arXiv:2601.09755v1 [cs.NE] 13 Jan 2026

面向实时机器人技术的异构计算平台。

Jakub Fil[1]* Yulia Sandamirskaya[2] HectorGonzalez[3] LoïcAzzalin[2]Stefan Glüge[2] Lukas Friedenstab[1] Friedrich Wolf[1] Tim Rosmeisl[3]Matthias Lohr mann[3] Mahmoud Akl[3] Khaleel Khan[4] LeonieWolf[1]Kristin Richter[1] Holm Puder[1] Mazhar Ali Bari[1] Xuan Choo[5]Noha Alharthi[6] Michael Hopkins[1] MansoorHanif Christian Mayr[4]Jens Struckmeier[1] Steve Furber[7]

2026年1月16日

摘要

工业4.0已实现机械设备（运营技术）与软件（信息技术）及互联网的紧密集成，构建起由传感器、数据和算法组成的网络，以实现高效可靠的生产。如今，社会5.0这一新理念正在兴起，城市基础设施将配备各种设施，以提高可靠性、效率和安全性。机器人技术将在实现这一愿景的过程中发挥关键作用；这一愿景由NEOM倡议开创，旨在打造一座人与机器人共同生活的智慧城市。本文探讨实现这一愿景所需的计算平台。我们展示了如何将神经形态计算硬件（以搭配事件相机使用的Loihi2处理器为例）用于感知和实时感知交互，并将其与本地人工智能计算集群（图形处理器）结合，以实现高级语言处理、认知和任务规划。我们通过一项交互任务展示了这种混合计算架构的应用：人形机器人与人类共同演奏乐器。我们的设计核心在于高效、无缝地集成不同组件，确保软硬件协同作用，从而最大限度地提升整体性能和响应能力。我们提出的系统架构凸显了异构计算架构在推动机器人自主性和交互智能方面的潜力，也预示着在复杂实时应用中，此类集成系统将成为常态。

1 引言

在不久的将来，认知城市将成为现实。它们将经过精心设计以实现可持续性，由可再生能源供电，并以数字景观为特征，机器人与人类共存其中 [Golubchikov and Thornbush, 2020]。这种无缝人机交互的愿景需要一种独特的解决方案，利用节能且低延迟的计算能力分析海量数据。本项目旨在研究将脑启发硬件与高密度图形处理器（GPU）解决方案相结合，用于实时社交机器人技术。本文提出结合两种脑启发方法：用于大规模脑建模的 SpiNNaker2 [Gonzalez et al., 2024]，以及英特尔的研究芯片 Loihi2 [Orchard et al., 2021]，后者用于边缘端的实时感知与驱动。此外，我们将动态视觉传感器（DVS）摄像头 [Lichtsteiner et al., 2008]与英特尔的 Loihi2 芯片相结合，并训练了一种用于跟踪人手的算法——这是首个部署在该系统上的实时机器人视觉工作负载。我们计划进一步扩展这些神经形态计算方法的能力，加入用于促进演示编排、运行大语言模型及其他深度学习算法的 NVIDIA DGX图形处理器系统，同时执行 Spaun [Eliasmith, 2013, Choo, 2018], 的专用版本。Spaun 使用脑启发计算机制实现认知架构。将这些要素、人形机器人和功能性脑模型 Spaun 相结合，可对传感器流进行实时处理，从而完成复杂的记忆任务。这种多层面方法支持实时语音处理和高级动作预测；对于控制一台负责演奏乐器的人形机器人而言，这些能力至关重要，并旨在创建一个面向智能通用机器人的交互式演示

*通讯作者–jakub.fil@waiys.com， [1]WAIYS GmbH， [2]苏黎世应用科学大学， [3]SpiNNcloud Systems GmbH， [4]德累斯顿工业大学， [5]应用脑研究， [6]NEOM， [7]曼彻斯特大学。

（GPR）。该项目包括实现一个交互式演示器，重点展现自然人机通信，这是具备人工智能能力的人形 GPR 的首项功能。为实现既定目标，我们还设计了一座高度专业化的可持续超级计算微型数据中心，配备创新的直接热液冷却系统，使其能够在定制化环境中为实时自主系统提供类似大脑的能力。

实现实时人工智能认知平台对于管理未来认知城市中的车辆、无人机、机器人及其他认知解决方案至关重要。除了处理海量数据外，这些平台还必须支持不同系统与设备之间的无缝通信。此类认知基础设施的运行需要能够适应动态城市环境的高级算法，从而实现预测性维护、快速异常检测和高效资源分配 [Shenavarmasouleh et al., 2022]。长期以来，人们一直认为，此类解决方案需要一种将大规模深度学习与符号推理和结构化知识相结合的混合架构，这也支持了这样一种观点：有效的人工智能需要多样化的异构系统[Garcez and Lamb, 2023年, Marcus, 2020]。本文提出了一种协作方法，将神经形态计算、深度学习和符号人工智能相结合，构建一个实时、低延迟且节能的认知人工智能平台。为展示如何将这些要素付诸实践，我们提出了一个创新用例，演示机器人与演奏特雷门琴音乐序列的人类之间的实时交互。

我们相信，认知城市和机器人学的未来依赖于异构计算系统的集成，以满足多样化的处理需求。高密度图形处理器负责较长时间尺度上的计算密集型任务，而边缘端节能的神经形态硬件则擅长实时、低延迟的决策。通过使用类脑算法，这种组合能够实现性能优化，同时降低能源消耗。此外，我们认为还必须考虑冷却和热再利用等其他基础设施要素，以实现智慧城市的可持续运行。我们的创新型直接热液冷却技术能够与这些系统相辅相成，增强热管理，并通过高效热回收进一步降低功耗 [Struckmeier, 2018]。这些技术共同构成了一个可持续、节能的平台，为未来互联城市和智能机器人所需的先进能力提供支持。

2 个结果

2.1 展示概述

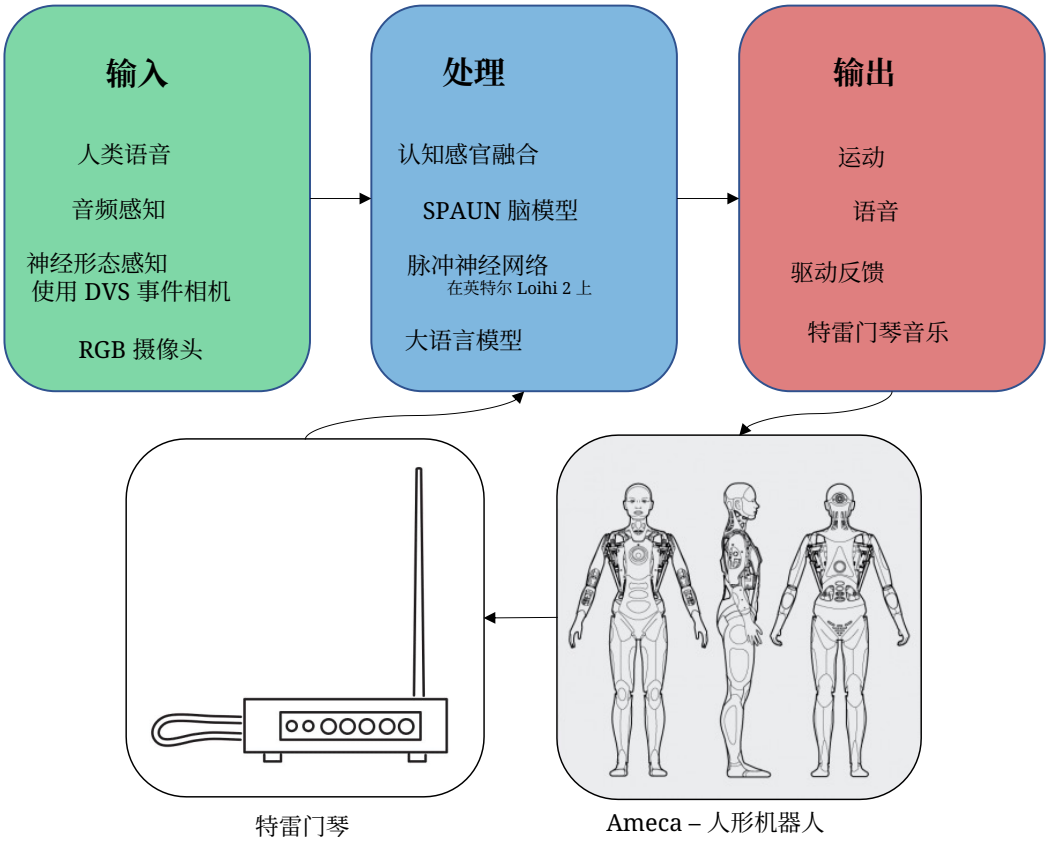

图 1：互动音乐展示关键组件概览。

机器人正日益融入我们的日常环境，而其有效沟通的能力也成为一个至关重要的考量因素。音乐无疑是最有力的沟通方式之一，常被誉为一种普遍语言，能够促进人类之间的连接 [Mehr et al., 2019]。共同演奏音乐需要许多与现实世界中具身沟通相似的技能：将长期记忆和规划与即时协调、轮流以及多模态信号交换结合起来。在此背景下，我们以演奏乐器作为示范性任务。我们教授人形机器人 Ameca 使用特雷门琴演奏乐曲；这是一种无需身体接触即可演奏的电子乐器。通过这一演示，我们探索了在基于 DGX 的集群环境中、利用多个图形处理器运行 Spaun 2.0 脑模型的能力。此外，我们还展示了英特尔研究芯片 Loihi2 的实时神经形态计算能力。重要的是，这项工作首次展示了在计算集群上运行的认知模型与神经形态平台实时工作负载的集成，从而为先进的实时视觉系统奠定了基础，使机器人能够持续感知并与其环境互动，同时使用大型世界模型或知识库。

此用例的核心围绕 Ameca 展开。这是一款由 Engineered Arts 打造的人形机器人，代表着人机交互研究领域最先进的平台之一。其设计融入了精密的面部机构，配备 32 个电机，专门用于实现细腻且富有表现力的面部表情模仿 [Singh et al., 2022]。这一能力是机器人传达多种情绪的关键。通过提供类似人类交流中的非语言线索，这对于让人机互动更加自然至关重要 [Paulus et al., 2013]。例如，Breazeal [2003] 指出，为机器人配备动态且富有表现力的面部特征至关重要，不仅有助于传达清晰的情感信号，也能减轻恐怖谷效应 [Mori et al., 2012], 从而促进更加自然的人机交互。

此外，Ameca 配备了灵活灵巧的手臂，可执行复杂的操作任务。在本研究中，机器人被设置为模仿人类参照对象通过特雷门琴演奏出的音符。研究采用传统的音符预处理技术，并结合先进的深度神经网络（DNN）模型，以确保准确识别声音。此外，集成 Spaun 2.0 系统后，机器人得以执行认知记忆任务，这些任务会影响机器人的决策过程以及系统的状态转换。Spaun 2.0 是世界上最大的功能性脑模型，包含多达 650 万个神经元；我们通过 NVIDIA DGX A100 系统部署了该模型。研究还使用神经形态芯片 Loihi2 快速处理设备端的视觉信息，从而使机器人能够与同步演奏的人类实时保持一致。

本项目的总体目标是提出一种受大脑启发的处理层级，将大规模神经网络模型的认知与内存能力，同 Loihi2 等神经形态硬件在边缘端的能效和实时处理能力结合起来。这种前所未有的组合，将云端和边缘端的类脑系统结合在一起，有望实现认知城市所需的许多复杂应用场景，同时也有助于从整体上降低大规模机器人应用场景的能源消耗。

2.2 音乐自动机与音乐演奏机器人

人工智能领域和整个社会长期以来都对音乐演奏机器人抱有兴趣。早期的例子包括伊斯梅尔·贾扎里——一位12世纪的穆斯林工程学与控制论先驱 [Dirik, 2020]。他制造了一台音乐自动机，形似一艘船，上面有四名自动演奏者，可以演奏预先编排的音乐序列。近年来，人们也尝试制造能够演奏乐器的机器人，例如钢琴 [Roads, 1986], 小提琴[Park et al., 2015], 或打击乐器 [Weinberg and Driscoll, 2006]。与这些自动机不同，真正的音乐演奏不仅包括忠实地再现音符和旋律，还包括演奏的力度变化，使之能与其他演奏者相协调，或与演奏者个人情绪的起伏相呼应，或体现独奏表演中乐器的发声过程。这是一个动态的互动过程。

特雷门琴被认为是最难演奏的乐器之一，是一种无需身体接触即可演奏音乐的电子乐器（见图 2）。特雷门琴通过检测演奏者双手置于两根天线之间所产生的干扰来发声。右侧的天线影响声音的音高，而左侧的天线控制音量。演奏特雷门琴需要根据产生的声音不断调整双手的位置。

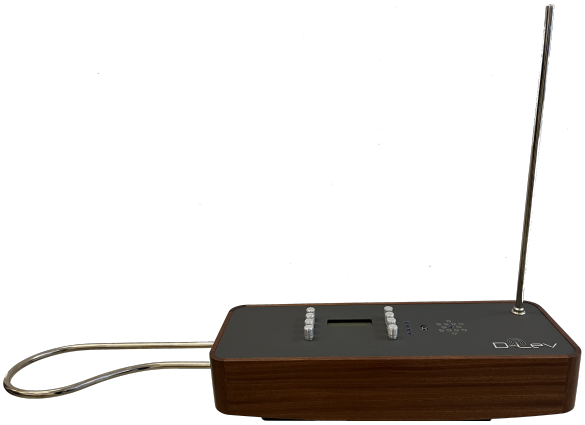

图 2：本研究中使用的乐器——特雷门琴。

此前已有一些研究致力于让机器人使用特雷门琴演奏音乐，例如 iCub [Mizumoto 等人, 2009] 以及 HRP-2 Promet [Wu 等人。2011]这些研究人员基于特雷门琴的音高，开发了前馈和反馈式机械臂控制方法，使机器人能够以独奏形式与特雷门琴互动。由于缺少可活动的手指，只能演奏非常简单的旋律，而且这些机器人没有采用专业认可的人类演奏技巧，例如 Carolina Eyck 提出的那种技巧。

与前文所述的方法不同，之前的方法仅关注音高波动，而忽略了由音量天线控制的扩展动态范围；我们训练 Ameca机器人使用自己的双手控制特雷门琴，并利用大脑模拟器做出决策。关于这一点，我们将在本文接下来的部分中进行更详细的介绍（见第 4.2.1 节）。这些先前的方法还忽略了类似人类的运动，例如颤音；在颤音中，细微的频率变化会在基音之上产生谐波耦合音。之前的方法没有包含由机器人自身执行的自动高级校准，而这对于自主系统来说是必不可少的，尤其是对于容易受到湿度和房间几何结构等外部因素影响的乐器而言。最后但同样重要的是，这些机器人都没有配备逼真的面部表情，无法配合演奏过程并提高观众参与度。事实上，这些方法都不是为独立运行而设计的，也无法与公众互动。此外，本文所述的方法

与之前的方法不同，涉及机器人与用户之间的协作演奏。我们的系统将音符和音量转换为双手的运动。后端软件向机器人提供需要演奏的音符及其持续时间，以控制右手。有关音量的信息则用于控制左手。系统会在适当的状态下向机器人提供一组音符和一组音量值，从而使机器人能够演奏长篇且复杂的乐曲。[1] [2]

该机器人所使用的训练数据集包含特雷门琴演奏大师 Carolina Eyck 演奏的音乐作品。对于每一首音乐作品，数据集都包含使用 ZED2 立体相机录制的视频、来自 iniVation 动态视觉传感器（DVS）的数据以及音频。该数据集旨在生成可在将旋律转换为机器人运动的过程中用作参考的数据。通过 ZED2 相机捕获的图像用于执行身体跟踪。为此，我们使用了由 ZED2 相机的制造商 Stereolabs 提供的身体跟踪 API 中的动态姿态估计算法。该系统能够按照 COCO18 标准检测人体上的 18 个关键点，从而准确预测人体四肢的位置和运动。此外，我们还使用了 Stereolabs 的物体检测 API 来检测目标物体，进一步提升录制数据的质量。随后，我们使用生成的身体姿态录制数据训练机器人执行相同的运动，从而复现特雷门琴演奏者的动作。为此，我们使用了 Ameca 专用的操作系统 Tritium，该系统可根据参考录制数据为机器人编程运动序列。

除了音乐表演之外，得益于部署在 GPU 集群上的语音与语言模型，机器人还可以在舞台上与人以及观众交谈。Ameca 可以围绕演奏特雷门琴展开对话，也可以讨论其语言模型能力范围内的任何其他话题。通过与机器人互动，用户可以要求机器人进行独奏表演，也可以启动教学模式或二重奏模式。在这些模式下，用户可以使用显示在第二台特雷门琴前方平板电脑上的图形用户界面。该界面显示从 C4 到 C5 的八个音符，用户可以参与互动游戏，为 Ameca 演奏的歌曲伴奏。

3 讨论

3.1 面向实时机器人技术的异构计算

社交机器人技术的发展需要一种计算范式，以协调高速信号处理与控制和较慢的长期认知推理之间相互冲突的需求。在此背景下，将神经形态处理器与高密度 GPU 架构相结合的异构计算平台，可能是一种很有前景的解决方案。此类系统对于实现无缝的人机交互至关重要，因为它们能够应对从即时感官反馈到复杂记忆任务的多样化计算需求。

我们的工作有一项关键发现：软件算法与硬件架构之间的协调至关重要。将仿脑硬件（如 Loihi2）与传统高性能 GPU 集群相结合，表明匹配良好的软硬件接口不仅有益，而且是提升效率的关键。这种协调能够最大限度地降低反馈回路中的延迟；当机器人必须处理传感器输入、执行决策算法并实时做出反应时，这一点至关重要。机器人会在多个时间尺度上运行，从快速的感知—行动循环，到更长期的规划与学习；正如 Zhao 和 Sentis 所展示的那样 [2019]，各处理单元之间的交互必须确保没有任何单一组件成为瓶颈。这种异构计算方法对工业界以及更广泛的神经形态计算和机器人学领域都具有重要意义。这些群体日益青睐既能提供先进技术，又具备可扩展性和能效的解决方案。稳健的异构计算平台能够降低功耗和运营成本，同时提升系统在动态真实环境中的性能，从而满足市场需求。此外，社区推动的硬件接口和软件协议标准化工作，正为未来奠定基础，使不同系统能够有效通信，进一步降低开发风险并促进创新。

不可或缺这一愿景的核心是开发通用通信协议，以协调
交互多种设备与计算节点之间的连接。与传输无关的协议 ol

[1]训练过程：https://youtube.com/shorts/bj6NaeokJbA [2]最终视频演示：https://youtu.be/OYW9qQl93aU

通用通信—地址事件表示（AER），由 Rast 等人提出[2015], ，可作为借鉴，通过利用基于事件的数据编码，构想一种面向未来系统的低开销、节能标准。这一点对于城市机器人学尤为重要，因为车队管理既要求实时协同决策，也要求严格的能效。尽管人们曾多次尝试连接事件型传感器与计算单元（例如 Pedersen 等人也曾进行相关尝试 [2024], ），但迄今尚无协议成为标准，即使相关工作正致力于推动其从学术界走向工程应用。随着脉冲神经网络和神经形态硬件日益进入商业领域，亟需一种易于使用的标准，以连接基于事件的设备与更传统的网络和计算资源。鉴于 AER 的灵活性及其过往的成功，它很可能成为底层高级协议；其软件和硬件实现的速度与可靠性，将是任何异构系统的关键。稳健的协议至少必须满足两项要求：1）针对非常近距离、高带宽的传感器到计算单元的连接，以最小开销实现脉冲的原始异步传输，不附带时间戳；2）针对易受数据损坏、延迟及其他不确定因素影响的较远距离连接，采用更安全、更审慎的传输方式，并包含参考标头、时间戳，以及可能需要的有效载荷和纠错机制。

展望未来，建立统一的表征方式和通用通信协议，对于社交机器人技术的发展至关重要。这将确保随着机器人逐渐成为认知城市不可或缺的组成部分，不同系统能够“使用同一种语言”，促进高密度基于云的模型与基于边缘的神经形态处理器之间无缝协作，从而实现更完善的系统集成，并打造更具韧性和适应能力的机器人基础设施，使其能够随着城市环境日益复杂的需求不断演进。因此，通信协议的标准化有望成为关键一步，推动实现这样的未来：集成认知系统将构成智慧可持续城市的支柱。

3.2 异构计算：智慧城市的未来

未来的城市景观将由集成的异构计算系统塑造，这些系统能够弥合高性能数据分析与实时、节能决策之间的差距。随着智慧城市不断发展，它们将越来越依赖混合架构，将高密度 GPU 集群强大的计算能力与神经形态处理器快速、低延迟的响应能力结合起来。这种双重策略能够同时处理用于战略规划的大型模型和历史数据集，以及对动态城市管理至关重要的即时传感器数据。

许多学者对智慧城市的基础要素进行了探索。例如， [Caragliu 等人、2011] 以及 [Albino 等人，2015] 提出了全面的框架，强调可扩展性、韧性，以及将先进技术融入城市基础设施的变革潜力。他们的见解凸显了这样一种必要性：系统不仅要管理日常城市功能，还要预见未来挑战并对此进行适应。

在这些基础性见解的基础上可以清楚地看出，未来城市环境的多层面挑战将要求采用能够覆盖广泛运行时间尺度和决策过程的异构计算架构。例如，对于实时交通管理、环境监测和应急响应至关重要的边缘传感数据即时处理，必须与支撑城市规划的长期分析相结合。此外，有效的自主车队管理，包括服务机器人、配送无人机和自动驾驶车辆，依赖于能够快速响应动态变化并持续学习以优化性能的系统。然而，传统深度学习模型往往难以实现持续学习，并且经常遭受灾难性遗忘，正如 Liu [2017]所指出的。一种有前景的替代方案是采用脑启发方法，例如 BitBrain [Hopkins 等人， 2023], 其天生能够从少量数据中快速学习，并根据需要在一段时间内保留、调整或遗忘这些知识。与此同时，在公共安全、医疗保健和市政服务领域促进有效的人机协作，需要将上下文感知响应与预测性决策能力相融合的计算框架。这些用例的多样性凸显了采用灵活且能够适应智慧城市不断演变的复杂性的异构计算解决方案的重要性 [Zanella 等人，2014年，Batty， 2012]。

此外，热管理领域的创新——例如直接热液冷却——不仅优化了集中式和分布式计算资源的热性能，也为创新性的城市能源循环利用铺平了道路。例如，先进数据中心产生的多余热量可以输送到区域供热网络，为住宅建筑、公共设施乃至城市温室供暖。在一些开创性项目中，余热已被用于预热

生活用水，或为机器人群和自动驾驶车辆提供节能冷却，从而同时降低能源消耗和运营成本。高性能计算与可持续能源再利用之间的协同，是设计未来智慧城市的基石（见第 4.1.4 节）。

总而言之，我们认为，异构计算架构、标准化通信协议与先进能源管理解决方案的融合，为未来智慧城市的发展提供了稳健蓝图。这种一体化方法不仅能帮助城市规划者优化资源分配、增强基础设施韧性，还能让智能机器人更高效地运行，从而确保未来的城市既智慧又可持续。

3.3 脑启发方法与机器学习相结合

为了增强机器人的决策子系统，我们采用了一种脑启发方法——Spaun 2.0（语义指针架构：统一网络）。这是一个由 650 万个神经元组成的大规模认知神经模型 [Choo, 2018]。该模型用于认知记忆任务和机器人的控制过程，因此 Ameca 能够通过一组预定义场景与用户互动，例如独自演奏、教授乐器演奏方法、交谈等。在本项目中，我们使用 Nengo 将 Spaun 模型部署在图形处理器系统上。Nengo 是一个开源神经模拟平台，可帮助研究人员在传统计算机和神经形态硬件上构建、模拟并部署大规模认知模型 [Bekolay 等人,2014]。展望未来，我们设想可用一台配备 5 个 SpiNNaker2 模块的服务器替代 Spaun 模型的执行，从而进一步降低能源足迹。有关我们对该模型的使用及其基于 GPU 的实现的更多详情，请参见第 4.2.1 节。

尽管 Nengo 和 Spaun 由不同模块和计算范式之间复杂而巧妙的交互构成，这些模块和计算范式受到脑连接和群体编码的启发，但其大部分表征能力源自一种向量符号架构（VSA）类型的选择，而这种架构支撑着 Spaun 的运行。向量符号架构所提供的可能性范围确实远远超出这一种选择；人们或许并未充分认识到这些理念有多么灵活 [Kleyko 等，2022年、2023年，Schlegel 等， 2021] 也未充分认识到它们能够多么有效地应用于与社交机器人技术相关的实际机器学习和人工智能问题，例如 Renner 等人提出的视觉场景识别 [2024] 和分类（如 Ge 和 Parhi 所述） [2020], ，以及许多其他认知架构，例如：ACT-R [Ritter 等人， 2019], SOAR [Laird， 2019], 自适应共振理论[Grossberg， 2013],或 DAC [Moulin-Frier 等人， 2017]。

除了具备理想的鲁棒性水平和对“绑定问题”的可行解决方案——如果人工智能系统要与生物智能竞争，这两点都是其面临的关键问题——它们可以说也非常适合事件型传感器和计算资源，因此应能充分实现神经形态计算在延迟和能耗方面的优势。尽管目前尚未通过演示得到证明，但针对精心选择的一组 VSA 机制进行直接、稀疏、二进制、事件驱动的实现，其速度应显著快于使用 Nengo 的 Spaun 实现，并且能效更高；后者采用了大规模率编码神经元群以及相关的编码/解码计算。这是因为，直接受树突计算原理启发的、极其简单且高度并行的局部稳态机制和结构可塑性机制 [Ahmad and Hawkins, 2016, Branco and Häusser, 2010, Kastellakis et al., 2015, Larkum and Nevian, 2008, London and Häusser, 2005] 已被证明能够自主生成稀疏的二进制模式，而这些模式非常适合相应的 VSA 机制，几乎不需要中间计算或数据移动 [Hopkins 等人，2023]。在更为传统的 ANN 环境中，节省的资源甚至更加明显，因为数百或数千轮反向传播，以及每轮计算导数所必需的浮点计算，都极其耗时耗能

3.4 神经形态计算的重要性

如今大多数计算系统都基于经典的冯·诺依曼计算机体系结构。在这种体系结构中，处理器或中央计算单元（中央处理器）按顺序执行基本计算，每次从内存中提取一个变量（或至少提取连续内存中的紧密数据块）。借助以数GHz频率运行的中央计算机时钟进行控制，这种方式每秒可以执行数十亿次操作。如今为人工智能提供支持的基于神经网络的算法需要大规模并行计算，因为其计算基底由数百万个神经元和数十亿个连接组成。当这类算法在中央处理器上实现时，每次计算都需要将变量移出内存并写回内存，从而浪费大量时间和能源。现代图形处理器通过引入由大量计算单元组成的并行系统，缓解了这一问题的一部分，但这些计算单元仍然需要从独立的内存单元中提取数据

并按顺序进行处理。这使得具有适当网络架构的神经网络算法能够比在中央处理器上运行得快得多。然而，将变量传入和传出内存仍会浪费大量能源；尤其是当神经网络的结构无法轻易映射到图形处理器上的本地内存系统（缓存）时。

神经形态技术是计算架构发展的下一个根本性步骤，其结构灵感源自生物大脑中的神经网络。生物大脑经过进化，能够高效地将实时感知与控制同长期记忆结合起来，而这些记忆是在不同时间尺度上形成并保存的。人脑仅需 20–30 瓦即可运行，对 860 亿多个神经元和数万亿个突触连接进行“计算”。这些网络构成了大规模并行系统，由基本计算单元和电路组成；每个单元和电路都独立于其他部分“跳动”，也就是说，不存在触发同步计算的中央时钟，整个系统支持一种效率更高、可在硬件中复现的异步计算模式。此外，与大脑中计算单元和内存不可分割的情况相同——二者均由神经元之间的突触连接及其内部状态实现——神经形态硬件系统也具有局部内存，它与处理单元位于同一位置（例如 SpiNNaker 或 Loihi 等数字神经形态器件），或嵌入混合信号器件的计算超大规模集成电路中，或集成于存内计算系统中。因此，神经形态处理器采用细粒度并行进行计算，不依赖特定的神经元网络拓扑结构（而图形处理器则适用于批量、密集的基于矩阵和张量的计算）。这一特性使其能够实现低延迟、高能效处理 [Sandamirskaya 等人，2022]。Renner 等人 [2024] 表明，对于规模较大的网络，Loihi 在能量延迟积（EDP）方面的效率最高可达 171 倍。其他研究表明，使用 SpiNNaker2 模拟基于 EGRU 的语言模型仅需 0.39 瓦，而英伟达 A100 图形处理器执行同等任务则需要 60 瓦 [Nazeer 等人， 2024]。同时，研究还表明，第一代 SpiNNaker 可用于对大脑的不同区域进行低延迟模拟。SpiNNaker 实现了感觉皮层的实时模拟，其表现超越了已发表的高性能计算最佳成果 3 倍，并超越图形处理器的成果 2 [Rhodes 等人， 2019]。

在考量能效时，采用其他计算范式（例如神经形态计算）的重要性尤为明显。在我们的项目中，我们选择使用 Loihi2，而不是采用 Nvidia Jason Nano 等基于图形处理器的系统来构建手部检测子系统。Jason Nano 的最大功耗为 5–10W，具体取决于配置；而 Loihi2 仅需 4mW 即可驱动计算。单个 SpiNNaker2 芯片的能源消耗约为 1 至 2.5W，每块 48 节点板卡的功耗则在 48–120W 范围内 [Rhodes et al.，2019]。就本文而言，我们估计在 SpiNNaker2 上运行 SPAUN 需要十块 48 节点板卡，而完整的 DGX 系统则配备 8 个 A100 图形处理器。DGX A100 单元的功耗约为 6.5KW，因此 SpiNNaker2 实现的功耗将比其低 5.4 至 13.5 倍。另一个优势是，在 SpiNNaker 平台上，模型可以实时运行；而使用 DGX 设备时，模型只能以实时速度的 45% 运行。虽然这足以满足我们的使用需求，但对于实时应用而言，这可能会带来问题。

神经形态硬件开辟了全新的算法空间。在这一空间中，基于事件的网络，或具有不同拓扑结构的脉冲神经网络，能够以比传统算法快几个数量级、能效高得多的方式，解决感知、记忆形成（即学习）、规划和控制任务。新的计算方式与信息表示方式——例如基于脉冲事件和脉冲时序依赖可塑性的时序计算、向量符号架构和超维计算的脉冲实现，或基于神经场吸引子动力学的计算——标志着计算理论与技术进入了新时代。即使使用小型模型，这些方法也能执行强大的计算，从而实现模式的快速识别与跟踪、动态记忆形成以及实时控制。生物启发式学习和推理机制还将在许多其他方面带来新的机遇，例如利用少量数据进行快速、持续的学习；由于当前人工智能模型的构建和训练方式，这始终是它们难以实现的能力。对于传统人工神经网络（人工神经网络）而言，“绑定问题”也已被证明极难解决。其他不那么显而易见、但同样重要的优势还包括：即使数据受损或传感器退化，也能保持稳健运行 [Hopkins 等人， 2023]。这些能力对于实现未来的泛在机器人系统尤为重要，因为这类系统需要大量智能计算，并且必须以可持续的方式实时执行。尽管生物启发式、类脉冲神经网络和神经形态算法带来了许多计算与推理方面的机遇，但迄今为止，对它们的研发投入与更传统的人工神经网络和大语言模型相比微乎其微，因此在大多数情况下，它们的绝对性能仍然落后。一个突出的难题是工具链和框架：它们应支持新思想的试验与共享，而传统人工智能领域在这方面已发展得十分成熟

。随着这些新思想吸引更多关注和投资，我们无疑会看到竞争环境逐渐趋于平衡。已经取得的可喜成果几乎肯定会与新的成果相结合，使其竞争力大幅提升，同时保留工程优势——前提是那些既得利益者不会试图阻碍这场竞争；这些利益者依赖高度依赖内存、耗能巨大的人工神经网络。

4 方法

4.1 硬件概览

4.1.1 Ameca——通用型社交机器人

Ameca 是一款社交人形机器人，能够展现栩栩如生的人类面部表情 [Singh 等人， 2022]。它被誉为全球最先进的社交机器人。该机器人由机器人公司 Engineered Arts 开发；该公司是娱乐型人形机器人的领先设计商和制造商。

Ameca 使用专有操作系统 Tritium，可通过 Python 脚本扩展其功能。为了做出运动和反应方面的决策，它配备了缓冲系统，并连接到负责状态转换的 Spaun 模型。Ameca 通过电机、传感器、摄像头和麦克风等硬件组件收集环境数据。机器人可以通过安装在胸部的扬声器进行交流；此外，还配有外接麦克风，用于用户与机器人之间的互动。该麦克风直接连接到机器人的内部计算机——Intel NUC，以减少背景噪声并提高传输音频的质量。



图 3：Ameca——由 Engineered Arts 制造的人形社交机器人。

4.1.2 微型数据中心

微型数据中心（MDC）为展示系统提供后端。为尽量降低服务器通风在舞台区域产生的声学噪声，MDC 采用了隔音机箱结构。MDC 由两个机架组成，包含以下组件：

- 3NVIDIA DGX A100 服务器，
- 配备双 CPU AMD Epyc 7313 的数据服务器，
- 用于显示视觉内容的视频服务器，
- 用于管理后端访问和前端设备网络集成的网络交换机，以及服务器的智能平台管理接口（IPMI）。
- 外围基础设施包括：电气与控制基础设施、消防与通风系统，以及音频系统的放大器和调音台。

4.1.3 图形处理器

本项目使用了三套 NVIDIA DGX A100 系统，每套均配备 8 个 A100 图形处理器。每套系统均专用于特定功能：

- DGX 1：大语言模型与语音识别。第一台 DGX 系统运行着拥有 130 亿个参数的大语言模型（LLM），用于与机器人进行低延迟常识对话。大语言模型的输入由 OpenAI 的 Whisper 语音识别引擎提供。

• DGX 2：Spaun 2.0 模型与音频处理。DGX 2 用于运行 Spaun 2.0 模型。该模型经过调整，可使用 7 个图形处理器在 A100 GPU 架构上运行（见第 4.2.1 节）。两台特雷门琴的音频输入通过 USB 接口传输至 DGX-2。

• DGX 3：处理用于舞台监测和灯光管理的感官输入。第三台 DGX 负责检测人员及其相对于舞台的位置。传感器输入由 3 台摄像头提供（2 台 RGB 摄像头、1 台深度摄像机）。每台摄像头均用于指定任务（见第 4.2.5 节）。

4.1.4 神经形态计算

SpiNNaker SpiNNaker 系统 [Furber 和 Bogdan， 2020] 是一个数字神经形态平台，旨在模拟大规模脉冲神经网络。曼彻斯特大学建成的这一设施在单个集群中搭载了超过 100 万个 ARM968 核心，从而实现了无与伦比的并行处理能力，与标准处理器架构相比，大幅提升了计算速度。这使得曼彻斯特的 SpiNNaker 神经形态集群成为全球最大的实时脑模拟器，直到最近被基于 Loihi2 的 Hala Point 系统超越。然而，德累斯顿工业大学正在建造的新型 SpiNNaker2 集群，其容量超过了这两个系统，创下了全球最大类脑超级计算机的新纪录。

SpiNNaker2 为了释放前沿的实时自主能力，我们将在项目后续迭代中充分发挥 SpiNNaker2 的优势。它是神经形态基底领域中最灵活的神经架构。SpiNNaker2 是专为大规模基于事件的异步处理而设计的加速器 [Mayr 等人， 2019]。因此，它能大幅降低能耗，在同等配置下的功耗仅为图形处理器方案的十分之一。

SpiNNaker2 微芯片包含 152 个基于 ARM 的处理单元，并通过片上网络（NoC）互连。芯片共有 19 MB 片上 SRAM，另配备 2 GB LPDDR4 内存。与第一代 SpiNNaker 芯片相比，其独特之处在于专用加速器可执行指数函数、随机数生成和乘加（MAC）运算。这些特性使其能够高效实现稀疏性感知人工神经网络、符号人工智能和脉冲神经网络 [Gonzalez 等人， 2024]。SpiNNaker2 的设计体现了生物学启发，其芯片架构尤为明显。分布式处理单元以异步方式运行，能够实现大规模并行的基于事件的稀疏计算，并按需控制功耗。与峰值性能相比，芯片在闲置期间会尽可能降低功耗。每个处理单元均可自主运行，并能通过动态电压频率调节（DVFS）分别控制各单元的功耗，从而按需实现精细的功耗优化。轻量级片上网络负责在芯片内分发信息，并以最低延迟支持数据输入输出流。类脑加速器可提升生物启发网络的计算性能。因此，该系统原生支持传统人工智能与脉冲网络的混合模型。

可持续超级计算能力与高温液冷设计未来几年人工智能应用和硬件预计将呈指数级增长，因此研究高能效解决方案，以减少 CO_2 排放和运营成本，具有重要意义。这一方法的支柱之一是使用神经形态硬件。另一项措施是采用高温液冷，从而实现冷却基础设施的最高效运行，并支持余热再利用。一个参考案例是与德累斯顿工业大学和 SpiNNcloud系统合作，为 SpiNNaker2 设计的 100% 液冷基础设施。由 Cloud&Heat 和 WAIYS 设计的专利液冷技术不仅比传统空气冷却运行时安静得多，还利用较高的液体温度，实现余热再利用，用于设施冷却，甚至工业过程热应用 [Struckmeier, 2018]。我们此前的数据中心开发经验表明，神经形态服务器的冷却回路消耗的能源约占整个集群所需总能源的 1%，其性能显著优于任何空气冷却解决方案。当该技术与其他互补冷却方法结合使用时，集成方案相比传统空气冷却最多可将能源消耗降低 70%，凸显了采用多样化冷却策略以实现能效用例的重要性 [Cloud&Heat Technologies GmbH，2019]。

英特尔 Loihi2——感知流的实时处理 Loihi 芯片最初由英特尔于 2018年发布，同时推出了英特尔的神经形态研究社区计划 [Davies et al.]。Loihi2 于 2021年发布，同时推出了 Lava——一个用于开发神经启发式应用的开源软件框架 [Orchard et al.，2021]。Loihi2 是一种神经形态芯片，可将脉冲神经网络计算推广至具有几乎任意拓扑结构的大规模异步图，并新增以脉冲形式发送分级信息的能力。这极大拓展了算法设计的可行方案空间。Lava 提供了一种思考神经形态算法的新方式：将其视为由一系列交互进程组成的系统，每个进程都随时间展开，并与运行于神经形态硬件或传统硬件上的其他进程交换脉冲消息，或与物理世界的接口（传感器、机器人控制器或屏幕）进行交互。

本文使用英特尔技术，重点在于展示这种新兴计算硬件如何助力机器人应用。神经形态硬件以基于事件的异步方式执行计算，其方式从根本上不同于传统计算，并具备细粒度并行和片上学习能力，因此能够实现一种新型机器人功能：在快速实时控制回路中，将复杂的“智能”处理与人工神经网络相结合，用于目标检测、跟踪和定位，以及运动规划。这种快速感知与认知能力将使机器人在动态、以人为中心的环境中更加安全、高效。在演示中，我们专注于这一处理过程中的感知部分，展示快速的实时视觉处理。

在我们的演示中，实现实时视觉的一项关键组件是神经形态事件相机——动态视觉传感器（DVS）。DVS 与神经形态计算硬件尤其匹配：DVS 的事件和神经形态芯片中的脉冲都是异步信号，因此视觉计算能够兼具能效与速度。我们将一款 DVS 传感器与英特尔 Loihi2 芯片集成，并训练了一种神经网络算法，根据 DVS 输出的异步变化事件跟踪人手。这是首次在 Loihi 上展示此类实时工作负载，为实现实时视觉铺平了道路；机器人需要实时视觉才能时刻感知周围环境。

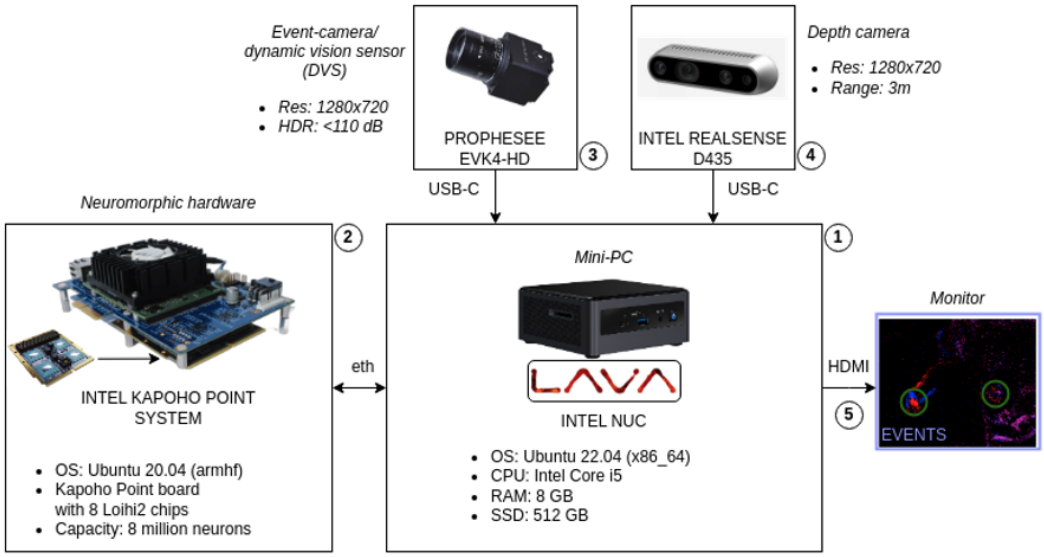

图4：实时神经形态手部追踪硬件层概览 .

该手部追踪系统由现成组件，以及英特尔最新的开发板 Kapoho Point 和神经形态研究芯片 Loihi2 共同构建而成，见图4。交付的硬件解决方案由若干关键组件组成：

- Kapoho Point 系统——英特尔的神经形态计算开发板，由 8 个 Loihi2 芯片、一个英特尔 FPGA 系统和一个用于处理外围设备的底板组成。其中一个外围设备是以太网接口，可与 NUC 通信。以太网接口用于从 NUC 向神经形态芯片发送命令，并中继板载应用程序生成的任何数据。

- Prophesee Evaluation Kit4 HD（EVK4-HD）是一款事件相机，可输出具有 1280x720 空间分辨率的时空事件流，延迟为 $220\mu s$，并具备大于 110dB 的高动态范围。

- 深度信息由 Intel RealSense 深度相机 D435 捕获，分辨率为 1280x720，测距范围最高可达 3 米。

4.2 软件概述

本节将深入介绍实现独特互动音乐体验的软硬件架构，并探讨一些关键组件，这些组件使 Ameca 能够流畅地进行交流。机器人通过安装在机身内部的传感器，以及安装在舞台周围的额外传感器，获取各种环境线索。其中一部分额外感官输入来自安装在舞台周围不同位置的摄像头，另一部分则来自舞台上及其附近的额外麦克风。此外，机器人还会接收特雷门琴传来的数字信息，以控制其声音，并帮助用户参与互动二重奏模式。

图5展示了为本项目构建的所有软硬件组件，以及它们的集成方式。我们采用先进系统进行独特集成，以处理输入的数据流。该组合包含神经形态解决方案，例如用于低延迟手部检测的 Loihi2；采用 NVIDIA DGX A100 的高性能图形处理器，可在本地运行大型语言模型和语音识别模型的模拟；以及云解决方案。此外，该架构还包括舞台管理系统，用于控制舞台灯光和视频墙显示屏。我们还集成了多个音频处理元件，包括扬声器、麦克风和接口。

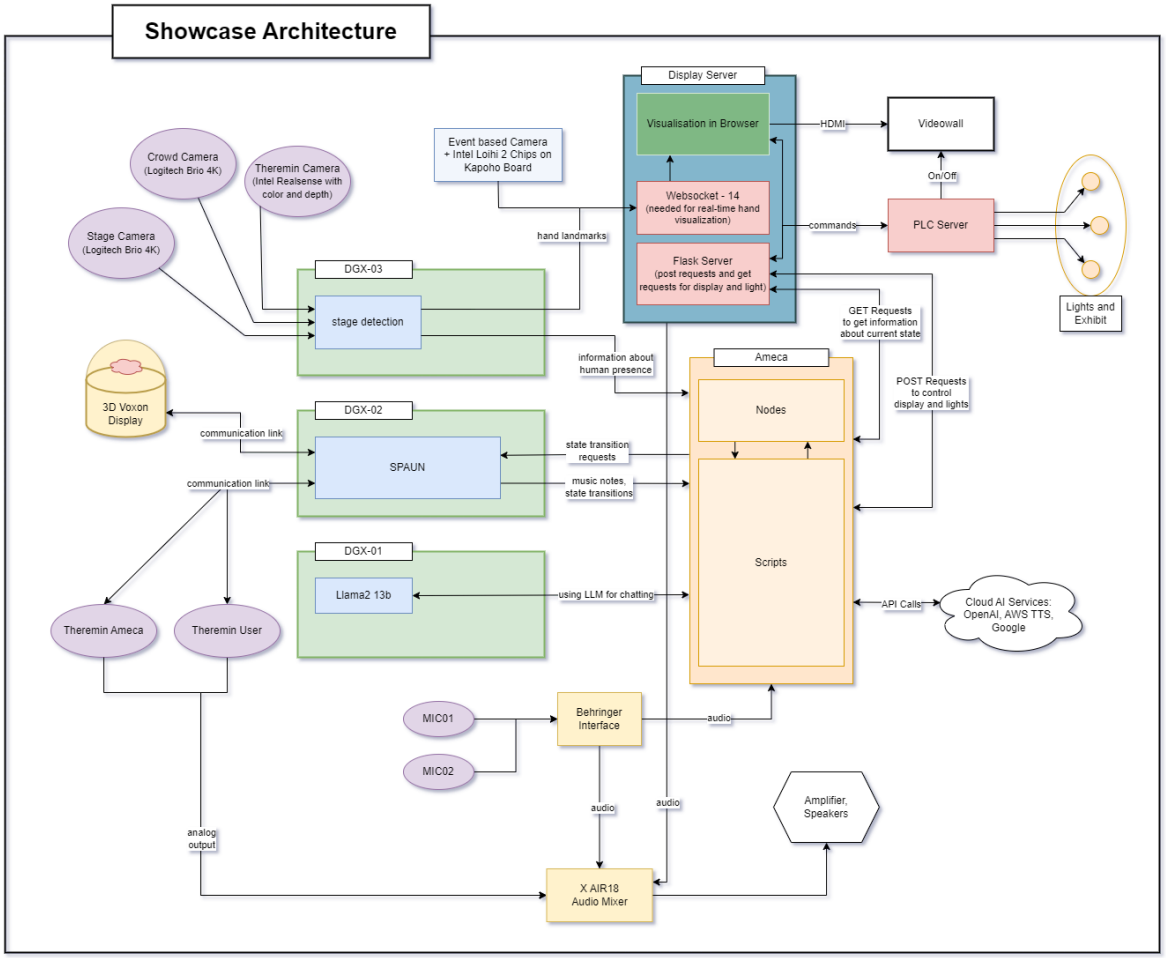

图5：互动音乐展示的软硬件集成架构概览。

4.2.1 Spaun 2.0 在认知记忆任务中的集成及其在决策中的作用

为模拟系统的状态转换，我们采用了 Spaun 2.0 模型。Spaun（语义指针架构：统一网络）是一种大规模认知神经模型，包含 250 万个神经元[Stewart 等人， 2012]，基于语义指针架构实现 [Eliasmith， 2013]。这一架构随后在 Spaun 2.0——世界上最大的功能性脑模型中得到扩展 [Choo，2018]。这一独特模型最多包含 650 万个神经元，用于执行认知记忆任务并进一步增强机器人的控制过程，从而使 Ameca 能够通过有限状态机（FSM）与用户互动。该模型能够执行 12 项认知任务，并已被证明能够重现自然认知智能体中观察到的行为和神经学数据。模型的功能模块对应于大脑的特定区域。作者证明，可以遵循人脑的组织原则建立模型来执行各种任务。Spaun 模型由 8 个模块组成：视觉处理、信息编码、转换、奖励评估、信息解码、运动控制、工作记忆和动作选择。Spaun 的动作选择模块负责跟踪 Spaun 的进展，并在执行每项任务时决定采取哪些动作。此外，动作选择系统还负责管理构成 Spaun 的各个网络之间的信息流。

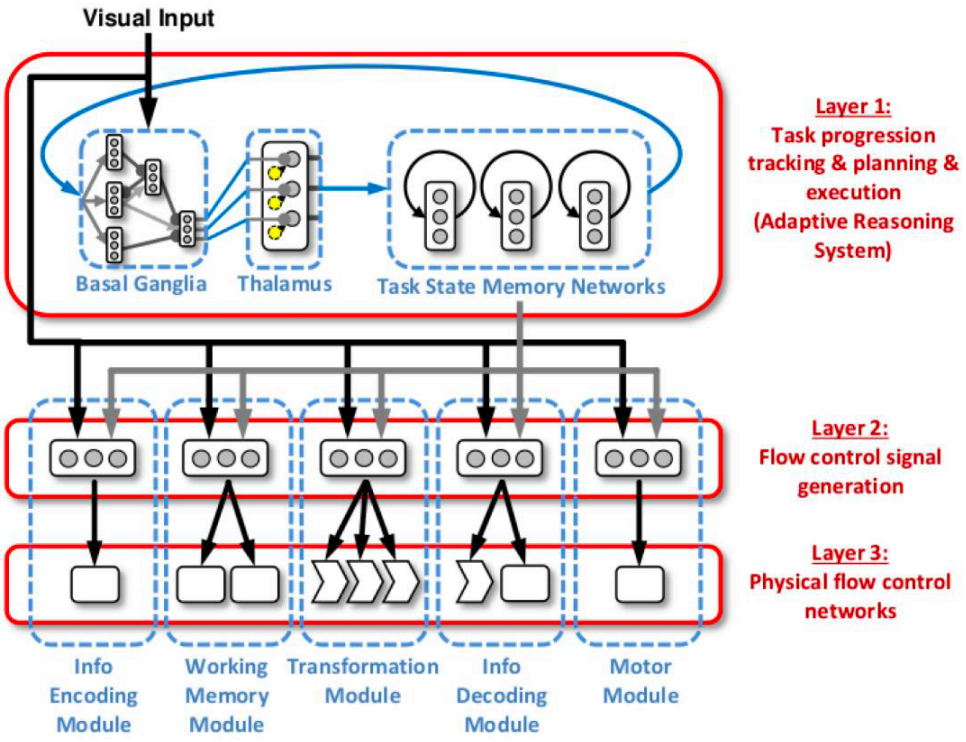

图 6: 三层动作选择层级示意图（图示源自 Choo [2018]）.

为了减少条件—结果对的数量和复杂度，Spaun 的动作选择模块采用三层层级结构实现，其中每一层负责动作选择系统的不同功能（见图 6）。层级的顶层接收来自视觉模块的输入，仅负责对 Spaun 任务进度进行高层跟踪、规划和执行。这一层由推理系统实现。层级的第二层将顶层的状态表示与视觉语义指针相结合，生成特定于模块的控制信号，用于控制 Spaun 各模块内部的信息流。动作选择层级的最后一层包含必要的物理网络，以实现由动作选择层级第二层输出所规定的信息流。状态机的行为，进而机器人的决策，由 Spaun 模型控制。这使机器人能够根据一系列与音乐演奏相关的基于状态的任务调整自身行为（见图 7）。

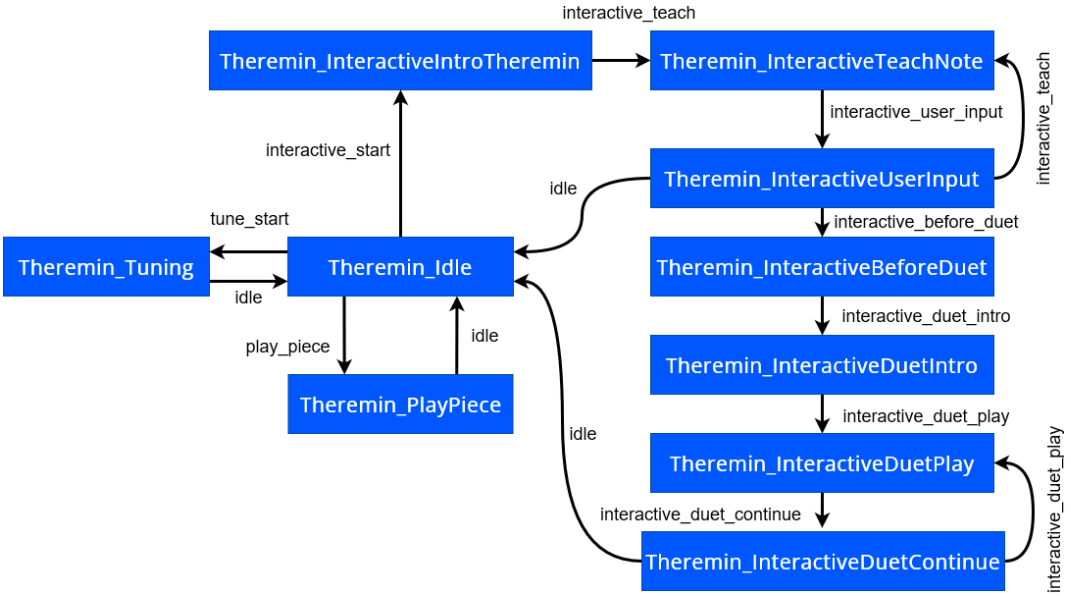

图 7：与特雷门琴交互相关的机器人状态转换概览 s.

Spaun 2.0 模型部署在英伟达图形处理器系统上。为实现实时多图形处理器部署，原始 Spaun 2.0 架构经过调整，将 Spaun 的每个模块划分为独立的“子网络”。随后，每个子网络都可在各自的 NengoOCL 模拟器中运行 [Bekolay 等人， 2014]。图 8 展示了 Spaun 2.0 系统架构如何分布在 7 个图形处理器上。未来，该系统可部署在液冷神经形态硬件（如 SpiNNaker2）上，以充分利用其能效优势并进一步降低延迟。原始 Spaun 2.0 架构包含 8 个模块（直观上可分布在 8 个图形处理器上），但由于奖励评估系统结构轻量，且与动作选择系统连接广泛，因此将奖励评估模块与动作选择模块合并。

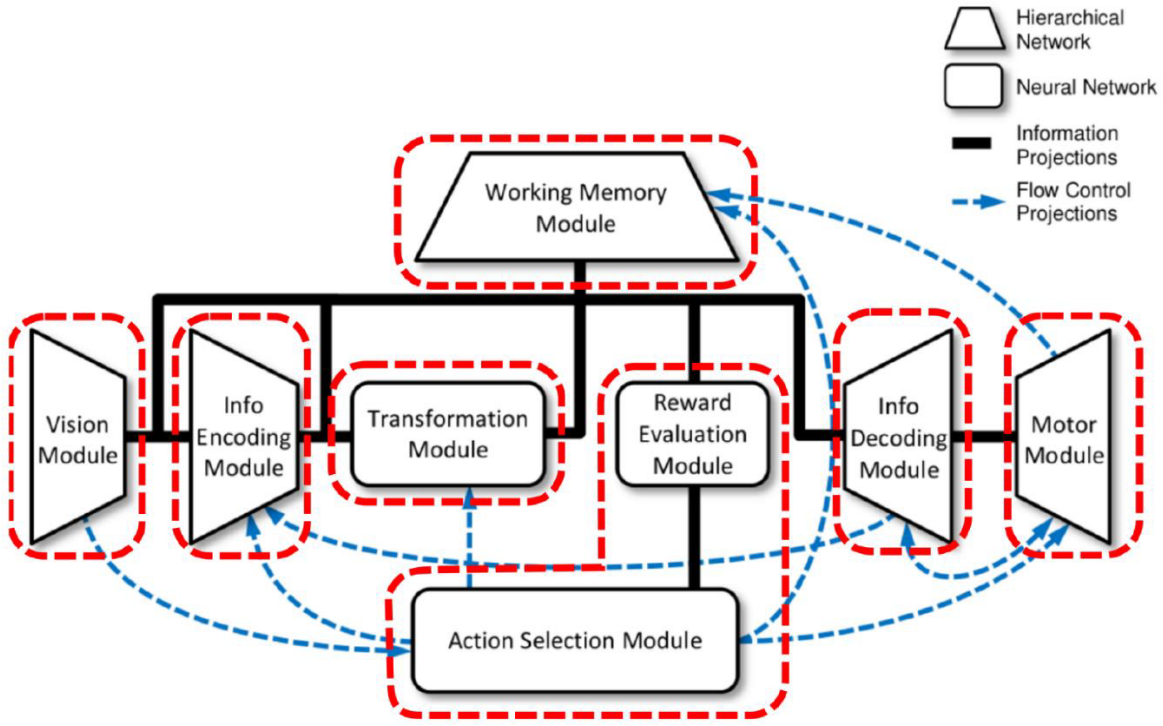

图 8：Spaun 2.0 模型架构在 7 个图形处理器上的划分。每个红色方框表示模型中一个运行于独立图形处理器上的部分（图示来自 Choo [2018]）。

4.2.2 对话引擎

要与社交机器人进行流畅、低延迟的对话，需要高性能的大语言模型。为此，我们采用了 Meta 推出的 LLaMA 系列中的一个最新模型 [Touvron 等人，2023]。在本次展示中，我们使用了具有 130亿个参数的 LLaMA 模型变体，它能够快速计算响应，同时不牺牲响应质量。该模型部署在本地图形处理器系统上。

机器人能够进行多轮对话、回答一般领域的问题，以及根据提供给模型的上下文信息回答与本次展示相关的问题。为了让机器人能够与来自不同地区的用户交流，我们使用了翻译软件。与机器人的语音交流通过 Google Cloud 翻译 API 进行翻译。

4.2.3 语音生成

为了将语言模型生成的文本转换为逼真的语音，使机器人能够与用户交流，我们决定采用 Amazon Polly API [Werchniak 等人， 2021]。Amazon Polly 利用先进的深度学习技术生成类人语音，将书面文本转换为自然流畅的音频。

4.2.4 语音识别

我们使用 OpenAI Whisper 模型 [Radford 等人， 2022]，使机器人能够听取用户说出的指令。Whisper 是一种自动语音识别系统，在英语语音识别方面展现出接近人类水平的鲁棒性和准确性。Whisper 模型拥有 1550M 个参数，并在一台 NVIDIA DGX A100 机器上本地运行。

4.2.5 玩家检测与深度分析 ysis

为了增强机器人对周围环境的感知能力，我们在舞台周围战略性地布置了三个摄像头。每个摄像头都采用不同的方法来解读场景：

- 摄像头 1 对舞台前方的观众进行拍摄，并利用实时人脸检测统计观看舞台的人数。

- 摄像头 2 负责检测舞台上是否有人。这有助于了解舞台的占用情况，并确保 Ameca 表演期间遵守安全措施。
- 摄像头 3 配备了深度传感器，用于识别特雷门琴后方是否有人。这能让机器人判断是否要与特雷门琴演奏者互动。为此，我们使用了 Yang 等人 [2024]介绍的先进深度估计模型 Depth Anything。

4.2.6 意图检测

为驱动机器人的有限状态机并实现互动音乐体验，我们使用了 OpenAI [Brown 等人 2020]提供的 GPT-3.5 API 来进行意图检测。意图检测通过指定上下文以及从中提取的一个或多个意图来实现。可能的意图由名称和描述组成，并以键值对的形式存储在字典中。系统将这些记录与上下文及待检测意图的消息一并传递给大语言模型。回调函数负责处理输出；输出内容为检测到的意图名称，如果未能检测到任何意图，则为“None”。

4.2.7 手部追踪

为了在与 Ameca机器人进行互动特雷门琴二重奏时直观呈现用户的输入，我们决定开发节能、低延迟的手部追踪技术。跟踪人手对于许多机器人应用至关重要，例如模仿学习和人机协作；在展示人与机器人共同演奏乐器时，这项技术也发挥着关键作用。为此，我们与英特尔合作，采用了其最新的神经形态处理器 Loihi2 [Davies 等人] ，并将其与 DVS相机结合使用。由于事件相机仅在特定像素检测到场景亮度变化时才输出信息，因此固定安装的事件相机可以直接用作简单的运动检测器。当场景与相机之间发生相对运动时，各像素会独立报告感知到的亮度变化，从而产生稀疏、异步的时空数据流。有关硬件设置的更多详情，请参阅第 4.1.4 节。针对本项目，我们开发了三条并行路径，以实现可靠的手部追踪。

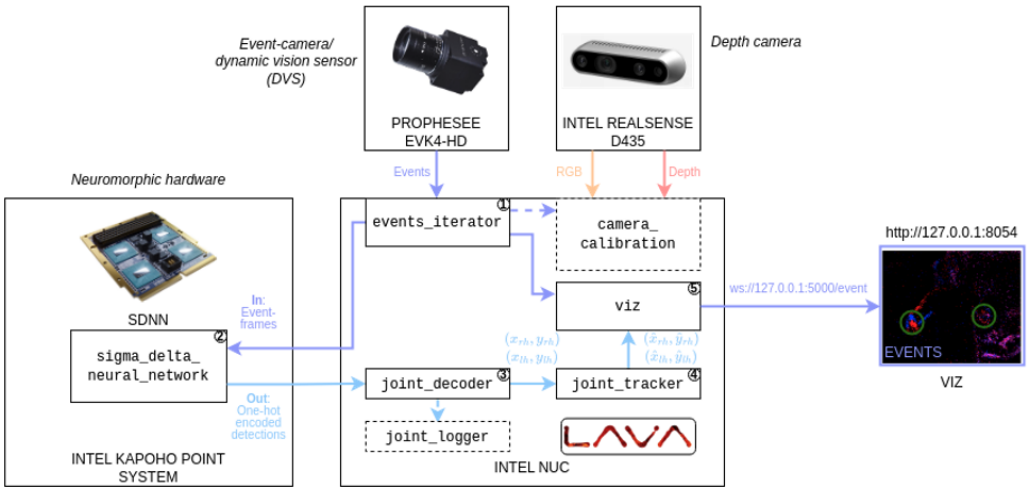

图 9：神经形态手部追踪软件层概览。

第一条路线采用经过训练的深度脉冲神经网络（DSNN），利用 DHP19 基于事件的数据集跟踪手部。该数据集包含移动人物各关节的自动生成标签，这些标签由运动跟踪系统获取 [Calabrese 等人， 2019]。我们使用该数据集训练了一个兼容 Loihi2、用于跟踪手部的脉冲神经网络。以脉冲为计算基础，神经形态硬件在运行适当的基于脉冲的算法时，能达到最高效率。鉴于最先进的基于神经网络的运动追踪技术针对不同类型的硬件（例如图形处理器）进行了优化，因此需要新的神经网络架构，才能充分发挥神经形态计算平台的优势。Σ-Δ 量化是一种

能在传统硬件和神经形态硬件上更高效地运行卷积神经网络的有效方法。在 Σ-Δ神经网络中，通过增量编码过程和Σ解码过程实现脉冲发放。编码器计算人工神经网络神经元最新非线性激活值之间的差异。当差异超过预设阈值时，编码器便向下一层发送一个脉冲。解码器累积这些脉冲，直至恢复原始值，从而对人工神经网络的激活过程进行离散近似。

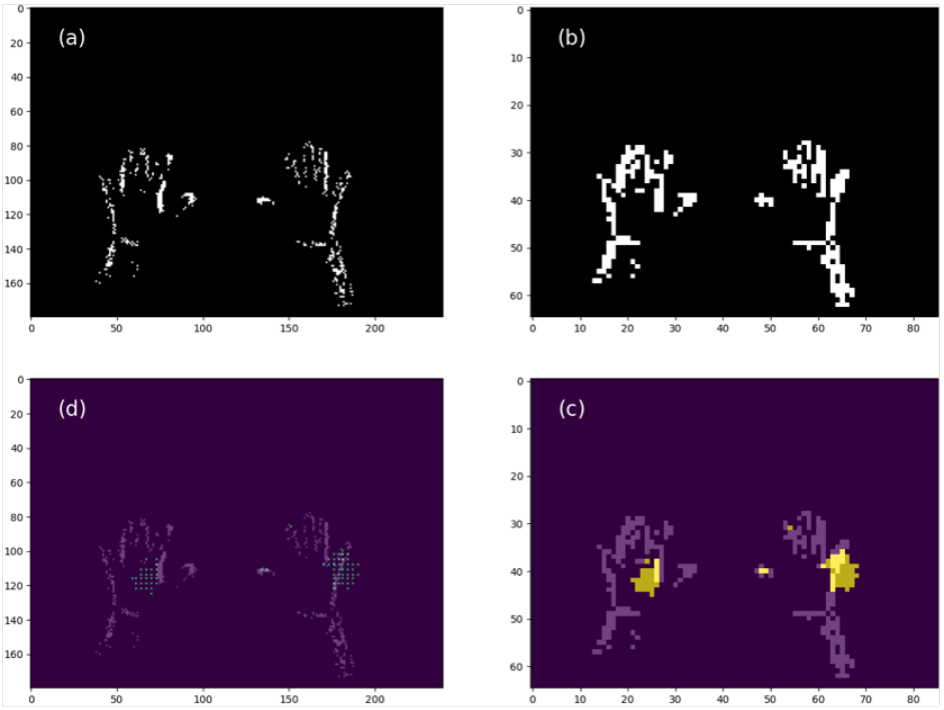

图 10：双手同时左右移动（挥手）的跟踪示例：(a) (240x180) 捕捉活动双手的事件帧；(b) (86x65) 下采样后用于在芯片上运行的事件帧；(c) 多峰动态神经场的输出叠加在下采样后的事件帧上；(d) 放大后的动态神经场输出叠加在原始事件帧上。

所部署的第二种计算依托动态神经场理论 [Schöner 等人， 2015] 以改进基于深度学习的跟踪。这种神经形态算法还可用于滤除事件流中的相干运动，从而构成生成手部运动假设的另一种机制。动态场理论（DFT）是用于建模人类认知的数学和概念框架。该方法已被用于为机器人开发认知架构（空间语言、行动规划、学习），并被用作对神经形态芯片进行高级编程的计算范式之一。在这一框架中，神经元群体（层）形成吸引子动力学，并产生经过稳定的空间模式，从而滤除噪声，并对干扰物存在时移动物体的跟踪提供时间滤波。

在此应用中，我们决定按顺序配置这些方法，由动态神经场滤波器稳定深度神经网络的输出。视觉场景的进一步滤波（前景—背景分割）基于传统 RGB-D相机提供的深度信息进行，本例中使用的是英特尔的 RealSense（见图 10）。

作者贡献

作者根据贡献进行分组：T.R.、X.C.、C.M.、H.G.、M.A.、K.K.、C.E.、M.A.——负责初始编程、特雷门琴研究和 Spaun 集成。M.H.、S.F.——协助撰写论文。Y.S.、L.A.、S.G.——开发 Loihi2 子系统。J.F.、L.F.、F.W.——负责展示项目的编程和进一步开发。L.W.、H.P.、J.S.、K.R.、N.A.、M.H.——负责展示项目的监督和最初构想。

致谢

我们还要感谢 NEOM 智慧城市倡议发起并资助特雷门琴演奏机器人应用案例的创建。感谢英特尔神经形态研究社区和英特尔神经形态计算实验室为 Loihi2 实现提供支持。这项工作背后的一些想法最初由作者在卡波卡恰研讨会“迈向神经形态智能”上进行讨论。

此外，我们谨向同事和合作者致谢，并感谢他们所作的贡献：Chris Eliasmith 教授、Carolina Eyck、Nourchene Ferchichi、Matthieu Amiguet、Travis Dewolf、Eric Schulz、Eric Wallin、Mike Davies 和 Morgan Roe。

参考文献

Subutai Ahmad 和 Jeff Hawkins。神经元如何处理稀疏分布式表征？关于稀疏性、神经元和活性树突的数学理论。2016年 01。

Vito Albino、Umberto Berardi 和 Rosa Dangelico。智慧城市：定义、维度、绩效与倡议。《城市技术期刊》，22:2015，2015年 02。doi：10.1080/10630732.2014. 942092。

Michael Batty。智慧城市、大数据。《环境与规划*B*：规划与设计》，39(2)：191–193，2012年。doi：10.1068/b3902ed。URL https://doi.org/10.1068/b3902ed。

Trevor Bekolay、James Bergstra、Eric Hunsberger、Travis DeWolf、Terrence C. Stewart、Daniel Rasmussen、Xuan Choo、Aaron Voelker 和 Chris Eliasmith。Nengo：用于构建大规模功能性脑模型的 Python 工具。*Frontiers in Neuroinformatics*，7，2014年。ISSN 1662-5196。doi：10.3389/fninf.2013.00048。URL https://www.frontiersin.org/journals/neuroinformatics/ articles/10.3389/fninf.2013.00048。

Tiago Branco 和 Michael Häusser。单个树突分支是神经系统中的基本功能单元。*Current Opinion in Neurobiology*，20(4)：494–502，2010。ISSN 0959-4388。doi：https: //doi.org/10.1016/j.conb.2010.07.009。URL https://www.sciencedirect.com/science/article/ pii/S0959438810001170。信号传导机制。

Cynthia Breazeal。迈向社交型机器人。《机器人学与自主系统》，42(3):167–175，2003年。ISSN 0921-8890。doi：https://doi.org/10.1016/S0921-8890(02)00373-1。URL https://www.sciencedirect.com/science/article/pii/S0921889002003731。社会交互机器人。

Tom B. Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared Kaplan、Prafulla Dhari- wal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell、Sandhini Agarwal、Ariel Herbert-Voss、Gretchen Krueger、Tom Henighan、Rewon Child、Aditya Ramesh、Daniel M. Ziegler、Jeffrey Wu、Clemens Winter、Christopher Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benjamin Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。《语言模型是少样本学习者》，2020年。

Enrico Calabrese、Gemma Taverni、Christopher Awai Easthope、Sophie Skriabine、Federico Corradi、Luca Longinotti、Kynan Eng 和 Tobi Delbruck。DHP19：动态视觉传感器三维人体姿态数据集。载于*2019*年*IEEE/CVF*计算机视觉与模式识别会议研讨会（*CVPRW*），第 1695–1704 页，2019年。doi：10.1109/CVPRW.2019.00217。

Andrea Caragliu、Chiara Del Bo 和 Peter Nijkamp。《欧洲智慧城市》。《城市技术期刊》，18(2)：65–82，2011年。doi：10.1080/10630732.2011.601117。网址：https://doi.org/10. 1080/10630732.2011.601117。

Feng-Xuan Choo。*Spaun 2.0*：扩展世界上最大的功能性脑模型。博士论文，2018年。

Cloud&Heat Technologies GmbH。《利用 Cloud&Heat 冷却系统和余热利用降低数据中心二氧化碳排放与成本的潜力》。白皮书，2019年12月。网址：https://www.cloudandheat.com/wp-content/uploads/2023/03/2019-12-16_CloudHeat- Whitepaper-Cost-saving-Potential_EN.pdf。访问日期：2025年2月20日。

Mike Davies、Andreas Wild、Garrick Orchard、Yulia Sandamirskaya、Gabriel A. Fonseca Guerra、Prasad Joshi、Philipp Plank 和 Sumedh R. Risbud。《利用 Loihi 推进神经形态计算：成果综述与展望》。*IEEE*会刊。Mahmut Dirik。《贾扎里：控制论与机器人学领域的杰出发明家》。《软计算与人工智能杂志》，1(1):47–58，2020年。C. Eliasmith。《如何构建大脑：生物认知的神经架构》。牛津认知模型与架构系列。牛津大学出版社，2013年。ISBN 9780199794690。网址：https://books.google.pl/books?id=eh1pAgAAQBAJ。Steve Furber 和 Petruţ Bogdan 编。《*SpiNNaker*：脉冲神经网络架构》。now publishers，波士顿-代尔夫特，2020年。ISBN 978-1-68083-652-3。doi：10.1561/9781680836523。网址：http: // dx.doi.org/10.1561/9781680836523。Artur d'Avila Garcez 和 Luís C. Lamb。《神经符号人工智能：第三次浪潮》。*Artificial IntelligenceReview*，56(11):12387–12406，2023年。doi：10.1007/s10462-023-10448-w。网址：https://doi.org/10. 1007/s10462-023-10448-w。

Lulu Ge 和 Keshab K. Parhi。基于超维计算的分类：综述。*IEEE*电路与系统杂志，20:30–47，2020年。URL https://api.semanticscholar.org/CorpusID: 216080530。

Oleg Golubchikov 和 Mary Thornbush。智慧城市战略与规划智慧发展中的人工智能和机器人学。智慧城市，3(4):1133–1144，2020年。ISSN 2624-6511。doi: 10. 3390/smartcities3040056。URL https://www.mdpi.com/2624-6511/3/4/56。

Hector A. Gonzalez、Jiaxin Huang、Florian Kelber、Khaleelulla Khan Nazeer、Tim Langer、Chen Liu、Matthias Lohrmann、Amirhossein Rostami、Mark Schöne、Bernhard Vogginger、Timo C. Wunderlich、Yexin Yan、Mahmoud Akl 和 Christian Mayr。SpiNNaker2：面向基于事件的异步机器学习的大规模神经形态系统，2024年。

Stephen Grossberg。自适应共振理论：大脑如何学会有意识地关注、学习并识别不断变化的世界。《神经网络》，37：1–47，2013年。

Michael Hopkins、Jakub Fil、Edward George Jones 和 Steve Furber。BitBrain 与稀疏二进制巧合（SBC）内存：用于神经形态计算架构的快速、稳健学习与推理。*Frontiers in Neuroinformatics*，17，2023年。ISSN 1662-5196。doi：10.3389/fninf. 2023.1125844。URLhttps://www.frontiersin.org/journals/neuroinformatics/articles/10. 3389/fninf.2023.1125844。

George Kastellakis、Denise J. Cai、Sara C. Mednick、Alcino J. Silva 和 Panayiota Poirazi。树突内的突触聚集：一种新兴的记忆形成理论。*Progress in Neurobiology*，126：19–35，2015年。ISSN 0301-0082。doi：https://doi.org/10.1016/j.pneurobio.2014.12.002。URL https: // www.sciencedirect.com/science/article/pii/S0301008214001373。

Denis Kleyko、Dmitri A. Rachkovskij、Evgeny Osipov 和 Abbas Rahimi。《超维计算（又称向量符号架构）综述，第一部分：模型与数据变换》。*ACMComputing Surveys*，55(6):1–40，2022年12月。ISSN 1557-7341。doi: 10.1145/3538531。URL http://dx.doi.org/10.1145/3538531。

Denis Kleyko、Dmitri Rachkovskij、Evgeny Osipov 和 Abbas Rahimi。《超维计算（又称向量符号架构）综述，第二部分：应用、认知模型与挑战》。*ACM Computing Surveys*，55(9):1–52，2023年1月。ISSN 1557-7341。doi: 10.1145/3558000。URL http://dx.doi.org/10.1145/3558000。

John E Laird。《Soar认知架构》。MIT出版社，2019年。

Matthew E Larkum 和 Thomas Nevian。通过树突信号传导机制实现突触聚集。*Current Opinion in Neurobiology*，18(3):321–331，2008。ISSN 0959-4388。doi: https: //doi.org/10.1016/j.conb.2008.08.013。URLhttps://www.sciencedirect.com/science/article/ pii/S095943880800086X。信号传导机制。

Patrick Lichtsteiner、Christoph Posch 和 Tobi Delbruck。一种 $128\times$ 128 120 db 15 μs 延迟异步时间对比视觉传感器。*IEEE*固态电路期刊，43(2):566–576，2008。doi: 10.1109/JSSC.2007.914337。

Bing Liu。终身机器学习：持续学习的一种范式。计算机科学前沿，11(3)：359–361，2017年。doi：10.1007/s11704-016-6903-6。URL https://doi.org/10.1007/ s11704-016-6903-6。

Michael London 和 Michael Häusser。树突计算。神经科学年度评论，28：503–32，2005年。URL https://api.semanticscholar.org/CorpusID:6546452。

Gary Marcus。人工智能的下一个十年：迈向鲁棒人工智能的四个步骤。*arXiv* 电子预印本，文章编号 arXiv:2002.06177，2020年2月。doi：10.48550/arXiv.2002.06177。

Christian Mayr、Sebastian Hoeppner 和 Steve Furber。SpiNNaker 2：用于脑模拟和机器学习的千万核处理器系统，2019年。

Samuel A. Mehr、Manvir Singh、Dean Knox、Daniel M. Ketter、Daniel Pickens-Jones、S. Atwood、Christopher Lucas、Nori Jacoby、Alena A. Egner、Erin J. Hopkins、Rhea M. Howard、Joshua K. Hartshorne、Mariela V. Jennings、Jan Simson、Constance M. Bainbridge、Steven Pinker、Timothy J. O'Donnell、Max M. Krasnow 和 Luke Glowacki。《人类歌唱的普遍性与多样性》。《科学》，366(6468):eaax0868，2019年。doi: 10.1126/science.aax0868。URL https://www.science.org/doi/ abs/10.1126/science.aax0868。

Takeshi Mizumoto、Hiroshi Tsujino、Toru Takahashi、Tetsuya Ogata 和 Hiroshi G. Okuno。特雷门琴演奏机器人：基于特雷门琴音高模型，开发具有前馈和反馈式机械臂控制功能的机器人特雷门琴演奏器。发表于*2009*年 *IEEE/RSJ*智能机器人与系统国际会议，第2297–2302页，2009年。doi: 10.1109/IROS.2009.5354473。

Masahiro Mori、Karl F. MacDorman 和 Norri Kageki。《恐怖谷 [来自该领域]》。*IEEE*机器人与自动化杂志，19(2):98–100，2012年。doi: 10.1109/MRA.2012.2192811。

Clément Moulin-Frier、Tobias Fischer、Maxime Petit、Grégoire Pointeau、Jordi-Ysard Puigbo、Ugo Pattacini、Sock Ching Low、Daniel Camilleri、Phuong Nguyen、Matej Hoffmann 等。Dac-H3：一种主动式机器人认知架构，用于获取并表达关于世界与自身的知识。*IEEE*认知与发展系统汇刊，10(4):1005–1022，2017年。

Khaleelulla Khan Nazeer、Mark Schöne、Rishav Mukherji、Bernhard Vogginger、Christian Mayr、David Kappel 和 Anand Subramoney。基于 SpiNNaker 2 神经形态芯片的语言建模，2024年。

Garrick Orchard、E. Paxon Frady、Daniel Ben Dayan Rubin、Sophia Sanborn、Sumit Bam Shrestha、Friedrich T. Sommer 和 Mike Davies。《使用 Loihi 2 进行高效神经形态信号处理》，2021年。

Hyeonjun Park、Wonse Jo、Kyeongmin Choi、Hwonjae Jung、Yura jargalbaatar、Bum-Joo Lee 和 Donghan Kim。关于小提琴演奏机器人的音质研究。《*Procedia Computer Sci- ence*》，56:496–501，2015年。ISSN 1877-0509。doi：https://doi.org/10.1016/j.procs.2015.07.241。URL https://www.sciencedirect.com/science/article/pii/S1877050915017226。第十届未来网络与通信国际会议（FNC 2015）/第十二届移动系统与普适计算国际会议（MobiSPC 2015）附属研讨会。

Dietrich Paulus、Viktor Seib、Julian Giesen 和 Dominik Grüntjens。通过具有面部表情和同步唇部动作的机器人脸部增强人机交互。2013年 01。

Jens E. Pedersen、Steven Abreu、Matthias Jobst、Gregor Lenz、Vittorio Fra、Felix Christian Bauer、Dylan Richard Muir、Peng Zhou、Bernhard Vogginger、Kade Heckel、Gianvito Urgese、Sadasi- van Shankar、Terrence C. Stewart、Sadique Sheik 和 Jason K. Eshraghian。神经形态中间表示：面向可互操作类脑计算的统一指令集。《自然·通讯》，15(1):8122，2024年。doi: 10.1038/s41467-024-52259-9。URL https: //doi.org/10.1038/s41467-024-52259-9。

Alec Radford、Jong Wook Kim、Tao Xu、Greg Brockman、Christine McLeavey 和 Ilya S utskever。《通过大规模弱监督实现稳健的语音识别》，2022年。

Alexander Rast、Alan B Stokes、Sergio Davies、Samantha V Adams、Himanshu Akolkar、David R Lester、Chiara Bartolozzi、Angelo Cangelosi 和 Steve Furber。用于通用 AER 通信的传输无关协议。载于《神经信息处理》，《计算机科学讲义》第 9492 卷，第 675–684 页，美国，2015年11月。Springer Nature。ISBN 978-3-319- 26560-5。doi: 10.1007/978-3-319-26561-2_79。

Alpha Renner、Lazar Supic、Andreea Danielescu、Giacomo Indiveri、Bruno A. Olshausen、Yulia Sandamirskaya、Friedrich T. Sommer 和 E. Paxon Frady。《利用谐振器网络进行神经形态视觉场景理解》。*Nature Machine Intelligence*，6(6):641–652，2024年6月。ISSN 2522-5839。doi: 10.1038/s42256-024-00848-0。URL http://dx.doi.org/10.1038/s42256-024-00848-0。

Oliver Rhodes、Luca Peres、Andrew G. D. Rowley、Andrew Gait、Luis A. Plana、Christian Bren- ninkmeijer 和 Steve B. Furber。在神经形态硬件上进行实时皮层模拟。英国皇家学会哲学汇刊*A*辑：数学、物理与工程科学，378(2164):20190160，2019年12月。ISSN 1471-2962。doi: 10.1098/rsta.2019.0160。URL http: //dx.doi.org/10.1098/rsta.2019.0160。

Frank E Ritter、Farnaz Tehranchi 和 Jacob D Oury。Act-r：一种用于认知建模的认知架构。*Wiley* 跨学科评论：认知科学，10(3):e1488，2019年。

Curtis Roads。Tsukuba 音乐机器人。*Computer Music Journal*，10(2):39–43，1986年。ISSN 01489267、15315169。URL http://www.jstor.org/stable/3679483。

Yulia Sandamirskaya、Mohsen Kaboli、Jorg Conradt 和 Tansu Celikel。用于机器人学的神经形态计算硬件与神经架构。*Science Robotics*，7(67):eabl8419，2022年。

Kenny Schlegel、Peer Neubert 和 Peter Protzel。向量符号架构的比较。*Artificial Intelligence Review*，55(6):4523–4555，2021年12月。ISSN 1573-7462。doi：10.1007/ s10462-021-10110-3。URL http://dx.doi.org/10.1007/s10462-021-10110-3。

Gregor Schöner、John Spencer 和 DFT Research Group。动态思维：动态场理论入门。牛津大学出版社，2015年11月。ISBN 9780199300563。doi: 10.1093/acprof: oso/9780199300563.001.0001。URL https://doi.org/10.1093/acprof:oso/9780199300563.001. 0001。

Farzan Shenavarmasouleh、Farid Ghareh Mohammadi、M. Hadi Amini 和 Hamid Reza Arabnia。具身人工智能驱动的智慧城市运营：简明综述，第3章，第29–45页。约翰·威立父子有限公司，2022年。ISBN 9781119748342。doi: https://doi.org/10.1002/9781119748342.ch3。URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119748342.ch3。

Shweta Singh、Devansh Chaudhary、Amar Deep Gupta、Bhanu Prakash Lohani、Pradeep Kumar Kush- waha 和 Vimal Bibhu。人工智能、认知机器人学与意识的本质。载于*2022*年第三届智能工程与管理国际会议（*ICIEM*），第447–454页，2022年。doi：10.1109/ICIEM54221.2022.9853081。

Terrence C. Stewart、Xuan Choo 和 Chris Eliasmith。Spaun：一种基于脉冲神经元的感知-认知-行动模型。载于认知科学学会，第1018–1023页，2012年。

Jens Struckmeier。用于建筑物供暖和/或制备热水的供暖系统及方法，2018年4月。网址：https://patents.google.com/patent/US9958882B2/。美国专利。

Hugo Touvron、Louis Martin、Kevin Stone、Peter Albert、Amjad Almahairi、Yasmine Babaei、Nikolay Bashlykov、Soumya Batra、Prajjwal Bhargava、Shruti Bhosale、Dan Bikel、Lukas Blecher、Cris- tian Canton Ferrer、Moya Chen、Guillem Cucurull、David Esiobu、Jude Fernandes、Jeremy Fu、Wenyin Fu、Brian Fuller、Cynthia Gao、Vedanuj Goswami、Naman Goyal、Anthony Hartshorn、Saghar Hosseini、Rui Hou、Hakan Inan、Marcin Kardas、Viktor Kerkez、Madian Khabsa、Isabel Kloumann、Artem Korenev、Punit Singh Koura、Marie-Anne Lachaux、Thibaut Lavril、Jenya Lee、Diana Liskovich、Yinghai Lu、Yuning Mao、Xavier Martinet、Todor Mihaylov、Pushkar Mishra、Igor Molybog、Yixin Nie、Andrew Poulton、Jeremy Reizenstein、Rashi Rungta、Kalyan Saladi、Alan Schel- ten、Ruan Silva、Eric Michael Smith、Ranjan Subramanian、Xiaoqing Ellen Tan、Binh Tang、Ross Taylor、Adina Williams、Jian Xiang Kuan、Puxin Xu、Zheng Yan、Iliyan Zarov、Yuchen Zhang、An- gela Fan、Melanie Kambadur、Sharan Narang、Aurelien Rodriguez、Robert Stojnic、Sergey Edunov 和 Thomas Scialom。《Llama 2：开放基础模型和微调聊天模型》，2023年。

Gil Weinberg 和 Scott Driscoll。与拟人化打击乐演奏者的机器人人机交互。第 1229–1232 页，2006 年 04 月。doi：10.1145/1124772.1124957。Andrew Werchniak、Roberto Barra Chicote、Yuriy Mishchenko、Jasha Droppo、Jeff Condal、Peng Liu 和 Anish Shah。探索合成音频在训练关键词检测器中的应用。载于*ICASSP 2021 - 2021 IEEE*声学、语音与信号处理国际会议 *(ICASSP)*，第 7993–7996 页，2021年。doi：10.1109/ICASSP39728.2021.9413448。Yan Wu、Polake Kuvinichkul、Peter Cheung 和 Yiannis Demiris。迈向拟人化机器人特雷门琴演奏者。第 235–240 页，2011 年 01 月。doi：10.1109/ROBIO.2010.5723333。Lihe Yang、Bingyi Kang、Zilong Huang、Xiaogang Xu、Jiashi Feng 和 Hengshuang Zhao。Depth Anything：释放大规模无标注数据的力量。*arXiv:2401.10891*，2024年。Andrea Zanella、Nicola Bui、Angelo Castellani、Lorenzo Vangelista 和 Michele Zorzi。面向智慧城市的物联网。*IEEE*物联网期刊，1(1):22–32，2014年。doi：10.1109/JIOT. 2014.2306328。

Ye Zhao 和 Luis Sentis。第 2 章——采用串联弹性驱动的高延迟机器人系统的分布式阻抗控制。载于 Qingbin Gao 和 Hamid Reza Karimi 编，《时滞系统的稳定性、控制与应用》，第 23–51 页。Butterworth-Heinemann，2019 年。ISBN 978-0-12-814928-7。doi：https://doi.org/10.1016/B978-0-12-814928-7.00002-0。URL：https: // www.sciencedirect.com/science/article/pii/B9780128149287000020。