Lasse Scherffig

From Action Capture to Interaction Gestalt

How can moving a small physical object on a table, observing *apparent motion* on a computer screen and pressing a finger onto one part of the object be experienced as one integrated action: clicking on something? This text proposes a novel answer to this question that goes back to the beginning of interactive computing and before, re-activating ideas from cybernetics to arrive at a new understanding of screen-based interaction.[1]

In this paper, I will briefly introduce cybernetics and its role in interactive computing and show how at the transition from analogue to digital computing screen-based interaction was introduced. I will then explain how screen-based interaction is subject to questions regarding the perception of motion that were first raised by gestalt psychology, explain how these questions relate to the idea of direct manipulation and how we might have to rethink the gestalt of an interface as an effect of interaction.

Cybernetics and Interaction

While cybernetics played an important role in the formation of early computer science, in recent times it has mainly been discussed in the humanities.[2] Here, the focus often is on the epistemological implications of this discipline, which, from its very beginning, proclaimed it would erase the boundaries between animal and machine, living and non-living systems.[3] One of its core tenets is the application of negative feedback to the description and control of any process that can be described as goal-directed behaviour.[4]

Negative feedback implies that the output of a system is fed back to its input as a negative quantity, resulting in a system that operates on the difference between its output and a desired goal. Systems using negative feedback hence use their own deviation from a given goal as a means of correcting this error.

Cybernetics to some extend can be understood as a science undertaking a "totalization" of feedback control.[5] Its importance for answering the question about *clicking*

1 For a much broader and more detailed development of this argument see Lasse Scherffig, *Feedbackmaschinen. Kybernetik und Interaktion*, Köln: Kunsthochschule für Medien Köln, 2017.

2 Kathryn Hayles, *How we Became Post-human. Virtual Bodies in Cybernetics, Literature, and Informatics*, Chicago, IL: Chicago University Press, 1999; Claus Pias, Zeit der Kybernetik. Eine Einstimmung, in: Claus Pias (ed.), *Cybernetics/Kybernetik – The Macy Conferences. Volume II*, Zürich/Berlin: Diaphanes, 2004, pp. 9–41.

3 This is already apparent in the title of Norbert Wiener's *Cybernetics. Or Control and Communication in the Animal and the Machine*, Cambridge, MA: MIT Press, 1948.

4 Arturo Rosenblueth, Norbert Wiener, Julian Bigelow, Behavior, Purpose and Teleology, in: *Philosophy of Science* 10 (1943), pp. 18–24.

5 Peter Galison, The Ontology of the Enemy. Norbert Wiener and the Cybernetic Vision, in: *Critical Inquiry* 21.1 (1994), pp. 228–266, p. 233.

1 The Product Integraph, an electro-mechanical analogue computer built at the Servomechanisms Laboratory.

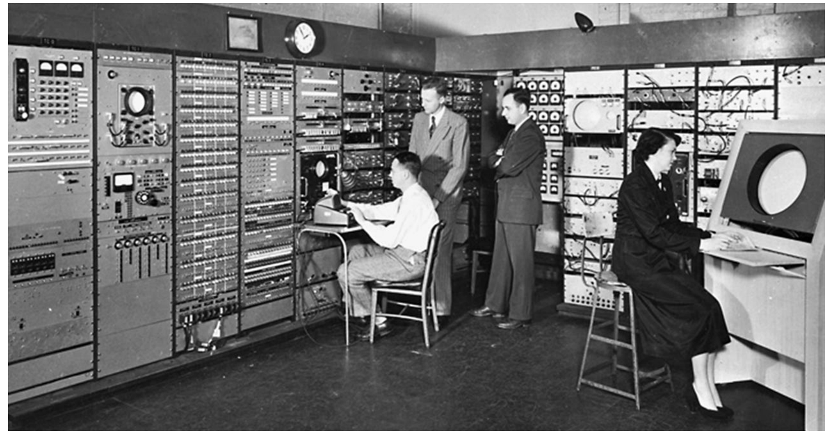
on something is two-fold: on one hand, the idea of negative feedback can help us to understand how in clicking on something, hands on the table and motion on the screen come together in one integrated sensorimotor act that, although involving a variety of distinct processes and locations, is perceived as one. On the other, historically, the MIT Servomechanisms Laboratory was behind much of the rise of negative feedback and thus the emergence of cybernetics as a field, as well as the construction of the first digital computer that was interactive in today's sense.

This computer started as an analogue computer for flight simulation, the Aircraft Stability and Control Analyzer (ASCA). The machine originally was planned as a continuation of the laboratory's successful work in analogue computing. Especially Vannevar Bush's famous Differential Analyzer had made clear that analogue computing can be applied to a variety of problems,[6] paving the way for the idea of building a flight simulator for arbitrary (existing and future) airplanes as "a cockpit or control cabin connected, somehow, to an analog computer"[7] (fig. 1).

Analogue computing, in this context, denotes a form of computation where a physical system is built in analogy to a phenomenon under study.[8] At the Servomechanisms Laboratory, during the early twentieth century, this practice led to the construction of a series of feedback-based electro-mechanical devices to study the dynamics of the electrical power grid and other complex systems (fig. 1). Likewise, the ASCA was conceived as an electro-mechanical system whose kinetic and electrical dynamics would resemble the dynamics of flying. Crucially, this meant that the cockpit would be an integral part of the computer – as the motion of its instruments and controls would be inseparable from the motion of computation.

However, halfway through its construction, the computing part of the machine was turned into a digital computer, because the project leads had realized the future potential of this emerging technology.[9] This change meant that a digital computer was to take over the role of an electro-mechanical device intrinsically connected to an environment (a cockpit, in this case). It hence had to be a special kind of digital computer: a computer that operates in real-time and allows for the exchange of data with its environment while operating.

6 David A. Mindell, *Between Human and Machine. Feedback, Control, and Computing before Cybernetics*, Baltimore: Johns Hopkins University Press, 2004, pp. 157–158.
7 Kent C. Redmond, Thomas M. Smith, *Project Whirlwind. The History of a Pioneer Computer*, Bedford, MA: Digital Press, 1980, p. 32.
8 Charles West Churchman, *Operations research. Eine Einführung in die Unternehmensforschung*, München: Oldenbourg Verlag, 1971, pp. 151–152.
9 Redmond 1980 (as fn. 7), pp. 27–44.



2 ASCA 1947 (left) and Whirlwind 1950 (right), a cockpit whose moving parts are part of a computational process versus the shape of computation to come.

During the construction of this novel machine, however, the task of building a versatile and fast digital computer became so dominant, that the engineers involved in the project increasingly neglected the cockpit portion of the system. As this cockpit still constituted a system of analogue instruments and moving parts, it later became clear that connecting these instruments to a digital computer posed fundamental problems that had never been dealt with before: "These problems were not impossible, but neither did established solutions exist. The digital computer was too new."[10]

In consequence, the simulator's cockpit was scrapped in late 1948 and the result was named Whirlwind the first interactive computer ever built and no longer a flight simulator (fig. 2).[11]

Reciprocal Visibility

What was *too new* to make a digital ASCA possible comes down to two questions: how to make digital data and processes visible to human viewers, and how to make the viewer's actions and reactions, in turn, *visible* to the computer?

Both problems are rooted in the nature of digital computation: the visibility of a representation in analogue computing is determined by the relationship between a physical system and the system it was made to model. Bush's Differential Analyzer, like the other analogue computers of the Servomechanisms Laboratory, was not so much a computer that solved differential equations as it was "an elegant, dynamical, mechanical model of the differential equation" that did "kinetically act out the mathematical equation".[12]

10 Ibid., p. 49.
11 Ibid., p. 60, pp. 43–44.

12 Larry Owens, Vannevar Bush and the Differential Analyzer. The Text and Context of an Early Computer, in: *Technology and Culture. The International*

Likewise, the ASCA would have been an electro-mechanical model of the aerodynamics of flight. This model would seamlessly integrate the instruments and controls of the cockpit, as well as any human action bearing on them, simply because the motion of instruments and controls would be part of *acting out* the computation. In contrast, digital computation has no *a priori* relationship to the systems it models.[13] It is marked by discrete states, represented by the symbols of a formalism, and "carefully chosen rules that dictate how one symbol succeeds another".[14] In order to be visible, digital computation must be translated into representations that "stand in an arbitrary relationship to the objects they represent."[15]

In addition, digital computation from the very beginning was conceptualized as a time and context free process. The idea of the Turing Machine (and equivalent definitions of computation) assumes that computation starts with a fixed input, operates on this input according to a fixed set of rules and terminates after a finite number of steps (or goes into an infinite loop of repetition).[16] Hence "[t]uring machines cannot handle the passage of time".[17]

Originally conceptualized as a machine in constant dialogue with a crew of flight operators in training, Whirlwind had to deviate from this assumption. The fact that indeed almost every computer we use today does so – by constantly waiting for new input from its environment while producing output that may affect future inputs – has only relatively recently been acknowledged by theoretical computer science.[18]

During the transition from an analogue ASCA to a digital Whirlwind, both problems were addressed pragmatically. The problem of the visibility of digital data was approached by establishing the mode of representation that is still dominant today: the computer drew arbitrary symbolic representations on the screen. To that end, Whirlwind's data registers were linked to the x/y-position of the electrode beam of a cathode-ray tube (CRT, fig. 3).[19] By so doing, the discrete states of machine computation were translated into representations that are readable by human observers, and the computer screen was introduced.

Within the project, the establishment of this new form of connecting people and computation was not seen as a great leap. Robert Everett, one of Whirlwind's engineers, simply noted later: "One of the things that I think we did first was to connect a visual display to a computer."[20] It was understood as something *I think we did first* because the engineering practice of the Second World War had already established the possibility of thinking (and building) this connection. With the Williams Tube a combination of digital computation and CRT was already in use. As the Williams Tube was a form of digital memory that *drew* zeros and ones onto a CRT screen in order to store them for a few milliseconds, it was not intended to be looked at by a human observer.[21] But in analogue radar technology, CRTs

Quarterly of the Society for the History of Technology (1986), pp. 63–95, p. 75.

13 Gerard O'Brien, Jon Opie, The Role of Representation in Computation, in: *Cognitive Processing* 10.1 (2008), pp. 53–62.

14 Ibid., p. 56.

15 Ibid., p. 58.

16 Georg Trogemann, Jochen Viehoff, *Code@Art. Eine elementare Einführung in die Programmierung als künstlerische Praktik*, Wien/New York: Springer, 2005, p. 85.

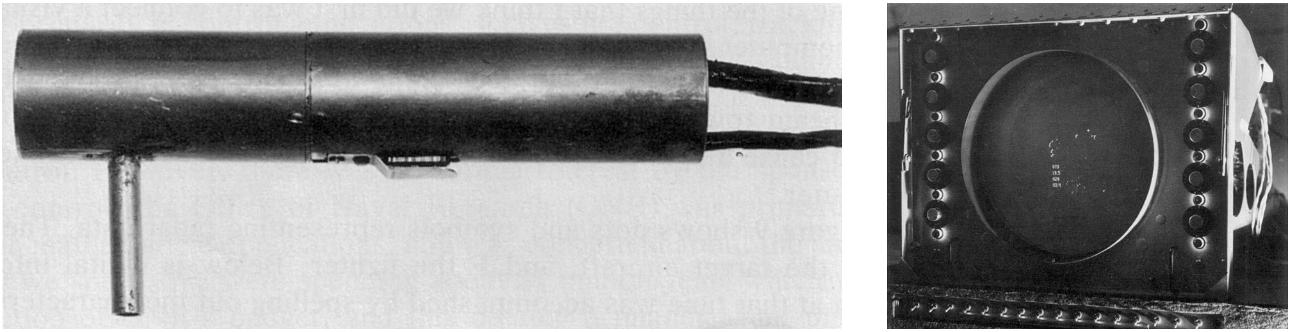
17 Peter Wegner, Why Interaction is More Powerful Than Algorithms, in: *Communications of the ACM* 40.5 (1997), pp. 80–91, p. 83.

18 Ibid.

19 Robert Everett, Whirlwind, in: J. Howlett, Gian Carlo Rota, Nicholas Metropolis (eds.), *A History of Computing in the Twentieth Century*, Orlando: Academic Press, 1980, pp. 365–384, p. 365.

20 Ibid, p. 375.

21 Claus Pias, *Computer Spiel Welten,* Dissertation, Weimar: Bauhaus-Universität, 2000, pp. 55–56.



3 Light-gun and symbolic representations on an early screen of Whirlwind.

had already been employed as visual displays.[22] Finished after the war, even radar CRTs leftover from the war could be used in the construction of Whirlwind.[23] The project thus simply had to connect both pre-existing practices (CRT-based digital memory and analogue radar displays) to create the arrangement of computation and representation we now refer to as the computer screen.

The problem of the *visibility* of human action to the process of computation was addressed by interrupting this process. A light-gun allowed for a literal *handling* of computation, as it made it possible to touch symbolic representations by pointing at them (fig. 3). This was achieved by placing a light sensor at the tip of the gun that would interrupt the computer's drawing process. As Whirlwind did not draw a rasterized image (organized in rows and columns of pixels), but drew one representation after the other, interrupting this process entailed that the light picked up at the moment of interruption would be emitted from the very object the gun was pointed at. It could thus be interpreted as a selection to be taken into account for further computation.[24]

With this setup, Whirlwind was ready to become the origin of SAGE, the Semi-Automatic Ground Environment air defence system – the largest computer built to date, managing American air defence until 1983.[25] More importantly, however, it established a feedback loop between screen-based representation and action. In consequence, the visibility of what was represented on screen became subject to interactions between motor activity and visual perception.

Direct Manipulation

The closing of this loop, in which action would be taken on a screen-based representation that in turn would react to that action, preconfigured how we interact with computers

22 Axel Roch, Die Maus. Von der elektrischen zur taktischen Feuerleitung, in: *Lab. Jahrbuch 1995/96 für Künste und Apparate*, Köln: Verlag der Buchhandlung Walther König, 1996, pp. 166–173, p. 170.
23 Everett 1980 (as fn. 19), p. 379.

24 C. R. Wieser, *Cape Cod System and Demonstration*, Cambridge, MA: MIT Lincoln Laboratory, 1953, p. 2.
25 Redmond 1980 (as fn. 7), p. 206.

until today. It established a remarkably stable dispositive of interaction, sustainably structuring large parts of the field of human-computer interaction (HCI), which would later refer to the combination of (mostly screen-based) representation with the capability to act on these representations as an *interface*. Nevertheless, it took the field until the 1980s to conceptualize the closed loop between representation and action as *direct manipulation*.

This discussion initially was framed by cognitive science and computational theories of the mind that treat interaction as a process of rule-based problem solving. For Ben Shneiderman, who introduced the term "direct manipulation", the phenomenon can accordingly be explained by assuming a difference between non-physical "semantics" of human problem solving and the physical "syntax" of representation and action at an interface.[26] While, according to this view, any form of HCI has to mediate between these two domains, direct manipulation reduces the difference between them by having users act in the world of semantics as opposed to syntax: direct manipulation, the argument goes, allows a writer to, for instance, directly interact with a paragraph of text (by marking it with the mouse) as opposed to decomposing high-level semantic intentions into low-level abstract commands whose syntax is largely unrelated to the paragraph itself and the act of manipulating it.[27]

Later, Edwin Hutchins, James Holland and Donald Norman expanded on Shneiderman's work, providing a seminal discussion of direct manipulation from a cognitive science perspective.[28] Starting from the assertion that "[w]e see promise in the notion of direct manipulation, but as of yet we see no explanation of it",[29] they develop an explanation that follows Shneiderman's path by distinguishing between the physical reality of an interface and the non-physical "model-world" of what it represents.[30] Direct manipulation, in this view, implies acting with the metaphors of that model-world, while well-chosen metaphors align this model with a user's problems. It is thus a function of the cognitive or information processing "distance" between the model-world and intention.[31]

Surprisingly, however, this does not seem to account for the whole phenomenon. Direct manipulation for the authors seems to possess a qualitative or experiential component that is hard to grasp in the terms of cognition and problem solving. In addition to cognitive *distance*, direct manipulation relies on emotional *engagement*, resulting from the feeling of being causally effective in that world – a phenomenon that cannot be understood in terms of goal-directed problem solving. The authors thus admit that direct manipulation seems like an "atavistic […] return to concrete thinking".[32] It may, however, be precisely the messy concrete thinking of our hands engaged in syntactic activities (or sensorimotor loops) that can help us to understand direct manipulation, as will become apparent later in this paper.

26 Ben Shneiderman, Direct Manipulation. A Step Beyond Programming Languages, in: Noah Wardrip-Fruin, Nick Montfort (eds.), *The New Media Reader*, New York, NY/London: W.W. Norton & Company, 2001, pp. 486–498.

27 Ibid. This argument alone is enough to cast doubt on the supposed directness of direct manipulation, as manipulating a paragraph with the mouse still presupposes a decomposition into low-level hand movements and button presses – only that this low-level syntax is *different* from, say, a command line interface.

28 Edwin L. Hutchins, James D. Hollan, Donald A. Norman, Direct Manipulation Interfaces, in: *Human-Computer Interaction* 1 (1985), pp. 311–338.

29 Ibid., p. 316.

30 Ibid., p. 317.

31 Ibid., p. 311.

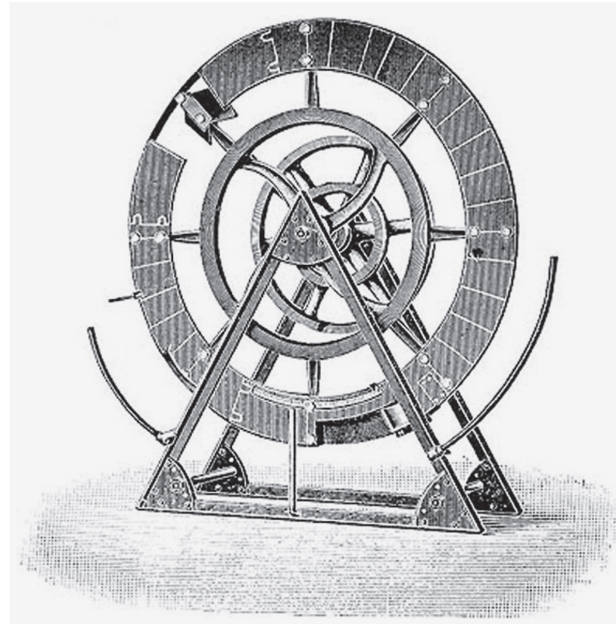
32 Ibid., p. 337.

Gestalt and Apparent Motion

The motion we see on computer screens is what the psychology of perception calls *stroboscopic* or *apparent* motion, an illusionary impression of motion created by the succession of static frames. Historically, the systematic investigation of apparent motion is closely connected to gestalt psychology,[33] as one of the texts defining the field is Max Wertheimer's (still untranslated) *Experimental Studies about the Perception of Motion*.[34]

Wertheimer's seminal study tries to understand how stroboscopic stimuli that are objectively not moving create the subjective *percept* of motion. For the study, Wertheimer employs the Schumann Tachistoscope as a stroboscope (fig. 4). This device uses rotation to quickly cover and uncover stimuli. A setup using two stimuli, a and b, and a prism allows Wertheimer to use the apparatus in a way that, to a viewer, presents both stimuli in quick alternating succession.

Focusing on those cases of apparent motion that do not yield a perfect illusion of seeing moving objects but, for instance, fractured and partial motion percepts[35], Wertheimer arrives at a remarkable conclusion that ultimately reverses the relation of movement and object as it was understood by his contemporaries (fig. 5). These, he argues, assume that the perception of motion presupposes the perception of a moving object, understanding the moving object as a primary and its motion as a secondary feature



4 The Schumann Tachistoscope.

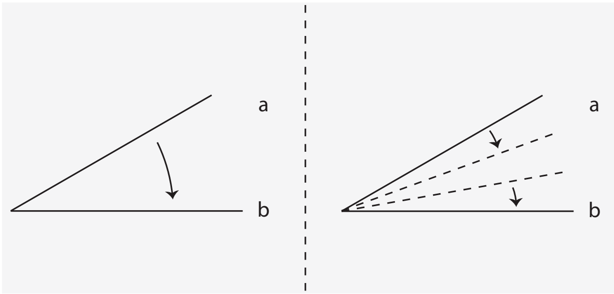
ascribed by perception. Wertheimer, instead, sees motion, named "pure φ" or "pure motion", as a primary object of perception, even reconstructing the identity of moving objects as a limiting case of motion.[36] In this view, perception of motion happens directly and immediately, preceding and enabling the perception of the gestalt of an object. The latter is hence conceived as a "short-circuit" of motion perception as a "duo-in-uno" when, for instance, two lines, a and b, in

33 Robert M. Steinman, Zygmunt Pizlo, Filip J. Pizlo, Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution, in: *Vision Research* 40 (2000), pp. 2257–2264.

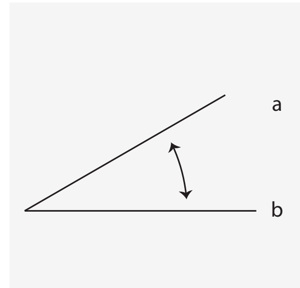
34 Max Wertheimer, Experimentelle Studien über das Sehen von Bewegung, in: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 61 (1912), pp. 161–265.

35 Ibid., p. 191.

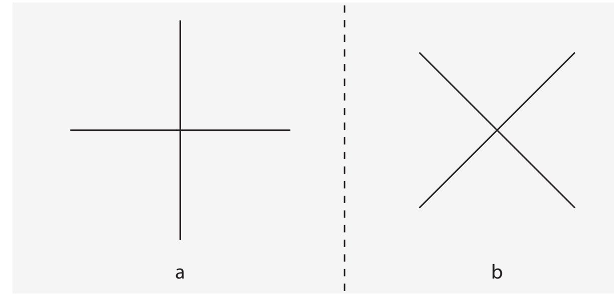
36 Ibid., p. 221.



5 Apparent motion of a line from a to b (left) and partial apparent motion if the time-interval between frames a and b becomes too long (right), as described by Wertheimer.



6 Two lines in rapid succession forming an angle composed of two sides, as described by Wertheimer.



7 Ambiguous motion as described by Linke.

rapid succession lead to the perception of one object: an angle composed of two sides (fig. 6).[37]

In special cases, the stroboscopic stimuli causing the perception of apparent motion may be ambiguous. Whenever, for example, two or more concurring interpretations of one stimulus are possible, their perception becomes multi-stable: Subjects perceive one possible percept at a time, while perception alternates between possibilities over time. This was first demonstrated by Paul Linke using a cross that is rotated by 45° from stimulus to stimulus and that, as a bistable stimulus, can be perceived as clockwise or counterclockwise rotation (fig. 7).[38] Termed “ambiguous motion”, this effect was later studied by Paul von Schiller, who tried to isolate the factors that determine which possible percept is perceived at a time, trying to establish the laws of how ambiguous motion is disambiguated to distinct percepts.[39] During this study, von Schiller made a remarkable observation: His subjects were able to control the perceived direction of motion most effectively by actively moving their hands and heads. This, he writes in a footnote, constitutes a case of motor activity having a gestalt influence on visual perception.[40] Because the experimental systems of experimental psychology of that time, such as the tachistoscope, only allowed for the precise control of the presentation of stimuli without connecting it to human action, this effect seemed too hard to control for him to warrant further investigation.[41]

37 Ibid., p. 251.
38 Paul von Schiller, Stroboskopische Alternativversuche, in: *Psychologische Forschung* 17 (1933), pp. 179–214, p. 180.
39 Ibid.

Action Capture

During the past decades, the methods of experimental psychology have changed significantly in favour of quantitative research that relies on a universal experimental system, enabling not only the precise control of the exposure of stimuli but also the measurement of human action. This system is fundamentally structured by the interactive

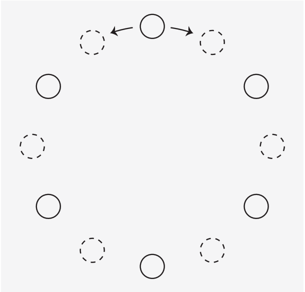
40 Ibid., p. 196.
41 Ibid., p. 195.

computer and was, for instance, pioneered in the famous experimental studies by Douglas Engelbart[42], which led to the decision to replace the light-gun for computer input with the mouse.[43] Today, "the computer [...] has taken over practically all the experimental procedures used to examine the perception of space and time".[44]

It nevertheless took until 1994 for the first publication to present quantitative evidence of the influence described by von Schiller.[45] Since then, a series of studies have shown that if ambiguous motion is coupled to physical motion, then the bodily movement *captures* its perception by influencing it in the direction of motion. This capture effect is strongest for voluntary self-motion and has therefore been named "action capture"[46] (or "priming by actions"[47]), to do justice to the fact that the influence is caused by whole actions, comprised of intentions, motor planning and execution. In most studies analysing the effect, computers are used to couple the movement of the hands with ambiguous motion stimuli presented on a screen (fig. 8). Action capture, one could hence argue, has been mostly studied as inter-action capture.

The concept of action capture not only holds for visual stimuli and the motion of our hands; it also has been shown that the perception of ambiguous auditory[48] and tactile[49] stimuli can be captured by the movement of the hands, eyes[50] and the walking body[51]. For the interface this implies that we may have to understand interfaces as perceived in action. Their visual (and even tactile and auditory) qualities are influenced by our actions, and interface design may have to take into account that it is not only the functioning of an interface that depends on its use, but also its perceptual qualities.

Research examining action capture has shown that it is facilitated by a close physical and temporal distance of the action and stimulus[52], as well as a correspondence of the axes and orientation of motion between both.[53] More importantly, the correspondence of stimulus and action that drives the effect is not an *a priori*. It is context dependent, as it can be influenced by expectations: when, for instance, a button with a right arrow is pressed, perception of apparent rotation is captured in the clockwise direction, because we have learned that pushing a round object to the right will most likely cause it to rotate in the clockwise direction.[54] The effect, in addition, can be modified and even reversed by training.[55] And finally, it is already present when actions are merely planned and not yet carried out.[56]



8 A typical ambiguous motion stimulus as used in experiments. Stroboscopic motion of the circles is presented on a computer screen at the same time as subjects perform physical motion on an input device, such as a keypad, knob, or mouse.

42 As a dispositive it structures the presentation of stimuli, the measurement of responses, the design and statistical analysis of experiments, and by that "the nature of the questions that can be addressed". Nicholas J. Wade, Dieter Heller, Scopes of Perception. The Experimental Manipulation of Space and Time, in: *Psychological Research* 60.4 (1997), pp. 227–237, p. 235.

43 William K. English, Douglas C. Engelbart, Melvyn L. Berman, Display-Selection Techniques for Text Manipulation, in: *IEEE Transactions on Human Factors in Electronics* 8.1 (1967), pp. 5–15.

44 Wade, Heller (as fn. 42), p. 235.

45 G. Ishimura, S. Shimojo, Voluntary Action Captures Visual Motion, in: *Investigative Ophthalmology and Visual Science (Supplement)* 35 (1994), p. 1275.

46 Ibid.

47 Andreas Wohlschläger, Visual Motion Priming by Invisible Actions, in: *Vision Research* 40 (2000), pp. 925–930, p. 929.

48 Bruno H. Repp, Günther Knoblich, Action Can Affect Auditory Perception, in: *Psychological Science* 18.1 (2007), pp. 6–7.

49 Olivia Carter, Talia Konkle, Qi Wang, Vincent Hayward, Christopher Moore, Tactile Rivalry Demonstrated with an Ambiguous Apparent-Motion Quartet, in: *Current Biology* 18 (2008), pp. 1050–1054.

50 Ibid.

51 Yoshiko Yabe, Gentaro Taga, Treadmill Locomotion Captures Visual Perception of Apparent Motion, in: *Experimental Brain Research* 191.4 (2008), pp. 487–494.

52 G. Ishimura, Visuomotor for Action Capture, in: *Investigative Ophthalmology and Visual Science (Supplement)* 36 (1995), p. 357.

53 Wohlschläger 2000 (as fn. 47), pp. 927–929.

54 Ibid., p. 928.

55 Ishimura 1995 (as fn. 52), p. 357.

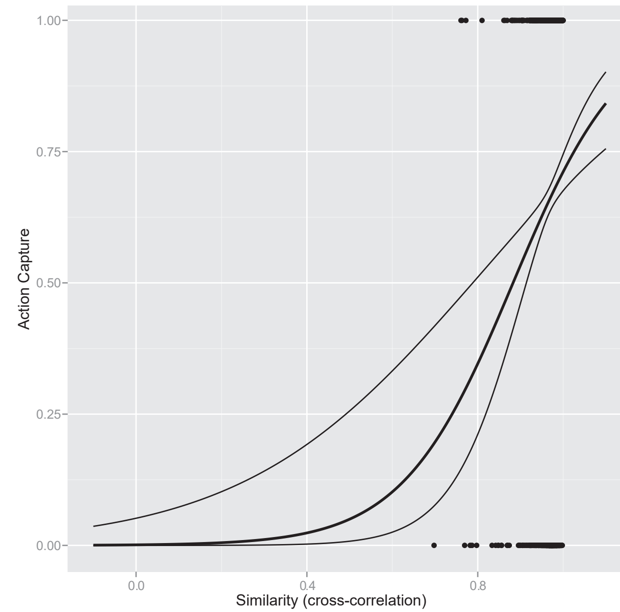
56 Wohlschläger 2000 (as fn. 47), p. 929.

Accordingly, action capture seems to not only depend on how much an action corresponds with what we perceive (in terms of spatio-temporal distance and orientation), but rather seems to depend on how much a possible percept corresponds to the result we expect an action to have. If I expect (or plan) my hands to be involved in causing clockwise rotation, I am more likely to perceive an ambiguous rotation as clockwise.

This corresponds to early findings from human factors indicating that the speed and error rate of actions at an interface depend on the "compatibility" of stimulus and response.[57] This compatibility has from the beginning been understood as an acquired relationship, for which it holds that "stimulus and response sets are optimally matched when the resulting ensemble agrees closely with the basic habits or expectancies of individuals".[58]

In order to further analyse the significance of action capture for screen-based interaction, I have conducted a study linking earlier research on action capture to the compatibility of mouse action and computer response.[59] Assessing compatibility, however, is a messy task, since basic habits or expectancies do not translate well into experimental protocols. But the computational tools of cybernetics – the field whose heritage still defines the way interactivity works – at least provide ways of measuring a non-semantic similarity of stimuli and response, understood as the cross-correlation of a time series of measurements.

Coupling an ambiguous motion stimulus to mouse movements, the experiment measured how subjects moved



9 Logistic regression of the relationship between a) the similarity of mouse motion and a perfect rotation on screen, in the direction subjects were asked to move the mouse, and b) the coincidence of perceived direction of ambiguous motion with the direction of mouse motion (action capture). The more similar the motion of the hand is to a perfect rotation in the correct direction, the more likely is action capture. For details see Scherffig 2017 (as fn. 1), pp. 259–261.

the mouse as they were asked to perform circular motion while looking at an ambiguous rotation on screen. It thus relied on a paradigmatic case of interaction, linking the motion of the mouse with apparent motion on screen, while making their interrelation measurable by using ambiguous motion that can be captured by the body's activity. In a series of trials, mouse motion was recorded together with the perceived direction of rotation of each trial, ascertained through questions (fig. 9).

57 Paul M. Fitts, Charles M. Seeger, S-R Compatibility. Spatial Characteristics of Stimulus and Response Codes, in: *Journal of Experimental Psychology* 46.3 (1953), pp. 199–210, p. 199.
58 Ibid., p. 208.
59 Scherffig 2017 (as fn. 1), pp. 257–262.

The result is as simple as it is statistically significant: the more similar the mouse motion is to the motion subjects were asked to perform, the higher the likelihood of action capture (the significance level of their correlation is $p<0.0001$, see also fig. 5). In other words, the more similar the action of the hand on the mouse and the reaction on the screen seem to be, the more the former captures the latter. Action capture, therefore, seems to quantitatively incorporate our actions into what we perceive. Our perception seems to *calculate* with our actions and their expected results.

Interaction Gestalt

We have seen the idea of calculating with one's actions before. It is the idea of comparing actions with changes in perception, or – in the technical terms of cybernetics – the idea of using the difference between a system's output and some goal to determine its future output.

Past and contemporary discourse in physiology, the psychology of perception and cognitive science has often identified this goal with expectations or predictions[60]: We constantly compare the change in perception induced by our activities with expected change. This is what becomes apparent (and even measurable) as action capture: our predictions regarding a physical action capture the way we perceive the results of that action.

Action capture thus suggests that at the interface, too, it is our actions that determine what we perceive. If we move a mouse and observe its cursor and the on-screen reactions to pressing a finger onto it, the motion of our hands on a physical object on a table and the perception of *apparent motion* on a computer screen is fused *internally* into a sensorimotor unity that goes beyond the mere correlation of both.

This is exactly what was observed by computer scientist Dag Svanæs.[61] Conducting experiments in which subjects interacted with abstract interactive systems consisting of black and white squares[62] he analysed their behaviour in correlation with their verbal descriptions of it, paying attention to the way the abstract black and white squares slowly became perceived as objects:

> *The objects described by the subjects in the experiments existed for them only through interaction. The objects emerged as a result of the interplay between the intentions of the users, the users' actions, and the feedback given by the system.*[63]

Observing the interaction with a simple system that would eventually be perceived as a switch, he notes:

> *When the subjects said 'It is a switch', they did not come to this conclusion from a formal analysis of the State Transition Diagram of the example. Nor did they conclude it from the visual appearance of the square, as the squares all looked the same. The switch behavior*

60 See, for instance, Karl J. Friston, Christopher Thornton, Andy Clark, Free-Energy Minimization and the Dark-Room Problem, in: *Frontiers in Psychology* 3 (2012), pp. 1–7; Jack M. Loomis, Distal Attribution and Presence, in: *Presence* 1.1 (1992), pp. 113–119; or Sarah-Jayne Blakemore, Chris D. Frith, Daniel M.Wolpert, Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli, in: *Journal of Cognitive Neuroscience* 11.5 (1999), pp. 551–559.

61 Dag Svanæs, *Understanding Interactivity. Steps to a Phenomenology of Human-Computer Interaction*, Dissertation, Trondheim: Norges Teknisk-Naturvitenskapelige Universitet, 2000.

62 Ibid., pp. 128–132, pp. 108–110.

63 Ibid., p. 230.

> *slowly emerged from the interaction as the square repeated its response to the subject's actions.*[64]

This implies that the hands on the mouse in dialogue with the computer's response yield the emergence of perceptual units having "gestalt properties" as their perception, once emerged, is "direct and immediate" and not a cognitive interpretation of action and perception.[65]

Svanæs therefore suggests treating the objects that compose an interface as "interaction gestalts", entities that are "similar to visual gestalts in that they are wholes, and not compositions of analytical elements".[66] The form or gestalt of an interface, understood as *interaction gestalt,* can be seen as a perceptual or experiential whole that is based on action and perception as *duo-in-uno* – as a limiting case of the loop of human action and machine reaction.

From a sensorimotor perspective, the elements that make up an interface are hence not so much the discrete entities that they have been designed and programmed to be; instead they are the results of being used. Buttons, in this sense, look like buttons because they are used as such – *and* the other way round. Their form does not imply or communicate their function. Instead, their (subjectively experienced) form and function are interdependent and are the result of their use and its context.

This suggests a *cybernetic* model of the interface and interaction, implying that what we see is enacted by how we react to it. According to this model, we can indeed understand direct manipulation in terms of *distance* end *engagement*. But distance would be reduced to simple spatio-temporal distance of stimulus and response and the perceptual similarity between both. Or more generally, it would be redefined as the distance of predicted and actual reaction, which is the negative feedback at the heart of cybernetics. Engagement, in turn, would become being engaged in sensorimotor loops that are continuously learned and exercised, forming the objects they deal with within this cyclical process. If today's touch-based interaction on mobile phone screens seems to constitute a return to Whirlwind's combination of screen-based representation and the possibility of touching it, this may be understood in light of HCI's constant effort to minimize distance and maximize engagement in these literal terms.

The simple need to establish *reciprocal visibility* between computation and human environment thus introduced a dispositive of interaction in which bodily movement at the computer screen, its predicted and its observed results together are integrated into coherent perceptions of interaction, that form the gestalt of the interface. This is the integration of hand and screen-based representation that allows us to speak of *clicking on something* while we steer a physical mouse on a table and watch apparent motion on a screen. What, according to this view, creates interfaces, is interaction.[67]

64 State Transition Diagrams are formal graphical representations of how a system of discrete states (such as combinations of black and white squares that can switch their color) can transition from one state to another. In Svanæs' experiments these diagrams describe the actual behavior of the systems used, as opposed to the perceived behavior described by his subjects. Ibid., p. 206.

65 Ibid., p. 244.

66 Ibid.

67 As in: not only enables and shapes their functioning as interface but also their appearance.

Figures

1 MIT Museum. Reprinted with permission from MIT Museum.

2 Copyright Computer History Museum. All Rights Reserved (left), The MITRE Corporation. All Rights Reserved (right). Reprinted with permission from The MITRE Corporation.

3 The MITRE Corporation. All Rights Reserved. Reprinted with permission from The MITRE Corporation.

4 Collection Jirí Hoskovec. Reprinted with permission by Simona Hoskovcová.

5–9 Author's figure.