

# 从普适计算到人机境融合计算

史元春 喻 纯 石伟男  
清华大学

关键词：普适计算 人机融合 交互服务

## 引言

20世纪90年代初，马克·维瑟（Mark Weiser）提出未来的普适/泛在计算（Ubiquitous Computing，同义术语还有 Pervasive Computing）<sup>[1]</sup>将使计算脱离桌面，在掌上人机融合（移动）计算的基础上，建立起我们今天称之为“人机境三元融合”的交互环境：更多的计算和感知能力将遁形于物理世界，构成持续服务于用户的分布式系统。近年来，信息技术不断朝此趋势发展，移动计算、云计算、物联网、大数据和人工智能等技术支撑越来越多不断成型的普适计算应用场景，如智慧城市、智能交通、工业物联网、智能家居等，人机境三元融合的计算平台将成为人类社会经济活动不可或缺的基础设施。普适计算场景是由人机境异构资源之间的数据融合关系构成的，其应用开发面临任务、资源、组合关系、编程需要专业领域知识等多方面的挑战，并且需要体现以人为中心的计算服务能力，而人又是三元中难以定义、只能适应的一方，更为复杂，这些复杂性需要靠操作系统来管理和屏蔽。

## 人机境融合计算新模式

随着信息技术的不断发展，人类正在迈入一个万物智能互联的“人机境”三元融合时代。与“普适计算”时代相比，这一新时代注重物理空间、信息空间和社会空间的有机融合，在交互广度上，扩展了人机交互的时空边界，用户在丰富的人机境场景中访问信

息，交互具有适应性和连续性；在交互深度上，泛在计算传感丰富的情境和用户数据，让用户表达交互意图和获取信息反馈的方式更加简便、高效。

在这一新型环境中，传统的“用户主动寻找所需功能”的交互模式受应用数量、复杂性和交互方式多样性的影响，使用成本较高。在这种情况下，我们持续研究以“人”为中心的新型交互模式——“系统主动向用户推荐交互服务”。“交互服务”是与界面无关的计算功能的基本单元，它不受当前以应用为单位的框架限制，而是在操作系统的统一管理下，根据情境信息和用户交互行为，推理意图，提供多模态的信息反馈，为用户提供更自然的交互体验。

以“人”为中心的新型交互模式的实现，主要面临如下几个方面的挑战：

1. 人机境之间关系是动态变化的。人机境之间的关系通过“情境”来描述，包含用户交互时的状态、所处的物理环境和数字环境中的信息等内容，是交互服务推荐的基础和最主要的参考信息。然而，在人机境三元融合的环境中，物理环境布置、可用的传感器数量以及用户状态等的动态变换将产生无穷多种精细复杂的情境，这些情境无法通过单纯列举和针对性设计算法的方式完成识别。因此，人机境三元融合的交互系统需要具备自动发现和识别情境的能力，从而根据不同的情境推荐不同的交互服务。

2. 人机界面的形式是多样并存的。和传统使用场景中从功能逻辑直接映射的交互界面不同，人机境融合场景下的人机界面具有多模态、分布式的特点，并不存在统一的范式。由于人机境情境信

息的不同，指定的信息模态、需要的交互形式、可用的呈现设备、用户当前的状态等对展现的信息和可行的交互方式有不同程度的约束限制。对交互服务的开发者而言，针对不同交互情境分别进行交互界面的适配是无法完成的，因此交互服务开发者应该只专注于提供服务功能，而操作系统则要根据情境信息自动优化生成完整、高效、一致、社会可接受的交互服务界面。

3. 人机境环境中资源是分布异构的。人机境三元融合环境中的交互服务由分布异构的传感器、处理器和执行器等资源共同支持，实现需要完成的功能。这些资源之间的组合和协同关系复杂，并不适合在上层的交互服务应用中直接控制，而是需要操作系统定义合适的资源抽象层，屏蔽底层实现的复杂性，结合情境信息、系统状态、交互任务，为上层的交互服务提供统一的访问接口，实现多种资源的有效管理和调度。

应对以上挑战，我们提出了 NUIX (灵犀)

自然人机交互操作系统，它提供了交互情境的自发现和识别、人机交互界面的自动生成、异构交互资源的统一管理为核心能力，面向人机境三元融合环境下的交互管理，支撑以“人”为中心的交互服务推荐的新型交互模式。NUIX 克服了传统交互模式的局限性，为用户提供更加便捷、高效、个性化的交互体验。

## NUIX自然人机交互操作系统

人机交互在计算机操作系统的发展过程中发挥了关键作用。随着计算设备和系统的发展，从主机计算到个人计算，再到移动计算时代，人机交互方式从命令行界面发展到图形用户界面，再到自然交互界面，偏向机器的规范化、机械化的交互方式越来越少，偏向于人的“自然”行为的交互方式则越来越多。伴随自然交互而来的则是人机交互在操作系统中的占比和所需开发工作量



图1 NUIX自然交互操作系统框架

的逐步上升。据统计，用户界面开发占数百万移动应用开发工作量的60%以上，而这一比重在可预见的未来还会进一步提高。

考虑到人机交互的重要价值和应用端做好交互所需的高昂成本，NUIX的核心思想是在系统层对交互资源进行智能管理：从终端用户的角度，用户可以自如地表达交互需求，NUIX智能识别用户所处情境的信息，推理用户交互意图，并根据交互资源呈现信息；从交互服务开发者的角度，开发者无须考虑底层对用户情境、行为的感知以及界面的呈现等通用功能的实现，而是专注于自身的数据存储和业务逻辑，为用户提供更优质的数据服务。

NUIX的系统框架如图1所示。作为面向新型终端和人机境融合场景的交互系统，NUIX包含了交互资源管理、交互情境感知、交互意图推理、交互路径优化、自然人机接口等关键技术，为交互服务开发者提供开发平台，为终端用户提供更自然的交互体验。下面针对系统中的主要模块进行介绍。

## 传感层

传感层主要解决交互资源分布异构的挑战。在人机境三元融合环境中，分散着海量的传感资源，提供不同模态的传感数据以支撑上层应用对用户行为的全面感知，为实现高质量的人机交互提供基础。然而，对上层应用而言，在使用传感资源时面临着传感设备的发现与连接、不同传感资源间的数据传输同步、根据系统状态进行参数设置和功耗优化等一系列通用问题。要解决这些问题不仅需要巨大的工程开发量，也增加了上层应用对底层硬件的管理成本，加大了应用间的耦

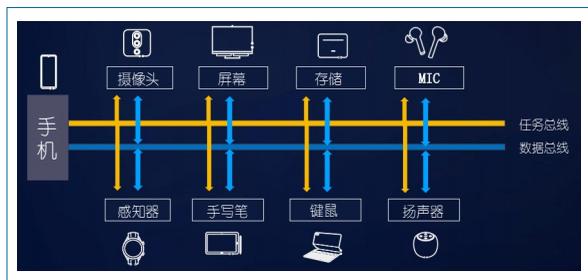


图2 鸿蒙分布式软总线示意

合关系，增加了资源冲突等情况出现的几率。

针对分散于人机境三元融合环境中的传感资源，未来交互系统的发展趋势是构建统一的交互资源的传感层，为环境中不同的传感资源提供统一的抽象接口，通过分布式共享总线完成对不同传感资源的调用，并由系统自动完成传感资源的调度与功耗优化。鸿蒙系统中的“分布式软总线”功能<sup>[7]</sup>（见图2）和泛在操作系统矽琇（XiUOS）<sup>[9]</sup>中的负责传感资源管理的“传感器框架”，都是在统一管理传感资源基础上的传感层的实现方案，它们均为设备间的无缝互联提供了统一的分布式通信能力，使得上层应用能够快速发现并连接设备，高效地传输任务和数据。

## 感知层

情境感知是人机境三元融合环境下的重要能力，用于建模人机境之间的动态变化关系，将传感层的多模态数据转化为可理解的“情境”语义，构建交互意图推理的基础。现有的情境感知与识别工作一般针对预设情境进行识别，情境数量有限（局限在几个到几十个），且扩展能力较差<sup>[8]</sup>，无法满足真实人机境三元融合环境下复杂多变的情境识别要求。为了实现真实场景下的情境感知，自动发现和识别交互情境是智能交互系统必须具备的能力，但目前的情境自动发现识别工作在自动发现的精度、个性化识别程度与实际部署性能等方面都有所不足<sup>[5,6]</sup>。

NUIX感知层根据用户多模态交互行为数据实现对交互情境的自动发现和识别，并在此基础上推理出用户的交互行为语义，为交互意图的识别提供基础信息。如图3所示，NUIX的主要创新在于：在用户日常使用过程中不断积累个性化数据，建立融合时空概率特性的智能特征提取算法，提取传感信号中与情境高度相关的特征，对特征向量进行聚类以实现新情境的自动发现；新情境发现后，在用户的自然交互过程中获取标注信息，扩展可识别情境的集合。当用户对新情境进行标注后，算法将该情境的历史数据全部视为已知数据，基于这些数据，

对不同用户训练个性化分类器，使自动发现的情境能被迅速、有效识别。

人机境三元融合环境中虽然有丰富的多模态传感资源，但相对于人体精细的行为表达而言其感知带宽仍然较窄，无法从数据中直接重构出用户的交互行为。在 NUIX 行为感知模块中，我们对用户交互行为表达能力建模，利用用户行为能力的先验知识指导传感器融合推理，通过挖掘数据内蕴交互行为与意图的可计算性，解决交互行为传感数据的稀

疏问题，并具有泛化能力强、计算存储效能高的优点。以头部动作追踪任务为例，我们提出的基于声场与惯性传感器（IMU）融合的头部追踪方法，在具有相对距离连续感知能力的基础上，偏航角追踪性能比之前的最好方法提升了 2 倍<sup>[10]</sup>。

### 认知层

推理用户交互意图是 NUIX 的核心，其中蕴含着系统对用户的认知，是交互服务推荐的基础。用

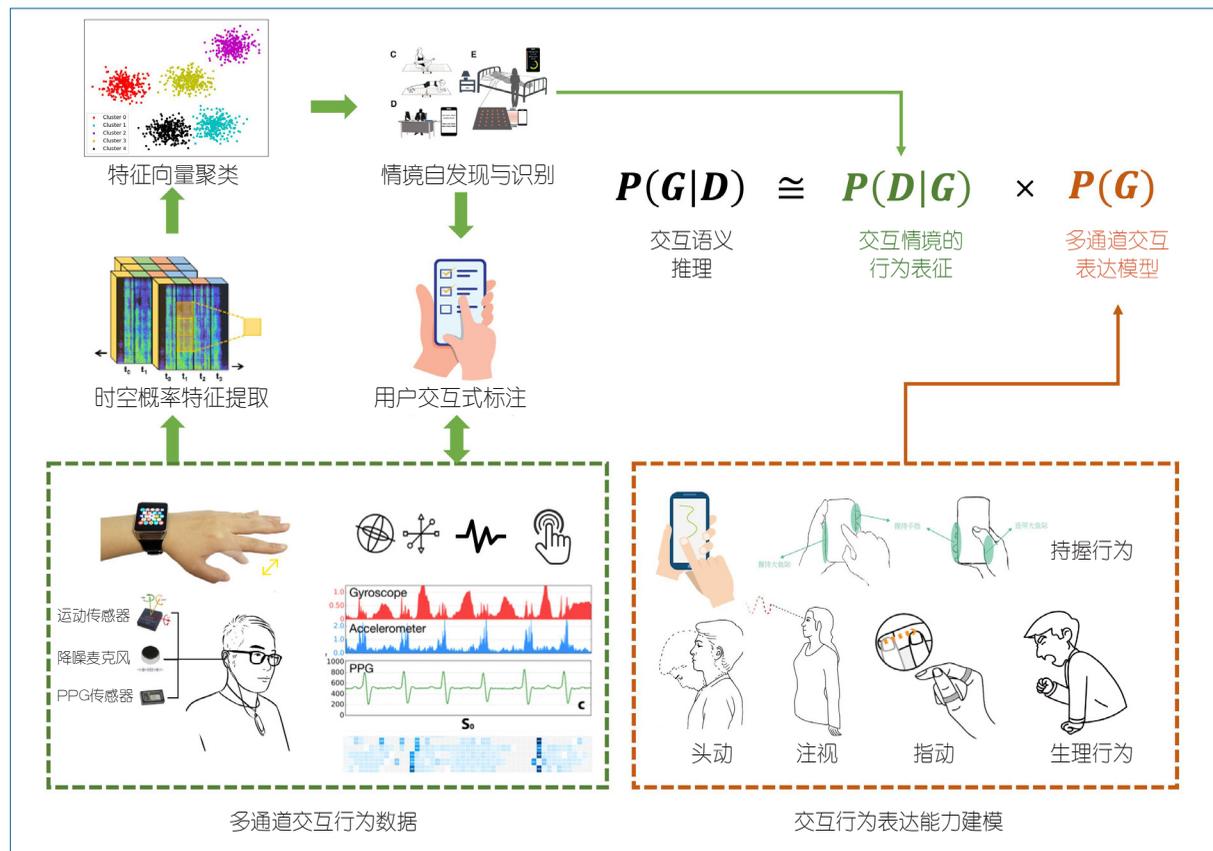


图3 交互情境自动发现和交互语义推理流程



图4 基于贝叶斯推理的自然交互理论与计算框架

户通过自然行为（包括动作、语言等）向系统表达交互意图。在人机境三元融合场景下，行为的模糊性和多样性给推理带来了巨大挑战：首先，用户的有意交互行为与无意行为混杂，在时间上缺少清晰的分界；其次，自然行为的表达受限于人对运动的控制能力，行为数据携带了由误差导致的噪声；最后，自然行为还受到情境、用户心理、用户习惯等复杂因素的影响，不同个体表达意图的方式具有差异性。

全面准确地理解“人”的行为规律是完成交互意图推理的基础，因此在人机交互领域有大量工作研究如何对人的运动控制能力进行建模，通过采集特定交互任务下人的行为传感数据，构建数据驱动的行为模型，提供交互意图推理的先验<sup>[3,4]</sup>。近年来，越来越多的研究人员意识到，人的心理模型在交互行为表达中占有重要的地位，是交互意图推理需要

考虑的重要因素。以芬兰 Aalto University 的工作<sup>[2]</sup>为代表，通过强化学习算法对用户交互决策机制进行建模，将人的认知行为决策白盒形式化为最优监督控制问题，能够较好地模拟出人在不同条件下的适应性行为。

基于自然动作、生理心理模型之间的相互作用机理，NUIX 在认知层构建了基于贝叶斯推理的交互意图推理计算框架。如图 4 所示，用户表达意图的过程以交互任务和交互情境为先验，在用户的心理生理模型的指导下产出自然的交互行为；对交互意图解析则是对这一前向表达过程所进行的逆向推理：以感知层提供的情境语义、用户的交互任务和生理心理模型作为先验知识，弥补人通过自然行为表达交互意图时的信息缺失，使用贝叶斯推理准确地解释用户的自然交互行为，推理用户的交互意图，从而为用户推荐合适的交互服务。该计算框架具有

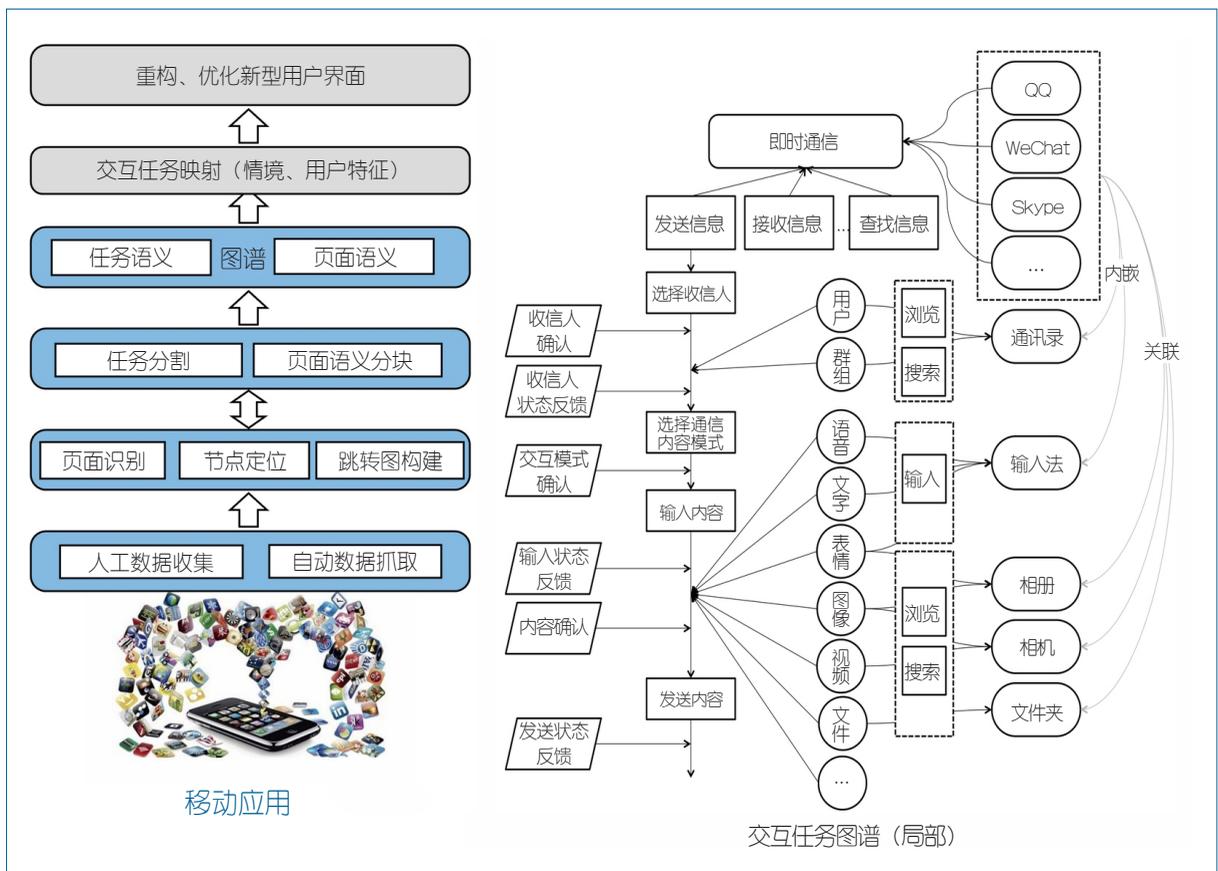


图5 交互服务知识图谱构建方法与示例

较强的通用性和可拓展性，且在不同交互任务和交互情境的意图推理中都取得了优于传统推理方法的效果<sup>[11, 12]</sup>。

## 服务层

交互服务是人机境三元融合环境下基本的计算和功能单元，是人机境环境下的“APP”，需要由交互操作系统统一管理。在丰富多样的情境下，终端用户对交互服务的偏好不同，不满足于预设的交互界面，期待系统能根据当前场景自适应生成自然高效的交互界面；对服务提供者和开发者而言，针对所有的情境进行交互界面的开发适配是不现实的。因此，在人机境三元融合场景下，交互服务仅由其本身的数据内容和业务逻辑描述，其交互界面基于当前情境动态生成，实现对交互服务的语义理解和对交互路径的优化。

通过充分解析服务之间的语义关联，可以帮助系统针对性地组织服务内容，完成符合用户需求的服务推荐和呈现。如图5所示，NUIX基于现有的移动应用，以自动化和人工标注相结合的方法解析出现有应用中蕴含的语义信息，对服务语义进行建模，构建出交互服务和交互任务的知识图谱。该知识图谱不仅覆盖海量的交互服务资源，也对交互服务推荐和用户交互界面生成提供语义基础。

为了在交互界面上实现自然高效的交互目标，交互界面需要为用户提供最优的交互路径。交互路径是用户完成交互目标所需的步骤过程，是人机境三元融合环境中信息交换的方式，其长度决定了信息交换的效率和质量。基于情境感知数据，根据用户即时信道状态构建约束条件，确定可行的交互路径集合，求解用户信息交换过程所需的最小信息量，通过剪枝生成优化的交互路径，可以提升信息交换效率、逼近理论极限，兼顾交互的自然性与高效性。

在交互服务知识图谱和路径优化的基础上，结合情境信息和交互任务信息，服务层可以完成对海量服务的管理，充分调动人机境三元融合环境中的多种智能终端设备，完成高质量的服务界面信息呈

现。人对信息的感知是多通道的，而人机境融合环境中的信息来源丰富、形式多样，人对交互信息的接受情况受多方面因素制约影响。在NUIX中生成服务界面时，需要结合人在交互时的生理心理机制，量化人的多模态信息感知能力，同时利用场景中可用的多设备构建多通道输出能力，联合优化出情境适应性强、可理解性强的多模态呈现界面。

## NUIX Studio测试研发平台

人机境三元融合场景具有多变性，每个场景拥有不同的软硬件配置、传感与显示设备、用户群体，广大开发者针对个例场景进行定制开发，不仅配置场景的人力时间成本极高，同时产生大量重复的工程实现，导致人力和资源的严重浪费。NUIX Studio是NUIX系统配套的研发平台，该平台利用虚拟现实环境模拟不同交互设备的能力，提供面向人机境三元融合场景的交互服务集成开发测试的平台和API。

NUIX Studio平台的功能技术目标：(1) 易于使用和测试。由于该平台运行于虚实融合的环境中，用户需要与不同种类的交互资源进行交互，因此需要提供一种通用自然的交互方式：对于虚拟设备，平台需要模拟完整的交互响应机制；对于现实世界的设备，平台需要实现虚拟世界和现实世界之间的数据和交互方式的互联互通。(2) 具有容错能力。平台应能有效地处理来自真实世界和虚拟世界的故障，无须开发者自行调试交互设备的功能性。(3) 具有可扩展性。平台应能支持在系统中灵活地增加修改设备、场景、用户的数量和种类等，且在性能上应保持可接受。(4) 支持多用户协同。与现实世界一样，当多用户同时与NUIX系统交互时，平台上每个正在运行的交互场景与设备实例都应该保持数据和响应上的同步一致。

针对以上目标，平台利用混合现实方法，融合虚拟终端、模拟互联设备，与真实场景、真实用户交互，在确保交互场景有效性与可用性的前提下，极大降低了开发成本；同时，利用众包技术向开发者提供将场景、终端、交互技术集成到

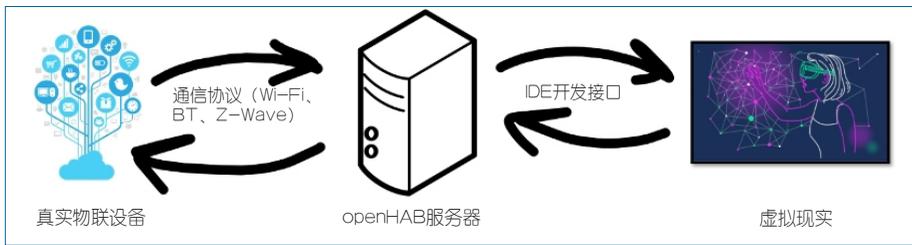


图6 NUIX Studio的基本结构

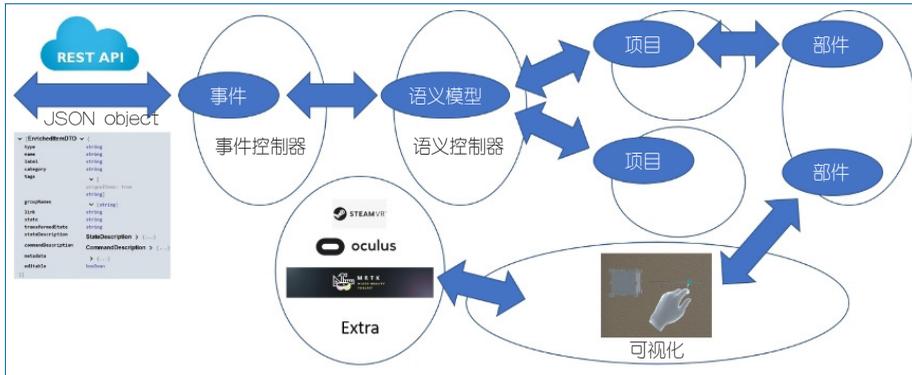


图7 NUIX Studio支持的物联应用架构

以将他们的交互服务技术集成到 NUIX 系统中验证交互效果。NUIX Studio 支持的物联应用架构如图 7 所示。

## NUIX赋能的新应用

NUIX 是一个综合平台，不仅包括智能人机交互的功能抽象，也可指导自然人机交互的集成和部署。目前，NUIX 已在安卓（包括鸿蒙）平台上得到了具体实现。该平台通过构

建系统级服务程序，能够实时处理智能手机的传感器数据，推断用户所处的空间位置、行为活动等情境信息，并支持基于多模态自然交互动作的感知和意图识别。

系统的方法，逐步丰富开发接口与资源，实现开发过程的高效复用。为保证开发测试的有效性以及交互服务的可用性，系统做了两方面的创新：一是结合三维扫描和重建技术，在虚拟世界中渲染还原现实场景中的视触觉观感，最大程度上保证用户行为的一致性；二是打通虚拟设备与真实设备的连接壁垒，两者共享输入输出，保证交互过程的可实现性。此外，系统针对多人多场景的灵活性需求，实现场景的同步控制，解决了复杂场景多人异地协同开发的难题；提供多层次可定制的开发接口支持，同时集成了包括注视、手势在内的自然交互技术接口，极大地降低了人机境三元融合环境下交互服务开发的难度和成本。

如图 6 所示，NUIX Studio 平台的体系结构有三层，其中 NUIX 中的传感层与现实世界中的设备相连接，从现实世界的设备中收集传感数据；openHAB 服务器所在的集成层分析来自虚拟现实和现实世界的的数据，将二者之间的数据进行同步集成；虚拟现实所在的可视化层通过模拟触摸、视觉、手势和其它交互执行与数字物联网（IoT）设备的交互，该层为交互服务的开发人员提供 API，

安卓版 NUIX 的服务层对应于自研的交互代理模块，维护移动应用界面交互语义，并通过 RPA（机器人流程自动化）模式实现交互路径的操作。该系统已经支持了拍拍助手和慧说读屏两款安卓应用的运行，分别针对适老化和信息无障碍两个场景提供自然交互手段。此外，NUIX 中的关键模块也可单独提取出来应用。例如，荣耀在 2022 年 11 月份发布的气息唤醒技术就直接使用了 NUIX 中基于风噪特征的近距离说话检测技术，在智能手机上首次实现了基于自然交互动作的唤醒语音交互技术。

可视化

## 总结

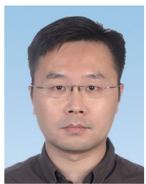
人机境三元融合给自然的人机交互带来了新的机遇，同时也带来了新的挑战。在人机境三元融合时代，用户主动寻找服务已不再适用，以“人”为

中心的交互服务推荐将成为新的交互模式。在交互变得越来越重要、越来越复杂的情况下，单独的应用服务已无法完成对交互相关资源的处理，应转变为由操作系统统一管理的模式。NUIX 自然人机交互操作系统围绕交互情境的自发现和识别、人机交互界面的自动生成、异构交互资源的统一管理 etc 核心能力构建，基于对人机境三元融合环境下交互资源的管理，为用户提供自然高效的交互体验，也为服务提供者提供低成本、高质量的服务开发体验。



史元春

CCF 会士, CCF 人机交互专委会主任, CCF 夏培肃奖获得者。清华大学教授、人工智能研究院智能人机交互中心主任, 青海大学校长 (清华大学对口支援岗位)。主要研究方向为人机交互、普适计算等。  
shiyu@tsinghua.edu.cn



喻 纯

CCF 杰出会员, CCF 人机交互专委会秘书长。清华大学计算机系副教授, 国家级青年人才项目获得者。主要研究方向为智能交互、人机协同。  
chunyu@tsinghua.edu.cn



石伟男

CCF 专业会员。清华大学计算机系助理研究员。主要研究方向为智能交互。  
swn@tsinghua.edu.cn

## 参考文献

- [1] Weiser M. The computer for the 21st century[J]. *ACM SIGMOBILE Mobile Computing and Communications Review*. 1999, 3(3): 3-11.
- [2] Oulasvirta A, Jokinen J P P, Howes A. Computational rationality as a theory of interaction[C]//*Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022: 1-14.
- [3] Yi X, Yu C, Shi W, et al. Is it too small?: Investigating the performances and preferences of users when typing on tiny QWERTY keyboards[J]. *International Journal of Human-Computer Studies*. 2017, 106: 44-62.
- [4] Goel M, Findlater L, Wobbrock J. WalkType: using accelerometer data to accommodate situational impairments in mobile touch screen text entry[C]//*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012: 2687-2696.
- [5] Xu X, Gong J, Brum C, et al. Enabling Hand Gesture Customization on Wrist-Worn Devices[C]//*Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, 2022, Article 496: 1-19.
- [6] Wu J, Harrison C, Bigham J P, et al. Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition[C]// CHI '20: CHI Conference on Human Factors in Computing Systems. 2020:1-14.
- [7] HarmonyOS. 技术特性[OL]. 2022.12.16. <https://developer.harmonyos.com/cn/docs/documentation/doc-guides/harmonyos-features-0000000000011907>.
- [8] Vaizman Y, Weibel N, Lanckriet G. Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 1(4): 1-22.
- [9] Cao D G, Xue D L, Ma Z Y, et al. XiUOS: an open-source ubiquitous operating system for industrial Internet of Things (opens new window)[J]. *Sci China Inf Sci*, 2022, 65(1): 117101, <https://doi.org/10.1007/s11432-021-3294-y>.
- [10] Wang Y, Ding J, Chatterjee I, et al. FaceOri: Tracking Head Position and Orientation Using Ultrasonic Ranging on Earphones[C]//*Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022: 1-12.
- [11] Wang Y, Yu A, Yi X, et al. Facilitating text entry on smartphones with QWERTY keyboard for users with Parkinson ' s disease[C]//*Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021: 1-12.
- [12] Yan Y, Yu C, Ma X, et al. Eyes-free target acquisition in interaction space around the body for virtual reality[C]//*Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018: 1-13.