

# 人是怎样定位声音的

申 兵 辉

(中国农业大学基础学院应用物理系 北京 100094)

人能利用双耳定位声源并能从听觉世界中不和谐音调中辨别出个别的声音。早在 120 年前,瑞利就了解部分定位过程。他发现,如果声源位于听者的右方,则听者的左耳就被头部所遮挡。因此,右耳得到的声强大于左耳得到的声强,这种差异是判断声源方位的重要线索。如今人们用图 1 所示的声音定位设备进行了有趣的物理学、心理学和生理学研究。研究表明,根据声音的多种信息,包括强度、时序和频谱,人的大脑能从听到的声音再现声学景观的三维图像。

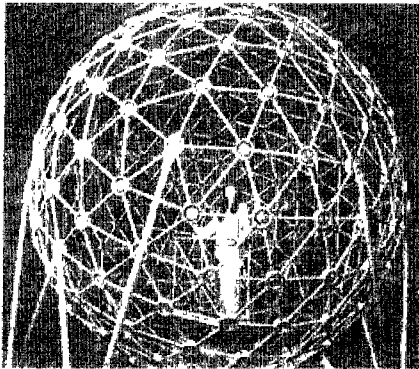


图 1 位于俄亥俄州代顿市的声音定位设备。球体直径 5 米,装有 277 个喇叭,封装在 6 立方米的消音室内,墙壁对频率大于 70Hz 的声波有强吸收作用

## 一、耳间级差

耳间级差 (ILD) 是指两耳处声强级的差值,用分贝表示。这种效应的大小可以通过计算球面两个相对的极点间的声强来估算。如图 2 所示,在听觉频谱范围内 (通常为 20~20000Hz), ILD 是频率的强相关函数。这是因为声波的波长比头部直径大时,存在明显的衍射。在 500Hz 的频率下,波长为 0.69 米,约为人类头部直径的 4 倍。因此,只要波源在 1 米以外,当频率低于 500Hz 时,ILD 较小。除此之外,头部的散射随频率的增大而迅速增强,在 4000Hz 下,头部的阴影将不可忽略。

ILD 的利用或多或少地依赖于中心神经系统对

这种差异的敏感程度。用进化论的术语来说,可以认为中心神经系统的灵敏度在一定程度上反映了 ILD 值的大小,但事实上并非如此。心理声学实验表明,中心神经系统对所有的频率具有相同的灵敏度。不管频率如何变化,最小可测到的 ILD 的变化约为 0.5 分贝。因此,ILD 大于 1 分贝时,它是定位声音的强有力的武器。球形脑模型显然过于简化了些,人类的头部包含有各种各样的二次散射物质,这些物质有可能形成 ILD 的高频依赖性的结构。不容置疑,这种结构能够为定位声音提供额外的线索。事实的确如此,不过,这是本文的另外一个话题,将在后面讨论。

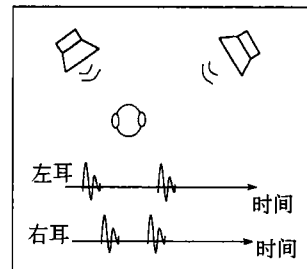


图 2 对于声源位于两个耳朵及鼻子构成的平面内的情形,计算所得的耳间级差。图中三条曲线分别对应于与聆听者正前方成 10°、45°和 90°的方位角

在长波极限下,球形脑模型正确地预示了 ILD 将变得小到可以忽略。所以,如果声音只由 ILD 来定位,这将很难定位频率在 500Hz 以下的声音。如果是这样的话,将无法解释瑞利的发现:他可以很容易地定位单色稳态低频声波,例如 256Hz 或 128Hz。因为他知道,声音的定位不能完全依赖于 ILD,最终他于 1907 年得出结论,耳朵一定能够觉察到两耳处波的相位差。

## 二、耳间时差

对于单色波,相位差等效于到达两耳的波形特征(如波峰和由正到零的交叉点)的时间差。对于频率为  $f$  的单色波,相位差  $\Delta\phi$  对应于耳间时差 (ITD)  $\Delta t = \Delta\phi / (2\pi f)$ 。在长波极限下,由球体对声波的

衍射公式可以得到,耳间时差 $\Delta t$ 表示成方位角(左-右) $\theta$ 的函数:

$$\Delta t = \frac{3a}{c} \sin\theta \quad (1)$$

其中 $a$ 为头部半径(约为0.0875m), $c$ 是声速(340m/s)。因此, $3a/c = 763\mu\text{s}$ 。

心理声学实验表明,人类对于500Hz的正弦声调可以定位到很高的精度。听者对前方( $\theta$ 接近于0)的声源,甚至可以觉察到只有 $1-2^\circ$ 的 $\Delta\theta$ 的差异。这种情况无法由ITD来解释,因为 $1^\circ$ 方位角的差异只相当于ITD的 $13\mu\text{s}$ 。一个神经中枢系统,其联合突触时间延迟为毫秒量级,对如此短的时间差成功地进行编码似乎是不可能的。但是,听觉系统却以某种方式提前进行编码了。耳机实验可以证明这种能力,实验中ITD可以在ILD之外独立进行。在此情形下,大脑成功的关键是并行处理,处理中心是位于中脑的上位橄榄体,它能够对两耳中的信号完成互相关操作。显然,双耳系统是利用多个神经元传递信息的。

耳机ITD实验给聆听者一个奇特的体验:映像的位置处于左侧或右侧依赖于ITD的符号,但映像似乎位于聆听者的头内。这样的映像被称为“侧向的”及不定域的。

利用耳机,可以测量可感觉的最小ITD的变化对ITD自身的函数关系。对于实际声源,用这些ITD数据及(1)式,可以预测最小可探测到的方位角的变化 $\Delta\theta$ 与 $\theta$ 的函数关系。实验发现,结果与预测值吻合,从而证实了大脑依赖于ITD来定位声音。

与其他相敏系统一样,当波长与两个探测器间距可比拟时,双耳相位探测器使用ITD受到相位模糊的影响,见图3。与之等价的时域观点是,为了避免相位模糊,波的半周期必须大于两耳间的时间延迟。当此延时正好为半周期时,到达两个耳朵的信号正好反相,为完全模糊。如果周期再短些,处于延时与双倍延时之间时,ITD将导致一个明显的错误,这就是在头部一侧的真实的声源被定位到与之相反的另一侧。与其得到这样一个错误的结果,倒不如对ITD没有任何反应。事实上,双耳系统正是用此方法解决了这个问题:双耳系统对1000Hz到1500Hz频率的声波彻底失去其对ITD的灵敏性,而这个频段正好是使耳间相差变得模糊的频段。

两耳差异可以概括为,双耳生理系统对任何频率的ILD的幅度敏感,当频率高于3000Hz时,ILD

变得更大更可靠。所以,对高频段,它是最有效的。与之相反,双耳生理系统仅在低于1500Hz的低频段才有能力从ITD获得相位信息。对处于中频段(例如2000Hz)的正弦声波,两种线索都不好。因此,人类对于这个频率范围的声波定位能力是极其有限的。

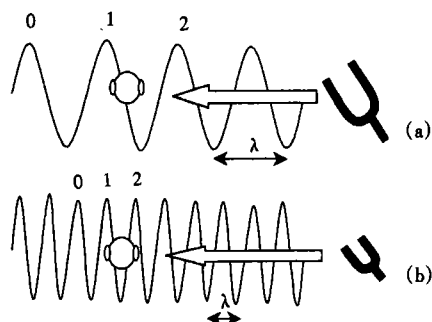


图3 耳间时差,只在长波段给出有用的定位信息。(a)信号来自右方,波形特征如第一号波峰先到达右耳。由于波长大于头部直径的两倍,不会与其他峰混淆。(b)信号仍然来自右方,但其波长小于头部直径的2倍,结果,到达右耳的周期2的每个特征紧跟到达左耳的周期1的相应特征。聆听者自然会断定声源位于左边,这与实际情况正好相反

### 三、两耳差异信息的不精确性

两耳间的时间和强度差异是定位声源的有用的线索,但它们都有很大的局限。在球形脑近似下,这种不精确性是显而易见的。这是因为对于在中分面(两耳间连线的中垂面)上运动的声源来说,到达两耳的信号相同,耳间差异为零。具有假想的球形脑的聆听者不能辨别声源是在前、在后还是在头顶上部。聆听者可以探测到左右运动的声源相对于他只有 $1^\circ$ 的位移,却不能判断声源的前后!这类定位困难与我们平常的经验不符。这个模型还有另外一个问题:如果某一声调或宽带噪声通过耳机被听到,并具有ITD或ILD或二者兼有,结果就像预期的那样,聆听者就会有偏侧的印象—或左或右,根据前面所提到的,声音映像出现在头内部,而且还有弥散与失真。这种感觉,同样与真实的经验(感觉声音在外部)不相符。解决这些问题需要用到另外的声音定位信息,那就是解剖学传递函数。

### 四、解剖学传递函数

来自空间不同方向的声波会被聆听者的外耳、头部、肩部以及上躯干所散射。这种散射将会引起对两耳处信号的声学滤波。这种滤波可用一个复响应函数——解剖学传递函数(ATF)来描述。ATF使从后面来的声波在1000Hz附近被放大,而前方来的波在接近3000Hz时被放大。频率高于4000Hz时,现代物理知识

波长小于 0.10m, 头部的细节, 尤其是外耳, 成为重要的散射体。在 6000Hz 以上的区域, 头上不同的细节呈现明显不同的 ATF, 但仍表现出一些相同的特征。大多数情况下, 随着声源从低于头部向高于头部移动, 在高频方向有一个明显的峰——谷结构。图 4 显示了波源位于前方、后方及头的正上方时的 ATF 谱。对高于头部的声源, 7000Hz 附近的峰是定位的重要信息。这样, 聆听者使用方向依赖性滤波, 解决了诸如前-后混淆及高度的确定等方面的问题。

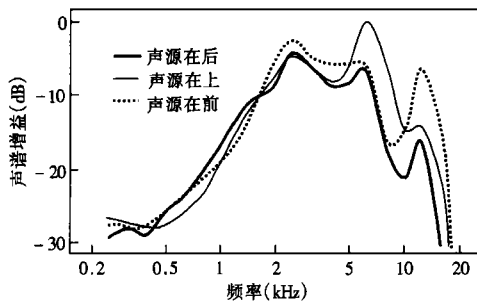


图4 解剖学传递函数。显示了一个人体模型的左耳听到的小型喇叭的声谱。

更进一步的实验表明, 利用虚拟现实技术得到的精确的 ATF 可使声源映像位于头外。但在 ATF 的应用中存在一个明显的问题, 聆听者无从知道某个特征谱是来自方向依赖性滤波还是原来声源本身的一部分。例如, 一个信号在 7000Hz 附近有一个峰, 并不一定来自头顶上部的声源——它可能来自刚好在此频率附近有较强功率的信号源。

这类在声源谱与 ATF 之间出现的混淆直接出现在窄带声源上, 例如带宽为几个半音程的纯音音调或噪声带就是如此。要解决这个问题, 聆听者只能通过调整头部使声源脱离中分面来实现。幸运的是, 日常生活中绝大多数的声音都是宽带的, 因此, 根据频谱的成分, 聆听者可以同时声音进行辨别和定位。目前对于这种定位过程是怎样工作的仍然不是十分清楚。

### 五、实验技巧

通过一些耳机实验, 我们已经了解了有关声音定位方面的知识。因为实验者用耳机能够精确地控制声源, 所以即使用猫、鸟和啮齿动物所做的实验, 也都是让这些动物戴上微型耳机进行的。对于各种频率的声调和各种成分的噪声, 利用 ILD 和 ITD, 通过耳机实验了解了大量有关两耳系统基本接受能力的知识。但是, 由于目前技术条件的限制, 还不能对

每个耳朵的 ATF 进行精确的描述, 因此, 大量声音定位方面的问题尚未解决。例如, 声学测量需要将微型探针型麦克风插入聆听者的耳道内距离鼓膜只有几毫米的地方。一旦麦克风和耳机本身的传递函数通过反转滤波得到补偿, 用这些麦克风测得的传递函数就可对真实情况进行精确的模拟。

精确的滤波要求快速的专门的数字信号处理器与实验用计算机相连接。通过一个电磁头部追踪系统可将聆听者头部的运动计入在内。头部追踪系统包含一个固定的发射机, 它的三个线圈产生低频磁场, 同时, 接收机中也有三个线圈, 固定在聆听者的头部。追踪系统记录头部 6 个自由度的运动, 每秒 60 次。根据头的运动, 用来控制的计算机向快速数字处理器发送指令, 使它再次滤波以保证听觉场景的稳定和真实。这种虚拟现实技术能够合成一个令人信服的声学环境, 并且为受控实验开辟了广阔的领域。

### 六、制造误差

耳机实验能够人为产生自然界不存在的条件, 使我们可以理解不同定位机制的作用。例如, 引入一个向左的 ILD 及一个向右的 ITD, 可以研究对两种定位机制进行对比研究。实验发现 ILD 在高频段占主导地位而 ITD 在低频段起主要作用。

宽带声音的精确合成可使实验更逼真。在合成中制造误差, 例如维持 ATF 幅度不变而不断减小 ITD, 会使映像不断靠近脸部, 同时形成一个模糊点, 一直延伸到耳道最后进入头部。这个过程还可以逆向进行, 即不断地增大 ITD 值, 则映像也沿原来的路线返回。

对声音进行偶然或人为的不精确的合成, 会出现很多不同的效应。有几个一般性规律: 错误将导致映像尺寸变大, 或将映像移到头内, 或产生映像在后面的感觉。避免前-后混淆需要声音合成的极高的精度。降低精度常常会使听者感觉声源在脑后, 如果进一步降低精度, 则会使听者感觉声源位于头的内部。

### 七、场所及反射

实验通常是在消声的房间进行的, 其中所有的声音从声源直线传到聆听者。但是日常生活中, 都存在墙壁、天花板、地板以及其他巨大物体对声音的反射, 这些反射会使波形发生畸变。因此, 房间的反射和回声不可避免地声音定位造成不利的影响, 对 ITD 尤其如此。

ITD 特别容易受到干扰,因为它依赖于两耳间信号的干涉,而回声不包含任何有用的相干信息。如果在一个大房间中,反射波很强,那么 ITD 这信息就变得不可靠了。

相反地,ILD 要好一些。首先,就像耳机实验中显示的那样,两耳强度的比较不依赖于两耳间信号是否相关,这样,当计算 ILD 时,不用考虑神经中枢系统时间选择的细节。当然,ILD 的精度受到室内驻波的不良影响,但这时 ILD 的第二个优势又表现出来:几乎每个反射面都有一种特性,这就是随着频率的增高,声学吸收越来越强;结果,与直射波相比反射波的能量就变得很小了。在存在强回声的环境下进行的实验,并使用 8000Hz 以上频率时,发现聆听者都能正确地响应。对人体模型应用 ILD 和 ITD 进行的测试,经统计决策理论分析表明,如果聆听者完全依赖 ILD 而不用 ITD 进行判断,就可以理解实验中观察到的定位错误模式。这种对定位信息重新加权的策略是完全无意识的。

### 八、优先效应

还有另一个聆听者无意识使用的策略,就是处理室内的失真的定位信息。他们根据最早到达的声波做出定位判断,这称为优先效应。因为最早到达的声波是直射波,超前于随后到达的传递错误信息的反射波,从而具有正确的定位信息。如果没有对最先到达的直射波产生优先效应,定位是不可能的,此时没有任何可用的 ITD 信息,又由于驻波的缘故,声音的强弱不再依赖于声源的远近了。

优先效应的作用常常被看作一扇神经中枢门,起始声波到达时开启,积累定位信息约 1ms,然后关闭,切断后面的定位信息。这种作用似乎颇具戏剧性,尽管一些实验有利于聆听者关注后来的信息,但会被优先效应阻止。另外一个模型将优先效应看作是对支持早期声音的信息进行强加权,因为后来的声音也不能被彻底排除在定位计算之外。

利用一个标准的家用立体声系统再现单声道声源,同样的信号被送到两个扬声器,可以很容易演示优先效应。站在两个扬声器中间,聆听者听到来自前方的声音。向左边的扬声器移动半米会感觉到声音完全来自那个扬声器。对此结果的分析是,每个扬声器向两个耳朵发送信号,每个扬声器产生一个

ILD 和一个 ITD,如图 5 所示。由于优先效应,先来的声音(来自左边的扬声器)占了优势,聆听者感觉声音来自左边。不过,尽管声音似乎单独来自左边的扬声器,但右边的扬声器继续对音量和空间广度感产生影响。通过突然断开右边的扬声器可以验证这种感觉,差别立刻会出现。看来,优先效应并不单纯依赖于两耳差异,它对位于中分面内声源的结构滤波引起的谱差异也有作用。

一个多世纪的工作之后,仍有许多有关声音定

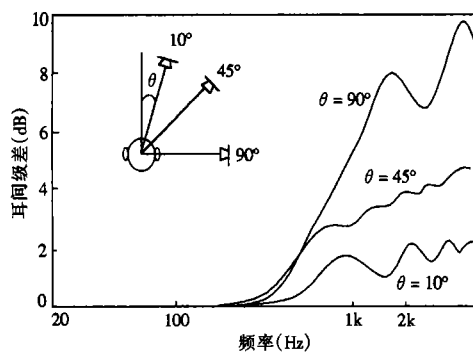


图5 优先效应。聆听者感觉声源位于左边

位方面的问题没有弄清楚,从而为心理声学 and 听觉心理学保留了一个活跃的研究领域。近年来,有关感性的观测报告、两耳处理系统的生理学数据、神经中枢模型的文章呈逐渐增多的趋势。有理由相信,我们会不断深入地了解声音定位的机理。应用有关神经信号处理的新思想建立起的神经中枢模型,将会从根本上改进我们的认识。10年前普遍认为,一般大量的声音定位和个别的优先效应可能是两耳系统早期阶段相互作用的直接结果,与上位橄榄体中的情形一样。近年的研究认为,这个过程分布得更广,利用上位橄榄体这样的大脑外围中心将有关 ILD、ITD、谱、到达的次序等信息发送到更高一级的中心,在那里对接收到的信息进行自恰性和真实性鉴定,或许还会将它们与视觉上得到的信息相比较。因此,声音定位并不简单,需要大量计算。然而,问题变得越复杂,我们研究它的手段就越多越完善。灵活合成真实声源的物理技术,同时探查不同神经区域的心理学实验,大脑成像的更快更精确的方法,以及更实用的计算模型,对人是怎样定位声音的问题,总有一天会给出一个圆满的答案。