中国社会科学院创新工程学术出版资助项目

语音学：标音、产生、声学和感知

[德]亨宁·雷茨　[荷兰]阿拉德·琼曼　著
曹梦雪　李爱军　译

中国社会科学出版社

中国社会科学院创新工程学术出版资助项目

语音学：标音、产生、声学和感知

[德]亨宁·雷茨 [荷兰]阿拉德·琼曼 著
曹梦雪 李爱军 译

中国社会科学出版社

图字：01－2015－0358

图书在版编目（CIP）数据

语音学：标音、产生、声学和感知／（德）亨宁·雷茨，（荷兰）阿拉德·琼曼著；曹梦雪等译．—北京：中国社会科学出版社，2018.10

ISBN 978－7－5203－3298－9

Ⅰ.①语… Ⅱ.①亨…②阿…③曹… Ⅲ.①语音学—教材 Ⅳ.①H01

中国版本图书馆 CIP 数据核字(2018)第 230313 号

出 版 人 赵剑英
责任编辑 张 林
特约编辑 宋英杰
责任校对 李 莉
责任印制 戴 宽

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 http://www.csspw.cn
发 行 部 010－84083685
门 市 部 010－84029450
经 销 新华书店及其他书店

印 刷 北京明恒达印务有限公司
装 订 廊坊市广阳区广增装订厂
版 次 2018 年 10 月第 1 版
印 次 2018 年 10 月第 1 次印刷

开 本 710×1000 1/16
印 张 22
插 页 2
字 数 339 千字
定 价 96.00 元

凡购买中国社会科学出版社图书，如有质量问题请与本社营销中心联系调换
电话：010－84083683

前　言

语音学一般可以分为三个分支领域：发音语音学研究言语产生的方式，需要了解发音器官的生理特性；声学语音学研究言语的声学特性，如频率、强度和时长，需要了解声波的知识；听觉语音学研究言语的感知，需要了解听觉系统和大脑存储的功能。因此，语音学是一门跨学科的科学，包括语言学、生物学、物理学和心理学等。此外，语音学专业的学生还必须熟悉语音标音，并会用一套符号来“记录”言语声。

一些语音学教材主要介绍发音语音学和语音标音，还有一些教材只是关注声学或听觉语音学。以我们的教学经验来看，某些语音模式从发音的角度能给出更好的解释，但另一些模式也许从听觉角度讲更容易理解，所以语音学教材应该全面介绍各个方面的知识，这也是我们写这本教材的初衷。

这本书介绍了语音学的四个领域：发音语音学、声学语音学、听觉语音学以及语音标音。适用于与语音相关的各个学科的学生（包括语言学、言语病理学、听力学、心理学和电子工程学）。虽然本书是一本导读性的教材，但书中有关语音学某些领域的内容还是较一般性的导读教材更为详细，可以满足不同读者（或者教师）的不同需求。由于本书采用了一种分步写作方式，读者完全可以跳过前面章节（比如技术性较强的章节）而直接进入任何一章。当然，一些技术性内容也是不可或缺的（比如语谱图的理解），要求读者只需具备高中以上水平的一些数学和物理知识就可以了，我们在技术性概念的解释方面给出了很多实例以帮助大家理解。此外，附录A和B提供了更多的技术性信息，帮助大家更好地理解正文的内容。本书采用模块化的方式综合覆盖了语音学的各个领域，其丰富的信息足以帮助读者完成深入了解这门复杂的交叉学科的挑战。

语音学作为一门研究言语的科学，并不是面向某一种语言的，但是由于本书是用英语撰写的，所以书中的例子以英语为主。当然，书中也给出了许多其他语言的例子来说明那些英语中没有的现象，不过读者并不需要具备这些语言的相关知识。

这本书反映了许多言语科学家的观点和研究成果，我们很荣幸成为他们中的一员多年来与他们一起讨论语音学问题。因此，我们首先要感谢 Aditi Lahiri 和 Joan Sereno，没有他们的持续支持和指导，本书将不可能完成。我们谨以此书献给他们。此外，我们还要感谢我们的老师们和导师们：Sheila Blumstein，Philip Lieberman 和 James D. Miller，是他们最先激发了我们对语音学研究的兴趣。还有那些使我们的兴趣日益浓厚的学生们：Mohammad Al-Masri，Ann Bradlow，Tobey Doeleman，Kazumi Maniwa，Corinne Moore，Alice Turk，Travis Wade，Yue Wang，Ratree Wayland。还有初级班和高级班中的很多同学们，通过他们提出的问题，我们才确认哪些话题是需要特别说明的重点。我们对 Ocke-Schwen Bohn，Vincent Evers，Carlos Gussenhoven，Wendy Herd，Kazumi Maniwa，Travis Wade，Ratree Wayland 和 Jie Zhang 等人也表示感谢，他们对本书此前的版本提出了宝贵的建议。在此，还要特别感谢为本书提供插图的 Regine Eckardt。最后我们也对 Wim van Dommelen，Fiona Mclaughlin，Simone Mikuteit，Joan Sereno，Craig Turnbull-Sailor，Yue Wang 和 Ratree Wayland 为我们提供的录音资料表示感谢。当然，他们都不必对本书中的任何谬误承担责任。

我们要特别感谢 Aditi Lahiri，由她资助的莱布尼茨奖为本书各个阶段的研究提供了大量的资金支持。康斯坦茨大学和堪萨斯大学提供了差旅支持和学术休假，使我们能够专注完成本书，我们对此一并表示感谢。

亨宁·雷茨
阿拉德·琼曼

目　录

1　**关于本书** …… (1)
1.1　初识语音学 …… (2)
1.2　本书结构 …… (7)
1.3　术语 …… (8)
1.4　演示及练习题 …… (8)
练习 …… (8)

2　**发音语音学** …… (9)
2.1　喉的发声 …… (9)
2.2　基本发音术语 …… (10)
2.3　辅音的发音 …… (13)
2.3.1　发音部位 …… (14)
2.3.2　发音方法 …… (14)
2.3.3　其他分类方案 …… (17)
2.4　元音的发音 …… (19)
练习 …… (22)

3　**语音标音** …… (24)
3.1　辅音 …… (28)
3.1.1　爆发音 …… (29)
3.1.2　鼻音 …… (29)
3.1.3　擦音 …… (30)
3.1.4　塞擦音 …… (31)

3.1.5 近音 …… (31)
3.2 元音 …… (33)
3.3 附加符号及其他符号 …… (37)
3.4 通用美语的标音 …… (39)
3.4.1 辅音 …… (39)
3.4.2 元音 …… (47)
练习 …… (49)
注释 …… (50)

4 辅音和元音的发音部位与发音方法 …… (51)
4.1 辅音 …… (52)
4.1.1 唇音 …… (53)
4.1.2 舌冠音 …… (54)
4.1.3 舌面音 …… (57)
4.1.4 咽喉音 …… (58)
4.2 其他发音方法 …… (59)
4.3 元音 …… (60)
4.4 次要发音 …… (63)
练习 …… (65)
注释 …… (65)

5 发音器官的生理机能 …… (66)
5.1 声门下系统:肺、支气管和气管 …… (66)
5.1.1 声门下系统的解剖结构 …… (66)
5.1.2 肺的运动 …… (68)
5.1.3 肺的容量与其随时间变化的控制 …… (70)
5.1.4 响度与肺部气压 …… (71)
5.2 喉的结构和功能 …… (72)
5.2.1 喉的解剖结构 …… (72)
5.2.2 声带振动 …… (76)
5.2.3 响度与喉部信号 …… (85)

5.2.4　声区 …… (87)
5.3　声道 …… (88)
5.3.1　咽 …… (89)
5.3.2　鼻腔通道和软腭 …… (89)
5.3.3　口腔通道 …… (91)
练习 …… (94)

6　气流机制与发声类型 …… (95)
6.1　气流机制 …… (95)
6.1.1　喉头气流机制 …… (95)
6.1.2　软腭气流机制 …… (97)
6.2　发声类型 …… (98)
6.3　爆发音的浊（带音性）、清（不带音性）与送气 …… (100)
6.4　常见语音与特殊语音 …… (106)
练习 …… (108)
注释 …… (108)

7　基础声学 …… (109)
7.1　声波 …… (109)
7.1.1　声波是气压的变化 …… (109)
7.1.2　声波的起源和传播 …… (110)
7.1.3　声速 …… (112)
7.1.4　声波中的相对位置 …… (113)
7.1.5　纵波与横波 …… (114)
7.2　声波的测量 …… (115)
7.2.1　麦克风 …… (115)
7.2.2　波形图 …… (117)
7.3　声学维度及其测量单位 …… (118)
7.3.1　频率 …… (118)
7.3.2　振幅 …… (124)
7.3.3　相位 …… (130)

练习 …… (132)
注释 …… (133)

8 语音的分析方法 …… (134)
8.1 声学信号的数字化 …… (134)
8.1.1 时域和振幅域的数字化 …… (135)
8.1.2 采样率 …… (136)
8.1.3 量化精度 …… (139)
8.2 声学信号的类型 …… (141)
8.3 声学信号的分析 …… (145)
8.3.1 傅里叶变换 …… (145)
8.3.2 通过谱可以观察到哪些信息 …… (154)
8.3.3 谱分析中的加窗 …… (154)
8.3.4 谱的其他表示形式:断面谱和语谱图 …… (160)
8.3.5 LPC 谱 …… (163)
练习 …… (165)
注释 …… (166)

9 言语产生的声源-滤波理论 …… (167)
9.1 共振 …… (168)
9.1.1 圆柱体声管的共振频率 …… (168)
9.1.2 非圆柱体声管的共振频率 …… (171)
9.2 阻尼 …… (174)
9.3 滤波器 …… (174)
9.4 发音器官的声源和滤波器 …… (178)
9.4.1 声道滤波器 …… (178)
9.4.2 唇和鼻孔的辐射 …… (179)
9.5 共振峰 …… (181)
9.5.1 共振峰频率 …… (181)
练习 …… (184)
注释 …… (185)

10　**语音的声学特征** …… (186)
10.1　元音 …… (186)
10.2　辅音 …… (190)
10.2.1　(央)近音 …… (190)
10.2.2　擦音 …… (193)
10.2.3　爆发音 …… (196)
10.2.4　鼻音 …… (198)
10.2.5　边近音 …… (200)
10.2.6　塞擦音 …… (201)
10.3　小结 …… (203)
10.4　可变性与不变量 …… (203)
10.4.1　声学不变量理论 …… (205)
练习 …… (210)
注释 …… (210)

11　**音节和超音段** …… (211)
11.1　音节 …… (211)
11.2　重音 …… (214)
11.3　音长 …… (219)
11.4　声调和语调 …… (221)
11.4.1　声调 …… (222)
11.4.2　语调 …… (225)
练习 …… (229)
注释 …… (229)

12　**听觉的生理学及心理物理学** …… (230)
12.1　外耳 …… (231)
12.2　中耳 …… (232)
12.2.1　中耳内的压力增加 …… (232)
12.2.2　中耳内的声音衰减 …… (233)
12.2.3　鼓室中的压力均衡 …… (234)

12.3 内耳 …… (235)
12.3.1 耳蜗中的压力波 …… (237)
12.3.2 作为振动体的基底膜 …… (238)
12.3.3 共振理论 …… (238)
12.3.4 对共振理论的反驳 …… (239)
12.3.5 行波理论 …… (239)
12.4 基底膜的结构 …… (241)
12.4.1 外毛细胞 …… (241)
12.4.2 内毛细胞 …… (242)
12.4.3 基底膜上的频率编码 …… (243)
12.4.4 耳声发射 …… (244)
12.5 听觉频率标度 …… (245)
12.5.1 线性标度 …… (246)
12.5.2 对数标度 …… (246)
12.5.3 美标度 …… (247)
12.5.4 巴克标度 …… (248)
12.5.5 等效矩形带宽(ERB)标度 …… (249)
12.6 听觉响度标度 …… (251)
12.7 听觉时间标度 …… (252)
练习 …… (253)
注释 …… (254)

13 **言语感知** …… (255)
13.1 元音 …… (256)
13.1.1 外部规整与内部规整 …… (258)
13.2 辅音 …… (261)
13.2.1 近音 …… (261)
13.2.2 擦音 …… (262)
13.2.3 鼻音 …… (263)
13.2.4 爆发音 …… (264)
13.3 言语感知的运动理论的贡献 …… (268)

13.3.1　范畴感知 …… (269)
13.3.2　言语是“特别”的吗？ …… (274)
13.4　语言学经验在言语感知中的作用 …… (279)
13.5　总结 …… (283)
练习 …… (284)
注释 …… (284)

附录 A …… (285)
A.1　质量、力和压强 …… (285)
A.2　能量、功率和强度 …… (288)
A.3　分贝 …… (291)
A.3.1　均方根振幅 …… (291)
A.3.2　均方根振幅与响度 …… (295)
A.3.3　用 dB 值进行计算 …… (298)
注释 …… (300)

附录 B …… (301)
B.1　物理学术语 …… (301)
B.2　数学符号 …… (304)
注释 …… (305)

附录 C …… (306)
C.1　共振峰频率值 …… (306)
C.2　基频值 …… (307)

参考文献 …… (308)

术语 …… (321)

1 关于本书

语音学是研究言语（speech）的一门跨学科的科学，研究内容包括四个领域：

- 言语是如何被记录的（即**语音标音 phonetic transcription**）；
- 言语是如何产生的（即**言语产生**或**发音语音学 speech production** or **articulatory phonetics**）；
- 言语的声学特性是怎样的（即**声学语音学 acoustic phonetics**）；
- 言语是如何被听音人感知的（即**言语感知**或**听觉语音学 speech perception** or **auditory phonetics**）。

本书介绍语音学这四个相关领域的知识，每个领域都与其他学科有所关联但是各自却又具有独特的方法。例如，言语标音以内省（有监督的）观察、仔细听辨和开口说话为基础；言语产生和声学语音学与生理、解剖以及物理学相关；而言语感知则更多涉及心理学问题。本书试图用常见的例子而不是“技术性”很强的术语向读者展示一些重要概念，因为我们坚信理解是获得知识的关键。

本书不仅面向语音学和语言学专业的学生，同时也面向心理学、计算机科学、医学、言语病理学和听力学等相关领域的学生。实际上，本书适合所有想了解我们如何说和听的学者阅读。语音学作为言语科学不是针对某一种语言的，由于原书是由英语撰写的，所以书中的例子以英语为主，但是我们也给出了其他语言的例子来说明英语中不存在的现象，不过读者并不需要具备这些语言的相关知识。

1.1　初识语音学

这一节将介绍语音学中的一些基本概念，在本书中我们将对这些概念进行详细的解释。这些概念术语如图1.1所示，从左至右包括：可以让我们说话的解剖结构、这些结构产生的声学信号以及让我们能够听到这些信号的听觉解剖结构。

在言语产生过程中具有重要作用的解剖器官由三个主要部分组成（见图1.1左侧）：**肺（lungs）**、**喉（larynx）**和**声道（vocal tract）**，声道本身包括嘴、鼻和咽。

肺用来呼吸，是产生言语声的主要能源。肺部流出的气流首先经过颈部的喉，而声带就位于喉部。声带在气流中振动，产生言语的音高：说话时，喉部的声带通过或快或慢的振动而产生旋律。这个重要的过程就称为**发声（phonation）**，由声带振动产生的言语声叫作**浊音（带音）（voiced sound）**。我们常说的“某人失声了”并不是指某人完全不能说话了，哑了，而是声带不能振动了，产出的是耳语声。声带之间的区域是很多言语声产生时的声源，这个区域有一个特别的名称叫作**声门（glottis）**。声道（嘴、鼻和咽）是语音产生过程的核心结构，这个过程叫作**发音**或**调音（articulation）**，语音产生过程中涉及的结构叫作**发音器官**或**调音器官（articulator）**。其中舌是最重要的发音器官，正如名词“mother tongue”和“language”（来自拉丁文里的“lingua”，即舌“tongue”）所体现的，我们的祖先早已知道它的重要性。

由于喉在发音系统中起到分割的作用，我们将喉上部的言语器官称为**喉上系统（supralaryngeal system）**，将喉下部的言语器官称为**声门下系统（subglottal system）**。

人类的**发音器官（vocal apparatus）**所产生的言语声在空气中以**声波（sound wave）**的形式传播，其本质上是气压的轻微扰动。在**波形图（oscillogram）**上，我们可以用图形来表示这些轻微的扰动，水平 x 轴表示时间，竖直 y 轴表示每个时间点上的压强（请见图1.2a中所示短语“How do you do?”的波形）。许多人第一次看到声波可能都会感到吃惊，因为词与词之间既没有像印刷的文字那样所体现出的停顿，也没有字母

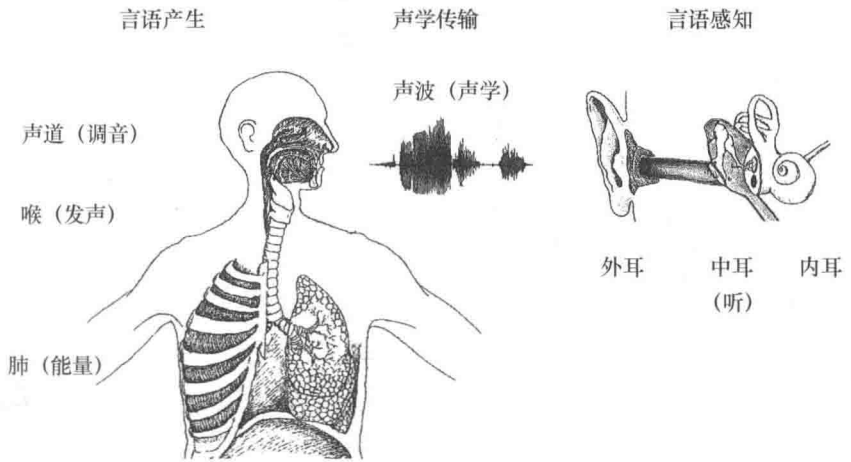


图 1.1 言语产生、声学传输和言语感知的主要成分

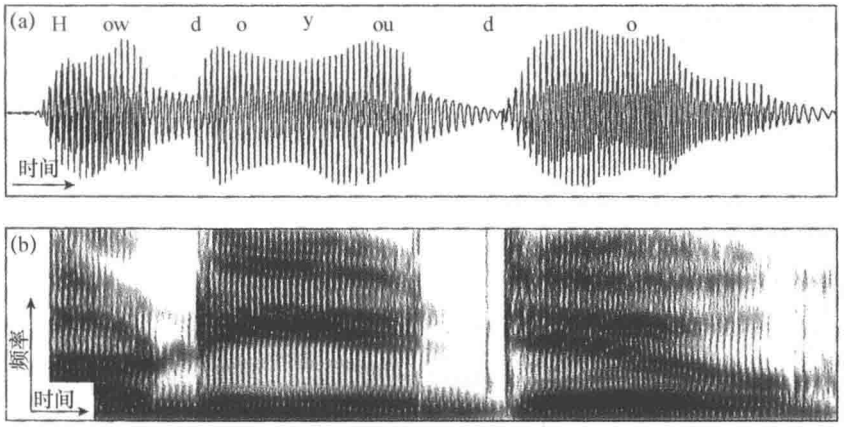


图 1.2 短语“How do you do?”的（a）波形图和（b）语谱图

与字母之间的分隔标记。其实，说话人并不会在词与词之间停顿（如“How - do - you - do”），因此言语声是连续的，正常情况下发音就像“Howdjoudo”，其中的“dj”类似于单词“jungle”的起始发音。这种发音的连续性以及词与词之间鲜有间断的现象，是成人学习第二语言面临的难点之一：母语者说话很快，会使词与词之间的边界变得含糊——这也是任何一种语言的母语者都具有的特点：发音器官连续地从一个声音

运动到另一个声音，一个词连着另一个词。因此，对这样一串声音的图形表示对于我们分析实际的发音是很有帮助的。

声音越响，它的声压变化就越大，波形图中的**振幅（amplitude**，即图 1.2a 中纵轴位移）就越高，就像海浪变高一样。如果声波在一定时间间隔内有规律地重复自己，即具有**周期性（periodic）**，那么我们便可以在波形图上观察到有规律的振动。如果声音没有规律，波形图上的信号显示就没有规则。没有声音时，波形图上显示的是一条水平线。因此，我们可以说波形图是声音波形的再现。

用**语谱图（spectrogram）**的方式分析信号和表达信号通常是进一步了解言语信号传递的声学信息的一个非常有效的途径（如图 1.2b 所示的语谱图，与图 1.2a 为同一句）。语谱图上，横轴通常表示时间，与波形图对应，纵轴表示不同音高区域的能量（更准确地应称为频带），频率沿纵轴不断递增。此外，语谱图上颜色的深浅表示音强的大小，颜色越深表示音强越大。

图 1.3 是另外一个例子，展示了伦敦大本钟（Big Ben）发出的钟声的前半段旋律。图 1.3a 的波形图显示有 4 个声学事件，但是没有进一步的分析我们并不能区分大本钟到底敲了什么音符。通过图 1.3b 的语谱图，有经验的学者可以看出这是一段钟声，而不是什么其他的比如吹号

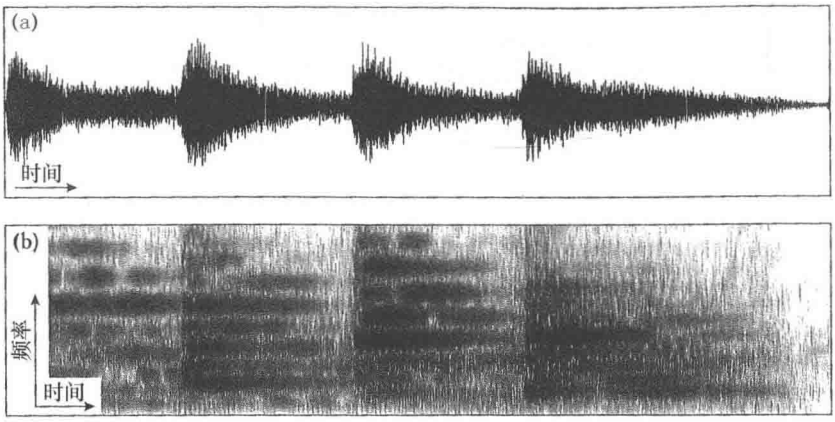


图 1.3　大本钟敲击的第一段旋律的（a）波形图和（b）语谱图

的声音。从图中我们还可以推测出钟的频率（即我们感知到的音高）。通过比较图 1.2 和图 1.3，我们很容易发现言语声的信号要比大本钟的信号复杂得多。

言语声最终到达听者的**耳朵**（**ear**，见图 1.1 右侧图）。耳不仅包括一个长在我们头外部的结构，我们称之为外耳，还包括听觉器官的核心部分，就是长在头内部的内耳。声能可以通过中耳里的一个机械系统从外耳传递到内耳，在内耳中的一个充满液体的腔体中把空气载声波（airborne sound wave）转换成压力波（pressure wave）。最终，大脑通过分析内耳感觉神经产生的信号，将其转化为对言语的感知。尽管我们不能直接观测到这个过程，但我们可以提出一些言语感知方面的理论，通过一些巧妙的实验加以验证。这有点类似于航天学家，他们可能从来也没有实际看到过某个星球，但却可以提出一些针对该星球的理论。遗憾的是，我们的感知并不像信号的物理特性那样易于测量，仅仅通过观察波形图或语谱图是不行的。比如，通过测量信号的振幅来表示声波有多“高”很容易，但振幅与人感知到的信号有多“响”并不是直接相关的。这种感知效应大家都有体会。比如在高速路上开车时听音乐，若停车休息几分钟后再启动车子，音乐声就显得特别大。无论在高速行驶还是在静止的车子里，音乐声的物理振幅都是相同的，但是人们感知到的声音大小却取决于背景噪声以及我们暴露于这种环境中的时间。

所有的活动——言语的产生、传输和感知——都是与声波紧密相连的，而且是实时变化的（run in real time）：当 DVD 停止播放时，图像可以停滞在那里，而声音却消失了。但是，我们是否可以将言语声描绘记录在纸上以对它们进行讲解呢？波形图和语谱图都是把信号记录在纸上的方式，但是不好理解，从图中推测某人说话的内容非常复杂。通常，我们用文字记录听到的内容，这么做的前提是我们知道如何拼写这种语言，记录的内容也可能与发音不太一致。比如，英语单词“cough，though，through，thorough”，都有“ough”这几个字母，但是发音却很不相同。因此，正则法（orthography）通常不是表示单词发音的好方法，言语声需要用特殊的符号即**国际音标（International Phonetic Alphabet：IPA）**来记录。其中一些符号看起来跟它们的书写形式相似，但这些语音上的 IPA 符号反映的是发音而不是字母，用方括号括起来以示区别。

本书里，我们使用双引号表示字母。例如，英语单词“ski”的 IPA 表示为［ski］，单词“cough，though，through，thorough”用 IPA 可以表示为［kɔf，ðoʊ，θɹu，ˈθʌɹə］，这种用语音符号记录发音的方法就是语音标音（phonetic transcription）。尽管语音标音看起来像外语，但这些单词底层的声音序列显然是有很大差异的，这反映了这些词在某种特殊的英语方言中的发音。大家一定要牢记描写声音的 IPA 符号与很多语言中使用的书写字母是不同的。

回来再看图 1.2a 中短语“How do you do?”的波形图，它真实记录了构成这个语音信号的气压扰动。通过观察这段波形，我们可以发现，语音并不是一串孤立的音段序列。这一点与打印出来的文字不同，文字是由一串孤立的字母组成的词串，空格把词与词隔开，而语音信号却是连续的、时刻变化的信息流。音段标音是一个相对主观的过程，反映了标音人对一串声音的理解和认识。但是，即使是发一个单一的声音也并不是一件简单的事情，例如“supper”中的辅音“p”，在几分之一秒的时间里，需要双唇、舌和喉的不同肌肉群精确协同工作。虽然这些动作最终产生的是一个具有不同成分的复杂的声学结构，但我们仍然会将其感知为一个**音段（sound segment）**；另外，即使从语流中去除了某个音段，相邻音段仍然会留下其发音的痕迹，去除的音段仍然可以被感知到。本书中，我们将介绍人们是如何产生、分析、感知和标记这种声音的。

此外，还有一些其他的语音特性与一个以上的音段相关，这些特性作用于一个音段以上的范围，通常被称为**超音段（suprasegmental）**。这里涉及**音节（syllable）**这个相当重要的概念，音节将几个声音组织在一起。在我们说话的时候，音节有轻有重，例如，作为形容词的“contrary”，在童谣“Mary，Mary，quite contrary”里，它第二个音节重读，但是“contrary”作为名词时，第一个音节重读，如短语“On the cóntrary，I said the opposite”。词重音可以改变词义，如“desért”重音在第二个音节上是“放弃、遗弃”的意思，而重音在第一个音节上时，désert 是“沙漠”的意思。显然重音对于语句的理解十分重要，但是重音不会在正则文字上表示出来（不过当重音改变时，人们可以感觉到元音音质上的变化）。另外一个超音段现象是说话时赋予一句话的**语调（intonation）**或者旋律。如“It is 10 o'clock.”这句话，音高在句末下降，而

问句“It is 10 o'clock?”的句末音高抬高以表示惊讶。两种情况的音段相同只有语调的差异。也有一些语言，如果单个音节的“语调”改变就会改变其词义，最好的例子就是汉语，比如“ma”的音高为平调、升调、降升调和陡降调的时候，分别对应“妈、麻、马和骂”不同的意思。这种对音高的使用被称为声调。对那些母语没有声调的人来说，这种音高的作用是很奇怪的，但是世界上确实有很多人都使用音高来区分词义。

1.2 本书结构

本书涵盖了语言学的四个领域：言语的标音、产生、声学和感知。我们不会将这四部分割裂开来，因为每部分之间都有一些交叠。这也体现了我们对语音学中言语的认识：为了理解言语，我们必须知道如何记录声音、声音是如何产生的、语音的声学相关量是什么以及听者是如何感知言语声的。为了同时了解这四个问题，我们必须提前对各个领域有所了解，因此，本书在某种程度上会分别对四个领域进行介绍。当不得不在详细介绍某一术语之前提及该术语时，我们会在其第一次出现的时候给出简要的描述。此外，技术性细节需要更长篇幅的解释，我们在保证正文部分可读性的前提下把相关内容放到了附录部分，尽管附录中的一些内容其实对深入理解相应知识至关重要。

第 2 章将描述发音器官的结构，这部分比较容易观测到，即喉部的发声（phonation）和声道内的发音（articulation）；第 3 章将介绍英语语音 IPA 标音的原则；第 4 章将系统介绍世界语言中许多辅音的标音；第 5 章将详细讨论呼吸系统和喉的解剖以及生理特性；第 6 章将解释不同气流机制产生语音的方法；第 7—8 章将介绍声音的基本物理特性，分析和测量声音的方法，以及各种利用计算机分析语音的方法；第 9 章在前两章的基础上将进一步介绍言语产生的声学理论。第 7—9 章的“技术性”相对较强，我们会尽量把其中的原理传达给那些没有太多数学基础的读者。第 7—9 章中介绍的概念和方法都会在第 10 章中得到应用。有趣的是辅音用一些发音术语更易描写，而元音则用声学术语更容易描述，因此我们就采用了一种发音和声学混合的描写法。本书最后三章也遵循相似的原则：言语不是一串孤立的音段而是听者最终感知到的具有更大结构

的连续语流。由于对这些更大的声音结构（言语必不可少的成分）的理解需要对每一个成分（即音段）进行理解，所以我们将这部分内容放在本书较为靠后的第 11 章。归根结底，言语因为可以被感知而存在。即使是儿童，通常也都是先进行言语感知而后进行产出。我们将听觉和感知放在本书的最后，因为这两个领域需要对语音声学有基本了解。其中，第 12 章将介绍听觉器官的结构，第 13 章将介绍一些言语感知方面的研究成果。附录将对一些技术细节进行详细解释，以飨那些对此感兴趣的读者。总之，全书 13 章描述了言语如何标记、如何产生、其声学特性是什么以及言语是如何被感知的。

1.3 术语

本书在介绍术语的时候，会采用**黑体**印刷。本书中的度量单位采用一般读者都比较熟知的公制单位（metric unit），对一些日常生活中的举例将同时使用公制（米和公里等）和英制（Imperial system，如英尺和磅）。

1.4 演示及练习题

每章后面设有练习题，以检查和加强读者对相关概念的理解。本书对应的网站上（http://www.blackwellpublishing.com/phonetics），放有更多的练习，特别是涉及声音和图形（如波形图和语谱图）使用的例子。网站也链接引用了本书，以便本书的内容能得到持续的更新、补充以及修改。

练　习

1. 列出并简要定义语音学的四个领域。
2. 言语产生中涉及的三个主要领域是什么？简要描述它们在言语产生中的作用。
3. 波形图是如何再现声音波形的？波形图与语谱图有什么区别？
4. 说明使用 IPA 符号而不是正则法来表示单词发音的理由。
5. 超音段特征的定义是什么？请举例说明。

2 发音语音学

语音学的重要内容之一是对声音的描写。其中的一种描写方式是对发音的描写，即对言语声的产生过程进行发音描写。首先，任何声音的产生都需要一个能源，产生言语声的能源是气流。对于世界上语言中的绝大部分语音来说，气流由肺部产生，这部分内容在5.1节中会有详细的介绍。气流从肺部产生以后，穿过气管，再经过喉（喉头），也即声带所在的部位。喉的构造以及它在语音形成的复杂过程中所起到的作用会在5.2节中详细讨论。咽腔（咽喉）、口腔（口）和鼻腔（鼻）等位于发音系统上半部的器官的生理特性会在5.3节中进行介绍。本章将介绍较易观察的发音器官，包括喉部的发声以及声道内的发音。对发音过程的这些描写有助于解释语音标音（见第3章）中所用到的术语。

2.1 喉的发声

喉（喉头）位于咽（咽喉）的底部，气管的顶端，由软骨、肌肉、韧带和其他组织构成。有些发音人的喉（从体外看是“喉结”）比较明显，它会在吞咽东西和说话时上下运动。但是这种上下运动并不是言语产生的主要原因，起关键作用的是喉中两条小肌肉的运动，这两条小肌肉便是声带（vocal folds 或 vocal cords）。而肺部所产生的气流须从它们之间通过。声带可以分开，也可以闭合，还可以紧闭。当它们分开时（在正常呼吸时），气流可以无阻碍地穿过两条声带之间的空隙。当声带紧闭时，气流无法通过，因此可以防止食物进入气管。当声带闭合时，肺部产生的气流会引起声带振动，这种振动被称为**浊嗓音（voicing）**或发声。需要注意的是，**浊音**或**带音（voiced sound）**的产生需要依靠声带的振动（声

带闭合），如元音和浊辅音；而发**清音**或**不带音**（**voiceless sound**）时声带不振动（声带分开）。你可以将手指轻轻放在喉结上来感知发音时声带是否振动，比如你将“zip”的起始音“z”拉长为“zzzzzzzz”，这是一个带音，发音时声带振动，你的手指应该能够感觉到这种振动。用同样的方法将“sip”的起始音拉长为“ssssssss”，它是一个不带音，发音时声带不振动，所以你的手指应该感觉不到任何振动。还有一种方法是用手罩住耳朵，然后发这些音。发带音（“zzz”）时，你应该会觉得头嗡嗡地响，而发不带音（“sss”）时则没有。这种嗡嗡声就是由声带振动引起的。

声带之间的空间叫作声门。通过调整声带，声门可以发出耳语声、气嗓音以及其他音质的声音（见6.2节）。

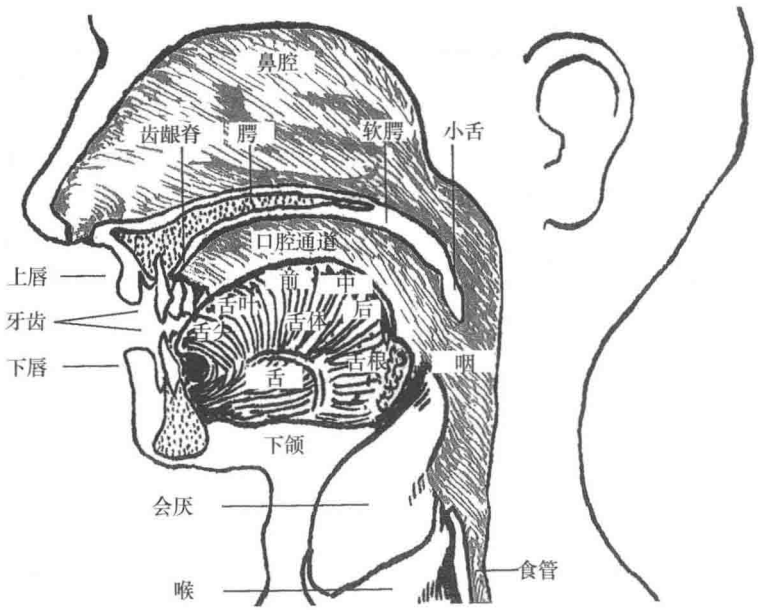


图2.1　声道

2.2　基本发音术语

图2.1为喉及喉上发音器官的侧面图（也称为剖面图）。我们将喉部

以上的气流通道总体称为声道，将喉部以上的器官统称为喉上器官。这些气流通道包括咽腔（咽喉）、口腔（口）和鼻腔（鼻）。

声道中参与发音的器官叫作发音器官。声音产生的基本原理是一个发音器官接近或接触另一个发音器官，这也是描述和产生语音的基本原理。通常，口腔下表面的发音器官会朝上面的发音器官运动。以下，我们将按照由声道前部到后部，也就是从唇到喉的顺序，来描述主要发音器官或喉上器官。

1. **唇（lip）**。上下唇都能参与发音。有唇参与的发音叫作**唇音（labial sound）**。
2. **齿（teeth**，主要指上门齿 upper incisors）。有齿参与的音叫作**齿音（dental sound）**。
3. **齿龈脊（alveolar ridge**，有时称为牙龈脊 gum ridge）。紧靠上齿背部的小突起，突起的程度因人而异。涉及这一部位的发音叫作**齿龈音（alveolar sound）**。
4. **腭（palate）**。口腔顶部前端坚硬的骨质部分。有时也被称作**硬腭（hard palate）**。发音过程中涉及这一部位的发音叫作**腭音（palatal sound）**。
5. **软腭（velum）**。口腔顶部后端柔软的肌肉。涉及这一部位的发音叫作**软腭音（velar sound）**。软腭在发音时还有另外一个作用：软腭可以抬升，使鼻腔关闭并与声道中其他发音器官隔离。用口腔呼吸时即是此种情况，空气只能通过口腔呼出。这种鼻腔的闭合被称为**软腭闭合（velic closure）**或**腭咽闭合（velopharyngeal closure）**。软腭抬高发出的音叫作**口音（oral sound）**。当软腭下降时，口腔和鼻腔的通道打开，鼻呼吸即是如此，气流可以从口腔和鼻腔通过。此种方法发出的音叫作**鼻音（nasal）**或**鼻化音（nasalized sound）**（鼻音和鼻化音的介绍见 3.4.2.2）。若气流在冲出口腔时受阻而只能从鼻腔通过，产生的音叫作**鼻塞音（nasal stop）**。
6. **小舌（uvula）**。小舌是悬在软腭下面的一小块楔形物。对着镜子，张大嘴，将舌压低，就能看到它。医生用木片压着你的舌，让你说“aaa”的时候，也能看见。小舌发出的音叫作**小舌音（uvula sound）**。

7. **咽（pharynx）**。小舌和喉之间的腔体叫作咽，也就是人们常说的**咽喉（throat）**。咽腔后壁可以被认为是声道上表面的发音器官。这里发出的音叫作**咽音（pharyngeal sound）**。

以上就是声道上表面的发声器官。在介绍声道下表面发声器官之前，应该先讨论一下通常我们所说的“喉头”。

8. **喉（larynx）**。通常为所有的带音提供声源。产出言语声时，喉中的声带同样可以形成声道中的最窄收紧点，因此我们也可以将喉看作是调音器官，用这种方法发出的音叫作**喉音（glottal sound）**。

组成声道下表面的调音器官包括：

1. 下唇，可主动靠近或接触上唇或上齿，分别发出双唇音和唇齿音。
2. 下齿，参与某些齿音的发音。
3. 舌，最重要的调音器官，固定在下颌上的一条肌肉。舌可分为以下六个区域，从前到后依次为：
   a）**舌尖（tongue tip）**。舌最前端的部分。舌尖（舌的顶点）发出的音叫作**舌尖音（apical sound）**。
   b）**舌叶（tongue blade）**。是舌尖后的一小段。当声道处于自然状态时，舌叶位于齿龈脊下方。**舌叶**（也称 **lamen**）发出的音叫作**舌叶音（laminal sound）**。
   c）**舌体（tongue body）**前部。舌体前半部实际上是舌的中部。当舌处在自然状态时，这部分位于硬腭的下方。
   d）舌体中部。在自然状态下，舌体中部大致位于硬腭和软腭的下方。
   e）舌体后部。舌体后部位于软腭下方。也被称为**舌背（tongue dorsum）**。
   f）**舌根（tongue root）**。舌正对着咽后壁的部分。用舌根发出的音叫作**舌根音（radical sound）**。
4. **会厌（epiglottis）**。跟舌一起向下向后运动的软骨结构。吞咽时，舌向后移动，会厌向下弯曲并盖住声门，把食物和液体送入食道。关于会厌在发音过程中所处的状态，大家的意见并不统一。有人认为会厌属于被动发音器官，只与舌一起联动，另一些人认为会厌是主动发音器官，例如，发希伯来语中一些辅音时便是如此。

以上是对声道的发音器官和结构的基本介绍。前文说过，语音产生的基本原则是一个发音器官接近或接触另一发音器官。那些运动的发音器官叫作**主动发音器官（active articulator**：包括唇，舌尖，舌叶，舌体前，舌体中，舌体后，舌根，会厌，软腭和喉），那些静止不动的发音器官叫作**被动发音器官（passive articulator**：包括唇，齿，齿龈脊，硬腭，软腭，小舌和咽壁）。

上述提到的发音器官中，有两个既可以是主动发音器官也可以是被动发音器官。唇既是主动发音器官又是被动发音器官：下唇是典型的主动发音器官，上唇通常是被动发音器官。下唇经常向上运动，与上唇接触。软腭有时是主动发音器官，有时是被动发音器官。区分口音和鼻音时软腭是主动发音器官：软腭抬高，与咽壁接触，气流不能通过鼻腔，形成口音；软腭下降，鼻腔通道打开，形成鼻音或鼻化音。软腭也可以是被动发音器官，当舌的某个部分（主动发音器官）接近或接触软腭时，形成软腭音。此外，小舌的振动并不是肌肉运动的结果，而是在肺部气流带动下的颤动，就像树叶随着风抖动一样。因此小舌也被认为是被动发音器官。除了唇、软腭和小舌，所有的被动发音器官都是硬质的结构。

2.3 辅音的发音

了解了每个发音器官和声道结构之后，我们现在可以对特定的言语声进行描述。言语声的两个基本分类之间的区别也就是**辅音（consonant）**和**元音（vowel）**之间的区别。语言学家常把辅音缩写为 C，元音缩写为 V，单词“milk”可表示为 CVCC。将言语声划分为元音和辅音的标准是它们的发音和产出方式。发元音时，口腔是打开的，气流通过不受阻碍，同时声带通常是振动的。但发辅音时，气流常受到各种影响，可能存在以下的任意一种情况：

1. 阻塞，形成（口）塞辅音；
2. 阻碍，受阻碍较大时，形成擦音，受阻碍较小时，形成近音；
3. 气流改道从鼻腔通过，形成鼻辅音。

另外，声带可能振动，也可能不振动，振动时形成浊辅音，不振动时则形成清辅音。

下面我们将言语声分为辅音和元音两类。我们先来介绍辅音。为了发一个辅音，气流从声道通过时必须受到某种阻碍。因此，我们可以根据阻碍产生的位置和程度来对辅音做进一步分类。这两个方面用语言学术语来讲就是**发音部位（place of articulation）**和**发音方法（manner of articulation）**。本章将介绍英语辅音的发音部位和发音方法。其他语言的情况将在第4章中介绍。

2.3.1　发音部位

按照声道由前向后，英语中可区分如下几种发音部位：

1. **双唇音（bilabial）**。发双唇音时，上下唇都参与发音，例如英语中“peak，beak，meek”的首音。
2. **唇齿音（labiodental）**。下唇与上门齿接触，比如“fine，vine”的首音。
3. **齿音（dental）**。齿辅音的发音涉及舌的舌尖或舌叶以及上门齿，例如“thigh”和“thy”的首音。
4. **齿龈音（alveolar）**。齿龈音涉及舌的舌尖或舌叶以及齿龈脊，如“tip，dip，sip，zip，lip，rip，nip”中的首音（对卷舌音变体“r”的讨论，见3.1.5）。
5. **龈后音（postalveolar）**。龈后音涉及舌的舌尖或舌叶以及齿龈脊后部。发音时，舌被抬向齿龈脊及硬腭前部，例如“sheep，genre，cheap，jeep”的首音。
6. **腭音（palatal）**。腭音涉及舌前部和硬腭，如“yes”的首音。
7. **软腭音（velar）**。软腭音涉及舌后部和软腭，例如“coal”和“goal”的首音，以及“sing”的尾音。
8. **喉音（glottal）**。英语有一个在声门处发出的喉音，发这个音时，气流从打开程度较大的声带间穿过，例如“heat”的首音。

通过上面的例子我们可以看出，每一个发音部位都对应若干个音，而它们之间可以通过清浊及发音方法来做进一步区分。

2.3.2　发音方法

对辅音进行分类的另一个参考维度是阻塞程度。阻塞的程度与发音

的方法直接相关，所以我们将这个维度称作发音方法。发辅音时，气流可能被完全阻塞或被部分阻碍，也可能转而从鼻腔通过。因此，辅音的发音方法可分为以下几类：

1. **爆发音（plosive）或塞音（stop）**。发爆发音时，发音器官相互靠近形成收紧（即成阻），短暂地完全阻塞口腔中的气流。对（口）塞音来说，软腭抬高，阻塞鼻腔，气流既不能从口腔通过也不能从鼻腔通过。从肺部而来的气流在口腔中聚集，在成阻位置后方形成压力。当发下一个音时，发音器官打开阻塞，气压会被释放，而产生一个小的声音爆破（见图 2.2b 中 A 处的“冲直条”）。这个爆破的声音可被认为是一个小的爆发，因此塞辅音也叫爆发音。由于塞音有短暂的气流阻塞，因此它们不能被单独感知。为了听到气流的释放，通常需要与元音搭配。在英语中，“buy，die，guy”的首音是浊爆发音，“pea，tea，key”的首音是清爆发音。英语爆发音的发音部位有三个：双唇音（“pie，buy”），发音时双唇闭合；齿龈音（“to，do”），发音时舌尖或舌叶抵住齿龈脊；软腭音（“cot，got”），发音时舌后部抵住软腭。
2. **鼻音（nasal）**。发鼻音（或鼻塞音）时，气流从鼻腔中通过。鼻音与爆发音类似，口腔完全闭塞，气流不能从口腔流出。但是因为发鼻音时，软腭下垂，气流可以从鼻腔流出，最终形成鼻塞音。英语的鼻音是浊音，它们与爆发音所用到的发音部位相同：双唇音（“me”），齿龈音（“no”）和软腭音（“long”的尾音）。也就是说，鼻音在口腔中的调音与同部位的爆发音一致，唯一的不同是发鼻音时软腭下垂。
3. **擦音（fricative）**。发擦音时，一个发音器官接近另一发音器官，在二者之间形成一条狭窄的通路。肺部的气流在相当大的压力下，被迫从狭窄的通路中通过，产生了噪声湍流，发出的音具有嘶嘶的音质（如图 2.2b 中点 D 处的深色高频区）。在发这些音时，因为气流没有被完全阻塞，所以我们可以持续并独立地发出一个擦音。英语擦音的发音部位有五个：唇齿音（“fine，vine”），发音时下唇靠近上门齿；齿音（“thigh，thy”），发音时舌尖或舌叶与上门齿间形成一条狭窄的通道；齿龈音（“sip，zip”），发音时舌

尖或舌叶抬高接触齿龈脊；龈后音（如“pressure”和“measure”中的“s”），发音时舌尖或舌叶向着齿龈脊的后部抬高，再向硬腭延伸；喉音（“high”），发音时气流通过呈打开状态的声门。

4. **塞擦音（affricate）**。塞擦音是两种发音方法的结合，即按顺序发塞音和擦音。由于塞音和擦音基本上是由同一发音部位发出的（是**同器官的，homorganic**），所以我们认为塞擦音只表现为一个发音单元，是一次发音运动的结果。在英语中，塞擦音是龈后爆发音与龈后擦音的结合，如“chin”和“gin”的首音。发音时，首先舌尖或舌叶抬高以迅速接触齿龈脊后部，然后舌尖与上腭的接触松开，形成同一发音部位的擦音。
5. **近音（approximant）**。发近音时，一个发音器官接近另一发音器官，收紧的程度虽不足以产生摩擦，但远比元音大而比擦音小。近音兼有元音和辅音的部分特性，因此有时候近音也被称为**半元音（semi-vowel）**。发近音时，气流并没有被完全阻碍，所以这一点和元音类似；而尽管收紧程度并不严重，但是气流会受到部分阻碍，这与辅音类似。近音在世界语言中的表现和辅音十分相像，例如，近音不单独出现，而元音可以单独成词，如英语中“I”或“awe”。近音又分为**央近音（central approximant）**和**边近音（lateral approximant）**。央近音是由口腔中部（正中矢状面）的一股“央”气流形成的。英语的央近音涉及三个发音部位：唇-软腭音（“way”），发音时，舌后部向软腭靠近，同时撮唇或者圆唇；齿龈音（“rye”），舌尖抬向齿龈脊；腭音（“you”），舌前部抬高趋近硬腭。发边近音时，中间的通路被阻塞，气流从阻塞部分的两侧流出。英语有一个边近音，发音部位为齿龈音，如“lie”。发边龈近音时，舌尖与齿龈脊完全接触，但气流并未因此受阻，而是转从舌的两侧通过。

综上所述，我们可以从清浊、发音部位和发音方法这几个维度来描述辅音。清浊表示一个音是带音还是不带音。按发音部位划分，英语的辅音可分为双唇音、唇齿音、齿音、齿龈音、龈后音、腭音、软腭音和喉音。若按发音方法分，则可分为爆发音、鼻音、擦音、塞擦音和近音。除此之外，再加上口/鼻和央/边这两个维度，辅音的描述就完整了。口/鼻表示一

个音是口音还是鼻音，央/边表示气流是从口腔中部还是口腔两侧流出。

例如，“zest”的起首辅音可以描述为“浊齿龈央口擦音”。英语中的大部分语音都是央音（只有“l”是边音）。另外，大部分音都是口音，发音时软腭闭合（只有“m”，“n”，“ng”是鼻音）。因此，央音和口音通常被认为是默认的属性，不用作特别说明，如此“z”的描述可以简化为“浊齿龈擦音”。

有一点需要注意，就是描述辅音的发音通常只描述被动发音器官，而默认主动发音器官是舌。在发音过程中，舌会接触或接近被动发音器官。例如，如果舌抬高接触齿龈脊，就称为齿龈塞音。类似地，如果舌抬高并在齿龈脊处形成一条狭窄的通路，就叫作齿龈擦音。尽管在上述过程中运动的是舌，但是在发音描述中却没有提到舌。这样就避免了繁复的术语，例如“舌－齿龈”等术语。但是在下列情况中，主动和被动发音器官都要予以说明：

1. 舌既不是主动发音器官也不是被动发音器官，如发双唇音或唇齿音时；
2. 需要精确描述时。例如，舌尖接触齿龈脊和舌叶接触齿龈脊都可以简单描述为齿龈音；若必须区分这种不同的情况，则应分别称为舌尖－齿龈音和舌叶－齿龈音。

2.3.3 其他分类方案

在结束辅音这一节之前，我们有必要多了解一些其他的辅音分类术语。爆发音、塞擦音和擦音都伴有气流的严重阻塞，因此人们有时将它们统称为**阻音（obstruent）**。近音可以分为**滑音（glide）**和**流音（liquid）**。滑音发音时从一个位置滑到另一个位置，流音的音质不尖锐。在英语中，唇－软腭近音（“way”）和硬腭近音（“you”）是滑音，而央齿龈近音（“rye”）和边齿龈近音（“lie”）是流音。此外，鼻音和近音有时统称为**响辅音（sonorant consonant）**。响音包括了声道没有被完全阻断且没有湍流的音，因此元音也属于响音。发响音时，在收紧部位后方不存在气压的增加，这一点与阻音不同。

有时我们可以将发音部位进行更为宽泛的分类：唇音，发音时涉及唇；**舌冠音（coronal sound）**，发音时涉及舌尖、舌叶或舌体前部；**舌面**

音（dorsal sound），用舌面发音；**咽喉音（guttural sound）**，是咽腔中的发音或者用喉发的音。这种分类不算太精确，但是反映了一个事实：世界上很多语言在以上四组发音区域内部，分别只存在一组对立，并且这四组发音区域是由三组相对独立的肌肉控制的。

目前讨论的各种发音类型都可以在波形图和语谱图上进行辨识，如图2.2给出的“conceptualizing”。塞音（图2.2中的点A和点F）在波形图上振幅为0，在语谱图上对应空白段，这些区域对应肺部的气流被声道中的收紧阻塞的时间段。清擦音（点D）是由声道中严重的收紧产生的，因此在波形图上呈现气压随机变化的无规则模式，在语谱图上表现为能量在高频区的集中（纵轴顶部）。与此形成对比的是，浊擦音（点K）在波形图上有较低的振幅，代表气压的变化较小，在语谱图上表现为较弱（颜色较浅）的条纹。在语谱图上，浊音最容易辨别，其表现为低频区（纵轴底部）有规则的黑色条纹。我们可以通过和元音对比来了解近音的声学特征。元音受阻较小，在波形图上振幅较大，在语谱图上表现为多个频率段上的黑色带状条纹（点B，E，H，J和L）。由于近音比元音的收紧程度大，所以尽管其表现模式与元音相似，但是其振幅较小，如波形图和语谱图上点I所示。鼻音在语谱图上通常表现为低频区有规律的黑色条纹，代表浊音的特性；另外，高频区可能还会存在较弱的能量带（点C），也可能不存在能量带（点M）。

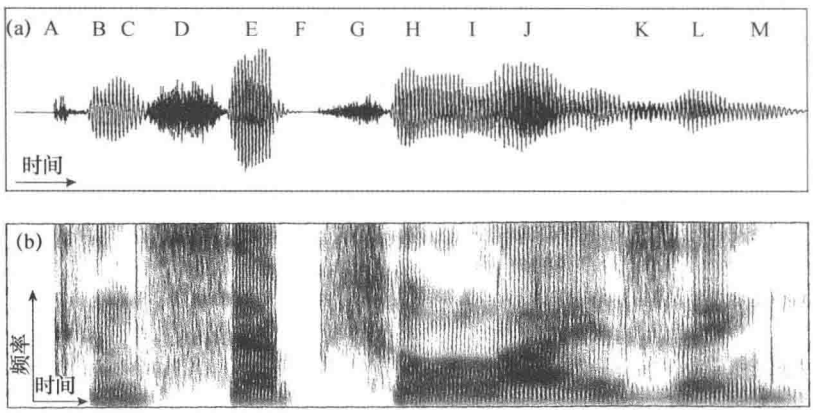


图2.2　单词“conceptualizing”的（a）波形图和（b）语谱图

塞擦音在语谱图上既表现出了塞音的特征也表现出了擦音的特征，如点 G 所示。清塞擦音的阻塞段在波形图上表现为低振幅，在语谱图上表现为空白段。除阻后紧接着一个擦音部分，在波形图上表现为气压的不规则波动，在语谱图上表现为高频区能量的集中。

2.4 元音的发音

发元音时，发音器官之间不像发辅音时那样靠近，因此气流相对不会受到阻碍。这一过程中，声带几乎一直保持振动，承担声源的角色。通过改变声道的长短和形状，我们可以发出不同的元音。舌的形状是决定声道形状的首要因素，而声道的形状可以进而影响声门处产生的声音。但是，我们对元音的发音特征的描写只关注舌最高点的位置，而不是描述舌整体的形状。另外，双唇的位置也会影响元音的音质。总之，元音的分类涉及三个主要的维度：

1. **前后（frontness）**。大部分元音发音时，舌都有一定的拱度，拱的方向可能朝向口腔的前、中或者后部。这种水平维度被称为元音的前后。
2. **高低（height）**。通过舌体在下颌内的上下运动，可使口腔变窄或变宽。这种纵向的维度叫作元音的高低。虽然舌通常与下颌一起联动，但是舌与下颌的运动是相互独立的。
3. **圆唇（lip rounding）**。发元音时，可以圆唇，也可使唇保持中性状态，即不圆唇或者展唇。

英语元音可分为前元音和后元音两组。每一组中，元音又可以分为高、中、低三个层级。英语与其他许多语言不同，在英语中，没有任何两个词可以仅靠元音的圆唇度来进行区分。但是德语与英语不同，德语中有很多词是仅凭元音的圆唇度来区分的，如“l̲esen”（意为“读”）中的“e”是前中不圆唇元音，而“l̲ösen”（意为“松开”）的“ö”为前中圆唇元音（见 10.1 节）。

我们将以“b̲eat”中的元音为例来进一步说明如何根据前后、高低、圆唇这几个参数来描述元音的音质。发这个元音时，舌体抬高并且相对靠前，舌尖轻轻接触下门齿的后部，舌的中部至前部抬高。单词“b̲it，

bet，bait，bat”中元音的发音情况与以上所描述的情况相同。发这五个元音时，舌的最高点都在口腔前部，所以我们称这些元音都是前元音。

这五个前元音的差别在于舌位的高度。“beat”中的元音，舌位抬高靠近硬腭。“bit”中的元音，舌位的最高点稍低一点。而“bait，bet，bat”中元音的舌位最高点又依次降低。实际上，发“bat”中的元音时，舌在口腔中很平。元音的高低是纵向维度，所以在描述元音的高度时，多使用**高（high）、中（mid）、低（low）**这样的词汇。例如，“beat”和“bit”中的元音是高元音，“bet”和“bait”中的元音是中元音，“bat”中的元音是低元音。“beat”与“bit”以及“bait”与“bet”之间的区别主要是元音的紧和松的区别，我们在3.2节中将作详细说明。有时，高和低也用**闭（close）**和**开（open）**来表示，因为对高元音来说，口腔通道相对闭合，对低元音来说，则相对打开。英语中的所有前元音都是不圆唇的。

如果从前后、高低和圆唇三个方面描述元音，“beat”中的元音为前高（或闭）不圆唇元音，“bat”中的元音为前低（或开）不圆唇元音。“bit”，“bait”和“bet”中的元音则分别为前高不圆唇（松）元音，前中不圆唇（紧）元音，以及前中不圆唇（松）元音。

对后元音的描述可采用同样的方法。以“boot”中的元音为例，发音时，舌位的最高点抬高接近口腔后部的软腭。由于舌位最高点在声道的后部（关于这点，见3.2节），这些元音叫作后元音。对“boot”中的元音来说，舌位接近声道的后表面，“cook，coat，caught，cot”中的元音也是相同的情况。（这里有一点需要说明，对很多说美式英语的人来说，“caught”和“cot”中的元音是一样的。他们在说这两个词时，其中的元音和“father”中的元音是同一个，因此，这种美式英语中的后元音只有四个而不是五个。）

看来，英语的不同变体存在差异，因此在描述语音时有必要说明我们描述的具体是哪一种英语。在本章以及后续章节中，我们所描述的是美国北部的英语，也就是通用美语（General American，GA）。我们经常可以在国家广播中听到通用美语，这种英语基本没有地域口音标记。通用美语不包括南部和东部海岸地区的美式英语，特别是新英格兰和纽约的美式英语。说通用美语的人说“cot，palm，father”这三个词时，词中

的元音是一样的，都是后低元音。在通用美语内部也存在地域性的口音差异。最明显的例子是，美国中西部和西部人以及加拿大人所说的“cot”和“caught”中的元音都是后低元音。除非特别说明，请大家记住，我们所说的英语都指这种通用美语，在我们对语音进行描述和标音时，也是如此。

大部分口音的差异都体现在元音上。英国国家广播电台一直使用标准发音（received pronunciation，RP，也叫作标准南部英式英语）。标准发音中的元音比通用美语多，例如“pot，palm，father”这三个单词就分别包含三个不同的元音。另外，标准发音不卷舌，即非元音前的“r”不发音。

我们再来讨论后元音，它们同样也可以根据高低（开口度）来进行区分，“boot”和“cook”中的元音是高（闭）元音，“cot”中的元音是低（开）元音。“coat”和“caught”中的元音可以描述为后中元音。“coat”中的元音是紧元音，而“caught”中的元音是松元音。

英语中，我们还剩三个元音没有提到。它们发音时都处在水平方向上前-后范围的中点，因此叫作**央元音（central）**。其中最常见的一个元音是非重读元音，例如“about”的起首元音。该元音发音时，舌位高点近似位于高低和前后的中点，因此称为中央元音（mid central vowel）。这个非重读的元音发音位置不高不低，不前也不后，处于“中性”位置，因此我们将这个元音称作**中性元音（schwa）**。另一个央元音是如“fur”中的元音，发“ur”这个音的时候，中性元音和辅音部分在发音时间上几乎交叠，所以不能把这个音描述成一个中性元音加上一个“r”，而是应将它们当作一个音段，描述为卷舌音色（r音色）央元音。最后，“cut”中的元音为低央元音，它与中性元音类似，但是要重读，且比中性元音略低。

还有一些音通常也被称为元音，我们还没有提到，如“buy，cow，boy”中的元音。这些音比较特殊，常被称作**二合元音（diphthong）**。二合元音在发音过程中，音质有显著的变化，因此我们可以将二合元音描述为由一个元音移动到另一个元音的发音，而通常第一个元音较为重要。为了与二合元音相区别，我们将没有明显发音动程的元音称作**单元音（monophthong）**，或纯元音。

英语中的二合元音描述如下。“buy”中的二合元音，发音器官由低央不圆唇元音移向前高不圆唇元音；“cow”中的二合元音则从同一个低央不圆唇元音向后高圆唇元音移动；“boy”的二合元音从后低圆唇元音向前高不圆唇元音移动。如何判断二合元音是由两部分组成的呢？最简单的方法就是延长其发音。二合元音的第一部分或第二部分可以分别延长，但是二合元音整体不能延长。虽然我们把二合元音描述为两个元音的组合，但是有些二合元音中的第一个元音并不是英语中的单元音，例如“buy”和“cow”中的二合元音都是以低央不圆唇元音开始的。

为了与“bait”和“boat”中的元音区分，“buy”，“cow”和“boy”中的二合元音有时候被称为“真”二合元音。“bait”和“boat”中的元音虽有一定的发音运动，但是动程没有那三个“真”二合元音那么明显。在英语中，“bait”和“boat”中的元音被叫作**二合元音化（diphthongized）**的元音。也就是说，发这些元音时，发音器官的运动使元音的音质发生了明显的变化，这些变化延续至发音结束，被叫作**出渡（offglide）**。“bait”和“boat”中的元音可以分别描述为前中不圆唇二合元音化元音和后中圆唇二合元音化元音。

总结一下本章的内容：元音可以从三个维度（高低、前后和圆唇）来描述；二合元音是一类特殊的元音，由两个元音组合而成。我们将在6.2节中讨论与音质相关的其他维度。介绍过辅音和元音的发音描写之后，下一章我们将介绍如何用一套符号来描写任意一段口语，这个过程也就是语音标音。

练　习

1. 描述发声的过程。
2. 舌尖音、舌叶音、舌面音在发音时分别涉及哪些发音器官？这几种音之间有什么不同？
3. 被动发音器官和主动发音器官之间有什么不同？举例说明为什么有的发音器官既是主动发音器官又是被动发音器官？
4. 下列词中划线部分的发音部位和发音方法是什么？
   “junk”，“did”，“shy”，“my”，“yes”。

5. 本章中介绍了几种划分元音的发音维度，请给出它们的名称和相应的定义。
6. 根据上述维度为下列单词中的元音分类：
“sat”，“look”，“weep”，“cup”，“odd”。

3 语音标音

虽然我们通过对语音进行发音描述可以较好地反映具体语音的特点，但是用发音术语来描写每一个音是相当烦琐的。例如对英语单词“speech”的发音描写如下：一个清齿龈擦音，接一个清双唇爆发音，再接一个前高不圆唇元音，最后是一个清龈后塞擦音。因此，为了更有效地描述语音，语音学家发明了一套符号，即国际音标（IPA），使用时，IPA 符号应置于闭合的方括号（[ ]）内。为了适应各种新记载的语音，成立于 1886 年的国际语音协会（首字母缩写也为 IPA）不断对这套符号进行修订。理想情况下，任何人只要接受过训练，都可以用这套符号对任何一种语言进行**标音（transcribe）**。接受一些正规的训练，可以使记音者的听音更加客观，并尽可能不受母语和方言的影响。专业人员应该可以根据最终的标音结果重构语音，但重构的还原度如何当然取决于标音的细致和准确程度。

对口语语音最真实的表征形式是录音。但是声音不能打印在纸上，而波形图或语谱图这种图像式的表征方式也无法“朗读”出来。在用书面符号表示言语声的方法中，最详尽的是**严式标音（narrow transcription）**。严式标音尽可能多地描写语音的细节，反映说话人的方言特点及个人特征。对语言学习者来说，当他们翻阅词典查找某一单词怎么读时，严式标音可能并不一定十分有帮助，因为对一种给定的语言来说，一个单词可能有若干种严式标音。在这种情况下，一种理想化的标音即**宽式标音（broad transcription）**更为合适。宽式标音没有严式标音那么细致，只包括正确产出这个单词所必需的几个音，而不标记语言相关或说话人相关的特征。例如，英语中，单词起首位置的“p”发音时会**送气（aspiration）**，在除阻之后可以听到一股气流冲出，字典里只记为 [p]。

而法语中，这个音除阻后没有气流冲出（也就是不送气），但是字典里也记为［p］。这两处所标的都是清双唇爆发音，而与语言相关的送气或不送气特征则不必在宽式标音中标出（不过一本好的字典应该在引言部分说明该语言中各种音的产生方式）。严式标音或宽式标音中所表示的语音叫作**音素（phone）**，有时也叫作音段（sound segment）。

实际上，在严式标音和宽式标音之间，还可能存在若干中介的标音层级，标音的精细程度取决于标音的目的以及标音员的经验和训练。假设我们要标记的是一种未知的语言，由于不知道哪种语音特性更重要，所以我们需要标出尽可能多的语音细节。这种标音极为细致，而且不包含词汇边界，因为我们对该语言的词汇还一无所知。这种级别的标音叫作**印象标音（impressionistic transcription）**或**通用语音标音（general phonetic transcription）**。印象标音和严式标音总是需要有（现场的或录制的）语音信号才可以完成，而宽式标音则可以不依赖于语音信号。

如此多种类的标音方法可能会让初学者感到困惑，但是在实际生活中，我们会根据不同的需求选用不同级别的标音方式。这就好比司机和徒步者所需的地图不同：徒步者需要地图上能够标出斜坡，而司机则不需要；但如果地图上不标台阶的话，司机就麻烦了。若在一张地图上把所有这些信息都标出来，那就密密麻麻没法看了。所以一张理想的地图通常是按需绘制的。标音也是这个道理：根据描写所需要达到的精细程度，以及对该语言发音情况的了解程度，标音的结果详略有异[1]。

相较于严式标音，宽式标音已略去了很多细节，但是若采用**音位表征（phonemic representation）**来描写语言，结果将更加简略。音位表征比较抽象，代表的是语音在人脑中的心理表征。音位标音关注的是语言中的**音位（phoneme）**，而不是具体的语音实现。为了将音位这种心理表征与音素这种物理实现加以区分，音位在标音中用双斜线（//）表示。以“handbag”为例，其发音可以实现为“ha［mb］ag”，这一宽式标音表示“d”音被删去，而“n”继承了其后的双唇音“b”的发音部位。相较而言，其音位标音为“ha/ndb/ag”，表示这个词在说话人心理词典中的储存形式。这种研究语音的抽象表示及其之间关系的科学是**音系学（phonology）**。对于罗马书写系统的使用者来说，他们对抽象表征与物理实现之间的差异并不陌生：字母“A”对他们来说是一个抽象的概念，因为这

个字母可以有多种物理实现，它可以被写成“A”，“a”，“ɑ”，“ɐ”，“*a*”等。

由于掌握严式标音需要进行大量的训练，因此，我们在这里将重点关注宽式标音，只在有必要的时候补充严式标音的相关内容。不过我们必须要注意一点，标音是语言相关的，人们对语音的感知难免会受母语的影响，这一点我们将在13.4中介绍。在一门语言中，一个音可能会被实现为不同的语音形式，但对母语者而言，他们并不会意识到各个语音之间的差异。而对另一门语言的使用者而言，这种语音上的差异就可能对应两个不同的音位，该语言的使用者可以很轻易地听出这种区别。以“l”和“r”为例，在英语中它们是两个不同的音，但日语母语者常觉得它们是同一个音。另一方面，没有经过正规训练的英语母语者听不出升调的“ma”（“麻”）和降升调的“ma”（“马”）有什么不同，他们会觉得声调很难掌握，但是一个说汉语普通话的三岁小孩儿却可以轻易地感知到两个音之间的差异。

语音标音的一般原则是，对某一特定的语言来说，应使用唯一的符号表示每个**区别性声音（distinctive sound）**。但是怎么确定哪些是有区别性的声音呢？母语者可能从来没想过他的母语中有多少区别性声音。那么初次接触一种语言时，该如何记录呢？我们通常从测试**最小对立体（minimal pair）**开始。具体的做法是，只改变一个音，看看这个词是否变成了另一个词（可自己判断，也可求助母语者）。以英语为例，我们由已知词汇“tip”开始，若把首音“t”换成“d”，这个词就会变成一个新词“dip”，如果这确实是一个新词，这时我们就可以说，在英语中“t”和“d”是有区别性的声音，不然就没法区分“tip”和“dip”了。这些有区别性的音位是区别意义的最小音段。换句话说，如果用［d］替换了［t］没有产生新的单词，那么［d］和［t］就不是两个相互有别的音位。在词的每个位置我们都可以进行最小对立体测试。“tip”中的“i”可以变成“o”，“tip”就变成了一个新词“top”。最后一个音“p”换成“n”，“tip”就变成了一个新词“tin”。有些音位只在单词的某些位置出现。例如，“long”中的浊软腭鼻音不会在英语单词起首出现，“hat”中的喉擦音从不出现在词尾。因此，一个单词中，往往不是所有的音段都可以通过上述替换而产生新的单词，所以最小对立体测试也并不是总能奏

效。不过，该测试足以帮助我们辨认某种语言中的音位。

让我们再来看“tɪp”的首音。这个音的发音有多种严式标音的标法。例如，发音时可以用舌叶接触齿龈脊（清舌叶－齿龈爆发音），也可以用舌尖接触齿龈脊（清舌尖－齿龈爆发音），还可以用舌尖或舌叶接触齿背（清齿爆发音）。虽然发出的这些“t”略有不同，但这种差异在英语中并不会起到区分两个单词的作用。也就是说，用舌叶－齿龈爆发音所发的“tɪp”，用舌尖－齿龈爆发音所发的“tɪp”，和用齿爆发音所发的“tɪp”，在英语中并不会构成相互区别的三个单词。这些不同版本的“t”叫作音位/t/的**音位变体（allophone）**。音位变体是同一个音位的不同实现形式，它们并不会导致意义上的变化。音位在词中出现的位置不同，就可能会实现为不同的音位变体。例如，英语中“l”有两个变体，其中亮“l”（light l）出现在词首位置，暗“l”（dark l）出现在词末位置（见3.1.5）。音位出现的**语境（context）**也会导致不同音位变体的出现。这里的语境是指与这个音位紧邻的前面或后面的音（见3.4.1.2中关于逆向协同发音的讨论）。若我们要研究某个语言中只跟语言相关的问题，这些音位变体之间的差异是没有必要在标音中体现出来的。但是，如果对于一种语言来说，上述所提到舌叶－齿龈爆发音和舌尖－齿龈爆发音是对立的两个音位，那么在标音时就必须将这种差异表示出来。

在音位标音中，人们重视音位上的对立，主要关注那些能够区分不同意义的音段。表示音位的符号放在两条斜线之间。例如，“tɪp”的音位标音为/tɪp/（“i”上的圆点没有了，这个变化很重要，将在稍后解释）。在宽式标音中，表示音段的符号放在方括号内，代表实际所发的音素。例如，“tɪp”的宽式标音为［tɪp］。这看起来好像只有括号的形状改变了，但事实上，说话人只是将心理词典中所存储的/tɪp/实现为了［tɪp］。而“handbag”/hændbæg/则会像前文所说的那样被实现为［hæmbæg］。音位标音和宽式标音使用的符号相同，初学者可能会将二者混淆，不过这倒是省去我们记两套符号的烦恼。但是要特别注意的是，宽式标音中的一个符号可能同时代表多种发音。例如，［t］可以代表许多种发音，正如上文中提到的舌叶－齿龈音、舌尖－齿龈音和齿音。在严式标音中，法国人说的英语“tɪp”可能要标成［t̪ɪp］，有些英国人说的则可能要标成［tʰɪp˺］——方括号表示该标音描写的为实际发音，而标音中所包含的

大量的对发音细节的描写，则反映出该标音为相当严格的严式标音。

下面两节中，我们将对英语的辅音和元音的标音进行概述。

3.1　辅音

表 3.1 是英语辅音的 IPA 符号，表 3.2 给出了对应这些辅音的例词。下面我们将详细介绍这些符号。

表 3.1　英语辅音的 IPA 符号。从左到右表示发音部位，从上到下表示发音方法。同一格中左边的符号表示清辅音，右边的符号表示浊辅音。

	双唇	唇齿	齿	齿龈	龈后	硬腭	软腭	喉
爆发音	p　b			t　d			k　g	
鼻音	m			n			ŋ	
擦音		f　v	θ　ð	s　z	ʃ　ʒ			h
塞擦音					t͡ʃ　d͡ʒ			
近音				ɹ		j	w	
边近音				l				

表 3.2　对应表 3.1 中 IPA 符号的英语单词的例子。表 3.1 中的 IPA 符号对应格中划线字母的发音。例如，“this”的首音是符号［ð］。

	双唇	唇齿	齿	齿龈	龈后	硬腭	软腭	喉
爆发音	pat　bat			tip　dip			came　game	
鼻音	mat			not			long	
擦音		fine　vine	thistle　this	sue　zoo	shoe　measure			ham
塞擦音					cheap　jeep			
近音				read		you	will	
边近音				lead				

从表 3.1 我们可以看出，有一些 IPA 符号看起来跟英语字母一样。但这并不意味着这些符号代表的音段与字母的发音相同。例如，［m］与字母“m”的发音相同（如“mill”），而［j］则与字母“y”的发音相同

(如“yoke”),[j] 与字母“j”的发音并不相同(如“joke”)。除了与英语字母相同的符号外,IPA 也有很多不属于用于书写英语的罗马字母的符号,如 [ʃ, ŋ]。

表 3.1 是按照这样的语音规则排列的:发音方法(见 2.3.2)为行,从阻塞最严重的辅音开始(爆发音和鼻音)到阻塞最小的辅音(近音)排列;发音部位(见 2.3.1)为列,从最左的唇到最右的喉。每一格中(发音部位和发音方法)清音(见 2.1 节)在左,浊音在右。下面我们将逐个介绍英语中的辅音,并会在稍后详细讨论其中的一些音。在第 4 章和第 6 章,我们将再介绍一些其他的音和相关细节,并覆盖世界上更多语言的情况。

3.1.1 爆发音

爆发音(或口塞音)使用的符号为 [p, b, t, d, k, g],掌握英语拼写的人都应该熟知这几个符号。但是正如上文提到的,这些符号并不是字母,而是从发音的角度分别代表一系列发音部位、发音方法和清浊不同的音。爆发音中的最小对立体很常见(如首音:“pie – buy”,“tie – dye”,“coast – ghost”;尾音:“rope – robe”,“sight – side”,“back – bag”);发音部位的对立也容易找到(如“by – die – guy”,“pair – (to)tear – care”,“bib – bid – big”,“ape – eight – ache”)。

3.1.2 鼻音

鼻音的符号为 [m, n, ŋ]。[m] 是双唇鼻音,[n] 是齿龈鼻音,会说(或会写)英语的人应该都知道。这两个 IPA 符号看起来和字母一样,发音也与字母一致。通过软腭鼻音的音标 [ŋ] 我们可以更清楚地明白这些符号所代表的是声音而不是英语字母。这个 IPA 符号并不属于英语拼写字母。这个符号是在“n”的右脚上加一个反向弯钩(这个钩弯的方向很重要,弯钩方向不同,其代表的 IPA 符号也不一样)。双唇鼻音和齿龈鼻音的最小对立体很常见(如“might – night”,“tumor – tuner”,“beam – bean”)。双唇鼻音、齿龈鼻音和软腭鼻音三种发音部位的最小对立体较难找,因为软腭鼻音不出现在英语的词首位置。但是,中间和词末的例子还是有的(例如,“simmer – sinner – singer”,“sum – son – sung”)。

如果你无法区分“king”和“kin”中最后的一个音（[ŋ] 和 [n]），可以在发这两个音时，把一根手指放进嘴里。发“kin”时，舌在口腔前部，发“king”时，舌很靠后，手指碰触到舌时，位于口腔内更靠后的位置。“kin”以齿龈鼻音 [n] 结尾，“king”以软腭鼻音 [ŋ] 结尾。鼻音和对应的口塞音是同一发音部位的音，舌的形状应完全相同。但是发鼻塞音时软腭下降，发口塞音时，清音的声带不振动。

3.1.3　擦音

从表 3.1 中能看到，擦音的发音部位要多一些，并且有些新的符号。唇齿清擦音 [f] 在英语单词中很常见。发这个音的浊音变体 [v] 时，发音器官的位置相同，但是声带振动。我们可以用上一章中提到的方法（见 2.1 节）来感受声带的振动。这两个音的最小对立体的例子有，起首位置：如“fan – van”；中间位置：如“surface – service”；末尾位置：如“leaf – leave”。

英语中有两个齿擦音，清齿擦音 [θ] 和浊齿擦音 [ð]。英语中齿擦音的清浊最小对立体不多（例如，“thigh – thy”，“teeth – teethe”）。但是清浊齿擦音与别的音在起首、中间、末尾位置的对立体倒是很常见（例如，“thigh – tie”，“both – boat”，“though – dough”，“worthy – wordy”，“breathe – breed”，等等）。

清齿龈擦音 [s] 和浊齿龈擦音 [z] 有很多最小对立体（“seal – zeal”，“lacy – lazy”，“ ice – eyes”）。前面提到过，发这些音时，有人用舌尖，有人用舌叶，但是两种方法的声学表现几乎没有差别，所以这种个人差异在宽式标音和大多数严式标音中都不描写。

龈后擦音有清音 [ʃ] 和浊音 [ʒ]，这两个音只有“Confucian – confusion”这一对最小对立体。[ʃ] 比较常见（例如，“sure”，“machine”，“rush”），[ʒ] 在词首或词末位置只出现在外来词中（例如，“genre”，“massage”）。[ʒ] 出现在词中位置的例子有“usual”。通常 [ʃ] 和 [ʒ] 的非 IPA 符号分别为 [š] 和 [ž]。

英语中有软腭塞音和软腭鼻音，但是没有软腭擦音，清的浊的都没有。这是语言的一种“空白”（gap）。英语曾经是有这些音的，但是后来“丢失”了，我们从一些词的拼写中还可以看出来，如“laugh”。现在这

个词发的是唇齿擦音，但是拼写上仍然是软腭音“g”后面加“h”，说明这个音曾经是软腭擦音。

还有一个清擦音，叫喉擦音［h］。这个音的发音部位在声门，发音时声道其他部分的位置比较自由，常取决于后面元音的位置，一些语言学家认为［h］是其后接元音的清音变体。

3.1.4 塞擦音

塞擦音分为清龈后塞擦音［t͡ʃ］和浊龈后塞擦音［d͡ʒ］。IPA 没有专门表示塞擦音的音标，只是用连音符（［͡］）把塞音和擦音连接起来。不为塞擦音编写专用音标的原因是：（1）这种表示方法可以很清楚地描写塞擦音，进而可以减少必须记忆的符号数量；（2）关于塞擦音是否是一个单一的音段还存在争论。有时更多是在争论塞擦音究竟是一个音段，还是塞音和擦音两个音段的组合体。例如，英语“cheap”的首音（［t͡ʃ］）和“jeep”的首音（［d͡ʒ］）被认为是塞擦音，是因为英语通常不允许在词的起首位置有塞音后接擦音的辅音丛。然而，“pants”和“hands”末尾的复数标记“-s”即［ts］和［dz］则被认为是爆发音和擦音的辅音丛，这是因为擦音并不是这两个词的一部分，而是复数标记“-s”附加在词末尾塞音上的结果。德语中［t͡s］更像是一个音（类似英语中的［t͡ʃ］）。因此，德语的［t͡s］可以出现在词的起首如“Ziel”［t͡siːl］（“目标”），也可以出现在词中如“Konzept”［kɔnˈt͡sɛpt］（“概念”）以及词尾如“Schutz”［ʃʊt͡s］（“保护”）。由于塞擦音的定义在一定程度上与这种语言的结构行为有关，有些语音学家总是将塞擦音看作是两个音段的序列，而不是一个音素或音位。当塞擦音被当作一个音的时候，它必须是同部位调音（homorganic arituclation，也即同一发音器官的发音，见2.3.2；例如［ks］不是塞擦音，因为其塞音和擦音的发音部位不同），而且必须是爆发音在前擦音在后，不能反过来（［st］不是塞擦音）。通常［t͡ʃ］和［d͡ʒ］的非 IPA 符号分别为［č］和［ǰ］。

3.1.5 近音

到目前为止，我们所讨论的音的特点都是口腔通路完全闭塞（爆发

音、鼻音和塞擦音），或者存在非常强烈的收紧（擦音）。英语以及大部分语言中的近音都是浊音，口腔内的通路打开较大。通路的正中矢状部位也可以形成阻塞，例如“leek”或“little”中的边近音［l］，但是在这种情况下，舌至少有一边是压低的，允许气流通过。

当发［l］这个音时，仔细听或者对着镜子观察自己口腔的变化，你会注意到“like”中的［l］和“ball”中的［l］的音质有所不同。若发这两个音时吸气，即用吸气气流发音，你应该能感觉到口腔里很凉，但是凉的部位不同，并且这两个音听起来也不一样。英语中，发词尾的/l/（例如，“ball”或“felt”）时，舌后部会向软腭靠近。受软腭处这一次要收紧的作用，/l/听起来具有比较灰暗（dark）的音质，所以叫作“**暗 l（dark l）**”。美国英语的很多方言并不区分词尾的“暗 l”和词首的“亮 l”，而只有“暗 l”，表示为［ɬ］（见 3.4.1.6 中关于软腭化的讨论）。

英语中另一个比较常见的近音是“r”。这个音其实有很多变体，包括齿龈音［ɹ］和**卷舌（retroflex）**变体［ɻ］。齿龈音变体经常发成“翘舌 r”（bunched r），舌尖向舌体后回收，在靠近齿龈或龈后的部位形成主要收紧。而卷舌变体会将舌尖，或舌尖下表面，或舌叶靠向上腭的龈后区域（详细介绍请参考 Laver 1994：299－302）。英语的这两个音没有最小对立体，因此它们可以看作是同一个音位的两个音位变体。这个音位常用符号/r/表示，而实际上/r/在 IPA 标音系统中代表的是齿龈颤音（见 4.2 节）。尽管这个音标描写的是另一种发音方法，但是在宽式标音中常用它来表示“r”音，这么标仅仅只是为了便于书写记录，但一定要牢记，这个音标在 IPA 里代表特定的发音方式。

英语只有一个腭音，即近音［j］。这个音可出现在词首（如“you”）和词中（如“bayou”）位置。［j］还出现在某些辅音和元音［u］之间，如“few，beauty，pure，cute”。很多英式英语者常把这个音发作一个由齿龈塞音到元音的滑音，如“due”。英式英语里，“due”和“do”发音不同，而在通用美语里，这两个单词发音相同。

音标［w］（“way”）在 2.3.2 中被描写为唇－软腭近音，在表 3.1 中却将其列为软腭近音。唇－软腭近音涉及**双重发音（double articulation）**，在这一过程中，唇（圆唇）的发音和软腭收紧（舌后部向软腭抬高）同等重要。IPA 表（表 3.1）无法将双重发音表示出来，所以［w］

有时归为软腭音，有时则归为唇音，本书将其归为软腭音。也有些书把［w］同时列在软腭音和唇音的位置，这样不太合适，因为很容易导致误解，让人以为一个符号代表了两个不同的音，这违反了 IPA 的基本原则。还有一种常见的做法就是把这个音在表之外单列出来。

3.2 元　　音

图 3.1 列出了英语元音的 IPA 符号，图 3.2 给出了包含这些元音的例词。

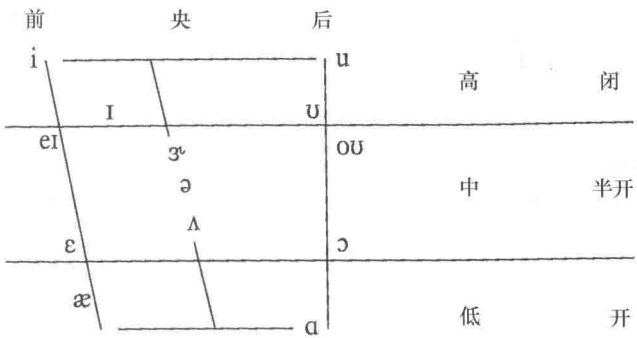


图 3.1　IPA 符号表示的元音四边形，包括美式英语中的单元音和二合元音化的元音。

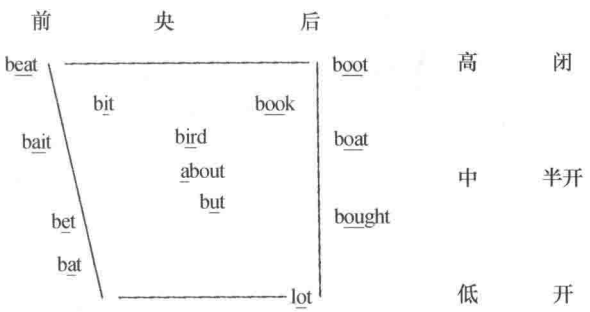


图 3.2　对应美式英语中单元音和二合元音化元音的单词举例。图 3.1 中的 IPA 符号表示对应位置词中划线字母的发音。例如，符号［æ］表示“bat”中的元音。

在2.4节中已经介绍过，元音可以从如下几个维度来分类：高低（高、中、低），前后（前、央、后）和圆唇（圆唇、不圆唇、展唇）。元音还可以根据松或紧来归类。在图3.1中，将元音画在一张图表内，大致是按照理想的舌位高低和前后来排列的。这种呈现形式叫作**元音四边形（vowel quadrilateral）**。但元音四边形究竟是一种发音表示还是一种声学表示，大家观点不一，我们将在10.1节对这个问题进行讨论。

辅音的发音部位和发音方法自己对着镜子就能看到，但与辅音不同，元音的舌位较难观察，并且发音人的个体差异也比较大。此外，不同的英语变体中，元音的音质差异很大。例如，澳大利亚英语的发音人说“today”时，听起来有点像美式英语发音人说“to die”。我们这里主要讨论通用美式英语，但是读者应清楚，他们所发的某些元音听起来可能与给出的例子稍有不同，舌和唇的位置也可能不太一样。在10.1节中，我们会介绍更客观的元音声学表示，与这里讨论的发音描写相对应。

发英语中的前高不圆唇元音［i］（如“feet”中的元音）时，舌体向前抬向硬腭。而“fit”中的元音音质则不同：元音位置稍低且更靠后，发音更短，在元音四边形中用［ɪ］的位置表示。“feet”和“fit”中的元音时长不同，也就是这两个元音的**音长（quantity）**不同。英语中，这种音长上的不同常伴随有**音质（quality）**的差异：“feet”中的元音不仅比“fit”中的元音要长，听起来也有一些不同。我们可以通过以下方法来测试：说“feet”这个词时，把元音缩短（或者使用电脑程序把元音缩短），缩短后的音并不像“fit”，而是像加快了的“feet”；如果把“fit”中的元音延长（延长但不改变舌位），听起来也不会是“feet”中类似元音“i”的声音。这说明这两个音主要是声学音质（即在元音四边形中的位置）不同，音长的差异是次要的。这也就是我们倾向于用音质来区分英语元音的原因，如图3.1所示。

除了用“长”和“短”来描述英语元音，还可以用**松（lax）**和**紧（tense）**来进行分类。元音的松紧最早用来描写声道肌肉的紧张程度。这里的松紧是相对于发中性元音（neutral schwa）时声道的状态而言的，“松”代表肌肉更松弛，“紧”代表肌肉更紧张。但是有时候两组音的区别并不在于肌肉的紧张程度，因此，我们用“松”和“紧”这一组术语来描述的是元音在发音和声学方面的不同。元音的松紧主要是音质和音

长的差异：松元音不会在元音四边形的边缘位置，并且音长比相应的紧元音短。确定一个元音是松元音还是紧元音，简单的方法是看这个音是否只能出现在闭音节中。**开音节（open syllable）**是指以元音结尾的音节（如“sea”），**闭音节（closed syllable）**则以辅音结尾（如“seat”）。紧元音可以出现在开音节和闭音节的重读音节中，而松元音只能出现在闭音节中。所以，元音［i］被归类为前高不圆唇紧元音，而元音［ɪ］被归为前高不圆唇松元音。

由于我们主要根据音质来区分［i］和［ɪ］，所以在宽式标音中不必标出时长的区别。但是，在严式标音中，时长的差异就值得注意。声音（元音或辅音）的音长在 IPA 中用**音长符号（length mark）**（像冒号的两个小三角形）表示。在严式标音中，“feet”标为［fiːt］，体现了元音的音长特点。元音［ɪ］在英语中多为短元音，“fit”标为［fɪt］。

对后元音而言，我们也能观察到类似的模式。发“suit”中的后高圆唇元音时，舌体后收抬向软腭。此外，发这个音时，人们通常会缩拢双唇（即圆唇）。［u］是紧元音。“feet”和“fit”在元音音质上可以构成前高元音的最小对立体，“suit”和“soot”是后高元音最小对立体。“suit”中的元音是紧元音，比“soot”中的元音高，位置更靠后，“soot”中的元音是松元音。以上所说的元音位置的差异在元音四边形中有所表示。“soot”中的元音用希腊字母的第二十个字母［ʊ］（upsilon）表示，是后高圆唇松元音。因为高元音（［i］和［u］）所对应的松元音（［ɪ］和［ʊ］）更接近元音四边形的系统中央，因此松元音被称为**央化（centralized）**的元音。

如果舌位比发“fit”时再低一些，就发出了“fate”，是一个前中不圆唇紧元音。若发音时没有发音动程，这个前中元音应标为［e］。但是，大部分美式英语的［e］都有一个二合元音化的元音变体（见 2.4 节）。这个元音在发音临近末尾时，舌位会稍稍滑向［ɪ］的位置，因此把这个音标成［eɪ］更为准确。若舌位再下降一点，则会得到“fed”中的元音。这个元音是前中不圆唇元音，用希腊字母的第五个字母［ɛ］（epsilon）表示。“fed”的标音为［fɛd］。

舌位若继续降低，就形成了前低元音，如“fat”中的元音。这个音虽是长元音，但它不出现在开音节中，所以也归为松元音。前低不圆唇

松元音用古英语字母［æ］（ash）表示。“fat”则标为［fæt］。

我们发现后元音具有与前元音类似的模式。我们刚才已经提到后高元音［u］和［ʊ］的关系与前高元音［i］和［ɪ］的关系类似。后高圆唇松元音［ʊ］（如“pull”中的元音）的舌位降低就会发出位置更低的元音［o］，如“pole”中的元音，这是一个后中圆唇紧元音。［o］和［e］一样，在美式英语中是典型的二合元音化元音，舌位在元音末尾会略微向［ʊ］滑动，因此标为［oʊ］。舌位若再低一点，就会发出后中圆唇松元音［ɔ］如“caulk”。若舌位再降低，就会发出“cot”中的后低不圆唇紧元音［ɑ］。虽然［ɔ］和［ɑ］在英式英语和美国东海岸英语中是相互区分的，但是，我们在2.4节也提到过，美式英语的很多变体并不区分这两个音，将它们都发成后低不圆唇紧元音［ɑ］。因此，“cot”和“caught”是**同音词（homophone）**，发音相同（［kɑt］）但拼写不同。

在图3.1中，有三个央（就前后而言）不圆唇元音。单词“bird”中的元音是中－央不圆唇**卷舌化（rhotacized）**紧元音或叫“r音色”紧元音（［ɝ］）。美式英语中卷舌化元音出现在词重音位置，如“bird”［bɝd］。卷舌用小钩［˞］表示，这个符号可以加在任何一个元音IPA音标上。例如，“bar”可以标成［ˈbɑ˞］。在英式英语的很多变体中，“r”不发音，“bird”应标成［bɜd］。然而，美式英语中没有不卷舌元音［ɜ］。中－央不圆唇松元音［ə］只在非重读位置出现，例如“aloof”［əluf］。因其位于元音四边形的中央，这个音也叫作“中性元音”（neutral vowel）。最后，［ʌ］也是中－央不圆唇松元音，如“cut”［kʌt］。这个元音与中性元音的音质相近，很多美式英语者不区分二者的发音。但有的人在重读位置发［ʌ］，在非重读位置发［ə］。

英语中有三个“真”二合元音，［ɔɪ］，［aɪ］和［aʊ］。在发这些音时，舌位的运动比二合元音化元音［eɪ］和［oʊ］明显。“boy”中的二合元音以后中元音［ɔ］开始，以前高元音［ɪ］结束，IPA符号［ɔɪ］体现了这种变化。“buy”中的二合元音以央低元音［a］开始，以前高元音［ɪ］结束。央低元音［a］在通用美语中不单独出现，位置比后低元音［ɑ］靠前。这个二合元音用IPA符号［aɪ］表示。“how”中的二合元音也是以央低元音［a］开始，并以后高元音［ʊ］结束，用IPA符号表示为［aʊ］。请注意图3.1中并没有标出上述“真”二合元音。因为在这

些音的产出过程中存在发音的变化，所以二合元音不能划归进任何一个单元的范畴中。图 3.3 试图表示出二合元音的动态特征，在元音四边形中，用箭头表示了每个二合元音的起点与终点。

总结一下，美式英语的元音包括前元音［i，ɪ，eɪ，ɛ，æ］，后元音［u，ʊ，oʊ，ɔ，ɑ］，央元音［ɝ，ə，ʌ］，以及二合元音［ɔɪ，aɪ，aʊ］。其中音质最相似的松元音和紧元音可以组成松紧元音对。英语有三对这样的音：［ɪ，i］，［ɛ，eɪ］和［ʊ，u］。其中，松元音要短一些，舌位低一些，位置也更接近中央。英语中的紧元音有［i，eɪ，ɑ，ɔ，oʊ，u，ɝ，aɪ，aʊ，ɔɪ］，松元音有［ɪ，ɛ，æ，ə，ʌ，ʊ］。与辅音相反，元音的拼写形式与实际产出的音素基本不相关（例如，“meat”，“meet”，“Pete”：［i］；“to”，“two”，“too”，“shoe”，“through”，“flew”，“hue”，“Hugh”：［u］）。

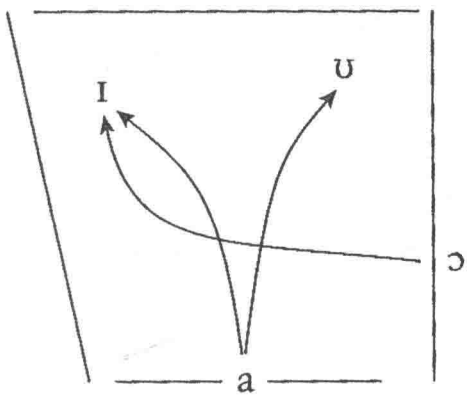


图 3.3 美式英语中二合元音的元音四边形。每个二合元音发音的起点与终点分别用箭头的尾和尖指示。

3.3 附加符号及其他符号

我们前面已经提到，在宽式标音中，每一个音都有一个与之相对应的符号。我们可以将标音的精确程度表示成一个连续统，一端是我们上文在整理英语发音时所介绍过的宽式标音，另一端就是严式标音。严式标音会尽可能精确地描写一个说话人在给定情况下说出的某句话。因此

严式标音会捕捉尽可能多的语音细节。宽式标音通常只使用一般音标符号，而严式标音还包含**附加符号（diacritic）**，以描写更多的细节。附加符号是一种额外的符号，可以加在辅音和元音的音标符号上，使原有对发音的描写更加精细。我们前面举过“tip”的例子，这个词的首音在宽式标音中标为［t］，而严式标音需要借助附加符号来区分不同的音位变体（例如，发音时是用舌尖还是舌叶，接触部位是在齿龈脊还是齿）。用舌尖发音的齿龈辅音用［t̻］表示，而用舌叶发音的用［t̺］表示。齿辅音可以用［t̪］表示，还可以再增加别的附加符号表示舌尖音或者舌叶音，如［t̻］表示舌尖音，［t̺］表示舌叶音。严式标音可以十分有效地描写发音的特异性，这种特异性可能是由于方言的差异产生的，也有可能缘于语音发展不同阶段的特点，抑或是缘于言语障碍或外国口音的发音特点。

对一句话的宽式标音可能不止一种。有时，一个词所处的语境不同则其发音也不同。例如“get”中的元音在不同语音环境中有很多音位变体，最常见的是［ɪ］和［ɛ］；例如，“You must g[ɛ]t a coffee”和“G[ɪ]t me a coffee”。用哪个音标来表示要取决于该元音音位所处的语境。为了强调某一对音位的对立，我们有时也会使用不同的标音符号。例如我们把“beat”和“bit”标成［i］和［ɪ］表明这两个音之间主要是音质的不同。若为了表明这两个音的音长也不一样，这两个元音可以标作［iː］和［ɪ］。音长符号［ː］表示［i］比较长。而若要表明时长是这两个元音的主要区别，这两个音就应标为［iː］和［i］。这几种标法都符合 IPA 的标音原则，只要保证系统中其他类似的元音对立都采用统一的标准即可。

我们只讨论了描写英语单个语音的符号。还有一些符号可以用于表示超越单个语音范围的特点，例如重音。**主重音（primary stress 或 main stress）**通过在负载重音的音节前加符号［ˈ］来表示，如“speaking”［ˈspikɪŋ］，“request”［ɹiˈkwɛst］；**次重音（secondary stress）**用［ˌ］表示，如［ˌmækəˈɹoʊni］。有时连母语者也不能确定一个词中的主重音或次重音落在哪里。而另一方面，重音落在错误的音节上听起来会很奇怪，如把“telephone”［ˈtɛləˌfoʊn］（主重音在第一个音节上）发成［ˌtɛləˈfoʊn］（主重音在最后一个音节上）。这说明说话者知道重音应该在哪个音节上，

但却又经常不能明确地意识到这一点。

3.4 通用美语的标音

上文陈述了宽式标音的基本原则。前面已经讲过，由于音位出现的语境或重音位置等因素的不同，人们在说话时会产出不同的音位变体。例如，英语的“暗 l”和“亮 l”取决于音位/l/在单词中出现的位置（见 3.1.5）。接下来我们将进一步明确音位变体出现的规则、音位的变体以及其标音符号。但是请注意，我们不可能介绍英语中所有的情况，本节只想让大家能多了解英语标音的一些规范流程和标音方法。如果想更进一步了解英语（或其他语言），请参看相关语言培训教材。

本节我们着重关注通用美语的标音，并进一步解释前文提到过的原则。我们的目的是介绍美式英语中音位变体的标音方法，包括常用的标音规范和指南。标音时采用准确明了的标音规范，可以免去文字说明的繁复，减轻标注员的负担。严式标音必须包含发音细节，而详细音位变体的标音规范可以显著提高标音的效率。接下来，我们还是来分别讨论辅音和元音。

3.4.1 辅音

首先，我们来讨论与辅音密切相关的送气、浊化和清化。

3.4.1.1 送气，浊化和清化

清爆发音在除阻后常能听到气流喷出，这种爆发音叫作**送气（aspirated）**爆发音。这种送气是除阻之后所产生的一段清音。在发送气爆发音时，除阻后声带仍然是分开的，所以有一股气流从打开的声门处冲出。在这一时段所产出的音叫作送气音（6.3 节对送气音有更详细的讨论）。要检查一个音是否送气，可以在发这些音时，把手放在嘴前。发送气音［p^h］时，应该能够感到一小股气流，而不送气音［p］则没有。单看送气部分，其发音就好比一个短［h］，正因如此，严式标音将送气表示为爆发音后加一个附加符号［^{h}］，例如“top”［$t^hɑp$］。在英语中，送气只是音位变体，只在特定语境中出现。

在下列情况中，通用美语的清爆发音要送气：

1. 在词首位置，即词的第一个音，例如“top” [t^hɑp] 和“today” [t^hə'deɪ]。
2. 在重读元音前。试比较“appeal” [ə'p^hil] 中送气的/p/和“apple” ['æpəɬ] 中不送气的/p/。“appeal” 中的“p”在重读元音前，因此是送气音。

除了以上两种情况，清爆发音大多是不送气的。例如在/s/后，清爆发音就不送气，如“stop” [stɑp]。在一些语言中，送气爆发音和不送气爆发音是不同的音位。对这些语言来说，不管是宽式标音还是严式标音都必须标出送气特征。例如，在泰语中，[tam] 的意思是“捣碎”，而 [t^ham] 的意思是“做”。

送气与否也会影响其后所跟近音的嗓音状态。送气会使声带迟后振动，接在送气爆发音后的近音会部分**清化（voiceless 或 devoiced）**。清化用符号 [̥] 表示，如“please” [pl̥iːz] 或“quick” [kw̥ɪk]。清音 [l̥] (“please”) 和浊音 [l] (“lease”) 的区别可以通过感受声带的振动来体会。具体做法是将手指放在喉部，发“lease”的/l/时能够感觉到声带的振动，发“please”的/l/时感觉不到声带的振动。

当浊爆发音和浊擦音位于词或短语末尾时会部分或完全清化。浊爆发音和浊擦音的完全清化在宽式标音和严式标音中用相应的清音表示即可，例如，将清化的/z/标为 [s]。但是近音就不能这么标了，因为近音没有对应的清音。所以在严式标音中，(完全或部分) 清化用符号 [̥] 表示。但是这样会造成一些混淆：[z̥] 表示部分清化（如果是完全清化，就会用 [s]），而 [m̥] 所表示的既可能是部分清化，也可能是完全清化。

虽然/h/是典型的清音，但是当它不在词首时，也可以浊化。浊化的/h/表示为 [ɦ]。例如词对“head-ahead”的严式标音为 [hɛd-ə'ɦɛd]。

3.4.1.2 协同发音

协同发音（coarticulation）指的是产生言语声时的发音重叠现象。也就是说，在发一串音时，一个音的发音动作与其前接或者后接音的发音动作重合。协同发音分为两种，逆向协同发音和顺向协同发音。**逆向协同发音（anticipatory coarticulation）**是指一个音的发音受到了其后接语音的影响。逆向协同发音也叫**“从右至左的”协同发音（“right-to-**

left" coarticulation)，因为发音时是后面的音段（右边）影响前面的音段（左边），是一种从右向左的方式。逆向协同发音也叫作**反向协同发音（regressive coarticulation）**，因为这是一种“反向的”对前音的影响。以软腭爆发音（[k, g]）为例，当它们位于前元音之前时，发音的位置会前移，如“keel”。换个角度解释，在发音时由于预见到后面会出现舌位靠前的前元音［i］，发［k］时舌与软腭的接触点便产生了明显的前移。在严式标音中，这种发音上的前移用符号［˖］表示，如［sk˖iː］。

再举一个齿龈辅音的例子。除了擦音以外，所有齿龈辅音都会受后面齿音的协同发音的影响。例如，齿龈爆发音［t］在齿音之前会发成齿音，如“width”，严式标音为［wɪt̪θ］，附加符号［̪］表示发音部位为齿。这就是一个逆向协同发音的例子，齿龈音受后面齿音的影响发成了齿音。元音也会受到逆向协同发音的影响，如3.4.2.2中将会讨论到的元音鼻化现象，当元音后接鼻音时，由于预见到鼻音会出现，软腭通道在发元音时就会打开，使鼻辅音前的元音发生鼻化。

顺向协同发音（perseverative coarticulation）是指前一个音的影响效应延续到后面音的情况。这种协同发音的产生是由于发音器官的惯性作用，也称作**顺承性协同发音（carry-over coarticulation）**或**“从左至右的”协同发音（“left-to-right” coarticulation）**，作用时由先出现的音扩散至后出现的音，为一个音对其后接语音产生的影响，这种协同发音也被叫作**正向协同发音（progressive coarticulation）**。上一节提到的近音在清送气爆发音后被清化的现象就是一个顺向协同发音的例子：起首爆发音的清音段影响了其后的近音，如“please”［pl̥iːz］。

3.4.1.3 辅音除阻

当爆发音位于词末或短语末尾时，人们的发音（也可以不发音）有多种变体。在2.3节中我们谈到，爆发音的特点是除阻后有一个短暂爆破。我们讨论这一特点时，主要指位于音节起始或词首位置的爆发音。但词末的爆发音可以突然除阻（像词首爆发音一样）；也可以逐渐除阻，这时我们可能听不到除阻；也可以根本不爆破。例如，词末浊爆发音可以实现为浊音，可以（部分）清化，也可以实现为无闻除阻（即，成阻之后，单词在“结束”时不再打开声道），如“rib”的发音可能是

[ɹb]、[ɹb̚] 或 [ɹb˺]。[ɹb˺] 中的附加符号 [˺] 表示无闻除阻。词末清爆发音的除阻可以实现为清音，送气音或无闻除阻，如“up”可以实现为 [ʌp]、[ʌpʰ] 或 [ʌp˺]。浊爆发音和清爆发音在词末处于近音后时经常不除阻，如“bulb”[bʌɬb˺] 或“harp”[hɑɹp˺]。

当一个爆发音处于词尾**辅音丛（consonant cluster**，即词尾的一串辅音）的末尾时，若其发音部位与前面的辅音相同则爆发音多数情况下会除阻，如“bump”[bʌ̃mp] 或 [bʌ̃mpʰ]（附加符号 [˜] 表示鼻化，见 3.4.2.2）。如果不除阻，就很难区分 [bʌ̃mp˺] 和 [bʌ̃m]。类似地，若末尾两个辅音发音方法相同，例如都是爆发音，那么第二个辅音多数情况下会除阻，如“picked”[pʰɪkt]，为与“pick”[pʰɪk] 相区别。但是在这样的词末爆发音组合中，第一个爆发音通常不除阻，例如“apt”中的双唇爆发音 [p] 不除阻，即使除阻了，这个音也听不到：当 [p] 的双唇闭合以后，处于自由状态的舌移向其后 [t] 的闭塞位置；当双唇分开想为 [p] 除阻时，气流却被 [t] 的闭合阻碍了，[t] 的发音掩蔽了 [p] 的除阻，所以双唇爆发音 [p] 的除阻为无闻除阻，标记为 [æp˺t]。从声学上看，元音 [æ] 到 [p] 的过渡信息可以显示出 [æ] 后面跟的是双唇音，但是除阻则由 [t] 完成。正因如此，[æp˺t] 听起来和“at”非常不一样，“at”发成 [æt] 或 [æt˺]，没有任何双唇音的影响。

3.4.1.4　闪音和拍音

在齿龈爆发音中能观察到的另一种变化是闪音。在通用美语中，**闪音（flap，或拍音 tap）**很常用，这一点与英式英语差异显著。闪音是由舌尖向齿龈脊弹一下而发出的，很短并且是浊音。图 3.4 罗列了音位/t/的三种发音。图 3.4a 表示的是“attack”中的/t/，是送气清爆发音 [tʰ]：从图上可以看到一段很长的无声闭合段（约 90 ms），后面紧随着除阻爆破段以及较长的送气段（约 60 ms）。图 3.4b 则实现为浊爆发音（[d]）：浊闭合段（约 50 ms）后紧跟一个（很弱的）除阻爆破，没有送气段。图 3.4c 实现为闪音（[ɾ]）：波形图显示一段很短的浊闭合段（约 35 ms），没有明显的爆破段也没有送气段。当齿龈爆发音在元音或央近音后并位于非重读元音前时通常实现为闪音，标为 [ɾ]。例如在通用美语中，“latter”和“ladder”都标成 [ˈlæɾɚ]。值得注意的是，当齿龈爆发音位于重读音节前时，人们不会将其实现为闪音：比较一下没有闪音的“pho-

netician”［ˌfoʊnəˈtʰɪʃə̃n］和有闪音的“phonetics”［fənˈɛrɪks］的区别。

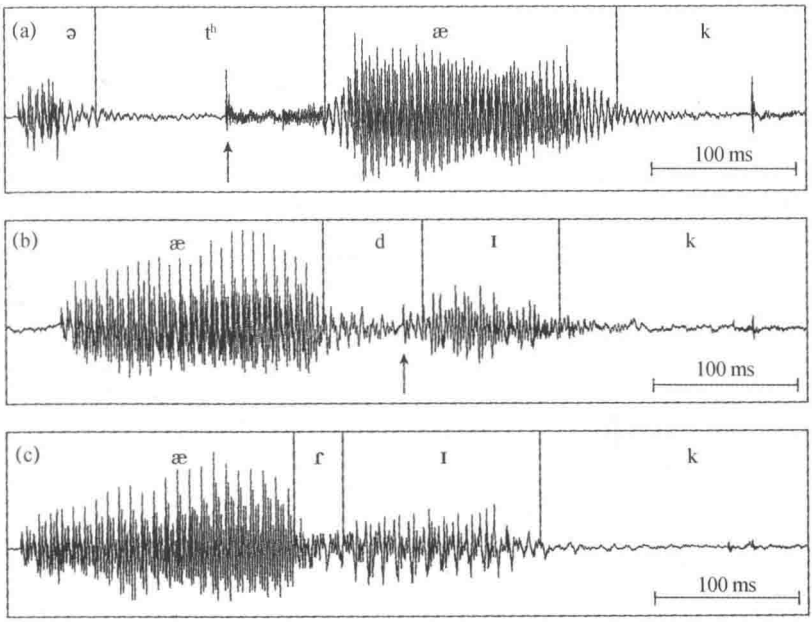


图 3.4 齿龈爆发音的各种变体的波形图。（a）“attack”，送气清爆发音［əˈtʰæk］；（b）“attic”，浊爆发音［ˈædɪk］；（c）“attic”，含有闪音［ˈæɾɪk］。音段以竖线表示，（a）和（b）中的箭头表示单词中部爆发音的除阻。

3.4.1.5 喉爆发音

我们在 3.4.1.4 中讲到，齿龈爆发音在某些语境中会发成闪音。此外，清齿龈音/t/在通用美语和英式英语中还经常会发成喉爆发音（也称作“喉塞音”）。在英语中，喉爆发音［ʔ］是/t/的音位变体，在发音时，声带会短暂而紧紧地闭合。在某些英式英语的变体中，一些单词中的/t/会被喉爆发音取代，如“butter”［ˈbʌʔə］。当两个相邻的词，第一个词以元音结尾，下一个词以元音开始时，喉爆发音也经常用来起到分隔单词的作用，比如“the only”［ðəˈʔõʊ̃nli］，“she eats”［ʃiˈʔits］。另外，不少人还会在词末清爆发音前插入一个喉爆发音，如“cup”［kʰʌʔp］或“pack”［pʰæʔk］。也就是说，他们在元音结束时会闭合声带，而不是保持声带打开来发后面的爆发音。最后，［ʔ］也可以出现在“mountain”

[ˈmɑ̃ʊ̃ʔn̩]，“sentence”[ˈsɛ̃ʔn̩s]，“button”[ˈbʌʔn̩]（见3.4.1.7中对[n̩]和其他音节性辅音的讨论）或“camps”[kæ̃ʔps]等词中。

3.4.1.6　软腭化

在3.1.5中，我们简要提到过齿龈边近音/l/有两个音位变体，“亮l”和“暗l”。“亮l”发音时舌尖顶住齿龈脊，舌体较低。一般来说，/l/在前元音之前是明亮清晰的“亮l”，如“leaf”[lif]。“暗l”的舌尖置于下门齿附近，舌体后部向软腭抬起，所以这个音位变体也叫作“软腭化的l”。一般来说，元音后面的/l/是“暗l”。人们在音标符号中间添加附加符号[~]，以表示软腭化：[ɫ]。

3.4.1.7　音节性辅音

当中性元音[ə]位于词末鼻音（[m, n, ŋ]）或流音（[l, ɹ]）之前时，[ə]可能会被省略，其后的辅音会表现出音节的特点（如“able”，[bl]是单词的第二个音节）。这种词末辅音叫作**音节性辅音（syllabic consonant）**，在辅音音标下加一条小竖线[̩]表示。它们承担了音节的全部时长。当相邻的辅音发音部位相同时，这一现象最为明显。例如，“sudden”[ˈsʌdn̩]，发齿龈爆发音[d]时，舌尖与齿龈形成闭合，之后软腭下降并发出[n]。因为[d]和[n]的发音部位相同，当[d]除阻后，舌尖不必移动位置即实现为鼻音[n]，这种现象叫作**鼻爆破（nasal plosion）**。**边爆破（lateral plosion）**的发音过程与之类似，先形成齿龈爆发音的闭塞，之后舌两侧下降，发出边音，如“channel”[ˈt͡ʃæ̃nɬ̩]。由于在发这两个辅音时发音部位一直维持收紧，中间无元音插入，词末的辅音就表现出了元音的特点。值得注意的是，当两个辅音的发音部位不同时，中性元音[ə]也可能被省略，例如，“motion”[ˈmoʊʃn̩]和“uncle”[ˈʌ̃ŋkɬ̩]。

3.4.1.8　增音

在3.4.1.3中，我们描述了除阻后音素可能部分无闻的情况。与之相反的一种情况是音素的**增音（intrusion）**。在某些语言中，增音是为了破坏辅音丛（如在荷兰语中，“melk”[mɛlk]“牛奶”经常发成[ˈmɛləkˌ]），增音也可能由协同发音所致。在英语中，当鼻辅音出现在清擦音前时会发生增音现象，如“something”。此时，会在鼻音和清擦音之间插入与鼻音

发音部位相同的清爆发音。具体过程如下。为了发出鼻音，口腔通路必须完全闭合（类似口塞音），并且软腭要下降。接下来发擦音，软腭抬高，同时阻塞以爆发的形式解除，形成擦音湍流。如果软腭先闭合，而口腔内的除阻延迟片刻，就会产生一个与口塞音相同的发音动作：口腔中形成阻塞，软腭抬升，然后除阻。这样，就插入了一个口塞音。例如，“chance”[t͡ʃæ̃ns] 可以读作 [t͡ʃæ̃nts]，“length”[lɛ̃ŋθ] 可读作 [lɛ̃ŋkθ]，“something”[ˈsʌ̃mθɪ̃ŋ] 可读作 [ˈsʌ̃mpθɪ̃ŋ]。图 3.5 所示的波形图展示了没有爆发音增音和有爆发音增音的对比。当增加了清爆发音 [t, k, p] 后，波形图上有一个明显的清闭合段，并且紧接一个明显的除阻爆破。

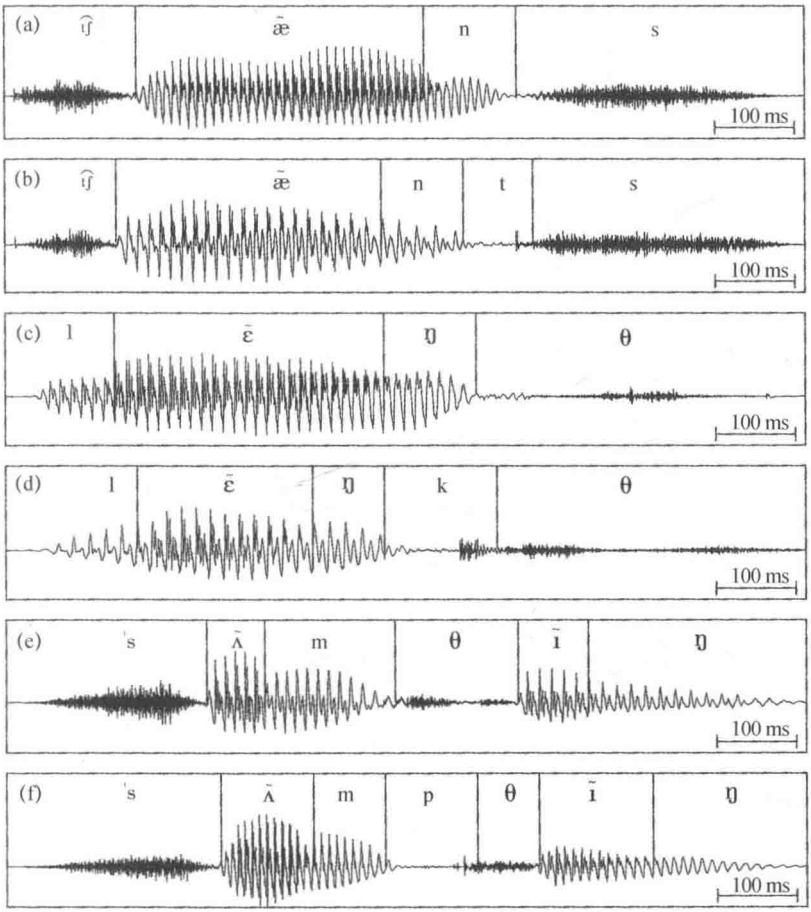


图 3.5 单词“chance”，“length”和“something”的波形图。图（a），（c），（e）中没有增加爆发音，（b），（d），（f）中增加了爆发音。

3.4.1.9　时长

当一个词的词尾与其后接单词词首为同一个辅音时，这两个辅音通常会合并为一个稍长的音：试比较“summer”[ˈsʌ̃mɚ]和“some more”[sʌ̃ˈmːɔɹ]，“top air”[tʰɑpˈɚ]和“top pair”[tʰɑˈpːɚ]。这种情况在单词内部也可以发生，试比较“meanest”[ˈmĩnəst]和“meanness”[ˈmĩnːəs]。图3.6给出了这三对词的波形图。图3.6b，3.6d，和3.6f中的辅音明显比图3.6a，3.6c和3.6e中的长。这种情况在标音时用音长符号[ː]表示，这个符号在3.2节讨论元音音长时已经介绍过。

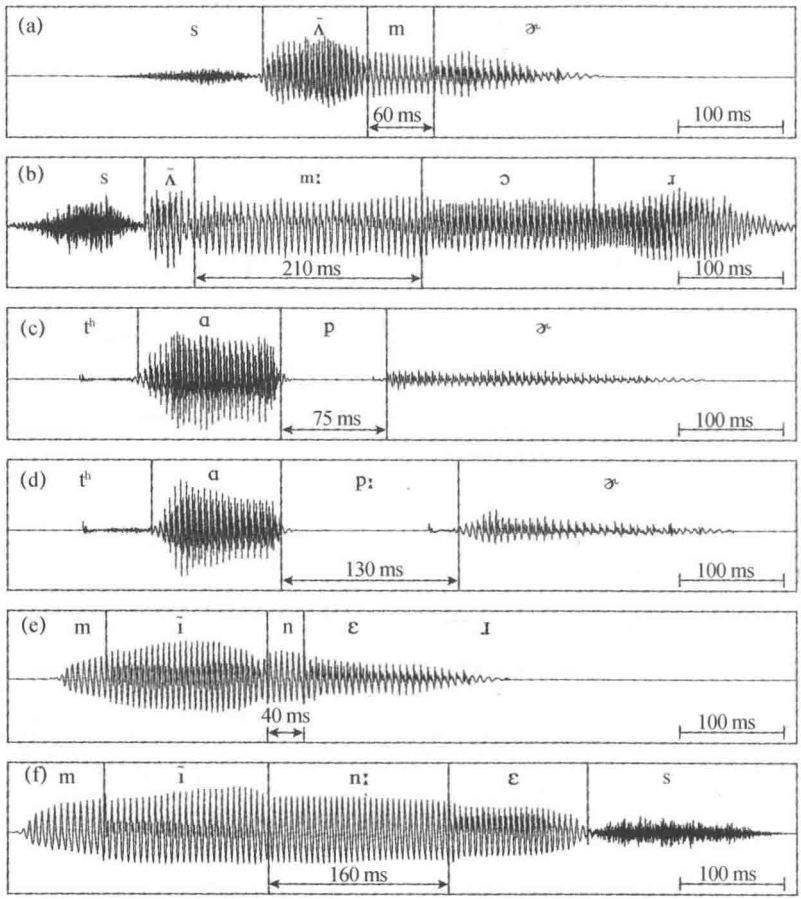


图3.6　(a)“summer”，(b)“some more”，(c)“top air”，(d)“top pair”，(e)“meaner”和(f)“meanness”的波形图。(a)，(b)，(e)和(f)中鼻音的时长需要通过频谱图来确定，这里并未给出。

3.4.2 元音

我们在3.2节已经讨论过，英语的不同方言中存在大量的元音变体。影响元音音质的因素本身就很值得研究（例如，Hughes 和 Trudgill，1996；Wolfram 和 Schilling-Estes，1998）。我们将在9.5.1中对元音的音质进行发音和声学上的描写，并将在10.1节中讨论其他语言的元音。在这里，我们只讨论元音标音中的两个关键问题：元音的时长以及后接辅音对其音质的影响。

3.4.2.1 时长

我们在3.2节中说过，元音有长有短。音长符号［ː］可以用来表示长元音和长辅音。另外，元音出现在重读音节中可以被延长，出现在非重读音节中则会缩短，试比较元音［aɪ］在非重读音节（如“insight”［ˈɪnˌsaɪt］）中和在重读音节（“incite”［ˌɪnˈsaɪt］）中的时长（见图3.7），［aɪ］在后一个词中明显更长。这种延长可以用音长符号［ː］表示，若延长后没有“真正的”长元音那么长，可用半音长符号［ˑ］表示。我们在3.2节中讨论过，音长不同大多伴随着音质的差异。因此，不但重读音节中的元音要比非重音音节中的元音长，长短元音的音质也可能会有所不同。

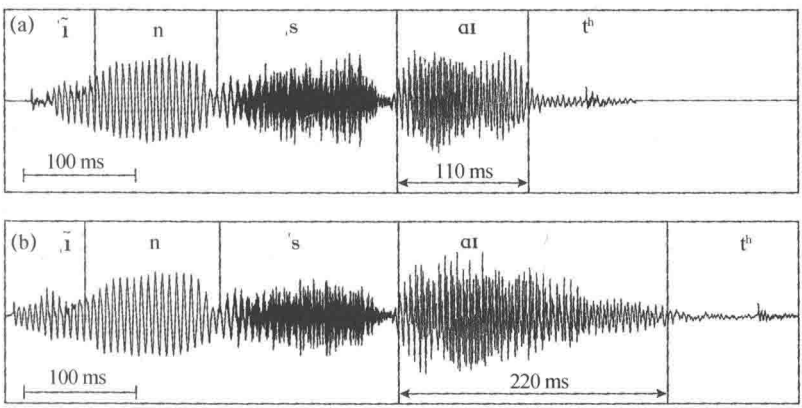


图3.7 （a）“insight”和（b）“incite”的波形图

在3.2节中讨论松元音和紧元音时，我们已经介绍过开音节和闭音节的区别。通过观察我们会发现，开音节中的元音明显比闭音节中的元音

长，这种现象叫作**开音节延长（open syllable lengthening）**。例如，“bee”中的元音比“beat”中的元音长，给人的感觉就像“beat”中的元音被后面的辅音缩短或截短了一样。这种元音的延长现象同样可以用音长符号[ː]表示。

元音在浊辅音之前通常要比在清辅音前长一些，如“bat”和“bad”中的元音（见图3.8）。就人们已经调查过的语言来说，元音在浊辅音前延长的情况非常普遍，而在英语中，清浊辅音前的元音时长差异则要更加明显。事实上，辅音的清浊可以通过其前接元音的音长来推测，尤其是当结尾辅音不除阻的时候（有些人所发的词末爆发音有无闻除阻的现象）。在听不到词末辅音发音的情况下，我们仅凭元音的音长就可以区分“bat”和“bad”。

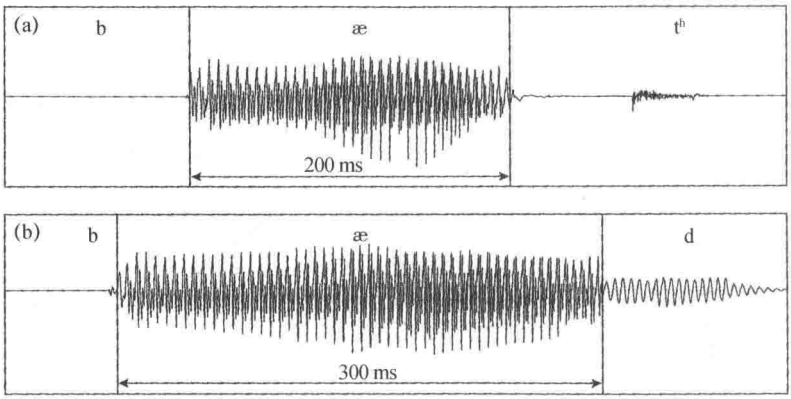


图3.8　（a）“bat”和（b）“bad”的波形图

正如我们前面所讲的，在严式标音中，所有音长上的差异都可以用音长符号[ː]标记，如果只是“半长”，则用符号[ˑ]表示。

3.4.2.2　后接辅音的影响

元音除了在浊辅音前会变长，在鼻辅音前还会发生鼻化，这在3.4.1.2中已经讨论过了，属于逆向协同发音。例如“ban”中的元音，由于预见到其后接有鼻辅音，软腭在发元音时就开始下降了。开始发元音的时候软腭抬高，元音前半部是口音，然而，后半部分受到后接鼻音

的影响，软腭下降，发生鼻化，而此时口腔仍保持打开。在严式标音中，在元音上方添加［˜］表示鼻化，试比较“ban”［bæ̃n］中的鼻化元音和“bad”［bæd］中的口元音。有些语言中确实有真正的鼻化元音，也就是说，在这种语言中存在鼻化与非鼻化元音的最小对立体。例如，在孟加拉语中［tɑt］的意思是“温暖”，而［tɑ̃t］的意思是“织布机”。在这种情况下，这种音位叫作**鼻元音（nasal vowel）**而不是**鼻化元音（nasalized vowel）**。

当元音出现在/ɹ/和/l/之前时，元音的性质会受到很大的影响，产生一些因舌体后收而形成的音位变体。当元音出现在音节末尾的/ɹ/前时，元音间的对立会明显减少，因为［i］和［ɪ］、［u］和［ʊ］、［eɪ］和［ɛ］之间的差异在/ɹ/之前基本中和了。因此，单词“bear”可以被标成［bɛɹ］或者［beɪɹ］。/ɹ/对前接元音音质的影响被称为**r音色（r-coloring）**或**卷舌音化（rhotacization）**。当二合元音化元音［eɪ］或者［oʊ］出现在/l/前时，它们可能会失去后滑，变成单元音，例如单词“bail”［beɬ］和“bowl”［boɬ］。

美式英语的标音我们就简单介绍到这里。上面概括的规则应该可以描述大多美式英语者的发音了。但是这些规则并未穷尽所有可能，只是让大家了解到通用美语和其他一些英语变体中的各种现象。掌握标音规则可以帮助大家理解一些产生言语声的方法。同时还要注意，虽然标音的方法有许多种，但标音的一致性非常重要。一旦采用了某一种标音规则，标相同的音时就要沿用这套规则，因为不同的标音符号代表不同的发音。

练　习

1. 严式标音和宽式标音的区别是什么？宽式标音和音位标音的区别是什么？请分别说明各种标音方法的应用场景。
2. 解释音素、音位及音位变体之间的区别。
3. 在通用美语中，紧元音和松元音在哪三个方面有所不同？
4. 送气音是怎样实现的？为什么英语中送气音是音位变体？用本书以外的例子来论证你的观点。

5. 逆向协同发音和顺向协同发音的区别是什么？请举例说明。
6. 写出通用美语中，下列音标表示的单词：

   [əˈnʌf]，[ˈd͡ʒʌŋgl̩]，[ˈkwaɪət]，[ˈlɪsn̩]，[ˈkɹi t͡ʃɚ]，[ˈd͡ʒʌstəˌfaɪ]，

   [əˈbaʊt]，[ˈfi t͡ʃɚz]，[ˈfeɪljɚ]，[ˈfoʊɾoʊ]。
7. 给出下列单词的宽式标音：

   “judge”，“literature”，“language”，“phonation”，“whistle”，“youth”，“television”，“peaceful”，“fountain”，“rather”。

注　释

1. 对不同标音方法的详细讨论，请见 Laver（1994：n. 18）和 *Handbook of the International Phonetic Association*（1999，pp. 28 – 30）。

4 辅音和元音的发音部位与发音方法

在第3章中，我们用清浊、发音部位和发音方法描述了英语的辅音，用前后、高低和圆唇描述了英语的元音。人们用不同的维度对辅音和元音进行描写，似乎表明辅音和元音在发音上基本没有共同点。但是，对辅音和元音的发音描写主要是基于收紧的位置（辅音的“发音部位”、元音的“前后”）和收紧的程度（辅音的“发音方法”、元音的“高低”）的。因此，我们认为辅音和元音都可以采用发音部位与发音方法来描述。

英语的言语声是世界所有有记载的言语声的一个子集。据估算，目前世界上有6000－7000种语言。然而，在这些语言中，许多语言都只有少数人在使用，处于濒临消失的危险边缘。但是，所有的语言都是同等珍贵的。某种语言可能只有少数人在使用，但这并不意味着这种语言是“奇怪”或者“反常”的——这只是历史、政治或经济发展的结果。最近的调查显示，有超过一半的目前仍在使用的语言将在一百年内消失。幸运的是，目前已有记载的语言可以帮助我们了解世界各地语音的丰富特点。本书最后一页的国际音标（IPA）表中的IPA符号可以给世界上很多语言标音。显然，在世界语言中有很多非英语的语音，这些语音我们在前一章介绍英语的辅音和元音时并没有提到过。必须要说明的是，IPA最初是由西欧的科学家设计的，因此无论是其中用单个符号来表示的那些音，还是以由左及右的书写顺序表示一个发音序列的方式，反映的都是当时那些西欧科学家所用语言的原貌。因此，目前的IPA符号是一系列符号任意混合的结果，其中包括罗马字母和希腊字母，或以这些字母修改（如倒置）所得的符号，以及一些从其他书写或符号系统借用而来的符号。这种选择符号的方式有一个缺陷：缺少对发音部位和发音方法的组织原则的内在表示。例如，所有用于表示擦音的符号并没有共用某个

特定的图形特征，使其能与其他符号类别区分开来；同样地，所有双唇音的符号也没有任何共同点。尽管如此，目前的这套 IPA 符号已经能够用来表示世界语言中的大部分语音了。

在本章中，我们将简要概述世界语言中许多不同类型的辅音和元音，并介绍相应的 IPA 标音符号。在收集例子的过程中，我们发现了两个非常珍贵的资源：Laver（1994）以及 Ladefoged 与 Maddieson（1996）发表的资料。有关这些语音的音频例子，请访问本书列出的网站和链接。另外，网站“sowl”（sounds of the world's language，http://www.phonetics.ucla.edu/vowels/contents.html）[1]也是非常有用的资源。

4.1　辅音

表4.1 列出的辅音 IPA 符号可以用来标记世界语言中的许多辅音。由于大部分语言与英语使用同样的发音部位，所以向现有的表格中增加符号不会引起太大的问题。不过，许多语言也有其他的发音方法。例如，英语只有齿擦音没有齿爆发音，而其他语言则可能有齿爆发音，例如法语。一些音在英语中只是音位变体，但在其他语言中则可能是不同的音位。而且，人们也发现了存在齿爆发音和齿龈爆发音对立的语言。

表4.1 的辅音表中有三种单元格：有语音符号的单元格、没有符号的单元格和有阴影的单元格。没有符号的单元格表示这种音可能存在，但是目前在世界语言中还没有发现。例如，发音部位“唇齿”列和发音方法“颤音”行交叉的空白单元格表示还没有语言记载有唇齿颤音。不过，由于没有解剖学或空气动力学因素可以解释人们为什么不能发出这个音，并且随着不断有新的语言被记载，人们最终可能会发现这个音确实是存在的（那时，就必须增加一个新的 IPA 符号）。与此相反，阴影单元格则表示这种发音被认为是不可能实现的。作为练习，读者可以尝试去思考为什么这些阴影单元格中的发音是不可实现的。例如，请根据辅音表判断，为什么没有语言会存在咽鼻音？

我们现在将讨论在介绍英语时没有提到的辅音，按照表中从左到右也就是发音部位由唇到喉的顺序进行。目前，我们的讨论限定在前面所介绍的英语中的发音方法，也就是爆发音、鼻音、擦音、塞擦音和近音。

在本章末尾，我们再介绍其他的发音方法（颤音、拍音和边擦音）。跟前一章一样，这里讨论的音都是由肺部产生的外呼气流（即肺部气流）产生的。而对于使用其他的气流源或者其他方向气流的发音，我们会在 6.1 节中进行讨论。

表 4.1 基于 IPA1993 所列的肺辅音的 IPA 符号，2005 年经过修订。这里将“会厌”放在发音部位列，“塞擦音”放在发音方法行，把唇－软腭近音放在软腭部位下。

	双唇	唇齿	齿	齿龈	龈后	卷舌	硬腭	软腭	小舌	咽	会厌	喉
爆发音	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ	ʔ
鼻音	m	ɱ		n		ɳ	ɲ	ŋ	ɴ			
颤音	ʙ			r					ʀ			
拍音/闪音		ⱱ		ɾ		ɽ						
擦音	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ	h ɦ
边擦音				ɬ ɮ								
塞擦音		p͡f		t͡s d͡z	t͡ʃ d͡ʒ			k͡x				
近音		ʋ		ɹ		ɻ	j	w ɰ				
边近音				l		ɭ	ʎ	ʟ				

4.1.1 唇音

表 4.1 的第一列表示双唇音。英语中有双唇爆发音［p，b］和鼻音［m］，但是没有双唇擦音。双唇擦音在世界语言中实际上是很常见的。很多班图语（Bantu）中有双唇擦音。例如，北索托语（Northern Sotho），一种在南非使用的尼日尔－刚果语（Niger-Congo），区分清双唇擦音［ɸ］和浊双唇擦音［β］（如［ˈɸeta］“通过”、［ˈβeta］“掐住”）。

西班牙语中存在浊音［β］，它是音位/b/处在元音之间时的音位变体。例如，短语“Barcelona and Valencia”发作“［b］arcelona y［β］alencia”，短语“Valencia and Barcelona”发作“［b］alencia y［β］arcelona”（表示“和”的“y”的发音是［i］）。我们可以看到，处在短语的起首时，两个城市名的首音都发作［b］，但是处在浊音后时，就发作浊双唇擦音［β］。也就是说，西班牙语的说话人根据语境，即根据其前后的音位区分［b］和［β］。

自双唇稍向后移，就是唇齿发音部位。唇齿爆发音还没有被发现，可能是因为上齿和下唇之间比较难以建立稳定的完全闭塞。唇齿鼻音很

少见。特科库库雅语（Teke-Kukuya，一种在刚果使用的尼日尔－刚果语）是目前有记载的唯一一种有唇齿鼻音音位［ɱ］的语言，例如［ɱfɨ̂］“眼睛”和［mfɨ̂］“尿”（［ˊ］和［ˋ］分别表示高调和低调，见11.4.1）。唇齿鼻音在英语中作为双唇音［m］或齿龈音［n］在［f, v］前的音位变体出现，如单词“symphony”［ˈsĩɱfə̃ni］和“envelope”［ˈɛ̃ɱvəloʊp］。实际上，大部分唇齿鼻音是与后面的唇齿擦音协同发音的结果。

英语中有唇齿擦音［f, v］。很多语言要么有双唇擦音，要么有唇齿擦音，只有少数语言存在这两种音的对立。一个例子是埃维语（Ewe，一种在加纳使用的尼日尔－刚果语）有双唇擦音（［ɸ, β］）和唇齿擦音（［f, v］）的对立，见表4.2。

表4.2　埃维语中双唇擦音和唇齿擦音的例子

双唇音	注释	唇齿音	注释
［ɸu］	骨头	［fu］	羽毛
［βu］	船	［vu］	撕破

在世界语言中，清唇齿塞擦音［p͡f］比较少见，发音时双唇塞音释放后紧接唇齿擦音。这个音只在德语和比贝语（Beembe，一种在刚果使用的尼日尔－刚果语）中被证实存在。清唇齿塞擦音与唇齿擦音［f］非常相似，但是德语中存在最小对立体，如“fad”［faːt］“无味”和“Pfad”［p͡faːt］“小路”。荷兰语中存在浊唇齿近音［ʋ］，对一些荷兰语发音人来说，［ʋ］和唇齿擦音［v］对立，如“wol”［ʋɔl］“羊毛”和“vol”［vɔl］“满的”。通常，多数语言都含有双唇爆发音、双唇鼻音和唇齿擦音。

4.1.2　舌冠音

表4.1中的下一列标为齿/齿龈/龈后。这三个标签在一起形成了一栏，这是因为它们之间的对立非常少见，齿音和齿龈音之间的对立尤其

少见。例如，大部分语言中有齿爆发音或齿龈爆发音或龈后爆发音，但不会三种都有。当不存在对立时，发音部位（齿、齿龈、龈后）和舌位（舌尖或舌叶）的变化范围相当大。龈后是一个宽泛的概念，表示任何落在齿龈脊到硬腭之间区域的主要发音。从跨语言的角度来说，齿爆发音比齿龈爆发音更常见。虽然英语和德语有齿龈爆发音，但是几乎其他的印欧语言都有齿爆发音。然而，在标这些语言的齿爆发音时，通常不用附加符号 [̪] 来表示齿音，而是直接用齿龈音的符号来描写，并按照惯例将这些音发为齿音。例如，法语“tous”（“所有的”）中的齿爆发音标作 [tu]，而不使用齿音的附加符号。

英语有齿擦音 [θ, ð]。在英语中，齿爆发音以音位变体的形式出现在齿擦音之前，例如“eighth” [eɪt̪θ]。齿龈爆发音 [t, d]、齿龈鼻音 [n] 和齿龈擦音 [s, z] 在英语中很常见。在德语中有清齿龈塞擦音 [t͡s]（如“Zahl” [t͡sa:l]“数字”）。保加利亚语既有清塞擦音，又有浊塞擦音（如 [t͡sar]“沙皇”，[d͡zar]“柏油”）。英语有齿龈近音 [ɹ] 和齿龈边近音 [l]，在 3.1.3 和 3.1.4 中，我们还介绍过英语中的龈后擦音 [ʃ, ʒ] 和龈后塞擦音 [t͡ʃ, d͡ʒ]。

顺着声道再向后移，下一个发音部位是卷舌。在发卷舌音时，舌尖向后卷向硬腭的前部。很多南亚语言都存在卷舌（retroflex）辅音。例如，孟加拉语（Bangali，一种在孟加拉国和印度使用的印欧语）中有对立词，如 [ʈan]“曲调”、[t̪an]“拉”、[ɖan]“权力”和 [d̪an]“慈善”。在 IPA 的符号中，卷舌辅音的符号与它们对应的非卷舌辅音的符号非常相似，不过卷舌辅音的符号都有向下延伸的笔画并且有一个向右的弯钩，例如齿龈音 [t] 和对应的卷舌音 [ʈ]。（在 IPA 的早期版本中，卷舌是通过在音标下的一个小圆点 [̣] 来表示的，例如 [ṭ]，而现在的 IPA 用这个附加符号表示“耳语声”，见 6.2 节。）

马拉雅拉姆语（Malayalam）是有卷舌鼻音 [ɳ] 的一种语言，如 [ˈpaɳam]“钱”。马拉雅拉姆语的爆发音及鼻音在以上所讨论的三个发音部位都存在对立，即齿音、齿龈音和卷舌音，见表 4.3。

表 4.3 马拉雅拉姆语中，爆发音和鼻音存在三个发音部位对立的音位。

齿音	注释	齿龈音	注释	卷舌音	注释
[kuṯːi]	刺伤	[kutːi]	钉	[kuʈːi]	儿童
[paɲːi]	猪	[kanːi]	处女	[kaɳːi]	连成串

如果一个语言不存在这些音位的对立，那么对这种语言的母语者来说，这些音可能只是相同音位的不同音位变体，听起来非常相似（见 13.4 节）。当然，如果说话人从小所接触的语言环境中存在这种音位对立，那么他们会认为这些音是明显不同的，有着本质的区别。

波兰语中有卷舌擦音［ʂ, ʐ］，如［kaʂa］“荞麦”和［kaʐa］“薄纱”。汉语普通话中有清音［ʂ］，如［ʂɑ̃ŋhaj］“上海”[2]。

我们在 3.1.5 中曾提到英语中的“r”音经常发作卷舌辅音。IPA 中“卷舌 r”的符号是［ɻ］，发音时，舌尖向后卷起，位于齿龈脊后方，但不与其产生接触。在英语中，卷舌爆发音有可能作为音位变体出现在卷舌音［ɻ］前。例如，试着在说“tea”［ti］和“tree”［tɻi］这两个单词时，只做准备动作，把发音器官保持在起首［t］的位置，比较发这两个［t］时舌的位置，你便会注意到这两个单词之间的区别。确切地说，“tree”中的［t］由于预见到随后有［ɻ］，经常会发成卷舌音［ʈ］。如果你的发音也是这样的，那么你应该能注意到舌尖会向后有一定程度的卷起。这个发音与孟加拉语这样存在卷舌爆发音的语言中的卷舌音［ʈ］十分相似。

在德拉威语（Dravidian）中，存在卷舌边音［ɭ］。例如印度南部的泰米尔语（Tamil）中，既有齿龈边音又有卷舌边音，如［baːlaː］“一个人名”和［meːɭaː］“节日”。

除了腭近音［j］以外，腭（palatal）辅音并不常见。捷克语有腭爆发音，并且清腭爆发音［c］和浊腭爆发音［ɟ］构成对立，如“tělo”［ˈcɛlo］“身体”和“dělo”［ˈɟɛlo］“枪”。西班牙语及法语中有腭鼻音，例如西班牙语“año”［ˈaɲo］“年”、法语“agneau”［aˈɲo］“羊肉”。腭擦音中，清擦音［ç］和浊擦音［ʝ］都很罕见。德语的音位系统中有清腭擦音，例如“ich”［ɪç］“我”。希腊语有浊腭擦音［ʝ］，例如［ˈʝɛri］

“老人”，对立的是［ˈçɛɾi］“手”。在包括英语在内的很多语言中，腭擦音经常作为腭近音的音位变体出现。在英语的许多方言中，单词“you”和“Hugh”的首音不同。说话人会把“you”［ju］发为浊腭近音以与清腭擦音“Hugh”［çu］相区别。意大利语有腭边近音［ʎ］，例如“figlio”［ˈfiʎo］“儿子”。

4.1.3 舌面音

软腭音（velar）可以通过使舌靠近或者接触软腭产生。因为存在协同发音，所以清软腭爆发音［k］的发音范围可以相当大，像英语“key”［ki］的发音，［k］的发音部位会比它在“coup”［ku］中靠前 1 cm。除了［k］以外，英语中还有浊软腭爆发音［g］。

软腭鼻音［ŋ］在音节尾位置比在音节首位置更为常见，但是在柬埔寨、老挝、泰国以及越南语中，软腭鼻音在音节首尾都会出现，例如泰语中的［ŋāːn］“工作”（［ˉ］表示中间调，见 11.4.1 的介绍）。波斯语是一种既有清软腭擦音［x］，又有浊软腭擦音［ɣ］的语言，例如［xæ̃m］“弯曲”和［ɣæ̃m］“悲伤的”。在像德语及希腊语这样既有硬腭擦音又有软腭擦音的语言中，发音部位通常依元音而定。例如，德语中的清软腭擦音［x］出现在后元音之后，而清硬腭擦音［ç］则出现在前元音之后，如表 4.4 所示。

表 4.4 德语中由元音决定的清软腭擦音［x］和清腭擦音［ç］

前元音	注释	后元音	注释
siechen［ˈziçən］	衰败	suchen［ˈzuxən］	搜索
rächen［ˈʁɛçən］	报仇	Rochen［ˈʁɔxən］	射线

在瑞士德语中，有罕见的软腭塞擦音［k͡x］，如［tːaŋ k͡x］“坦克”。

英语中有唇－软腭近音［w］，为了制表方便，我们把它放在了软腭音一栏（见 3.1.5）。软腭近音［ɰ］也很少见，在确认它是一个音位的语言中，人们对它的描述都不够细致，很难区分该音位到底是软腭近音［ɰ］还是唇－软腭近音［w］。

在中瓦几语（Mid-Waghi，一种在巴布亚新几内亚使用的巴布亚语言）中有浊软腭边音［ʟ］，例如［aʟaʟe］“眩晕的”。

小舌（uvular）爆发音产生时有一个比软腭部位更靠后的收紧，通过将舌后部抬向小舌来完成。阿拉伯语中有一个清小舌爆发音［q］，如［qalb］“听”。达吉语（Dargi，一种在俄罗斯使用的高加索语言）有浊小舌爆发音［ɢ］，如［ɢaraʔul］“观察”。浊小舌爆发音［ɢ］很少见，这或许是因为在喉部附近收紧的同时很难维持发浊音。发这个音时，喉以上的气压很快就会和声门下的气压相同，这时将不再有气流从肺中流出，声带振动也随之停止。

小舌鼻音［ɴ］可以在因纽特语中找到，这是一种在格陵兰和加拿大所使用的爱斯基摩－阿留申语（Eskimo-Aleut），比如［eɴina］“旋律”。发现一个真正有浊小舌擦音和清小舌擦音的语言也许并非易事，因为在产出擦音的时候小舌经常会振动，这样就产生了小舌颤音而不是小舌擦音。法语有浊小舌擦音，例如“roux”［ʁu］“红的”；清小舌擦音为［ʁ］的音位变体，例如“lettre”［lɛtχ］“信件”。

4.1.4　咽喉音

咽喉音包括咽音、会厌音和喉音。咽音由咽的中部产生，会厌音由咽的较低部位产生，而喉音由声门处产生。咽音（pharyngeal）产生时，舌根向咽部后缩，这种音出现在很多闪族语（Semitic languages）中。例如，现代希伯来语的东方方言中既有清咽擦音［ħ］，如［ħaːli］“我的情况”；也有浊咽擦音［ʕ］，如［ʕor］“皮肤”。不过这其中还存在一些争论，例如，阿拉伯语和希伯来语中的咽音究竟是真正的咽音还是会厌音呢？对于会厌音（epiglottal）而言，会厌是主动发音器官，发音时成阻位置在靠近咽的后壁处形成一处收紧。

有这样一个罕见的例子，在高加索地区属于阿古尔语（Agul）的布基汉（Burkikhan）方言中，既有咽音又有会厌音。这个语言中有一个清会厌爆发音［ʡ］，如［jaʡar］“中心”；同时，也包含对立的清咽擦音［ħ］、浊咽擦音［ʕ］和清会厌擦音［ʜ］，如［muˈħar］“谷仓”、［muˈʕar］“桥”和［mɛˈʜɛr］“乳清”。在达吉语中还有浊会厌擦音

[ʕ]，如［ʕuʃːaː］“你”。喉音（glottal）产生时，会在声带处成阻。我们在3.4.1.5中将英语中的喉爆发音看作/ʔ/的音位变体。喉爆发音在夏威夷语中具有音位属性，如［ʔaʔa］“挑战”。此外，在前面讨论英语语音时，我们还介绍了清喉擦音和浊喉擦音［h，ɦ］。

4.2 其他发音方法

现在我们来看看表4.1中所列的那些在英语中不起到音位区分作用的其他发音方法，即颤音、拍音/闪音和边擦音。

发**颤音（trill）**时，一个发音器官靠近另一个发音器官，气流通过两个发音器官之间形成的空隙使其中一个发音器官产生振动。这种振动是由空气动力引起的，即一个发音器官会在气流的冲击下产生激烈的振动，这种振动并不是由肌肉的收缩产生的。如表4.1中所示，这种由气流引发的振动可以出现在不同的发音部位：唇、舌尖和小舌都可以颤动；颤音可以是清音也可以是浊音。一个浊双唇颤音［ʙ］看似奇怪，但是却很容易发出来。仅需将嘴唇放松地碰在一起，然后吹气就可以引起振动。浊齿龈颤音［r］是将舌尖或者舌叶置于齿龈脊下方而产生的。如果你儿时没有学过这些发音，那么对你而言，这些发音或许会有些困难，因为舌尖不仅要收紧以接近齿龈脊，同时也要足够放松，才能在气流的作用下保持颤动。西班牙语有齿龈颤音，如“perro”［ˈpɛro］“狗”。小舌颤音［ʀ］是通过小舌的颤动产生的，就像漱口的感觉一样。在一些荷兰语的变体中可以找到这个音，例如“rok”［ʀɔk］“裙子”的首音。

拍音（tap）或者**闪音（flap）**仅仅是肌肉一次收缩产生的，即一个发音器官接触到另一个发音器官而产生的一个比塞音更为短暂的闭合。IPA最近增加了一个新的唇齿闪音符号，这个闪音出现在很多非洲中部和非洲东南部的语言中。在唇齿闪音的发音过程中，下唇收缩时经过上齿，接着迅速向前短暂地碰触并经过上齿。这一发音的IPA符号是［ⱱ］，如莫诺语（Mono）中的［áⱱárá］“明智的”（［á］代表了高调的元音［a］，见11.4.1）。在美式英语中，当齿龈爆发音处在元音或央近音之后并且在非重读元音之前时，这类音经常会实现为闪音，例如“latter”［ˈlæɾɚ］

或“party”［ˈpɑɹɾi］（见3.4.1.4）。在西班牙语中，齿龈拍音（一个短暂的闭合）和齿龈颤音（若干短暂的闭合）形成对立，如“pero”［ˈpɛɾo］“但是”与“perro”［ˈpɛro］“狗”。闪音需要肌肉活动并且是一种一次性的运动，颤音则是由空气流动产生的，是发音器官的一种被动的重复性的颤动。

印地语（Hindi，一种在印度使用的印欧语言）有卷舌闪音［ɽ］，如［gʰoɽa］“马”。

尽管边音是典型的近音，但是边音也可以由不同程度的收紧产生。其中最常见的是**边擦音（lateral fractive）**。祖鲁语（Zulu，一种在南非使用的尼日尔－刚果语言）存在清边擦音、浊边擦音以及浊边近音的对立，如［ɬàɬá］“切断”，［ɮálà］“玩”和［lálà］“躺下”。

4.3　元音

图4.1为元音的IPA符号。这些符号是按照3.2节中介绍的元音四边形组织排列的。当符号成对出现时，左侧的符号代表不圆唇元音，右侧的符号代表与左侧符号相对应的圆唇元音。这个图中共有28个元音，我们在介绍英语元音时（见3.2节），已经讨论过其中的13个元音。在世

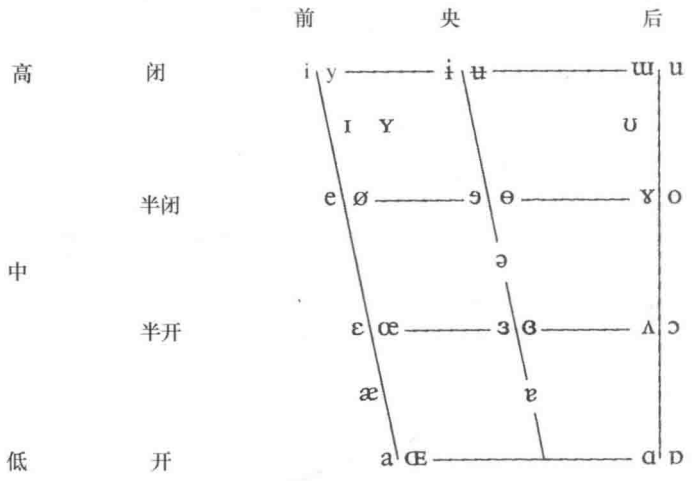
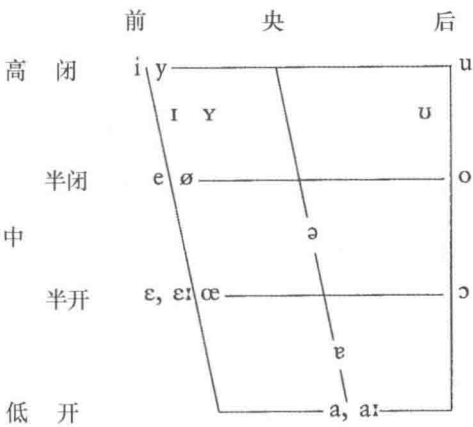


图4.1　元音标音所使用的IPA符号（2005版）

界语言中，每种语言所含的元音的数量相差很大，从 3 个到 40 个不等。元音数目多的语言通常包括长元音和短元音、口元音和鼻元音或者普通元音（plain vowels）与咽化元音（pharyngealized vowels）。在世界语言中，包含/i，e，a，o，u/这五个元音的语言是十分常见的。相较而言，英语所包含的元音是比较多的。然而，虽然英语具有高、中、低元音及前、央、后元音，但是英语元音在音长或者圆唇度上不存在对立。



例子		注释
bieten	[biːtn̩]	“提供”
bitten	[bɪtn̩]	“乞求”
Wüste	[vʏstə]	“沙漠”
wüsste	[vʏstə]	“知道”
Ruhm	[ʁuːm]	“声望”
Rum	[ʁʊm]	“朗姆酒”
beten	[beːtn̩]	“祈祷”
Betten	[bɛtn̩]	“床”
bäten	[bɛːtn̩]	“他们乞求”
Goethe	[gøːtə]	“歌德”
göttlich	[gœtlɪç]	“神圣的”
Gote	[goːtə]	“哥特人”
Gott	[gɔt]	“上帝”
Maße	[maːsə]	“测量”
Masse	[masə]	“民众”

图 4.2 德语元音、包含相应元音的例词以及每个例词对应的 IPA 转写符号（注意：尽管［a/aː］是前元音的符号，在德语 IPA 图上，［a/aː］通常放在中央或更靠后的位置）。

德语是一种会使用音长和圆唇度构成对立的语言。如图 4.2 所示，德语有 17 个元音，它们的 IPA 符号也在图上给出。图中清楚表明，德语中有前圆唇元音和前不圆唇相对立（[i, y]，[ɪ, ʏ]，[e, ø]，[ɛ, œ]），也有长短元音的对立（[ɛ, ɛː]，[a, aː]）。值得注意的是，尽管德语有几对长元音和短元音（如 [i, ɪ]，[e, ɛ]，[y, ʏ]，[ø, œ]，[u, ʊ]，[o, ɔ]），但是事实上，每组音之间的区别主要是音质上的差异，这种差异应用不同的音标符号（而非音长附加符号）加以区别。另一方面，[ɛ, ɛː] 和 [a, aː] 这两组符号则表明，每组音之间的主要区别实际上是音长而不是音质。当考察图 4.3 所表示的德语的元音声学空间时，上述差别一目了然。图中所示的数据是通过声学分析获得的（发音与元音音

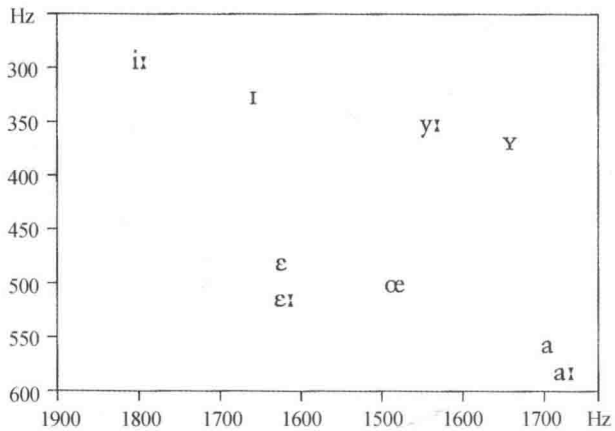


图 4.3　德语的元音声学空间。成对出现的元音体现出音长和圆唇度的区别。

质的声学描述之间的关系将在 9.5.1 中讨论)。元音 [ɛ] 与 [ɛː]、[a] 与 [aː] 彼此非常接近，这表明每对元音在音质上差异很小。然而，[iː] 与 [ɪ]、[yː] 与 [ʏ] 彼此之间的音质区别就很明显。图中同样显示了前圆唇元音和前不圆唇元音 [i, y]，[ɪ, ʏ] 以及 [ɛ, œ]。由此可见，圆唇元音与不圆唇元音之间的音质区别非常明显。在世界语言中，通常后元音是圆唇的，前元音是不圆唇的。然而，除了德语以外，很多语言都有前圆唇元音，例如荷兰语、芬兰语、法语、匈牙利语、韩语、汉语普通话和瑞典语。

利用英语和德语，我们已经介绍了图 4.1 所列的 28 个符号中的 18 个。接下来，我们将简要地讨论剩下的元音。在前元音中，我们唯一没有说到的就是 [œ]。它紧挨在 [a] 的右边，因此我们知道 [œ] 是元音 [a] 对应的圆唇元音。就像你可以仅通过圆唇而不改变其他发音器官的位置就把 [i] 发成 [y] 一样，你同样也可以采用这种方法把 [a] 发成 [œ]。元音 [œ] 通常很少见，可以在阿姆施泰滕 (Amstetten) 的巴伐利亚 (Bavarian) 方言中找到这个音 (Traunmüller, 1982)。

图 4.1 中，央高元音的区域有元音 [ɨ] 和 [ʉ]。央高元音 [ɨ] 在美式英语的一些方言中存在，例如复数形式 “houses” [ˈhaʊzɨz]。基于元音的发音比辅音更加连续这一事实，试着慢慢地调整你的发音器官由

[i] 向 [u] 转变，这需要将舌向后缩，但同时应保持舌位高低不变且唇形不变圆。在这过程中，你会经过 [ɨ] 这个元音。同样地，如果你反方向由 [u] 向 [i] 发音，舌向前运动而不降低，你会经过 [ʉ] 这个元音，在挪威语中就有这个音。元音 [ɨ] 和 [ʉ] 的描述分别是央高不圆唇元音和央高圆唇元音，而它们通常分别被称为“杠 i”（barred i）和“杠 u”（barred u）。

在图 4.1 的中心区域，也有一些符号。我们已经讨论了其中的元音 [ə]，它是一个弱化的非重读元音。我们也简要地介绍了元音 [ɜ]，在英式英语的一些方言中有这个元音，如“bird”中的元音，其中的“r”是不发音的。还有一些央 - 中元音，但是人们并不清楚它们是否在一些语言中起着对立的作用。虽然如此，我们还是可以通过在与其对应的前元音和后元音之间插值来逼近它们的音质。最后，在德语中有元音 [ɐ]，如“Becher”[ˈbɛçɐ]“烧杯”，词尾的 [əʁ] 会发生元音化，变为 [ɐ]。这个音的音质实际上和美式英语“but”中的 [ʌ] 十分相似。

在后元音中，我们首先看 [ɯ]。它的位置在 [u] 的左边表明它是一个后高不圆唇元音，也就是说，它对应的是一种不圆唇的 [u] 的发音。这个元音很常见，在诸如日语以及韩语等语言中都存在。在图上更低一点的位置，是与 [o] 对应的后中高不圆唇元音 [ɤ]，在泰语和越南语等语言中有这个音。最后，后低圆唇元音 [ɒ] 与不圆唇元音 [ɑ] 相对应，[ɒ] 是英式发音中的一个元音，例如“pot”。

4.4 次要发音

到目前为止，我们所描述的声音大都只具有一个收紧特征。然而，声音也可以由位于两个不同发音部位的同步的收紧产生。当发生这种情况时，可以按照收紧程度来区分这些声音。一种情况是两个收紧的程度相同，另一种情况是收紧程度一个比另外一个更为显著。当两个收紧的程度相同时，两者均为主要发音；当一个收紧比另一个更显著时，较显著的为主要发音，较不显著的为次要发音。在前面我们已经讨论过了唇 - 软腭音 [w]，发音时唇和软腭的收紧程度相同，像这样的发音涉及两个主要的发音器官，是一种双重发音（见 3.1.5）。有些辅音在发音时

存在**次要发音（secondary articulation）**，如单词“soup”词首辅音的圆唇动作。单词起首擦音的主要收紧位于齿龈脊，但是由于预见到后接的圆唇元音［u］而产生了圆唇的次要发音，发出了唇化擦音［sʷup］。因为次要发音通常被描述成在辅音发音的基础上叠加一个类似元音的发音，所以我们首先讨论了有关元音的描述。想将一个次要发音与一个两音段序列区分开来是很困难的，例如“soup”［sʷup］与“swoop”［swup］。它们之间的区别与发［s］和［w］时的相对时序有关。在次要发音中，这两个发音姿态同时出现；而在一个两音段的序列中，［w］的发音起始较晚，与［s］的后过渡段相连但有部分重叠。

通常人们认为有四种次要发音：

1. 唇化（labialization）正是上文所描述的那种次要发音，通过附加符号［ʷ］表示。我们可以将唇化看成是在主要发音上叠加类似［u］或者［w］的发音。唇化的出现常依赖于一些相邻的音段，如［sʷup］。不过，唇化也可以用于构成对立，例如契维语（Twi，一种在加纳使用的尼日尔－刚果语言）中的［àkʷá］“迂回曲折”与［àká］“某人咬了”。

2. 硬腭化/腭化（palatalization）叠加了类似［i］或者［j］的发音，发音时舌面在硬腭处形成一个次要收紧。俄语在很多发音部位都存在普通辅音（plain consonants）与腭化辅音的对立，如［sok］“果汁”和［sʲok］“他抽打了”。

3. 软腭化（velarization）需要叠加一个像后高元音［ɤ］那样的发音，将舌后部抬向软腭形成次要收紧。软腭化可以通过附加符号［~］来表示，例如英语中位于元音后的“暗 l”（dark l），用［ɬ］；或者可以用附加符号［ˠ］表示。软腭化在马绍尔群岛语（Marshallese，一种在马绍尔群岛使用的奥斯楚尼西亚语言）中构成对立，如［mˠatʲ］“鳗鱼”和［matʲ］“眼睛”。

4. 咽化（pharyngealization）需要将舌后部降低，并使舌根向咽腔后壁收缩。咽化通常与软腭化用相同的附加符号来表示，因为世界语言中一般不会有这两种次要发音的对立。有时也会用浊喉擦音的符号当作附加符号［ʕ］以表示咽化。许多阿拉伯语的方言会区分普通辅音与一系列被称为**重辅音（emphatic consonant）**的咽化辅音。例如约旦的阿拉伯语方言中［tuːb］“悔改”与［tʕuːb］“砖”对立。尽管重读被认为是这

类辅音的特质，但这里主要感知到的还是相邻元音的音质变化。

在这四种次要发音中，唇化是世界语言中最常见的。尽管次要发音在一些语言中具有音位地位，但它们在 IPA 中仍用附加符号来表示。

本章中，我们尝试从辅音和元音的发音部位及发音方法的角度向大家展示了世界语言的多样性。当然，这些还不能囊括世界语言中的所有声音。但是，本章应该加强了我们对英语之外的其他语言中各种言语产生方式的认识。至于哪些声音听起来有一些“奇怪”，要取决于你的母语是哪一种语言。下一章将详细介绍由肺到双唇以及鼻孔的气流，而第 6 章将介绍由非肺部气流所产生的其他辅音。

练　习

1. 在国际音标表中选择几个灰色的单元格，请试着解释为什么那些位置的发音是不可能实现的。
2. “emphasis”中的［m］与“empire”中的［m］在发音方面有哪些不同？为什么？
3. 在国际音标表中，为什么将齿音、齿龈音和龈后音放在同一列中？
4. 解释一下拍音和颤音的发音方法有什么不同，请各举一例。
5. 为什么一些在音长上存在对立的元音是通过附加符号（如［ɛ］和［ɛː］）表示，而另一些则是通过单独的符号（如［ɪ］和［i］）表示？
6. 什么是次要发音？总结并举例说明四种次要发音的特点。

注　释

1. 译者注：原书所列的链接为 http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants，该链接已发生变化。
2. 译者注：“上海”的汉语普通话读音应为［ʂɑŋxai］。

5 发音器官的生理机能

在1.1节中我们讲到，肺是言语产生的主要能量来源，而喉则是重要的声源。语音的发音过程主要是在声道中完成的。现在我们将更加详细地描述这些发音器官以理解它们的功能和它们对言语产出的影响。

5.1 声门下系统:肺、支气管和气管

世界语言中的大多数语音，包括英语中的所有语音都是通过从肺部产生的气流形成的。要理解**外呼肺气流（egressive pulmonic airstream）**的产生，首先要了解肺的解剖结构和功能。

5.1.1 声门下系统的解剖结构

我们用来呼吸的声门下系统（见图5.1）是由位于**胸部（thorax）**的两片**肺叶（pulmone）**组成的。肺叶通过左右**支气管（bronchi）**与**气管（trachea）**相连。在肺内部，支气管分裂为称作**小支气管（bronchia）**的更小的分支，并近一步分裂成近3000万个**肺泡（alveolli pulmonis）**（Weibel，1984）。对人类至关重要的空气与血液之间的气体交换就发生在肺泡里。为了进行气体交换，肺泡进化出了非常薄的膜，从而可以把血液（肺泡外的）和空气（肺泡内的）分隔开。

肺是海绵状的弹性纤维，不含任何肌肉。肺部大约有25%的弹性都源于肺组织本身的弹性，这种特性与袜子相似。其余的弹性则来自肺泡周围血液中水分子的**表面张力（surface tension）**，这部分弹性起到了支撑肺部结构的作用。肺组织本身的弹性和肺泡的表面张力一起构成了**弹性回缩力（elastic recoil force）**，可以让肺收紧。

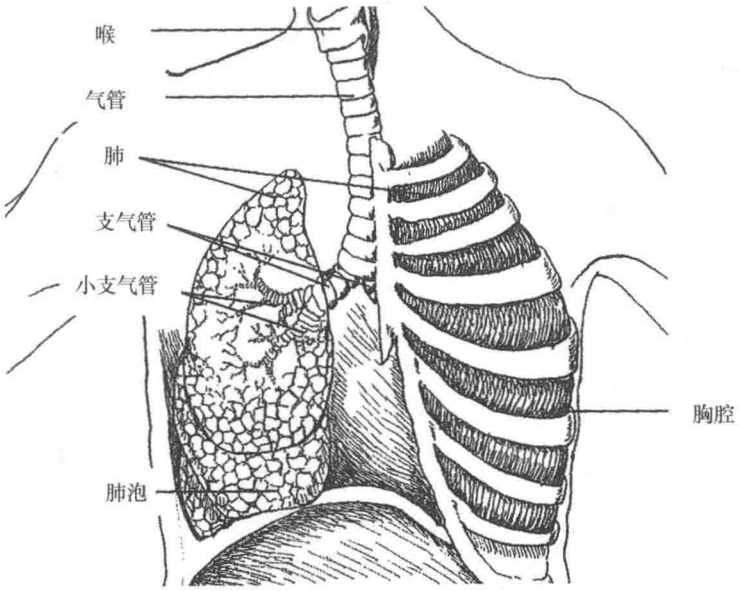


图5.1 声门下系统。在左边，为了显示相关的结构，隐去了组织和肋骨。

既然肺本身没有肌肉，那么我们是如何进行呼吸的呢？更确切地说，我们是怎样吸入和呼出空气的呢？吸气会使肺部扩张，这样肺内部的气压就比体外的气压低，空气便可以通过气管进入（吸入）肺里了。在物理中这个现象被称为**波义耳定律（Boyle's Law）**——在给定的温度下，气体的压力和其体积的乘积是一个常数：

$$\text{体积} \times \text{压力} = \text{常数}$$

或者说，压力与体积成反比：

$$\text{体积} = \frac{\text{常数}}{\text{压力}} \quad \left(\text{或 } V = \frac{\text{const}}{p}\right)$$

这表明随着肺体积的增大（即 V 变大），肺内部的气压会减小（即 p 变小），进而导致气流的**内吸（ingressive**，即吸气）。反之，肺的收缩会导致气流的**外呼（egressive**，即呼气）。也就是说，肺体积的缩小增大了肺内部的气压，从而使气流被呼出。肺就是靠这样交替的扩张和收缩来吸气和呼气的。但是肺并不能自主地完成这些动作，因为肺本身没有任何肌肉。

肺在胸腔中的悬挂方式（见图5.2）是使其具备扩张和收缩能力的重

要因素。两片肺叶都被它们各自的**脏层胸膜（visceral pleura）**（即**肺胸膜，lung membrane**）所包围。这些膜通过**胸膜间腔（interpleural space）**和同与肋骨相邻且处于**膈膜（diaphragm）**上方的**壁层胸膜（parietal pleura）**相连。脏层胸膜和壁层胸膜的组织非常光滑而且几乎完全气密，这对人们实现正常的呼吸和说话都至关重要。

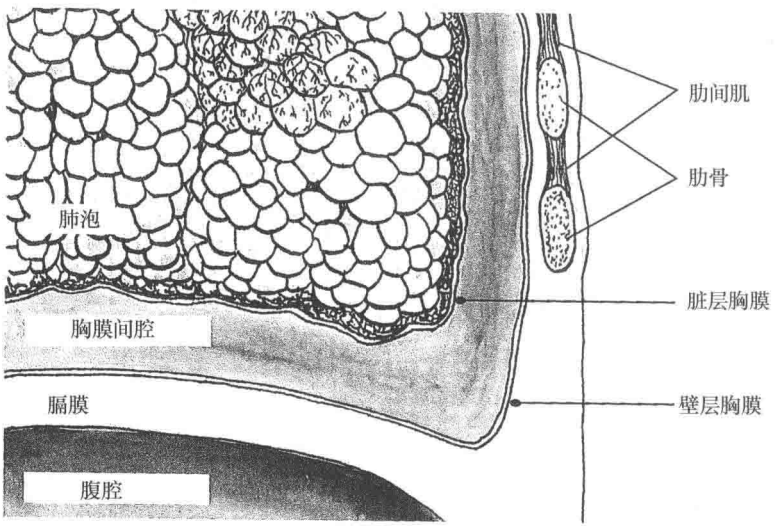


图5.2　一张肺部的特写，详细展示了肺膜和相邻的组织结构。

5.1.2　肺的运动

由肋骨和膈膜形成的腔体被称为**胸腔（thoracic cavity）**。胸腔的运动通过脏层胸膜、壁层胸膜和胸膜间腔传送到肺部。胸膜间腔位于脏层胸膜和壁层胸膜之间，其内部充满了润滑液。胸膜中的特殊细胞会不断地从胸膜间腔吸收气体，从而使液体中的压力比肺内部的低。这便引起了脏层胸膜和壁层胸膜之间的**胸腔联动（pleural linkage）**，使肺可以随着胸腔和膈膜的运动而运动。这个系统可以通过下面这个简单的例子来加深理解：在两块玻璃片（模拟两个胸膜）之间滴几滴水，这两块玻璃片便可以相互摩擦运动，但我们却很难将它们分开。

5.1.2.1 吸入

肌肉的活动可以改变胸腔的大小。当一个人进行**胸式呼吸（thoracic breathing）**的时候，**肋间外肌（external intercostal muscle）**的收缩可以让肋骨稍微外凸，从而使胸腔扩大。膈膜本身是由肌肉构成，可以用于辅助呼吸。处于静息状态时，膈膜略微向上凸起。当膈膜被拉紧时，它自身会变平，向下扩大了肺所在的腔体，这就产生了**腹式呼吸（abdominal breathing）**。人们呼吸的方式会存在个体差异，有人喜欢采用胸式呼吸，有人则倾向于腹式呼吸，这种差异与性别无关（Zemlin，1998：93）。在言语产出的过程中，一般情况下，上述两种运动都会发生：激活吸气肌肉，会同时使肋间肌向外收缩并拉紧膈膜，从而扩大胸腔，扩张肺的体积。

5.1.2.2 呼出

呼气过程中的运动与上面提到的吸气过程恰好相反。在呼气的过程中，几乎不需要任何肌肉的力量，因为肺会在弹性回缩力的作用下收缩。由于在吸气的过程中，肋骨和/或膈膜在肌肉运动的作用下比静息位置下更向外扩张，在呼气的过程中，它们会被动地恢复至原位。另外由于受重力的作用，在呼气的过程中胸腔会被自然地下拉，腔体内的空间会缩小。因此，直到肺恢复到静息位置，也就是肺的收缩力和胸腔的扩张力相当的时候，呼气（即被动呼气）都不需要任何肌肉力的作用。如果肺继续收缩，**肋间内肌（internal intercostal muscle）**就会起作用，使肋骨相互靠拢。人们对通过肋间内肌作用的主动的呼气的控制不如由肋间外肌调控的被动呼气那么精确。由于语音的响度取决于肺部的气压（见5.1.4），因此与主动呼气相比，我们可以在被动呼气的过程中更精确地控制响度。

正如我们所描述的那样，呼出是自动发生的。所以我们可以很自然地把肺比作一个气球：当将充气后的气球中的气体释放时，气球会迅速地将气体放空，并且这一气体流出的过程是非常猛烈和突然的。肺也是如此，不过在呼气阶段，在肺恢复到静息位置之前，用于辅助吸气的肌肉可以起到控制的作用，使肺中的气体不会过于迅速地被呼出。这说明在呼气过程中，用于辅助吸气的肌肉也仍然保持在激活状态。用于辅助呼气的肌肉只有在肺部收缩得比静息状态还要小时才会被激活以挤出肺

里的空气。总的来说，在言语产出的过程中，辅助吸气的肌肉和辅助呼气的肌肉共同作用，从而保持了气压的相对稳定。

5.1.3　肺的容量与其随时间变化的控制

两片肺叶加起来的最大容积约有 6 L。在静息状态时，两片肺叶中的气体体积大约为 3 L。在安静呼吸的情况下，两片肺叶会主动吸入大约 0.5 L 的气体。这称作**潮气量（tidal volume）**或者**静息容量（rest capacity）**（见图 5.3）。在一次深呼吸的过程中，肺可以通过吸气肌肉的作用吸入体积等同于**吸气储备容量（inspiratory reserve volume）**的气体使肺部容积达到最大。在呼气的过程中，尽管辅助吸气的肌肉可以起到一定的控制作用，但在弹性回缩力的作用下，呼气会使肺被动地恢复到静息位置。当肺的收缩超过了静息位置之后，在呼气肌肉的作用下，空气可以被主动排出，最多可以排出体积等同于**呼气储备容量（expiratory reserve volume）**的气体。最后有 1 L 左右的**残留量（residual volume）**不能被排出，一直储存在肺泡中。最大吸气容量和最大呼气容量之间的差值被称为**肺活量（vital capacity）**。

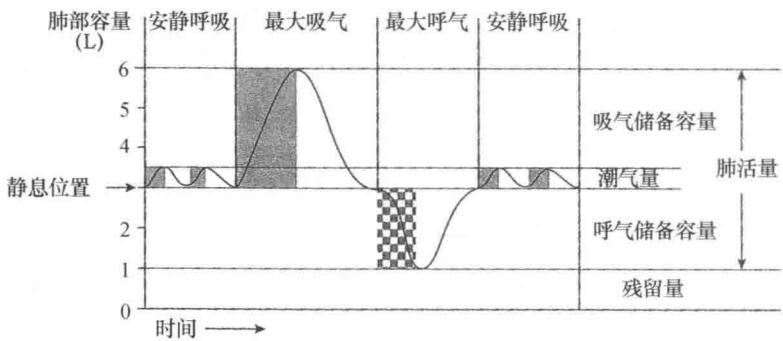


图 5.3　肺部的气体容量在呼吸循环中的变化。灰色部分表示吸气肌肉（肋间外肌和膈膜）处在激活状态；黑白方格部分表示呼气肌肉（肋间内肌）处在激活状态。

在安静的情况下，人每分钟呼吸 10 到 20 次，吸入或呼出的气体约为 0.5 L。在说话时，呼吸的节律会降低。然而，安静状态下的呼吸和说话时相比，最主要的区别并不是呼吸节律，而是呼吸序列的组织安排。在

安静呼吸时，呼吸循环中大约40%用于吸入，余下的60%用于呼出。但是，在说话时，呼吸循环中只有10%被用于吸入，而其余90%都被用于呼出（见图5.4）。当一个人说话时，辅助呼吸的肌肉组织可以起到延长呼出段的作用。

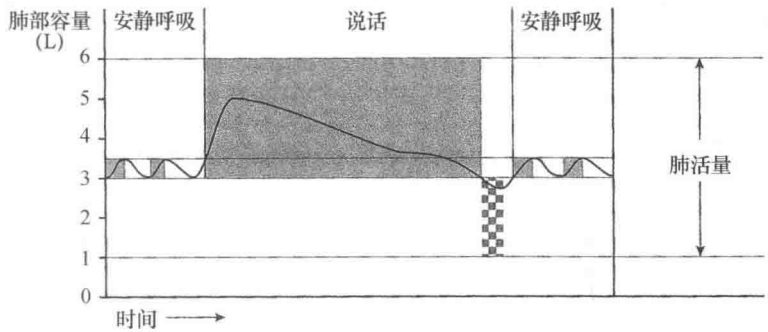


图5.4 在呼吸和说话时肺部的气体容量。灰色部分表示吸气肌（肋间外肌和膈膜）处在激活状态；黑白方格部分表示呼气肌（肋间内肌）处在激活状态。

5.1.4 响度与肺部气压

我们在前面讲到，在说话的过程中，肺部的气压保持相对稳定。无论产生什么语音，不同的肌肉群都会使肺部的气压保持在基本相同的水平。这一点非常重要，因为语音信号的声压级（见7.3.2）是衡量语音响度的一个重要参数。语音信号声压级的增加量与声门下气压增加量的平方成正比。也就是说，声门下气压增大为以前的两倍就可以使声压级增大到以前的四倍。因此，要想控制语音的响度，我们必须对气压有高水平的控制能力。这样的控制主要是由肋间外肌来实现的。

这种气压变化调控对音节重音的相关理论有着直接的影响（见11.1节）。重音的实现似乎并不是简单地依靠增加肌肉紧张程度来增大音节的响度的。重音的变化还可能与其他的很多因素有关。例如，增大声带闭合的速度会使听音人在感知上觉得语音信号变得更“响”了（见5.2.3）。呼气是一个吸气肌肉（在肺到达静息位置过程中处在激活状态）与呼气肌肉（在肺超过静息位置之后处在激活状态）之间复杂交互作用的过程。如果只是简单地增大肌肉的活跃程度，会产生什么样的结果呢？

在呼气的第一阶段，较强的肌肉活性会增加吸气肌肉的紧张度，这样可以使肺扩张，进而使肺部的气压降低，使得产出的声音听起来比较柔和。只有当肺部收缩超过静息位置之后，呼气肌肉才会被激活，这会引起响度的增加。也就是说，像“重音”这样的现象并不是一种简单的从一个概念（重音）到生理现实（肌肉拉紧）的一对一的映射。相反，一个简单抽象的概念揭示出的是一个高度复杂的运动过程。这个过程中涉及不同肌肉群的激活状态，而且其激活与否，要取决于肺的收缩与静息位置之间的相对关系。

总的来说，在吸气时，胸腔主动扩张，气流通过气管、支气管和小支气管进入两片肺叶。尽管一些感叹词是通过内吸的气流产生的（大多数情况下是用来表现惊讶或者确认，而不是用来表达实际的语言意义），但是在语流中没有语音可以在吸气阶段产生。在说话时，从肺部流出的气流通过气管和喉进入声道。这个过程可以说是被动的，因为直到肺恢复到它的静息位置之前，肺中的气流都可以自动地流出。在言语产出的过程中，我们需要采用主动的力来控制这些被动流出的气流。而当肺收缩超过了静息位置之后，我们还需要主动的力来支援气体的呼出。我们需要精确地控制气管内的气压，因为语音信号的音强就取决于此。就像前面提到的，喉的作用对言语产出至关重要，下一节我们将对此作详细的介绍。

5.2　喉的结构和功能

喉位于颈内气管的顶部，声带就位于喉中。在言语产出过程中，如果声带振动，该语音就被称作浊音。如果声带不振动，产出的语音就是非浊的或者被称作清音（见 2.1 节，但关于一些语言中爆发音清浊的定义请另见 6.3 节）。声带的振动使产出的言语具有一种特性，这种特性被感知为**音高（pitch）**。音高使得词或者句子具有了一定的“旋律”。下面的部分将介绍喉的解剖结构和声带振动的原理。

5.2.1　喉的解剖结构

喉是一个可以调节的软骨管道结构，通过这个管道，气流可以被吸

入或呼出。喉位于气管的顶部，止于**舌骨（hyoid bone）**。舌骨是一个马蹄铁形的环形骨，它位于颈部的上端，支撑着舌根（见图5.5a和b）。

气管由许多个软骨环组成，这些软骨环后侧带有开口，看起来像一个个小马蹄铁。半环状的软骨后侧直接与**食管/食道（esophagus/gullet）**相连。这些软骨的组织结构和**气管肌（trachealis muscle）**一起构成了一个密闭的环。在这些软骨环之间，有一系列韧带组织结构包围着整个气管，使其气闭。

位于气管上方的喉与气管一样，由软骨、肌肉和韧带组成。喉的框架由**甲状软骨（thyroid cartilage）**构成，包括两片在前端结合在一起的软骨。对女性来说，这两片软骨的夹角是90°；但是对男性来说，这两片软骨的夹角是80°（Zemlin，1998：176－177；见图5.5c）。甲状软骨的前端是可见的，特别是在男性的颈部，可以看见这两片软骨形成的尖角，通常被称为**喉结（Adam's apple）**。在构成甲状软骨的两片软骨之间是一个三角形的凹槽，该凹槽可以从外面摸到（见图5.5a）。位于甲状软骨内部的声带正好处在这个三角形凹槽的下方，声带的振动为言语的发声提供了条件。

喉的第二个软骨结构是**环状软骨（cricoid cartilage）**。在环状软骨环形结构的后部有一个扁平的面（见图5.5a），这使其看上去像一枚戒指。环状软骨两边各有一个延伸部分（在图5.5中看不见）使其与**甲状软骨下角（cornu inferior）**相连，这样就提供了一个水平方向的轴，使甲状软骨和环状软骨可以围绕这个轴相对转动。这样的转动由**喉外肌（external laryngeal muscle）**控制，转动方式与摇摆运动类似，可以使声带产生一定程度的拉紧，这对控制声带的振动速率（即我们感知到的音高）十分重要。

喉的第三个软骨结构是**杓状软骨（arytenoid cartilage）**。两块杓状软骨像两个金字塔一样附着在环状软骨的上部（见图5.5a和c）。每个杓状软骨都有一个延伸部分，叫作**声带突（vocal process）**，声带突与声带相连。通过拉紧**环杓后肌（posterior cricoarytenoid muscle）**，杓状软骨会变得倾斜，进而使得声带向上、向两边分开，这个过程叫作**外展（abduction）**。这样的外展可以使声带完全分开，例如在呼吸的时候，或者在发清音的时候，声带就处于完全分开的状态。通过拉紧**环杓侧肌（lateral**

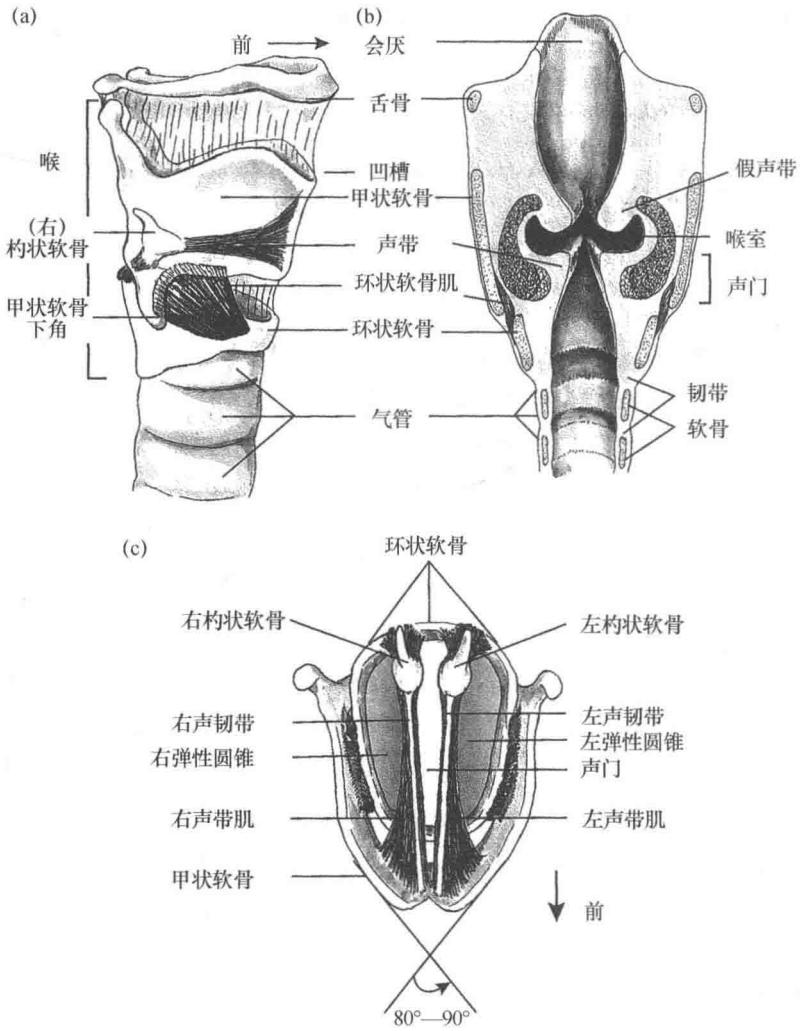


图 5.5 从三个角度显示喉部解剖结构：（a）从右侧看的侧面（矢状面）图，（b）从前面看的纵剖面（冠状面）图，（c）从上面看的横剖面（水平面）图。在（a）的半透视图中，可以看到位于甲状软骨内部的声带和杓状软骨。

cricoarytenoid muscle），杓状软骨会向下、向内移动，使声带**内收**（**adducted**）合拢。如果声带只是稍微分开并处于拉紧状态，那么它就可以在肺气流的作用下开始振动，产生浊音。拉紧**横杓肌**（**transverse arytenoid muscle**）和**斜杓肌**（**oblique arytenoid muscle**）可使声带完全闭合，例如在吞咽的时候，为了避免食物进入气管，就会完全闭合声带。声带的完

全闭合也可能出现在言语产出过程中，例如产生一个喉爆发音（见3.4.1.5）或者一个声门气流的时候（见6.1.1）。

会厌（见图5.5b）是一块鞋拔形状的软骨，位于舌骨中间，处在喉入口的顶部。尽管我们还没有完全了解会厌的功能（Zemlin 1998：107－108），但是一些研究者（例如，Moore 和 Dalley 1999：1054）认为会厌可以在人吞咽的时候起到阻止食物进入气管的作用。这种说法的依据是会厌所处的位置。会厌就像喉顶部的一个小盖一样，当食物碎块到达喉部时，食物自身的压力可以使这个“小盖”被动地关闭。此外，甲状会厌肌的收缩可以使会厌主动地关闭。

喉内部的两片**声带（vocal folds，**有时也称作 vocal cords）在言语产出的过程中起到至关重要的作用。两片声带的前端分别与喉结的三角形凹槽下方的两片甲状软骨的汇合处相连。从这个汇合处起，每片声带都延伸到其对应的杓状软骨上（见图5.5a 和 c）。每片声带都包含**声带肌（vocalis muscle）**，声带肌附着在**声韧带（vocal ligament）**上，而声韧带则会从甲状软骨伸展到对应的杓状软骨上。声韧带是密闭的**弹性圆锥（conus elasticus）**的上部；弹性圆锥向环状软骨的内部延伸，将喉分成上部和下部。声带的肌肉可以靠收缩来改变弹性圆锥的弹性张力，但这也会使声带的形状从比较厚的唇形变成拉紧的薄带状，这是影响声带振动特性的另外一种方式。

在过去，人们认为声带的振动就像两根拉紧的小提琴的弦一样。但是现在的研究认为，声带的振动是一个基于肌肉组织特性的更为复杂的运动。这样看来，声带的术语使用“vocal folds”要比“vocal cords”更为恰当。因为无论是从外形还是机械原理的角度来说，声带都与“瓣状物”（flaps）或“褶状物”（folds）更为相似，而不是“弦”（cords）。

假声带（false vocal folds）位于真正的声带的上方，在假声带和声带之间有**喉室（ventricles）**相隔（见图5.5b）。假声带是不可以被拉伸的，不过人们认为假声带具有保护功能。当一些有病态的人说话时，假声带会像在风中飘动的两片纸一样，使发出的声音带有比较粗的音色。

声带中间的空隙（从声带的下边缘到上边缘）叫作声门。提起声门，人们经常把它看做是一个器官，例如“打开声门”，但实际上，声门不是一个器官。声门从本质上讲就是言语产出过程中所使用到的声带之间的

整个区域。需要指出的是，即使是在声门打开到最大时，声带也会挡住气管直径大约50%的范围，因此人类的呼吸效率并不是很高。

总的来说，喉位于气管的上部，通过韧带组织和肌肉组织部分地悬挂于舌骨上，舌骨本身则通过韧带组织和肌肉组织附着于舌根。构成喉主体的甲状软骨位于环状软骨的上方。甲状软骨和环状软骨之间可以相对地转动。随着这种转动，声带可以或拉紧或放松。位于环状软骨顶部的杓状软骨可以产生倾斜，以改变声带的位置。杓状软骨向下、向里倾斜可以使声带内收合拢；杓状软骨向上、向外倾斜可以使声带外展分开。声带的拉紧和放松可以影响其自身振动的速率（见5.2.2）。声带的振动是由肺部气流和声门相互作用而产生的。声门可以打开，例如在呼吸或者发清音的时候；也可以闭合，例如在吞咽或者在发喷音的时候（见6.1.1）；还可以部分打开，这时声带内收到刚刚可以在肺部气流的作用下振动的位置。下一部分将详细地描述这种振动方式。

5.2.2　声带振动

声带的振动是一个非常复杂的运动过程。通常来说，声带打开的过程是自下而上、由后向前的。声带闭合的过程同样也是自下而上的，不过在水平方向上，声带的运动从中间开始，同时向前向后逐渐关闭。通常声带并不能完全闭合，尤其是女性，因为在声带后端、位于杓状软骨旁边的位置会留有一个小的三角形缺口。

前人提出了很多的理论试图去解释说话过程中的声带振动原理。尽管许多理论已经被现在的研究否定，我们还是在下面简要介绍，以向大家展示声带振动理论的发展过程。学者们曾经认为发声是源于以下的原因之一。

1. 振动弦理论：声带在气流中像琴弦一样地振动。
2. 神经时值理论：中枢神经系统的神经脉冲直接控制声带振动。
3. 空气动力学理论：气流吸入压的下降。
4. 肌弹性理论：声带的弹性。

振动弦理论（vibrating string theory）（Ferrein，1741）认为声带就像小提琴的弦一样，依靠自身的振动产生声音。然而这种假说并不合理，因为一根弦的振动需要借助共鸣体的扩音作用才可以使人清楚地听到。

例如，电吉他没有共鸣箱，所以如果不借助放大器的帮助，仅靠拨动琴弦产生的声音就很难被人听到；而对于声乐吉他来说，由于它有共鸣箱，所以琴弦振动的声音就可以被清楚地听到了。在第 9 章中我们将提到，声道具有共鸣性质，但其实声道只是一个简陋的共鸣器，需要借助空气喷流通过声带时产生的巨大能量才能使人们听到声道的共鸣特点。仅靠声带自身完成的振动（没有气流经过声带）是不能被人耳所听到的。

神经时值理论（neurochronaxic theory）（Husson，1950）认为声带的振动依靠的是肌肉的急速收紧和放松。然而这种假说同样是不合理的，因为在说话的过程中声带每秒钟开合 100 至 400 次，而在人身体内没有任何一个肌肉群可以像声带一样如此急速地运动。在唱歌的时候声带的振动可以变得更快，而一个正在啼哭的小孩的声带，每秒钟可以振动多达 2000 次。没有肌肉可以依靠主动运动达到如此高的振动速率。此外，即便是负责激活肌肉的神经，每秒钟最多也只能产生 1000 个脉冲。因此，对于在浊音的产出过程中声带所产生的快速振动，人们还需要寻找其他的解释。

我们现在知道，声带的振动不是依靠肌肉的急速收紧和放松，也不是依靠共鸣体的作用，而是更像一个木管乐器的吹嘴的工作原理。例如双簧管：演奏时，吹入的一股气流被周期性地阻断，由此产生的一段段空气脉冲就形成了一个音符。声音产生的原因并不是声带本身的振动，而是声带振动对通过声带的气流所产生的影响，就像双簧管，吹嘴对气流的周期性阻断才使得吹嘴处的振动被人们听到。

5.2.2.1 *伯努利效应和空气动力学理论*

我们可以把声带想象成是在一个通风良好的走廊里的两扇自由的摆动门。它们可以处于关闭的状态（内收），或者被墙钩固定处于永久打开的状态（外展），或者在流动的风中自由地摆动。风的气流越强，门开得就越大。在 5.1.4 中我们讲到肺部的气压会保持在一个相对恒定的水平，这说明通过气管和喉部的气流也是相对稳定的。如果气流是恒定的，那么摆动门（声带）可以打开并保持在某一特定的角度，但是门却不可能不停地前后摆动。正因如此，长期以来大家都不明白为什么稳定的气流可以使声带来回地打开和闭合，而不是仅仅让它们保持在稍微打开的状态。对于这个问题 van den Berg 和其同事（1957）基于**伯努利效应（Bernoulli effect）**给出了合理的解释。下面将详细介绍伯努利效应。

当液体或者气体（例如空气）稳定地通过一个管子时，所有的分子都会以大致相同的速度沿着液体或者气体流动的方向移动。这种流动状态被称为**层流（laminar flow）**。更准确地说，分子仍然在进行着大量的前后移动，但是总体上是沿着一个主要的方向稳定移动。这就好比嘉年华的游行队伍，大家不停地沿不同的方向跳舞移动，但是总体来说，人们还是在随着队伍行进的方向移动。

如果这个管子变窄，会发生什么呢（见图5.6）？这就和位于喉部的两片声带之间的情况差不多。流入和流出管子的气流是恒定的。由于空气分子在狭窄通道前后的流动速率相同，所以当分子通过狭窄通道的时候，它们必须移动得更快。

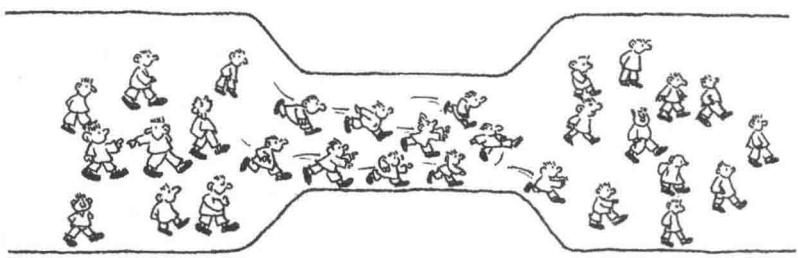


图5.6　层流中的分子。在狭窄通道里的分子为了避免拥挤，必须比在通道前后的分子移动得更快。

图5.7更准确地描述了这个过程。在这个“快照”中，在狭窄通道前后的通道里的分子（用四个黑色的脚印表示）四个排成一行地前进。在狭窄通道内部，虽然仍是四个分子，但是这次它们排成两行，每行两个分子，因为这个狭窄通道的宽度只有前后通道的一半。现在来考虑下一时刻的情景，这里用白色脚印代表。在狭窄通道前后的通道里的分子都比前一时刻前进了一步。在狭窄通道内部的分子也必须前进，因为在稳定的气流中不会出现阻塞。这就意味着在狭窄通道里排成两行的分子与前一时刻相比必须前进两步。也就是说，为了保证稳定向前的移动，要求在狭窄通道前后的通道中有同等数量的分子向前移动。但是在狭窄通道内部，分子必须移动得更快。因此，在通过狭窄通道时气流的速度会比进入之前和通过之后更大。

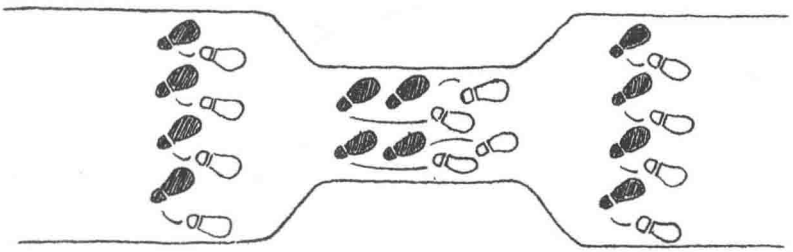


图 5.7 空气在层流中通过狭窄通道的速度。图中的脚印代表分子。

这个现象是相对比较容易解释和理解的。但是有一点看上去似乎与上面我们分析的现象相矛盾，就是狭窄通道内部的气压比狭窄通道前后的气压要低。并且，这似乎有悖常理，因为人人都知道，在交通堵塞的时候，狭窄通道内部的拥挤程度（因此压力同样）看上去是最高的。

但是这样的比较是不恰当的。例如我们可以观察到，当拥挤的车辆从两条车道并入一条车道然后又回到两条车道时，拥挤是从狭窄通道之前开始的，而且最拥挤的地方也是在狭窄通道之前。在狭窄通道内部，尽管车辆移动得很缓慢而且靠得更近，但车辆确实是在往前移动的。在狭窄通道过后，车辆就会散开并且加快移动。层流中的情况与上面的例子有所不同，因为在狭窄通道之前、之后和其内部的分子密度是一样的，而且分子在狭窄通道内部的速度比通道之前和之后的要快。因此，拥挤会在狭窄通道中产生并导致最大压力的这种经验性认识，与气流通过一根管子的狭窄部分的情况是不一样的。那应该怎样解释这种现象呢？

我们之前已经提到，尽管分子的运动事实上是杂乱无章的，但总的来说它们是按一定的速度向一个方向移动的。它们会前后、左右或者上下移动。而全部空气分子的移动速度和方向是指这些零散移动的总和。我们把这种现象与通过街道的嘉年华游行队伍进行类比。每一个演员都自由地四处走动，他们甚至有可能高高跃起，但他们都仍然继续朝着一个方向在移动。一些在游行队伍边缘的演员有可能会突然冲入观众群，这就是他们对游行路线边缘所施加的压力。同样地，空气分子会冲撞管壁，所以也会对管子的内壁产生压力。当嘉年华游行队伍的移动速度加快时，参与游行的个体向左或右移动的时间就会减少，因此游行队伍边

缘处的个体冲入观众群的次数就会减少。通过狭窄通道的气流与这个情况完全相同。单个分子沿某一个方向的移动速度加快了，所以其向上、向下、向左或向右的移动次数就会减少。因此它们冲撞管壁的次数也会减少，所以对管壁施加的压力也会降低。在狭窄通道内部的分子加快向前移动的速度会导致通道内部的气压比通道之前或之后（也就是声带以下和声带以上）的气压要低。也就是说，通过狭窄通道的气流的压力实际上并没有增大而是减小了。这个现象被称为伯努利效应。由于这是流动气流的一个特点，所以被称作一种**空气动力学效应（aerodynamic effect）**。因此，由于两片声带中间的压力相对较低，声带才会被吸在一起，而不是相互分开。

图5.8简单地展示了空气动力伯努利效应的原理。在图中，两片纸分别悬挂在两根临近的铅笔上，如果向两片纸之间吹气，两片纸就会彼此贴近。这表明由于空气在两片纸之间快速通过，通道中间的气压降低了。因此，这两片纸彼此接近，而不是像气压增大那样被分开。

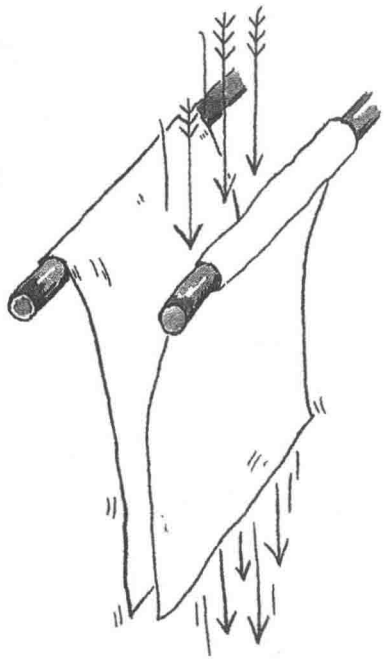


图5.8　伯努利效应的演示。当空气流过两片纸中间时，两片纸向彼此靠近。

5.2.2.2 *声带振动的肌弹性理论*

用空气动力学的伯努利效应来解释声带振动行为是理解声带振动机制的一项重要突破。运用伯努利效应，我们可以解释为什么尽管从肺部流出的气压会把声带冲开，但两片声带还是可以如此迅速地闭合。通过空气动力学来解释声带振动，涉及以下几点：声带最初是闭合的，然后它们被声门下气压吹开，最后它们因为伯努利效应被吸在一起。一旦声带闭合，这个开合循环又会重复开始。但这种解释至少有一个缺点：根据这种观点，声带振动的速率取决于气流的速度，也就是说声带振动的速率取决于声门下和喉上空气压力的关系。由于声门下气流是相对恒定的，所以声带振动的速率也应该保持相对的恒定。但是，在言语产出的过程中，声带振动的速率会不停地改变。因此还需要其他的机制来对声带的振动作出全面合理的解释。

影响声带振动速率的第一个重要因素是声带的长度：长的声带比短的声带振动得慢。这与弦乐乐器的原理相似：大提琴的琴弦比小提琴长，所以产生的音调比小提琴的要低。由于男性的声带通常比女性的长，所以男性的嗓音通常比女性的要低。我们可以在青春期时观察到声带长度影响振动速率的例子。在青春期，男孩的声带长度（和厚度，见下文）变化很快。在这段时期内，他们还没有学会怎样正确地控制变化了的生理结构，这便是所谓的“变声期”。

影响声带振动速率的第二个因素是声带的弹性。即使没有伯努利效应，声带在打开之后也会因为其本身的弹性回缩力而恢复原位。这与被拨动的吉他琴弦一样：一旦琴弦离开了静止位置，它会在弹性回缩力的作用下返回到该位置，然后会向相反的方向移动，如此往复。这种声带振动模式会在声带拉得很紧、发音人使用音高非常高的假声时发生。而在这一过程中，并没有伯努利效应的参与。

此外，声带振动的速率取决于它们的弹性张力：拉紧的声带比松弛的声带振动得更快，因为它们会被更大的力拉回到静止位置。对于吉他的弦来说也是这样：如果弦被拉紧，振动得就会快一些（也就是，每个单位时间内振动的次数更多），因此会产生更高的音调——实际上，人们正是利用这种效应给吉他调音。声带的紧张程度可以通过由环甲肌控制的环状软骨和甲状软骨之间的相互运动来调整。但是转动会导致甲状软

骨的凹槽和杓状软骨之间的距离增加，这会使声带被稍稍拉长，进而使声带振动的频率下降。不过声带紧张程度的增加要占主导作用，所以最后这两个软骨旋转运动的综合结果还是声带振动频率的增加。在言语产出过程中，声带振动速率的微调就是通过这种机制实现的。

影响声带振动速率的第四个因素是质量，这与声带的“厚度”有关。这同样也可以与弦乐乐器相类比：当吉他弦的长度和拉紧程度保持不变的情况下，一根较粗的弦比一根细的弦产生的音调要低。同样地，较厚的声带的振动速率要比较薄的声带低。参与振动的那部分声带的厚度在一定程度上是由生理因素决定的，但是也可以通过声带本身的肌肉组织加以额外的调节。声带可以从较厚的唇形变成很薄的带状，这样就改变了它们振动的速率。

上述这四种把肌肉性声带组织的振动速率与声带的长度、弹性、紧张程度和质量联系起来的效应被称作**肌弹性效应（myoelastic effect）**。

到目前为止，我们谈到的影响声带振动的因素包括以下四个方面。

1. 伯努利效应使声带可以在正常振动的过程中闭合。这种空气动力学效应取决于气流的速度。随着气流速度的增加，与气流方向相垂直的方向上的气压就会降低。由于气流速度又取决于声门下和喉上的气压差，所以声带的振动速率也受声带上下气压差的影响。
2. 较长的声带的振动速率比较短的声带要慢。声带的长度差异主要取决于生理因素。
3. 声带在拉紧时比松弛时振动得更快。声带的紧张程度主要取决于环状软骨和甲状软骨旋转或者类似摇摆的运动。通过控制声带的紧张程度，人们可以在日常言语产出时对声带振动的速率进行微调。
4. 厚的声带比薄的声带振动得慢。声带的厚薄程度一部分由生理因素决定，但同时也可以通过肌肉本身来进行调节。

把声带看成一个空气动力学和肌弹性相结合的系统，为声带振动的原理及声带振动速率的变化提供了一个合理的解释。然而，至少还需要另外三个理论才能完全地解释声带实际的振动特征（Broad，1973）。如果把声带想象成一个弹性系统并受到伯努利效应的空气动力作用，它们就如同处在通风良好的过道里的两扇不断开合的门。正如前面提到的：空气使这两扇门一开一合，而铰链处的紧张程度以及门本身的质量和大小

决定了摇摆的速率。处在通风良好的过道里，只有当铰链处弹簧的紧张程度很高时，摇摆的门才能完全闭合；事实上，两扇门会周期性地前后摆动，但永远不会完全关闭。但是这与声带振动的情况并不一样：当声带振动时，左右声带会在相对较长的一段时间内以及相对较长的振动范围内保持相互的接触。这与不断摆动的门完全不同。下面的小节会介绍可以解释这种差异的一些其他的理论。

5.2.2.3 *声带振动的双质量理论*

最近的研究表明，与摆动开合的两扇门不同，声带的结构并不是一个不可再分的整体，正因如此，声带才可以振动。事实上，声带的上部和下部可以实现相对独立但却又相互关联的运动。图5.9展示了详细的过程。首先，肺部气流的压力迫使声带的下部分开，但是声带的上部此时仍然保持闭合（图5.9b）。只有在稍后的阶段，声带的上部才会被打开（图5.9d）——部分是受气压的作用，部分是因为它们随着声带下部组织的打开而被拉开。受自身质量加速度的影响，声带的上部的打开程度会继续变大，但与此同时，声带的下部已经开始相互靠拢（图5.9e）。再过一会儿，声带的下部已经完全合拢，但是上部仍然还处在逐渐闭合的阶段（图5.9g）。声带的上部和下部都在进行开合动作，但是它们开合的时

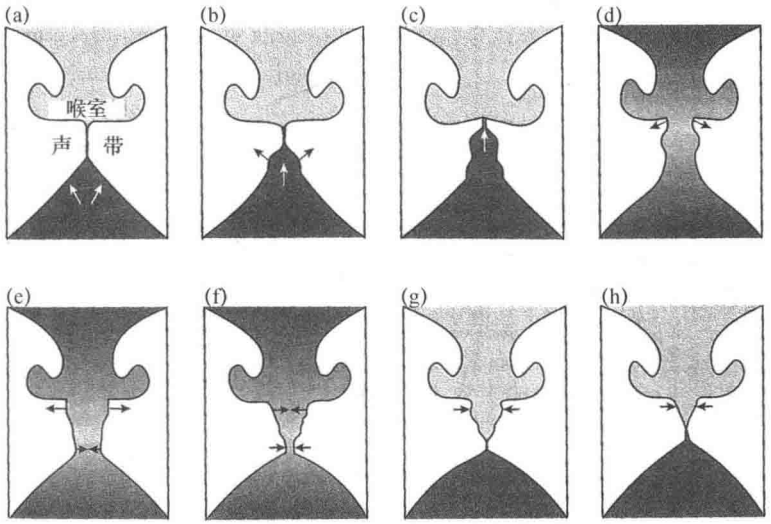


图5.9 声带振动的不同阶段。这里显示的是简化的喉部纵剖面图（细节请见5.2.2.5）。在每一幅图里，气管位于底部，咽位于顶部。

间段并不一样：声带的上部的动作滞后于下部的动作。正是声带的上部的动作的滞后使得声带可以完全闭合，遮住气管。这个原理被称作声带振动的**双质量理论（two-mass theory）**（Ishizaka 和 Flanagan，1972）。这并不是说只有两个质量参与了运动，可以是任何的数量，但两个质量是能解释这种运动行为的最小数量。

5.2.2.4　*黏弹力理论、体层被覆层理论与流动－分离理论*

上文介绍的各种不同的机制可以较好地解释声带的振动原理。但是数学建模研究揭示，空气动力学和肌弹性效应再加上双质量理论都不能充分解释振动的细节问题。声带振动还有另外两个特点。首先，声带的外缘是由很多层不同的组织构成的，这些组织共同构成了黏稠的表面（见图5.10）。因此，声带的表层对气流的反应和声带里层的肌肉主体对气流的反应不一样，当声带在气流中振动的时候，声带的表层处于“飘动”的状态。这与晾衣绳上随风飘动的床单类似，声带外缘的这种运动会产生一种拉力，从而使声带得到某种程度上的拉伸。这种效应被 Broad（1979）称为**黏弹力（muco-viscose）**理论，后被 Hirano（1974）描述为**体层被覆层（cover-body）**理论。

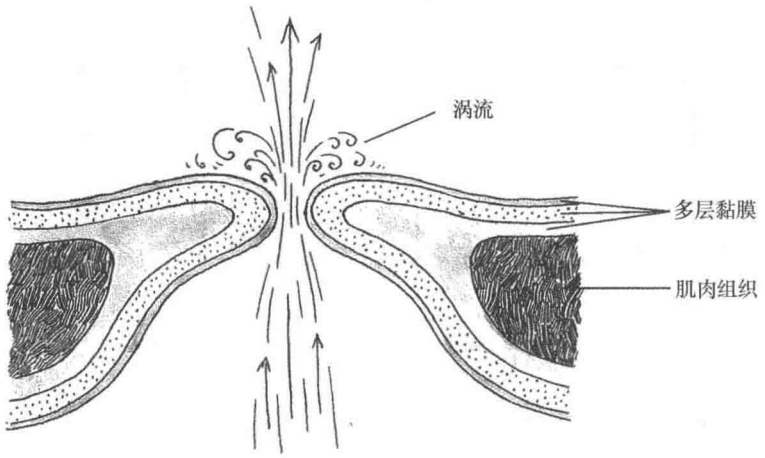


图5.10　气流通过声门时声带的唇状部分的特写

其次，**流动－分离理论（flow-separation theory）**（Ishizaka 和 Matsudaira，1968）指出，在声带外缘处发生的气流的突然改变会导致一定数

量的湍流（也称为**涡流，eddy**）。这与秋天有时能在屋角看到的涡流相似，涡流会使落叶围成圈打转。这些很小的涡流会迫使两片声带的唇状部分分开，进而影响声带的运动情况（见图 5.10）。如果没有这些力的作用，声带就不会具有其独特的振动模式。

5.2.2.5 *声带振动的一个循环*

总结起来，下面这些因素可以完整地描述声带的振动。

1. 空气动力效应解释了声带为什么可以在正常振动的情况下迅速地闭合。
2. 肌弹性效应解释了声带为什么可以被打开，以及声带为什么可以在伯努利效应不适用的情况下仍然保持振动。
3. 双质量理论解释了声带为什么可以完全地闭合。
4. 黏弹力理论和流动 - 分离理论解释了声带振动的细节特征。

总的来说，声带振动的一个循环可以采用以下的描述（见图 5.9）。开始时，声带因为内收的杓状软骨完全闭合（a）。声带因为自身的肌肉力量以及环状软骨相对于甲状软骨的位置而稍稍拉紧。肺部气流的压力把声带的下部推开，而此时声带的上部仍然合在一起（b）。当声带的上部被打开时（c），气流开始从声带之间流出。这时伯努利效应开始发挥作用，使两片声带被拉向一起，但是在声带自身质量的惯性作用下，声带的上部仍然处在继续分开的过程中（d）。在伯努利效应和声带的弹性回缩力的作用下，声带的下部开始逐渐内收，但是在质量的惯性以及其黏弹力结构（黏弹力结构在流动 - 分离理论提到的涡流的作用下使声带的外缘分开）的作用下，声带的上部仍然在逐渐地分开（e）。声带的下部最后几乎闭合（f），这意味着此时的伯努利作用非常强，把声带的下部紧紧地拉到一起。声带的上部在弹性回缩力和声带的下部的拉力的作用下逐渐靠拢（g）。（注意这种靠拢不可能是伯努利效应的结果，因为这时气流已经被阻断。）最终，声带的上部也闭合，下一个循环开始（h）。

5.2.3 响度与喉部信号

受上文所描述的声带振动机制的影响，从肺部呼出的气流在声带处会被频繁地打断。由此所产生的喉部信号是所有浊音的基础。也就是说，喉部信号就是**声源信号（source signal）**，声源信号会在之后的传播中被

声道滤掉一部分，这就像有色玻璃过滤阳光一样。喉部信号是影响说话人声音响度大小的原因之一。要增加声音的响度可以采用两种方式：首先，可以增加声带的打开程度，这样可以增大流过声带的气流量；其次，可以增加气流的压力，这样在一个声门开合循环中，可以有更多的空气分子被推送过声门。大多数人可能会选择第二种方式来增加声音的响度。以上这些增大声音响度的方法有两个主要的缺点。首先，这些方法需要消耗更多的气流和能量。其次，由于有更多的气流以较高的压力通过声带，声带的紧张程度会加大，这使声带更容易变干，导致声音嘶哑。其实还有一种更有效的增加声音响度的方法。

响度的大小也取决于声带开合的活力（vigor）。这与连着水管的水龙头的表现相似。如果水龙头被慢慢地关紧，流过水管的水量会逐渐地减少，直到水龙头被完全关紧。但是，如果水流被突然中断（就像一些洗衣机所做的那样），水管就会产生一个突然的抖动。这种抖动不是取决于水流量的多少，而是取决于水流改变的突然程度，通常是关水比开水快。这与声门信号的情况是一样的：语音信号波形中最大的峰值就是由声带的快速闭合产生的，而不是由它们的（较慢的）打开产生的。在声带振动的过程中，声带开合的活力会影响语音响度的大小。声带闭合的活力可以通过增加声带的弹性力来增加。增加声带闭合的活力并不一定会使每秒内产生更多的振动周期，因为声带可以增加保持闭合状态的时长，从而不改变一个完整的合－开－合循环的整体时间（见图 5.11）。

因此，较高的声带开合速度（不是振动的速率）会使感知到的响度增加。受过训练的发音人和歌手在发音过程中大概会利用到这个原理。因此他们喉部的信号看上去不像图 5.11a，而是像图 5.11b。增大声带开合速度的另外一个结果是声带保持打开状态时间的减少。这意味着尽管响度增加了，但是需要的气流量却变少了，因此需要的能量也就变少了。人们在强调一个单词内的某一个音节时，会在不知不觉间利用这个效应（Sluijter, 1995）。另外，对重音程度的细微调节也可以通过调节声带开合的速度来实现。

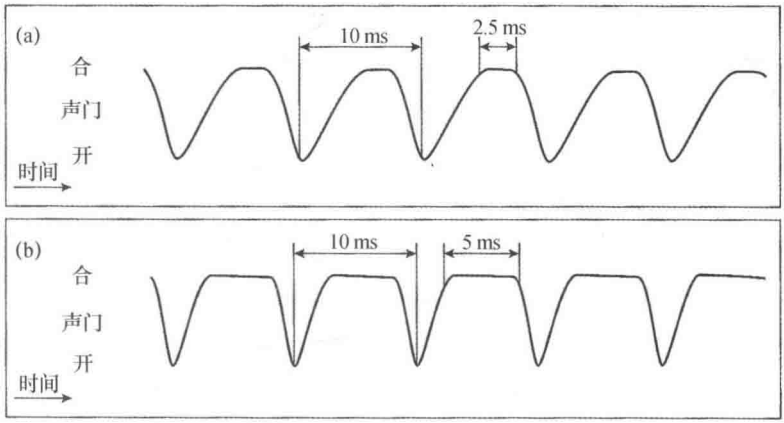


图5.11 声带较慢地闭合（a）与较快地闭合（b）时，声门打开的程度。注意在（a）和（b）中的合－开－合循环的时长是一样的，但是在（b）中，声带保持闭合状态的时间更长。

5.2.4 声区

声区（register）的概念与声带振动的不同方式有关。关于这个概念的描述和分类有很多不同的甚至是相反的意见（见 Laver，1980：93－95）。在本书中，我们只描述与言语产出相关的三种**发声类型（types of phonation）**（见 6.2 节）。

第一种，声带的“正常”振动被称为**胸声区（chest register）**或者**常态嗓音（modal voice）**。这就是前面几页所描述的声带振动。

如果杓状软骨没有完全内收，在一个声门的开合循环中，声带并不是完全闭合的，总会有一条狭窄的缝隙。这个狭窄的缝隙会激发伯努利效应使气流快速流过。因此在声门处会出现摩擦的噪声，使语音听起来粗而沙哑，具有**气嗓音（breathy voice）**的音色。这种音色可能与发音人自身的发音特点有关，也可以由感冒引起，但是在某些语言中，一些语音就采用这样的发声方式（见 6.2 节）。

当声带完全松弛、变厚时，它们就像面对面悬挂的两个口袋。在这种状态下，声带并没有真正地振动。而事实上，气流像气泡一样“冒”过松散而紧邻的声带，产生一种**紧喉嗓音（creaky voice）**，这时声带振动得非常缓慢（每秒振动 30－50 次）。在这种振动模式下，不会产生伯

努利效应，因为两片声带并没有分开。这种紧喉嗓音有时会出现在一些发音人或一些方言的浊音话语的末尾。而在有些语言中，这种紧喉嗓音被视为一种音位。

总的来说，声带的振动可以被描述为一个复杂的三维的运动过程。声带是以一种波浪状的运动方式自下而上、从后往前地逐渐打开，然后再自上而下、从中间同时往前往后地逐渐闭合。下面的这些参数对声带的振动很重要。

1. 杓状软骨可以使声带打开或闭合，所以我们可以在言语产出的过程中产生带音和不带音。而只有杓状软骨内收的时候声带才会振动。
2. 环状软骨和甲状软骨之间的相对摆动可以改变声带的紧张程度。这样就可以快速并准确地调节声带振动的速率。
3. 声带的质量和紧张程度的改变可以使声带产生各种不同的振动类型。

婴儿的声带每秒大概振动（打开和闭合）400 次，年轻女性每秒大概 220 次，年轻男性每秒大概 130 次。随着年龄的增长，女性的声带振动速率会降低，而男性的则会增加，这样性别之间的差异就会逐渐减少。

声带的运动非常复杂，然而这种复杂性对言语的产生却非常重要。但问题是，虽然这个复杂的声源信号的响度和音高是可变的，但它却只有一种音质：它与其他的任何语音的音质都不相似。然而，这个在声带处产生的声源信号却是构成所有浊音的声学原材料。正是因为这种原材料“内涵丰富”，声源信号才可以在喉上系统的过滤作用下，使信号的特定部分得到加强或减弱，最后再从唇和/或鼻辐射出去。下一节将介绍声道的解剖特征。

5.3　声道

声道大致可以分为三个部分：咽、鼻腔通道和口腔通道（见图 2.1）。每一部分又可以分为许多子结构，下面的小节将进行详细的介绍。与肺和喉相比，声道包含更多的骨骼、韧带和肌肉，所以会有更多与之相关的术语，但是这些术语对理解声道的作用并没有太大的帮助。因此，本

节对这些细节的描述进行了压缩，不过依然力图向读者较为完整地介绍与言语产出过程有关的结构。

5.3.1 咽

喉咽（laryngopharynx）是喉的顶部与食管交汇的区域。这部分的一个重要的结构是会厌（见5.2.1），它构成了喉的上部。会厌最上面的部分标志着喉咽部分的顶端，再向上的部分被称为**口咽（oropharynx）**，它构成了口腔通道的后部。**鼻咽（nasopharynx）**是位于软腭上方的区域，即鼻腔通道的后部。从某种意义上来说，咽是一个像拉伸的“X”形一样的四向交叉口，在这里鼻腔通道和口腔通道相交然后向下分成喉和食道。所有这四个通道可以同时打开。软腭和舌后部可以分别打开鼻腔通道和口腔通道，会厌和声带可以打开从喉到气管的通道。食道只有在吞咽食物的时候才会被打开。实际上，吞咽动作是一个复杂的过程，尤其是食物必须得经过咽部的交叉口从口腔通道运进食道。正常情况下，为了使呼吸畅通，从鼻腔通道到喉的通道是保持打开的。食物的传送和喉通道的关闭是通过三块**括约肌（constrictor muscle）**以及腭咽肌、茎突咽肌和咽鼓管咽肌来实现的。三块括约肌位于咽部周围，不会阻塞口腔通道，其余的三块肌肉则共同构成了咽的内部纵层。在吞咽和说话时，这后三块肌肉会抬高喉部并且缩短口咽。这对发某些塞辅音时产生的喉头气流特别重要（见6.1.1）。

如果口咽被拉长，那么它自然就会变窄。这一过程可以通过使舌根后收或者拉紧括约肌来实现。当口咽通道变得足够窄时，就可以在这个部位产生擦音。如4.1.4中提到的，很少有语言使用这样的擦音，但是将咽化作为一种次要发音方式还是比较普遍的。

5.3.2 鼻腔通道和软腭

咽的上部是由融合到鼻腔通道的鼻咽组成的。每个人的鼻腔通道的大小和形状都会有很大的不同（Bjuggren 和 Fant，1965；Dang 等，1994）。而且鼻腔通道是不可移动的，仅靠自身并不能产生语音。但是软腭的降低或抬升可以使鼻腔通道与声道的其他部分相通或相隔，这样可以使鼻腔通道对改变语音信号的声学特点起到一定的作用（见10.2.4）。鼻

(塞)辅音(见2.3.2)、鼻元音(见3.4.2.2)和鼻化元音(见3.4.1.2)都具有鼻音特征。

鼻腔通道是由黏膜覆盖的骨结构。**鼻中隔(nasal septum)**将鼻的通路分成两个**鼻孔(nares)**,末端经**外鼻孔(nostrils)**软骨与大气相通。鼻中隔通常是弯曲的,这导致两个鼻孔并不对称。在鼻中隔后部,两个管道止于伸入鼻咽的**内鼻孔(choanae)**。每个鼻孔的后部被分成三个与骨质**鼻甲(concha)**相隔的通道(**鼻道,meati**)。大多数空气会流到最底部的那个通道中,而下鼻甲将这个位于底部的通道与其他通道隔开。下鼻甲由很多层含有血管的黏膜覆盖。当覆盖在鼻甲上的黏膜受到感染时,它们就会发生肿胀。例如在感冒的时候,黏膜的肿胀将会引起鼻子堵塞。鼻内部所有的表面都被黏膜所覆盖,在几种腺体的共同作用下,它们可以保持湿润。部分腺体位于几对腔体内,而在产生鼻音时这些腔体会产生共鸣。弯曲的鼻道有一项生理功能,它可以使呼吸时吸入的气体变暖。从某种程度上讲,鼻腔形成了一种类似空调的系统,可以阻止干、冷、热或者脏的空气到达敏感的肺泡。鼻腔与口腔相连,但是鼻腔的阻尼很大,这样的声学系统无法使太多的声音能量从鼻孔辐射出去。

通过降低软腭可以打开**腭咽闭合口(velopharyngeal port)**使鼻腔通道与口咽相连。软腭并不是一个简单的高低升降的“门”。事实上,软腭的运动是一个复杂的过程。我们可以通过在镜子前交替地发口音和鼻音(如[a]和[ã])来观察软腭升降的过程:软腭向前鼓起,改变其厚度,从而被抬高(或者降低)。软腭的升降通常还伴随着口咽侧壁和舌后部的共同运动。

总的来说,鼻腔通道的大小和形状具有很大的个体差异。血管或者肿胀的黏膜可以使通道变窄,而侧面的腔体可能通畅,也可能被黏液填满。此外,腭咽闭合口的打开程度可以通过软腭降低的程度来控制,而软腭降低的程度也反过来会影响软腭的形状和舌后部被拉伸的程度。

正如前面提到的,当软腭下降时,气流通过鼻腔流出产生的语音通常被称为鼻音。因此,如果要确定语音信号中鼻化度(nasality)的大小,似乎只需要简单地测量从鼻腔流出的气流或者腭咽闭合口的打开程度就可以了。但是如果说话人的鼻子堵塞(因为感冒或者用手指堵住鼻孔),那么其产出的语音听起来的鼻化度是非常高的。在这种情况下,完全没

有气流从鼻腔通道流出。类似地，一些语音尽管在发音时腭咽闭合口打开的角度较大，但其鼻音程度却比腭咽闭合口打开角度较小的语音要低（Maeda，1993：156）。因此，对于鼻化度的定义不能单靠测量鼻腔通道打开程度或者通过鼻腔的气流量，而需要依赖一种不同的测量方法。

第一个改进是考虑从鼻腔通道和口腔通道流出气流的关系。如果口腔通道的开口较大，那么从鼻腔通道流出的气流对发音的影响就较小；如果口腔通道开口较小，则从鼻腔通道流出的气流对发音的影响就较大。因为当口腔通道开口较大时，更多的声音能量是从口腔通道释放的。通过计算鼻腔通道和口腔通道气流的相对关系，我们可以用鼻腔通道气流相对于口腔通道气流的百分比来估算出语音信号的**鼻流量（nasalance）**：

$$\%\,\text{鼻流量} = 100 \times \frac{\text{鼻腔通道气流}}{\text{鼻腔通道气流} + \text{口腔通道气流}}$$

这就解释了为什么产生一个低元音（例如［ɑ］）时，尽管腭咽的打开程度较大，鼻腔通道气流也比较多，但听起来却没有鼻音成分（因为口腔通道的气流很大）；而产生一个高元音（例如［i］）时，口腔通道开口较小，鼻腔通道的气流也比较少，但听起来却更像夹带着“鼻音”。

第二点，上文提到通过阻塞鼻腔通道可以导致产出的语音中带有鼻音成分。这个效应可以通过考察鼻腔与口腔通道的偶合关系进行解释（见10.2.4）。即使没有气流从鼻孔流出，鼻腔依靠其较大的容积也可以较明显地改变声道的共鸣腔系统。从喉部释放出来的“声音能量”有一部分流入了鼻腔，而没有通过口腔通道流出。即使没有气流从鼻腔流出，鼻腔也会弱化从口腔通道释放出的声音能量，从而导致语音听起来口音成分更少而鼻音成分更多。总的来说，腭咽闭合口的打开使得语音听起来带有鼻音音色，但是这必须与声道的整体结构联系起来看。如果鼻腔通道没有被堵塞，我们可以通过测量鼻腔通道的气流量来估计腭咽闭合口的开口大小。进而，只要声道其他部分的结构保持不变，我们也可以由此推知语音信号的鼻化度大小。例如，通过测量多个［i］的气流状况，我们就可以知道哪个［i］的鼻化度更大，但是在［i］上测量到的气流量结果不能简单地与［a］产出时的气流量相比。

5.3.3 口腔通道

我们在2.2节中介绍了说话所需要的主动的和被动的发音器官。现在

我们将补充介绍在言语产出过程中所涉及的口腔通道的一些解剖学知识。

口（包括双唇）是一个非常灵活的三维肌肉结构——在面部至少有12块肌肉与口区域的构成有关，其中部分肌肉是成对的。这里不会逐个列出所有肌肉及其活动状况，但是会强调一个现象：双唇可以改变自身的形状，例如从较突出的圆唇变为展唇。有一点应该引起注意，任何唇部活动都会涉及好几块肌肉的协作，其中的一些还是较大的肌肉。所以，唇部的运动会比声道中其他参与言语产出的肌肉慢一些，这仅仅是因为唇部的运动需要更大范围的肌肉组织参与。

口腔通道内另外一个比较重要的结构是下颌，它可以在**颞下颌关节（temporomandibular joint，TMJ）**处滑动和转动。这个关节位于耳垂的前面，在张口和闭口的时候可以用手指触摸感觉到。当用手指触摸时，我们可以感受到下颌的打开并不是简单地在关节处转动，而是下颌向前向下的移动。颞下颌关节内的**关节盘（articular disc）**将它分为上部和下部，上部主要进行滑动，而下部则主要进行升降运动。

通过降低下颌来张口，主要是受重力的作用：下颌自身的重量会将颌往下拉。这个动作是由舌骨上方的**舌骨上肌（suprahyoid muscle）**和舌骨下方脸颊两侧的**舌骨下肌（infrahyoid muscle）**共同完成的。因此，在这些肌肉作用下，下颌的主动打开会将喉部上提，因为这些舌骨肌的收缩会减少下颌与喉部之间的距离。

口的闭合主要是由**咬肌（masseter）**、**颞肌（temporal）**和**翼外肌（lateral pterygoid）**来实现的。咬肌从下颌的最下端一直延伸至下巴。颞肌从颞下颌关节起横跨头骨的整个侧面，是头两侧面积很大的肌肉。如果牢牢地将口闭合，便能很容易地看到并触摸到这两块肌肉的运动。使口闭合需要用到的很大（而且很强）的肌肉，与使口打开的重力并不对称，这种不对称性表明下颌的主要功能是咬碎食物。对言语产出而言，这种不对称性似乎不起作用。也就是说，使一个发音器官从闭合变为打开的过程与从打开到闭合的过程所用的时间并无不同，因此这也无法与肌肉的强度联系起来。换句话说，在一个 C_1VC_2 的序列中（例如，任意一个辅音后面跟有任意一个元音，其后又跟有其他任意一个辅音），从 C_1 到 V 的转换时间（从闭到开）并不会比从 V 到 C_2 的转换时间（从开到闭）长。这种在发音中缺少不对称性的现象实际上是由另外一个因素决

定的：口腔通道自身的打开和闭合主要是通过舌而不是下颌来实现的。

在口腔通道中，舌是言语产出过程中最重要的肌肉系统。舌是一个由很多肌肉组成的非常复杂的系统。舌的表面覆盖有一层黏膜，可以使所有肌肉一起协同运动，因此我们很难独立描述某个肌肉对舌运动的影响。大致来讲，有两组肌肉群：（1）外部肌肉：**颏舌肌（genioglossus）**、**舌骨舌肌（hyoglossus）**、**茎突舌肌（styloglossus）**和**腭舌肌（palatoglossus）**；（2）内部肌肉：**舌上纵肌（superior longitudinal）**、**舌下纵肌（inferior longitudinal）**、**舌横肌（transverse）**和**舌垂直肌（vertical）**。每组肌肉群内的肌肉都成对存在，分别位于舌的左部和右部。简单来讲，外部肌肉是较大的肌肉，它们使舌下压或后缩；内部肌肉是较小的肌肉，它们可以改变舌的形状：使舌尖或者舌两侧的边缘上卷（舌上纵肌、舌下纵肌），使舌变窄变长（舌横肌），使舌变平变宽（舌垂直肌）。

外部肌肉和内部肌肉在质量上的差异会使舌不同部分的运动速度有所不同，舌尖运动最快，舌根运动最慢。我们来通过下面这个例子观察在产生齿龈爆发音时语速快慢对发音器官运动情况的影响。当语速快时，舌尖首先向后稍微卷起，然后再向前移动来产生齿龈爆发音。当语速慢时，舌尖基本沿直线移向齿龈。发音运动过程中的这种不同引发了人们对于发音运动策略的讨论，讨论认为发音的运动策略可能因语速的快慢而产生变化（如，Payan 和 Perrier，1997；McClean，2000），这对言语的大脑表征有着深远的影响。也就是说，如果不同的语速需要不同的运动指令序列，那么在大脑中储存的“言语”就必须含有很详细的发音细节，包括对快语速和慢语速情况下的不同的运动指令。如果快语速与慢语速情况不需要不同的运动指令，那么大脑中的“言语”可能不包含具体的发音细节，而是以一种更加抽象的形式进行储存的。一些研究者认为，我们可以通过不同的运动质量来解释快慢速的不同运动轨迹：舌尖质量轻，所以可以快速移动；而舌体则比较重。要达到齿龈脊，舌必须向前移动而且舌尖必须抬高。对慢语速来说，这些移动较慢，这样舌体和舌尖可以以相同的速度向目标位置移动。在快语速时，较轻的舌尖（稍微往上往后卷起）比将舌尖缓慢向前推至目标位置的较重的舌体更快地达到目标位置。也就是说，不同语速下的不同的运动轨迹不是由不同的运动引起的，而仅仅是由舌的不同质量在加速运动时所产生的惯性所致。

在不同语速条件下是否存在不同的发音规划，或者不同的运动轨迹是否只是移动的质量不同所导致的结果，现在都还没有定论。但是，对发音运动的研究和解释对于探索言语在大脑中的存储和加工机制的模型有所帮助。

与鼻腔相似，口腔通道本身也存在着很大的个体差异。口腔的大小因人而异，而且硬腭可以是扁平的或者半球形的，其形状也会有很大的差异。此外，每个人舌的灵活程度也各不相同：有些人可以很容易地使舌的中间形成一个凹槽，但是另一些人却做不到。这对语音的产出有直接的影响。同样的一个发音位置，对一个人来说可能比较“舒服”，但另一个人可能就觉得很“不自然”。尽管两个人都可以产出同样的语音，但他们却很可能是通过不同的发音位置来实现的。

总的来说，看似简单的发音动作，比如降低软腭，可能会涉及肌肉系统非常复杂的交互作用，从而产生相当复杂的声学结果。而且对一个人来说很“自然”的发音动作对另一个人来说也许会是非常困难的。

以上介绍了发音器官的生理特征，现在我们需要研究言语产出的功能。与这章介绍气流、喉和声道的顺序一样，我们将首先在第 6 章介绍不同的气流机制。在本章所介绍的概念基础上，我们在第 9 章会介绍喉与声道的交互作用。为了理解喉和声道的作用效应，我们需要了解更多的声学知识，学习应该怎样测量声学效果，并探索共鸣的起因和结果。在更详细地描述言语声之前，第 7 到第 9 章将为我们介绍上述这些知识。

练　习

1. 请描述在吸气和呼气过程中，肋间外肌、肋间内肌、膈膜以及弹性回缩力的作用。
2. 在声带振动和语音产出的过程中，环杓后肌和环杓侧肌起到了什么作用？
3. 为什么振动弦理论和神经时值理论不能完全解释声带振动的原理？
4. 构成肌弹性效应的四个因素是什么？这些因素是如何影响声带振动速率的？
5. 增加响度的三种不同的方法是什么？最有效的方法是哪一种？为什么？

6 气流机制与发声类型

介绍了呼吸与喉部系统的基本情况之后，我们便可以采用一种与英语不同的方式来描写涉及这些系统的语音类别。特别是一些语言中的语音是由非肺部气流产生的，有一些语言还存在其他的发声类型，这些使得人类语言的声音清单得以扩展，而不仅是清音（带音）和浊音（不带音）之分。

6.1 气流机制

语音产生过程中有三种基本的气流机制：**肺部气流（pulmonic airflow）**、**喉头气流（glottalic airflow）**和**软腭气流（velaric airflow）**。肺部气流产生于肺部，是言语产生的基本气流，在正常呼吸中，吸气（内吸气流）和呼气（外呼气流）交替进行（见5.1.2）。肺部产生的外呼气流是言语产生最为常见的方式，称作外呼肺部气流。外呼气流发生在吸气之后，这一过程不需要肌肉力量参与，弹性回缩力会自动减小肺部体积，从而使肺部气体流出至声道外部的空气中。

由外呼肺部气流产生的口塞音也叫爆发音。然而，塞音和其他阻音（obstruent）也可由非肺部气流机制产生。同发爆发音时一样，通常其口腔中也存在收紧，但气流的产生和调节方式有所不同。后面会讲到另外三种阻音，其中两种由喉头气流机制产生，一种由软腭气流机制产生。下面将详细介绍这些机制。

6.1.1 喉头气流机制

喉部（而非肺部）是喉头气流的能量源。发音时，声门处于关闭状

态，肺中气体维持在声门以下，可通过抬高或降低喉部来移动声道中声门以上的气体。

抬高喉部产生的阻音为**喷音（ejective）**。例如发齿龈塞喷音时，声门关闭，并在齿龈脊处形成一处收紧。然后喉部上抬，这好比一个活塞，压缩了喉部和齿龈收紧点之间的气体。而舌根的收缩会进一步压缩这部分气体。然后，齿龈收紧点放松，压缩的空气得以释放，发出齿龈塞喷音［tʼ］（如表6.1所示，喷音用附加符号［ʼ］表示，加在肺部气流阻音符号之后）。因此，喷音通过外呼的喉头气流产生。当齿龈处的收紧放松以后，声门处的收紧也随之放松，声带恢复振动。喷音主要出现在非洲、高加索、北美、中美以及南美的众多语言中，在一些语言中还存在擦喷音和塞擦喷音。

表6.1　非肺部气流辅音的IPA符号：喷音，浊内爆音和龈音

喷音	浊内爆音	龈音
［ʼ］如：		
［pʼ］双唇音	［ɓ］双唇音	［ʘ］双唇音
［tʼ］齿音	［ɗ］齿音/齿龈音	［ǀ］齿音/齿龈音
		［ǃ］齿龈音（龈后音）
		［ǂ］腭龈音
		［ǁ］齿龈边音
［cʼ］腭音	［ʄ］腭音	
［kʼ］软腭音	［ɠ］软腭音	
［qʼ］小舌音	［ʛ］小舌音	
［sʼ］齿龈擦音		

通过降低喉部产生的塞音称为**内爆音（implosive）**。发内爆音的初始阶段与喷音相同：声门关闭，口腔中形成收紧。之后的不同之处是此时喉部降低而非抬高，从而扩大了声门和收紧部位之间的容积。根据波义耳定律，此时气压减小，当口腔收紧放松后，空气被吸入口腔，形成内

吸的喉头气流。但是由于喉部以下气压的作用，声门在降低的过程中难以维持紧闭的状态。因此喉部在降低的过程中通常会泄漏部分气体，从而引起声带振动。实际上，喉部与收紧部位之间容积的增加不会对气压有太大改变。虽然名为内爆音，但在发音时却并没有真正的内爆动作。这种内爆的音质是源于“声道形状与声带振动模式的复杂变化”（Ladefoged，2006：117）。由于声带没有完全闭合，在气体泄出时会发生振动，因此内爆音通常为浊音。如表6.1所示，浊内爆音的IPA符号是在相应的肺部气流塞音符号（例如［d］）上附加一个弯钩（例如［ɗ］）。目前的IPA版本没有清内爆音的符号，但可用表示清化的附加符号［̥］表示，例如［aɓ̥a］（同样也可以在清辅音上面添加一个钩形符号来表示清内爆音，例如［ɓ̊］）。内爆音主要存在于非洲西部的语言中，例如谢列尔－塞因语（Seereer-Siin，塞内加尔使用的一种尼日尔－刚果语），在三个发音部位均有清浊之分，如表6.2所示。

表6.2 谢列尔－塞因语中的浊内爆音与清内爆音（McLaughlin，2005）

	双唇音	齿龈音	腭音
浊	［aɓira］他挤了奶	［aɗega］他切了	［aʄaxa］他咬碎了
清	［aɓ̥ira］他们挤了奶	［aɗ̥ega］他们切了	［aʄ̥axa］他们咬碎了

6.1.2 软腭气流机制

发**龈音（clicks）**时涉及软腭气流机制，许多非洲南部的语言中都有龈音。软腭气流完全产生于口腔，因此也叫**口腔气流（oral airstream）**机制。发音时，通常在口腔后部形成一处完全的收紧，一般在软腭，有时也可在小舌；同时前部也会形成一处收紧，使两个收紧之间夹存一团空气。随后舌体下降，两处收紧之间的容积增加。随着前部收紧的打开，外部气流涌入，产生类似咂嘴的声音，后收紧点随后打开。龈音的发音部位由前收紧点决定，其IPA符号见表6.1。口腔后部的发音动作和前收紧点的具体位置决定了龈音的伴随特征（Ladefoged和Maddieson，1996），例如龈音的产生可能会伴随软腭后部的发音、不带音、送气或气嗓音（见6.2节对气嗓音的讨论）等特征。齿龈音［ǀ］及其伴随特征可分别

表示为［kǀ］、［kǀʰ］和［g̈ǀ］。此外，腭咽部分可以打开，形成鼻伴随龈音，如［ŋǀ］。

总之，根据气流源及其方向，塞音可分为五种，即爆发音（或口塞音）、鼻音（或鼻塞音）、塞喷音、内爆音以及龈音，如表6.3所示。

表6.3　根据气流源及其方向划分的五种爆发音

	气流源	气流方向	气流机制
爆发音	肺	外呼	肺部外呼
鼻音	肺	外呼（仅通过鼻腔）	肺部外呼
喷音	声门	外呼	喉外呼
内爆音	声门	内吸	喉内吸
龈音	舌	内吸	软腭内吸

6.2　发声类型

发声这一术语主要是指声带的振动。我们在5.2节曾讨论过，喉部的肌肉结构很复杂，因此可以有多种调控声带的方式。此外，除了声带的姿态，还有很多其他因素可以决定发声类型，包括声门开口的形状和大小、气流速度以及声带是否振动等。喉部可以有多种状态，其中有五种为人类语言的常用状态，这五种发声类型或者说是声门状态请见表6.4。

表6.4　不同发声类型及其相应的声带姿态

	声带姿态	声带振动
清音	分得很开	否
耳语声	分得较开	否
气嗓音	轻微分开	是
浊音	紧密合拢	是
紧喉嗓音	后部：紧密合拢	否
	前部：松弛合拢	是

带音由声带全部或大部分长度范围的正常振动产生，也称常态嗓音，

存在于世界上所有的语言之中，详见5.2.2中的说明。简单来讲，一开始声带相互靠拢，维持一种松弛的接触，然后在肺部气流、弹力与空气动力的作用下发生振动：肺部气流将声带吹开，而弹力与空气动力又使其闭合。

清音（不带音）（voicelessness）是指没有任何声带振动的情况。此时声带分得很开，使非湍流（层流）可以通过声门。与其他发声类型相比，在不带音的产生过程中，声门的分开程度相对较大，类似于正常呼吸时声门打开的程度。

发**耳语声（whisper）**时要求两片声带更为接近。此时声带呈部分内收的状态，在其后部靠近勺状软骨处留有开口。高速气流通过该窄口时会产生湍流噪声，成为耳语声的声源，引起口腔共鸣。发耳语声需要高速气流，因此耳语声不是一种高效的说话方式。如果用耳语声去说一个长句子，很快就会将气息用尽。发耳语声时声带不振动，标音时用附加符号[̥]表示，例如[ḁ]（注意，在旧版的IPA中，该符号表示卷舌音，见4.1.2）。发耳语声时，我们无法获取那些靠声带是否振动来区分的语音。这点很好证明：你的搭档能区分你用耳语声所发的最小对立体吗（例如含清擦音的“Sue”与含浊擦音的“zoo”的耳语声）？如果可以，很可能是由于耳语的“Sue”中擦音的时长要长于耳语“zoo”中的擦音。在一些南亚和美国土著语言中，“正常”（中性）元音与耳语声元音确实会构成音位对立。

耳语声元音也称作清化元音或者清元音。日语和韩语中均有受语音环境制约的清化元音的例子。这些语言中，元音处于清塞音语境时会变为“清化元音”、“清元音”，或者“耳语化元音”。具体而言，高元音处在清阻音之间时会变为清音。例如日语区分[kika]（气化）与[kiga]（饥饿）。但这究竟是元音清化，还是元音被完全删除后听音人靠相邻音段的协同发音信息进行的“填补”，目前还不完全清楚。

气嗓音也称哞声（murmur），发音时声带微微分开（见5.2.4）。虽然在发声时声带会产生振动，但声带不会完全闭合，因此不会对肺部气流造成阻碍，产出的声音带有气嗓或者沙哑的音质。在英语中，两个元音之间的浊音[ɦ]有时会带有这种气嗓音质，如“ahead”。气嗓音的IPA附加符

号是在音标下加两个点，如［a̤］。南亚的一些语言中存在气嗓音与常态嗓音的对立，如泰国南部的尖竹汶高棉语（Chanthaburi Khmer）就区分中性元音和送气元音，如［kat］是“切”的意思，［kat̤］则表示“他或她”。

最后，紧喉嗓音也可称作**喉化音（laryngealization）**、**声门化音（glottalization）**、**吱嘎声（creak）**、**喉气泡声（glottal fry）**或**口气泡声（vocal fry）**（详见5.2.4）。发紧喉嗓音时，声带的后部紧闭，同时前部放松，从而产生一种低频的振动。当发声达到声带可维持振动的频率的下限时，就会产生紧喉嗓音。通常可以在一段较长语流的末尾处观察到紧喉嗓音。紧喉嗓音用IPA附加符号［˷］表示，例如［a˷］。南亚和中美洲的一些语言中存在紧喉嗓音与常态嗓音的对立，例如扎拉巴马萨特克语（Jalapa Mazatec），一种在墨西哥使用的奥托曼格安语（Otomanguean）就存在常态嗓音、气嗓音和紧喉嗓音的三重对立，例如［já］，［ja̤］和［ja˷］分别表示“树”“他穿”和“他携带”（Kirk 等，1993；［á］表示元音［a］为高调，见11.4.1）。

6.3　爆发音的浊(带音性)、清(不带音性)与送气

在一段元音－爆发音－元音（VCV）的语音序列中，一系列的发音事件与爆发音的产出有关，主要包括：口腔成阻，声门打开的程度和时长，口腔除阻以及后接元音声带振动的起始。这些事件之间的时长关系在很大程度上决定了爆发音的清（带音）、浊（不带音）以及送气的程度。根据以上音姿的时长，爆发音可分为三类：浊爆发音、清爆发音和清送气爆发音。下面将依次进行介绍。

1. 浊爆发音（voiced plosive）。整个VCV序列都保持带音，声带振动贯穿第一个元音、辅音的闭合段、辅音的除阻以及第二个元音。只有当闭合时间过长导致没有足够气流维持声带振动的时候，声带才停止振动。但此时声带仍保持在相同位置，并可在除阻后继续振动。带音性闭合段的时长通常较短，并且除阻与后接元音的起始之间没有延迟（见图6.1a），这种爆发音称为**全浊爆发音（fully voiced plosive）**。

2. 清爆发音（voiceless plosive）。第一个元音结束后的闭合段不发声，此时声带停止振动。这通常是由于声带彼此分开，也可能是因为声

带完全闭合并产生了一个喉塞音。但是在除阻时（或除阻后很短的时间内），声带回到振动位置，继续产生带音。通常不带音性闭合段的时长较长，除阻和后接元音之间几乎没有延迟（见图6.1b）。这类爆发音也称为**清不送气爆发音（voiceless unaspirated plosive）**或**普通爆发音（plain plosive）**。

3. 清送气爆发音（voiceless aspirated plosive）。同样，闭合段不发声。与清不送气爆发音的情况相同，声带彼此分开，在元音结束时停止振动。但在发送气爆发音时，除阻时刻，声带并不闭合。由于声带在除阻时并未完全闭合，因此在一段时间内，气流会通过打开的声门及无收紧的口腔流出。这段无阻气流会产生送气噪声，与发［h］时类似。这一送气阶段是清送气塞音的特征。通常不带音性闭合段的时长较长，并且除阻与后接元音的起始之间具有明显的延迟（见图6.1c）。

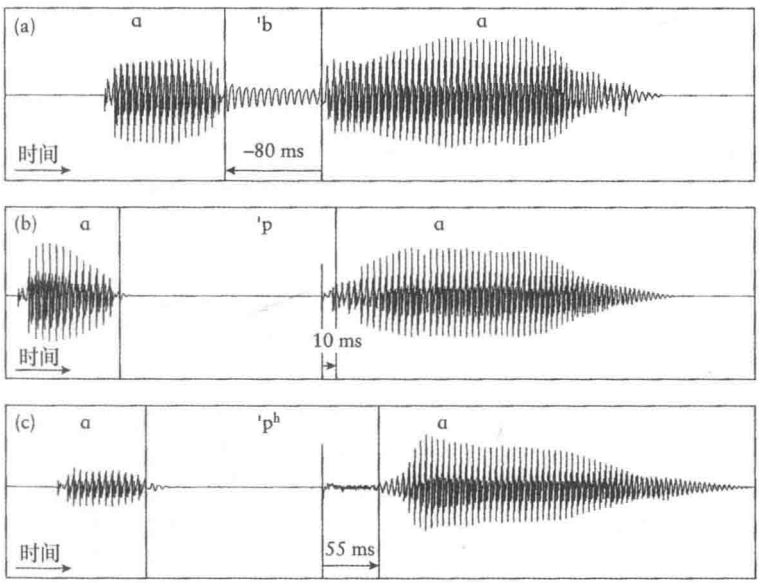


图6.1 三段发音（a）［ɑˈbɑ］，（b）［ɑˈpɑ］和（c）［ɑˈpʰɑ］的波形图。每个图中，最左边的标记表示辅音闭合段的起始点，最右边的标记表示元音嗓音的起始点。（b）和（c）中间的标记表示辅音的除阻，每个波形下面均给出了VOT的测量值。

嗓音起始时间（voice onset time，VOT）这一参数便于量化上述的一些发声与送气模式。VOT 是指从辅音除阻到嗓音起始之间的时长，用毫秒（ms）表示。VOT 涉及两个发音动作：一个是口腔动作，即辅音的除阻，一个是喉部动作，即声带开始振动。虽然 VOT 最初仅用来描述起始爆发音（initial plosive，Lisker 和 Abramson，1964），但按现今的惯例，我们将更宽泛地使用 VOT 来对爆发音进行描写[1]。

就 VOT 而言，全浊爆发音（见图 6.1a）的嗓音起始时刻先于辅音除阻。根据惯例，此时 VOT 用负值表示，描述时说这个音具有“负 VOT”（或“lead VOT”）。例如图 6.1a［ɑˈbɑ］中爆发音的 VOT 为 -80 ms。如果位于词首，这段区间可称为**前浊音（prevoicing）**段。对于清不送气爆发音（见图 6.2b），辅音除阻与嗓音起始几乎同时发生（前后相差一般不超过 20 ms），因此这类爆发音的 VOT 小于 20 ms，如［ɑˈpɑ］，被称为具有“短延迟”（短正值）VOT。对于清送气爆发音（见图 6.1c），辅音除阻与嗓音起始之间具有明显的延迟，称 VOT 为正值或“长延迟”（长正值）VOT，如图 6.1c［ɑˈpʰɑ］中爆发音的 VOT 为 55 ms。此外，通过元音开始部分振幅的衰减可以看出，送气在元音部分还延续了 20 ms。

总之，VOT 有时是区分浊爆发音（负 VOT）、清爆发音（短正值 VOT）以及清送气爆发音（长正值 VOT）的一种简单的方法。但请切记，VOT 只是区分爆发音的嗓音与送气状态的众多参数之一。最明显的一个问题是 VOT 不能描写闭合段的时长信息。不过具有负 VOT 的爆发音除外，在这种情况下 VOT 等于闭合段的时长。

VOT 常用来描写既定语言内部或不同语言之间的发声差异，比如很多语言区分全浊爆发音（VOT 为负值）与清不送气爆发音（VOT 约为 0 ms）。但是，关于清浊辅音在音系上的对立，不同语言的划分标准并不一定相同。对比荷兰语和英语中音系上对立的清、浊爆发音的语音实现（关于音系学和语音学区别的简单介绍见第 3 章）就可以说明这一点。荷兰语（还有西班牙语、法语等诸多语言）中，音系上的清浊对立表现为声带在辅音闭合段是否振动（见图 6.2a 和 b）。而在英语中，词首爆发音在音系上的对立表现为清不送气爆发音（短正值 VOT）与清送气爆发音（长正值 VOT）的对立（见图 6.2 c 和 d）。因此，尽管荷兰语和英语中都有音系上对立的浊、清爆发音，但在语音上的实现方式却有所不同。荷

兰语通过闭合段中声带的振动与否来实现，而英语则通过 VOT 的长短来实现。实际上，一个具有短正值 VOT 的爆发音在荷兰语中归为清爆发音，而在英语中则对应浊爆发音。当母语为荷兰语这类语言的发音人说英语这类语言时，通常会在清浊特征的感知上产生混淆。

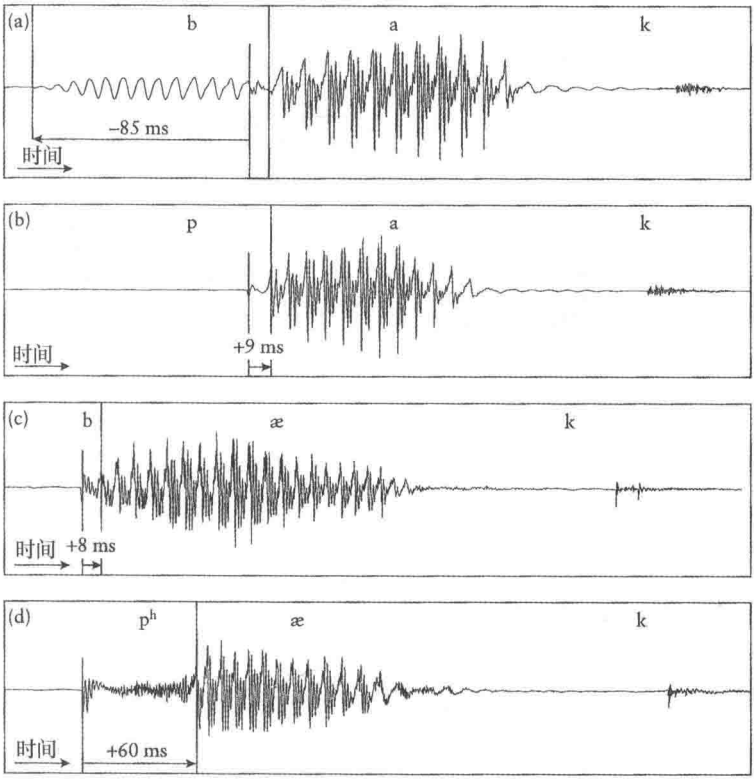


图 6.2　音系上对立的位于音节起始位置的浊爆发音与清爆发音的波形图和 VOT 的测量，（a，b）为荷兰语男性发音，（c，d）为英语男性发音。（a）中有三条标记竖线，最左边的标记表示辅音嗓音的起始点，中间的标记表示除阻点，最右边的标记表示元音嗓音的起始点。（b，c，d）中，左边的标记表示除阻点，右边的标记表示元音嗓音的起始点。

在英语和其他语言中存在较长的不带音性的闭合段与一个短的（正值）VOT 组合在一起的语音环境，这通常出现在位于词首的“擦音 + 口塞音”的序列中（例如图 6.3“stop”）。若按是否存在带音进

行标音，这个音应当标为［t］。若将 VOT 时长作为区分清浊爆发音的标准，则应标为［d］。不过，在这种情况下仍应标为［t］，因为其较长的闭合段会使人们将其感知为“清”音。在一定程度上，听音人会根据信号中的闭合段时长和 VOT 来计算“不带音性”，而并非仅利用其中某一条标准。

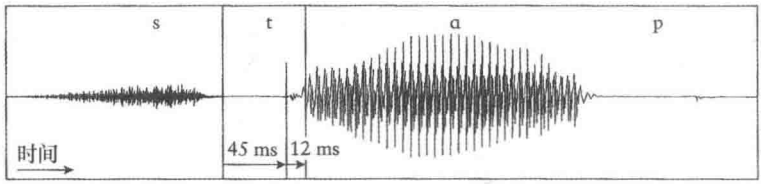


图 6.3　英语发音人所发单词“stop”的波形图。最左边的标记线表示闭合段的起始点，中间标记线表示除阻点，最右边的标记线为元音嗓音的起始点。闭合段时长为 45 ms，VOT 为 12 ms。

还有一些语言区分全浊、清不送气和清送气爆发音三种情况，例如泰语中的最小三项对立组［dam］（黑）、［tam］（去敲击）和［t^ham］（去做）（见图 6.4a－c）。韩语中也存在三种对立，但与泰语有所不同。韩语区分松、送气和紧（或强化）爆发音（见图 6.4d－f）。这种区分出现在双唇、齿龈、硬腭及软腭等发音部位。清松爆发音［p，t，c，k］的送气程度由中到强，位于词首位置时，VOT 值在 20 ms 到 70 ms 之间。送气爆发音［p^h，t^h，c^h，k^h］总是强送气清音，其 VOT 值在 70 ms 至 140 ms 之间。最后，紧或强化爆发音［p^*，t^*，c^*，k^*］为不送气清音，其 VOT 较短，在 5 ms 至 25 ms 之间。比起松爆发音，发这组爆发音时需要更强程度的肌肉活动和气压（详见 Cho 等，2002）。此时，喉部更紧张，经常使后接元音产生喉化的紧嗓音音质。另外还应注意，送气和紧辅音的闭合段时长比普通辅音更长。

对于另外两类爆发音而言，VOT 并不是有效的测量手段：

- **浊送气爆发音（voiced aspirated plosive）**。与浊爆发音相同，声带在辅音闭合段便开始振动。除阻后，声带便开始以放松的状态振动，此时气流可高速通过声门。因此，元音部分同时具有送气与带音特征，与

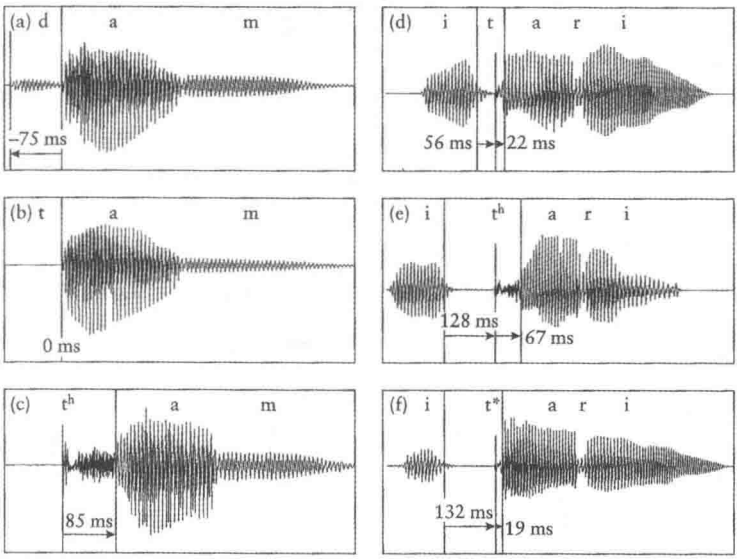


图 6.4 泰语中（a）全浊（b）清不送气（c）清送气爆发音；韩语中（d）松（e）送气和（f）紧爆发音的波形图、闭合与 VOT 的测量。

气嗓元音相同（见图 6.5a）。浊送气爆发音用与（清）送气附加符号对应的浊送气附加符号［ʱ］来表示，如［ɑˈbʱɑ］。遗憾的是，此时 VOT 并不能有效地描写浊送气爆发音的情况，因为这时的区别特征为带音与送气的组合，无法仅通过时间上的测量反映出来[2]。

• **清预送气爆发音（voiceless pre-aspirated plosive）**。预送气出现在元音后接清爆发音的序列中，指的是预见到后接辅音的不带音性后，元音嗓音的提早结束。描述预送气爆发音的测量参数为**嗓音结束时间（voice offset time）**，该参数是指元音嗓音结束点与后接爆发音闭合段对应的静音段起始点之间的间隔（见图 6.5b）。清预送气爆发音存在于北欧的几种语言中，包括芬兰语、盖尔语和冰岛语。我们可以将送气符号置于爆发音之前表示清预期送气爆发音，如冰岛语中的［ˈlaʰka］（“盐水”）与［ˈlaga］（“制作”）。

在孟加拉语中，送气是区分爆发音的一种附加方式。它有全浊爆发音（负 VOT）和清爆发音（正 VOT），闭合段有短与长之分，也有送气与不送气之分。对大部分发音部位而言，几乎所有这些方式均可互相组

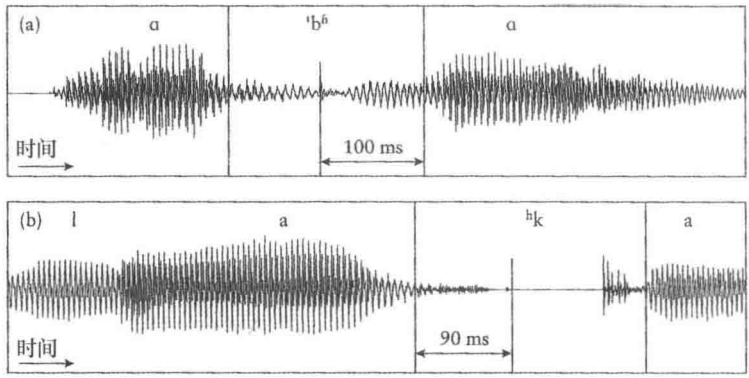


图 6.5　（a）孟加拉语中的浊送气爆发音的［ɑˈbʱɑ］与（b）冰岛语中的预送气爆发音的［ˈlaʰka］的波形图

合，总共可形成 8 种不同的爆发音。而西孟加拉语可在 4 个发音部位上构成 32 种不同的爆发音（见表 6.5）。

表 6.5　孟加拉语中的爆发音。孟加拉语使用送气、音长和嗓音来区分 4 个发音部位的爆发音。

		爆发音（清/浊）			
送气	音长	双唇音	齿音	卷舌音	软腭音
不送气	短	p/b	t̪/d̪	ʈ/ɖ	k/g
	长	pː/bː	t̪ː/d̪ː	ʈː/ɖː	kː/gː
送气	短	p^h/$b^ɦ$	$t̪^h$/$d̪^ɦ$	$ʈ^h$/$ɖ^ɦ$	k^h/$g^ɦ$
	长	p^hː/$b^ɦ$ː	$t̪^h$ː/$d̪^ɦ$ː	$ʈ^h$ː/$ɖ^ɦ$ː	k^hː/$g^ɦ$ː

6.4　常见语音与特殊语音

本章以及第 3 章和第 4 章介绍了多种类型的语音。IPA 辅音表总共包括 83 个符号，分别表示 59 个肺部气流辅音、14 个非肺部气流辅音以及 10 个“其他”辅音。需注意，这是目前所记录的世界上所有语言的全部辅音集合，没有一个语言包括所有的哪怕是其中大部分的音位。对于特

定音位的出现概率，UPSID（UCLA 语音音段清单数据库，1984）是一个很有价值的参考资源。Ian Maddieson 在 1984 年出版的《语音模式》一书中对 UPSID 进行了详细的描述与分析。该书出版时，UPSID 包括了大约 317 种语言的信息[3]。这些语言是按照遗传学方式进行采样的，因此每一个语系只列举一种语言。这些信息主要基于参考语法，其中几种语言则基于语音田野调查。

表 6.6 为常态辅音清单，即数据库中出现频率最高的辅音。

该表列出了一些常见的辅音。英语中包含其中大部分的辅音，只缺少龈后鼻音和喉爆发音。喷音、内爆音及咂音不属于世界语言中最常见的辅音。就爆发音而言，UPSID 中 92% 的语言有普通清爆发音，67% 有普通浊爆发音，29% 有清送气爆发音，16% 有清喷音，11% 有浊内爆音，还有其他一些少数类型。在 UPSID 的语言中，爆发音最常见的发音部位为双唇音、齿音/齿龈音和软腭音（均为 99% 左右），接下来是龈后音（19%）和小舌音（15%）。而在 UPSID 中，93% 的语言有一个或多个擦音，最常见的形式是/s/，出现于 84% 的语言中，接下来是/ʃ/（46%）和/f/（43%）。

表 6.6 UPSID 中最常出现的辅音（根据 Maddieson，1984）

	（双）唇音	齿音/齿龈音	龈后音/腭音	软腭音	喉音
爆发音	p b	t d		k g	ʔ
鼻音	m	n		ŋ	
擦音	f	s	ʃ		ɦ
塞擦音			t͡ʃ		
近音		ɹ, l	j	w	

对喜欢详细分析这些发音模式的读者来说，Maddieson 的书提供了大量的参考信息。虽然表 6.6 只列举了世界各国语言中最常出现的辅音，但我们可以将这些音在给定语言中出现的概率看作随语系变化的函数。例如，虽然/s/是 20 个最常见的辅音之一，出现于 UPSID 中 84% 的语言中，但大部分澳大利亚语言却根本没有擦音。最后强调一点，UPSID 列出的是音位出现的频率值而非声音出现的频率值。也就是说，尽管世界语言

中的很多音在英语中并不构成对立（如鼻元音或者紧喉元音），但它们的确经常在英语中出现。

本章对语音的发音和标音进行了总结。开始部分主要关注了英语，后续小节则通过考察其他的发音部位、发音方法、发声类型及气流机制，阐释了世界上各种语言中语音的丰富性。下一章将介绍声学和听觉的一些基本概念，为后面讨论言语声学和言语感知打下基础。

练　习

1. 哪两类塞音是由喉头气流产生的？阐述一下每类塞音的产生过程。
2. 哪类塞音是由软腭气流产生的？这些塞音是如何产生的，如何确定其发音部位？
3. 简述以下发声类型：清音（带音）、浊音（不带音）、耳语声、气嗓音和紧喉嗓音。
4. 为什么浊塞音的闭合段比清塞音短？
5. 解释如何通过 VOT 的测量来区分全浊、清不送气和清送气塞音，并举例说明。

注　释

1. Mikuteit 和 Reetz（2007）提出了应用更广的术语“收紧后时间”（After Closure Time，ACT）和“叠加送气”（Superimposed Aspiration，SA），可应用于描述更为广泛的爆发音。由于并不确定该术语是否会被学术界采用，本书中我们沿用 VOT 这个术语，尽管它并不完全适用于非起始爆发音，也不能覆盖所有送气的情况。
2. 上文提到的术语 ACT 和 SA 则同样可以被用于描写此类情况。
3. 该数据库的链接请见本书网页。

7 基础声学

我们介绍了言语声是如何产生的以及应该怎样进行语音符号的标注，也详细介绍了产生言语声所用到的器官和机制。最终，所有这些活动综合起来促成了言语声的产生。描写声音的科学被称为声学。本章将介绍声学中的基本概念，重点是声波方面的概念，包括频率、振幅和相位等，此外还将介绍这些物理量的测量和表示方法。

7.1 声波

声波是指声学能量从某个声源（比如说话人或者扩音器）到一个声音接收者（比如一个听音人或者一个话筒）之间的传播方式。下面介绍什么是声波以及声波是如何产生和传播的。

7.1.1 声波是气压的变化

声波是沿各个方向扩散的微弱的气压差。气压在天气预报中是一个很常用的概念，它指测量到的某些区域的空气压力的高低（气象图中以H或者L表示），例如，以百帕（hPa）、毫巴（mb）、毫米汞柱（mmHg；托，Torr）或英寸汞柱（insHg）来度量。其中平均大气压为1013 hPa，等于1013 mb、760 Torr或者29.92 insHg[1]。

从本质上讲，大气压和声波是相同的。大气中，气压的变化可能需要几个小时或者几天的时间才能实现，而声波引起的气压变化每秒钟可以发生成百上千次。大气中，气压在1013 hPa上下以±50 hPa（即5000 Pa）的范围波动，而人耳可以感知到的声波压力波动的变化范围为20 μPa至20 Pa（正常言语信号的范围是600 μPa至2 Pa）。此外，空气

中大气压在数公里或者数英里的范围内都保持不变，而声波的压力则只会在几米或者几厘米的范围内产生变化。

气压的变化通常会与当前的大气压相叠加，例如声波引起的气压变化就是如此。假设大气压是1010 hPa（101000 Pa），声压，即声波的振动范围为 8 mPa（0.008 Pa），那么在声波中，气压的变化范围就在100999.996 Pa到101000.004 Pa之间。因为声波是在接近静止的大气压中做快速的振动，所以当我们测量由声波引起的气压变化时，气象学中所说的大气压可以忽略不计，而最终得到的是一个 ±4 mPa 的气压变化。

7.1.2　声波的起源和传播

设想一个充满空气的密闭空间，例如一间客厅。我们在一张纸上用大量的随机分布的点来表示房间里的空气分子（见图7.1a）。纸上的这些点当然不会运动，但是在真正的房间里，空气分子会一直运动，每秒钟发生上百万次的相互碰撞和反弹。空气分子不断运动的结果是一个平均的气压。这样的平均气压通常会受到两种方式的影响。如果房间变小而空气分子的数量不变的话，那么由于分子要挤在一个更小的空间里，气压会变大；类似地，如果想要在一张更小的纸上呈现相同数量的点，那么点分布的密度就会增大。而另外一种情况是，如果有更多数量的空气分子分布在相同大小的空间内，气压就会增大（见图7.1b）；同样地，增加纸上点的数量也可以增大点分布的密度。这两种情况都证明了波义耳定律（见5.1.1），即在给定温度下，气压和体积成反比关系。

现在设想一个位于房间中央的空气球（见图7.1c）。当用气筒给气球充气时，气球开始膨胀，占据一定的空间，从而将周围的空气分子挤开。由于空气具有弹性，所以被气球挤开的分子移动到了相邻的区域，从而增大了这些区域的气压（见图7.1d），而房间中其他位置的气压则保持不变。这种气压差进而使空气分子从气压高的区域向周围气压低的区域移动（见图7.1e）。尽管气球保持原地不动，但很快地，离气球越来越远处的气压会得到增大。

如果此时将气球中的空气释放（见图7.1f），原来气球所占的区域会出现一个不含任何空气分子的真空空间，而周围区域的空气分子会立刻占领这个空间。因此，就像那些气压升高的区域一样，气压的降低也会

向整个房间传播。如果对气球进行反复的充气和放气，空气分子会随着气球的膨胀或缩小而不断地前后运动（见图 7.1g）。正如上面所说明的那样，这些气压上的波动不仅在气球的周围发生，还会在气球的远处发生。

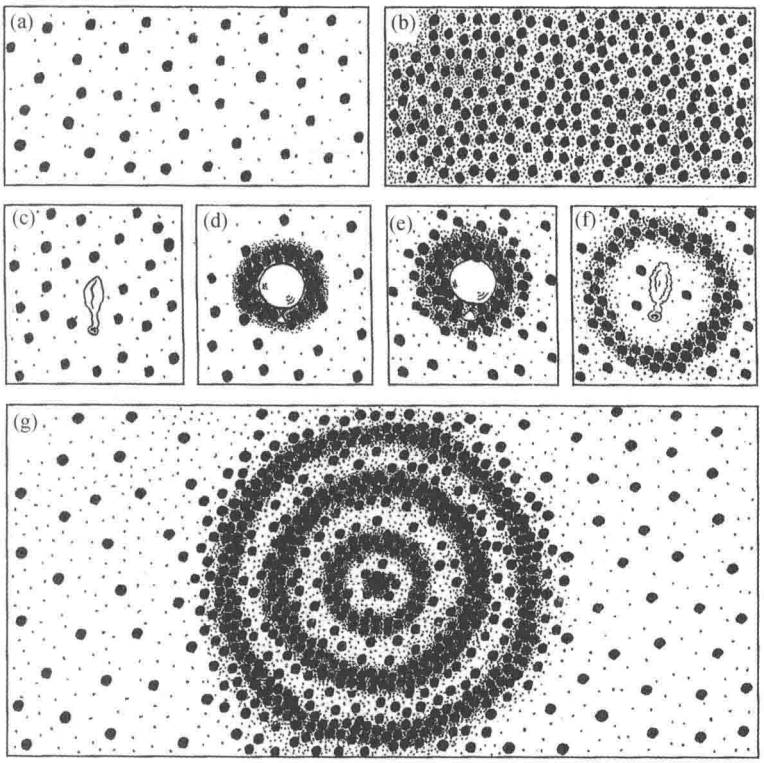


图 7.1 一个房间里的空气分子的状态（a）低气压（b）高气压。迅速地给一个气球充气、放气可以产生一个压力波（c－g）。

空气分子的这种振动与德国传统聚会时的一种叫作“schunkeln”的活动十分相似。在这种活动中，人们排成一行，有节奏地从左向右，边唱歌边摆动（见图 7.2）。如果第一个人向右侧靠，那么几秒后他右侧相邻的那个人也会做同样的动作，接着下一个相邻的人继续，如此往复。不久，第一个人摆回来再向左摆动，然后是第二个人、第三个人，等等。参与者们只是左右摆动，并不离开他们的位置，但是摆动运动本身却在整行人中传递着。

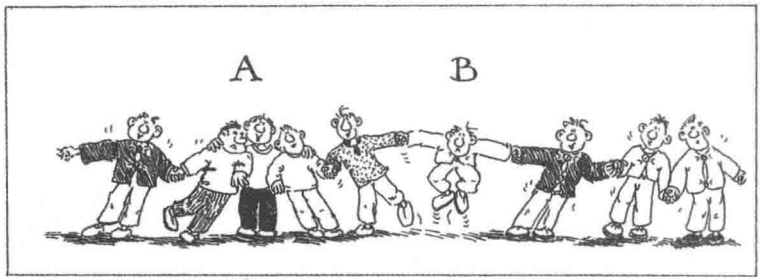


图7.2　聚会的客人由左向右摆动

这种运动在其他方面也与声波中空气分子的运动特点有相似之处。例如，参与者可以选择只在一个很小的距离内摆动，这就像气球只是被轻微地充气和放气，使每个空气分子只会移动较小的一段距离。但是参与游戏的人也可能会向外更大地摆动，这就像气球被充了更多的气，空气分子的移动距离也就更长了一些。此外，人们还可以快速或者慢速地进行摆动，而这与他们摆动距离的大小无关。同样地，空气分子可以快速或者慢速地运动，而这也与它们的运动范围无关。

但是分子振动的范围和其振动速度之间是有相关性的[2]：为了在给定的时间内移动得更远，分子就必须更快地移动，因为在相同时间内需要移动更长的距离。因此分子振动的速度依赖于其移动的距离。在计算语音信号的“能量”时这种关系非常重要，请参见7.3.2。

我们来总结一下上面的内容，目前我们探讨了声波的三个独立的特性：气压差传播的距离要远远大于一个分子的运动范围，这个运动范围在变化幅度或者变化速率上存在差异。

7.1.3　声速

空气分子的扩散运动（或者说声波的传播）需要时间：一个分子撞击相邻的分子，然后这个分子再与其相邻的分子碰撞，如此往复。如果知道了运动的距离和所花费的时间，我们就可以求出其运动的速度。除了气压差的大小和比率之外，声速是第三个可以测量的特性。

速度是按照物理上标准的“米每秒”来测量的，等于运动距离除以其运动的时间：

$$速度\left[\frac{m}{s}\right]=\frac{距离[m]}{时间[s]}$$

必须明确的一点是，声波的传播速度并不等于分子的运动速度，后者与分子振动速率相关（即分子通过静止位置的运动次数）。事实上，声波传播的速度只与介质（空气、气体或者固体物质）的特性有关，而与声波的特性（分子振动的大小和速率）完全无关。

因此，声波是由气压的变化构成的。在某个空间中，气压变化传播的速度称作**声速（speed of sound）**。声速与介质（通常为空气）的密度和弹性有关，在21℃的干燥空气中，声速约为344 m/s（约1238 km/h）[3]。“打破声障”指的是超过这个特定的速度。气体中，声速随着温度的增加而增加（因为高温下分子运动加快），当分子变轻的时候声速也会增加（较轻的分子更易运动，因此更快），但是声速与气压无关。

7.1.4 声波中的相对位置

由于声波从一个位置移动到另一个位置需要一定时间，因此在一个房间内，在声音传播路线上的两个不同的点的气压在相同时刻不必相等。为了理解其中的道理，我们可以回忆一下上面提到的摇摆的例子：在任何时刻，总有某些人离得较近（见图7.2和图7.3的A点），而另外一些人相距较远（见图7.2和图7.3的B点）。这说明A点（在单位空间内有更多人）的压力比B点（在单位空间内有更少人）大。在同一个声波中，两个特定状态之间的差异叫作**相位（phase）**。

尽管我们不能觉察出声波之间的相位差（因为我们在某个时刻只会处于某一个位置上），但我们的双耳却能感知到声波的相位信息。尽管我们处在一个位置上，我们的双耳却分开在两边，左右耳之间的距离也可视为图7.3中A和B两点之间距离。大脑能够分析左耳和右耳接收的信号之间的相位差，这个信息有助于大脑确定声音是从哪个方向来的。

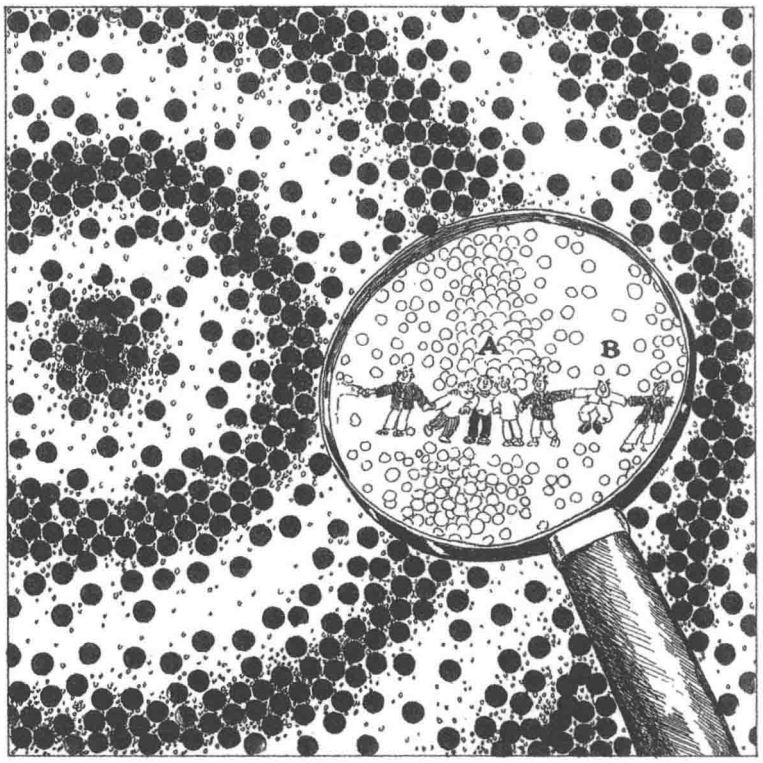


图 7.3　声波中的气压在 A 点高于 B 点

7.1.5　纵波与横波

上面我们介绍的声波属于**纵波（longitudinal wave）**。在这类波中，分子振动的方向与波本身的运动方向是一致的。在图 7.2 和图 7.3 中的摆动示例中可以清楚地看到：参与者们沿水平方向左右摆动，因摆动而产生的波动也同样是沿水平方向传递的。另外一种波的分子的振动方向与波的传递方向垂直，称为**横波（transverse wave）**。“人浪波”或者叫“墨西哥波”就是横波的一个例子，它类似于观看体育赛事的观众自发形成的人浪，见图 7.4。做人浪时，一个人站起来高举两臂，紧挨着的人随后也站起举臂，一个接着一个不断地进行；与此同时，第一个人又坐下，第二、第三个人也相继坐下，如此往复。结果，每个人都垂直地站起坐下，而人浪却是沿水平方向移动的，即与观众的运动方向垂直。横波在内耳中起到很大的作用，使声波转换为神经脉冲（见 12.3.5）。

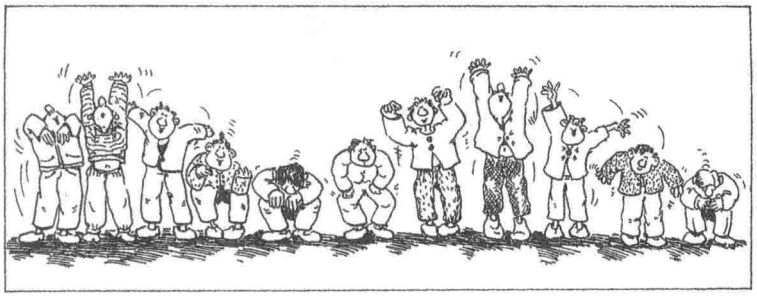


图7.4 横波的例子："人浪波"（"墨西哥波"），参与者一个接着一个地站起或蹲下。

现在我们已经介绍了声波的传播原理，那么下一个问题就是：怎么测量声波？在声学语音学中，我们描述和分析的语音信号，从物理上说就是声波。

7.2 声波的测量

声波是很小的、快速辐射的气压变化。为了研究这些波，我们必须测量它们。众所周知，麦克风是可以把气压变化转换成电信号的设备（见7.2.1），一旦声学信号通过麦克风拾音后，就可以用波形图以图形化的方式展现（见1.1节和7.2.2）。

7.2.1 麦克风

麦克风（microphone）会把声波的气压变化转化为电信号，以便进行进一步的处理。麦克风的构造中，有一层膜包在一个隔腔上，形成一个几乎气密的封闭空间（见图7.5a），这与鼓的结构类似。当声波（即一段连续的气压高低不同的区域）到达麦克风时，由于膜下面的密闭气室内的气压保持恒定，所以膜被迫地向外或向内运动。如果膜外的气压比膜内的高，膜就向内运动（见图7.5b），反之，膜便向外运动（见图7.5c）。那么如何测量这种变化呢？如果将一只很小的铅笔置于膜的中央，旁边放一张纸，随着膜的上下振动，铅笔就会在纸上画出一条线来（见图7.5d）。膜移动得离静止位置越远，线画得就越长，换言之，线的长度取决于气压变化的最大尺度。

随着膜不停地往返运动，铅笔就能画出一条线来。但是这种画法有一个重要的缺陷，它只能记录隔膜离开静止位置的偏离量的大小，而不能记录某一个特定的偏离量所发生的时刻。但是如果我们匀速移动这张纸，铅笔就可以在纸上画出一条连续的线来（见图 7.5e）。这样的线不仅显示了气压变化的大小，也可以观测到某个气压出现的时刻。

当然，麦克风不会使用笔纸来记录膜的振动。膜的振动会使麦克风产生很小的电信号，电信号被放大后可以通过闪存或者硬盘、计算机或者其他的存储设备来进行存储。在读取声音的时候，电信号再被转换成声波，通过将放音设备连接到电子放大器和喇叭（或者耳机）上，使声音再现。

麦克风膜下面的隔腔不应该是完全密封的，否则膜也会受到大气压差的影响。大气压的变化范围（±5000 Pa）相对于声压的变化范围（最大 ±20 Pa）来说是非常大的，这意味着大气压的任何变化，都可能给麦克风的膜带来巨大的影响。因此，膜下的隔腔内有一个非常小的孔，使隔腔里面的气压可以适应这个很大但缓慢的大气压变化。在声波引起的快速气压变化中，几乎没有空气分子有机会闯过这个小孔。所以，对于声波来说，密闭的隔腔里的气压是相当稳定的。在 12.2.3 中，我们会看到人耳的机制也是如此。

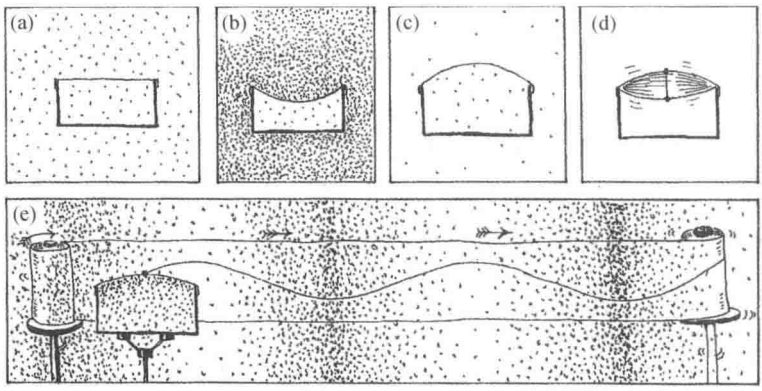


图 7.5　麦克风的原理（具体解释见文中说明）

7.2.2 波形图

麦克风产生的电信号可以用图形表示，得到的图形与图 7.5e 中膜上铅笔画出来的类似。这种图示以声压为纵轴、以时间为横轴的时候，我们称之为**波形图（waveform）**（见图 7.6）。通过波形图，我们可以观察到平时只能通过听来获得的声学事件。

波形图是一种将声学信号图形化表示的方法。尽管波形图可以忠实地呈现声波的气压变化，但我们还需要进行进一步的语音学分析和表示来分离信号中的重要特征。我们将在8.3 节中重点讲述有关内容。我们下面将对波形图中可以直接读出的参数做详细的介绍。

对于语音学研究而言，把声波记录下来并表示为波形图是非常重要的。一个声音，从本质上说是具有瞬时性的，例如说话的声音只是在说话的那一刻存在。声波一旦产生，会马上传播并渐渐消失。在 1877 年爱迪生发明留声机之前，对语音学的研究者来说，某个声音只能听一遍，或者他们会尝试尽可能像地去重复原始的发音。但是，两个发音之间永远不会是完全相同的。人们必须“身临其境”，例如去参加音乐会、演讲会才能听到某位歌唱家的演唱，或者某管弦乐队的演奏，或者某人的演讲。但是录音技术使我们可以重复不断地播放、重现或者储存这些录下的声学事件。

即便我们录下了某个声学事件，它只会在每次播放的时候暂时性地存在。这一点与图像不同，图像的观测或者研究不受时间限制，而声学事件稍纵即逝。但是波形图是对一个声学事件的图形化表示，它使我们可以像看图像一样地观察声学事件。通过实现将语音以图形化的方式进行表示，我们可以使以前仅靠聆听瞬时声音信号所无法实现的语音学研究成为可能。

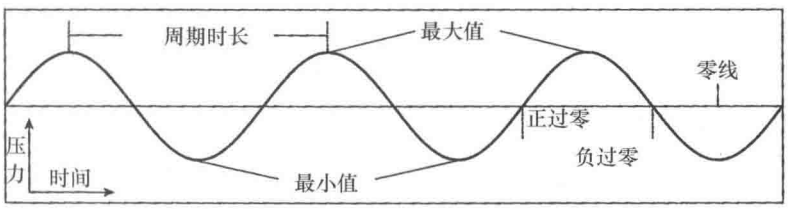


图 7.6 一个简单的对称信号对应的波形图

7.3　声学维度及其测量单位

本节将介绍用于定义声学信号的主要维度，即频率、振幅和相位，它们对应的测量单位分别是赫兹（Hz）、分贝（dB）和度（°）[4]。在7.1节，我们已经介绍了声学的基本概念，声波的主要特性由下面几点决定：

1. 气压改变的速率；
2. 气压改变的幅度；
3. 一个声波内各点的相对位置或者两个声波之间的相对位置（即相位）；
4. 声波的传播速度。

下面我们将更加详细地介绍这些特性。

7.3.1　频率

在波形图上，最大和最小的气压点与代表平均气压值的中线（零线）的距离最远（见图7.6）。波形图中气压的偏移是向上画还是向下画是任意的，因为同一个声波在同一个位置可以产生两个镜像，画图的方向主要取决于当时麦克风是向上还是向下放置的。我们会用一个向上的偏移来表示波形图中气压的最大值。

7.3.1.1　周期时长

在图7.6的波形图中，我们很容易得到两个最大值之间的距离。图中的横轴代表时间。如果波形为周期波，即相同的气压变化一次一次地重复出现（就像钟摆那样的振动），则相邻两个最大值之间的时间间隔就称为**周期（period）**。

测量周期时长的时候也不一定必须取两个最大值之间的距离，两个最小值之间，或者两个正过零点或者两个负过零点之间的时间段都是一个周期。正过零点或负过零点指的是曲线向上运动或者向下运动经过零线时的信号点。取点时的唯一的要求是这些点必须具有相同的相位。换言之，它们的位置必须存在周期上的对应关系。

如果空气分子以较高的速度振动（即同一时间段内振动的次数更

多），则只要图 7.5e 中的纸的运动速度不变，那么曲线最大值之间的距离就会更近；如果空气分子振动得更慢，则最大值之间的距离就会更远。换言之，缓慢的振动对应较长的周期（见图 7.7a），而快速的振动对应较短的周期（见图 7.7b）。

为了测量波形图上的时间间隔，仅仅通过测量纸上的距离是不够的，我们还必须考虑纸的运动速度。如图 7.7 中所示，图 7.7a 和图 7.7c 上的波形看起来一样，但其实它们所代表的是两个不同的信号。相反，图 7.7b 和图 7.7c 看起来不同，但它们表示的却是相同的信号。实际上，图 7.7a、图 7.7b 与图 7.7c 的差异在于时间的比例不同，即纸移动的速度不同。

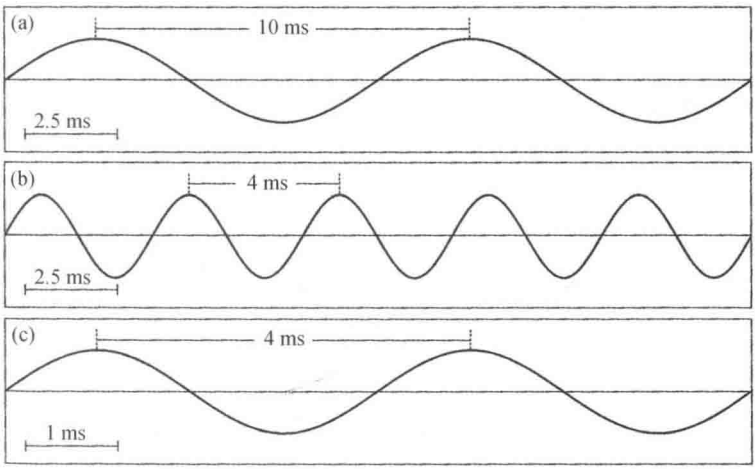


图 7.7 不同时间坐标比例对应的波形图。波形（a）和（b）表示的是两个不同的信号，但波形（b）和（c）表示的则是相同的信号。（a）和（b）的时间坐标比例相同，但都与（c）不同。（a）和（c）看起来一样，只是由于不同时间坐标比例造成的。

在图 7.7a 和图 7.7b 中，1.5 cm 代表 2.5 ms，而在图 7.7c 中 1.5 cm 则代表 1 ms。也就是说，在图 7.7a 中，最大值之间的间隔 6 cm 所代表的时间间隔为 10 ms（6 cm 的距离 × 2.5 ms/1.5 cm 的标尺 = 10 ms 的时间间隔）；但在图 7.7c 中，6 cm 则代表 4 ms（6 cm × 1 ms/1.5 cm = 4 ms）。

因此，当我们研究一个波形图时，时间坐标比例是非常重要的。

那么画语音信号的理想坐标比例是什么呢？对于这个问题我们没有唯一的答案，正确的坐标比例是由画图时不同的目的所决定的。在某些情况下，我们需要在一个视图中显示整个词或者短语的波形图，那么时间窗长一点就比较好；在其他一些情况下，例如需要察看声波某部分的细节时，则时间窗短一些为宜。这很像我们开车旅行时使用行车路线图的情况。司机会选择较大比例尺的地图来察看从一个地方到另一个地方的路线。这时，在图上可以观察到起点和终点的位置关系，但却看不出行车路线的具体细节。而如果想察看路线的具体细节，则必须选用一张比例尺更小的地图。这样每条街道都一清二楚，不过总体路线的方向却又看不到了。如果司机同时拥有两种比例的地图，那就可以确保万无一失了。观察波形图的时候亦是如此。

7.3.1.2　周期时长与周期频率

在前面的章节中我们介绍了“周期性”的概念是用来描述等间隔内重复出现的事件的。周期事件可以通过给定的周期时长来描述（如每 20 分钟），或者也可以通过事件在单位时间内出现的次数来描述（如每小时 3 次）。“出现率”或者“单位时间内出现的次数”被称为**频率（frequency）**。用于测量频率的单位是“每秒钟出现的周期数”（或者周期每秒［cps］），称为**赫兹（Hz）**。当描述一个信号的时候，人们经常会使用频率这个概念。频率是理解很多声波特性的基础。

目前，我们介绍了三种用来描述声波中两个气压最大值之间，或者任何两个同相位点之间关系的方法：

1. 每秒钟出现的次数，用赫兹（Hz）表示；
2. 连续出现的两点之间的时间间隔，用秒（s）表示；
3. 它们的（空间）距离，用米（m）表示。

这三种测量方法之间有什么关系呢？例如一个钟摆，往返摆动一次需要 1 秒，则其振动频率就是 1 赫兹（1 Hz），此时一个振动周期持续 1 秒（1 s）。如果一个振动的频率是每秒 10 个周期（或者 10 Hz），则每个周期就是十分之一秒（1/10 s = 0.1 s）；如果振动频率为 1000 Hz，则一个周期为千分之一秒（1/1000 s = 0.001 s）。频率（赫兹）和周期时长（秒）之间的转关系可以通过以下公式来表示：

$$\text{频率}[\text{Hz}]=\frac{\text{周期数}}{1[\text{s}]}=\frac{1}{\frac{1}{\text{周期数}}[\text{s}]}=\frac{1}{\text{周期时长}[\text{s}]}$$

或者可以用f代表频率，T代表周期，简写作：

$$f=\frac{1}{T}$$

为了避免在表示较高频率的时候出现过多的零，1000 可缩写为“千”（kilo），则 1000 Hz 可计为 1 kHz。类似地，“毫”（milli）是千分之一的缩写：一般人们不说一个周期的时长是千分之一秒，而说周期时长是 1 毫秒（1 ms）。换言之，如果一个周期性振动的频率为 1 kHz，则每个周期的时长为 1 ms。人们常误以为如果 100 Hz 信号的周期时长是 10 ms，那么 200 Hz 信号的周期时长一定是 20 ms。其错误在于周期时长的增加会导致频率的减小。而 200 Hz 信号的周期时长应为 5 ms。

人类语音的频率范围在 30 Hz 至 8 kHz 之间，对应的周期时长为 33.3 ms 到 0.125 ms。

7.3.1.3 周期频率与波长

在前一节中，我们了解到周期时长（s）与周期频率（Hz）有关。现在我们来计算房间里传播的声波中，两个最大气压点之间的距离（即一个周期的“长度”）。

对于一个 100 Hz 的信号，其每个周期持续为 10 ms。在介绍声速的 7.1.3 中，我们提到速度就是简单的距离（米）与时间（秒）的关系：

$$\text{速度}[\frac{\text{m}}{\text{s}}]=\frac{\text{距离}[\text{m}]}{\text{时间}[\text{s}]}$$

公式中，声速是已知的（约 340 m/s）。在这个例子中，一个周期的时长也是已知的，10 ms = 0.01 s。未知量是声波中两个最大气压点之间的距离。将上面公式做变换后，就可计算出距离：

$$\text{速度}[\frac{\text{m}}{\text{s}}]\times\text{时间}[\text{s}]=\text{距离}[\text{m}]$$

这个公式说明：在通常情况下运动的距离越远，就意味着运动所需要的时间越长，或者运动的速度越快，即运动的距离取决于运动速度的快慢和运动时间的长短。如果将这个公式应用于声音的传播，则我们需要将“速度”换成“声速”，“时间”换成“周期时长”，“距离”换成

“波长”，进而得到：

$$\text{声速}[\frac{\text{m}}{\text{s}}] \times \text{周期时长}[\text{s}] = \text{波长}[\text{m}]$$

或者用 c 代表声速，T 代表周期时长，用希腊字母 λ [ˈlæmdə] 代表波长，则有：

$$c \times T = \lambda$$

将上面例子中的数值（声速 340 m/s，周期时长 10 ms）代入此公式，我们可以得到：

$$340 \frac{\text{m}}{\text{s}} \times 0.01 \text{ s} = 3.4 \text{ m}$$

因此，在上例中，当房间里的声波的频率为 100 Hz 的时候，波形的两个最大气压点之间的距离是 3.4 m。如果可以将空气“凝固”，那么每隔 3.4 m 我们就可以观测到一个最大值。同理，当声波的频率为 1 kHz 的时候，两个最大气压点之间的距离为 34 cm，对于 50 Hz 的声波，该距离为 6.8 m。

此外，我们可以将用于表示频率和周期时长之间关系的公式 $f = 1/T$（见 7.3.1.2）变换为 $T = 1/f$。这样我们就可计算声速 c，周期时长 T，频率 f 以及波长 λ 所有这些量之间的关系了：

$$c \times T = c \times \frac{1}{f} = \frac{c}{f} = \lambda, \frac{c}{\lambda} = f, \text{ 和 } \lambda \times f = c$$

最大值之间的距离叫作振动的**波长（wavelength）**。如公式所示，波长取决于声速和声音的频率。在声速不变（即说明温度、湿度等条件也保持不变）的情况下，高频声音的波长小，低频声音的波长大。

7.3.1.4　基频的时域表示

目前为止，我们已经讨论了信号的频率。像图 7.6 中那种简单的信号只有一个频率，然而语音信号则完全不同。像我们在 1.1 节中看到的那样，语音信号是一个由很多具有各自频率的信号组成的复合信号。然而，即使是在一个复合信号里，也存在一个“周期”，使信号一遍一遍地重复出现。与图 7.6 中简单的“上下上”模式不同，一个周期性的语音信号的形状会比较复杂，然而其形状看起来依然会不断重复，我们称之为一个“（复合）周期”。复合信号的“周期频率”是一个“基本”频率，叫

作**基频（fundamental frequency）**或者记作 F_0（读作："F 零"）。对于一个浊语音信号而言，其**基本周期（fundamental period）**（$T_0 = 1/F_0$）是由声带周期性的开合引起的。

一个（语音）信号的基频就是我们所感知到的音高[5]。在言语表达过程中，音高的变化引起的是语音学中所说的语调。因此跟踪基频随时间变化的情况非常重要，这可以通过**基频曲线（F_0 contour）**或者**音高曲线（pitch contour）**（见图 7.8b）来表示。与图 7.8a 中的波形图一样，基频曲线的横轴代表时间，但纵轴代表频率，以赫兹表示。图 7.8 所示的是一个英语疑问语气短语的基频曲线，短语的末尾音高被抬高。在波形图上短语末尾的部分，我们可以观察到，单个周期变短了，因此波形变得更加密集，也就是说频率升高了。

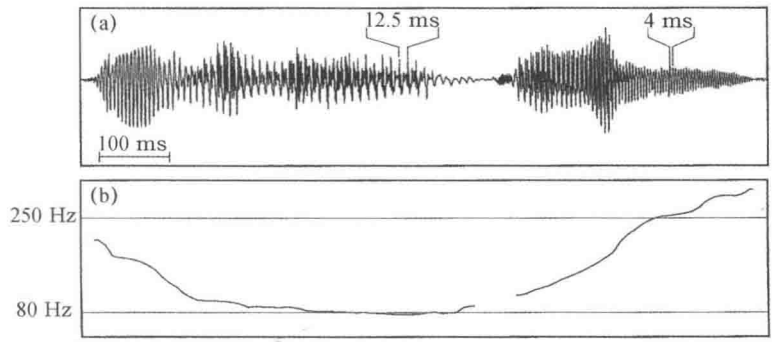


图 7.8 一位男性英语发音人所发的疑问语气短语的（a）波形图和（b）基频曲线

看起来计算一个信号的基频很容易，计算机程序只需测量周期时长然后将其转换为赫兹表示即可。然而，声带振动的每个周期并不是完美的周期信号；每一个周期与其前后的周期都在长度和形状上有着微小的差别。因此，尽管我们通常会说"周期性声带振动"，这种振动实际上应该是**准周期（quasi-periodic）**的。这些微小的不规则的变化让对语音信号基频的计算变得异常复杂。虽然人们提出了许多计算基频的方法，但没有一种理想的方法可以计算出一个信号真正的音高。所有的方法都存在误差，从而可能会导致计算出错误的基频值、无法计算出基频

值、或者将不带音的延伸部分计算为基频值等情况。因此，即使我们使用计算机来计算基频值，参照语音波形图来检测基频值的准确性还是很有必要的。

7.3.2　振幅

在7.3.1中，我们看到空气分子振动的频率就是声波的周期频率。在波形图中，我们可以用两个对应点之间的距离（即一个周期的时长）来间接地表示声波的周期频率。此外，每个空气分子的振动与麦克风的膜的运动直接相关。当没有声波影响时，麦克风的膜处于静止位置（即其平衡状态），其位置对应为波形图的**零线（zero-line）**（见图7.9）；偏离零线的偏离量称为**位移（displacement）**。在波形图中，零线与最大位移之间（在纵向范围上）的距离称为**振幅（amplitude）**。一个最大值（零线上方）与紧邻的一个最小值（零线下方）之间的距离被称为**峰峰振幅（peak-to-peak amplitude）**。

峰峰振幅较高表示分子振动的距离范围较大，也就是气压最大值和最小值之间的差异较大。这意味着感知到的声学信号会更响。（对于物理上的“振幅”和主观感知到的“响度”之间的区别请见12.6节。）需要注意的是，信号在每个时刻都有位移，位移可以在零线上方（正值），或者在零线下方（负值），或者等于“零”；而振幅或者位移等于“零”并不代表没有气压，而是说明其恰好到达了环境气压（例如1013 hPa）。

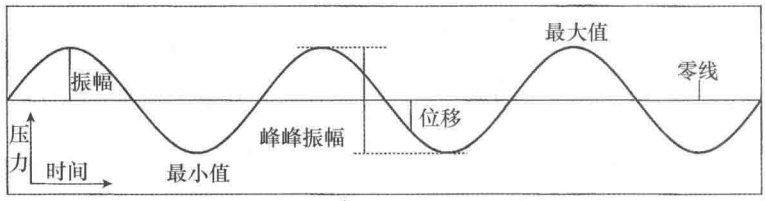


图7.9　通过一个简单对称的信号的波形图来说明与振幅相关的术语

到目前为止，我们所讨论的波形图都是通过横轴来表示时间并利用其计算周期时长和周期频率（见7.3.1.2）的，但是我们从来没有提过纵轴上用于表示振幅的量度和单位。我们提到过，声波是气压的微小波动，

可以用“帕斯卡”来度量。这些波动在被麦克风拾音后会被转化成电压或者电流。这样如果使用经过校准的麦克风和电放大器，我们就可以在波形图上简单地用“压力”作为纵坐标振幅的量度，用“帕斯卡”作为其单位。

然而在通常的录音中，麦克风是没有经过校准的。麦克风的录音级别经常会因为手动或自动的调整而发生变化。因此，麦克风的输入信号和其存储在录音设备里的数据是不一致的。在回放的过程中，我们也经常会通过调整放大器的“音量”旋钮来调节放音的响度。总之，经过麦克风拾音之后，在信号的存储、播放或者复制的过程中，我们便无法再获得声音信号原始的帕斯卡值了。

尽管如此，只要录音级别或者音量设置在一段录音中保持不变，信号振幅较高的部分和振幅较低的部分之间的关系就会保持不变。换言之，虽然我们无法求得绝对振幅的值，但是振幅之间的相对关系是完全可以计算的。通常情况下，我们会采用某个参考值来计算振幅之间的相对关系。

若在记录语音时采用声波的实际气压值，则还存在另一个问题，那就是声波的实际气压值的变化范围很广（从 20 μPa 到 20 Pa），最大振幅是最小振幅的一百万倍。人耳对振幅的感知特点令人惊讶，人耳能很轻易地分辨几微帕的振幅差，而当信号的振幅增大到几毫帕的时候人们却无法察觉相同大小的振幅差了。换言之，在安静的环境下我们可以听到针落地的声音，但在摇滚音乐会上我们却做不到。

对于较小的信号我们能够感知到其微小的变化，而当信号较大时，我们只能感知到其较大的变化。这种能力还存在于人类其他的感知模式中，例如对亮度和力度的感知。利用“对数”我们可以很好地逼近这种关系，我们可以通过以下公式简单地计算两个声压之间的关系：

$$\log\left(\frac{\text{振幅}}{\text{参考振幅}}\right)$$

对数关系表示的是两个值之间的比率而非绝对差异。对数关系可以反映上述现象，即人们在感知上认为振幅较小信号的较小的差异与振幅较大信号的较大的差异是相同的。附录 A.3 对这个公式以及计算信号“响度”的方法做了非常详细的解释。

我们采用**贝尔（bel）**表示通过以上公式计算两个振幅值之间关系所得的结果。由于贝尔值通常很小，所以我们会将所得到的值乘以 10。因此 1 贝尔等于 10 个**分贝（decibels，dB）**。

因此，在计算信号的相对振幅时，我们采用下面的计算公式：

$$10 \times \log\left(\frac{\text{振幅}}{\text{参考振幅}}\right)[\text{dB}]$$

需要注意的是，这里的（分）贝不是一个单位（即并不涉及像帕斯卡、伏特或其他一些维度的单位）而只是一个用来计算数值的公式。实际上，这样做的优点十分显著，因为我们可以简单地通过（分）贝去比较压强、电压、电流、计算机中储存的数字或者任何其他维度的数值之间的关系。

正如 7.1.2 中提到的，在频率保持不变的情况下，与振幅较小的信号相比，振幅较大的声音信号中的空气分子在相同时间内要跨越更大的距离。这会加速空气分子的振动（分子的振动速度与声速无关；见 7.1.3）。我们可以用钟摆的运动来解释这一现象：无论摆离平衡点较小的距离（低振幅）还是较大距离（高振幅），钟摆的摆动频率都不会改变。但是，摆离平衡点越远的钟摆，摆动的速度会越快。因此，摆动幅度越大的钟摆能量也越大，该能量称为钟摆的“动能”[6]。如果在最低的位置截住钟摆，我们可以清楚地感觉到钟摆动能的大小。

总的来说，钟摆（或者空气分子的振动）的能量与其速度的平方成正比。因此，在计算响度时通常对振幅值取平方（附录 A.2 会更详细地介绍能量、强度和功之间的关系）。在计算 dB 值的时候，也是如此。通过声压值来计算声音能量的公式为[7]：

$$10 \times \log\left(\frac{\text{振幅}^2}{\text{参考振幅}^2}\right) = 20 \times \log\left(\frac{\text{振幅}}{\text{参考振幅}}\right)$$

如果使用 dB 计算声压值，并将参考振幅设为人能感知到的最小振幅值 20 μPa，那么这个公式所表示的就是**声压级（sound pressure level，SPL）**，用 $\mathbf{dB_{SPL}}$ 表示。如果把参考振幅设为某个人能感知到的最低振幅，那么这个公式所表示的就是**声级（sound level，SL）**，用 $\mathbf{dB_{SL}}$ 表示。在 CD 上或者声音文件中，最小的数值是数字录音的“1”个单位（见 8.1.3），所以在分贝公式中的参考振幅就是“1”。在这种情况下，所得

的分贝通常没有下标，或者是以最高的级别（“满量程”，full scale）作为参考，得到值是负值，用 $\mathbf{dB_{FS}}$ 表示。

均方根（root mean square，RMS）振幅是用来计算某一时段的振幅值的常用方法，它可以很好地反映感知到的信号响度。计算均方根振幅的具体方法可参考附录 A.3.1。在计算的过程中，最正确的做法是同时指明所使用的计算方法（例如均方根，RMS）和参考量（例如声压级，SPL），例如记作“$dB_{RMS\text{-}SPL}$”。而在实际应用中，由于瞬时的 dB 值是针对一个值计算的，而某个时间段的 dB 值通常是基于均方根振幅计算的，所以在下标中通常不会示明计算方法（也就是说，经常写作“dB_{SPL}”而非“$dB_{RMS\text{-}SPL}$”）。

使用不同的 dB 标度看起来可能会有一些混乱，但这样做会有一个重要的优点。dB 值所表示的是两个测量值的比值，而在计算比值的过程中，这两个测量值的单位被消去了，因此我们可以很方便地对比从不同物理域所测得的 dB 值（帕斯卡、瓦特、厘米、计算机中储存的数字等等）。不同的标度之间的区别是测量单位的不同，而在各自域中的关系是相同的。

归纳起来，**音强级（intensity level）**的计算公式为：

$$10 \times \log\left(\frac{\text{音强}}{\text{参考音强}}\right)[\mathrm{dB_{IL}}]$$

声压级的计算公式为：

$$20 \times \log\left(\frac{\text{振幅}[\mathrm{Pa}]}{20\mu\mathrm{Pa}}\right)[\mathrm{dB_{SPL}}]$$

由于大多数情况下人们不会用校准后的设备去测量振幅，所以声压级经常是针对某个随意设定的参考值计算出来的。这个值可以是某个系统的最小单元（例如计算机中的一个比特，bit），因此公式可以简化为：

$$20 \times \log(\text{振幅})[\mathrm{dB}]$$

上式中的振幅通常为均方根振幅。

用比值代替绝对数值还有一个优点。正如前面提到过的，在录音的时候人们首先会选择某个录音级别，在回放的时候人们也会将音量调整到某一个舒适的收听级别。这种设置录音和回放级别的过程当然会改变

绝对响度的值，但这不会改变大声和小声段落（loud and soft passages）之间的相对关系。因此，通过采用 dB 标度，我们可以计算某个信号相对振幅的特性，而不必去参考录音或者回放时的级别设置。

此外，由于计算的是两个值之间的关系，因此 dB 值可以就某个系统能够容许的最大振幅来计算（目前为止，我们一直使用的是系统能够处理的最小值）。这样，公式中的“振幅”就会比“参考振幅”要小，因此得到的 dB 值便是负值。这种 dB 级别的表示方法通常被用于专业设备。例如，“-3 dB”表明其振幅比系统所能容许的最大值小“3 dB”。

最后，我们要清楚地认识到“0 dB”并不代表没有振幅（就像 0 ℃不代表没有温度），只是表明已经达到了“参考振幅”：

$$20 \times \log\left(\frac{\text{参考振幅}}{\text{参考振幅}}\right) = 20 \times \log(1) = 0$$

第二点，振幅增加一倍不代表 dB 值增加一倍，而是增加 6 dB，因为：

$$20 \times \log\left(\frac{2 \times \text{振幅}}{\text{振幅}}\right) = 20 \times \log(2) \approx 20 \times 0.3010 \approx 6 \text{ dB}$$

7.3.2.1　表示振幅随时间的变化

通过计算一个语音信号的均方根振幅可以大致测量这个信号整体的响度。但是人们通常感兴趣的是一句话中响度的变化，例如比较某个音节是否比另一个更响（或者更重）。为此，我们需要求出的是随时间推移的均方根振幅的变化，而不是只对信号的整体求一个值。因此，我们需要一种可以在信号层面体现均方根振幅随时间推移变化的图形。

用一种简单方法就可以解决这个问题。首先，选择一小段信号并计算其均方根振幅；然后选择这段后面相同大小的一段并做相同的计算；一直重复这个过程直至覆盖整个信号（见图 7.10a）。换言之，这个过程并不是一次性地对“整个”信号进行计算，而是采用一个**窗（window）**沿着信号移动，并在每个窗内计算其对应的均方根振幅并画图。这就犹如你坐在火车上，透过车窗看外边远处的风景：虽然从车窗中只能看到一段风景，但随着火车的移动，整个风景都会尽收眼底。通过这种方法画出来的振幅值会形成一条曲线，称为**振幅曲线（amplitude contour）**。

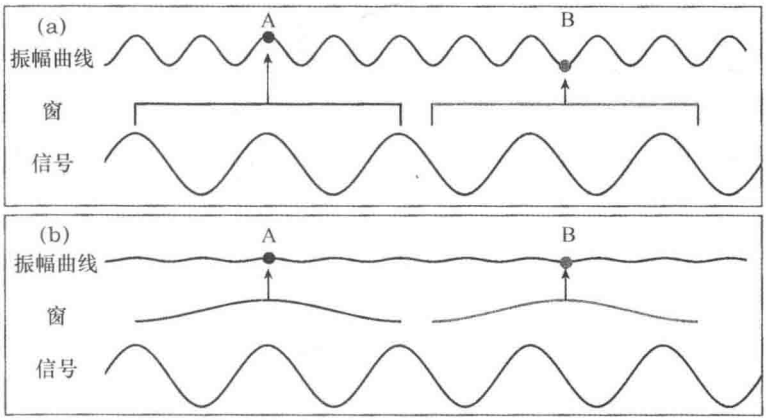


图7.10 采用（a）矩形窗（b）“平滑”窗（只给出A点和B点两个窗的位置，但是在两点之间可以有任意的叠加的窗）计算振幅曲线。

计算所得的振幅曲线会受到窗的形状和宽窄的影响。如图7.10a所示，使用矩形窗所得到的振幅曲线会有较大的起伏，因为在计算的过程中，有时有5个最大或最小值在窗里（点A），而有时则只有4个最大或最小值在窗里（点B）。因此，尽管信号的振幅保持不变，计算所得的dB值却随着窗在信号中相对位置（即它的相位）的变化而变化。我们可以通过“压平”窗的边缘来减小窗边缘处的振幅对计算的影响，如图7.10b所示。

但是，即便使用扁平窗我们也不能完全避免计算所得振幅曲线的不规则。如果窗太宽[8]，就不能追踪到振幅的快速变化，因为这时的平均值是在一个较长的时段上计算出来的；如果窗太窄，窗就会像海浪中的一条小船一样随着波形上下移动。通常，在时间上的高分辨率（窄窗）与好的平滑效果（宽窗）之间可以找到一个平衡点。按照一般的经验，窗的宽窄主要是由信号的周期时长决定的：为了获得有用的振幅曲线，窗宽应该覆盖2至3个周期。

图7.11对以上的考量做了很好的说明。图7.11a是英语单词“stair”的波形图，图7.11b是对应其采用较窄的窗计算所得的振幅曲线。从图7.11b中可以看出，信号在除阻前A点处有较小的振幅，而在除阻点B点处振幅有所增大；然而，元音的均方根振幅（C点处）却表现出了信

号中没有的变化，这仅仅是由于窗太窄的缘故。元音部分的这些变化看起来是随机的，这是由于信号中元音部分周期时长的微小变化所致。如果采用较宽的窗（见图 7.11c），这种振幅上的变化就消失了。但是，这也导致我们无法观察到 A 点处对应的静音段和 B 点处振幅的突然变大。没有一种窗宽的选择是“最好”的，窗的宽窄必须根据研究的需要来选择，这与 7.3.1.1 中路线图的例子类似。

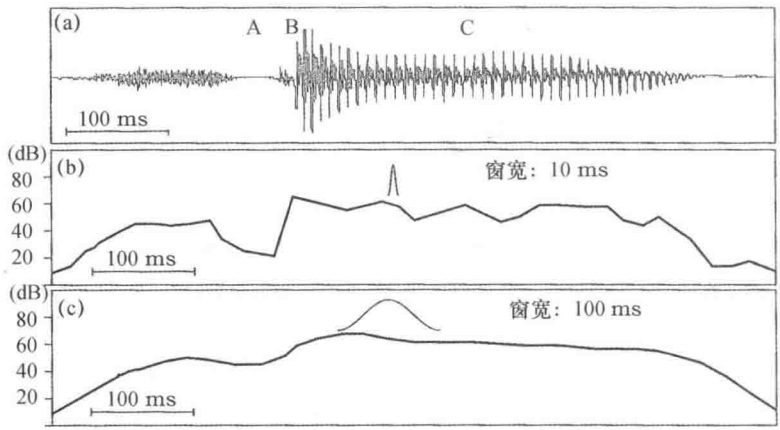


图 7.11　（a）波形图（b）和（c）为英语单词“stair”［stɛɹ］在不同窗宽下的两种振幅曲线。（b）的窗宽为 10 ms；（c）的窗宽为 100 ms。

7.3.3　相位

频率和振幅是一个波形的根本特性。但是还需要借助另外一个维度我们才能确定波形中特定点的确切位置，或者比较两个波形的相互位置关系。这个维度便是相位。从某种意义上说，相位可以告诉我们一个周期里每一个时间点的具体位置。这些点的位置在几何上用**度**（°）表示。那么一个周期和几何上的角度有什么关系呢？

前面我们讲过，波形就是一个由不断重复的周期所组成的序列。这个过程可以比喻为模拟手表上的一圈一圈往复运动的指针。分针的位置在 0 至 60 之间（60 和 0 其实是同一个位置），用几何上的度表示，则其对应的是 0°至 360°。为了标明圆周上某个特定的点，我们可以借助其离开起始位置的角度来表示，称为**相角（phase angle）**。同样地，一个周期

中的某个位置也可以用相角表示，值在0°至360°之间。

角度不仅能用来表示特定点在一个周期中的位置，还可以用来比较两个波形之间的相对位置。图7.12a和b中的两个波形的频率和振幅都相同，但两个波形并不完全相同，因为两个信号之间存在**相移（phase shift）**：图7.12b中的波形比图7.12a中的要超前70°，因为图7.12b的波形比图7.12a的波形先越过零线。

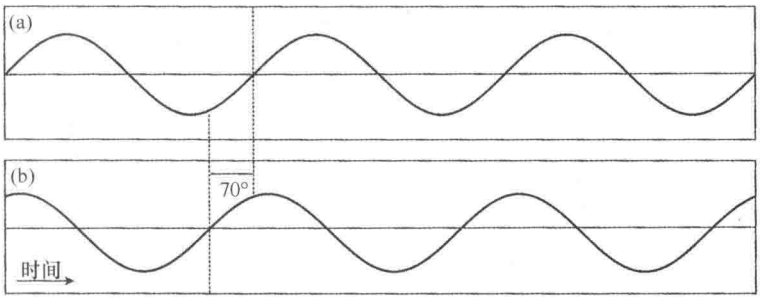


图7.12 两个振幅和频率相同的信号之间的相移。注意，图中没有标示时间标度，因为相移与信号的频率无关。

如果想对两个（简单）波形进行全面的描述，需要的参数包括它们的频率、振幅和相位。这三个参数足以从数学上全面地描述简单的波形了。

总结起来，一个信号的基频（以赫兹来测量）是由它的周期时长（通常以毫秒来测量）所决定的，基频会被感知为这个信号的音高。周期时长和周期频率之间成反比关系，周期时长减小则频率增大。信号的振幅通常被感知为响度。由于通常很难用帕斯卡值来表示绝对的振幅，振幅一般都以其与参考值之间的相对关系表示。由于响度的感知与振幅的对数值相关，通常以分贝标度来表示振幅值。最后，信号的相位关系（以度来测量）是描述信号的第三个维度。

在下一章，我们将介绍利用振幅、频率和相位分析信号的方法。由于这些操作都是在计算机上完成的，我们将介绍如何在计算机中表示信号。这也可以帮助大家理解数字录音机是如何录制信号的，因为它与计算机表达信号的原理相同。第8章对不同信号类型的介绍将覆盖语音信

号分析（第9章的主要内容）和声学建模（第10章的主要内容）所需要的所有基本内容。

练　习

1. 举例说明纵波和横波之间的区别。
2. 假设你是一位电影导演。剧中，你的女英雄驾驶的直升机在空中悬停，其战斗成员刚刚炸毁了10 km以外的一个目标。如果你计划用一个很大的闪光爆炸来表现这一情景，那么为了使影片更加真实，观众应该在多长时间后听到爆炸声？
3. 请补足缺失的数据：

周期时长		频率	
(ms)	(s)	(Hz)	(kHz)
1	0.001	1000	1
	0.05		
10			
			0.05
		1	
4			
		400	
			3
	0.1		

4. 什么是基频？听音人是如何感知基频的？为什么计算一个语音信号的基频非常困难？
5. 在波形图上如何表示振幅？听音人是如何感知振幅的？
6. 分贝标度与描述绝对振幅的标度的差异是什么？分贝标度的优点是什么？
7. 描述一个正弦信号需要哪三个维度的参数？每个维度的参数是如何测量的？

注 释

1. “h”代表百，“m”是“毫”，代表千分之一，“μ”是“微”，代表百万分之一。这些缩写在附录 B 中有详细的说明。
2. 注意，在一个波形里，分子偏离或偏向它们的静止位置的速度与分子在空间中的移动速度（即声速）并不相同（见 7.1.2）。而术语“速率”则指的是分子每秒钟通过静止位置的运动次数（见 7.3.1）。
3. 为了简便，我们在以后的计算中都采用 340 m/s。
4. 维度是我们所测量的对象。度量单位是在某一特定维度中表示该测量的一个值。例如，我们可以用米、英尺或者码这些单位来测量长度这个维度。维度是一个绝对存在的特性（即每个物体都有一个长度），而单位则是这个维度所附属的某个相对任意的特征（例如，长度可以以米、英尺、厄尔等来度量）。
5. 在基频和音高之间存在着技术上的差异。基频是物理量（定义为声带振动的频率），而音高则是音调高度的主观感知量。在这本书中，我们并不区分基频和音高。
6. 动能 $=\frac{1}{2}$质量 $\times$ 速度^2 或者 $E_{kin}=\frac{1}{2}mv^2$。
7. 如果不了解计算平方和对数的数学规则，那么很有必要知道：$(a^2/b^2)=(a/b)^2$，以及 $\log(a^2)=2\times\log(a)$。
8. “窗”是时间上的一段范围，可以长也可以短。但我们通常说窗的大小用“宽”或者“窄”。

8 语音的分析方法

在这一章，依据在语音描述中会用到的各种声学参数，我们将介绍一些语音学研究中常用的分析方法。由于这些分析通常都是借助计算机完成的，所以我们将首先讨论声学信号是如何输入到计算机中的。

8.1 声学信号的数字化

正如我们在前一章所介绍的，声音是随时间传播的声压变化序列。声压的变化可以发生在任意时间，并可在一个很大的范围内取任意值。这也就是说，在空间中存在着无限个任意的声压值，并且这些值可以出现在无限个任意的时间点上。因此，语音信号被称为是**连续的（continuous）**。在录音的过程中，麦克风的膜会随着声音信号的声压变化而运动，并产生与声压相一致的**模拟（analog）**电信号。麦克风膜的运动和由此产生的电信号都是连续的。然而，数字计算机却只能表示有限的**离散（discrete）**状态。当计算机读入连续信号以后，连续的信号会被转换为有限的**数字化的（digital）**数值。两个数字化的数值点之间的距离可以非常小，但从原理上讲，数字计算机不能表示所有可能的数值。因此，经过计算机处理，连续的模拟信号就被转换成了离散的数字信号。

我们可以通过一个时钟的例子简单地说明模拟状态和离散状态之间的区别（见图 8.1）。模拟时钟用指针表示时间。指针会连续走过表盘一周中的每一点并表示任何可能的时间。因此，即使是慢速移动的分针也可以通过它所在的位置表示每一秒钟。但是由于指针的每一个位置都代表不同的时刻，所以当我们的视角不垂直于表盘时，看到的时间就会有（稍许的）偏误。在模拟时钟上并不存在一个所谓的“安全区域”，无法

保证即使观测的位置稍微变化后，钟上的示数仍然表示相同的时间。

数字时钟用不同的数字表示时间。只有那些可以用数字表示的时间点才能在数字时钟上有所显示。例如，图 8.1b 中的数字时钟不能表示比分钟再细分的时刻（秒钟）。但是，即使从不同的角度看，我们都可以很容易地读到数字的值。这种高可靠性的数值表示方法是数字表示的主要优点。但数字表示也存在缺点，它不能表示所有的数值，一些数值的中间值也会缺失。数字化的数据有一个内在的风险：如果数据被损坏，我们就再也不能进行读数了，信息会完全丢失。（例如，如果数字显示组块的中间水平段出现故障，那么“8”可能就会被显示成“0”）

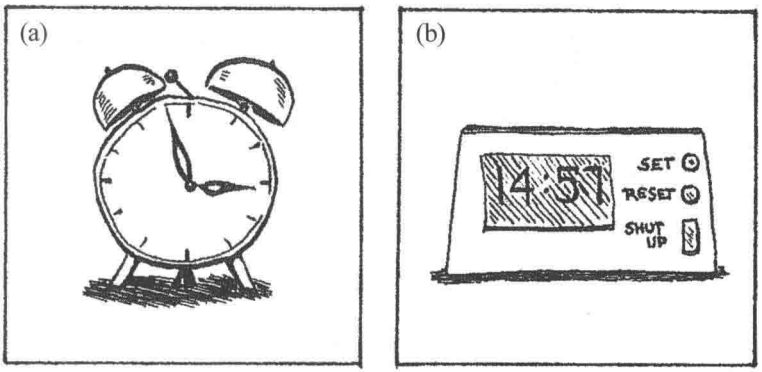


图 8.1 （a）模拟时钟和（b）数字时钟

8.1.1 时域和振幅域的数字化

在语音信号被计算机数字化的过程中，模拟世界中的两个维度会被转换为离散的数值。首先，连续变化的气压值被转换成离散的数值。其次，这种对气压值的转换是在离散的时间点上完成的。这可以用气象学家记录某个月的室外温度的例子来说明其原理。她用一个（模拟的）水银温度计、一支笔、一张纸和一只手表来观察。每一分钟，她都会观测温度计，并把那一时刻的温度和时间都记录在纸上。这其实就是一个数字化的例子：时间是连续的，但是气象学家只记录了每一分钟所观测到的测量值。而且，尽管温度计是模拟的，但是气象学家通过水银柱高度所读出并记录的观测值却是离散的。采用这种测量方法，不可避免地会

丢失一些信息，因为这样不可能记录水银柱在每一个时刻的每一个可能的高度。但另一方面，这种方法可以使她方便地存储、拷贝或者传递所记录到的数据。一旦数据被数字化，人们便可以更容易地获取并通过不同的方法来处理和存储它们。

有两种方法可以提高数字化的质量：第一是提高测量的频度，第二是提高测量的精度。气象学家的例子可以用来说明两者的优劣。气象学家测量得越频繁，温度曲线的表示就会越完整，但是她需要更多的纸来记录和存储数据。气象学家测量得越精确，每一个测量值的长度就越长，这同样也导致她需要用更多的纸。因此，我们需要找到频度和精度之间的一个平衡点。

计算机在将语音信号数字化的过程中也需要考虑同样的问题。为了得到最忠于原信号的表示，我们希望能尽可能地多次测量，并在每次测量时都采用最大的精度值。但是这样会导致数据量过大。因此，主要的问题是，在测量语音信号时需要多频繁多精确？在下面的小节中（见8.3.5）我们将看到测量过于频繁其实会对测量结果产生负面影响。

8.1.2　采样率

我们再来看一下气象学家的例子。如果她只观测几个时间点的值，那么她便无法记录到观测点之间的温度变化。另一方面，如果她观测得非常频繁，则绝大多数的温度变化都会被记录下来，但是由此产生的数据量会很大，往往没有必要。那么，为了保证不丢失有变化的观测值，在观测中应该采用怎样的测量频度呢？

我们通过一个简单的声学信号来探究这个问题。假设每秒钟从信号中采集10000个**样本（sample）**（即**采样率 sampling rate** 为10000 Hz或者10 kHz）：

$$\frac{10000\text{ 个样本}}{1[\text{s}]}=10000[\text{Hz}]=10[\text{kHz}]\text{采样率}$$

也就是说每100 μs（100微秒）就采样一次（即测量一个值）：

$$\frac{10000\text{ 个样本}}{1[\text{s}]}=\frac{1\text{ 个样本}}{\frac{1}{10000}[\text{s}]}=\frac{1\text{ 个样本}}{0.0001[\text{s}]}=\frac{1\text{ 个样本}}{100[\mu\text{s}]}$$

现在我们将这个**采样（sampling）**过程应用到图 8.2 中的一个 200 Hz 的信号上。一个 200 Hz 的信号的周期时长是 5 ms（毫秒）或者 0.005 s（见 7.3.1.2）。由于每 100 μs 采样一次，那么这个信号在它每个 5 ms 的周期内将被采样 50 次（换句话说，在这个信号的一个周期中有 50 个采样点），从而我们可以得到非常精确的数字化表示（见图 8.2a）：

$$\text{每周期测量次数} = \frac{\text{周期时长[s]}}{\text{两个样本之间的间隔[s]}}$$

在此处：

$$\frac{5\ \text{ms}}{100\ \mu\text{s}} = \frac{0.005\ \text{s}}{0.0001\ \text{s}} = 50$$

但是如果每秒只测量 300 次（即采样率为 300 Hz，两个采样点之间间隔 3.33 ms），那么我们将无法再由数字化的信号重构原始的波形（见图 8.2b）：将采样点连接起来不能得到原始信号（见图 8.2b 中的虚线）。如果采样率为 2 kHz，每个周期测量 10 次（见图 8.2c），虽然其数字化表示的精度不如采样率为 10 kHz 的那组，但是它也能相当好地表示波形了。通过连接各个采样点，我们可以较好地重构原始信号。

理论上讲，采样频率至少应为信号中最高频率分量的两倍，以保证可以将这个信号正确地数字化，这叫作**奈奎斯特准则（Nyquist criterion）**。**奈奎斯特频率（Nyquist frequency）**是采样频率的一半。在这个 200 Hz 的例子中，从理论上讲，如果想满足将信号数字化的要求，则采样频率至少应为 400.1 Hz。但在实际操作中，采样率应该大于这个理论上的最小值。

我们从前面的讨论中了解到采样率依赖于被采样信号的频率。如果被采样信号的频率高于奈奎斯特频率，采样后就可能会出现“伪”信号，这种效应被称作**混叠（aliasing）**。如图 8.2b 所示，信号的频率为 200 Hz，而采样率只有 300 Hz，所以产生了混叠效应。通过连接采样点重构的信号与原始信号并不一样，而是像一个频率更低的信号（如图 8.2b 中的虚线所示）。虽然这个由数字化的信号再生的信号并不是原始信号的一部分，但它还是有声的。我们无法鉴别该频率是否为一个错误的信号，因为在数字化的数据中没有证据表明它不属于原始的信号。换言之，再生的信号听起来并不是那种像调节收音机时产生的失真的声音；它听上去

与其他信号一样，是一个清晰正常的声音，只不过它不再是原始信号的一部分了。为了防止这种**欠采样（undersampling）**现象的发生，在对信号进行数字化之前，需要用**抗混叠滤波器（anti-aliasing filter）**对信号进行处理。抗混叠滤波器的作用是抑制那些高于奈奎斯特频率（即高于采样频率一半的那些频率）的频率分量。通常，抗混叠滤波器是集成在计算机的音频硬件中的，由软件自动控制，因此它对计算机用户来说是“不可见的”。

理论上讲，采样率应该恰好高于信号中最高频率分量的两倍。但遗憾的是，我们无法通过一个简单的标准来确切地判定到底采样率为多少时才能实现对语音信号的忠实表示。在语音实验室中，研究者们常用的采样频率有 10、11.025、16、20、22.05 或者 44.1 kHz。CD 的采样率为 44.1 kHz，数字录音机（DAT）的采样率是 48 kHz，而超级音频 CD（SA－CD）和音频 DVD 的采样率都高达 96.2 kHz。这些采样率从一定程度上说是计算机科学发展的历史产物。22.05 kHz 是 44.1 kHz 的一半，

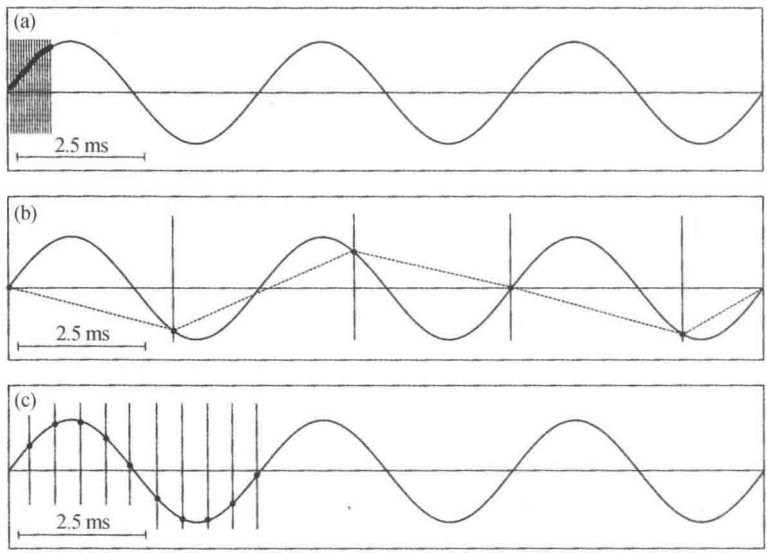


图 8.2　在时域对信号做数字化。将一个 200 Hz 的信号数字化的时候采样率分别为（a）10 kHz，（b）300 Hz 和（c）2 kHz。（b）图将每个测量值之间用虚线连接起来以表示混叠的信号。

11.025 kHz 则是其四分之一。尽管看起来采样率越高越好，但那样会需要更多的存储空间，并会导致在运算过程中消耗更多的时间，却并不能得到更精确的结果。

8.1.3 量化精度

影响数字化过程的第二个参数是振幅域的分辨率。在理想情况下，振幅的分辨率应该在能表示人耳可以感知到的最小振幅的同时，也能表示响度极大的信号。我们可以用分贝来非常方便地测量振幅（见 7.3.2）。那么多大的振幅范围（以 dB 表示）才能恰当地表示一个语音信号呢？

现代语音学的经验是在对信号进行数字化的时候将振幅的变化分为 65536 个步长。这个数值看上去很古怪，但它是 2 的 16 次方，其数学表示可写为 $2^{16}=65536$。这是 16 位二进制数字能够表示的最大整数值（这个数值有利于数字设备的计算），所以人们有时会提到“16 比特（bit）精度”。这一精度可以提供 96 dB 的振幅范围，我们可以通过 7.3.2 中提到的振幅的 dB 公式得出这个结果：

$$20\times\log\left(\frac{65536}{1}\right)\approx 20\times 4.816\ \text{dB}\approx 96.3\ \text{dB}$$

这样的振幅范围对于大多数情况来说都足够了。事实上，很少有麦克风或者录音场景能覆盖整个范围。大多数情况下，录音的范围在 35 dB 至 50 dB 之间。

图 8.3 表示的是将一个信号数字化的过程。图 8.3a 中的竖线表示对信号的测量是在时域上离散的点处进行的，在这些点处，可以测得信号的位移。然而，测量所得到的值只能用图 8.3a 中的水平线来表示。在每一个测量点处，离测量点最近的一条水平线的值被记为测量所得到的值。只有这些值才是计算机能存储的数据（见图 8.3b）。当重构这个信号时，每个样本的测量值会一个接一个地再生出来。在计算机屏幕上以图形表示的时候，通常会用线将测量点连起来（图 8.3c 中的实线），使这个信号看起来更像连续的模拟信号。但我们必须记住的是，数字信号只是由一系列离散的数据点组成的。

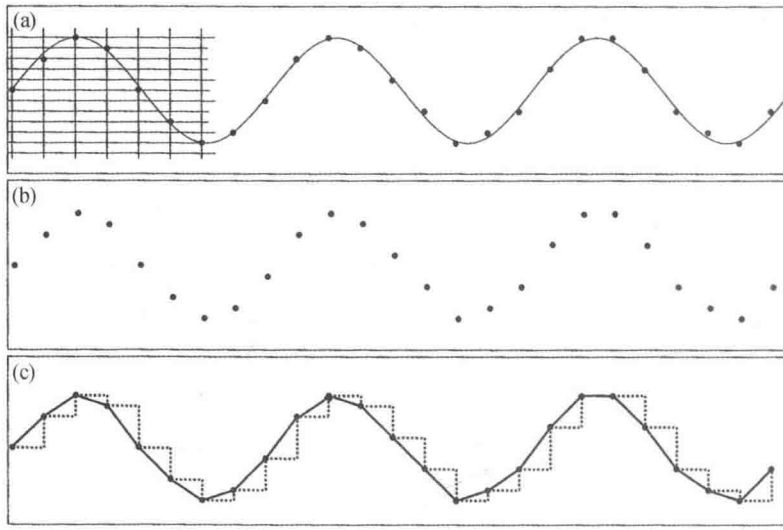


图 8.3　在振幅域和时间域将信号数字化的过程

图 8.4 展示了（语音）信号由模拟到数字的转换（模数转换）的完整过程。作为数字化的第一步，抗混叠滤波器保证了信号不会含有高于奈奎斯特频率的频率分量。在离散的时间点处，滤波后的信号被转化成离散的采样值（已量化）。这一步是真正的数字化过程，得到的是语音信号的采样点。采样率（例如 22.05 kHz）标志着采样过程在时域的详细程度，而量化精度（例如 16 bit）则标志着对振幅位移变化的取值次数。采样所得到的样本是一串数字，并可以被无限次地拷贝、传送或存储。每当将数字信号再生的时候，每个离散的样本值就会被转换成对应的电流和电压。这个数字到模拟的转换（数模转换）产生的是一个阶梯状的信号，如图 8.3c 中的虚线所示。通过使用**重构滤波器（reconstruction filter）**，这个阶梯状的函数就可以被转换为“平滑”的模拟信号了。

在上面的内容中，我们介绍了将信号数字化的基本概念。在下面的一节中，我们将介绍不同种类的简单信号并讨论一些可以将复杂的语音信号转换成这些简单信号的谱分析方法，从而使我们可以实现在波形图中无法实现的对语音信号的更为详细的研究。

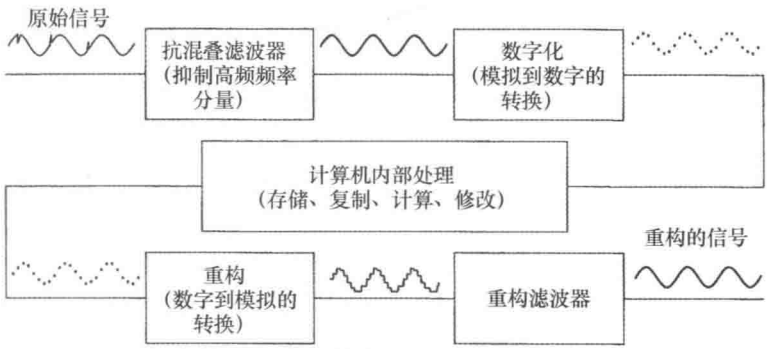


图 8.4 计算机录制和再生信号的过程

8.2 声学信号的类型

频率、振幅和相位是所有声学信号的属性。图 8.5a 至 8.5e 展示了五种不同类型的信号，我们将在下一节中讨论这些信号。

由一个简单的周期性振动（例如音叉的振动）产生的声学信号被称为**纯音（pure tone）**或者**正弦波（sine wave）**（见图 8.5a）。之所以说这种声学信号“纯”，是因为从物理的角度来看，这种信号是最简单的一种信号。任何音叉的振动或者钟摆的摆动都可以用正弦函数来表示，并且这个正弦函数完全由振动的频率、振幅和相位来确定。正如 8.3.1 中将要介绍的那样，所有信号都可以表示为许多正弦函数的组合。

包括语音信号在内的许多信号都不是“纯”的正弦信号，而是**复合信号（complex signal）**（见图 8.5b）。复合信号是周期性的，也就是说信号每个周期都会重复自己的形状。此外，我们可以通过叠加一定数量的正弦信号得到复合信号，我们将在下一节中对此作详细的介绍。

除了频率、振幅和相位之外，声音还有第四种决定性特征：音质，也叫**音色（timbre）**。声音的音色可以区分两个音高和响度完全相同的声音。不同乐器之间的声音差异是说明“音色”作用的很好的例子：即便响度一样，小号演奏出来的“中音 C”与钢琴演奏出来的“中音 C”听起来也有很大的不同。同样地，元音［ɑ］与［i］听起来也不一样。两个复合信号间的音色差异，部分取决于组成这个信号的各个正弦信号

(或者说正弦分量),以及这些信号的频率、振幅和相位关系(见8.3.1)。

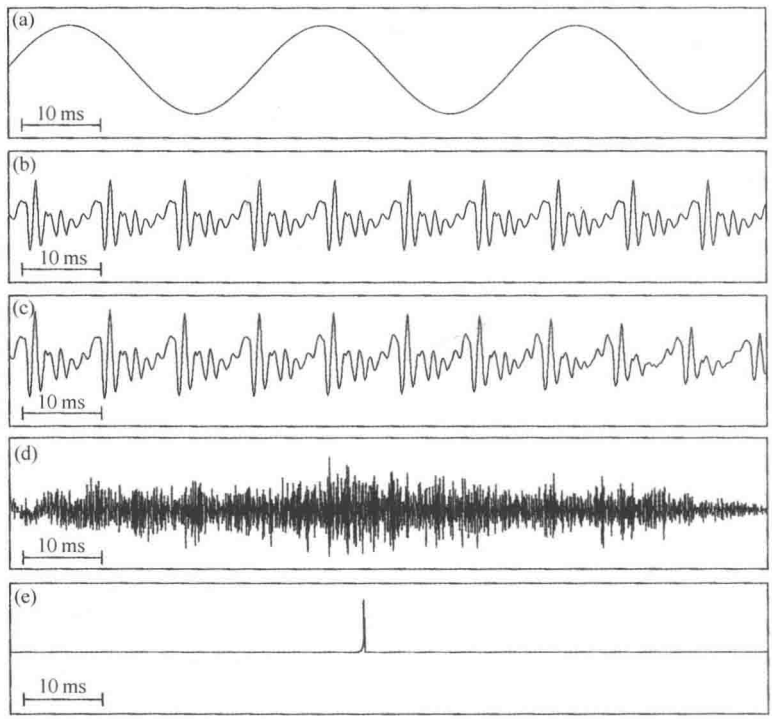


图8.5　五种不同类型的信号:(a)纯音、(b)周期信号、(c)准周期信号、(d)噪声信号和(e)脉冲信号。只有(a)和(b)是真正的周期信号,(c)为准周期信号,(d)和(e)为非周期信号。

与我们上面所看到的周期信号相比,带音信号的周期性并不那么完美。实际上,语音信号中存在着连续但微小的变化,所以在任何两个周期之间,正弦分量的频率、振幅和相位关系都不会完全相同。因此,这类信号被称作**准周期信号(quasi-periodic signal)**。一个准周期信号(见图8.5c)可能看起来和听起来都像一个复合周期信号(见图8.5b),但是信号内部存在的连续变化能够反映其特质。例如,在图8.5b和图8.5c中,其左侧的信号是一样的,然而在仅仅几毫秒之后,位于右侧的信号就开始显现出差别了。这个变化是一种渐变,所以很难确定变化具体是

从哪一个点开始产生的。由于很难找到对准周期信号的确切描述，准周期信号的一小段就被看作是具有周期性的。我们会在8.3.3中对这一判断标准作更详细的讨论。对于准周期信号来说，一定程度的周期性是存在的，即一小段信号与其相邻的信号看起来是相似的。虽然准周期信号的两个“准周期”永远不会完全相同，但我们还是可以给准周期信号指定一个周期频率（尽管从严格的物理意义上来讲这并不正确）。

然而，对于以下的两类信号来说，我们无法确定其周期频率：**噪声（noise）**信号，例如在开启装有汽水的易拉罐时，罐内气体流出所产生的咝咝声（见图8.5d），以及**脉冲（impulse）**信号，例如用锤子钉钉子时所发出的声音。在这些情况下，声音极不规则，信号中没有周期性甚至准周期性地重复自身形状的部分。我们可以说噪声信号具有一定的振幅和音色，但是它没有周期频率，信号的不同部分之间唯一的共同点就是其不规则性。脉冲信号是更为极端的非周期信号，其信号中没有一段是相似的。脉冲信号是由一个突然出现的振幅很大的无规则信号构成的，其前后部分都是无声段。因此，一个脉冲信号也不会有周期频率。

语音信号的波形（见图8.6）含有复合的周期和准周期片段，也有噪声和脉冲。图8.6a是句子“Almost everyone knows Blinda hasn't found out yet”［ʔɑ̆mɔstɛvɹʌnoʊs bəlĩnɾʌfiæznfɑ̃ʊndaʊjɛː］的波形，时长为2.4秒。图中有一些部分是振幅较高的准周期信号，主要来自带音（最可能是元音）的声门脉冲；另一些部分振幅很小，可能来自停顿、塞音或者其他能量较小的音。图8.6b是“hasn't found”［fiæznfɑ̃ʊnd］片段的时域放大图，图中显示的时间只有0.5秒。左边喉擦音［fi］与前低元音［æ］相融合，从而在波形图上表现为因摩擦而引起的振幅降低和波形变黑。浊擦音［z］可以被看作是一段相当黑的（由摩擦引起的）并且还表现出声带振动调制特点的（由嗓音引起的）信号。在音节性鼻音［n］后没有看到闭合或者爆破的现象，这说明发音人丢弃了爆发音［t］，并同时发出唇齿擦音［f］，显示为一段很低的振幅。二合元音［ɑ̃ʊ］中，［ɑ̃］比［ʊ̃］的振幅略大，而［n］显示出较“平滑”的波形，看上去与“hasn't”中的［n］相似。最后的［d］则完全脱落，只有在图8.6d那样的分辨率中才可能看到爆破段。这里由于空间有限，我们就不再显示了。所有的这些细节在图8.6a上就更难观测到了。在图8.6c中，我们放大显

示了二合元音［ãŏ］的第一部分的 50 ms，进而我们可以清楚地观测到带音的每一个周期。这些周期虽然看起来有一些“不规则”，但是每个周期的形状一次一次被近似地重复，这就是（准）周期性的本质。最后图 8.6d 显示了一段 10 ms 长的信号。在这样的视图下，我们看不到明显的周期性（因为只能看到一个周期），信号看起来也比图 8.6b 中的更为平缓。通过以上这些同一信号的不同波形图我们可以看出，选择不同的时间尺度，所观测到的信号的模样也随之改变。在不同的时间尺度下，我们可以揭示信号不同的细节信息。另一方面，如果观测尺度太小，信号的一些信息也会被掩盖，例如在图 8.6d 中我们就观测不到信号的周期性。看上去二合元音［ãŏ］的振幅在图 8.6b 至图 8.6d 中在逐渐变小，但这只是一种视觉上的幻觉，其实三幅图中所显示的峰峰振幅都是相同的。

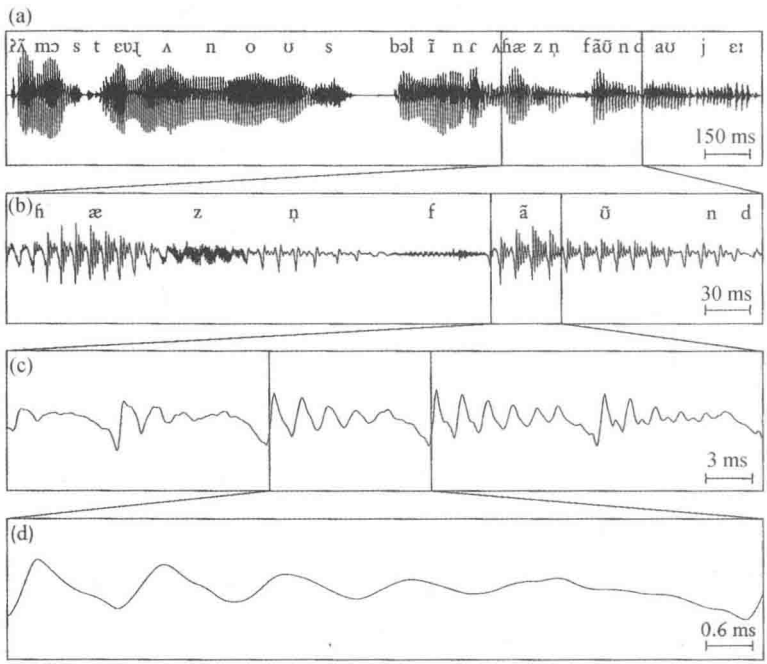


图 8.6　语音信号的波形图：（a）短语“Almost everyone knows Blinda hasn't found out yet”，时长为 2.5 s；（b）短语“hasn't found”，时长为 500 ms；（c）二合元音［ãŏ］中的［ã］，时长为 50 ms；（d）元音［ã］的一个周期，时长为 10 ms。图中的竖线标记了放大音段的位置。四幅图中纵轴的标度相同。

8.3 声学信号的分析

通过波形图，我们可以很容易地辨识图8.5中所表示的五种信号类型（纯音、复合周期信号、准周期信号、噪声和脉冲）。因此，利用波形图，我们可以对信号进行简单的声学分析（见2.3.3），也可以测量信号的时长（见6.3节）。但是如果我们想要区分语音的音质（例如区分［ɑ］和［i］），波形图的作用就十分有限了。尽管这两个元音听起来差别很大，但在波形图上看起来却非常相似（见图8.7）。波形图中确实包含了区别两个语音信号的声学信息，因为波形图是声学信号的完整表示。但是，我们从波形图上却无法直接观察到这些信息。在这一节，我们将介绍和讨论一些跟踪和量化这类波形图中不易观察到的信息的重要方法。

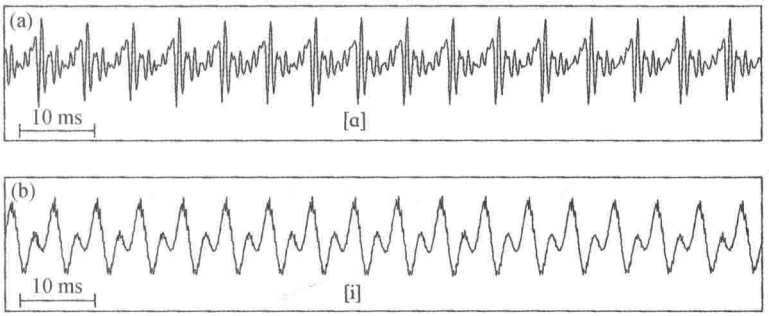


图8.7 元音波形图：（a）元音［ɑ］，（b）元音［i］。

8.3.1 傅里叶变换

在8.2节中我们提到，复合信号可以通过纯音信号的叠加来得到。在1822年，傅里叶提出，任何周期信号（不管这个信号有多么复杂）都可以分解成一系列正弦信号的组合。下面的这一部分，我们将首先介绍正弦信号是怎样相加的，以及这些正弦信号相加所得到的和是如何构成任意一个周期信号的。

8.3.1.1 信号相加：傅里叶合成

一个声学信号是一系列连续变化的气压值。当两个声波同时出现时

(例如来自两个不同的扩音器)，它们的声压值就会相互叠加。也就是说，我们可以简单地通过将两个波形的位移相叠加来实现波形的合并。

两个信号相加以后会产生一个新的信号，这个信号每一个时间点的位移都是原来两个信号在该点处的位移之和（见图 8.8）。在某一时间点，如果两个原信号的位移位于零线的同侧，则相加所得的位移值就大于两个原信号中的任何一个（见图 8.8 中的点 A)；如果两个原信号的位移位于零线的两侧（一正一负)，则相加所得的位移值就小于两个原信号中的任何一个（见图 8.8 中的点 B)；如果两个原信号的位移位于零线的两侧但位移大小相同，则相加所得的位移值为“0”（见图 8.8 中的点 C)。

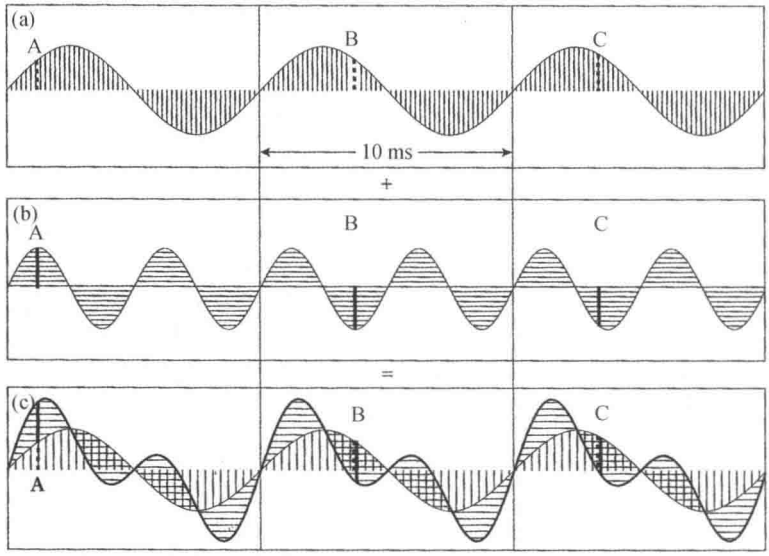


图 8.8　(a) 频率为 100 Hz 的正弦信号，(b) 频率为 200 Hz 的正弦信号。这两个信号的振幅和初始相位都相同。(c) 两个信号相加的结果为深色曲线，阴影表示的是每个信号在叠加过程中的贡献。

如上一段中讲到的原理所述，合并后信号的峰峰振幅并不一定是两个原始信号峰峰振幅的和（如图 8.9 所示)。在图 8.9a 中，两个频率、振幅和相位都相同的正弦信号相叠加，由于这两个信号的振幅最大值和振幅最小值都出现在相同的时间点，所以相加后所得信号的振幅是原来

每个信号的两倍。

在图 8.9b 中，两个信号的相位相差 180°。现在这两个信号的位移大小相同，但却分别位于零线的两侧，因此它们相加的和永远都是“0”。所以两个信号相互抵消，通过叠加所得到的信号是一条零线。这个效应在“降噪耳机”中有所应用。降噪耳机会使用一个内置的麦克风来录制周围的环境噪声信号，然后将其相位移动 180°，再与声音信号同时播放。由于噪声信号和声音信号中的噪声成分相互抵消，（在理想情况下）人们就听不到环境噪声了。

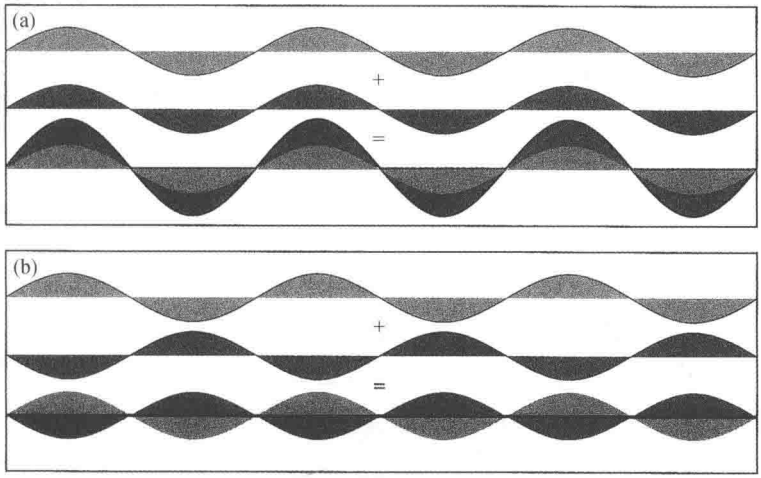


图 8.9 频率和振幅完全相同的两个正弦信号相加。（a）相位相同的两个信号相加，（b）相位相反的两个信号相加。

从图 8.8 中我们看到，通过叠加两个频率不同、振幅不同的正弦信号就已经可以得到一个十分复杂的信号了。这种通过叠加正弦信号构造复合周期信号的过程叫作**傅里叶合成（Fourier synthesis）**。通过叠加具有特定频率、振幅和相位的正弦信号，我们可以通过傅里叶合成构造出任意的周期信号。由此可见正弦信号在声学上的重要性，它们是所有复合声学信号的基础。

8.3.1.2 信号分解：傅里叶分析

上面讲到的过程也可以通过**傅里叶分析（Fourier analysis）**逆向进

行：任何一个周期信号都可以被分解为若干个具有不同的频率、振幅和相位的正弦信号。换言之，对任何一个给定的周期信号而言，只有特定的正弦信号的组合才能构成这个信号，我们可以通过傅里叶分析找到这个组合。

图 8.10 描述了傅里叶分析的过程。通过傅里叶分析，位于图中“前景”的信号被分解为四个正弦信号，即如果把这四个正弦信号叠加起来，我们就可以得到前景中的这个信号。在前景的图中，水平坐标从左到右表示时间，纵坐标从下到上表示振幅。因此，前景的图是一个波形图。在图 8.10 中，正弦信号分量的频率信息以由前向后的“景深”表示，左侧的纵轴与频率轴垂直，其高度表示相应正弦分量的振幅，而纵轴所处的位置代表了相关正弦信号的周期频率。

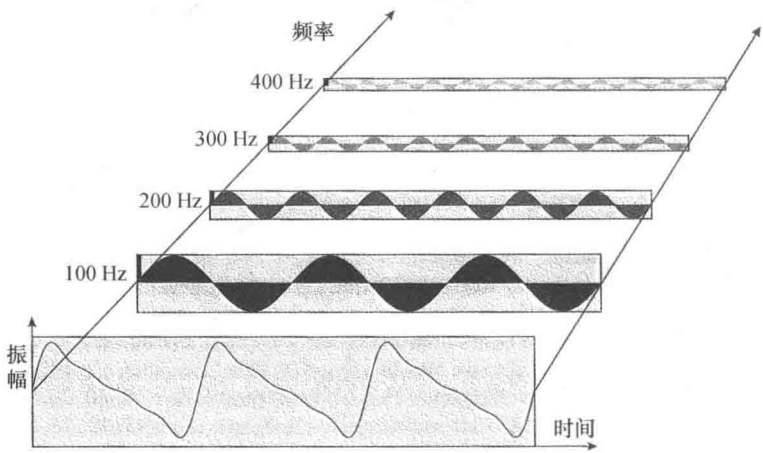


图 8.10　四个频率递增振幅递减的正弦信号相加。在上面的三维表示中，相加的结果是“前景”中的波形，而每个正弦信号和它们对应的频率在图中以“景深的纵深”表示。

在图 8.11 中，我们将表示正弦信号频率和其振幅的图形进行旋转，让频率轴旋转为从左到右递增的方向。因此，图中纵轴表示振幅，横轴表示频率。需要注意的是，谱并不能表示每个正弦分量之间的相位关系。因此，这样的图示应称为**频率谱（frequency spectrum）**。此外，频率谱通常并不表示正弦分量的振幅（例如，以帕斯卡表示），而表示其分贝级

别。这一类的图称为（**功率，power**）**谱**（**spectrum**）。在数学中，旋转90°叫作变换，因此我们将这个过程称为**傅里叶变换**（**Fourier transformation**）。

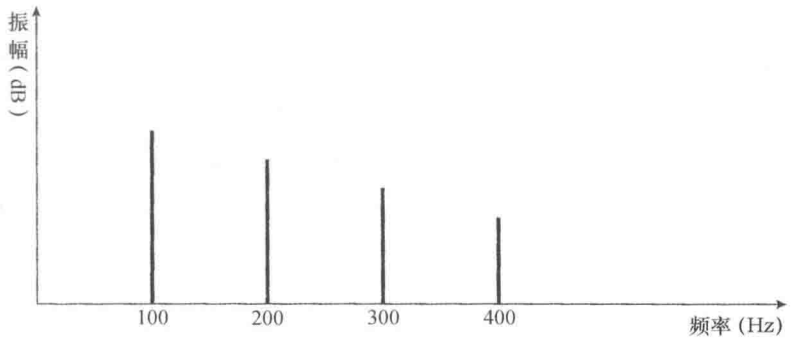


图8.11 频率功率谱。从左到右表示图8.10中随着“景深的增加”，各个正弦信号的频率和振幅值。

8.3.1.3 谐波频率

一个复合周期信号的频率谱的频率分量的频率值，通常是最低频率分量值的整数倍。例如，如果一个复合信号的周期频率（F_0；见7.3.1.4）为100 Hz，那么它所包含的正弦分量的频率只能是100、200、300 Hz等。为什么这样呢？

解释这一现象时，一个重要的因素就是所分析的信号应为周期信号。也就是说，每一个周期，波形都重复自己，与前一个周期保持一致。100 Hz的纯音信号每10 ms重复一次——也就是说10、20、30……ms后的信号都是完全相同的（见图8.12a）。一个200 Hz的信号，每5 ms重复一次（5、10、15、20、25、30……ms；见图8.12b）。如果同时考虑这两个信号，我们会发现它们每10 ms“相遇”一次。也就是说，两个信号在100 Hz信号的周期时长的整数倍的时刻相位相同。这个时长也是相加所得信号的周期时长（见图8.12c）。这个基本的周期（T_0）所对应的基频（$F_0 = 1/T_0$）称为这个复合信号的**第一谐波**（**first harmonic，H1**）。200 Hz的分量称为第二谐波H2，等等。也就是说，一个复合信号的所有频率分量都是基频的谐波：F_0 = H1，$n \times F_0$ = Hn。基频是所有谐波的频

率的最大公约数，而基频的周期时长是所有谐波的周期时长的最小公倍数。图 8.12d 中的竖线表示图 8.12c 中所示复合信号的每个谱分量。因此，这种特殊的频率谱称为**线谱（line spectrum）**[1]。

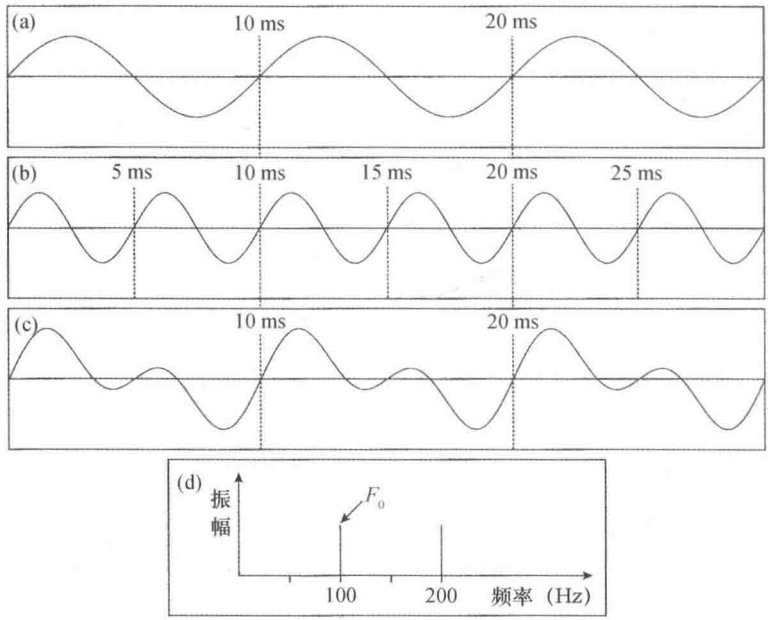


图 8.12　（a）一个 100 Hz 和（b）一个 200 Hz 的正弦信号，（c）它们相加所得的信号，（d）相加所得信号的谱。

下面看一下 100 Hz 和 150 Hz 这两个正弦信号相加的结果（见图 8.13）。100 Hz 的信号每 10 ms 重复自己一次（10、20、30……ms；见图 8.13a），而 150 Hz 的信号每 6.7 ms 重复自己一次（6.7、13.3、20……ms；图 8.13b）[2]。每 20 ms 两个信号“相遇”一次，也就是说每间隔 20 ms 它们的相位就是相同的，对应的频率是 50 Hz[3]。因此，尽管这个复合信号包含的正弦分量的频率分别是 100 Hz 和 150 Hz，我们所得到的周期频率却是 50 Hz。通过对这个信号做傅里叶分析，我们得到的只有 100 Hz 和 150 Hz 两个正弦分量（见图 8.13d）。而作为最小公约数的 50 Hz 在谱图上却完全没有显示。有趣的是，听话人确实将 50 Hz 的基频感知为这个信号的音高。例如，电话通常不能传递说话人的基频信息，

因为电话只能传输频率在 300 Hz 至 4000 Hz 之间的信号，而我们的听觉系统却能够从高次谐波的信息中重新计算基频，从而将其感知为音高。所以人们可以“听”得出电话另一端说话人的音高。

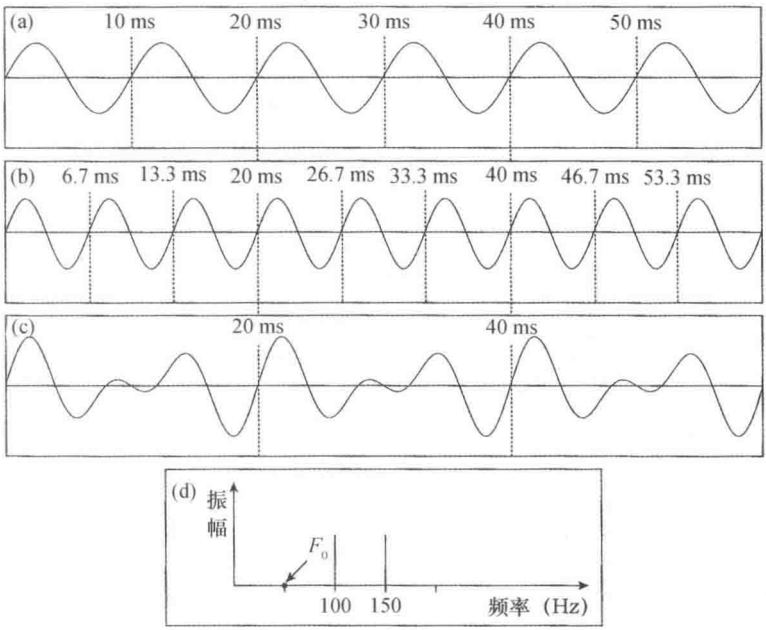


图 8.13 （a）一个 100 Hz 的正弦信号和（b）一个 150 Hz 的正弦信号，（c）它们相加所得的信号，（d）相加所得信号的谱。值得注意的是，听话人感知到的信号音高是 50 Hz，尽管这个频率在信号中和它的谱上都不存在。

应该强调一点是，谱分析（例如傅里叶变换）不会给信号增加任何信息。最好的情况也就是信号中的信息被全部保留。例如，对于一个给定的复合信号，通过傅里叶变换我们可以得到其正弦分量和它们的相位关系。但如果将傅里叶变换的结果用频率谱表示，相位关系就丢失了。换言之，频率谱所提供的信息比信号本身的要少。傅里叶变换改变的仅仅是信号中所包含信息的表达形式。与基于时间维度的波形图相比，有些信息在谱表示中更易观察。但无论如何，变换的过程永远不会向信号中增加任何信息。

8.3.1.4　离散傅里叶变换和快速傅里叶变换

严格地讲，傅里叶变换只能用于分析连续信号。有一种基于计算机分析的方法可以将傅里叶变换作用于时间上离散的值，称作**离散傅里叶变换（discrete Fourier transformation，DFT）**。

当人们最初开始使用计算机做 DFT 的时候，运算的时间很长。不久人们便提出了一种可以减少 DFT 运算时间的算法，称为**快速傅里叶变换（fast Fourier transformation，FFT）**，是 DFT 的一种特殊形式。FFT 的主要特性是其计算的点数总是 2 的幂，比如 16、32、64、128、256、512、1024、2408、4096……点。尽管现代计算机的计算能力有了巨大的提高，足以在可接受的时间内完成 DFT 运算，但是大多数计算机程序使用的还是 FFT。

总结而言，尽管人们最常用的谱分析方法是傅里叶变换，但它只是众多谱分析方式中的一种。DFT 是对离散数据的傅里叶变换（例如，在电脑中进行运算），而 FFT 只是计算机中高效实现 DFT 的一种算法。另外，FFT 或者 FFT 谱通常指的就是功率谱。

8.3.1.5　非周期信号的傅里叶变换

在对傅里叶变换的讨论中，为了保证使用傅里叶变换的合理性，我们都假设分析的信号为周期信号。但是当声学信号（例如语音信号）不再是完美的周期信号而只是准周期或者非周期信号时怎么办呢？在语言信号中永远不会出现完整的周期性，那么到底如何将傅里叶分析用于语音信号呢？在这一部分，我们将回答这个问题。

对周期信号来说，基频是信号中所有正弦分量的最大公约数。基频本身并不一定会出现在复合信号的谱中（如 8.3.1.3 中所述）。例如，如果两个正弦信号的频率分别是 100 Hz 和 101 Hz，那么叠加所形成的信号的基频就是 1 Hz，但是这个基频并不是该信号的正弦分量。理论上讲，一个复合信号的正弦分量的频率可以无限地接近，而其基频也就会无限地低。这样的话，作为无限低的频率的倍数，任何频率都可能出现在谱上，进而该信号的谱可以由无数个位置上无限接近的谐波构成。因此，不同于图 8.14a 和图 8.14b 所示的那种线谱，我们得到是一种**连续谱（continuous spectrum）**（见图 8.14c 中的右侧图示）[4]。因此每一个信号都可以看成是具有无限低的基频的周期性信号，这也说明了傅里叶变换的合理性。（这个论断看似有

点“狡猾”，但确实是数学中常用的一种论证方法。除了我们从不会在现实生活中遇到无限小的东西以外，这种论断是没有任何问题的。）

对于一个准周期信号、一个非周期的噪声信号，或者一个脉冲信号来说，所有的频率分量都可能会出现。在准周期信号的谱上我们可以观察到谐波信息（见图 8.14c），但是噪声信号和脉冲信号根本不存在谐波结构（见图 8.14d 和图 8.14e）。

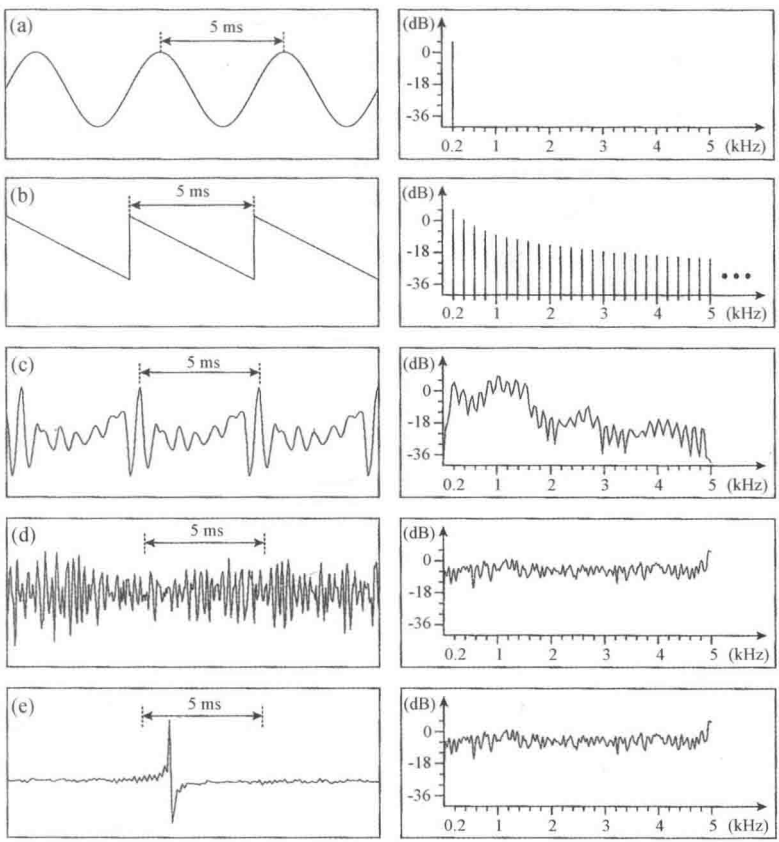


图 8.14 不同信号的波形图（左）和谱图（右）的例子：（a）正弦信号（纯音），只有一条唯一的谱线；（b）锯齿波信号是一个复合周期信号的例子，它的正弦分量为线谱；（c）元音［ɑ］是一个准周期信号的例子，它具有用连续谱表示的（伪）谐波；（d）清擦音［s］是非周期信号（噪声）的例子，它的谱是连续的；（e）爆发音［t］是脉冲信号的例子，与非周期信号一样，它的谱也是连续的，但这两种信号的区别之处在于其频率分量的相位关系（在频率谱中无法显示）。

频率谱并不能表现噪声信号和脉冲信号之间的差异：两种信号的频率分布有可能是相同的，而其差异主要体现在相位特征上。对脉冲信号来说，所有的（正弦）频率分量只有在（波形图中的）某一个时间点上是“同相位的”，即振幅的最大值同时出现；那个点之后，所有的频率分量便相互抵消。但是对于噪声信号而言，所有频率分量的相位相互之间是随机的关系，所以其波形图看起来完全是一个随机的位移序列。

8.3.2　通过谱可以观察到哪些信息

一个频谱可以清楚地反映信号中存在哪些频率以及它们的凸显性如何。图 8.14a 显示了一系列波形图和它们所对应的谱。图 8.14a 是一个正弦信号的例子。在它所对应的谱中，我们只能看到一条谱线，它代表的就是这个正弦信号的频率。图 8.14b 是一个周期性的锯齿波信号，它的谱中显示了信号的基频和信号的高次谐波，并且谐波的振幅随频率的增大而减小。因为这个锯齿波是一个完美的周期信号，所以其谐波在谱中表现为一条一条的竖线。这种锯齿波虽然简单，但却可以很好地近似描写声带振动的谱。图 8.14c 表示的是元音［ɑ］（从英语单词“cot”中提取的元音）。这个声音是一个准周期信号，因此它具有连续的谱。图 8.14d 表示的是清擦音［s］，它的谱特点是：高频区域的能量高，低频区域的能量低。此外，由于噪声信号是非周期信号，所以它的谱中观察不到任何谐波。图 8.14e 描述的是单词“cot”中的爆发音［t］，它也是一个非周期信号，并且是脉冲信号。与噪声［s］不同，脉冲信号［t］的频率分量是同相位的，所以它们的最大值会同时出现——尽管我们在频率谱上观察不到这些。

目前为止，我们已经将语音信号的基本元素都展示给大家了。一些语音信号在波形上比在谱上更容易区分，例如区分［s］（噪声）和［t］（脉冲）；而对另一些信号而言，在谱上比在波形上更易识别，例如区分不同的元音。

8.3.3　谱分析中的加窗

在对傅里叶变换的讨论中，我们首先讨论了周期信号，然后将其应用扩展到了准周期和非周期信号。在完美的周期信号中，所有的周期都

是相同的，因此一个周期就足以描述整个信号的特性了。在图 8.15 中，我们首先选择一个周期（图 8.15a），然后将其从信号中截出（图 8.15b），将其复制（图 8.15c），再得到一个重构的与原始信号相同的周期信号。因此，通过从原始信号中截取一个周期的信号，我们便可以通过傅里叶变换计算并得到原始信号的谱。图 8.15d 中得到的谱信息显示这个复合信号由 100、200、300 和 400 Hz 的分量构成。如 7.3.2.1 中所介绍的，截取一段信号的操作叫作**加窗（windowing）**，通过分析所产生的是一个**短时谱（short time spectrum）**。

正好对应一个周期的谱，通过一条条竖线表示了它的基频和谐波（见图 8.15d）。但是在语音信号中，我们无法确定一个周期的时长，因为即使在最好情况下，语音信号也只是准周期信号。此外，FFT 只能在一个特定的长度内进行运算（见 8.3.1.4），通常不能对应一个信号的周期时长。如果窗宽没有恰好覆盖一个周期的长度，会发生什么呢？

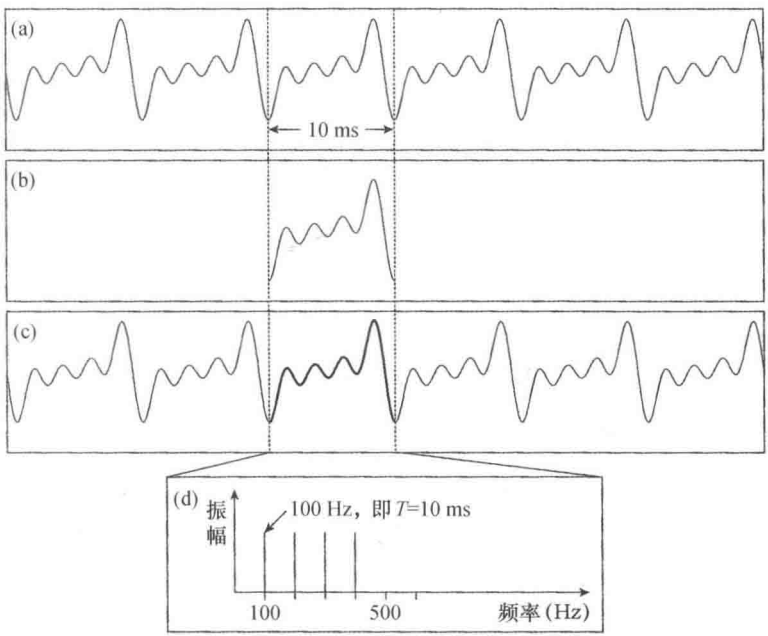


图 8.15 一个周期信号（a）可以基于它的一个周期（b）来完全重构（c）。谱（d）显示了信号（c）的谐波，与原始信号（a）的谐波完全一致。

图 8.16a 中的信号与图 8.15a 中的信号相同，其周期时长为 10 ms。在图 8.16b 中，我们截取了 12.8 ms 窗宽的信号。通过上文所述的过程，利用这一段信号重构的周期信号（见图 8.16c）与原始信号就不同了。由于重构的信号是通过复制截取的信号所得，所以其周期的时长为窗宽（12.8 ms）。重构信号中包含了一个原始周期（10 ms）之外的信号，它的谱中也含有其他不属于原始信号的频率分量（见图 8.16d）。谐波出现在窗宽大小的整数倍（12.8 ms，即 78 Hz）处，谱线在原始信号谐波（100、200、300 和 400 Hz）的附近较高。换言之，加窗使信号的谱产生了失真。那么这种失真可以避免吗?

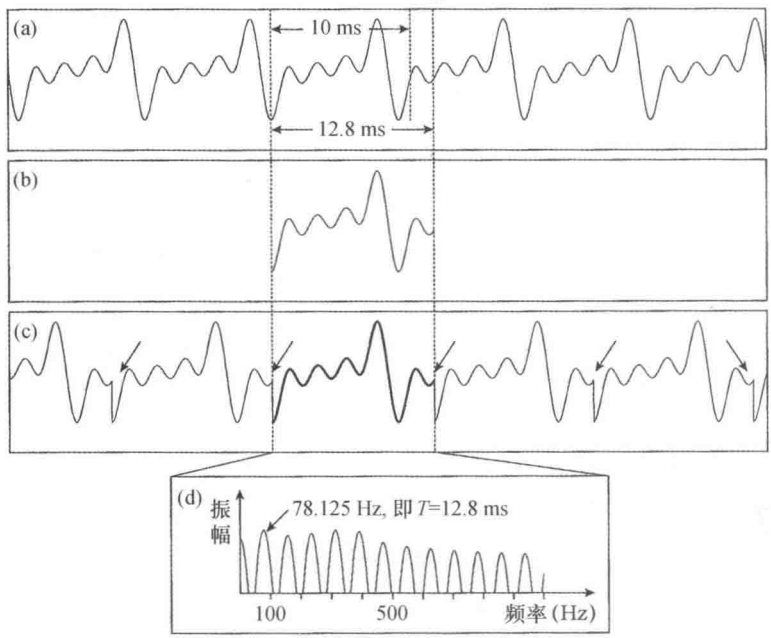


图 8.16　从周期信号（a）中截取一个信号的片段（b），与信号的周期不符。（c）为由这个片段重构而成的信号，其周期时长与所截片段的时长相符。

失真的问题是由截取的信号在窗边缘突然的跳变导致的（见图 8.16c 中的箭头所示）。在保持窗的中心振幅不变的前提下，我们可以通过减小窗信号边缘的振幅来减弱这种失真（见图 8.17）。（在 7.3.2.1 中，我们已利用这种窗来计算振幅曲线了。）

使用这种窗函数（见图 8.17b）得到的是图 8.17c 中的信号，通过复制之后得到图 8.17d 中的信号。分析后得到的谱（见图 8.17e）显示原始信号具有的谐波是 100、200、300 和 400 Hz。与图 8.16d 相比，其他频率所导致的失真被大幅降低了。窗的形状会影响谱的质量，但人们通常需要有所妥协：如果窗的边缘坡度太陡，信号中就会出现突然的跳变；另一方面，如果窗的边缘坡度太缓，就会抑制太多信号，导致分析过程中只有很少的一部分信号保持原始的振幅。理想的窗形应该是在边缘处坡度很缓，并仍保持分析窗口内大部分信号的振幅与原始信号一致。

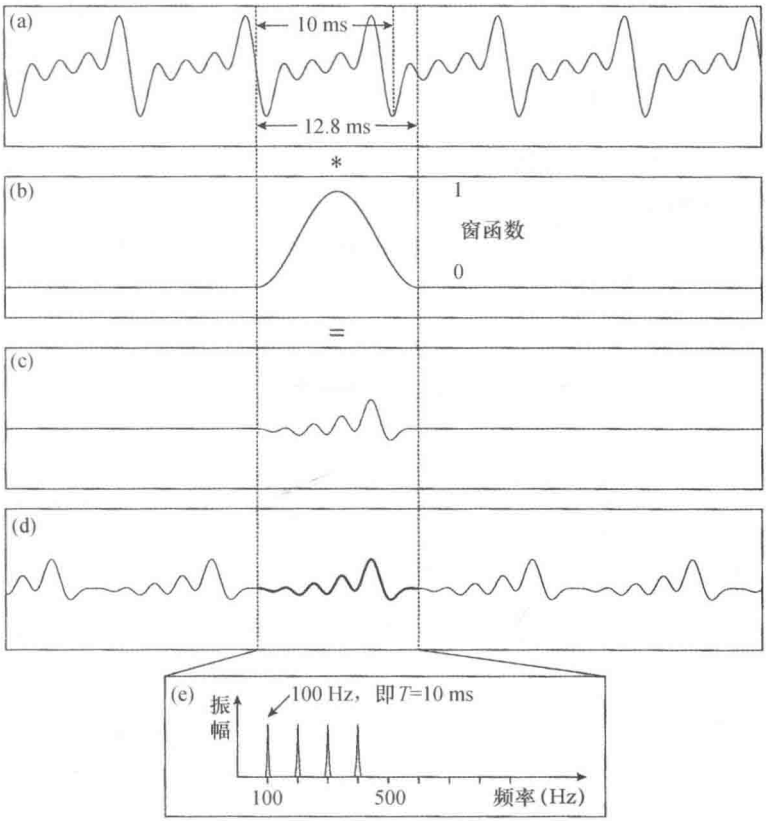


图 8.17 选择合适的窗函数以减少信号分析中的失真。

现有的窗有很多种（Harris，1978），都以窗的形状或窗的发明者来命名。语音分析中常用的窗有汉宁（Hann）窗、海明（Hamming）窗、

布莱克曼（Blackman）窗以及凯泽（Kaiser）窗，它们对大多数的语音信号都适用。一般来讲，使用较宽的窗（如汉宁窗和海明窗）可以保证在“正确”的频率附近获得较窄的频率带宽，但同时也会带来一定的失真；而使用较窄的窗（如布莱克曼窗和凯泽窗），产生的失真就相对较小。不过在实际的语音分析中，这些窗的分析结果非常接近。

通过在谱分析中使用窗函数我们还能解决另一个问题，这对 FFT 来说尤为重要。加窗以后，位于窗边缘处的信号振幅会减小到零或者接近零，所以在边缘处可以根据需要添加一些额外的“零值采样点”。从本质上讲，这样做并不会影响谱的计算，因为加窗信号的周期是保持不变的。我们可以将这个技巧用于 FFT（见 8.3.1.4），使得本来只可对采样点数是 2 的幂（……512、1024、2048……）的信号进行分析的 FFT 应用于任意长度的窗宽。例如，如果通过加海明窗提取出 300 个采样点（见图 8.18b），我们可以通过增补 212 个零值采样点使其达到 512 个点（见图 8.18c），此时就可以满足使用 FFT 的条件了（见图 8.18d）。

需要明确的一点是，通过选取适当的窗函数，我们可以减少但却不能完全消除振幅的突然跳变对信号的影响。

8.3.3.1　窗宽和谱分辨率之间的关系

傅里叶变换实际上把窗宽视为周期时长（周期时长与频率成反比；见 7.3.1.2），然后从原始信号中“取出”这段周期，并计算这个周期的振幅和相位，对应的是（窗宽的）第一谐波的信息。然后，对第二个、第三个……周期重复这个过程，得到第二、第三等谐波的振幅和相位。因此，**谱分辨率（spectral resolution）**即频率值之间的距离与窗宽（以毫秒表示）成反比：10 ms 的窗宽对应的谱分辨率是 100 Hz，20 ms 的窗宽对应的为 50 Hz，等等。因此，谱分辨率完全是由以毫秒表示的窗宽所决定的，而不是由数字化信号的采样率决定的。较高的采样率可以揭示信号的高频信息（截止到奈奎斯特频率；见 8.1.2），但并不能增加谱分辨率。只有通过增大窗宽（即，使频率值之间的间距更小）才能增大谱分辨率。

8.3.3.2　时域和频域分辨率之间的关系

如果窗宽较窄，那么所得的谱代表的是一小段语音。因为语音信号是在不断变化的，所以看起来选较窄的窗似乎很合理。但不幸的是，窗

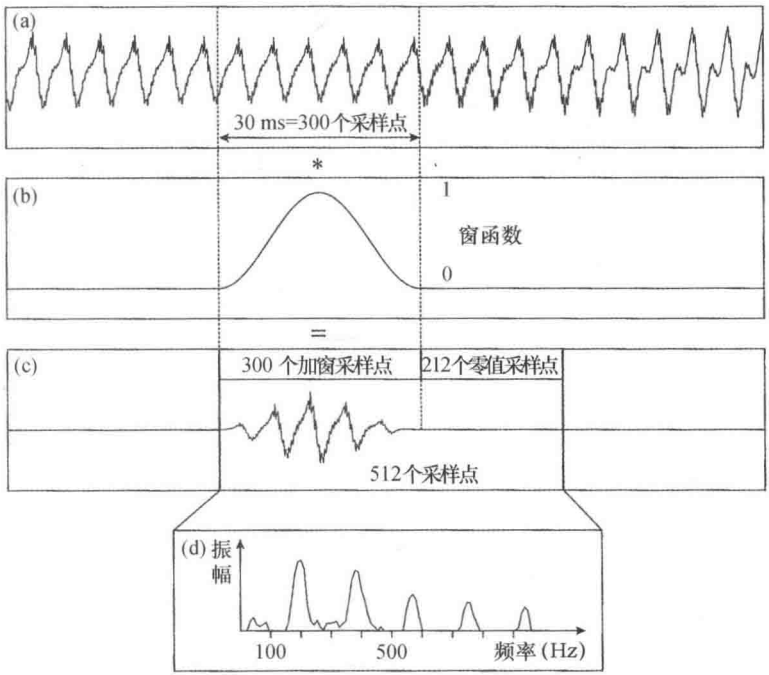


图 8.18 采样率为 10 kHz 的语音信号（a），乘以 30 ms 的海明窗（b），得到一个窗宽为 300 个采样点的信号（c）。为了得到最近的 2 的幂（512 点），需增加 212 个零值采样点。最后，通过快速傅里叶变换（FFT）得到频率谱（d）。

越窄，谱分辨率就越低（频率值之间的距离越大）。例如，如果采用 1 ms 的窗宽，谱上频率值之间的距离就为 1 kHz。通过增加窗宽，谱分辨率也会随之增加。但是由于语音信号是在不断变化的，较宽的窗可能会涵盖信号的不同部分（例如同时涵盖一个元音和一个爆发音的信息），这样会导致最终所得的谱是这两个不同部分的平均。通常情况下，较窄的窗（即时域分辨率高）对应较低的谱分辨率，较宽的窗（即时域分辨率低）对应较高的谱分辨率。

图 8.19 描述了这个原理，我们对同样的语音信号做了两次分析。图 8.19b 中的谱是通过较宽的窗计算得到的，每个频率的信息都能得到较为详细的表达。然而由于语音信号的不同部分都参与了计算，所以窗内的每一个频率分量的信息都有所失真。而图 8.19c 中的谱是通过较窄的窗计算得到的，在这样的谱中，只有几个频率分量的值可以得到区分。通常，

选取 20 ms 的窗宽可以在较好的时域和较好的频域分辨率之间找到平衡。

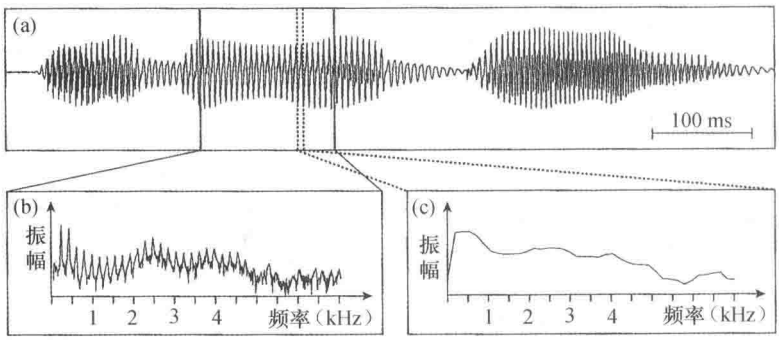


图 8.19　窗宽对谱（频率）分辨率的影响。左边的谱（b）是由窗宽为 102.4 ms 的窗计算得到的，右边的谱（c）是由窗宽为 2.56 ms 的窗计算得到的。

8.3.4　谱的其他表示形式：断面谱和语谱图

为了研究一段较长的具有高时域分辨率的语音信号，采用较宽的窗宽是不恰当的，因为这样会损失时域分辨率。我们需要通过采用其他的方法而不是选取较宽的窗宽来处理这样的信号。例如，在图 8.20 中，一系列短时谱一个挨一个地像断面一样排列在一起。在图中，横坐标表示频率，纵坐标表示振幅，时间由另一个沿景深方向“伸向图里”的坐标轴表示。图中清楚地表示了谱随时间变化的情况。由于这是一种三维的表示方法，且一系列谱看起来像瀑布一样，所以我们将这种表示方法称为**三维表示（3D-representation）**或者**瀑布式显示（waterfall display，断面显示）**。尽管这种表示形式相当好地表现了谱随时间变化的过程，但是我们不容易观察到某个频率在给定点的振幅值。此外，如果断面谱的表示中包含了太多的谱细节，我们就很难看清楚每一个频率分量的信息，某些部分的信息也可能被其他部分所掩盖。为此，我们常采用另外一种表示方式，从而可以清楚地显示不同频率区域的能量分布随时间变化的情况。现在我们就来介绍这种表示方法。

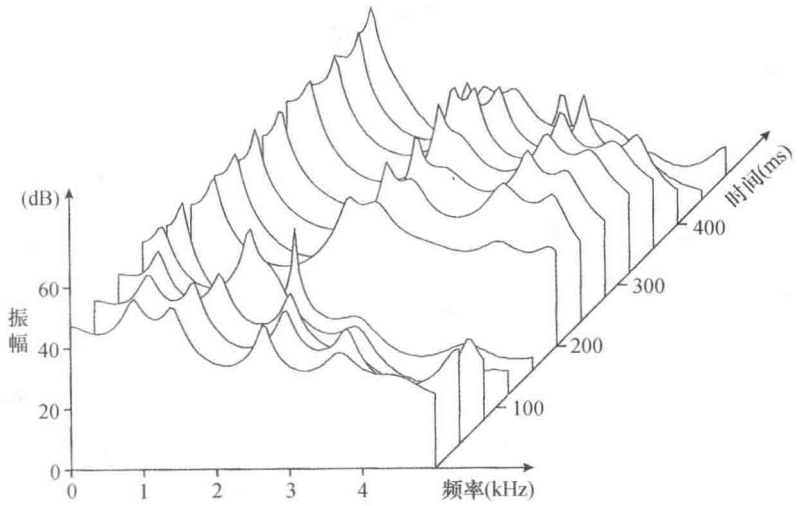


图 8.20 一系列谱的断面显示。每一个“断面”代表一个时间窗内的谱。

图 8.20 中的一系列谱看起来像“绵延的山脉”。如果像看地形图那样从上方俯视，就可以得到**语谱图（sonogram 或者 spectrogram）**（见图 8.21）。在语谱图的表示中，横轴从左至右表示时间（与波形图一样），纵轴从下至上表示频率，振幅的大小是由不同的灰度表示的：振幅越大颜色就越深，越小就越浅。通过语谱图，我们可以清楚地观察到语音信号随时间变化的情况。语谱图可以同时提供信号的频率、振幅和时间信息。

对于时域和频域的分辨率问题，在 8.3.3.2 中所得到的结论同样适用于语谱图：要么时域分辨率高而对应的频域分辨率低，要么频域分辨率高而对应的时域分辨率低。根据这样的分辨率关系，我们可以获得两种语谱图：**窄带语谱图（narrow-band spectrogram）**（见图 8.21b）和**宽带语谱图（wide-band spectrogram）**（见图 8.21c）。前者的频域分辨率高，后者的时域分辨率高[5]。在窄带语谱图中，水平的黑线可以清楚地表示每一条谐波的信息。但是，在波形图中可以观察到的［d］的除阻点（见图 8.21b 中的箭头）在语谱图中只显示为非常模糊的过渡。另一方面，在宽带语谱图上，我们可以清楚地看到除阻点的信息（见图 8.21c 中的箭头），但是各个谐波的信息却看不到了。此外，由于宽带语谱图的时域分辨率高，它能够显示每一次声门脉冲，即图中那些垂直的条纹。

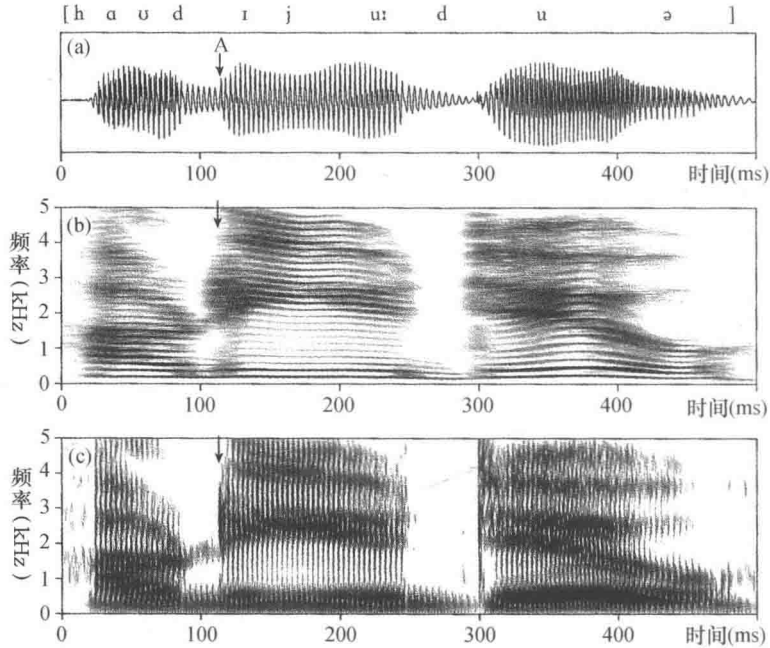


图 8.21　(a)“How do you do?”的语音信号(需要注意的是,发音人在发第一个 do 的时候并没有圆唇,而将其中的元音实现为一个前高不圆唇元音),(b)窄带语谱图,(c)宽带语谱图。

语谱图和波形图都可以表示语音信号中的信息,但是它们确实存在许多区别。有时,人们认为语谱图所包含的信息要多于波形图,但是在 8.3.1.3 的讨论中我们已经说明,没有任何一种表示方法可以提供比声学信号本身更多的信息——并且波形图已经是信号的完整表示了。看起来语谱图好像可以更为精确地表示某些信息,但应该清楚的是,我们是在计算语谱图的过程中通过使用一些参数才使它看起来是现在的样子。此外,改变灰度效果的程度也会影响语谱图中信息的可见性。一些很小的频率值在一种表示中可能被显示为白色(因而不可见),在另一种表示中却可能被显示为浅灰色。而元音音质的变化则完全是由语谱图中灰度的明暗变化所反映的。

我们应该牢记,通过计算机产生语谱图需要设置一系列参数,而这些参数往往会严重影响语谱图的样子。简单地点一下鼠标我们就可以得到语谱图,但这并不能够保证语谱图可以以最佳的状态展示数据信息。想要正确地设置各种不同的参数,需要谨慎的尝试和丰富的经验。

8.3.5 LPC 谱

还有很多其他的谱表示方式，每一种都通过其各自的参数设置来决定谱的计算和表达。这里我们来介绍语音学分析中常用的 LPC 谱：**线性预测编码（linear predictive coding，LPC）**谱。线性预测编码是利用数学方法来确定线性估计因子（预测因子）的过程。在这里，我们不会详细地介绍 LPC 复杂的数学过程（Markel 和 Gray，1976），我们只介绍其中的原理以说明如何使用 LPC 分析。对于 LPC 计算过程的详细介绍可以参看 Ladefoged（1996：181－214）。这个方法最早是为了减少在电话通话中的数据传输量而发展起来的，其目的是提高电话线的利用率。从图 8.22 中，可以清楚地看到我们可以更为有效地对语音信号进行编码：语音信号中的一小部分与后面的一部分会非常相似，后面的这部分与其后的一部分也十分相似，以此类推。基于一段信号，我们可以大概预测出下一段信号。这种预测过程在数学上可以被定义为可解的线性方程系统。一旦建立好线性方程系统，我们就不用传输信号本身，而只需传输解方程系统所需的经过编码的数学参数就足够了。

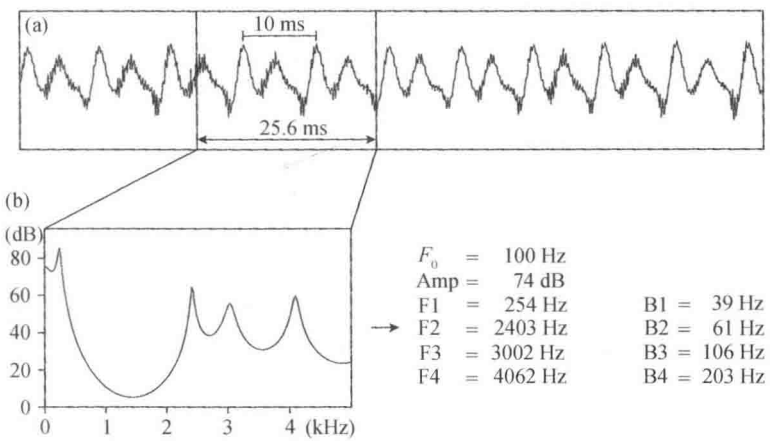


图 8.22 利用 LPC 分析，语音信号可以被转换为声带的振动频率（F_0）信息，以及一组描述声道的参数（F1、B1、F2、B2 等）和信号的振幅（Amp）。利用这些参数，我们能够预测产生语音的声道的谱（b）。

简单地讲，LPC 谱试图从谱的表示中去除声带的振动所带来的周期

性。如果从发音的角度来表述，LPC 谱试图把喉和声道分离，从而只在谱中显示声道的作用。(从这个意义上说，LPC 应用了言语产生中的声源滤波理论；见 9.4 节。) 从输出结果看，LPC 谱与 FFT 谱不同：FFT 谱表示了声带振动和声道对语音信号的共同作用，而 LPC 谱中则没有那么多细节，它只表示了声道的信息。形象地讲，LPC 只用“山峰”来描述谱信息，而不去描述山脉绵延的细节。这些“山峰”可以通过它们的位置(称为“中心频率”，即山峰最高点所对应的频域坐标) 以及峰宽 (称为“带宽”) 来进行参数化的表示。例如，位于低频区的陡峰具有较低的中心频率和较窄的带宽，而位于低频区的缓峰同样具有较低的中心频率但其带宽较宽。

LPC 的分析过程对语音学研究非常有意义，因为它的目的是把声道信息和声带振动信息分离开。因此，LPC 分析过程为我们提供了一种可以得到“干净”的声道谱的简便方法，这对元音的描写十分有效。此外，这种方法还有一个优点就是它可以进行“逆向”运算。LPC 不但可以分析语音，它还可以被用来合成语音。利用 **LPC 再合成 (LPC re-synthesis)** 的方法，我们可以在声道描写参数、振幅参数和基频参数的基础上对语音信号进行重构。这些参数经调试之后可以用于生成不同的 (人工) 说话人。

除了窗宽，LPC 计算中另一个重要参数就是“极点数”。极点数决定了计算过程中所能够覆盖的谱“峰”的个数，而“峰”的位置是一种对“共振峰”的预测。共振峰可以描写元音和其他声音的声学特征，是语音的声源滤波理论中的一个非常重要的概念 (见 9.4.1)。“峰”的个数总是极点个数的一半，即 10 个极点的 LPC 谱上最多对应 5 个峰。LPC 分析希望根据能量在特定频带的出现情况，尽量把这些峰分布到频率轴上。由于将信号数字化以后，信号的最高频率受限于奈奎斯特频率 (见 8.1.2)，所以这里会涉及极点个数与采样率之间的关系。例如，如果采样率为 10 kHz，在 0 至 5 kHz 之间 (这个范围对于语音信号特别是元音来说非常重要；见 10.1 节) 最多只能分布 5 个峰。如果采样率为 44.1 kHz，那么这 5 个峰会分布在 0 至22.05 kHz的范围内。结果可能导致只有少数或几乎没有峰会出现在对于元音非常重要的 5 kHz 以下的频段。因此，在采用较高的采样率时，最好相应地增加极点的个数。按一

般的经验规律，极点的个数等于采样频率（以千赫兹为单位）加上2，但不应大于24；也就是说10 kHz对应12个极点，44.1 kHz对应24个极点。这个规律对于10 kHz（或者11.025 kHz）采样率的情况效果很好，因为这个频率范围的大部分能量主要来源于语音信号。对于较高采样率而言，8 kHz以下的能量大部分来自语音信号，而其他的部分则主要是背景噪声。但不幸的是，我们很难预测这些“非语音”部分的信息在谱上需要多少极点来表示。因此，在语音分析中，为了简化分析过程，应采用不高于16 kHz甚至不高于10 kHz的采样率。换言之，正如我们在本章的开始所提到的那样，对于某些分析过程，较高的采样率可能并不比较低的采样率好。

本章我们介绍了各种表示和分析语音信号的方法。波形图显示了语音波形所有的声压变化，这种图形表示包含了进入我们人耳的语音信号的全部信息。然而，有一些信息我们很难通过波形图来获得，例如信号高频能量和低频能量的分布情况。如果用傅里叶谱来表示同一个信号，我们便可以从中观测到每个频率分量了。但这样做的一个缺点是所得的傅里叶谱中同时包含了声道和声带的信息。LPC谱则可以很好地表示声道的信息，尤其是元音的声道信息。但是我们需要调整某些参数才能获得LPC谱。利用语谱图，我们可以对频率分量进行时域上的分析，但这同样需要调整某些参数。通常情况下，以上这些谱分析方法可以在选取20 ms窗宽的时候达到时域分辨率和频域分辨率的平衡。

上一章介绍了基本的声学术语，这些术语可以被用于任何声学信号尤其是语音信号。下一章，我们将研究言语发音的声学特性。

练　习

1. 什么是采样率？在模数转换的过程中，如何确定最小采样率才能在重构原始信号时避免发生混叠现象？
2. 什么是混叠？为什么会出现混叠？如何避免混叠？
3. 基频和谐波频率之间的关系是什么？如果一个信号的基频是150 Hz，那么它的第一谐波、第二谐波和第三谐波分别是多少？
4. 在频谱分析的过程中，为何选择合适的窗形和窗宽非常重要？

5. 在（1）宽带语谱图、（2）窄带语谱图、（3）功率谱以及（4）波形图中如何表示基频?
6. 如果采样率为 16 kHz，那么窗宽为 25 ms 的窗内有多少个采样点?

注　释

1. 图 8.11 也是一个线谱图。
2. $150\ \text{Hz} = \frac{150}{1\ \text{s}} = \frac{1}{0.006666\cdots\text{s}} \approx \frac{1}{6.7\ \text{ms}}$。
3. $\frac{1}{20\ \text{ms}} = \frac{1}{0.020\ \text{s}} = 50\ \text{Hz}$。
4. 计算机中的“连续谱”是由离散的频率值和振幅值组成的，在数学概念上它是不连续的。但是我们在计算机分析中使用“连续谱”这个术语是为了将其与谱线只出现在基频整数倍处的线谱相区别。
5. 语谱图的命名直接与其所使用的滤波器的技术特性（见 9.3 节）相关。为了产生一个窄带语谱图，所用的滤波器只允许较窄范围内的频率通过；对于宽带语谱图，所用的滤波器允许较宽范围内的频率通过。

9 言语产生的声源 - 滤波理论

第3、4章介绍了发音器官的发音方式以及语音的标音方法，第5、6章对语音的产生进行了详细阐述，第7、8章介绍了声学及语音分析基础。在这一章中，我们会将以上知识应用于语音产生的声学方面。首先介绍一个重要概念，即语音可以被分解为一个或多个声源（source）信号，以及滤波器（filter）对这些信号的调制。

语音的声源极为有限，要么是产生浊音（带音）的声带，要么是声道中产生噪声的某个收紧部位。这种噪声可以是持续的，例如擦音；也可以是瞬时的，例如爆发音。在发颤音的时候，“颤动”的发音器官也可以作为声源。例如，颤音［r］的声源是颤动的舌尖。同一时刻可有多个声源并存，例如浊擦音，其咽部声源在声门，口部声源则在发音部位的收紧处。虽然声源数量不多，但是通过对声源进行滤波，就可以产生多种语音。这就是声道对喉部声源或声道自身声源的作用，即声道对声源进行滤波，从而赋予语音不同的特性。

这种情况就好比小号手吹出的声音，其音高或高或低，音调或响亮或轻柔，但始终带有小号的“声音”特点。号手可在号口处使用弱音器，并在演奏中移动它的位置，以改变音质。这样做并不能改变音高，仅在一定程度上改变了响度，相较而言，音质的改变则更为重要。弱音器实际上对声源，即小号的音调，进行了滤波。对于音响设备的调节也是如此，调节“低音”或“高音”旋钮并不会改变声音的音高或节奏，但相应操作对低频或高频区的增强或衰减可以改变声音的音质。

本章讨论的是调音如何使声音具有不同的声学特性。语音的声学特性取决于声源信号与特定形状声道的滤波特性。我们首先会介绍与描述和分析语音信号相关的几个概念，即共振、阻尼和滤波。

9.1　共振

用叉子敲击酒杯，酒杯便开始振动发声。如果将叉子快速移至耳旁，则可以听到叉子振动的声音，但它比酒杯的声音消失得更快。如果不敲击酒杯，而是唱出音符，当音符的音高与酒杯发出声音的音高相同时，酒杯会以同样的频率开始振动。这种一个物体对另一物体作出反应并随之一起振动的现象叫作**共振（resonance）**。每一个物体都有一个偏好的频率，这个频率是其因共振而产生振动时的振动频率，即**共振频率（resonance frequency）**。本例中的共振频率就是酒杯或叉子受到敲击后振动最强烈的频率，它取决于物体的特性（如材料、体积）。

不只是物体，管中的空气同样具有共振频率。我们可以通过对着瓶口吹气来加以证明。长气柱（空瓶）的共振频率较低，短气柱（矮瓶或满瓶）的共振频率较高。振动物体的特性决定其共振频率，同样，空气柱的共振频率由声速与空气柱振动区域所对应的管长决定。声道便是与声源信号发生共振的声管，其共振频率对语音的产生作用重大，因此我们将对其进行详细的介绍。下面我们首先介绍最简单的共振器——圆柱体声管，它大致等同于发中性元音［ə］时所对应的声道。

9.1.1　圆柱体声管的共振频率

想象一个圆柱体声管，一端开口，另一端封闭（见图 9.1），它的共振频率是多少呢？开口端的气压与周围大气压（图 9.1 中 A 点）相同，越往里移动，气压就越会增大（或者减小）。因此在封闭端（图 9.1 中 B 点），气压会达到最大值（或最小值）。气压的增大和减小如图 9.1 下半部分所示。图中的过零点对应于大气压，波形的最大（峰峰）振幅对应于声管中气压的最大变化（最大值与最小值之差）。最大变化点叫作振动的**反节点（anti-node）**，最小变化点（大气压处）叫作**节点（node）**。图 9.1 中的管长为单次振动波长（即从正弦波的最大值点到过零点的距离）的 1/4。换句话说，声波的波长等于管长的 4 倍。根据 7.3.1.3 中所述，声波的波长与频率相关（$c/\lambda = f$），因而我们可得出声管的共振频率为：

$$\frac{\text{声速}[\text{m/s}]}{4\times\text{声管长度}[\text{m}]}=\text{共振频率}[\text{Hz}] \quad \text{或者} \quad \frac{c}{4l}=f$$

因此，根据声速（空气中约为340 m/s）和管长，我们可以计算出圆柱体声管内空气柱的共振频率。例如，若管长为20 cm，则共振频率为：

$$\frac{340\ \text{m/s}}{4\times 0.2\ \text{m}}=\frac{340}{0.8\ \text{s}}=425\ \text{Hz}$$

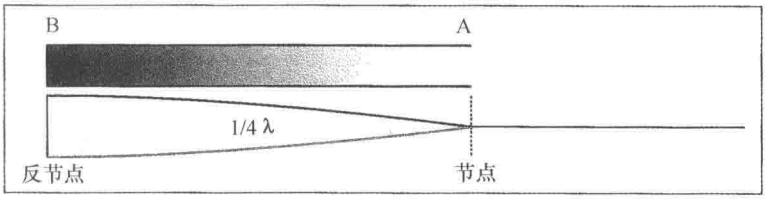


图9.1 一端开口的圆柱体共振声管内的气压变化。上：灰色越深代表气压变化越大；下：黑色线标记了气压最大值，浅色线标记了气压最小值。

管乐器的发声原理也是如此。同样地，声管的长度决定声音的频率。例如，长笛演奏者通过堵塞笛管上的小孔来产生低音符。堵住小孔后，管长增加了，共振频率的波长也随之增加，从而产生了低音。但像法国号这类的乐器与长笛不同，它并没有孔，而是一根长管。这似乎表示法国号只能产生一个乐符，但实际上演奏者依然可以演奏出多种曲调。这说明声管可能还有其他的共振频率。这些频率是怎样出现的呢？

以上讨论中，对于一端开口的管子，开口端为节点（大气压），闭合端为反节点（气压变化最大），因此波长的四分之一刚好可以匹配一个管长。但是，对于开口端为节点、闭口端为反节点的情况，不仅可以匹配1/4波长，还可以匹配3/4波长（见图9.2a）、5/4波长（见图9.2b）等。换言之，共振频率匹配的波长是管长的4、4/3、4/5倍等。由于1/4波长的奇数倍均与管长匹配，可与这些波长的波发生共振的声管称为**四分之一波长共鸣器（quarter-wavelength resonator）**。

因为频率与波长是反相关的关系（$c/\lambda=f$），当管长为1/4波长的3倍、5倍等时，其对应的共振频率一定变为原来的3倍、5倍等。我们只需要将管子的第一共振频率乘以奇数，便可得出其共振频率为：

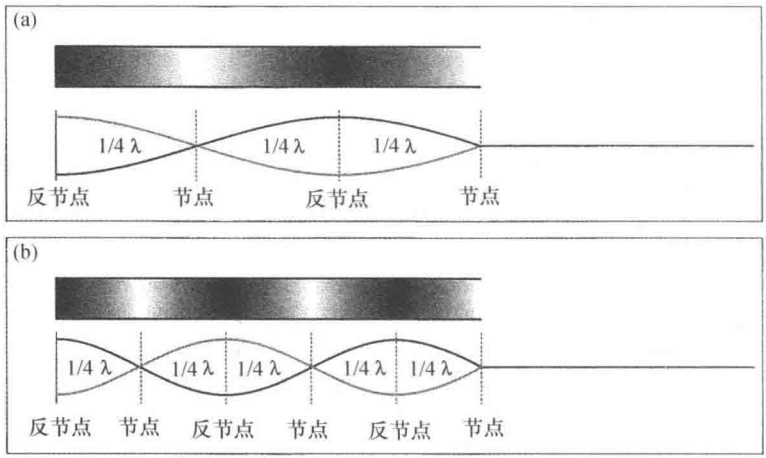


图 9.2　一端开口的高次频率的圆柱体声管

$$f_k = (2k-1) \times \frac{c}{4l}$$

其中，f_k表示第 k 个共振频率，当 k 为 1，2，3…时，$2k-1$ 为奇数 1，3，5…。

上例中，第一共振频率为 425 Hz，按照上面公式我们可算出第二共振频率（$k=2$，$c=340$ m/s，$l=0.2$ m）：

$$f_2 = (2\times2-1) \times \frac{340 \text{ m/s}}{4\times0.2 \text{ m}} = 3\times425 \text{ Hz} = 1275 \text{ Hz}$$

依此类推，第三共振频率为 2125 Hz，等等。对着管口稍微吹气通常便会产生第一共振频率，改变吹气方式便可产生更高的共振频率。

当圆柱体声管两端开口时，两端均为大气压，反节点位于声管中央。因此，1/2 波长刚好可匹配一个管长（见图 9.3a）。与上面所述原理类似，两端开口声管的共振频率为：

$$\frac{\text{声速[m/s]}}{2\times\text{声管长度[m]}} = \text{共振频率[Hz]} \quad \text{或者} \quad \frac{c}{2l} = f$$

类似地，这个管子还可以匹配 2/2、3/2 波长，等等（见图 9.3b 和 c），称为**二分之一波长共鸣器（half-wavelength resonator）**。也就是说，该共鸣器的频率所对应的波长是管长的 2/1、2/2、2/3 倍等。因此，共振频率可由公式得出：

$$f_k = k \times \frac{c}{2l}$$

一端开口的声管（四分之一波长共鸣器）和两端开口的声管（二分之一波长共鸣器）都与言语产生有关。此外，两端封闭的声管同样与之相关，见 10.2.2。两端封闭的共鸣管与两端开口的声管具有相同特性，都是半波长共鸣器。

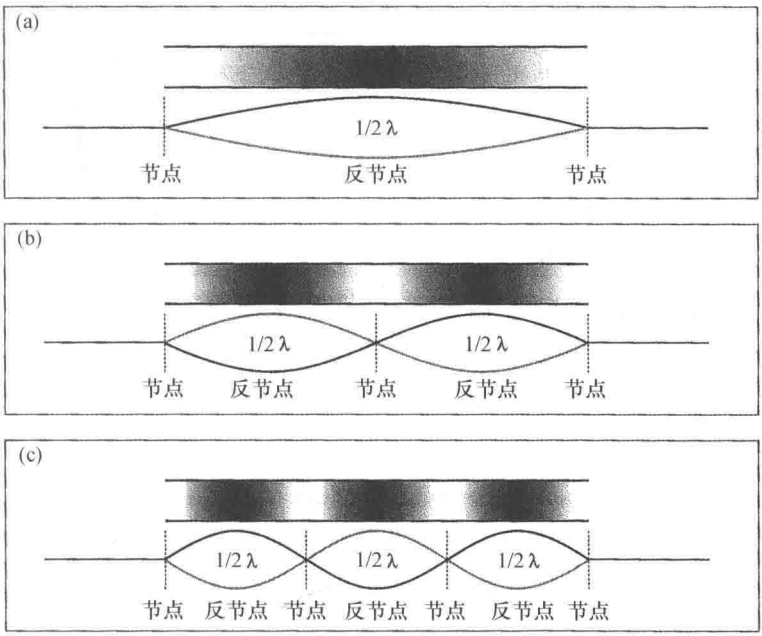


图 9.3　两端开口圆柱体声管的共振频率

9.1.2　非圆柱体声管的共振频率

共振频率不仅取决于管长与声速，还取决于声管的形状。为帮助理解，我们来详细研究一下空气分子的特性。当声波在声管里振动时，根据 7.1.2 中阐述的原理：每个空气分子会前后摇摆，但并不移进或移出声管。我们可以再次将这比作来回摇摆的一行人，人们的运动与两端封闭声管中的分子振动相似（见图 9.4）。位于反节点“B”的人会受到交替的挤压或拉伸（即“近邻密度”变化最大），但他们无须移动。只有位于反节点之间的人才会来回摇摆，位于节点“A”处的人（即反节点之间

的中央位置）运动得最剧烈。

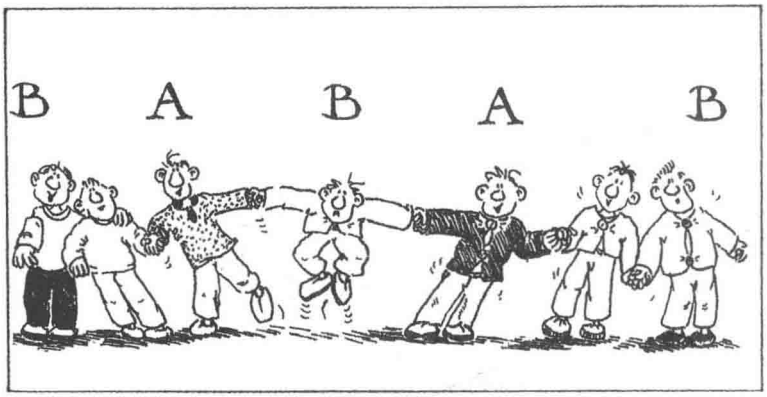


图 9.4　两端封闭声管中的分子摆动（气压情况与图 9.3b 所示相反）

如果在节点处存在收紧，这些人（或分子）的运动就会受到干扰，那么摆动（振动）的时间会变长，导致共振频率降低。如果在节点处存在放松，则人（或分子）变得更易于运动，进而共振频率提高。如果在反节点处存在收紧，达到最大或最小密度所需的人（或分子）数会变少，振动会变快，从而频率升高。同样，如果在反节点处存在放松，则达到最大或最小密度需要的人（或分子）数会变多，从而频率降低。如图 9.5a – d 所示，由于在闭口端均有反节点，开口端均有节点，上述情况适用于推测任何开口或闭口声管的共振频率。

另一方面，如果收紧部位位于声管内部的某处，则会对不同的共振频率产生不同的影响。以单端开口声管的第一、第二共振频率为例，声管中央的收紧点（见图 9.5f）对第一、第二共振频率不会有太大的影响，因为收紧点位于节点与反节点的中间[1]。如果收紧点靠前（见图 9.5g），第一共振频率会稍微降低，因为收紧点更接近节点；第二共振频率则会提高，因为收紧点在反节点位置。反之，如果收紧点靠后（见图 9.5h），则第一共振频率会稍微升高（收紧点靠近反节点），而第二共振频率会降低（收紧点在节点位置）。

类似地，我们可以算出声管形状对第三或更高共振频率的影响，且对两端开口或封闭的声管同样适用。这种解释共振频率改变的方法称作

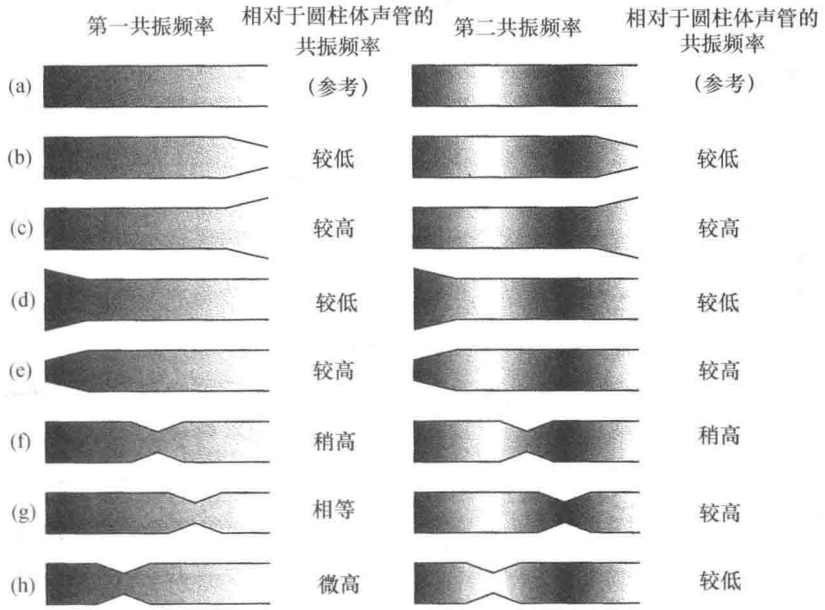


图 9.5 声管形状对共振频率的影响

干扰理论（perturbation theory）（Chiba 和 Kajiyama，1941）。通过该理论，对于任意位置收紧的声管，我们都可以较容易地计算出其共振频率相对于无收紧的圆柱声管是高还是低。将干扰理论应用于元音的产出，如果收紧部位已知，我们就可以根据该理论预测每个共振频率相对于中性元音的增减变化（见 9.5.1）。

对于存在收紧部位的声道，还有一种计算共振频率的方法，就是将其看作是多段简单圆柱体声管的组合，这些圆柱体声管可以是一端开口、两端开口或者两端封闭的。10.2 节将使用该方法计算几种辅音声道形状的共振频率。

正是声道与共振频率的这种依赖关系使人类得以产出各种各样的言语声。在说话时，通过嘴、舌以及下颌的连续运动可以改变口腔的形状，进而改变口腔的共振频率。这可以使口腔以不同的方式对声源信号进行滤波，从而产出不同的言语声。

我们在本章开始时提到，酒杯与叉子的振动时长不同。下一节将引入阻尼的概念来描述这一现象，这一概念与共振频率密切相关。

9.2　阻尼

每个物体都有特定的共振频率。两个物体不仅共振频率不同，而且它们在达到同等激活程度后的振动时间也会有所不同。例如，酒杯可以长时间地振动，而叉子在受到敲击后的振动却十分短暂。因此我们需要另一个参数来描写这种共振特性：**阻尼（damping）**。这一参数可提供振动的时长信息。

阻尼是对波形衰减速度的测量，取决于振动物体的材质。例如，敲击一只玻璃酒杯，敲击的能量转化为玻璃的振动。酒杯的弹性特征决定它仅以一个频率振动，因此振动会持续很长时间，直至敲击的能量转化为声学能量。而叉子则以若干频率振动，这些频率分布于“固有频率”或“中心频率”周围（见9.3节）。能量分布在这些频率上，一部分会相互抵消，因而振动会快速消失。也就是说阻尼（振动消失的快慢）与物体的振动匹配共振频率的精度反相关。低阻尼总是对应高精度的共振频率和持久的振动，而高阻尼则对应低精度的共振频率以及较短的振动时间。

使物体（或空气柱）振动不仅可以通过敲击来实现，还可以将其置入其他振动环境中以使其振动。一位训练有素的歌唱家可以唱出与酒杯共振频率相同的音符使其振动。歌声所具有的几乎完美的周期波可以使酒杯振动，这一现象称作**交感共振（sympathetic resonance）**。歌唱家的歌唱频率结合酒杯的共振特性，可以增大在共振频率附近的频率的振幅。同样的原理，可将声道看作是具有一个或多个共振频率的声管，该声管还有一定程度的阻尼，而阻尼的大小则取决于声道的形状。在声源的频率分量中，与距离共振频率较远的频率分量相比，接近声道共振频率的频率分量在通过声道时受到的阻碍较小。这是一种滤波现象，将在下一节中进行介绍。

9.3　滤波器

当一个具有多种不同正弦频率分量的声音通过声管时，只有与共振频率对应的分量得到保持，其他频率均被衰减。声管自身并不产生频率

（向瓶中吹气时便是这样），而是对现有（声源）信号的某些频率进行**滤波（filter）**。我们可以把滤波器比作一只筛子，只允许细沙通过，更大的石粒则留在筛子里。同样地，对于声道而言，滤波器允许某些频率通过，而其他频率则难以通过或无法通过。

在5.2.2中讲到，位于喉部的声带进行的是一种复杂的振动，从而产生复合信号。根据8.3.1.3中所述，一个复合信号包含除基频以外的多个高次谐波频率。声道的形状允许处在某些频域的谐波较容易地通过，而将其他频域的谐波从信号中滤除。这就可以解释为什么喉部的声源信号可以被调置为［ɑ］或［i］了。这两个音的喉部声源信号相同，但发音器官位置的改变使声道的形状有所不同。因此，滤波器，即声道形成的声管系统，发生了变化，进而使一些频率容易通过，另一些不易通过。这一过程决定了产出信号的性质，而这一信号会进一步经双唇辐射而出。

为理解怎样从声学原理方面对此进行解释，我们下面将详细讨论滤波器的声学特性，并介绍一些与声道声学表现接近的复杂声管系统。在第10章，我们将利用这些知识从声学的角度对语音进行描写。

由于滤波器会影响信号中不同频域的振幅（从而影响强度），因此用谱来描述滤波器的特性（即共振频率和阻尼特点）非常合适。

图9.6为一个滤波器的谱，该滤波器抑制高频分量，允许低频分量通过。这类滤波器叫作**低通滤波器（low-pass filter）**。它有两个重要的参数，即截止频率和陡度。**截止频率（cut-off frequency）**是一种频率边界，指频率分量刚好被滤波器的抑制作用影响时，该分量所处的频率。图9.6所示滤波器的截止频率为100 Hz，当然也可是其他频率。**陡度（steepness）**是指大于截止频率的频率分量的强度的**衰减（attenuate）**速率。

基于物理上的限制，低通滤波器不会使所有截止频率以下的频率分量完全无衰减地通过，也不会完全消除截止频率以上的所有频率分量。滤波器的**通带（pass-band）**（频率分量可无障碍通过）与**阻带（stop-band）**（频率被衰减）之间是逐渐过渡的。因此，滤波器的截止频率定义为频率分量的强度衰减3 dB时所对应的频率（关于强度与振幅之间的差异见7.3.2和附录A.2）。对于高于截止频率的分量，滤波器并不是按每赫兹衰减一定值的方式工作，而是以频率每翻倍或减半一次的间隔进行

衰减。由于频率的“翻倍”或者“减半”与音阶抬高或者降低一个倍频程相同，高于截止频率部分的陡度以**分贝每倍频程（decibels per octave，dB/oct）**来测量，即信号振幅在每个倍频程（对于低通滤波器，每次频率翻倍；对于高通滤波器，每次频率减半）减小的分贝值。图 9.6 中滤波器的陡度（即斜率）为每倍频程 6 dB。

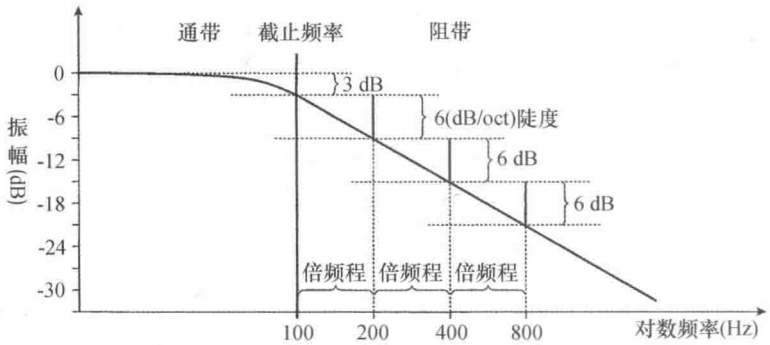


图 9.6　斜率为每倍频程 6 分贝的低通滤波器。频率用对数标度表示（见 12.5.2）。

对于**高通滤波器（high-pass filter）**，图形正好相反：大于截止频率的频率分量几乎无阻碍地通过（即高频分量通过滤波器），而截止频率以下的频率则按滤波器的陡度发生衰减。也就是说，低通滤波器衰减高频，高通滤波器则抑制低频。

低通和高通滤波器组合而成的滤波器叫**带通滤波器（band-pass filter）**（图 9.7 为一个带通滤波器，中心频率为 800 Hz）[2]。在这样的滤波器中，高通滤波器衰减低频分量，而低通滤波器衰减高频分量，只有在两个滤波器重叠区域的频率分量可以通过带通滤波器。这类滤波器的一个重要特征是两个截止频率之间的宽度。由于滤波器在截止频率处会将振幅衰减 3 dB，这一特征称为**3 dB 带宽（3dB bandwidth）**，这一特征用赫兹表示，因为它表示的是谱上以赫兹度量的滤波器的“宽度”（见图 9.8）。宽的带宽表示滤波器允许较大的频率范围通过，反之，窄的带宽表示滤波器几乎衰减所有频率，仅有很小的一段频率范围除外。带通滤波器的另一个重要参数是**中心频率（center frequency）**，是滤波器的“中心”，中心频率处的频率衰减最小，频率分量可最大限度地通过滤波器。

图9.8中，中心频率为 x 的带通滤波器具有较宽的3 dB带宽，而中心频率为 y 的滤波器的3 dB带宽较窄。

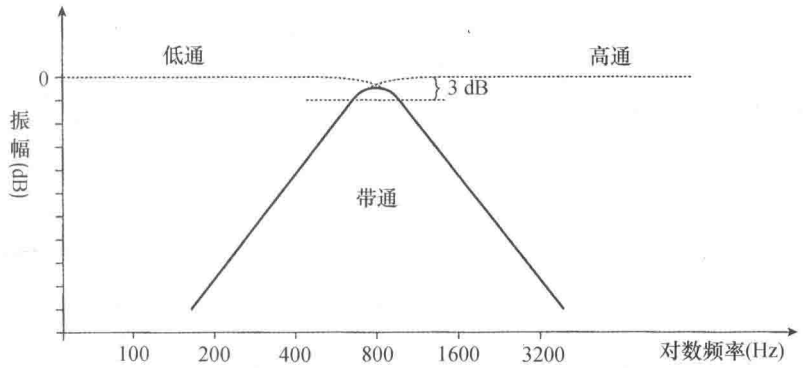


图9.7 一个低通滤波器和一个高通滤波器（虚线）叠加构成的带通滤波器（实线）

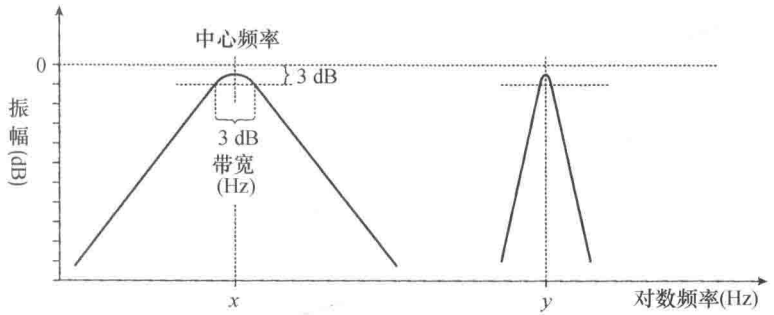


图9.8 中心频率和3 dB带宽不同的两个带通滤波器。左侧滤波器的中心频率为 x 且3 dB带宽较宽，右侧滤波器的中心频率为 y 且3 dB带宽较窄。

目前为止，用声学模型来描述发音器官所需的全部内容，至少对于具有相对打开的声道和闭合的腭咽口（velopharyngeal port）的浊音，已经介绍完毕。(1) 声带的复杂振动会产生一个具有许多谐波的声源。(2) 声源信号通过声道时，由于声道系统的共振频率和阻尼特性，某些频率与其他频率相比衰减较小，即声道对声源信号进行了滤波。(3) 信号最终从双唇扩散至空气中，并通过环境中气压的变化进行传播。下面我们将更为详细地研究这三部分的内容。

9.4　发音器官的声源和滤波器

如果声带处于合适的位置，外呼气流就能使其振动，产生如图 9.9a 所示的波形。这个信号的谱特性是其谐波以 12 dB/oct 的速率衰减，即谐波的振幅每倍频程降低 12 dB。这里需要注意，以 dB 表示的信号振幅是相对于参考值而言的，这种情况下，参考值为声带完全闭合时位于声带后方的气流，接近于 0。因此，声门打开时的气流与声门闭合时的气流之间的关系可以产生第一谐波振幅为 164 dB 的信号[3]。假定声带的振动频率约为 250 Hz，那么高一个倍频程的谐波频率大约为 500 Hz，其振幅降低 12 dB，即 152 dB。再高一个倍频程（第四谐波大约为 1000 Hz）的振幅为 140 dB，依此类推（见图 9.9b）。因此，16 kHz 谐波的振幅为 92 dB。即使喉部产生的第一谐波的声级远低于 164 dB，对于谱衰减为 12 dB/oct 的信号，在 16 kHz 处的强度依然会降低 66 dB。换言之，尽管基频只有 250 Hz，喉部产生信号在高频分量中仍携带相当大的能量。声带的复杂运动（见 5.2.2）产生了复合信号，而其能量则主要由远大于基频的频率负载。声源信号的这种“丰富性”使我们可以在声道滤波的作用下，利用相同的声源信号产出多种不同的语音。

9.4.1　声道滤波器

发中性元音［ə］时，声道可以近似为一个圆柱体声管，一端开口（双唇），在声门处几乎闭合。声管的声学特性由长度和宽度决定。我们可以将声道视为四分之一波长共鸣器（见 9.1.1）。声管的弯曲对声学特性几乎没有影响：无论声管是长直圆柱体还是螺旋体，均不会影响其声学特性。假设声道长度（即声管长度）为 17 cm，声速为 340 m/s，我们便可计算出声道产生的中性元音的共振频率：

$$f_k[\mathrm{Hz}] = (2k-1) \times \frac{340\ \mathrm{m/s}}{4 \times 0.17\ \mathrm{m}} = (2k-1) \times 500\ \mathrm{Hz}$$

也就是说，该形状声道的共振频率为 500 Hz、1500 Hz、2500 Hz、3500 Hz、4500 Hz，等等。声道的共振频率非常重要，称为**共振峰频率（formant frequency）**，分别按编号命名为 F1、F2、F3 等。但要注意，该

编号序列与基频（F_0）无关：基频（F_0）是声带振动（声源）的特性，而共振峰频率（F1、F2、F3 等）是声道（滤波器）特性。

从声学角度看，声道包括口腔和鼻腔的软组织，所有频率范围内的能量都会被衰减，因此声道是一个“糟糕”的声管。只有共振频率自身的衰减程度较小，在谱上表现为最大值。前面提到，阻尼越大，带宽就越宽，也就是说声道的每个共振频率都覆盖较大的范围。用声学术语来说，声道是一个带通滤波器系统，可由中心频率和 3 dB 带宽描述。

9.4.2 唇和鼻孔的辐射

空气分子一旦离开口和鼻，它们的运动便与房间中的庞大气体相遇。这一现象可比作一群穿过一条长走廊（气管）的行人。走廊上有一扇门（声门），门的开关交替进行。这使人们在通过门之后以组为单位前进，每组的人数取决于门打开的时长和宽度。人们继续前行，通过宽窄不一的走廊（声道），一直走到走廊的终点（双唇），这时，他们遇到了静止的一大群人（房间里的庞大气体）。在进入人群时，他们会撞到站在那里的几个人，然后这几个人又撞到与他们相邻的人，如此继续下去。就这样，一个小规模的运动便在人群中扩散开来。当遇到一大群人时，一小组人比较容易融入人群，而若是一大组人则难度就会增大。在本例中，小组人对应高频，大组人对应低频。

这一现象表现为，当信号离开声道进入外部空气时，高频部分会受到 6 dB 每倍频程的“照顾”。也就是说，声音由双唇投射出口腔时，高频振幅比低频增加大约 6 dB/oct。

图 9.9 是声道作为滤波器对喉部信号进行滤波的原理示意图。如图所示，喉部信号有许多高次谐波（见图 9.9a），其衰减速率为 12 dB/oct（见图 9.9b）。信号通过声道，（假定）声道的共振频率为 1500 Hz 和 3500 Hz（即声道可由以该频率为中心频率的带通滤波器系统表示；见图 9.9c）。信号最后由双唇离开声道，高频增幅为 6 dB/oct（见图 9.9d），并最终以声波的形式在空气中传播（见图 9.9e）。改变声道“管子”的形状，便可以改变滤波器的中心频率和带宽，进而能衰减不同的频率区域。因此，声道特定的共振频率决定了语音信号特定的谱形状，从而产生了相互区别的语音。

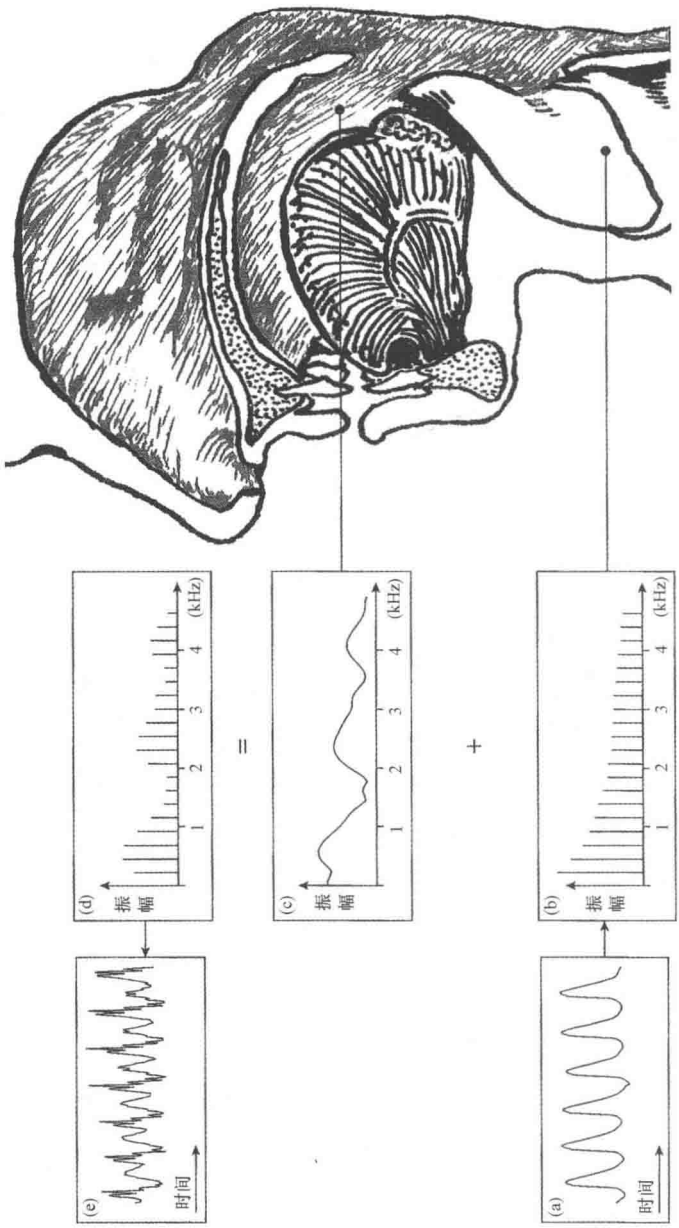


图9.9（a）喉部信号，（b）喉部信号的谱，（c）声道滤波器的谱，（d）语音的谱，（e）语音信号。

9.5 共振峰

共振峰（见9.4.1）是声道自身的特性，与是否存在喉部声源信号无关。这一点可通过一个小实验加以说明。例如，将发音器官调整到发［o］的位置，但不要从肺部呼出任何气流。此时，用手指轻敲喉部或者下颌，便可以听到［o］的发音。如果将发音器官调整到［ɑ］，轻敲喉部便可听到［ɑ］。声道的形状决定了共振峰，而与是否存在声源信号无关。声源可以是声带的振动，也可以是对喉部的敲击。

即使存在声源信号，共振峰的位置也与声带振动的基频无关。共振峰由发音器官的位置决定，并不属于语音信号本身的特性，因此共振峰频率并不总与喉部信号的谐波相对应（见图9.10）。例如，共振峰可以是400 Hz，而声源信号的基频是90 Hz，因此基频的谐波（F_0 = H1，H2，H3，H4，等等）是90，180，270，360，450，540，630，… Hz，但没有恰好为400 Hz的谐波（图9.10a）。但是，第一共振峰频率是400 Hz，因为它只由声道形状决定。对于谐波密集的信号（基频值低），单个谐波接近或者与共振峰一致的可能性较大。对于较高的基频，共振峰频率接近谐波的可能性较小，共振峰频率可能在声学信号中并不存在。这就是听高音时（例如女歌唱家和儿童；见图9.10b）难以区分不同元音的原因，因为人们此时难以辨别相应信号中的共振峰。

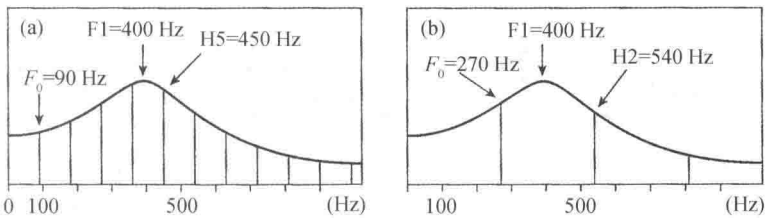


图9.10 （a）低基频与（b）高基频信号的谐波和声道频谱

9.5.1 共振峰频率

发音器官的位置决定了共振峰所处的位置。由于共振峰的频率取决

于声道的形状，因此，关于发音器官的位置如何影响共振峰的频率，我们可以根据干扰理论（见 9.1.2）归纳出一些普遍规律。

我们以几个元音为例进行说明。例如，元音［ɑ］的收紧位于咽部，接近图 9.11 中的 A 点。对于 F1，收紧点的气压更接近反节点而不是节点，根据干扰理论，收紧点接近反节点，共振峰频率抬高（相对于中性元音，图中用“+”表示），因此［ɑ］的 F1 较高。而对于 F2，咽部收紧点更接近节点，根据干扰理论，收紧点靠近节点，共振峰频率降低（图中用“-”表示），因此［ɑ］的 F2 较低。元音［i］的收紧点接近硬腭，接近图 9.11 中的 B 点。对于 F1，该点接近节点（双唇处）而不是反节点，因此我们可以预测其 F1 较低。对于 F2，该点接近反节点，因此我们可以预测其 F2 较高。这些预测可通过比较［ɑ］和［i］（或任何其他元音）与中性元音［ə］的共振峰频率值进行验证（见附录 C.1 的英语元音共振峰频率均值表）。按照这样的思路，图 9.11 展示了利用干扰理论所得的一些预测结果：相对于由无收紧的圆柱体声管所产出的类似中性元音的音而言，收紧点的位置对 F1、F2 和 F3 取值高低的影响，以及这种影响的程度。

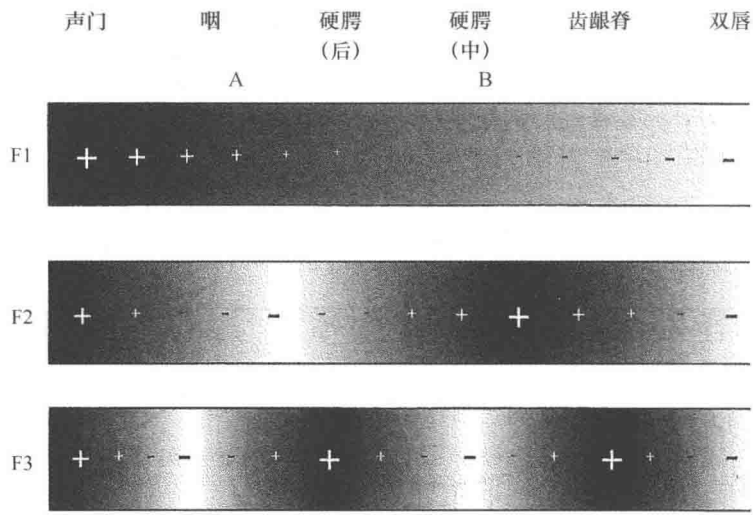


图 9.11　收紧位置对前三个共振峰频率 F1 – F3 的影响。随着收紧点沿声道的移动，每个共振峰相对于中性元音或者抬高（+），或者降低（-）。

一般来说，在元音四边形中，低元音的 F1 高，而高元音的 F1 低，前元音的 F2 高，而后元音的 F2 低。注意这里的高低前后是指元音四边形中的位置，反映了理想化的舌位，即从发音角度的描述。另一方面，共振峰频率的高低反映了从声学角度得到的测量结果——这便是低元音具有“高”共振峰频率的原因。

共振峰的频率值可作为对元音进行粗略分类的基础，适用于不同的说话人、语言和方言。Peterson 和 Barney（1952）对男性、女性和儿童进行了大样本研究，Hillenbrand 等（1995）也进行了重复实验，结果表明，不同的元音之间有非常一致的共振峰（分布）模式。这两个研究中男性、女性和儿童的共振峰频率均值请见附录 C.1，而 Hillenbrand 等（1995）研究中的男性数据请见图 9.12。该图示可以清楚地体现出上述关于元音类型与共振峰位置的关系原则。但是，显然不同发音人之间存在着个体差异，因此共振峰的值并不足以决定一个元音的音质。

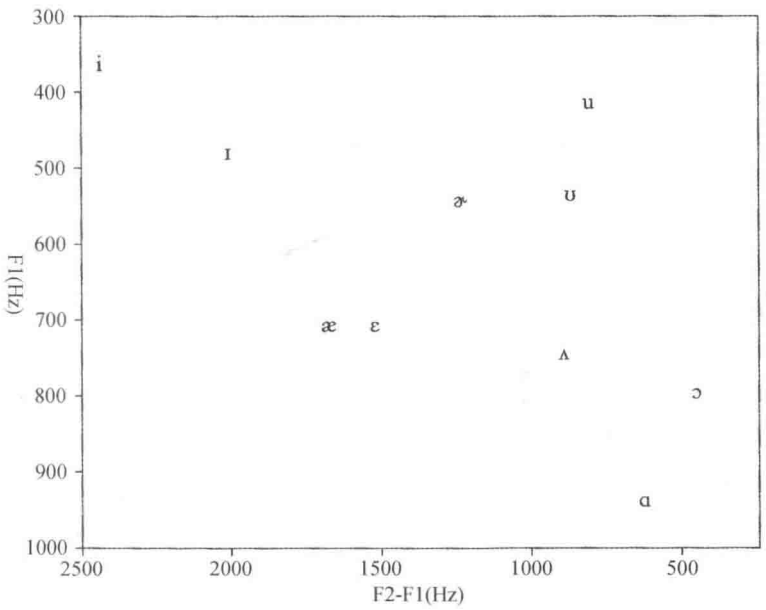


图 9.12 50 位男性发音人产出的美式英语元音的共振峰频率均值，坐标平面由 F1 和 F2－F1 构成。

以上介绍了发音器官的位置与共振峰频率之间的关系，现在我们回到对元音的讨论。在2.4节中，我们从高低、前后以及圆唇这三个发音维度对元音进行了讨论。尽管3.2节中介绍的元音四边形是讨论英语或其他语言元音的一种便利的方式，但我们仍要注意，元音四边形的维度和标注与高低和前后并没有直接的对应关系。例如，在高低维度，元音［i］和［u］都被描述为高元音，这似乎意味着在发这两个元音时，舌的最高点都接近口腔上缘，且程度相当。但是，Ladefoged（1967，2006）和其他学者都相继指出，实际情况并非如此。［u］的最高点与前高松元音［ɪ］更为相似。实际上，人们对元音的描写至少在一定程度上依赖于听觉印象。其实，早在1928年，美国语音学家Russell就指出了这个事实："语音学家根据声学事实进行思考，并通过生理学上的想象来表达他们的观点。"

现在，元音四边形被认为与元音的声学和听觉特性更为相关，而非发音特性。图9.12中所展示的美式英语元音被画在一个叫作声学空间的平面中。图中的维度自然表示的是声学特性，纵轴为第一共振峰F1，横轴为第二共振峰与第一共振峰的差值F2－F1。这种表示方法基于声学特性而非发音部位的特性，其元音分布特点与3.2节中的元音四边形非常相似。在10.1节中我们将对元音进行更详细的声学描述。

在下一章，我们将讨论元音的声学特征，并将声源－滤波理论扩展至辅音的产出，从而也对辅音的声学特征进行探讨。

练　习

1. 声门声源谱的基频为125 Hz，第一谐波的振幅为62.5 dB。那么比第四谐波高一个倍频程的分量的频率和振幅是多少？
2. 四分之一波长共鸣器的管长为25 cm，计算其第一、第二和第三共振频率，并计算管长相同的二分之一波长共鸣器的这三个共振频率。
3. 声音通过低通滤波器时会发生什么变化？通过高通滤波器时呢？这与言语产生有怎样的关系？
4. 基频和声道形状会影响共振峰频率吗？请作相应的解释。
5. F1和F2都很高的元音是什么？F1高、F2低的呢？F1低、F2高的呢？

F1 和 F2 都很低的呢?

6. 元音的声学描述和发音描述之间有什么不同?

注 释

1. 节点与反节点“中间”的相位为 45°，但节点与反节点的气压（或速度）差在 30°达到一半，因为 sin(30°) = 0.5。
2. 注意，滤波器的“通过带宽”是指不受阻碍通过滤波器的频率范围，而“带通”滤波器是指滤波器的形状。
3. 考虑到在正常谈话的响度约为 65 dB，飞机发动时的响度约为 120 dB，这个数值看上去非常高。但事实上，因为响度的 dB 值通常是在距声源 1 米处测量的，而强度的大小与距离的平方成反比（见附录 A.2），所以声源处的值要高很多。

10　语音的声学特征

本章将通过语谱图和相应的谱来探讨并说明各种语音的主要声学特征。我们将首先介绍元音，然后再介绍辅音。

10.1　元音

发元音时，声道的打开程度相对较大，气流通过基本不受阻碍，产生的声学信号比较响亮。此外，发元音时，声带通常保持振动。元音的主要声学特征是由语谱图上共振峰的位置（即前三个共振峰的频率）来决定的。我们已在9.5.1中讨论过，声道的形状决定共振峰频率的位置。发音器官位置的改变会改变声道的形状，从而改变共振峰频率的位置。同样的共振峰频率值可以通过不同的发音部位来产生，因此共振峰频率的位置是元音音质而非发音部位的关键决定因素。对一个人或者一组声道长度相同的发音人来说，每个元音都对应其特定的共振峰频率模式，例如图10.1分别给出了一位英语男性发音人产出的元音［i］和［ɑ］的语谱图及线性预测编码（LPC）谱。

语谱图的纵轴表示频率，横轴表示时间；颜色深浅表示强度的大小，颜色越深强度越大。图中所示两个元音的语谱图有几个相似之处：沿频率轴都有几条深色带状区域，这些深色带状区域对应的是共振峰频率，反映了声道的共振特性，对应于LPC频谱上的“峰”。

两个元音之间最主要的差异当然还是共振峰频率的位置。例如图10.1中，元音［i］的第一共振峰频率约为240 Hz，第二共振峰频率约为2450 Hz，第三共振峰频率约为3200 Hz。相比之下，元音［ɑ］的第一共振峰频率约为810 Hz，第二共振峰频率约为1250 Hz，第三共振峰频率约

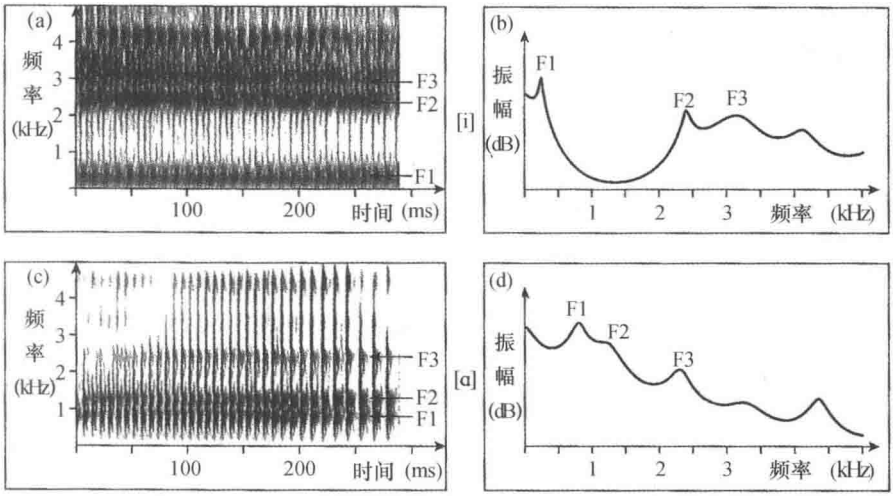


图 10.1 （a，c）英语元音［i］和［ɑ］的语谱图，（b，d）英语元音［i］和［ɑ］的线性预测（LPC）谱。发音人为男性。

为 2400 Hz。这样的差异反映了这两个元音发音时声道形状的差异。图 10.2 抽象地画出了美式英语中所有单元音前三个共振峰的频率，取值为 50 位美式英语男性发音人的平均值（Hillenbrand 等，1995）。

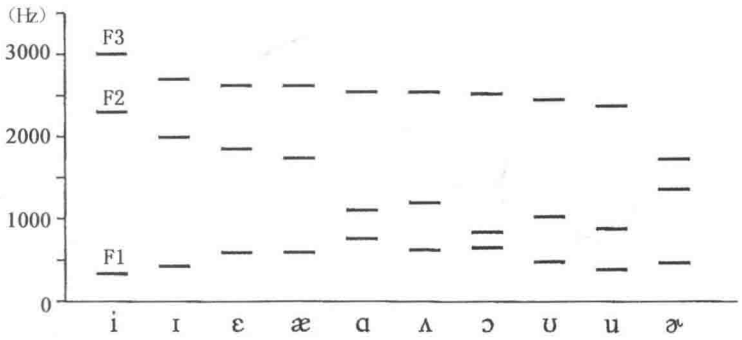


图 10.2 单元音前三个共振峰 F1 – F3 的平均值的抽象表示，数据取自 50 位美式英语男性发音人。

元音在水平轴上从前元音到后元音排列，此外，前元音又按照高度递减的顺序从高元音［i］到低元音［æ］排列，后元音按高度递增的顺

序由低元音［ɑ］到高元音［u］排列。经过这样的排列，描写元音的两个主要维度，高低和前后，与声学的相关性就非常清楚了。元音的高度与第一共振峰的频率成反比：元音越高（舌位越高），F1 越小。此规律对前后元音均适用。此外，若两个元音的相对高度相近，例如［i］和［u］，则它们的第一共振峰频率也相近。元音的前后反映在 F2 上，或者更准确地说是反映在 F2 与 F1 的差值上。从图 10.2 中可以清楚地看到，前元音的 F2 – F1 差值较大，而后元音则相对较小。

当元音的音质发生变化时，第三共振峰频率（F3）的变化没有 F1 和 F2 的大。不过元音［i］是个例外，F3 相当高。总体来说，英语元音的 F3 不太能够反映元音的音质，因此，在描述元音时大多只描写第一共振峰和第二共振峰频率。但是在很多语言中，F3 是元音音质的一个非常重要的特征，例如荷兰语、法语、德语和瑞典语都有前不圆唇元音和前圆唇元音。圆唇使声道变长，进而使所有的共振峰降低（见 9.5.1），并且使 F3 靠近 F2，以此来区分前不圆唇元音和前圆唇元音，如图 10.3 中展示的德语的例子。

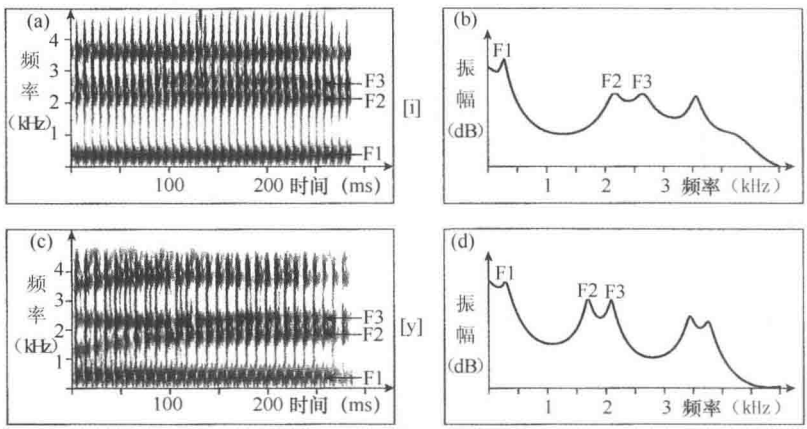


图 10.3　语谱图和 LPC 谱：（a，b）前高不圆唇元音［i］，（c，d）与［i］对应的圆唇元音［y］。发音人为一位男性德语母语者。

一般情况下，我们认为 F3 以上的共振峰频率不是辨认元音的重要线索，因为它们与元音音质的变化关系不大。事实上，对元音来说，F4 和

F5 等高次共振峰频率与说话人的个人特点更为相关，其反映的更多是说话人的个体特征，而不是元音本身的特征。因此，元音的 F1 – F3 反映了元音音质的**声学线索（acoustic cue）**。声学线索包括一个或者多个声学特征，从而可以提供用以确定某一特定音段的唯一信息，例如元音的高低、前后以及圆唇。

发鼻元音和鼻化元音时需要两个共鸣腔，即口腔和鼻腔。这两个腔之间复杂的相互作用，以及鼻腔严重的阻尼，使口元音与鼻（化）元音有所不同。与口元音相比，鼻元音通常表现为共振峰带宽较宽，振幅较小，并且其存在一个低频的鼻音共振峰以及一个或多个反共振峰（详见 10.2.4 中对鼻化的声学相关量的讨论）。图 10.4 中为同一发音人所发的口元音［ɑ］和鼻元音［ɑ̃］的谱图，注意在图 10.4b 中在 200 Hz 附近，第二谐波的能量有所增加，在 450 Hz 附近，谱图出现凹陷。

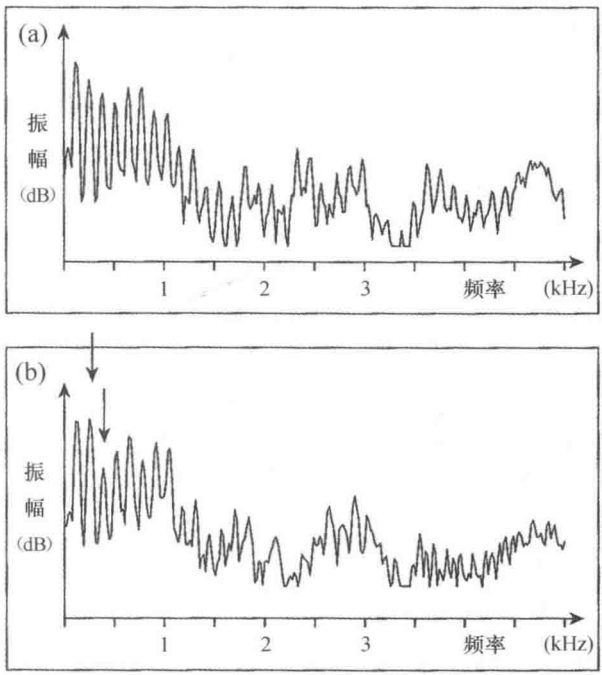


图 10.4 FFT 谱：（a）口元音［ɑ］和（b）鼻元音［ɑ̃］，发音人与图 10.3 相同。（b）中箭头指示 200 Hz 附近能量增大，450 Hz 附近能量减小。

时长也可以作为辨认元音的一个线索。许多语言都存在以时长作为主要差异的元音对立。例如，德语中［ɪ］和［iː］的区别在于时长以及谱特征的不同。爱沙尼亚语区分短元音、长元音和特长元音。英语的紧元音比松元音有更长的**稳定状态（steady-state）**。稳定状态段是共振峰频率相对稳定不受前后音段影响的那一段，通常位于元音中部。不过英语的［æ］是一个特例，它虽是松元音（不出现在开音节中），但时长较长。人们发现，当紧元音夹在两个辅音之间时，共振峰从辅音过渡到元音（即“入渡”，transition into）以及从元音过渡到辅音（即“出渡”，transition out of）的部分比较对称；而对松元音而言，其入渡较长，而出渡则要更长。通常，英语的松紧元音主要是音质（共振峰模式）的不同，而不是时长。［i］缩短以后听起来还是［i］，［ɪ］加长以后听起来还是［ɪ］。这就是我们在3.2节说过的，为什么英语的紧元音和松元音用两个不同的符号［i］和［ɪ］来表示，比用符号［iː］和［i］或［ɪː］和［ɪ］来表示更加妥当。

二合元音（见2.4节和3.2节）的特点是发音时共振峰模式随着发音上形状的变化而改变。这种改变主要影响前两个共振峰，在语谱图上的表现非常明显。图10.5给出了英语的二合元音［aɪ］、［ɔɪ］以及[aʊ]的语谱图。声学上，二合元音通常用起始频率和结束频率来表示。值得注意的是，二合元音结束点的共振峰频率值通常与相应的单元音（［ɪ，ʊ］）一致（见图10.2），并且［ɔɪ］起始点的共振峰频率值与单元音[ɔ]一致。二合元音［aɪ］和［aʊ］的起始元音比单元音［ɑ］还要靠中央一点，所以相应的F2起始值要稍微高一些。

10.2　辅音

我们在发所有的辅音时都需要对气流大小进行调控，发近音时气流量中等，发爆发音时气流量显著。我们按发音方法来介绍辅音的声学特征，首先要介绍的是近音，因为它与元音最为相似，最后介绍爆发音。

10.2.1　（央）近音

发近音时，两个发音器官相互趋近，但不阻碍气流，因此近音的声

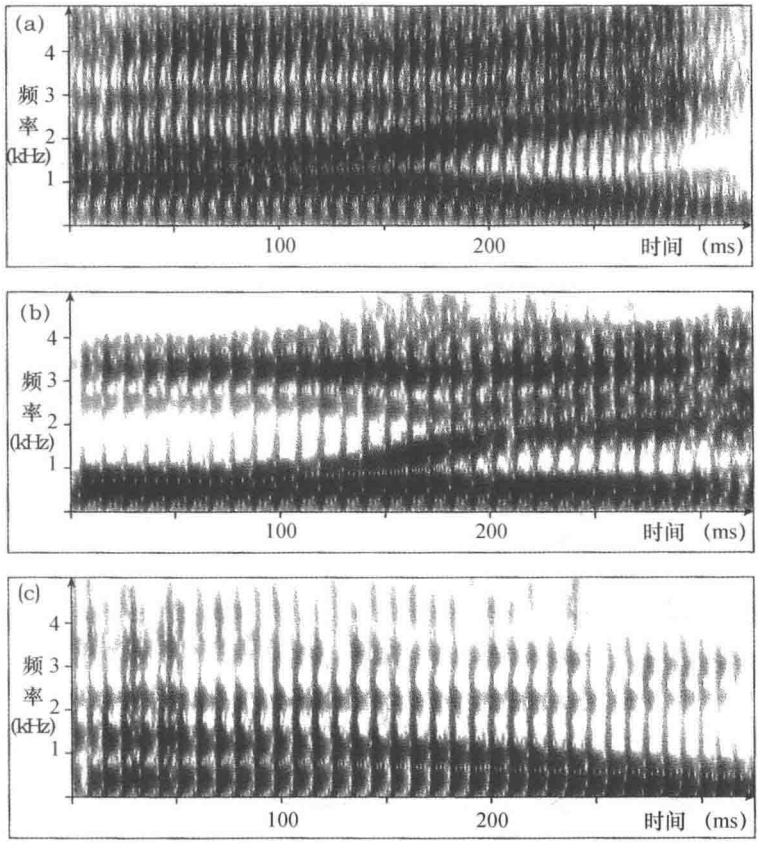


图 10.5 英语二合元音语谱图（a）［aɪ］（b）［ɔɪ］以及（c）［aʊ］，发音人为男性。

学特性与同部位的元音十分相似。近音的共振峰模式比较清晰，但是由于近音的收紧度比元音稍大，其共振峰模式不如元音明显，而且稳定段较短，声学能量较低。例如硬腭近音［j］，发音时舌叶接近硬腭，与发元音［i］时非常相似，但发［j］时舌叶比［i］更接近硬腭，气流阻碍比［i］严重，因此虽然［j］的频率谱与［i］相似，但谱的颜色没有［i］那么深。图 10.6a 是［ɑjɑ］的语谱图。硬腭近音的特征表现为 F1 低，F2 和 F3 高，与［i］相似。近音的入渡及出渡十分明显，这体现在频率范围和时长的变化上。当然频域过渡的范围是由邻近的元音决定的，如图 10.6b 所示，若［j］夹在两个共振峰相似的元音中间，如［iji］，过渡段就很短。

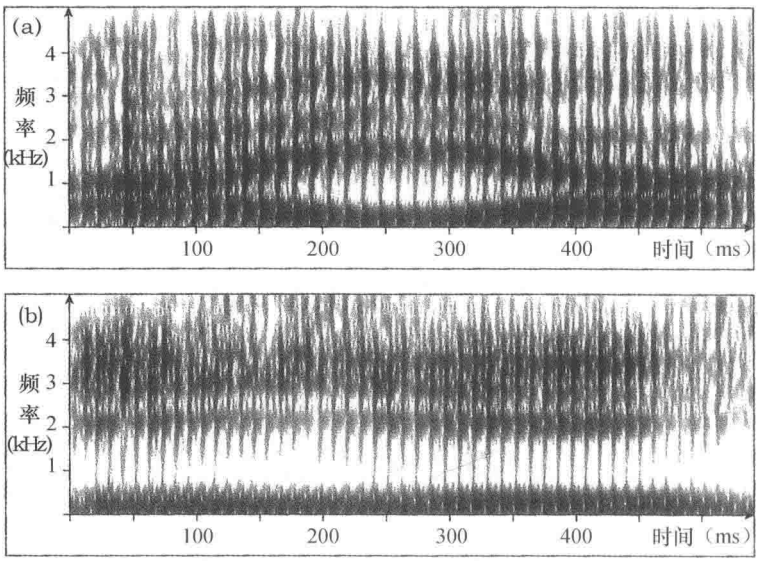


图 10.6　（a）［ɑjɑ］的语谱图和（b）［iji］的语谱图，发音人为英语母语男性。注意辅音段的 F3 几乎与 F2 “重合”。

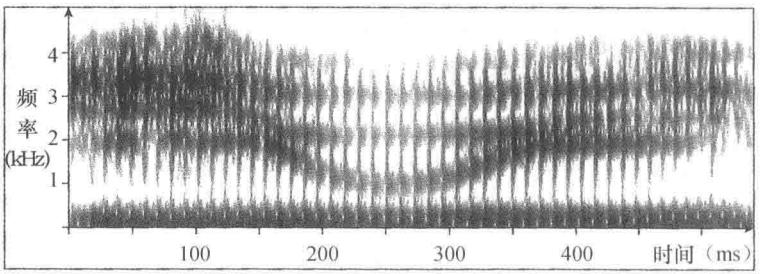


图 10.7　［iwi］的语谱图，发音人为英语母语男性。

在发唇－软腭近音［w］的时候，舌后部接近软腭且双唇变圆，与发元音［u］非常相似。图 10.7 为［iwi］的语谱图，对应的唇－软腭近音 F1 和 F2 较低，并且靠得比较近，F3 稳定在 2300 Hz 左右，与元音［u］相似。同样，共振峰过渡也是由相邻的元音决定的。

卷舌近音［ɻ］发音时同时有三个收紧位置，圆唇、舌尖在齿龈脊附近形成的收紧以及舌根后缩引起的咽腔变窄，这些都会使 F3 降低。所以［ɻ］的特点是 F3 很低，接近 F2。正如图 10.8 中所示，近音的 F3 大约为

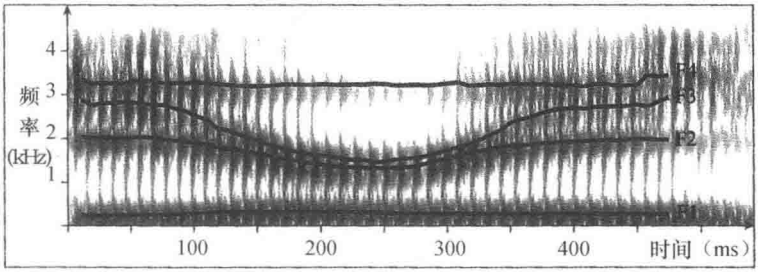


图 10.8 ［iɻi］的语谱图。发音人是英语母语女性。图中经过平滑处理的曲线表示共振峰轨迹。

1800 Hz。［ɻ］的 F3 是所有英语语音中最低的，即使是女性都能低于 2000 Hz。

简而言之，近音的特点就是比元音短一些弱一些，并且共振峰过渡比元音长一些。

10.2.2 擦音

发擦音时口腔内有一处很窄的收紧。擦音的声源通常是气流通过这个收紧点（擦音不同位置不同）时产生的湍流（浊擦音还具有声门处声源）。我们可以从四个方面来定义擦音的性质：摩擦噪声的谱特点、噪声的振幅、噪声的时长以及邻近元音向擦音、擦音向邻近元音过渡的谱特性。

擦音谱的整体形状很大程度上取决于收紧点前部的口腔的大小和形状。我们在第 9 章介绍了言语产生的声源 - 滤波理论，并通过元音的声学特征对该理论做了进一步的说明。我们也可以用声源 - 滤波理论来解释辅音的声学特征。由于发阻音时收紧程度很大，阻音的共振频率可以按前腔和后腔单独进行计算，而不用考虑两者之间复杂的交互作用。图 10.9 给出了一个最简单的模型来表示辅音的发音构型。该模型应包含三个声管，一个模拟前腔，一个模拟收紧，最后一个模拟后腔。

三声管模型由四分之一波长共鸣器和二分之一波长共鸣器组成（参见 9.1.1）。后腔近似于两端闭合的二分之一波长共鸣器：一端是

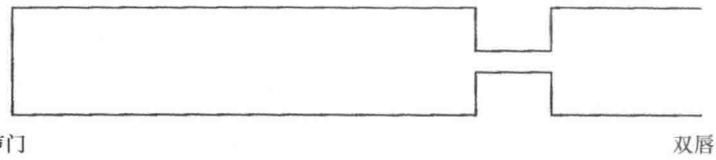


图 10.9　用三个声管表示阻音发音构型的声道模型，从左至右分别对应后腔、收紧和前腔。

声门，另一端是口腔内的收紧。连接后腔和前腔的短管代表收紧，它也是一个二分之一波长共鸣器：短管是两端开放的，与较小的收紧面积相比，两端的开口很大。前腔一端闭合，为收紧，另一端开放，为双唇，是一个四分之一波长共鸣器。

下面我们用这个模型来计算共振峰频率。这里不考虑第一共振峰，我们可以通过**赫姆霍兹共鸣器（Helmholtz resonator）**得到第一共振峰的值。赫姆霍兹共鸣器是一种共鸣系统，该系统具有一个较宽的后腔，并且一端完全闭合（即声门处），另一端有很窄的收紧。赫姆霍兹共鸣器的共振频率由后腔和收紧处的相对空气容积决定，一般很低（低于 1 kHz）。

以龈擦音［s］为例，根据 X 光和核磁共振（MRI）的声道研究，假设声道长 16 cm，后腔长 11 cm，收紧部分长 3 cm，前腔长 2 cm。由于后腔是一个二分之一波长共鸣器，据 9.1.1 中提出的公式：

$$f_k = k \times \frac{c}{2l}$$

根据 $l = 11\ \text{cm} = 0.11\ \text{m}$，声速 $c = 340\ \text{m/s}$，可得后腔的共振频率：

$$f_k = k \times \frac{340\ \text{m/s}}{2 \times 0.11\ \text{m}} = k \times \frac{340}{0.22\ \text{s}} = k \times 1545\ \text{Hz}$$

当 $k = 1$，2，3，4 时，所得共振频率分别是 1545 Hz，3090 Hz，4635 Hz 和 6180 Hz。

由于前腔为四分之一波长共鸣器，使用 9.1.1 节的公式：

$$f_k = (2k - 1) \times \frac{c}{4l}$$

根据 $l = 2\ \text{cm} = 0.02\ \text{m}$，可得前腔的共振频率：

$$f_k = (2k-1) \times \frac{340\ \text{m/s}}{4 \times 0.02\ \text{m}} = (2k-1) \times \frac{340}{0.08\ \text{s}} = (2k-1) \times 4250\ \text{Hz}$$

当 $k=1$，2 时，所得共振频率分别是 4250 Hz 和 12750 Hz。

在为共振峰频率编号时，上边计算得到的最低的共振频率应标为 F2（F1 由赫姆霍兹共鸣器产生），接下来就是 F3，等等。因此，对于［s］，这个模型预测的 F2 为 1545 Hz，F3 为 3090 Hz，F4 为 4250 Hz，F5 为 4635 Hz，等等。但是一定要弄清楚所得的共振峰频率是与前腔有关还是与后腔有关，因为阻音的谱主要以前腔共鸣为主。在收紧处产生的摩擦噪声或者爆破摩擦主要激励收紧点以前的声腔。我们上面计算所得的 F2 和 F3 都是后腔的共振，而 F4 是前腔共振的结果，所以 F2 和 F3 的影响并没有 F4 大。因此，我们预测［s］的谱能量集中区在 4250 Hz（及以上）。

通常情况下，前腔越长，谱能量的强弱对比越强，因为前腔好比“共振室”（resonating “chamber”），容积越大影响越大。此外，发擦音［s，z，ʃ，ʒ］时气流会与牙齿相碰，所以这几个音的谱的形状更为显著。因此，龈擦音和龈后擦音的谱非常清晰、明显，而唇齿音和齿（间）擦音的谱则相对平坦。特别值得注意的是，［ʃ，ʒ］在 2500 ~ 3500 Hz 处存在中频的谱峰。发齿龈音［s，z］时，前腔比发［ʃ，ʒ］时要短，因此其谱峰处于高频区，在 4000 Hz 至 7000 Hz 之间。此外，发以上四个擦音时，气流都会与牙齿相碰，因此会产生很强的高频湍流。相对而言，［f，v］和［θ，ð］的谱比较平坦，但最近有一些研究认为唇齿音可能有比齿间音还要高一些的谱峰频率（接近 8000 Hz）。

对于收紧点较为靠后（前腔更长）的辅音，如软腭擦音、小舌擦音以及喉擦音，其能量主要集中在低频区，即第一和第二共振峰所在区域。软腭擦音的能量集中区对应邻接元音的 F2，高频区能量很少。小舌擦音和喉擦音的主要区别在于第一共振峰的频率不同，这两个音的收紧点都较为靠后，但喉擦音的 F1 频率比小舌擦音要高。

浊擦音有两个声源：收紧位置产生的湍流噪声以及声带振动产生的低频能量。浊擦音的语谱图与清擦音相似，区别在于浊擦音有因声带振动而产生的低频能量，并且由于声带的振动耗费了部分气流能量，浊擦音的高频能量较小。

从摩擦噪声的振幅来看，[s，ʃ] 的振幅要明显大于 [f，θ] 的振幅（要大 10～15 dB），因此人们有时将 [s，ʃ] 以及它们所对应的浊音 [z，ʒ] 称为强擦音或咝音（sibilant），而称 [f，θ，v，ð] 为弱擦音或非咝音（non-sibilant）。两组擦音中涉及的四种不同发音部位均可以用**相对振幅（relative amplitude）**加以区别。相对振幅是指某个频率范围内（齿龈音和龈后音对应 F3，唇齿音和齿间音对应 F5），擦音与其后接元音的振幅差值。这种计算方法表明，龈后音的能量集中区相当于其后接元音 F3 所在位置，而齿龈音的能量集中区在更高的位置。此外，对非咝音而言，唇齿音的 F5 区域的摩擦振幅要高于齿间音。

噪声时长也可以帮助我们区分咝擦音和非咝擦音，[s，ʃ，ʒ] 比 [f，θ] 长。噪声时长也可用于区分浊擦音和清擦音，清擦音的噪声比浊擦音长。

最后还要说明一点，有研究指出擦音和元音之间共振峰的过渡，特别是第二共振峰的过渡，可以作为判断发音部位的线索。最近的一项研究发现，发音部位在口腔中越靠后，元音 F2 的起始位置就越高。

总之，擦音具有相对较长的噪声段和十分稳定的声学特征。

10.2.3　爆发音

发爆发音时，口腔中的收紧点完全闭合，气流受到短暂的阻塞。如我们在 2.3.2 中所讨论的，气压会在阻塞的后方积聚。发音器官之间阻塞的解除以及发音器官向下一个发音的后续运动构成了爆发音发音部位的两个主要特征：爆破频率以及共振峰过渡。图 10.10 是处在两个元音之间的爆发音的语谱图，图中显示了爆发音的许多声学特征。起始元音与爆发音爆破之间的闭合段是爆发音的一个主要特征，这一段是由声道的完全闭合而产生的。语谱图上的这段“空白”对应于辅音的闭合段。清爆发音的闭合段没有能量，全浊爆发音的闭合段只在频率非常低的频域有能量（称为**嗓音横杠**或**浊音杠，voice bar**）。

爆破频率（burst frequency），即除阻爆破的频率谱，随收紧前部的声道长度的变化而变化，这与擦音摩擦噪声的主频率类似。将前面预测擦音共振频率的公式稍作调整，我们便可以预测爆破频率。双唇音的收紧位于双唇处，在收紧的前部不存在任何有效的腔体。由于唇部收紧通常

会降低共振峰的频率（参见 9.1.2），双唇爆破的能量一般集中在低频区（500 ~ 1500 Hz）。对齿龈爆发音而言，收紧位于齿龈处，前腔较短，爆破频率相对较高，为 2500 ~ 4000 Hz。软腭音的爆破频率比齿龈音低，因为收紧位于软腭，收紧前部的声道明显长于齿龈音。软腭音的爆破频率介于双唇音和齿龈音之间，为 1500 ~ 2500 Hz。图 10.10 中，除阻爆破清晰可见，爆破频率也基本符合上面的描述。我们对软腭爆发音和小舌爆发音之间的声学差异了解很少，它们的发音都靠后，小舌爆发音的爆破频率更低。

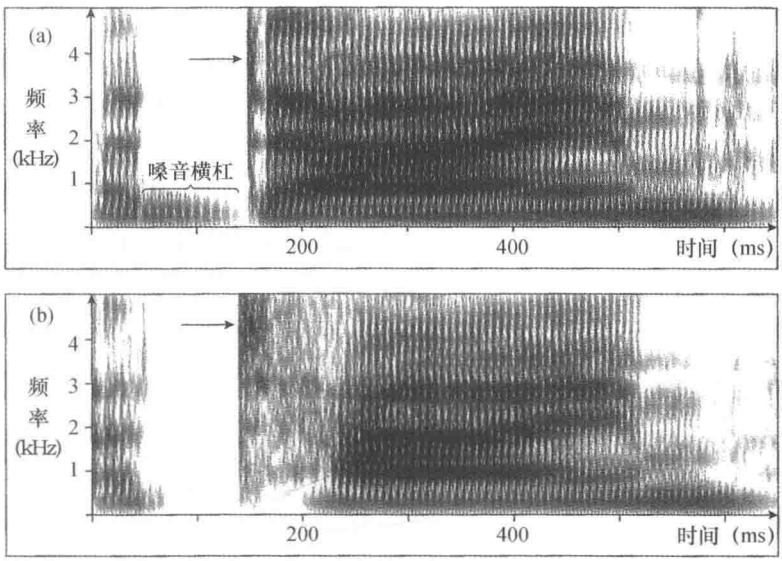


图 10.10 不同语段的语谱图（a）“a dime”，两个元音间为浊爆发音（b）“a time”两个元音间为清爆发音。发音人是母语为美式英语的女性，箭头指向爆破能量最集中的区域。

第二共振峰和第三共振峰从起始元音到闭合段的过渡，以及从除阻爆破到其后接元音的过渡同样可以为发音部位的特征提供线索。共振峰的过渡是相邻元音与爆发音之间发音运动的声学表现，一般在 50 ms 内就可完成。因为收紧期间 F1 的频率约为 200 Hz，所以所有爆发音的 F1 都很低。由于双唇音除阻时能量集中在低频，元音过渡到爆发音时，F2 和

F3 会下降至闭合段，之后再由低频爆破段向后接元音的共振峰频率抬升。同样，由于齿龈音除阻时能量通常集中在高频，其前接元音的 F2 和 F3 会先向闭合段抬升，然后再从高频的爆破下降至其后接的元音。最后，对于软腭爆发音而言，其爆破频率通常在中频段（2000 ~ 2500 Hz），在该频率范围内，其 F2 和 F3 非常接近。元音向软腭爆发音过渡时，F2 向闭合段抬高，F3 向闭合段下降；接着，F2 和 F3 分别从爆破段下降和上升并过渡到其后接元音。发软腭爆发音时 F2 和 F3 的这种互相趋近的变化叫作**软腭夹（velar pinch）**。共振峰的过渡为发音部位提供了线索，但过渡的趋向则由相邻元音决定。例如，若拿［ɑtɑ］和［iti］相比，前者第二共振峰的过渡要陡得多（［ɑ］的 F2 较低，而［i］的 F2 较高，与［t］位于高频的能量集中区更为接近）。

最后，振幅和时长反映的是辅音的发声特点而不是发音部位的特点。与浊爆发音相比，英语的清爆发音通常具有较长的闭合段，较强的爆破，以及较长的（正值）VOT。

总结起来，爆发音的主要特点是：具有闭合段、除阻爆破以及快速变化的共振峰过渡。

10.2.4　鼻音

发鼻塞音时，口腔内有一个完全闭合的收紧，气流在口腔中暂时被完全阻塞。不过，软腭下垂使得气流可以通过鼻腔流出，如此产生的声音称为**鼻哞音（nasal murmur）**。鼻音需要两个共鸣腔：口腔和鼻腔。它们之间的组合方式比较复杂。作为辅音的一类，鼻音有如下几个特性。（1）第一共振峰很低，这是因为所有分支腔体的总长度（包括咽腔、鼻腔和口腔）长于口音。这个很低的 F1 有时被称作**鼻共振峰（nasal formant）**。（2）所有的共振峰都比较弱（振幅低），因为通向鼻腔的开口较窄，阻碍气流的进入。（3）共振峰带宽更宽，这是由于口腔壁和鼻腔壁吸收了部分的声音能量。（4）最后，存在**反共振峰（anti-formant）**，反共振峰是由口腔和鼻腔的耦合作用而形成的（见下文）。

对于给定的发音部位，发口塞音和鼻塞音时，口腔中的状态非常相似。例如，［d］和［n］都需要在齿龈脊形成完全的闭塞，因此它们有一些相同的声学特性。不过，两者发音的唯一不同在于发［n］的时候软腭

下降，这构成了两者间主要的声学差异。如图 10.11 所示，由于加入了敞开的鼻腔，我们需要用双管模型来模拟声道。一个声管在声门处闭合，在鼻腔处打开，与大气相通；另一个声管可视为一个侧声管，一端在小舌处打开，另一端在口腔内的收紧处闭合。

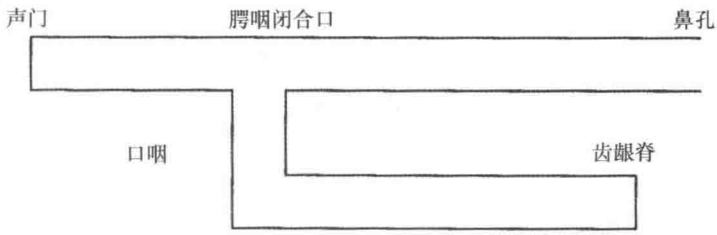


图 10.11 齿龈鼻塞辅音的声道形状示意图

在鼻音的谱上，我们无法直接观察到口腔的共振，因为口腔与空气不相通。但是，口腔和鼻腔之间的相互耦合或者相互作用产生了反共振峰。也就是说，鼻音中与口腔共振频率相近的共振频率都被吸收或者消除了，进而被从整体的谱中“减去”了，这是因为部分由喉部产生的脉冲气流的声学能量在口腔中引起了共振。这些反共振峰在谱上表现为明显的波谷，并且其频率因发音部位的不同而改变。口腔后部越短（即发音部位越靠后），共振频率越高，输出的谱上的反共振峰频率也越高。一般来说，双唇音［m］的反共振峰频率在低频区（750 ~ 1250 Hz），齿龈音［n］位于中频区（1450 ~ 2200 Hz），而软腭音［ŋ］位于高频区（高于 3000 Hz）。图 10.12 展示了这些特征。

在结束对鼻音发音部位的线索的讨论之前，我们必须说明一点，反共振峰是很难测量的，我们甚至很难准确判断反共振峰的位置。而个体鼻腔大小和形状的差异会使原本因口腔大小的不同而造成的个体差异更加显著。而相邻两个音段间的共振峰过渡也有可能包含发音部位的信息，鼻辅音相应过渡段的特点以及相应的解释，与同部位的爆发音十分相似（见 10.2.3）。

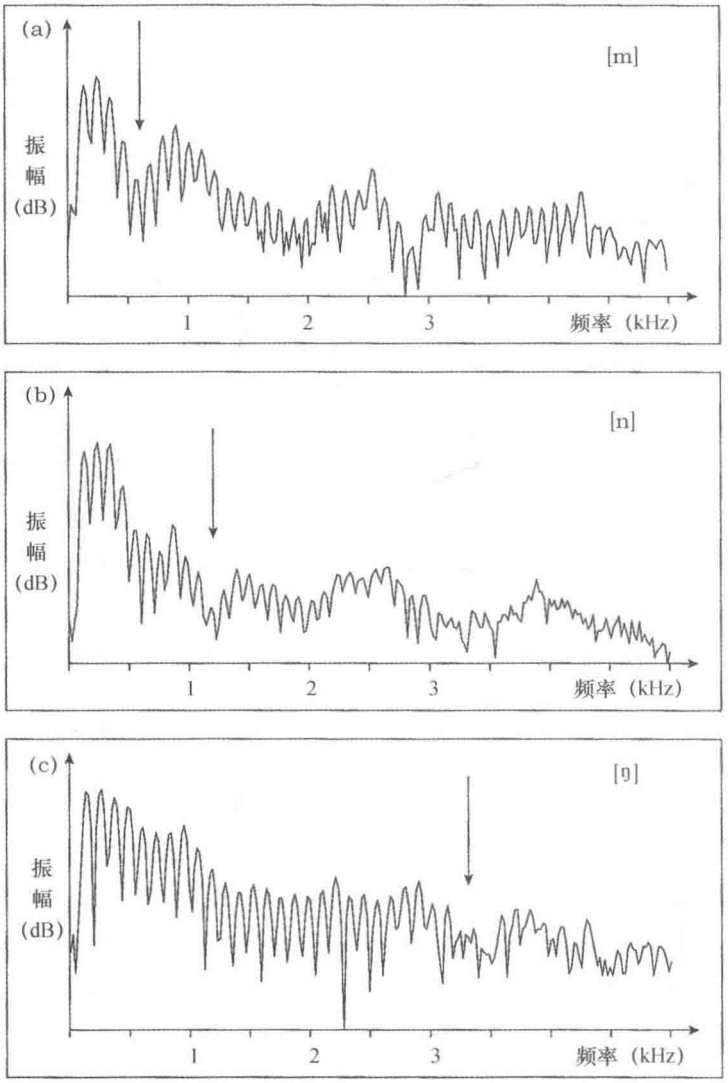


图 10.12　语段中鼻辅音中部的 FFT 谱 (a) [ama], (b) [ana], (c) [aŋa]。箭头指向反共振峰的位置，在谱中表现为凹陷点。

10.2.5　边近音

齿龈边近音 [l] 的声学特征与鼻音非常相似。发 [l] 的时候，舌的一部分与口腔上部接触。具体而言，舌尖与齿龈脊相碰，同时舌的一侧或者两侧下降，因此舌侧仅仅是趋近口腔顶部。[l] 具有稳定段，它

的谱具有共振峰模式。男性发的［l］的共振峰频率（F1 到 F3）的平均值分别约为 340 Hz、1200 Hz 和 2800 Hz，见图 10.13 所示。发齿龈边近音［l］时，空气从收紧部位的任意一侧通过，而舌体上方留存有一小部分气体。这种情况同鼻音类似，我们依然可以将声道看作一个主声管和一个侧声管的组合。发龈边近音［l］时，主要声管自声门位置起，延伸至口腔开口处，而那一小部分气体可以用一个较短的侧声管来模拟。

这两个声管之间的耦合作用同样会产生反共振峰。侧声管产生的共振峰会抵消主声管产生的共振峰。由于侧声管较短，反共振峰的频率很高，成年男性为 2000 ~ 2300 Hz。

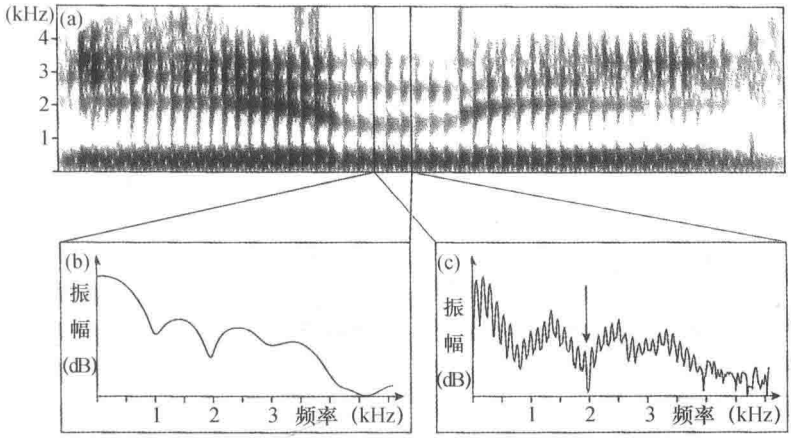


图 10.13 （a）语段［ili］的语谱图和（b）LPC 谱，（c）［l］中部的 FFT 谱。发音人为英语母语男性。FFT 谱中的箭头表示 2000 Hz 附近的反共振峰。

10.2.6 塞擦音

塞擦音兼具爆发音和擦音的声学特点。塞擦音与发音部位相同的擦音和爆发音的声学特征相似，只是摩擦段的时长有所不同，塞擦音比擦音要短一些。图 10.14 给出了［apfa］、［atsa］、［atʃa］以及［akxa］的语谱图。

作为辅音的一个分类，塞擦音与擦音可以用**上升时间（rise time）**来加以区分。上升时间指的是，从辅音起点到摩擦噪声达到最大振幅所需

的时长。擦音的振幅是逐渐升高的，因此上升时间较长。相较而言，塞擦音的上升时间较短，爆破之后，摩擦噪声很快达到最大振幅。

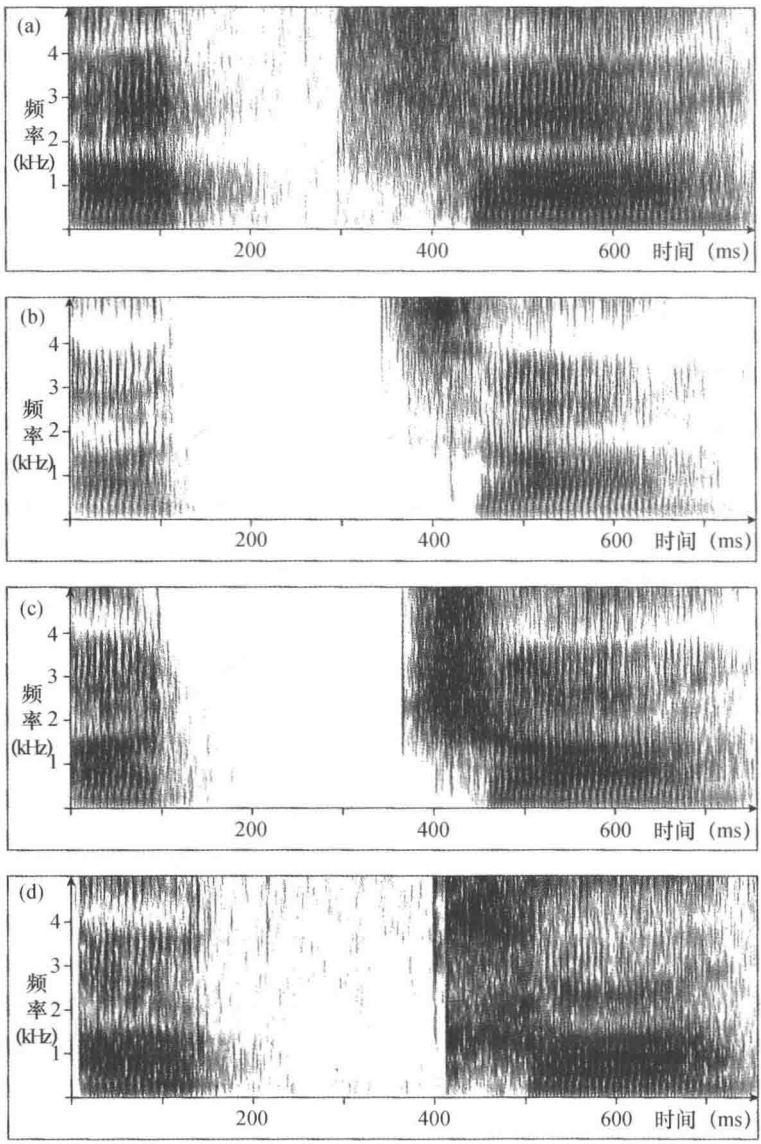


图 10.14　语谱图：(a) [apfa]，(b) [atsa]，(c) [atʃa]，(d) [akxa]，发音人为一位讲瑞士德语的人。

10.3 小结

通过对语音声学特征的讨论，我们知道，发音方法以及嗓音状态是比较好辨认的特征，而确定发音部位则比较困难。发音方法方面，元音和近音的特点比较突出，没有噪声（爆发或摩擦）段。元音与近音的区别在于，元音的稳定段较长且稳定性较强。爆发音和塞擦音都具有一个没有（高频）能量的闭合段。口塞音在除阻后的能量增长比塞擦音更为迅速，而塞擦音的上升时间要比擦音短。最后，鼻音的第一共振峰的频率很低，且具有反共振峰。

带音具有低频能量，因为发音时声带会振动。对元音、近音、擦音和鼻音而言，这一浊音（带音）特点通常贯穿整个发音。而爆发音的浊音（带音）特点只在闭合段最为明显。

对于口塞音和擦音等收紧比较严重的音，发音部位的特征主要取决于声道收紧部位的前腔长度（即声道的前部）。因此，发音部位相同的擦音和塞音具有某些相同的声学特征。双唇音的能量一般集中在低频区，齿龈音的能量集中在高频区，而软腭音的能量集中在中频区。鼻音的情况稍有不同，代表鼻音发音部位特征的反共振峰的位置是由小舌和口腔收紧之间的声道长度决定的，因此软腭鼻音的反共振峰的频率最高。

有一点要清楚，虽然我们讨论的声学特征代表了不同的发音方法、发音部位和嗓音特征，但并不是每一个语音都具备所有这些特征。对于某一个特定的语音而言，出现哪些特征以及程度如何，是由诸多因素决定的，包括个体差异、语音环境和语速等。这一现象引起了研究者们的许多争论：对于特定语音而言，有哪些声学特征是不变的，哪些特征是随语境改变的？本章的最后一节将援引一些以往的研究数据，来讨论连续的物理信号与离散的语音学范畴之间的映射关系。

10.4 可变性与不变量

正如本章前面的内容所讨论的，对于辅音而言，特别是爆发音，最

为困难的是与发音部位有关的线索。与爆发音发音部位有关的线索包括除阻爆破的频率和共振峰过渡的特征，但这两个线索的作用一直是争论的焦点。我们来考虑一下图 10.15。

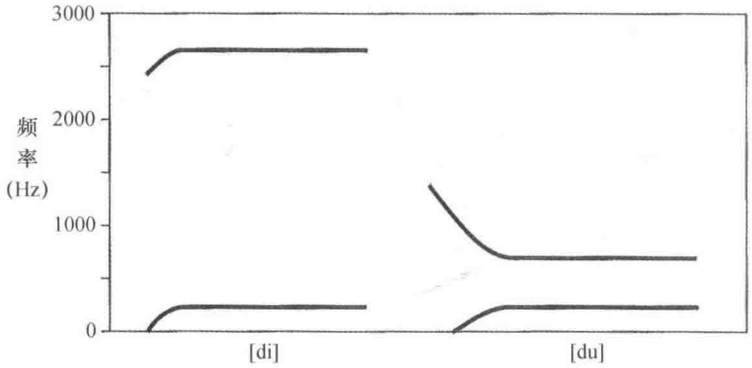


图 10.15　［di］和［du］抽象化的 F1 和 F2 模式（Delattre 等，1955）

这张经典的图（引自 Delattre 等，1955）为两个音（合成语音）的语谱示意图。英语母语者将这两个音分别感知为英语［di］和［du］（注意：这些例子没有除阻爆破，是对自然语音的简化表示）。到底是语谱图中哪一部分使得听音人把这两个音节的起始音都感知为齿龈爆发音的呢？我们先来看［di］，语谱图分为如下几个部分：呈上升趋势的 F1 过渡表明起始音是一个爆发音；F1 和 F2 的稳定段表明其接后元音是［i］。现在仅剩的可以判定辅音发音部位的线索是上升的 F2 过渡。然而再看［du］，我们马上就知道，上升的 F2 过渡不可能是判定齿龈音发音部位的可靠线索。音节［du］上升的 F1 过渡对应爆发音，共振峰稳定段对应［u］，但是现在表示发音部位的部分，即 F2 的过渡，却在另一个不同频率范围表现为明显下降的趋势。

以上采用的观察角度可以得出这样一个观点：大部分的声学线索都是高度语境相关的。也就是说，没有（或极少的）声学线索是稳定的、一致的或不变的。由此以来，决定发音部位的 F2 过渡的线索随元音环境的不同而产生剧烈的变化。也就是说，［d］的感知依赖于后接元音，同一个［d］因为后接元音不同，需要用两个不同的声学线索来表示。据

此，Liberman 和他的同事认为，语音信号“缺少不变量”，也就是说声学特征与语音感知之间不存在一对一的简单映射关系，而不同的声学线索却可能导致相同的语音感知结果（比如［di-du］的例子）。

这里要强调的是，此处所说的缺少不变量仅限于声学领域，该讨论与语音感知过程中是否存在不变量无关。当听音人将不同的声学线索映射到固定的音系范畴时，不变量显然发挥了十分重要的作用。问题的关键是，这种不变量作用存在于哪个层次上。上述例子展示了语音线索的语境相关性，这一特点指出，不变量并不存在于的声学信号本身。

10.4.1 声学不变量理论

与 Liberman 的结论不同，其他的一些学者（最著名的是 Stevens 和 Blumstein，1981）认为不变的声学特性是确实存在的。导致这种分歧的原因之一可能与双方分析语音信号的方法有关。

Liberman 研究声学不变量的问题时，分析语音信号的最佳方法是语谱图。正如 8.3.4 中讨论过的，语谱图描述了频率随时间的变化。当考虑爆发音 - 元音音节时，其语谱图上有三个突出的特性：除阻爆破段，共振峰过渡段和共振峰稳定段。在研究辅音感知时，人们自然会将这些属性看作独立的特征来处理。在 20 世纪 50 年代，相关研究确实是这样做的。比如说，设计一组不同频率位置的除阻爆破特征，以此来研究它如何影响对发音部位的判定。除阻后衔接不同组合模式的第一、二共振峰的稳定段，以代表不同元音。实验结果表明爆破频率和所感知的发音部位之间并没有一致的对应关系。这些发现进一步肯定了如下观点：语音感知所依赖的线索很大程度上是语境相关的。

在研究爆发音 - 元音音节时，Stevens 和 Blumstein 使用了一种新的谱分析技术，即 LPC 分析技术，这一技术直到 20 世纪 60 年代才被广泛使用。在 8.3.5 中我们介绍过，在 LPC 分析中，谱信息是在时间窗上的积分。因此在 LPC 谱中，除阻爆破、共振峰过渡以及稳定状态都不是独立的事件，整个谱所展示的是所有这些信息在时间上累积而得到的一份“快照”。分析爆发音的谱所用的时间窗一般为 10 ~ 25 ms。一般认为这种表示方法更接近听觉系统。

对 CV 序列的前 25 ms 进行的 LPC 谱分析发现，谱的整体形状包含了

发音部位的线索，这些线索并不因语音环境和声道大小的改变而变化。图 10.16 给出了各个发音部位所对应的整体形状模式。双唇辅音的谱可能是下降的，表示能量集中在低频区；也可能较为平坦，表明没有能量特别集中的区域。齿龈音的谱呈上升的模式，表示能量集中在高频区。而软腭音的谱显示能量一般集中在中频率区，具有“紧凑”的谱形状。

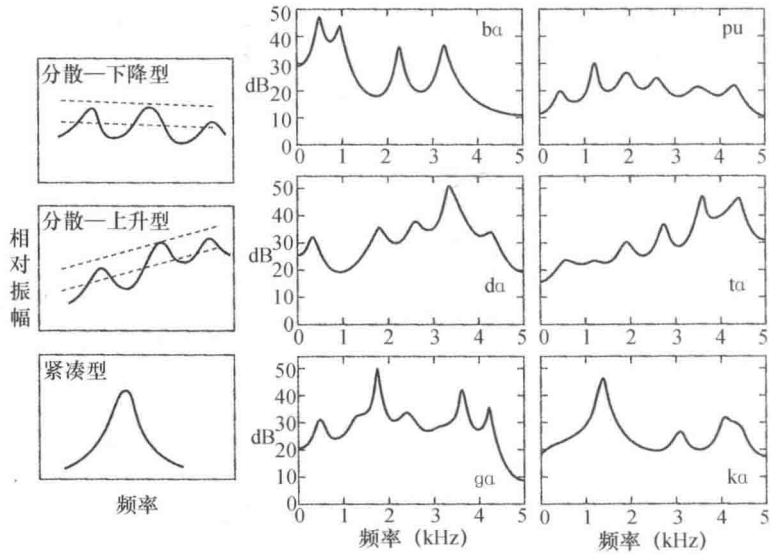


图 10.16　中间和右边两栏为双唇、齿龈和软腭爆发音的 LPC 谱，根据 CV 音节的前 25 ms 计算得出。左边的一栏给出了用“分散 - 下降型”（diffuse-falling）、“分散 - 上升型”（diffuse-rising）以及“紧凑型”（compact）模式分别表示的双唇、齿龈和软腭爆发音的谱形状。

经许可，我们在此使用了 Sheila E. Blumstein 和 Kenneth N. Stevens 发表的图示。见 *The Journal of the Acoustical Society of America*, 66, 1001（1979）。版权属于美国声学学会，1979。

之后，Blumstein 和 Stevens 重新找了一组发音人，录制了 CV 和 VC 音节，分析了辅音的 LPC 谱，并根据谱模式对辅音的发音部位进行分类。结果显示，对于全部的 1800 个样本（由 4 位男性和 2 位女性提供），发音部位的判断正确率可以达到 85%（Blumstein 和 Stevens，1979）。

此后对不变量的研究主要集中在对更具动态的线索的研究上，即比

较辅音起始和后接元音起始之间高频能量与低频能量在分布上的变化。这种方法是对前面方法的改进，并且依然能保持基本的认识，比如双唇音的能量集中在低频区，齿龈音的能量集中在高频区。采用这种动态的评价方法，英语、法语和马来西亚语中的唇、齿和齿龈爆发音的正确区分率达到了91%（Lahiri 等，1984；也可参见 Kewley-Port，1983 中关于动态方法的内容）。

结合声学动态不变量线索，也有研究探讨对鼻音发音部位的区分。研究关注鼻音－元音音节中鼻音－元音边界附近谱的变化。实验计算了鼻音最后两个声门脉冲以及除阻后进入元音的前两个脉冲，建立了一个低频到高频范围内能量相对变化的度量标准，分别对应［m］和［n］的反共振峰位置。这个度量标准表明，［m］的特点是低频能量快速增加，而［n］的特点是高频能量快速增加。研究考察了三名发音人，包括了五种元音环境，结果鼻音发音部位的分类正确率为89%（Kurowski 和 Blumstein，1987）。

最后，人们也将声学不变量线索应用于区分爆发音和近音（主要针对滑音［w］和［j］）的发音方法的研究。除了共振峰起始频率和共振峰过渡段时长有差异，爆发音－元音组合（如［bɑ］）与滑音－元音组合（如［wɑ］）的差别还体现在辅音除阻及其后接元音嗓音起始之间的能量相对变化。由于爆发音有一个由完全闭塞到除阻的过程，爆发音到元音之间的能量变化相当大，比只有中等收紧程度的近音大得多。Mack 和 Blumstein（1983）计算了除阻后第一个音高周期与包含除阻的周期的能量比值，发现爆发音的比值都要大于滑音。若将 1.37 作为分界值，爆发音和滑音的辨识率可达92%，［d，g］与［j］的辨识率可达90%。

总之，我们上面提到的这些研究说明声学信号中有可能存在对应发音部位和发音方法的不变模式。关键是，这些模式并不特指某一声学特征，因为 Liberman 和他的同事已经明确指出，单一的声学特征通常与语音环境高度相关。而动态线索受到说话人和语音环境的影响较小，因为它整合了某个时间段内的若干声学特征。

在更近的一段时期内，有一些研究提出使用统计的方法来对辅音进行发音部位的分类。为了降低语境的影响，Forrest 等（1988）用**谱矩**（**spectral moments**）来区分词首的清阻音。操作时，从阻音的起始部分

开始，每隔 10 ms 就进行一次快速傅里叶变换（FFT）。每个 FFT 被视为一种随机概率分布（即一组可以进行统计分析的数据），从而得到前四项谱矩，即均值、方差、偏度以及峰度。频率的均值按照其振幅来衡量，反映了频率维度上能量集中区的位置。方差表示了谱能量的分布范围，偏度指出了谱的斜率（即整体能量分布的倾向），峰度则表示了分布的“峰值程度”。因此，谱矩度量标准既能表示局部（谱峰）信息又能表示整体（谱形状）信息。对谱矩进行统计性判别分析可以帮助我们对发音部位进行归类[1]。通过分析爆发音前 20 ms 的谱矩信息（发音人为 5 名男性和 5 名女性），发音部位的平均正确区分率为 89%。

另一种统计方法基于**音轨方程（locus equation）**。Lindblom（1963）首先发现，在爆发音 - 元音音节中，浊爆发音后接元音的 F2 起始频率与此元音中点的频率有相关性，浊爆发音的这个特点可用音轨方程进行定量描写。音轨方程定义为“F2 过渡段起始值与其协同发音的元音 F2 中点‘目标频率值’之间的线性回归拟合”（Sussman 等，1993：1256）。也就是说，每发一个 CV 音节，就会产生一个数据点，这个点的坐标分别是元音起始位置的 F2 值以及此元音中间位置的 F2 值。把这些点都画在坐标系中，横轴和纵轴分别为元音起始位置的 F2 和元音中间位置的 F2，此时根据所有的数据点可以得一条最优拟合线，使所有点都尽可能地靠近这条线。音轨方程可以描述爆发音与后接元音的协同发音程度。若协同发音程度最大，则在前接的爆发音中就已经达到了元音 F2 的目标值，元音中点的 F2 值便确定了元音起始位置的 F2 的值，此时音轨方程的斜率为 1，见图 10.17 中实心数据点的拟合线。如果不存在协同发音，则辅音的 F2 保持不变，与其后接元音的音质无关，音轨方程的斜率为 0，见图 10.17 中空心数据点的拟合线。

由音轨方程可知，双唇音的协同发音程度最大，斜率最陡；软腭音协同发音的程度居中，齿龈音的协同发音程度最小，斜率最小，见图 10.18。

音轨方程构成了对语音的动态描写，因为它表示了语音信号中不同位置（元音起始点和中点）的 F2 之间的关系。Sussman 等（1991）研究了英语浊爆发音，他们通过每个发音人的音轨方程对三种发音部位进行了判别分析（基于斜率和 y 截距），归类的正确率为 100%。

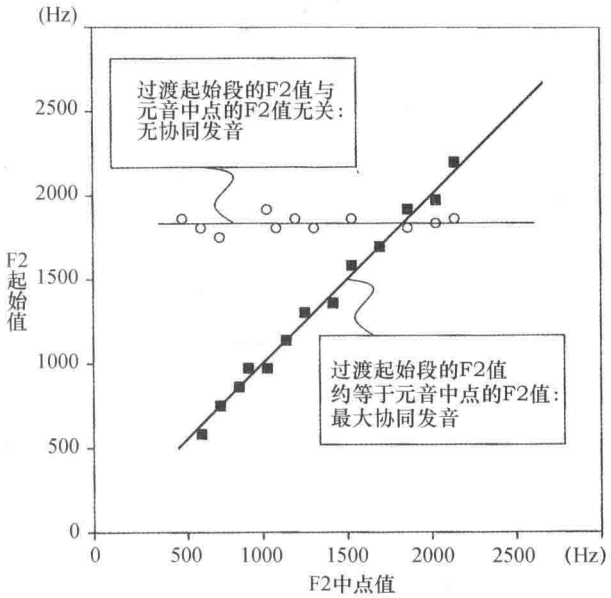


图 10.17 “辅音到元音”的最大和最小协同发音的音轨方程的示意图

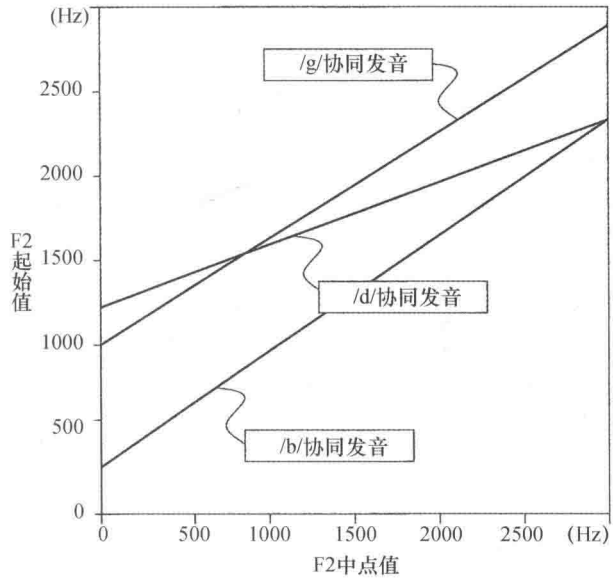


图 10.18 德语［p，t，k］后接元音［i，a，u］时，元音过渡段的 F2 值与元音中点的 F2 目标之间的回归线。数据基于 10 位德语发音人。

总之，本章讨论了不同语音类别的声学相关量。语谱图（例如图10.6）显示，有些相关量随语境的不同而变化。关于声学不变量究竟是否存在及其本质是什么，相关的争论还将持续一段时间。虽然最近用统计度量标准进行分类的正确率很高，但我们仍然不能确认听音人是否是基于类似的信息来辩认发音部位的。我们将在第13章再来讨论这个问题。

练　习

1. 荷兰语和德语中，哪些共振峰对于区分元音重要？每个共振峰都提供了哪些信息？
2. 语谱图上口元音和鼻元音有什么不同？为何有这些不同？
3. 语谱图上单元音和二合元音有什么区别？
4. 区分擦音的四个声学特征是什么？[s，z，ʃ，ʒ] 与 [f，v，θ，ð] 在这几个特征上有何区别？
5. 对于收紧程度大的语音，我们可以用一个三声管模型来模拟其声学属性：一个后腔、一个收紧部位和一个前腔。假设声管总长16 cm（声管长13 cm，收紧长3 cm），请根据以下辅音的腔体配置计算其相应的F2，F3和F4：
   “see”中的/s/（后腔长度是11 cm）
   “she”中的/ʃ/（后腔长度是10 cm）
   请分别预测这两个辅音的能量集中区的范围。
6. 辨别塞音的发音部位的声学特征是什么？塞音清浊的标志是什么？
7. 如何理解语音信号缺少不变量的观点？

注　释

1. 统计性判别分析基于一系列（声学）参数将个体分组。

11 音节和超音段

前面我们关于语音的发音特点和声学特征的讨论只关注了元音和辅音，然而口语却很少是由孤立的语音构成的。通常我们需要将元音和辅音组合成更大的单元结构，例如音节、短语和句子。我们同样可以采用发音和声学特点来描述这些更大的语音单元。这些对于超出单个元音和辅音结构的语音研究属于超音段（suprasegmental）或者**韵律（prosody）**的研究。“韵律”这个术语有时被用作“语调”（intonation）这个术语的同义词。但是，正如我们在这章将要讨论的一样，语调只是包罗万象的“韵律”的一个方面。所以我们将“韵律”看作“超音段”的同义词，而语调只是其中的一部分。

超音段特征主要包括重音（一个音节的“凸显性”）、音长、声调（在一个音节上基频的高低或者其曲折变化）和语调（在一个短语内基频的变化）。这些特征和描写音段属性的特征是相互独立的。尽管超音段这个术语体现了个体音段特征与更大语言单位特征的不同，但是把韵律结构看作是语音信号不可分割的一部分是非常重要的。

在介绍超音段的概念之前，我们有必要先讨论一下音节的概念，因为某些超音段特征，例如重音和声调，都是基于音节而言的。尽管大多数人对什么是音节有天生的语感，但是我们却很难用语音标准来定义一个音节。

11.1 音节

尽管音节通常是由元音和辅音组合而成的，但也可以只由元音（如

英语单词“eye”［aɪ］）或者只由辅音（如英语单词“cotton”［kɑtn̩］中的音节性鼻音）构成。音节通常用希腊符号 σ 表示。为了描述上的方便，我们可以说一个音节至少必须包含**音节核（nucleus）**，并且最多只能包含**音节首（onset）**、音节核和**音节尾（coda）**。通常情况下，音节的音节核是元音，音节首由元音前面的辅音构成，音节尾由元音后面的辅音构成。音节核和音节尾共同构成**韵基（rime）**（见图 11.1）。

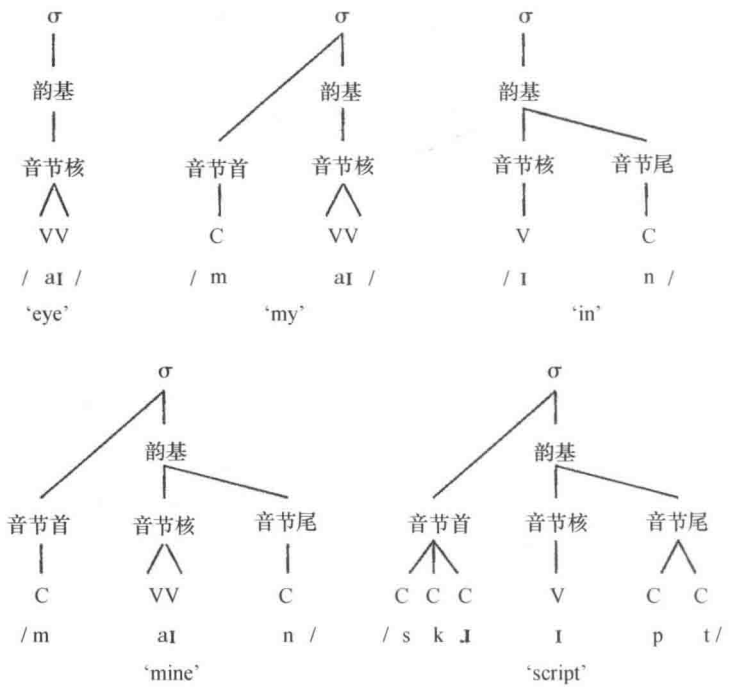


图 11.1　不同音节类型的构成成分（σ 代表“音节”）

大多数人都能够轻易地感知到音节的存在。例如，英语母语者一般都认为单词“yesterday”［ˈjɛs. tɚ. deɪ］包括 3 个音节（依据 IPA 惯例，音节与音节之间用符号［.］相隔）。但是，其他的一些情况则存在不少争议。例如“mirror”或者“fire”这样的单词，一些人认为这类词包含两个音节，另外一些人则认为只有一个音节。音节的定义应该不仅能够适用于划分明确的情况，也要能适用于上述这种“存在争议”的情况。

相关研究对音节的语音学定义主要从发音和听觉两个方面着手。早

期从发音角度出发的定义认为，每个音节都是呼吸肌肉群一次收缩运动的结果。但是，后来的研究表明音节和肌肉活动之间并不是一一对应的：某些双音节词的产出只需要呼吸肌肉群的一次收缩；相反，某些单音节词的产出则需要两次肌肉收缩。（见 5.1.4 中讨论的在胸部的肌肉活动与语音的响度大小之间建立联系的困难。）

从听觉方面来定义音节依据的是**响度（sonority）**的概念。响度是指声音相对的响亮程度。即使所有语音的时长、重音和音高都一样，一些语音也会比其他语音更响一些。例如，元音比辅音的响度高；元音之间相比较来说，[ɑ] 的响度比 [i] 高。当语音按照响度大小排列时，就形成了**响度等级（sonority hierarchy）**。人们对响度的感知主要取决于两个因素，一个是声道开口度大小（声道开口度越大，声音越响），另一个是嗓音状态。响度等级从高到低的顺序为：元音→近音→鼻音→擦音→爆发音。在每一类发音方法的内部，可以对响度等级进行更细致的划分，例如低元音比高元音响，浊爆发音比清爆发音响。

如何用响度等级来定义音节呢？音系学家提出，一个音节对应一个响度峰（如 Selkirk，1984）。**响度原则（sonority principle）**规定，响度从音节首开始增加，在音节核处达到最大，然后开始下降一直至音节尾结束。英语单词如“past”，“pastor”和“pastoral”完全符合响度原则，分别有一个、两个和三个响度峰（见图 11.2）。但是，一些英语单词并不符合响度原则，大多数不符合这个原则的是以“[s]＋爆发音”这样的辅音丛开始的单词，如“speak”[spik]。由于听感上 [s] 比其他的清爆发音和浊爆发音都响，所以“speak”就应该有两个响度峰，但是这个单词只有一个音节。或许最恰当的观点是将响度原则看作是对很多语言调查之后得出来的一个统计事实，而不是一个语法事实。很多语言中都存在不

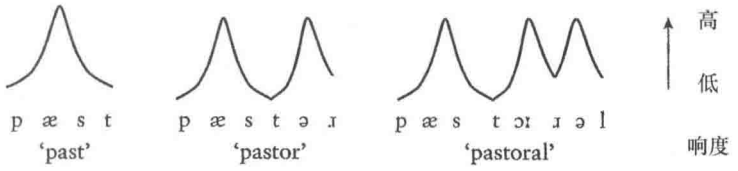


图 11.2 英语单词“past”，“pastor”和“pastoral”的理想化的响度等级

符合响度原则的情况，比如［mr, ft, rt］这样的音节首和［rj］这样的音节尾。尽管如此，把“响度”作为声音“理想化”的凸显度（凸显度大致跟声道的打开程度对应）的排序原则，并在特定的语言中允许存在一些例外，确实为划分单词的音节提供了一种规矩的方法。

总的来说，音节这个概念在语感上非常明确，在音系理论中也起着非常显著的作用，但音节依然缺少一个综合性的语音定义。

11.2　重音

重音是音节的一个属性，它使音节更加显著。重音通常划分为三个等级，从强到弱依次是：主重音、次重音和非重音。例如，在单词“phonetician”［ˌfoʊ.nə.ˈtɪ.ʃn̩］里，第三个音节负载主重音（用一个标在上角的竖线［ˈ］来表示），第一个音节负载次重音（用一个标在下角的竖线［ˌ］来表示），第二个和最后一个音节是非重音音节。从发音的角度来说，重读音节的发音通常需要身体使出更大的力气。这可以体现在喉部肌肉（特别是环甲肌）活动的加强、声门下气压的增大或者频谱斜率的调整上，也就是高频能量相对低频能量的增加（Sluijter 等，1997；见5.2.3）。从声学的角度来说，重音的四个主要声学相关量分别是基频、时长、强度和共振峰模式。（目前频谱斜率的作用还没有得到深入的研究，所以下面的讨论暂不涉及。）英语中的重读音节通常比非重读的基频高、时长长、强度大、音质弱化（即弱化为中性元音）少。重读音节的特征可以通过上述一个或者多个相关量来反映。在一些语言中，重音还可以通过基频的降低来实现（如孟加拉语；转引 Hayes 和 Lahiri，1991），或者通过更复杂的模式来实现（如威尔士语；转引 Williams，1983）。

对于英语重音声学相关量的了解要归功于一系列关于最小对立双音节的听辨研究，如名词“the **re**cord”［ˈɹɛ.kɚd］和动词“to re**cord**”［ɹə.ˈkɔɹd］这样的最小对立体。在英语中，双音节名词的重音通常在第一个音节上，有时候被称为“扬抑格”（trochaic）重音模式；而双音节动词的重音则通常在第二个音节上，即“抑扬格”（iambic）重音模式。从对最小对立体的语音转写中，我们可以清晰地看到它们之间存在音质上的区别：重读的第一个音节包含的是一个完全实现的元音［ɛ］，而非

重读的第一个音节包含的是一个弱化的元音［ə］。图 11.3 显示了名词“record”和动词“record”的波形图、语谱图，以及强度和基频曲线。在这个例子中，重音是由这四个声学相关量共同体现的。重读音节相对于非重读音节而言，表现为基频升高、时长增长、强度加强以及共振峰模式的变化。就基频对重音的作用而言，不仅仅可以通过抬高基频来影响重音，基频向任何一个方向的变化都可能影响重音。也就是说，一个重读音节的基频与对应的非重读音节的基频相比，既可以抬高也可以降低。用合成语音（音段相同但是音段自身重音的声学相关量呈系统性地逐渐变化）做语音刺激的感知实验表明：对重音而言，基频和时长可能是比强度更相关的声学线索。就共振峰而言，动词第一个音节的元音的第二共振峰（F2）比较低，而名词第二个音节的 F2 比较低。元音弱化现象在英语中很普遍，英语里很多非重读音节都弱化为［ə］。其他语言里就不一定是这样的情况。例如，在西班牙语的最小对立体“miro”［ˈmi. ro］（“我看”）和“miro”［mi. ˈro］（“他看过”）中，元音/o/的音质就不会因为重音的不同而发生改变。

需要注意的一点是，重音的声学相关量还承担着其他的功能。基频同时也起到声调和语调的作用，时长在依靠发音长短来区分音位的语言中非常重要，共振峰模式是元音音质的主要相关量。重音的某个声学相关量的相对重要性可能部分取决于这个声学相关量在表征其他区别特征时所担任的角色。

除了音节以外，**音步（foot）**也是以往重音讨论中的一个重要成分。我们可以将音步定义为由一个重读音节和这个重读音节之后的所有非重读音节构成的单位（其他关于音步的定义参见 Hayes，1995）。人们认为重读音节和非重读音节的交替出现构成了有节奏的凸显峰。在单词“information”［ˌɪn. fɚ.ˈmeɪ.ʃn̩］中，重读音节［ɪn］和［meɪ］后面都跟有一个非重读音节。这样，单词“information”包含两个音步［ɪn. fɚ］和［meɪ.ʃn̩］。凸显音节和非凸显音节交替出现的语言被称为**重音计时（stress-timed）**语言。在重音计时语言中，又分为固定重音和自由重音两种。在固定重音的语言里，主重音总是出现在相同的音节上。例如，在捷克语中，重音总是出现在单词的第一个音节上；在波兰语中，重音几乎总是负载在单词的倒数第二个音节上；在法语中，重音几乎都负载在

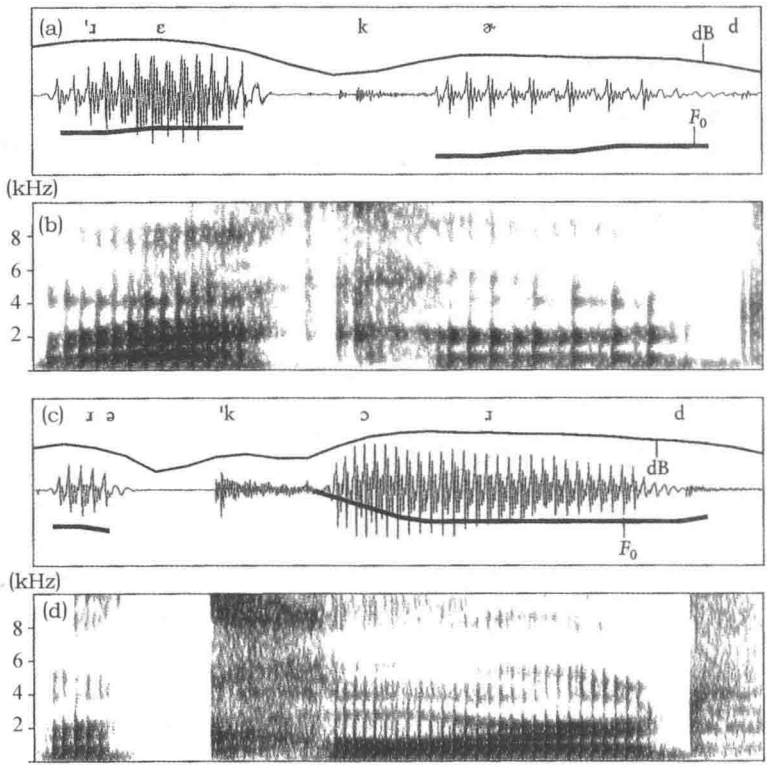


图 11.3　名词“record”［ˈɹɛkɚd］和动词“record”［ɹəˈkɔɹd］的波形图、语谱图、强度和基频曲线（图 a，b 为名词；图 c，d 为动词）。

单词的最后一个音节上。在自由重音的语言中，例如英语和德语，重音并不总是固定在单词的相同位置上，不同单词中重音的位置各不相同（转引 Hayes，1995）。

在句子层面上，交替出现的重音可以产生不同的节奏。例如，短语“**Ed**dy's **play**ing a trom**bone**”中包括三个音步：“**Ed**dy's”、“**play**ing a trom”和“**bone**”。人们认为，在诸如英语和其他日耳曼语这样的重音计时语言中，语言学意义上的节奏是等长的重音间隔，具有**等时性（isochrony）**。也就是说，在重音计时的语言中，重音之间的间隔有等长趋势。换句话说，刚才的例句中，每个音步（“**Ed**dy's”、“**play**ing a trom”、“**bone**”）无论音节数量的多少，其时长都是相等的。但是，如图 11.4 中的波形图所示，重音间隔的等时性还没有得到充分的实验证明。

在诸如法语和西班牙语这样的**音节计时（syllable-timed）**的语言中，与重音计时的语言（重音计时语言中的重读音节被认为是以固定的时间间隔出现的）不同，所有音节，不论是重读音节还是非重读音节，都被认为是以固定的时间间隔出现的。人们认为语言学意义上的节奏缘于音节之间间隔的等时性。也就是说，对于音节计时的语言而言，重读音节和非重读音节的时长相差不大。确实，实验证明，西班牙语（音节计时语言）中的重读音节和非重读音节的时长差异要比英语（重音计时语言）中的小。西班牙语中非重读音节不发生弱化这一事实，也是造成其重读音节和非重读音节之间的间隔相等的原因之一。但是，音节时长也不总是恒定不变的。

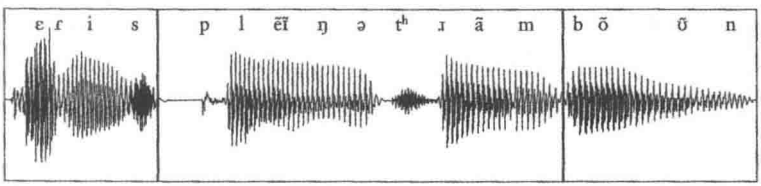


图 11.4 句子“Eddy's playing a trombone”的波形图。从标注出来的每个音步的时长来看，这并没有证明等时性的存在。

我们应该注意到，日语是第三种节奏类型，即**莫拉计时（mora timing）**。日语中有一个单位叫**莫拉（mora）**。在日语中，音节和莫拉之间的区别是，一个音节末尾的辅音自己就构成一个莫拉（一个长元音的后半部分也构成一个莫拉）。根据传统的理论，每个莫拉的时长是大致相等的。例如日语单词/ha-ʃi/（“边”）和/ke-k-ko-ŋ/（“婚姻”），两个单词都包含两个音节，但是第一个单词只有两个莫拉，而第二个单词则有四个莫拉（为了方便起见，莫拉之间用短横线相隔表示）。一些时长测量支持了以莫拉为基础的解释，也就是说，/ke-k-ko-ŋ/的时长确实是/ha-ʃi/的两倍。但是，最近更细致的声学研究并没有能够证明莫拉是一个直接的时长单位（Warner 和 Arai，2001）。

根据句子结构的不同，句子里的同一个单词可以负载或者不负载重音。在句子“I crave fame **and** fortune”中，强调“and”，“and”就是以**强**

形式（strong form）即**非弱化形式（unreduced form）**［ænd］说出的。相比之下，同样的句子，“I **crave** fame and fortune”，也可以不强调“and”，“and”就是**弱形式（weak form）**即**弱化形式（reduced form）**［ənd］。功能词（冠词、连词、介词和代词）通常比实词（如名词和动词）更容易出现这种强－弱两读的变化。从语音上来说，这种强弱模式上的不同可以使元音音质产生变化，甚至会导致元音或者辅音的丢失，例如“and”的弱读形式有［ənd］、［ən］、［n̩d］和［n̩］。

在句子层面上，一个单词的重音也有可能随相邻单词重音的变化而变化。例如，单词“fourteen”出现在句子“I'll turn fourteen tomorrow”中时，“fourteen”是第二个音节重读；但是当出现在句子“I am fourteen years old”中时，通常重读“fourteen”的第一个音节。这种现象被称为**重音移位（stress shift）**或者**节奏返转（rhythm reversal）**。我们发现在句子“I am fourteen semesters behind”中，“fourteen”也是重读第一个音节。不过与上一例句不同，这一例句中“fourteen”这个单词后面紧接的是一个非重读音节，这说明重音移位不是因为两个重读音节相邻而触发的；相反，重音移位是在短语层面上发生的。正如 Carr（1999）指出的那样，重音移位是指“当在同一个短语结构内，若一个次重音加主重音的序列之后存在另一个主重音音节，则交换次重音和主重音的顺序”（Carr，1999：111）。

当单词进入句子中时，**音高重音（pitch accent）**或者**核心重音（nuclear accent）**可以使某些重读音节比其他重读音节更加凸显。音高重音就是一段话中使某个特定音节相对更加凸显的音高结构。一个负载了音高重音的音节比其他任何没有负载音高重音的音节读得更重。在英语或者其他重音计时语言中，音高重音音节一定是一个重读音节。在语调短语（见 11.4.2）中，最后一个重读音节比其他重读音节更为凸显，所以这是一种特殊的音高重音类别，被称为核心重音。这个承载核心重音的音节就是人们通常所指的“重读音节”。核心重音的位置由篇章的属性来决定，即说话人想要对听话人强调的部分（例如对话中的新信息或者对比信息）。

11.3 音长

音长的声学参数是时长。时长受很多因素的影响，包括从音段层面到句子层面的各种因素。目前人们所了解的局部因素或全局因素对音段时长的影响大都来自语音合成方面的研究。早期的合成语音（通过计算机生成的语音）有时人们根本听不懂，那是因为合成规则没有恰当地调整音段的时长。这就促使研究者们进一步详细地确定和量化影响时长的各种因素。表 11.1 列出了一些影响时长的因素，这个表是根据 Klatt 的一个类似的表制作的，Klatt 是现代语音合成研究的先驱之一。

表 11.1 影响音段时长的参数（参照 Klatt 1973）。Klatt 为下面的每个陈述都提供了参考文献，我们在这里就省略了。

1 插入停顿原则	在每个句子内部的主句前面或者在有逗号标记的边界处插入小停顿。
2 分句句末延长	延长停顿前的音节中的元音或者音节性辅音的时长。在这个韵（即元音和停顿之间）中的任何辅音也都会得到延长。
3 短语末尾延长	如果音节性音段（包括元音和音节性辅音）是短语末尾的最后一个音节，则将其延长。在复杂名词短语中或者当主语 - 动词 - 宾语的顺序被打乱时，名词短语或者动词短语的边界可以延长。代词的时长很少会被延长。时长的延长在听感上十分重要。
4 非词尾音段缩短	非词尾音节中音节性音段的时长会被缩短（这条规则存在争议）。
5 多音节缩短	多音节词中的音节性音段的时长会被轻微缩短（这条规则也存在争议）。
6 非词首辅音缩短	如果辅音处于非词首的位置，则缩短其时长。
7 非重读缩短	非重读音段比重读音段短而且压缩更多。
8 强调延长	被强调的元音，其时长明显延长。
9 元音后的语境	同一个单词中，如果元音之后是清辅音，则元音时长缩短。这种缩短效应在短语边界或者分句边界处更加明显。
10 辅音丛缩短	辅音 - 辅音序列中的音段缩短（与词边界无关，但不能跨短语边界）。
11 爆发音送气导致的延长	如果一个重读的元音或者响音前面有清爆发音，则该元音或响音的时长会相应地延长。

从发音角度来说，时长取决于发音器官运动的时间长短。音段、音节、音步和句子的时长受到很多因素的影响，既包括局部因素也包括全局因素。一个影响时长的主要因素是音位本身的音长对立，被称作**对立音长（contrastive length）**。许多语言都有长元音和短元音的对立。在像阿拉伯语、丹麦语、爱沙尼亚语、埃维语、芬兰语和日语这样的一些语言里，音长是某些元音对立的主要特征，长元音和短元音的时长之比在1.3∶1至2∶1之间，也就是长元音比对应的短元音长30%至100%。在这些情况下，长元音用音长符号［ː］表示，如［iː］和［i］。但是，在像荷兰语、英语和瑞典语这样的语言里，长元音和短元音的对立主要是音质上的差异，其次才是时长。在这些情况下，长元音和短元音的对立用两个不同的语音符号来表示，如［i］和［ɪ］。

尽管辅音音长对立没有元音音长对立现象普遍，但该现象在很多语言中的确存在。意大利语是一个经常被提到的有辅音长短对立的例子，如最小对立体“fato”［ˈfato］（“命运”）和“fatto”［ˈfatːo］（“已做”）。这些长的辅音被称作**长辅音（geminate）**。其他有详细记载的具有单爆发音和长爆发音对立的语言包括阿拉伯语、孟加拉语、爱沙尼亚语、芬兰语、意大利语、卢干达语、瑞士德语和土耳其语。这些语言中的长辅音通常是对应的短辅音的两倍或者三倍长。长爆发音可以通过增加闭合段的时长来实现（如瑞士德语），或者通过增加嗓音起始时间（VOT）的时长来实现（如塞浦路斯希腊语）。

影响音长的另一个因素是固有时长差异。当其他因素不变时，低元音比高元音长20～25 ms。这个差异是可以被感知到的，因为心理物理学研究证明：音段固有的时长差异已经大于听音人能感知到的两个音之间的最小时长差异**（最小可觉差 just noticeable difference，JND）**。人们发现很多语言都存在音段固有时长的差异，这一特点很可能具有语言普遍性。一般认为，可以解释舌位的高低与时长之间关系的说法是，发低元音时发音器官的移动程度更大。低元音向后接辅音过渡的时候，由于辅音在发音时声道更窄，对应的舌位较高，舌需要移动较大的距离，因此需要更多的时间以达到发音目标。对辅音来说，唇音比齿龈音或者软腭音更长：一种解释就是双唇没有舌的运动活性大，因此运动得更慢一些。

还有其他一些局部因素会影响音段的时长。其中的一个因素在3.4.2.1中已经提到过了，就是后面跟有浊辅音的元音比后面跟有清辅音的元音要长。其他影响音段时长的因素包括元音前后的辅音的发音部位和发音方法。更大范围的因素包括一个音节中音段的数量和一个单词中音节的数量。总的来说，音段的时长随着音段和音节数量的增加而减少。例如，元音［u］的时长在单词“fruitier”中最短，在“fruity”中要长一些，在“fruit”中最长。这正好也是等时性的一个例子，在较长的单词中缩短音段的时长，以保证短语层面上重读音节之间的时长大致相等。

除了局部因素，全局的影响因素包括音段或者音节在整个句子中的位置。我们可以通过测量并比较语流中位于不同位置的同一音节的时长来观察这种影响。在短语末尾的音节比在短语起始或者短语中间位置的音节要长得多，这种现象被称为**短语末尾延长（phrase-final lengthening）**。

最后一个影响音段音长的全局因素是**语速（speaking rate）**。语速是某个话轮的整体速度。相同的一句话（“How do you do?”）可以以慢速或者快速的方式说出来。语速是基于每秒的音位数或者音节数来计算的。对英语来说，平均语速是每秒4到6个音节。但是，随着语速的增加，通常会伴随音位产出的变化。所以，语速的计算是基于理想化的正则发音还是实际发音，其结果是不同的。

在计算语速的时候，话语中的有声停顿（如“呃，”“哦，”等）以及话语之间标注话轮转换的无声停顿都必须考虑在内。语速的加快或减慢会缩短或增加音段的时长和停顿的时长。但是，语速的变化对所有音段和停顿时长的影响是不同的。总的来说，语速变化对时长较长的音段影响大于时长较短的音段，例如元音的时长变化要比辅音的大。除了时长方面受到影响之外，语速的变化还会影响 F_0 的范围，语速的增加通常会缩小 F_0 的变化范围。

11.4 声调和语调

基频的变化是词重音的一个重要声学线索，不过还有很多其他的局部或全局因素会影响基频的变化。在7.3.1.4中提到过，音高是基频的听

觉相关量，而声带振动是基频的发音相关量。影响基频的最显著的因素是性别，男性的平均基频值为 130 Hz，女性的平均基频值为 220 Hz（当然，这些值可能随发音人文化的不同而产生变化）。另一个影响基频的因素是说话人的情感状态。例如，人在生气、害怕或者高兴的时候，基频就会显著升高。而且不同元音的**固有基频（intrinsic fundamental frequency）**也不一样[1]。总的来说，高元音的基频比低元音的基频高（大概高 15 Hz）。这种基频差异完全在人的感知范围内，因为在理想情况下，人对于基频的最小可觉差大约是 1 ~ 2 Hz。如果是真实语音的声音刺激，则相应的最小可觉差是 10 Hz（Klatt，1973）。

世界上的许多语言中都发现存在语音固有基频的差异。一种解释认为，这种差异是受喉部和舌之间肌肉连接的影响。根据“舌带动”（tongue-pull）理论（Lehiste，1970），发高元音时，舌的抬高使喉部也升高，这导致环状软骨和甲状软骨之间形成了一个倾斜，从而使声带拉紧，最终引起了基频的升高。根据基于生理学的“喉部依附”（laryngeal dependency）理论（Ladd 和 Silverman，1984），发高元音时引起的高度的肌肉紧张也引起了喉部肌肉的高度紧张。还有，另外一种“声学耦合”（acoustic coupling）理论（Flanagan 和 Landgraf，1968）认为，收紧的声道使气流速度增加，产生了伯努利效应，这样喉部以上的气压降低，从而增加了气流通过声门的压力，并最终提高了声带的振动频率。最后，在“说话人的主动控制”下，基频可以增加元音在舌位高低上的区分性（Kingston 和 Diehl，1994）。

与影响时长的局部条件因素相似，影响基频的一个条件因素是，当元音前面是清阻音的时候，元音音段的基频比前面是浊阻音时要高（大约高 5 ~ 10 Hz）。

11.4.1　声调

性别是导致说话人之间基频差异的主要原因，但是影响同一个说话人基频的最主要的因素是具有音位性质的基频对立，被称作**对立声调（contrastive tone）**或者**词调（lexical tone）**。声调的定义是可以区别词义或者语法意义的音高变化。使用声调区分词义的语言叫作**声调语言（tone language）**。

声调语言可以包含**平调（register tone）**或者**曲折调（contour tone）**。在平调语言中，每个音节的音高在说话人音域范围内所对应的相对高度可以提供区分词义的线索。这些语言通常可以有至多4个相互区别的声调。非洲的大部分语言属于平调语言（包括伊博语、绍纳语、约鲁巴语、祖鲁语，等等）。例如，约鲁巴语有三个平调：高调，如［ó.bɛ́］（“他跳”）中的第一、第二个音节；中调，如［ó.bɛ̄］（“他向前”）中的第二个音节；低调，如［ó.bɛ̀］（“他请求原谅”）中的第二个音节。需要注意的是，根据IPA的惯例，高调用一个在元音上方的左斜重音符号［ˊ］表示；中调用一个在元音上方的长音符号［ˉ］表示；低调用一个在元音上方的右斜重音符号［ˋ］表示。

在曲折调语言中，词义是通过音高的曲折变化而不是音高水平来区分的。每一个声调的音高曲线都有一个特定的形状和特定的变化方向。东南亚的许多语言（包括汉语普通话、泰语和越南语）都是曲折调语言。例如，汉语普通话有四个声调：高平调（阴平）、高升调（阳平）、低降升调（上声）和高降调（去声）。图11.5显示了每个声调的音高曲线。阴平确实是在说话人调域的高调域，而且整个调形保持水平；阳平的起点虽然较阴平低，但是仍然属于说话人的高调域，整个调形是上升趋势；上声是从说话人的低调域开始，向下降得更低，然后再上升，直至调尾；去声的起点和阴平一样高，但是之后就急速地下降。

表11.2给出了四个声调的标音样例。标为“调值”的一栏表示声调通常采用的数值标记方法，1和5分别表示说话人调域的最低值和最高值，2、3、4是平均划分说话人的调域所得（例如，“35”表示一个声调曲线从调域中部开始一直上升到调域的最高部）。IPA有两种标注声调的符号集：一种是“声调符号”，另一种是“声调字母”。两种符号都列在IPA表的“声调与词重音”子表中。声调符号应标注在元音的上方。在表11.2中，声调符号这一栏显示了用于表示汉语普通话声调的四个符号。而声调字母是附加在音节之后的，由一条竖线和竖线前的笔画构成。竖线代表调域，前面的笔画代表在这个调域中的音高变化的趋势和高度。

从历史演变来说，由一种非声调语言演变成一种声调语言的过程叫作**声调发生学（tonogenesis）**。在这里我们仅仅讨论其中一种词汇声调发展的方式。这种演变与前面提到的一种音段条件因素有关。之前说到，

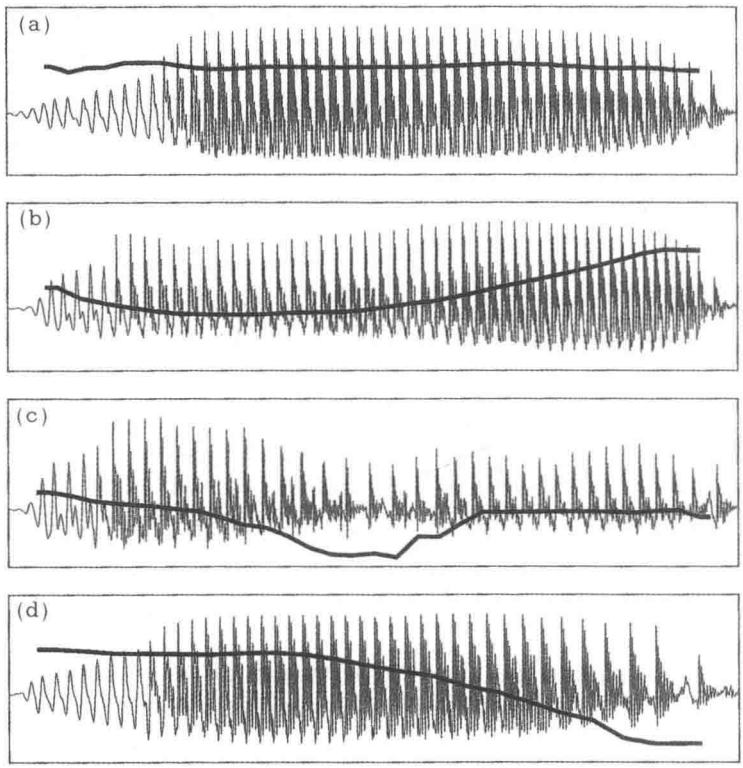


图 11.5　由一名成年男性朗读的汉语普通话里［ma］的四个声调的波形图和音高曲线（用黑线表示）

在清阻音之后的元音的基频比在浊阻音之后的高。研究（转引 Hombert，1978）表明，一些声调对立的出现来源于主要特征丢失之后，听话人对原次要特征进行的重新理解。例如，在阻音有清浊之分的语言中，浊阻音后面的元音的音高有所降低。正如前面提到的，这种现象可能是说话人为了加强清浊对立而主动产生的。但是，当这种语言逐渐丢失这种爆发音中的清浊对立之后，原本处于次要地位的音高差异现在就变成了区分词义的主要特征。进而，一种声调语言就产生了，在元音上出现了高调和低调两种声调。这种原来只是清浊对立的副产物的音高差异就变成了声调对立，成为这种语言音位系统中的一部分。

表 11.2 汉语普通话里的四个声调和三种标音方法

	调形	调值	声调符号	声调字母	例字
阴平	高平	55	[mā]	[ma ˥]	妈
阳平	高升	35	[má]	[ma ˧˥]	麻
上声	低降升	214	[mǎ]	[ma ˨˩˧]	马
去声	高降	51	[mà]	[ma ˥˩]	骂

11.4.2 语调

语调是在比单词更大的单元上对音高的区别性特征的应用。语调可以通过标示句法结构的边界或者区分不同的句型，如陈述句、疑问句和命令句来传递语言学信息。语调也可以传递非语言学信息，例如厌倦、不耐烦或者礼貌。在这里，我们只讨论语调的语言学功能。

很多语言都会通过一个降调（就像在陈述句中一样）来表明句法单元的边界，如一个短语或句子的结束。许多语言都会用升调来表示是非疑问句（这样的疑问句通常用“是”或“不是”来回答）或者其他询问信息的问句。尽管英语的陈述句和疑问句通常在单词顺序上有所不同（“This is Jane.”与“Is this Jane?”），但是如果使用不同的语调，一个陈述句既可以表示陈述也可以表示疑问。例如，用降调来表达“We've been invited to a party tomorrow.”是告知某人有关邀请这件事情，但是用升调来表达“We've been invited to a party tomorrow?”就表示说话人对于存在这一邀请事实感到很惊讶。

图 11.6 显示了陈述句和是非疑问句的音高曲线。可以注意到，陈述句中，基频有一个贯穿整个句子的总体的下降趋势，这个现象叫作**下倾（declination）**。相比之下，在是非疑问句中，基频有一个比较明显且很陡的上升趋势，特别是在句子的末尾。

在英语陈述句中经常能观察到的整体音高下倾也会因为局部的强调而受到影响。请看下面的句子，斜体字是被强调的部分：

(a) We will buy our *first* home tomorrow.

(b) We will buy our first *home* tomorrow.

(c) We will buy our first home *tomorrow*.

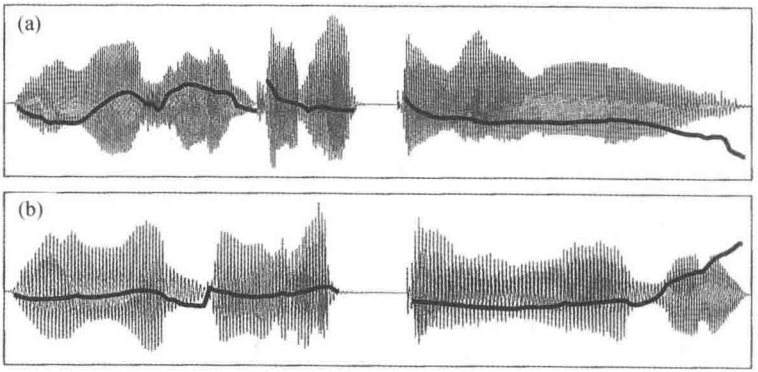


图 11.6　句子“We've been invited to a party tomorrow.”读作陈述句（a）和是非疑问句（b）的波形图和音高曲线

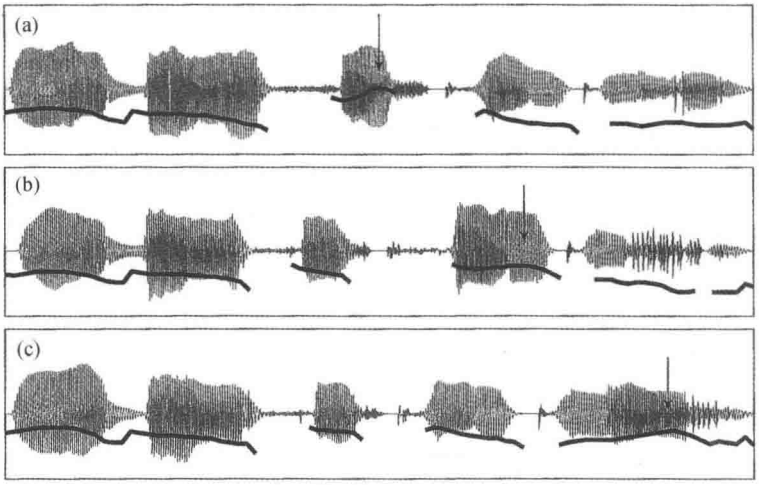


图 11.7　句子“We will buy our first home tomorrow.”分别强调（a）“first”、（b）“home”、（c）“tomorrow”（箭头指示的部分）时的波形图和音高曲线。

图 11.7 显示了这三种读法的音高曲线。语调曲线在**调核音节（tonic syllable）**处发生了明显的变化，这个调核音节是短语中负载了主重音的音节（即例句中的斜体音节）。

到现在我们讨论的句子都只有一个**语调短语（intonational phrase）**，但是句子“We will buy our first home if the bank will let us.”包含了两个语

调短语。如图 11.8 所示，第二个语调短语“if the bank will let us”的起始处出现了基频的抬升，这表明了第二个语调短语的存在。

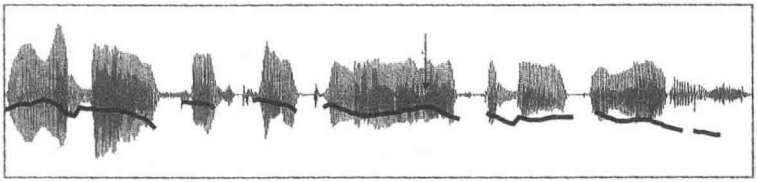


图 11.8 句子“We will buy our first home tomorrow if the bank will let us.”（包含两个语调短语）的波形图和音高曲线。箭头标明了第二个语调短语“if the bank will let us”的起始处。

图 11.9 显示了可以在句法结构上构成最小对立体的句子的音高曲线。句子“Almost everyone knows but Linda hasn't found out yet.”包含两个并列分句，句子“Almost everyone knows Belinda hasn't found out yet.”在音段上与前一个句子几乎一样，但是这句话包含一个主句和一个嵌入分句。这两个句子音高曲线的不同之处在于，两个并列分句间的边界的“降 - 升”模式比主句和嵌入句之间的更加明显。在这两句话中，单词“knows”的基频峰值相同，但是包含并列分句的句子中的“knows”的基频最低值和“lin”的基频峰都低了很多（大约低 5 Hz）（Cooper 和 Sorensen，1977）。研究也表明，听话人在划分句子结构的时候可以利用这类基频特征信息。

到目前为止，我们对重音、声调和语调的讨论似乎一直是彼此孤立的。但是，一旦我们研究的范围大于单词这个层级，重音、声调和语调的声学特征就开始相互作用。就像我们之前提到的，语调可以使某些单词更加凸显。此外，我们并不是说一种语言只能用音高来表示声调或只能用音高来表示语调，或者一种语言只能是重音语言或只能是声调语言（见 Hyman，2006）。在大多数声调语言中，陈述句通常是降调（也就是说，这些语言使用语调）。而且，在有一些语言中，例如皮拉罕语（Pirahaã）、玛雅语（Ma'ya）和库拉索帕皮阿门托语（Curaçao Papiamentu），既有重音又有声调。

最后，我们来讨论一种既不是声调语言又不是重音语言的语言类型。

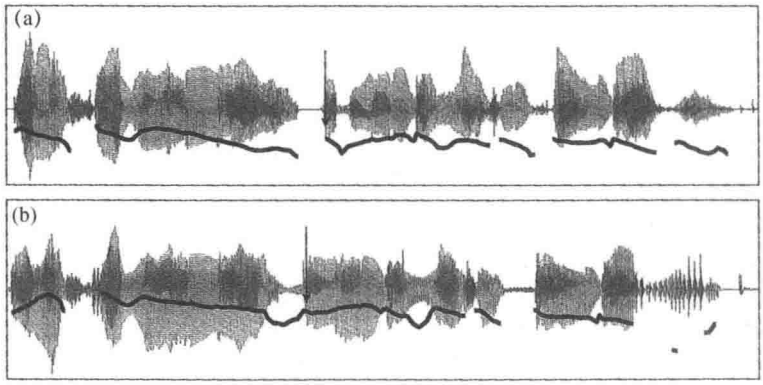


图 11.9　句子（a）“Almost everyone knows but Linda hasn't found out yet.”和（b）句子“Almost everyone knows Belinda hasn't found out yet.”的音高曲线。这两个句子的不同的句法结构体现在［lin］开始部分（箭头指示部分）不同的语调模式。

再强调一下，在这一章里介绍的对立和分类并不总是界限分明的。这种语言类型就是音高重音语言，日语就是一个很典型的例子。音高重音语言同声调语言相似的地方在于单词的每个音节（对日语而言是莫拉）都负载一个特定的声调。但是，和声调语言不同，在音高重音语言里，只要知道重音在单词中的位置，就可以知道整个单词的声调模式。请参见表 11.3 中的日语例子。

重音所在的位置用附加符号［'］来表示，它的位置是不可预测的，每个单词的重音位置都需要靠学习来掌握。但是，一旦知道了重音的位置，整个单词的声调模式就可以由两个简单的规则推出：第一条规则是，重读莫拉和其之前的所有莫拉都是高调（H），重读莫拉之后的所有莫拉都是低调（L）；第二条规则是，（在日本东京地区的方言里）单词的第一个莫拉如果没有重读就是低调。一旦知道了重音的位置，表 11.3 中的所有声调模式都可以用这两条规则来解释。像汉语普通话这样的声调语言就不同于此，说话人必须学习单词中每个音节的音高模式。

表 11.3 音高重音语言日语（东京方言）的单词的声调模式

声调模式	单词	词义
H – L	[sóɽa]	天空
L – H	[kawá]	河
L – H – L	[kokóɽo]	心
L – H – H	[otokó]	男人

在这一章里，我们讨论了几个主要的超音段特征，重音、音长、声调和语调以及与它们相对应的声学相关量，基频、时长、振幅和共振峰频率模式。从听感上来说，听觉系统依赖区分语音信号的四个维度是音高、时长、响度和音质（音色）。在下一章里，我们将讨论人的听觉系统的最基本的结构和功能，以及听觉系统和语音信号之间的非线性转换关系。

练　习

1. 韵律是什么？用来描述韵律的特征有哪些？
2. 为什么很难给音节下定义？什么是响度等级？响度等级对定义音节有怎样的帮助作用？
3. 重音是什么？重音的声学相关量是什么？
4. 简述一下重音计时、音节计时和莫拉计时语言之间的不同。
5. 平调和曲折调语言的区别是什么？
6. 语调和重音有什么不同？

注　释

1. 有时候叫作“固有音高”。不过固有基频比固有音高这个说法更好，因为声带的振动频率增加了（这不属于感知现象），而 F_0 的增加不一定会引起感知音高的增加。

12 听觉的生理学及心理物理学

由于历史的原因，人们是根据言语声的产出方式来对言语声进行分类的。使用这种基于语音产生的方法来研究语音的思路，是由当时有限的测量和表征语音的手段决定的。产生言语的器官相对比较易于接触和研究，人们对言语产出中的一系列事件的研究是通过自省和仔细的观察得到的。但是对感知研究而言，人耳太小而且不易研究。要研究活体人耳的内部情况，人们需要先进的技术设备。而且，涉及“听”这个任务的大部分过程都与大脑的加工有关，大脑最终将感知到的“声音”转换为“言语”。因此，对听觉过程的科学研究还相对比较年轻，而且我们对很多领域的了解都还不够充分。

随着最新的**脑成像技术（brain imaging technique）**（包括脑电图 electro-encephalograhy，EEG；脑磁图 magneto-encephalography，MEG；功能性磁共振成像 functional magnetic resonance imaging，fMRI）的出现，我们现在可以在一定程度上观察听音人在听到某个语音（[ɑ] 或者 [i]）时的大脑活动。我们也可以通过一些间接的方法来研究与听觉过程有关的其他各个方面，但是只能得到一些推论：例如，我们可以询问被试听到的是 [ɑ] 还是 [i]，然后把这个结果与语音信号的声学特征联系起来。

人们发现，一个音位可以通过不同的发音动作来产生。这说明，除了发音方面的描写外，对语音的听觉和感知方面的描写也非常重要。例如，不同的说话人可以通过截然不同的舌位来产出元音，而圆唇元音（如 [u]）的产出并不一定真正存在圆唇的动作：请站在镜子前面，尝试在不圆唇的情况下发 [u]。这表明在言语感知中起主要作用的是听觉目标（例如谱中能量的分布）而不是发音目标（例如舌位）。换句话说，尽管我们可以通过发音特征来描写语音，而且现有的发音范畴总体而言也

十分有效，但是使用听觉或感知范畴可能更有优势。

下面的小节主要介绍关于人耳和听觉的基本知识。听觉器官由一系列结构组成：外耳、中耳和内耳。内耳包含相对简单的知觉细胞，这些细胞被包围在稀薄的液体中。这些知觉细胞所激发的神经脉冲将由大脑再做进一步分析。

12.1　外耳

人们通常所说的“耳”只是实际听觉器官的一小部分。如图 12.1 所示，**外耳（external ear）**包括可以看见的**耳廓（auricle/pinna）**，**外耳道（meatus/ear canal）**和**鼓膜（tympanic membrane/ear drum）**。

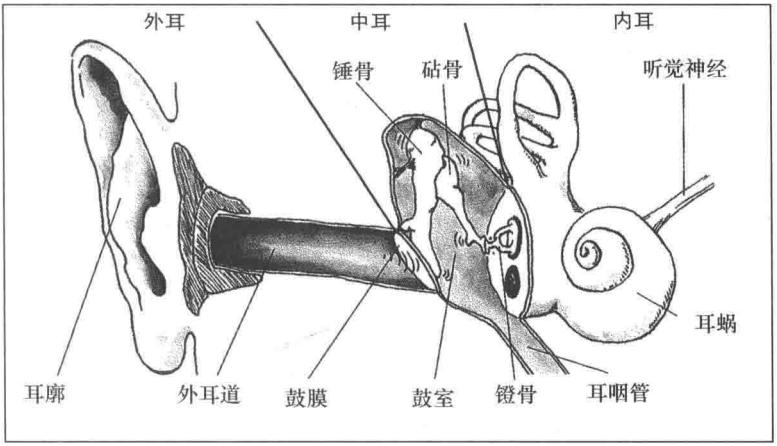


图 12.1　外耳、中耳、内耳和它们的主要结构

耳廓可以帮助人们确定声源的位置，特别是判断声源来自头的前方还是后方。我们可以通过一个简单的实验来证明这个原理。当我们戴上头戴式耳机时，由于耳机屏蔽了耳廓的定位功能，我们无法获得声源相对于头部的位置信息。因此，耳机里播放的音乐就好像是从“头里面”的某个地方传出来的。

声波通过耳廓之后，就到达了耳道。这个通道的内部覆盖着一层很

薄的耳蚕（earwax；医学上称作“耵聍”cerumen），它是一种有毒的黏稠性的物质，可以防止耳道干燥并同时防止昆虫进入耳道。如果耳道被堵塞，例如，被耳蚕堵塞，听力就会大大减弱。

耳道大约长 2.5 cm，宽 8 mm，一端封闭。如果把耳道模拟为一个一端开口的管道（即一个四分之一波长共鸣器；见 9.1.1），那么它的第一共振频率大约为 3400 Hz[1]。但是由于耳道有很高的阻尼，高于或低于这个共振频率的频率也可以被耳道传导。从声学角度来看，耳道是一个滤波器，它可以使 2 kHz 到 5 kHz 之间的频率相对于其他频带放大 15 dB。我们在第 10 章提到过，这个频率范围对语音信号来说是非常重要的。

耳道的另一端被鼓膜封闭。鼓膜是一层很薄且具有弹性的薄膜，它在外耳和中耳之间形成了一个气密的边界。受到声波带来的气压变化的影响，鼓膜可以向内或者向外弯曲，就像 7.2.1 中所描述的麦克风里的薄膜一样。鼓膜的这些运动会被传递至中耳。在非常高的声压级别下，鼓膜的不同区域会以不同的相位开始振动。这种**局部振荡（partial oscillation）**减弱了鼓膜的传输能力。这意味着在某个特定的时间点，声波的压力变化会使鼓膜的一些区域向内弯曲，而另一些区域向外弯曲。这就抵消了传输到中耳的一部分力。因此，这种机制可以保护人耳的内部，以避免其受到高声压级的损伤。

12.2　中耳

中耳（middle ear）包括**鼓室（tympanic cavity）**（见图 12.1）。鼓室是一个几乎气密的空间，里面含有人体中可以活动的最小的骨骼：**听小骨（ossicles）**。其中，**锤骨（malleus/hammer）**和鼓膜相连，它把鼓膜的运动传输至**砧骨（incus/anvil）**，然后砧骨再把运动传输至**镫骨（stapes/stirrup）**。因此，这三块听小骨可以将声波从外耳传至内耳。锤骨和镫骨与肌肉相连，从而可以减弱听小骨对声音的传导（见 12.2.2）。

12.2.1　中耳内的压力增加

从外耳传至鼓膜的声波会通过中耳里听小骨的机械运动传递至内耳。内耳中充满了稀薄的液体。这些听小骨把弹性介质（空气）中的压力变

化转化为几乎不能压缩的液体（水）的压力变化。在这个转换过程中，这些听小骨的作用就像一个圆锥体，它们把施加在面积较大的鼓膜上的力传输到镫骨下面面积较小的表面上（见图 12.1）。施加在这两个表面上的力基本相等，但是由于镫骨下的面积非常小，所以镫骨处的压力就比鼓膜上的压力大得多（见附录 A.1）。鼓膜表面面积和镫骨底面面积的比率使压强增加了大约 17 倍。此外，锤骨比砧骨稍长，所以这两块听小骨一起产生了杠杆效应，而较大的砧骨上的运动被传递到较小的镫骨上形成了更加有力的运动。对人类来说，砧骨长度和镫骨长度之比大约为 1∶1.3。表面积大小的变化和杠杆效应的结合使得中耳具有 $17 \times 1.3 \approx 22$ 的放大因子（amplification factor）。换句话说，传导至内耳的压力变化比声波最初的空气压力变化大约强 22 倍。如果用分贝来表达，这个放大因子大概为 $20 \times \log(22/1) \approx 27$ dB。

由空气进入液体时，受到的压力一定会增加，凡是曾跳入过泳池的人都会有所体会。如果以比较大的表面进入水中，也就是，如果入水时腹部先接触到水面，入水就不会太深；事实上，水会把跳水的人“反射”出来。但是如果入水时手指先接触水面，也就是入水时以非常小的面积接触水面，入水就可以很深。声波的传输也是一样：在中耳的作用下，鼓膜的大表面变成了镫骨上的小表面。如果不通过中耳的机械原理增大压强级别，那么大部分声波就会被内耳的液体表面反射出去。

听小骨的杠杆效应对频率在 800 ~ 2500 Hz 的声波最为有效，而对于频率在 2500 Hz 以上的声波的作用要弱一些，因为在这些频率上，鼓膜会产生一些局部振荡。而对于 800 Hz 以下的频率，听小骨会连接到其支撑结点和肌肉上，使声音的传播受到抑制。

12.2.2 中耳内的声音衰减

中耳的另外一个作用就是“音量控制”。支撑锤骨和镫骨的肌肉可以稍微拉紧，这样就可以起到抑制低频能量的作用。尽管这种抑制机制（可以达到 20 dB）反应迅速，但是这并不能十分有效地保护人耳免受突然出现的强噪声的刺激。要使中耳里的肌肉拉紧，大脑必须发送相应的神经脉冲——而要发送这个神经脉冲，必须要等内耳对噪声进行加工之后才能实现。也就是说，这种对声音信号的抑制只有在高强度的声音信

号到达而且很有可能已经损伤了内耳之后才会产生。

图 12.2 给出了中耳这一工作机制的原理，横轴表示时间。图中展示了声音级别的突然增大，以及随后大脑向肌肉发出神经脉冲的过程。在图 12.2a 中，信号先开始（A），肌肉随后才拉紧（B）。

图 12.2b 表示中耳的肌肉可以在说话人自己开始说话之前就拉紧。在这种情况下，肌肉可以在说话人自己的声音到达内耳之前（图 12.2b 中的 A 点）被激活（图 12.2b 中的 C 点）。中耳肌肉的这种抑制机制使说话人可以对自己产出的声音形成自我保护。

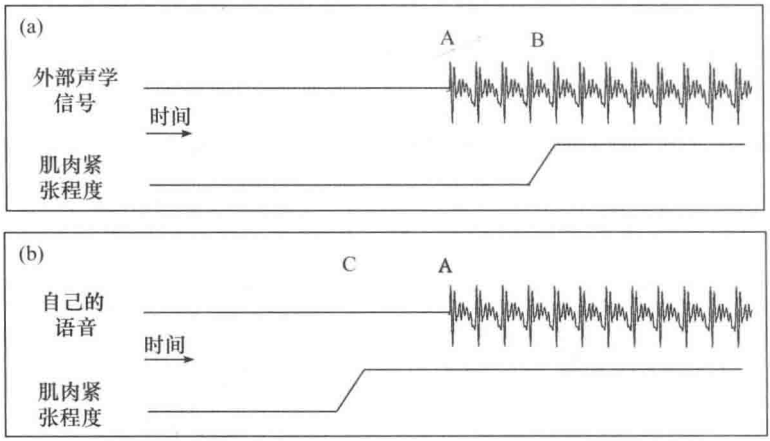


图 12.2　中耳肌肉主动减少音量的原理示意图

中耳肌肉的另一个功能是抑制声音信号的低频成分。这样，听话人可以把更多的注意力集中在高频，这对判断声源的位置非常重要。这个机制可以帮助听话人在嘈杂的环境中有效地辨认出某个人的声音。

12.2.3　鼓室中的压力均衡

中耳内的听小骨被封装在一个无菌的空间里，使其免受病菌感染。一种理想的方案是使中耳耳腔保持气密。但是，如果鼓室是完全气密的，随着每天大气压的变化，这会导致严重的问题。因此，中耳的鼓室通过**耳咽管（Eustachian tube）**与鼻咽相连（见图 12.1）。尽管耳咽管没有参与从外耳到内耳的声音传播过程，但是它使人耳可以适应大气压的缓慢

变化。

在7.1.1中，我们解释了声波与大气压变化的原理相同，只是声波的振幅更小而且比大气压的变化快得多。如果鼓室完全封闭，当天气发生变化时，由于大气压变化比声波的压力变化大得多，鼓膜就会被向内推或者向外拉。这就是大家都了解的“压耳”效应。例如在我们开车下山的时候，伴随着海拔的快速变化，气压会产生变化，我们便可以感受到这种效应。此时就是耳咽管发挥作用的时候。耳咽管的管壁通常是松弛地挨在一起的，封闭耳咽通道。但是，如果通道两端存在很大的压力差，空气就会通过耳咽管被压入或者压出鼓室。（打哈欠和吞咽的动作可以使通道主动打开。）最终，鼓室的气压可以与耳道内的大气压相适应，从而防止耳膜受损。声波的气压变化非常小，不会穿过耳咽管，但足以使鼓膜的形状发生改变。

中耳的听小骨可以在声波到达内耳的**卵圆窗（oval window）**的过程中增大声波的压力（见12.3节）。卵圆窗的功能非常明显：没有卵圆窗，镫骨的运动就无法传输到内耳内部的淋巴液中。

12.3 内耳

与听觉有关的内耳结构是**耳蜗（cochlea）**。在这个骨质的、螺旋状的结构中，声波被转换成神经脉冲，通过听觉神经传至大脑。

耳蜗的形状像蜗牛壳。**蜗底（base）**附着在中耳上。相比较而言，蜗底更厚更窄，**蜗顶（apex）**更薄更宽。这样的结构特点对其功能有着非常重要的意义，见12.3.5中的说明。如果我们可以将耳蜗展开（如图12.3），成人的耳蜗长约3.5 cm。耳蜗内有两条通道，上面的是**前庭阶（scala vestibuli）**，下面的是**鼓阶（scala tympani）**，中间以**基底膜（basilar membrane）**相隔，听觉神经就是从这里开始的。这两条“走廊”在**蜗孔（helicotrema）**内的蜗顶相连。**前庭膜（Reissner's membrane）**把前庭阶和**蜗管（scala media，也称“中阶”）**完全分开（见图12.4）。前庭阶和鼓阶充满了**外淋巴液（perilymph）**，外淋巴液中含有大量的钠（钠离子）；而蜗管内又充满了**内淋巴液（endolymph）**，内淋巴液体中含有大量的钾（钾离子）。前庭阶通向中耳鼓室壁里一个柔韧性很强的膜，

位于卵圆窗处，就在镫骨的下面。鼓阶通向一层柔韧性很强的膜，叫作**圆窗（round window）**，并延伸到鼓室（见图 12.3 和图 12.1）。

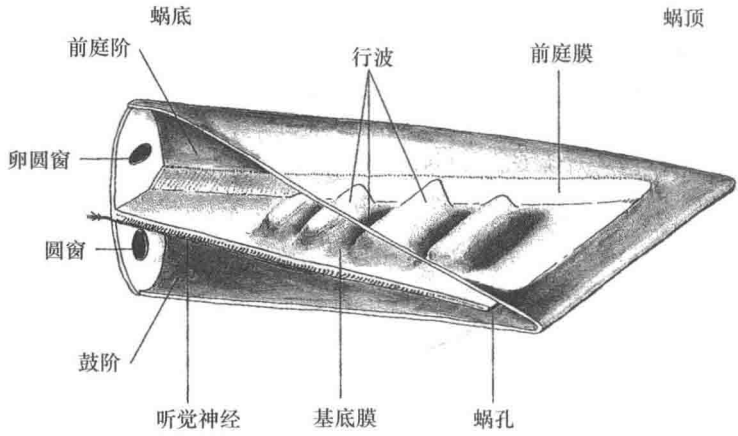


图 12.3　“展开”的耳蜗以及行波的示意图

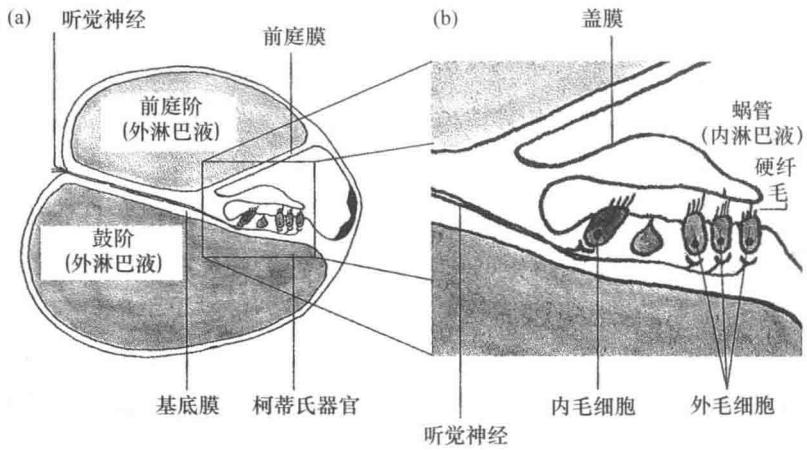


图 12.4　柯蒂氏器官：（a）整个耳蜗的横断面，（b）局部放大图。

在讲解基底膜结构和实际的听觉器官之前，下面将首先介绍当声波到达耳蜗各阶以及基底膜时会发生什么情况。

12.3.1 耳蜗中的压力波

声波的压力变化，通过位于中耳内的听小骨加强后，经由卵圆窗到达耳蜗。在耳蜗处，声波以纵压力波的形式在内耳液体中通过前庭阶由蜗底移向蜗顶，速度大约是1500 m/s（见7.1.5）。声波经由蜗孔继续传递，然后通过鼓阶返回耳蜗底。同时，压力波也在基底膜上传播。在蜗底，压力波通过圆窗的薄膜再一次进入中耳的鼓室。

圆窗的功能是把声波从耳蜗中“释放”出来。这是必须的，因为外淋巴液和内淋巴液实际上是不可压缩的。如果不释放声波，内耳的淋巴液就会变成僵硬的物质，这样即便卵圆窗具有较强的韧性，这种变化也会阻止卵圆窗对声波进行响应。因此，圆窗可以调节耳蜗内部的压力：卵圆窗的向内运动会导致圆窗的向外运动，反之亦然。

由前庭阶、蜗孔和鼓阶构成的通路只允许声学信号中较低的频率的相对较慢的压力变化通过。对于较高的频率及其相应的较快的压力变化，小的蜗孔就像一个不可穿透的障碍。这就像慢速移动和快速移动的人群要通过一扇狭窄的门的道理一样：移动慢的人群（像较低的频率一样）有足够的时间来穿过这道门，但是移动快的人群就会在狭窄入口处造成阻塞。但是压力变化从卵圆窗传至圆窗并不需要经过蜗孔，因为基底膜是一个有弹性的结构，它本身就会在压力波的影响下振动（见下文）。因此，基底膜可以绕过蜗孔直接把压力从前庭阶传至鼓阶。

卵圆窗和圆窗的存在对于中耳的另外一个功能有非常重要的意义。中耳的听小骨不仅可以在声波通往卵圆窗的过程中增大声波的压力（见12.2.1），同时它们还会把压力波不对称地传输至耳蜗，即只传输到前庭阶而不传输到鼓阶。如果没有听小骨，声波经过鼓膜之后会自由地在鼓室中的空气内传播，然后几乎同时到达卵圆窗和圆窗。这样卵圆窗的膜和圆窗的膜就会向内或者向外移动相同的距离。虽然这样可以增大或者减小耳蜗中的压力，但是压力波却无法沿着上面所讲到的路径进行传播了。并且由于前庭阶和鼓阶中的压力相互抵消，基底膜也就无法振动了。

一些由于听小骨缺失或失效而导致的失聪就是因为缺少12.2.1中讲到的压力调节机制。而导致这种失聪的另一个重要原因是耳蜗内卵圆窗和圆窗之间的压力传播不对称性的缺失。如果基底膜不能振动，内耳就不能产生听力所必需的神经脉冲，我们将在下面的内容中进行讲解。

12.3.2　作为振动体的基底膜

在耳蜗内，基底膜的振动对把压力波转变为神经脉冲的过程非常重要。想象一下，基底膜两侧都是相同的不可压缩的液体，而且压力几乎是同时出现在基底膜的上面和下面，所以基底膜就不能像麦克风上的膜一样上下振动。那么，控制基底膜振动的机制是什么呢?

12.3.3　共振理论

在19世纪，研究者发现基底膜由一些独立的条状体组成，它们通过组织连接在一起。在蜗底的条状体比在蜗顶的更硬一些。由此，Helmholtz（1863）提出这样的理论：基底膜像一架“倒置的钢琴”，不同的“琴弦”分别与压力波的不同频率产生共振。高频使基底膜的蜗底振动得更显著，因为基底膜的蜗底比蜗顶更硬更窄。相反，低频使蜗顶的振动更加强烈，因为顶部的基底膜更软更宽。只要通过测量基底膜振动的位置，神经细胞就能够判断声音信号的振幅和信号所含的频率。根据这种理论，人耳就像在进行一种机械化的傅里叶分析一样，频率是通过基底膜振动的位置来编码的，振幅是通过基底膜在该点的位移来决定的。

在20世纪初，Von Békésy（1928）尝试用实验的方法来验证Helmholtz的理论。他将显微镜镜片插入到死去的动物的耳蜗里，然后通过机械地移动卵圆窗来模拟声波。这种方法可以直接观察到基底膜的运动情况。由于这些运动非常细微，Von Békésy使用对应于响度级别在120至160 dB的高振幅来进行测量，这使得基底膜可以产生大约0.5 μm的振动，这样才刚刚可以被显微镜观察到。他发现，正如Helmholtz所预测的，基底膜确实在不同的频率下振动位置不一样：高频时蜗底附近振动，低频时蜗顶附近振动。但是，他并没有观察到基底膜的条状体像共振理论预测的那样单独振动，而是观察到了**行波（traveling wave）**。这种波与抖动桌子上的桌布时出现的波类似（抓住桌布的一端，然后快速地上下抖动）。这种振动在基底膜的整个组织上传递，在某个频率（称作**临界频率 critical frequency**）的对应位置达到位移极值，然后又逐渐消失（见图12.3）。

这些位移的极大值与基底膜各部分的弹性程度的差异有关。我们将

在 12.3.4 中来解释这一点，并在 12.3.5 中讨论行波是如何产生的。

12.3.4 对共振理论的反驳

如9.2 节讲的一样，固体不仅有自己的共振频率，而且还有一定程度的阻尼。例如红酒杯这样的物体，能使酒杯达到共振的频率范围很小，因此酒杯的阻尼作用也很小，由共振产生的振动可以维持很长时间。另一方面，对于阻尼很大的物体，例如叉子，能够使它达到共振的频率范围较大，而由共振产生的振动也不会延续太长的时间。基底膜的阻尼程度很高：一旦声音消失，基底膜就停止振动，我们对声音的感知也随之消失。可是，如果阻尼程度很高，共振频率的准确程度就会相应地降低（即很大范围内的频率都可以使这个物体产生振动）。这和人耳听觉系统的能力并不匹配，因为在最佳条件下，人耳可以区分小于 1 Hz 的频率差异。要感知如此微小的差异，基底膜必须对频率有很强的选择性。这意味着，基底膜需要有很小的阻尼，并且可以长时间地保持振动。但我们都知道，事实并不是这样的。这表明基底膜的频率选择特性不是由共振引起的。

12.3.5 行波理论

基底膜的运动不是由通过液体快速传播的压力波（如 12.3.1 中的描述）决定的，而是由横向传播的“表面波”（与人浪波类似，见 7.1.5）决定的。这种波与水面上的波相似。想象一个湖，在完全没有风的时候，像镜子一样的湖水表面就如同完全安静状态下的基底膜。当空气从气压比较高的区域向气压比较低的区域移动时，风便产生了。空气压力的这种微小的增加更早到达离风源比较近的水分子。这些在较高压力下的分子会向下移动，以避开这种压力，而处于原来的较低压力下的分子会向上移动。这些压力变化最终在水面上形成了表面波，然后开始以横波的形式传播，比水表面以下的纵压力波要慢得多。与之前一直讨论的压力波不一样，表面波的速度和振幅取决于信号的频率和水域的深度。

我们可以在海滩上观察到这种效应。离海滩较远的海面的波浪没有近处的明显。当波浪接近海滩时，它们会变得更高（直到最终瓦解）。在

开阔的海域，海水很深，水分子不会受到海底的阻碍，因此它们可以相对自由地上下运动。但当波浪接近海滩时，由于水分子不能完全自由地上下运动，波浪的速度就会减慢。波浪以比较慢的速度传播，这意味着波本身会变得比较短。然而，尽管在浅水区随波浪运动的水分子会减少，但是这些波浪的能量依然保持不变。因此，水分子的运动距离会变得更长，所以波浪就会变得更高。

在蜗顶附近，耳蜗变得比较薄，使得液体深度变浅，这与海滩处海水的情况类似。这与行波（表面波）所产生的效应类似。此外，基底膜在蜗顶附近更加柔软，这使声波的速度变得更慢，从而导致了更大的位移量。由于耳蜗“深度”的不同和基底膜“柔软性”的变化，声波会到达基底膜上的不同位置。高频使位于卵圆窗后的蜗底产生较大的位移；而低频会向着蜗顶的方向继续移动，根据自身频率的不同在传播路径上的不同位置达到位移极值。

按照这种方式，基底膜像一个机械的傅里叶转换器一样工作，但这并不是由共振原理得出的结论，这与 Helmholtz 的观点有所不同。不过，需要注意的是，在有复合信号或者噪声的情况下，例如语音信号，每个行波的频率分量会相互重叠，此时基底膜的运动是非常复杂且不可预测的。

行波会影响整个基底膜，从蜗底一直到蜗顶。由于基底膜的两侧都附着在耳蜗上，从蜗底到蜗顶的纵向移动会引起基底膜的横向运动。声波沿着基底膜的中部运动，引起基底膜中部的组织上下移动。而且，由于基底膜的两侧是固定在耳蜗壁上的，基底膜会在宽度范围内发生扭曲。在下一节中我们会介绍，基底膜的这种扭曲运动是耳蜗神经最终产生神经脉冲的原因。

以上所陈述的这些事实解释了在基底膜上存在行波的原因。但是，人们所观察到的行波的振幅却要比只考虑基底膜本身机械特性得到的振幅高得多。而且随着声音响度的增加，行波振幅增加的程度与声波本身振幅的增加程度并不一致。这表明基底膜的运动并不只是一种像被抖动的桌布那样的被动振动，而确切地讲应该是一个主动运动过程的结果。只有在彻底了解基底膜的结构以后，我们才能理解这样的主动运动过程以及将运动转换为实际的神经脉冲的过程。这些内容将在下一节中进行介绍。

12.4 基底膜的结构

基底膜包括**柯蒂氏器官（organ of Corti）**，它与听觉神经相连（见图12.4）。柯蒂氏器官由**盖膜（tectorial membrane）**、**毛细胞（hair cell）**和其他大量的支持细胞构成。与基底膜不同，盖膜没有覆盖耳蜗的整个宽度，只有部分与基底膜重叠。盖膜的一边与基底膜相连，距离前庭膜的连接点很近。另一边由约20000个**外毛细胞（outer hair cell）**支撑着，这些外毛细胞沿着基底膜组成三束（见12.4.1）。这些外毛细胞的底端依附在基底膜上，它们的顶端是水晶“毛”束，也就是**硬纤毛（stereocilia）**，与盖膜相连。毛细胞的主体部分几乎完全被柯蒂氏器官里与鼓阶相连的外淋巴液包围，只有纤毛本身被蜗管中的内淋巴液包围。

12.4.1 外毛细胞

在蜗底处的外毛细胞大约有25 μm长，而蜗顶的外毛细胞大约是蜗底的4倍长，并且比蜗底的更为灵活。除了基底膜弹性上的差异（该差异会使不同频率在基底膜上产生不同的位移）以外，外毛细胞增加了另一个区分不同频率的机械源。在蜗底较硬较短的毛细胞对高频更加敏感，而在蜗顶的更长更灵活的毛细胞在低频下更易弯曲。除了这个机械特点以外，外毛细胞的电特性似乎在蜗底部分对高频的反应更灵敏，而在蜗顶部分对低频的反应更灵敏。

每当基底膜在行波的作用下产生运动的时候，基底膜也会同时发生扭曲，就像12.3.5中讲到的一样。这种运动会使盖膜与基底膜发生碰触，进而使外毛细胞的硬纤毛发生偏转，并使小离子通道得以打开，从而引出硬纤毛所在的蜗管里的内淋巴液，使其到达处在外淋巴液中的毛细胞体内。这一过程会引起放电，从而使外毛细胞收缩，促使盖膜进一步向下运动，使柯蒂氏器官中内淋巴液的流动增强。外毛细胞的这种放大效应可以通过大约1800个**传出（efferent）**神经纤维来调节，这些纤维从大脑传输神经脉冲[2]。外毛细胞很像由大脑控制的一些小肌肉，它们可以改变基底膜的弹性以及它们对偏转运动的反应强度。也就是说，外毛细胞就像一个调节器，它们可以在大脑的控制下主动地改变基底膜的弹性

特点。外毛细胞只有少数几个去向大脑的神经通路，但是这些通路只包含反应非常慢的细胞，而且人们一般认为它们不会将声音信号的任何细节信息传入大脑。

外毛细胞可以增加柯蒂氏器官中淋巴液的流动，但是它们也可以抑制基底膜的运动，也就是减小基底膜的振动。基底膜振动幅度的增加程度与声波振幅的增加程度有所不同。这种主动的机制，与中耳听小骨上的肌肉控制机制相比，可以更快更精确地抑制某个特定的频率区域。将声音信息传给大脑的过程实际上是由内毛细胞完成的。

12.4.2　内毛细胞

内毛细胞（inner hair cell）的作用非常重要。虽然内毛细胞的数量不超过3500个，但是内毛细胞却负责将基底膜的机械运动转换为神经脉冲。内毛细胞在柯蒂氏器官中组合成一束。与外毛细胞不同，内毛细胞不会触碰盖膜。事实上，它们会对柯蒂氏器官中内淋巴液流动速度的变化作出反应。大约有30000个传入神经纤维由内毛细胞伸入大脑。也就是说每一个内毛细胞都同大约10个神经细胞相连。

内毛细胞的主细胞体上带有伸出基底膜之外的很细的像茎一样的硬纤毛，就像附着在根茎顶部的细毛一样。硬纤毛的顶端与细小的纤维相连，这些纤维弯向一侧的时候，会在硬纤毛中打开细小的“陷阱”，进而使硬纤毛可以与毛细胞体内的外淋巴液相接触。

每当基底膜在行波的影响下产生运动的时候，基底膜也会同时发生扭曲，就像12.3.5中讲到的一样。基底膜的位置和机械特性可以使柯蒂氏器官里的内淋巴液中产生运动。而外毛细胞可以加强内淋巴液的这种运动，进而使内毛细胞的硬纤毛前后来回运动。内淋巴液向上靠近基底膜的运动，会在包含了硬纤毛的外淋巴液与内淋巴液中的毛细胞之间打开小离子通道。因此，神经细胞中钠、钾离子的浓度突然发生改变，进而引起了它们所带电荷的突然改变。这实际上就是神经脉冲，被称作神经细胞的“激发”。与外毛细胞的情况不同，这种放电是通过不同类型的神经元传输到大脑的，并最终产生被我们称为“听觉”的输入。

当硬纤毛向相反的方向运动时，离子通道就关闭，使细胞可以自行

充电。毛细胞总是在基底膜达到位移极值时激发，而基底膜的位移与声波是同相位的[3]。因此，毛细胞的激发与声波是同相位的。

要了解内毛细胞的机制，就必须深入了解神经细胞是怎样传递信息的。在休息状态时，神经细胞带有轻微电荷，就像小号电池一样。神经脉冲是神经细胞的一次突然的完全的放电。对内毛细胞来说，这是由小离子通道引起的。在硬纤毛顶端附近打开的小离子通道会使细胞电荷与蜗管里的内淋巴液电荷之间产生短路，进而导致放电。在放电之后，细胞会逐渐恢复电荷，在一定时间后，它们便又可以再次激发了。这段时间间隔，称作**不应期（refractory period）**。不应期的长短取决于神经的类型，但是决不会短于 1 ms。这些细胞只处于两种状态，即“满”或者“空”，也就是说，它们要么是处于休息状态，要么处于最大激发幅度的状态。换句话说，它们要么“开”要么“关”，没有中间状态。而知觉的强弱印象是由神经细胞激发的速率而不是激发的强度所编码决定的。

12.4.3 基底膜上的频率编码

由此看来，内毛细胞所提供的频率信息是通过这些信息在基底膜上的**音频响应（tonotopic）**位置进行编码的（蜗底处的位移对应高频；蜗顶处的位移对应低频）。与此同时，神经细胞的激发速率也对频率信息进行编码。根据基底膜达到位移极值的区域的不同，神经细胞会以不同的频率激发。从这种意义上来说，基底膜的复杂的主动控制运动可以被描述为一种机械的傅里叶变换。而且，随着年龄的增加，人们对频率的感知会发生改变，即便原先对音高很敏感的人也是如此[4]。例如人们不再把 440 Hz 感知为基准音 a^1，而是将其感知为一个较低的调。这种改变可能是由随着年龄的增长，基底膜硬度的增加造成的，因此同样的频率便会在基底膜的另一区域引起最大位移。

另一种理论认为，声音信号的音高信息是通过神经细胞的激发速率（而不是通过它们在基底膜上的位置）来传播的。神经细胞的激发与声波的极大值同步，因此神经脉冲可以反映一个周期的时长。这一点可以用来说明音高（或者谱分量的频率）不是根据基底膜的不同位置来进行编码的，而是根据神经细胞不同的激发速率来编码的。关于这种理论，一个明显的反证是神经细胞激发的速率无法超过每秒一千次，这意味着人

们无法感知到任何超过1000 Hz 的频率。这个反驳确实是有效的，但是这只考虑了一个神经细胞的情况。然而，基底膜上几乎全部的毛细胞都能被所有的频率刺激，这样若干细胞可以对相同的频率同时激发[5]，当一个神经细胞因为需要时间恢复电荷而不能足够快地激发时，另一个细胞会在声波极大值处同时激发。考虑所有神经细胞的总体模式，可以说神经脉冲确实以和声音信号相同的周期进入大脑。也就是说，周期性在许多神经细胞的总体激发模式中得到了保留。

单个细胞的激发与声波极大值的对齐，以及由不同神经细胞构成的总体的同步模式存在于大约5 kHz 以下的频率中。这与人类只能感知最高约5 kHz 的旋律的事实是一致的；在5 kHz 以上，人们仍然能感知音调，但是无法再把不同的频率排列到音调标度中去了。这一点也佐证了声音信号的音高信息是通过神经细胞的激发速率（而非它们在基底膜上的位置）来传输的这一理论。

这种理论的支持者认为基底膜的振动与出现在正常言语中的声波压力水平没有对应关系，所以也就无法对位移进行明确的区分。这表明频率信息不可能通过基底膜的位移极值来编码，而只能通过神经细胞的激发周期来编码。

从本质上讲，内耳把宽带输入信号（即声波，包含30 Hz 至18000 Hz 之间的信息）编码成大量的窄带神经通路。把声波转换成很多并行的神经通路有利于使系统具有一定的冗余度，这样单个神经细胞的失效就不会导致信息传输的整体性失败。

听觉是一个复杂的过程，它把声音信号转换为神经脉冲。即使借助现在的先进技术，对神经信息的研究依然是一项非常复杂的工作，因为听觉末梢的神经位于大脑的深处，并且只有在活生命体中才能对其进行研究。

12.4.4　耳声发射

一般认为，神经首先把声音刺激（投射到耳朵的脉冲）传输到大脑，这个传输过程在内耳中进行；然后大脑才告诉外毛细胞做出反应。这样，会使基底膜产生**耳声发射（oto-acoustic emission）**，即外毛细胞的变化会引起基底膜某种特定的运动（就像桌布突然被拉紧时产生的运动）。这种信号最早是由 Kemp（1978）测量到的，他把很短的信号脉冲投射到人耳

里，在一些延迟之后，人耳产生了这些脉冲的回声。有趣的是，这种回声的出现时间比预期的要晚。预期时间是根据声音从外耳传至内耳，再从内耳传回需要经过的距离计算出来的（即穿过耳道，在被鼓膜和中耳的听小骨传输后，经过卵圆窗、耳蜗，再通过圆窗，经过中耳的空气到达鼓膜，最后通过耳道返回）。而且，回声的振幅比预期的纯粹缘于机械反射的情况要高。这种回声是人耳自身反应的结果，发生在声音感知的几毫秒之后。几年之后，人们又测量到了自发的耳声发射。这意味着即使在没有声音刺激的情况下，基底膜也可能自发地运动。从某种意义上来说，这种效应是外毛细胞的一种“痉挛”。这种自发的耳声发射进一步证明了基底膜是可以进行主动控制的，并不只是在声波的刺激下才做出被动的反应。因此，基底膜是部分主动的，所以内耳的作用也便不再只是简单地以被动的方式转换声音刺激了[6]。

这也是反驳把基底膜认为是傅里叶变换器的另一个证据，因为基底膜的运动很明显是由从大脑传来的反馈过程控制的。

总的来说，前面的内容很清楚地表明，听觉系统对声学信号进行了许多转换，这使通过听觉系统神经的激发模式得到的对信号的表征会与最初的声学输入信号有很多不同。因此区分**物理刺激（physical stimulus）**和**听觉感知（auditory sensation）**是非常重要的。到目前为止，我们讨论了声波的物理特性：音长、频率和振幅。但是，音长、频率和振幅的物理性质与听感上的时长、音高和响度并不一致。因此，人们使用**听觉标度（auditory scale）**来表示这些听感的程度。这些标度不同于物理意义上的“维度”：两个维度是彼此独立的，但是主观的感觉会相互影响。例如，一个“更响”的声音可能听起来音高也“更高”，尽管发生变化的是信号的振幅而不是频率。

在下面的小节里，我们将讲解一些听觉标度。这些标度可以比线性标度更好地反映听觉系统的行为。但是必须注意的是，这些标度仍然只是我们对听觉系统的时间相关且复杂的表征的一个粗糙的近似的反映。

12.5 听觉频率标度

当在图上展示一个信号或者一次分析的结果时，所得的图像不仅取

决于信号本身而且还取决于图的标度。我们可以用许多不同的标度来表示频率和振幅，对于谱和语调曲线的表征也是如此。乍看起来，我们可能会觉得用不同的标度来表示一个信号的频率很奇怪，因为频率可以用很清楚的物理术语来定义。事实上，采用不同标度的原因并不是出于对严格的物理维度的描写，而是出于对人类感知频率的心理基础的考量。不同的频率标度用来表示不同的感知特性。

12.5.1　线性标度

到目前为止，我们是用**线性（linear）**标度来表示频率值的。用这种标度，100 Hz 和 103 Hz 之间的距离与 500 Hz 和 503 Hz 之间的距离是相等的。也就是说，不管频率所处的频率范围如何，相同的频率差总是用纸上相同的距离来表示。这是符合我们的常识的，因为人们可以同样好地感知到 100 Hz 和 103 Hz 之间的差异以及 500 Hz 和 503 Hz 之间的差异。与响度或重量的感知不同（见附录 A.3.2），人们可以很好地感知到 1000 Hz 以下的绝对差异。在线性标度上的两个频率之间的距离完全是由这两个频率值之间的差异所决定的。相同的频率差异在线性标度上的距离也是一样的，与频率所处的频域无关。

12.5.2　对数标度

现在来考虑音高的感知。一个 200 Hz 的音调被感知为是一个 100 Hz 音调音高的两倍。这个比例在音乐里被认为是一个八度音（一个倍频程）。如果想让感知到的音调提高一个八度，那么频率就必须得再次翻倍，也就是从 200 Hz 变为 400 Hz。注意并不是从 200 Hz 到 300 Hz 这样增加 100 Hz。所以，以 100 Hz 为基准，要得到一个高两个八度的音，必须将频率值乘以 4。

要使感知的音高以因子 2 增长，相对应的物理频率则需要以因子 4 来增长，这种关系可以用一个**对数（logarithmic）**关系来表示。用对数表示（音乐）音高的频率域，则两个八度之间的感知距离总是相同的，而其物理距离不同。在这种表征方式下，220 Hz 和 440 Hz 之间的距离与 1760 Hz 和 3520 Hz 之间的距离是相等的，即都增加了一个八度，尽管前

者两频率值之间相差 220 Hz，后者两频率值之间相差 1760 Hz。

在西方音乐中，一个八度被分为 12 个**半音（semitone）**，计算两个频率之间的距离时，通常是基于半音这种对数标度。由于八度之间表示 2 的倍数关系，所以计算使用的是以 2 为底数的对数：

$$\text{以 semitone 为单位的距离} = 12 \times \log_2\left(\frac{\text{频率“X”[Hz]}}{\text{频率“Y”[Hz]}}\right)[\text{semitone}]$$

大多数便携式计算器都没有以 2 为底数的对数（$\log_2$）的按键，只有以 10 为底数（log 或者 $\log_{10}$）或者以 e 为底数的自然对数（ln 或者 $\log_e$）。所以下面的转换公式也许有用：

$$12 \times \log_2\left(\frac{\text{测量频率[Hz]}}{\text{参考频率[Hz]}}\right) \approx 39.863 \times \log_{10}\left(\frac{\text{测量频率[Hz]}}{\text{参考频率[Hz]}}\right)$$

$$\approx 17.312 \times \ln\left(\frac{\text{测量频率[Hz]}}{\text{参考频率[Hz]}}\right)$$

例如，标准的参考基音 a¹（音乐会音高）的频率是 440 Hz，中音 C 的频率是 261.6 Hz，它们之间的距离用半音表示如下：

$$39.863 \times \log_{10}\left(\frac{261.6[\text{Hz}]}{440[\text{Hz}]}\right) \approx -9[\text{st}]$$

对数标度有一个共同的主要特点：物理值之间的比值相同，则在标度表示上，它们之间的距离也相同，这与频率在标度的哪个区域无关。对数标度很常用，因为在不同的领域，例如音高或者响度，我们的主观感知是建立在对不同值之间相对关系的感知之上的，而不是建立在不同值之间的绝对差异之上的。

12.5.3 美标度

早期关于频率差异的感知研究让 Stevens 和 Volkman（1940）引入了“悦耳的”**美标度（mel scale）**。这个标度是通过将人的感知与 1000 Hz 的正弦音相比较来对正弦音进行分类的。为了实现这个目的，需要被试指出什么样的测试正弦音是参考音的两倍或者一半。被感知为只有一半高的音被设定为 500 mel，被感知为两倍高的音被设定为 2000 mel。通过进一步比较，就可以获得对美标度的精确划分。与线性或者对数标度不同，美标度是完全基于人的感知的。一美被定义为 1 kHz 正弦音感知音高的千分之一。原则上讲，这个标度只能通过汇总每个听音人的印象来获

得，而不能只通过对物理频率（以赫兹表示）的计算来获得。不过，到目前为止所发现的美的值可以通过下面的经验公式来近似计算[7]：

$$1\ \text{mel} = \frac{1000}{\log_{10}(2)} \times \log_{10}\left(\frac{f[\text{Hz}]}{1000} + 1\right) \approx 3322 \times (\log_{10}(1000 + f[\text{Hz}]) - 3)$$

例如，音乐会音高 a¹，频率值为 440 Hz，对应音高为 526 mel：

$$3322 \times (\log_{10}(1000 + 440) - 3) \approx 526\ \text{mel}$$

尽管美标度只是基于人们对正弦音之间的感知关系，是一种经过精细调整的对数标度，而且美标度不能反映听觉对更复杂信号的感知，但是美标度已被广泛应用于自动语音识别领域中。

12.5.4　巴克标度

和美标度一样，**巴克标度（Bark scale）**也是基于人对音高的感知。更确切地说，这个标度是基于由 Fletcher（1940）提出的**临界频带（critical bands）**理论。与完全基于对正弦音的主观感知的美标度不同，巴克标度是基于对更复杂信号的感知。巴克标度考虑到：两个频率相近且同时发生的信号，如果它们的频率比较低，那么它们会被感知为两个不同的音；但如果它们的频率比较高，那么它们就会被感知为同一个音。这种现象可以通过临界频带理论来描述。这个理论认为，在人的音高感知中，频率范围被划分为不同的滤波频带，低频区域带宽窄一些（区分更细一些），高频区域带宽宽一些。在一个临界频带内的所有能量都会被投射到这个频带的中心；也就是说，一定频率范围内的声音能量被整合成了一个感知量。需要注意的是临界频带并不是都包含相同的频率范围。感知低频的一个临界频带频率范围比感知高频的一个临界频带频率范围要窄得多。

因为这些特点，人们认为巴克标度比美标度更准确地反映了人的音高感知。音高感知值（以巴克表示）和频率（以赫兹表示）之间的关系可以大致用下面的公式来表示：

$$1\ \text{Bark}_{\text{CB}} = \frac{26.81 \times f[\text{Hz}]}{1960 + f[\text{Hz}]} - 0.53$$

例如音乐会音高 a¹，频率是 440 Hz，用巴克标度表示，其音高大约是 4.4 Bark_{CB}。

$$\frac{26.81 \times 440[\text{Hz}]}{1960 + 440[\text{Hz}]} - 0.53 \approx 4.385\ \text{Bark}_{\text{CB}}$$

如果要将一个完整的谱（例如 FFT 的结果）转换为巴克值，第一步就是将（线性的）频率值划分成不同的临界频带。这些临界频带就成为带通滤波器，它们的带宽和陡度取决于这些频带的中心频率和声压级（以 dB_{SPL} 表示）。

要将线性的傅里叶谱转换成巴克谱，需要针对给定的某个中心频率和振幅，来确定其对应的临界频带的特定形状和宽度。我们需要把这个频带内的所有振幅值代入计算，最终得到这个特定中心频率的唯一的振幅值。最后将这个振幅值赋给由上述公式计算所得的巴克频率值。这个过程包括两个部分：第一，将临界频带的窗口在谱上移动，整合每个临界频带内的能量；第二，沿着频率标度，转换频率之间的距离。不过在使用巴克标度时，通常只进行第二个步骤。但是，这里需要明确，若以这种方式使用巴克标度，其方法与上面提到的完整的过程是有区别的。如果只使用上面提到的公式把频率值转换为巴克值，那么会产生下面的效果：在线性赫兹标度上相互分立的值（见图 12.5a）在巴克标度上也是分立的（见图 12.5b）。但是如果将频率范围划分为一系列临界频带，那么这两个在赫兹标度上相互分立的值在巴克标度上就可能会合并为一个值（见图 12.5c）。这与感知数据是一致的，因为人类也是无法区分这些频率的。但是，由于信号的振幅会影响需要合并的具体频率，完整的转换过程是相当复杂的。

12.5.5 等效矩形带宽（ERB）标度

等效矩形带宽标度（equivalent rectangular bandwidth scale，ERB scale）和巴克标度非常相似，但是决定临界频带的方法更加细化，这样可以更好地体现复杂声音的本质特点（Patterson，1976）。这个标度应该是可以最真实地表示频率感知情况的方法了。与巴克标度一样，它也包括两个部分，估计临界频带以及将频率映射到 ERB 标度上。下面的经验公式可以把赫兹转换为 ERB 值（此处不考虑临界频带内的能量分布）：

$$1\ \text{Bark}_{\text{ERB}} = 25.72 \times \log_{10}\left(\frac{312 + f[\text{Hz}]}{14675 + f[\text{Hz}]}\right) + 43;\ \text{当}\ 100\ \text{Hz} \leqslant f \leqslant 6500\ \text{Hz}$$

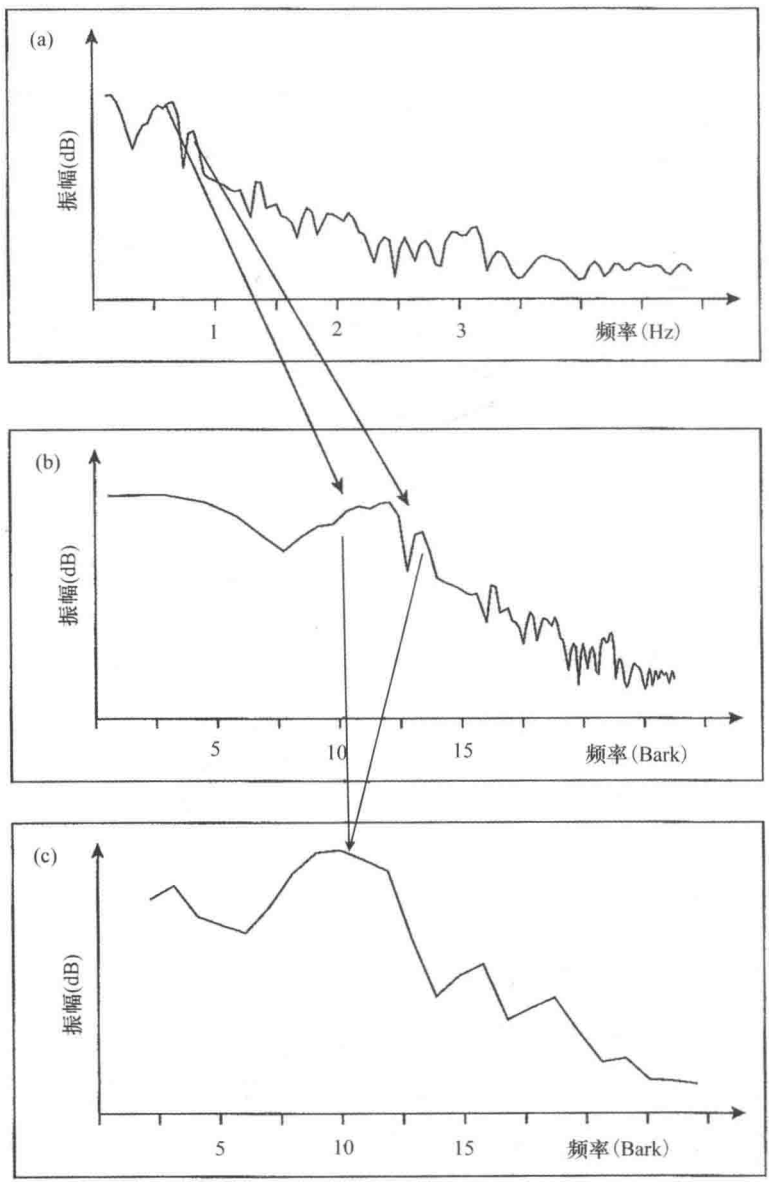


图 12.5　将线性谱（a）被转换成巴克标度。（b）只转换频率，（c）同时考虑临界滤波器。箭头表示将两个分立的频谱峰由（a）映射到（b），并在（c）中合并为一个。

根据这个公式，音乐会音高 a¹（频率为 440 Hz）相当于 9.5 $Bark_{ERB}$：

$$25.72 \times \log_{10}\left(\frac{312 + 440[\text{Hz}]}{14675 + 440[\text{Hz}]}\right) + 43 \approx 9.5\ \text{Bark}_{ERB}$$

巴克和 ERB 标度只在临界频带的形状和频率标度的转换上存在细微的差异。

使用这么多标度似乎有些令人困惑，但其实每一种标度都有其存在的原因。线性标度展示了人可以感知到相同的频率差异（至少在 1 kHz 以下如此）。对数标度展示了人可以将相同的半音差异感知为音乐上相同的音程。美、巴克和 ERB 标度反映了人对频率值的主观感知。但是必须注意的是，巴克和 ERB 标度不仅包括了对赫兹值的非线性转换，还包含了临界频带的概念。

听觉标度的目的在于能更好地表示人对声音信号的感知。Mp3 播放器的原理实际上就是基于听觉感知实验的。这些播放器不是以线性的方式来储存数字声音的，它们会计算谱信息然后只储存人能够听到的那些部分，这也是基于临界频带的概念。因此，原始的信号和储存在这种播放器中的信号是不同的，但是对大多数人来说，它们听起来是一样的。

在语音学上，人们依然在使用线性（语谱图的或者谱的）表示方法，因为人们对它们很熟悉而且线性表示是一种可靠的可视化的方法。但是，使用听觉标度则可能会改变我们对（言语）信号表征的模型，也会改变我们对语音特征的假设。遗憾的是现在我们还不知道哪种“听觉”谱可以更好地表示言语信号。考虑到在内耳中发生的复杂的转换以及之后大脑对声音的加工，我们看待语谱图对语音的表示方式可能与我们听觉系统感知语音或任何其他声学信号的方式截然不同。

12.6 听觉响度标度

对于响度来说，dB_{SPL} 标度（见 7.3.2 和附录 A.3.2）把所有频率都视为是等强的，也就是说它是基于线性的频率标度。但是，心理物理实验证明，人们对很低或很高的频率的感知没有对 2 ~ 4 kHz 范围内的频率的感知好（低于这一范围的频率多数都被中耳抑制，而高于这一范围的频率又会受到耳道共振特性的衰减）。而且对不同频率范围的减弱程度取

决于声音整体的强度，进而对不同强度的声音构成不同的曲线（见图12.6）。考虑到这一效应的标度是**方（phon）**标度。方的定义是与一个1 kHz的正弦波相同的声压级（dB_{SPL}）。其他频率的响度与在1 kHz处感知到的响度有关。对于每个强度，这个方法都会产生一个**等响度标度（equal loudness scale）**。把几个被试的感知结果平均之后，就得到了常用的dB_A标度。dB_A标度考虑了上述的效应。响度为40方的曲线是一条特殊的曲线，被称为一**宋（sone）**。一宋是一个参考值，它的响度大约和在1米的距离处正常谈话的响度相当。宋值每翻一倍，就等于增加10方，2宋等于50方，4宋等于60方，8宋等于70方，以此类推。

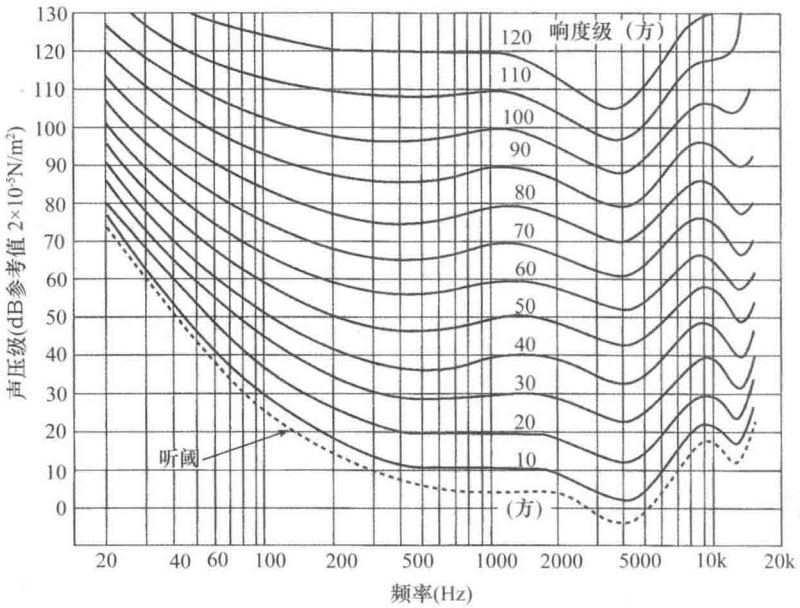


图12.6　根据ISO 226：2003标准，不同响度水平的频率敏感度（该图的使用经过ISO允许）[8]。

12.7　听觉时间标度

许多语言都有长短元音和/或长短辅音之分，或者持阻时间或嗓

音起始时间（VOT）的长短之分。但是时间长短的感知取决于许多因素：语言本身的特点、语音在单词中的位置（音素在单词末尾或者短语末尾会被延长）、单词的长度（较长单词的音段较短）、语速或者语境。单念的单词通常比在短语环境中的长很多，但是大多数听音人却感觉不到这种差异。即使单念的音长比短语中该单词音长的两倍还长，人们也通常将它们感知为是等长的。正如 11.3 节中提到的那样，“语速”也是很难定义的。听觉所感知到的“长”或“短”通常被表示为用毫秒度量的一段时间范围（例如，德语中的一个较长的闭合段通常大于 40 ms，一个较短的闭合段要小于 20 ms）。尽管绝对的数值不尽相同，世界上存在长短对立的语言非常常见。音长上的对立可能会因为语言的不同而有所差异，因为有的语言会使用其他附加手段来区分语音的“长”与“短”（例如，英语的“长”元音通常是二合元音化的；而德语中非低元音的长短对立是通过谱特性体现的）。总的来说，各种语言似乎不太依赖绝对的音长差来区分对立，相反，人们依赖于时长的相对差异。

练　习

1. 我们已经对语音进行了发音方面的描述，为什么还需要从听觉或者感知的角度对语音进行描述？
2. 中耳的作用是什么？如果没有中耳，为什么声音从外耳传输到内耳的效率会降低？
3. 为什么听小骨、卵圆窗和圆窗对内耳中的声音通路非常重要？
4. 共振理论或者行波理论是否能更好地解释基底膜的振动？请给出理由。
5. 关于音高传递原理的理论有哪些？请给出每种理论的支持证据和反驳证据。
6. 美（mel）、巴克（Bark）和 ERB 标度与线性和对数标度有什么不同？它们相互之间又有什么不同？

注　释

1. $f=\frac{c}{4l}$: $\frac{340\ \text{m/s}}{4\times 0.025\ \text{m}}=3400\ \text{Hz}$。
2. 神经像“单行道”一样，要么是将电脉冲从大脑里传出（传出神经纤维），要么是将电脉冲传入大脑（传入神经纤维）。
3. 但是，这并不表示毛细胞是在声压波的极大值处激发。因为声压波在基底膜上被转换为行波，所以它不一定与声波具有相同的相位。
4. 完美的音高感知是直接根据感知到的音调来决定音符的能力。例如，在听到 440 Hz 的音调时，一个具有完美音高感知能力的人就能判断它是基准音 a¹。这种能力可以在 3 至 6 岁之间通过音乐训练来培养。
5. 尽管根据不同的频率分量，行波的极大值会出现在基底膜上的不同位置，但是每个频率都会刺激基底膜上很大的一片区域。
6. 医院用耳声发射来检查婴儿的听力。婴儿的耳内如果存在耳声发射的回声，则表明中耳、内耳以及周围神经加工系统都可以正常工作。如果没有这种回声，就说明其中某个部位存在问题，需要进一步检查。通过这样一个两分钟的、完全无害的测试，我们就可以评估只有几天大的婴儿的听力是否正常。
7. 经验公式表示了两类值之间的关系，但并不会声明这两类值之间的函数关系。原则上讲，我们也可以通过查表来得到其对应值。经验公式只是代替了表格，使计算更加方便。
8. 译者注：该图也被称作等响度曲线。

13 言语感知

前面章节介绍了发音姿态可以产生对应的声学特征，因而由发音器官产生的声音可以用频率、音强和时长来进行描述。我们还介绍了听觉系统的结构以及其对输入信号的各种转换。但是，要了解言语是如何被感知的，除了上述这些因素以外，我们还必须考虑语言学因素。语言学因素对我们理解口语信息至关重要，其中包括词汇、句法、语义和语用等信息。言语感知的研究主要是要解释听音人分析声学模式的方式以及听觉系统是怎样解读声学信息从而识别说话人想要传递的信息的。在这一章里，我们主要关注对感知单个语音至关重要的声学特性和听觉约束的种类。可以说，大多数关于言语感知的研究都已经在单个音段或者音节的层面讨论过上述问题。直到最近，研究者们才开始对词或者短语这样更大的单元进行感知研究，进而考虑意义和句法范畴的作用。

我们可以将言语感知看作是一个把连续声学信号转换为离散单位序列的过程。多年来，研究者们提出了很多可能的候选单位，包括特征（被各种语言用来以不同方式进行组合并构成音位系统的语音特征）、音段、音节和单词。然而对于什么是言语感知的基本单位这一问题，人们并无定论。其实，很有可能是在不同层面的分析中，这几种候选单位都分别起到了关键的作用。暂且不管这些语言学单位确切的本质是什么，人们通常认为它们是在时间上先后出现的。但是，我们却很少能在声学信号中观察到这种**线性**（**linearity**）特点。这种线性的问题是指人们很难根据感知到的单个语音单位来对语音信号进行切分。尽管言语产出的过程在一定程度上很可能是一种规划好的离散单位（或者目标）的序列，但该序列在生理上却实现为发音器官的近乎连续的运动。因此，人们对声学事件的感知，例如对有序的音位序列的感知，并不总是线性有序的。

虽然线性特征的缺乏为言语感知理论带来了挑战，研究者们依然就在感知单个语音或者某类语音时所涉及的声学特征的本质逐渐达成了一致。我们将首先讨论元音的感知线索，然后将从发音方法的角度对辅音的感知线索进行综述。需要说明的是，关于言语感知的文献数量繁多，如果深入全面地综述，足以编成一本手册，因此本章将着重讨论一些主要且成熟的发现。

13.1　元音

元音属于感知特征最为显著的语音之一。元音通常是浊音，并且时长较长；由于发音时发音器官的收紧程度相对较小，元音相对比较响亮。元音音质最重要的声学线索是共振峰频率的位置。我们在 9.5.1 中讲到，每个元音的产生都对应一个特定的发音器官的位置，进而每个元音都具有一个特定的共振峰频率位置。初看上去，元音的感知似乎简单明了：要感知一个元音，听音人只需提取前两三个共振峰的频率，并将其映射至内化的（internalized）元音范畴即可。例如，如果知道第一和第二共振峰的频率分别是 300 Hz 和 2300 Hz，那么一个英语母语者会将这种共振峰频率模式感知为元音/i/。同样地，如果第一和第二共振峰的频率分别为 800 Hz 和 1100 Hz，那么英语母语者可能会将其感知为/ɑ/。

然而，语音环境、语速、发音人的体型等因素均会影响元音共振峰频率的具体模式。语音环境（即元音前后的实际音段）和语速的改变都会导致**元音不到位（vowel undershoot）**。在这种情况下，共振峰频率没有达到“标准”值，即在单独发音时所测得的值；在单独发元音时，发音器官处于“最佳”位置。例如在 CVC 音节中，由于起首或结尾辅音的实现需要调整元音的发音构型和/或发音速度，所以实现元音发音的发音器官无法达到标准位置。也就是说，发音器官会偏离目标位置，进而共振峰频率也会偏离标准值。但并非所有情况都会表现为共振峰频率过低，例如高元音（F1 低）发音不到位的体现是偏高的 F1，也就是说，由于 F1 的目标频率值低，发音不到位会导致得到一个比较高的频率值。

如果考虑到说话人的因素，元音的感知便更加复杂了。声道长度是决定共振峰频率位置的主要因素，不同说话人的声道长度不尽相同，因

此不同说话人所发的同一个元音会具有不同的共振峰频率。例如，同样为元音/i/，可能男性发音人的 F1 为 300 Hz，F2 为 2300 Hz，而女性发音人的 F1 为 400 Hz，F2 为 2800 Hz。但是即使是同一方言区的一组男性发音人，发音人之间共振峰频率的差异也是相当明显的。如图 13.1 为 45 位男性所发英语单元音的 F1 与 F2 值（Hillenbrand 等，1995）。如果每个元音范畴均可由第一共振峰和第二共振峰唯一确定，那么不同范畴的元音之间就不应该有重叠。但是，在一些元音范畴之间，我们可以观察到明显的重叠。

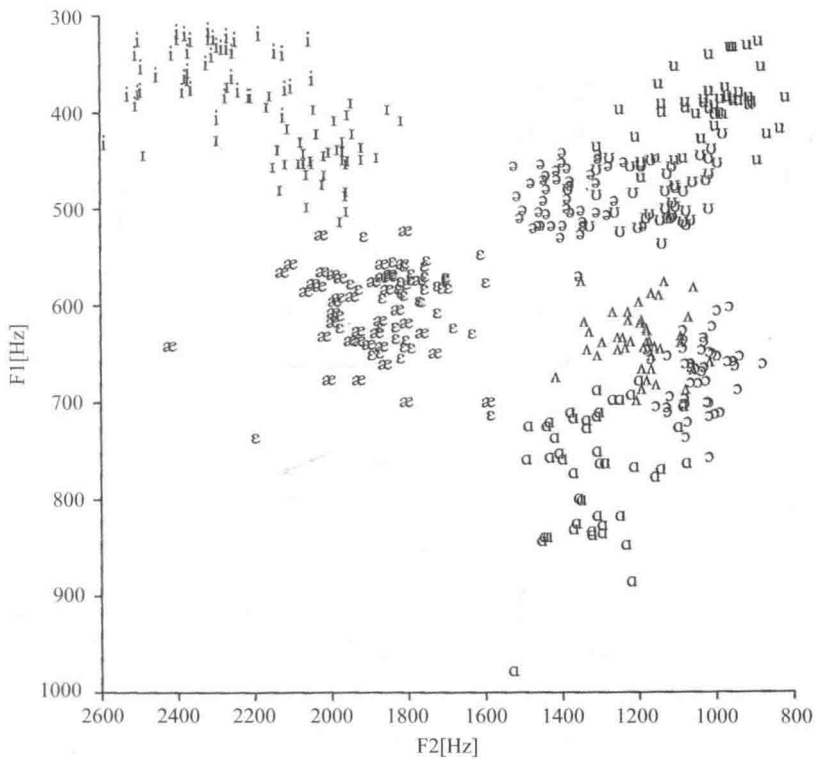


图 13.1 Hillenbrand 等（1995）所绘制的元音空间图，发音人为 45 位成年男性。

听话人须把这些不同的声学模式归为同一个音位范畴，这一过程被心理学家称作**等同分类（equivalence classification）**。人们假设听话人可

以对声道长度引起的变化进行补偿，这一过程被称作声道**规整（normorlization）**。声道归整的基本思想是：听话人利用声学信号的某些特征来估算说话人声道的长度，进而确定该说话人元音空间的频率参数。规整是去除声学特征的系统共变的过程，这样便可以减少或消除元音范畴之间的明显重叠。

13.1.1　外部规整与内部规整

声道归整有很多种模型，其中**外部规整（extrinsic normalization）**和**内部归整（intrinsic normalization）**是两种主要的方法。根据单纯的外部规整模型，听话人可以通过单个说话人所发元音的分布信息确立参考框架，也就是说听话人可通过前后语音环境来校准特定说话人的元音空间，从而在感知该说话人所发的后续元音时进行参照。Ladefoged 和 Broadbent（1957）的经典研究最先为这一理论提供了证据。对于“Please say what this word is”这句话，他们合成了六个版本，每个版本都具有不同的 F1 和 F2 范围。同时，他们还创建了四个基于“bVt”结构的目标词（即辅音/b/后面跟元音再跟辅音/t/，如“bit，bet，bat”等）。当单独听辨以上测试词时，随着元音 F1 的增加，听音人感知到的单词依次为“bit”，“bet”，“bat”，“but”。然而，当把这些单词放在之前那句话（负载句）的后面时，对特定目标词的感知则随负载句中共振峰频率范围的变化而变化。例如，目标词“bit”在单独听辨时被感知为“bit”，但如果前面负载句的 F1 相对较低，则会被感知为“bet”。因此这其中存在一个对比效应，如果负载句的 F1 较低，听话人会将目标的 F1 感知得较实际值高一些（例如低元音），从而将/ɪ/感知为/ɛ/。

这些结果为某些外部归整理论提供了有力的证据，但同时也引出了人们的疑问：哪些种类的前置信息可以最为有效地校准说话人的元音空间；各种前置信息的校准程度又如何？然而当随机混合不同说话人的声音时，元音感知的正确率依然非常高，这一事实可能是反驳外部归整理论的最有力的证据。事实上，上述对混合说话人情况下元音的感知与将前置信息变为同一个说话人的三个**顶点元音（point vowel）**/i，u，ɑ/时的元音感知同样准确（Verbrugge 等，1976）。在校准元音空间的过程中，这些顶点元音提供的信息非常有用，因为它们代表了发音空间的极限位

置，代表了声学空间的极限频率，而发音器官位置的中等幅度内的改变不会显著影响顶点元音的共振峰频率，并且，只有顶点元音的共振峰模式与声道面积函数具有唯一的对应关系。但是，Verbrugge 等（1976）证明，提供顶点元音似乎并不能提高元音感知的准确程度，也就是说，不论目标元音前面是否有顶点元音，听话人都能极为准确地辨识元音。

任何外部归整理论都回避了一个问题：如果对元音的感知取决于其前置的元音，那么成功的感知是如何开始的呢？一种可能的解释是只需正确感知一个元音即可。这个元音最可能是/i/，因为 Nearey（1978）指出，由于［i］独特的共振峰模式，任何人［i］的发音总能被感知为元音/i/。也就是说，根据一个说话人发出的一个元音/i/，听话人就能马上估算出该说话人的声道长度。这是**点规整（point normalization）**方法的一个例子，因为该方法只要求知道说话人元音空间中的一个已知点即可。点规整的一个变体是**范围规整（range normalization）**，该方法至少需要两个已知元音。

与外部规整相对的一种方法是内部规整。根据内部规整的观点，所有用于辨认一个元音的必要信息都包含在该元音之中。该方法认为，当恰当地评估了元音的声学特性之后，元音范畴之间的重叠便不再存在，或至少可以显著降低。这种方式通常希望获得与人类听觉系统所得到的类似的元音表征。前面已经提到，听觉系统会对语音信号进行一系列的非线性变换，特别是在频域和振幅域（见 12.4.3）。传统上，人们通常在线性频率标度的声学空间中表示元音。最近，研究者开始使用更加基于听觉的标度来表示频率（如 Bark 或 ERB）和响度（如方或宋）。采用这些标度似乎的确可以在很大程度上降低在传统标度中所观察到的元音范畴之间的重叠现象（如 Syrdal 和 Gopal，1986）。但要注意，单纯将线性标度转换为对数等其他标度并不会改变标度中各点的顺序：在线性标度上“较低”的值，在对数标度上依然“较低”，只是各点之间的距离发生了变化。当转换标度后，各点之间平均距离（可能代表元音的频率）的增大并不意味着元音之间就变得更加分散，因为只是测量距离的方法改变了，而各元音之间的相对位置并未发生变化。

除了使用更基于听觉的标度，研究者同样探索了针对单个共振峰频率位置所提供信息的不同组合方法。例如，除了传统参数 F1 和 F2，Miller

（1989）增加了 F_0 和F3，并在三维空间中表示元音，三个坐标轴分别是 $x = \log(F1/F_0)$、$y = \log(F2/F1)$、$z = \log(F3/F2)$。这种方法基于共振峰比值理论，即元音音质取决于共振峰频率之间的比值而非各共振峰频率的绝对数值。其观点为，对于给定的元音，声道长度的变化会使该元音所有共振峰的频率都以相同的比率发生变化。因此，计算比率可以排除声道长度的差异。这种方法同样可以很好地划分元音的范畴。

在本节的开始，我们指出，规整是从语音信号中去除特殊信息的过程，其目的是得到典型的、无重叠的抽象范畴。但最近的研究表明，保留特殊信息的范畴会具有一定优势。**范例理论（examplar-based theory）**认为，一个感知范畴（例如辅音或元音）包含该范畴的所有已知范例。一个范例由一组声学或听觉特性与一组范畴标签之间的关联构成。由于范例保留有全部相关的声学（例如详细的谱）、语言学和标记（例如语言学范畴、说话人性别）信息，人们可参照以上任意一个维度对新项目进行归类。该理论假设，构成每个语音范畴的“范例云”一般会在由上述所有信息所构成的更高维度的空间中占据相对独立的区域。Johnson（1997）利用一个基于范例理论的模型模拟了对元音的辨认任务，结果表明，即便是不进行任何形式的说话人规整，元音辨认的准确率也很高。关于范例模型对语音信号中其他变量的适应程度仍有待讨论（见 Johnson，2003）。

最后，值得注意的一点是，目前为止所提到的元音表征方法均基于元音中点处的共振峰频率。元音中点距元音两侧的音段最远，因而这一部分受语境效应的影响（如协同发音）最小。也就是说，在元音中点测量所得的共振峰频率可以提供对该元音“最纯正”的表征。但是，如果只关注元音的稳定段，其他潜在的重要信息就会丢失。特别是元音时长、元音共振峰的入渡和出渡等动态特征，均可以为元音音质的感知提供重要的线索。一些实验（如 Strange 等，1976；又见 Macchi，1980）确实证明，与独立元音相比，CVC 环境中的元音的感知结果更为准确。这说明听音人感知元音时所利用的信息不仅限于稳定段，他们同样还会利用动态的信息。Hillenbrand 和 Nearey（1999）的感知实验进一步证明了动态信息的重要性。该实验要求听音人识别合成的“hVd”（例如“heed，hid，head”）刺激。有两种合成版本，一种是对自然语音的高度仿真，另

一种则不包括共振峰过渡信息，共振峰的频率值被设置为元音在“最稳定状态”的测量值。结果显示，对不含过渡信息的刺激的元音识别率（74%），显著低于保留谱变化信息的刺激（89%）。

13.2 辅音

由于辅音的声学特征更富于变化，人们通常认为辅音感知比元音感知更为复杂。声道长短、语音环境、语速和重音均可以影响辅音的声学特征。我们将从近音开始介绍，因为近音与元音最为相似。

13.2.1 近音

发近音时，发音器官的收紧程度很小，舌靠近被动发音器官，但并不会过于接近。所以，近音的声学特征与元音在很多方面都有相似之处。从听感上讲，近音共振峰的入渡和出渡是发音部位的重要线索（见10.2.1）。现已证明，第二共振峰过渡的起始频率是区分/w/和/j/的主要线索。由于发/w/时有一个位于双唇的收紧（另外还有位于软腭的收紧），所有共振峰的频率都会比较低（见10.2.1），较低的F1和F2是主要的感知线索。由于/j/是在硬腭处收紧，前腔较短，较高的F2是主要线索。而/ɹ/与/l/之间的差异似乎体现在许多线索中。这两个音最显著的区别是F3的位置：/ɹ/的F3很低，明显低于/l/的F3。不过，研究显示，F2的起始频率（/ɹ/低，/l/高）和F1的过渡时长（/ɹ/长，/l/短）同样有助于对/ɹ/与/l/差异的感知。

关于近音发音方法线索的研究，主要集中在对爆发音/b/与近音/w/的区分上。共振峰（特别是F2）的过渡时长是区分两者的一个重要线索：/b/的过渡段短，而/w/的过渡段较长。利用语音合成技术，延长/b/F2的过渡时长，并保持其他参数不变，便可以将一个合理的/b/变为一个可以被接受的/w/。只要时间参数是区分差异的主要或唯一因素，那么语速变化就很可能会影响人们对差异的感知，**范畴边界（category boundary）**的变化就体现了这一点。针对一个或多个声学参数（本例中是F2的过渡时长）的改变，如果某一个取值使听音人有50%的情况将其感知为一个范畴（例如/b/），而另50%的情况将其感知为另一个范畴（例如

/w/），那么该取值处就是这两个范畴的范畴边界。（关于范畴边界的详细讨论请参见13.3.1。）事实证明，/b/和/w/的范畴边界的确会随语速的变化而移动：在实验中，通过缩短起始辅音后的元音的时长来模拟语速的加快，则将辅音判断为/w/的情况也随之增加（Miller 和 Liberman，1979）。他们认为，之所以会发生上述现象，是因为元音的缩短使过渡的时长显得相对更长了。这个例子体现了声学线索的本质属性——语境相关。

但是，将F2的过渡时长作为区分/b/与/w/的唯一线索，这可能有一些把问题简单化了。10.4.1中讲到，塞音与滑音除阻快慢的差异既会影响振幅也会影响过渡时长。例如，/b/在除阻时，振幅瞬间增加，而/w/在除阻时，振幅则变化缓慢。自然语音中，过渡时长、振幅以及其他多种细微线索均有助于我们对/b/和/w/进行区分。当所有这些线索被同时用作区分判断的时候，语速的影响就会大幅减少甚至完全消失（Shinn 等，1985）；而在有背景噪声时，语速效应似乎又会重新发挥作用（Miller 和 Wayland，1993）。

13.2.2　擦音

擦音相对较长，而且在其时长范围内都十分稳定。擦音有两个决定性特征，一是摩擦噪声本身，二是擦音向后接元音的过渡。从声学上讲，我们可以通过谱峰位置和相对振幅（见10.2.2）来辨别擦音的发音部位。

一些研究从感知上考察了谱峰位置的作用。其中最著名的是 Heinz 和 Stevens（1961）的研究。他们通过合成技术，系统地将擦音的谱峰从2 kHz逐渐变为8 kHz。听辨结果表明，峰值在3 kHz以下的擦音被感知为/ʃ/；峰值在4.5～6.5 kHz之间的擦音被感知为/s/；峰值在6.5 kHz以上的则被感知为/f/或/θ/。但是，听音人无法对/f/和/θ/进行区分。在研究周围元音对辅音/s/和/ʃ/感知的影响时，相关研究同样表明，通过改变谱峰的位置可成功地生成/s－ʃ/的连续统（例如 Mann 和 Repp，1980；Whalen，1981）。总之，这些研究表明，谱峰的位置是区分咝音/s，ʃ/与非咝音/f，θ/的可靠感知线索，而谱峰同样可以在感知上区分/s/与/ʃ/。

最近的研究表明，相对振幅同样是确定擦音发音部位的感知线索。Hedrick 和 Ohde（1993）证明，擦音的振幅与元音起始段的振幅之间的相

对关系影响听音人对/θ/、/s/和/ʃ/的感知。但是，在感知非咝擦音内部的差异时，相对振幅是否是线索之一，目前还不清楚。

至于摩擦噪声的其他特征在感知上的重要程度如何，相关证据错综复杂。Behrens 和 Blumstein（1988）的研究表明，感知擦音时，仅依靠噪声振幅这一个线索并不可靠。另外，噪声的时长似乎也不是擦音发音部位的重要感知线索。

最后，Harris（1958）及 Nittrouer（2002）的研究证明了过渡信息的作用。Harris 录制了由擦音和元音构成的音节，涵盖了英语中所有的擦音以及四个元音。然后她将一个音节的擦音段与另一音节的元音段进行**交叉拼接（cross-splicing）**[1]，以构成听辨实验的刺激。结果表明，对咝音的辨认是基于摩擦噪声的：将噪声段/s，z，ʃ，ʒ/与任一元音段拼接，感知的结果依次还是/s，z，ʃ，ʒ/。但/f，θ/的感知却主要取决于元音段。例如，当与/f/后的元音段拼接时，刺激中的/f/或/θ/噪声会被感知为/f/；当与其他任意元音段拼接时则都会被感知为/θ/。Harris（1958）根据这些结果推断，听音人首先根据摩擦噪声来区分/s，ʃ/与/f，θ/，然后再根据元音段区分/f/与/θ/。

这一观点得到了进一步的证实。当只给听音人听擦音与元音的过渡时，听音人对/f，θ/的感知显著高于机会水平，但对/s，ʃ/的感知却并非如此。而且，去掉过渡信息并不会影响听音人对/s，ʃ/的感知。总之，现有证据表明，共振峰过渡在非咝擦音的感知中发挥着作用。

发音时，声带是否振动是区分浊擦音和清擦音的主要依据。声学上，就是看是否存在低频能量。尽管通常清擦音要长于浊擦音，但时长本身似乎并不足以区分擦音的清浊。最后，擦音与其他阻音的区别主要体现在起始部分的特征上：擦音的上升时间较长，其次是塞擦音，爆发音的上升时间最短。只要存在起始特征，即使擦音的时长被明显地缩短，听音人也依然能够将其感知为擦音（Jongman，1989）。

13.2.3 鼻音

鼻塞辅音具有很多标志性的特征，包括鼻音哞音段的共振峰与反共振峰模式，以及鼻音起始和末尾处的共振峰过渡。鼻音发音方法的线索相对明显，而关于发音部位的线索则更为复杂。鼻音的哞音段包含了几

个关于鼻音发音方法的线索，包括带宽较宽的弱共振峰、低频共振（约300 Hz）即鼻共振峰，以及反共振峰（见10.2.4）。哞音段可以为鼻音发音方法的感知提供充分的线索。当口腔元音后接有鼻音时，软腭通常在元音的后半段就开始下降。由于软腭通道打开，元音频谱中的高次共振峰能量变弱，这同样为后接鼻音提供了感知线索。

发音部位的感知线索同时存在于哞音段和共振峰的过渡中。早期研究（例如Liberman等，1954；Malecot，1956）认为，携带有关发音部位的线索的是过渡段而非哞音段。而较新的研究则表明，听音人会同时利用这两种信息源。具体而言，虽然仅通过哞音段或共振峰过渡来判断发音部位看上去是可行的，但如果给听音人呈现几个包含哞音段和过渡段的脉冲，对发音部位的辨认率会大大提高（Kurowski 和 Blumstein，1984；Repp 和 Svastikula，1988）。

13.2.4　爆发音

声学上，爆发音的特点表现为一系列的事件：辅音闭塞、除阻爆破和共振峰过渡（包括前一音段向爆发音的过渡和爆发音向后一音段的过渡）。这些特征均可以用频率、振幅及时长进行分析。嗓音起始时间（VOT）是另一个用于区分爆发音的时间参数。从发音方法的角度来看，爆发音独树一帜，因为其发音特点包括闭合段、除阻爆破和共振峰的快速过渡。研究表明，只听CV结构音节的前10～20 ms，听音人便可以判断出听到的辅音部分是爆发音。爆发音的清浊可以有多条感知线索。相比浊爆发音，清爆发音通常具有更长的闭合段，没有或只有短暂的声门脉冲，具有更加强烈的除阻爆破，具有正值VOT，并且其后接音段的 F_0 和F1的起始频率更高。

大多数有关爆发音感知的研究所关注的都是发音部位的线索。有两种声学特征包含有关发音部位的信息，一是除阻爆破的频率，二是共振峰过渡模式。在过去的50年里，很多研究者都试图去弄清楚这些特征（或单独或共同作用）是否可以并在多大程度上可以帮助听音人判断爆发音的发音部位。早期研究表明，爆破频率本身并不是判断发音部位的可靠线索。通过合成的语音刺激，研究者可以只改变人工生成的语音的一个声学参数，而使其他参数保持不变。Liberman等（1952）就合成了一系

列除阻爆破，频率从 360 Hz 到 4320 Hz。这些爆破之后接有一系列双共振峰（F1 和 F2）模式，这些共振峰模式代表了英语中的七个不同元音。听音人需判断这些合成刺激的起始辅音是/p/、/t/还是/k/。听辨结果请见图 13.2。尽管爆破特征总能被感知为爆发音，但对其发音部位的感知却会在很大程度上受到元音环境的影响。大致而言，高频的爆破总能被感知为/t/；频率很低的爆破大多被感知为/p/。这一发现与我们在 10.2.3 节中关于发音部位声学线索的讨论一致。但对处在 1400～2000 Hz 的中频爆破，当后接元音是/ɛ，a，ɔ/时，会被感知为/k/，而当后接元音是/i，e，o，u/时，则会被感知为/p/。爆破频率似乎并不是区分爆发音，特别是软腭爆发音，发音部位的可靠线索。

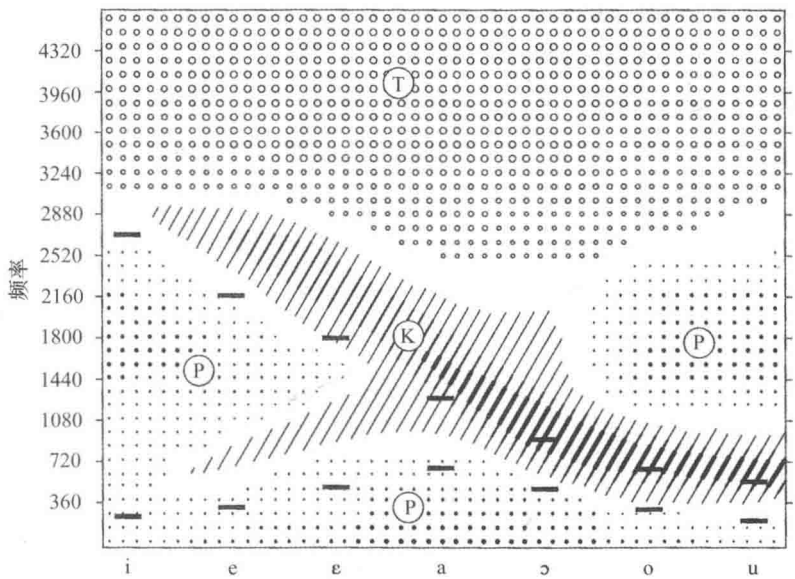


图 13.2 听音人在爆破感知研究中对合成刺激的感知结果，引自 Liberman 等（1952）。纵轴代表频率（单位为 Hz）。空心圆圈、实心圆圈和斜线分别表示感知为/t/、/p/、/k/的爆破频率。浓度表示感知程度。横线表示相应元音的共振峰频率（F1 和 F2）。

后续的实验通过对自然语音进行交叉拼接来合成刺激，以对发音部位的感知进行研究。例如，Cole 和 Scott（1974）录制了英语元音/i，ɑ，u/之前的清爆发音和浊爆发音，然后将辅音从一个元音环境中拼接至另一个元音环境。例如将/ti/中的爆破和送气段“t(i)”拼接至从/tu/中取出的元音段“(t)u”，反之亦然。Cole 和 Scott（1974）的研究表明，不论元音环境是否改变，双唇音和齿龈音的辨认率均非常高，但对软腭音的感知就要差许多。利用相同的方法，Dorman 等（1977）研究了 9 个不同元音前的唇、齿龈及软腭浊爆发音，结果表明对发音部位的辨认率很低。

一些研究只向被试提供了自然产出的“爆发音－元音”音节中的除阻爆破段，结果表明，这种情况下听音人对发音部位的辨认准确率在中等至良好之间（例如 Winitz 等，1971；Kewley－Port 等，1983）。Tekieli 和 Cullinan（1979）的研究表明，对于后接 8 个不同元音的情况，当仅呈现清爆发音的前 10 ms 时，听音人对发音部位的辨认率明显高于机会水平。这些研究总体表明，爆破频率可能是感知塞辅音发音部位的一个必要但不充分的线索。

同样，也有大量实验探讨了共振峰过渡在发音部位的感知线索中所扮演的角色。作为对上述探索爆破频率作用的研究的补充，Liberman 等（1952）合成了一系列双共振峰模式（无爆破），并系统地改变 F2 过渡段的取值。感知结果显示，大多数听音人将上升的 F2 感知为双唇爆发音，而对于下降的 F2 模式，则根据后接元音的不同，分别感知为齿龈爆发音或软腭爆发音。

关于共振峰过渡在感知中所起的作用，另一种评估的方法是将自然产出的“爆发音－元音”音节中的爆破段去除，只保留共振峰过渡。相关研究结果表明，对于浊爆发音，发音部位的辨认率在中等至良好之间，而对清爆发音发音部位的辨认率却要差得多（例如 LaRiviere 等，1975；Dorman 等，1977）。

在讨论爆破和共振峰过渡对辨认爆发音发音部位的贡献的研究中，目前要数 Smits 等（1996）的研究最为细致和全面。他们对荷兰语中的爆发音/p，t，k，b，d/进行了研究。在研究中，这些爆发音的后接元音为/i，y，a，u/。他们利用数字化的交叉拼接与滤波技术，合成了“只有爆破”“无爆破”及“交叉拼接”的刺激。“只有爆破”和“无爆破”刺激

的平均辨认率请见表 13.1。

表 13.1 “只有爆破”和“无爆破”刺激在所有元音环境中的发音部位辨认正确率（百分比）（引自 Smits 等，1996）

	辨认正确率						
刺激	[p]	[t]	[k]	均值	[b]	[d]	均值
只有爆破	80	50	91	74	65	57	61
无爆破	95	64	47	69	96	87	92

该研究通过交叉拼接来合成刺激，即将给定发音部位的爆发音的爆破段（例如/pa/的爆破段“p(a)”）与其他发音部位的爆发音中除爆破段以外的部分（例如“(t)a”或“(k)a）拼接在一起。如此得到的刺激所包含的线索与感知冲突，因此被称为**冲突线索（conflicting-cue）**信号。在这种情况下，爆破段指征一个发音部位，而共振峰过渡则指征另一个发音部位。对于清爆发音的交叉拼接刺激，49% 的辨认与爆破段相关，43% 的辨认与过渡段相关。并且，/k/的爆破特征比/p/和/t/的爆破特征对部位的指征作用更强，而过渡段对/p/、/t/、/k/发音部位的指征作用同等重要。对于浊爆发音的交叉拼接刺激，只有 26% 的辨认与爆破段相关，而有 74% 的辨认与过渡段相关。/d/的爆破比/b/对发音部位的指征作用更强。

总之，这一研究证实了将爆破和过渡段作为爆发音发音部位线索的重要性，并加深了我们对这一问题的理解。尽管线索的可靠性会随不同的发音部位、元音环境以及说话人而变化，但仍有据可循。一般而言，在前元音环境中，爆破段是感知发音部位的最强线索，而在后元音环境中，过渡段则是最重要的线索。此外，清、浊爆发音相比，清爆发音的爆破段对其发音部位的指征作用更强，而浊爆发音的过渡段则能为对其发音部位的辨认提供更多信息。

我们在 10.4.1 中提到，爆破频率和共振峰过渡频率等单独的声学特征通常会表现出很强的语境依赖性，因此，一些研究者考虑将全局而非局部的特征作为发音部位的感知线索。例如我们在 10.4.1 中提到，Lahiri 等（1984）发现，辅音起始部分与后接元音起始部分相比，高频能量相对于低频能量在分布上的变化是一个稳定的线索。该研究中的感知实验

证实，听音人在进行语音判断时，会综合运用爆破起始部分的谱与元音起始部分的谱这两条线索。

13.3　言语感知的运动理论的贡献

我们在10.4节中谈到，为研究音位对立的相关量，Liberman及其同事进行了大量而仔细的实验。但在语音环境变化的条件下，他们未能找到稳定的声学相关量。既然缺乏声学上的不变量，那人们到底是利用哪些其他渠道的信息来实现对语音的感知的呢？Liberman及其同事推断，听音人一定是利用了发音方面的信息。他们据此建立了**言语感知的运动理论（the motor theory of speech perception）**，通常简称为运动理论。根据运动理论，言语感知与发音（而非声学）更为相关。该理论认为言语感知是以发音为参照的：发音运动及其知觉效果共同调节着声学刺激和相应的感知之间的关系。运动理论的基本假设是：听音人可以通过分析引起当前听觉模式的发音模式来解析其当前所接受到的听觉模式。因此，产生语音的发音过程才是语音的不变线索。这一理论引人入胜的特点之一是其直接将言语感知和言语产出联系了起来。而该理论提出的“言语产出和言语感知过程有很多共同之处”这一假设，从直观上讲，要比两者无关说更有吸引力。早期的发音研究为该理论提供了支持：尽管同一个语音存在许多声学上的变体，但发音器官的位置和运动可以将这些声学变体归为一类（Liberman，1957）。然而，同样有研究发现，同一个语音可由非常不同的发音动作来实现（MacNeilage，1970）。所以，发音运动序列跟语音信号自身一样多变。因此，Liberman（1970）对运动理论进行了相应的修改：对语音进行分类的是控制发音器官的运动指令，而不是发音运动本身（Liberman，1970）。因此，听音人对声道运动指令效果的了解构成了感知过程的单位（Liberman，1982）。这一理论被进一步扩展为**修正的运动理论（revised motor theory）**，即对**预期音姿（intended gesture）**的感知（即使相关音姿并没有实现）促成了对语音的感知（Liberman和Mattingly，1985）。根据这种观点，听音就是对引起具有协同发音的并且高度多变的语音信号的预期音姿的感知。

另一个从音姿角度解释言语感知的是**直接现实理论（direct realism）**（Fowler，1986；2003）。直接现实理论是一个通用的感知理论，而不仅限

于听觉感知。就言语感知来说，运动理论和直接现实理论都认为，感知的直接对象是发音姿态。但是，运动理论中的感知对象是预期音姿，或者是产生这些姿态的指令，而直接现实理论则认为感知对象是实际的声道运动。其道理与人类的视觉感知相似：眼睛感知到的是物体，而不是光线的物理反射；同样，人耳感知到的是发音姿态，而不是语音信号的频率、振幅和音长的物理变化。这一观点要求声学信号可以被还原为唯一的声道运动，但实际情况似乎并非如此，因为对于一个给定的声学信号，我们可以通过多种发音方式来实现其声学特点。

最初的运动理论没有解释人类如何通过语音信号感知发音位置和发音动作，以及儿童是如何学会这种转换的，这是该理论受到的主要批评。即使在修正的运动理论中，相关研究者也没有提出实际的机制。最近的神经学研究或许可以为运动理论提出的言语感知中的运动募集（motor recruitment）思想提供支持。Rizzolatti 及其同事发现，猴子有一组神经元——**镜像神经元（mirror neurons）**，当猴子观察到的事件与内部产生的类似行为匹配时，这组神经元便会被激活。具体而言，当猴子实际执行一个动作时，比如撕纸，或者当它们只听到撕纸的声音时，这些镜像神经元都会放电（例如 Rizzolatti 和 Arbib，1988；Kohler 等，2002）。这一发现令人振奋，可以激发人们对言语的产出与感知之间关系的进一步研究。至于镜像神经元的概念是否可以证明运动理论，现在下定论还为时尚早。例如，对撕纸声有反应的神经元可以由联想学习产生，这个联想学习机制是一个为了快速应对环境状况而进化出来的系统的一部分。目前，可以肯定的是，运动理论激发了大量的研究，其中既包括该理论的支持者也包括该理论的批评者，这些研究都有利于人们深入探索言语感知的本质。

13.3.1 范畴感知

Liberman 及其同事最著名的发现当属**范畴感知（categorical perception）**。范畴感知是指将均匀分布于一个物理连续统中的刺激感知为一个或另一个范畴，而不是随物理值的变化而变化的函数。在言语感知中，范畴感知现象是指感知某些对立的方式。范畴感知要求测试听音人辨认和区分语音的能力。一般而言，与**辨认（identifying）**声音（标记声音；即这是一个“升 C 调”，那是一个“B 调”）相比，人们更擅长**区分**

(discriminating) 声音（即听出一对声音的差异）。例如，听音人可以区分出数百种不同的音高，但能辨认出来的却少于 10 种。

我们来举个例子以便清楚地说明这一点。考虑一下英语中音节首爆发音的清浊对立。浊爆发音的 VOT 通常为较短的正值，而清爆发音的 VOT 为较长的正值（见 10.2.3）。通过合成的语音或者经过调制的自然语音，我们可以生成一系列语音刺激，使其 VOT 呈系统性地改变。第一项的 VOT 值可以为 0 ms，下一项的 VOT 值为 10 ms，以此类推，直至最后一项的 VOT 值为 80 ms。这样便生成了一个清浊连续统，例如从/dɑ/到/tɑ/，共包含 9 个刺激项。再将每个刺激项重复多遍（例如 10 遍），共构成 90 个刺激项，然后对这 90 项进行**随机（randomized）**排序。随机化可将刺激以任意顺序排列。之后，将经过随机排序的刺激呈现给听音人，让他们辨认听到的是/dɑ/还是/tɑ/。也就是说，向听音人呈现的刺激的 VOT 可能第一项为 40 ms，第二项为 0 ms，第三项为 60 ms，等等。我们预计听音人会将 VOT 为 0 ms 和 80 ms 的刺激分别辨认为/dɑ/和/tɑ/。那么其他刺激项的辨认结果如何呢？一个合理的猜测是：辨认为/dɑ/的数目随 VOT 的增加呈线性下降的趋势，如图 13.3 所示。但是，这种辨认实验的典型结果通常更像图 13.4 中所示的样子。

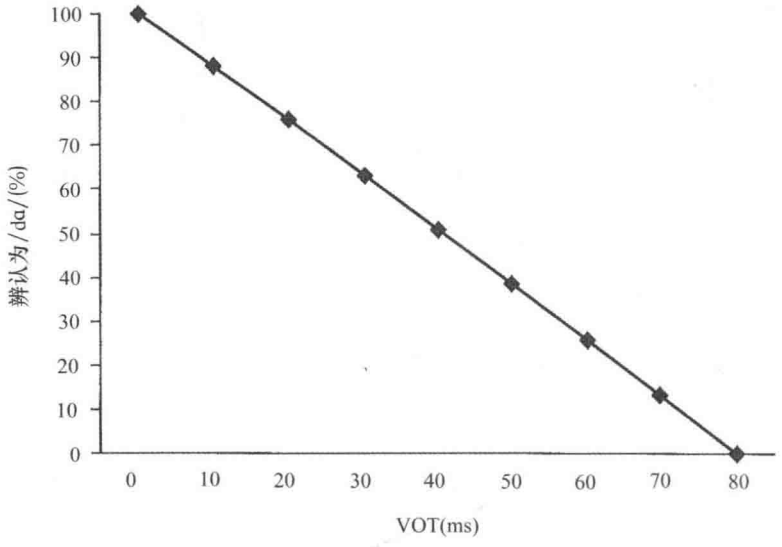


图 13.3　辨认实验的预期结果，其中 VOT 呈系统性变化。

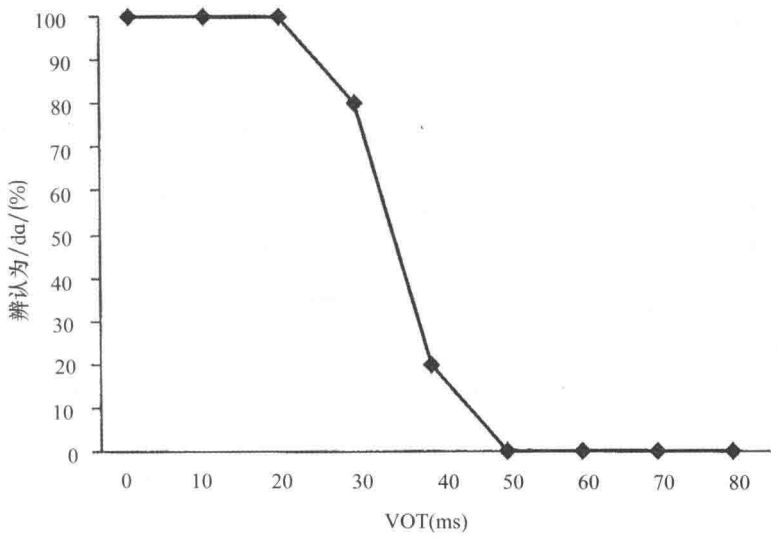


图 13.4 辨认实验的实际结果，其中 VOT 从/dɑ/到/tɑ/呈系统性地变化，每次增加 10 ms。

从/dɑ/到/tɑ/的变化不是渐变，而是突变或者说是范畴性的变化。听音人把这些刺激划分为两个边界分明的范畴“da”和“ta”。两个范畴（或音位）之间的范畴边界是/dɑ/（或/tɑ/）的辨认率为 50% 的 VOT 值，大约为 35 ms。注意对于划为同一范畴的刺激，其辨认结果几乎没有差异。例如，VOT 为 0 ms 和 10 ms 的刺激均被辨认为/dɑ/。但是，当刺激处于边界两侧时，辨认的结果却有很大的差异，即便两者的物理距离与其他的刺激对之间是相同的（VOT 差值同样为 10 ms）。也就是说，VOT 为 30 ms 的刺激在 80% 的情况下被感知为/dɑ/，而 VOT 为 40 ms 时只有 20% 的情况被感知为/dɑ/。

除了很陡的辨认函数以外，范畴感知还具有特征区分函数。由于范畴感知的一个前提是区分能力不如辨认能力，因此还必须进行区分实验（否则这种感知模式不能称为范畴感知）。区分实验有多种操作方法，这里我们只关注最常用的操作方法，即 **ABX 任务**，也叫 **odd-ball 任务**。在这种特定的区分实验中，每个实验项会呈现 A、B、X 三个刺激。其中 A 与 B 不同，而 X 与 A 或 B 相同。听音人的任务是判断 X 与 A 相同还是与

B 相同。例如，在选择 A 和 B 时，两者之间总要隔有连续统中的一项（称为“二步区分”，因为 A、B 之间须有两步的距离），因此，对于之前讨论的 0 ~ 80 ms的 VOT 连续统，A 和 B 的组合共有 7 种可能（VOT 的值分别为 0 ms 和 20 ms，10 ms 和 30 ms，20 ms 和 40 ms，30 ms 和 50 ms，40 ms 和 60 ms，50 ms 和 70 ms，60 ms 和 80 ms）。因此，实验项的三个刺激可以是 0 – 20 – 20，听音人必须判断最后一个 VOT 为 20 ms 的刺激听起来更像第一个 VOT 为 0 ms 的刺激，还是第二个 VOT 为 20 ms 的刺激。同样，每个实验项重复 10 次左右，随机呈现给听音人进行区分。如果两个刺激很容易区分，听音人的准确率应达到或接近 100%。如果两个刺激难以区分，对于给定的刺激，听音人会在一半的情况下将其判断为 A，一半的情况下将其判断为 B，因而准确率为 50%。区分实验的结果见图 13.5。

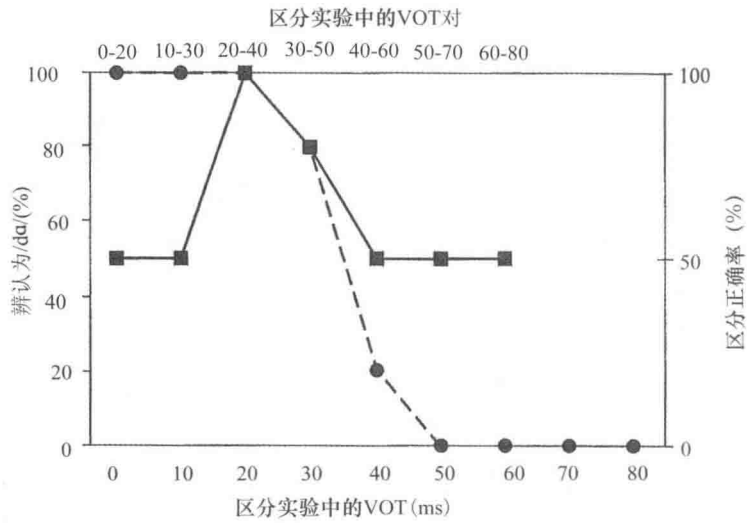


图 13.5　二步区分实验的结果（实线），VOT 连续统中的刺激与辨认实验相同。虚线表示辨认实验的结果。

区分函数两端平坦，中间具有峰值。其中，确实有一对刺激非常容易区分，就在区分函数的峰值位置。该刺激对的 VOT 分别为 20 ms 和 40 ms。也就是说，听音人可以很容易地将这对刺激区分开来。而回顾一

下辨认实验，这对刺激分别被划归到了两个不同的范畴“da”和“ta”。在图中同样可以看到，对其他所有刺激对的区分准确率都在50%左右。这表明听音人只能在一半的情况下区分这些刺激对中的两项刺激。也就是说，听音人在一半的情况下认为刺激对中的两项刺激相同，而另一半的情况下则认为它们不同。因此，从本质上讲，听音人只是在猜测。对这些刺激的区分表现处于机会水平。也就是说在只有两种反馈类别（“相同”或“不同”）的情况下，通过猜测所得的准确率为50%。值得注意的是，对于所有那些区分率为机会水平的刺激对，它们在辨认实验中均被归于相同的范畴。

总之，这些结果表明，人们的区分能力确实不如辨认能力。这与本小节开始关于音高感知的说法非常不同。人们对辨认同一范畴内的刺激所得的辨认结果并无差异。此外，人们对同一范畴刺激的能力区分处于机会水平（50%），而区分音位边界两侧的刺激时，正确率就非常高。这种辨认和区分模式是范畴感知的特点。除了对清浊的感知，人们对发音部位的感知（例如使用F2过渡段呈系统性变化的/bɑ-dɑ-gɑ/连续统）同样也是范畴化的。人们对爆发音的范畴感知是一个非常稳定的现象，研究者们在涉及不同语言及听音人的研究中都可以不断地重复这一现象（详尽综述请见Repp，1984）。

与爆发音的感知相反，元音的感知似乎是连续的。如果通过逐步等值地提高F1来创建一个元音的连续统，例如从/ɪ/到/ɛ/，那么，人们对这个连续统的感知便不是范畴化的，而更像图13.3所示的逐渐的连续的变化。以上这一事实，即爆发音是范畴感知而元音不是，最初被当作是支持运动理论的证据。人们将感知看作是对产出的模仿。由于爆发音/b，d，g/分别是由不同范畴的发音姿态产生，因此对它们的感知也是范畴化的。而元音的产出则更具连续性，所以人们对元音的感知也是连续性的。

但是，之后的研究表明，爆发音与元音的感知差异或许与发音上的区别没有关系。这两类音在时长上也有差异——元音较长，而爆发音较短；并且具有动态特征是爆发音更重要的一个特点。当元音缩短至与爆发音时长相当时，对元音的感知同样会变为范畴化的。

对范畴感知的最初的理解认为，听音人只能区分属于不同范畴的刺激，这种观点可能太过绝对了。最初的结论认为听音人不能区分属于同

一范畴的不同刺激，这一观点可能与实验任务的限制有关。利用另两个反应测试量度——**反应时（reaction time）**和**优度评级（goodness rating）**，相关研究表明，听音人确实可以感知到同一范畴内不同刺激项之间的差异。在研究中，特别是在心理语言学研究中，研究者们经常用反应时来测量大脑加工的工作量。通常，刺激越模糊，被试（即实验中的听音人）便会需要更多的加工时间，作出某种反应（例如按键）的时间就越长。优度评级表示被试对给定示例的优劣程度的判断。例如，被试须判断听到是否为/tɑ/，并由“好”至“不好”作出不同程度的评价。或者，可以画一条线，一端为/tɑ/，另一端为/dɑ/，被试须在这条线上用叉来标记呈现项与/tɑ/或/dɑ/的接近程度，即判断刺激（呈现项）与目标（/tɑ/或/dɑ/）的匹配程度。最初的辨认结果显示，人们似乎认为同一范畴内的刺激都是完全相同的（例如100%都为/dɑ/）。但是，反应时和优度评级的测量证明了范畴实际上是具有内部结构的：刺激越靠近音位边界，辨认得就越慢，其优度评级也越低。最近使用眼动和脑电图（EEG）测量的研究进一步证明了范畴内部的细节在言语感知中的重要性（例如McMurray 等，2002；Sharma 和 Dorman，1999）。因此，这类研究促使我们对范畴感知的最初观点重新进行审视。对于不会触发音位范畴对立的声音而言，忽略它们之间细微的声学差异或许可以使言语感知过程更加高效，但这并不意味听音人意识不到那些差异。从这个意义上讲，区分能力不如辨认能力的范畴感知现象是等同分类的一种形式（见13.1节）。

13.3.2　言语是“特别”的吗？

成人对爆发音的感知表现出范畴化模式，这一发现自然激发了人们对范畴感知能力的来源的兴趣。具体而言，范畴感知究竟是由于听觉系统的结构而与生俱来的属性，还是受语言环境的影响而产生的呢？这个问题是争论的焦点。研究者意识到，在新生儿中探索爆发音感知的本质是解决该问题的一个合理的方式，因为新生儿缺少语言接触。但是，直到20世纪60年代，才发展出可以测试婴儿的实验技术。这项技术被称为**高振幅吮吸（high amplitude sucking，HAS）**范式，由 Eimas 等（1971）在一经典研究中首次引入。简单来说，当向婴儿呈现新鲜刺激时，他们吮吸的频率会增加，该范式便利用了这一特点。实验中，婴儿含着奶嘴，奶嘴与传感器相

连，由此实验员便可以监测婴儿吮吸奶嘴的频率了。通过扩音器向婴儿反复呈现给定的刺激，随着刺激的新鲜感逐渐消失，婴儿的吮吸频率会下降，这一过程叫作习惯化（habituation）。关键问题是，再向婴儿呈现一个不同的刺激时，婴儿会做出怎样的反应？如果婴儿认为新刺激与之前听到的不同，我们预期吮吸的频率会增加；如果婴儿认为新刺激与之前的刺激没有差别，则我们预期吮吸的频率会保持稳定甚至进一步下降。于是，我们便可以让婴儿以这种方式来完成一个区分任务。

Eimas 及其同事测试了 4 个月大的母语为英语的婴儿，考察英语音节首爆发音的清浊对立感知。该研究合成了一个/ba-pa/连续统，VOT 的值

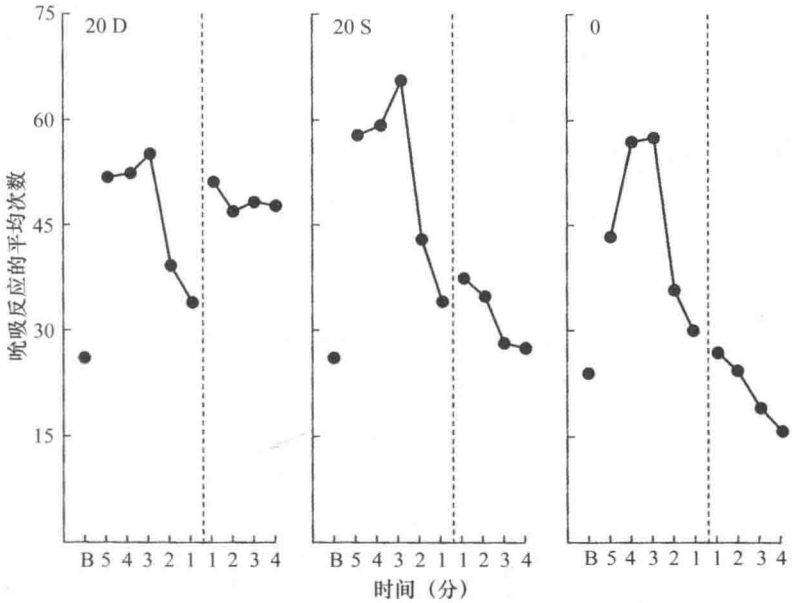


图 13.6 Eimas 等（1971）的结果。在不同时间和实验条件下，4 个月大的婴儿吮吸反应的平均次数表示为时间与实验条件的函数。虚线表示刺激出现变化的时刻，或者在控制组（右图）中表示本应出现刺激变化的时刻。左图表示对 VOT 从 20 ms 变为 40 ms 的反应；中图表示对 VOT 从 0 ms 变为 20 ms 的反应；右图表示对 VOT 不发生改变的反应。B 表示吮吸频率的基线。时间的测量以刺激变化点为参考，表示了变化前的 5 分钟及变化后的 4 分钟内的情况。

经美国科学促进协会（the American Association for the Advancement of Science）许可后重新制图。

从 0 ms 开始以 20 ms 的间隔逐步递增至 80 ms。如图 13.6 所示，这些婴儿的表现与成人十分相似。左边的图表示婴儿能够区分对成人来说属于不同范畴的两个刺激，婴儿可以清楚地将 20 ms 的 VOT 与 40 ms 的 VOT 归为两个不同的范畴。中间的图表示婴儿不能区分对成人来说属于同一范畴的刺激，即婴儿不能区分 0 ms 的 VOT 与 20 ms 的 VOT。同样，婴儿也不能区分 60 ms 的 VOT 与 80 ms 的 VOT，两者对成人来说均为清音。最后，右边的图表示的是控制组的情况，该组实验证实，当呈现的刺激不变时，吮吸的频率会持续下降。

虽然有时人们认为这些结果证明了婴儿是以范畴化的方式感知爆发音的清浊的，但如果要真正证明范畴感知，还需要区分和辨认测试的综合结果。由于几乎不可能获得婴儿的辨认数据，因此在解释区分测试结果时一定要谨慎。尽管如此，婴儿的感知结果表明，成人的范畴感知与语言经验无关，这一行为似乎是与生俱来的。除了清浊对立，婴儿还对多种语言学对立表现出与成人相似的感知特点。

以上我们了解到，人们感知某些语言学对立似乎并不需要语言经验。而这一发现引发了研究者们的争论：人类的言语感知是否利用了专为言语加工而进化出的机制呢？有时人们将这一争论称作**“言语是特别的”（speech is special）**之辩。该争论围绕两个相互竞争的假设，Kuhl（1989）将它们称为**“特殊机制说”（special mechanism account，SMA）**和**“普遍机制说”（general mechanism account，GMA）**。根据“特殊机制说”，言语感知是人类独有的专门针对言语的特殊机制直接作用下的结果；根据“普遍机制说”，言语感知是基于普遍的听觉和认知机制的，不需要内在的专门针对言语的处理。争论中的关键证据通常包括人类对言语声和非言语声感知的对比，以及人类与动物对言语声感知的对比。

非言语信号在某些重要方面可以模拟言语信号。例如，尽管非言语信号不会被感知为言语，但是非言语信号可以用来模拟言语中频率和振幅的快速变化。通过进行非言语信号的实验，我们可以检验“言语感知需要专门针对言语的特殊机制”这一论断的可靠性。很明显，如果有任何发现表明非言语感知与言语感知在某种程度上是相似的，那么这一论断的可靠性就会被削弱，从而使言语感知的普遍听觉机制模型得到支持。同样，如果发现动物感知言语的机制与人类相似，则也会加强对“言语

感知涉及的是跨物种的普遍听觉机制”这一论断的支持。下面我们将简单介绍一下这两个领域的研究。

13.3.2.1 非言语感知

以往的研究已经针对很多语言学上的对立创建了相应的非言语版本的刺激，包括辅音的清浊、发音部位及发音方法。例如，在创建模拟爆发音清浊连续统的非言语版本时，要考虑到 VOT 体现的是两个事件的时间关系，即除阻爆破和嗓音起始。通过系统地改变两个声学事件的间隔，便可以创建一个 VOT 连续统的非言语版本。通过控制两个纯音的相对起始时间，便得到一个音调起始时间（tone onset time，TOT）连续统，图 13.7 展示了其中的几项。该连续统模拟了 VOT 连续统中的关键特征（两个事件之间的时间关系），但同时又不会被感知为言语。

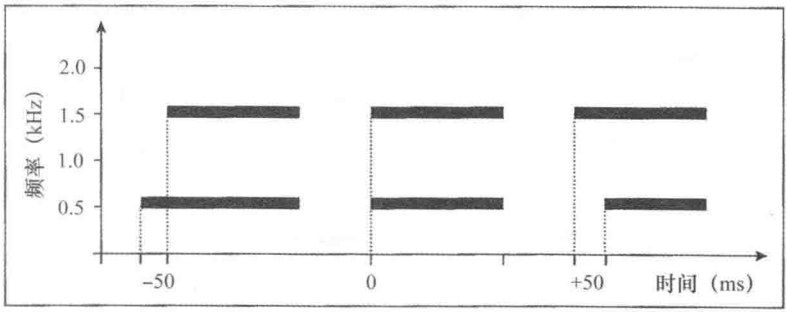


图 13.7 Pisoni（1977）设计的刺激。图中为 VOT 连续统的音调起始时间模拟版本的三个刺激的示意图。每个刺激都包括一个 500 Hz 和一个 1500 Hz 的音调。两个音调的相对起始时间呈系统性地变化。左图表示低调早于高调 50 ms，模拟 VOT 为 -50 ms 的情况；中图表示低调和高调同时开始，模拟 VOT 为 0 ms 的情况；右图表示低调晚于高调 50 ms，模拟 VOT 为 +50 ms 的情况。

研究表明，听音人对这些非言语连续统的感知与相应的言语连续统的感知模式非常相似。例如，Pisoni（1977）发现，听音人划分 VOT 和 TOT 连续统的结果表现出相似的范畴感知特点。类似的情况在由不同发音方法（如/b-w/和/ɹ-l/）以及由不同发音部位（例如/d-g/）构成的非言语连续统的实验中都有所发现。以上发现说明言语加工属于一种普遍机制。

13.3.2.2　*动物的感知*

Kuhl 和 Miller（1975）较早就开始了对动物范畴感知的研究。他们比较了人和栗鼠（拉丁美洲的有毛哺乳动物，听觉系统与人类非常相似）对 VOT 连续统的感知。首先通过电击逃避训练让栗鼠学会判断连续统的两端。研究人员为了通过训练使栗鼠对一端的刺激做出反应，教它们从笼子内的一边逃到另一边以躲避弱电击。而对另一端的刺激，研究人员训练栗鼠原地不动，以抑制这种穿越笼子的反应，并将喝水作为抑制成功的奖励。当它们学会区分这两种极端刺激之后，实验员便加入连续统中的其余刺激以供栗鼠进行辨认。结果表明，栗鼠的清浊边界与人类非常相似。Kuhl 和 Miller 从三个发音部位（双唇音、齿龈音和软腭音）比较了栗鼠对清浊的感知。他们发现动物和人的感知之间存在非常紧密的相关性：在每个发音部位，随着人类的 VOT 边界逐渐移至更长的 VOT 值，动物的边界也同样如此。

关于动物感知的研究还包括其他语言学上的对立（例如发音部位和方法）以及各种不同的物种（例如恒河猴、猕猴及日本鹌鹑）。一般而言，这些研究的结果表明，人类和动物对语言学对立的感知方式有许多相似之处。

非言语感知和动物言语感知实验的结果能告诉我们什么呢？解释这些结果时一定要谨慎。“非言语感知在一定程度上与言语感知类似”这种表述，似乎表明，在考虑基本的言语感知过程时，无须引入专门针对言语的机制这一概念。一般的听觉加工就能应对言语及其他听觉信号的感知。同样，研究发现动物对言语的感知与人类相似，这说明基础的言语感知现象可以利用一般的听觉机制来理解。例如，有人假设清浊边界的 VOT（20 ms 或更长）直接反映了听觉系统的时间分辨率。Hirsh（1959）证明，只有当两个听觉事件相隔 20 ms 以上时，听音人才能判断它们的时间顺序。如果两个事件的起始时间间隔小于 20 ms，听音人就不能确定其先后顺序。就 VOT（或 TOT）而言，一个事件是除阻爆破，另一个事件是嗓音起始。Hirsh 的发现说明，如果除阻与嗓音起始之间相隔不到 20 ms，那么听音人会认为它们是同时开始的，进而会使以像英语这样的语言为母语的听音人将它们都感知为浊爆发音。但如果嗓音起始在除阻之后 20 ms 或更长时间才发生，听音人就能够区分这两个事件，辨别出嗓

音滞后于除阻，从而产生清音的感知结果。因此，嗓音边界可能源于听觉系统普遍的特征和局限性。最近关于非言语声的研究确实表明，基本的听觉敏感性会影响范畴学习的程度。对于两个 TOT 分布而言，当区分边界为 20 ms 时，与 40 ms 的情况相比，被试可以更快地学会区分它们。对后者的区分想必会更加难以学习，因为 40 ms 的 TOT 不对应任何区分结果中的已知峰值，并且其中一个分布的峰值 TOT 大约为 20 ms，是一片已知的敏感性增强区域（area of increased sensitivity）（Holt 等，2004）。这个发现说明可学习性（learnability）也许决定了语言声音的清单（inventory）。

最后，我们并没有说在任何言语或非言语感知任务中，动物都能够和人类有同样表现。在更高级的等同分类行为中，动物与人类之间所表现出的紧密的相关性就不复存在了。例如，人们发现婴儿可以觉察到元音/a/或/i/的声音与相应的一个人在发这些音的情景之间的关联性，但目前并未发现动物具有类似的模式。

13.4 语言学经验在言语感知中的作用

到目前为止，我们一直假定说话人和听话人的母语是相同的。但双语或三语现象在很多文化中十分常见，说话人可能同等程度地精通两种或三种语言，或将其中一种作为主要语言。此外，还有很多人由于生态、政治、教育、经济乃至旅游等原因离开家乡，因此，在世界上的很多区域，会话中的部分或所有参与者使用某一种第二语言进行交流的现象非常普遍。关于非母语产出和感知的研究形成了一个较新的领域，称为**第二语言习得语音学（phonetics of second language acquisition）**。针对非母语言语感知的研究通常被称为**跨语言言语感知（cross-language speech perception）**研究。下面我们将简要综述这一领域中的一些问题。

当成人学习第二语言（L2）时，母语（L1）已形成并稳定了很多年。因此，L1 系统对 L2 感知的影响程度是一个基本的问题。在早期关于跨语言感知的研究中，范畴感知范式占主导地位。这些研究探索了不同母语背景的听音人在划分 VOT 连续统的清浊边界时是否会表现出母语差异。具体做法是将一个取值范围很广（从 -150 ms 到 +150 ms）

的/bɑ/ -/pɑ/的 VOT 连续统呈现给母语为英语、西班牙语和泰语的听音人。我们在 6.3 节中讨论过，对于词首爆发音而言，英语存在负值或短正值 VOT 与长正值 VOT 的对立，西班牙语存在负值 VOT 与短正值 VOT 的对立，泰语存在负值、短正值与长正值三种 VOT 的对立。辨认实验结果表明，当呈现相同的连续统时，英语听音人判断的双唇音清浊边界在 +20 ms，西班牙语听音人在 -5 ms，而泰语听音人则判断出两个边界，一个在 -20 ms，另一个在 +40 ms。进一步的区分实验的结果同样表明，听音人只能区分属于不同范畴的刺激：英语和西班牙语的听音人都表现出两个对立的范畴，而泰语听音人则表现出三个对立的范畴。总之，听音人划分连续统的方式是由母语经验决定的（Abramson 和 Lisker，1970；Williams，1977）。之后关于多种语言及不同辅音对立的研究都验证了这些结果。

这些结果引出一个问题，随着听音人第二语言经验的积累，这些感知模式是否会发生改变呢?

人们在加工 L2 时可以不受 L1 影响或者使 L1 适应 L2 的加工，研究者们认为这种能力反映了大脑的**可塑性（plasticity）**，即大脑根据经验重组其结构和功能的潜力。可塑性这个概念与**关键期假说（critical period hypothesis）**密切相关。根据该假说，语言习得过程中存在一个特殊的机会窗，在此期间习得的语言可达到母语水平（Lenneberg，1967）。关于关键期是否存在或时间长短等问题，目前学界还未达成一致的观点，不过很明显，无论是婴幼儿还是成人，其感知系统都具有高度的可塑性。在理想的二语学习状态下，即接触二语的时间非常早，习得者可以达到母语或接近母语的水平。沿这一主线所开展的研究构成了双语（bilingualism）研究的一部分。但是，研究表明，即便二语学习者能够区分很难的非母语语音对立，他们所使用的声学线索也可能与母语者有所不同（Underbakke 等，1988）。

对于稍大一点的儿童，足够的语言暴露可能会改变其 L1 的模式。研究表明，说西班牙语的儿童（8 ~ 10 岁和 14 ~ 16 岁）在美国生活的时间越长，他们的清浊边界就越接近英语使用者。但是，即便西班牙语儿童在美国生活三年以后，其清浊边界的 VOT 值还是比单语美国儿童的短很多（Williams，1979）。成人也是如此，针对某一个具体的较难掌握的 L2

音位的对立，短期的训练可以明显地改善感知的结果。让被试暴露在由多个发音人在不同语音环境下所发的目标语音中，通过 10 至 20 个小时的训练，被试对复杂语音对立的感知能力甚至发音能力就可以得到明显的改善（Logan 等，1991；Wang 等，1999）。

儿童和成人在语音范畴上都表现出了一些可塑性（malleability），而 2 至 4 个月大的婴儿则对很多细微的声学差异非常敏感，无论这些差异是否在其周围的语言中起到区分音位的作用。也就是说，婴儿可以范畴化地区分母语中的音位以及其他许多非母语的音位。但这种敏感性似乎会在一岁末的时候变弱。因此，在前 12 个月中，母语就像一个过滤器，强化母语中的语音对立，而弱化非母语的语音对立。例如，印地语和英语环境中的婴儿都可以区分齿音/t/和卷舌音/ʈ/。这两个音在印地语中是对立音位，而在英语中不对立。但是，以印地语为母语的成人可以继续区分这两个音，而以英语为母语的成人则无法区分。事实上，说英语的婴儿在 10～12 个月大时就不再能区分这两个音了（Werker 和 Tees，1984；见 Cheour 等，1998 通过 EEG 得出的相同结论）。

有关元音感知的研究结果显示，母语的影响早在年龄更小的时候就已经开始了。Kuhl（Grieser 和 Kuhl，1989；Kuhl，1991）关于“感知磁吸效应”（perceptual magnet effect）的研究表明，同将元音/i/的非原型范例与原型范例的变体进行区分相比，成人和婴儿都更难以将元音/i/的原型范例与周围声学空间中存在的原型范例的变体进行区分。尽管在这两种条件中，原型和非原型范例与原型范例的变体之间的听觉距离是相同的。根据 Kuhl 的说法，原型就像磁铁，它会吸引范畴内的其他成员，因而缩小了感知空间。之后的实验表明，磁吸效应是由周围环境的语言决定的。6 个月大的美国婴儿对英语中的原型/i/表现出磁吸效应，但对瑞典语中的原型/y/却没有。相反，6 个月大的瑞典婴儿对瑞典语中的原型/y/ 表现出磁吸效应，对英语中的原型/i/则没有（Kuhl 等，1992）。因此，对元音感知而言，语言经验的效应早在婴儿 6 个月大时就已经有所体现。

再来讨论成人，单语成人并不是无法区分所有的非母语音位。尽管有一些非母语的对立区分起来很难，但成人对其他一些非母语对立的区分能力却几乎可以达到母语水平。已有研究表明，L1 和 L2 语音清单之

间的关系可以为预测听音人对 L2 音位的感知方式提供主要依据。目前主要有两种 L2 语音范畴习得模型，一种是感知同化模型（perceptual assimilation model，PAM；Best，1994），一种是语音学习模型（speech learning model，SLM；Flege，1995）。Best 和 Tyler（2007）强调，感知同化模型是不熟悉目标语言的无经验的听音人对非母语的语音感知模型。对非母语对立的感知是通过比较它们与母语音系范畴的音姿的相似程度来实现的。根据对音素对立体中的语音所采用的感知同化方式，感知同化模型可以对区分能力的表现做出具体的预测。当两个非母语音素被感知为母语中的两个不同音位的范例时（“双范畴”同化），则区分度非常好；如果它们被感知为母语中同一音位的范例，且优度相当（“单范畴”同化），则区分度会很低；如果它们被感知为母语中同一音位的范例，但优度不同（“范畴优度”差异），则区分度居中。除了以上不同的同化程度以外，还有另一种可能，即对立体的一项或两项未被感知为母语中任何音位的范例（“未归类”）。感知同化模型预测显示，人们对“归类 - 未归类”对立音位的区分度非常好，但对两个未归类音素的区分则取决于每个音素映射至特定母语音位的程度。

日本的英语学习者对/ɹ/和/l/的习得是一个经典的例子。日语中有一个类似于/ɹ/的音（通常表现为卷舌拍音/ɽ/），但是没有/l/，日语母语的学习者很难区分这两个音（Strange，1992）。根据感知同化模型，这两个音被当作日语里/ɽ/范畴中的两个不典型的范例，属于单范畴同化，因而难以分辨。与此不同的是，英语母语者可准确地区分祖鲁语中对立的清边擦音/ɬ/和浊边擦音/ɮ/。英语中没有这两个擦音，这属于双范畴同化，英语母语者将祖鲁语中的/ɬ/和/ɮ/分别映射到了英语中的/s/和/l/上（Best 等，2001）。

语音学习模型特别关注的是对 L2 发音的最终习得，并假设 L2 的发音错误主要源于感知上的偏误。与感知同化模型一样，母语与 L2 的语音关系在非母语语音感知中起到十分重要的作用。与感知同化模型不同的是，语音学习模型并没有具体说明这种相似性的本质是声学上的还是发音上的。语音学习模型区分“陌生”音素与“相似”音素。L2 的陌生音素在母语中没有对应的发音，因此与 L1 中的任何音素都不同。例如，对母语为英语的法语学习者来说，法语中的前高圆唇元音［y］就是一个陌

生音素。L2 的相似音素与对应的 L1 中的音素相似但不等同。例如法语中的/t/，虽然它与英语中的/t/不完全相同，因它是齿音且不送气，但它明显与英语中的/t/非常相似。语音学模型预测，等同分类的过程会阻碍 L2 学习者对相似但不陌生的语音建立一个新的范畴。因此该模型预测，L2 学习者对相似音素的发音不会达到母语水平；但是，如果有足够的语言暴露，L2 学习者应该能够正确地产出陌生的音素。

Flege 及其同事的研究表明，人们对陌生音素和相似音素的习得确实存在差异（如 Flege，1987）。母语为英语的高水平法语学习者可以掌握法语中的陌生音素/y/，但却无法掌握相似音素/u/。有趣的是，不仅 L1 的范畴会影响 L2 的语音，对 L2 范畴的习得同样会影响 L1 语音的产出。例如，在美国生活了十年以上的法语母语者所产出的英语和法语/t/的 VOT 值都介于两种语言的母语标准值之间（Flege，1987）。

在感知同化模型和语音学习模型中，语音相似性是预测听音人如何对待 L2 语音的关键。但是，现在还是缺少一个可以独立于 L2 感知难度之外的能够明确判断语音相似性的实证评估方法（见 Strange，2007）。

总的来说，单语成年听音人对很多非母语的对立语音的感知存在困难。但是，感知的难易程度并不统一。成年听音人遇到的感知困难，不能只通过对比 L1 和 L2 的音位清单来进行预测。研究表明，该过程应该考虑其他的一些因素，包括标记音位对立的具体声学线索、L1 和 L2 中音位变体的分布情况、说话人的年龄、对 L2 的暴露程度、L2 的输入质量以及 L1 和 L2 的使用程度。尽管如此，感知系统也表现出了可观的可塑性，即便是成人，其语音范畴也可以发生改变。

13.5 总结

听话人从语音信号中提取信息的过程使他们可以理解说话人的信息。这是一个复杂的过程，而且人们目前还不能完全理解其工作机制。尽管我们对单个音段的声学线索已经了解得比较透彻，但至于这些线索究竟是变化的还是稳定的，现在仍存有争议。并且，人们关于言语感知的基本单位也还未达成一致。另外一个问题是，我们还不清楚言语感知究竟是基于一般的听觉机制，还是基于专门针对言语加工进化而来的机制。

最后，虽然感知系统表现出了很强的可塑性，但是母语影响着我们感知语音的方式。

练　习

1. 支持和反对外部规整的观点有哪些？外部规整与点规整和范围规整有什么不同？
2. 擦音发音部位的感知线索是什么？这些线索位于音节的什么位置？
3. 爆发音发音部位及发音方法的感知线索是什么？哪些线索是受语境制约的？请举例说明。
4. 言语感知的运动理论是什么？请给出支持和反对这种理论的证据。
5. 要验证范畴感知必须进行的两种实验是什么？用由“ba”至“pa”构成的连续统来说明这些实验的作用。范畴感知和连续感知有什么不同？
6. 哪些证据能证明范畴感知是人类语言的内在特征？这些证据是怎样收集的？与之相反的证据有哪些？
7. 感知同化模型和语音学习模型是什么？它们有什么不同？这些模型是否可以预测美国的韩语学习者难以分辨后圆唇元音［u］和后不圆唇元音［ɯ］？请给出你的解释。

注　释

1. 将附有相关信号部分的盒式音频磁带的两个相对应的片段互换并连接在一起。现今，这种拼接操作已经被数字化了，利用电脑的“剪切与粘贴”操作就可完成。

附 录 A

A.1 质量、力和压强

质量（mass）是由组成物体的物质所决定的，具体而言，是由组成物体的分子所决定的，例如液态分子和气态分子。一个物体的质量是恒定不变的，它不随物体位置的变化而变化。同样一个物体，无论是在地球上，还是在别的行星上，抑或是在零重力空间，它的质量都不会发生变化。质量是以**克（gram，g）**来衡量的。

一个物体的**重量（weight）**是它施加在支撑面上的力的大小。它不仅依赖于物体本身的质量，还受物体所在行星上的万有引力所影响。万有引力会使物体朝着行星的中心做加速运动。即使物体在受到一个支撑面的支撑作用的时候，这种加速度也依然存在。不过此时由于物体的运动被支撑面所“限制”，物体不会发生移动；结果，物体会对该表面产生力的作用。这个力就是物体的重量。一个物体在地球上的重量比在月球上的大，因为地球上的万有引力比月球上的要强。人们以**牛顿（newton，N）**来衡量力的大小，所以重量也以牛顿来计。1 牛顿（1 N）就是约 102 克的物质施加在支撑它的表面上的力的大小；这种力的大小与支撑面的大小无关。在地球上，1 kg 的物质的重量大约是 10 N。

在地球上，将一个物体的质量（按 kg 计）乘以地球的万有引力（9.81 m/s^2），就得到了该物体施加在支撑面上的力（按 N 计）。其中，9.81 m/s^2这个值表示了地球上万有引力的大小：当一个物体从飞机上竖直掉落时，它的速度每秒钟会增加 9.81 m/s（不考虑空气阻力的作用）。换句话说，这个物体以每秒 9.81 米每秒的速度加速：

$$\frac{9.81\ \mathrm{m/s}}{\mathrm{s}} = 9.81 \times \frac{\mathrm{m}}{\mathrm{s} \times \mathrm{s}} = 9.81\ \frac{\mathrm{m}}{\mathrm{s}^2}$$

在日常对话中，我们通常不注意质量（“物质的存在”）与重量（物质做加速运动而产生的力）的区别，因为人们一般不会往返于多个行星之间。例如，体重秤虽然真正测量的是人对体重秤施加的力，但是却错误地用 kg（也就是用质量的单位）来表示人的重量。（用磅来显示重量其实掩盖了物理量的本质，因为磅既可以表示质量又可以表示重量，这样更是加重了概念的混淆程度。）我们可以用一个秤验证这一点。虽然秤测量的是力的大小，但其刻度盘上显示的通常是质量的单位“kg”。如果我们把这个秤放到电梯上，然后我们站到秤上，那么当电梯加速上升的时候，秤显示的力要比电梯开始下降的时候大。换句话说，人产生的重量（作为质量的函数，以牛顿计）随着电梯的上升或下降而发生改变。但实际上人的质量是不随电梯运行的方向而改变的。因此，体重秤上所标记的测量单位“kg”是错误的，因为它实际上测量的是以牛顿（N）计的力，而不是以千克（kg）计的质量。然而，由于我们很少在电梯上或者是其他行星上使用体重秤，所以人们通常不会被注意到这一错误。

压强（pressure）是施加在一个给定表面上的力。如果一个较大力施加在一个较小的表面上，压强就较大；如果同样的力施加在一个较大的表面上，压强就要小一些。这种效应可以用一个圆锥来说明（见图 A.1）。如果圆锥以底面放置，那么它的重量就会被分散到较大的表面上，因此圆锥施加在该表面的压强就较小（见图 A.1a）。然而如果把圆锥上下颠倒，让它尖的一头接触支撑面，它的全部重量会施加在这块较小的区域，因此圆锥施加在该表面的压强就较大（见图 A.1b）。在这两种情况下，圆锥体是相同的，具有相同的质量和重量，对于支撑面所施加的力也是相同的。因此，只是支撑面大小的差异造成了压强的不同。压强是以**帕斯卡（pascal，Pa）**来衡量的。

一帕斯卡是一牛顿（N）的力作用在一平方米的表面上所产生的压强（1 Pa = 1 $\mathrm{N/m^2} \approx 0.0209$ lbs/sqft ≈ 0.000145 psi）。因此，“大气压强”这一术语指的并不是大气的重量（也就是，所有大气的质量对地球表面所施加的推力），而是大气对单位面积表面所施加的力。换句话说，每一个大气分子都有给定的质量；存在于这些大气分子的质量和地球之间的万

有引力使这些分子各自都具有了一定的重量。这个重量使分子对地球表面一个很小的部分施加力的作用，这个施加在特定表面范围内的力也就对该表面施加了压强。

下面这个例子可以帮助我们厘清压强、力和质量之间的关系。假设有一本质量为 760 g 的书，它的尺寸为 12 cm × 20 cm × 4 cm（见图 A.1c），若将其平放在桌子上（因此，它的支撑面是 12 cm × 20 cm），它对桌子施加的压强是多大？

$$\text{压强[Pa]} = \frac{\text{力[N]}}{\text{面积[m}^2\text{]}} = \frac{\text{质量[kg]} \times \text{地球加速度[m/s}^2\text{]}}{\text{宽[m]} \times \text{高[m]}}$$

$$= \frac{0.760\ \text{kg} \times 9.81\ \text{m/s}^2}{0.12\ \text{m} \times 0.2\ \text{m}} \approx \frac{7.5\ \text{N}}{0.024\ \text{m}^2} \approx 310.6\ \text{Pa}$$

如果将这本书立在桌面上（见图 A.1d），它的支撑面就减小为 12 cm × 4 cm 了。此时书对桌子施加的压强是：

$$\text{压强[Pa]} = \frac{\text{力[N]}}{\text{面积[m}^2\text{]}} = \frac{0.760\ \text{kg} \times 9.81\ \text{m/s}^2}{0.12\ \text{m} \times 0.04\ \text{m}}$$

$$\approx \frac{7.5\ \text{N}}{0.0048\ \text{m}^2} \approx 1553\ \text{Pa}$$

由此可以看出，如果同样大小的力（这里都是相同的重量 7.5 N）作用在一个较小的表面上（书立在桌面上），力对该表面所施加的压强就大一些。如果同样的力作用在较大的表面上（书平放在桌面上），压强就小一些。该原理对于理解中耳的工作机制（见 12.2.1）十分重要。

总结而言，压强的大小（以帕斯卡计）依赖于力（以牛顿计）和表面的面积（以平方米计）；力的大小则由质量（以千克计）和行星的万有引力共同决定。

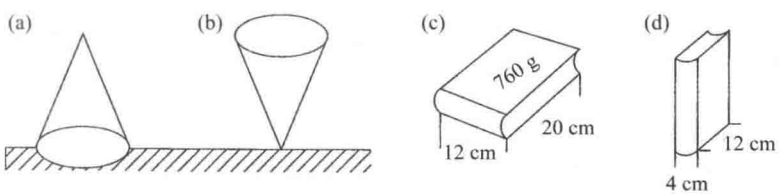


图 A.1　相同重量的物体所施加的压强随支撑面面积的变化而变化。

A.2　能量、功率和强度

能量、功率和强度的概念是紧密相关的，所以人们经常将它们相互混淆。此外，人们在日常生活中对这些术语的使用有时与它们在物理上的定义并不一致。

能量（energy）是一种基本属性。它可以被转化成多种不同的形式，比如说热能、电能、动能（即运动的能量）、势能（即位置的能量），也可以通过波进行传播，但是能量不能凭空地“产生”或“消失”。例如，在将电能由放大器通过电缆传至扬声器的过程中，电能会被转化为热能（即温度）和声能（即声波）。电能向热能的转化是一种我们不希望看到的效应。然而，传递给扬声器的能量中大约只有1%的能量被转化为声能，其余的都被转化为热能。由于这一过程会产生如此多的热能，当电能过多时，扬声器就会被“烧毁”。尽管如此，电能可以转化为声学能量。因此，声波是携带能量的。虽然振动的空气分子的动能在能量的传输中起到了一定的作用，但声波并没有在空间中传输任何的物质，这一点在7.1.2中有所阐述。

能量是物质或波的属性[1]。能量的测量单位是**焦耳（joule，J）**。该单位可以由千克（kg），米（m）和秒（s）这三个单位推导得出[2]：

$$\begin{aligned}\text{能量}[\mathrm{J}] &= \text{质量}[\mathrm{kg}] \times \text{速度}^2[\mathrm{m^2/s^2}] \\ &= \text{质量}[\mathrm{kg}] \times \text{加速度}[\mathrm{m/s^2}] \times \text{距离}[\mathrm{m}] \\ &= \text{力}[\mathrm{N}] \times \text{距离}[\mathrm{m}]\end{aligned}$$

能量的单位，以及它与力和距离的关系并不十分直观。让我们来考虑下面这个例子。一个平放在地面上的物体比它平放在桌面上时所具有的势能要小一些。在这两种摆放位置下，该物体的重量（以牛顿，N计）几乎是相同的，都是它的质量（以kg计）乘以地球上的万有引力。事实上，物体放在桌面上时的重量要稍微小一些，因为此时该物体离地心更远一些。确切地说，位于地面上的物体与位于桌面上的物体之间的主要差异是按米计的高度。势能的计算方法很简单：用物体的重力（以牛顿$=\mathrm{kg}\times\mathrm{m/s^2}$计）乘以高度（以米计）即得。这恰好就是上面所提到的公式：能量=力×距离（或者，能量=质量×速度^2）。

与质量一样，能量也是一个与时间无关的维度。一定量的物质，例如一升的煤油，会拥有一定量的能量。该能量可以以较慢的速度释放，例如煤油灯中的煤油可以用好几个小时，但也可以以很快的速度释放，例如若煤油爆炸，则该过程不到一秒。两种情况下，一升煤油中蕴含的能量完全相同——唯一的不同在于单位时间内所释放的能量的多少。这个表征“单位时间内能量”的维度就是**功率（power）**。如果在短时间内转化了大量的能量，就对应较大的功率；如果用较长时间来转化同样多的能量，那么其功率就较低。

下面的情境将有助于我们理解这一概念。想象有一个人，把 1000 kg 的物体从一层搬到房子的四层。由于该物体在四层的势能比它在一层时要大，这一过程需要一定的能量才可实现。需要的这部分能量只取决于高度、地球的万有引力以及物体的质量。如果用一周的时间去搬运这 1000 kg 的物体，就不需要很大的功率，因为能量会被分散到很长的时间中。但是如果在一个小时内将同样质量的物体搬到房子的四层，就需要很大的功率。功率以**瓦特（Watt，W）**来度量，表示的是单位时间内的能量：

$$\text{功率}[\text{W}]=\frac{\text{能量}[\text{J}]}{\text{时间}[\text{s}]}=\frac{\text{力}[\text{N}]\times\text{距离}[\text{m}]}{\text{时间}[\text{s}]}$$

值得注意的一点是，瓦特这个单位被人们用来描述其他形式的能量。例如电功率，就是用单位瓦特来度量的，就像下面这个电灯泡的例子。与一个 50 W 的灯泡相比，一个 100 W 的灯泡可以将 2 倍的电能转化成光能和热能。（需要注意的是，我们日常所说的“一个灯泡耗能 100 W”从物理学的角度上来说是不正确的，因为能量不能被消耗；这 100 W 的能量被转化成了光能和热能。）为了计算在给定时长内转化的总能量，需要将功率（以瓦特计）与时长（以小时或秒计）相乘。这就引出了“千瓦时”（kWh）这个度量能量的单位：

$$\text{功率}[\text{W}]\times\text{时间}[\text{s}]=\frac{\text{能量}[\text{J}]\times\text{时间}[\text{s}]}{\text{时间}[\text{s}]}=\text{能量}[\text{J}]$$

“1 kWh”只是“360 万焦耳”的另一种写法：1 kWh = 1000 Wh = 1000 × 3600 Ws = 3600000 J = 3.6 MJ。

想象一个位于房间中某一点处的声源。理想情况下，这个“0 维”

的扬声器发射出的声音的功率在各个方向上的传播速度是相同的。某种程度上讲，声波的传播，就像一个球体表面渐渐远离扬声器（球心）的过程。球体与扬声器的距离越远，它的表面就越大。离声源中心越远，声波的**强度（intensity）**就越弱。图 A. 2 中展示了这一点。在离扬声器距离为 r 处，声音遇到了用边长为 a 的正方形代表的表面。该表面的面积是 $a \times a = a^2$。如果距离加倍，变为 $2 \times r$，则正方形的边长也加倍变成 $2 \times a$，这意味着面积变为了原来的 4 倍：$(2 \times a) \times (2 \times a) = 4 \times a^2$。换句话说，距离加倍对应着表面的面积变为原来的 4 倍；也就是说，表面的面积以距离增长的平方增加。对于一个固定大小的表面，这就意味着声音的强度（在给定时间内到达表面的能量）以表面到声源距离的平方下降。在距离为 $2 \times r$ 处，影响面积为 $a \times a = a^2$ 的表面的功率只有在距离为 r 的地方影响同样表面面积 a^2 的强度的四分之一。强度这个维度并没有自己的单位；它可以简单地表示为功率（以瓦特计）每表面（以平方米计）：

$$\text{强度} = \frac{\text{功率}[\mathrm{W}]}{\text{面积}[\mathrm{m}^2]} = \frac{\text{力}[\mathrm{N}]}{\text{距离}[\mathrm{m}] \times \text{时间}[\mathrm{s}]} = \frac{\text{质量}[\mathrm{kg}]}{\text{时间}^3[\mathrm{s}^3]}$$

理论上讲，声波可以传播无限远，虽然那样它的强度会变得无限小。然而事实上，声波的声能会持续不断地转化为热能，所以经过一段时间以后，声波完全消失，因为它的能量全部转化为了热能。同样的原理也被应用到隔音中：声能并没有被“消灭”，而是转化成了热能。因为声音中所包含的能量很小，所以吸声材料并不会因此变热。

总结而言，能量被包含于物质或者波中。能量可以转化成不同的形式，但却不会消失。功率反映了在一段时间内出现的能量的多少。强度是影响一个表面的功率。

需要注意的是，声波的能量是很小的。音响系统的放大器大约需要

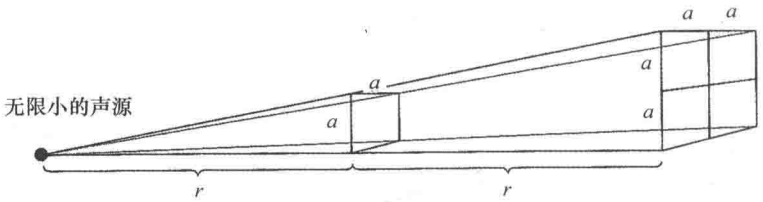


图 A. 2　表面的面积随距离的平方而增加。

产生 1 瓦特的电功率，才能使声音的响度满足客厅的需求。由于扬声器所产生的能量只有大约 1% 被转化成了声能，在客厅中的声音的功率大约只有10 毫瓦。即使在一个喧闹的迪厅中，声音的功率也只能点亮一盏小台灯。

A.3　分贝

一个对声音非常重要的度量标准（或者更准确地说是计算规则）是分贝（dB）。分贝在这本书中经常出现。当你购买高保真音响设备或者洗碗机的时候，或者在你关注噪声污染的讨论时，总会出现“dB”这一术语以表示声学信号的响度，或者衡量设备处理声学信号的能力。在后一种情况下经常会引入信噪比（S/N）这个术语，它用分贝来表示。这一术语可以很好地说明分贝所表示的是两个数值之间的关系：具体来说，是信号和噪声之间的关系。然而，关于这一量度的由来，以及对它的“行为”的理解，都比较困难。由于分贝的概念不是很容易掌握，我们将对其进行详细的介绍。

我们将会按几个小的步骤讨论这个概念，而在每一个步骤中，我们将会用一些例子来作详细说明。在后文中会出现很多的公式，它们不应该分散我们的注意力，而是应该为我们更好地理解这一部分的内容服务。

A.3.1　均方根振幅

振幅是一种粗略地描述信号响度的量度。图 A.3a 所示的信号仅仅在一个时刻达到其最大的振幅，在其他时刻的位移量都非常小（关于振幅的定义请见 7.3.2）。虽然图 A.3b 所示信号与图 A.3a 所示的信号具有相同的振幅，但该信号在很长的一段时间内都维持在最大振幅附近。所以平均来说，图 A.3a 所示信号不如图 A.3b 所示信号的响度大。因此我们需要另外一种方法来描述信号的响度，以取代测量最大振幅这种方法。

一种描述信号响度的方法是将信号在波形图中显示的所有位移相加。然而，这种方法是行不通的。平均来说，一个信号上下摆动的幅度是很接近的，因为每当空气密度达到最大值之后，都会出现一个与之对应的最小值。将这些正负位移量简单地相加就总会得到一个趋近于零的和。

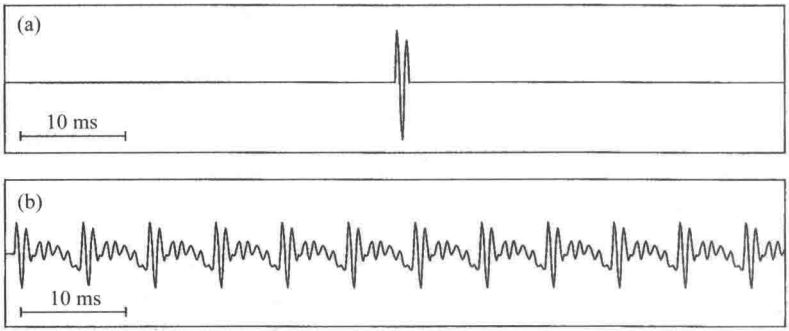


图 A.3　振幅相同的两个信号

我们需要考虑的第二点是这样一个事实：把一个空气分子稍稍挪离其静止位置是很容易的，但是如果想让它远离静止位置，就需要更多的能量。这一事实必须反映在计算响度的方法中。解决以上两方面问题的一个比较简单的方法是：（1）把所有正负幅值相加，并使其和不为零；（2）将偏离值平方，使偏离量较大的部分获得较大的值。这就是简单地对每一个气压值做自乘[3]。

以下两个例子可以说明取平方的过程。如果声波某一点上的气压（相对于周围环境的大气压强而言）是 -10 Pa，也就是说，位移量是 -10 Pa，那么平方后的值就是 $(-10\ \text{Pa}) \times (-10\ \text{Pa}) = 100\ \text{Pa}^2$。如果气压是 0.01 Pa = 10 mPa，平方以后的值就是 $0.01\ \text{Pa} \times 0.01\ \text{Pa} = 0.0001\ \text{Pa}^2 = 0.1\ \text{mPa}^2 = 100\ \mu\text{Pa}^2$。

将所有平方以后的位移值相加就得到了一个与信号能量相关的量度。越多的位移被考虑进来，相加所得的和就越大。所以说，这个和必须被参与相加的位移的数目除——换句话说，最后计算的是“平均和”即“算术平均值”。而由于每一个位移量已经被平方了，我们需要对得到的计算结果做开方来进行修正。最后，我们就得到了“平方数的平均值的根”即“均方根”（RMS）：

$$\text{均方根振幅} = \sqrt{\frac{\text{所有位移的平方和}}{\text{位移的数目}}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

（x_i 是每一个独立的位移值，n 是位移的数目；符号 $\sum$［求和符号］和

x_i[表示变量]的概念将会在附录 B.2 中作进一步讨论。)

下面这个例子(见图 A.4 和与之相关的计算)可以帮助我们理解这个公式。这是一个十分简单的信号,可以用 10 个位移值(x_1, x_2, …, x_{10})表示。

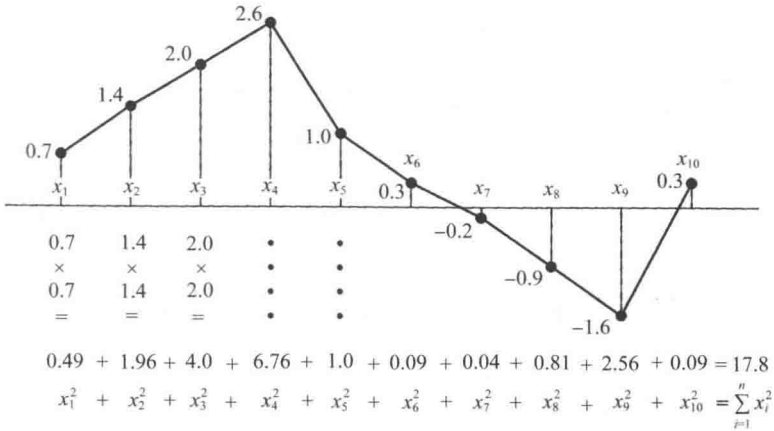


图 A.4　对一个有 10 个采样点的信号的平方和的计算

把这 10 个值代入公式中(平方和的值是 17.8, 10 是数值的数目),得到:

$$\sqrt{\frac{\sum_{i=1}^{10} x_i^2}{10}} = \sqrt{\frac{17.8\ \text{Pa}^2}{10}} = \sqrt{1.78\text{Pa}^2} \approx 1.3\ \text{Pa}$$

在这个例子中,得到的均方根振幅("平均幅值")大约为 1.3 帕斯卡。

由于位移值被平方了,较大的位移比较小的位移更加突显。如果两个信号的振幅相等,那么拥有更多大位移值的信号就会拥有更大的均方根振幅。要注意,拥有很多大位移值的信号含有高频分量,因为高频意味着某一现象发生得很频繁。下面这个例子可以很好地阐明这一点。一个振幅为 10,拥有 8 个位移值 0, 5, 10, 5, 0, −5, −10 和 −5 的信号,其均方根振幅为(见图 A.5a):

$$\sqrt{\frac{0^2+5^2+10^2+5^2+0^2+(-5)^2+(-10)^2+(-5)^2}{8}}$$

$$=\sqrt{\frac{0+25+100+25+0+25+100+25}{8}}=\sqrt{\frac{300}{8}}\approx 6.1$$

另一个有着同样振幅（10）的信号，其位移值分别是 0，10，0，-10，0，10，0 和 -10，则其均方根振幅为（见图 A.5b）：

$$\sqrt{\frac{0^2+10^2+0^2+(-10)^2+0^2+10^2+0^2+(-10)^2}{8}}$$

$$=\sqrt{\frac{0+100+0+100+0+100+0+100}{8}}=\sqrt{\frac{400}{8}}\approx 7.1$$

因此，具有高频分量的信号就比振幅相同但具有较低频率分量的信号拥有更大的均方根振幅。乍一看这个论断有些奇怪，但其原因却非常明了：如果一个信号中经常出现大的位移值（也就是说频率高），那么很明显需要更多的能量将空气分子挪离其静止位置并再挪回去。也就是说，两个信号相比，达到同一位移更为频繁的信号具有更大的均方根振幅。

均方根振幅为我们提供了一种比较信号平均振幅的度量方法。例如，图 A.3a 中信号的均方根振幅为 0.25 Pa，而图 A.3b 中信号的均方根振幅为 1.10 Pa——虽然这两个信号的最大振幅都是 2.70 Pa。

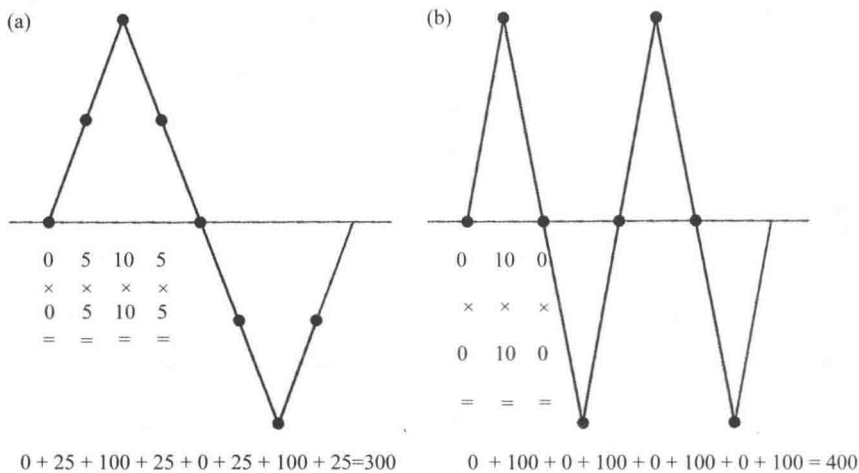


图 A.5　（a）低频信号的平方和计算与（b）高频信号的平方和计算

A.3.2　均方根振幅与响度

均方根振幅是一个粗略衡量响度的量度[4]。然而，Weber 和 Fechner 早在 19 世纪建立的观点认为，人们感知到的通常是相对变化而非绝对变化。这一现象与知觉感知的特点非常相似。例如，2 kg 的重量似乎比 1 kg 要重很多，然而同样的 1 kg 的重量差在我们试图提举 51 kg 和 50 kg 的物体时却根本感觉不到（或者只能感觉到很细微的差别）。

这一现象对响度也同样适用：在一个安静的环境中，即使是很轻微的气压变化我们都能感觉到，但是在一个嘈杂的屋子中，需要很大的压强变化才能使我们有相同的感觉。例如，人们在均方根振幅为 0.4 mPa 的信号中感知到的气压变化比均方根振幅为 0.2 mPa 的信号中要响两倍。然而，对于均方根振幅为 20.4 mPa 和 20.2 mPa 的信号，我们却感知不到二者的差异。换句话说，我们感知不到绝对的差异，而只能感知到相对的差别。

相对的量度通常以比例表示：2 kg 和 1 kg 相比是 2 kg/1 kg = 2；因此 2 kg 就是 1 kg 的两倍重。然而，51 kg 的重量和 50 kg 相比是 51 kg/50 kg = 1.02，因此 51 kg 几乎不比 50 kg 重。同样的道理，一个均方根振幅为 0.4 mPa 的信号和均方根振幅为 0.2 mPa 的信号相比是 0.4 mPa/0.2 mPa = 2。而一个均方根振幅为 20.4 mPa 的信号和均方根振幅为 20.2 mPa 的信号相比是 20.4 mPa/20.2 mPa ≈ 1.01。因此它们的响度几乎没有差异。虽然从绝对量上来看，这两种情况下的差异是相同的（0.2 mPa），但比例反映了我们感知到的两种声音在响度上的差异。

当使用比例时，量度的单位不会出现，比值不会表明所使用的是什么单位（例如：磅、帕斯卡、微帕、厘米、英寸、伏特或者是其他任何单位）。注意到在上文的比例关系中，单位 kg 和 Pa 都消失了。比例就是一个没有任何特定度量单位的数字。这一事实在后续的应用中非常有用。

要想达到相同程度的感知效果，与小数值的情况相比，较大的数值需要产生更大的变化。现在我们考虑将一个重量乘以 10（例如从0.1 kg 变为 1 kg），然后再乘以 10（从 1 kg 变为 10 kg）。做两次这样的乘法的效果让重量变为了原来的 100 倍（从 0.1 kg 到 10 kg）而不是 20 倍。如果将这个重量值再乘以 10（从 10 kg 变为 100 kg），那么这三次乘法就让

原来的重量乘了1000倍，以此类推。具体总结如下：

	重量	系数	乘以10的次数
100克 =	0.1 kg	1	0
	1 kg	10	1
	10 kg	100	2
	100 kg	1000	3
	1000 kg	10000	4

虽然这个表格中的重量只被乘了4次，但它已经从0.1 kg增加到了1000 kg，也就是说，它被放大的系数是1万。与此同时，我们对这一变化的感觉是它的重量只产生了4倍的变化（如果我们可以举起1000 kg的重量）。

这一效应同样适用于响度。下面的表格包含了可以使听音人对响度的感知以系数10增长的气压变化。第一行的信息是人类在理想的安静条件下可以感知到的最低的响度值（20 μPa）；最后一行是人类可以容忍的最大的响度值。这样就得到了下表：

大气压强	系数	乘以10的次数
20 μPa = 0.00002 Pa	1	0
200 μPa = 0.0002 Pa	10	1
2 mPa = 0.002 Pa	100	2
20 mPa = 0.02 Pa	1000	3
200 mPa = 0.2 Pa	10000	4
2 Pa	100000	5
20 Pa	1000000	6

因此，如果一个听音人有6次感到声音的响度变为了原来的10倍，那么气压的振幅实际上已经变为原来的100万倍了。

如前文所述，影响特定感知的因素不是它的绝对数值，而是它与初始值对比的相对值。与只测量以帕斯卡计量的气压的绝对数值相比，更有效的一种方法是计算其与初始值的关系，或者说，计算从初始值变为当前值所需要乘以10的次数。后面的这种“计算乘法次数”的原理，恰好就是对数函数的功能。例如，在计算器中输入1000，然后按下“log”键，就会得到3，表示10需要乘3次才能得到1000[5]。

我们现在已经介绍了计算信号响度需要的所有因素。为了得到响度

值，我们需要计算信号的振幅与可感知的最小振幅的比值，然后对得到的值取对数：

$$\log\left(\frac{\text{均方根振幅}}{\text{可感知到的最小均方根振幅}}\right)$$

这个计算结果的单位是贝尔（Bel）。这一单位重点表示了为了得到分子（分数线上面的数），分母（分数线下面的数）需要乘以 10 的次数。因为这个结果的值通常非常小（6 贝尔是人耳可以容忍的最大响度与可以感知到的最小响度的比值），得到的结果总是做乘以 10 的处理。于是得到的单位就是“十分之一贝尔”，因此被称作分贝（dB）（也就是说，6 贝尔就是 60 分贝）。

在这一阶段，我们必须考虑一个已经在描述均方根振幅概念的时候提到的效应：较大的大气压变化比较小的变化更为显著。由于这个原因，当我们计算响度的时候要将振幅值平方，对 dB 值的计算也是如此。除此之外，分贝值的计算也可以不以“可感知到的最小均方根振幅”20 μPa 为参考而采用其他的参考值。理论上讲，任何振幅值均可作为参考值；这样当然会使计算得到的 dB 值发生变化，但它并不影响公式的本质。在语音学中，计算响度的最常见的公式是：

$$10\times\log\left(\frac{\text{均方根振幅}}{\text{均方根参考振幅}}\right)^2=20\times\log\left(\frac{\text{均方根振幅}}{\text{均方根参考振幅}}\right)$$

注意到这个公式只是包含了一个叫作“均方根参考振幅”的值作为普适参考值。这意味着事实上上述计算无论是基于声音的压强（以帕斯卡计），麦克风的输出（以伏特计），波形图中的距离（以厘米或英寸计），抑或是其他任何维度的任何度量单位，都是不重要的，因为度量单位在计算分贝值的过程中被抵消掉了。如果我们通过上述方法来计算声压值，并且将参考的幅度值设为可感知的最小均方根振幅（20 μPa），那么这个公式就可以表示声压级（SPL），用 dB_{SPL} 来表示。

当我们使用 dB 值的时候，必须指明分贝的种类，因为分贝只表示一个比值的对数值乘以 10 以后的结果。例如，对于声压级来说，我们应指明是 dB_{SPL}（分贝，声压级），或者对于均方根振幅来说，应指明是 dB_{RMS}（分贝，均方根振幅），或者对于将人耳听觉系统对高振幅和低振幅敏感程度的差异考虑在内的，表示与人对响度的主观印象相对应的强度来说，

指明是 dB_A。

同时存在不同的 dB 值标度似乎有些让人迷惑，然而这一特点有一个重要的优势。由于 dB 值表示的是两个值之间的比值，而这两个值的度量单位在计算过程中消失了（见上文），我们很容易比较从不同物理领域所获得的 dB 值。例如，声压级是空气压强变化的比值，然而在麦克风中，气压被转化成了电压（伏特）或者电流（安培）。通过计算 dB 值，我们便可以比较麦克风和空气压强的数据，虽然它们所表示的是完全不同的物理维度。在图形表示中，数值都以厘米或英寸来衡量；若使用 dB 标度，这些值便也可以直接和空气压强的差异进行比较了。

dB 标度的一个重要的优势是它利用了比值而不是绝对数值。一个人在用录音机录音时要先通过调节旋钮来选择合适的录音级别。然后在回放时，这个人会调节音量控制旋钮以获得舒服的听音级别。设置录音级别和回放级别的过程当然会改变响度的绝对数值，但是这并没有改变大声和小声段落之间的相对关系。因此，dB 标度可以让我们在完全不知道录音和回放的级别设置的情况下，计算出信号的相对振幅特性。

最后，很重要的一点是，与“米”是“长度”的单位不同，“dB”并不是一个度量单位，它是一种计算的约定。然而，就像“长度”可以用不同的单位（例如米、英寸、英里）表示一样，比例值也可以表示为不同形式的 dB 值，例如 dB_{SPL} 或 dB_{RMS}。

A.3.3　用 dB 值进行计算

推导 dB 标度的方法有一些复杂，因为它表示的是相对于参考值的比例，并且在计算公式中应用了对数。用 dB 值进行计算的时候，有一些与像“米”或者“英寸”这样的单位不同的地方。这一部分的内容展示了在用 dB 值进行计算时需要特别注意的两个方面。

第一，因为计算是基于比例的，所以不存在绝对的零点。这就意味着像“声压级是 0 dB_{SPL}”这样的说法并不表示没有声音的压强（声音的压强在任何时候都是以 1013 hPa 的大气压强为基础的）。这句话只是意味着声音的压强等于可感知到的最小压强 20 μPa，但它并不是 0 μPa[7]。这一点与用摄氏度或华氏度测量温度的道理是类似的。“温度是 0 ℃”并不表示没有温度，而只是意味着温度等于某一个给定的参考温度。

像温度值一样，dB 值也可以是负数。也就是说，如果声音的压强比参考值低的话，它与参考值的比值就小于 1，则其对应的对数值就是负的。例如：

$$20\times\log\left(\frac{10\ \mu\text{Pa}}{20\ \mu\text{Pa}}\right)=20\times\log(0.5)\,\text{dB}_{\text{SPL}}$$

$$\approx 20\times(-0.3010)\,\text{dB}_{\text{SPL}}\approx -6\ \text{dB}_{\text{SPL}}$$

第二，如果声音的压强变成原来的两倍高，这并不意味着对应的 dB 值也要乘以 2。事实上，把声音压强加倍会使对应的声压级增加大约 6 dB[8]：

$$20\times\log\left(\frac{\text{两倍的声音压强}}{\text{给定的声音压强}}\right)=20\times\log\left(\frac{2\times\text{给定的声音压强}}{\text{给定的声音压强}}\right)$$

$$=20\times\log(2)\,\text{dB}\approx 20\times 0.3010\ \text{dB}\approx 6\ \text{dB}$$

因此，若声音压强加倍，则对应的声压级就增加了约 6 dB。如果声音压强变成了原来的 4 倍，对应的声压级就增加大约 12 dB；如果它变为原来的 8 倍，声压级就增加大约 18 dB，以此类推。每一次声音压强加倍，对应的声压级就增加约 6 dB。对于声音压强降低的情形，声压级相应的变化规律也是成立的：

$$20\times\log\left(\frac{\text{减半的声音压强}}{\text{给定的声音压强}}\right)=20\times\log\left(\frac{0.5\times\text{给定的声音压强}}{\text{给定的声音压强}}\right)$$

$$=20\times\log(0.5)\,\text{dB}\approx 20\times(-0.3010)\,\text{dB}\approx -6\ \text{dB}$$

于是，若声音压强减半，则对应的声压级就减少大约 6 dB；若声音压强减少到原来的四分之一，则对应的声压级就减少大约 12 dB，以此类推。

另外一个与“正常”的直觉相反的现象是两个信号相加时 dB 值的变化。当我们将信号相加的时候，相位扮演了关键的角色。一方面，如果两个信号同相位，并且声压级同为 0 dB（也就是和参考值相同），那么这两个信号相加所得的信号的声压级就是 6 dB。另一方面，两个反相位的信号相加则会完全相互抵消。即使它们各自的声压级是 70 dB，它们相加所得之和的声压级却是“无穷”，因为它们相加起来（会相互抵消）得到的声音压强的值是“0”，而事实上“0”的对数没有定义。

通常来讲，当我们用 dB 值进行计算的时候，如果可以将计算过程想象为是“信号”而不是单纯的数值在相加，那么将有助于我们理解上述特点。下面是最后一个例子。我们考虑两个相位相同的信号，第一个信号的声音压强是 70 dB，第二个是 58 dB。将两者相加以后，会得到一个

声压值小于 72 dB 的声音信号，正如下面公式中的推导所示：

$$70\ \text{dB} = 20 \times \log(3162)\text{；}\ 58\ \text{dB} = 20 \times \log(794)\text{；}$$

$$20 \times \log(3162 + 794) = 20 \times \log(3956) \approx 72\ \text{dB}$$

正如前文所述，dB 标度的推导和计算并不是十分简单，可能不是每个人都能马上理解。如果确实没有理解，不用着急，因为虽然 dB 这一标度在语音学中的应用十分广泛，但是理解 dB 标度对于理解这本书而言并不是必要的。我们只需要记住 dB 标度是一个普适的标度，它可以用来方便地表示振幅的差异。在全书中用到 dB 标度的地方，增加 6 dB 总是意味着声音压强变为了原来的 2 倍。

注　释

1. 携带能量的是声波本身，而不是参与声波传输的分子。这一点在电磁波中是显而易见的，比方说光，即使在通过真空这种没有任何物质的介质时也可以携带能量。
2. 通常“卡路里”是度量能量的单位，1 cal≈4.19 J。
3. 对位移值取平方的一个更有说服力的原因是，振动分子的速度事实上受到声波的振幅和频率这两者的影响。这一点在 7.3.2 中已经有所提及。
4. 与其他任何感知量一样，响度也是一种主观的印象，不能通过物理测量直接推导得出（见 12.6 节）。然而，主观印象是可以从物理值中估算出来的，我们在这一章中就是这样做的。
5. 当计算器的面板中出现像“6.9077…”这样的值时，则其采用的是所谓的“自然”对数，而不是在这里使用的以 10 为底数的对数。为了把一个自然对数值转化成以 10 为底数的对数值，需要将这个数乘以 0.4343：$6.9077 \times 0.4343 \approx 3$。
6. 不必对计算幂值和对数值的数学原理进行深入探讨，我们只需了解 $(a^2/b^2) = (a/b)^2$，以及 $\log(a^2) = 2 \times \log(a)$ 即可。
7. $20 \times \log(20\ \mu\text{Pa}/20\ \mu\text{Pa}) = 20 \times \log(1)\,\text{dB}_{\text{SPL}} = 20 \times 0\ \text{dB}_{\text{SPL}} = 0\ \text{dB}_{\text{SPL}}$。
8. 在强度的 dB 标度中，强度加倍意味着增加了 3 dB。

附 录 B

B.1 物理学术语

在物理学中，人们会将被测量的维度和描述某个维度的度量单位加以区分。

维度（dimension），比如说速度、距离或时间，都是用拉丁语单词的缩写来表示的。例如，**c** 是“celeritas”（速度）的缩写，**v** 是“velocitas”（速度矢量，有方向的速度）的缩写，**l** 是“longitudo”（距离）的缩写，**s** 是“spatium”（位移矢量，有方向的距离）的缩写，以及 **t** 是 tempus（时间）的缩写。

这些缩写应该与度量这些维度的**单位（unit）**区分开。这些单位通常用方括号括起来：**m** 表示米，**s** 表示秒，**Hz** 表示赫兹，等等。因此距离 **l** 就以米［m］来度量，时间 **t** 用秒［s］来度量。表 B.1 列出了一些重要的维度和它们所对应的测量单位。

为了避免出现太多的零，人们经常会采用比例系数来表示数值。这些系数都是希腊或意大利语单词，放在单位之前。比如说，“1000”就可以用字母 **k**（“kilo”）来指示。我们可以用“1 kg”来替代“1000 g”这种写法。最常用的比例系数都在表 B.2 中列出。需要注意的是“billion”在传统的英式英语和美式英语中的意思是不同的。

需要强调的一点是，在计算机术语中，像“kilo”，“mega”，“giga”或者“tera”等单位并不是以千（1000）为基数的，而是以计算机的“自然”倍数 1024 为基数[1]。因为这些单位与千基本上是相等的，所以我们采用相同的希腊字母和单词来表示它们——但是它们以大写字母来表示：在计算机术语中，**K** 代表“kilo”（$= 1024 = 2^{10} \approx 10^3$），**M** 代表“mega”

（$=1024\times1024=2^{20}\approx10^{6}$），**G** 代表“giga”（$=2^{30}\approx10^{9}$），**T** 代表“tera”（$=2^{40}\approx10^{12}$）。其中，“m”这个字母可能会导致一些混淆：小写字母 m 表示“毫”（千分之一），而大写字母 M 既可以表示“百万”（“mega-”，1 百万 = 1000000）也可以表示基于计算机的“兆”（“Mega-”，$2^{20}=1048576$）。

举几个例子：

（1）距离 **l** 可以用千分之一米来表示（毫米［**mm**］）。

（2）压强用 **p** 来表示；**1 μPa** 是一百万分之一帕斯卡。

（3）一秒钟振动一千次的频率 **f** 对应 **1 kHz**。

表 B.1　物理单位：名称、推导、维度、符号及其单位

维度		符号	单位	公式
距离		l	米	[m]
质量		m	千克	[kg]
时间		t	秒	[s]
面积		$A=l^2$	平方米	$[m^2]$
频率	每秒的周期数	$f=\frac{1}{t}$	赫兹	$[Hz]=\frac{1}{[s]}$
速度	单位时间的距离	$c=\frac{l}{t}$	米每秒	$\frac{[m]}{[s]}$
加速度	单位时间的速度变化	$a=\frac{\Delta c}{t}$	米每平方秒	$\frac{[m]}{[s^2]}$
力		$F=m\times a^2$	牛顿	$[N]=\frac{[kg\times m]}{[s^2]}$
压强	单位面积上的力	$p=\frac{F}{A}$	帕斯卡	$[Pa]=\frac{[kg]}{[m\times s^2]}$
能量（功）	力乘以距离	$E=m\times c^2$ $=P\times t$ $=F\times l$	焦耳 瓦特秒 牛顿米	$[J]=[Ws]$ $=[Nm]$ $=\frac{[kg\times m^2]}{[s^2]}$
功率	单位时间做的功	$P=\frac{E}{t}$	瓦特	$[W]=\frac{[J]}{[s]}$ $=\frac{[kg\times m^2]}{[s^3]}$
强度	单位面积上的功率	$I=\frac{P}{A}$	瓦特每平方米	$\left[\frac{W}{m^2}\right]=\frac{[kg]}{[s^3]}$
（级别）			分贝（公式，不是单位）	[dB]

除了这些物理维度之外，一些维度可以表示主观的印象，例如响度、音高以及音色。理论上讲，这些维度只能通过实验中听音人的主观判断来度量。然而，一旦我们进行了这些实验，就可以在对应的主观维度和单位（例如以巴克表示的感知音高）的基础上发现可以逼近物理上的维度和单位（例如以赫兹度量的频率）的公式。表 B.3 中包含了最常见的主观维度单位。我们在 12.6 节介绍了一些用来逼近相应物理维度的公式。

表 B.2　比例系数：缩写、希腊或意大利语来源、英式或美式英语名称、数字表达和以 10 为指数的幂的表达

缩写	希腊语/意大利语	英式英语	美式英语	数字	幂
t	Tera	Billion	Trillion	1000000000000	10^{12}
g	Giga	Milliard	Billion	1000000000	10^{9}
M	Mega	Million	Million	1000000	10^{6}
k	Kilo	Thousand	Thousand	1000	10^{3}
h	Hecto	Hundred	Hundred	100	10^{2}
da	Deca	Ten	Ten	10	10^{1}
d	Deci	Tenth	Tenth	0.1	10^{-1}
c	Centi	Hundredth	Hundredth	0.01	10^{-2}
m	Milli	Thousandth	Thousandth	0.001	10^{-3}
μ	Micro	Millionth	Millionth	0.000001	10^{-6}
n	Nano	Milliardth	Billionth	0.000000001	10^{-9}
p	Pico	Billionth	Trillionth	0.000000000001	10^{-12}

表 B.3　物理维度和级别以及其对应的单位和主观等价量

物理		感知	
维度/级	单位	维度	单位
音长	秒［s］	时长	–
基频	赫兹［Hz］	音高	mel bark_{CB} bark_{ERB}
声压级	dB_{SPL} dB_{RMS}	响度	dB_{A} phon sone

B.2　数学符号

在数学上，我们常用一个简化符号将序列（在数学术语中，称为数列）中所有数相加，来表示系统和。例如，如果我们需要将气压计测量所得的所有压强值相加，一种方法是把这些测量值写在一张表格中（见表 B.4）。

表 B.4　声压测样表，测量时间和测量值

测量数	日期和时间	值（百帕 hPa）
1	2004 年 5 月 2 日，上午 6 点	1015
2	2004 年 5 月 3 日，上午 6 点	1023
3	2004 年 5 月 4 日，上午 6 点	1031
4	2004 年 5 月 5 日，上午 6 点	927
5	2004 年 5 月 6 日，上午 6 点	986

为了计算这一序列的平均压强，我们需要把所有测量值相加（见表 B.5）。

表 B.5　五个气压测量值之和

第 1 个测量值	第 2 个测量值	第 3 个测量值	第 4 个测量值	第 5 个测量值	和
1015 +	1023 +	1031 +	927 +	986 =	4982

最后，必须将得到的和除以测量值的数目：

$$\frac{4982}{5}=996.4$$

在更一般的情形下，这一过程可以用下式表示：

$$\text{平均气压}=\frac{\text{第 1}+\text{第 2}+\text{第 3}+\text{第 4}+\text{第 5 测量值}}{5}$$

我们应该把这一过程推广到更一般的情形，使它可以适用于任何数量的测量值而不仅仅是 5 个。用我们日常生活的语言去描述这一过程：

“把所有的测量值相加，再将得到的和除以测量值的数目。”

在这样简单的情形中，这句话很好理解。但是如果涉及更复杂的计算，用日常用语表示的公式就会显得很不清晰了。这就是在数学表示中要使用符号的原因，例如用“+”来表示“加”，用“×”来表示“乘”。“把所有数相加”这一运算有自己的数学表示符号，用“sum”这一单词的缩写“S”所对应的希腊字母 Σ 表示。当然，在使用这个符号时，要说明求和的是哪些数字，以及“所有”的具体含义。因为测量值（或者其他任何一系列数值）都可以用序列编号（第 1 个测量值，第 2 个测量值，等等）来表示，我们只需要说明索引值（也就是它在序列中占据的位置）以及测量值的总数就足够了。索引值通常用字母 i（即“index”）来表示，而数值的总数一般用 n（即“numerus”）来表示。

在数学表示方法中，“将所有测量值相加”（或者，更精确地说“把序号由‘1’到‘n’的测量值全部加起来”），可以用以下符号表示：

$$\sum_{i=1}^{n} x_i$$

将这个和除以测量值的数目（n）就可以表示成：

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

对于上面的例子来说，我们只需要把五个真实的测量值填入这个抽象的表达式中：

$$\frac{\sum_{i=1}^{5} x_i}{5} = \frac{1015 + 1023 + 1031 + 927 + 986}{5} = 996.4$$

这种表示方法的一个重要的好处是可以避免对计算方法表达的混淆。相对而言，日常生活用语的描述通常容许多种解释，从而可能会导致错误的出现。

注　释

1. 数字计算机以“2”为基数进行运算。1024 这个数是由“2”自乘 10 次得到的：$1024 = 2^{10} = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$。

附 录 C

C. 1 共振峰频率值

表 C. 1 美式英语 10 个元音的共振峰频率值，发音人分别是儿童、成年女性、成年男性；以及这三组发音人的均值。数据来自 Peterson 和 Barney（1952；P&B），以及 Hillenbrand 等（1995；HGCW）的研究。在计算共振峰值的时候，Peterson 和 Barney 测量的是宽带语谱图，而 Hillenbrand 等测量的是 LPC 谱。尽管这两个研究的结果存在一些差异，但每个元音都具有相似的趋势。

	性别	研究	元音									
			i	ɪ	ɛ	æ	ɑ	ʌ	ɔ	ʊ	u	ɚ
F1	儿童	P&B	360	534	700	1017	1030	855	694	560	432	569
		HGCW	452	513	740	718	992	738	836	571	493	586
	女性	P&B	310	441	608	863	864	758	587	469	378	503
		HGCW	437	484	727	676	921	760	804	519	460	524
	男性	P&B	267	392	526	664	718	631	568	437	307	489
		HGCW	343	429	588	591	756	621	656	469	380	475
	$\bar{x}$		**362**	**464**	**648**	**749**	**886**	**729**	**679**	**504**	**405**	**522**
F2	儿童	P&B	3178	2744	2616	2334	1383	1592	1064	1402	1193	1806
		HGCW	3073	2556	2279	2497	1689	1538	1303	1506	1404	1721
	女性	P&B	2783	2474	2334	2049	1229	1409	915	1162	961	1641
		HGCW	2761	2369	2063	2335	1526	1416	1188	1229	1106	1588
	男性	P&B	2323	2034	1803	1930	1309	1181	1023	1123	992	1379
		HGCW	2294	1993	1854	1727	1091	1192	836	1023	876	1360
	$\bar{x}$		**2744**	**2367**	**2154**	**2142**	**1370**	**1395**	**1024**	**1231**	**1066**	**1556**

续表

	性别	研究	元音									
			i	ɪ	ɛ	æ	ɑ	ʌ	ɔ	ʊ	u	ɚ
F3	儿童	P&B	3763	3604	3564	3366	3188	3328	3263	3332	3250	2194
		HGCW	3702	3409	3297	3289	2937	3126	2951	3076	2992	2154
	女性	P&B	3312	3063	2999	2832	2783	2768	2736	2685	2666	1977
		HGCW	2273	3057	2953	2973	2832	2901	2834	2829	2735	1930
	男性	P&B	2937	2569	2481	2420	2442	2377	2403	2245	2239	1709
		HGCW	3001	2687	2604	2595	2535	2548	2521	2435	2355	1711
	$\bar{x}$		**3346**	**3065**	**2991**	**2902**	**2790**	**2863**	**2774**	**2764**	**2703**	**1930**

C.2　基频值

表 C.2　美式英语 10 个元音的基频值，发音人分别是儿童、成年女性和成年男性。数据来自 Peterson 和 Barney（1952；P&B），以及 Hillenbrand 等（1995；HGCW）的研究。标记为 $\bar{x}$ 的一列显示了三组数据的平均值。由表中数据，我们可以清楚地观察到儿童的基频值最高，而成年男性的基频值最低。此外，我们也可以观察到高元音［i，ɪ，u，ʊ］的基频平均值比低元音［ɛ，æ，ɑ，ʌ，ɔ］的要高。

	性别	研究	元音										
			i	ɪ	ɛ	æ	ɑ	ʌ	ɔ	ʊ	u	ɚ	$\bar{x}$
F_0	儿童	P&B	272	269	260	251	256	261	263	276	274	261	**249**
		HGCW	246	241	230	228	229	236	225	243	249	236	
	女性	P&B	235	232	223	210	212	221	216	232	231	218	**221**
		HGCW	227	224	214	215	215	218	210	230	235	217	
	男性	P&B	136	135	130	127	124	130	129	137	141	133	**131**
		HGCW	138	135	127	123	123	133	121	133	143	130	
	$\bar{x}$		**209**	**206**	**197**	**192**	**193**	**200**	**194**	**208**	**212**	**199**	

参考文献

Abramson, A. S. , and Lisker, L. (1970) . Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia, pp. 569 – 573.

Behrens, S. , and Blumstein, S. E. (1988) . Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics* 16, 295 – 298.

Békésy, G. von (1928) . Zur Theorie des Hörens: Die Schwingungsform der Basilarmembran. *Physikalische Zeitschrift* 22, 793 – 810.

Best, C. T. (1994) . The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman and H. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press, pp. 167 – 224.

Best, C. T. , and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn and M. J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 13 – 34.

Best, C. T. , McRoberts, G. W. , and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109, 775 – 794.

Bjuggren, G. , and Fant, G. (1965). The nasal cavity structures. *Quarterly Progress and Status Report*, Royal Institute of Technology, Stockholm, pp. 5 – 7.

Blumstein, S. E. , and Stevens, K. N. (1979). Acoustic invariance in speech

production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America* 66, 1001 – 1017.

Broad, D. J. (1973). Phonation. In F. D. Minifie, T. J. Hixon, and F. Williams (eds.), *Normal aspects of speech, hearing and language.* Englewood Cliffs, NJ: Prentice-Hall, pp. 127 – 167.

Broad, D. J. (1979). The new theories of vocal fold vibration. In N. Lass (ed.), *Speech and language: Advances in basic research and practice*, vol. 2. New York: Academic Press, pp. 203 – 256.

Carr, P. (1999). *English phonetics and phonology: An introduction.* Oxford: Blackwell.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., and Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience* 1, 351 – 353.

Chiba, T., and Kajiyama, M. (1941). *The vowel: Its nature and structure.* Tokyo: Tokyo-Kaiseikan.

Cho, T., Jun, S., and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30, 193 – 228.

Cole, R. A., and Scott, B. (1974). Toward a theory of speech perception. *Psychological Review* 81, 348 – 374.

Cooper, W. E., and Sorensen, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America* 62, 683 – 692.

Dang, J., Honda, K., and Suzuki, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *Journal of the Acoustical Society of America* 96, 2088 – 2100.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27, 769 – 773.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics* 22, 109 – 122.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303 – 306.

Ferrein, M. A. (1741). *De la formation de la voix de l'homme.* Recueil de l'Académie Royale des Sciences de Paris, 402 – 432.

Flanagan, J. L., and Landgraf, L. (1968). Self-oscillating source for vocal tract synthesizers. *IEEE Transactions Audio and Electroacoustics* AU – 16, 57 – 64.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47 – 65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Baltimore, MD: York Press, pp. 233 – 278.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics* 12, 47 – 65.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84, 115 – 123.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14, 3 – 28.

Fowler, C. A. (2003). Speech production and perception. In A. Healy and R. Proctor (eds.), *Handbook of psychology*, vol. 4: *Experimental psychology.* New York: John Wiley & Sons, pp. 237 – 266.

Grieser, D., and Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology* 25, 577 – 588.

Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet (1999). Cambridge: Cambridge University Press.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE* 66, 51 – 83.

Harris, K. S. (1958). Cues for the discrimination of American English frica-

tives in spoken syllables. *Language and Speech* 1, 1 – 7.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago, IL: University of Chicago Press.

Hayes, B., and Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9, 47 – 96.

Hedrick, M. S., and Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America* 94, 2005 – 2026.

Heinz, J. M., and Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America* 33, 589 – 596.

Helmholtz, H. von (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* Braunschweig: Vieweg.

Hillenbrand, J. M., and Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America* 105, 3509 – 3523.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97, 3099 – 3111.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica* 26, 89 – 94.

Hirsch, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America* 31, 759 – 767.

Holt, L. L., Lotto, A. J., and Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America* 116, 1763 – 1773.

Hombert, J. M. (1978). Consonant types, vowel quality and tone. In V. Fromkin (ed.), *Tone: A linguistic survey.* New York: Academic Press, pp. 77 – 111.

Hughes, A., and Trudgill, P. (1996). *English accents and dialects: An introduction to social and regional varieties of British English.* London: Edward Arnold.

Husson, R. (1950). Étude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée. *Éditions de La revue scientifique*, Paris, 1 – 91.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology* 23, 225 – 257.

Ishizaka, K., and Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell System Technical Journal* 51, 1233 – 1268.

Ishizaka, K., and Matsudaira, M. (1968). What makes the vocal cords vibrate? *Proceedings of Sixth International Congress of Acoustics*, Tokyo, pp. B1 – 3.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson and J. W. Mullennix (eds.), *Talker variability in speech processing*. San Diego, CA: Academic Press, pp. 145 – 166.

Johnson, K. (2003). Speaker normalization in speech perception. In D. B. Pisoni and R. E. Remez (eds.), *The handbook of speech perception*. Oxford: Blackwell, pp. 363 – 389.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85, 1718 – 1725.

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America* 64, 1386 – 1391.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America* 73, 322 – 335.

Kewley-Port, D., Pisoni, D. B., Studdert-Kennedy, M. (1983). Perception ofstatic and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* 73, 1779 – 1793.

Kingston, J., and Diehl, R. (1994). Phonetic knowledge. *Language* 70, 419 – 454.

Kirk, P. L., Ladefoged, J., and Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In A. Mattina and T. Montler (eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*. University of Montana Occasional Pa-

pers in Linguistics, no. 10. Missoula, MT: University of Montana, pp. 435 – 450.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America* 53, 8 – 16.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science* 297, 846 – 848.

Kuhl, P. K. (1989). On babies, birds, modules, and mechanisms: A comparative approach to the acquisition of vocal communication. In R. J. Dooling and S. H. Hulse (eds.), *The comparative psychology of audition: Perceiving complex sounds*. Hillsdale, NJ: Lawrence Erlbaum, pp. 379 – 419.

Kuhl, P. K. (1991). Human adults and infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93 – 107.

Kuhl, P. K., and Miller, J. D. (1975). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63, 905 – 917.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606 – 608.

Kurowski, K., and Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *Journal of the Acoustical Society of America* 76, 383 – 390.

Kurowski, K., and Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America* 81, 1917 – 1927.

Ladd, D. R., and Silverman, K. (1984). Vowel intrinsic pitch in connected speech. *Phonetica* 41, 31 – 40.

Ladefoged, P. (1967). *Three areas of experimental phonetics*. Oxford: Oxford University Press.

Ladefoged, P. (1996). *Elements of acoustic phonetics*, 2nd edn. Chicago, IL: University of Chicago Press.

Ladefoged, P. (2006). *A course in phonetics*. Boston, MA: Thomson Wadsworth.

Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America* 29, 98 – 104.

Ladefoged, P., and Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America* 76, 391 – 404.

LaRiviere, C., Winitz, H., and Herriman, E. (1975). Vocalic transitions in the perception of voiceless initial stops. *Journal of the Acoustical Society of America* 57, 470 – 475.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley & Sons.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America* 29, 117 – 123.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology* 1, 301 – 323.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist* 37, 148 – 167.

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1 – 36.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1952). The role of

selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology* 65, 497 – 516.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied* 68, 1 – 13.

Lindblom, B. (1963). On vowel reduction. *Report no.* 29, *Speech Transmission Laboratory*, The Royal Institute of Technology, Sweden.

Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word* 20, 384 – 422.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89, 874 – 886.

Macchi, M. J. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. *Journal of the Acoustical Society of America* 68, 1636 – 1642.

Mack, M., and Blumstein, S. E. (1983). Further evidence of acoustic invariance in speech production: the stop-glide contrast. *Journal of the Acoustical Society of America* 73, 1739 – 1750.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review* 77, 182 – 196.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In M. Huffman and R. Krakow (eds.), *Phonetics and phonology: Nasals, nasalization, and the velum*, vol. 5. New York: Academic Press, pp. 147 – 167.

Malecot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language* 32, 274 – 284.

Mann, V. A., and Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ] – [s] distinction. *Perception & Psychophysics* 28, 213 – 228.

Markel, J. D., and Gray, A. H. (1976). *Linear prediction of speech.* Berlin: Springer-Verlag.

McClean, M. D. (2000). Patterns of orofacial movement velocity across variations in speech rate. *Journal of Speech, Language, and Hearing Research* 43, 205 – 216.

McLaughlin, F. (2005). Voiceless implosives in Seereer-Siin. *Journal of the International Phonetic Association* 35, 201 – 214.

McMurray, B., Tanenhaus, M. K., and Aslin, R. A. (2002). Gradient effects of within-category variation on lexical access. *Cognition* 86, 33 – 42.

Mikuteit, S., and Reetz, H. (2007). Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50, 247 – 279.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America* 85, 2114 – 2134.

Miller, J. L., and Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics* 25, 457 – 465.

Miller, J. L., and Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics* 54, 205 – 210.

Moore, K. L., and Dalley, A. F. (1999). *Clinically oriented anatomy.* Philadelphia, PA: Lippincott, Williams & Wilkins.

Nearey, T. M. (1978). *Phonetic feature systems for vowels.* Bloomington, IN: Indiana University Linguistics Club.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America* 112, 711 – 719.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America* 59, 640 – 654.

Payan, Y., and Perrier, P. (1997). Why should speech control studies based on kinematics be considered with caution? Insights from a 2D biomechanical model of the tongue. *Proceedings of Eurospeech* 97, *Rhodes, Greece,*

pp. 2019 – 2022.

Peterson, G. E. , and Barney, H. E. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175 – 184.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61, 1352 – 1361.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (ed.), *Speech and language: Advances in basic research and practice*, vol. 10. New York: Academic Press, pp. 243 – 335.

Repp, B. H. , and Svastikula, K. (1988). Perception of the [m] – [n] distinction in VC syllables. *Journal of the Acoustical Society of America* 83, 237 – 247.

Rizzolatti, G. , and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences* 21, 188 – 194.

Selkirk, E. O. (1984). On the major class features and syllable theory. In M. Aronoff and R. T. Oehrle (eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*. Cambridge, MA: MIT Press, pp. 107 – 113.

Sharma, A. , and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America* 106, 1078 – 1083.

Shinn, P. C. , Blumstein, S. E. , and Jongman, A. (1985). Limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics* 38, 397 – 407.

Sluijter, A. M. C. (1995). Phonetic correlates of stress and accent. Doctoral dissertation, Leiden University, The Netherlands.

Sluijter, A. M. C. , Van Heuven, V. J. , and Pacilly, J. J. A. (1997). Spectral balance as a cue to linguistic stress. *Journal of the Acoustical Society of America* 101, 503 – 513.

Smits, R. , Ten Bosch, L. , and Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants.

I. Perception experiment. *Journal of the Acoustical Society of America* 100, 3852 – 3864.

Stevens, K. N., and Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas and J. L. Miller (eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum, pp. 1038 – 1055.

Stevens, S. S., and Volkman, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology* 53, 329 – 353.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. Tohkura, E. Vatikiotis-Bateson, and U. Sigasaka (eds.), *Speech perception, production, and linguistic structure*. Tokyo: Ohmasha, pp. 197 – 219.

Strange, W. (2007). Cross-language similarity of vowels: Theoretical and methodological issues. In O. S. Bohn and M. J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 35 – 56.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America* 60, 213 – 224.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90, 1309 – 1325.

Sussman, H. M., Hoemeke, K. A., and Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *Journal of the Acoustical Society of America* 94, 1256 – 1268.

Syrdal, A. K., and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79, 1086 – 1100.

Tekieli, M. E., and Cullinan, W. L. (1979). The perception of temporally segmented vowels and consonant-vowel syllables. *Journal of Speech and Hearing Research* 22, 103 – 121.

Traunmüller, H. (1982). Der Vokalismus im Ostmittelbairischen. *Zeitschrift für Dialektologie und Linguistik* 49, 289 – 333.

Underbakke, M., Polka, L., Gottfried, T. L., and Strange, W. (1988). Trading relations in the perception of /ɪ/ – /l/ by Japanese learners of English. *Journal of the Acoustical Society of America* 84, 90 – 100.

Van den Berg, J., Zantema, J. T., and Doornenbal, P., Jr. (1957). On the air resistance and the Bernoulli effect of the human larynx. *Journal of the Acoustical Society of America* 29, 626 – 631.

Verbrugge, R. R., Strange, W., Shankweiler, D. P, and Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America* 60, 198 – 212.

Wang, Y., Spence, M., Jongman, A., and Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106, 3649 – 3659.

Warner, N., and Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109, 1144 – 1156.

Weibel, E. R. (1984). *The pathway for oxygen. Structure and function in the mammalian respiratory system.* Cambridge, MA: Harvard University Press.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49 – 63.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s] – [ʃ] distinction. *Journal of the Acoustical Society of America* 69, 275 – 282.

Williams, B. (1983). Stress in modern Welsh. Doctoral dissertation, University of Cambridge.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics* 21, 289 – 297.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics* 26, 95 – 104.

Winitz, H., Scheib, M. E., and Reeds, J. A. (1971). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51, 1309 – 1317.

Wolfram, W., and Schilling-Estes, N. (1998). *American English.* Oxford: Blackwell.

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology.* Boston, MA: Allyn and Bacon.

术　语

3-dB bandwidth 3-dB 带宽
3D-representation 3D 表示（3 维表示）

abdominal breathing 腹式呼吸
abduction 外展
ABX-task ABX 任务
acoustic cue 声学线索
acoustic phonetics 声学语音学
Adam's apple 喉结
adduction 内收
aerodynamic theory 空气动力学理论
affricate 塞擦音
aliasing 混叠
allophone 音位变体
alveolar 齿龈音
alveolar ridge 齿龈脊
alveoli pulmonis 肺泡
amplitude 振幅
amplitude contour 振幅曲线
analog 模拟
anticipatory coarticulation 逆向协同发音，也称预期性协同发音
anti-aliasing filter 抗混叠滤波器
anti-formant 反共振峰

anti-node 反节点
anvil 见 incus 砧骨
apex (of the cochlea) 耳蜗顶
apex (of the tongue) 舌尖
apical 舌尖音
approximant 近音
articular disc 关节盘
articulation 发音，也称调音
articulator 发音器官，也称调音器官
articulatory phonetics 发音语音学
arytenoid cartilage 杓状软骨
aspiration 呼吸
attenuated 衰减
auditory phonetics 听觉语音学
auditory scale 听觉标度
auditory sensation 听觉感知
auricle 耳廓

band-pass filter 带通滤波器
Bark scale 巴克标度
base (of the cochlea) (耳蜗的) 基底
basilar membrane 基底膜
bel 贝尔
Bernoulli effect 伯努力效应
bilabial 双唇音
bilingualism 双语
Boyle's law 波义耳定律
brain imaging technique 脑成像技术
breathy voice 气嗓音
broad band spectrogram 见 wide-band spectrogram 宽带语谱图
broad transcription 宽式标音

bronchi 支气管
bronchia 小支气管

carry-over 顺向
categorical perception 范畴感知
category boundary 范畴边界
center frequency 中心频率
central approximant 央近音
central vowel 央元音
centralized 央化
chest 见 thorax 胸部
chest register 胸声区
choanae 内鼻孔
click 咂音
closed syllable 闭音节
closed vowel 闭元音
coarticulation 协同发音
cochlea 耳蜗
coda（syllable） 音节尾
complex signal 复合信号
concha 外耳
conflicting-cue 冲突线索
consonant cluster 辅音丛
consonant 辅音
constrictor muscle 括约肌
context 语境
continuous signal 连续信号
continuous spectrum 连续谱
contour tone 曲折调
contrastive length 对立音长
contrastive tone 对立声调

conus elasticus 弹性圆锥
cornu inferior 甲状软骨下角
cornal 舌冠音
cover body 体层被覆层
creaky voice 紧喉嗓音，也称嘎裂声
cricoid cartilage 环状软骨
cricothyroid muscle 见 external laryngeal muscle 环甲肌
critical bands 临界频带
critical frequency 临界频率
critical period hypothesis 关键期假说
cross-language speech perception 跨语言言语感知
cross-splicing 交叉拼接
cut-off frequency 截止频率

damping 阻尼、衰减
dark l ([ɬ]) 暗 l
dB 分贝
dB/oct 分贝/倍频程
dB_A 等响度分贝
dB_{FS} 满量程分贝
dB_{SL} 声级分贝
dB_{SPL} 声压级分贝
decibel 见 dB 分贝
declination 下倾
degree [°] 度
dental 齿音
devoicing 不带音、清化
DFT 见 discrete Fourier transformation 离散傅里叶变换
diacritic 附加符号
diaphragm 膈膜
digital 数字（的）

dimension 维度
diphthong 二合元音
diphthongized 二合元音化的
direct realism 直接现实理论
discrete 离散
discrete Fourier transformation 离散傅里叶变换
discrimination 区分
displacement 位移
distinctive sound 区别性声音
dorsal 舌面音
double articulation 双重发音

ear 耳
ear canal 见 meatus 耳道
ear drum 见 tympanic membrane 鼓膜
eddy 涡流
efferent 传出
egressive 外呼
egressive pulmonic airstream 外呼肺气流
ejective 喷音
elastic recoil force 弹性回缩力
emphatic consonant 重辅音
endolymph 内淋巴液
energy 能量
epiglottis 会厌
equal loudness scale 等响度标度
equivalence classification 等同分类
equivalent rectangular bandwidth scale 等效矩形带宽标度
ERB 见 equivalent rectangular bandwidth scale 等效矩形带宽标度
esophagus 食管，也称食道
Eustachian tube 耳咽管，也称咽鼓管

examplar-based theory 范例理论
expiratory reserve volume 呼气储备容量
external ear 外耳
external intercostal muscle 肋间外肌
external laryngeal muscle 喉外肌
extrinsic normalization 外部规整

F_0见 fundamental frequency 基频
F_0 contour 基频曲线
false vocal folds 假声带
fast Fourier transformation 快速傅里叶变换
FFT 见 fast Fourier transformation 快速傅里叶变换
filter 滤波、滤波器
first harmonic 第一谐波，也称一次谐波
flap 闪音
flow-separation theory 流动-分离理论
foot 音步
formant 共振峰
Fourier analysis 傅里叶分析
Fourier synthesis 傅里叶合成
Fourier transformation 傅里叶变换
frequency 频率
frequency spectrum 频率谱
fricative 擦音
frontness (of vowels)(元音的) 前后
fully voiced plosive 全浊爆发音
fundamental frequency 基频
fundamental period 基本周期

geminate 长辅音
general mechanism account 普遍机制说

general phonetic transcription 通用语音标音
genioglossus muscle 颏舌肌
glide 滑音
glottal 喉音
glottal fry 喉气泡声
glottalic airflow 喉头气流
glottalization 喉化
glottis 声门
GMA 见 general mechanism account 普遍机制说
goodness rating 优度评级
gullet 见 esophagus 食道
guttural 咽喉音

hair cell 毛细胞
half-wavelength resonator 二分之一波长共鸣器
hammer 见 malleus 锤骨
hard palate 硬腭
HAS 见 high amplitude sucking 高振幅吮吸
height（of vowels）（元音的）高低
helicotrema 蜗孔
Helmholtz resonator 赫姆霍兹共鸣器
Hertz 赫兹
high amplitude sucking 高振幅吮吸
high vowel 高元音
high-pass filter 高通滤波器
homophone 同音词
homorganic 同部位、同器官的
hyoglossus muscle 舌骨舌肌
hyoid bone 舌骨
Hz 见 Hertz 赫兹

identifaction 辨认
implosive 内爆音
impressionistic transcription 印象标音
impulse 脉冲
incus 砧骨
inferior longitudinal muscle 舌下纵肌
infrahyoid muscle 舌骨下肌
ingressive 内吸
inner hair cell 内毛细胞
inspiratory reserve volume 吸气储备量
intended gesture 预期音姿
intensity 强度
intensity level 音强级
internal ear 内耳
internal intercostal muscle 肋间内肌
International Phonetic Alphabet 国际音标
interpleural space 胸膜间腔
intonation 语调
intonational phrase 语调短语
intrinsic fundamental frequency 固有基频
intrinsic normalization 内部规整
intrusion 增音
IPA 见 International Phonetic Alphabet 国际音标
isochrony 等时性

JND 见 just noticeable difference 最小可觉差
just noticeable difference 最小可觉差

L1 第一语言
L2 第二语言
labial 唇音

labialization 唇化
labiodental 唇齿音
laminal 舌叶音
laminar flow 层流
laryngealization 喉化
laryngopharynx 喉咽
larynx 喉
lateral approximant 边近音
lateral cricoarytenoid muscle 环杓侧肌
lateral fricative 边擦音
lateral plosion 边爆破
lateral pterygoid muscle 翼外肌
lax vowel 松元音
length mark 音长符号
lexical tone 词调
line spectrum 线谱
linear predictive coding 线性预测编码
linear scale 线性标度
lip rounding 圆唇
lips 唇
liquid 流音
locus equation 音轨方程
logarithmic scale 对数标度
longitudinal wave 纵波
low vowel 低元音
low-pass filter 低通滤波器
LPC 见 linear predictive coding 线性预测编码
LPC re-synthesis 线性预测编码再合成（LPC 再合成）
lung membrane 见 visceral pleura 肺胸膜
lungs 肺

main stress 见 primary stress 主重音
malleus 锤骨
manner of articulation 发音方法，也称调音方法
mass 质量
masseter muscle 咬肌
meati 鼻道
meatus（外耳）道
mel scale 美标度
microphone 麦克风
mid vowel 中元音
middle ear 中耳
minimal pair 最小对立体
mirror neurons 镜像神经元
modal voice 常态嗓音，也称中性嗓音
monophthong 单元音
mora 莫拉
mora timing 莫拉计时
motor theory of speech perception 言语感知的运动理论
muco-viscose theory 黏弹力理论
myoelastic theory 肌弹性理论

nares 鼻孔
narrow transcription 严式标音
narrow-band spectrogram 窄带语谱图
nasal formant 鼻共振峰
nasal murmur 鼻哞音
nasal plosion 鼻爆破
nasal septum 鼻中隔
nasal sound 鼻音
nasal stop 鼻塞音
nasal vowel 鼻元音

nasalized 鼻化
nasalized vowel 鼻化元音
nasopharynx 鼻咽
neurochronaxic theory 神经时值理论
node 节点
noise 噪声
normalization 规整
nostrils 鼻孔
nuclear accent 核心重音
nucleus（syllable）音节核
Nyquist criterion 奈奎斯特准则
Nyquist frequency 奈奎斯特频率

oblique arytenoid muscle 杓斜肌
obstruent 阻音
odd-ball task odd-ball 任务
offglide 出渡
onset（syllable）音节首
open syllable 开音节
open syllable lengthening 开音节延长
open vowel 开元音
oral airstream 口气流
oral sound 口音
oral tract 口腔
organ of Corti 柯蒂氏器官
oropharynx 口咽
oscillogram 波形图
ossicles 听小骨
oto-acoustic emission 耳声发射
outer hair cells 外毛细胞
oval window 卵圆窗

Pa 见 pascal 帕斯卡
palatal 腭音
palatalization 腭化
palate 腭、硬腭
palatoalveolar 见 postalveolar 龈后音
palatoglossus muscle 腭舌肌
parietal pleura 壁层胸膜
partial oscillation 局部振荡
pascal 帕斯卡
pass-band 通带
peak-to-peak amplitude 峰峰振幅
perilymph 外淋巴液
period 周期
periodic 周期性
perseverative coarticulation 顺向协同发音，也称保持性协同发音
perturbation theory 干扰理论
pharyngeal 咽音
pharyngealization 咽化
pharynx 咽
phase 相位
phase angle 相角
phase shift 相移
phon 方
phonation 发声
phone 音素
phoneme 音位
phonemic 音位的
phonetic transcription 语音标音
phonetics of second language acquisition 第二语言习得语音学
phonology 音系学
phrase-final lengthening 短语末尾延长

physical stimulus 物理刺激
pinna 见 auricle 耳廓
pitch 音高
pitch accent 音高重音
pitch contour 音高曲线
place of articulation 发音部位，也称调音部位
plain plosive 普通爆发音
plasticity 可塑性
pleural linkage 胸腔联动
plosive 爆发音
point normalization 点规整
point vowel 顶点元音
postalveolar 龈后音
posterior cricoarytenoid muscle 环杓后肌
power 功率
power spectrum 见 spectrum 功率谱
pressure 压力、压强
prevoicing 前浊音，也称前带音
primary stress 主重音
progressive coarticulation 顺向协同发音
prosody 韵律
pulmones 见 lungs 肺
pulmonic airflow 肺气流
pure tone 纯音

quality of a sound 音质
quantity of a sound 音长
quarter-wavelength resonator 四分之一波长共鸣器
quasi-periodic signal 准周期信号

r-coloring 卷舌音色，也称 r-音色

radical 舌根音
randomization 随机
range normalization 范围规整
reaction time 反应时间
reconstruction filter 重构滤波器
reduced form 见 weak form 弱化形式
refractory period 不应期
register 声区
register tone 平调
regressive coarticulation 逆向协同发音
Reissner's membrane 前庭膜
relative amplitude 相对振幅
residual volume 残留量
resonance (frequency) 共振 (频率)
rest capacity 静息容量
retroflex 卷舌
revised motor theory 修正的运动理论
rhotacization 卷舌化
rhythm reversal 节奏返转
rime (syllable) (音节的) 韵基
rise time 上升时间
RMS 见 root mean square 均方根
root mean square 均方根
round window 圆窗

sample 采样、样本
sampling rate 采样率
scala media 蜗管, 也称中阶
scala tympani 鼓阶
scala vestibuli 前庭阶
schwa 中性元音, 也称中央元音

secondary articulation 次要发音
secondary stress 次重音
semitone 半音
semi-vowel 半元音
short time spectrum 短时谱
sine wave 正弦波
SL 见 sound level 声级
SMA 见 special mechanism account 特殊机制说
soft palate 软腭
sone 宋
sonogram 语谱图
sonorant 响音
sonority 响度
sound level 声级
sound pressure level 声压级
sound segment 音段
sound wave 声波
source signal 声源信号
speaking rate 语速
special mechanism account 特殊机制说
spectral moment 谱矩
spectral resolution 谱分辨率
spectrogram 语谱图
spectrum 谱
speech is special 言语是特别的
speech perception 言语感知
speech production 言语产生
speed of sound 声速
SPL 见 sound pressure level 声压级
stapes 镫骨
steady state 稳定状态

steepness 陡度
stereocilia 硬纤毛
stirrup 见 stapes 镫骨
stop 塞音
stop-band 阻带
stress shift 重音移位
stress-timed 重音计时
strong form 强形式
styloglossus muscle 茎突舌肌
subglottal system 声门下系统
superior longitudinal muscle 舌上纵肌
suprahyoid muscle 舌骨上肌
supralaryngeal system 喉上系统，也称声门上系统
suprasegmental 超音段
surface tension 表面张力
syllabic consonant 音节性辅音
syllable 音节
syllable-timed 音节计时
sympathetic resonance 交感共振

tap 拍音
tectorial membrane 盖膜，也称柯蒂氏膜
teeth 牙
temporal muscle 颞肌
temporomandibular joint 颞下颌关节
tense vowel 紧元音
thoracic breathing 胸式呼吸
thoracic cavity 胸腔
thorax 胸部
thyroid cartilage 甲状软骨
tidal volume 潮汐量

timbre 音色
TMJ 见 temporomandibular joint 颞下颌关节
tone language 声调语言
tongue blade 舌叶
tongue body 舌体
tongue dorsum 舌背
tongue muscles 舌肌
tongue root 舌根
tongue tip 舌尖
tonic syllable 调核音节
tonogenesis 声调发生学
tonotopic 音频响应
trachea 气管
trachealis muscle 气管肌
transcription 见 phonetic transcription 标音
transverse arytenoids muscle 横杓肌
transverse muscle (of the tongue) 舌横肌
transverse wave 横波
traveling wave 行波
trill 颤音
two-mass theory 双质量理论
tympanic cavity 鼓室
tympanic membrane 鼓膜
types of phonation 发声类型

undersampling 欠采样
unit 单位
unreduced form 见 strong form 非弱化形式
unvoiced 见 voiceless sound 清化
uvula 小舌
uvular 小舌音

velar 软腭音
velar pinch 软腭夹
velaric airflow 软腭气流
velarization 软腭化
velopharyngeal closure 腭咽闭合
velopharyngeal port 腭咽口
velum 软腭
ventricles 喉室
vertical muscle (of the tongue) 舌垂直肌
vibrating string theory 振动弦理论
visceral pleura 脏层胸膜
vital capacity 肺活量
vocal apparatus 发音器官
vocal c(h)ords 见 vocal folds 声带
vocal fold vibration 声带振动
vocal folds 声带
vocal fry 见 glottal fry 气泡声
vocal ligament 声韧带
vocal process 声带突
vocal tract 声道
vocalis muscle 声带肌
voice bar 嗓音横杠，也称浊音杠
voice offset time 嗓音结束时间
voice onset time 嗓音起始时间
voiced aspirate plosive 浊送气爆发音
voiced plosive 浊爆发音
voiced sound 浊音，也称带音
voiceless aspirate plosive 清送气爆发音
voiceless plosive 清爆发音
voiceless pre-aspirated plosive 清预送气爆发音
voiceless sound 清音，也称不带音

voiceless unaspirated plosive 清不送气爆发音
voicelessness 清音、不带音性
voicing 见 phonation；voiced sound 发声、带音
VOT 见 voice onset time 嗓音起始时间
vowel quadrilateral 元音四边形
vowel undershoot 元音不到位
vowels 元音

waterfall display 断面显示，也称瀑布式显示
waveform 波形
wavelength 波长
weak form 弱形式
weight 重量
whisper 耳语声
wide-band spectrogram 宽带语谱图
windowing 加窗
windpipe 见 trachea 气管

zero-line 零线

国际音标（修订至 2018 年）

辅音（肺部气流）

参考 IPA2018 及《方言》2007 年第 01 期《国际音标(修订至 2005 年)》制表

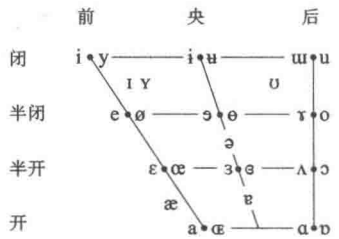
	双唇	唇齿	齿	齿龈	龈后	卷舌	硬腭	软腭	小舌	咽	喉
爆发音	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
鼻音	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
颤音	ʙ			r					ʀ		
拍音/闪音		ⱱ		ɾ		ɽ					
擦音	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
边擦音				ɬ ɮ							
近音		ʋ		ɹ		ɻ	j	ɰ			
边近音				l		ɭ	ʎ	ʟ			

单元格中左侧的音标为清辅音，右侧的音标为浊辅音。阴影区域表示不可能产生的发音。

辅音（非肺部气流）

咽音		浊内爆音		喷音	
ʘ	双唇音	ɓ	双唇音	’	例如：
ǀ	齿音	ɗ	齿音 / 龈音	pʼ	双唇音
ǃ	龈（后）音	ʄ	硬腭音	tʼ	齿音 / 龈音
ǂ	腭龈音	ɠ	软腭音	kʼ	软腭音
ǁ	齿边音	ʛ	小舌音	sʼ	龈擦音

元音



成对出现的音标，右边的为圆唇元音。

其他符号

ʍ	清唇-软腭擦音	ɕ ʑ	龈-硬腭擦音
w	浊唇-软腭近音	ɺ	浊龈边闪音
ɥ	浊唇-硬腭近音	ɧ	同时发 ʃ 和 x
ʜ	清会厌擦音		
ʢ	浊会厌擦音	若必要，可用连音符将两个音标连接起来以表示塞擦音及双重发音。	t͡s k͡p
ʡ	会厌爆发音		

超音段

ˈ	注重音	ˌfoʊnəˈtɪʃən
ˌ	次重音	
ː	长	eː
ˑ	半长	eˑ
˘	超短	ĕ
|	小（音步）组块	
‖	大（语调）组块	
.	音节间隔	ɹi.ækt
‿	连接（间隔不出现）	

声调与词重调

平调		曲折调	
ě 或 ˥	超高	ě 或 ˩˥	升
é	˦ 高	ê	˥˩ 降
ē	˧ 中	ě	˧˥ 高升
è	˨ 低	ě	˩˧ 低升
ȅ	˩ 超低	ě	˧˥˧ 升降
↓	降阶	↗	整体上升
↑	升阶	↘	整体下降

附加符号 如果是下伸符号，附加符号可以加在上方，例如：n̥

◌̥	清化	n̥	d̥	◌̤	气声性	b̤	a̤	◌̪	齿化	t̪	d̪
◌̬	浊化	s̬	t̬	◌̰	嘎裂声性	b̰	a̰	◌̺	舌尖性	t̺	d̺
◌ʰ	送气	tʰ	dʰ	◌̼	舌唇	t̼	d̼	◌̻	舌叶性	t̻	d̻
◌̹	更圆	ɔ̹		◌ʷ	唇化	tʷ	dʷ	◌̃	鼻化		ẽ
◌̜	略展	ɔ̜		◌ʲ	腭化	tʲ	dʲ	◌ⁿ	鼻除阻		dⁿ
◌̟	偏前	u̟		◌ˠ	软腭化	tˠ	dˠ	◌ˡ	边除阻		dˡ
◌̠	偏后	e̠		◌ˤ	咽化	tˤ	dˤ	◌̚	无闻除阻		d̚
◌̈	央化	ë		◌~	软腭化或咽化	ɬ					
◌̽	中-央化	e̽		◌˖	偏高	e̝	（ɻ̝ = 浊龈擦音）				
◌̩	成音节	n̩		◌˗	偏低	e̞	（β̞ = 浊双唇近音）				
◌̯	不成音节	e̯		◌˕	舌根偏前	e͒					
◌˞	r 音色	ɔ˞	a˞	◌˔	舌根偏后	e͓					

语音学：标音、产生、声学和感知

《语音学：标音、产生、声学和感知》向读者全面介绍了与言语和声音研究有关的内容。两位作者通过悉心专业的引导使读者可以通晓这一领域的主要议题：

- **语音标音：言语声是如何被记录的**
- **言语产生：言语声是如何产生的**
- **声学：言语声及言语声的声学特征**
- **言语感知：言语是如何被听音人感知的**

这本教材首次对语音学相关领域的内容进行了全面的覆盖，涵盖了对声带振动以及人耳工作原理的介绍，为教师提供了一份灵活而引人入胜的教学材料。每一章最后的练习可以让学生们有机会将他们所学的知识应用于实践。而与本书配套的网站（www.blackwellpublishing.com/phonetics）也提供了一系列声音文件，可以更为生动地解释书中相关的文字内容。

这本书是专为语言学专业的本科生或言语科学、应用语言学等相关专业的学生设计的教材。

ISBN 978-7-5203-3298-9
9 787520 332989 >

定价：96.00元