

动态听觉定位因素及其在空间声重放的作用*

谢菠菘[†]

(华南理工大学 物理与光电学院声学研究所 广州 510641)

2024 年 7 月 29 日收到

2024 年 9 月 14 日定稿

摘要 听觉定位机理是生理和心理声学的重要研究内容,在空间声技术方面有重要的作用。听觉定位是多因素(如双耳时间差、双耳声级差、个性化谱因素等)协同作用的结果,但各种因素所提供的信息有一定的冗余。听觉侧向定位机理已比较明确,主要是双耳时间差和双耳声级差的作用,特别是低频双耳时间差起主导作用。但前后和垂直定位机理的一些问题有待解决与澄清。虽然目前已确认个性化谱因素和头部转动带来的动态听觉定位因素提供了前后和垂直定位的信息,并且自从 1940 年 Wallach 提出了动态因素的经典假设以来,一直有不少实验从各方面证实了动态因素对听觉定位的作用;但总体上,动态因素的作用特别是对垂直定位的作用常常被低估,甚至较普遍地认为个性化谱因素起了完全主导的作用。听觉定位机理作为空间声技术的心理声学基础,如果假设前提不正确,将无法解析一些重要的实验事实,也无法解决一些普遍存在的问题。目前空间声技术面临的许多问题和误区多数都和动态因素密切相关。文章综述了有关动态听觉定位因素的研究与发展,特别是近年国际上和作者课题组的系列工作,并将其用于分析和解决空间声重放(包括双耳重放、多通路三维空间声和 Ambisonics)的一些重要的共性问题,文中最后讨论了存在问题与研究前景。

关键词 听觉定位, 动态因素, 空间声, 头相关传输函数

PACS: 43.66, 43.38

DOI: 10.12395/0371-0025.2024222

CSTR: 32049.14.11-2065.2024222

Dynamic auditory localization cue and its role on spatial sound reproduction

XIE Bosun[†]

(Acoustic Lab., School of Physics and Optoelectronics, South China University of Technology Guangzhou 510641)

Received Jul. 29, 2024

Revised Sept. 14, 2024

Abstract The mechanism of auditory localization is an important topic on psychoacoustics and physiological acoustics. It also plays a vital role on spatial sound technique. Auditory localization is the consequence of cooperation of multiple cues, but the information provided by various cues is somewhat redundant. The mechanism of lateral localization is relatively clear. Both interaural time difference and interaural level difference contribute to lateral localization, especially the former at low frequency is dominant. In contrast, some problems on the mechanism of front-back and vertical localization should be addressed and clarified. It is already known that both individualized spectral cue and dynamic cue caused by head turning provide information for front-back and vertical localization. Since Wallach proposed the classical hypothesis in 1940, there have been a series of experiment that validated the effect of dynamic cue. However, the contribution of dynamic cue, especially the contribution to vertical localization, is often underestimated, and the spectral cue is generally regarded as the dominant one. The mechanism of auditory localization is essential to spatial sound technique. However, based on inappropriate psychoacoustic hypothesis, it is impossible to explain some important experimental results and solve some general problems, resulting in misunderstanding in the development of spatial sound technique. Recently, many problems in spatial sound technique are closely related to dynamic cue. In this article, the researches and recent progresses on dynamic auditory localization cues, especially a series of related works by international and the author's groups, are reviewed. The mechanism of dynamic auditory localization cue is applied to address some general problems and clarify the

* 国家自然科学基金项目(12174118)资助

[†] 通讯作者: 谢菠菘, phbsxie@scut.edu.cn

misunderstanding on spatial sound, including Ambisonics, multichannel sound and virtual auditory display. Some problems to be solved and prospects are also discussed.

Keywords Auditory localization, Dynamic cue, Spatial sound, Head-related transfer function

引言

听觉定位是人类(以及许多动物)感知外界信息的能力之一。利用声源发出的声音,听觉系统可以估计出声源的空间位置(包括方向和距离)。听觉定位不但可以帮助人类结合视觉寻找目标,还可以帮助人类探索周围的环境,避免潜在的危险。

听觉定位的探索与研究已有相当长的历史。自从瑞利在 100 多年前提出听觉方向定位的双因素理论以来,听觉定位一直是生理和心理声学的重要研究领域,同时也是空间声技术、人工听觉、声学信息的仿生探测、机器人听觉等现代技术的基础^[1]。特别是近二三十年来,空间声技术在通信、计算机与互联网、虚拟与增强现实等前沿领域应用实现的快速发展,听觉定位的研究与应用也面临新的挑战。

听觉方向定位是多个因素综合作用的结果,这些因素的相互作用使方向定位机理非常复杂^[1]。经过多年的研究,声源侧向定位机理已比较明确,主要是低频双耳时间差(ITD)和低频双耳声级差(ILD)因素。另一方面,早在 1940 年 Wallach 就提出动态因素提供了前后和垂直方向定位信息的假设^[2],一些早期的研究也已经表明动态因素和个性化谱因素都提供了定位的信息。近年来垂直定位机理特别是动态定位因素是心理声学领域的一个重要研究方向。这一方面是由于声学实验、信号处理和计算机技术的发展,可以更精确地控制各种实验条件,得到更精细的实验规律与结果。更为重要的是,随着空间听觉机理研究和三维空间声技术的发展,需要重新审视动态因素对定位的作用及其与谱因素的复杂相互作用,有一些重要的问题需要解决与澄清。特别是过去曾经有相当长的时期,动态因素对定位的作用被低估,个性化谱因素的作用被过分强调。由于听觉定位机理是空间声的心理声学基础,上述问题导致的不正确的心理声学假设,加上一些商业利益的驱动,给空间声发展带来了较大的误区。

本文在回顾听觉方向定位机理与空间声的一些基本概念后,详细评述动态因素定位的基本原理、最新研究进展及其在空间声重放中的重要应用,特别是探讨了忽略动态定位因素所带来的问题,并展望了今后的发展方向。

1 听觉方向定位机理与空间声技术

1.1 方向定位机理回顾

声源辐射的声音经直达和反射途径传输,形成空间声场。空气中的声场由时域声压作为空间位置 and 时间的函数,或者频域声压作为空间位置和频率的函数描述。空间声场包含有声源和环境的声学信息。倾听者进入声场后,声波被头部、耳廓等生理结构散射,最后被双耳鼓膜接收,并转换为鼓膜的机械振动。鼓膜的机械振动经中耳传输后,在内耳进行频率分析并转换为神经脉冲,再经听觉神经传送到高层神经系统处理,最终形成听觉事件或感知,包括声源定位等各种空间听觉感知。从声源到听觉感知包含了物理、生理和心理过程。其中,声源到鼓膜的传输是物理过程,将声源和环境的声学信息转换为双耳声压信号。中耳、内耳与高层神经处理是生理过程,涉及神经生理学的机理。而听觉事件和感知的最后形成主要是心理反应的过程。相应地,方向定位机理的研究可以在生理声学和心理学两个不同层面进行。生理声学层面主要研究中耳、内耳和高层神经系统中声学信息的传输与处理,是解释各种听觉机理的最终一步^[3]。但听觉定位的神经生理学机理非常复杂,虽然研究已取得相当的进展,但受到目前神经生理学发展的限制,离最终目标还相差较远。心理学层面则基本上绕过神理学过程,直接(主要通过实验)研究声场或双耳声压所包含的物理因素与听觉感知之间的关系^[1]。在现阶段的研究中,特别是空间声等技术的发展中,心理学占有重要的地位^[4]。当然,生理声学与心理学是紧密关联的,一方面生理声学的许多结果需要通过心理学方法验证;另一方面,心理学的研究为进一步的生理学研究提供基础。

作为听觉物理过程最简单的情况,在自由场单一点声源的情况下,头部、耳廓等生理结构对声波的散射作用及最终的双耳声压由头相关传输函数(HRTF)所决定。HRTF 定义为自由场情况下从点声源到双耳的归一化频域声学传输函数^[5]:

$$H_a = H_a(r_s, \Omega_s, f, a) = \frac{P_a(r_s, \Omega_s, f, a)}{P_0(r_s, f)}, \quad (1)$$

其中, r_s 是声源相对头中心的距离; Ω_s 是声源相对头中心的方向; f 是频率; a 为个性化生理参数; α 为 L 或 R, 表示左耳或右耳; P_α 表示声源在左耳或右耳产生的声压; P_0 是头移开后点声源在原头中心位置处产生的自由场频域复数声压。由于在大约 14 kHz 以下的频率, 声波在耳道的传输近似是一维的, 双耳声压 P_α 的测量点可以选择从封闭耳道口到鼓膜之间的任意位置^[6]。

一般情况下, HRTF 和声源的距离、方向以及频率有关。由于不同个体的头部、耳廓等生理结构的形状、尺寸不同, HRTF 还与个体有关, 这里用一组参数 a 表示个性化生理参数 (不考虑个性化时将略去该组参数)。在声源距离 $r_s \geq 1.0$ m 的远场情况下, HRTF 近似与距离无关。头相关脉冲响应 (HRIR) 是 HRTF 的时域表示, 由逆傅里叶变换与 HRTF 联系。

由 HRTF 可以严格计算出双耳声压及其所包含的与自由场听觉定位有关的物理因素, 并通过心理声学实验探讨这些因素与听觉定位的关系。在一定的条件下, 也可采用简化的头部模型对听觉定位因素进行分析。本文附录 A 给出计算各种定位因素的方法, 并规定了所采用的坐标系。

大量的物理和心理声学研究表明^[7], 环境反射声的强度和延时满足优先效应的条件下, 声源方向定位与自由场的情况类似, 主要由直达声的信息所决定。当然, 增加反射声的强度会影响方向定位的准确性以至难以定位。自由场声源的方向定位因素主要包括双耳时间差 (Interaural Time Difference, ITD)、双耳声级差 (Interaural Level Difference, ILD)、谱因素 (Spectral Cue) 和动态因素 (Dynamic Cue)^[1]。

ITD 是指声源到双耳传输的时间差。当声源位于中垂面时, 它到双耳的距离相等, ITD 为零。但声源偏离中垂面时, 其到左右耳的距离不同, 因而存在与侧向声源方向有关的传输时间差。图 1 是按式 (A8) 和刚性球形头部模型 (半径 $a = 0.0875$ m) 的 HRTF 数据计算得到的左半水平的 ITD, 计算的频率范围是 $f \leq 1.5$ kHz。随着声源偏离正前方 $\theta_s = 0^\circ$ 或正后方 $\theta_s = 180^\circ$ 时, ITD 增加, 并在侧向 $\theta_s = 90^\circ$ 达到最大。同时, ITD 是 (近似) 前后对称的。

另一方面, 当声源偏离中垂面时, 由于头部对声波的阴影和散射作用, 特别在高频, 与声源异侧耳处的声压受到衰减, 而与声源同侧耳的声压有一定的提升, 形成与声源方向和频率有关的 ILD。图 2 是按式 (A9) 和刚性头部模型计算得到的左半水平面的 ILD, 图中以 ka 作为与频率有关的参数, k 为波数, a 为头半径。当声源偏离正前方 $\theta_s = 0^\circ$ 和正后方

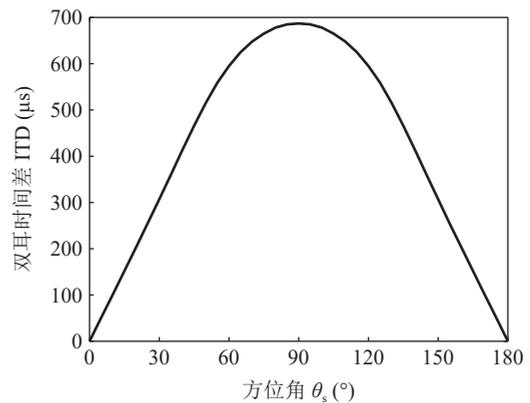


图 1 左半水平面的双耳时间差

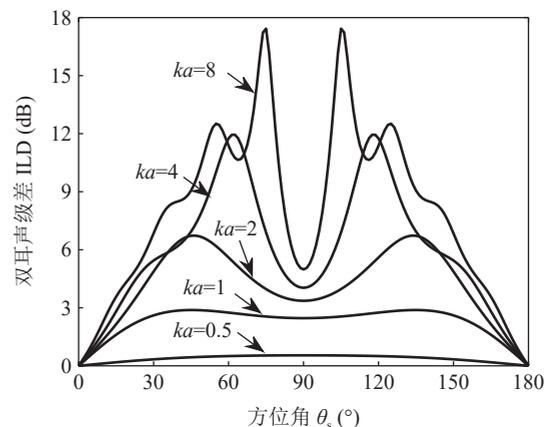


图 2 左半水平面的双耳声级差

$\theta_s = 180^\circ$ 时, $ILD \neq 0$, 并且是声源方位角与频率的复杂函数。ILD 也是 (近似) 前后对称的。

低频 ITD (主要是 1.5 kHz 以下的双耳相延时差) 与高频 (2~3 kHz 以上) 的 ILD 是声源侧向定位的主要因素, 这就是瑞利有关听觉定位的双因素理论, 并被大量的心理声学实验所证实。

但 ITD 与 ILD 提供的信息并不足以完全确定声源的空间位置。略去头部弯曲表面的影响, 将双耳看成是自由空间的两点。根据对称性可知, 有无限多个空间点集组成一个锥形表面, 称为混乱锥 (Cone of Confusion), 如图 3 所示。在混乱锥上不同位置声源所产生的 ITD 是相同的。类似地, 如果采用简化的刚性球形头部模型, 在声源距离远大于头部半径的远场近似下, 同样存在无限个空间点组成的集合, 这些点上的声源所产生的 ILD 也相等。虽然实际双耳并非自由空间两点, 头部也并非一个球体, 但近似对称性引起的 ITD 和 ILD 与空间方向的非单值的函数关系依然存在^[8]。这使得 ITD 和 ILD 提供的信息最多能确定声源所在的“混乱锥”, 并没有提供前后和垂直定位的准确信息。

生理结构对入射声波的散射改变了每个耳所接

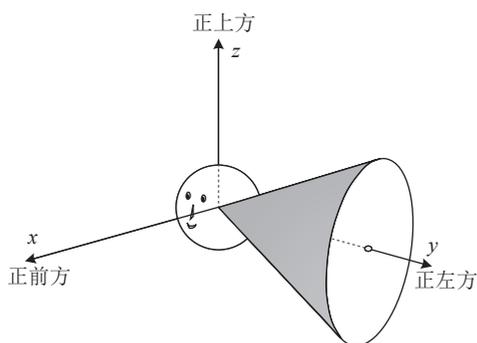


图3 混乱锥

收到声波的频谱,这种频谱改变是和声源方向有关的,并称为谱因素(Spectral Cue)。谱因素带来了声源定位信息。特别是耳廓对高频声波的散射(5~6 kHz以上,耳廓尺度与声波波长相比拟)所带来的谱因素提供了前后和垂直定位的信息^[9]。另外,当声源偏离中垂面后,躯干(特别是肩部)对声波散射带来的低频(3 kHz以下)谱因素也可能带来弱的垂直定位信息^[10]。必须指出,由于不同个体的耳廓等生理结构、尺寸的差异,对入射声波的散射不同,高频因素是具有非常个性化特征的定位因素。图4是根据中国人样本数据得到的10名受试者对正前方声源的左耳HRTF幅度谱。

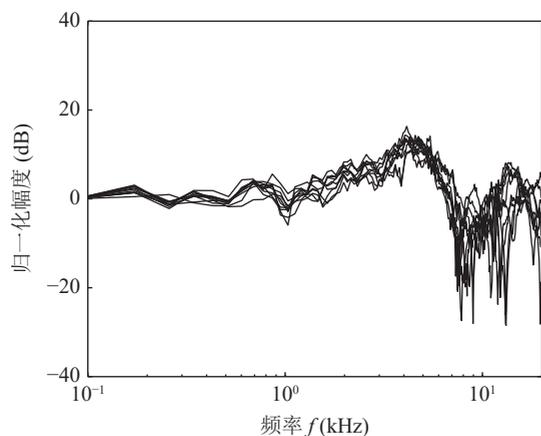


图4 10名受试者对正前方声源的左耳HRTF幅度谱

另一方面,在声源产生的声场中,当倾听者头部转动后,声源相对倾听者的方向发生了变化,引起双耳声压及其包含的ITD、ILD和谱因素都发生变化,这些变化与声源方向和头部转动的方式、角度有关,从而带来了前后和垂直定位信息。因此谱和动态因素是分辨混乱锥上的声源位置,特别是前后和垂直定位的重要因素。

听觉方向定位是上述多个因素综合作用的结果。不同因素对方向定位的贡献不同,作用的频率范围也不同,并且它们之间还存在相互作用。这使

得方向定位不但和声源信号的类型有关,且其机理非常复杂。这也是多年来听觉定位一直是心理声学领域前沿问题的原因之一,特别是前后和垂直定位的详细机理与应用是目前的研究重点。总体上,多个因素的协同作用增强定位。但各因素提供的信息有一定的冗余,当部分因素缺失甚至冲突时,听觉系统还可根据主导的因素或者一致性好的因素进行定位。如果缺失或冲突的因素过多,则定位精确性下降以至不能定位。例如,当声信号包含有低频成分时,低频ITD对侧向定位起主导作用而可忽略冲突的ILD因素^[11]。但另一方面,对于谱因素和动态因素在前后和垂直定位中的作用及其冗余性,有不少问题需要澄清。

听觉距离感知的机理更加复杂,也是多个因素共同作用的结果^[12]。自由场情况下包括随距离变化的响度因素、HRTF因素(特别是非中垂面 $r_s < 1.0$ m近场声源产生的随距离变化的低频 $f < 3$ kHz的ILD)、空气吸收引起的声波高频谱衰减(弱因素,10多米的距离范围内可忽略)。在反射环境下,直达与反射声能比也是一个重要的距离感知因素。

1.2 空间声的原理与心理声学

听觉定位心理声学的一个重要应用领域是空间声技术。空间声的目的是检拾或模拟、传输或记录、重放声音的空间信息,给倾听者产生期望的空间听觉感知。国际上多年来的研究已发展了多种不同的空间声技术,并得到实际应用。从物理上看,声音空间信息的重放可以在空间声场的层面实施,也可以在双耳声压的层面实施。相应地,根据其物理原理,空间声技术可分为三大类^[4]:

(1) 第1类,物理声场的精确重构。这类系统是在一定的空间区域内实现物理声场的精确重构,产生一个与期望或目标声场完全相同或尽可能接近的物理声压分布。倾听者可以在重放声场中获得期望的声音空间信息,产生相应的空间听觉感知。这类系统的代表包括高阶Ambisonics和波场合成,前者是基于声场的多极(空间谐波)展开与逐阶逼近重构^[13],后者是基于声场的惠更斯原理与基尔霍夫-亥姆霍兹积分^[14-15]。

(2) 第2类,心理声学物理声场的近似重放。这类系统并不追求物理声场的精确重构,在一定条件下可能是目标声场的一种粗略近似重放。利用特定的心理声学原理,在一定程度上可以得到与目标声场类似的空间听觉事件或感知。传统的两通路立体声和各种多通路声是这类系统的典型代表^[4,16]。

(3) 第 3 类, 双耳与虚拟听觉重放。双耳声压(信号)包含了声场的听觉信息。通过精确重放双耳声信号也可以重放声音的空间信息, 从而产生相应的空间听觉事件或感知。传统的双耳检拾与重放以及虚拟听觉重放就是这类系统的例子^[5]。其中后者采用信号处理的方法模拟声源到双耳的声学传输(包括直达和反射声的途径), 从而模拟出期望的双耳声信号, 并(主要)用耳机重放。虚拟听觉重放可再细分为稳态重放与动态重放两大类。前者假定倾听者头部固定不动, 信号处理只模拟声源到双耳的稳态传输过程。而后者采用头部跟踪器实时检测倾听者头部的运动, 并刷新声源到双耳传输的模拟, 从而带来动态听觉信息。

对于第 1 类空间声系统, 虽然其物理概念是完美的, 但受声场空间采样与重构理论的限制^[4], 如果要在整个可听声频率范围(上限至 20 kHz)内精确重构物理声场, 则需要非常多的扬声器而难以实现(有源噪声控制也有类似问题)。实际中通常只能在一定的频率范围(上限在 2~4 kHz 的量级)、一定的听音区域精确重构物理声场, 这就需要利用空间听觉的心理声学原理产生各种不同的听觉事件与感知。对于第 2 类系统, 其本来就是基于各种空间听觉的心理声学原理而发展的。事实上第 2 类与第 1 类系统之间并没有严格的区分界限, 它们之间有一个连续的过渡过程, 由低阶 Ambisonics 过渡到高阶 Ambisonics 就是一个典型的例子。而第 3 类系统, 空间听觉的心理声学原理也经常用于其信号处理的简化。

因此, 空间听觉(包括方向定位)的心理声学原理在上述三类系统起了重要的作用, 成为分析、设计和评价系统的基础。各种不同定位因素提供的信息的冗余性给空间声的发展带来方便与挑战。空间声不一定要精确重放所有信息, 可以产生部分主导信息而略去冗余信息, 使系统得到简化。特别是近年研究与应用正朝着面向目标的三维空间声技术发展^[17], 面向目标的声系统结构可以灵活应用各种不同原理的空间声技术。因而正确理解和应用三维空间听觉定位机理(包括前后和垂直方向定位机理)对目前空间声的发展特别重要, 否则容易产生误区。

2 动态定位因素的假设与实验

2.1 Wallach 有关动态定位因素的假设

从物理上看, 人类头部可作三个自由度的转动, 即绕垂直、前后和左右轴的转动。当声源方向固定不动时, 任一自由度的转动都会引起双耳声压的改

变。早在 1920 年代至 1930 年初, 已有研究注意到动态因素对前后定位的作用^[1]。例如在 Young 的实验中^[18], 采用一对喇叭形状的声波接收器, 并通过一对管道连接到耳道; 喇叭形状接收器是不可动的。结果表明, 去除动态因素会发生前后混乱。

1940 年, Wallach 定量分析了头部转动引起的双耳因素(主要是 ITD)的变化, 对动态因素与声源方向定位提出了以下假设^[2]: (1) 头部绕垂直轴转动提供了前后定位信息; (2) 头部绕垂直轴转动也提供了垂直偏离水平面信息; (3) 头部绕前后轴转动提供了分辨上下方向的附加信息。

事实上, 头部转动引起双耳声压变化, 从而引起 ITD、ILD 和谱因素的变化, 它们都可能是潜在的动态定位因素。这些潜在的动态定位因素可按附录 A 的方法用 HRTF 模型精确计算得到。如果只考虑头部转动引起的低频 ITD 变化, 则可以采用一种简单的几何模型分析, 得到 Wallach 假设的数学表述^[19-20], 附录 A 也给出了简单的推导。但从逻辑上看, 无论是 HRTF 模型还是简单的几何模型, 只能说明动态因素提供了可能的前后和垂直定位信息。多年来方向定位机理的一个研究重点, 就是要通过心理声学检验听觉系统能否利用这些信息进行定位, 以及各动态因素之间及其与谱因素之间的交互作用。

听觉定位机理的心理声学实验需要控制双耳声压包含的各种物理因素, 如不同频率范围的 ITD、ILD、谱因素和动态因素等。传统的心理声学实验是在真实声源与环境下进行的, 但不易精确灵活控制双耳声压的物理因素。利用空间声重放系统, 特别是上述第 3 类的虚拟听觉重放系统, 可以方便地通过信号处理控制双耳声压的各种物理因素。这是一种“虚拟现实”环境下的心理声学实验方法。空间声重放系统不但是空间听觉心理声学原理的一个实际应用, 近年也作为实验工具逐渐广泛用于听觉的科学研究。

为了检验头部绕垂直轴转动在前后和垂直定位中的作用, 在 Wallach 的经典实验中^[2], 多个声源(扬声器)布置在前半水平圆弧上, 每次只有一个声源发声。如果发声的声源是固定的, 与头部转动无关, 则受试者感知的方向与实际的声源方向一致。另一方面, 如图 5 所示, 如果头部绕垂直轴的转动控制声源开关, 使得发声的声源随头部的方向变化。当声源方向变化的角速度是头部转动角速度的两倍, 则 ITD 随头部转动的变化率与水平面后方镜像位置的声源的结果是一致的, 倾听者会产生后方镜像位置虚拟源的感知, 而不是运动虚拟源。当声源方向变

化的角速度与头部转动的角速度相同,使倾听者一直面对声源时,ITD 是固定的(零值),不随头部转动而变化。这与正上方位置声源的情况一致,倾听者会产生上方虚拟源的感知。该实验结果说明,动态因素主导前后定位,同时也提供垂直偏离水平面的定位因素。根据实验结果,Wallach 也总结出最小位移原理(Principle of Least Displacement),即“听觉偏爱于感知最小的源在空间的运动”^[2,21]。

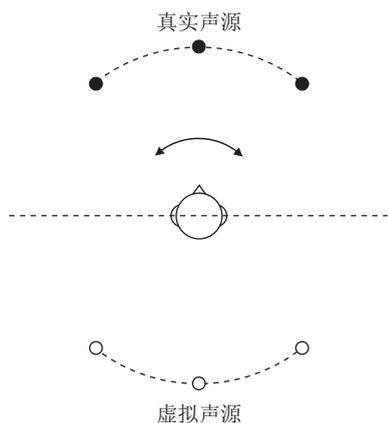


图5 Wallach 的经典实验示意图

2.2 有关动态定位因素的实验

在 Wallach 的实验之后,不断有实验验证了头部转动对前后定位的作用,例如 Thurlow 和 Runge 在真实声源定位实验中证实头部绕垂直轴转动可以明显减少前后混乱,但头部绕前后轴转动只能略为减少前后混乱^[22]。Blauert 的专著总结了一些经典的实验及其结果^[1]。而近二十多年以来,由于空间声技术的发展需要,国际上用不同的方法重新开展更精确的实验,取得了更确切的实验结果。

Bronkhorst 采用真实声源和虚拟听觉重放的方法,证实对持续时间较长的信号,受试者头部转向源方向时前后混乱率很低;但对于持续时间短的信号,受试者的头来不及转动,前后混乱率高^[23]。Wightman 和 Kistler 采用真实声源和动态虚拟听觉重放系统,并以高斯白噪声作为信号,对比了限制和允许受试者头部转动情况下的定位。结果表明,与限制头部转动的情况相比较,头部转动几乎完全消除了前后混乱^[24]。Brimjoin 和 Akeroyd 的实验是 Wallach 经典实验的现代版,采用光学摄像检测倾听者头位置和计算机控制声源发声,以得到更好的控制精度,其结果也与 Wallach 的经典实验类似。但该实验同时表明,随着信号中 500 Hz 以上成分能量的增加,镜像位置的虚拟源将会变得不稳定。因而,头部绕垂直轴转动的动态因素主要对低频前后定位有贡献^[25]。

为了检验动态 ITD、ILD 和稳态谱因素在前后定位的作用,Macpherson 进行了两组实验^[21]。第 1 组实验中,多个目标声源(扬声器)均匀地布置在水平面的圆周上,每次选定一个目标声源发声。实验对比了受试者头部固定(面向前方)和头部连续转动时的感知声源方向。其中头部转动的情况下,通过跟踪器检测头部取向,只有头部的取向落入一定角度窗口范围内,选定的目标声源才发声。第 2 组实验中,采用耳机动态虚拟听觉重放的方法,并使用球形头部模型的 HRTF 数据(去除耳廓引起的谱因素)。实验信号包括宽带(0.5~16 kHz)、低频带(0.5~1.0 kHz)和 4~16 kHz 噪声信号,以包括或排除不同频带的定位因素。实验结果表明,与动态 ILD 变化比较,低频动态 ITD 变化是主要的前后定位因素,且只在谱因素不起作用时起主导作用。

为了进一步检验动态 ILD 的贡献,Pöntynen 和 Salminen 在自由场真实声源和动态虚拟听觉重放环境下进行水平面的定位实验^[26]。其中真实声源的实验方法与 Macpherson 的方法类似。实验信号为频率 0.5 kHz, 2 kHz, 4 kHz, 8 kHz 正弦信号。正弦信号可以排除谱因素的影响,而不同的正弦信号频率可以包括或排除不同频带的定位因素。在动态虚拟听觉重放环境下,也可以通过信号处理的方法排除动态 ITD 或 ILD 因素。实验结果提示动态 ILD 主要是在高频、ILD 随头部转动单调变化的条件下才对前后定位有贡献,但在自由场真实声源且窄带信号的情况下并非如此。而 Pöntynen 等采用真实声源(扬声器阵列)和基于振幅矢量的信号混合方法,并采用 500 Hz 低通、4 kHz 高通、低通与高通组合、全频带的棕噪声四种信号,以包括或排除不同频带的定位因素。心理声学实验结果表明,当谱因素和动态因素提供的前后定位信息有冲突时,听觉上会感知到前后分离的两个虚拟源^[27]。

Yost 最近利用水平面真实声源进行了实验^[28]。原始信号是中心频率分别为 0.5 kHz 和 5 kHz 的 2 倍频带宽噪声,然后对原始信号进行带宽从 0.5 至 2 倍中心频率的带通滤波,并对带通滤波信号幅度谱进行 ± 1.5 dB 的随机涨落,或者对其中心频率 0.05 倍的涨落。实验结果表明,随机信号谱会增加定位的前后混乱,但头转动会减少这种混乱。该实验表明谱和动态因素的信息有一定的冗余性。

继 Wallach 的假设和实验之后,早期只有少量实验研究动态因素在垂直方向定位中的作用^[22,29-30]。但是这些早期实验并不能完全排除稳态高频谱因素的影响。因而动态因素在垂直方向定位中的作用早

期没有得到严格的实验证实,其结论也常被忽视。在后来的实际空间声应用中也经常误认为谱因素是主要甚至唯一的垂直定位因素。

Perrett 和 Noble 用真实声源进行中垂面和侧垂面的定位实验^[31]。用不同的方法对可能的定位因素进行控制,包括:头部固定或转动而去除或引入动态因素;将开放的短管插入耳道口而破坏高频谱因素;采用滤波器控制噪声信号频带(低通、高通和宽带),从而保留或去除不同频带的谱或动态因素。实验结果表明,破坏或去除高频谱因素后,稳态情况将不能垂直定位。但引入头部绕垂直轴转动的动态因素后,只要信号包含低频(2 kHz 以下)成分,垂直定位变得可能。因而该实验严格证实了头部绕垂直轴转动的低频动态 ITD 对垂直定位的作用。另外,Perrett 和 Noble 的实验还发现,即使头部固定不动,仍然可以在中垂面和侧垂面进行上下定位,这可能是躯干低频散射也提供了分辨上下的信息。当然,头部绕前后轴的转动会进一步增强上下定位。

Toshima 和 Aoki 将人工头安装在机械装置上进行现场检拾,得到的双耳信号经放大后用耳机重放,所用信号是高斯白噪声^[32]。头部跟踪器实时检测受试者头部三个自由度的转动,并通过计算机控制机械装置上人工头的转动,从而在双耳信号中引入动态因素。对比动态和稳态(人工头不动)的定位结果,动态因素确实改善了前后和垂直定位。

Martens 等让行走的听力正常受试者佩戴上双耳助听器,并交换佩戴助听器的左右耳机^[33]。此时,左右耳的动态因素与正常听觉情况下是完全相反的,相当于对实际声源作左右、前后和上下空间反演后的虚拟声源产生的动态因素。实验中倾听者感知到空间位置反演(而不是实际声源位置)的虚拟源,因而证实了动态因素在前后和垂直定位中的作用。该实验采用的是语言信号,一般情况下语言信号的功率谱在 4 kHz 以上的高频已经有较大的衰减,因而对前后和垂直定位起主要作用的应该是动态因素而不是谱因素。

McLachlan 等在虚拟听觉重放和自由场真实声源环境下研究了头部小转动($\pm 10^\circ$)时,谱因素的动态变化对前后和垂直定位的贡献^[34]。采用白噪声作为信号。控制的实验条件包括以下组合:(1)头部固定/绕垂直轴转动/绕左右轴转动;(2)不同的信号条件(去除或保留动态 ITD、动态 ILD、动态谱因素)。结果表明,头部绕垂直轴转动(引起大的动态 ITD 变化)明显减少前后混乱,而头部绕左右轴转动(引起大的动态谱变化)的效果不大。即使存在动态 ITD

因素,谱因素也能改善定位。受试者在头运动前利用谱因素定位,但利用小的头运动的动态因素时不会利用动态谱因素。

为了定量研究动态因素和谱因素对垂直定位的贡献,本课题组以虚拟听觉重放系统为平台,研究了以下三类条件组合下的双耳极方位角 $\theta = 0^\circ$ 的中垂面以及 $\theta = -45^\circ$ 垂直平面的虚拟源定位^[35]:(1)稳态/动态重放(无/有动态因素);(2)无耳廓/非个性化/个性化 HRTF 双耳信号合成处理(无谱因素/非个性化谱因素/个性化谱因素);(3)全频带/1.5 kHz 低通滤波粉红噪声。图 6 为中垂面、全频带粉红噪声信号时,8 名受试者、每名受试者 4 次重复定位的散点图,其中横坐标表示目标虚拟源方向(双耳极仰角),纵坐标表示感知虚拟源方向。理想情况下,定位的散点应集中在图中对角线附近。在稳态重放的情况下,无耳廓 HRTF 和非个性化 HRTF 都给出了混乱的结果;而个性化 HRTF 给出了略有改善的结果。在动态重放的情况下,即使是无耳廓 HRTF 也表现出感知极仰角随目标极仰角变化的趋势;非个性化 HRTF 已经可以得到较为理想的结果;而个性化 HRTF 可以得到理想的结果。统计学上的方差分析也证明了上述结论。

另一方面,对稳态重放,普遍出现了前后和上下混乱的情况。无耳廓 HRTF 的前后和上下混乱率分别为 49.4% 和 36.1%;非个性化 HRTF 可以分别将前后和上下混乱率减少到 39.1% 和 33.0%,而个性化 HRTF 可以进一步将前后和上下混乱率分别减少至 26.3% 和 15.6%。另外,稳态重放还出现了少部分头中定位的感知。但对于动态重放的情况,三种 HRTF 的前后混乱率都非常低,在 4.7% 至 0.0% 之间;而对无耳廓、非个性化和个性化 HRTF,上下混乱率分别为 24.0%, 10.4%, 1.0%。因而,动态因素和谱因素(特别是个性化谱因素)对前后和上下定位都有作用,但动态因素对前后定位的作用更为突出,动态因素与谱因素的协同作用基本消除了上下混乱。

为了区分前后、上下混乱和垂直偏离水平面的感知仰角畸变,可以对图 6 中出现前后、上下混乱的数据进行空间反演,消除混乱后再绘制出感知极仰角的散点图。动态重放情况下混乱率本来就不高,空间反演后结果变化不大。对于稳态重放、无耳廓和非个性化 HRTF 的情况,空间反演后极仰角定位的结果还是有一定的错误。对于个性化 HRTF 的情况,结果如图 7 所示。因而,个性化谱因素已经可以在一定程度上产生垂直偏离水平面的虚拟源感知,虽然存在一定的前后和上下混乱率。

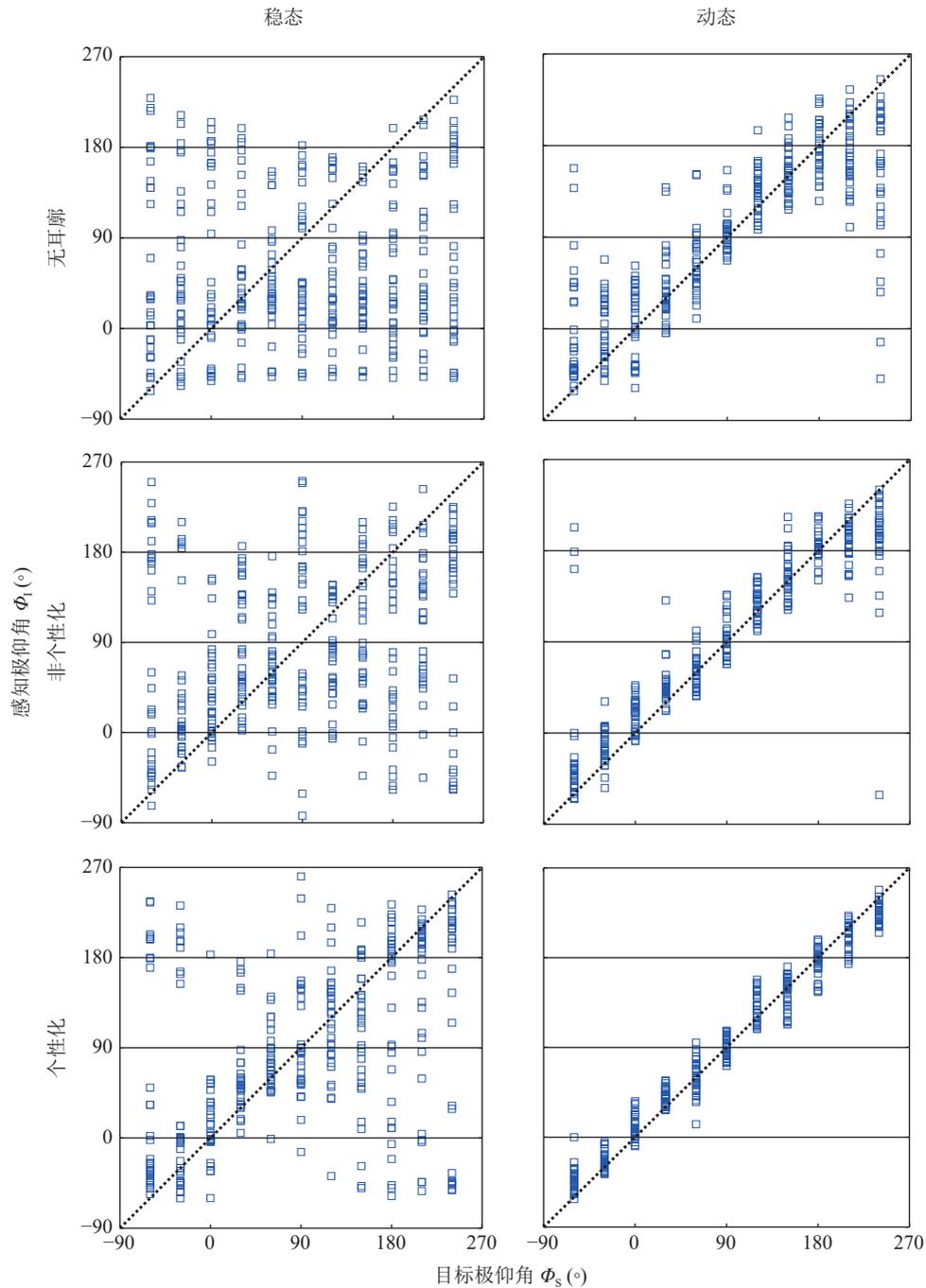


图6 不同条件下中垂面上虚拟源定位的散点图(全频带粉红噪声, 8名受试者, 每名受试者4次重复定位)

另外, 采用 1.5 kHz 低通粉红噪声信号(完全排除高频谱因素)的实验也表明, 在动态重放的情况下, 采用三种 HRTF 都可以一定程度上产生垂直偏离水平面的虚拟源感知。而对极方位角 $\theta = -45^\circ$ 垂直面的定位实验结果也和中垂面的情况类似。因而, 该实验定量证实了低频动态因素(主要是动态 ITD)和稳态高频谱因素对前后、垂直定位的作用, 以及它们所提供信息的冗余性和交互作用。

上述实验研究涉及动态和谱因素之一缺失(或者谱因素不准确)的定位情况。本课题组进一步研

究了动态和谱因素冲突情况下的中垂面前后和垂直定位问题^[20]。实验也是基于动态虚拟听觉重放系统平台。通过人工修改信号处理使用的 HRTF 数据, 产生了具有冲突动态 ITD 和谱因素的双耳信号。实验信号包括全频带和 3.0 kHz 低滤波波粉红噪声, 以探索不同频带动态定位因素的贡献。心理声学的实验结果表明, 动态 ITD 和稳态谱因素分别对低频和高频的垂直定位有贡献。对全频带信号(粉红噪声), 冲突的动态 ITD 和谱因素通常会导致不同仰角的两个分开的虚拟源, 分别对应信号的低频和高分成分,

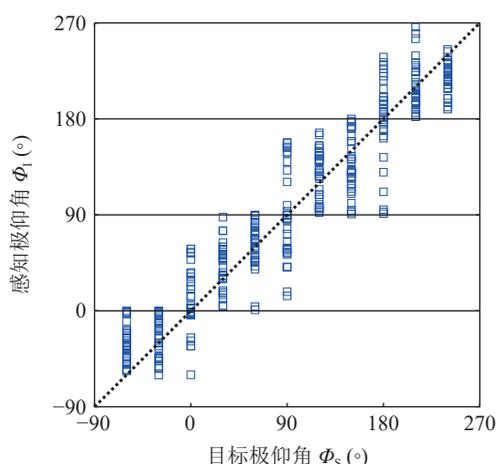


图7 空间反演消除混乱后的定位散点图(稳态重放,个性化HRTF,粉红噪声)

不会产生空间上融合的听觉事件。

关于提供可靠动态因素所需的头部转动的程度问题, McAnally 和 Martin 的实验研究指出^[36], 小的头转动可以减小定位的前后混乱, 而大约需要 16° 到 32° 的头转动才能明显地减少垂直定位错误和几乎完全消除混乱。

最后要提及的是动态听觉定位因素与运动声源的感知问题。对空间位置固定的声源, 头部沿某一方向的转动可以引起双耳声压和各种听觉定位因素的动态变化。但从相对运动的角度, 当头部固定时, 声源沿相反方向的运动也会引起类似的双耳声压与各种听觉定位因素的变化。前者通常产生空间位置固定的声源感知, 而后者通常产生运动声源的感知。因而高层神经系统需要组合倾听者的非听觉空间信息而区分两种情况。这些非听觉空间信息可能是前庭提供的倾听者自运动信息^[21], 或者是来自其他多模态感知的类似自运动的空间信息^[24,37], 这方面需更深入地研究。

2.3 动态定位因素的实验结果总结

以上诸实验已跨越了近百年的历史, 不同实验得到的结果有一些差异, 但总体上基本证实了 Wallach 假设的动态因素对前后和垂直定位的作用, 并且其结果和信号的类型(频谱)有关。综合上述实验结果, 特别是近年的现代实验结果, 并结合过去对稳态谱因素的实验研究, 可以总结出以下规律:

(1) 动态因素与稳态谱因素都对前后和垂直定位有贡献, 但其作用的频带不同。动态因素主导低频定位; 稳态谱因素主导高频定位。两类因素的协同作用会增强定位。

(2) 动态因素主要是头部绕垂直轴转动引起的

低频 ITD 变化, 其对前后和垂直偏离水平面的定位有重要贡献。头部绕前后轴转动引起的低频 ITD 变化对分辨上下方向有一定的贡献。高频动态 ILD 的贡献相对弱, 高频谱因素动态变化的贡献应该可以忽略。

(3) 稳态谱因素主要是耳廓等引起的 $5\sim 6$ kHz 以上高频谱因素, 它对前后和垂直定位有重要贡献。躯干等的散射带来的 3 kHz 以下的低频谱因素对分辨上下方向可能有一定的贡献。

(4) 动态和稳态谱因素提供的信息有一定的冗余。当其中之一缺失(或不准确)时, 单靠另一因素仍可能在一定程度上定位, 但错误增加。具体地, 当存在动态因素时, 缺失谱因素仍可一定程度上定位, 增加非个性化谱因素即可较好地定位。但动态因素缺失时, 需要准确的个性化谱因素才能一定程度上定位。

(5) 当动态因素和谱因素提供的信息有冲突时, 听觉上可能会感知到两个分开的虚拟源, 不会产生空间上融合的听觉事件。

(6) 前庭或多模态感知可能提供了区分倾听者运动与声源运动的信息。

特别值得注意的是, 上述第 4 条规律表明低频动态 ITD 提供的信息比较稳定, 而个性化高频(波长短)谱因素的稳定性较差, 因而可以推测人类听觉应该更依靠相对稳定的低频动态因素进行定位。从生物进化的角度看, 该推测应是合理的。但也正是因为不同因素提供信息的冗余性, 在不适当的心理声学分析中, 很容易因为一种因素的作用而错误地否定了另一因素的作用。

3 多扬声器重放与动态因素

3.1 多扬声器重放前后和垂直定位信息基本分析

如前所述, 从物理原理上看, 多扬声器的空间声系统主要包括第 1 类物理声场的精确重构技术和第 2 类心理声学与物理声场的近似重放技术。目前这两类技术正在向三维空间声重放的方向发展。所面临的一个共同且重要的问题是如何在重放中产生前后和垂直定位信息。

对影院、厅堂等大尺度空间声重放的应用, 多数采用第 2 类技术, 即把布置在不同方向上的多个扬声器分组, 利用一个或方向相近的一组扬声器近似产生期望的声场(期望声源方向的离散空间粗略采样), 从而近似产生特定方向定位和其他期望的空间听觉信息。

对于家庭等小尺度空间声重放的应用, 两类方法都需要采用两个或更多扬声器声波相关叠加的方

法(产生某一扬声器方向上虚拟源感知的情况除外),在局部区域重构目标声场或者期望的空间听觉(包括声源定位)信息。但受空间采样定理的限制^[4],第1类系统实际上很难在2~4 kHz以上的高频范围内准确重构目标声场,第2类系统也不能准确产生高频的声源定位因素(目标虚拟源与某一扬声器方向重合的情况除外)。因而两类系统都要依靠合成定位的心理声学原理,也就是利用各种听觉定位信息的冗余性,在多扬声器产生的相关叠加声场中产生主导的低频方向定位信息,从而产生相应的听觉定位感知。

当信号含有1.5 kHz以下的低频成分时,低频ITD是侧向定位的主导因素;头部转动引起的动态ITD变化是低频前后和垂直定位的主要因素。因而只要在多扬声器重放中较准确地产生期望的低频ITD及其动态变化,即可在一定程度上产生期望的空间虚拟源定位感知。由于语言和一些音乐等实际声音信号的功率谱在4 kHz以上时高频成分已有较大衰减,这时多扬声器重放也不会产生明显冲突的高频定位因素,从而影响定位感知。这就是多扬声器重放产生不同空间方向虚拟源的基本心理声学原理^[4,19]。相应地,在多扬声器重放的设计和评价中,首先应对重放产生的ITD及其动态变化进行分析,并和单声源的情况比较。当然,进一步将重放产生的其他高频因素(如ILD、谱因素、双耳响度谱等)与目标声场的情况比较,可以对重放虚拟源的感知质量和其他属性(如音色染色)进行分析^[38]。近年一些研究的主要问题是动态因素在多扬声器重放中的作用得不到重视,期望多扬声器能合成高频谱因素,并作为前后和垂直定位的主导因素,这很容易引起误解。有关动态因素在听觉定位中的作用的最新研究进展需要我们重新考虑与审视多扬声器重放中前后和垂直定位问题。

事实上,在多通路声发展的早期,就有研究通过分析多扬声器重放产生的低频ITD及其随头部绕垂直轴转动的变化率,得到了水平面多扬声器重放的合成虚拟源定位公式^[39]。基于Wallach的假设和简化的头部模型,谢兴甫分析了空间多扬声器产生的低频双耳时间差及其随头部绕垂直、前后轴转动的变化率,并与单声源的情况比较,得到了多扬声器三维空间声重放合成虚拟源定位的一组公式^[40]。本课题组对该组公式进行了修正与推广,使其适用于中垂面扬声器布置的情况(见附录A)^[19,41]。另一方面,Gerzon早年的研究^[42],通过对多声源叠加声场的速度与能量矢量的分析,得到相应的合成虚拟源定位理论。Gerzon的速度矢量分析方法与上述简化头

部模型的ITD及其动态变化的分析基本是等价的。但ITD及其动态变化分析的心理声学意义更加明显。

简化头部模型分析略去了头部对声波的散射作用,因而只适合700 Hz以下的低频情况。更精确的分析应采用附录A给出的HRTF模型计算多扬声器重放产生的ITD及其动态变化(以及其他定位因素),并和目标单声源的情况比较(见文献[4]及附录A)。以上简化和HRTF模型计算是多扬声器空间声重放分析与设计的心理声学基础。

3.2 Ambisonics 重放分析

Ambisonics 是第1类基于物理声场精确重构原理空间声系统的典型代表。早年的研究中,一阶的Ambisonics 是作为第2类基于心理声学物理声场近似重放的多通路声的一种信号馈给法而发展的^[4,43-45],是普通前方立体声信号的指向性传声器检拾在水平面和三维空间推广。其后的研究表明,Ambisonics 的本质是一种声场的空间谐波分解(多极展开)、逐级逼近理想目标声场的一种方法^[4,13,46]。作为一类重要的重放方法,Ambisonics 及其应用是目前空间声研究的一个重点。

Ambisonics 可细分为水平面和空间 Ambisonics,前者采用水平面声场的贝塞尔-方位角傅里叶级数分解方法,后者采用三维空间声场的球贝塞尔-球谐函数分解方法,逐级逼近理想目标声场。以空间 Ambisonics 为例,假设目标声场由位于 (r_s, Ω_s) 的点声源产生。重放时 M 个扬声器布置在半径 r_0 的球面上,第 i 个扬声器的方向为 Ω_i ,信号振幅为 A_i 。如果将目标声源在 origin 附近产生的自由场声压和 M 个扬声器产生的叠加声压都按球贝塞尔-球谐函数分解,令其相匹配并截断到 $(L-1)$ 阶,则可得到第 i 个扬声器信号的振幅为前 $(L-1)$ 阶目标声源方向球谐函数的线性组合^[13]。

$$A_i(\Omega_S) = A_{\text{total}} \sum_{l=0}^{L-1} \sum_{m=0}^l \sum_{\sigma=1}^2 D_{lm}^{(\sigma)}(\Omega_i) \Xi_l(kr_s, kr_0) Y_{lm}^{(\sigma)}(\Omega_S), \quad (2)$$

其中, A_{total} 是与重构声压总振幅有关的常数, $Y_{lm}^{(\sigma)}(\Omega_S)$ 是一组关于目标声源方向的实数值球谐函数,如果采用附录A给出的球坐标系,可写为

$$Y_{lm}^{(\sigma)}(\Omega_S) = \begin{cases} N_{lm} P_l^m[\cos(90^\circ - \varphi_S)] \cos(m\theta_S), & \sigma = 1, \\ N_{lm} P_l^m[\cos(90^\circ - \varphi_S)] \sin(m\theta_S), & \sigma = 2, \end{cases}$$

$$N_{lm} = \sqrt{\frac{(l-m)!(2l+1)}{(l+m)!2\pi A_m}}, \quad A_m = \begin{cases} 2, & m \neq 0, \\ 1, & m = 0, \end{cases} \quad (3)$$

而

$$\Xi_l(kr_s, kr_0) = \frac{h_l(kr_s)}{h_l(kr_0)}, \quad (4)$$

式中, $P_l^m(\cdot)$ 是缔合勒让德多项式, $h_l(\cdot)$ 是 l 阶第二类球汉克尔函数, $\Xi_l(kr_s, kr_0)$ 是与频率或波数 k 有关的滤波函数, 用于补偿实际扬声器距离与目标声源距离不同造成的球面波弯曲波阵面的差异, $D_{lm}^{(\sigma)}(\Omega_i)$ 是一组 Ambisonics 的解码矩阵系数。对 $(L-1)$ 阶 Ambisonics 重放, 如果给定扬声器布置, 并且其方向满足稳定性的要求的条件, 数目满足

$$M \geq L^2, \quad (5)$$

则 $D_{lm}^{(\sigma)}(\Omega_i)$ 可由扬声器方向的球谐函数矩阵使用伪逆的方法求出^[13]。

式 (5) 给出了 $(L-1)$ 阶空间 Ambisonics 重放所需要的最少扬声器数目。另一方面, 虽然理论上 Ambisonics 可以逐级逼近目标声场, 但根据声场的空间采样定理, $(L-1)$ 阶空间 Ambisonics 重放能够在半径为 r 的球形区域内精确重构目标声场的上限频率为^[4]

$$f_{\max} = \frac{(L-1)c}{2\pi r}, \quad (6)$$

其中, $c = 343$ m/s 是声速。如果试图在 $r = 0.0875$ m 的平均头部半径范围, $f_{\max} = 20$ kHz 的可听声频率上限重构目标声场, 根据式 (5) 和式 (6) 可以估计出, 需要 $(L-1) = 32$ 阶空间 Ambisonics 重放, 并且至少需要 1000 个量级以上的扬声器。这样的系统非常复杂, 是难以实现的。实际中只能利用前面几阶的 Ambisonics 重放, 以换取对系统的简化。但根据式 (6), 第一、二和三阶空间 Ambisonics 在平均头部半径范围内重构目标声场的上限频率分别为 0.62 kHz, 1.25 kHz, 1.87 kHz, 远未达到高频谱因素起作用的频率范围。所以实际的 Ambisonics 重放不能产生正确的前后和垂直定位谱因素。

实际的 Ambisonics 重放应该产生正确的低频定位因素, 主要是 1.5 kHz 以下的 ITD 及其随头部转动的动态变化, 同时避免产生明显冲突或者误导的高频定位因素。利用不同定位因素所提供信息的冗余性, 可以产生侧向、前后和垂直虚拟源的定位感知。早期的理论分析表明^[40], 均匀扬声器布置的一阶空间 Ambisonics 系统可以产生期望的低频 ITD 及其随头部绕上下、前后轴转动的动态变化, 附录 A 虚拟源定位公式 (式 (A5) 一式 (A7)) 给出了一致的结果, 即无论是头部固定、绕上下轴转动、绕前后轴转动的情况, 感知虚拟源方向都与目标虚拟源方向一致。

采用速度矢量分析方法也可以得到类似的结论^[42]。这些传统的理论方法已用于一阶空间 Ambisonics 的分析和设计。

上述传统分析方法忽略了头部对声波的散射作用, 其结果只适用于 $f < 0.7$ kHz 的低频。更严格的分析应该采用 HRTF 计算 ITD 及其动态变化 (见附录 A 及文献 [19])。作为例子, 分析 28+1 个扬声器布置的空间 Ambisonics 重放的情况。按附录 A 的球坐标系统, 28 个扬声器分别布置在 $\varphi = -45^\circ, 0^\circ, 45^\circ$ 仰角面, 三个仰角面的扬声器数目分别为 8, 12, 8, 方位角间隔分别为 $45^\circ, 30^\circ, 45^\circ$ 。另有一个扬声器布置在 $\varphi = 90^\circ$ 的正上方。如果根据式 (5) 估计, 该扬声器布置最多可作四阶空间 Ambisonics 重放。但由于扬声器是非均匀布置的, 因而最多可作三阶 Ambisonics 重放。根据式 (2) 一式 (4) 得到各扬声器的信号振幅并作远场近似后, 用式 (A8) 给出的相关法和 KEMAR 人工头的 HRTF 可求出重放产生的 ITD 及其动态变化, 相关法计算的上限频率为 1.5 kHz。计算结果表明, 对目标方向在中垂面的情况, 第一至三阶重放产生的 ITD 都是零。图 8 为头部绕垂直轴向右转动 $\Delta\theta = 10^\circ$ 后, 第一至三阶重放的动态 ITD 变化, 并和目标单声源的情况比较。为了方便表示, 图中已采用 $\Phi - 90^\circ$ 作为坐标。随着阶数的增加, 动态 ITD 变化更接近真实声源的情况, 三阶重放的结果已和单声源的情况一致。采用全频段粉红噪声的虚拟源定位实验及相应的统计分析也表明, 三阶重放已经可以在中垂面产生不同仰角的虚拟源, 其前后混乱率为零, 上下混乱率为 4.8%。由其他一些实验结果可知, 虽然 Ambisonics 重放的定位性能随其阶数增加而改善, 但三或四阶重放已经可以产生较为理想的虚拟源定位效果^[47]。

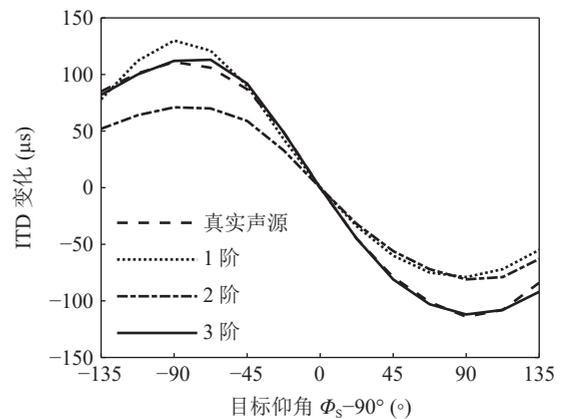


图 8 头部绕垂直轴向右转动 $\Delta\theta = 10^\circ$ 后中垂面目标声源的 ITD 变化 ($\Phi_s - 90^\circ = -180^\circ, -90^\circ, 0^\circ, 90^\circ$ 分别表示正下、正前、正上和正后方向)

事实上,根据式(6)也可直接得到上面的结论。式(6)表明,三阶空间 Ambisonics 可在头部尺度区域、 $f \leq 1.87 \text{ kHz}$ 的频率范围内准确重构目标声场,这已经覆盖了低频 ITD 及其动态变化作为定位因素的频率范围。当倾听者进入声场后,自然会得到正确的低频定位因素。这种方法可以推广至 Ambisonics 非中心倾听位置的定位分析,只要将式(6)的 r 替换为包含非中心位置头部的球形区域半径即可。当然,如果进一步考虑重放的音色等感知属性,三阶空间 Ambisonics 也不一定能得到完全理想的感知效果。这是因为在式(6)给出的上限频率之外不能准确重构目标声场,因而引起音色失真,需要增加音色均衡的方法^[38]。

上面的例子已足以说明低频动态定位因素在 Ambisonics 声重放中的作用,如果忽略这一因素,将会带来较大的误区。

3.3 多通路声重放的前后和垂直定位

多通路声是第 2 类心理声学物理声场近似重放技术的典型代表,它在传统的两通路立体声的基础上发展而成。多通路声在向三维空间重放发展,国际上已经发展了 9.1、11.1、22.2 等多种多通三维空间声技术^[16]。为了适应伴随视频重放的应用,并考虑到不同方向空间信息的听觉分辨率,这些多通声系统多采用前方密、后方和上方疏的非均匀扬声器布置。

两通路立体声采用合成定位的心理声学原理产生虚拟源^[1,4,48],即采用分立-对信号馈给方法,通过改变布置在前方一对左右扬声器信号的通路声级差,从而改变倾听者的低频双耳时间差,产生布置扬声器之间的不同方向的虚拟源。作为两通路立体声的推广,应用中也希望能在多通路空间声重放中利用一对布置在垂直平面的扬声器合成它们之间的虚拟源感知。但前后、垂直定位的心理声学机理和侧向定位不同,因而需要不同的分析。

在各种实际的多通路空间声扬声器布置中,垂直方向扬声器之间仰角间隔通常不少于 $20^\circ \sim 30^\circ$ 。受空间采样定理限制,这是不能准确合成 $5 \sim 6 \text{ kHz}$ 高频谱因素的^[4],采用 HRTF 的计算和双耳听觉模型分析可以证实这一点^[49-50]。与前面讨论的 Ambisonics 情况类似,只能依靠合成的低频 ITD 及其动态变化因素。

本课题组首先研究了中垂面一对前后对称布置扬声器的合成定位问题^[51],证明了通过调节扬声器信号的通路声级差,可以产生期望的(头部绕前后和

垂直轴转动)的低频动态 ITD 变化因素,从而产生两扬声器之间不同方向的虚拟源。采用低频信号的虚拟源定位实验也证实了理论分析,其结果与式(A5)一式(A7)的虚拟源定位公式预测的结果基本一致。因而从逻辑上看,该研究不但是听觉定位机理在多通路空间声重放的一种应用,同时也是 Wallach 有关动态定位因素假设的一个间接证明。

本课题组进一步研究了中垂面上其他一对相邻扬声器的合成定位问题^[19]。如图 9 所示,五个扬声器布置在中垂面上,其中扬声器 0 和 2 的布置是上下对称的,扬声器 2 和 4 的布置是前后对称的。根据式(A5)一式(A7)可以证明,采用一对相邻扬声器(0, 1), (1, 2), (2, 3) 或 (3, 4), 通过调节两扬声器信号的声级差,可以产生合适的低频动态 ITD 变化,因而可以产生扬声器之间的虚拟源听觉感知。采用前后对称布置的一对扬声器(2, 4)的情况和上面的例子相同。采用上下对称布置的一对扬声器(0, 2), 通过调节两扬声器信号的声级差,不能产生合适的动态 ITD 变化,因而不能产生扬声器之间的合成虚拟源。而采用 HRTF 的更精确的分析,得到类似结果。对其他垂直平面的扬声器布置的分析与此类似。虚拟源定位实验也得到与分析基本一致的结果,并且与过去的一些允许头部转动的实验结果基本吻合^[16]。

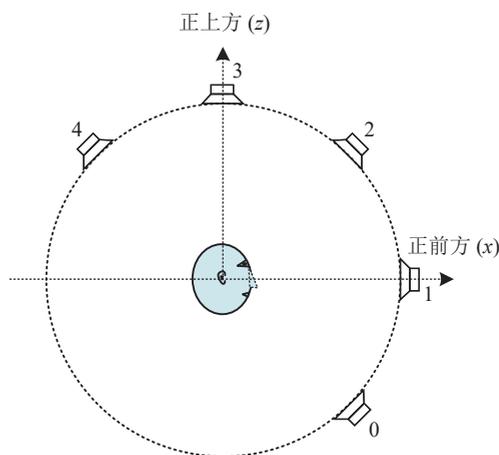


图 9 中垂面扬声器布置

采用两个相邻扬声器和分立-对信号馈给方法可以在中垂面产生垂直定位的动态 ITD 变化因素。但头部绕垂直轴和前后轴转动引起的动态 ITD 变化并不能给出完全一致的结果,表现在低频虚拟源定位式(A6)和式(A7)的结果并不完全自治,其感知虚拟源质量还可改进。参考前述(全局) Ambisonics 信号馈给法的主要思路,本课题组提出了一种中垂面上三扬声器的局域 Ambisonics 信号馈给法^[52]。采用图 9 中的扬声器 0, 1, 2, 并将各扬声器信号振幅取为

零阶和一对一阶仰角谐波的适当线性组合。图 10 是该信号馈给的曲线,其中三个扬声器分别布置在 $\varphi = 0^\circ$ 和 $\pm 45^\circ$ 。采用低频虚拟源定位式 (A5)~式 (A7) 的分析表明,当头部绕垂直和前后轴转动时,该扬声器布置和信号馈给可以产生一致的动态 ITD 变化信息,虚拟源定位实验也表明,重放可以产生三个扬声器布置之内以及略超出布置之外 ($\varphi = \pm 60^\circ$) 范围的垂直虚拟源感知。另外,利用类似方法也可以设计前方左前、右前、左前上、右前上四个扬声器布置条件下的信号馈给,产生扬声器布置之内的水平和垂直虚拟源^[53]。这种方法适合于常用的伴随视频多通路三维空间声扬声器布置,产生与视频配合的虚拟源感知。

上面的例子可以说明低频动态因素在多通路声重放中垂直合成定位的作用,并可用于设计扬声器布置与信号馈给。

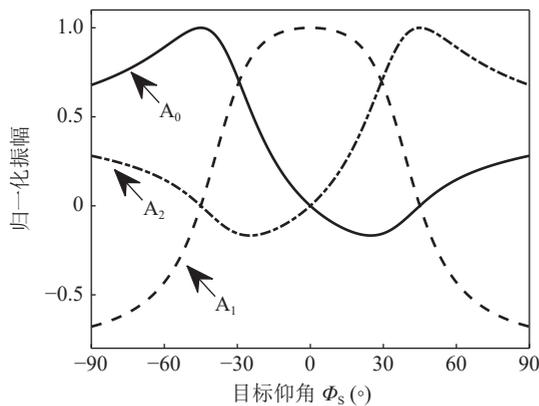


图 10 中垂面上三扬声器的局域 Ambisonics 信号馈给

4 虚拟听觉重放与动态定位因素

4.1 虚拟听觉重放的基本方法

虚拟听觉重放是第 3 类空间声重放系统的典型,其目标是尽可能精确模拟双耳信号,并(多数是)在耳机重放中产生期望的听觉事件。这种重放技术的硬件结构相对简单,功耗小,特别适用于虚拟现实、计算机、各种手持播放设备的声重放,是空间声重放技术的一个新热点。

虚拟听觉重放中最简单但具有普遍意义的情况是产生目标位置的自由场虚拟源。对稳态虚拟听觉重放,假设倾听者头部固定不动。根据式 (1) 给出的 HRTF 的定义,如果已知目标声源位置的一对 HRTF,将单路 E_0 输入信号用 HRTF 滤波,即模拟声源到双耳的声学传输过程,从而合成双耳信号^[5-6,54-55]。

$$E_\alpha = H_\alpha(r_s, \Omega_s, f)E_0, \quad \alpha = L, R. \quad (7)$$

图 11 是稳态虚拟听觉重放模拟自由场声源产生双耳信号的方块图,其中增加了一对滤波器 F_L 和 F_R ,用于对耳机到耳道非理想传输特性的均衡处理。

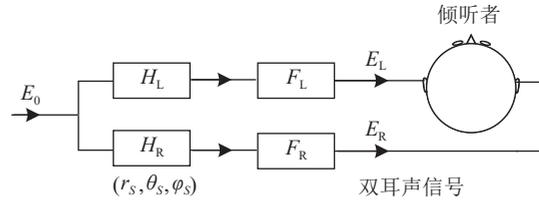


图 11 稳态虚拟听觉重放模拟自由场声源产生双耳信号的方块图

对于动态虚拟听觉重放,各种头部跟踪器实时检测到头部运动(例如,头部的三个自由度转动)的瞬时位置后,信号处理系统动态地刷新式 (7) 中合成双耳信号的 HRTF 数据,从而在双耳信号中模拟各种动态信息^[5]。

除了直接产生虚拟源与其他各种听觉感知外,虚拟听觉重放的一个特殊应用是 Ambisonics 和各种多通路声信号的耳机虚拟重放^[6]。这些信号原本是为扬声器重放而设计的,如果直接向下混合成两通路信号并用耳机重放,将得不到正确的声音信息。可以将各原始的扬声器信号用相应扬声器方向的 HRTF 滤波并叠加,模拟出多个扬声器到双耳的传输,从而将信号用虚拟扬声器重放出来。假设有 M 个(虚拟)扬声器,第 i 个虚拟扬声器信号为 A_i ,其位置为 (r_i, Ω_i) ,到双耳的 HRTF 为 $H_\alpha(r_i, \Omega_i, f)$,则双耳信号为

$$E_\alpha = \sum_{i=0}^{M-1} H_\alpha(r_i, \Omega_i, f)A_i, \quad \alpha = L, R. \quad (8)$$

上述稳态和动态虚拟听觉重放的方法都可以用于产生虚拟扬声器。

除了直接用耳机重放外,双耳信号经过串声消除处理后还可以用一对布置在前方的左右扬声器布置重放,即听觉传输重放^[5,56]。如果将上述产生多个虚拟扬声器的方法与串声消除结合,理论上可以用一对左右的真实扬声器产生多个不同方向的虚拟扬声器,从而实现多通路声的少量扬声器重放。

虽然采用耳机的虚拟听觉重放和采用扬声器的听觉传输重放理论上可以精确重构双耳声压,但实际中都存在着一定缺陷,且这些缺陷都和动态听觉定位因素密切相关。为克服这些缺陷,近二十年国际上在学术研究和产业应用方面投入了大量的工作^[5],但一直没有很好地解决,甚至有可能给研究带来了误区,这本质上是低估了动态定位因素的作用导致

的。以下将综合讨论该问题。

4.2 耳机虚拟听觉重放的动态因素与谱因素问题

大量的实验与应用表明,采用耳机稳态虚拟听觉重放普遍存在的主要问题是^[5]:(1)头中定位(重放中感知自由场虚拟源在头内或头表面附近);(2)感知虚拟源前后、上下混乱,垂直方向定位错误(感知仰角上升)。

以上问题对中垂面上目标虚拟源的情况特别明显,这些问题已经影响到系统的实际应用。另一方面,已有心理声学实验结果已表明^[57-59],动态虚拟听觉重放可以很好地克服上述问题,几乎完全消除了头中定位和前后、上下混乱^[60],也可以明显减少垂直方向的定位错误。因此可以推断,耳机稳态虚拟听觉重放的问题主要源于重放中动态因素的缺失,而高频谱因素的稳定性较差。

事实上,在自然听觉环境中,头部自觉和不自觉的转动会带来动态听觉信息。高层神经系统将综合稳态和动态信息形成自然的听觉定位感知。当部分信息缺失或严重错误时,听觉系统对信息综合处理的过程就有可能“出错”,从而产生头中定位的不自然的听觉感知,同时也会产生前后、上下混乱,以及垂直方向的定位错误。按照第2节的讨论,动态和稳态谱因素所提供的信息有一定的冗余。当信号包含1.5 kHz低频成分时(实际的信号大部分如此),在动态因素的基础上增加非个性化谱因素即可得到较好的定位感知。

但另一方面,动态虚拟听觉重放系统的硬件结构比较复杂,应用上更希望在稳态虚拟听觉重放框架内解决上述问题。但从近二十年的大量研究来看,实际效果却是有限的,较好情况下也只是部分地改进。这是因为动态因素缺失时,需要准确的个性化高频谱因素才能一定程度上实现定位,而个性化高频(波长短)谱因素的稳定性较差。在信号处理与重放过程中任何一个环节的误差都有可能破坏个性化谱因素的信息,从而影响最后的定位感知效果。因而需要准确的个性化HRTF处理和个性化耳机到耳道传输均衡,使耳机重放在鼓膜处产生的声压与目标声源的情况相同,才能取得一定的效果。在Wightman和Kistler有关稳态重放的经典实验中^[55],采用探针传声器测量得到的鼓膜附近的个性化HRTF数据,并采用个性化的耳机到耳道(鼓膜附近位置)均衡处理,产生三维空间虚拟源的听觉感知。但即使在如此精确的信号处理条件下,仍存在一定的定位错误。例如,在前方范围,低、中、高仰角的

平均定位混乱率分别为10%,7%,38%(自由场真实声源的情况分别为4%,1%,16%)。Wenzel等进一步的实验结果表明^[61],采用非个性化HRTF会增加混乱率,全空间的平均前后和上下混乱率分别为31%和18%(自由场真实声源的情况分别为19%和6%)。另外,在鼓膜附近测量HRTF存在一定的危险,实际应用有一定的困难。

改善稳态虚拟听觉重放的主要困难在于个性化HRTF的获取。相较其他方法,实验测量是获取个性化HRTF的最准确的方法。近二十多年,已有多个课题组建立了真人受试者的测量远场HRTF数据库,部分数据库在互联网上公开^[62]。在2006年,本课题组已建立了52名中国人受试者的HRTF数据^[63]。不同距离的近场HRTF测量技术较远场HRTF复杂,且工作量大,目前国际上只有少量的测量人工头近场HRTF数据库^[64-66],以及本课题组的56名真人受试者的近场HRTF数据库^[67]。

数值计算也是一种准确获取HRTF的方法。它通过光学等方法扫描得到头部、耳廓等几何外形边界,然后通过边界元等方法求解散射问题以及求解波动方程的数值解,得到相应的HRTF数据^[68-69]。

目前HRTF的测量与数值计算技术已相对成熟,但其结果仍然存在误差。以远场HRTF测量为例,仔细设计测量系统和实验,可减少测量误差。本实验室的第二代测量系统,在8 kHz和12 kHz以下的频率范围,不同方向的平均谱失真分别小于0.5 dB和3 dB^[70]。而对国际上多个不同课题组的测量结果^[71-72],较好的情况下,在4.5 kHz,12 kHz以下的频率范围,不同方向平均谱失真分别是2~3 dB和3~6 dB的量级;而在12 kHz以上的频率范围,不同方向的平均谱失真可大于10 dB。一些课题组的测量系统设计粗糙,误差更大。HRTF数值计算的误差也和测量类似。

HRTF的测量或计算误差主要出现在高频,而个性化HRTF的幅度谱差异也主要在5~6 kHz到12~14 kHz的高频段(14 kHz以上的HRTF幅度谱对听觉定位的贡献较少)。Engel等最近的实验研究指出,在3 kHz以上的频段,不同受试者的个性化HRTF平均幅度差异可达6 dB的量级^[73]。但如果测量或计算HRTF的幅度谱误差(或者加上虚拟听觉重放信号处理的误差)接近个性化HRTF幅度谱的差异,个性化HRTF的处理就失去了意义,导致稳态虚拟听觉重放出现上述问题。

另一方面,个性化HRTF的测量需要特殊的设备,并且测量或数值计算也非常耗时,在大规模的应

用中并不一定可行。因此近二十多年,有大量关于个性化 HRTF 定制工作^[74-76],即通过(适当数量)生理参数测量(早期是用传统的尺等工具,近年是用计算机视觉等方法)预测或者匹配 HRTF,或主观选择估算个性化 HRTF。这类方法已有商业应用,但对于解决稳态虚拟听觉重放问题的实际效果却相当有限。为此,许多研究试图改进从生理参数预测个性化 HRTF 的数学方法,从早期的简单线性预测到近几年的机器学习、人工智能等数学方法^[77],但效果依然有限。

事实上,各种定制方法的本质是基于现有(足够数量个体)的测量(或数值计算)HRTF 和生理参数数据库进行统计分析,从而得到 HRTF 和生理参数之间的映射关系^[5]。测量 HRTF 本来就存在一定的高频误差。而头部、耳廓等生理结构及其对高频声波的散射作用非常复杂,并非少量生理参数就能够准确预测,即使最好的统计预测方法也不能完全解决这一问题,各种预测方法都不可避免带来一定的误差。例如,有些研究中各种定制 HRTF 全频带(甚至 8 kHz 以下)平均谱失真在误差 5~6 dB 量级^[78]。较好的情况下,全频带范围的平均谱失真也在 3 dB 的量级。当定制误差与测量误差的叠加达到个性化 HRTF 差异的量级时,个性化 HRTF 定制就失去了意义。有许多认为“明显改善”的定制方法也只是将谱失真“减少”了 1 dB 的量级甚至更小,这对听觉感知的意义不大。有些研究心理声学实验设计和数据分析并不合理(甚至连最基本的统计检验和方差分析也没有),最后得出的结论当然存疑。

总体上,稳态虚拟听觉重放的主要问题在于失去了动态因素,这时就更加依赖仅存的个性化谱因素。而个性化谱因素的贡献本来就不及动态因素,且稳定性较差。HRTF 测量、计算、特别是定制误差,以及信号处理环节的误差,很容易使个性化 HRTF 失去意义。过去的研究在这方面存在较大的误区,动态虚拟听觉重放应该是最终解决上述问题的方法。只能采用稳态重放时,尽量采用准确的 HRTF,保留剩存的信息。

发展更高效、准确的 HRTF 测量和计算方法有助于改善上述问题。例如,可以利用不同方向 HRTF 之间的相关性,通过少量方向的 HRTF 测量而准确恢复全空间位置的 HRTF 数据。本课题组利用空间主成分分析的方法,通过对足够(30 名以上)受试者的基线 HRTF 数据进行分析,然后通过 73 个方向的远场 HRTF 测量数据即可恢复 493 个方向的 HRTF 幅度^[79]。考虑到 HRTF 取决于多个变量的特

性,采用 HRTF 张量分解方法可以进一步提高测量效率。已经证明^[80],从大约 30 个方向的远场 HRTF 测量数据即可在 $f < 14$ kHz 的频率范围恢复全空间方向的远场 HRTF 幅度,甚至可以恢复不同距离的近场 HRTF 幅度,避免了近场 HRTF 测量的困难^[81]。上述几种从少量方向 HRTF 测量恢复全空间 HRTF 方法的平均相对误差大约是在 -15~20 dB 的量级,已经满足听觉要求的精度。也可以采用 HRTF 聚类的方法^[82],因为大部分人的 HRTF 可分为 7 类,用类中心 HRTF 处理对动态重放已足够。当然,还可以发展其他 HRTF 获取方法,这方面有待研究。

4.3 耳机虚拟听觉重放中距离感知问题

除了产生不同方向的远场虚拟源外,也期望能在耳机虚拟听觉重放中产生不同距离的近场虚拟源感知。这可通过在合成双耳信号中控制各种距离感知因素而实现。这种方法不但在虚拟现实等方面有重要的实际应用,也可作为听觉距离感知研究的重要实验工具。

对于模拟自由场虚拟源距离感知的情况,可在式(7)和图 11 的双耳信号处理中采用不同距离的近场 HRTF 处理,再乘以模拟声压随声源距离衰减的 $1/r_s$ 因子。但稳态虚拟听觉重放经常会产生头中定位而影响距离感知的判断,如果用作听觉距离感知的实验工具,经常会出现较大的争议性的结论。本课题组最近对比研究了稳态和动态的虚拟听觉重放对绝对距离感知效果的影响^[83]。心理声学实验结果表明,在稳态虚拟听觉重放条件下,受试者不同程度地报告了存在头中定位。而动态虚拟听觉重放系统能够产生头外化的听觉感知。统计分析表明,动态虚拟听觉重放下的距离感知效果与稳态重放存在明显的差异。虽然动态因素并非距离感知因素,但动态因素的缺失会引起头中定位,妨碍了进一步对感知距离的控制。因而,应选用动态虚拟听觉重放作为实验工具。同时,在实际的应用中也应采用动态虚拟听觉重放产生不同的距离感知。

本课题组研究了动态双耳 Ambisonics 重放自由场虚拟源方向和距离信息的简化信号处理方法^[84]。将式(2)的 Ambisonics 信号馈给与式(8)的虚拟扬声器方法相结合,用动态 HRTF 滤波处理将 Ambisonics 的扬声器重放信号转换为双耳重放信号并用耳机重放。心理声学实验结果表明,五阶动态双耳 Ambisonics 重放可产生不同方向和 1.0 m 以下不同近场距离的自由场虚拟源的听觉感知。该方法仅需固定距离的 48 个均匀空间方向的远场非个性化

HRTF 处理, 实现了信号处理的简化。值得注意的是, 在式 (7) 中取 $r = 0.0875 \text{ m}$ (平均头部半径), 可以估算出, 五阶双耳 Ambisonics 只能在 3.1 kHz 以下的频率范围内准确重构双耳声压, 虽然未达到高频谱因素明显起作用的频率范围, 但已经包括了低频 ITD 及其动态变化以及随距离变化的低频 ILD 因素的主要作用频率范围。这进一步说明了动态因素在虚拟自由场声学方向定位和距离感知中的作用。

4.4 两扬声器听觉传输重放的虚拟源分布

前方两扬声器的听觉传输重放中, 理想的串声消除理论上能产生空间任意方向目标声源相同的双耳声压, 因而也期望其能产生三维空间虚拟源^[56]。但只有在很苛刻的实验条件下 (个性化 HRTF 处理、消声室重放和限制头部移动和转动), 部分倾听者会感知到后方甚至垂直虚拟源的效果^[85]。更多情况下^[86-87], 感知虚拟源被限制在前半水平面的范围内, 后半水平面和水平面外的目标虚拟源会出现在前半水平面的同一混乱锥方向附近。近二十年在改进听觉传输重放信号处理方面有不少工作, 但实际的感知效果并不明显。

上述问题也是由动态定位因素的错误所引起。在两扬声器听觉传输重放中, 虽然理论上可以精确重构目标双耳声压, 但这只对特定的倾听位置、头部固定的情况下有效。由于 5~6 kHz 的高频谱因素具有个性化特性且稳定性较差, 即使听觉传输重放信号处理中采用了个性化的 HRTF, 只要倾听者偏离 $1/4 \sim 1/2$ 波长的位置, 双耳声压所包含的高频因素就完全失效。另一方面, 稳态的两扬声器听觉传输合成中, 不但没有考虑倾听者头部转动带来的动态定位因素, 前方扬声器布置带来了错误的动态定位信息 (提示虚拟源在前半水平面)。因此上述前方两扬声器稳态听觉传输重放的问题是低频动态信息错误、高频谱信息不稳定引起的。本课题组最近在动态虚拟听觉重放平台上模拟前方两扬声器的听觉传输重放, 通过人工控制和补偿动态因素, 并用心理声学实验证实了上述分析^[88]。

因而前方两扬声器稳态听觉传输重放的虚拟源方向畸变是其固有的局限, 是物理与听觉原理限制所造成的, 不可能通过信号处理的方法完全解决, 除非采用动态的听觉传输重放。

4.5 动态因素与人工听觉

听觉障碍除体现为单侧或双侧可听阈的提高 (甚至听力完全缺失) 外, 还可能包括空间听觉能力

的下降甚至缺失。这影响了患者寻找目标、发现和避免周围潜在危险的能力, 且空间听觉能力下降带来的一大问题是干扰环境下语言获取能力的降低。人工听觉技术 (包括助听器与人工耳蜗等) 是对听觉障碍的治疗手段。双耳助听器可看成是一种特殊的双耳检拾与重放技术。听觉障碍患者高频听力损失通常更为严重, 而助听器常用的耳背式传声器检拾得到的谱因素和耳道口检拾不同, 因而助听器不易给患者提供可用的高频谱信息。但另一方面, 由于双耳助听器的传声器通常固定在头部表面且随头部一起运动, 因而检拾到的是动态双耳信号, 有可能保留动态定位信息。当然, 目前双耳助听器的一些信号处理可能会破坏这些信息。目前有关动态因素在听觉障碍和人工听觉技术方面应用的研究刚开始发展^[89-91]。

5 总结与展望

听觉定位机理是心理和生理声学的传统且前沿的研究领域。100 多年前已有研究开始认识到动态因素在听觉定位中的作用, Wallach 提出动态因素的经典假设也有 90 多年的历史, 但在很长时间内, 动态因素对听觉定位特别是垂直定位的作用得不到足够的重视。近二十年, 一方面由于实验技术的发展, 另一方面也由于三维空间声重放技术的发展要求, 动态因素及其与其他因素的相互作用成为听觉定位机理研究的重要前沿领域。

大量实验研究已经表明, 动态因素和个性化谱因素都对前后和垂直定位有贡献, 但其作用的频率范围不同, 动态因素 (主要是动态 ITD 变化) 主要作用在低频, 而谱因素主要作用在高频, 且稳定性较差。一方面两种因素的协同作用增强定位, 另一方面两种因素提供的信息有一定的冗余。但如果不同频段的两种因素发生冲突, 有可能会产生空间位置分裂的两个听觉事件。

近年有关动态因素对听觉定位作用的研究已取得相当大的进展, 但有关动态因素及其与其他定位因素 (特别是个性化谱因素) 的相互作用的许多细节问题还有待深入研究。而动态听觉因素与其他多模态感知信息的相互作用问题 (特别是与受试者自运动带来的非听觉空间信息的相互作用) 更值得深入研究。这方面的研究对揭示人类对多模态信息感知的规律非常重要。

近年空间声技术快速发展, 动态听觉定位机理在空间声技术中有重要的作用。受系统复杂性的限

制, 实际的多扬声器空间声重放系统(包括第 1 类基于物理声场重构系统和第 2 类基于心理声学原理的系统) 很难准确产生正确的高频谱因素。因而需要产生正确的低频动态定位因素, 利用不同因素提供信息的冗余性而产生前后和垂直方向的空间听觉感知。对动态因素的分析应成为设计和评价多扬声器三维空间声的一个重要部分。而对于虚拟听觉重放, 包括稳态耳机重放和经过串声消除后的扬声器听觉传输重放, 其许多重要的问题与缺陷都与重放中不能产生正确的动态因素有关。近年来这一问题在研究和技术开发中带来了较大的误区, 需引起注意。解决这些问题的最终办法是采用动态重放, 当然采用个性化处理可进一步改善重放的空间感知效果。

对空间声系统的客观评价是一个重要的研究方向。由于各种实际的空间声技术都利用了心理声学的原理, 纯物理的评价方法是不够的。目前主要采用一些基于心理声学实验得到的数据或者经验模型进行评价, 并且有些模型已考虑了动态因素分析^[92]。但由于定位因素的冗余性, 不同因素及其相互作用对最后感知和评价的贡献也是值得深入研究的。

最后, 在空间声重放的应用中, 涉及动态因素的研究目前主要在心理声学的层面上进行。这对于实际应用可能暂时是足够的, 但从听觉基础研究的角度, 研究应进一步深入到听觉生理声学及神经生理机制的层面。目前国际上已开始这方面的工作。这方面的研究不但对探索人类听觉机理有重要的意义, 其结果可能对进一步了解空间声感知的本质及设计有重要的意义。

致谢: 作者感谢北京邮电大学管善群教授, 华南理工大学余光正教授、饶丹副教授、孟庆林副教授, 广州大学张承云教授, 以及历届研究生的合作研究。感谢余光正教授和杜善文博士生帮助绘制了本文的插图。

参 考 文 献

- Blauert J. Spatial hearing: The psychophysics of human sound localization. Revised edition. Cambridge, MA, USA: MIT Press, 1997
- Wallach H. The role of head movement and vestibular and visual cue in sound localization. *J. Exp. Psychol.*, 1940; **27**(4): 339–368
- Gelfand S A. Hearing: An Introduction to psychological and physiological acoustics. 5th edition. London, UK: Informa Healthcare, 2010
- Xie B S. Spatial sound principle and applications. USA: CRC Press, 2022
- Xie B S. Head-related transfer function and virtual auditory display. 2nd edition. USA: J Ross Publishing, 2013
- Møller H. Fundamentals of binaural technology. *Appl. Acoust.*, 1992; **36**(3-4): 171–218
- Litovsky R Y, Colburn H S, Yost W A, *et al.* The precedence effect. *J. Acoust. Soc. Am.*, 1999; **106**(4): 1633–1654
- Sugiyama K, Sakaguchi T, Aoki S, *et al.* Calculation of acoustic coefficients between two ears using spheroids. *J. Acoust. Soc. Jpn.*, 1995; **51**(2): 117–122
- Wightman F L, Kistler D J. Monaural sound localization revisited. *J. Acoust. Soc. Am.*, 1997; **101**(2): 1050–1063
- Algazi V R, Avendano C, Duda R O. Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.*, 2001; **109**(3): 1110–1122
- Wightman F L, Kistler D J. The dominant role of low-frequency interaural time difference in sound localization. *J. Acoust. Soc. Am.*, 1992; **91**(3): 1648–1661
- Kolarik A J, Moore B C J, Zahorik P, *et al.* Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. *Attent. Percept. Psychophys.*, 2016; **78**: 373–395
- Daniel J, Moreau S. Further study of sound field coding with higher order Ambisonics. The AES 116th Convention, Berlin, Germany, 2004: 6017
- De Vries D. Wave field synthesis. New York, USA: Audio Engineering Society, 2009
- Ahrens J. Analytic methods of sound field synthesis. Berlin, Germany: Springer, 2012
- International Telecommunication Union. Multichannel sound technology in home and broadcasting applications: ITU-R Report BS. 2159-7. Geneva, Switzerland: ITU, 2015
- Herre J, Hilpert J, Kuntz A, *et al.* MPEG-H audio—The new standard for universal spatial/3D audio coding. *J. Audio Eng. Soc.*, 2014; **62**(12): 821–830
- Young P T. The role of head movement in auditory localization. *J. Exp. Psychol.*, 1931; **14**(2): 95–124
- Xie B S, Mai H M, Rao D, *et al.* Analysis of and experiments on vertical summing localization of multichannel sound reproduction with amplitude panning. *J. Audio. Eng. Soc.*, 2019; **67**(6): 382–399
- Xie B S, Liu L L, Jiang J L, *et al.* Auditory vertical localization in the median plane with conflicting dynamic interaural time difference and other elevation cues. *J. Acoust. Soc. Am.*, 2023; **154**(3): 1770–1786
- Macpherson E A. Cue weighting and vestibular mediation of temporal dynamics in sound localization. The 21st International Congress on Acoustics, Montreal, Canada, 2013
- Thurlow W R, Runge P S. Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.*, 1967; **42**(2): 480–488
- Bronkhorst A W. Localization of real and virtual sound source. *J. Acoust. Soc. Am.*, 1995; **98**(5): 2542–2553
- Wightman F L, Kistler D J. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 1999; **105**(5): 2841–2853
- Brimijoin W O, Akeroyd M A. The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, 2012; **3**(3): 179–182
- Pöntynen H, Salminen N H. Resolving front-back with head rotation: The role of level dynamics. *Hear. Res.*, 2019; **377**: 196–207
- Pöntynen H, Santala O, Pulkki V. Conflicting dynamic and spec-

- tral directional cues form separate auditory images. The Audio Engineering Society 140th Convention, Paris, France, 2016: 9582
- 28 Yost W A. Randomizing spectral cues used to resolve for front-back reversals in sound-source localization. *J. Acoust. Soc. Am.*, 2023; **154**(2): 661–670
- 29 Pollack I, Rose M. Effect of head movement on the localization of sounds in the equatorial plane. *Percept. Psychophys.*, 1967; **2**: 591–596
- 30 Thurlow W R, Mergener J R. Effect of stimulus duration on localization of direction of noise stimuli. *J. Speech Lang. Hear. Res.*, 1970; **13**(4): 826–838
- 31 Perrett S, Noble W. The effect of head rotations on vertical plane sound localization. *J. Acoust. Soc. Am.*, 1997; **102**(4): 2325–2332
- 32 Toshima I, Aoki S. Sound localization during head movement using an acoustical telepresence robot: TeleHead. *Adv. Rob.*, 2009; **23**(3): 289–304
- 33 Martens W L, Cabrera D, Kim S. The 'phantom walker' illusion: Evidence of the dominance of dynamic interaural over spectral directional cues during walking. In: Principles and applications of spatial hearing. Singapore: World Scientific Publishing, 2011: 81–102
- 34 McLachlan G, Majdak P, Reijnen J, et al. Dynamic spectral cues do not affect human sound localization during small head movements. *Front. Neurosci.*, 2023; **17**: 1027827
- 35 Jiang J L, Xie B S, Mai H M, et al. The role of dynamic cue in auditory vertical localization. *Appl. Acoust.*, 2019; **146**: 398–408
- 36 McAnally K I, Martin R L. Sound localization with head movement: Implication for 3-d audio displays. *Front. Neurosci.*, 2014; **8**: 210
- 37 Yost W A, Pastore M T, Pulling K R. Sound-source localization as a multisystem process: The Wallach azimuth illusion. *J. Acoust. Soc. Am.*, 2019; **146**(1): 382–398
- 38 刘阳, 谢菠荪. Ambisonics 声检拾与重放音色的双耳听觉模型分析与实验. *声学学报*, 2015; **40**(5): 717–729
- 39 Bernfeld B. Simple equations for multichannel stereophonic sound localization. *J. Audio Eng. Soc.*, 1975; **23**(7): 553–557
- 40 谢兴甫. 三维立体声场的 (数学) 分析. *声学学报*, 1988; **13**(5): 321–328
- 41 饶丹, 谢菠荪. 多通路三维空间环绕声系统. *声学学报*, 2005; **30**(2): 163–170
- 42 Gerzon M A. General metatheory of auditory localization. The AES 92nd Convention, Vienna, Austria, 1992: 3306
- 43 Gerzon M A. Ambisonics in multichannel broadcasting and video. *J. Audio Eng. Soc.*, 1985; **33**(11): 859–871
- 44 Xie X F. The 4-3-N matrix multi-channel sound system. *Chinese Journal of Acoustics*, 1982; **21**(2): 201–218
- 45 谢兴甫. 立体声原理. 北京: 科学出版社, 1981
- 46 Xie B S, Xie X F. Analyse and sound image localization experiment study on multi-channel planar surround sound system. *Chinese Journal of Acoustics*, 1996; **15**(1): 52–64
- 47 Bertet S, Daniel J, Parizet E, et al. Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources. *Acta Acust. united Acust.*, 2013; **99**: 642–657
- 48 Leakey D M. Some measurements on the effects of interchannel intensity and time differences in two channel sound systems. *J. Acoust. Soc. Am.*, 1959; **31**(7): 977–986
- 49 Pulkki V. Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning. *J. Audio Eng. Soc.*, 2001; **49**(9): 753–767
- 50 Baumgartner R, Majdak P. Modeling localization of amplitude-panned virtual sources in sagittal planes. *J. Audio Eng. Soc.*, 2015; **63**(7-8): 562–569
- 51 Rao D, Xie B S. Head rotation and sound image localization in the median plane. *Chin. Sci. Bull.*, 2005; **50**(5): 412–416
- 52 Yi K L, Xie B S. Local Ambisonics panning method for creating virtual source in the vertical plane of the frontal hemisphere. *Appl. Acoust.*, 2020; **165**: 107319
- 53 张驰, 谢菠荪. 多通路空间声的前方四扬声器局域 Ambisonics 信号馈给法. *应用声学*, 2024; **43**(1): 64–75
- 54 Wightman F L, Kistler D J. Headphone simulation of free-field listening, I: stimulus synthesis. *J. Acoust. Soc. Am.*, 1989; **85**(2): 858–867
- 55 Wightman F L, Kistler D J. Headphone simulation of free-field listening, II: psycho-physical validation. *J. Acoust. Soc. Am.*, 1989; **85**(2): 868–878
- 56 Bauck J, Cooper D H. Generalized transaural stereo and applications. *J. Audio. Eng. Soc.*, 1996; **44**(9): 683–705
- 57 Wenzel E M. What perception implies about implementation of interactive virtual acoustic environments. The AES 101st Convention, Los Angeles, CA, USA, 1996: 4353
- 58 Zhang C Y, Xie B S. Platform for dynamic virtual auditory environment real-time rendering system. *Chin. Sci. Bull.*, 2013; **58**(3): 316–327
- 59 Brimijoin W O, Boyd A W, Akeroyd M A. The contribution of head movement to the externalization and internalization of sounds. *PLoS One*, 2013; **8**(12): e83068
- 60 Begault D R, Wenzel E M, Anderson M R. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio. Eng. Soc.*, 2001; **49**(10): 904–916
- 61 Wenzel E M, Arruda M, Kistler D J, et al. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 1993; **94**(1): 111–123
- 62 Majdak P, Zotter F, Brinkmann F, et al. Spatially oriented format for acoustics 2.1: Introduction and recent advances. *J. Audio Eng. Soc.*, 2022; **70**(7-8): 565–584
- 63 Xie B S, Zhong X L, Rao D, et al. Head-related transfer function database and its analyses. *Sci. China Ser. G.*, 2007; **50**(3): 267–280
- 64 龚玫, 肖峥, 曲天书, 等. 近场头相关传输函数的测量与分析. *应用声学*, 2007; **26**(6): 326–334
- 65 余光正, 谢菠荪, 饶丹. 人工头近场头相关传输函数及其特性. *声学学报*, 2012; **37**(4): 378–385
- 66 Marschall M, Bolanos J G, Prepelita S T, et al. A database of near-field head-related transfer functions based on measurements with a laser spark source. *Appl. Acoust.*, 2023; **203**: 109173
- 67 Yu G Z, Wu R X, Liu Y, et al. Near-field head-related transfer-function measurement and database of human subjects. *J. Acoust. Soc. Am.*, 2018; **143**(3): EL194–EL198
- 68 Katz B F G. Boundary element method calculation of individual head-related transfer function. I. rigid model calculation. *J. Acoust. Soc. Am.*, 2001; **110**(5): 2440–2448
- 69 Gumerov N A, O'Donovan A E, Duraiswami R, et al. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.*, 2010; **127**(1): 370–386

- 70 余光正, 刘昱, 谢波菘. 近场头相关传输函数的多声源快速测量系统设计与验证. *声学学报*, 2017; **42**(3): 348–360
- 71 Katz B F G, Begault D R. Round robin comparison of HRTF measurement systems: Preliminary results. The 19th International Congress on Acoustics, Madrid, Spain, 2007
- 72 Andreopoulou A, Begault D R, Katz F G. Inter-laboratory round robin HRTF measurement comparison. *IEEE J. Sel. Top. Signal Process.*, 2015; **9**(5): 895–906
- 73 Engel I, Daugintis R, Vicente T, *et al.* The SONICOM HRTF database. *J. Audio Eng. Soc.*, 2023; **71**(5): 241–253
- 74 Iida K, Ishii Y, Nishioka S. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. *J. Acoust. Soc. Am.*, 2014; **136**(1): 317–333
- 75 Iida K, Shimazaki H, Oota M. Generation of the amplitude spectra of the individual head-related transfer functions in the upper median plane based on the anthropometry of the listener's pinnae. *Appl. Acoust.*, 2019; **155**: 280–285
- 76 Pelzer R, Dinakaran M, Brinkmann F, *et al.* Head-related transfer function recommendation based on perceptual similarities and anthropometric features. *J. Acoust. Soc. Am.*, 2020; **148**(6): 3809–3817
- 77 McMullen K, Wan Y. A machine learning tutorial for spatial auditory display using head-related transfer functions. *J. Acoust. Soc. Am.*, 2022; **151**(2): 1277–1293
- 78 Torres-Gallegos E A, Bustamante F O, Cosio F A. Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. *Appl. Acoust.*, 2015; **97**: 84–95
- 79 Xie B S. Recovery of individual head-related transfer functions from a small set of measurements. *J. Acoust. Soc. Am.*, 2012; **132**(1): 282–294
- 80 赵童, 谢波菘, 朱俊, 等. 基于个体张量分解与模态的头相关传输函数重构. *声学学报*, 2023; **48**(1): 215–224
- 81 Zhao T, Xie B S, Zhu J. Reconstruction of individualized near-field head-related transfer functions from a small set of far-field data based on tensor decomposition. *J. Acoust. Soc. Am.*, 2023; **154**(4S): A256
- 82 Xie B S, Zhong X L, He N N. Typical data and cluster analysis on head-related transfer functions from Chinese subjects. *Appl. Acoust.*, 2015; **94**(1): 1–13
- 83 燕子翔, 谢波菘, 朱俊. 稳态和动态虚拟听觉重放对近场距离感知的影响. *应用声学*, 2024; **43**(3): 469–477
- 84 谢波菘, 刘路路, 江建亮. 重放自由场虚拟源距离信息的动态双耳 Ambisonics 方法. *声学学报*, 2021; **46**(6): 1223–1233
- 85 Takeuchi T, Nelson P A, Kirkeby O, *et al.* Influence of individual head related transfer function on the performance of virtual acoustic imaging systems. The AES 104th Convention, Amsterdam, The Netherlands, 1998: 4700
- 86 Nelson P A, Orduña-Bustamante F, Engler E, *et al.* Experiments on a system for synthesis of virtual acoustic sources. *J. Audio Eng. Soc.*, 1996; **44**(11): 990–1007
- 87 Gardner W G. 3-D Audio using loudspeakers. Doctoral dissertation, Massachusetts, USA: Massachusetts Institute of Technology, 1997
- 88 Liu L L, Xie B S. Analysis and experiment on the limitations of static and dynamic transaural reproduction with two frontal loudspeakers. *Arch. Acoust.*, 2021; **46**(2): 213–228
- 89 Akeroyd M A, Whitmer W M. Spatial hearing and hearing aids.

ENT Audiol. News, 2011; **20**(5): 76–79

- 90 谢波菘, 孟庆林. 听觉障碍患者的空间听觉. *应用声学*, 2018; **37**(5): 607–612
- 91 Gessa E, Giovanelli E, Spinella D, *et al.* Spontaneous head-movements improve sound localization in aging adults with hearing loss. *Front. Hum. Neurosci.*, 2022; **16**: 1026056
- 92 Braasch J, Clapp S, Parks A, *et al.* A binaural model that analysis acoustic space and stereophonic reproduction system by utilizing head rotations. In: Blauert J. (eds) The technology of binaural listening. Berlin, Heidelberg: Springer, 2013

附录A 方向定位因素与 Wallach 假设的分析

在空间听觉和空间声的研究中, 不同文献采用的坐标系有所不同。为统一和讨论方便, 本文用到两种坐标系。第一种是逆时针方向的球坐标系。如图 A1(a) 所示, 三维空间位置由原点在头中心的球坐标 (r, θ, φ) 决定, 其中 $0 \leq r < +\infty$ 是声源的距离; $-90^\circ \leq \varphi \leq 90^\circ$ 是仰角; $-180^\circ < \theta \leq 180^\circ$ 是方位角。 $\varphi = -90^\circ, 0^\circ, 90^\circ$ 分别表示正下方、水平面和正上方; 在水平面上, $\theta = 0^\circ, 90^\circ, 180^\circ$ 分别表示正前、正左和正后方向。

第二种是双耳极坐标系。如图 A1(b) 所示, 三维空间位置由双耳极坐标 (r, Θ, Φ) 表示, 其中 r 的意义同前。 $-90^\circ \leq \Theta \leq +90^\circ$ 是双耳极方位角, 即声源与中垂面的夹角, $\Theta = -90^\circ$ 和 $+90^\circ$ 分别表示正右和正左方向, 图 3 给出的混乱锥上 Θ 是常数。 $-90^\circ < \Phi \leq 270^\circ$ 是极仰角, 即声源方向矢量在中垂面投影与正前方的夹角, 在中垂面上, $\Phi = -90^\circ, 0^\circ, 90^\circ, 180^\circ$ 分别表示正下、正前、正上和正后方向。两种坐标之间的变换关系为

$$\begin{aligned} \sin \Theta &= \sin \theta \cos \varphi, \quad \cos \Theta \sin \Phi = \sin \varphi, \\ \cos \Theta \cos \Phi &= \cos \theta \cos \varphi. \end{aligned} \quad (\text{A1})$$

如果只考虑低频 ITD 及其动态变化, 则可以用简单的模型进行计算, 得到 Wallach 假设简单数学表述与解释^[19-20]。在低频情况下, 略去头部对声波的散射作用, 将双耳看成是自由空间中相距 $2a$ 的两点。对位于 $r_s \gg a$ 远场距离, 方向为 (θ_s, φ_s) 的点声源, 按照简单的几何关系, 可以计算出声源到双耳的距离差, 除以声速 c 后, 得到 ITD 为

$$\text{ITD}(\theta_s, \varphi_s) = \frac{2a}{c} \sin \theta_s \cos \varphi_s. \quad (\text{A2})$$

如图 A2(a) 所示, 当头部绕垂直轴沿顺时针方向向右转动一个微小的角度 $\delta\theta$ 后, 仰角 φ_s 不变, 方位角变为 $\theta_s \rightarrow \theta_s + \delta\theta$, 由式 (A2) 得到 ITD 随 $\Delta\theta$ 的变化率为

$$\frac{d\text{ITD}(\theta_s, \varphi_s)}{d(\delta\theta)} = \frac{2a}{c} \cos \theta_s \cos \varphi_s. \quad (\text{A3})$$

当头部绕前后轴沿顺时针方向向右转动一个微小的角度

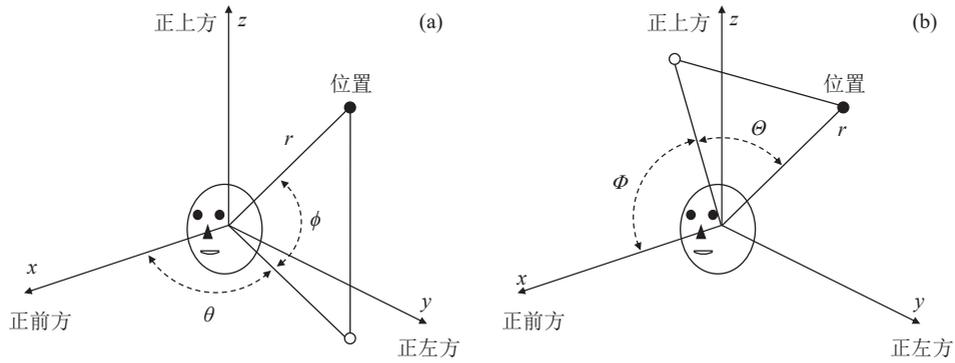


图 A1 本文采用的坐标系 (a) 球坐标系; (b) 双耳极坐标系

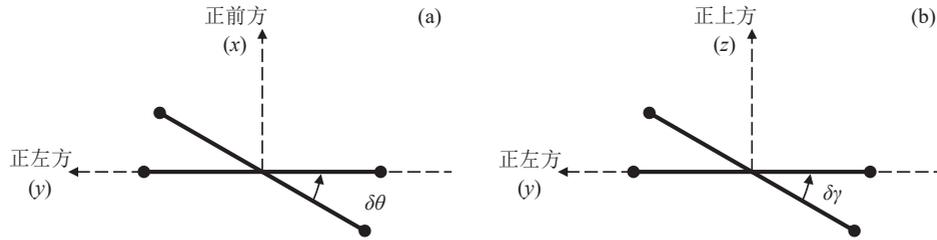


图 A2 头部绕垂直和前后轴转动 (a) 绕垂直轴转动; (b) 绕前后轴转动

$\delta\gamma$ 后, 声源相对头部的方位角和仰角都发生改变, 经过一些简单的几何运算, 可以得到 ITD 随 $\delta\gamma$ 的变化率为

$$\frac{d\text{ITD}(\theta_s, \varphi_s)}{d(\delta\gamma)} = \frac{2a}{c} \sin\varphi_s. \quad (\text{A4})$$

式 (A2) 一式 (A4) 表明, 在 ITD 的基础上增加头部绕垂直和前后轴转动引起的 ITD 变化率信息可以唯一地确定声源的空间方向, 即 ITD 的动态变化提供了前后和垂直定位信息。例如, 对水平面 $\varphi_s = 0^\circ$ 的特殊情况, 由式 (A2) 得到 $\text{ITD} = 2a\sin\theta_s/c$ 。由于 θ_s 和 $180^\circ - \theta_s$ 方向的 ITD 是相同的, 因而 ITD 不足以分辨前后镜像方向的声源。增加了头部绕垂直轴转动的信息后, 由式 (A3) 得到 ITD 变化率是 $2a\cos\theta_s/c$ 。前后镜像方向的 ITD 变化率符号是相反的, 因而提供了前后的信息。

类似地, 对中垂面 $\theta_s = 0^\circ$ 或 180° 的特殊情况, 由式 (A2) 可得 $\text{ITD} = 0$ 。增加头部绕垂直轴转动的信息后, 由式 (A3) 得到 ITD 变化率是 $\pm 2a\cos\varphi_s/c$, 其中前中垂面 $\theta_s = 0^\circ$ 取正号, 后中垂面 $\theta_s = 180^\circ$ 取负号。由于前后中垂面 ITD 变化率符号相反, 因而提供了前后定位信息, 并且 ITD 变化率的幅度正比于 $\cos\varphi_s$ 。在水平面 $\varphi_s = 0^\circ$, ITD 变化率最大; 随着声源向上或向下偏离水平面, ITD 的变化率减少; 在正上、正下方 $\varphi_s = \pm 90^\circ$, ITD 的变化率为零。因而头部绕垂直轴转动也提供了垂直偏离水平面的信息。但另一方面, 由于上下对称, $\pm\varphi_s$ 方向头部绕垂直轴转动引起的 ITD 的变化率是相同的, 需要增加头部绕前后轴转动引起的 ITD 变化率 (式 (A4)), 由于 $\sin\varphi_s$ 在 $\pm\varphi_s$ 方向的符号相反, 因而可以辅助确定声源的上下位置。以上正是

Wallach 假设的数学表述。

对多扬声器重放的情况, 假设 M 个扬声器布置在远场距离的球面上, 倾听者的头中心与球心 (坐标原点) 重合, 第 i 个扬声器的方向为 (θ_i, φ_i) , 信号的归一化振幅为 A_i (同相或反相)。同样采用上述简单的模型, 可计算出双耳叠加声压以及低频 ITD。假设低频 ITD 及其动态变化是方向定位的主要因素, 将它们与上述单声源的情况比较, 多扬声器重放的感知虚拟源方向由以下三个公式决定^[19,40]。

头部固定不动、考虑重放产生 ITD 的情况:

$$\sin\theta_l \cos\varphi_l = \frac{\sum_{i=0}^{M-1} A_i \sin\theta_i \cos\varphi_i}{\sum_{i=0}^{M-1} A_i}, \quad (\text{A5})$$

头部绕垂直轴转动、考虑重放产生的 ITD 变化率的情况:

$$\cos\theta_l' \cos\varphi_l' = \frac{\sum_{i=0}^{M-1} A_i \cos\theta_i \cos\varphi_i}{\sum_{i=0}^{M-1} A_i}, \quad (\text{A6})$$

头部绕前后轴转动、考虑重放产生的 ITD 变化率的情况:

$$\sin\varphi_l'' = \frac{\sum_{i=0}^{M-1} A_i \sin\varphi_i}{\sum_{i=0}^{M-1} A_i}. \quad (\text{A7})$$

如果这三个公式得到一致的结果, 则重放可以产生稳定的空间虚拟源。对水平面前方两扬声器布置的特殊情况, 式 (A5) 将简化为普通立体声重放的正弦定理^[4,48]。

利用 HRTF 可以更精确地计算 ITD 及其动态变化, 以及 ILD、每个耳的声压幅度谱等。相关法计算 ITD 是使双耳声压或 HRTF 的归一化互相关函数最大的双耳间延时, 在频域的计算公式为^[4]

$$\Psi_{LR}(r_S, \theta_S, \varphi_S, \tau) = \frac{\int H_L^*(r_S, \theta_S, \varphi_S, f) H_R(r_S, \theta_S, \varphi_S, f) \exp(j2\pi f\tau) df}{\left\{ \left[\int |H_L(r_S, \theta_S, \varphi_S, f)|^2 df \right] \left[\int |H_R(r_S, \theta_S, \varphi_S, f)|^2 df \right] \right\}^{1/2}},$$

$$\text{ITD}(r_S, \theta_S, \varphi_S) = \max\{\Psi_{LR}(r_S, \theta_S, \varphi_S, \tau), |\tau| \leq 1 \text{ ms}\}. \quad (\text{A8})$$

相关法得到的是一定频率范围 (由式 (A8) 对频率的积分区间决定) 的“平均”或“加权”ITD。

ILD 可定义为

$$\text{ILD}(r_S, \theta_S, \varphi_S, f) = 20 \log_{10} \left| \frac{H_L(r_S, \theta_S, \varphi_S, f)}{H_R(r_S, \theta_S, \varphi_S, f)} \right| \quad (\text{dB}). \quad (\text{A9})$$

每个耳的声压对数幅度谱可直接由 HRTF 计算:

$$S p_\alpha(r_S, \theta_S, \varphi_S, f) = 20 \log_{10} |H_\alpha(r_S, \theta_S, \varphi_S, f)| \quad (\text{dB}),$$

$$\alpha = \text{L, R}. \quad (\text{A10})$$

至于动态因素, 当头部转动后, 分别由声源相对头部的方向计算出 ITD、ILD、 $S p_\alpha$ 并和头部转动前的情况比较, 即可得到相应的动态变化。

对多扬声器重放的情况, 可由 HRTF 先计算出各扬声器在双耳产生的叠加声压, 然后用上述方法计算各种定位因素:

$$P'_L(\theta_S, \varphi_S, f) = \sum_{i=0}^{M-1} \frac{1}{4\pi r_i} H_L(\theta_i, \varphi_i, f) E_i,$$

$$P'_R(\theta_S, \varphi_S, f) = \sum_{i=0}^{M-1} \frac{1}{4\pi r_i} H_R(\theta_i, \varphi_i, f) E_i. \quad (\text{A11})$$