CHAPTER TWENTY-NINE

Controllers & Inputs

Masters of Puppets

DANIEL ROTH, JEAN-LUC LUGRIN, SEBASTIAN VON MAMMEN, & MARC ERICH LATOSCHIK

Picture vibrant cities of the future or battlefields of the past—we can dive right into these fantastic scenarios by means of modern technology. The digital characters that populate these and various other digital worlds are distinct. Some are referred to as "agents" and are driven by the machine—they wander about and behave based on algorithmic instructions. Others are "avatars" and are driven by human movement, like puppets: "The interaction of the player with the video-game is the puppetry. Puppetry describes how the player starts approaching the video-game until eventually the game being played is the outcome of the actions of the player" (Calvillo-Gámez, Cairns, & Cox, 2015, p. 47). As the puppet Pinocchio's creation was famously characterized by his maker Gepetto: "The legs and feet still had to be made. As soon as they were done, Geppetto felt a sharp kick on the tip of his nose. 'I deserve it!' he said to himself. I should have thought of this before I made him. Now it's too late!" (Collodi, 1881/2012).

Although Collidi's Pinocchio and Gepetto seem to be disconnected entities, it is Gepetto's intentions and physical actions, in combination with his senses, that *control* his wooden companion. To become the puppeteer of an avatar, breathing life into its digital body, *controls* are a necessity for every interactive system and are, in contrast to other components, "universally applicable to every style of game" (Rogers, 2014, p. 163). As the linking channel between the user's intentions and the avatar behavior, controllers and inputs are a crucial component of avatar-based systems. A *controller* in this sense can have two distinct meanings, either in terms of a physical input device (e.g., a joystick) or in terms of a part of a system. The *input* is the information the physical device is sensing and delivering to the system.

THE USER AS THE CONTROLLER

Just as this book treats avatars as an assemblage—as a system of interconnected parts—avatars are, in turn, part of the game as an even more complex system. By definition, a system is a collection of components linked together and organized to be recognizable as a single unit (Englander & Englander, 2003). Accordingly, a system can be characterized by its interconnectivity, the structural organization and the behavior of its components. In systems theory, a controller (as part of a control system) obtains, in combination with other components (*actuator* and *process*) a desired system response. Consider, for instance, using a microwave stove. It typically requires the user to set a timer to a given time (the input, via the controller). The stove (actuator) will then heat the food (process) until the timer runs out. The set time is the input of the system, whereas the timer represents the controller. Such a simple system, an *open-loop* control system, does not use feedback to determine if the goal is reached, and does not compensate for any disturbances. Its input-output relationship can also be described as a cause-effect relationship (Dorf & Bishop, 1998). Alternatively, the controller might be informed by the system's state, i.e., the temperature of the heated food. In modern control systems, this feedback control is an essential part of most systems. In so-called *closed-loop* feedback control systems, the actual output and feedback is measured using a sensor and compared to the desired output response (Dorf & Bishop, 1998). In our microwave example, the food temperature can be measured by a sensor and compared to the reference input (desired temperature) by the controller. The controller may consequently increase or shorten the remaining heating time.

Applying this logic to controlling avatars in games or other digital environments, the human becomes part of a real-time interactive system (RIS), a system that senses external events (user actions), processes the input, and provides corresponding outputs of this interaction loop at real-time speeds (Englander & Englander, 2003). In a way, we can describe any RIS, including games and interactive digital environments, as combinations of actuators and sensors to perceive the users' actions, the users themselves as controlling (and consuming) entities, rules that determine system behavior based on its state and any given inputs, and actuators that provide feedback to the users. Specifically, the user can control various attributes of component sets of the system (von Mammen, 2016), for example, by pressing a key to move an avatar. The keyboard provides the information that drives the system behavior, e.g., the movement of the avatar, and a screen provides the output to the user. It is the user's task to close the loop of information flow, to utilize the feedback to adjust his control actions and thus achieve a given intention or goal (Figure 29.1). Taking on the job of the controller is typically the challenge that brings about the fun in videogames (Koster, 2013).

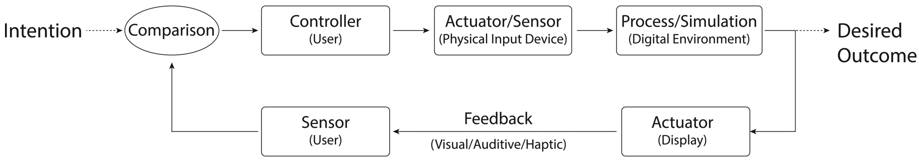

Figure 29.1. Closed-loop control (real-time interactive) system for avatar control.
(Source: author, adapted from Dorf & Bishop, 1998; Englander & Englander, 2003)

The human player is the entity responsible for driving its avatar-counterpart to perform an action (i.e., deliver input), and the computer or console is the hardware and software that is responsible for simulating the avatar action and its respective visual feedback. Human and computer are interacting parts of one system. The system's design defines the input-output structure of the loop of interaction. In our keyboard example, the human (as controller) interacts with other parts of the system using a mechanical controller (a physical device) as actuator. This notion brings us to the second meaning of the term controller in our context, the controller as an *input device*.

THE CONTROLLER AS AN INPUT DEVICE

While it may be intuitive to think about a game interface in terms of things that happen on a screen (cf. Limperos & Stevens, this volume), devices used to influence or directly control on-screen events are equally important. Here, a controller is an input device for communicating from the user to the computer—a technical construct that measures activity of the user in one or more physical or physiological dimensions (degrees of freedom), and transforms this information into digital data (input) digestible by a computer system.

Controlling an avatar, then, refers to the central, dynamic relationship between the user and the system, in which the input device acts as a medium through which the user's intentions are conveyed and the user-avatar relationship in part unfolds. This medium can channel various forms of user activity depending on the affordances and capabilities of the input device, the game or digital environment and its platform, and the user's intentions. Input devices can be purely passive, if they continuously create data without any physical interaction with the device (e.g., controlling the avatar's eyes through eye-tracking sensors that sense the user's gaze direction), or purely active, if physical interactions with the device are required (Bowman et al., 2004).

In general, we can distinguish between two forms of input signals for controlling an avatar: *discrete* input, such as a signal generated by a brief keyboard button press, and *continuous* input, such as the held horizontal position of a joystick or a maintained mouse position. For the sake of illustration, consider the controls in sports games such as the *FIFA*, *NBA*, or *NFL* series by Electronic Arts in which passing the ball from one digital ball player to another is discrete action evoked by a discrete input (a single button press), whereas steering the ball player around a digital football field requires continuous input (holding a joystick position). Discrete inputs report a single data value (e.g., button pressed: yes or no) that can change over continuous time, whereas continuous inputs report a continuous value (e.g., axial position of the joystick) changing over continuous time. In this regard, our definition holds a flaw, as continuous *digital* input devices are not actually time continuous or value continuous. Rather, they capture a precise measure of discrete values over discrete sampling periods (frequency of data) which can be considered quasi-continuous. These frequencies (typically measured in Hertz) can vary to large extend and can impact on the overall processing time for the input.

The time between when a user provides input and this input is processed until the appropriate feedback is created can be critical for the user experience of both continuous and discrete input. If the period takes too long, the user or player might experience latency or "lag" (see Johnson, this volume), or a crucial input sample might be missed leading to system misinterpretations of the user's actual intentions. For example, to make Mario perform a higher jump (double jump) in *Super Mario 64* (1996) the player needs to press the jump button twice in a narrow time window. If the avatar control system's input sampling frequency is too low, it is not capable of detecting both button presses and the process does not result in the desired outcome due to the system's limitation. Such deviations may result in a degraded game experience, e.g., poor response times and "sticky" controls, which may lead to player frustration. In general, latency in digital environments can result in decreased efficiency or negative training effects, degraded vision, degraded performance, breaks in presence, and can be a cause for cyber sickness (Jerald, 2015). In these ways, accurate translation of both discrete and continuous inputs is essential for a successful interaction.

CONTROLLING AVATARS THROUGH CONTROL SCHEMES

Considering the dynamics noted above, interaction relies on specific mappings of physical actions to digital data and a system response to perform a task; these mappings are known as interaction techniques (Foley, van Dam, Feiner, & Hughes, 1990), and a combination of interaction techniques and their semantic mappings can be described as a control scheme. Technically speaking, a control scheme is the conceptual framework around which system components and feedback are structured; in games and other digital environments they include the schematic mapping of multiple user actions to multiple avatar mechanics, as well as a translation interface between the (human) physical behaviors and the (avatar) digital behaviors. Let us refer to the double jump example once again. While if we tried to physically jump twice, the height of our second jump would not stack atop the height of our first jump. However, when controlling Mario, we press a button twice and the avatar's behavioral mechanic is executed, i.e., its animation state changes. This abstraction of the intended physical behavior (jumping) as it is mapped onto a controlling device (Gamepad Button A) together with the semantic mapping of power ("push button twice") define the design of the control scheme. Control schemes are determined by the general design, the expressivity, and the complexity of the controller device, which can be measured by the degrees of freedom and input device affordances.

A BRIEF HISTORY OF CONTROLLERS AND CONTROL SCHEMES

While a more detailed history of game controllers and their influence on games warrants a tome of its own (see Cummings, 2007, to start), it is worth offering a short review. Early controllers of *Tennis for Two* (1958) *or Spacewar!* (1961) were custom-developed digital control boxes consisting of discrete digital button and knob inputs that evolved to paddles and joysticks, and eventually to gamepads. With the 1990s and 2D/3D first-person shooters such as the *Doom* (1993) and the *Quake* (1996) series, new control schemes evolved that used the keyboard keys (e.g., arrow keys or WASD) to steer and rotate. Later, mouse control allowed controlling avatar movements and the camera's rotational perspective (e.g., "mouselook" in *Quake*) or selecting and navigating in real-time strategy games.

Beyond mainstream formats, less usual control devices were developed, including rhythm pads for dance games, light guns, and even drums for music games. Striving for more naturalness in control, touch and multi-touch approaches link the human finger input with control schemes such as, double tab, timed tab, hold and drag, touch and hold, and swipe (Rogers, 2014), to give the user greater control over the applications. Interfaces that leverage touch, gesture, gaze, speech, or handwriting to control an application are often referred to as "natural user interfaces," or NUIs. Fostering the transportation of emotion, algorithms and sensors today offer possibilities for real-time facial control of an avatar (Weise, Bouaziz, Li, & Pauly, 2011). New motion-controllers such as Leap Motion are especially interesting for avatar control, as they enable systems to sense the spatial (biological) motion of the human hand and to transfer the input to avatar behavior. Similarly, full-body motion controllers (e.g., Microsoft Kinect) are utilized to detect the body's pose, as in in the dance training game *Kinect Dance Central 3* (2012) where the player's dancing moves are mapped to the avatar. Modern algorithms and frameworks can detect gestures, poses, and social signals from behavioral data and other modalities. Marker-based motion tracking techniques allow for more precision in motion control by using multiple reference points placed on the human source body, and sometimes these reference points can be reduced in numbers by applying inverse kinematic algorithms. Compared to other input device types, motion tracking establishes a more direct translation between the movement of the user in the physical world and avatar movement in the digital world by utilizing the numerous degrees of freedom of the human body. Given compatible kinematic structures between the user and the avatar (that is, matching movements and forces), it almost allows a real 1:1 mapping between user (source) actions and avatar (target) movements, mirroring the original movements in real-time.

By means of such controllers and control schemes, the user "puppeteers" the avatar which results in a *sense of control*, making the user feel responsible for the avatar's behavior (Calvillo-Gámez et al., 2015). In the physical world, the "sense of intending and executing actions, including the feeling of controlling one's own body movements, and, through them, events in the external environment" (Tsakiris, Prabhu, & Haggard, 2006, p. 424) is understood as *agency*, which only evolves from voluntary actions of control.

AVATAR EMBODIMENT AND VIRTUAL BODY OWNERSHIP

Controlling an avatar and exploring digital environments essentially creates a connection between the character and the user. In the physical world, we experience *embodiment*—a sense of being present with our own body and having a sense of it (corporal-awareness, self-awareness; Longo, Schüür, Kammers, Tsakiris, & Haggard, 2008). Through embodiment, we experience body ownership, or the perception of our "own body as the source of sensations" (Tsakiris et al., 2006, p. 424). Under certain circumstances, it is possible to trick the human mind into perceiving an *illusory body ownership* for other objects and accept them as part of one's own body. First experimentally investigated inducing an illusion of owning a rubber hand as part of one's body (Botvinick & Cohen, 1998) researchers found this phenomenon extended to digital bodies (Slater, Pérez Marcos, Ehrsson, & Sanchez-Vives, 2008). By utilizing motion-tracking technologies for kinematic avatar control, the level of perceived control over the avatar (and therefore the level of agency) may be improved. As agency is a concept strongly related to the illusion of body ownership, and the coupled avatar can represent the human behaviors to a large degree, the system can induce an *illusion of virtual body ownership*, ranging from the assumption of single body parts to whole, digital bodies (Slater et al., 2009). Researchers started to investigate driving factors of virtual body ownership illusions. Reviewing the findings, processes such as the identity and synchronicity between visual and motor perception (sensuo-motor coherence), synchronicity between visual and tactile stimuli (visuo-tactile), as well as the visualization of the environment from an avatar's first-person perspective are main factors for the illusion of virtual body ownership, which may be influenced by the character's realism (Maselli & Slater, 2013).

The importance of the first-person perspective emphasizes the interplay between controls and camera. In contrast to two-dimensional displays, virtual reality displays (especially head mounted displays) directly link the user's head position and orientation to the camera perspective in the digital environment. The user literally steps into the shoes of the avatar, evoking vivid, immersive experiences. The technical realization of virtual body ownership illusions poses a challenge. It was shown that increased end-to-end latencies reduce motor performance and body ownership (Waltemate et al., 2016). Therefore, engineers must solve current issues of latency, lag, and jitter (Rogers, 2014) and regain control over timeliness and the game/simulation loop.

THE FUTURE IN AUTONOMY AND HYBRID SYSTEMS

Perceptually real controls turn *steering* Mario through a digital world into *being* Mario in a digital world. Consumer VR and behavioral controller technologies open a large design space for future development of digital worlds populated by avatars, agents, and even hybrids (Roth, Latoschik, Vogeley, & Bente, 2015). Part of the human control will then be in the hands of the machine, modifying interactions and behaviors for the sake of better virtual rapport (Gratch et al., 2007) and nonverbal synchrony (Roth et al., 2015). Avatar systems can help to understand, assess, and train communicative impairments (Georgescu, Kuzmanovic, Roth, Bente, & Vogeley, 2014) and the appropriate AI may transform to adequate behaviors. Hybrid forms of "ourselves" enable constant embodied conversation and co-presence, whether connected or disconnected from our digital "selves" (Gerhard, Moore, & Hobbs, 2004). Social artificial intelligence (AI) will allow for multiple, fully registered interactions at the same time and for transporting messages by means of newly learned behavioral patterns not accessible in the physical world. Gestures, speech, and social reactions will be learned and adapted across intercultural differences, serving and mediating even among large groups of people.

With the rising complexity of hybrid avatar/agent systems, we will soon reach the limits of natural inter-human communication. To this end, brain-computer interfaces and implants may provide solutions for new dimensions of sensing and display. Avatars will be the "interface" to our "selves" in these systems, but for puppeteering multiple selves in hybrid systems, metaphorically, we again find ourselves in another exciting era of *Pong*.

With individual programs controlling the digital agents in future digital worlds, the theme of self-organization will play an increasingly important role. While each avatar or agent in a system may only act based on locally provided information, many of them together may have an impact on the digital world that only a few could not possibly achieve (Schmeck, Müller-Schloer, Çakar, Mnif, & Richter, 2010). Harnessing this distributed constructive power by only a few (or even only one player or avatar) motivates the field of human-swarm interaction (von Mammen, 2016). It investigates according interfaces that provide the means to instruct, influence, or inspect groups of agents that otherwise act autonomously. In the end, it is not only a pragmatic but also a philosophical and ethical question: to what degree human society will trust the machine to puppeteer, and to what degree we take control ourselves.

To ensure that we can engage with digital worlds of considerable complexities, we need to promote the incorporation of multimodal and neural sensing and control as components of avatars. We need to master and innovate the "avatar" as the "interface" in computer-mediated communication and games. This implies that avatars need to become perfectly "natural" to increase behavioral realism, social presence, and trust (Bente, Rüggenberg, Krämer, & Eschenburg, 2008). Part of this naturalness lies in the user's opportunity to detect and react to subtle cues of behavioral control. When translated to the domain of embodied agents and hybrid systems, the underlying social AI needs to be developed as well. Beyond mere reactive AIs that seek perfection in mimicking natural presentations, AIs will need to serve and promote the users' individuality and consider the multifaceted, open-ended ways of human communication. Not unlike organic computing systems (Schmeck et al., 2010), such AIs may draw from expert knowledge and learn models based on large data sets, but they also need to learn continuously during runtime. To foster the rapid and diverse representation of individual behaviors, AIs for avatar-agent systems necessarily need adapt information about the human controller's intuition and self-awareness. In symbiosis with the human controller and as part of one control system, machines need to process and learn signals that remain unconscious for the human and their underlying neural correlates (Roth et al., 2015).

Returning to the puppet and puppeteer:

> "'There he is,' answered Geppetto. And he pointed to a large Marionette leaning against a chair, head turned to one side, arms hanging limp, and legs twisted under him. After a long, long look, Pinocchio said to himself with great content: 'How ridiculous I was as a Marionette! And how happy I am, now that I have become a real boy!'"

The avatar of the future might strike us in the very same way, opening doors to new worlds of experiences, understanding, and communication. Similar to Pinocchio's awareness of himself as a living entity, the input and control elements of avatar systems allow us to experience worlds as different entities. It is our intentions and intuition rather than the patterns of performing actions that have to control these entities in future avatar-based digital environments.

REFERENCES

Bente, G., Rüggenberg, S., Krämer, N. C., & Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based Collaborations. *Human Communication Research, 34*(2), 287–318.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature, 391*, 756–756.

Calvillo-Gámez, E. H., Cairns, P., & Cox, A. L. (2015). Assessing the core elements of the gaming experience. In *Game User Experience Evaluation* (pp. 37–62). Cham, Switzerland: Springer.

Collodi, C. (2012). The adventures of Pinocchio: The tale of a puppet. (Trans: C. D. Chiesa). La Maddalena, Italy: Aonia. (Original work published 1881).

Cummings, A. H. (2007). The evolution of game controllers and control schemes and their effect on their games. In *Proceedings of The 17th Annual University of Southampton Multimedia Systems Conference* (Vol. 21).

Dorf, R. C., & Bishop, R. H. (1998). *Modern control systems*. Menlo Park, CA: Addison-Wesley.

Englander, I., & Englander, A. (2003). *The architecture of computer hardware and systems software: An information technology approach*. Hoboken, NJ: Wiley.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (2nd Ed.). Boston, MA: Addison-Wesley Longman Publishing.

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., & Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Frontiers in Human Neuroscience*, *8*.

Gerhard, M., Moore, D., & Hobbs, D. (2004). Embodiment and copresence in collaborative interfaces. *International Journal of Human-Computer Studies, 61*(4), 453–480.

Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R. J., & Morency, L.-P. (2007). Can virtual humans be more engaging than real ones? In *International Conference on Human-Computer Interaction* (pp. 286–297). Berlin: Springer.

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Williston, VT: Morgan & Claypool.

Koster, R. (2013). *Theory of fun for game design*. Sebastopol, CA: O'Reilly Media, Inc.

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition, 107*(3), 978–998.

Maselli, A., & Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience*, *7*, Article 83.

Norman, D. A. (2010). Natural user interfaces are not natural. *Interactions*, *17*(3), 6–10.

Rogers, S. (2014). *Level up! The guide to great video game design*. West Sussex: John Wiley & Sons.

Roth, D., Latoschik, M. E., Vogeley, K., & Bente, G. (2015). Hybrid avatar-agent technology—A conceptual step towards mediated "social" virtual reality and its respective challenges. *i-Com - Journal of Interactive Media, 14*(2), 107–114.

Roth, D., Waldow, K., Stetter, F., Bente, G., Latoschik, M. E., & Fuhrmann, A. (2016). SIAM-C—A socially immersive avatar mediated communication platform. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (pp. 357–358). New York: ACM.

Schmeck, H., Müller-Schloer, C., Çakar, E., Mnif, M., & Richter, U. (2010). Adaptivity and self-organization in organic computing systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, *5*(3), no. 10. New York: ACM.

Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, *3*(2), 214-220.

Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition*, *15*(2), 423–432.

von Mammen, S. (2016). Self-organisation in games, games on self-organisation. In *Games and Virtual Worlds for Serious Applications (VS-Games), 2016 8th International Conference on* (pp. 1–8). IEEE.

Waltemate, T., Senna, I., Hülsmann, F., Rohde, M., Kopp, S., Ernst, M., & Botsch, M. (2016). The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. New York: ACM.

Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011). Realtime performance-based facial animation. In *ACM SIGGRAPH 2011 Papers* (p. 77:1–77:10). New York: ACM.