



工业和信息化部“十二五”规划教材

声信号处理基础

曾向阳 杨宏晖◎编著



西北工业大学出版社



工业和信息化部“十二五”规划教材

SHENGXINHAO CHULI JICHU

声信号处理基础

曾向阳 杨宏晖 编著

西北工业大学出版社

【内容简介】 本书以声信号为中心,以基本数字信号处理原理为基础,以声学工程应用为背景,以 Matlab 为实现工具,较全面地介绍了声学基础知识、声信号的来源与特点、声信号分析处理方法、数字信号处理基本理论、声信号处理算法的 Matlab 实现及其在语音信号分析、水下目标识别、厅堂音质虚拟、噪声与振动控制等领域的应用,同时也对近年来的热门信号处理算法做了简要的介绍。

本书既包括了通用数字信号处理教材中的核心内容,又强调了声信号的特点和处理方法的应用,注重基础知识、由浅入深、简明易懂,针对性和实用性较强,可作为高等学校声学或信号处理相关专业的本科教材,亦可供相关研究人员参考。

图书在版编目(CIP)数据

声信号处理基础/曾向阳,杨宏晖编著. —西安:西北工业大学出版社,2015.9

ISBN 978-7-5612-4620-7

I. ①声… II. ①曾…②杨… III. ①声—信号处理 IV. ①TN912.3

中国版本图书馆 CIP 数据核字(2015)第 220893 号

出版发行:西北工业大学出版社

通信地址:西安市友谊西路 127 号 邮编:710072

电 话:(029)88493844 88491757

网 址:www.nwpup.com

印 刷 者:陕西向阳印务有限公司

开 本:787 mm×1 092 mm 1/16

印 张:13.625

字 数:329 千字

版 次:2015 年 9 月第 1 版 2015 年 9 月第 1 次印刷

定 价:36.00 元

前 言

声信号处理在语音通信和识别、声学定位、电声设计、噪声控制等众多领域的应用越来越广,但目前国内尚未发现适用于高等学校本科教学的专业教材。数字信号处理方面的书籍很多,不过大多以傅里叶变换、 z 变换、滤波器设计方法等经典理论或高级信号处理算法为主要内容,较少涉及与声学应用紧密相关的问题。因此,笔者结合本单位教学需要,于2006年着手编写本书。此后8年中经历了两版讲义和8届本科教学实践,对教材内容进行了多次修改和补充,形成了本书终稿。

本书以声信号为中心,以基本数字信号处理原理为基础,以声学工程应用为背景,以Matlab为实现工具,较全面地介绍了声学基础知识、声信号的来源与特点、声信号分析处理方法、数字信号处理基本理论、声信号处理算法的Matlab实现及其在语音信号分析、水下目标识别、厅堂音质虚拟、噪声与振动控制等领域的应用,同时也对近年来的热门信号处理算法做了简要的介绍。

全书共分8章。第1章简要介绍了声的产生和传播基本理论。第2章介绍了声信号的特点、分类、有关概念以及声信号处理的发展及应用情况。第3章着重介绍数字声信号处理基本理论,包括声信号的数字化、时移、折叠、卷积等基本处理过程,以及线性时不变离散系统的有关概念,并介绍了 z 变换及其主要性质。第4、5两章重点介绍了数字信号处理的核心理论,包括离散傅里叶变换、快速傅里叶变换和FIR、IIR滤波器的设计方法以及这些理论在声信号处理中的应用实例。第6章介绍了声信号处理在语音信号处理、水声信号分析、听觉虚拟现实、噪声与振动信号分析等领域的应用。第7章针对课内实验,给出了前述几章有关算法的Matlab实现方法和程序示例。最后一章介绍了近10年来声信号处理领域出现的一些新理论和新算法,包括自适应信号处理、小波分析、Hilbert-Huang变换等,可作为选修内容。此外,本书在主要章节后附有一定量的习题。

本书既包括了通用数字信号处理教材中的核心内容,又强调了声信号的特点和处理方法的应用,注重基础知识、由浅入深、简明易懂,针对性和实用性较强,可作为高等学校声学或信号处理相关专业的本科教材,亦可供相关研究人员参考。

本书1~4、6~8章由曾向阳教授编著,第5章由杨宏晖副教授编著。本书在资料整理、制图与排版过程中,得到了西北工业大学航海学院王强、蔡怀珍、王蕾、马慧颖、陈玲、朱建斌、陈幸幸等研究生的协助,在此表示衷心的感谢。

如有错漏之处,望读者不吝指正。

编著者

2015年6月

目 录

第 1 章 声学基础理论	1
1.1 声波及其产生机理	1
1.2 常用声学参数	2
1.3 波动方程与声传播规律	4
1.4 声波的主观感知与声学效应	7
习题 1	10
第 2 章 声信号处理概论	11
2.1 信号与声信号	11
2.2 声信号处理的发展	16
2.3 声信号处理的应用	18
习题 2	19
第 3 章 数字声信号处理基础理论	20
3.1 声信号的数字化	20
3.2 数字声信号处理系统	24
3.3 典型数字信号	26
3.4 数字信号的基本运算	30
3.5 线性时不变离散时间系统	32
3.6 卷积	34
3.7 z 变换	35
习题 3	45
第 4 章 DFT/FFT 及其应用	47
4.1 周期序列	47
4.2 离散傅里叶级数	49
4.3 离散傅里叶变换	51
4.4 快速傅里叶变换	63
4.5 FFT 在数字声信号处理中的应用	75
习题 4	82

第 5 章 数字滤波器设计方法及应用	83
5.1 滤波器概述	83
5.2 IIR 数字低通滤波器设计	85
5.3 FIR 数字滤波器设计	96
5.4 数字声信号处理中的典型滤波器及应用	119
习题 5	123
第 6 章 声信号处理的典型应用	124
6.1 语音信号分析	124
6.2 水声信号分析与处理	138
6.3 声信号处理在听觉虚拟现实中的应用	144
6.4 噪声与振动信号分析	146
习题 6	151
第 7 章 声信号处理的 Matlab 及 DSP 实现	152
7.1 声信号的计算机表示	152
7.2 典型声信号处理算法的 Matlab 实现	155
7.3 Matlab 在声信号处理中的应用实例	162
7.4 声信号处理的 DSP 实现	176
习题 7	180
第 8 章 声信号的高级处理方法	182
8.1 离散随机信号分析	182
8.2 小波分析	191
8.3 自适应信号处理	199
8.4 高阶统计量分析	204
8.5 Hilbert - Huang 变换及其应用	208
参考文献	211

第 1 章 声学基础理论

声波是声信号的载体,为了更深入地认识声信号并对其加以利用,本章首先简要介绍声波的产生、传播、效应等方面的内容。

1.1 声波及其产生机理

声波源于物体振动,人的语音来源于喉内声带的振动,扬声器发声来源于纸盆的振动,机械噪声来源于机器部件的振动。凡是发出声音的物体都可称为声源。声源不一定是固体,液体和气体同样可能由于振动而发声,如波涛声和汽笛声就是由流体振动产生的。

声音是听觉系统对声波的主观反应,而声波的产生来源于声源诱发的振动在媒质中的传播。因此,产生声波的必要条件是声源和媒质(空气、水等)。真空中没有媒质存在,因而在真空中不能传播声音。声波在媒质中向四面八方的传播,只是媒质振动的传播过程,媒质本身并没有向前运动,它只是在其平衡位置附近来回地振动,而所传播出去的是物质的运动形态,这种运动形式叫波动。声音是机械振动的传播,因此这种传播过程是一种机械性质的波动,故而称为声波。

以科学的方法研究物体振动及其辐射声波始于 17 世纪伟大的科学家伽利略。从 1638 年起,他深入研究了单弦振动和发声的关系;提出了频率的概念(当时称之为振动数),研究了同样调谐的弦间共振现象,与 11 世纪我国北宋科学家沈括定义的应声极为相似;他还得到了音调与弦长有关的结论。之后,不少人对弦的振动做了大量的工作。1660 年,物理学中著名的弹性定律的发现者胡克试图获得音调与频率之间的关系,他发明了现在还常用于教学演示的用齿轮在纸边转过以确定频率高低的方法。对音调和频率的关系进行彻底研究的是法国科学家索沃,他还发现了拍音,并提出了“Acoustics”这个术语(中文“声学”是沈括提出的)。无穷级数发明者——英国著名数学家泰勒于 1713 年第一次求得弦振动的初步解析解,只有基频,这也是牛顿运动方程 $F=ma$ 第一次用于连续介质中质点的运动。由于当时缺乏偏微分方程的求解工具,没有得到弦的全解。后来,伯努力、达朗贝尔和欧拉完成了这方面的工作。

在研究弦的振动时,18 世纪的数学家不能接受多个频率的振动可以叠加的理论。1822 年发表的傅里叶级数解决了这个问题,其意义极为深远。1787 年,德国科学家克拉尼发表了用沙子显示振动分布的克拉尼图形,成为研究固体振动的重要实验手段。1759 年,拉格朗日(意大利科学家)研究了管风琴和其他管乐器的发声。1734—1735 年,欧拉和伯努力研究了棒的振动,后来瑞利对他们的方法进行改进,得到四阶偏微分方程,发表于其著作《声的理论》中。板的振动比较难解,早在 1784 年,克拉尼发表他的图形时,就把沙子撒到振动着的板上显示其波节,观者都赞叹其美丽多变,但一直到 1850 年,板的振动求解问题才由基尔霍夫完成。

人的发声系统很晚才受到物理学家的注意。英国惠斯通(阻抗电桥的发明者)提出发原音的谐音理论。他认为声带振动发出基音和大量的谐音,气流带其经过口腔时,被振动加强。另

一理论是 1829 年英国维李斯提出的,他认为是气流通过声带成为一股一股的喷气,喷气进入口腔激发其共振而形成声。亥姆霍兹在 1862 年发表的伟大声学著作《音的感知》中用他的共鸣器对此做了透彻的研究,并且判断两种发声理论各有其合理处,这已被后来的研究所证明。

1.2 常用声学参数

1. 声压

定量描述声波的基本物理量是声压,它是媒质受扰动后产生的逾量压强,是空间位置 and 时间的函数,单位是压强的单位:Pa(帕), $1\text{ Pa} = 1\text{ N/m}^2$ 。

声场中某一瞬时的声压值称为瞬时声压,在一定时间间隔内最大的瞬时声压为峰值声压。一定时间间隔内,瞬时声压对时间取均方根值称为有效声压,即

$$p_e = \sqrt{\frac{1}{T} \int_0^T p^2 dt} \quad (1-1)$$

式中, T 代表平均的时间间隔。

介质质点速度是求声能量所必需的参量,它是一个矢量。已知声压,可以由运动方程求出质点速度,即

$$\mathbf{v} = -\frac{1}{\rho_0} \int \text{grad } p dt \quad (1-2)$$

在 x 方向,有

$$v_x = -\frac{1}{\rho_0} \int \frac{\partial p}{\partial x} dt \quad (1-3)$$

描述声压的基本参量是幅度、相位、频率、波长等。例如,一列纯音声波,在数学上可以表示为 $p = p_a \sin(\omega t + \varphi)$,则 p_a 是该声波的幅度, $\omega = 2\pi f = 2\pi/T$, ω 为角频率, f 为频率, T 为周期, φ 为相位。一个周期内声波的长度称为波长,它与频率成反比,满足 $\lambda = c_0/f$,其中 c_0 是声波传播的速度。

声速是各种声学问题研究中一个十分重要的参数。1635 年,法国伽桑地用枪声做了声速的测量,假设发枪的火花传播不需要时间,得到的结果是 478.4 m/s 。法国梅森认为他的结果太高,对枪声测试做了认真分析,重复了试验,得到了 450 m/s 的值。还有很多人重复了该试验。大约 100 年后,法国科学院组织了 1738 年大气中(无风时)的声速测量,用加农炮声得到的结果折合到摄氏零度是 322 m/s 。现代常用值是 $(331.45 \pm 0.05)\text{ m/s}$ 。当然,声速并不是一个固定值,它受到媒质特性的影响,如温度、风速等。不同介质中声速也不同,如水中声速远大于空气中声速。

2. 声能量与声能密度

在一个足够小的体积元内,体积、声压和密度分别记为 V_0 , p , ρ_0 ,则声扰动的能量为动能和势能之和,有

$$\Delta E = \Delta E_k + \Delta E_p = \frac{V_0}{2} \rho_0 \left(v^2 + \frac{1}{\rho_0^2 c_0^2} p^2 \right) \quad (1-4)$$

单位体积里的声能量称为声能量密度,其表达式为

$$\epsilon = \frac{\Delta E}{V_0} = \frac{1}{2} \rho_0 \left(v^2 + \frac{1}{\rho_0^2 c_0^2} p^2 \right) \quad (1-5)$$

以上方程对所有形式的声波都成立,具有普遍意义。对于平面波,有

$$\Delta E = V_0 \frac{p_a^2}{\rho_0 c_0^2} \cos^2(\omega t - kx) \quad (1-6)$$

单位体积里的平均声能量为平均声能密度,有

$$\bar{\epsilon} = \frac{\overline{\Delta E}}{V_0} = \frac{p_a^2}{2\rho_0 c_0^2} = \frac{p_e^2}{\rho_0 c_0^2} \quad (1-7)$$

3. 声功率与声强

单位时间内通过垂直于声传播方向面积 S 的平均声能量称为平均声功率或平均声能量流,它与平均声能密度的关系是

$$\bar{W} = \bar{\epsilon} c_0 S \quad (1-8)$$

单位面积上的平均声功率称为声强,有

$$I = \frac{\bar{W}}{S} = \bar{\epsilon} c_0 \quad (1-9)$$

求声强大小的另一种方法是

$$I = \frac{1}{T} \int_0^T \text{Re}(p) \text{Re}(v) dt \quad (1-10)$$

声强的单位为 W/m^2 。需要特别注意的是,声强是一个矢量,它表示声场中声能流的运动方向。

4. 声压级、声强级和声功率级

直接应用声压、声功率等描述实际生活中的各种声音,数值变化范围非常宽,而用对数标度以突出其数量级的变化则相对明了。另一方面,人耳对声音的接收,并不是正比于强度的绝对值,而是更近于正比其对数值。因此,在声学中普遍使用对数标度来量度声压、声强、声功率等声学参量,分别称为声压级、声强级和声功率级,单位用分贝(dB)表示。

声压级的符号为 L_p ,其定义为,将待测声压的有效值 p_e 与参考声压 p_0 的比值取以10为底的常用对数,再乘以20,即

$$L_p = 20 \lg \frac{p_e}{p_0} \quad (\text{dB}) \quad (1-11)$$

在空气中,参考声压 $p_0 = 2 \times 10^{-5} \text{ Pa}$,这个数值是正常人耳对1 kHz声音刚刚能够觉察到的最低声压值。也就是说,低于这一声压值,一般人耳就不能觉察到此声音的存在了,亦即听阈声压级为0 dB。式(1-11)也可以写为

$$L_p = 20 \lg p_e + 94 \quad (\text{dB}) \quad (1-12)$$

人耳的感觉特性,从可听阈的 $2 \times 10^{-5} \text{ Pa}$ 的声压到痛阈的 20 Pa ,两者相差100万倍,而用声压级来表示则变化为 $0 \sim 120 \text{ dB}$ 的范围,使声音的量度大为简明。由此可以看出,声压值增大10倍,相当于声压级增加20 dB;声压值增大100倍,相当于声压级增加40 dB。一般微风轻轻吹拂树叶的声音约为14 dB,在房间中高声谈话(相距1 m处)为68~74 dB,飞机强力发动机的声音(5 m远)约为140 dB。一个声音比另一声音的声压大一倍时,声压级增加约6 dB,一般人耳对于声音强弱的分辨能力约为0.5 dB。

声强级 L_I 的定义为,待测声强 I 与参考声强 I_0 的比值取常用对数再乘以10,即

$$L_I = 10 \lg \frac{I}{I_0} \quad (\text{dB}) \quad (1-13)$$

在空气中,参考声强 I_0 取为 10^{-12} W/m^2 ,它是与空气中参考声压 $p_{\text{ref}} = 2 \times 10^{-5} \text{ Pa}$ 相对应的声强。这样式(1-13)又可写成

$$L_I = 10 \lg I + 120 \quad (\text{dB}) \quad (1-14)$$

式中,声强 I 的单位为 W/m^2 。

$$I = \frac{p_e^2}{\rho_0 c_0} \quad (1-15)$$

式中, $\rho_0 c_0$ 是空气的特性阻抗。在计算参考声强时,空气的特性阻抗取值为 $400 \text{ Pa} \cdot \text{s/m}$,代入式(1-13),有

$$L_I = 10 \lg \frac{I}{I_0} = 10 \lg \left(\frac{p^2}{\rho_0 c_0} \times \frac{400}{p_0^2} \right) = L_p + 10 \lg \frac{400}{\rho_0 c_0} = L_p + \Delta L_p \quad (1-16)$$

大多数情况下, $\Delta L_p = 10 \lg (400/\rho_0 c_0)$ 的值很小,因此,声压级 L_p 近似等于声强级 L_I 。例如,空气中在一个大气压下, 0°C 时, $\rho_0 c_0 = 428 \text{ Pa} \cdot \text{s/m}$, $\Delta L_p = -0.29 \text{ dB}$; 20°C 时, $\rho_0 c_0 = 415 \text{ Pa} \cdot \text{s/m}$, $\Delta L = -0.16 \text{ dB}$ 。可见,都可以认为 $L_p \approx L_I$ 。

声功率 W 也可用“级”来表示,称为声功率级,其定义为

$$L_W = 10 \lg \frac{W}{W_0} \quad (\text{dB}) \quad (1-17)$$

这里 W 是指声功率的平均值 $\bar{W}$,对于空气媒质参考声功率 $W_0 = 10^{-12} \text{ W}$,这样式(1-17)可写为

$$L_W = 10 \lg \bar{W} + 120 \quad (\text{dB}) \quad (1-18)$$

由式(1-9)的声强与声功率的关系,以及空气中声强级近似地等于声压级,可得

$$L_p \approx L_I = 10 \lg \left[\frac{W}{S} \frac{1}{I_0} \right] \quad (1-19)$$

将 $W_0 = 10^{-12} \text{ W/m}^2$ 代入式(1-19),便得到

$$L_p \approx L_I = L_W - 10 \lg S \quad (1-20)$$

这就是空气中声强级、声压级与声功率级之间的关系。需要注意的是,式(1-20)应用的条件必须是自由声场,即除了声源发声外,其他声源的声音和反射声的影响均应小到可以忽略的程度。在自由场和半自由场中测量机器噪声辐射声功率的方法就是基于上述理论的。

1.3 波动方程与声传播规律

在声场中,描述声场时间、空间变化规律和相互联系的数学方程即波动方程,它是各种声学理论研究的基础。为了使研究的问题得到简化,这里仅讨论小振幅声波的情况,相应的波动方程称为线性波动方程。推导该方程的前提条件:媒质中不存在黏滞性;媒质在宏观上是均匀的、静止的;声波在媒质中的传播为绝热过程。

1. 波动方程

声波的扰动要满足的三个基本物理定律:牛顿第二定律、质量守恒定律、物态方程。由此可以得到理想流体媒质中的三个基本方程:运动方程、连续性方程和物态方程。在一维空间

(如 x 方向), 这三个方程分别为

$$\rho \frac{dv}{dt} = -\frac{\partial p}{\partial x} \quad (1-21)$$

$$-\frac{\partial(\rho v)}{\partial x} = \frac{\partial \rho}{\partial t} \quad (1-22)$$

$$dP = c^2 d\rho \quad (1-23)$$

式中, p, v, ρ 分别为声场中某一点由于声扰动引起的声压、质点振速和介质密度; P 为总的声压; c 为声波的传播速度。

小振幅声波满足如下条件: ① 声压远小于媒质中的静态压强; ② 媒质质点振速远小于声波的传播速度; ③ 质点位移远小于声波波长; ④ 媒质密度增量远小于静态密度。自然界中的绝大多数声波可归为小振幅声波, 它是线性声学研究的对象。

在一维空间中, 对于小振幅声波, 上述三个方程分别可以进一步简化为如下形式:

$$\rho_0 \frac{\partial v}{\partial t} = -\frac{\partial p}{\partial x} \quad (1-24)$$

$$-\rho_0 \frac{\partial v}{\partial x} = \frac{\partial \rho'}{\partial t} \quad (1-25)$$

$$p = c_0^2 \rho' \quad (1-26)$$

式中, ρ_0, c_0 为没有声扰动时介质的密度和声波的传播速度。消去以上 3 式中的任意两个变量, 如质点振速和介质密度变量, 剩下的两个式子分别对 x 和 t 求导, 综合整理后可以得到

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \quad (1-27)$$

这就是一维声场中的波动方程。

在三维空间中, 式(1-24) ~ 式(1-26) 可以推广表示为如下形式:

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\mathbf{grad} p \quad (1-28)$$

$$-\text{div}(\rho_0 \mathbf{v}) = \frac{\partial \rho'}{\partial t} \quad (1-29)$$

$$p = c_0^2 \rho' \quad (1-30)$$

其中

$$\mathbf{grad} p = \frac{\partial p}{\partial x} \mathbf{i} + \frac{\partial p}{\partial y} \mathbf{j} + \frac{\partial p}{\partial z} \mathbf{k}, \quad \text{div}(\rho_0 \mathbf{v}) = \frac{\partial(\rho_0 v_x)}{\partial x} + \frac{\partial(\rho_0 v_y)}{\partial y} + \frac{\partial(\rho_0 v_z)}{\partial z}$$

由此可以导出三维波动方程为

$$\nabla^2 p = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \quad (1-31)$$

其中, ∇^2 为直角坐标系中的拉普拉斯算子, 有

$$\nabla^2 = \text{div}(\mathbf{grad} p) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

2. 声波的吸收、反射与折射

在同一介质中, 声波的能量会由于介质的吸收而逐渐衰减。通常利用吸声系数 α 来描述各种材料(结构)的吸声能力, 定义为材料(结构)吸收的声能(含透射声能)与入射到材料(结构)声能的比值。用公式可以表示为

$$\alpha = \frac{E_a}{E_i} = \frac{E_i - E_r}{E_i} \quad (1-32)$$

式中, E_i, E_a, E_r 分别表示入射声能、被吸收声能和反射声能。

吸声系数不仅与材料的性质有关, 也与入射角度和频率有重要的关系。这方面的详细知识可以查阅声学书籍。

空气作为最典型的传声介质, 其吸收作用主要取决于空气的相对湿度和声波的频率。在常温下, 湿度越大, 空气吸声影响反而越小; 而声波频率越高, 吸收作用越明显。

若设空气中的声强吸收系数为 2ξ , 声波初始能量为 W_i , 在空气中传播时间为 t 后, 能量便衰减为 $W_i e^{-2\xi t}$, 空气对声能的这种吸收作用相当于一个低通滤波器, 如图 1-1 所示。

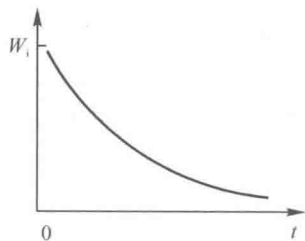


图 1-1 空气对声能的吸收作用

在声波传播过程中, 还可能遇到各种各样的障碍物, 它们会对声波的能量及其传播方式产生影响。声波在两种媒质的分界面上会发生反射、透射(对垂直入射声波)和折射(对斜入射声波)现象。要获得入射波、反射波、透射波(或折射波)之间的定量关系, 需要用到边界条件。在无限大的分界面上, 有两种声学边界条件, 它们是声压连续条件和法向质点振速连续条件, 表达式为

$$p_1 = p_2 \quad (1-33)$$

$$v_1 = v_2 \quad (1-34)$$

式中, p, v 为分界面上的声压和质点振速; 下标 1 和 2 分别表示媒质 1 和媒质 2。

对于一维斜入射平面波问题, 入射波声压和质点振速为

$$p_i = p_{ia} e^{j(\omega t - k_1 x \cos \theta_i - k_1 y \sin \theta_i)} \quad (1-35)$$

$$v_{ix} = -\frac{\cos \theta_i}{\rho_1 c_1} p_i \quad (1-36)$$

反射波声压和质点振速为

$$p_r = p_{ra} e^{j(\omega t + k_1 x \cos \theta_r - k_1 y \sin \theta_r)} \quad (1-37)$$

$$v_{rx} = -\frac{\cos \theta_r}{\rho_1 c_1} p_r \quad (1-38)$$

在媒质另一侧的入射波声压和质点振速为

$$p_t = p_{ta} e^{j(\omega t - k_2 x \cos \theta_t - k_2 y \sin \theta_t)} \quad (1-39)$$

$$v_{tx} = -\frac{\cos \theta_t}{\rho_2 c_2} p_t \quad (1-40)$$

在分界面上, 有以下边界条件

$$p_i + p_r = p_t \quad (1-41)$$

$$v_{ix} + v_{rx} = v_{tx} \quad (1-42)$$

由此可以获得声波反射与折射定律

$$\theta_i = \theta_r \quad (1-43)$$

$$\frac{\sin \theta_i}{\sin \theta_t} = \frac{k_2}{k_1} = \frac{c_1}{c_2} \quad (1-44)$$

在有些研究领域, 如建筑声学仿真研究中, 界面反射还常常进一步区分为镜面反射和

散射。

3. 声波的叠加

设有两列同频率、相差固定的平面声波,分别为

$$p_1 = p_{1a} \cos(\omega t - \varphi_1), \quad p_2 = p_{2a} \cos(\omega t - \varphi_2) \quad (1-45)$$

合成声场的声压为

$$p = p_a \cos(\omega t - \varphi) \quad (1-46)$$

其中

$$p_a^2 = p_{1a}^2 + p_{2a}^2 + 2p_{1a}p_{2a} \cos(\varphi_2 - \varphi_1) \quad (1-47)$$

$$\varphi = \arctan \frac{p_{1a} \sin \varphi_1 + p_{2a} \sin \varphi_2}{p_{1a} \cos \varphi_1 + p_{2a} \cos \varphi_2} \quad (1-48)$$

合成声场的平均声能密度为

$$\bar{\epsilon} = \bar{\epsilon}_1 + \bar{\epsilon}_2 + \frac{p_{1a}p_{2a}}{\rho_0 c_0^2} \cos \psi \quad (1-49)$$

由式(1-49)可以看出,两列声波叠加后的平均声能密度出现极大、极小相互交错的现象,称为声波干涉现象。对于不同频率的两列固定相差声波,有

$$\bar{\epsilon} = \bar{\epsilon}_1 + \bar{\epsilon}_2 \quad (1-50)$$

对于具有相同频率的两列相位随机变化的声波,有

$$p_1 = p_{1a} \cos(\omega t - \varphi_1), \quad p_2 = p_{2a} \cos(\omega t - \varphi_2) \quad (1-51)$$

合成声场的声压为

$$p = p_a \cos(\omega t - \varphi) \quad (1-52)$$

其中

$$p_a^2 = p_{1a}^2 + p_{2a}^2 + 2p_{1a}p_{2a} \cos(\varphi_2 - \varphi_1) \quad (1-53)$$

$$\varphi = \arctan \frac{p_{1a} \sin \varphi_1 + p_{2a} \sin \varphi_2}{p_{1a} \cos \varphi_1 + p_{2a} \cos \varphi_2} \quad (1-54)$$

合成声场的平均声能密度为

$$\bar{\epsilon} = \bar{\epsilon}_1 + \bar{\epsilon}_2 \quad (1-55)$$

对于多列这样的声波,有

$$p_e^2 = p_{1e}^2 + p_{2e}^2 + \cdots + p_{ne}^2 \quad (1-56)$$

在实际的场合中,多人讲话发出的声音、多台机器发出的噪声、不同车辆发出的交通噪声的叠加都可以看作无规相位声波的叠加。

1.4 声波的主观感知与声学效应

1.4.1 声波的主观感知

声波的接收包括主观感知和机器接收两方面。人的听觉灵敏度和视觉灵敏度都是很惊人的。正常听觉阈是 10^{-12} W/cm^2 , 如果鼓膜的平均面积是 0.66 cm^2 , 则人耳能感觉到的声波是

$6.6 \times 10^{-17} \text{ W}$, 该声音持续响 0.1 s 即可被人耳感知, 对应的声能量是 $6.6 \times 10^{-18} \text{ J}$ 。

人耳听觉问题不仅是物理学家的事, 也引起了很多生理学家和心理学家的注意。关于听觉频率范围已做过很多工作, 法国科学家萨瓦于 1830 年前后求得听觉低限频率是 8 Hz , 高限为频率是 $24\,000 \text{ Hz}$ 。后来, 西倍、毕奥、科尼希、亥姆霍兹等继续做了不少研究, 认为低频在 $16 \sim 32 \text{ Hz}$ 之间, 人与人之间的差异明显, 但高频差异更大, 而且高频极限随年龄的增加而降低。现在通常认为 $20 \sim 20\,000 \text{ Hz}$ 是可听声频率范围。声强的听觉低限是托普勒与著名的统计物理学家波尔兹曼在 1870 年前后合作研究得到的。他们巧妙地用光学干涉的方法测量空气密度的最大变化, 据此求出最大的声场强度是 10^{-7} W/m^2 , 这比现在接受的值大得多, 但已经足以说明听觉灵敏度的突出。

首先提出听觉理论的是电路定律的提出者——德国物理学家欧姆, 他于 1843 年指出一个乐音具有基波和频率为整数倍的谐波, 谐波结构决定乐音的音色。人耳听音时就像谐波分析器一样, 可把声音的基波和各谐波分解, 如 1822 年提出的傅里叶级数。欧姆听觉定律引发了大量生理声学和心理声学的研究工作, 其中最伟大的是亥姆霍兹, 他根据大量的听音实验推论得出了耳内共振理论。按照此理论, 耳蜗基底膜可以对不同频率的声波产生共振, 这样就可以解释欧姆定律。亥姆霍兹共鸣器就是在这些实验中发现的, 至今仍有重要的应用价值。在亥姆霍兹时代, 由于生理学知识有限, 不少结论和理论是由物理现象推论而来的。后来, 也就是近百年之后, 诺贝尔奖获得者匈牙利生理学家贝开西用生理学方法验证了亥姆霍兹的主要工作, 也指出了一些不准确的地方。

1.4.2 典型声学效应

在对人耳听觉机理的研究中, 发现了一系列的现象或效应。这些效应对于认知人耳的听觉机制, 解释日常生活中的各种主观听觉能力具有重要的价值。以下对最常见的几种效应进行介绍。

1. 掩蔽效应 (Masking Effect)

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) 所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闻阈。实验表明, $3 \sim 5 \text{ kHz}$ 时的绝对闻阈值最小, 即人耳对此频段的微弱声音最敏感; 在低频和更高频区绝对闻阈值要大得多。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闻阈称为掩蔽闻阈 (或掩蔽门限), 被掩蔽弱音必须提高的分贝值称为掩蔽量 (或阈移)。

已有实验表明, 纯音对纯音、噪音对纯音的掩蔽效应具有不同的结论。

纯音间的掩蔽: 对处于中等强度时的纯音最有效的掩蔽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。

噪音对纯音的掩蔽: 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 17 dB , 且较平坦; 超过 500 Hz 时大约每 10 倍频程增大 10 dB 。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, 若该纯音刚好能被听到时的功率等于这一频

带内的噪声功率,则这一带宽称为临界频带宽度。临界频带的单位是巴克(Bark),1 Bark 即 1 个临界频带宽度。临界频带与线性频率之间存在非线性映射关系。通常认为,20 Hz~16 kHz 范围内有 24 个子临界频带。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。

掩蔽效应可以分为频域掩蔽和时域掩蔽。

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 1 000 Hz 的音比另一个 900 Hz 的音高 18 dB,则 900 Hz 的音将被 1 000 Hz 的音掩蔽。而若 1 000 Hz 的音比离它较远的另一个 1 800 Hz 的音高 18 dB,则这两个音将同时被人耳听到。若要让 1 800 Hz 的音听不到,则 1 000 Hz 的音要比 1 800 Hz 的音高 45 dB。一般来说,低频的音容易掩蔽高频的音;在距离强音较远处,绝对闻阈比该强音所引起的掩蔽阈值高,这时,噪声的掩蔽阈值应取绝对闻阈。

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 3~20 ms,而滞后掩蔽却可以持续 50~100 ms。

2. 哈斯效应(Haas Effect)

实际中多个声源同时存在的情况非常普遍。若两个声源强度相等,先后到达人耳,如果其中一个延迟时间在 30 ms 以内,听觉上将感到声音好像只来自未延迟的声源,并不会感觉到经过延迟的声源的存在。当延迟时间超过 30 ms 而未达到 50 ms 时,听觉上可以识别出已延迟的声源存在,但仍然感觉声音来自未经延迟的声源。只有在延迟时间超过 50 ms 后,听觉上才会感到延迟声成为一个清晰的回声。这种现象称为哈斯效应,有时也称为优先效应(Precedence Effect)。

3. 鸡尾酒会效应(Cocktail Party Effect)

鸡尾酒会效应,也称选择性关注(Selective Attention),是指人的一种听力选择能力,在这种情况下,注意力集中在某一个人的谈话之中而忽略背景中其他的对话或噪音。该效应揭示了人类听觉系统中令人惊奇的能力,使人们可以在噪声环境中开展谈话。鸡尾酒会效应能够让多数人将很多其他无关的声音屏蔽掉,只选择听自己关注的那个声音。

4. 双耳效应(Binaural Effect)

双耳效应是人们依靠双耳间的强度差、时间差和音色差判别声音方位的效应。

如果声音来自听音者的正前方,由于声源到左、右耳的距离相等,从而声波到达左、右耳的时间差(相位差)、音色差为零,此时感受出声音来自听音者的正前方,而不是偏向某一侧。声音强弱不同时,可以感受出声源与听音者之间的距离。

由于双耳的生理结构复杂,且对称分布于脑部两侧,使得对于不同空间方位的声波到达双耳存在一系列差异,双耳正是利用这些线索完成了空间声定位等复杂的工作。双耳差异主要包括:

(1) 双耳时间差(Inter-aural Time Difference, ITD)。

由于左右耳之间有一定的距离,因此,除了来自前方和正后方的声音之外,由其他方向传来的声音到达两耳的时间就有先后,从而造成时间差。如果声源偏右,则声音必先到达右耳,后到达左耳。声源越是偏向一侧,则时间差也越大。实验证明,当声源在两耳连线上时,时间差约为 0.62 ms。

(2) 双耳声级差(Inter-aural Level Difference, ILD)。

两耳之间的距离虽然很近,但由于头颅对声音的阻隔作用,声音到达两耳的声级就可能不同。如果声源偏左,则左耳感觉声级大一些,而右耳感觉声级小一些。当声源在两耳连线上时,声级差可达到 25 dB 左右。

(3) 双耳相位差(Inter-aural Phase Difference, IPD)。

声音是以波的形式传播的,而声波在空间不同位置上的相位是不同的(除非刚好相距一个波长)。由于两耳在空间上的距离,所以声波到达两耳的相位就可能有差别。耳朵内的鼓膜是随声波而振动的,这个振动的相位差也就成为人耳判别声源方位的一个因素。当然,频率越低,相位差定位感觉越明显。

(4) 双耳音色差。

声波如果从右侧的某个方向上传来,则要绕过头部的某些部分才能到达左耳。已知声波的绕射能力同波长与障碍物尺度之间的比例有关。人头的直径约为 20 cm,相当于 1 700 Hz 声波的波长,因此频率为 1 000 Hz 以上的声波绕过头颅的能力较差,衰减越大。也就是说,同一个声音中的各个分量绕过头部的能力各不相同,频率越高的分量衰减越大。于是左耳听到的音色与右耳听到音色就存在差异。只要声音不是从正前方(或正后方)来,两耳听到音色就会不同,这也是人们判别声源方位的一种依据。

关于双耳效应的研究,不仅有益于进一步探索人类的听觉机制,在实际应用中,对于声源定位、空间听觉、3D 立体声等领域也具有重要的意义。

习 题 1

1. 试分析声吸收和声反射对声信号的传播可能造成何种影响。
2. 通过思考和查阅文献,分析鸡尾酒会效应和双耳效应可以在哪些研究领域得到应用。

第2章 声信号处理概论

2.1 信号与声信号

2.1.1 信号

信号(Signal)与消息(Message)、信息(Information)相似,在生活中被广泛使用,但同时也是具有特定意义的科学概念。三者既有密切的联系,又有所不同。将它们联系起来介绍,有助于理解信号的物理意义。

首先来看看信息与消息的区别。在通信工程中,一般将语言、文字、图像、数据等统称为消息。基于消息,人们可以对某个事物的状态有所了解,如果由此能消除人们对该事物状态的不确定性,则认为从中获得了信息。因此,信息被看作是用以消除信宿对信源发出何种消息的不确定性的东西。简单地说,信息是指有新内容、新知识的消息。信息与消息的关系是内容与形式的关系。消息是信息的载体,其形式是多样的、具体的,信息则是指包含在各种具体消息中的抽象内容。这也是信息论的创建者——贝尔实验室申农的基本观点。

再来看信号与消息的关系。所谓信号,是带有信息的某种物理量,如电信号、光信号、声音信号等。因此,信号是指消息的表现形式,而消息则是信号的具体内容。消息的传送一般都不是直接的,而必须借助于一定形式的信号才能便于传输和进行各种处理。例如,古代人利用点燃烽火台而产生的滚滚狼烟,向远方军队传递敌人入侵的消息,这属于光信号;当我们说话时,声波传递到他人的耳朵,使他人了解我们的意图,这属于声信号;遨游太空的各种无线电波、四通八达的电话网中的电流等,都可以用来向远方表达各种消息,这属于电信号。人们通过对声、光、电信号进行接收,才知道对方要表达的消息。因此,可以认为信号是运载消息的工具,是消息的载体。

在物理学中,信号可以理解为随时间或空间变化的某种物理量。因此,在数学上,信号可以用包含一个或几个独立变量的函数(Function)来表达,但它又与通常所说的时间或空间函数不同:

(1)信号一般可用函数来描述,但函数并不一定都是信号。信号是实际的物理量,而函数则可能只是一种抽象的数学定义。

(2)信号的值一定是单值,而函数的值则可能是多值。

信号也可以用函数的曲线图形,即信号的波形(Wave)来表示。如电信号的相位随时间的变化情况,既可以用函数来表达,也可以画出波形来表示。如用 $\sin t$ 表示的交流电相位的波形如图 2-1 所示。

除了上述两种直观的描述方法以外,信号的特性还可从两方面来描述,即时域特性(Time