AI as Modality in Human Augmentation: Toward New Forms of Multimodal Interaction with AI-Embodied Modalities

Radu-Daniel Vatavu
MintViz Lab, MANSiD Research Center
Ştefan cel Mare University of Suceava
Suceava, Romania
radu.vatavu@usm.ro

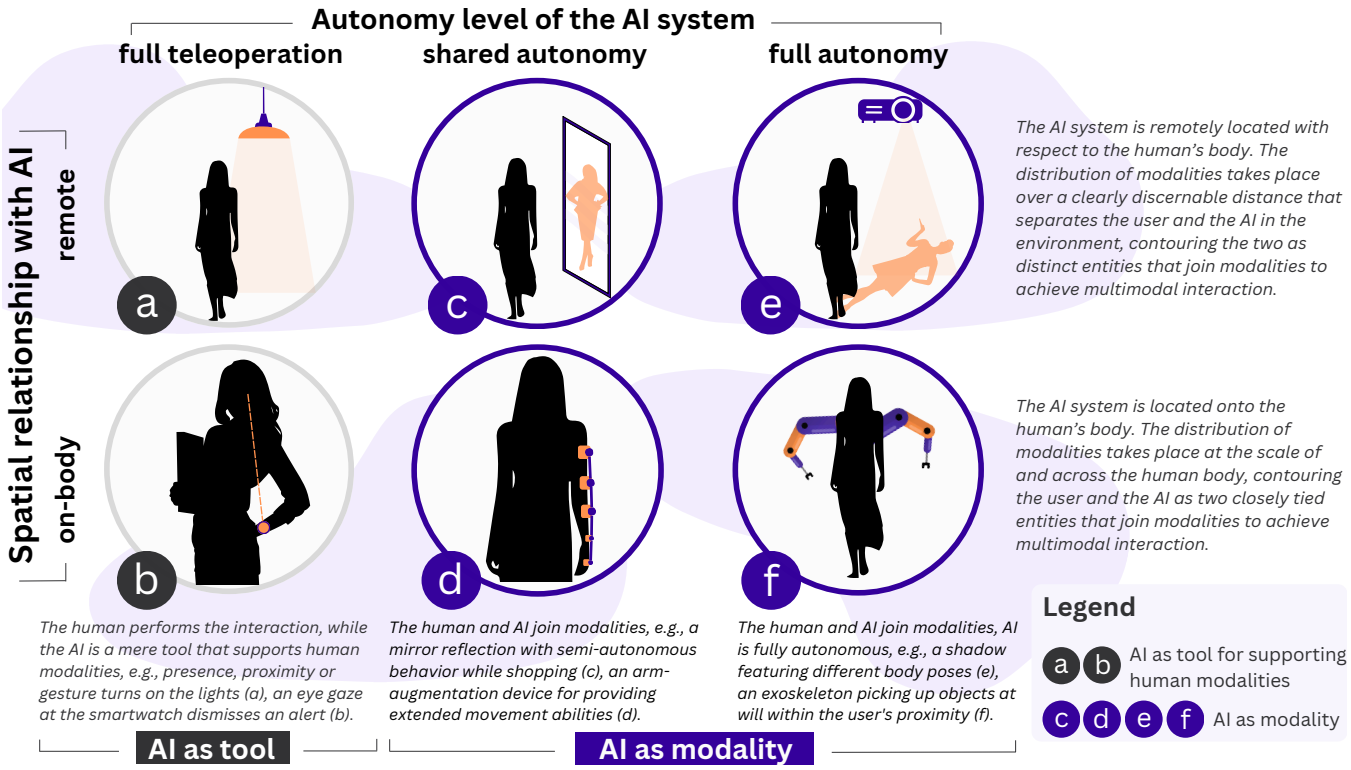

Figure 1: An overview of the "AI as modality" vision (c-f), where various AI embodiments deliver distinctive modalities from those of humans toward new forms of multimodal interaction, in contrast to the prevailing paradigm of "AI as tool" (a,b), where the AI merely supports human modalities. From top to bottom and left to right: (a) a smart environment detects users' presence or gestures to turn on the lights; (b) sensing eye gaze fixation in the direction of the smartwatch dismisses a notification; (c) a semi-autonomous mirror reflection of the user showcases various body poses during shopping; (d) a semi-autonomous arm-augmentation device detects the user's arm movements and performs fine adjustments to assist with power or precision grips; (e) a fully autonomous shadow, projected next to the user, provides a complementary modality for multimodal interaction in a smart environment assisted by a digital self; (f) a fully autonomous exoskeleton picks up objects in the user's proximity, independent of the user's will and actions, offering a complementary modality of a distinct dexterity than that of the human.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICMI '24, November 4–8, 2024, San Jose, Costa Rica
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0462-8/24/11
https://doi.org/10.1145/3678957.3678958

Abstract

AI techniques have always been central to multimodal interaction, being leveraged for sensing user intention, action, and emotion and for generating adaptive feedback. In this process, AI has played the role of enabling technology toward making interactions with computer systems and computerized environments more efficient, accessible, and natural through the prism of multimodality. In this work, we argue that AI can play yet another key role in multimodal interaction by implementing distinctive modalities complementing

those accessible by humans on their own. The result is multimodal interaction through human-AI integration, where multimodality is distributed across users and AI embodiments of various form factors, from robotic assistants to smart wearables to ambient devices. The new concept of "AI as modality," both contrasting and complementing the established "AI as tool," is particularly relevant for recent interaction paradigms involving human-computer integration and human augmentation. In this paper, we outline the vision of leveraging AI for delivering distinctive modalities toward the next generation of multimodal interactive experiences.

CCS Concepts

• **Human-centered computing → Interaction design theory, concepts and paradigms**; **Ubiquitous and mobile computing theory, concepts and paradigms**; **Interaction paradigms**.

Keywords

Artificial intelligence; AI; Multimodal interaction; Autonomous computer systems; Human augmentation; Wearable computing

ACM Reference Format:

Radu-Daniel Vatavu. 2024. AI as Modality in Human Augmentation: Toward New Forms of Multimodal Interaction with AI-Embodied Modalities. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '24), November 4–8, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3678957.3678958

1 Introduction

Over the last decades, Artificial Intelligence (AI) has become key to shaping our interactions with computer systems and environments of various kinds, from personal assistants [7] to robotic systems [32] to smart objects [2] and computerized environments [5] featuring adaptive interactions, context awareness, and behavior reflective of ambient intelligence [23]. The recent advances in generative AI techniques [15] have led to increased opportunities in interactive computer systems, ranging from sensible conversations delivered through large language models [30] to deep learning (DL)-powered generation of photorealistic visual content [39].

In the field of multimodal interaction, AI techniques have always been key to securing research advances, being leveraged for sensing natural user input, such as whole-body movements [1], gestures [37], voice [18], and eye gaze [28], for the fusion of heterogeneous signals [17], and for delivering effective feedback, e.g., by fission of audio and haptics [14]. From the early beginnings of ICMI, machine learning (ML) techniques have been recognized as a core component to drive developments in multimodal interaction. For example, the MLMI[1] workshop, which debuted [4] only a few years after ICMI, made the connection between ML and multimodal interaction explicit, and ICMI and MLMI were organized conjointly in 2009 [11] only to fuse two years later [8]. Many other examples abound. For instance, in his 2014 review of multimodal interaction, Turk [33] noted that "Fundamental improvements based on machine learning techniques are necessary for improved performance, personalization, and adaptability" (p. 193). A couple of years later, as part of the keynote speech at ICMI '20, Louis-Philippe Morency revisited multimodal research through the prism of multimodal ML, highlighting how "a broad and impactful body of research emerged in artificial intelligence under the umbrella of multimodal, characterized by multiple modalities" [19, p.1]. Beyond ML being core to multimodal interaction research and practice, the most recent call for papers of the ICMI '24 conference explicitly centers on AI as a key component of designing multimodal interactions: "ICMI is the premier international forum that brings together multimodal artificial intelligence (AI) and social interaction research."[2]

In this dynamic context of AI intertwining social interaction research, the current perspective on the use of AI, mostly ML/DL, in the ICMI community has primarily been that of *enabling technology* that accommodates and supports human modalities. While this perspective has worked well for many application areas of multimodal interaction, it falls short for emerging areas of computing involving new environments and technology for human augmentation. For example, the human-computer integration paradigm [13,20,38], emphasizing symbiosis and fusion between humans and computer systems, raises new questions about control and agency where integration complements interaction, providing distinctive nuances to what is to be human: "In designing the future of computing, it is no longer sufficient to think only in terms of the interaction between users and devices. We must also tackle the challenges and opportunities of integration between users and devices" [20, p. 1].

Contrary to the established perspective of "AI as tool" for *accommodating and supporting human modalities,* our goal in this paper is to put forth the vision of "AI as modality," where AI embodiments *generate new modalities, complementary to those accessible by users on their own,* toward new forms of multimodal interaction distributed across humans and AI alike. This includes semi-autonomous and fully autonomous AI embodiments that, from the vantage point offered by human augmentation, feature new modalities through, e.g., a sixth finger [27], an extra pair of arms [40], a video-projected body limb [34], and so forth; see Figure 1 for our proposal of a structured space of "AI as modality" involving the level of AI autonomy and the spatial relationship between the user's body and the AI embodiment. For example, an arm-augmentation device (Figure 1d) can implement semi-autonomous behavior for following the user's movements and enable powerful grasps and precise grips. In doing so, the device offers a distinctive modality that complements the user's movements with enhanced force and dexterity. A fully autonomous shadow projected next to the user's body (Figure 1e) can stretch and elongate in the environment, affecting overlaid digital devices, e.g., turn devices on/off. Due to its autonomous behavior, the shadow offers an additional modality that enhances the user's presence, proximity, and movement in that environment. These examples surface nuances of multimodal interaction through AI embodiments, not easy to conceptualize or address within the mainstream paradigm of "AI as tool," where AI systems merely support human modalities (Figures 1a and 1c).

In this context, our paper advocates for revisiting the commonly accepted notion of modality by proposing the new paradigm of "AI as modality" toward new opportunities for multimodal interaction distributed across humans and embodied AI systems. We believe that revisiting multimodal interaction in the context of human-computer integration for human augmentation by exploring novel

[1] International Workshop on Machine Learning for Multimodal Interaction

[2] https://icmi.acm.org/2024

types of modalities generated by AI embodying various form factors represents an exciting road toward a new generation of multimodal interaction design. In this paper, we describe this vision, present accompanying examples, and outline a roadmap for the future.

2 Modality in Human-Computer Integration and Human Augmentation

Multimodal interaction techniques enable users to employ various modes to interact with computer systems and, at large, within the inhabited computer-supported environment. Such interactions are mediated by multimodal UIs that, according to Turk [33], "seek to leverage natural human capabilities to communicate via speech, gesture, touch, facial expression, and other modalities, bringing more sophisticated pattern recognition and classification methods to human–computer interaction" (p. 189). According to this perspective, there are two objectives of multimodal interfaces, following Dumans *et al.* [12]. First, multimodal interfaces both support and accommodate users' perceptual and communicative capabilities. Second, they integrate computational skills of computers in the real world by offering more natural ways of interaction to humans.

In this context, multimodal interaction has focused on what is human and supporting existing human modalities. However, this established paradigm represents merely a historical consequence of how "modality" and "multimodal interaction" have been defined and understood in the community. For instance, in their work on multimodal integration, Blattner and Glinert [6] considered modality according to two orthogonal measures: input *vs.* output and human *vs.* computer: "Human input modalities refer primarily to our sensory abilities, whereas computer output modalities refer to information encodings and the nature of our interaction with them" (p. 14). Many other researchers have explicitly considered modality in relation to human modalities only. In his overview of challenges and perspectives in multimodal interfaces, Sebe [29] considered by modality "a mode of communication according to *human* senses and computer input devices activated by *humans* or measuring *human* qualities" (p. 24), emphasis ours. For technical surveys of multimodal interaction, we refer readers to Turk [33], Jaimes and Sebe [16], Dumas *et al.* [12], Oviatt [24], and Oviatt *et al.* [25] for complementary perspectives. However, is the mainstream paradigm of "AI as tool," where AI primarily supports and accommodates human modalities, still suitable for emerging application areas and contexts involving human-computer integration [13,20]? We challenge this status quo and propose that the future of multimodal interaction can transcend these limitations by radically redefining multimodal interaction through the integration of modalities delivered by AI embodiments, namely "AI as modality."

To move beyond the established paradigm, the concept of modality must be broader. To this end, we seek support in the perspectives of modality as information [22] and modality as experience [3], respectively. Even if these perspectives were introduced with humans at the center, e.g., "multiple sensing modalities give us a wealth of information to support interaction with the world and with one another." [33, p. 189], they provide a starting point for expanding beyond human senses, qualities, and evolutionarily acquired modes of communication and action. Since interaction always involves exchange of information, we employ a definition of modality from Nigay and Coutaz [22] as "the type of communication channel used to convey or acquire information" (p. 172). Furthermore, since interaction always generates a form of experience, we also employ a complementary definition, from Baltrusaitis, Ahuja, and Morency [3], that looks at modality as "the way in which something happens or is experienced" (p. 423). These definitions are sufficiently broad to extend beyond human perception and action, and we adopt them in this work. Next, we present our vision of "AI as modality," where multimodal interactions are designed to distribute across humans and AI embodiments.

3 Toward New Forms of Multimodal Interaction with AI-Embodied Modalities

Our discussion so far has highlighted that extending the semantic scope of modality across both humans and AI embodiments is needed to match technology advances in human-computer integration. In this section, we show how this extension can lead to new forms of multimodal interactions and advance multimodal interface design. Figure 1 provides a visual illustration of the "AI as modality" *vs.* "AI as tool" paradigms, structured along two dimensions:

(1) *The autonomy level of the AI embodiment:* full teleoperation, shared autonomy, and full autonomy. In full teleoperation, the human performs the interaction and the AI provides support for human modalities. In the shared and full autonomy categories, the human and AI join modalities.
(2) *The spatial relationship between the human and the AI embodiment*: remote in the environment or on the body. In the remote category, the distribution of modalities takes place over a clearly discernible distance that separates the user and the AI embodiment, contouring the two as distinct entities that join modalities. In the on-body category, the distribution of modalities takes place at the scale of and across the human body, contouring the user and the AI as two closely tied entities engaged in multimodal interaction.

Although other dimensions may also be useful to characterize practical opportunities enabled by AI as modality, such as interaction scale [20], we focus on just these two for the distinctive vantage points they bring to our discussion, i.e., human-environment interaction, where the AI is embodied within the environment, and on-body interaction, where the AI embodies a wearable form factor and, possibly, integrates within one's body image.

We start our discussion with Figures 1a and 1b that depict use cases where the AI system supports interactions performed exclusively through human modalities, i.e., "AI as tool." For example, the *remote full teleoperation* category subsumes systems within the environment, external to the user, that integrate some form of AI to sense, react, and assist the user within that environment. Common examples include natural interaction, where specific input modalities are used to implement remote control, such as proximity or a gesture flick to turn on the lights, as depicted in Figure 1a. The role of the AI system in this case is that of supporting technology for recognizing user action and providing a proper response. Similarly, the *on-body full teleoperation* category subsumes systems designed to be worn, which integrate a form of AI. Common examples include smart wearables, such as a smartwatch that dismisses a notification upon eye gaze fixation, as illustrated in Figure 1b. Just like in the

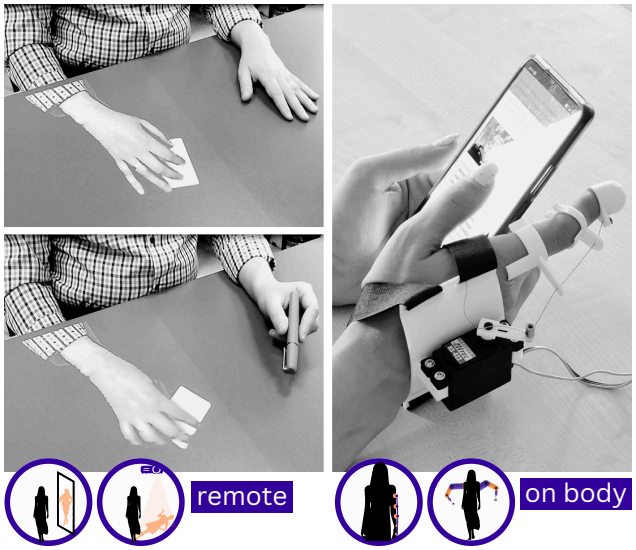

Figure 2: Examples of multimodal interaction distributed across humans and AI embodiments in scenarios of human-computer integration involving technology for the environment (left) and the body (right). On the left, a virtual hand projected within the environment enables bimanual interaction across the physical and the virtual. On the right, a finger-augmentation device puts the finger into specific poses to deliver information through kinesthesia. In both cases, multimodal interaction emerges through the AI embodiment.

previous category, the role of the AI system is to accommodate and support, as enabling technology, existing human modalities.

Reconceptualizing multimodal interaction by incorporating modalities delivered through AI embodiments becomes relevant when AI systems, designed for human augmentation, expose semi-autonomous or fully autonomous behavior, as depicted in Figures 1c to 1f. Various combinations across the two dimensions of our framework give rise to novel forms of multimodal interaction, as follows.

Multimodal interaction in human augmentation through AI embodied within the environment. In this category, AI systems within the environment provide modalities that complement the user's own toward achieving a task through multimodal interaction distributed across the human and the AI. For example, a video projection of the user's silhouette can stand as a new modality to reach within the environment beyond the physical limitations of one's body and engage in new interactions. The "Shadow Reaching" interaction technique of Shoemaker *et al.* [31] was designed to facilitate manipulation of digital content on wall displays over large distances by making use of a perspective projection applied to a shadow representation of the user. While in the original technique the shadow is fully controlled by the user, it is easy to imagine a semi-autonomous AI driving the shadow representation that would feature proactive behavior to best assist the user, such as shadow elongation toward specific targets taking the form of feedforward that is aligned to the environment's characteristics [21]. Beyond shared autonomy, a fully autonomous AI embodiment would expose independent behavior from the user and, thus, determine a different user experience. For example, Vatavu [34] played with the concept of hybrid bodies where a virtual arm was video-projected onto a tabletop and aligned with the user's physical body. The virtual hand plays movements that are independent of those performed by the physical hand, resulting in bimanual interactions that require two actors, the user and the AI embodiment. For example, the physical hand would hold a pen and the virtual hand a digital piece of paper, enabling interactions that take place across the boundary of the physical and virtual worlds and across humans and AI embodiments, respectively; see Figure 2, left. In these examples, multimodality surfaces from human augmentation delivered through elements of the environment, external to the user's body.

Multimodal interaction in human augmentation through AI embodiments close to the body. In this category, the AI system is integrated into a form factor that is in a close relation to the user's body, from where it provides the extra modality. For example, the sixth finger prototype of Prattichizzo *et al.* [27] was designed to enhance physical objects manipulation and enlarge the workspace of humans through increased hand dexterity. The sixth finger device is driven through an object-based mapping algorithm by sensing the hand's movements. The result is grasping abilities of objects larger than possible to grasp with the five fingers alone. Another example is JIZAI ARMS [40], a supernumerary robotic limb system consisting of detachable arms, designed to enable social interaction between multiple wearers, such as an exchange of arms, and explore possible interactions in a cyborg society. Lastly, Finger-hints [9] is a finger-augmentation device designed for kinesthetic feedback, which operates by placing the finger into a state of hyper-extension, trading off user agency for information delivery through the user's body; see Figure 2, right. The processes implemented by these devices are forms of multimodal interaction, where the extra modality is provided through AI embodiment.

4 Next Steps for AI as Modality

As technological advances in human-computer integration and augmentation continue, there is a need for a roadmap to achieve the vision of "AI as modality" within the ICMI community. We envision several stages in this regard. First, understanding how multimodal interaction encompasses modalities originating from both humans and AI embodiments is crucial. This may need revisiting established models of multimodal interaction, such as CASE [22] and CARE [10], by integrating aspects related to AI embodiment, autonomy, user agency, and so forth. Second, new conceptual frameworks are needed for multimodal interaction in human-computer integration, including frameworks that capitalize on human abilities mediated by computer technology [35], paradigms that incorporate multimodal interaction featuring natural *vs.* non-natural modalities [36], and philosophical and cultural frameworks for examining physical-digital environments designed to amplify, augment, mediate, and extend human sensorimotor abilities and intelligence [26]. Third, new opportunities arise in designing multimodal interaction across humans and AI embodiments for various application areas and contexts of use, including understanding the user experience of such novel interactions. These directions promise an exciting journey for the ICMI community in redefining multimodal interaction toward the next generation of multimodal interactive experiences.

Acknowledgments

This work was conducted within the MintViz Research Laboratory. This work is also supported by the NetZeRoCities Competence Center, funded by European Union - NextGenerationEU and Romanian Government, under the National Recovery and Resilience Plan for Romania, contract no. 760007/30.12.2022 with the Romanian Ministry of Research, Innovation and Digitalization through the specific research project P3, Smart and Sustainable Buildings.

References

[1] Aishat Aloba, Julia Woodward, and Lisa Anthony. 2020. FilterJoint: Toward an Understanding of Whole-Body Gesture Articulation. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20)*. ACM, New York, NY, USA, 213–221. https://doi.org/10.1145/3382507.3418822

[2] Alexandru-Tudor Andrei, Laura-Bianca Bilius, and Radu-Daniel Vatavu. 2024. Take a Seat, Make a Gesture: Charting User Preferences for On-Chair and From-Chair Gesture Input. In *Proc. CHI Conf. on Human Factors in Computing Systems*. ACM, USA, Article 555, 17 pages. https://doi.org/10.1145/3613904.3642028

[3] Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2 (2019), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

[4] Samy Bengio and Hervé Bourlard. 2004. *Proceedings of the 1st International Workshop on Machine Learning for Multimodal Interaction, Lecture Notes in Computer Science*. Springer, Berlin. https://doi.org/10.1007/b105752

[5] Laura-Bianca Bilius, Alexandru-Tudor Andrei, and Radu-Daniel Vatavu. 2024. From Smart Buildings to Smart Vehicles: Mobile User Interfaces for Multi-Environment Interactions. In *Proc. of the Int. Conf. on Development and Application Systems*. IEEE, USA, 152–155. https://doi.org/10.1109/DAS61944.2024.10541208

[6] Meera M. Blattner and Ephraim P. Glinert. 1996. Multimodal Integration. *IEEE MultiMedia* 3, 4 (1996), 14–24. https://doi.org/10.1109/93.556457

[7] Michael Bonfert, Maximilian Spliethöver, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. In *Proc. of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. ACM, New York, NY, USA, 95–102. https://doi.org/10.1145/3242969.3242995

[8] Hervé Bourlard, Thomas S. Huang, Enrique Vidal, Daniel Gatica-Perez, Louis-Philippe Morency, and Nicu Sebe. 2011. *Proc. of the 13th Int. Conf. on Multimodal Interfaces*. ACM, New York, NY, USA. https://dx.doi.org/10.1145/2070481

[9] Adrian-Vasile Catană and Radu-Daniel Vatavu. 2023. Fingerhints: Understanding Users' Perceptions of and Preferences for On-Finger Kinesthetic Notifications. In *Proc. CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, USA, Article 518, 17 pages. https://doi.org/10.1145/3544548.3581022

[10] Joëlle Coutaz, Laurence Nigay, Daniel Salber, Ann Blandford, Jon May, and Richard M. Young. 1995. Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The Care Properties. In *IFIP Advances in Information and Communication Technology*. Springer, Boston, MA, USA, 115–120. https://doi.org/10.1007/978-1-5041-2896-4_19

[11] James L. Crowley, Yuri Ivanov, Christopher Wren, Daniel Gatica-Perez, Michael Johnston, and Rainer Stiefelhagen. 2009. *Proc. of the 2009 Int. Conf. on Multimodal Interfaces*. ACM, New York, NY, USA. https://dx.doi.org/10.1145/1647314

[12] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. 2009. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In *Human Machine Interaction: Research Results of the MMI Program*, Denis Lalanne and J. Kohlas (Eds.). Springer, Berlin, Heidelberg, 3–26. https://doi.org/10.1007/978-3-642-00437-7_1

[13] Umer Farooq and Jonathan Grudin. 2016. Human-Computer Integration. *Interactions* 23, 6 (oct 2016), 26–32. https://doi.org/10.1145/3001896

[14] Darren Guinness, Annika Muehlbradt, Daniel Szafir, and Shaun K. Kane. 2018. The Haptic Video Player: Using Mobile Robots to Create Tangible Video Annotations. In *Proc. of the 2018 ACM Int. Conf. on Interactive Surfaces and Spaces (ISS '18)*. ACM, New York, NY, USA, 203–211. https://doi.org/10.1145/3279778.3279805

[15] Priyanka Gupta, Bosheng Ding, Chong Guan, and Ding Ding. 2024. Generative AI: A systematic review using topic modelling techniques. *Data and Information Management* June (2024), 100066. https://doi.org/10.1016/j.dim.2024.100066

[16] Alejandro Jaimes and Nicu Sebe. 2007. Multimodal Human–Computer Interaction: A Survey. *Computer Vision and Image Understanding* 108, 1 (2007), 116–134. https://doi.org/10.1016/j.cviu.2006.10.019

[17] Paul Pu Liang, Yun Cheng, Ruslan Salakhutdinov, and Louis-Philippe Morency. 2023. Multimodal Fusion Interactions: A Study of Human and Automatic Quantification. In *Proc. of the 25th Int. Conf. on Multimodal Interaction (ICMI '23)*. ACM, New York, NY, USA, 425–435. https://doi.org/10.1145/3577190.3614151

[18] Tiantian Liu and Feng Lin. 2024. A New mmWave-Speech Multimodal Speech System for Voice User Interface. *GetMobile: Mobile Comp. and Comm.* 27, 4 (2024), 31–35. https://doi.org/10.1145/3640087.3640096

[19] Louis-Philippe Morency. 2022. What is Multimodal?. In *Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22)*. ACM, New York, NY, USA, 1. https://doi.org/10.1145/3536221.3565369

[20] Florian Floyd Mueller, Pedro Lopes, Paul Strohmeier, Wendy Ju, Caitlyn Seim, Martin Weigel, Suranga Nanayakkara, Marianna Obrist, Zhuying Li, Joseph Delfa, Jun Nishida, Elizabeth M. Gerber, Dag Svanaes, Jonathan Grudin, Stefan Greuter, Kai Kunze, Thomas Erickson, Steven Greenspan, Masahiko Inami, Joe Marshall, Harald Reiterer, Katrin Wolf, Jochen Meyer, Thecla Schiphorst, Dakuo Wang, and Pattie Maes. 2020. Next Steps for Human-Computer Integration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3313831.3376242

[21] Andreea Muresan, Jess Mcintosh, and Kasper Hornbæk. 2023. Using Feedforward to Reveal Interaction Possibilities in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 30, 6, Article 82 (sep 2023), 47 pages. https://doi.org/10.1145/3603623

[22] Laurence Nigay and Joëlle Coutaz. 1993. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 172–178. https://doi.org/10.1145/169059.169143

[23] Anton Nijholt. 2004. Multimodality and Ambient Intelligence. In *Algorithms in Ambient Intelligence*, W.F.J. Verhaegh, E. Aarts, and J. Korst (Eds.). Springer, Dordrecht, 23–53. https://doi.org/10.1007/978-94-017-0703-9_2

[24] Sharon Oviatt. 1999. Ten Myths of Multimodal Interaction. *Commun. ACM* 42, 11 (nov 1999), 74–81. https://doi.org/10.1145/319382.319398

[25] Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, and Antonio Krüger (Eds.). 2017. *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations*. ACM and Morgan & Claypool, USA. https://doi.org/10.1145/3015783

[26] Bogdan Popoveniuc and Radu-Daniel Vatavu. 2022. Transhumanism as a Philosophical and Cultural Framework for Extended Reality Applied to Human Augmentation. In *Proc. of the 13th Augmented Human Int. Conf. (AH '22)*. ACM, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/3532525.3532528

[27] Domenico Prattichizzo, Monica Malvezzi, Irfan Hussain, and Gionata Salvietti. 2014. The Sixth-Finger: A Modular Extra-Finger to Enhance Human Hand Capabilities. In *Proc. of the 23rd IEEE Int. Symp. on Robot and Human Interactive Communication*. IEEE, USA, 993–998. https://doi.org/10.1109/ROMAN.2014.6926382

[28] Pernilla Qvarfordt. 2017. *Gaze-informed Multimodal Interaction*. ACM and Morgan & Claypool, USA, 365–402. https://doi.org/10.1145/3015783.3015794

[29] Nicu Sebe. 2009. Multimodal Interfaces: Challenges and Perspectives. *J. Ambient Intell. Smart Environ.* 1, 1 (2009), 23–30. https://doi.org/10.3233/AIS-2009-0003

[30] Murray Shanahan. 2024. Talking about Large Language Models. *Commun. ACM* 67, 2 (jan 2024), 68–79. https://doi.org/10.1145/3624724

[31] Garth Shoemaker, Anthony Tang, and Kellogg S. Booth. 2007. Shadow Reaching: A New Perspective on Interaction for Large Displays. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 53–56. https://doi.org/10.1145/1294211.1294221

[32] Hang Su, Wen Qi, Jiahao Chen, Chenguang Yang, Juan Sandoval, and Med Amine Laribi. 2023. Recent Advancements in Multimodal Human–Robot Interaction. *Front. Neurorobot.* 17 (2023), 21 pages. https://doi.org/10.3389/fnbot.2023.1084000

[33] Matthew Turk. 2014. Multimodal Interaction: A Review. *Pattern Recognition Letters* 36 (2014), 189–195. https://doi.org/10.1016/j.patrec.2013.07.003

[34] Radu-Daniel Vatavu. 2022. Possi(A)bilities: Augmented Reality Experiences of Possible Motor Abilities Enabled by a Video-Projected Virtual Hand. In *Proc. of the 27th International Symposium on Electronic Art (ISEA '22)*. ISEA International, Rotterdam, 825–828. https://doi.org/10.7238/ISEA2022.Proceedings

[35] Radu-Daniel Vatavu. 2022. Sensorimotor Realities: Formalizing Ability-Mediating Design for Computer-Mediated Reality Environments. In *Proc. of the 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '22)*. IEEE, USA, 685–694. https://doi.org/10.1109/ISMAR55827.2022.00086

[36] Radu-Daniel Vatavu. 2023. From Natural to Non-Natural Interaction: Embracing Interaction Design Beyond the Accepted Convention of Natural. In *Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23)*. ACM, New York, NY, USA, 684–688. https://doi.org/10.1145/3577190.3616122

[37] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as point clouds: A $P recognizer for user interface prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*. ACM, New York, NY, USA, 273–280. https://doi.org/10.1145/2388676.2388732

[38] Sang won Leigh, Harshit Agrawal, and Pattie Maes. 2018. Robotic Symbionts: Interweaving Human and Machine Actions. *IEEE Pervasive Computing* 17, 02 (apr 2018), 34–43. https://doi.org/10.1109/MPRV.2018.022511241

[39] Weihao Xia and Jing-Hao Xue. 2023. A Survey on Deep Generative 3D-aware Image Synthesis. *ACM Comput. Surv.* 56, Article 90 (nov 2023), 34 pages. https://doi.org/10.1145/3626193

[40] Nahoko Yamamura, Daisuke Uriu, Mitsuru Muramatsu, Yusuke Kamiyama, Zendai Kashino, Shin Sakamoto, Naoki Tanaka, Toma Tanigawa, Akiyoshi Onishi, Shigeo Yoshida, Shunji Yamanaka, and Masahiko Inami. 2023. Social Digital Cyborgs: The Collaborative Design Process of JIZAI ARMS. In *Proc. of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, USA, Article 369, 19 pages. https://doi.org/10.1145/3544548.3581169