



## 四、混音

1. 混音的介绍：需要同时播放多重声音（简单加法混音、求平均值、查表求对数或更复杂的数学方法）、交由操作系统或者DirectX Audio、FMOD等音频引擎处理

## 数字音频技术发展简史

### 一、频率调制（FM）合成技术

早期PC音效使用频率调制FM合成技术，它使用某种算法来重建声波，**并不存储实际的样本数据**，而是用数学近似方法重建不同形状、频率和音量的声波，对于长笛等呈正弦波的声波等效果较好，对于其它发声体效果非常差。

### 二、MOD格式（波表合成的初步形式）

1. Amiga的MOD格式普及：MOD不同于上面的FM合成采用数学近似方法，它**使用真实的声音样本**、体积小。最初支持4轨迹，每轨迹可以在任意时间、音调播放一个样本。
2. MOD格式由一组采样、曲谱和时序信息组成，告诉播放器以何种音高去播放某条音轨上的某个样本，并附带一些演奏效果如颤音等。它是**波表合成（Wave Table Synthesis）技术**的初步形式。
3. 优点：介于WAV（纯正样本数据文件）和General MIDI（纯正时序信息文件）之间的混合体，**比较灵活**。缺点：格式变化太多，**mod格式的不统一导致了商业上的无作为**。
4. 当下MOD的应用：各种软件的注册机、算号器。MOD的制作需要用到音序器，一般叫做Tracker

### 三、波表合成技术

1. 什么叫波表合成：录制下来每种乐器的声音，**以数字化形式存放**，通过选择哪件乐器演奏、音量、音速、混合和平移参数等来奏响记录下来的声波来合成出新的声音。
2. 分为**软波表**（用软件代替声卡上的波表合成器，通过CPU的实时运算来回放MIDI音效）和**硬波表**（**取样文件集成在声卡上**，通过声卡上的微处理器来处理发声）。
3. 波表合成的**本质**：对采样声音的**调制**。对采样声音进行调制的过程是**动态地改变音频信号的某个参数**的过程，这些参数包括：音量（振幅调制或震音）、音高（调频或颤音）以及滤波器截止频率。调制的作用是**实时控制参数**，使声音产生变化，因而不需添加更多的样本也能得到丰富的声音效果，以起到节省内存的目的。

## 四、MIDI格式

1. MIDI格式：MIDI音乐，电脑音乐的代名词，指**使用电脑和数字音乐设备制作出来的音乐**。存储MIDI信息的文件格式，即标准MIDI格式（Standard MIDI File Format, SMF），后缀名一般为“mid”。
2. MIDI格式是一种**通信标准**，来作为音乐设备之间、音乐设备与PC之间传输MIDI信号的约定规范。
3. MIDI音乐文件可以看做是一张乐谱，记录了何时演奏某个音符，何时停止某个音符，以及各个音符的音色、力度、声像、表达、延音等一系列控制信息。MIDI文件回放的质量依赖于所采用的音色库的质量。MIDI的制作也需要用到音序器。

## 五、DLS标准

DLS标准：允许单个乐器有多个样本，样本按照八度划分，使得MIDI更进一步，允许MIDI音乐自由增加新的乐器声音。

## 六、CD音频标准（红皮书音频）

1. CD音频标准（红皮书音频）：缺点：存储空间大、不能构建动态音乐。
2. **动态音乐**：根据进程实时合成复杂音乐；改变音乐或巧妙的将正在播放的音乐过渡到其他完全不同的声音，增加沉浸感。

## 音频引擎

### 一、FMOD资源加载方式

1. FMOD\_CREATEAMPLE（完全加载解码后的数据）
2. FMOD\_CREATECOMPRESSEDAMPLE（完全加载压缩后的数据）
3. FMOD\_CREATESTREAM（流式）

### 二、数据驱动理念

1. 数据驱动来源于FOXPRO中的表驱动思想；把程序中的控制信息从程序中分离出来，形成**抽象控制程序与控制信息表**两部分。
2. 抽象控制程序是可以重用的，在具体应用中的不必改变；
3. 控制信息表是可变的，其内容不同程序的运行结果就不同。
4. 数据驱动理念特色：支持事件逻辑、**声音设计与编程分开**

### 三、FMOD主要特色

无需编程、跨平台、音频格式兼容多平台、支持网络监听

### 四、OpenAL

1. 是自由世界的跨平台音效API，设计用来编写三维音效
2. 从层次来说，OpenAL比FOMD更加底层，FMOD可以调用OpenAL进行三维音效实现
3. 运行OpenAL的程序需要安装OpenAL的驱动（主要是openal.dll）

## 第二章 声学基础知识

---

### 预备知识

## 一、机械波

1. 什么是机械波：机械振动在介质中传播形成机械波
2. 产生条件：有做机械运动的物体作为波源、有能传播机械振动的介质
3. 分类：横波（质点的振动方向与波的传播方向垂直，凸起部分为波峰，凹下部分叫波谷）、纵波（质点振动方向与波的传播方向在一条直线上，质点分布密的叫密部，疏的叫疏部）

**注意：液体和气体不能传播横波！**

4. 机械波的传播过程：传播的是振动形式和能量、质点只在各自的平衡位置做振动，不随波迁移、介质中质点的振动周期和频率都与波源的振动周期和频率相同、由波源向远处的各质点都依次重复波源的振动

5. 描述机械波的物理量：

①波长 $\lambda$ ：两个相邻的，在震动中相对平衡位置的位移总是相等的质点间的距离

横波：两个相邻的波峰或者波谷之间的距离

纵波：两个相邻的疏部（密部）中央间的距离

②周期和频率：

波的频率由振源决定，在任何介质中传播波的频率不变

波从一种介质进入另外一种介质，唯一不变的频率，波速波长都发生变化

③波速：单位时间内波向外传播的距离

$$v = s/t = \lambda/T = \lambda f$$

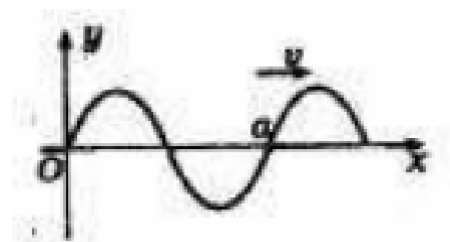
波速大小由介质决定，介质中质点的弹力越大，对运动反应越灵敏，对波的传播速度越大。

通常，固体对机械波的传播速度较大，气体对机械波的传播速度较小

6. 波的图像：

坐标轴：取质点平衡位置的连线为x轴，表示质点分布的顺序，取过波源的振动方向为y轴，表示质点位移

意义：在波的传播方向上，介质中质点在某一时刻相对各自平衡位置的位移



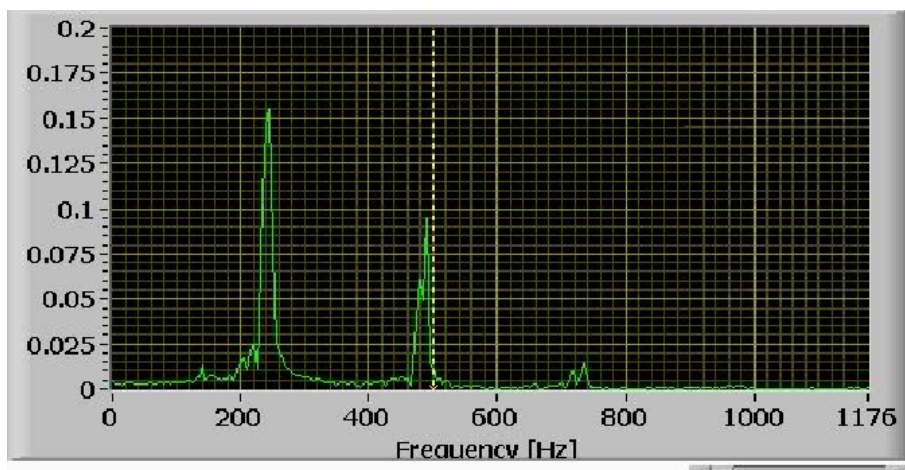
## 二、声音的分类

1. 乐音：好听悦耳的声音  
乐音三要素：音高、响度、音色
2. 噪声：嘈杂刺耳的声音

## 三、纯音、基音、泛音、复合音、音色

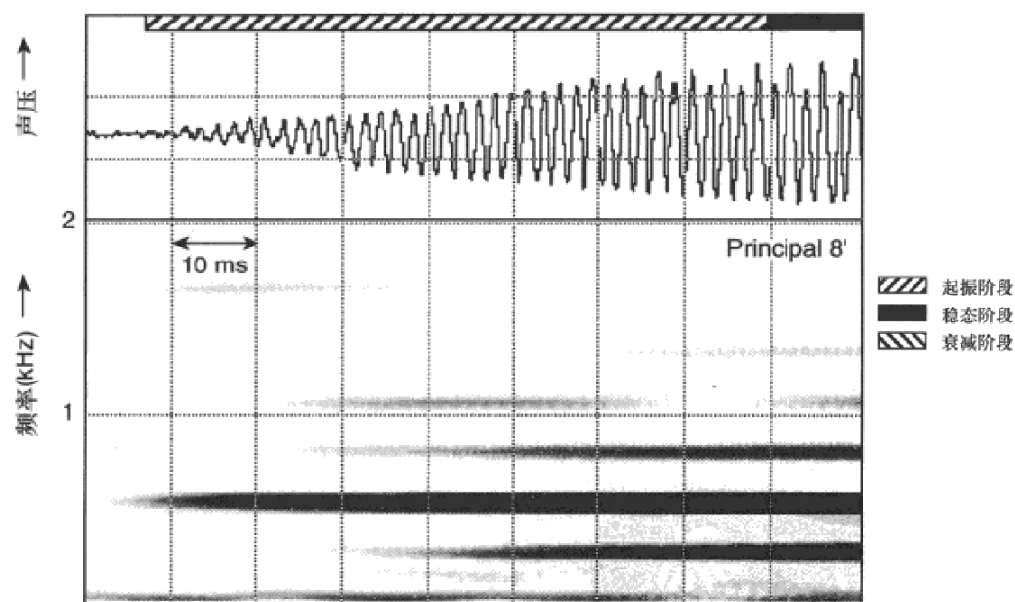
1. 纯音：只有一种频率的声音，声音可以分解为若干个不同频率纯音的叠加（傅里叶变换）
2. 基音：一般的声音都是由发音体发出的一系列频率、振幅各不相同的振动复合而成的。这些振动中有一个**频率最低的振动**，由它发出的音就是基音
3. 泛音：假设某个音的基频为 $f$ ，则频率为 $2f$ 的音为第一泛音，频率为 $3f$ 的音为第二泛音

4. 复合音：基音和泛音结合在一起而形成的音，日常我们听到的都是复合音



在上面的图中，基音是最大的波浪，泛音是最大波浪中的小波浪

#### 四、乐音的起振、稳态、衰减三阶段



可以看到，稳态阶段并非完全不变，幅度/包络仍然随着时间变化，衰减阶段对音色感知所起的作用最弱。

乐音的起振阶段对音色其非常重要的作用，第一次反射声对直达声的染色效应往往发生在起振阶段结束后，起振阶段不易受到环境声学特性影响

当乐音的起振和衰减阶段去除后，听者很难辨别音色

## 声音的产生和传播

### 一、声波、声场

1. 声波：向前推进着的空气振动称为声波
2. 声场：有声波传播的空间叫声场
3. 声音传播的实质：声音传播是指物体振动形式的传播

# 声波的描述

## 一、描述声波的基本物理量

1. 位移：物体离开静止位置的距离称为位移，最大的位移叫振幅，振幅的大小决定了声音的大小
2. 相位表：用来指示混音中左右声道之间的相位相关性。测度范围：-1~+1。+1表示左右声道完全同相，即输出的是完全相同的信号（单声道）。0表示左右声道之间存在一些完全不同的内容（立体声）。-1表示左右声道完全反向。正的读数意味着混音在相位上比较正常，负读数表示混音在相位上存在问题。
3. 周期：质点振动每往复一次所需要的时间，单位为秒。
4. 声波频率：一秒钟内媒质质点振动的次数，单位为赫兹Hz。（<20为次声、<500为低频声、500~2000为中频声、>2000为高频声、>20000超声）
5. 波长：声波两个相邻密部或两个相邻疏部之间的距离叫做波长，或者说声源每震动一次，声波的传播距离。 $\lambda=c/f=cT$   $f=1/T$
6. 声速：振动在媒质中的传播速度，声速还会随着环境的温度湿度有一些变化，温度高声速快、湿度高声速快

## 二、声波的物理量度

1. 声压：介质中的压强相对于无声波时的压强的改变值，单位为牛顿/米<sup>2</sup>或帕斯卡。瞬时声压：某一瞬间的声压。有效声压：在一定时间间隔中将瞬时声压对时间求均方根值得到有效声压。听阈声压：人耳能听到的最低声压，1000赫纯音的声压，其值为 $2 \times 10^{-5}$ Pa，就是听阈声压。痛阈声压：正常人的痛阈声压为20Pa。
2. 声压级：该声音的声压与参考声压的比值取以10为底的对数再乘上20：

$$L_p = 20 \lg \frac{p}{p_0} \qquad p_0 = 2 \times 10^{-5} p_a$$

声压级的单位：分贝

3. 峰值电平表：显示当前时刻的信号电平大小。**最高刻度是0db**，最低的电平由数字系统的量化深度决定（16-bit量化的音频信号的最低电平为-96db）。一般最低刻度的显示为大约-70db
4. 平均值电平表：可以被认为是响度电平表。因为**人耳感受到的声音响度与声音的平均值相关而非峰值相关**
5. 声强：在垂直于声波传播的指定面上，单位时间内通过单位面积的声能，单位为W/m<sup>2</sup>。声强越大表示声音越大。正常人耳对1000赫纯音的可听声强是 $10^{-12}$ 瓦/米<sup>2</sup>，这就是**基准声强（ $10^{-12}$ 瓦/米<sup>2</sup>）**。声强可衡量声音的强弱，越大听到的声音越响。与离开声源的距离有关，离声源越远则声强越小。
6. 声强与声压的关系： $I = P^2 / \rho c$ ， $\rho$ —空气密度，kg/m<sup>3</sup>； $c$ —声速，m/s； $\rho c$ —空气的特性阻抗，瑞利，Pa·s/m（声压P易测，I不易测量）。**声压只有大小没有方向，声强是矢量，方向就是声音传播的方向**
7. 声强级：该声音的声强与参考声强的比值取以10为底的对数再乘10：

$$L_I = 10 \lg \frac{I}{I_0}$$

$$I_0 = 10^{-12} \text{ W/m}^2$$

声强级的单位：分贝

**在单一声源的情况下，且与声源和任一边界相距足够远时，声强级和声压级可以等效。当存在反射声时，不等，但变化保持不变。当问题与声音大小相关时，用声压级来度量**

8. 声功率：声源在单位时间内辐射的总能量，单位是瓦。意义：声功率是衡量声源声能量输出大小的基本物理量；**声功率可用于鉴定各种声源，它是一个恒量，与声源的远近没有关系。**波阵面：波源发出的振动在介质中传播经过相同所到达的各点组成的面。同一波阵面上各点的振动位相相同。平方反比定律：当声源向各个方向辐射时，声强随着距离平方减小的规律
9. 声功率级：该声音的声功率与参考声功率的比值取以10为底的对数再乘10：

$$L_W = 10 \lg \frac{W}{W_0} \quad W_0 = 10^{-12} \text{ W} \quad \text{声功率级的单位：分贝}$$

10. 频段：对于人耳可听到的连续声谱，要求出每一频率成分是不可能也是不必要的，通常把可听声划分为若干个部分，每一部分用它的中心频率为代表，讲这样划分出来的频率段叫做频段（频带）

$$11. \quad n = \log_2 \frac{f_{\text{上}}}{f_{\text{下}}} \quad \begin{array}{l} f_{\text{上}} \text{——上限频率} \\ f_{\text{下}} \text{——下限频率} \\ n \text{——倍频程数} \end{array} \quad \text{当} n=1 \text{ 时称为倍频程，当} n=1/2 \text{ 时称为} 1/2 \text{ 倍频程以此类推。各}$$

倍频程的中心频率是上下限频率表的几何平均值：  $f_m = \sqrt{f_{\text{上}} \cdot f_{\text{下}}}$

12. 响度级：通过采用响度级概念使人耳对频率的响应与客观量声压级联系起来，并定量的描述这种关系。

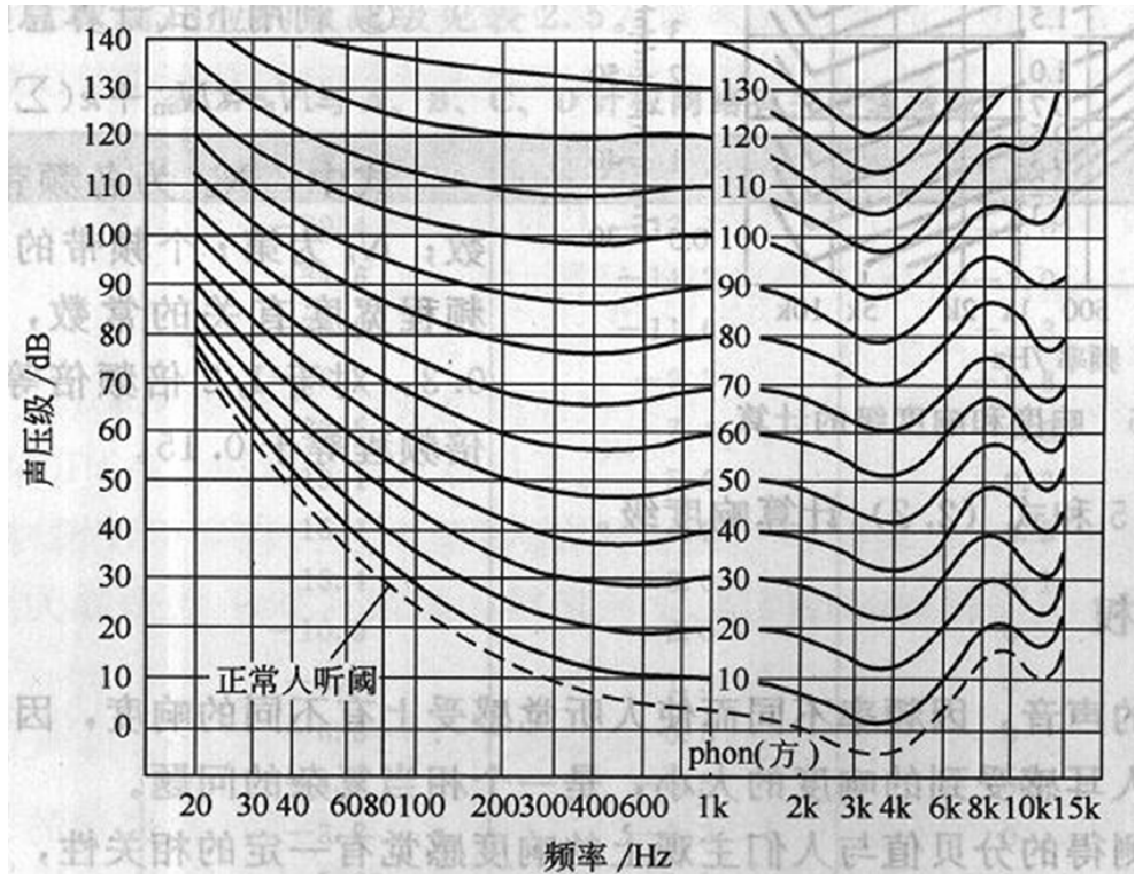
考虑到人耳的频率响应特性，依照声压级的概念，引入一个既和声压级有关，又与频率有关的物理量（主观量）——响度级来衡量声音响度，它的单位是 phon（方）。

“phon”就是以1kHz纯音为基准声音，任何声音如果听起来和某个1kHz纯音一样响，那么这个1kHz纯音声压级的分贝值就是该声音的响度级phon值。等响曲线：人耳听觉范围内一系列响度相等（听觉上大小相同）的声压级与频率关系曲线。

13. 等响曲线：在听觉实验中，如果把某个频率的纯音与一定声压级的1000Hz的纯音很快地交替比较，当听者感觉两者一样响时，把该频率的声压级标出时，便可连出一条曲线。这条曲线即为等响曲线。



等响曲线是鲁滨逊和达逊提出的，被ISO所推荐采用，所以又称ISO等响曲线。



等响曲线提供的信息：

①可看出各个频率的声音在不同的声压级时，人们主观感觉的声音强度大小。例如：60dB的1kHz的声音，响度级为60phon；等响度的4kHz的声音，其声压级是52dB；100Hz等响度的声压级为67dB；30Hz等响度的声压级是90dB。

②人耳对不同频率分布的声音敏感性不同。对低频声较为迟钝，对于频率越低的声音，人耳能听到相同的声音大小，它的声压级就越高，人耳对2 - 5KHz声音敏感——噪声控制中，首先应降低中高频噪声。

注：在3.5kHz处有一个突起（对该频段信号比较敏感），主要源于人外耳道的共振频率，婴儿啼哭的中心频率就在这个频段上。

③高声压级时，曲线较为平坦。这说明声音强度达到一定程度后，声压级相同的各频率声音几乎一样响，与频率关系不大。也即高声压级时，人耳分辨高低频声音的能力变差。

④图中最下一条曲线是人耳刚能听到声音的响度，最上一条曲线为痛觉曲线120方（即1000 Hz声音，声压级120dB），高于此响度级人们听不到声音，但耳朵发痛。

14. 响度：定量反映声音响亮程度（一声音比另一声音响了多少）的主观量。是人耳判别声音由轻到响的强度等级概念。单位为sone“宋”。1宋是指声音频率为1kHz,声压级为40dB,且来自听者正前方的平面波形的强度。响度与频率、声压级、波形都有关。

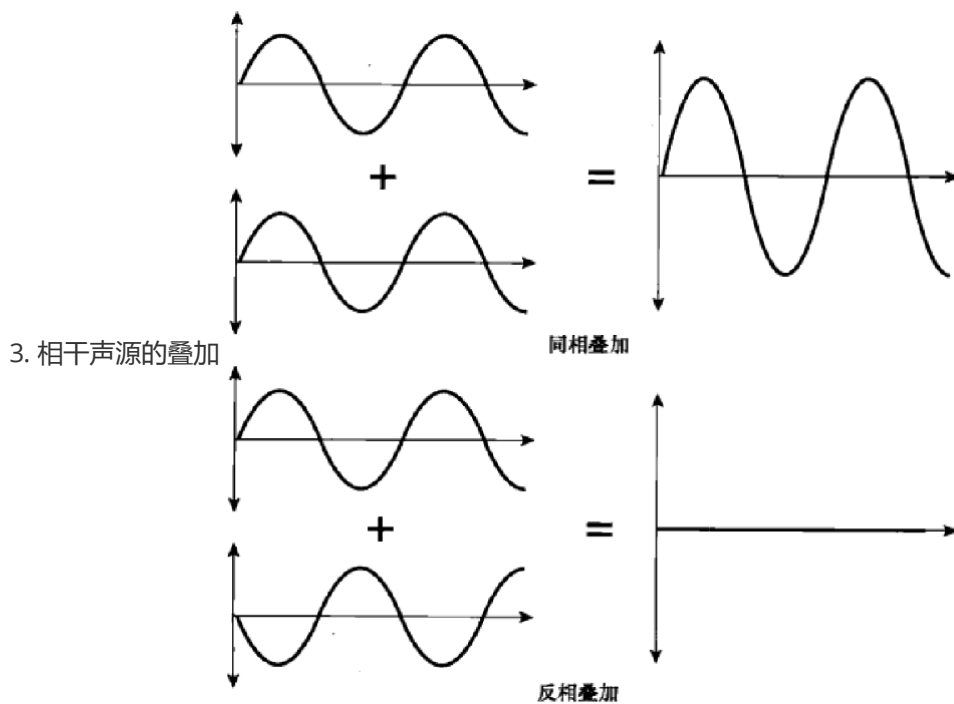
15. 音调：音调是听觉分辨声音高低时用于描述感觉的一种特性。音调与频率不成正比关系。主观上感觉音调的单位采用美（Mel）标度。一个高于听阈40db，频率为1Khz纯音产生的音调定为1000Mel。如果一个纯音听起来比1000Mel声音音调高出一倍，其音调为2000Mel

## 声波的传播特性

### 一、相干声源与非相干声源

1. 相干声源：声音来自相关联的几个声源（其它声源由同一声源产生）。如：通过简单反射产生其它声源，延迟声短，与原声非常近似，从而与原声有较强的相干性。声音来自同一电子信号源，由多只扬声器重放。
2. 非相关声源：声音来自不相关联的几个声源。如：来自同一声源，但多次反射后延时做够长，延时后声音的音调、大小和波形发生变化。来自同一声源，但多次反射后延时做够长，延时后声音的音调、大小和波形发生变化。来自不同声源，产生不同的波形和频率成分。





4. 非相关声源的叠加：满足总功率为每个声波的声功率之和

$$I = \frac{W}{S} \quad I = \frac{P^2}{\rho c} \quad \Rightarrow \quad \frac{W}{S} = \frac{P^2}{\rho c}$$

声功率与声压的平方成正比

$$P = \sqrt{(P_1^2 + P_2^2 + \dots + P_N^2)}$$

5. 相干声源与非相干声源叠加的区别：

1) 非相干声源的叠加只与所叠加信号的声功率有关，与声源的相对相位无关。

2) 非相干声源叠加后声压级的增量没有相干声源大。

两个相干声源：叠加后最大声压振幅可以是单个声压振幅的2倍。

两个非相干声源：叠加后最大振幅为一个声源声压振幅的 $\sqrt{2}$ 倍。

6. 声波的干涉现象：

两列波的频率相同，振动方向相同和具有恒定相位差的声波，合成声仍是同一频率的振动，在空间某一些位置的振动始终加强，在

另一些位置的振动始终减弱，这种现象称为声波的干涉现象。

能产生干涉现象的声波称为**相干波**。非相干声源不会发生干涉现象。

## 二、声波的反射与折射

入射声波、反射声波与折射声波的传播方向应该满足Snell定律：
$$\frac{\sin \theta_i}{c_1} = \frac{\sin \theta_r}{c_1} = \frac{\sin \theta_t}{c_2}$$

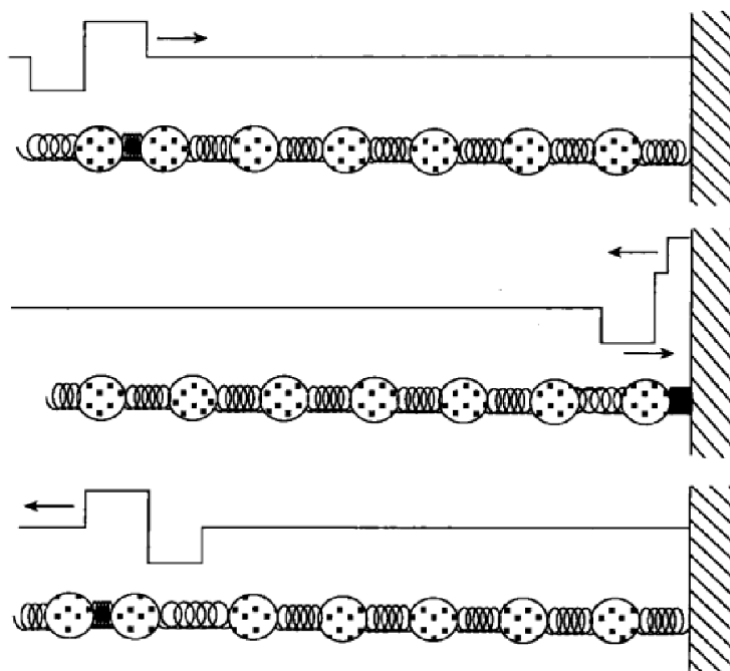
反射定律： $\theta_i = \theta_r$

折射定律：
$$\frac{\sin \theta_i}{\sin \theta_t} = \frac{c_1}{c_2} = \sqrt{\frac{T_1}{T_2}}$$

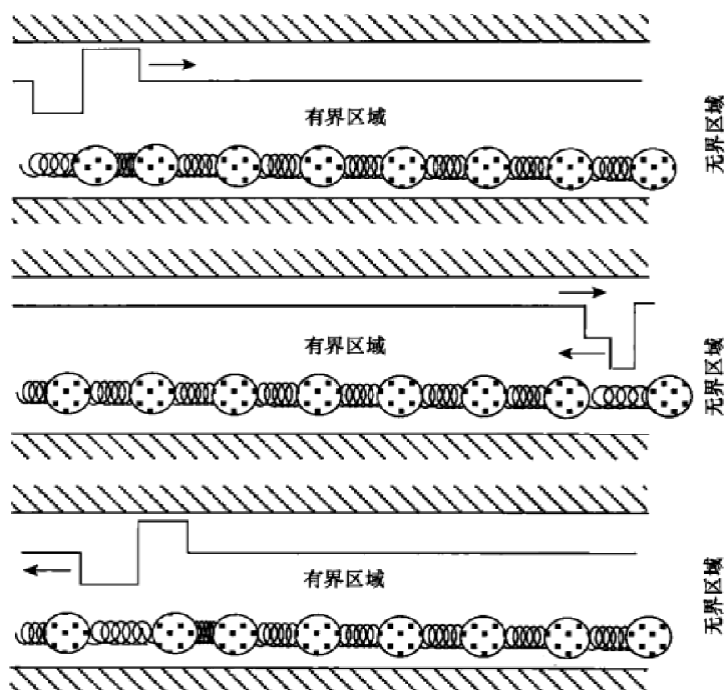
1. 由折射定律可知：声波的折射是由声速决定的。从冷空气传到热空气时折射角变大。逆风传播的声音难以听清。

2. 刚性边界的声反射：

1) 声波撞击某个固定的物体或刚性边界（反射边界上速度分量为0，压力分量为原来的2倍）



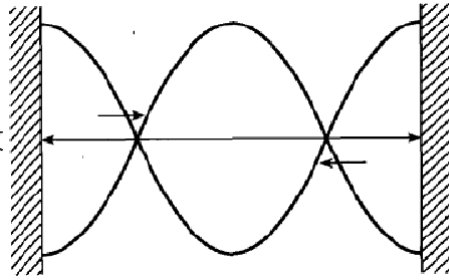
2) 从有界区域到无界区域的声反射（反射边界上，声压分量为0，速度分量为原来的2倍）：



3. 刚性边界内的驻波（共振模式）：当声波在反射面之间来回反射时，会形成驻波。当波长和反射面的间距成一定整数比例时，声波的疏密状态在两个反射面之间的位置是固定的，因而称为驻波（Standing waves）。驻波以正常速度传播，只是两列波叠加后波形不向前推进。例如，在听觉上，如在两墙中间位置，某个频率的声波可能因为振幅衰减而变得几乎听不到，而在紧贴墙面的位置，该频率的声波由于振幅提升而变得很响。能够产生驻波的频率称为共振的简正频率。通过扬声器播放20~600Hz不同频率的正弦波，然后在房间里来回走动，会发现某一频率的信号在房间内不

同位置会产生明显的提升或衰减。flowest数值越高，驻波所产生的效果越不明显，当达到500Hz

时，可忽略带来的问题。校正方法：扩散和吸收



### 三、声波的散射与衍射

1. 衍射：在声波传播过程中，偏离直线的传播方向而绕过障碍物继续向前传播的现象叫声波的衍射（绕射）。声波的衍射程度与波长有关。波长小，衍射弱
2. 散射：在声波传播过程中，遇到的障碍物表面较粗糙或者障碍物的大小与波长差不多，则当声波入射时，就产生各个方向的反射，这种现象称为散射。声波的散射既与障碍物的形状有关，又与入射声波的频率有关。波长相对障碍物尺寸大，散射往往是全方位的。波长相对障碍物尺寸小，散射方向性强；

## 时域与频域

### 一、频谱与频谱图：

在复合音中，不同频率成分的声波具有不同的能量，这种频率成分与能量分布的关系称为声音的频谱；各频率成分与能量分布关系的图形称为频谱图。

通常先测定出该声音各频率成分与相应的声压级，然后以频率为横轴，以声压级为纵轴进行绘图而得到。

### 二、周期性声波的频谱特性：

傅里叶分析方法：任一随时间作周期性变化的波，都可以分解为一系列不同频率、不同振幅和不同相位的正弦波。波形的每一个正弦波分量称为波形的分音。如果正弦波分量的频率之间存在倍数关系，这些分量称为谐波。最低频率的分量称为基波（一次谐波）。二次谐波的频率是基频的二倍，以此类推。其它正弦波是基频的奇数倍，幅度与频率成反比。更多高频叠加，将接近于理想方波。高频成分让方波具有陡峭的上升前沿和尖锐转角。人耳对声音的相对幅度比相位更加敏感。

### 三、非周期性声波的频谱特性：

周期性声波具有离散的频谱分量，在频谱中表现为具有一定间隔依次排列的线段形式。非周期性声波具有连续频谱

## 频谱分析

### 一、滤波器及其种类

1. 滤波器：滤波器是从完整的声音信号频谱中分离出部分频谱成分的声音信号处理器
2. 基本类型：低通滤波器：只通过截止频率以下的信号  
高通滤波器：只通过截止频率以上的信号  
带通滤波器：通过特定频带的信号（这个特定频带是上限频率和下限频率之间的频带，称为滤波器带宽）。  
带阻滤波器：只通过上限频率以上和下线频率以下的信号

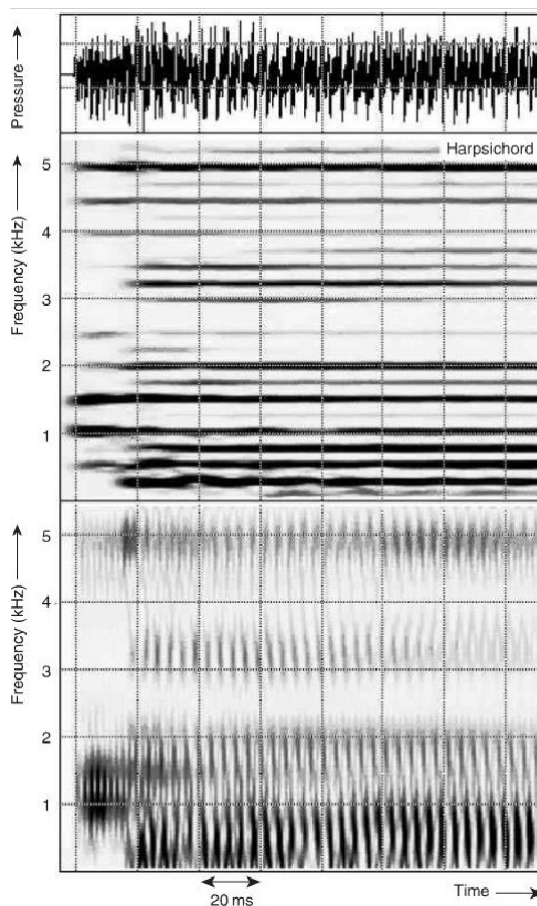
## 二、均衡器

1. 均衡器能够衰减或提升信号中某一频率范围内的成分，从而改变信号的音色。均衡器是建立在滤波（filtering）基础上的，因此，均衡器和滤波器这两个术语可相互替代。滤波器以一个特定频率作为参考；由几个滤波器所构成的设备成为均衡器。
2. 混音中使用的滤波器种类：通过式（Pass）  
搁架式（Shelving）  
参量式（Parametric），共通点：参考频率（Reference frequency）通常只有一个。
3. 参考频率称为截止频率，截止频率是均衡曲线衰减了3dB的频率点。
4. 问题：一旦信号通过了一个滤波器，那么当它再次通过相同的滤波器的时候就不会产生任何效果了吗？答案：滤波器并非完全除去信号的某一特定部分，而是对特定部分进行衰减。会产生影响效果。
5. 参量式滤波器：中心频率越高则滤波器的处理效果越不明显。  
当中心频率向高频方向移动时，参量式滤波器的带宽不变，会使得钟形频响曲线的形状变窄。  
原因：  
人耳对音调的感受是非线性的。  
在200~800Hz内的600Hz带宽等同于2个倍频程（24个半音）  
而相同的600Hz带宽如果处于10 000~10 600Hz范围内，则只相当于一个半音。**Q值越大，钟形频响曲线的形状越窄。Q值与均衡量相关联的两种方式：**  
**比例型Q值，Proportional-Q**  
**即均衡量在增大时频响曲线的形状会变窄，使声音显得有力。**  
**恒定Q值，Constant-Q**  
**即让均衡器的带宽在均衡量变化时保持基本恒定。其声音处理效果会更加柔和，听上去更和谐。**
6. 均衡器的作用：
  - ①实现频谱平衡  
修正被过分强调或者缺失的频率范围。
  - ②塑造声音的表现力  
让声音听上去变得瘦或胖，大或小，干净或肮脏，文雅或粗鲁，尖锐或圆滑。
  - ③实现声音的分离  
避免声音频率交叠所产生的遮蔽效应，声音频率交叠的遮蔽效应听起来是糟糕的。
  - ④提高声音的清晰度  
清晰度与声音的可懂度（能够被识别）或自然感紧密联系。
  - ⑤传达情感和情绪  
大脑将不同的频率与不同的感情色彩关联在一起，声音情感包括甜美、侵略性、圆润、柔和等。
  - ⑥用均衡创造趣味  
如对均衡进行自动化控制。
  - ⑦增强深度  
大脑认为黯淡的声音是从较远的地方发出的。  
如对音频使用低通滤波器，当高频信号衰减时，声音会显得更靠后。
  - ⑧提供更多实用性效果  
混响器经常包括基本的均衡器模块。
  - ⑨增强立体声效果  
如果立体声左右声道的内容完全一致，均衡处理后，所感受到的声像会变宽，声音更充实，体积更大。
  - ⑩进行精细的电平调整
  - ⑪去除不需要的内容  
比如干扰声音（隆隆声、嘶嘶声、嗡嗡声）
  - ⑫对录音效果进行补偿

## 三、声谱图

1. 声谱图（Spectrogram）：是一种将声音信号的时间特性和频率特性同时表示出来的常用方法。信号的频谱是时间的函数，坐标轴的横纵坐标分别表示时间和频率，信号的幅度大小用灰度表示，深

色表示较高幅度，浅色表示较低幅度。



2. 窄带滤波中，谐波成分清晰可见，但信号初始阶段有明显的灰度由浅变深的过程，是因为窄带滤波器有较慢的时间响应。

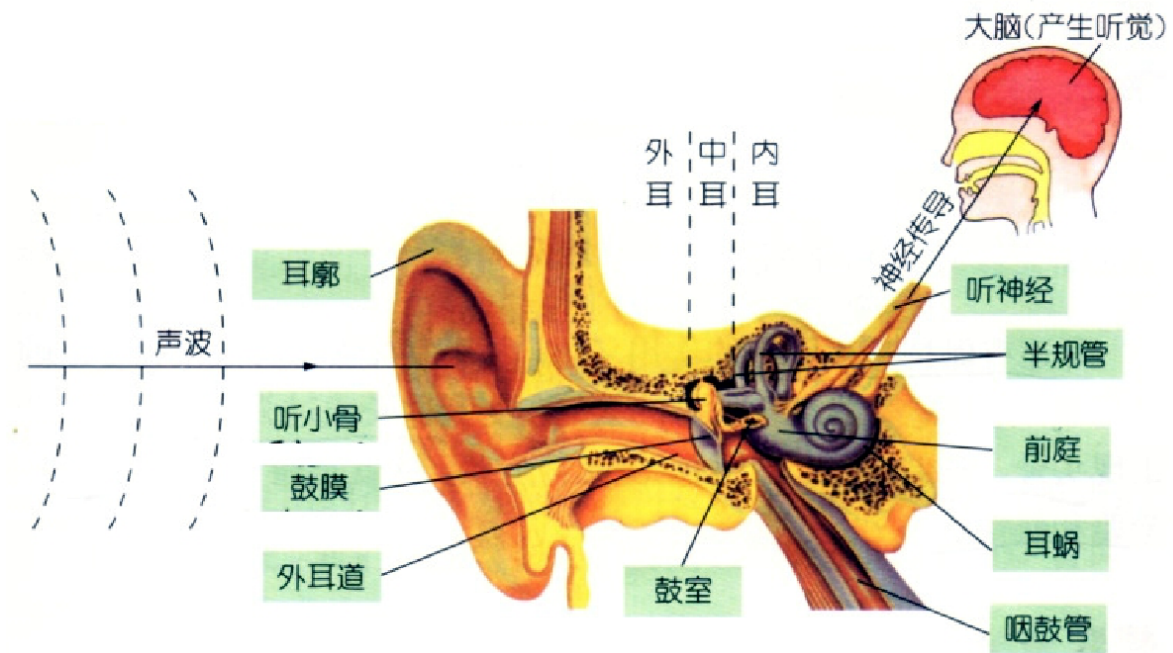
宽带滤波具有极好地时间响应，能够表现音高所固有的幅度的周期性变化规律（在声谱图中表现为纵向条纹），但不具备分解信号谐波成分的能力。

同时满足时间响应快和带宽窄两个特点的理想滤波器在物理上不可实现。

但人类听觉系统可兼顾两者。

## 第三章 听觉感知及其对声源的定位

### 听觉系统的构造



## 一、外耳

1. 耳廓：保护耳孔，卷曲状具有定向作用
2. 外耳道：长度约25mm，按声学理论，一段封闭的管道，对其管长4倍的声波具有共振放大作用
3. **外耳的作用：对声源定位、对声源放大（共振、头的衍射，总共20db的放大）**

## 二、中耳

1. 鼓室
2. 听小骨：鼓膜的机械运动通过听小骨传到耳蜗卵形窗，**在一定的范围内，听小骨实现声音的线性传递、放大声压，在特强声时，听小骨实现声音的非线性传递保护内耳**
3. 咽鼓管

## 三、内耳

半规管和前庭窗为本体感受器，与机体的平衡性有关

1. 半规管：半规管和前庭窗为本体感受器，与机体的平衡性有关
2. 前庭窗：感受静止的位置和直线的变速运动
3. 耳蜗：类似蜗牛壳，由蜗螺旋管**旋转两圈半**构成。功能：将机械振动转换为最终传输给大脑处理的神经电脉冲。
4. 基底膜：对输入声音信号进行频谱分析，**基底膜窄而薄的部位对高频响应最好，厚宽部位对低频响应最好。基底膜的作用：相当于多个频率响应交叠的带通滤波器，每个滤波器都有一定的带宽。**

## 四、部位学说

1. 临界频带：主观听感突然发生明显变化的频带宽度。开始察觉两个纯音信号从不平稳可分离转变为平稳可分离的频率差对应的带宽。频率差：  
<12.5Hz, 拍音（可以听到声音振幅有波动性变化）。

15Hz，听感从拍音变成一个音高不明确的不稳定的声音。  
继续增大，两个频率从混合在一起变为可分离、不平稳的听感。  
进一步增大，听感从不平稳变为平稳。

2. 音高感知的部位学说：

①可能性方法一：缺失基频现象实验。合成声音信号，去除基频成分，听音者感知音高不变。实验证明：在感知音高时，基频不是必须存在的；最低的频率成分不是感知音高的基本依据；

②可能性方法二：找出相邻谐波之间的最小频率差

③可能性方法三：找出所有频率成分的最大公约数。

事实证明，不管声音的基频是多少，它的前5到7次谐波一般都能够通过部位分析机理分辨出来。

3. 部位学说存在的问题：部位学说不能解释：

人类感知音高的极高精度；

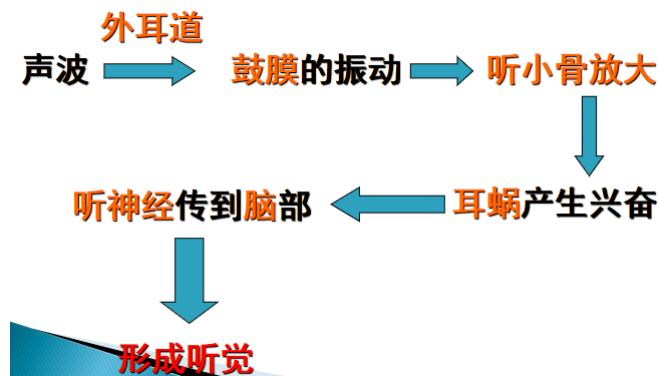
频率成分不能被听觉部位机理分辨的声音的音高感知；

具有连续（非谐波）频谱的声音的音高感知；

基频低于50Hz声音的音高感知。

在此频率范围，基底膜的振动模式不随频率的改变而改变。

## 五、听觉的形成过程



听觉方位：前方最准确，后方及两侧最不准确

## 双耳时间差

### 一、双耳时间差表达式

$$ITD = \frac{r(\theta + \sin \theta)}{c}$$

声源入射角为90°时，双耳时间差达到最大

### 二、双耳时间差的作用

1. 双耳时间差非常短，但是该值的变化决定低频声波的方向。
2. 当 $f$ 较小时，声波波长较长，时间差所造成的相位差有确定意义。双耳根据它来判定声源的位置。当 $f$ 较大时，时间差所造成的相位差甚至超过180°，使人不能判定是超前还是滞后，失去了声源定位的意义。

### 三、相位与相位表

1. 相位：①用来描述两个或者多个声波之间时间关系的物理量。②相似信号：两个完全一样的波形、来源于同一声源的波形。③相似信号的相位关系：同相 (in phase) / 相位一致 (phase-coherent)  
两个相似波形同一时刻发出  
异相 (out of phase) / 相移 (phase-shifted)  
两个相似波形不同一时刻发出，二者具有相位差  
反相 (phase inverted)



两个相似波形从同一时刻发出，振动方向正好相反，即波形按横轴对称。

反相 $\neq 180^\circ$ 相移。

2. 相位表：用来指示混音中左右声道之间的相位相关性。

3. 相位问题：具有相移或者反相的相似信号进行混音时，产生相位问题。信号时间差在35ms以内，梳状滤波器效应（combfiltering）

信号音调变化和音色染色。

两波形反相

波形抵消，造成电平衰减，如两波形振幅恰好完全相同，产生无声效果。

## 双耳声级差

由头部的遮蔽效应产生的

1. 有实验表明，两耳的强度比随着声源的入射角以正弦方式在0dB~20dB变化  
变化多少与频率有关。

2. 存在最小频率，当低于这个最小频率时，声级差对声源的定位作用不大。  
可以认为，当物体的尺寸为声波波长1/3时开始发生散射。

最小频率为：637Hz

3. 双耳声级差和时间差总结：双耳时间差ITD主要对低频声音的定位起作用。

双耳声级差IID主要对高频声音的定位起作用。

大约在700~2.8kHz听觉完成这两种定位方法的过渡。

双耳对这两个频率之间的声音信号定位能力不如其他频率的定位。

不足：

双耳效应只能解释前方水平方向上的声音定位，三维空间定位主要依赖于耳廓效应。

## 耳廓效应（单耳效应）

辐射到耳廓的声波形成反射声。

该反射声与外耳道的直达声由于延时量很小，形成梳状滤波效应。

由于声音到达方向不同，反射声与直达声时间（相位）和音量等方面发生微小差异，形成一种与音源方向相关的频谱特性，听觉系统据此来判断声音的空间方位。

这种定位方法适用于高于5kHz的频率范围。

定位效果因人而异

随着年龄的增大，存在学习过程，定位能力增强。

## 头部相关听觉传递函数

是一种频谱线索

耳朵所接收的声波与音源不完全相同——经过头与耳的多次折射，与音源位置有关

如果带上耳机，消除了这些线索，会产生声音来自头部之内的感受。

## 头部转动对定位的作用

头部转动解决声像定位不明确的一种有效方法。

头部转动，让声源位于正前方，时间差和强度差清零。

# 哈斯效应

## 一、定义

两个同声源的声波若到达听音者的时间差（时间差在5-35ms以内，人无法区分、时间差在35~50ms，人耳开始感知但听感仍是前导声源、大于50ms，人耳能分辨出前导声源和滞后声源的方位也就是能听到回声）

## 二、哈斯技巧

事件差在35ms以内，滞后声比前导声响度低10db以上。所以我们可以通过对单声道音轨在立体声重放提升声音的空间感使得声像听上去体积更大。注意这里需要将某一个音轨的声音延后5ms-35ms。

哈斯技巧只适用于对一到两个单声道音频进行处理，过多使用将会造成声像定位不准确，使得声像的重放清晰度变差

## 三、界外立体声技巧

如何让声音听上去好像是从人头的两侧发出？

方法：保持原始信号与复制信号时间上的重合，对复制信号进行反相

效果：听上去好像被虚幻的声音包围，而非简单来自左侧或者右侧

使用限制：只有听者头部完全位于左右扬声器中央位置时才突出，适用于时间很短的声源

## 四、多普勒效应

观察者听到的声音的音调，是由观察者接受到的频率，即单位时间接收到的完全波的个数决定的。当波源和观察者有相对运动，观察者接收到的频率会改变，在单位时间内，观察者接收到的完全波的个数增多，接受到的频率增大

## 五、遮蔽效应

一种频率的声音阻碍听觉系统感受另一种频率的声音的现象称为掩蔽效应。

MP3格式：只突出记录了人耳朵较为敏感的中频段声音，而对于较高和较低的频率的声音则简略记录。

## 六、鸡尾酒会效应

在嘈杂的声音中可把自己的听力完全集中在某个人的谈话上，而把其他人的声音都推到背景声中。

## 立体声重放

基于听觉对声源的定位机理

利用一对扬声器/耳机获得声像定位感。

### 一、时间差式立体声

拾音系统包括两只具有一定间隔的全指向性传声器，与音源保持一定距离。

由于传声器间隔小，因此可以认为拾取的声压级是一样的，只存在一定的时间差。

因此传声器拾取信号重放时只包含时间差定位信息。

### 二、声级差式立体声

拾音系统包含两只主轴形成一定夹角、放置在同一点的两只指向性传声器，分别指向声源的左右两侧。

由于放于同一点，不存在时间差。

由于具有指向性，因此，两只传声器拾取的信号强度不同，因此在重放时只包含声级差定位信息。

上述两种方法的区别在于记录原始声源的方法不同，而与声音的重放方式无关。

可将上述两种方法结合起来使用：

将指向性传声器以一定间距放置进行拾音，使拾取的立体声包含声级差和时间差信息

弱点：易产生不真实的定位信息，使听觉定位不准确

### 三、仿真头式立体声

将两个性能完全相同的全指向性传声器放在真人或仿真人头的外耳道内，然后将拾取的信号用耳机重放。

## 动态范围处理器

### 一、动态包络

决定音色的两个重要特性：频谱、动态包络

使用动态范围处理器来控制信号音量包络的微观动态，或者通过重塑它们的动态包络来达到改变音色的目的。

### 二、动态范围

动态范围（Dynamic Range）定义

一个系统所能容纳的最小和最大声音之间的差值。

像任何两个电平测量值的差值一样，直接用dB表示

压缩（Compress）信号的动态范围

默认为向下压缩（Downward compression），将大的声音变小。

向上压缩（Upward compression），将小的声音变大。

扩展（Expansion）信号的动态范围

默认为向下扩展（Downward expansion），将小的声音变得更小。

向上扩展（Upward expansion），将大的声音变得更大。

### 三、动态范围处理器的种类

压缩器（Compressor）

缩小高于门限的输入信号电平，使大的声音变小。

限制器（Limiter）

将任何超过门限的电平缩小到门限电平。

向上压缩器（Upward compressor）

将低于门限的输入电平进行提升，从而使小的声音变大。

扩展器（Expander）

对低于门限的输入电平进行衰减，从而让小的声音更小。

向上扩展器（Upward expander）

提升高于门限的输入电平，从而让大的声音更大。

噪声门（Gate）

对低于门限的信号按一个固定的数值（即增益变化范围）进行衰减。

闪避处理器（Ducker）

对所有高于门限的信号按一个固定的数值（即增益变化范围）进行衰减。

## 四、抽吸效应和喘息效应

抽吸效应是由可闻的电平快速变化产生的。比如大幅度的压缩和限制而产生的高电平快速变化。  
喘息效应是由噪声（或嘶声）电平变化产生的。通常由于使用了噪声门或扩展器，使得低电平的变化会引起该效应

## 压缩器

### 一、压缩器参数

门限 (Threshold)  
压缩比 (Ratio)  
增益 (Gain)  
Attack (建立时间)  
Release (释放时间)  
侧链 (Sidechain)

### 二、门限与压缩比的关系

门限决定了受到压缩的信号范围。压缩比决定了压缩的程度。  
二者通常会被结合在一起同时调整。

### 三、建立时间

建立时间越长，对声音原始的起振感保留越多。

较长的建立时间能够消弱信号的高频，使低频部分有所增强，音色变得更温暖。

建立时间不应影响其随后的各个动态包络部分。

快速建立时间带来低频失真

产生原因：  
低频信号周期长，足以让压缩器在每一个周期内产生反应。

失真量大，会带来难听的刮擦般的声音。  
失真量小，会对低频增加温暖感和清晰度。

### 四、释放时间与侧链控制

释放时间：很短的释放时间也会带来低频失真。较长的释放时间容易削弱信号的高频。

侧链控制：默认状态下，增益调整模块和侧链部分具有相同的输入信号。大部分压缩器还允许用户使用一个外部信号作为侧链部分的输入信号，在此情况下，输入信号是否被压缩则由侧链的外部信号电平大小决定。

## 第四章 信号与系统

---

### 信号的定义及其分类

## 一、信号的定义

信号可以描述范围极其广泛的物理现象。信号是消息的表现形式，是承载信息的物理量。信号可以分为确知信号与随机信号，也可以分为连续时间信号与离散时间信号.....

信号的物理种类可以是声音、图象、电、光等等。

## 二、信号的描述方法

常常借助于数学工具来描述和分析信号和系统。描述信号常用函数和波形两种方法。信号所包含的信息总是寄寓在某种变化形式的波形中。

## 三、信号的分类

### 1. 连续时间信号和离散时间信号

自变量连续可变的信号称为连续信号，自变量仅取一组离散值的信号称为离散信号。

### 2. 信号能量和功率

实际应用中，所考虑的信号直接与某一物理系统中具有功率和能量的一些物理量相关

### 3. 三类重要信号：

**能量信号——信号具有有限的总能量、功率信号——信号有无限的总能量，但平均功率有限、非能量信号且非功率信号——信号的总能量与平均功率都是无限的**

### 4. 周期信号与非周期信号

周期信号属于功率信号，通常用它的平均功率来表征。如果信号是非周期的，且能量有限则称为能量信号。

### 5. 奇信号和偶信号

根据信号函数是奇函数还是偶函数进行的分类。

### 6. 复信号和实信号

根据信号的取值进行的分类。

**任何一个信号都能分解成一个偶信号和一个奇信号之和**

### 7. 成谐波关系的复指数信号集

该信号集中的每个信号都是周期的，它们的频率分别为  $k\omega$ ，都是  $\omega$  的整数倍，因而称它们是成谐波关系的。信号集中信号的基波频率为  $\omega$ ，基波周期为  $T$ ，各次谐波的周期分别为  $2\pi/k\omega$ ，它们的公共周期是  $2\pi/\omega$ 。

## 连续时间与离散时间系统

### 一、系统

1. 系统是非常广泛的概念。通常将若干相互依赖，相互作用的事物所组成的具有一定功能的整体称为系统。它可以是物理系统，也可以是非物理系统。
2. 离散时间系统：输入信号与输出响应都是离散时间信号的系统。
3. 连续时间系统：输入信号与输出响应都是连续时间信号的系统。

## 二、时不变性

时不变性：如果一个系统当输入信号有一个时移时，输出响应也产生同样的时移。除此之外，输出响应无任何其它变化，则称该系统是时不变的( time-invariant system )。否则就是时变的( time-varying )。

## 三、线性（连续或离散）

如果某一个输入是由几个信号的加权和组成的话，那么输出也就是系统对这组信号中每一个的响应的加权和。(叠加性质)

# 离散时间系统LTI的卷积和

## 一、基本思想

如果能把任意输入信号分解成基本信号的线性组合，那么只要得到了LTI系统对基本信号的响应，就可以利用系统的线性特性，将系统对任意输入信号产生的响应表示成系统对基本信号的响应的线性组合。

# 混响

## 一、什么是混响

混响是对由房间界面的反射所形成的全部声音的统称。

## 二、预延时

到达人耳的直达声与第一次反射声之间的时间间隔。

提供了房间大小的信息

大房间，造成预延时更大。

提供声源到听众的距离信息

距离越近，预延时越长。

## 三、混响的比例

1. 临界距离：直达声与混响声强度相同的距离
2. 直混比：用于调整干信号和湿信号（同时包括早期反射声和混响声）的比例来控制混响声的大小。  
如果混响比干信号还要大，就会损害声音的可懂度和清晰度。

## 四、混响时间

用RT60这个值来度量混响会在多久后消失，即决定混响的时间长度。

## 五、密度与扩散

密度：用来控制早期反射声和混响声

反射声的密度越大，意味着房间越小。

因为小房间会迅速反射，并从附近的界面再次反射回来。

扩散：用于描述声音的分散状况。

## 六、卷积混响器

在房间内架设一对立体声传声器，然后记录很短的冲击信号的房间响应（即混响）。

用卷积混响器加载录制得到的冲击响应信号（IR， impulse response），与待处理信号进行卷积运算，最终将得到与原始声场所形成的混响非常相似的混响效果。

一个好的冲击响应能够产生非常令人信服的混响效果，可以媲美最好的数字混响器的处理效果。

问题：难以调整效果。