SOUND PARTICLES

音频知识3D小百科

Nuno Fonseca博士

卢峰 编译

www.soundparticles.com

前言

2020 年 8 月

沉浸式音频是音频行业的大势所趋。电影业、电视、虚拟现实/增强现实技术 (VR/AR)、电子游戏、音乐、博物馆等领域都已开始引进 3D 音频技术，让观众沉浸其中、改善体验。但这一趋势也带来了新概念和专业人士最初可能无法理解的新/旧技术，如 7.1.2 通道、幅度矢量合成(VBAP)、22.2 通道、对象、beds、B-格式、AmbiX、ACN、SN3D、头相关传递函数 (HRTF)、头相关脉冲响应 (HRIR) 等流行词。

在撰写这个小册子的过程中，我根据自己在音频工程学会 (AES) 大会中所进行的几个演讲，试图总结关于该主题最相关的信息，包括基于通道音频、基于对象音频、Ambisonics以及双耳音频。

祝您阅读愉快。欢迎来到 3D 音效的世界……

Nuno Fonseca
nuno.fonseca@soundparticles.com

关于 Nuno

Nuno Fonseca（博士）是 Sound Particles 公司的创始人兼首席执行官，是 Sound Particles 软件创始人，这款类似 3D 计算机生成图像 (CGI) 的音频制作软件曾用于许多知名影视作品之中，例如《星球大战 9》、《冰雪奇缘 2》、《海王》和《权力的游戏》。

Nuno 曾经担任大学教授（计算机科学与音乐技术领域），编写了葡萄牙语《Introdução à Engenharia de Som（声音工程导论）》，合著葡萄牙语《Desenvolvimento em iOS（iOS 开发）》，发表 20 多篇关于音频研究的论文。

他是音频工程学会 (AES)、电影电视工程师协会 (SMPTE)、计算机技术学会 (CAS)、音效剪辑者协会 (MPSE)、电影声音协会 (AMPS) 以及音频工程学会 (AES)“电影音频”技术委员会的成员。

中文版简介

我是无意中在领英上看到Nuno关于3D audio小册子的链接，一读之下，觉得这个小册子的内容组织的由浅入深，条理清晰，是难得的总结了环绕声和沉浸式音频的入门读物。虽然他的目标读者是录音混音专业人士，但是对我这种重放端的音频专业人员一样有教育意义，帮我们理清了很多的相关的概念。

我当时就想把它翻译为中文，接触了Nuno后，原来他已经安排了中文版的翻译，问我愿不愿帮忙审阅一下翻译稿，我立刻欣然答应了。

今年8月份，Nuno把翻译稿寄给我。我发现因为译者不是音频专业人士，有很多专业术语不正确，很多地方因为对技术不理解，表达的意思错误。所以，我帮手对译文进行了很多修正。

在这里我想对一些概念进一步说明一下：

- Pan/Panning/Panner：它的意思就是调整声像在空间的位置的处理/信息/设备。最初在立体声录音中，就是调整左右的相对位置，后来环绕声和3D音频出现后，泛指调整到空间任何位置。

- Bed：在基于目标的音频发行格式里，并不是所有声音都是基于目标的，很多还是混音于固定的通道，用基于通道的传统方法发行，这一部分称为bed。中文还没有正式的术语。

- Ambisonics：有中文翻译为“环境空间谐波”，大部分中文资料都是直接用英文术语。

- 麦克风的指向：当英文提到8字型麦克风指向两侧时，意思是它的两个主旁瓣指向两侧，而我们看到的棍状的麦克风整体从视觉上还是指向前方的。这种8字型麦克风里面拾音头的振膜是平行于棍状的麦克风整体的，和心型指向麦克风的振膜方向不同。

- 单声道（mono）：不管在中文还是英文中，这个词在不同情景下运用，有些许不同的意思。比如在基于目标的音频中，每个目标都是一个单声道音频，这里指它是一个独立通道的声音波形文件或音轨。又比如在ambisonics中提到的单声道成分，是指拾取所有的方向的声音，就好像混合立体声的左右通道变成了一个单声道音频一样。

- Decode / Encode：除了我们熟悉的音频编码格式里的codec会用到编码和解码，这个词也被借用描述对声音的空间位置的处理的不同阶段。比如基于目标的音频里，某一个音轨和metadata在一起encode了声目标，在重放阶段，会因应本地的扬声器设置，decode到各个通道去。又比如，Ambisonics的录音encode了所有的空间音频的信息，到了重放阶段需要因应本地的扬声器设置，decode到各个通道去。有时候（没有包含在这个小册子里），重放的扬声器是一个个阵列，主处理器会把含有空间信息的音频数据encode后，打包传送给扬声器阵列，扬声器因应自身阵列的情况，decode出阵列内各通道的音频内容。

- Render：这是经常在软件/网络专业上使用的词汇，指的是在音频重放阶段，把音频解码后，进行处理，如3D audio decode，应用音量调整，动态调整等，得到最终的各通道音频信号重放。有时翻译为合成或呈现。而进行解码，处理和重放的设备，称为renderer。

我还和Nuno讨论了为何小册子里缺少decode Ambisonics的部分。Nuno说，decode Ambisonics非常的复杂，远超出了这个小册子所覆盖的水平。因为目标的读者是录音混音专业人士，他觉得应该让他们明白Ambisonics encoding的概念，可以在DAW上看到并操作正确的波形。至于decode的工作，会有专门的插件来处理。

如果有读者了解正式的中文术语，或有任何意见，评论，可以提供给我们。

希望这个小册子可以让你有所得益。

卢峰

TCL鸿鹄实验室高级声学和音频专家

fengflu@hotmail.com

© 2021 Sound Particles Lda

目录

基于通道音频 1

- 立体声 1
- 中置扬声器 2
- 环绕扬声器 3
- 低频效果通道 / 重低音扬声器 4
- 其他水平布局（6.1/7.1） 5
- 沉浸式音频 5
- 优点与缺点 8

基于对象音频 9

- 基于通道音频与基于对象音频 9
- 对象 11
- 杜比全景声DOLBY ATMOS 12
- BEDS 14
- 优点与缺点 15

AMBISONICS 17

- M/S技术 17
- AMBISONICS 19
- AMBISONICS麦克风 20
- 高阶AMBISONICS (HOA) 21
- 多个AMBISONICS版本 22
- VR 应用 24
- 优点与缺点 25

双耳音频 26

- 方向感 26
- 头相关传递函数 (HRTF) 27
- HRTF 个人化 28
- 使用双耳音频 30
- 优点与缺点 31

总结 32

基于通道音频

许多音频格式都会预先定义音频通道数量和扬声器位置，如立体声或 5.1 通道，因此又被称为**基于通道音频**。在我们深入了解和探讨 3D 及沉浸式音效前，需要更好地理解一些基本的声音格式，以便为学习 3D 音频奠定基础，这一点至关重要。那么，我们先从最简单的格式开始。

立体声

立体声试图用两个扬声器重建**前方声像**。如果向这两个扬声器发送相同的信号，就会产生一个幻像声源，我们感觉声音就来自正前方。如果改变 2 个扬声器信号之间的增益关系，就可以让声音在左右扬声器之间“移动”。

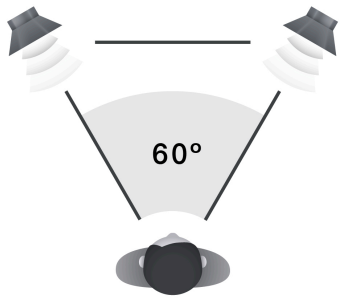

图 1 - 立体声布局：两个角度约为 60°的扬声器与人形成一个完美的等边三角形。

立体声设置中，我们会通常认为扬声器的角度应该为 60 度。为什么？

一方面，我们希望声像越宽越好。如果角度太小（将两个扬声器放的很近），声像就会太窄、太小。另一方面，如果我们扩大角度大幅超过 60 度（将两个扬声器分得很远），就会造成前方声像不连续。我们将失去中间的幻像声源，变成两个独立的音源，可以清晰感受到一个声音来自左侧扬声器，另一个声音来自右侧扬声器。左右两侧声音产生一个**中间空洞**，我们感到那个位置没有声音。

中置扬声器

随着电影行业的发展，专业人员决定在左右扬声器之间添加一个中置扬声器，这就引出了一个问题：“如果能在一对立体声扬声器之间创建出声像，为什么还需要在中间添加一个中置扬声器？”

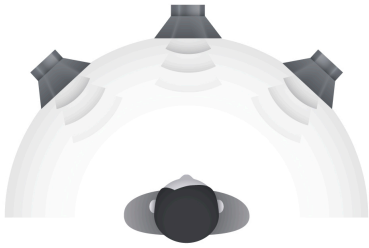

图 2 - 立体声布局 + 中置扬声器

有时，我和一些世界顶尖混音师交流，他们在著名专辑的混音中有利用环绕音，我甚至听到他们表示根本不用中置扬声器，因为在创作中不觉得缺少中间的声像。那么我们**为什么还**要中置扬声器？

任何事情**都不可能始终完美**，在空间音频世界中，这意味着并非所有人都能坐在房间中的最佳聆听位（完美聆听位置）。如果您处于最佳聆听位，左右扬声器可以在他们之间形成前方声像，但在很多时候，听者不在最佳聆听位。

例如在电影院，只有少数人能荣幸坐在**最佳聆听位**。大多数听众的座位过于靠两侧、靠前或靠后。左右扬声器位于投影屏幕边缘附近的后方。可以想象，坐在前排的听众左右扬声器之间的角度必然大于 60 度，这会造成前方声像不连续，形成中间空洞。同样，两侧的听众会更靠近其中一侧扬声器，会造成声像“失真”（不是音频失真，而是声音的方向感扭曲）。

在上述所有情况下，我们都需要中置扬声器，它就像一个**锚**，帮助我们感觉到更明确稳定的声像，尤其是帮助没有坐到最佳聆听位的听者。所以中置扬声器被广泛用于电影行业，因为在影院中，观众都分散在影厅中的不同位置，仅有少数幸运儿才能坐在最佳聆听位。

环绕扬声器

为了改善观影体验，人们开始在电影院两侧增设扬声器。电影不再仅仅提供前方声像，还会通过两侧的声音提供其他方向的声音，让聆听体验更真实，听众更能沉浸于场景之中（比如战争场面）。

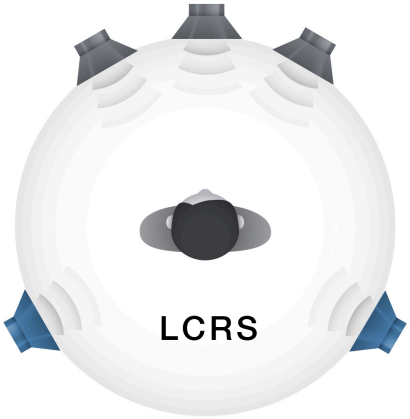

图 3 - LCRS 布局

其中一种早期格式是LCRS（左通道、中置通道、右通道、环绕声通道）。通过 LCRS，我们开始有个 5 个扬声器和 4 个音频通道的布局。在这种情况下，环绕声通道只是一个音频通道，即左环绕扬声器和右环绕扬声器重放完全相同的音频。

后来，5.1 通道等格式出现，开始独立的左环绕通道和右环绕通道，每边都有自己独立的音频通道。

在电影院，各环绕声通道并非由单个扬声器重放，而是各由**一组扬声器重放**。在配备 5.1 通道音频系统的电影院里，你会发现：两侧和后方的墙壁上安装着许多扬声器。但这些扬声器其实只重放两个音频通道。这意味着左墙的所有扬声器和后墙左侧的那些扬声器重放完全相同的音频通道，右侧扬声器也遵循同样原理，重放右环绕通道。

为什么我们需要设置如此多的扬声器，而不是每侧只摆放一个扬声器？

第一个原因：增加更多扬声器可以防止声像“**空洞**”。不错，现在的环绕声在空间位置的精确度降低了，变得更加**扩散**，几乎没有聚焦，但不会出现不连续的空洞，所有方向都有声音。

其次，**世界并不完美**，很多人都无法坐到最佳聆听位。通过这种扬声器阵列的方法，无论您坐在何处，周围都会有几个扬声器。如果电影院中每个环绕声通道只使用一个扬声器，就会导致部分听众感受到很大的“空洞”：对于后排听众，环绕扬声器在其两侧（可能会略微靠前），导致无法感知后方音效；两侧靠前听众可能会听到后方扬声器的声音，但却无法感知附近侧墙方向的声音。

低频效果通道 / 重低音扬声器

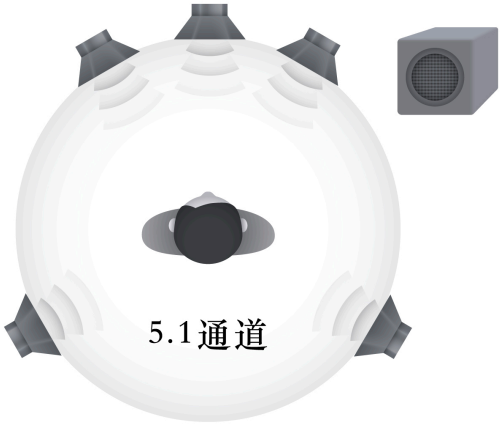

图 4 — 5.1 通道布局环绕

随着电影行业的发展，电影制作方会希望通过更低频率声音提供更震撼的效果，如爆炸或地震。**低频声音**的问题在于：人类对这些频率范围的声音不敏感。想要听到 80 赫兹以下的频率的声音，就必须将其加强，提高声压级 (SPL)，这就需要非常强大的扬声器重放这些频率的信号。

重放这些频率的声音就需要**大幅增加功率**，以下 2 种方法将可能达到期望结果。

第一种显而易见的解决方案是：使用**更大功率的扬声器**即更**强大的音响系统**重放低频声音。这是一个昂贵的解决方案，因为我们必须把所有扬声器替换为更大功率的新扬声器。还有更重要的，我们必须处理**动态范围**的问题。那个时代使用的是模拟系统，其**动态范围**比较窄，这意味着系统最大声音和背景噪音之间的范围较小。虽然提高功率使得系统可以重放低频声音，但也提高了系统的背景噪音水平，这在伴音音量很小的段落中更加明显，比如对话场景。

第二种同时也更有优势的方法：**只增加一个专门重放低频的扬声器，馈给它一个单独的通道**。为什么？第一，人对低频声音的方向不灵敏，这意味着使用单一的扬声器（虽然物理上声音来自单一方向）不会影响人的空间感或损害声像。第二，其他所有扬声器可以保持在原本的功率范围工作，不会产生动态范围的问题，如噪音水平增加。第三，通过拥有独立的音频通道 (LFE)，系统可以预设以比其他通道更高的增益重放该通道（即在不影响其他通道动态范围的情况下，增加低频的功率）。第四，电影院只需要再买一个扬声器即可，无需替换所有扬声器。通过这种方法，5.1 通道中的“1”就产生了： 5 个主要通道（左通道、中置通道、右通道、左环绕声和右环绕声）以及 1 个用于低频效果的通道LFE（名称中的“1”部分）。

需要注意的是，LFE 和**重低音扬声器**是不同的概念，LFE 是传输这些低频的通道名称，重低音扬声器（低音炮）是重放这些频率的扬声器名称。

其他水平布局（6.1/7.1）

除了 5.1 通道之外，通过增加额外通道产生了其他布局。

索尼在 90 年代创建了 SDDS，将其设置为 7.1 通道，但在前方设置 5 个扬声器：左侧、中间偏左、中置、中间偏右、右侧，同时保留 2 个环绕通道。由于在前方增加了扬声器，SDDS 主要应用于大屏幕影院，因为在这样的影院中，即使设置 3 个扬声器也不足以满足前排观众的需要。

之后，杜比创建了Dolby Digital Ex，首次应用于《星球大战》第 1 部。这是一种带有一个后置环绕扬声器的 6.1 通道格式，能够增强后方的声音感知。

最终，目前已经是“传统”的 7.1 通道出现了，这种布局包含 3 个前方通道和 4 个环绕通道，提高了两侧/后方的空间分辨率。

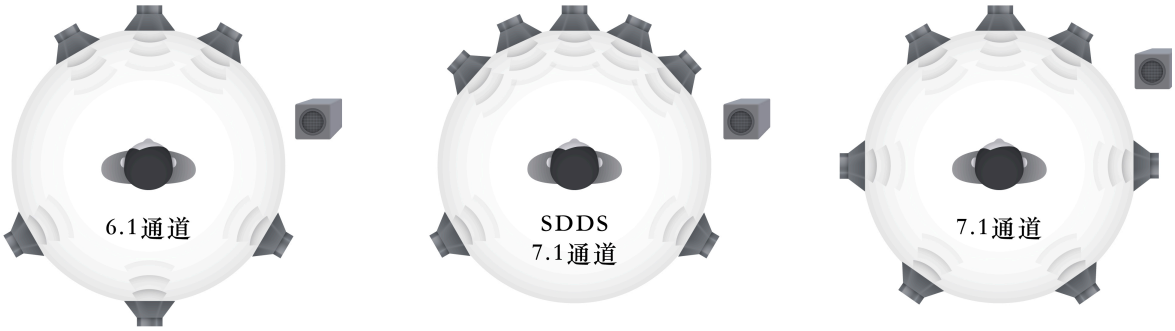

图 5 - 杜比数码 6.1 通道布局；索尼 7.1 通道布局；目前的 7.1 通道格式

沉浸式音频

“沉浸式音频”的概念通常是指：让听者感受水平面以外其他方向声音的音频系统。沉浸式系统不止将扬声器放在水平面（前方、两侧以及后方），还在例如天花板、甚至地板上增加扬声器，能产生声音在不同高度的方向感。

为了产生不同高度的声音方向感，出现了几种格式：Auro 11.1/13.1、IMAX 12.0、NHK 22.2等格式。

例如，Auro 11.1 / Auro 13.1 设置了 3 层扬声器：在水平面设置一个传统的5.1通道或 7.1 通道扬声器组，在仰角约呈 30°角的高度层增加一层 5.0通道的扬声器组，然后在听众正上方设置一个扬声器（它有个很形象的比喻："上帝之声"）。这意味着 5.1 + 5.0 + 1 = 11.1 或 7.1 + 5.0 + 1 = 13.1。

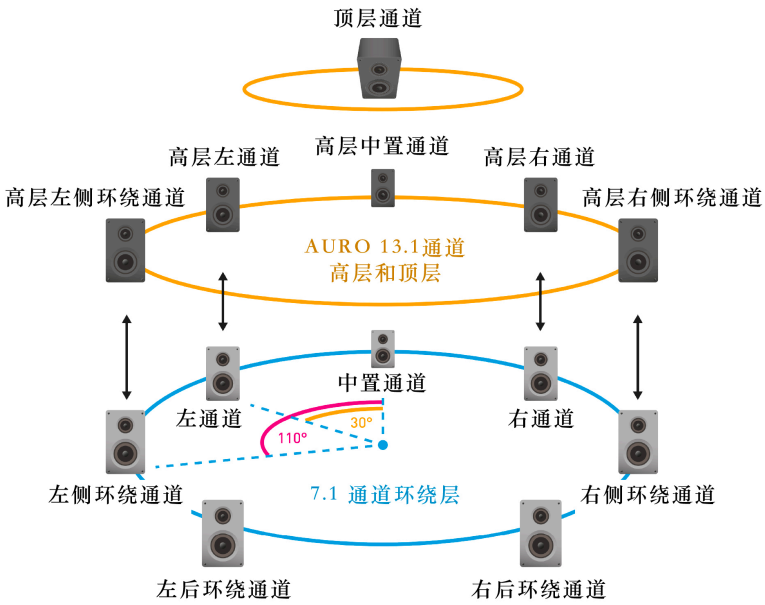

图 6 - Auro 13.1 通道（由 Auro 科技公司提供）

IMAX 12.0 采用的方法略有不同。

这种格式设置 12 个通道，其中在水平面上设置传统的 7 个通道，再在高层设置 5 个通道提供不同高度方向的声音，没有 LFE 通道（"12.0"中的 0）。

NHK 22.2 通道是由日本广播公司创建的格式，更进一步，包含 2 个 LFE 通道（左右各一个），在低层设置 3 个扬声器（放在前方低于水平面的位置，声音将从地下方向传出），在水平层面设置 10 个通道（前方 5 个，环绕声 5 个），在高层设置 8 个通道 + 1 个"上帝之声"(3.2 + 10 + 8 + 1 = 22.2)。

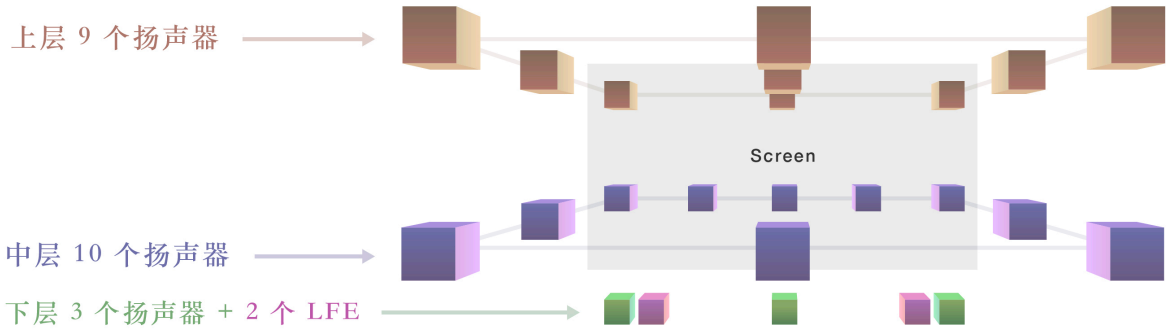

图 7 - NHK 22.2 通道布局

3D PANNING

在水平面布局，声音通常会在两个扬声器之间Panning，虚拟出扬声器之间不同方向的声音。同样的情况也适用于沉浸式音频布局。但现在我们考虑由 3 个扬声器形成的三角形 — 通过改变每个扬声器播放的声音强度，让听者可以感知来自扬声器三角形区域内任何位置的声音。这种Panning技术被称为幅度矢量合成 (VBAP)。

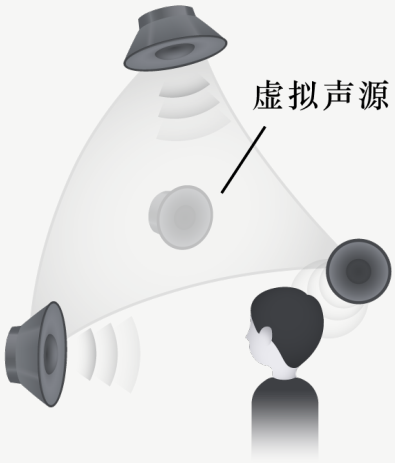

图 8 - 3D Panning

优点与缺点

基于通道音频有利有弊。主要优点在于，如果我们已经明确知道音频重放的扬声器布局，就可以打造一个**完美的制作+重放系统**。例如，如果我事先知道混音作品会被重放到一个每个扬声器都会放在规定的位置的5.1 通道系统。那就好了！我可以在录音室中使用完全相同条件的布局，制作一个完美的混音。

基于通道音频的主要缺点是**缺乏灵活性**，因为此类音频制作固定于某个特定的扬声器布局，如果需要改变输出扬声器布局，就需要重新制作混音。例如，如果某部电影发行的格式有立体声、 5.1 通道、7.1 通道、11.1 通道和 22.2 通道，就需要 5 种不同的混音，每种格式各一种。

这种情况就会带来一个问题："我们能否创建一种不依赖于输出通道布局的格式？我们能否创建一种格式，只需一次混音就能自动适应不同的扬声器布局？"答案是：可以，我们需要采用**基于对象音频**或**Ambisonics音频**。

基于对象音频

随着杜比全景声的出现，基于对象音频开始受到关注。但到底什么是基于对象音频？

基于通道音频与基于对象音频

在开始了解基于对象音频前，让我们先来分析如何制作并发行基于通道音频。

基于通道音频的基础是一个预先设定的扬声器布局 — 例如立体声或 5.1 通道，这意味着我们已经知道重放混音音频时的布局。据此，我们为这些通道创建一个混音版本后发行（例如经流媒体，存储在文件、DVD 中）。在重放过程中，只需要将每个通道的音频信号发送到相应的扬声器即可。

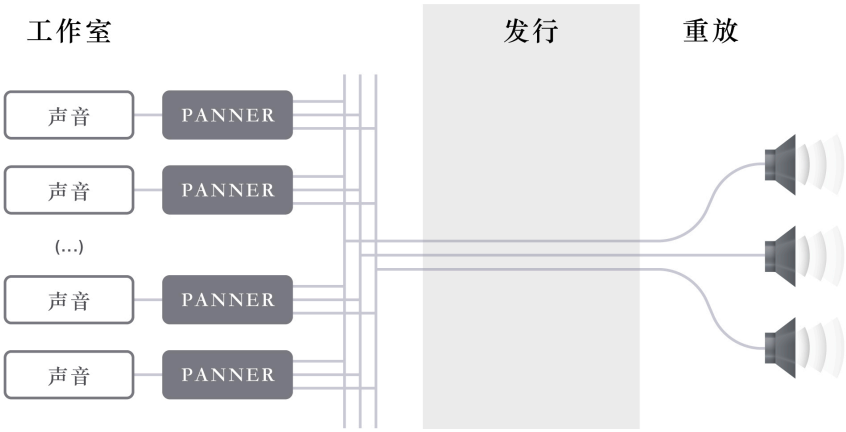

图 9 - 基于通道音频

在录音室，每个声音都要经过Panner，它控制在每个输出通道上应该分配的声音的相对幅度。例如在 5.1 通道混音中，如果我希望将某个声音设置在中置和右侧扬声器之间的某处，就会使用Panner进行控制，在中间和右侧通道引入正确幅度的信号，但不会引入到其余通道上。

之后，所有Panner输出的声音在发行前混音在一起（使用总线）：将所有panning到左通道输出的声音混合在一起，成为左通道混音，右通道也是一样，以此类推。

最后，我们发行的是混音后的信号（DVD、流媒体等），以供之后重放。

基于对象音频的概念略有不同。不同于在录音室中混合所有声音并发行最终的混音，基于对象音频，在发行时所有声音都是独立分开的，**在重放的时候才进行混音**。

在录音室中仍要使用Panner来定位声音，但不会将panning的信息应用到声音上： 您只需要指明声音应该定位在何处。经过发行后，在重放过程中，这些panning信息将根据重放系统的实际情况应用到声音上。

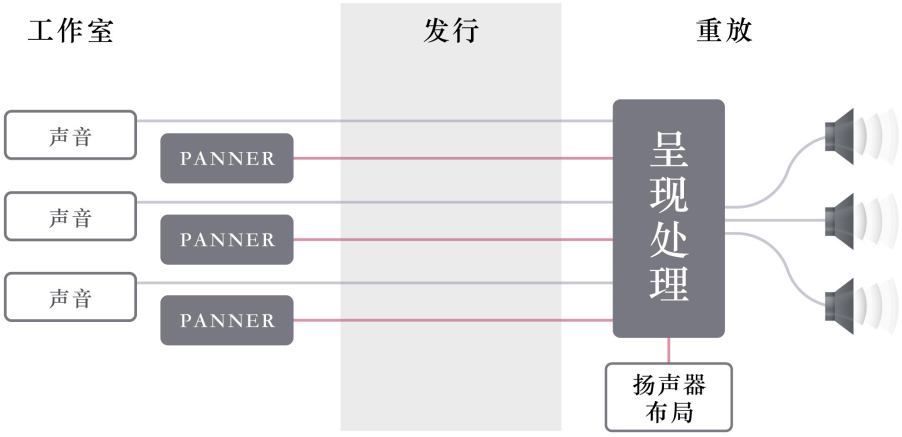

图 10 - 基于对象音频

从本质上说，在发行中，您是在**发布您的目标**，“我希望将这个声音出现在这里，另一个声音出现在那里。”重放音频系统知道其扬声器数目和每个扬声器位置的信息，它在重放中以这些信息为基础对所有的声音进行实时处理，确保将正确信号发送给每个扬声器，实现制作者的目标。

让我们来看一个实际应用的例子：假设您希望一个声音出现在“中置通道”和“右通道”之间。如果是基于对象音频，我们会发布这个声音的音频流以及理想的panning位置。然后在重放过程中，根据扬声器实际设置的不同，声音最终将会在正确的几个扬声器中“呈现render”。如果影院有 3 个前置扬声器（“左侧”、“中置”、“右侧”），声音将被发送到“中置”和“右侧”的扬声器，但如果电影院有 5 个前置扬声器（“左侧”、“中间偏左”、“中置”、“中间偏右”、“右侧”），声音将被发送到“中置”和“中间偏右”或“右侧”和“中间偏右”的扬声器，这取决于声音的定位是否更接近正中间。

另一个例子：一辆汽车从左侧经过的场景。如果是基于通道音频，将完全在录音室完成混音。最终在 7.1 通道系统中，声音会从“左环绕后置”开始播放，逐渐移动到“左环绕侧方”，并在“左前置”通道终止。如果是基于对象音频，假设电影院在左墙设置 5 个扬声器，声音将从最后面的 5 号扬声器开始播放，依次移动到 4 号扬声器、3 号扬声器、2 号扬声器、1 号扬声器，最终移动到左前置通道。如果影院在左墙上设置了 8 个扬声器，顺序将为 8→7→6→5→4→3→2→1→左前置通道，利用额外的扬声器达到更高的空间分辨率。

当然，虽然我们说，仅在重放过程中完成混音，显而易见，混音师需要使用自己的重放系统来重放混音，在录音室/配音阶段听到他指定的目标重放后的效果。

对象

为某个特定声音发布的panning信息不是一个静态值，而是一个**会随时间变化的值**。还是用汽车经过的例子：我们只有一个（单声道）声音，但这个声音的panning信息会随时间而变化。此外，除了声音的位置，还可以考虑其他信息。例如，基于对象音频系统也可以使用“**尺寸**”概念：声音是一个准确到点（仅使用 1 到 2 个扬声器重放声音），还是声音散开些，覆盖大一些的区域（在极端情况下，声音可以被所有扬声器重放）。还是用汽车经过的例子，对后面一种情况，重放的顺序不是：

8 → 7 → 6 → 5 → 4 → 3 → 2 → 1 → 左前置通道

而是使用以下顺序：

8765 → 7654 → 6543 → 5432 → 4321 → 321 左前置通道

此外，在某些情况下，混音师可能不关注声音的确切位置，而是侧重于确保单个扬声器重放某个特定声音[1]，他们会将重放系统设置为**选择离某个特定方向最近的扬声器**，只用该扬声器重放。

[1] 通常是为了避免两个扬声器由于和听者的距离不同产生的梳状滤波引起的音色变化。当然，这只适用于静态的声目标。

与每个声音一起发布的所有信息（例如位置、尺寸等）都被称为“**元数据**”（metadata关于数据的数据）。每个声音（及其相应的元数据）一起称为一个“**对象**”，其完整描述了一个在 3D 空间中移动的（单声道）声音对象。

在基于对象音频中，我们制作的是能够适应重放系统的音频：既可以使用10 个扬声器的系统也可以使用 50 个扬声器的系统来重放这些声音，利用额外的扬声器可达到更高的空间分辨率。

此外，由于最终我们在重放期间进行混音，甚至还可以让听者自己**启用**或**禁用**一些音频对象或beds。这种方式更适合电视广播应用，比如在体育赛事中，听众可以选择要不要解说，或者甚至根据他们喜欢的队伍来选择声音，切换为靠近 A 队或 B 队支持者附近的麦克风的声音。

杜比全景声Dolby Atmos

杜比全景声是使用基于对象音频的最为人所知的系统之一，因此，我们会将杜比全景声作为基于对象音频的一个示例进行研究。但还有其他基于对象音频格式，例如 **MPEG-H**、**AuroMax**、**DTS:X**。

杜比全景声既支持基于对象音频（**对象**），也支持基于通道音频（**beds**）。一个对象会有自己的音频（单通道音频剪辑）和相应的元数据，但杜比全景声也可以通过使用 7.1.2 通道的beds（7.1 通道 + 2 个高层通道）支持基于通道音频。我们首先来了解对象。

杜比全景声支持 128 个独立通道，这意味着我们可能有多达 118 个音频对象，以及为 7.1.2 通道bed预留的 10 个通道（更多关于bed的信息见后文）。在重放过程中，杜比全景声支持最多 **64 个独立扬声器**，这意味着电影院可以设置最多 64 个独立控制的扬声器，每个扬声器都可以重放与其他所有扬声器不同的音频。如果其中一些扬声器重放相同的音频信号，影院里甚至可以设置 64 个以上的扬声器。

杜比全景声影院与常规的 5.1/7.1 通道影院略有不同，如图 11 所示

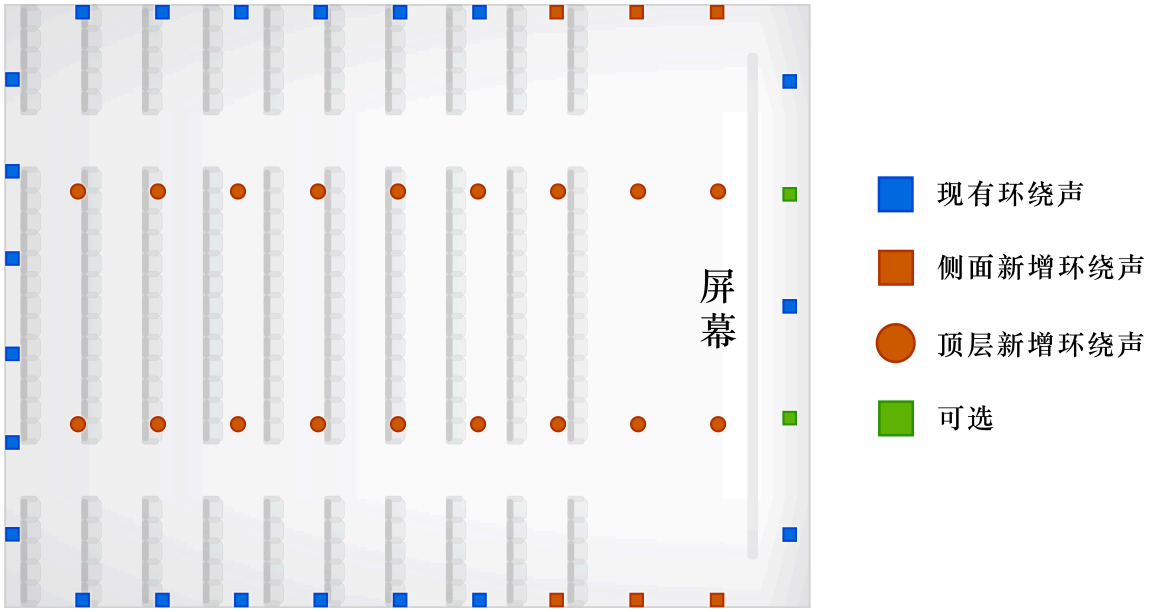

图 11 - 5.1/7.1 通道和杜比全景声影院之间的区别

最明显的区别是天花板上的扬声器。在杜比全景声中，电影院在天花板上安装 2 排扬声器，开始提供不同高度的声音。

在前方，将继续在屏幕后面设置“左前置”、“中置”、“右前置”扬声器。但在屏幕更大的影厅，可以选择在中置和左/右前置扬声器之间再安装 2 个扬声器。

此外，影厅内最多可以增加 4 个重低音扬声器。

最后，在靠近前置扬声器的两侧增加额外的扬声器。在传统影厅中，我们通常从影厅边长的 1/3 处开始设置环绕扬声器，只占据侧墙后侧的 2/3，这意味着屏幕和最前的环绕扬声器之间有一段空缺。在杜比全景声中，这段空缺将不复存在，通过使用额外的扬声器，确保声音可来自整个水平面所有方向。

Beds

在电影制作过程中，一些音频内容可能已经预先混音。例如，可能传送给终混[2]的配乐是已完成预混的，或者至少预混在各主音轨（stem）中（7.1 通道弦乐、7.1 通道铜管乐等）；或者想象已经用5.1 通道录制或编辑过的音效环境声部分。

除对象外，杜比全景声还支持基于通道的内容，这种内容被称为音频**Beds**。这意味着除对象外，我们还可以包含“常规”基于通道音频。杜比全景声使用 7.1.2 通道Beds，相当于传统的 7.1 通道（3 个前置通道 + 4 个环绕通道 + LFE）加上 2 个用于天花板的通道（一个在天花板的左侧，另一个在天花板的右侧）。例如，所有已有的 5.1 和 7.1 声轨都可以直接混合到 7.1.2 通道Bed。即使不使用对象，混音师仍然可以通过向头上方扬声器发送音频来提供不同高度的声音。

布局表达格式（如 7.1.2 通道）

多年以来，我们都使用 2 个数字的方式表示布局格式。例如：5.1 通道即 5 个主通道 + 1 个低频效果（LFE）通道。随着全新沉浸式格式的出现，我们增加了一个新数字，代表天花板上的通道数量。因此，7.1.2 通道意味着 7 个主通道（水平方向）+ 1 个 LFE 通道 + 2 个天花板通道；9.1.4 通道即意味着 9 个主通道（水平方向）+ 1 个 LFE 通道 + 4 个天花板通道等。

这种表达方式可能不够完整（有些布局在水平面（地面）以下有一层通道，有些布局可能在头部上方有 2 层通道等），但至少可以大体提供更多关于音响系统的信息。

[2] Re-recodring再录音，是电影后期制作中音频处理的最后阶段，又称为final mixing终混。将对白、音乐、音效等多种音源予以混合处理，按照发行格式的要求输出。

其实Beds不仅仅用于预混的音频材料。**许多电影制作在混音过程中都会使用数百条音轨**，某些情况下甚至超过 1000 条音轨。记得在我第一次参观环球影业的后期制作时，在配音阶段我曾听人说“这是我们的混音器，带有 512 个音效通道和 256 个对白及音乐通道”。虽然杜比全景声支持 118 个对象，但这还不足以处理所有音轨。但我们可以利用beds解决这个问题，因为我们可以在beds上**混合无限多的声音**。对于对象，每个声音都必须独立，才能控制其panning，但在beds中就不存在这个问题了，因为可以使用“常规”panning来放置不同方向的声音。因此，混音师可以对最重要/有意义的，更需要准确定位的声音使用对象通道，然后在beds上混合次要的声音。

杜比全景声包括一个 7.1.2 通道bed，但可以通过减少可用对象的数量来增加bed。例如，我们可能有 118 个对象 + 1 个bed，或 108 个对象 + 2 个beds，或 98 个对象 + 3 个beds等。但是因为一个bed上可以混合无限多的声音，就声音而言，使用多个beds并没有任何特别好处。

优点与缺点

基于对象音频的主要优点是：**能够适应任何扬声器布局**。我们可以通过将“方位目标”发送到重放系统，让这些方位目标完美适应实际的扬声器布局。还有，我们能够从系统中获得**最高空间分辨率**：通过增加更多扬声器提高重放的空间精确度。

但基于对象音频也存在一些缺点。首先，在发行时，我们需要包括**更多数量的音频通道**。在电影院，杜比全景声可能使用多达 128 个音频通道。但如果我们改用其他媒介（流媒体、蓝光碟），这 128 个音频通道就太多了，我们必须减少对象的数量，这意味着许多对象必须转化为beds的形式呈现，这也会失去对象的优势。其次，对象**存在一定数量的限制**，128 是一个很大的数字，大多数项目不需要用到所有 128 个通道，但在某些情况下，这个数字也很容易达到。例如，Sound Particles 完全能够呈现带有数千个声音对象的音频场景（我们能够同时呈现并播放多达 100 万个声音）。此外，有些好莱坞电影作品会使用 1000 多条音轨。音频beds能够解决以上大多数限制，但这样一来，就不是一种纯粹基于对象音频方法。

开拓之作 — 迪士尼的《幻想曲》

《幻想曲》是迪士尼公司在 1940 年制作的电影，米老鼠戴着巫师帽的经典场景就源于这部电影。在电影制作期间，Walt Disney 认为，音乐在这部电影中的角色至关重要，因此必须给观众带来更完美的声音体验。为此，迪士尼公司和美国无线电公司（RCA）的工程师们研发了“Fantasound”多扬声器音响系统，将其临时安装在几家电影院（部分系统甚至还在美国各地巡回使用），让观众在观看电影时获得更有趣的声音体验。因此，《幻想曲》被认为是**首部使用环绕声的电影**。

Fantasound 系统包括几种不同变化，其中一些系统会在天花板上设置扬声器，因此《幻想曲》不仅是首部使用环绕声的电影，还是**首部使用沉浸式音效的电影**。

从技术上来说，该系统使用 4 个声轨：3 个用于录制的音频的声轨（3 个独立音频信号）和 1 个用于控制信号的声轨，这个声轨将控制 3 个通道重放的方式：在某些时候，可以将通道 1 发送到左扬声器，通道 2 发送到右扬声器，通道 3 设置为环绕声；后面某些段落，可能将通道 1 在到前置扬声器重放，通道 2 在天花板的扬声器重放，通道 3 在后置扬声器重放等。尽管当时使用的是模拟技术，但这种技术也使《幻想曲》成为**首部使用基于对象音频的电影**。

80 年前的首部**环绕声、沉浸式**且使用**基于对象音频**技术的电影！显而易见，这一创举的背后，是具备超越时代远见的开拓者。

Ambisonics

Ambisonics的主要目标是：创建一个独立于输出重放布局的音频格式。如果您用Ambisonics录制声音，就无需关注重放过程中有多少扬声器或这些扬声器的位置，因为Ambisonics都能够适应。

尽管Ambisonics在上世纪 70 年代已经出现（Michael Gerzon 是该技术的奠基人），但在过去十年中，人们对这种格式的兴趣显著增加，这主要与 VR 应用有关。但在讨论Ambisonics之前，让我们先理解比Ambisonics更简单的类似技术：M/S技术。

M/S技术

虽然许多人会使用传统的 XY 立体声录音技术，即用两个心型指向麦克风（一个指向左前侧，另一个指向右前侧），但有的专业人士更青睐M/S（Mid/Side）录音技术。

M/S是一种被广泛使用的立体声录音技术，使用一个指向前方的心型麦克风（Mid中置，后面简称M麦克风）和一个指向两侧的 8 字型麦克风（Side两侧，后面简称S麦克风），如图 12 所示。在这种方法中，M麦克风将拾取主要来自前方的声音。将 8 字型S麦克风指向两侧，就可以拾取来自左侧和右侧的声音，而不是垂直方向的声音（如前方）。

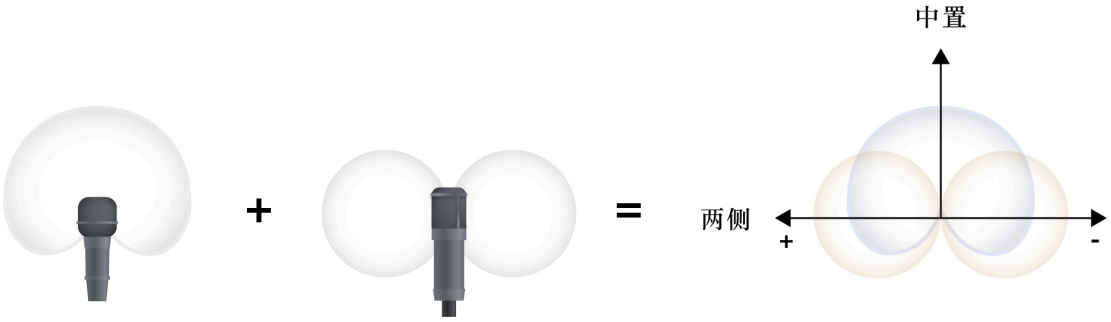

图 12 – M/S技术（留意S麦克风拾取的左右两侧的信号是反相的）

如您所见，我们使用一个麦克风来拾取单声道部分（M麦克风会拾取前方的声音），使用另一个麦克风来拾取非单声道部分（S麦克风不会拾取前方的声音）。

如果我们将M/S信号直接发送到 2 个扬声器会怎样？扬声器仍将重放普通音频，但无法听到立体声像。您需要首先将信号解码到左/右扬声器，达到最佳结果 — 即前方立体声像。

为了得到最终的立体声输出，我们需要通过以下方式混合来自麦克风的信号：

- 通过混合两种信号（M + S）得到左信号
- 通过混合M信号与S麦克风信号的反相版本得到右信号（M - S）

因此，如果您使用的是一对M/S麦克风，当有一个声音来自左侧，解码后这个声音将被通过左通道播放（M和S麦克风拾取到的声音具有相同的相位），解码到右通道会很小（由于M与S麦克风的信号相减后，这两者的信号将会相互抵消）。

利用将M/S转为L/R通道的方式，我们也可以将L/R通道转换为M/S通道[3]，这意味着我们可以在左-右通道和M/S通道之间来回转换。实际上，我们在**制作母带**（mastering）的过程中经常进行这样的转换。假设您需要对歌曲的主唱部分进行小幅压缩，但只有最终的立体声混音（无法访问单独的音轨）。制作母带的专业人员会将左-右通道转换为M/S通道，仅在M通道使用压缩，再将M/S通道转换回左-右通道，而不是将压缩用于整个立体声信号。在这种方式中，压缩主要应用于靠近中间的声音。

此外，我们会在一些**音效装置**中使用这种M/S转换，以增强立体声效果。如果您将某个立体声信号（左-右通道）转换为M/S通道，然后增加两侧通道的增益，将结果转换回立体声，最终的效果是增加所有非单声道部分的增益，增强信号的空间感[4]。

---

[3] 通过下面的等式可能更容易理解：

$$L+R=(M+S)+(M-S)=2M$$

$$L-R=(M+S)-(M-S)=2S$$

[4] 当然，这对纯单声道信号不起作用，因为其S通道是空白。

Ambisonics

从本质上来说，Ambisonics是加强型的M/S技术。让我们想象一下，除了使用一个指向两侧的 8 字型麦克风（左右轴方向）之外，再加上一个指向天花板的 8 字型麦克风（上下轴方向）和一个指向前方的 8 字型麦克风（前后轴方向），再用全向麦克风来代替心型指向中置麦克风，这样拾取的单通道信号包括所有角度的信号。这就是**Ambisonics**。

传统的Ambisonics格式被称为“**一阶Ambisonics（B-格式）**”，这种格式使用 4 个音频通道：

- W 通道 — 全指向性分量，类似使用一个全向麦克风拾取现场声音，这个通道就有来自所有方向的音频。
- X、Y、Z 通道 — 这 3 个通道对应 3 个 8 字型麦克风，每个麦克风指向不同的方向：左-右；前-后；上-下。

当我们谈论Ambisonics时，请记住它有多种不同版本，人们通常指的是“一阶Ambisonics，B - 格式“（后文中有更多信息）。

所以……Ambisonics仍然是**正常的音频信号**，而不是让 3D 音效发挥作用的某种神秘黑魔法。Ambisonics只是音频通道：带有所有方向分量的单声道及以下 3 个分量的音频通道：左-右；前-后；上-下。我们也可以说Ambisonics是M/S技术的 3D 版本。

就像我们不能直接用扬声器重放M/S信号（这样做无法得到理想的前方声像）一样，也无法直接用扬声器重放Ambisonics信号：我们有 4 个音频通道，可以单独听到各个通道的信号（各自的常规音频 — 不是类似噪音的异常信号），但在利用任何扬声器布局重放这些信号前都需要“**解码**”。

Ambisonics麦克风

图 13 - 带有 4 个话筒的四面麦克风：Sennheiser AMBEO VR 麦克风（由 Sennheiser 提供）和 RØDE NT-SF1（由 RØDE 麦克风提供）

从一开始大家就发现，用全向麦克风和 3 个 8 字型麦克风进行录音很麻烦，因为要考虑如何把各个麦克风的收音头尽可能靠在一起。不过，用一个带有 4 个心型指向收音头的**四面体麦克风**（如图 13 所示）就可以达到同样的效果。

通过这 4 个收音头的方位安排 — 左前上 (LFU)、右前下 (RFD)、左后下 (LBD) 和右后上 (RBU) — 我们可以得到 4 个信号，然后将其转换为常规的全方位通道和 3 个 8 字型通道。这个四面体麦克风（LFU、RFD、LBD、RBU）的收音头直接拾取的信号被命名为**“A-格式”**，传统信号（全指向通道 + 3 个 8 字型通道）被命名为**“B-格式”**。

不仅如此，通过一些数学运算，我们可以将 A-格式转换为 B-格式。例如，如果我们混合所有收音头拾取的信号，得到的信号就相当于全指向通道。如果用两个后向收音头的信号减去两个前向收音头的信号，得到的就是相当于前-后通道发出的信号，以此类推。

- W = LFU + RFD + LBD + RBU
- X = LFU + RFD - LBD - RBU
- Y = LFU + LBD – RFD - RBU
- Z = LFU + RBU – RFD - LBD

虽然各个收音头彼此接近，但这样短的距离还是会对某些频率产生一些影响，因为信号不会同时到达所有收音头。因此，在A-格式到B-格式的转换时，一些系统/插件可以进行更复杂的的信号处理，去补偿这些微小影响。

再次强调，虽然我们以 A-格式拾音，但应该尽快将其转换为 B-格式，因为 B-格式是Ambisonics的标准格式。

使用AMBISONICS麦克风时，也不要忘记立体声录音的经验

使用Ambisonics麦克风是一种非常有趣的录音方法，相比其他单声道或立体声麦克风，这种麦克风能够录制更多的空间信息。**但必须记住 2 个关键要点。**

首先，有时人们过于相信这种麦克风的效果。假设一部电影作品将以立体声形式发布，你不会天真到只需要在摄像机上放置一个立体声麦克风就够了。3D 声音制作也是同样的道理 — **这并非单纯只是在 360 度摄像机上添加一个Ambisonics麦克风那么简单。**你还是需要将麦克风放在靠近对白的地方（即使它们只是单声道），还是需要在混音过程中添加更好的音效，也需要进行后期制作等工作。

其次，那些能录制出色的管弦乐而获得格莱美奖的工程师，不会只放置一个 XY 立体声麦克风。通常他们会将几个麦克风放置在不同的位置，**将时间也作为空间工具使用。**再次强调，在沉浸式录音过程中，Ambisonic麦克风的效果更好（肯定优于 XY 和 M/S 立体声）。但是如果您希望更进一步，可能就需要将多个麦克风间隔式放置，同样将时间作为空间工具使用。

高阶Ambisonics (HOA)

那么，我们能否仅使用 4 个音频通道就得到很准确的 3D 声像呢？**不见得。**虽然Ambisonics仅使用 4 个音频通道就可以达成重放 3D 音效环境的任务，但它的空间分辨率并不太高。这意味着每个声音在方向上有少许模糊。但**高阶Ambisonics (HOA)** 能够解决这个问题。

在一阶Ambisonics (FOA) 常规的 4 个音频通道的基础上，为提高空间分辨率，我们在高阶Ambisonics (HOA) 中增加了更多的音频通道：二阶Ambisonics (SOA)有共9个、三阶Ambisonics (TOA)有共16个、四阶Ambisonics有共25 个、五阶Ambisonics有共36 个、六阶Ambisonics有共49 个、等等。

通过增加更多的通道，我们丰富了声像的细节。新增加的通道将继续携带音频信号，但会呈现更特殊的三维指向性（见图 14），而不是传统的一阶 8 字型图像。数学上称之为球谐函数，因为就像声音可以在其频域谐波中分解，3D 声场也可以在球坐标系用球谐函数分解。

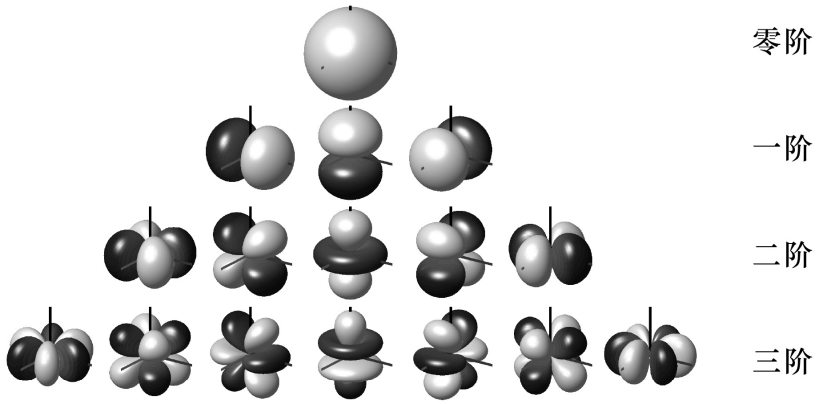

图 14 - 若干Ambisonics通道的极坐标图（由 Franz Zotter 提供）

图 15 - Eigenmike®麦克风，能够录制四阶 Ambisonics（由 MH Acoustics 提供）

虽然没有麦克风收音头可以直接具有这些特殊三维指向性，但可以使用带有大量收音头的特殊麦克风的信号中导出所需的信号，比如 Eigenmike®麦克风（见图 15）。

多个Ambisonics版本

当我们谈论Ambisonics时，很重要的一点需要考虑的是Ambisonics的“多版本”特点。否则就很容易犯错误，使用错误的格式或以错误的方式解析格式。为此，我们通过以下 4 个重要参数判断我们所谈论的Ambisonics版本：

- 阶数（一阶、二阶……）
- 格式（A、B……）
- 方位分量顺序（ACN、FuMa、SID）
- 归一化（SN3D、N3D、FuMa……）

我们首先需要确认使用的Ambisonics的**阶数**[5]是什么：是不是一阶？还是二阶或三阶等？这很容易识别，因为阶数决定了我们得到的音频通道的数量：

一阶	4 个音频通道
二阶	9 个音频通道
三阶	16 个音频通道
四阶	25 个音频通道
五阶	36 个音频通道
六阶	49 个音频通道
……	……
n阶	$(n+1)^2$个音频通道

我们需要确定的第二件事是：Ambisonics的**格式**是“B-格式”还是“A-格式”（其他格式都非常罕见）。直接从Ambisonics麦克风录制的音频属于“A-格式”(每个通道代表来自特定收音头的音频），但我们应该尽快将“A-格式”转换[6]为“B-格式”，因为“B-格式”是Ambisonics的通用格式，也是大多数软件能够识别并设计使用的格式。

我们需要注意的第三件事是**方位分量顺序**，即音频通道的顺序。这里的坏消息是：Ambisonics内部有很多不同的音频通道排序方式，但好消息是：常用的很少。最初，通道都是以习惯指定的方式排序（W、X、Y、Z），但随着高阶Ambisonics的出现，通道顺序的排列方法就显得尤为重要，要实现理想的伸缩性，用一致的方法适用于从通道数较少到有更多通道、场景更复杂的情况，因此就出现了 **ACN** 顺序。举例来说，在 ACN 中，一阶Ambisonics的通道顺序为 W、Y、X、Z；但在 Fu-Ma 中，顺序则为 W、

[5] 还有一个**混合阶数**的概念，但现在已经十分罕见，这种概念会提高水平面上的分辨率，减少高度层面的分辨率（甚至没有）。

[6] 您的Ambisonics麦克风应该包括一个能够将 A-格式转换为 B-格式的插件。

X、Y、Z。现代的Ambisonics平台使用 ACN 通道顺序，但使用旧版Ambisonics平台时要多加注意，因为旧版平台可能仍然使用之前的 **Fu-Ma** 顺序。还有一种 **SID**排序 方法，但目前几乎不存在了。如果您遇到Ambisonics方向错误的问题 — 应该从左侧播放的声音却跳至前方播放等 — 这很可能是通道顺序引起的问题。

第四个重要参数是**归一化**，即调整通道增益到目标水平。这其中同样也包含几种方法（**SN3D**、**N3D**、**maxN / Fu-Ma**[7]），但目前 **SN3D** 是最常见的方法，因为一个特殊原因：在 SN3D 中，第一个通道（全指向通道）始终是最高电平的通道，如果您对这个格式的信号进行混音或录制，只需要注意第一个通道是否削波（如果这个通道不削波，其他通道也不会削波）。其他的标准化方法（如 N3D）无法保证全指向通道为最高电平的通道，至少无法保证所有不同的声音方向都如此[8]。

在说到这些参数时，有人可能会提到“**AmbiX**”，这是Ambisonics的一种特殊音频文件格式，使用 B-格式、ACN、SN3D。所以当提到 AmbiX，可能并不是指这种文件格式，而是直接表示“B-格式、ACN、SN3D”这些参数。

所以要格外注意，确保您不会混淆不同版本的Ambisonics。

VR 应用

在Ambisonics中应用**旋转**方法很简单。假设您设置了一个 3D 场景，但是希望将这个场景旋转（例如向左旋转 60º或再增高 10º）。在Ambisonics中，旋转简化为应用**矩阵**，也就是说，逐一应用公式于所有现有的通道后得到新的旋转后的通道。

这个特点让Ambisonics能够完美适用于 VR 应用。为什么？当您戴上 VR 设备环顾四周，其实就是像是在旋转镜头角度：看侧面（**偏航角**）、上下看（**俯仰角**）或倾斜头部（**翻滚角**）。在这种情况下，VR 设备能够用Ambisonics的信号，根据头部旋转轻松进行信号旋转，获得与图像方向一致的声音。

---

[7] Furse-Malham 的简称

[8] 这里指，当Ambisonic microphone的方向变化或旋转处理后，不能保证所有通道都不会出现削波。

除此之外，Ambisonics还能够使用少量音频通道（4 个音频通道）重放 3D 音效，这适用于所有平台，尤其是网页和手机平台。因此在过去几年里，随着 VR 应用的复兴，公众对Ambisonics的兴趣也越来越高。

在旋转处理后，VR 应用能够直接将Ambisonics信号转换为**双耳音频**，用户能够使用耳机听到 3D 音效。我们将在下一章节中讨论更多关于双耳音频的内容。

优点与缺点

Ambisonics的最大优点是其能够适应所有扬声器布局的**灵活性**。第二，Ambisonics能够使用**少量音频通道**重放 3D 音频场景（可以从仅 4 个通道开始）。第三，Ambisonics可以**扩展使用**更多通道来提高空间分辨率。

Ambisonics的缺点是：**如果您希望达到高空间分辨率，就需要使用大量通道**，因为传统的“一阶”Ambisonics的空间分辨率不高。因此，您可能需要使用三阶（带有 16 个音频通道）或以上的Ambisonics，才能真正开始达到良好的空间分辨率。如果希望获得类似于基于对象音频的效果，甚至需要更高阶数的Ambisonics（如带有 49 个音频通道的六阶Ambisonics）。

双耳音频

说到沉浸式声音，大多数人首先想到的就是：房间中在前后左右上下摆满扬声器。然而，我们只需要一种特殊方法使用一个非常简单的设备就能够享受到沉浸式声音：耳机。那么，如何用耳机呈现高品质 3D 音效？

其实，我们可以体验到 3D 音效都是因为我们的两个耳朵！即使身处一个杜比全景声影院，周围 64 个扬声器带来美妙的声音体验，但最终，我的大脑只收听 2 个音频信号，来自我的左耳和右耳。因此，我们完全可以通过耳机重现全景 3D 音效，只要双耳可以继续听到和影院完全相同的声音信号，就会达到四周装满扬声器同样的效果，实现真正的沉浸式体验。

使用耳机重放的 3D 音效，被称为双耳（Binaural）音频，无需任何特殊类型的耳机。对于从事沉浸式音效工作的人而言，这是一个不可缺少的基本工具。

方向感

为了理解双耳音频的工作方式，我们首先要理解人类如何感知声音的方向。例如，假设您右侧的某处正在播放某种声音，如图中所示（图 16）。

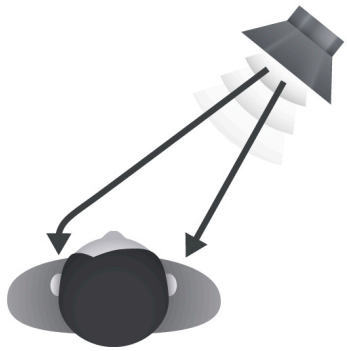

图 16 - 在右侧播放的声音

右耳听到的声音更大，左耳听到的声音更弱，这是一个基本的线索让大脑感知声音来自右方。在技术上，我们将这种线索称为耳间声强差 (ILD)。

还有，声音将首先到达右耳，随后才到达左耳，根据头部的大小和声音的方向，左耳可能比右耳有 1 毫秒内的延迟。这种双耳间的时间差被称为耳间时间差 (ITD)。

但这两种现象并不能解释一切问题。例如，我们如何探测前方、上方或后方声音之间的区别？在这 3 种情况下，双耳同时接收到声音，而且声音的大小一样。

为此，人体进化出“设计先进”的特殊构造，即耳朵。根据声音来源的方向，声音会在外耳（耳廓）的不同部位反弹，产生微小的反射，这种反射有助于大脑识别声音的方向。我们知道，将某个声音与其延迟的版本混合时，一些频率会增强，另一些频率会减弱。因此，我们的耳朵将作为一个小型均衡器，根据声音的方向改变其频率响应。来自肩部的反射也将补充外耳反射，再次轻微改变接收到的声音频率响应。

最后，头部对另一边的耳朵也会产生微弱的阴影效应。更高的频率由于波长较短，更难“倾斜”并进入阴影区域，这意味着离声源较远的耳朵收到的高频率声音也会轻微降低。

头相关传递函数 (HRTF)

正如我们所见，这些帮助大脑探测声音方向的因素，主要效应相当于增益、延迟和滤波。那么接下来的问题是：对指向人头部的一个特定声音方向，如何拾取（或表示）其整体效应？

我们知道，可以用脉冲响应表达由不同反射（更准确地说是经过滤波的反射[9]）在一个房间中形成的混响；我们同样也可以使用这个概念，表达特定声音方向在人耳的响应。这些脉冲响应又被称为头相关脉冲响应 (HRIR)，又名头相关传递函数 (HRTF)。HRIR 和 HRTF 几乎是一回事。唯一的区别是：HRIR 在时域（类似波形），HRTF 在频域（类似频谱）。

多年来，科学家使用以下 2 种方法收集了这些脉冲的集合，称之为 HRTF 数据集：

- 使用微型麦克风，将其放在真人的耳朵里，每只耳朵里有一个收音头；
- 使用带躯干或无躯干的仿真头，在仿真头上安装精确复制的假耳朵，放置麦克风收音头在假耳朵的内部（见图 17）；

图 17 - 双耳音频麦克风（Neumann KU 100 麦克风，由 Neumann 提供）

[9] 例如，混凝土反射声音的频率响应，与地毯反射声音的频率响应不同。

通过使用以上方法，科学家能够拾取到360° 环绕头部的一系列不同位置的左右脉冲响应。基于这些数据集，软件开发者可以创造出双耳音频插件和类似的软件（游戏音频引擎等），创造来自特定位置的声音幻像。

例如，对于直升机的声音（单声道），我希望听者感觉直升机的声音来自某个特定的方向，只需要使用 HRTF 数据集找到一个接近目标方向的记录位置即可（甚至可以在数据集位置之间用插值方法插入新位置）。然后利用这个位置的一对脉冲响应（一个左脉冲响应和一个右脉冲响应），将这一对脉冲响应用于直升机的声音，产生 2 种声音：左侧声音（直升机声音经过左脉冲响应）将被发送到左侧耳机，右侧声音（直升机声音经过右脉冲响应）将被发送到右侧耳机。顺便提一下，这种用脉冲响应处理音频信号的行为称为**卷积**。通过**卷积**，听众就会感知直升机来自于那个方向。

虽然大多数人在后期制作或重放过程中才会使用**双耳音频**技术，但也完全可以进行**双耳音频录音**。同样，将微型麦克风放在耳朵里[10]，或者在音乐厅的某个观众席上放置一个仿真头（如图 17 所示），就完全能够以双耳音频方式录制音频。然后，通过将这 2 个音频通道发送到耳机，您就可以听到身临其境般的 3D 音频场景。

HRTF 个人化

那么，这是否意味着我们每个人都可以通过耳机听到逼真的 3D 音效？很遗憾，这一点**并不对所有人适用**。

每个人的头部、耳朵和其他身体构造都是**独特的**，不尽相同，所以每个人耳朵对声音的影响方式都会略有不同。每个人的大脑都在多年成长的过程中，形成具有个人特征的聆听方式，不同于其他人。

这对于扬声器就并不是问题。声音从扬声器传出，虽然每个人的耳朵听到声音的方式略有不同，但大脑会认为没有不同（还是以其习惯的方式听取声音）。用耳机听其他人的

[10] 应尽量将这些麦克风拾音头放置在耳内，否则这些将无法捕捉到耳朵的真实响应。

HRTF存在的问题是，HRTF体现的是其他人听取声音的方式，和本人略有不同，会有些陌生。

如果听取从他人采集的 HRTF 数据集生成的双耳音频：一种可能，因为HRTF一致，我可以享受到完美 3D 音效的美妙体验；一种可能，我和原本采集 HRTF 数据的人之间存在的生理差异，导致我的大脑很难处理**冲突的HRTF信息**。

例如，对于相同的位置（中间正前方），图 18 显示了到达不同个体耳内声音的差异。如您所见，人与人之间存在巨大的差异，在某些情况下甚至可以超过 30 分贝！有些人的 HRTF 在 6 千赫兹频率时显著下降，其他人的 HRTF 则在 10 千赫兹以上频率时才下降。

如果您使用的 HRTF 与您耳朵的处理相似，大脑就会意识到这一点："我知道这种均衡！从那个方向传来的声音是这样的"一切都按原本的习惯进行。其他情况下，使用的 HRTF 在某种程度上类似于你的耳朵，但是声音来自于不同的方向："这种均衡听起来很熟悉……我认为声音来自那个方向"，大脑会给你发送错误信息。例如，应用来自正前方的 HRTF，但您的大脑却认为声音来自后面（这被称为前-后混淆）或不同的高度。有时候，HRTF 与大脑所感受到的信息大相径庭，导致大脑变得混淆："我没有探测到任何熟悉的均衡"，您的大脑不会给出任何特定的方向，这时就像普通立体声在耳机里播放一样，您会感觉声音来自头内部。

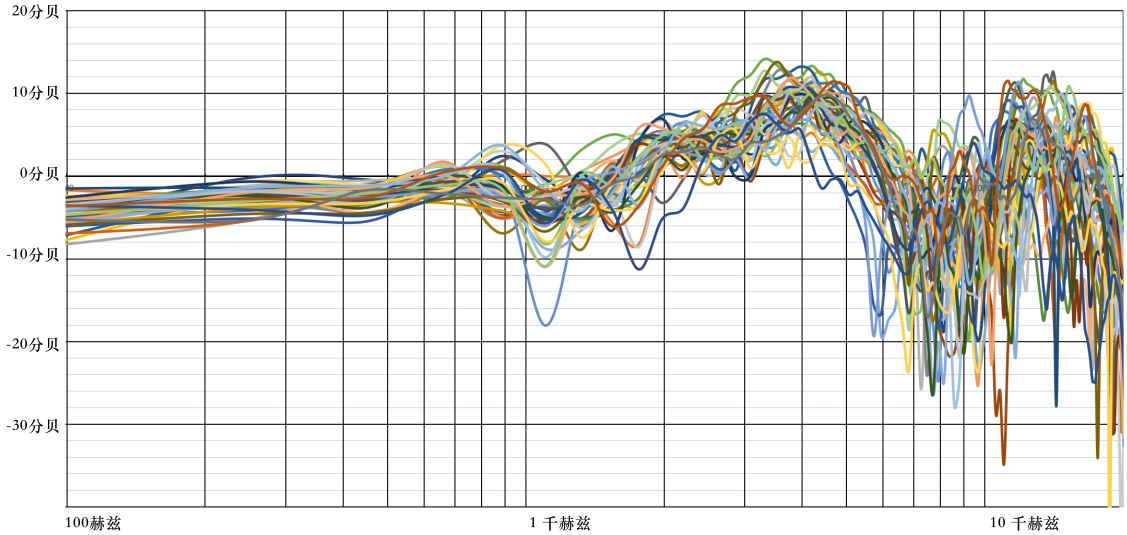

图 18 - 45 个人在完全相同位置下左耳的 HRTF：正前方（方位角 = 0，仰角 = 0）

所以，对特定的双耳音频声音，有些人会感觉非常真实的3D效果，另一些人感觉有一些真实的效果（比立体声更好），还有某些人完全不觉得有真实的效果。例如，当我听双耳音频时，大多数时候，所有前向的声音似乎都升高了，可能是因为我的耳朵与一般人略有不同。

许多公司都试图选择现有最佳的 HRTF 数据集，将其应用于其产品之中，这意味着能够满足最广泛听众的 HRTF 数据集。但是，即便使用极佳的 HRTF 数据集，也总有一部分听众无法充分享受到这个数据集的双耳音频。

所以我们需要某种 **HRTF 个人化方法**：或者向听众提供多个HRTF 数据集选项（选择来自与其身体特征相似的数据集）或使用**本人的 HRTF**。当然，只有少数人能在实验室中采集到自己的 HRTF，但一些公司也开始提供基于照片、视频甚至是人头部 3D 扫描的 HRTF 个人化服务。个人化 HRTF 数据集，使用一种被称为 **SOFA** 的文件格式。这种格式由 AES（音频工程学会）定义，可以让人们拥有自己的 HRTF 文件，在支持该格式的系统上使用。

使用双耳音频

再次强调，双耳音频仅在使用**耳机**重放时才可以正常播放，这一点至关重要。如果您使用立体声扬声器重放双耳音频，最终可能会得到类似立体声的声像，但无法达到 3D 体验或任何接近的效果。

第二个重点是，我们在上文中提到的所有不同方法（基于通道音频、基于对象音频以及Ambisonics）都可以**转换为双耳音频**。

例如，可以将 5.1 通道或 22.2 通道的信号转换为双耳音频，或将杜比全景声以双耳音频形式呈现，或将三阶Ambisonics信号转换为双耳音频。实际上，大多数 VR 应用都会使用这种从Ambisonics到双耳音频的转换。虽然我们可以通过扬声器体验 VR，但在大多数情况下，我们会将 VR 场景的Ambisonics音频信号转换为双耳音频，然后发送到 VR 头盔的耳机。

条形音响（SOUNDBARS）

为了使用耳机重放 3D 音效，双耳音频应运而生。但很多研究都试图通过扬声器重建双耳音频的 3D 音效。有些条形音响应用了这些概念来开发，让人们通过前置条形音响体验 3D 声音。

大多数条形音响都能够重放比普通立体声更具有沉浸感的声音，但比起完整的扬声器布局效果要差，尤其是没有坐在最佳聆听位的听者而言，更是如此。尽管如此，如果您想要体验 3D 音效却无法安装完整3D布局的扬声器组合，条形音响也是一个不错的选择。

优点与缺点

双耳音频的最大优点是：能够通过**非常简单的设备**重放 3D 音效 -只使用耳机 - 无需很多的扬声器组合；让您可以随时随地通过物美价廉的设备体验 3D 音效。

双耳音频最大的缺点是：它对许多人来说并没有那么有效，由于缺乏个人化，**会导致前/后混淆、声音升高**（尤其是前方声音）以及**“来自头内部”**的头中效应。其次，双耳音频不支持**扬声器重放**。

但正如之前所说，只要人们理解其在双耳音频方面的个人化需求，双耳音频就能够有效发挥作用。

总结

我们已经了解了有关 3D 音效的 4 种不同方式：基于**通道**音频、基于**对象**音频、Ambisonics（又称基于场景音频）以及**双耳音频**。每种方式都有利有弊，没有孰优孰劣之分，这意味着您需要谨慎进行选择。

如果重放使用和混音拥有一个固定的扬声器布局，最适合采用基于**通道**音频，但它不能灵活的转变适用于多种布局。基于**对象**音频可通过灵活布局提供最佳空间分辨率，但需要单独发布每个音频流，如果声源很多，可能就会造成问题。Ambisonics能够使用少量输出通道重现不错的3D 音效，但如果您希望达到更高的空间分辨率，就需要使用大量通道。**双耳音频**是一种适用于耳机重放音频的方法，但双耳音频不适用于扬声器，并且存在需要个人化的问题。但不要忘记还有**混合**方法，即设法将 2 个方法的好处相结合。例如，杜比全景声同时使用了beds（基于通道）和对象（基于对象），大多数 VR 应用同时使用Ambisonics和双耳音频。

相信到此您已经更好地理解了 3D 音效，包括 3D 音效如何运作以及我们如何使用 3D 音效，现有哪些类型的系统以及具体的用途。此外，您也了解了各种 3D 音频应用的领域，从电影业到电视，从电子游戏到音乐，从 VR 到其他多种应用，3D 音频几乎随处可见。如果您喜欢带有空间感的音频效果，现在的选择可谓多种多样……Sound Particles 软件支持以上大多数格式，更是会让您的制作锦上添花。

Sound Particles 是一款终极 3D 音频制作软件，多个好莱坞制片厂都曾使用该软件制作票房大片，如《权力的游戏》、《星球大战 9》和《冰雪奇缘 2》

通过虚拟麦克风，您可以呈现所有格式的音频，从单声道到 22.2 通道，从六阶 Ambisonics到任何自定义布局。

Sound Particles 是好莱坞创造史诗般声音效果的秘密武器。Sound Particles 的一大独特功能，是能够使用粒子系统在您周围产生数千种声音。

想创造一个“战场”？直接创建一个包含 10000 个粒子的粒子群（声源），从您的声音库导入 200 个战争相关的音效，添加随机运动和随机处理，然后使用虚拟麦克风捕捉结果 — 只需几分钟即可完成全过程。

立即尝试无时限演示

教师、学生及学校可免费使用 Sound Particles

Sound Particles SA
IDDNET, Rua da Carvalha, 570, 2400-441, Leiria, Portugal
www.soundparticles.com • info@soundparticles.com • +351 244 859 465