工业和信息化普通高等教育
“十二五”规划教材立项项目

多媒体信息处理

卢官明 焦良葆 编著

21世纪高等院校信息与通信工程规划教材
21st Century University Planned Textbooks of Information and Communication Engineering

Multimedia
Information Processing

人民邮电出版社
POSTS & TELECOM PRESS

精品系列

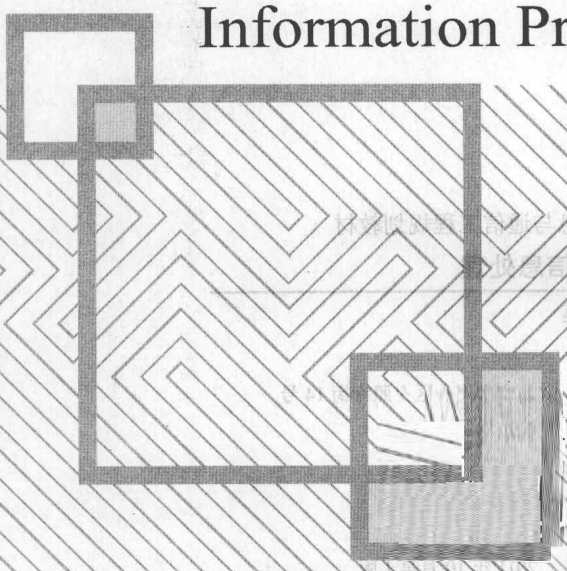
工业和信息化普通高等教育
“十二五”规划教材立项项目

多媒体信息处理

卢官明 焦良葆 编著

21世纪高等院校信息与通信工程规划教材
21st Century University Planned Textbooks of Information and Communication Engineering

Multimedia Information Processing



人民邮电出版社
北京

精品系列

前言

人类社会正进入信息社会的历史阶段。在我国，随着国家信息化发展战略的贯彻实施，信息化建设已进入了全方位、多层次推进应用的新阶段，各行各业的信息化进程不断加速。随着信息技术的发展，传统的信息处理方式和表现手段已经难以适应社会的需要。作为现代科学技术发展的最新成就，多媒体信息处理为媒体的集成和信息的传播提供了丰富的手段，在信息社会的发展过程中将起到极其重要的作用。多媒体信息处理正在给人类的学习、工作、生活和娱乐方式带来深刻的革命，其应用已渗透到社会生活的方方面面。随着人们在日常生活中对多媒体信息需求的不断增强，多媒体信息处理在近年来得到了迅速的发展，成为当前科学研究和应用开发的热点。

由于多媒体信息处理具有极强的应用价值和广阔的发展前景，因而得到了学术界和产业界的广泛关注。国内外大多数高等院校陆续开设了多媒体信息处理方面的课程，社会上各类继续教育机构还纷纷开展了相关技术的培训。虽然有关数字图像处理的教材很多，也不乏经典；有关音频信息处理或视频信息处理的书籍也有一些，但作为教材的不多，一般以专著的形式出版；有关多媒体信息处理的教材也有很多，但都是偏重多媒体软件设计技术、多媒体数据库技术以及一些具体的多媒体应用软件，如多媒体编辑和创作工具、多媒体网页制作等，而全面介绍多媒体信息处理的教材目前还鲜有见到。近年来，作者始终关注着多媒体信息处理相关技术的发展，并一直致力于该领域的教学与研究工作，同时，多次为本科生讲授多媒体信息处理课程，深感出版一本《多媒体信息处理》实有必要。

编写本教材的指导思想是：将音频、图像、视频等媒体信息的处理技术有机地整合在一起，揭示其内在的联系，以便让学生在有限的学习时间内掌握更系统、更全面的知识。出发点如下。

（1）图像与视频两者密切相关，动态的视频是由一系列静态的图像组成的，有关的视频处理技术是在数字图像处理技术的基础上发展起来的。

（2）语音、音频、图像以及视频信号的压缩编码原理在本质上也是相通的。

（3）音频、图像以及视频的文件格式基本具有相似的结构。

（4）多媒体信息处理技术本身就是计算机综合处理文本、图像、图形、动画、音频、视频等多种媒体信息的技术，它使人们能以更加自然的方式使用信息，并与计算机进行交互，使表现的信息图、文、声并茂。

本书的编著力图体现以下特点。

（1）取材精选，内容新颖。本书精选了多媒体信息处理领域所涉及的经典方法与关键技术内容，介绍了相关领域的最新研究成果及发展新动向，如数字水印技术、基于内容的多媒体信息检索技术等。

（2）重点突出，注重实用。本书以掌握基本原理、强化应用为重点，在强调基本概念、基本原理的同时，注重理论与实际应用相结合，列举了大量具有实际应用价值的 MATLAB 编程实例，使读者能较快地掌握多媒体信息处理系统的基本理论、方法、实用技术及一些典型应用，学以致用。

（3）条理清晰，通俗易懂。针对多数学生的学习特点，采用通俗易懂、深入浅出的方法讲解知识，逻辑性强、层次分明、叙述准确而精练、图文并茂，适合教学与自学。每章都附有小结与习题，有助于读者理解和掌握所学的知识要点。

本书既可作为高等院校电子信息工程、通信工程、电子科学与技术、计算机应用、广播电视工程等专业的高年级本科生或研究生的教材或教学参考书，也可供从事多媒体信息处理领域工作的研究与开发人员参考。

在本书的编著过程中，作者参考和引用了一些学者的研究成果、著作和论文，具体出处见参考文献。在此，谨向这些文献的著作者表示敬意和感谢！

本书的第 3 章、第 4 章、第 8 章的大部分内容由焦良葆、童莹编写，其余各章由卢官明编写，全书由卢官明统审、定稿。鉴于作者水平所限，加之多媒体信息处理涉及面广，相关技术发展迅速，书中难免存在不妥之处和个人之拙见，敬请广大读者批评指正，提出宝贵意见和建议。

作　者

2011 年 4 月

目　录

第1章 多媒体信息处理基础

本章学习目标

- 熟悉多媒体及多媒体技术的基本概念及特征。
- 了解声音的基本特性及主观感觉。
- 熟悉音频、图像、视频信号数字化的过程，掌握均匀量化的原理。
- 掌握彩色三要素、三基色原理及混色方法等色度学基本知识。
- 理解 RGB、YUV、YIQ、YC_bC_r、HSI/HSV 等颜色空间的表示及转换。
- 熟悉 ITU-R BT.601 建议的主要内容。
- 了解 MATLAB 在图像处理和分析领域的应用。

1.1 多媒体的基本概念

1.1.1 媒体的概念

要弄清什么是多媒体，首先要知道什么是媒体（Medium）。按传统的说法，媒体指的是信息的载体，如日常生活中的报纸、电视、广播、广告、杂志等，信息借助于这些载体得以交流传播。如果对这些媒体的本质进行详细分析，就可找到媒体传递信息的基本元素，如声音、图形、视频、图像、动画、文字等。它们都是表示信息的媒体。另外，在计算机、通信领域中，“Medium”曾被广泛译作“介质”、“媒介”，指的是信息的存储实体和传输实体。

根据信息被人们感知、加以表示，使之呈现、实现存储或进行传送的载体的不同，原国际电报电话咨询委员会（CCITT，现改称为国际电信联盟电信标准化部门 ITU-T）将媒体定义为以下 5 种。

- 感知媒体（Perception Medium）：感知媒体能够直接作用于人的感官，使人产生感觉。人类利用视觉、听觉、触觉、嗅觉和味觉来感受各种信息。因此感知媒体可以分为视觉媒体、听觉媒体、触觉媒体、嗅觉媒体和味觉媒体。
- 表示媒体（Representation Medium）：表示媒体是为加工、处理和传输感知媒体而人为构造出来的一种媒体。其目的是更有效地加工、处理和传输感知媒体。表示媒体包括各种编码方式，如语音编码、视频编码、图像编码等。
- 呈现媒体（Presentation Medium）：呈现媒体的作用是将媒体信息的内容呈现出来。

它分为两种：一种是输入类呈现媒体，是指获取信息的工具和设备，如键盘、鼠标、扫描仪、摄像机、光笔、话筒等；另一种是输出类呈现媒体，意指为人们再现信息的物理工具和设备，如显示器、扬声器、打印机等。

- 存储媒体（Storage Medium）：存储媒体是用于存储表示媒体的物理介质，用以方便计算机处理、加工和调用，这类媒体主要是指与计算机相关的外部存储介质，如磁盘、光盘、磁带等。
- 传输媒体（Transmission Medium）：传输媒体是用来将表示媒体从一处传输到另一处的物理媒介，如双绞线、同轴电缆、光纤、微波等。

在多媒体技术中所说的媒体一般指感知媒体。目前，在多媒体系统中经常处理的感知媒体是视觉媒体和听觉媒体，触觉媒体也开始被引入到虚拟现实系统中，嗅觉媒体和味觉媒体尚不能在计算机中处理。

1．视觉媒体

视觉媒体包括图像、视频、图形、动画、符号与文本等，它们是通过视觉来传递信息的。

（1）图像

图像在人类的感知中扮演着非常重要的角色，人类随时随处都能看见图像。图像（Image）是指采用各种观测系统获得的、能够为人类视觉系统所感知的实体，是指各种图片和影像的总称，包括照片、X 光片、遥感图片、电视画面、绘图等。

事实上，无论是图形、文字还是视频，最终都是以图像的形式出现的。但由于计算机中对它们的表示、处理、显示方法不同，一般又把它们看做不同的媒体形式。这里所说的图像特指位图（Bitmap）。

位图图像是用像素点来描述或映射的图，是指在空间和亮度上已经离散化了的图像。可以把一幅位图图像看做一个矩阵，矩阵中的任一元素对应图像中的一个像素点，相应的值对应该点的灰度（或颜色）等级，这是量化后得到的结果。这个数字矩阵的元素称为像素，存放于显示缓冲区中，与显示器上的显示点一一对应，故称为位映射图，简称位图。位图图像适合于表现层次和色彩比较丰富、包含大量细节的图像。

（2）视频

视频（Video）又称动态图像，是一组图像按时间顺序的连续表现。视频的表示与图像序列、时间关系有关。

（3）图形

图形（Graphics）是一种抽象化的图像，一般指用计算机绘制的几何画面，如直线、圆、圆弧、矩形、任意曲线和图表等。它不直接描述图像的每一点，而是描述产生这些点的算法。图形文件是以一组指令的形式存在的，这些指令描述一幅图中所包含的直线、圆、弧线、矩形的大小和形状，也可以用更为复杂的形式表示图像中曲面、光照、材质等效果，如 line(x1,y1,x2,y2,color) 和 circle(x,y,r,color) 分别是画线、画圆的指令。因此，图形也被称为矢量图形。在计算机上显示一幅图形时，首先要由绘图程序解释这些指令，然后将它们转变成屏幕上所显示的形状和颜色。由于在图形文件中只记录生成图的算法和图上的某些特征点，所以相对于图像文件的大数据量来说，它占用的存储容量较小，但这是以显示中的计算时间为代价的。图形的最大优点在于可以分别控制处理图中的各个部分，如在屏幕上移动、旋转、

放大、缩小等，仍保持图形特性，不同的物体还可在屏幕上重叠并保持各自的特性，必要时仍可分开，而这一点对于位图图像来说就十分困难。因此，图形主要用于表示线框型的图画、工程制图和美术字等。绝大多数 CAD 和 3D 造型软件使用矢量图形来作为基本图形存储格式。

（4）动画

动画（Animation）是动态图像的一种，它与视频的不同之处在于，动画中的图像采用的是计算机产生出来或人工绘制的图像或图形，而视频中的图像则是真实的图像。也就是说，动画是运动的图形，其实质是一幅幅静态图形的连续播放。动画的连续播放既指时间上的连续，也指内容上的连续，即播放的相邻两幅图形之间内容相差不大。

（5）符号与文本

符号（Symbol）是人类对信息进行抽象的结果。量的值可以用 1、2、3 等数值符号来表示，逻辑的真、假、大于、小于等也可以用专门的符号来描述。虽然我们看到的是“图像”，但由于大脑知识加工的结果，它们都被抽象成了特定的符号而被识别出来。由于符号是人类创造出来表示某种含义的载体，所以它与使用者的知识有关，是比图形更高一级的抽象。只有具备特定的知识，才能解释特定的符号。

由于符号具有明显的结构性，大脑可以识别这种结构，从而可以识别出由这一组符号所代表的信息。文本（Text）是用得最多的一种符号媒体形式，是人类创造出来用于记述信息的工具，由具有上下文关系的字符串所组成。

符号在计算机中用特定的代码表示，如 ASCII 码、中文国标码等。计算机在进行文字处理时，依据的就是对字符代码的识别，它是文本处理程序的基础，也是多媒体应用程序的基础。那些用图像方式显示的字符，虽然人可以识别，但由于没有使用字符代码，所以并不属于文本信息。

2. 听觉媒体

听觉媒体包括波形声音、语音和音乐。

（1）波形声音

波形声音实际上包含了所有的声音形式。自然界中的各种声音，包括人的说话声、音乐、天空的惊雷、山林的狂风、大海的涛声等，也包括各种噪声，可以用一种模拟的连续波形表示。通过对模拟的声音信号进行采样、量化和编码，可生成数字的波形声音数据。

对声音的处理，主要包括声音的采集、数字化、压缩/解压缩以及声音的播放。

（2）语音

语音（Speech）也叫话音（Voice），是人类为表达思想通过发音器官发出的声音，是人类语言的物理形式，它由一连串的音素组成。“一句话”中包含了许多音节及其相互间过渡过程的连接体。当发音系统产生一定的运动，并借助于介质点的振动而传播时，便形成了语音。

语音不仅是一种波形声音，而且还具有内在的语言、语音学内涵，可以经由特殊的方法而提取，即进行一次抽象。波形声音也可以表现和记录语音，但常把语音作为一种特殊的听觉媒体。对语音的符号化表示实际上就是对语音的识别，将语音转变为文本；反之，就是将文本合成为语音。对语音的符号化有助于对声音内容的进一步处理。

（3）音乐

音乐与语音相比，形式更为规范。事实上，音乐是符号化了的声音，这种符号就是乐谱。

在多媒体计算机中，MIDI（Musical Instrument Digital Interface）就是一种乐谱数字化描述的规范。任何电子乐器，只要有处理 MIDI 信息的微处理器，并有合适的硬件接口，都可以成为一个 MIDI 设备。在这里，乐谱完全由音符序列、定时以及被称为合成音色的乐器定义组成。当一组 MIDI 信息通过音乐合成器演奏时，合成器就会解释这些符号并产生音乐。

与波形声音相比，MIDI 数据不是声音而是指令，所以它的数据量要比波形声音少得多。

3. 触觉媒体

除了视觉、听觉之外，触觉在人类的信息传递和交流中同样起着十分重要的作用。我们的皮肤可以感觉环境的温度、湿度，也可感觉压力，我们的身体可以感觉振动、运动、旋转等，这些都是触觉在起作用。所以，触觉也可以作为传递信息的媒体。例如，去推一扇门，这扇门给予推门的手的反作用力，就传递了门是开或关的状态信息。当环境被确定时，也就提供了一个特定的触觉信息传递的范围。又如，当置身于驾驶室中时，身体所感到的振动是车辆行走时的振动，脚踩刹车的反作用力反映了环境对控制的反应等。很显然，当在信息系统中引入了触觉媒体后，人与环境在信息交互的方式上更进了一步。人类与外界环境的触觉交互主要包括位置跟踪、力量反馈等方面。目前对触觉媒体的研究还在初级阶段，本书不讨论触觉媒体信息的处理技术。

1.1.2 多媒体与多媒体技术

多媒体译自英文“Multimedia”，该词由 Multiple（多）和 Media（媒体）复合而成。在多媒体技术中所说的“多媒体”，主要是指多种形式的感知媒体。顾名思义，多媒体意味着非单一媒体，是两个或两个以上的单一媒体的有机组合，指的是文本、图形、图像、动画、视频、语音、音乐或数据等多种形态信息的处理和集成呈现。广义的“多媒体”，则同时也包括上述其他几种媒体。在现代社会中，多媒体信息的获取、处理、存储、传输和呈现等过程，是离不开计算机的。所以，人们指的多媒体，首先是计算机处理的数字化的多媒体。

需要指出的是，一般所说的“多媒体”，不仅指多种媒体信息本身，而且包含处理和应用多媒体信息的相应技术，因此“多媒体”常被当做“多媒体技术”的同义词。多媒体技术就是利用计算机技术把文本、图像、图形、动画、音频及视频等多种媒体有机地集成起来，使人们能以更加自然的方式使用信息，并与计算机进行交互，且使表现的信息图、文、声并茂。简言之，多媒体技术就是计算机综合处理声、文、图信息的技术，具有集成性、实时性和交互性。

1.2 音频信息处理基础

音频信息涉及人耳所能听到的声音信息，包括语音、音乐和自然声音。声音的传播携带了大量的信息，是人类传播信息的一种主要媒体。据统计，人类从外界获得的信息大约有 16% 是从耳朵得到的。在多媒体技术中，音频信息占有很重要的地位。例如在会议电视系统中，音频信息是优先级最高的信息通道。了解音频信息的相关知识对人们更进一步掌握多媒体技术是很重要的。

1.2.1 声音的基本特性

人们在日常生活中听到各种各样的声音，它们都是机械振动或气流振动引起周围传播媒质（气体、液体、固体等）发生波动的现象，通常将产生声音的发声体称为声源。当声源体产生振动时，引起相邻近的空气的振动。这样空气就随着声源体所振动幅度的不同，而产生密或稀的振动，空气的这种振动被称为声波。声波所及的空间范围被称为声场。声波传到人耳，经过人类听觉系统的感知就是声音。

声波可以用一条连续的曲线来表示，它可以分解成一系列正弦波的线性叠加，其数学表达式为

$$A(t)=\sum_{k=0}^{\infty}A_k\sin\left(2\pi kft+\phi_k\right) \tag{1-1}$$

式中，f为基音频率，它决定了声音音调的高低；A_k为第k次谐波分量的振幅，与声音的响度有关；kf为谐音的频率，与声音的音色有关；ϕ_k为第k次谐波的初始相位。

1．频率

频率是单位时间内信号振动的次数，一般用f表示，单位是赫兹（Hz）。

声波的频率范围相当宽，从10^{-4}～10^{12} Hz。人们把频率低于 20 Hz 的声波称为次声波；频率高于 20 kHz 的声波称为超声波，这两类声音是人耳听不到的。人耳可以听到的声音是频率在 20 Hz～20 kHz 之间的声波，人们称之为音频（Audio）信号。而人的发音器官发出的声音频率在 80～3400 Hz 之间，但人说话的信号频率通常在 300～3400 Hz 之间，人们把这种频率范围的信号称为语音（或话音）信号。

在多媒体应用领域，按照对声音质量的要求不同以及使用频带的宽窄，将音频信号通常分为以下 4 类。

（1）窄带语音

窄带语音，又称电话频带语音，其信号频带为 300～3400 Hz，用于各类电话通信。

（2）宽带语音

宽带语音信号的频带为 50～7000 Hz。它提供了比窄带语音更好的音质和说话人特征，用于电话会议、视频会议等。

（3）数字音频广播（Digital Audio Broadcasting，DAB）信号

DAB 信号的频带为 20～15000 Hz。

（4）高保真立体声音频信号

高保真立体声音频信号的频带为 20～20000 Hz。用于 VCD（Video Compact Disc，视频高密度光盘）、DVD（Digital Versatile Disc，数字通用光盘）、CD（Compact Disc，激光唱盘）、HDTV（High Definition Television，高清晰度电视）伴音等。

2．频谱

现实世界中有各种各样的声音。其中，频率单一、声压随时间按正弦函数规律变化的声音称为纯音。在自然界和日常生活中很少遇到纯音，大多数的声音是由多个频率成分组合而成的复合音，如语言、音乐或噪声。纯音可由音叉产生，也可用电子振荡电路或音响合成器

产生。复合音是由频率不同、振幅不同和相位不同的正弦波叠加形成的，它也是一种周期性的振动波。在复合音中频率最低的成分（分音）称为基音。频率与基音成整倍数的分音称为谐音（谐波）。频率为基音的 2 倍或 3 倍的分音分别称二次或三次谐音。

声音的频谱结构是用基音、谐音数目、各谐音幅度大小及相位关系来描述的。声音的音色就是由其频谱成分决定的，音调相同而音色不同的声音就是由于它们的谐音数目、谐音振幅及其随时间衰减的规律不同而造成的。各种乐器都有其特定的音色。

3. 声压及声压级

对于空气媒质，当没有声波时，空气处在平衡状态，其静压强一般等于大气压。当有声波传播时，媒质各部分能产生压缩和膨胀的周期性变化。压缩时压强增加，大于静压强，这时压强差为正；膨胀时压强减小，小于静压强，这时压强差为负。这一交变的压强差即为声压。声压的大小反映了声音振动的强弱，同时也决定了声波的幅度大小。更具体的描述变化部分压强，可以用瞬时声压、峰值声压和有效声压等。瞬时声压是某点的瞬时总压强减去静压强。在某一时间间隔中最大的瞬时声压，称为峰值声压。在一定时间内，瞬时声压对时间取均方根值，称为有效声压。对于周期波，在某一周期内的极大声压是这一周期中瞬时声压的极大绝对值；如所取时间等于整个周期，峰值声压就和极大声压相同。对于简谐波，峰值声压是声压的幅值，等于有效声压的 $\sqrt{2}$ 倍。如果没有特别说明，一般所称的声压指的就是有效声压，用电子仪器测量得到的通常是有效声压。声压一般用符号 p 表示，单位是帕（Pa）或微巴（μbar）。

实验表明，人们对声音强弱的主观感觉并不正比于声压的绝对值，而是大致正比于声压的对数值。另外，人耳能听到的声压范围非常之大，从能听到的最小声压 2×10^{-5} Pa 到能承受的最大声压 20 Pa，两者相差高达 100 万倍。所以，用声压的绝对值来表示声音的强弱显然是很不方便的。基于以上两方面的原因，常采用按对数方式分级的办法表示声音的强弱，这就是声压级。

声压级用符号 L_p 表示，单位是分贝（dB），可用式（1-2）计算，即

$$L_p = 20\lg\frac{p}{p_{ref}} \tag{1-2}$$

式中，p 为有效声压值；p_{ref} 为基准声压，一般取 2×10^{-5} Pa，这个数值是人耳所能听到的 1 kHz 声音的最低声压，低于这一声压，人耳就无法觉察出声波的存在了。

1.2.2 声音的主观感觉

当声波传播到人的听觉器官——人耳处时，耳膜受到相应的声压变化而对听觉神经产生刺激，该刺激通过神经系统传入大脑听觉中枢形成感觉，使人感到声音的存在。并非所有声波都能被人耳听觉所感知，甚至即使对人耳能感知到的声音，其感觉也各有不同，因为人的听感是一个非常复杂的物理—生理—心理过程。人对声音的感知有响度、音调和音色三个主观听感要素。人的主观听感要素与声波的客观物理量——声压、频率和频谱成分之间既有着密不可分的联系，又有一定的区别。声音的响度与声波振动的幅度有关，音调高低取决于声波的基音频率，音色由声波的频谱成分决定。

1. 响度

响度是人耳对声音强弱的主观感觉程度。在客观的度量中，声音的强弱是由声波的振幅（声压）决定的。但是，响度与声波的振幅并不完全一致，对于同一强度的声波，不同的人听到的效果并不一致，因而对响度的描述有很大的主观性。一般来说，在人类听觉的动态范围内，响度同声压级大体成比例，即声压级越大响度也越大，但这只对同一频率的声音来说是正确的。实验表明，声压级不是决定响度的唯一因素，另一个重要因素是频率。举一个极端的例子，频率极低的纯次声和频率极高的纯超声，无论其声压级有多大，我们都会觉得它“不响”。事实上，即使在可听声的频率范围（20 Hz～20 kHz）内，对于声压级相同而频率不同的声音，人们听起来也会感觉不一样响。对强度相同的声音，人耳感受 1～4 kHz 之间频率的声音最响，超出此频率范围的声音，其响度随频率的降低或上升将减小。

为了对响度进行计量，定义响度的单位为宋（sone）。国际上规定：频率为 1 kHz 的纯音在声压级为 40 dB 时的响度为 1 宋（sone）。

大量统计表表明，一般人耳对声压的变化感觉是，声压级每增加 10 dB，响度增加 1 倍，因此响度与声压级有如下关系

$$N = 2^{0.1(L_p-40)} \tag{1-3}$$

式中，N 为响度（单位为 sone），L_p 为声压级（单位为 dB）。

2. 响度级

人耳对声音强弱的主观感觉还可以用响度级来表示。响度级的单位为方（phon），一般用符号 L_N 来表示。以 1 kHz 的纯音为基准声音，将其他频率的纯音和 1 kHz 纯音相比较，调整前者的声压级，使得听者判断两个纯音一样响，则称该纯音的响度级（phon 值）在数值上与那个等响的 1 kHz 的纯音的声压级（dB 值）相等。例如，1 kHz 纯音的声压级为 0 dB 时，响度级定为 0 phon；声压级为 40 dB 时，响度级定为 40 phon，响度为 1 sone。

响度级 L_N 与响度 N 之间的换算公式为

$$L_N = 40 + 10\log_2 N \tag{1-4}$$

从响度及响度级的定义可知，响度级每增加 10 phon，响度增加 1 倍。声压级与响度、响度级的关系如表 1-1 所示。

表 1-1　　声压级与响度、响度级的关系

响度/sone	1	2	4	8	16	32	64	128	256
声压级/dB	40	50	60	70	80	90	100	110	120
响度级/phon	40	50	60	70	80	90	100	110	120

3. 等响度曲线

由于响度是指人耳对声音强弱的一种主观感觉，因此，当听到其他任何频率的纯音同声压级为 40 dB 的 1 kHz 的纯音一样响时，虽然其他频率的声压级不是 40 dB，但也定义为 40 phon。这种利用与基准音比较的实验方法，测得一组一般人对不同频率的纯音感觉一样响的响度级、声压级与频率三者之间的关系曲线，称为等响度曲线。图 1-1 所示为国际标准化

组织的等响度曲线，它是对大量具有正常听力的青年人进行测量的统计结果，反映了人类对响度感觉的基本规律。

曲线中的每一条等响度曲线对应一个固定的响度级值，即 1 kHz 纯音对应的声压级。

从对等响度曲线分析可得出如下结论。

① 对于某一确定的频率，响度级与人耳处的声压级有关。声压级提高，相应的响度级随之增大。对于 1 kHz 的纯音，响度级的值等于声压级的值。

② 人耳对频率在 3～4 kHz 范围内的声音响度感觉最灵敏，而对 100 Hz 以下的低频声不敏感。

③ 当响度级较小时，等响度曲线上各频率声音的声压级相差很大。例如，频率为 30 Hz 的声音达到 10 phon 响度级时，需有约 65 dB 的声压级；而对于频率为 10 kHz 的声音，相同的响度级只需约 20 dB 的声压级，两者的声压级差约 45 dB。

④ 响度级越高，等响度曲线越趋于平坦。当响度级高于 100 phon 时，等响度曲线逐渐拉平。这说明当声音达到一定强度（>100 phon），声音的响度决定于声压级，而与频率关系不太大。

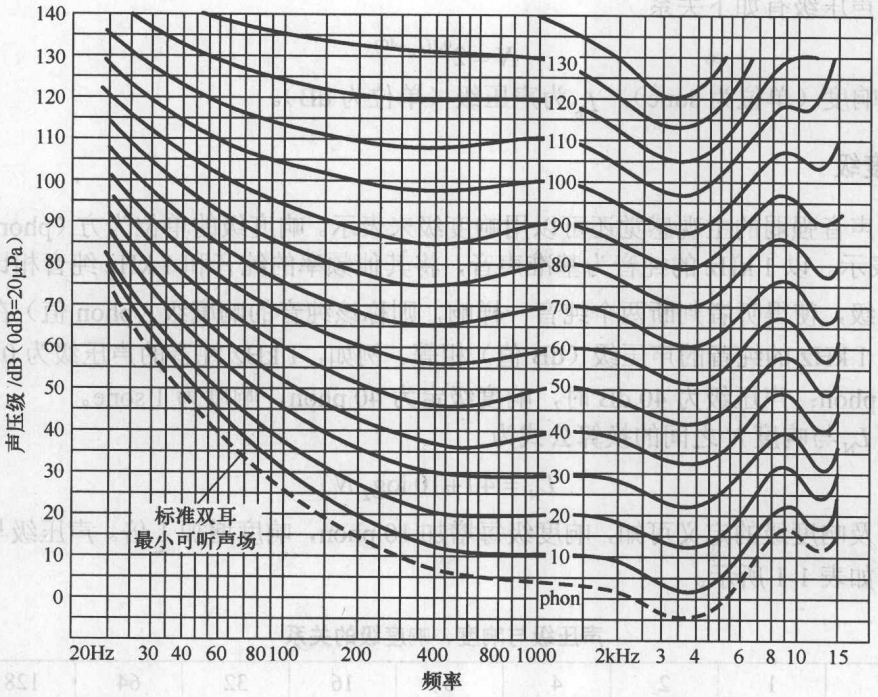


图 1-1 人耳听觉的等响度曲线

4. 听阈与痛阈

人耳对于声音细节的分辨与响度直接有关：只有在响度适中时，人耳辨音才最灵敏。正常人听觉的声压级范围为 0 dB～140 dB（也有人认为是−5 dB～130 dB）。固然，超出人耳的可听频率范围的声音，即使声压级再大，人耳也听不到声音。但在人耳的可听频率范围内，若声音弱到或强到一定程度，人耳同样听不到。当声音减弱到人耳刚刚可以听见时，此时的声音强度称为最小可听阈值，简称为“听阈”。一般以 1 kHz 纯音为准进行测量，人耳刚能听

到的声压级为 0 dB（通常大于 0.3 dB 即有感觉）。图 1-1 中最下面的一条等响度曲线（虚线）描述的是最小可听阈值。而当声音增强到使人耳感到疼痛时，这个听觉阈值称为“痛阈”。仍以 1 kHz 纯音为准来进行测量，使人耳感到疼痛时的声压级约达到 140 dB。

实验表明，听阈和痛阈是随频率变化的。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。

小于 0 dB 听阈和大于 140 dB 痛阈时为不可听声，即使是人耳最敏感频率范围内的声音，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 3～4 kHz 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（如小于 800 Hz）和高频区（如大于 5 kHz），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。

5. 音调

音调也称音高，表示人耳对声音调子高低的主观感觉。以客观的物理量来度量，音调与声波基频相对应，一般来说，频率低的调子给人以低沉、厚实、粗犷的感觉，而频率高的调子则给人以亮丽、明快的感觉。音调与频率有正相关的关系，但没有严格的比例关系，且因人而异。

6. 音色

音色主要是由声音的频谱结构决定的。一般来说，声音的频率成分（谐波数目）越多，音色便越丰富，听起来声音就越宽广、感人肺腑、扣人心弦、娓娓动听。如果声音中的频率成分很少，甚至是单一频率，音色则很单调乏味、平淡无奇。各种发声物体或乐器在发出同一音调的声音时，所发出的声音之所以不同，就在于虽然基波相同，但谐波的多少不同，并且各次谐波的幅度各异，因而具有各自的声音特色。我们每个人的声带和口腔形状不完全一样，因此，说起话来也各有自己的特色，使别人听到后能够分辨出谁在说话。

如果声音经传输后频谱有了变化，则重放出的声音音色就会改变。为了使声音逼真，必须尽量保持原来的音色。声音中某些频率成分被过分放大或压缩都会改变音色，从而造成失真。

在语音处理系统中，最重要的是保持良好的清晰度。适当减少一些低音和增加一些中音成分，特别是鼻音或喉音很重的人改变低频部分的音色，有利于达到改善语音清晰度的要求。

1.2.3 音频信号的数字化

自然界中的声音是一种连续变化的模拟信号，而计算机只能处理和存储离散的数据。如果要用计算机对音频信息进行处理，则首先要将模拟音频信号（如语音、音乐等）转换为数字信号。

音频信号的数字化一般需要完成采样、量化和编码三个步骤，如图 1-2 所示。采样是指用每隔一定时间间隔的信号样本值序列代替原来在时间上连续的信号，也就是在时间上将模拟信号离散化。采样后的信号在时间域上是离散的，但在幅度上仍保持着连续的特点，所以要进行量化。量化是用有限个幅度值近似原来连续变化的幅度值，把模拟信号的连续幅度变为有限数量、有一定间隔的离散值。编码则是按照一定的规律，把量化后的离散值用二进制数码表示。上述数字化的过程又称为脉冲编码调制（Pulse Code Modulation，PCM），通常由模/数（A/D）转换器来实现。