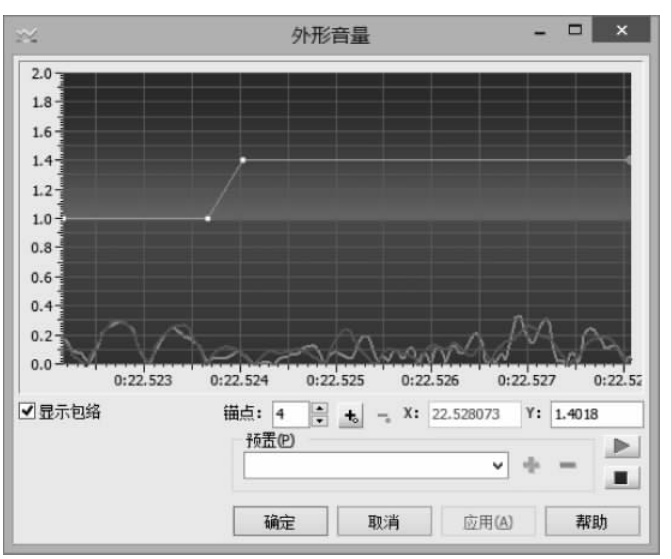


图 3-36 外形音量包络线的调整

图 3-36 中两个新添节点之间的斜线就是压缩音量与提升音量的过渡区。在节点上单击右键可消除该节点。

（5）动态调节。“外形音量”适用于分别改变声音的某一段局部波形，但不适用于改变混合在声音中的音量“台阶”。如果想把声音中的所有音量低的部分都提升上去，但这些波形并不存在于某一段中，可以使用“动态”调节。选择“效果”→“动态”命令，打开“动态”窗口，如图 3-37 所示。

“动态”窗口中 x 轴表示原音量值，y 轴表示调节值，黄色的斜线是“调节线”，和封套线一样可单击鼠标添加“节点”。从图 3-37 中可以看出 0 坐标点在正中间，调节线是在正负两区中间的线条，即表明声波的正负半周可以分别调整。如果要把音量大的声音压缩，把音量小的声音提升，使调节线可调节成如图 3-38 所示的形状。

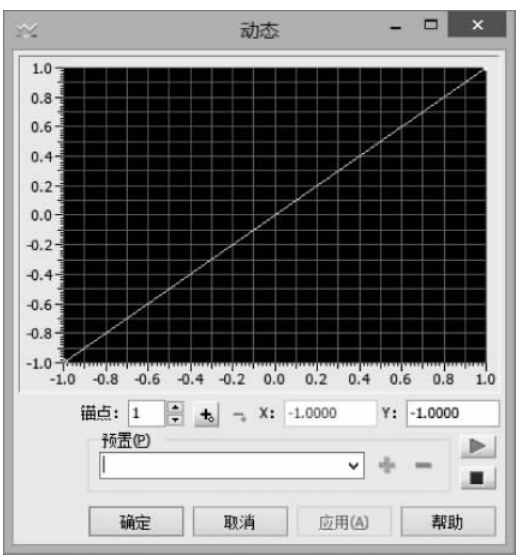


图 3-37　“动态”窗口

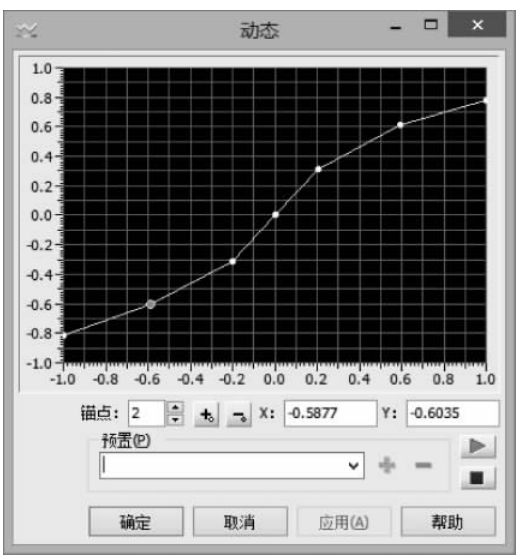


图 3-38　音量的动态调节

图 3-38 所示形状表明，原来音量为 1.0 的地方调到了 0.8，原来音量为－1.0 的地方调到了－0.8，原来音量为 0.2 的地方调到了 0.3，原来音量为－0.2 的地方调到了－0.3 等。如此就将声波中波幅高的部分降低了，声波中波幅小的部分提升了，使得声音整体音量保持较一致的水平，过渡自然。

3.5.5　混音

朗读声音调整好之后，可以为其添加背景音乐，实现配乐朗读的最终效果。混音的步骤如下：

（1）选择背景音乐。打开“雨的印记”背景音乐，在波形上右键单击，在弹出的快捷菜单中选择“选择全部”命令以选中整个波形。再次右键单击，在弹出的快捷菜单中选择“复制”命令复制整个波形。

（2）切换到“再别康桥 朗诵”声音波形，选择“编辑”→“混音”命令，打开“混音”对话框，如图 3-39 所示。可以单击右边的试听按钮 ▶ 确定混音开始的位置，并设置合适的音量大小。试听结束，单击“确定”按钮。

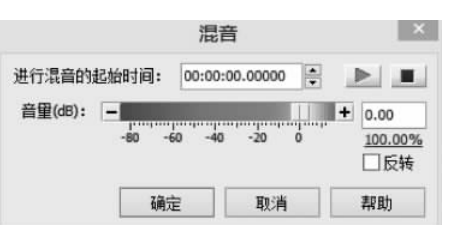


图 3-39　“混音”对话框

练习题

1. 名词解释

采样,采样频率,量化,声道数,编码

2. 单项选择题

(1) 下列要素中(　　)不属于声音的三要素。

A. 音调　　B. 音色　　C. 音律　　D. 音强

(2) MIDI 的音乐合成器有(　　)。

① FM　　② 波表　　③ 复音　　④ 音轨

A. 仅①　　B. ①②　　C. ①②③　　D. 全部

(3) 下列采集的波形声音中(　　)的质量最好。

A. 单声道、8 位量化、22.05kHz 采样率　　B. 双声道、8 位量化、44.1kHz 采样率

C. 单声道、16 位量化、22.05kHz 采样率　　D. 双声道、16 位量化、44.1kHz 采样率

(4) 在数字音频信息获取与处理过程中,下述顺序正确的是(　　)。

A. A/D 变换,采样,压缩,存储,解压缩,D/A 变换

B. 采样,压缩,A/D 变换,存储,解压缩,D/A 变换

C. 采样,A/D 变换,压缩,存储,解压缩,D/A 变换

D. 采样,D/A 变换,压缩,存储,解压缩,A/D 变换

(5) 一般来说,要求声音的质量越高,则(　　)。

A. 量化级数越低和采样频率越低　　B. 量化级数越高和采样频率越高

C. 量化级数越低和采样频率越高　　D. 量化级数越高和采样频率越低

(6) 用 Windows 自带的录音机录制的声音,默认保存的文件格式是(　　)。

A. WAV　　B. MP3　　C. AVI　　D. BMP

(7) 下列音频文件格式占用存储空间最大的是(　　)。

A. WAV　　B. MIDI　　C. CD-DA　　D. MP3

(8) 下列(　　)与数字化声音信号的质量无关。

A. 采样频率　　B. 量化位数　　C. 原始声音的质量　　D. 音强

(9) 声音是一种波,它的两个基本参数为(　　)。

A. 振幅、频率　　B. 音色、音高

C. 噪声、音质　　D. 采样率、采样位数

(10) 下述声音分类中质量最好的是(　　)。

A. 数字激光唱盘　　B. 调频无线电广播

C. 调幅无线电广播　　D. 电话

3. 填空题

(1) 通常人耳听力的频率范围是_____。

(2) 如果以 CD 激光盘音质(44 100Hz 的采样频率,16 位,立体声,172KB/s)记录一首 5 分钟的乐曲,那么其数据量为_____。

(3) 声音的三要素为_____、_____和音强。

(4) 音频数字化的过程为_____、_____和编码。

（5）常用的语音识别方法有三种，分别为_____方法、_____方法及利用人工神经网络方法。

（6）常用的录音方式有两种，即_____和_____。

4．问答题

（1）什么是声音？

（2）什么是采样频率？

（3）采样频率与声音还原频率存在什么关系？

（4）音频文件的数据量与哪些因素有关？

（5）回声效果是怎样产生的？

（6）阅读以下说明，回答下面三个问题。

在多媒体制作领域，音频素材是不可或缺的部分。可以利用外部声源设备通过声卡把声音输入计算机；通过软件对声音进行编辑、合成、音效处理等操作；通过音箱实现声音的输出。

问题1：声卡是连接计算机和外围声音设备的桥梁，请问声卡完成的主要功能是什么？声卡的位数表示什么？如果要通过声卡采集一台具备多种信号输出端子的CD机播放的音乐声音信号，应该如何连接CD机和声卡？

问题2：在使用音频处理软件录制声音素材时，常使用采样降噪法来降低声音素材中的噪音。请简要说明采样降噪法的基本原理。

问题3：Dolby AC-3数字音频编码技术提供了5.1声道的支持，其中的5指哪些声音通道?.1表示什么？

（7）阅读下列说明，回答下面三个问题。

在设计网络实时传输多媒体信息的应用系统时，必须准确计算媒体流的数据量，然后根据网络传输系统的实际情况来确定流式媒体的数据传输率等系统运行参数，从而在满足实时传输的条件下提供高质量的多媒体信息传输服务。假设你需要在1Mb/s带宽的网络上实现实时的立体声音频节目的播放，请考虑以下的应用需求，计算并解决问题。

问题1：如果系统设计的音频信号采样率是固定的44.1kHz，要实现实时的无压缩音频数据播放，在最好的质量下应该设置系统对音频信号的量化位数是多少？

问题2：如果系统设计的每个声道音频信号量化位数是固定的16位采样，要实现实时的无压缩音频数据播放，则：

① 在最好的质量下应该设置系统对音频信号的采样率是多少？

② 此时系统在保证不丢失频率分量的前提下能够传输的信号最高频率是多少？

问题3：如果应用系统需要实时播放CD音质的音频信号，那么必须选择使用或自行设计开发压缩编码器，定义压缩比＝压缩后的数据量/原始数据量，则选择使用的或设计开发的编码器的压缩比至少应该是多少？