

北京理工大学

本科生毕业设计(论文)

基于物理模型的碰撞与滚动声音实时模拟

Physically-based Simulation for Real-time Collision and Rolling  
Sound Synthesis

学	院：	计算机学院
专	业：	计算机科学与技术
学 生 姓 名：		王铎暉
学	号：	1120151912
指 导 教 师：		刘庆晖
校外指导教师：		刘天添

2019 年 5 月 23 日

## 原创性声明

特此申明。

## 关于使用授权的声明

本人签名:                      日 期:            年    月    日

指导老师签名:                日 期:            年    月    日

## 基于物理模型的碰撞和滚动声音实时模拟

### 摘要

在基于视觉的仿真与模拟研究被广泛应用于动画制作，虚拟环境搭建，人像建模，危险实验预模拟等领域的大环境下，研究针对其他感官系统（如听觉，触觉，嗅觉等）的仿真与模拟成为利用计算能力探索搭建更真实虚拟世界的新趋势。得益于现今游戏引擎（如 Unity3D, Unreal, Blender 等）和虚拟现实设备（如 Oculus Rift, HTC Vive 等）的普及，本工作提出了一种高效的基于物理仿真来生成针对刚性物体碰撞声音的方法，实时，且能够应用于虚拟现实环境中。我们通过利用弹簧质点阻尼模型来提取刚性物体的震动信息（利用正弦波表示），并合理抓取如碰撞力，碰撞点的物理信息来进行声音合成。

然后，本工作基于根据已记录声音进行实时碰撞声音模拟的方法，在虚拟现实环境中进行了一个经典心理学实验，主要是利用了 3D 声音能够提供给听者位置信息的功能，来测试声音对人认知物体间运动因果关系的影响：当一个运动的物体碰撞一个静止物体时，观察者会立刻认知到相应的因果事件，那就是运动的物体造成了这个静止物体向前运动。但是如果这个碰撞事件同时伴随着碰撞声音，和存在第二个物体运动的时间延迟，观察者对整个因果事件的判断就会受到影响。因为在真实世界动态改变声音各种参数和设置时间延迟比较困难，所以本工作利用了沉浸式的虚拟环境来进行探究，最终结果证明了碰撞声音发生位置的偏离也将削弱人对物体间运动因果关系的认知。该实验关于此因果关系感知的结论将从理论层面帮助我们在下一步工作中为基于物理的声音模拟提供更可靠的，针对大型复杂模拟环境的加速方案。

**关键词：** 声音合成，虚拟仿真测试，刚体模拟

## Physically-based Simulation for Real-time Collision and Rolling Sound Synthesis

### Abstract

Under the over-all situation that lots of research in the areas of simulation based on visual effects is massively utilized for animation production, virtual environments setup, human modeling, pre-simulated hazard environments and so on, research into simulation based on other sense organs (e.g., audio, haptic, olfactory feelings) has gradually become a new trend to build a more immersive virtual world by exploiting computing resources than only focus on visual parts. Recently, game engine such as Unity3D, Unreal Engine, and Blender combining consumer-grade virtual reality devices like Oculus and HTC Vive, which provide users high-quality immersive feelings in virtual environments(VEs), are increasingly popular. We take advantage of these in our work and present an efficient method to produce impact sound between rigid-body objects based on physical simulation and these sounds are synthesized in real time, which can also be put into VEs. By utilizing mass-spring-damping system to extract vibrations (sinusoid waves) on rigid-body objects, we can synthesis sounds integrating more physical information such as impact point and force.

Then, we conduct a classical psychological experiment in virtual environments based on the method of pre-recorded sounds. We mainly take advantage of 3D sounds that can offer position information to listeners to test influence of sound that affects humans' causal perception on launching events, which means that when a moving object collide with stationary one, an observer will immediately perceive a causal event that a moving object cause the stationary one to move forward. However, when the second object's motion is delayed, or is accompanied by a collision sound, such causal impressions attenuate and strengthen. We research into this experiments wholly in virtual environments because it is really difficult to dynamically setup time delay of objects' motions and parameters of sounds. The results of this experiment also provides us with theoretical support to accelerate the speed of synthesizing sound in real-time, complex environments.

**Keywords: Sound Synthesis, Virtually Simulated Test, Rigid-body Simulation**

## 目 录

第 1 章 引言 .....	1
1.1 声音实时模拟 .....	1
1.2 心理认知实验 .....	1
第 2 章 相关工作 .....	4
2.1 声音合成 .....	4
2.2 感知研究与测试环境模拟 .....	5
2.21 因果感知 .....	5
2.22 虚拟现实 .....	6
第 3 章 技术方法 .....	7
3.1 基本质点振动模型 .....	7
3.2 三维物体振动模型 .....	8
3.3 有限元分析 .....	9
3.4 声音输出 .....	11
3.5 声音合成加速处理 .....	11
3.51 模式压缩 (Mode Compression) .....	11
3.52 模式截断 (Mode Truncation) .....	12
3.6 滚动声音模拟 .....	13
第 4 章 测试及实验 .....	15
4.1 实验一： 2D 虚拟环境下物体运动的因果感知 .....	19
4.2 实验二： 3D 虚拟环境下物体运动的因果感知 .....	22
4.3 实验三： 3D 虚拟环境下碰撞声音位置被改变 .....	25
第 5 章 结果与讨论 .....	30
5.1 场景中实时声音合成 .....	30
5.2 心理学实验结果讨论 .....	31
结论 .....	33
参考文献 .....	34
致谢 .....	38

## 第1章 引言

### 1.1 声音实时模拟

当前游戏引擎下，大多数的交互引用都是利用的已经记录好的声音来给场景中模拟的物体交互提供声音的反馈(详见于[3])。它主要利用了这样的声音实时性很好且产生声音的过程非常快，使得能够比较好的应用于大型模拟场景中，但是这样的声音很难考虑到物体交互间比较复杂的物理性质，仅仅能够改变一下声音总体的大小和音高等。以碰撞盘子状物体发出的声音为例，盘子声音的大小和音色在敲击盘子边缘和中心时肯定是不同的。所以，基于模拟与仿真对声音进行建模的方法被提出且应用于生成合理的，考虑了较多物理因素(如碰撞力的大小/方向)的声音。本工作建立了对应刚体的弹簧质点模型并结合有限元建模<sup>[2, 4, 5]</sup>的方法来抓住刚体表面比较小规模的震动从而自然生成刚体之间的碰撞声音。本工作基于一个重要理论：即对于弹簧质点模型的波动方程而言，可以用一系列的正弦波形（无论谐波与否）的叠加（傅里叶合成）来近似表示合理的刚体碰撞声音<sup>[1]</sup>。我们利用从非线性系数二阶常微分方程组(Ordinary Differential Equation - ODE)转换为线性系数二阶常微分方程组的方法推导出了上述声音建模对应方程的解析解<sup>[4, 5, 6]</sup>，从而得到组成物体声音对应的一系列频率和大小。同时，利用第二部分的工作，我们提出了基于听者认知方面的加速方案，通过牺牲一定的声音质量，使得能够更快速地进行声音合成。本节工作主要贡献如下：

- 1) 该声音合成方法可以用于实时交互性的模拟场景，提供更加真实的碰撞声音以及连续型的滚动声音。
- 2) 本声音生成方法基于弹簧质点系统，利用其产生的正弦波形模拟碰撞声音。并将相应非线性常微分方程组线性化使得更加利于实现。
- 3) 利用了四面体来作为物体网格信息的基本元进行声音震动模型建模，使得弹簧质点系统建立更加精确。

### 1.2 心理认知实验

本节工作的目的主要是建立从人体感知层面对上述声音合成模拟工作提供加速方案

的理论支持，因此简化了碰撞声音的场景，限制在小球碰撞场景中。本节利用声音合成方案是对已记录的声音片段进行二次处理来适配碰撞的场景变化，二次处理包括声音频率，大小，音效的添加等。现考虑以下的视觉场景：一个 2D 的红球沿直线向一个 2D 蓝球移动直到这两个小球的边缘相互接触。在这两个小球接触后，红球停下运动，蓝球向前以直线运动远离红球。尽管在这样的场景中没有直接的信号提示表明这两个小球之间的运动存在某种因果关系，但观察者仍然会很可能认为是红球造成了蓝球向前运动(详见于[7])。这就是一个关于运动的因果认知的例子，这种立刻地，自动产生且不可磨灭地因果认知来源于低层次地认知输入<sup>[8]</sup>。它区别于高层次的认知干扰，这种低层次的认知描述了我们对于真实世界的理解是来源于对逻辑规则和自身认知知识的运用，而不是高层次知识比如物理定律，相关背景知识或者经验等[9]，这也是因果认知对印象建立进行阐述的重要特点<sup>[8,9,10]</sup>。因果认知对于动态时间的时空性质是非常敏感的<sup>[8,11]</sup>。比如，在上述的视频播放进行实验的条件下：(1)如果在红色小球停下和蓝色小球运动之间有一个暂时的延迟；(2)如果两个小球的边缘在碰撞的时刻是分开或者重叠；(3)又或者如果蓝色的小球开始运动的方向和之前红色小球运动的方向不一致，这种针对于红色小球造成蓝色小球向前运动的认知就会减少(详见[12,13,14])。这样的触发事件也能够被周围形状的运动所影响<sup>[14]</sup>，这表明人体认知系统能够快速处理时空信息和视觉内容信息来从感知输入中立刻建立因果印象<sup>[9,15]</sup>。

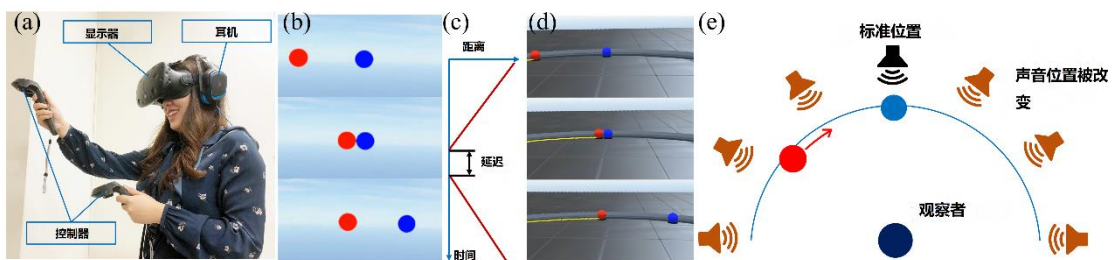


图 1-1 (a)：在 2D((b)) 和 3D((d)) 测试环境下虚拟现实相关设备的使用。

(b)：2D 环境下小球碰撞过程(直线轨迹运动)

(c)：在 2D 和 3D 测试环境下红球向碰撞点移动，停下，在一定延迟后，蓝球向前移动。

(d)：3D 环境下小球碰撞过程(圆形轨迹运动)

(e)：两球碰撞会发出碰撞声音，同时这个声音的位置可以被改变。

迄今为止,研究人员已经大量的利用了 2D 的利用视频播放的方式来检验因果感知事件<sup>[8, 12, 13, 14]</sup>。这主要取决于在真实世界中改变运动物体时空性质的困难性。具体而言,想要暂停真实世界中的碰撞事件且没有外界工具的介入是非常困难的:比如使用磁铁或者数字遥控器来来在透明轨道上以既定速度移动金属物体。然而,虚拟现实技术提供了方法在模拟的 3D 环境中来处理这样的难题。而且虚拟现实可以在 3D 环境中轻易改变声音的位置。前述工作已经表明当一个碰撞事件如果伴随一个表明碰撞物体之间关联的声音信息(如碰撞声音),观察者会留下出比无声碰撞事件更强的因果印象(即认为红色物体造成蓝色物体向前运动的可能性更大了)<sup>[16]</sup>。然后对于人体感知系统是否能对形成上述因果印象时,同时对声音位置信息进行正确解析仍然是未知的。比如腹语的情形,人们会认为声音是从表演者手里的娃娃发出的[17]。因此本工作想要解决如下问题:

1) 上述在 2D 环境中进行的经典小球碰撞实验能否延展到虚拟 3D 环境,得到相同的结论或更多新的发现。

2) 碰撞声音的位置信息能否影响对小球碰撞因果关系的认知。

因此本工作设计了三个对应的实验来解决上述问题。在每一个实验中,一个一开始运动的红色小球与一个一开始静止的蓝色小球碰撞,在 0-400ms 的延迟过后,蓝色小球开始向前运动(见图 1-1(b,d))。在实验一中,参与者能够在 2D 的环境下建立起该事件的因果联系,印象,无论是有没有碰撞声音的发出。第一个实验是 Guski 和 Troje [16]之前工作的再现,是被设计来确保以前他们工作所发现的结论可以被延申到虚拟环境(2D)中。实验 2,参与者完成同样的观察任务但是在 3D 虚拟现实环境中进行。实验二中的碰撞声音的位置仍然是在碰撞点的标准位置(ground-truth)上。实验二的目的是确保以前工作的结论能够延展到 3D 的虚拟现实环境中,实验三与实验二实验条件相同,除了改变了声音的位置信息(见图 1-1(e))。总的来说,本节工作具如下贡献:

1) 在虚拟现实环境中复刻了之前因果印象实验的工作证明了虚拟技术可以用作对检测人认知和心理具体价值

2) 检验了声音位置被改变的碰撞声音对人感知这样的碰撞事件因果关系的影响。

3) 测验了人在虚拟环境中对声音位置的估计情况。



## 第 2 章 相关工作

### 2.1 声音合成

对于利用计算机进行声音合成的方法比较广：如今范围用的最广就是利用采样的方法，将真实环境中采样的声音放入仿真环境中（广泛见于电影声音制作中 - Foley[18]）。该方法充分利用采样，实时性较高，且能够对收集的声音进行再处理（如改变频率，大小，添加环境音效等）。但对于大型复杂环境而言，当场景发生物体太多且复杂，模拟环境中需要存储的声音也会比较多，不仅使得采样工作耗时耗力，且高质量的采样声音存储占用内存也比较高<sup>[6]</sup>。因此，Cook, D. Brandon 和 Lloyd 等提出了在基于傅里叶合成基础上利用分析声音频谱，提取合理正弦波进行更底层的声音合成的方法来应用于交互环境中[1, 19]。该方法利用了频谱建模的方法，认为声音由正弦波和噪音组成，且正弦波的叠加能够表示出所合成声音绝大多数的性质。该方法快速且与声音设计师的开发思路相接近，问题在于这是一个基于数据收集的方法，它对声音采样的数量要求也不低。

为了合成声音更具有真实性：能够根据敲击位置的不同有不同的反馈，O'Brien 等人提出了从输入物体网格信息和材质信息来提取振动模式（振动模式被认为以正弦波的形式存在）从而生成声音的方法[21]，成功建立起声音物理性质和模拟之间的联系，使得声音的模拟能够细化到每一种材质，每一个物体。为了实现更快的声音合成速度，Raghuvanshi 和 Ming Lin 等提出了一系列的加速技术来实现可以用于实时场景，同时利用弹簧质点模型对物体表面建模来近似振动的情形，可以得到更易实现的解析解[6]。另一种加速手段，由 James, Dingzeyu Li 和 Changxi Zheng 等提出（见[4, 5]）：利用从频率域对振动模式进行处理，再利用傅里叶逆变换/拉普拉斯逆变换回到时间域。本工作同样利用弹簧质点模型对振动进行建模处理，但使用了更精确的方法-有限元分析<sup>[21, 22]</sup>来计算弹簧质点模型的相应参数，如胡克常数和质点质量等。利用 Sifakis 等提出的对刚体进行有限元分析离散化建模的方法[2]，结合物体四面体网格信息，我们可以建立对刚体表面振动更准确的模型来合成更自然的声音。对于实时场景而言，James 和 Ming Lin 等都提出需要将弹簧振动的非线性系数二阶常微分方程组转换为线性系数二阶常微分方程组，从而使得每个质点的每一个自由度对应的振动模式能够独立地被计算（详见于[4, 5, 6]）。

最后需要将声音合成系统放入模拟场景中，需要合适的物理引擎来处理对物体作用的实时碰撞力，本工作利用 Unity 游戏引擎<sup>[20]</sup>进行处理并使用其进行简单的场景测试。

## 2.2 感知研究与测试环境模拟

对于因果关系感知的联系是人体对动态世界进行感知的基本。它广泛地存在于日常生活中，比如 6 到 10 个月的婴儿就能够对视觉场景中的物体运动因果关系有认知<sup>[23, 24]</sup>。正如 Michotte 的经典工作[11]，物体间的碰撞或者说发射事件被认为是非常有价值的系统来检测因果联系的认知（详见于[8, 12, 13, 14, 16, 25]）。因此，这样的因果关系推断可以在认知层面上进行建模，2.21 和 2.22 节叙述了最近利用虚拟现实技术进行相应研究的方法和例子。

### 2.21 因果感知

近来，许多研究者已经提出了因果关系可以通过推断在视觉场景中观察到的物体之间联系的边缘概率来得到对应的感知建立，就是所谓的噪音牛顿框架 (Noisy Newton)<sup>[13, 26]</sup>。因为即使具有基本的物理知识，人对简单的可感知的变量的估计（如：距离，速度和时间等）是本能不确定的，对于他们的准确估计必须根据预估和提前观察。一些简单的运动例子：1. 人是会认为物体是缓慢运动的，或者越动越慢（而不是按照运动学的一直运动）。2. 对于一个碰撞事件而言，人会认为一个静止的物体在被撞后应该立即运动，这样这个碰撞才被认为是发生了（详见于[13]）。因此当通过视觉在观察这样的碰撞事件时，对于速度，物体位置的估计，时间延迟等都是基于人自己之前对整体场景有一定的预估和观察。

噪音牛顿框架更进一步的认为人体是有自己在神经回路中建立一套属于自己的物理模型<sup>[27]</sup>来对真实的物理规则进行近似和估计，从而生成一套随着时间变化的感知。所以对于基于感知和物理建模的碰撞事件，物体应该按照动量守恒定律进行运动，而对于没有感知与物理的建模，物体将随机运动。所以两个物体在碰撞前后的速度，位置，以及运动的时间延迟等都被用作重要因素来对这两个物体之间的因果联系进行判断。这一套理论提供了人体对运动事件因果感知的量化估计，它随时间和空间的变化而变化。比如对于上述碰撞事件而言，当两个小球之间的运动延迟和碰撞时两球距离拉大时，对于他们之间存在碰撞的因果联系的可能性估计就会降低<sup>[13]</sup>。

尽管噪音牛顿框架已经从视觉层面揭示了人类对物理世界的感知情况<sup>[26, 59]</sup>。但对于同时包含视觉和听觉信息的情况还需要进行建模及测试。对于听觉信息如声音位置信息等能否在建立上述因果联系时同样被人体感知仍然是不清晰的，因此本工作目的就是做出对物理世界进行感知层面建模复杂度与深度分析的第一步。

## 2.22 虚拟现实

虚拟现实技术已经展现了作为低成本高效的手段，可充分应用于模拟极端环境来测试和训练人类。比如，虚拟现实结合游戏系统被广泛应用来训练人进行在灾难逃生<sup>[29, 39, 31, 32]</sup>，医疗急救<sup>[33, 34]</sup>，灭火训练<sup>[35, 36, 37]</sup>，航空安全<sup>[38]</sup>，交通和火灾<sup>[39, 40, 41]</sup>等。这些针对训练设计的系统都非常有用，因为他们利用模拟的方法将本很危险的环境放置在了安全可控的环境中。

然而，以前的虚拟现实设备受限于计算能力和较差的声音与视频质量，造成上述的研究只能集中于简单的重复性过程训练。传统的虚拟现实系统也受限于需要大型测试环境和大量的动作捕捉设备，使得利用虚拟现实技术进行心理学方面的研究不够顺利。因为参与者需要经过大量训练来适应复杂的虚拟现实设备。因此，以前的心理学实验都仅仅基于视频播放和问卷调查等，虚拟现实技术仅仅是一种想要被用于进行更好模拟实验的设想<sup>[42, 43, 44]</sup>。

近来由于虚拟现实工业的大发展，许多消费级的虚拟设备比如 Vive, Oculus 等简化了虚拟环境的搭建，同时 GPU 计算能力的大量提升使得听觉和视觉体验得到极大提升。在过去的两年时间里，虚拟现实已经被广泛应用于实验性研究(模拟复杂环境)。比如地震的模拟<sup>[45]</sup>，视觉导航<sup>[46]</sup>，语义分析<sup>[47]</sup>，和机器人控制<sup>[48, 49]</sup>。各种虚拟现实平台被用于人机交互场景设计<sup>[50]</sup>以及自动驾驶模拟<sup>[51]</sup>。在认知研究方面，研究人员已经利用虚拟现实设备来研究人类的欺骗行为<sup>[52]</sup>，重力感知<sup>[25]</sup>，视觉认知调节能力<sup>[53]</sup>，触觉重定向能力<sup>[54]</sup>，以及位置感知能力<sup>[55]</sup>。本工作目的是建立起在 3D 虚拟环境中完整的心理学实验流程，证明虚拟现实技术在进行复杂心理学实验的可行性。

### 第 3 章 技术方法

本方法中，声音在刚性物体受外界力冲击下刚体物体的表面弹性振动产生的。本方法将这些振动都近似为正弦波的简谐振动形式<sup>[6]</sup>。这些振动能够搅乱物体周围空气从而进行波形的传递从而发出声音。如果声音的频率在 20-2200Hz 之间，将可以被耳听到。当给予这些振动合理的衰减模式(本方法采用理想的指数型衰减进行近似)，结合傅里叶叠加原理<sup>[1]</sup>，将随时间衰减的正弦波进行叠加，将得到与物体本身网格形状和材质适配的声音(如金属球声，砖块声等)。如今，游戏引擎如 Unity, Unreal 等已经能够提供高度成熟的刚体碰撞模拟器，因此我们将碰撞力的提取最终实现在了游戏引擎中，从而实现了整个声音系统的可视化。

对于物体表面振动波形的提取，本方法利用物体网格信息建立对应的弹簧质点模型，利用基本的弹簧振动来近似物体表面因碰撞产生的振动波形。因此，对于一个输入的由顶点和边组成的物体网格，我们将质点代替顶点，带阻尼的弹簧代替边从而建立起基于物体形状的弹簧质点系统<sup>[5,6]</sup>。下面，我将从单个质点的振动模型开始进行整个声音建模的方法解析。

#### 3.1 基本质点振动模型

对于每一个质点而言，它的运动是一个基本的弹簧阻尼模型(见图 3-1)，这里认为质点  $x$  只有一个自由度(即在一个方向运动)，但实际上三维物体中每一个质点应有三个自由度。因此一个具有  $n$  个质点的物体，对应振动模型的质点运动应用一个  $3n*1$  的向量进行描述。

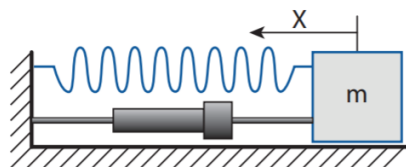


图 3-1：弹簧阻尼模型(一维)<sup>[5]</sup>

由于弹簧位置固定，图 3-1 所示的弹簧阻尼模型只考虑产生波形对与时间的变化情况。从而可以利用下述二阶常微分方程，获得质点简谐运动时合理的正弦波：

$$m\ddot{x} + c\dot{x} + kx = f(t) \quad (3-1)$$

$x$  描述质点位置， $m$  为物体质量， $k$  表示弹簧的刚度(stiffness)，也是这里的胡克常数； $f(t)$ 描述了跟随时间变化的外力； $\ddot{x}$ 则表示了  $x$  的二次导数，也就是加速度； $\dot{x}$ 则表示了  $x$  的一次导数，也就是速度； $c$  表示衰减模型，本工作中都需要设置合理的衰减系数，使得能够产生振动且最后停止(即欠阻尼情形)，那么  $c$  应该小于临界阻尼  $c_c=2\sqrt{km}$ 。为了得到式(3-1)的解析解，我们利用拉普拉斯逆变换，并认为  $f(t)$ 为狄拉克函数形式的单位冲击，得到  $x$  的解析解。

$$x(t) = \frac{1}{mw} e^{-\xi w_d t} \sin(w_d t) \quad (3-2)$$

$$w = \sqrt{\frac{k}{m}} \quad (3-3)$$

$$\xi = c/2mw \quad (3-4)$$

$$w_d = w\sqrt{1 - \xi^2} \quad (3-5)$$

$w$  是没有衰减的振动频率， $\xi$ 是衰减因子，因此 $w_d$ 是有阻尼下，衰减的振动频率。从公式中也可以得到，衰减频率必须在 $\xi < 1$ 的情况下才有意义，同时由这里也可以证明  $c < c_c$ 。利用傅里叶叠加的原理<sup>[1]</sup>，对式(3-2)中得到的每一个质点对应的正弦波进行叠加，就得到了物体产生的声音，从中我们可以看出：我们在这里是将力以狄拉克函数(Dirac Delta Function)形式，作为单位冲击来表示的。因此对于多个不同强度的任意外力，我们需要将他们进行叠加从而得到声音的强度信息，这也是在声音合成中唯一需要实时计算的步骤，利用的是如下时间积分。

$$x(t) = \int_0^t f(\gamma)/(mw_d) e^{-\xi w(t-\gamma)} \sin w_d(t - \gamma) d\gamma \quad (3-6)$$

### 3.2 三维物体振动模型

对于一个三维物体(含  $n$  个质点)而言，利用矩阵思维将式(3-1)进行拓展， $x(t) \rightarrow x_i(t), i = 1 \dots n$ ，得到如下式子：

$$M\ddot{u} + D\dot{u} + Ku = f(t) \quad (3-7)$$

其中 $u$ 表示质点相对平衡位置的偏离。那么对于由  $n$  个质点组成的物体， $u = (u_1, u_2, u_3 \dots \dots u_n)$ ，设平衡位置为 $X_i, i = 1 \dots n$ ，那么 $u_i(t) = x_i(t) - X_i$ 。 $M$ 为质点质量组成的质量矩阵(Mass Matrix,  $M$ 是一个对角矩阵)， $K$ 为胡克常数组成的刚度矩阵(Stiff Matrix,  $K$ 是一个对称矩阵)， $D$ 为 Rayleigh 衰减模型： $D = \alpha M + \beta K$ ， $f(t)$ 为外

力。很明显，这个方程是一个非线性系数二阶常微分方程组，要求解来得到质点位移 $\mathbf{u}$ 是非常复杂的。因此我们想要将其转换为线性系数常微分方程组，从而得到与式(3-1)所示的每一个质点偏离的解析解[2, 3]。核心思想是利用对 $\mathbf{K}$ 和 $\mathbf{M}$ 进行广义特征值分解：

$$\mathbf{KU} = \mathbf{MU}\mathbf{S} \quad (3-8)$$

通过式(3-8)的广义特征值分解得到了由特征向量组成的震动模型形状描述的矩阵 $\mathbf{U}$ 和由特征值组成的对角矩阵 $\mathbf{S}$ 。 $\mathbf{U}$ 决定了每一个质点得到的正弦波对最终合成声音的贡献程度， $\mathbf{S}$ 的对角元素即是 $\mathbf{w}^2$ 。根据特征值分解的性质， $\mathbf{U}$ 是满秩矩阵。所以可以用线性合成的方式表示式(3-7)中的质点偏离 $\mathbf{u}$ ， $\mathbf{u} = \mathbf{U}\mathbf{q}$ 。然后把它带入式(3-7)中，可以得到式(3-9)：

$$\mathbf{U}^T\mathbf{MU}\ddot{\mathbf{q}} + \mathbf{U}^T\mathbf{DU}\dot{\mathbf{q}} + \mathbf{U}^T\mathbf{KU}\mathbf{q} = \mathbf{U}^T\mathbf{f}(t) \quad (3-9)$$

再根据特征值分解的性质： $\mathbf{U}^T\mathbf{MU} = \mathbf{I}$ 和 $\mathbf{U}^T\mathbf{KU} = \mathbf{S}$ ，同时利用 Rayleigh 衰减模型 $\mathbf{U}^T\mathbf{DU} = \alpha\mathbf{I} + \beta\mathbf{S}$ ，那么式(3-5)的非线性方程形式最终可变为如下线性常微分方程形式：

$$\ddot{\mathbf{q}} + (\alpha\mathbf{I} + \beta\mathbf{S})\dot{\mathbf{q}} + \mathbf{S}\mathbf{q} = \mathbf{U}^T\mathbf{f}(t) \quad (3-10)$$

所以到式(3-10)为止我们得到了关于质点位移的线性常微分方程组，而且其中 $\mathbf{q}$ 向量中的每一个元素 $q_i$ 都是相互独立的，所以可以解出式(3-2)得到的每一个质点的位置偏离方程，且通过 $\mathbf{S}$ 矩阵，我们在预处理阶段便得到每一个正弦波的频率。需要合成声音的部分只剩下通过碰撞力得到每一个正弦波对应的振幅，从而合成声音最终的大小。

对于上述线性方程组的代码层次求解实现而言，本工作利用了 Eigen C++库<sup>[65]</sup>进行相应的矩阵层面的计算操作，包括广义特征值分解，Rayleigh 衰减模型，3.51 节的模式压缩等。

### 3.3 有限元分析

通过输入的物体网格模型得到的质量矩阵和刚性矩阵决定了极大程度上决定了最终声音合成的质量。因此本方法将利用有限元分析的离散化方法来获得更精确的质量矩阵和刚性矩阵。基于式(3-5)，将连续物体的动态振动方程进行离散化处理，对物体四面体网格(见图 3-2)进行逐一处理。首先利用 Tetgen[57]，基于 C++的物体网格转换器来将以三角网格组成的物体网格转换为由四面体网格组成的物体网格。

我们从一个四面体入手来阐述质量矩阵和刚体矩阵的建立情形：

1) 质量矩阵：对于一个四面体而言，其质量矩阵大小为  $12 \times 12$ ，由四个  $3 \times 3$  的对角矩阵分块组成，设这四个对角矩阵为  $B_{ij}, i, j = 1 \dots 4$ ，任意一个  $B_{ij} = 1/20 * \rho V (1 + \delta_{ij}) I_{3 \times 3}$  ( $\delta_{ij}$  为克罗内克符号函数，即只有  $i$  等于  $j$  时，值为 1，其余情况为 0， $\rho$  是物体密度， $V$  是四面体体积)。当建立好这一个四面体对应的质量矩阵时，将其元素按照对应行列加入  $3N \times 3N$  的质量矩阵即可。

2) 刚体矩阵：刚体矩阵的建立较为复杂，但建立流程与质量矩阵类似。在建立好单个四面体的  $12 \times 12$  的刚体矩阵后，再将对应元素按照对应行列加入  $3N \times 3N$  的刚体矩阵即可。总的刚体矩阵建立基于式 (3-11)，其中  $t$  是所有基于 FEM 离散化的质点的总受力情况， $u$  仍是质点偏离。对于单个四面体的刚体矩阵建立，需要从形变入手：

$$t = Ku \quad (3-11)$$

$$F = [x_2 - x_1 \ x_3 - x_1 \ x_4 - x_1] [X_2 - X_1 \ X_3 - X_1 \ X_4 - X_1]^{-1} \quad (3-12)$$

$$E = \frac{1}{2} (F + F^T) \quad (3-13)$$

$$t_j = \sum_{i \in \tau_j} K_i u \quad (3-14)$$

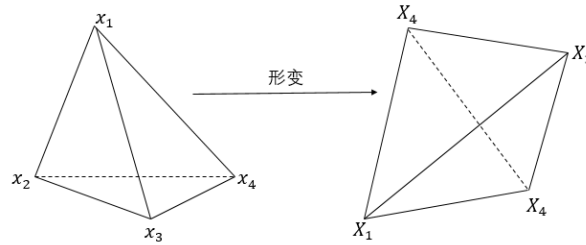


图 3-2：单个四面体网格及形变

按照 3.2 节参数设定，由于变形前后  $x_i$  与  $X_i$  的位置是可以确定的，那么式 (3-12) 中的变形梯度将是常数矩阵 ( $F$ )，说明衡量微小形变的张力张量矩阵也是一个常数矩阵。由于刚体表面的振动产生的声音再多数情形下是比较微小的，所以我们利用线性张力张量  $E$  来近似表示张力张量矩阵 (见式 (3-13))。然后利用一阶 Piolar-Kirchhoff 压力张量来表示因为形变产生的弹力，最后按照四面体结构进行线性叠加 (见式 (3-14)， $t_j, j = 1..4$ ，为四面体上质点所受内力， $K_i$  为四面体的刚体张量， $\tau_j$  为与质点  $j$  相关联的四面体)，将得到  $12 \times 12$  大小的单个刚体矩阵。

### 3.4 声音输出

当得到物体所有质点的振动正弦波(数量为  $3N$ )后,提取碰撞力信息来得到波形相应的振幅,然后利用傅里叶叠加的原理,将声音所有波形根据式(3-2)的衰减模型,依次叠加/混合在一起即可以得到的最终的声音。本工作利用 OpenAL 库对所有波形以 44100 的采样速率对所有波形采样 3 秒。需要注意的是,因为人对频率的感知范围是 20-22000Hz,所以我们在这里也只混合了处在人听觉范围内的提取到的波形。同时,本系统进行声音合成时只生成了单声道的声音(即左右耳声音一致),对于后续 3D 音效,混响音效等高级声音处理直接交给了 Unity 声音游戏引擎部分。

### 3.5 声音合成加速处理

本节工作利用了 Ming Lin 等人提出的声音合成速度提升的部分方法[6],对单个物体的声音合成进行了加速处理以达到实时的要求:

#### 3.5.1 模式压缩 (Mode Compression)

对多个振动模式(波形)进行压缩的原理基于 Zwicker 和 Fastl 等关于声音频率认知的工作[56]。他们发现两个非常相近频率的波形无论是同时发出还是邻近时间发出,听者都是不能够区分出他们的。他们将这种无法分辨的频率之间的距离称为差异阈值(Difference Limens to Change)。图 3-3 描绘了人的听觉系统对于这种相近频率的区分能力,该图表明了人的听觉系统相对于中心频率能分清其邻近频率声音的能力。需要注意,人体的听觉范围是 2-22KHz,从图中可以看出在 2KHz 时,已分不清 1Hz 的区别。本工作利用图三所示的线性近似的曲线对已经利用式(3-8)的特征值分解后得到的矩阵  $U$  和波形频率进行相应处理来实现模式压缩,即用具有中心频率的波形等效其相近的无法区分频率的波形。从图三也可以看出,利用上述近似的方法使得需要混合的模式的总数量小于 1000 个,从而与物体本身的网格结构独立开来。这是因为我们认为在 20 到 22000Hz 听觉范围下,人类最多能分清差距 1Hz 的不同频率,那么进行压缩后最坏的情况就是需要混合 1000 个正弦波,而不是优化前的  $3N$ 。

首先我们对矩阵  $U(3N \times 3N)$  所代表的物理性质进行具体分析:该矩阵每一行与物体顶点对应,每一列与不同的振动模式对应(也对应  $S$  矩阵的对角元素  $w^2$ )。所以对于一行而言列出了不同振动模式的振幅(根据式(3-6), $U$  与振幅线性相关),因此如果敲击了某一个质点,那么某一行上所有的振动模式就应该叠加



在一起从而发出声音。模式压缩的操作就是将  $U$  矩阵的列合并处理，混合更少的模式来产生声音从而实现加速。具体实现流程如下：

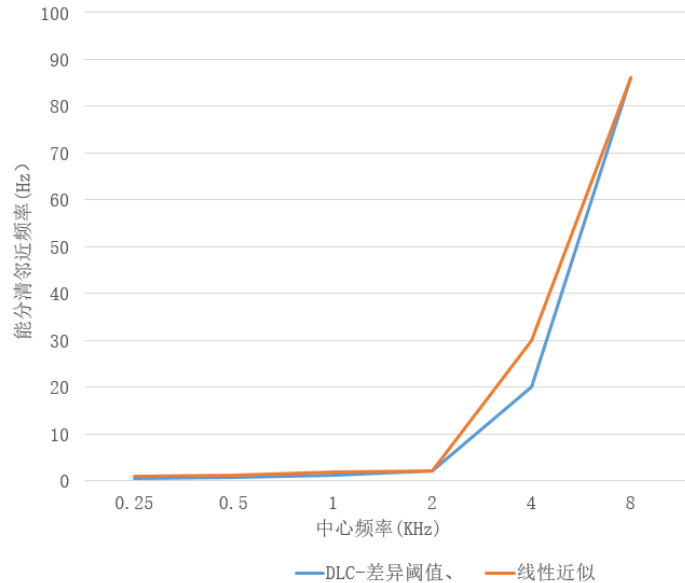


图 3-3: 对 DLC-差异阈值的线性近似

1) 将  $U$  矩阵每一列按照对应频率进行增序排列，但由于  $U$  矩阵由特征值分解得到，同时其列已经按照特征值 (角频率的平方  $\omega^2$ ) 由小到大的顺序进行排列，所以这一步其实可以省略。

2) 遍历频率，根据图三差异阈值的线性分段近似对  $U$  矩阵进行处理，比较当前列对应模式的频率和下一列对应模式的频率，如果他们是无法分辨的，那么用他们列元素之和代替。

从上述处理与分析中，可以明显看出模式压缩是一项预处理的加速策略，不需要放在实时输出声音进行，它充分利用人体的听觉感知能力，通过减少需要混合的模式数量实现性能优化。这样的优化不仅体现在实时声音输出时，声音采样相关的浮点数操作数量 (基于式 (3-6)) 的减少从而加速声音合成，同时也体现在对于一个物体对应声音信息的存储优化，因为需要混合的振动模式往往仅几百个，那么将极大减少  $U$  矩阵规模，使得一个物体声音信息的内存要求减少很多，如对于有五千顶点的物体，它的内存需求从 225MB 下降到少于 15MB (认为一个浮点数存储占据空间为 1 Byte)。

### 3.52 模式截断 (Mode Truncation)

模式截断的思路是当振动模式的振幅小于一定值时，停止对其进行采样，因此这项加

速策略时放在实时声音输出部分进行处理。因为碰撞非常短暂，所以对于振幅较大的波形短时间的采样是能够抓住声音音质本质的，至少从感知层面上来说是几乎没有区别的<sup>[4,6]</sup>。那么所谓三秒的采样对于一些衰减较快的波形将不是必须的。那么可以设定合理的阈值，将对最终声音贡献较小的波形提前去除，从而减少实时声音输出所需的浮点数操作数量，从而实现加速。本工作的实现是 Ming Lin 等工作的简化[6]，我们利用他们工作结论，设置阈值  $1e-5$ ，那么当波形振幅衰减到  $1e-5$  以下时，将不对这个波形进行采样，即不再将此波形混入最终的声音。

### 3.6 滚动声音模拟

本工作对滚动声音的模拟同样进行了尝试，参考了 Ming Lin 等的碰撞声音模拟的延展工作[6]和当今游戏引擎利用已记录声音[20]的两种实现思路。当使用基于物理的方法实现了上述的碰撞声音模拟时，滚动声音的模拟也已能够按同样方法实现，唯一的区别就是滚动声音的碰撞力获取更加连续，但当今游戏引擎的碰撞模拟器已经非常成熟模拟连续作用的力的分析。由于想要将滚动声音用于后续心理学实验中的小球运动，本工作也尝试了使用已经记录的持续时间非常短的滚动声音，通过周期循环播放的方式，根据式(4-1)和图 4-1 的方法控制声音音高和音量的方法来模拟实时的滚动声音。但为了尽可能的控制变量，减少实验中影响观察者对碰撞事件因果关系的判断因素，在最终的测试还是去掉了滚动声音。

图 3-4 展现了对小球实现滚动声音实时模拟的例子。直接利用之前碰撞声音模拟的方法是可以得到滚动声音的，但由于小球表面网格信息比较复杂，即使有 Unity 物理引擎实时提供外界力的信息，直接从网格信息来模拟滚动信息的计算仍较为繁复。但小球是一个完全对称的物体，这里采取的方法结合了上述两种方法，我们首先利用物理模拟的方法采样了当作用较小碰撞力时的碰撞声音，由于球类物体完全对称，因此这个碰撞声音可以通过周期循环播放的方式交给 Unity 游戏引擎声音系统进行实时调整形成滚动声音(如：速度不同相应声音的音高不同等)。本工作测试了在小球不同运动速度（5m/s, 8m/s, 10m/s）下的声音模拟效果(包括滚动和碰撞声音)。

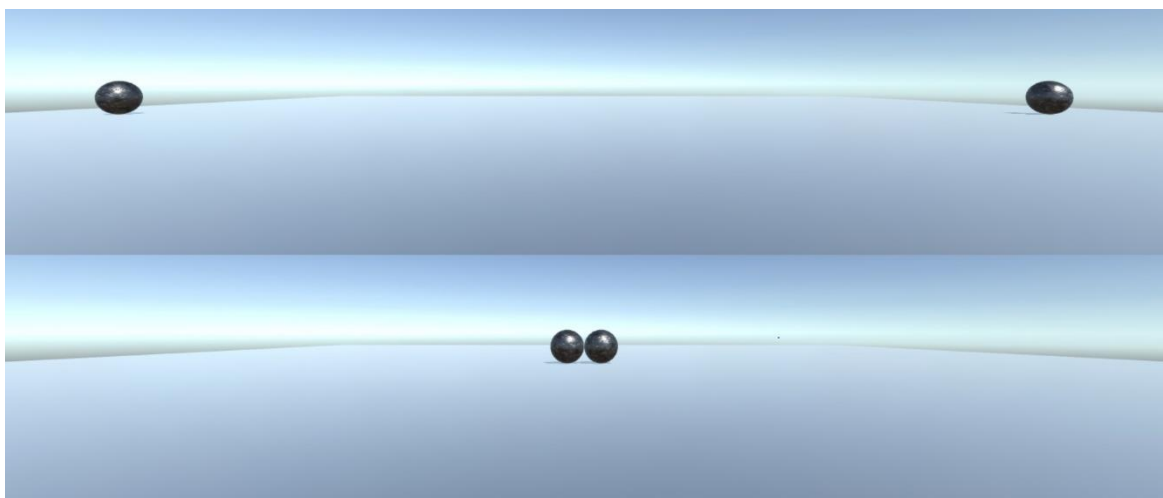


图 3-4：两个质量相等小球从静止两端匀速相向运动然后碰撞(带滚动声音)。

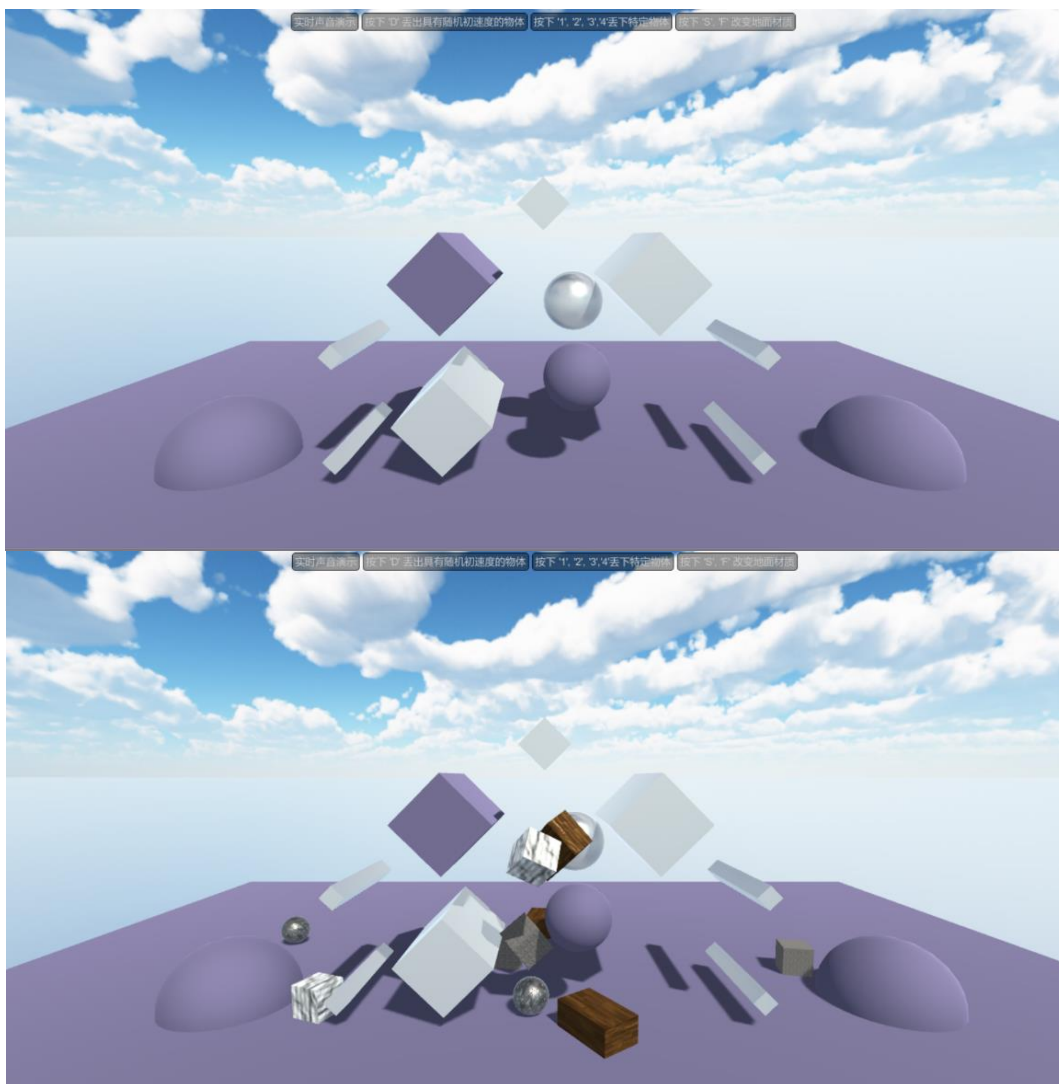


图 3-5：简单形状物体(立方体/球类)碰撞声音模拟，利用 Unity 引擎进行场景设计

## 第 4 章 测试及实验

本节将主要介绍心理学实验的实验设计，测试流程和实验结果。本实验所采用的实时声音合成方法基于利用已记录的声音，进行相应声音音高和音量的实时调整，并包括环境音效，3D 音效等添加，具体见实验设计部分。本实验没有采用基于物理的声音合成方法是因为这里的动态事件非常简单：仅仅是两个小球的碰撞。他们是形状均匀的物体，且为无动量损失，无摩擦的匀速运动和完全弹性碰撞。同时也不需要考虑他们的材质信息，因此在这里我们就利用了已记录的声音，用最普通的碰撞声音来描述整个碰撞事件。本实验设计没有引入滚动声音是为减少对被试者因果感知的影响因素，在总的三个实验设计中，我们核心是需要说明物体间运动的时间延迟和声音位置这两个变量对整个动态事件因果感知的影响情况，因此尽可能去除掉其他变量，对其他变量的考量可以查看第 5 节的结果讨论部分。

**参与人员(被试者)：**总共 36 名参与人员(男性 19 名；女性 17 名)被分配到了当前三个不同实验中。其中，十名参与人员(男性 7 名；女性 3 名)被分配参与实验一，十名参与者(男性 5 名；女性 5 名)被分配参与实验二，十六名参与人员(7 名男性；9 名女性)被分配参与实验三。所有的参与人员都是本科生或者研究生。所有的参与人员的视力与听力都是正常的，他们的平均年龄为 22.5 岁(标准差=2.3)，每一个参与者都是随机被分配到三个实验中的一个。

**测试声明：**当前进行的实验流程是符合心理学实验标准且被第三方机构审核通过的，也不用于任何商业用途。本实验的参与风险非常小且参与人员都是自愿参与的，且参与人员可以在实验的任何阶段中止测试。所有的实验都在参与人员的同意下进行，且没有收集任何参与人员的身份信息。

**设备与规程：**当前实验的虚拟环境设计和模拟都是在 Unity 游戏引擎中完成，使用的设备是 HTC Vive(见图 1-1(a))。在每一个实验中，被试者都会佩戴 Vive 头盔显示设备(HMD)来提供虚拟环境中的视觉输入。这个头盔由两块显示屏组成(一只眼睛一个)，每一块显示屏提供 1080x1200 的分辨率，相应刷新率为 90Hz。参与人员也会佩戴一对由挂在墙上的 Vive 跟踪器实时追踪的 HTC 控制器，使得他们参与者可以根据虚拟环境中的指导将所有的感知结果直接在虚拟环境中进行输出，利用射线指向对应的选择框进行选择即可。听觉的输入由一对 Logitech G430 游戏耳机提供，

它可以提供 7.1 声道的环绕声音音效输出。这个耳机可以提供稳定准确的声音位置信息，这对实验三是非常重要的。

为了去除掉其他的场景因素的影响，每个实验都是在一个安静的密闭测试房间进行。在整个实验中，参与人员坐在可以旋转 360 度的旋转椅上。在进行实验之前，参与人员首先接收指导如何带上显示器和耳机，以及如何使用控制器在虚拟环境中的具体操作。参与人员可以在实验中随时告知实验人员任何不适或者眩晕，且可以在任何时候中止这个实验。每一个实验的相应指导是通过虚拟环境中的 UI 窗口提供的。参与人员首先读这些指导要求然后利用控制器进行点击交互等来表示他们可以进入下一步。这样的实验设计是为了防止实验人员朗读知道要求造成的理解偏差。但是同时，如果参与人员有任何对实验指导要求的疑问都可在这期间向实验人员进行提问。

**实验总纲：**实验一和实验二都是被设计来复刻 Guski 和 Troje 等的主要结论（碰撞声音在上述小球碰撞事件中能够提升对因果关系的感知[16]。）是可以被延展到 2D 和 3D 的虚拟现实环境中的。对以前的实验复刻的主要目的是：1) 确保该心理学实验设计的结果是不会被使用 VR 的交互系统而影响的。2) 是为了确定以前关于声音和视觉的因果关系感知的结论可以延申到实时的 3D 虚拟环境中去。前两个实验的基本流程都是和 Guski 和 Troje 等工作类似的：参与人员首先观察红色物体（实验一是 2D 红色圆盘，实验二是 3D 红色小球）朝向静止的蓝色小球运动直到两个物体边缘相接。一旦这样的接触发生，红色物体就停下，在一段时间延迟过后，蓝色物体开始向前运动。每个实验中伴随相应碰撞声音的蓝色物体运动的时间延迟的范围为 0-400ms。可以参考图 1-1 (c) 中对于时间延迟的解释加深对实验理解。在每次观察上述的碰撞事件后，参与人员将对红色物体造成了蓝色物体向前运动这一事件的可能性程度进行输出/汇报（评级，分值表示程度）。在实验一的 2D 虚拟环境中，两个小球是沿着直线路径匀速运动的。在实验二的 3D 虚拟环境中，两个小球是沿着圆形轨迹匀速运动的。在实验二中，碰撞声音的位置是在标准位置，即碰撞点的位置。实验三与实验二除了声音的位置沿着圆形轨迹变化以外（图 1-1 (e)），其他实验参数设置都是一样的。由于是圆形轨迹，声音的位置在变化，但声音的音量并没有随着位置变化而变化。而如果是直线轨迹的运动，随着声音位置变化，如果与被试者的距离增加了，那么声音音量将会变小（图 4-1，实验一使用了线性模型来模拟 2D 碰撞声音；在实验二和实验三使用了对数模型来模拟 3D 虚拟环境中对应于声音位置和运动的声音。）。

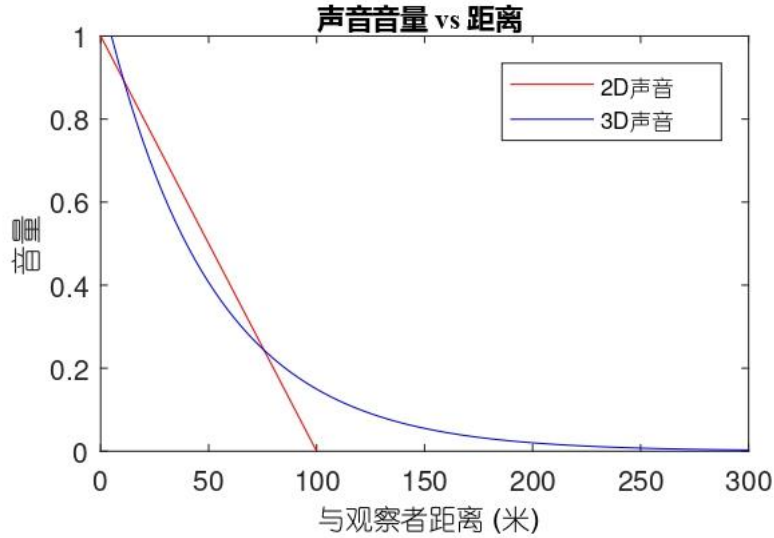


图 4-1 合成声音音量和声音与被试者的距离之间的关系。

如果在 3D 虚拟环境中仍让小球们沿着直线运动，改变声音的位置将同时影响碰撞声音音量。为了剔除掉多个变量的同时影响，我们设置了小球在 3D 虚拟环境沿着圆形轨迹运动。因此实验二也是一个实验三的对照实验，如果没有实验二，那么对于实验三中被试者对因果关系感知的评级是受碰撞声音位置的影响，还是物体运动轨迹改变的影响都将是不确认的。

**实验设计：**为了阻止实验之间延迟效应造成的影响，我们应用的是间实验设计 (between subjects design)，即每一个被试者只做了一个实验(实验一/二/三)。对于实验一和实验二，参与人员不休息地完成了以随机顺序排列的 36 个小测试。时间延迟的设置 0 到 400ms，每隔 50ms 设置一个，一共九个时间延迟设置情况，相同的设置会出现两次。因此构成了 18 个小测试情形，再加上碰撞声音有无的变量设置，总共生成上述 36 个小测试。实验一和实验二分别会花费大约三十分钟来完成。在参与人员在每一个小测试中观察了物体碰撞事件后，他们会被要求回答下列问题：是红色物体造成了蓝色物体的运动吗？并在虚拟环境中，利用数值滑杆进行评级，评级范围从”绝对不是”到”绝对是”（见图 4-2）。其中”绝对是”的情况对应于被试者有着红球造成了蓝球向前运动的强烈印象，“绝对不是”的情况对应于被试者有着红球一定没有影响蓝球运动的强烈印象。对于滑杆上基于文字的评级将在后续根据之前工作相同的实验设计<sup>[16]</sup>转换为对应的数字分数，具体公式如下：

$$\frac{s-1}{9} \leq \frac{p-p_{\min}}{p_{\max}-p_{\min}} \leq \frac{s}{9} \quad (4-1)$$

其中 $p$ 是从数值滑杆中读取的被试者评级情况， $p_{\max}$ 和 $p_{\min}$ 是对应的滑杆最大值和最小值。 $s$ 是被转换后的数字分数(范围 1-9)，度量单位长度为 1。在这里，分数为 1 对应于“绝对不是”的因果关系感知，分数为 9 对应于“绝对是”的因果关系感知。

在实验三中，同样设置了九种不同的时间延迟，但同时碰撞声音的位置也是被改变了的(见图 1-1(e)，声音分布从 $-90^{\circ}$ 到 $90^{\circ}$ 沿圆周轨迹分布，单位度量间隔为 $30^{\circ}$ ，因此包括没有声音情况一起共有八个声音条件设置条件)。上述测试会被重复两次，实验三总共有 144( $8*9*2$ )个小测试以随机顺序呈现，且分为两部分完成，即被试者可以在实验三中途休息，且可以在实验指导阶段和实验一/二一样询问相关的流程问题。在完成上述 144 个测试后，参与人员在实验三中还需要在数值化杆上完成对声音位置估计的实验。在这一部分中，所有的测试都是有碰撞声音的，所以在这一部分实验最终将会有 64 个测试，他们以随机顺序呈现给被试者。实验三这一部分的测试和实验一/二都是不同的，设计这一部分实验不是为了测验碰撞声音对因果关系感知的影响，而是为了检测：

- 1) 被试者对碰撞声音位置的估计准确与否。
- 2) 是否对碰撞声音注意力资源的不同分配(由于声音位置不同导致)能够真的影响对动态碰撞事件的因果关系感知。

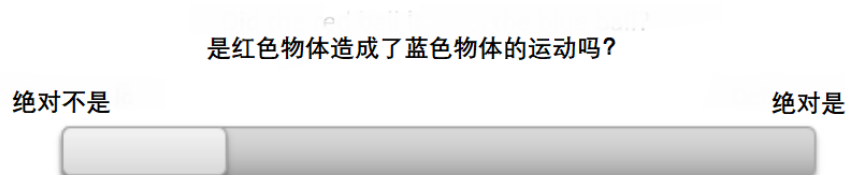


图 4-2: 对物体碰撞事件因果关系评级机制

每一个参与人员完成实验三的全部内容大概需要两小时。在 2D 和 3D 虚拟环境中物体大小和速度的设置也是与以前的经典实验设计所匹配的，2D 与 3D 虚拟环境中，每一个物体大小对应的视角(visual angle)为 $10.1^{\circ}$ 和 $10.7^{\circ}$ 。在所有的实验中，物体以 $13.2^{\circ}/\text{sec}$ 的角速度运动。碰撞的设计是完全弹性碰撞，即碰撞后的蓝球运动速度与碰撞前的红色物体运动速度相同，两小球质量相同。在每一个实验中，每一个物体的大小和材质都是一样的，只有物体颜色不同(蓝/红)。在实验一中，整个碰撞事



件的呈现持续约 10 秒，在实验二和实验三中，每一个碰撞事件的呈现持续约 15 秒。

#### 4.1 实验一： 2D 虚拟环境下物体运动的因果感知

在实验一中，无论场景中存在碰撞声音，参与者被要求在 2D 虚拟环境中观察了动态碰撞事件后进行相应的因果关系评分。进行实验一的目的是将在虚拟环境中的因果关系感知结果和以前工作中使用 LCD 投影<sup>[16]</sup>得到的结果进行比较，来验证是否一致，从而验证使用虚拟现实技术的可行性。在 2D 空间中，每个物体的运动方式都是直线运动，环境的搭建中没有任何阴影加载视觉渲染中（见图 1-1 (b)）。整个实验一的设计和实现方法与 Guski 和 Troje 等人的工作一致。唯一的比较小的不同是上述所说的呈现动态碰撞事件使用的硬件不同。同时我们也根据以前的研究结果设计了相同的问题给参与人员，在每次观察上述碰撞后，参与人员需要回答类似“是物体 A 造成了物体 B 运动”的问题。

**2D 声音合成：**整个声音合成的方法利用的是已记录声音，并结合 Unity 原生声音引擎进行复杂的音效设置。为了复刻原心理学实验设计中的碰撞声音<sup>[16]</sup>，声音的扩散音效在虚拟环境中被关掉了。同时，空间混合音效被设置为 2D 模式，意味着声音是单声道的，左右耳没有不同，混响效果也被关闭了。线性模型被使用来决定声音音量和声音与被试者距离之间的关系（见图 4-1）。上述效果合并在一起，整个 2D 声音的效果将为所有声道都是相同的音量，且没有额外的衰减情况，这样的声音最终呈现效果与之前工作的碰撞声音呈现相同。

当前工作中，我们使用的已记录的声音是保龄球碰撞的声音，在 2D 虚拟环境中声音持续时间大约为 10ms，它的音高 $P$ （或者说频率）与速度的关系由下列公式决定：

$$P = \frac{s_t}{s_0} + b \quad (4-2)$$

$s_t$  表示物体运动速度， $s_0$  时物体参考速度（如：小球最大线性速度）， $b$  是声音音高的 offset。碰撞声音音高按照式 (4-2) 的设置可以使得低速运动的物体有更低频率的碰撞声音，由于实验中小球运动速度 $s_t$ 为常数，那么此时碰撞声音音高保持为常数。



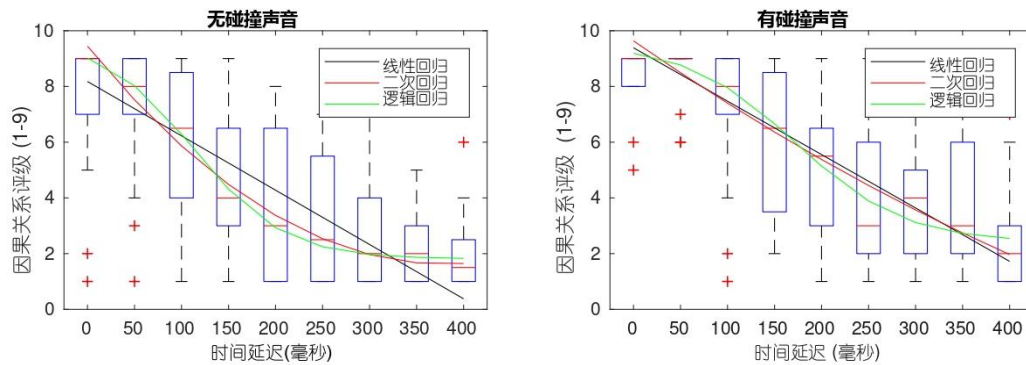


图 4-3: 对实验一的因果关系评级结果:左)无碰撞声音条件, 右)有碰撞声音条件

**训练阶段:** 参与人员通过在虚拟环境中读一系列实验指导要求后开始实验。首先, 实验人员提供给参与人员一系列对于动态碰撞事件的静态描述, 和对实验任务的简单描述。一旦参与人员完成了知道要求的阅读, 他们就可以开始一系列训练测试。根据 Guski 和 Troje 工作中的实验流程设计, 参与人员被告知了在训练测试中有一些比较差和比较好的碰撞事件(参与者能肯定两物体发生碰撞的即是比较好的碰撞事件), 也有一些情况一般的碰撞事件。训练环节是被用于帮助参与人员建立起个人对于上述事件因果感知的基本评级机制, 比如什么样的情形可以被参与人员个人感知为发生碰撞, 那些又是不可行的。在训练环节中没有告知参与人员是否一定有碰撞声音的出现。

共有六个训练小测试, 由三个不同时间延迟(0ms, 200ms, 400ms)和碰撞声音有无组合在一起。这些小测试以随机顺序呈现, 且标准的场景设置(比如没有时间延迟和碰撞声音的出现)是没有明确告知参与人员的。也没有任何反馈告知参与人员他们完成了每一个小的训练测试。只有当所有训练小测试都结束了, 他们才可以进入真正的实验阶段。可以看出, 本节工作是在尽可能不在实验人员的指导和干预下, 让参与人员自己观察训练测试中的场景, 自行地对上述动态碰撞事件产生初步的, 属于自己的因果关系感知。

**实验阶段:** 在完成了上述训练阶段后, 参与人员将要完成以随即顺序呈现的 36 个真正的测试。相应的激励性参数设置(如, 时间延迟和碰撞声音的出现与否)是与 Guski 和 Troje 之前的心理学实验设计相同的[16]。参与人员将会回答与训练阶段相同的问题, 然后给出相应的因果关系评级。评级分数按照式(15)转换为 1-9 的数字分数, 然后被 VR 系统后台记录成文本。参与人员在实验阶段将不会得到任何反馈。

**实验结果分析：** 实验一得出的结论与以前经典实验的结论一致，实验一中关于因果关系评级的中值在有无碰撞声音的条件下都随着时间延迟的增减而下降（见图 4-3，红色水平线表示因果关系评级结果的中值，蓝色柱状的顶部与底部相应表示评级结果在 75%和 25%的情况。虚线的延展到了最极端的被认为是异常值的数据，而红色的‘+’号标出了对应的异常值。图中横轴表示了碰撞和蓝色圆盘开始运动之间的时间延迟。黑线，红线和绿线分别对应的是线性回归，二次回归和逻辑回归对上述数据的拟合情况。）。因为被试者的评级分数在有无碰撞声音的条件下都倾向于更低的值，所以对于这个分数的分布遵从正态分布是不对的。因此，非参数统计被采用来分析评级分数的分布情况。本工作采用了 Friedman 假设检验来分析测出的评级分数，结果表明了有碰撞声音的巨大影响： $\chi^2(1) = 16.9, p < 0.001$ ，即碰撞声音有无使得实验结果有显著差异，因为卡方值( $\chi^2(k-1), k=2$ )大于了查表值，拒绝无差异的假设。这个结果是与以前通过该经典实验发现的基于视觉和听觉的因果关系感知匹配的，即碰撞声音加强了被试者对于此动态碰撞事件的因果关系感知。

利用线性，二次和对数回归模型基于评级分数的中值和时间延迟的关系进行描绘。可以从图 4-3 看出尽管统计得到的评级分数有一些偏差，但基于其中值的预测的误差是满足回归分析中假设的正态性分布的。对于每一个回归模型，我们计算了相关系数和贝叶斯信息准则(BIC，值越小越表明模型拟合更好)，他们都受到模型复杂度，如模型的自由参数的制约。在没有声音的条件下，评级分数显示出较强拟合二次回归( $r^2 = 0.98, BIC = -12.77$ )和逻辑回归( $r^2 = 0.995, BIC = -25.03$ )的趋势，线性回归的拟合较弱( $r^2 = 0.69, BIC = -0.04$ )。相反的是，在有碰撞声音的条件下，评级分数显示出较强拟合线性回归( $r^2 = 0.94, BIC = -6.00$ )的程度，二次回归( $r^2 = 0.92, BIC = -4.36$ )和逻辑回归( $r^2 = 0.96, BIC = -3.95$ )拟合程度也是不错的。尽管二次和逻辑回归模型在有碰撞声音条件下的拟合程度也很强，但整个变化曲线还是更趋近于线性回归模型，因为线性回归模型显示出最小的 BIC 值，我们可以得出在有碰撞声音条件下线性回归是最拟合的结论。这个结论也得到了来自逻辑回归模型斜率数据的支持：有碰撞声音( $b = 3.01$ )VS 无碰撞声音( $b = 3.60$ )。当前实验结果所得结论是与 Guski 和 Troje 等以前的工作是符合的，即因果关系评级分数在有碰撞声音条件下更拟合于线性回归模型，无碰撞声音更拟合于二次回归模型。

尽管该实验结果的总情况是与 Guski 和 Troje 的研究结论是一样的，但仍有一些细微的偏差。比如我们发现在时间延迟为 400ms 时，评级分数在有无碰撞声音的条

件下几乎没有明显区别(见图 4-3)，但 Guski 和 Troje 他们发现在有碰撞声音条件下评级分数大约是没有声音条件下评级分数的两倍(见[16];图 3, 794 页)。可能的原因是本工作进行的实验在进行实验指导时给参与人员提供关于因果关系进行判断的问题有所不同,我们给予的问题是“是红球造成了蓝球的运动吗? ”,而 Guski 和 Troje 给予的问题是“蓝球向前运动有多大的可能性是之前动态感知事件造成的?”这两个提问方式有巨大的不同:Guski 他们将关于动态碰撞事件究竟是声音还是红球的运动留给了参与人员自己进行理解,而本实验具体地提示了红球却没有强调声音有无的条件。由于 Guski 他们做实验时给被试者强调了声音存在的条件,很有可能导致了在有声条件下,被试者在有无声音延迟条件下都打出的评级分数更高的偏差。

总的来说,实验一的结果确认了碰撞声音在动态事件中因果关系感知的影响,与以前工作的发现是匹配的[16, 60, 61, 62, 63, 64]。这表明经典的心理学实验设计可以在虚拟现实环境中顺利进行。同时实验一的结果也比较强有力的表明了视觉和听觉感知的不同事件在发生的时间延迟小于/等于 200ms 时他们是能够被一起感知为同一事件的。

## 4.2 实验二：3D 虚拟环境下物体运动的因果感知

除了整个动态碰撞事件是在 3D 虚拟环境中模拟,实验二的其他实验设置与实验一是一样的。实验二的主要目的是决定是否之前该经典心理学实验的结论能够从 2D 延展到 3D 虚拟现实环境中。参与人员将能够看到以 3D 网格渲染并带有自然光照和阴影的红球与蓝球。带有明显砖块分解的浅灰色地面和有一定高度的圆形墙面被放置在虚拟环境中来模拟 3D 环境下小球按照圆形轨迹的运动。每一个小球的运动轨迹也通过一条黄色的在小球尾部的线进行了标注。这是为了给予参与人员对圆形运动轨迹的印象(见图 1-1(d))。

**3D 声音合成:** 在实验二中的 3D 虚拟环境不仅提供的是 3D 视觉效果,也同样提供了 3D 听觉感知:如利用 Unity 原生声音引擎来模拟适配声音位置和物体运动的 3D 声音。这是通过设置将空间混合音效设置为 3D 模式-使得具有多声道-以及利用 Microsoft HRTF Spatializer 提供的 7.1 环绕音效实现的。声音的音高仍然通过式(16)计算得到,同时混淆效果开启但范围取值较小(对应较小的测试房间)。对于声音音量的处理,使用了对数模型来决定音量和声音位置与被试者之间距离的关系(见图 4-1)。使用对数模型是为了抓住距离对左右耳相对音量的细致影响。对数模式的应用对

实验三非常重要：在每一个耳朵声音的不同都是被用来对 3D 环境中的声音信号进行定位的。在 3D 环境中的碰撞声音持续事件约为 6ms。

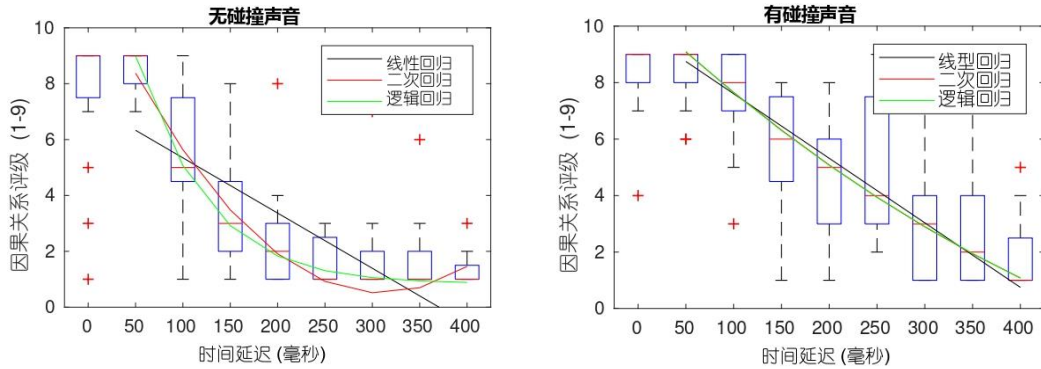


图 4-4：对实验二的因果关系评级结果：左)无碰撞声音条件，右)有碰撞声音条件

**实验流程：** 实验二同样也有训练环节，与实验一设计一样，由两个小的训练测试组成。除了在指导要求加了一步：确保参与人员在读完所有指导后旋转自身位置到合适的观察位置，其他的指导要求都和实验一是一样的。新添的指导要求中，参与人员被告知需要去寻找在他们左手方向的‘+’符号，从而使得他们能够面对静止的蓝色小球。一个开始按钮被固定在了‘+’符号下方，在按钮被按下之后，测试才开始。这个新添的指导要求是为了确保在 3D 虚拟环境中所有的参与人员观察每一次的碰撞事件都是从同一个角度观察的。

在训练阶段结束后，参与人员进入实验阶段，同样需要完成 36 个正式测试。除了实验开始前有两秒的暂停时间来给予参与人员足够时间准备以外，其他的时间延迟参数设置与实验一是一样的。和训练阶段一样，在碰撞事件呈现以前，参与人员需要按照‘+’符号的引导来调整他们的坐姿。

**实验结果分析：** 在实验二中，无论是有无碰撞声音的条件，因果关系评级分数的中值随着时间延迟的增加而下降。Friedman 检验的结果同样表明在有碰撞声音的条件下，评级分数大于没有碰撞声音条件下的评级分数： $\lambda^2(1) = 51.6, p < 0.001$ 。将实验一与实验二的结果进行对比，我们发现碰撞声音在 3D 虚拟环境中的影响要大于 2D 虚拟环境，这也可以从实验二中 $\lambda^2(1)$ 的值大于实验一中 $\lambda^2(1)$ 的值来证明。对数回归模型的斜率在有声音的条件下也比没有声音的条件下要小很多( $b = 0.32$  vs  $b = 2.66$ )，该差别要比实验一中的数据明显的多。

同时我们发现，在 3D 虚拟环境中，相应的因果关系感知评级分数直到时间延迟在大于等于 50ms 时，才开始下降，这和 Michotte 工作的经典结论是一致的[60]。因此，在实验二中进行的回归分析是从时间延迟为 50ms 开始的，而不是 0ms。其中，没有碰撞声音的条件下的回归分析结果是和实验一的结果类似的(图 4-4，红色水平线表示因果关系评级结果的中值，蓝色柱状的顶部与底部相应表示评级结果在 75%和 25%的情况。虚线的延展到了最极端的被认为是异常值的数据，而红色的‘+’号标出了对应的异常值。图中横轴表示了碰撞和蓝球开始运动之间的时间延迟。黑线，红线和绿线分别对应的是线性回归，二次回归和逻辑回归对上述数据的拟合情况。)：评级分数关于时间延迟的变化更拟合于二次回归模型 ( $r^2 = 0.97$ ,  $BIC = -8.83$ )和对数回归模型 ( $r^2 = 0.997$ ,  $BIC = -27.47$ )，而线性回归模型拟合程度较差 ( $r^2 = 0.72$ ,  $BIC = 7.56$ )。同时，有碰撞声音的条件下的回归分析结果也与实验一的结果类似：评级分数关于时间延迟的变化最拟合于线性回归模型，但同时二次回归模型 ( $r^2 = 0.995$ ,  $BIC = -23.19$ )和对数回归模型 ( $r^2 = 0.995$ ,  $BIC = -19.22$ )的拟合程度也是很不错的。

对比实验二(3D 环境)与实验一(2D 环境)的实验结果,我们发现如下明显不同:

1)在实验二中，关于因果关系的评级分数直到时间延迟在大于等于 50ms 才开始下降，而不是从 0ms(没有延迟)就开始下降。

2)在实验二中，对于有碰撞声音的条件下，二次回归模型和对数回归模型的 BIC 其实都要比线性回归模型的 BIC 要低。第一个不同可能是由于 3D 虚拟环境模拟的性质造成的，如物体的运动在 3D 环境的渲染，光照和阴影下更自然了，所以即使在 50ms 的延迟下，相应因果关系评级分数也是最高的。第二个不同的出现很有可能是因为二次回归模型和对数回归模型的斜率非常小，所以他们实际上非常接近线性回归模型的情况。尽管上述两个模型因为拥有一个额外的自由参数而受到了拟合上的计算惩罚，但他们模型本身是更精确的，所以得到了和线性回归模型几乎一样的 BIC 值，甚至更好。总的来说实验二的结果是和实验一的结果以及与 Guski 和 Troje 的结论是相符的：没有碰撞声音的条件下的因果关系感知评级分数是最拟合于非线性回归模型，而有碰撞声音的条件下的因果关系感知评级分数是最拟合于线性回归模型。

### 4.3 实验三：3D 虚拟环境下碰撞声音位置被改变

设计第三个实验的目的是为了检测是否碰撞声音的位置信息能够影响因果关系的感知。除了碰撞声音的位置是被改变：沿着圆周从视角 $-90^\circ$ 到 $90^\circ$ ，每间隔 $30^\circ$ 进行碰撞声音的放置。其他的实验设置都是和实验二一样的。此外，我们检验了参与人员是如何通过声音输入估计声音位置信息的。声音位置在 3D 环境中的改变遵循了下列公式所示的沿圆周的角度 $\theta$ 变化：

$$x_s = x_0 - r \sin(\theta) \quad (4-3)$$

$$y_s = y_0 \quad (4-4)$$

$$z_s = z_0 - r \cos(\theta) \quad (4-5)$$

其中 $(x_s, y_s, z_s)$ 是当前碰撞声音的位置信息， $(x_0, y_0, z_0)$ 是碰撞声音标准位置(ground-truth)，即物体碰撞发生碰撞点的位置。我们使用了 Unity 游戏引擎来实现物体声音位置的改变，同时声音音高也参照式(4-1)计算，声音音量的设置遵循了图 4-1 所示的对数模型。

**实验流程：**实验三正式开始前同样需要进行训练测试，一共有 12 个训练小测试由三种不同时间延迟(低延迟-50ms，一般程度延迟-200ms，高延迟-400ms)和四种不同声音的设置(没有声音，声音在 $-90^\circ, 0^\circ, 90^\circ$ )组成。需要注意的是这里在实验三训练阶段使用的时间延迟是与实验一和实验二不同的。这主要是因为之前的心理学实验工作没有发现关于因果关系的打分基本不会在 50ms 之前下降的结论，而我们利用实验一与实验二的结果，将新得到的结论用到了实验三的实验设置中，即直接从 50 ms 开始设置时间延迟条件。上述 12 个小测试是以随机顺序呈现的。

在完成训练测试后，参与人员随即如第 4 节实验设计部分所介绍的进行正式实验测试。整个正式测试环节被分为三个部分。需要参与人员在前两个测试部分回答的问题是和实验二一致的。在第三个部分，参与人员也需要利用控制器在虚拟环境中来表明他们对于 3D 虚拟现实环境中碰撞声音位置来源的判断。

**实验结果分析：**我们首先比较了在有碰撞声音且碰撞声音位置相互对称( $\pm 30^\circ$ ,  $\pm 60^\circ$ ,  $\pm 90^\circ$ )情形下的因果关系感知评级分数。这个结果比较是为了决定评级分数是取决于碰撞点前的位置还是碰撞点之后的位置。根据 Friedman 检验的结果表明：参与人员对上述对称情况的评级分数是几乎一样的，即不存在显著性差异： $\pm 30^\circ$  ( $\lambda^2(1) = 0.3$ ,  $p < 0.56$ ),  $\pm 60^\circ$  ( $\lambda^2(1) = 0.4$ ,  $p < 0.55$ ),  $\pm 90^\circ$  ( $\lambda^2(1) = 2.9$ ,

$p < 0.09$ )。因此，在对称偏离位置上的评级分数数据被总和在一起，与有碰撞声音但声音位置在 **ground-truth** 位置即  $\theta = 0^\circ$ ，和无碰撞声音条件下的数据进行比较(图 4-7，该图中所做的所有分析(线性回归，二次回归，逻辑回归分析)和图中符号与图 4-6 一致。)。我们发现评级分数在每一组声音位置有偏离与声音位置在  $\theta = 0^\circ$  时有显著差异： $\pm 30^\circ$  与  $0^\circ$  的  $\lambda^2(1)$  值为 14.66， $\pm 60^\circ$  与  $0^\circ$  的  $\lambda^2(1)$  值为 18.8， $\pm 90^\circ$  与  $0^\circ$  的  $\lambda^2(1)$  值为 34.22。尽管在  $\pm 90^\circ$  条件下与没有碰撞声音条件下的评级分数没有明显的统计学上的显著差异 ( $\lambda^2(1) = 0.4$ ,  $p = 0.5$ )，但  $\pm 30^\circ$  ( $\lambda^2(1) = 9.5$ ,  $p < 0.01$ ) 和  $\pm 60^\circ$  ( $\lambda^2(1) = 5.7$ ,  $p < 0.02$ ) 条件下与没有碰撞声音的评级分数存在显著性差异。上述结果表明如果将碰撞声音的位置直接放在被试者的正左边或者右边( $\pm 90^\circ$ )将没有什么影响(与没有碰撞声音类似)；而将声音放在相对于被试者左边或右边  $30^\circ$  和  $60^\circ$  时(相对于碰撞点衡量的角距离)将会对因果关系感知造成一定影响，但完全没有将碰撞声音放在 **ground-truth** 位置的影响大。

我们也对上述每一个声音位置条件下的因果感知评级分数进行了线性，二次和逻辑回归模型拟合分析。相关系数的平方( $r^2$ )和贝叶斯信息准则(BIC)值的数据被整理在表 4-1 中(二次回归和逻辑回归模型在每一种声音位置条件下拟合情况都是几乎一样的，随着声音位置越来越靠近碰撞点的标准位置线性回归模型的拟合情况逐渐提升,即  $r^2$  接近 1，评级分数与声音位置相关性越来越强)。

这些回归分析的结果与实验二中相应的分析结果是相符的：尽管当碰撞声音在 **ground-truth** 位置时，线性二次和逻辑回归对评级分数的拟合程度几乎是一样的，没有碰撞声音的条件下的因果关系感知评级分数仍最拟合于非线性回归模型；而有碰

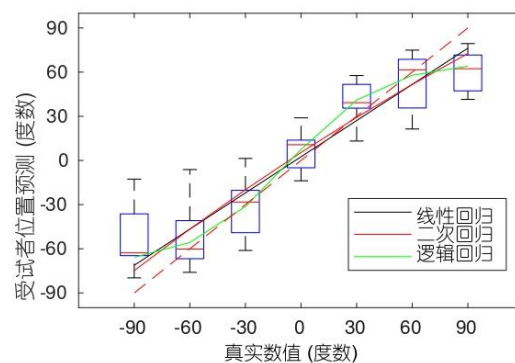


图 4-5：参与人员对碰撞声音位置的估计相对于真实位置(**ground-truth**)的情况



撞声音的条件下的因果关系感知评级分数仍是最拟合于线性回归模型的。有趣的是，随着碰撞声音的位置逐渐由 $\pm 90^\circ$ 的偏离位置变化到 ground-truth 的标准位置，线性回归模型拟合程度也越来越好。同时，在碰撞声音偏离位置为 $\pm 90^\circ$ 和 $\pm 60^\circ$ 的条件下，声音在参与人员视野领域右半部分明显线性拟合程度更好，即如 $+90^\circ$ 线性拟合程度大于 $-90^\circ$ 。尽管对于视野领域左右部分的因果关系感知评级分数没有统计上的显著性差异，但从拟合的角度上来说，其线性拟合程度在碰撞声音偏离位置为 $+90^\circ$ 和 $+60^\circ$ 条件下比在偏离位置为 $-90^\circ$ 和 $-60^\circ$ 条件下相对于 ground-truth 位置条件下下降的更快。这个差异没有在碰撞声音偏离位置为 $\pm 30^\circ$ 条件下发现。

我们也检验了参与人员是如何在 VR 环境中估计碰撞声音的位置的。图 4-5 描绘了相对于 ground-truth 位置的被试者预测碰撞声音位置的情况，其中黑线，红线和绿线分别表示了利用线性回归模型，二次回归模型，逻辑回归模型对上述数据的拟合情况。红色的虚线表示理想的估计情况。尽管被试者对于预测碰撞声音来自于哪个方向是基本准确的，他们对于当碰撞声音位置偏离较多时 $|\theta| \geq 60^\circ$ ，预测的偏差也较大。线性，二次和逻辑回归模型都在图六的位置预测中进行了分析。从图中也可以看出，关于碰撞声音位置的预测是最拟合于逻辑回归模型的（线性回归模型： $r^2 = 0.96$ ，二次回归模型： $r^2 = 0.96$ ，逻辑回归模型： $r^2 = 0.997$ ）。我们同时也比较了相对于碰撞声音在 ground-truth 位置偏离方向不同时，参与者预测声音位置的情况。有趣的是，通过比较 $\theta > 0$ 与 $\theta < 0$ 的数据，利用 t 检验，我们发现参与人员对位置的估计在方向上是有显著差异的： $t(143) = 0.4$ ,  $p = 0.03$ 。总的来说，实验三的结果表明人是能够在 3D 虚拟环境中对声音进行定位的，同时也会利用这样的位置信息来帮助感知系统建立起快速的对动态时间的因果关系感知。

在实验一中，我们强调了当前实验对参与人员提问问题的方式与以前 Guski 和 Troje 工作里的询问方式不同。然而在实验三的第三部分测试中，在进行碰撞声音位置预测测试前的实验指导环节，我们与他们之前实验一样向参与人员提示了碰撞声音存在的信息（包括位置可能的变化）。那么与他们的结论相似的是，我们也得到了关于当时间延迟为 400ms 时，因果关系评级分数在有碰撞声音的条件下是在无碰撞声音条件下分数的两倍的结论（图 4-6，需要注意与实验一和实验二不同，参与人员在进行因果关系感知的打分前被明确地告知了需要注意该实验中有可能有碰撞声音的存在。红色水平线表示因果关系评级结果的中值，蓝色柱状的顶部与底部相应表示评级结果在 75%和 25%的情况。虚线的延展到了最极端的被认为是异常值的数据，而红



色的‘+’号标出了对应的异常值。图中横轴表示了碰撞和蓝球开始运动之间的时间延迟。黑线，红线和绿线分别对应的是线性回归，二次回归和逻辑回归对上述数据的拟合情况。)。这个结果强有力的表明了本工作与之前 Guski 和 Troje 之前的工作存在的细微差别是由于关于判断因果关系的提问问题方式造成的。

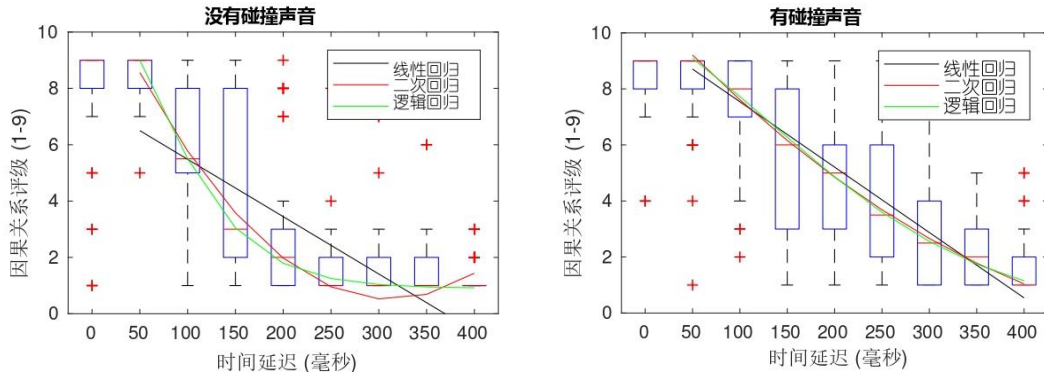


图 4-6：对实验三的因果关系评级结果：左)无碰撞声音条件，右)有碰撞声音条件

表 4-1：对于在每一个声音位置情况的评级分数上进行回归分析时  
相关系数的平方( $r^2$ )和贝叶斯信息准则(BIC)。

声音偏离角度( $\theta$ )	线性回归		二次回归		逻辑回归		
	$r^2$	BIC	$r^2$	BIC	$r^2$	BIC	Slope b
无声音	0.73	7.48	0.98	-11.57	0.998	-29.55	-0.32
-90°	0.67	9.20	0.96	-6.47	0.992	-14.83	-4.04
90°	0.82	5.74	0.97	-7.42	0.98	-6.05	-4.04
-60°	0.77	5.16	0.99	-15.04	0.99	-12.30	-4.63
60°	0.86	3.20	0.96	-4.45	0.97	-2.58	-3.54
-30°	0.88	3.99	0.94	0.64	0.99	-6.63	-5.39
30°	0.86	4.18	0.98	-8.85	0.99	-10.46	-4.15
0°	0.98	-13.04	0.99	-20.85	0.995	-18.57	-1.46

在实验三中关于对声音位置感知的结果将给予我们对基于物理声音模拟声音加速方案的理论支持。从表 4-1 中我们能明显看出当声音出现在观察者左手边/右手边时，声音对被试者注意力的吸引是比较低的，那么我们将不需要给予偏离被试者视角正前方的模拟声音同样的模拟时间来达到较高的声音质量，这样既能够加速大型场景中需要较多声音需要被模拟的情况，也能够最大程度减少总体声音质量的降低(具体见 5.2 节)。

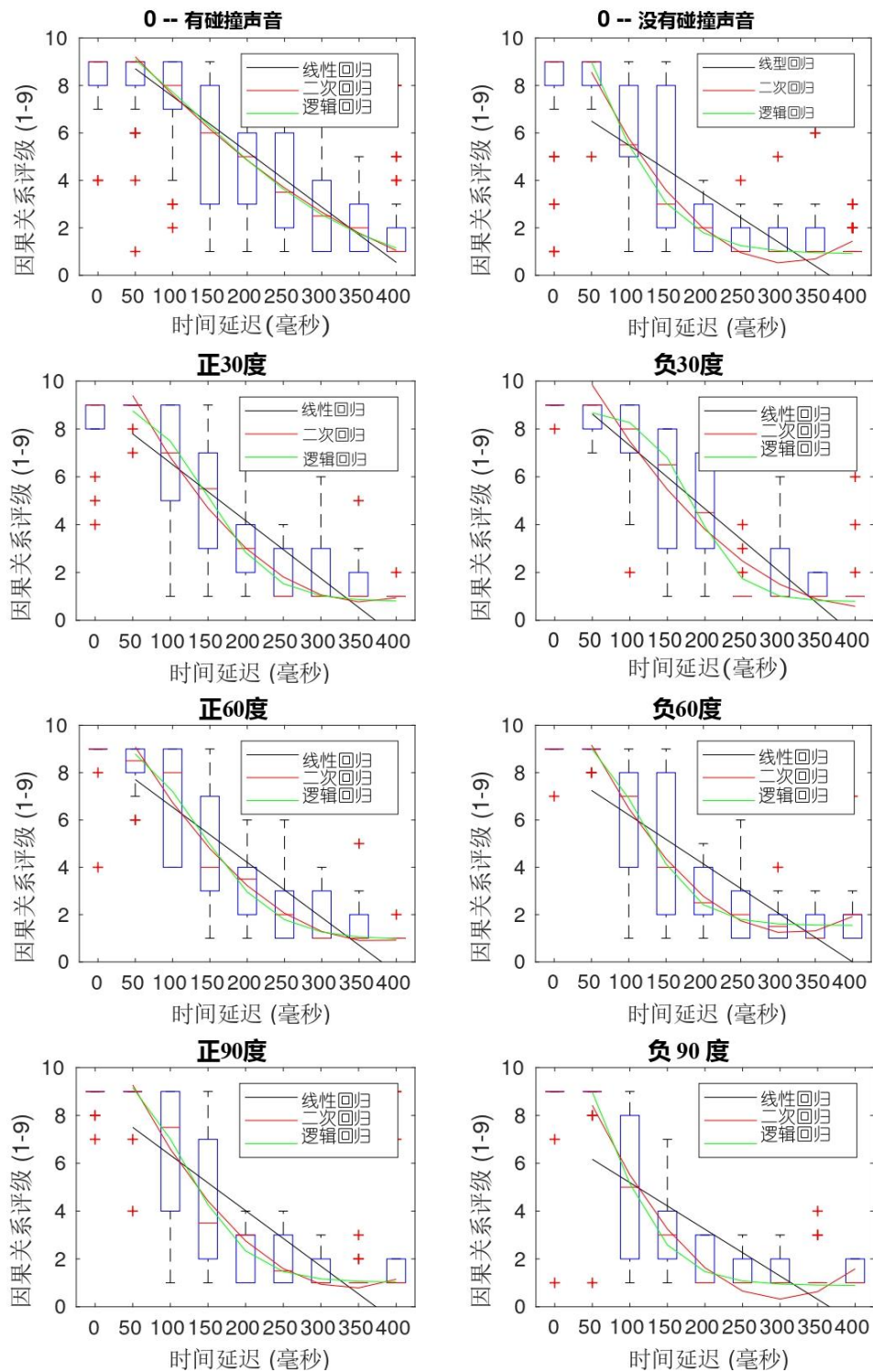


图 4-7：对实验三的因果关系评级：包括了所有声音条件(有无声音及声音位置)的分析。

## 第 5 章 结果与讨论

### 5.1 场景中实时声音合成

本小节主要介绍基于第三节的理论框架，对实时声音合成的具体实现情况：

(1) 预处理阶段-能够基于上述公式，对物体的网格信息进行处理并输入材质信息（杨氏模量和泊松比）后，通过弹簧质点系统建立相应的质量矩阵和刚体矩阵，通过广义特征值分解得到物体对应的频率和振动模式。

(2) 实时合成阶段-实现了模式压缩与模式截断的加速方案，利用 Unity 的游戏引擎提供碰撞力的信息，以实时决定混入的各个正弦波的振幅，从而产生根据敲击的不同力度不同位置而变化的声音。目前受限于对上述声音合成的可视化难度（作为一个完整的插件放入游戏引擎难度较大）。本工作当前只能在 Unity 中展现一些简单物体的声音合成如：球类物体，棍状物体，立方体物体等（见图 5-1）。但本工作已能够充分利用材质信息来合成对应的物体碰撞声音如：金属碰撞声，木材碰撞声，石头碰撞声等（图 3-5，包括软硬地面切换导致不同碰撞声音和金属球滚动声音。）。但我们已经充分显示了该基于物理的声音合成方法，能够被集成在游戏引擎中去来提供更加真实的声音。

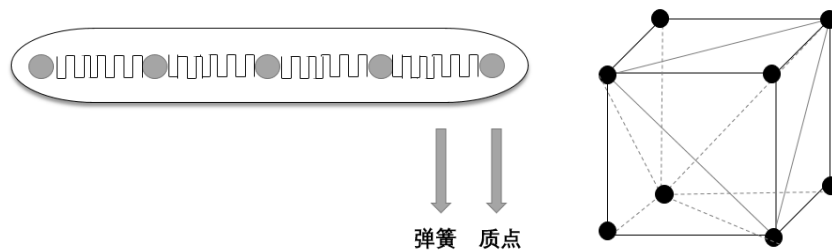


图 5-1：左) 棍状物体弹簧质点系统的建立，右) 立方体物体基于四面体网格信息的剖分。

同时由于目前测试物体较为简单，在场景中，我们是能够承受几十上百个物体同时进行声音合成而发声的，这充分说明了实时性。但对于多个复杂物体的声音实时合成，还存在着加速方案在未来的工作中去实现（见 5.2 节）。

## 5.2 心理学实验结果讨论

基于当前实验，我们证明了 1) 因果感知的经典结果[11, 13, 16]能够在虚拟现实被成功复现。2) 尽管有一些偏差，参与者能够感知到声音的虚拟位置。3) 碰撞声音的位置信息的偏离确实会衰减运动物体之间的因果感知印象。

对于实验三的进一步讨论：实验三的结果可以根据视觉信息在感知整合的普遍性进行理解<sup>[58]</sup>。一个常见的例子是腹语表演，表演者手边有一个玩偶仿佛也在发出声音[17]。如果表演者站在原理玩偶的另一边，观察者将非常清楚地认为这个声音发出者是表演者。这与第三个实验结果是匹配的，即当偏离角度为 $\pm 90^\circ$ 时，声音将对小球碰撞事件的因果判断没有任何影响。当声音的位置逐渐回到碰撞点的标准位置时，实验中的评级分数将逐渐递增。所以在未来的工作中将可以针对 $\pm 30^\circ$ 的偏离区间进行更加细致的分析，来测验位置信息更细致的影响情况。

实验三的结果也表明了物体运动的时间延迟和因果关系感知之间的关系可能不是单一的静态形式的(不仅仅是纯偏向于线性/二次关系)，将随着声音位置而改变。比如声音在实验三中场景设置的圆弧右边时，因果关系感知与时间延迟之间的关系更趋近于线性下降。而声音在圆弧左边时，他们之间的关系更趋近于二次。这很有可能是因为人对注意力资源的分配在视觉上是不均匀的，更多的注意力是被分配到了被试者朝向的物体上，更少的注意力分配到了之前看过的物体和精致的物体。但我们也要注意到这也有可能是 HMD 显示器在显示时存在一定的方向误差造成的。未来的工作也应该将被试者观察的方向作为变量纳入实验设计的考量范围内进行检测。

总的来说，当前实验结果展现了 VR 技术对研究人感知和认知的可行性。该技术提供了花费较低且能够提供更多复杂系统直观交互形式的新手段，可以被充分用于构建独特且在真实世界中难以模拟的环境来进行各种人体认知研究。然而对于是否一些基于其他感觉器官的认知研究是否也可以放入虚拟现实环境中仍然是不清楚的(比如，基于触觉反馈的一些研究，当前虚拟环境还没有高效的基于物理的触觉反馈交互模拟[25])。因此，下一步的工作应是测定利用 VR 技术对人类感知研究的长处与缺点。

当前的实验结果特别是实验三的实验结果，给未来的大型场景的声音合成提供了实验层次的理论性支持：

1) 注意力在感知上分配是不均匀的，无论是视觉还是听觉，人都倾向于给与其正

前方的动态物体最多的关注度，给予背对的物体最少的关注度。

2) 由于偏离角度为 $\pm 90^\circ$ 时，被试者对此声音对运动物体的发声贡献感知非常小，那么对于整个场景的多种声音而言，这样的声音的质量对场景总的声音效果影响也是非常小的。这也是 Ming Lin 等设计声音合成加速方案的重要理论依据[6]，他们提出认为观察者只对其正面面对的声音注意最多，因此如果只保持观察者正面对的发声物体的声音质量为最高时(其他方向的声音依次降低，在观察者背部的声音质量最低)，对场景中声音的总效果质量将是最少的。那么在未来工作中，我们可以利用这个重要结论在对单个物体声音合成进行加速处理的基础上，对场景中多个发生物体的声音合成进行加速处理，使得基于物理的声音合成方法能够真正的应用于复杂物体，复杂场景，能够成为一个比较完整的声音插件集成到游戏引擎中。

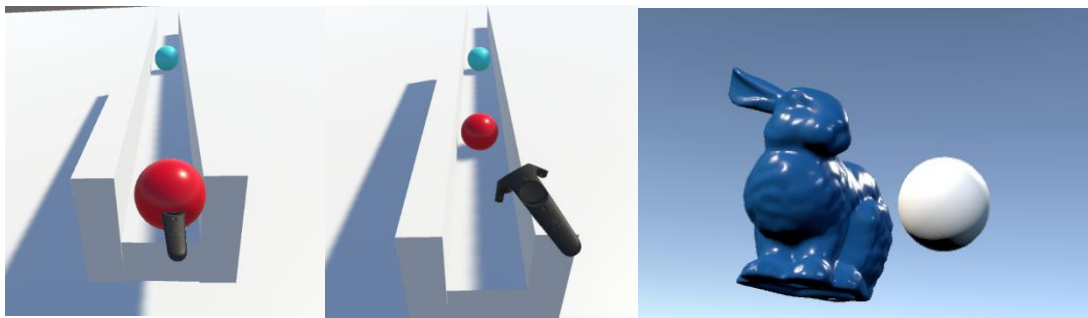


图 5-2：未来可能的实验设计：1) 被试者自由决定小球速度 2) 不同形状物体之间碰撞

当前的三个实验也有许多可以进行延展的地方，因为目前的所有实验中的动态事件都是被试者仅仅通过观看来进行感知的。下一步工作的可能方向会是将被试者的参与作为一个输入变量加入到当前虚拟测试实验中(如，被试者利用控制器自由决定运动红球初速度，图 5-2)，这也会影响到被试者对因果关系的感知。当被试者决定好初速度后，后续的实验中红球就以新的初速度进行运动。另一个潜在的影响变量将是根据物体移动速度改变声音的音高，由于本实验中两小球质量相等且遵循无摩擦的动量守恒，所以碰撞前后移动速度并无任何改变。当加入这一变量后，可以将实验设置延展到带有摩擦和带碰撞损失的场景中，这可以直接用本工作中基于物理的声音合成的方法来进行实现。最后一个潜在的影响变量是两碰撞物体之间的形状如果不同将造成被试者对碰撞点难以估测，这也会影响到被试者对因果关系的感知(图 5-2)。但具体的影响可以留在未来的工作继续探索了。

## 结论

本工作以计算机图形学和虚拟现实技术为基础，从听觉层面对计算机的实时模拟技术进行了一定探索，提供了未来有希望能够应用于工业级环境模拟/游戏开发的实时声音模拟方法。本工作的创新点主要在于从物理建模的角度而不是以前以采样为基础的声模拟方法，且通过心理学实验的搭建展示了实时声音模拟可以进行应用的场景，同时从感知层面和实验的结果提供了后续提升模拟实时性的加速方案。

本工作在实验设计方面，也提供了在虚拟现实环境中进行实验的比较详细的设计与实施方案，包括对变量的控制，针对参与人员的实验设计，相应训练流程设计以及实验数据统计的设计等。充分展现了虚拟现实技术提供的沉浸式环境能够从视觉和听觉的双重层面来进行许多以前在现实中难以设置实验条件的实验/训练/测试等。

在未来的工作中，应该将更多丰富的感知器官(如:嗅觉，触觉)引入到沉浸式的虚拟环境中从而建立起更完善的环境模拟技术。同时环境模拟也需要考虑更多的实时性和准确性，因此需要探索更多的计算加速，模型压缩等方法来进行模拟层面的加速。对于环境模拟真实性的研究也应该继续推进，这不仅需要进行更多的与真实环境对比的实验，也需要更多的真实数据收集和分析工作。比如对于声音模拟而言，只有采集了足够多的真实声音样本，我们才能更加具体的分析哪些声音是更容易实时模拟或者利于物理建模进行分析的，从而建立起更加成熟完整的，能够用于工业界声音模拟方案。当今电影工业和游戏工业主要还是在基于采样进行实时声音的模拟，本工作提出的方法如果在未来发展成熟将能够给他们提供更加丰富，方便调整的各种声音，将极大提升电影作品/游戏作品的质量和生产速度。

## 参考文献

- [1] Cook, P. R. (2002). Real sound synthesis for interactive applications[M]. AK Peters/CRC Press.
- [2] Sifakis, E., & Barbic, J. (2012, August). FEM simulation of 3D deformable solids: a practitioner's guide to theory, discretization and model reduction[R]. In Acm Siggraph 2012 courses (p. 20). ACM.
- [3] Liu, L., Wang, R., Xu, S., & Bai, Y. (2018, August). Development of AR/VR Project Based on Unity3D Engine. In Journal of Physics: Conference Series (Vol. 1069, No. 1, p. 012033)[J]. IOP Publishing.
- [4] Li, D., Fei, Y., & Zheng, C. (2015) [J]. Interactive acoustic transfer approximation for modal sound. ACM Transactions on Graphics (TOG), 35(1), 2.
- [5] James, D. L. (2016, July). Physically based sound for computer animation and virtual environments[R]. In ACM SIGGRAPH 2016 Courses (p. 22). ACM.
- [6] Raghuvanshi, N., & Lin, M. C. (2006, March) [J]. Interactive sound synthesis for large scale environments. In Proceedings of the 2006 symposium on Interactive 3D graphics and games (pp. 101-108). ACM.
- [7] Hume, D. (2006). A Treatise of Human Nature: Volume 1: Texts[M].
- [8] Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy[J]. Trends in cognitive sciences, 4(8), 299-309.
- [9] Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain[J]. Neuropsychology, 19(5), 591.
- [10] Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality[J]. The Quarterly Journal of Experimental Psychology, 44(2), 321-342.
- [11] Natsoulas, T. (1961). Principles of momentum and kinetic energy in the perception of causality[J]. The American Journal of Psychology, 74(3), 394-402.
- [12] Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: The role of the prior context[J]. Cognition, 132(3), 485-490.
- [13] Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects[J]. Psychological review, 120(2), 411.
- [14] Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events[J]. Psychological Science, 13(6), 493-498.
- [15] Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain[J]. Neuropsychology, 19(5), 591.
- [16] Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. Perception & psychophysics, 65(5), 789-800[J].
- [17] Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration[J]. Current biology, 14(3), 257-262.
- [18] Ament, V. T. (2014). The Foley grail: The art of performing sound for film, games, and animation[M]. Focal Press.
- [19] Lloyd, D. B., Raghuvanshi, N., & Govindaraju, N. K. (2011, February). Sound synthesis for impact

- sounds in video games[J]. In Symposium on Interactive 3D Graphics and Games (pp. PAGE-7). ACM.
- [20] Engine, U. G. (2008). Unity game engine-official site. Online[OL]. <http://unity3d.com>, 1534-4320.
- [21] Director-O'Brien, J. F. (2001, August). Synthesizing sounds from physically based motion[J]. In ACM SIGGRAPH 2001 video review on Animation theater program (p. 59). ACM.
- [22] O'Brien, J. F., Shen, C., & Gatchalian, C. M. (2002, July). Synthesizing sounds from rigid-body simulations[J]. In Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation (pp. 175-181). ACM.
- [23] Cohen, L.B., & Oakes, L.M. (1993). How infants perceive a simple causal event[J]. *Developmental Psychology*, 29 (3), 421.
- [24] Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality[J]? *Cognition*, 25(3), 265-288.
- [25] Ye, T., Qi, S., Kubricht, J., Zhu, Y., Lu, H., & Zhu, S. C. (2017). The martian: Examining human physical judgments across virtual gravity fields[J]. *IEEE transactions on visualization and computer graphics*, 23(4), 1399-1408.
- [26] Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction[J]. *Psychological review*, 116(4), 661.
- [27] Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference[J]. *Proceedings of the national academy of sciences*, 113(34), E5072-E5081.
- [28] Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. [J] *Trends in cognitive sciences*, 21(10), 749-759.
- [29] Mól, A. C. A., Jorge, C. A. F., & Couto, P. M. (2008). Using a game engine for VR simulations in evacuation planning[J]. *IEEE computer graphics and applications*, 28(3), 6-12.
- [30] Reznick, M., Harter, P., & Krummel, T. (2002). Virtual reality and simulation: training the future emergency physics[J]. *Academic Emergency Medicine*, 9(1), 78-87.
- [31] Tarnanas, I., & Manos, G. C. (2001). Using virtual reality to teach special populations how to cope in crisis: the case of a virtual earthquake[J]. *Studies in health technology and informatics*, 81, 495-501.
- [32] Xi, M., & Smith, S. P. (2014, March). Simulating cooperative fire evacuation training in a virtual environment using gaming technology[J]. In 2014 IEEE Virtual Reality (VR)(pp. 139-140).
- [33] Stansfield, S., Shawver, D., Sobel, A., Prasad, M., & Tapia, L. (2000). Design and implementation of a virtual reality system and its application to training medical first responders[J]. *Presence: Teleoperators & Virtual Environments*, 9(6), 524-556.
- [34] Andreatta, P. B., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., ... & Frankel, J. (2010). Virtual reality triage training provides a viable solution for disaster-preparedness[J]. *Academic emergency medicine*, 17(8), 870-876.
- [35] Backlund, P., Engstrom, H., Hammar, C., Johannesson, M., & Lebram, M. (2007, July). Sidh-a game based firefighter training simulation[J]. In 2007 11th International Conference Information Visualization (IV'07) (pp. 899-907). IEEE.
- [36] Cha, M., Han, S., Lee, J., & Choi, B. (2012). A virtual reality based fire training simulator integrated with fire dynamics data[J]. *Fire Safety Journal*, 50, 12-24.



- [37] Tate, D. L., Sibert, L., & King, T. (1997, March). Virtual environments for shipboard fire-fighting training[J]. In Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality (pp. 61-68). IEEE.
- [38] Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety[J]. IEEE transactions on visualization and computer graphics, 21(4), 529-538.
- [39] Backlund, P., Engstrom, H., Johannesson, M., & Lebram, M. (2007, July). Games and traffic safety-an experimental study in a game-based simulation environment[J]. In 2007 11th International Conference Information Visualization (IV'07) (pp. 908-916). IEEE.
- [40] McComas, J., MacKay, M., & Pivik, J. (2002). Effectiveness of virtual reality for teaching pedestrian safety[J]. CyberPsychology & Behavior, 5(3), 185-190.
- [41] Padgett, L. S., Strickland, D., & Coles, C. D. (2005). Case study: using a virtual reality computer game to teach fire safety skills to children diagnosed with fetal alcohol syndrome[J]. Journal of pediatric psychology, 31(1), 65-70.
- [42] Kozlov, M. D., & Johansen, M. K. (2010). Real behavior in virtual environments: Psychology experiments in a simple virtual-reality paradigm using video games[J]. Cyberpsychology, behavior, and social networking, 13(6), 711-714.
- [43] Olivier, A. H., Bruneau, J., Cirio, G., & Pettré, J. (2014). A virtual reality platform to study crowd behaviors[R]. Transportation Research Procedia, 2, 114-122.
- [44] Rovira, A., Swapp, D., Spanlang, B., & Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents[J]. Frontiers in behavioral neuroscience, 3, 59.
- [45] Li, C., Liang, W., Quigley, C., Zhao, Y., & Yu, L. F. (2017). Earthquake safety training through virtual drills[J]. IEEE transactions on visualization and computer graphics, 23(4), 1275-1284.
- [46] Zhu, Y., Mottaghi, R., Kolve, E., Lim, J. J., Gupta, A., Fei-Fei, L., & Farhadi, A. (2017, May). Target-driven visual navigation in indoor scenes using deep reinforcement learning[J]. In 2017 IEEE international conference on robotics and automation (ICRA) (pp. 3357-3364). IEEE.
- [47] Zhu, Y., Gordon, D., Kolve, E., Fox, D., Fei-Fei, L., Gupta, A., ... & Farhadi, A. (2017). Visual semantic planning using deep successor representations[J]. In Proceedings of the IEEE International Conference on Computer Vision (pp. 483-492).
- [48] Hertkorn, K., Roa, M. A., Brucker, M., Kremer, P., & Borst, C. (2013, November). Virtual reality support for teleoperation using online grasp planning[J]. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2074-2074). IEEE.
- [49] Yan, X., Khansari, M., Bai, Y., Hsu, J., Pathak, A., Gupta, A., ... & Lee, H. (2017). Learning grasping interaction with geometry-aware 3d representations[A]. arXiv preprint arXiv:1708.07303, 1.
- [50] Lin, J., Guo, X., Shao, J., Jiang, C., Zhu, Y., & Zhu, S. C. (2016, November). A virtual reality platform for dynamic human-scene interaction[J]. In SIGGRAPH ASIA 2016 virtual reality meets physical reality: Modelling and simulating virtual humans and environments (p. 11). ACM.
- [51] Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2018). Airsim: High-fidelity visual and physical simulation for autonomous vehicles[J]. In Field and service robotics (pp. 621-635). Springer, Cham.
- [52] Aravena, C., Vo, M., Gao, T., Shiratori, T., Yu, L. F., & Contributors, E. (2017). Perception Meets Examination: Studying Deceptive Behaviors in VR[J]. In CogSci.
- [53] Lin, J., Zhu, Y., Kubricht, J., Zhu, S. C., & Lu, H. (2017). Visuomotor Adaptation and Sensory

- Recalibration in Reversed Hand Movement Task[J]. In CogSci.
- [54] Azmandian, M., Hancock, M., Benko, H., Ofek, E., & Wilson, A. D. (2016, May). Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences[J]. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 1968-1979). ACM.
- [55] Bruder, G., & Steinicke, F. (2014, November). Threefolded motion perception during im-mersive walkthroughs[J]. In Proceedings of the 20th ACM symposium on virtual reality software and technology (pp. 177-185). ACM.
- [56] Zwicker, E., & Fastl, H. (2013). Psychoacoustics: Facts and models (Vol. 22)[A]. Springer Science & Business Media.
- [57] Si, H. (2015). TetGen, a Delaunay-based quality tetrahedral mesh generator[J]. ACM Transactions on Mathematical Software (TOMS), 41(2), 11.
- [58] Gibson, J. J. (1966). The senses considered as perceptual systems[M].
- [59] Michotte, A. (2017). The perception of causality[M]. Routledge.
- [60] Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony[J]. Perception, 9(6), 719.
- [61] Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration[J]. Behavioural Brain Research, 121(1), 69-79.
- [62] Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. Journal of Experimental Psychology Human Perception & Performance, 22(5), 1094.
- [63] Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception[J]. Nature, 385(6614), 308.
- [64] Pappas, T. N. (2000). Sensitivity to auditory-visual asynchrony and to jitter in auditory-visual timing[J]. Electronic Imaging.
- [65] Jacob, B., & Guennebaud, G. (2016)[M]. Eigen: C++ template library for linear algebra.

## 致谢

论文行文至此，内容已经基本结束，同时我的本科大学生活也即将画上句号。回首过去的满满四年，我在北京理工大学学习了诸多专业知识，结识了许多可亲可敬可爱的老师和同学。在这四年的时光里，我也逐渐地成长，成熟。

这篇论文继承了许多前人，老师们，同学们的心血，也花费了我大量的时间和精力，但我认为这一切都是值得的。在这里我想特别感谢我的指导老师们，以及参与本实验的各位同学，是你们的亲切帮助才使得我一步一地从零开始坚持到这个工作取得阶段性成果。在大学四年里，很幸运地能在自己喜爱的道路上自由探索，充分追寻科研的方向，这也离不开学校与学院的倾力支持，为我真正的插上了全球视野的翅膀，让我将想要出国留学的想法逐步变成了可以实现的目标。希望在以后的道路上能够继续以同样的饱满的热情，投入到新的工作和学习中，将北理人的精神发扬光大，不仅努力为自己的发展铺平道路，也希望未来能够做出一番事业来回馈母校。