DOI: 10.3969/j.issn.1009-6868.2017.06.001
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20171109.0856.004.html

虚拟环境中人和虚拟角色互动的关键技术

The Key Technologies of Human-Virtual Character Interaction in Virtual Environment

翁冬冬/WENG Dongdong
薛雅琼/XUE Yaqiong
（北京理工大学，北京 100081）
(Beijing Institute of Technology, Beijing 100081, China)

中图分类号：TN929.5 文献标志码：A 文章编号：1009-6868 (2017) 06-0002-004

摘要： 剖析虚拟环境中人和虚拟角色互动的关键技术，其中空间定位、动作捕捉、眼动追踪和语音输入实现虚拟角色对用户的多通道交互输入信号的接收，反馈控制系统建立从交互输入信号到呈现给用户的多通道反馈输出信号的映射，反馈信号渲染将多通道反馈输出信号以虚拟角色的表情、动作、语音等形式呈现给用户，同时力触觉反馈技术支持虚拟角色提供触觉通道的交互，使交互具有更多的物理性。

关键词： 人-虚拟角色互动；虚拟现实；多通道交互信号接收；反馈控制；多通道反馈信号渲染；力触觉反馈

Abstract: In this paper, the key technologies of human-virtual character interaction in virtual environment are analyzed. The reception of multi-channel interactive input signal of user is achieved by space positioning, motion capture, eye tracking and voice input. The mapping from interactive input signal to response signal is constructed by response control system. The multi-channel response signal is shown to user by response render technology with lifelike facial appearance, motion and voice. Moreover, force feedback technology adds the haptic interaction to human-virtual character interaction, and enhances the physicality of interaction.

Key words: human-virtual character interaction; virtual reality; reception of multi-channel interactive signal; response control; rendering of multi-channel response signal; force feedback

1 人和虚拟角色的互动

虚拟现实（VR）技术通过以计算机技术、人机交互为核心的高新技术生成逼真的视觉、听觉、触觉等多通道的、一定范围内的虚拟环境，用户可以借助必要的设备以自然的方式与虚拟环境中的物体交互，获得亲临等同真实环境的感受和体验。VR具有“3I”特性，即沉浸感（immersion）、交互性（interaction）和构想性（imagination），并且随着同人工智能技术的不断结合，VR系统表现出更多的智能性（intelligence），并逐渐向4I发展。

VR技术在发展的早期阶段多用于军事、航空航天、医学、培训、工业仿真、城市规划等严肃的专业领域，充分利用3I特性对现实世界进行高精度的模拟与呈现，辅助用户研究分析，解决复杂问题。随着Oculus Rift、HTC Vive和PlayStation VR等消费级VR产品的发布，VR技术逐渐从专业应用领域走向消费者市场，教育、娱乐等个人应用领域进一步蓬勃发展。

最简单的VR应用是虚拟电影院，在用户的视野前方设置一块可以播放电影的巨大屏幕，虽然播放的资源仍然是原先的传统电影，但是可以让用户享受到在电影院观影的感觉。360°视频则更进一步地不再固定用户的视角，给予用户更多自主选择观看视角的自由。严格来看，这两者不能算作真正意义上的VR应用，因为并没有充分体现出VR的3I特性，单单通过虚拟环境的360°环绕给人仿佛“身临其境”的感官错觉，缺乏空间感知和运动支持，实现的沉浸感是粗糙而脆弱的，交互性和构想性更是不曾体现。它们最多只能作为当前阶段缺乏成熟的制作VR互动体验的经验，为方便广大消费者了解VR技术而做出的折衷。

在真正的VR互动体验中，用户不再以一个旁观者的角度观看，而是走进故事中，亲身存在于虚拟世界里，以第一人称视角对虚拟世界进行感知与理解，通过四处走动加深对虚拟世界的空间感认知，同时通过自身活动影响周围的虚拟环境和虚拟角

收稿日期：2017-09-26
网络出版日期：2017-11-09
基金项目：国家高技术研究发展（“863”）计划（2015AA016303）；国家重点研发计划（2016YFB1001401）

色的发展，从而更深刻地感受到踏入一个全新世界的真实感。人和虚拟角色的互动就可以归入这一范畴。如果需要实现复杂的、接近真实世界中人-人交互的人和虚拟角色的互动体验，需要众多相关技术的支持。

2 人和虚拟角色互动的关键技术

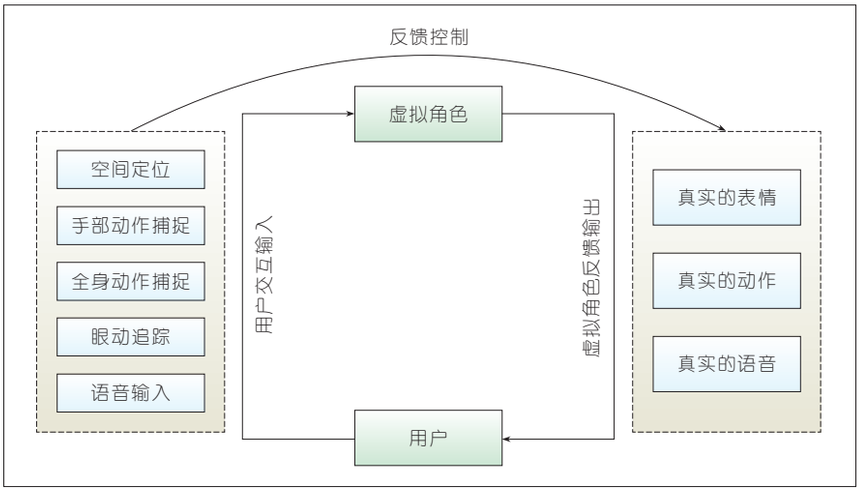
如图1所示，一个完整的人-虚拟角色交互环路及各环节中需要解决的关键技术包括：(1)虚拟角色对用户的多通道交互输入信号进行接收，即通过对用户的头部和手部进行空间定位，赋予虚拟角色感知用户当前位置和视线方向，以及打招呼、递东西等手部动作的能力；通过动作捕捉，对用户全身的动作进行跟踪；通过眼动追踪，更准确地了解用户注视点的变化；通过麦克风采集用户的语音输入。(2)虚拟角色根据用户的多通道交互输入信号，确定将要呈现给用户的多通道反馈输出信号，例如面对用户的打招呼动作，确定是否做出反应，以及做出怎样的反应。(3)虚拟角色将多通道反馈输出信号通过自身的表情、姿态、动作、语音等形式呈现给用户。

此外，在纯虚拟的交互基础上，可以进一步借助力触觉反馈技术为虚拟角色提供触觉支持，使交互具有更多的物理性，即用户在虚拟环境中看到虚拟角色的同时，可以在真实世界的相同位置处触摸到虚拟角色。

2.1 空间定位

为了给用户提供真实的包含视、听、触多通道信息的虚拟环境，需要实时检测用户头部的位置和视线方向，计算机能根据这些信息确定所要呈现的虚拟感官通道信息，并通过各通道的输出设备实时呈现。为了使用户在虚拟环境中体验到更高的沉浸感，同时赋予用户一定的手部交互能力，需要实时检测用户手部的位置和姿态并在虚拟环境中渲染，同时通过将手部与可交互的虚拟目标进行碰撞检测做出交互判断。

应用于VR领域的跟踪技术，通常以传感器为核心构建跟踪系统，根据选用的传感器种类不同，跟踪系统分为机械跟踪、电磁跟踪、超声波跟踪、光学跟踪和惯性跟踪。机械跟踪会极大地限制用户自身的运动，不适合人和虚拟角色存在丰富互动的应用。电磁跟踪和超声波跟踪易受工作环境中磁场、金属物体与刺激性声波脉冲的干扰，抗干扰性较差，且跟踪精度会随着跟踪范围的增大而迅速衰减，也不适合包含空间定位，允许用户在较大范围内走动交互的应用。而光学跟踪(包括广义的光学跟踪技术，例如激光扫描空间定位技术)和惯性跟踪能够实现较大的跟踪范围，且两者相结合的跟踪方案能够取长补短，光学跟踪提供高精度的空间定位，同时对惯性跟踪中随时间推移产生的较大累积误差进行校正，惯性跟踪克服光学跟踪在遮挡、画面模糊时定位失败的问题，同时高采样率确保能实时跟踪目标的快速运动，保证交互的实时性，是当前技术阶段中主流的空间定位解决方案。



▲图1 人-虚拟角色交互环路中的关键技术

2.2 动作捕捉

通过对用户的手部进行空间定位，虚拟角色能够感知用户手部做出的简单动作，例如：招手、递送物体，并做出相应的回应，但是手柄实现的简单空间定位无法完全跟踪手部丰富自由度的运动，且在交互过程中持续地手持设备会干扰人-虚拟角色交互的自然性。为此可以采用手部动作捕捉技术，无需手持设备，裸手或者穿戴数据手套就能跟踪全手运动，与虚拟角色自然互动。此外还可以进一步通过全身动作捕捉来实现用户全身的运动跟踪。

动作捕捉技术从实现方式上分场景深度解析方案和可穿戴方案两种。场景深度解析方案通过光学传感器接收到的光信号来分析场景的深度信息，进而确定手部或者全身的位置和姿态，无需用户手持或者穿戴专用设备就可实现动作捕捉，最大程度地减少真实世界的干扰，使人和虚拟角色的互动更加自然；但是获取的人体骨骼运动较为粗糙，限制了交互的真实感。相比之下，可穿戴方案通过在用户身上多个关键点处固定传感器或者标志点，测量该点的位置变化或者弯曲程度，反算用户的身体运动，实现的动作捕捉更为精细。

2.3 眼动追踪

通过对用户的头部进行空间定位，虚拟角色能够利用屏幕中心粗略地估计用户的注视点，做出与用户视

线接触等交互。若是需要更准确地确定用户的注视点，以便更好地理解用户的交互意图，可以将眼动追踪技术[1]集成至头戴式显示器中，例如FOVE头戴式显示器通过嵌入两枚小型红外摄像头，采集被红外发光二极管照亮的人眼的图像，利用角膜反射法[2]计算用户的眼球位置。

眼动追踪技术还能捕捉、记录反映用户一定情绪和认知过程的眼部活动[3-4]，例如表征情绪状态变化的瞳孔缩放和与心理负荷息息相关的眨眼频率。综合分析这些反映情绪和认知的眼动数据和其他通道获取的用户输入，能更好地判断用户的情绪和交互意图，使虚拟角色做出更加合适的反应。

除此之外，利用眼动追踪技术获取的眼球运动数据，还可以实现模拟人眼视觉的视网膜中心凹渲染技术。该技术可以只对视域中央的画面进行高分辨率渲染，视域边缘采用逐渐降低的分辨率，大幅降低硬件计算负担，避免精细地渲染整幅画面耗费大量计算资源，导致渲染帧率下降，人-虚拟角色互动实时性变差[5-6]。

2.4 语音输入

人们倾向于将交互对象拟人化，面对一个虚拟角色，尤其是类人形时，会自然地期待它能表现出类似人类的行为，而语音交互作为人-人交互中最重要交互手段之一，非常有必要实现于人-虚拟角色互动中。

在交互过程中，虚拟角色将采集到的用户语音输入，先通过语音识别转化为相应的文本内容，再通过语义理解进行基于上下文的交互意图判断，同时虚拟角色还可以通过对音调、响度等声学特征和语音内容进行分析，判断用户在与之互动时的情绪变化[7-8]，综合分析用户的交互意图和情绪，做出更为合适的反应。

2.5 反馈控制系统

反馈控制系统的作用在于根据获取到的多通道交互输入信号，确定将要呈现给用户的多通道反馈输出信号。传统方式是建立一个由交互输入到反馈输出的程式化的映射，一切按照预先写好的程序进行，一定的输入必然对应一定的输出，或者按照一定概率对应一系列输出中的一种，这种程式化的虚拟角色反馈极度单调、不自然，缺乏使人与之长期互动的吸引力。

为了建立丰富的、自然的虚拟角色反馈，感知-控制-行动模型（SCA）、并行转换网络模型（PaT-Nets）[9]、等多种虚拟角色行为控制模型被建立，且随着深度学习和大数据的不断发展，在不久的将来虚拟角色甚至能够以学习的方式自行建立起人-虚拟角色交互的反馈模型。

2.6 反馈信号渲染

反馈信号渲染的目标在于将包括表情、动作和语音在内的多通道反馈输出信号呈现给用户。为了使用户感到虚拟角色是真实的，吸引用户按照人-人交互的方式同虚拟角色互动，要求虚拟角色的表情、动作和语音都是接近真实的。

真实的表情和动作可以通过动作捕捉并录制获得，通过在动作录制者的脸上和身上粘贴或绘制标志点，并对标志点进行跟踪，即可利用跟踪数据来驱动虚拟角色做出同样的表情和动作，但是此方法获得动作和表情需预先录制，限制交互的丰富性。虚拟角色的表情和动作也可以通过深度学习进行训练，目前Google使用强化学习算法训练人工智能越过障碍物从起点跑至终点，已经成功地使人形模型自行学会了行走、跳跃等动作[10]，这种通过学习产生的动作和表情能够实时生成无需预先录制，同时有望做到非常接近真实的程度，但当前研究进展距离商业可应用还有一段不短的路要走。

传统的语音合成方案为参数化语音合成和拼接式语音合成，均利用已有的声音进行重组来合成新的语段音频。该方式产生的语音能基本接近人类表达的流畅度，但是听起来不自然，且由于无法产生可以自适应变化的语调和语速来反映说话者的情绪，很难让人产生“我在跟一个人说话”的感觉。为了获得真实的语音，一方面可以针对交互场景预先录制，此方法同样存在交互丰富性受限的问题；另一方面可以引入学习的手段，例如Google的WaveNet[11]利用真实的人类声音和相应的语言、语音特征来训练卷积神经网络，使其掌握不同语音、语言的模式，能够实时合成出更加接近自然人声的语音音频，并且模拟一定的语调、情感和口音，但是距离让用户无法区分是机器合成还是真人讲话尚有很大差距。

2.7 力触觉反馈技术

上述的关键技术已涵盖人-虚拟角色交互的整个环路，但是无论是用户的交互输入还是虚拟角色的反馈输出都完全虚拟，看得见摸不着，可能会发生用户的虚拟化身穿过虚拟角色身体造成临场感中断的现象。为了避免出现此类视觉穿透现象破坏人-虚拟角色互动体验，一方面可以通过巧妙的方式拉开人和虚拟角色之间的距离，但是遏制了视觉穿透可能性的同时也可能给交互带来距离感；另一方面则可以在纯虚拟的交互基础上，借助力触觉反馈技术为虚拟角色提供触觉支持，避免视觉穿透的同时对视觉、听觉双通道交互进行触觉通道的扩展。

力触觉反馈技术从实现机制上分为主动式和被动式两种。主动式力触觉反馈设备包括场景/桌面式、手持式和可穿戴式3种：场景/桌面式设备固定放置于桌面上或者立于交互场景中，通过电机驱动操纵杆或者线绳的方式来输出三维空间中的虚拟作用力，由于需要用户一直与设备接触以感受其产生的作用力输出，会损伤用户与虚拟角色互动时的沉浸

感，并且有限的工作范围也严重限制了人-虚拟角色交互的自由性，故不适合用于人和虚拟角色的互动体验；手持式设备，顾名思义需要用户时刻持于手中，通常以手柄、手持道具类设备出现，通过振动触觉、气动等技术模拟作用于手部的力，同样存在持续接触干扰交互自然性的问题；可穿戴式设备通常以触觉衣、臂带和手套等形式出现，利用振动、气动、肌肉电刺激、挤压、力矩操纵等技术模拟力触觉作用于人体的感受，目前高精度的触觉分布模拟的计算难关还未突破，只能对和虚拟角色的握手、拥抱等触觉交互进行较为粗糙的实现。

被动式力触觉反馈则是通过跟踪真实世界中一个和虚拟角色近似1:1对应的实物，并在其上精确地叠加虚拟角色，使用户在虚拟环境中看到虚拟角色的同时，在相同位置处触摸到与虚拟角色对应的实物。该方案利用真实物体本身的属性提供力触觉反馈，真实感更高，并且不存在持续接触的问题，更容易实现和虚拟角色的握手、拥抱等触觉交互，但是需要在真实世界中存在一个相似于虚拟角色的实物限制了该类技术应用的灵活性。

当前阶段，选用被动方案为静态的虚拟角色提供力触觉反馈更为自然真实。而随着传感驱动装置的小型化集成技术更加成熟，随着对触觉这一感官通道的研究更加深入，对触觉的模拟更加真实，使用被动方案或可穿戴式方案为动态的虚拟角色赋予物理性变得可行，其中被动方案的实现更加自然，接近真实世界中的交互；而可穿戴式方案更加灵活，可以方便调整为不同的虚拟角色提供力触觉反馈。

3 结束语

在虚拟环境中实现同虚拟角色的互动需要解决3个核心问题：如何实时精确地采集用户的多通道交互输入信号；如何建立虚拟角色的由交互刺激到反馈的“体现智能与情感的”映射；如何将虚拟角色的多通道反馈信号真实地呈现出来。就这3个核心问题，又需要解决一系列相关的关键技术。

对用户交互信号的采集是3个环节中最依赖于硬件设备发展的一环。目前，空间定位和动作捕捉领域已有较多相对成熟的技术与产品，更多在于针对应用特性选择适合的方案，光学和惯性相结合的空间定位与动作捕捉能提供较大范围内高精度的空间定位和实时性高、无惧遮挡的动作捕捉，是当前适合于人-虚拟角色互动应用的成熟且优秀的方案。眼动追踪方面，目前已有FOVE、七鑫易维、Tobii等几家公司完成了眼动追踪技术到头戴式显示器的集成。近场语音识别更是在借助深度学习以后识别准确率有了实质性提高，已经达到了初期的商业可用的阶段。综上所述，目前已有技术已经能够比较完整地实现用户交互信号的采集环节；而虚拟角色的由交互刺激到反馈的“体现智能与情感的”映射建立和多通道反馈信号的真实呈现，目前还处于研究阶段，需依靠认知心理学、人机交互、人工智能等技术的进一步发展。

虚拟角色的触觉支持是对上述各环节实现的增强，避免发生视觉穿透现象影响交互沉浸感。目前触觉的发展相对视觉和听觉还有很大差距，并没有一个完善的触觉解决方案，使用和虚拟角色近似1:1的实物来提供被动力触觉反馈只是一个权宜之计。还需待传感驱动装置的小型化集成更加成熟，或者对触觉这一感官通道的研究更加深入之后，才能利用主动/被动方案为动态的虚拟角色提供更加灵活、真实的触觉支持。

参考文献
[1] DUCHOWSKI A T. Eye Tracking Methodology: Theory and Practice[M]. London:Springer, 2003
[2] SIGUT J, SIDHA S A. Iris Center Corneal Reflection Method for Gaze Tracking Using Visible Light[J]. IEEE Transactions on Bio-Medical Engineering, 2011, 58(2):411.DOI: 10.1109/TBME.2010.2087330
[3] GAO Y, BARRETO A, ZHAI J, et al. Digital Filtering of Pupil Diameter Variations for the Detection of Stress in Computer Users[C]// Proceedings of the 11th World Multi-Conference on Systemics, Cybernetics and Informatics. USA:IEEE, 2007:30-35.DOI: 10.1109/TBME.2010.2087330
[4] ISHIMARU S, KAI K, KISE K, et al. In the Blink of An Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass[C]// Augmented Human International Conference. USA:ACM, 2014: 15. DOI: 10.1145/2582051.2582066
[5] PATNEY A, SALVI M, KIM J, et al. Towards Foveated Rendering for GazeTracked Virtual Reality[J]. ACM Transactions on Graphics, 2016, 35(6):179
[6] GUENTER B, FINCH M, DRUCKER S, et al. Foveated 3D Graphics[J]. ACM Transactions on Graphics, 2012, 31(6):164
[7] JUSLIN P N, SCHERER K R. Vocal Expression of Affect[J]. The New Handbook of Methods in Nonverbal Behavior Research, 2005: 65-135. DOI: 10.1093/acprof:oso/9780198529620.003.0003
[8] SCHERER K R. Vocal Affect Expression: A Review and A Model for Future research[J]. Psychological Bulletin, 1986, 99(2):143
[9] BADLER N I, WEBBER B L, BECKET W, et al. Planning and Parallel Transition Networks: Animation's New Frontiers[J]. Center for Human Modeling and Simulation, 1995: 91
[10] Google's DeepMind AI Just Taught Itself to Walk[EB/OL]. (2017-07-11)[2017-09-23]. http://www.businessinsider.com/google-deepmind-ai-artificial-intelligence-taught-itself-walk-2017-7
[11] WaveNet: A Generative Model for Raw Audio[EB/OL]. [2017-09-23].https://deepmind.com/blog/wavenet-generative-model-raw-audio/
[12] SHOJI M, MIURA K, KONNO A. U-Tsu-Shi-O-Mi: the Virtual Humanoid You Can Reach[C]// ACM SIGGRAPH 2006 Emerging technologies. USA:ACM, 2006: 34

作者简介

翁冬冬，北京理工大学光电学院副研究员；主要研究领域为虚拟现实、增强现实与新型人机交互技术；先后主持和参加国家级项目20余项；已发表被SCI/EI检索论文30余篇，获授权专利5项。

薛雅琼，北京理工大学在读硕士研究生；主要研究方向为虚拟现实与新型人机交互技术。