



BRILL

Television and the Origins of Visual Pattern Recognition AI in Japan: a ‘Quest for a Seeing Machine’

Hansun Hsiung | ORCID: 0000-0003-3001-3497
MLAC, Durham University, Durham, United Kingdom
hansun.hsiung@durham.ac.uk

Received 20 December 2023 | Accepted 8 September 2024 |

Published online 11 November 2024

Abstract

This article traces the genealogy of contemporary data-driven computer vision to developments at the Japan Broadcasting Corporation (NHK) in the 1960s. Specifically, it examines NHK's Visual and Auditory Information Science Unit and its role in the invention of the world's first deep convolutional neural network. The use of television to collect viewer behaviour data enabled modelling of eye-brain information processing, in particular mechanisms of feature extraction. This, in turn, linked Fukushima Kunihiko's formative work on signal compression to the development of a pattern recognition machine, resulting in the creation of the world's first convolutional neural network. Recovering this history is important for two reasons. First, it helps counter a trend of 'digital universalism' that covertly homogenizes local differences into a single culture of artificial intelligence in the Cold-War-era US. Second, it reveals the largely ignored role of television in the genesis of digital image technologies and AI more broadly.

Keywords

compression – computer vision – information technology – neural networks – television

In the spring of 2021, the Franklin Institute awarded its Bower Award for Achievement in Science to Fukushima Kunihiro (b. 1936). Established in 1824, the institute previously bestowed Bower awards on Stephen Hawking, Albert Einstein, Max Planck, and Thomas Edison. According to the prize committee, Fukushima's name deserved to be added to this pantheon for the following reason: credit is due to Fukushima for the 'invention of the first deep convolutional neural network' in 1979 – an achievement made possible by the application of 'principles of neuroscience to engineering' (Franklin Institute 2020).

Deep convolutional neural networks (deep CNNs or simply CNNs) should at least in name be familiar to those who have followed developments in second-generation AI since the 2010s. In particular, the application of deep CNNs to computer vision problems has radically expanded the capacity of image recognition and classification algorithms, increasing their accuracy from around 50 percent to over 95 percent (Krizhevsky et al. 2012). Since then, deep CNNs have become ubiquitous, embedded into the sociotechnical fabric of our everyday lives: AI visual pattern recognition is an essential part of our smartphones, cameras, and cars as well as our medical diagnostics and policing apparatuses. Indeed, before the emergence of the generative AI explosion that began with ChatGPT, computer vision was at the centre of discourse on the 'deep learning' boom, including trenchant critiques of algorithmic bias (Crawford & Paglen 2019). Japanese technologies played a central but overlooked role within this landscape, with NEC acting as a leading global supplier of facial recognition systems – including those which led to the wrongful arrest of an African-American man in Michigan in 2020 (Humphrey 2022: 72).

Why did the underlying models for contemporary visual pattern recognition AI first emerge in Japan? And what can this tell us more broadly about Japan's position within the global history of science and technology? The Franklin Institute's dual invocation of 'neuroscience and engineering' at the conceptual heart of deep CNNs initially suggests that we might easily fold Fukushima's 1979 contribution into the well-known narrative of American cold war cybernetics. As work in the history of science and media studies has shown, the cybernetic drive towards the machinic reassemblage of human cognitive faculties – 'learning machines', as Norbert Wiener (1954: 66) called them – ushered in a new form of biopower fundamental to contemporary forms of governance in which 'Big Data' is gathered and processed by networked systems of self-regulating intelligent technologies (Furuhata 2019; Halpern 2015; Halpern & Mitchell 2023; Kline 2015). Behind this stood the conceptual logic and defensive funding infrastructure of the Cold War (Edwards 1997; Erickson et al. 2013: 19–21; Galison 1994). The broader link between warfare and media has long

been stressed – one thinks of Virilio's work (1989), or else Kittler's (2010: 74, 193, 209, 221–222) trenchant characterization of media technologies as 'byproducts or waste products of pure military research'. More recently, detailed historical studies show that, before the 1990s, 'data-focused machine learning' research in the US existed on the peripheries of mainstream academic work on AI, enjoying little prestige if not 'widely lambasted' outright and routinely relying on military and intelligence funding (Jones 2023: 1360; 2018: 674). Image recognition and classification algorithms were a conspicuous example of this trend, emerging as the direct child of the US military imperative to process large quantities of photographic surveillance data (Dobson 2023).

The story of Fukushima Kunihiro in Japan, however, produces wrinkles in this US-centric narrative. Fukushima was certainly attuned to the context of cold war cybernetics research, most notably Frank Rosenblatt's Perceptron project sponsored by the Department of Defense. At the same time, precisely because of its Cold War alliance with the US, Japan allocated proportionately less direct spending on defence in its budget – less than 1 percent of its annual gross domestic product, compared with roughly 6 percent of the US budget in the 1970s – allowing it to funnel state income to other sectors of industrial growth (Maslow 2015: 761). Indeed, it was precisely during the first 'AI winter' of the 1970s – a period of decreased investment in AI-related research and development in the US – that Fukushima achieved his most significant innovations. This suggests that in order to understand the global unfolding of AI research – the sites and institutions that fostered its development, and the needs and desires to which it was attuned – we must look to particular structures of Japanese technoscience under high-speed economic growth that do not settle as easily into the model of military-to-civilian technology transfer. Through the story of Fukushima Kunihiro, we discover another site from which contemporary computer vision emerged: broadcast television and its attendant televisual technologies.

This article shows that broadcast television was a crucial site for the production of technoscientific knowledge on computer vision in post-war Japan. My argument can be taken in at least three senses. First, Japan's major television broadcasters functioned as an physical home for projects of machine learning. From 1959–1989, the pivotal years during which he developed the world's first deep CNN, Fukushima was a researcher at NHK, Japan's public broadcaster, established in 1926 on the model of the BBC. For most of that time (1965–1984), Fukushima worked in the Visual and Auditory Information Science Unit (VAISU; Shichō kagaku kenkyūshitsu 視聴科学研究室) of NHK's Broadcasting Science Research Laboratory (BSRL; Hōsō kagaku kisokenkyūjo

放送科学基礎研究所).¹ VAISU envisioned televisual technologies as a vital stepping stone to broader ‘bionics’ research, an ambition captured by the words of Fukushima’s colleague, Hiwatashi Kenji (1922–2016): ‘The most important thing is the organic union of physiology, psychology, and engineering, and one can say that our research unit is the first in the country to attempt to realize this goal’ (1965: 407).

Second, within these laboratory spaces, the television studio itself became a model for experimental setups. In the US, as James Dobson (2023) has recently noted, computer vision researchers took cues from the ‘soundproof control rooms and tape storage’ design of the television studio environment. This symmetry was all the more pronounced in Japan, where such research not only took place under the general aegis of broadcasters but was partly targeted to serve the imminent needs of NHK programming content. Moreover, the relationship between the studio and the laboratory rested on the centrality of televisual devices themselves as experimental instruments. Computer vision research required the mass capture and conversion of real-world images into electronic signals for processing. It also required new ways to visualize the physio-psychological process of human acts of seeing. Television cameras provided this link, serving as the means by which to scan and encode images as digital data for manipulation on computers, and to produce real-time visual data on human behaviour. In this manner, television technology functioned as the enabling apparatus for the machine simulation of human vision.

This second literal sense leads us into the third way in which broadcast television functioned as a site for producing computer vision. Central to the cybernetic project in general was the premise that the human nervous system, and thus its faculties of perception and cognition, could be understood in computational terms. Computer vision, in particular, could arguably be branded a ‘technomorphic physiology’ (Azar et al. 2021: 1101). In the early post-war decades, however, this premise was at best an abstract hope: fledgling neurosciences had yet to elucidate the workings of systems as essential as the human visual cortex, much less to offer any practical evidence that bionic theories could be computationally implemented. Well before he became involved in computer vision research, Fukushima (1966: 5) had already highlighted this state of affairs as a doctoral student, lamenting that insofar as ‘the physiological and psychological properties of the human visual system have still not been unravelled’, there was ‘far too little data’ for engineers to create functional

1 The name of VAISU in Japanese, Shichō kagaku kenkyūshitsu, literally translates into English as Visual and Auditory Science Unit. However, English-language publications consistently insert ‘Information’ into the translation, so I follow that practice here.

machine models. Broadcast television emerged to fill this lacuna. Specifically, it enabled researchers at NHK to harness the behaviours of viewers as data that might better clarify the physiological and psychological mechanisms underlying vision.

By situating NHK at the origin of deep CNNs and thus the origin of visual pattern recognition AI, my broader aim is twofold. Most fundamentally, attention to the conditions that shaped Fukushima's work poses a pluralizing resistance to a discourse of 'digital universalism' that sees the trajectory of contemporary information technologies as a vector for 'independence from local constraints' (Loukissas 2019: 9). As Xiao Liu (2019: 11) argues, digital universalism results in and 'overgeneralized media history and media theory', which tacitly attributes the origin of contemporary information technologies to certain North American and European centres and then transforms concepts from this limited geographic context into a global condition. One result is the inadvertent reification of a singular worldwide capitalism that aligns with Silicon Valley's ideological self-understanding. Recently, resistance to this tendency has taken the form of calls for pluralization and site-specificity, pointing to the varied geo-cultural factors that shape different 'markets, corporate forms, and institutional organizations' (Steinberg et al. 2024). For instance, Liu's work, focused on the post-socialist transition in China, shows how domestic resources, such as *qigong* and traditional Chinese medicine, were used to grapple with new information technologies and notions of the 'network' tied to Western liberal capitalism. Taking a cue from these studies, my article lays out a context for the emergence of deep CNNs at NHK that, though overlapping with and connected to the 'algorithmic culture' (Striphas 2015) of American cybernetics, ultimately developed in response to local configurations of ideas and institutions. Specifically, I suggest that visual pattern recognition AI developed contingently at a moment when the Japanese post-war developmental state, amidst a feeling of impending triumph over Western liberal capitalism, was re-evaluating the future meaning of television technology.

At the same time as we remain sensitive to local conditions, however, we should also seek to identify how the local reveals hidden insights for our broader understanding of the history of AI. Herein lies the second goal of this article: to argue for the central role of television in the genesis of contemporary AI overall. As Thomas Lamarre (2018: 5) reminds us, television 'has continuously mutated through a series of assemblings', taking on 'computational modalities and telecommunications technologies by making way for them in its very operations'. Lamarre's proposed genealogy of the 'screen-brain apparatus' centres on the period since the 1990s; his timeline roughly parallels that of Sheila Murphy's (2011) analysis of television's relation to the emergence

of ‘new media’, such as game consoles, personal computers and mobile phones. Yet, as my article demonstrates, the incorporation of television into experimental assemblages for both the understanding and simulation of human visual perception at the neural level began in Japan by the mid-1960s at the latest, forming one core part of NHK’s long-term research agenda. There is, therefore, the need for a much more expansive genealogy in which television, far from being only later on incorporated into computational and telecommunications assemblages, arguably functioned as a motor for some of the earliest attempts to practically implement models of neural computing that underlie core parts of contemporary AI. The exploration that follows sketches the basis for such a revised genealogy (Okazawa 2023).

The remainder of this article offers a microstudy of VAISU as a crucible for the emergence of contemporary data-driven computer vision. This microstudy is primarily based on NHK publications that reflect diverse scales of time and intended audience. At the smallest scale is the *NHK Giken geppō*, a monthly newsletter by and for NHK technical researchers containing summaries of work in progress, transcripts of internal lectures and roundtables (*zadankai*), book reviews, and short translations of relevant foreign literature. At the opposite end of the scale are a stream of retrospectives issued by NHK – its annual *Kenkyū nenpō*, its decennial *Kenkyūshi*, and its monumental thirty- and fifty-year histories. In between are various articles and occasional monographs by VAISU researchers. Additionally, my interpretations have as background a series of interviews with former VAISU employees – including Fukushima Kunihiko – and the personal papers of Hiwatashi Kenji.²

My story begins in Section 1 in the early 1960s, slightly before VAISU was formed, with the deployment of television as an experimental apparatus in NHK’s studies of viewer attention. Then, in Section 2, I show how VAISU research built on viewer attention studies to articulate the possibility of a ‘seeing machine’. Finally, in Section 3, I turn to Fukushima’s work, elucidating the links between viewer attention and his formative research on lossy compression. These links concerned, in particular, the computational physiology of

2 In the spring and summer of 2024, Okazawa Yasuhiro of Kyoto University, Maejima Masahiro of the National Museum of Nature and Science (Tokyo), and I conducted interviews with Fukushima Kunihiko and Itō Takayuki. We furthermore began cataloguing the personal papers of Hiwatashi Kenji in the possession of his son, Hiwatashi Yutaka. A special issue of *Jinbun gaku* featuring these materials is scheduled for publication in June 2025, but as we are currently still awaiting the review of the parties involved regarding sensitive information requiring redaction, direct quotations from the interviews could not be included in this article.

‘feature extraction’ – a concept at the heart of his invention of the world’s first deep CNN, dubbed the Neocognitron.

1 Audiences and Visual Fixation

We begin in June 1961, approximately four years before VAISU was founded. The start of the 1960s marked a new era of hope for NHK. Only one decade earlier, when announcements surfaced that NHK and Nippon Television were planning to launch public broadcasting services, the outlook had been pessimistic. SCAP’s Civil Communications Section, in a report to the US Department of State in the summer of 1951, predicted that ‘television for Japan could not be expected on any other than an experimental basis for at least 10 years’ (Boehringer 1951). Part of the problem was only temporary: the Occupation had held a monopoly over standard frequency channels used for television broadcasts, and denied Japanese entities their usage (Holthusen 1951). However, even after the Occupation ended in 1952, two enduring and intertwined problems remained. Despite pioneering efforts by Takayanagi Kenjirō (1899–1990) in the late 1920s and early 1930s to create ‘native’ television devices, the underlying technology – in particular, cathode-ray tubes – remained imported (Iida 2016: 120–122). This made television sets unaffordable as a consumer technology. Estimates on the eve of NHK’s launch of regular television broadcasting services in February 1953 forecast a maximum audience of ca. 3,000, television sets being ‘much beyond the purchasing power of the average Japanese’; broadcasts were accordingly limited to merely four hours of content per day (‘Television in Japan’, 1953).

By 1961, however, TV in Japan had defied all expectations. Economic recovery dovetailed with heavy investment in transistors, which solved the dual problem of cost and domestic manufacture. Sony’s release of the first fully transistorized television in 1959, combined with rising affluence among the population, spurred a rise in television ownership in Japan to the tune of an additional 1.5 million sets within a year – a number that exceeded the total sum of sets purchased in the previous six years (Ikuta 1960: 46). One year later, in 1961, this number increased by an additional 3 million (Bush 1961: 21). In addition, in the spring of 1959, Tokyo won its bid to host the 1964 Olympics, anticipating that it would be the first Olympic games to be broadcast worldwide via satellite. Together, these forces set the stage for NHK to pursue an agenda of aggressive development that would expand the range of programming content, now broadcast from 6 AM until 12 AM, and impress upon a global audience the prowess of Japanese TV.

Essential to this development agenda was a more thorough understanding of the national audience. Social-scientific fieldwork served as one means to obtain it. In 1960, NHK sent researchers into homes across Japan to conduct the first public iteration of its 'national time-use survey' (*kokumin seikatsu jikan chōsa*), focused in particular, but not exclusively, on when and how citizens engaged with radio and television in their everyday lives – at breakfast, lunch, and dinner; while doing housework and relaxing in the evening, or while commuting through public spaces (Mitsuya 2014; Yoshimi 2014). Yet another means for understanding audiences was in NHK's Technology Research Laboratory (Giken; Gijutsu kenkyūjo 技術研究所).³ Established in 1930, Giken had as its initial *raison d'être* the domestic production of radio and television devices, working in conjunction with electronics manufacturers around the country as well as Nippon TV. With this goal achieved, however, Giken researchers were mobilized in the 1960s to contribute to audience research.

It was there, in June of 1961, that the engineer Watanabe Akira began work on a 'television eye camera'. In principle, the television eye camera was an apparatus for tracking eye movement in order to determine points of attention, or 'visual fixation points' (*chūshiten*). Tracing eye movements had long been a theme in physiology and psychology since the previous century, and multiple experimental apparatuses already existed. But these apparatuses relied on a photographic principle. Watanabe's device, while retaining elements of photography, strove to maximize the affordances of *televsual* composition.

The contrast is best illustrated by Watanabe's own diagram. Prior photographic eye cameras corresponded roughly to the central section of Watanabe's diagram, as extracted and outlined in red in the bottom portion of Figure 1. First, the experimental test subject was presented with a visual stimulus. A miniature bulb (labelled 'lamp') was then used to shine a light onto the cornea of the test subject at an angle. The cornea acted as a convex mirror, reflecting the light. This light then travelled through a concave lens, leading the rays to converge on a point on the surface of the film. The different points inscribed onto the film served thereafter as a record of eye movement for researchers.

What, then, was different about Watanabe's television eye camera? The key was the use of television technology to split signals and then mix them electronically into composite images, allowing one to pinpoint the exact spatial

3 Since 1984, Giken has changed its full name to Hōsō Gijutsu Kenkyūjo 放送技術研究所, and rendered its official English name as 'NHK Science and Technology Research Laboratories'. As my article deals with pre-1984 events, I have opted for its historical name instead.

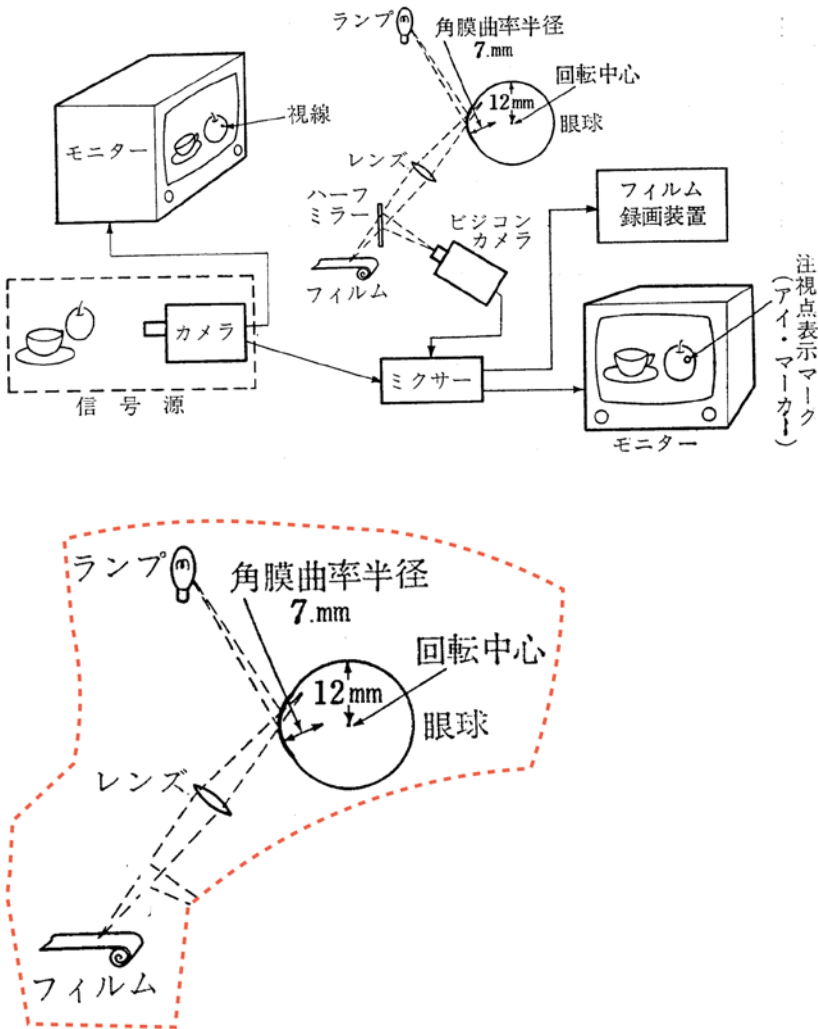


FIGURE 1 Top: Watanabe's schematic of his television eye camera
 Bottom: Standard photographic eye camera, modified from Watanabe's schematic. For English translations of the Japanese terms used in these diagrams, see Table 1.

SOURCE: WATANABE (1964: 11)

TABLE 1 English translation of Japanese terms in Figure 1 (left to right, top to bottom)

Japanese	English
モニター	TV monitor
視線	gaze
ランプ	lamp
角膜曲率半径	corneal curvature radius
回転中心	axis of rotation
眼球	eyeball
レンズ	lens
ハーフミラー	one-way mirror
フィルム	film
ビジコンカメラ	TV camera
フィルム録画装置	film recording apparatus
カメラ	TV camera
ミクサー	signal mixer
モニター	TV monitor
注視点表示マーク (アイ・マーカー)	Display mark of visual fixation point (eye marker)

location of the gaze relative to visual stimuli. First, Watanabe's apparatus used a television camera to broadcast a live feed to a television monitor, the TV images serving as visual stimuli for test subjects. The same signal from that television camera was also split and directed into a mixer. Meanwhile, Watanabe introduced a one-way mirror between the concave converging lens and the strip of film. On the one hand, the one-way mirror allowed light reflected from the test subject's cornea to pass through to a point on the film, as with previous eye cameras. On the other hand, the one-way mirror also reflected the light from the test subject's cornea into another television camera. The signal from this second television camera was also fed into the mixer. And here the core innovation: the mixer enabled the two signals – the first showing the original visual stimulus, and the other, the movement of the eye – to be *overlaid* as a composite image. Researchers could view that overlay in real time on the lower-right-hand monitor, and the overlaid feeds could be stored as a video recording for later consultation and analysis, as seen in Figure 2, in which the viewer's point of visual fixation appears as a bright white dot on the screen. In short, the television eye camera was an intermedial assemblage that combined principles of classical optics – lenses and mirrors – with the photochemical

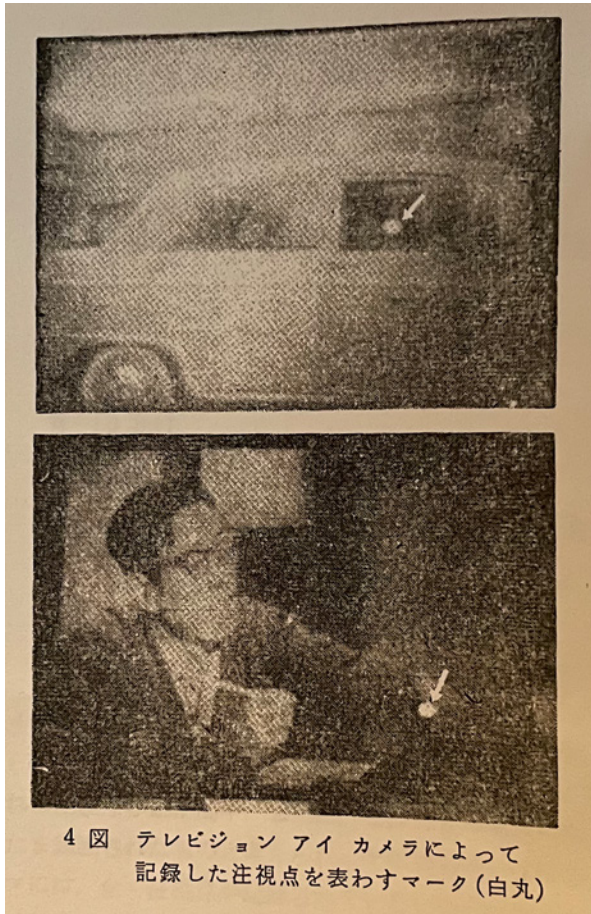


FIGURE 2 Composite overlays of points of visual fixation relative to television broadcasts
SOURCE: WATANABE (1963: 572)

principles of film recording, as well as the electrical principles of television. This allowed for real-time observation of eye movements relative to on-screen TV images, while also producing two durable inscriptions: the synchronic film that recorded the aggregate pattern of eye movements, and the diachronic video that recorded the overlaid feeds for later playback (Watanabe 1963, 1964).

A seemingly abstruse device, Watanabe's television eye camera was pursued as part of a practical project within the context of NHK's expansion of programming content and preparations for Olympics broadcasting. Experiments with the television eye camera offered comparative data on visual fixation points differentiated by program genre and viewer gender. In initial tests, viewers

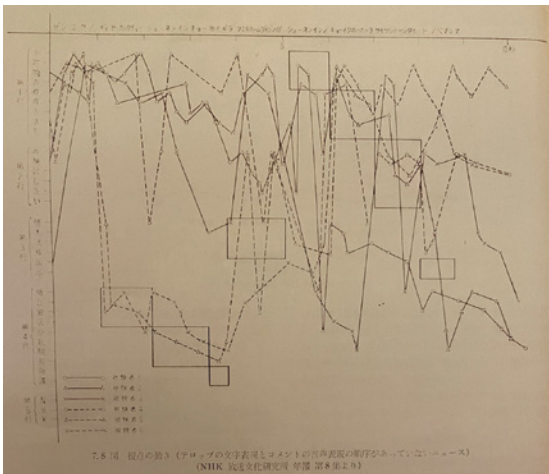
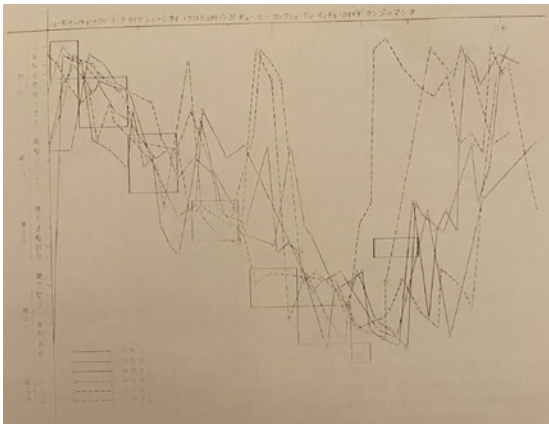
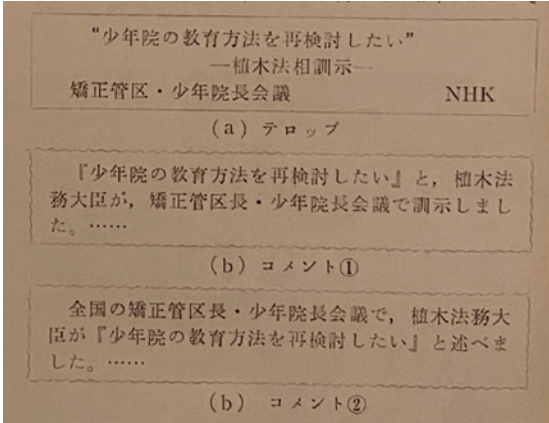


FIGURE 3
 Sample of results from Watanabe's experiment on visual fixation points in cases of simultaneous text and audio on television
 Top: transcript of telop text and two different audio tracks
 Middle: eye movements in case of alignment between telop text and the vocabulary and syntax of the audios
 Bottom: eye movements in case of alignment between telop text and audio vocabulary, but misalignment with audio syntax
 SOURCE: HIWATASHI AND WATANABE (1965: 16–17)

were subjected to news broadcasts, the drama *Gō-sutoppu monogatari*, and the nature documentary series *Shizen no arubamu*. Regardless of gender and genre, visual fixation tended to occur in the upper centre of the screen, leading to the recommendation that important content be positioned there. At a more granular level, Watanabe (1965) concluded that in the case of dramas, men's gaze lingered longer on the eyes of actors, whereas women were drawn to areas around the mouth and the juncture between torso and lower neck (*erimoto*). This data, he proposed, could better inform directors about their composition strategies for shots. In the case of news broadcasts, specific interest lay in understanding how to optimize new telop technologies for superimposing text onto screens, examining the relative effect of simultaneous on-screen text and oral delivery by a news announcer (Watanabe 1963: 575). Figure 4 illustrates one set of results. In this experiment, two test recordings were prepared. In the first recording, the news announcer read aloud the words in the box labelled 1. In the second recording, the announcer read aloud the words in box 2. For both recordings, the text in the box labelled (a) was superimposed on the bottom of the screen via telop. Although the informational content between the oral and written was in all instances identical, the first recording featured a close match of word order between oral delivery and on-screen text; the second, a reversal of word order. Watanabe's studies revealed that in the first instance, viewers' eyes roughly followed the text on screen, whereas in the second instance, their gazes became increasingly erratic, the eye jumping back and forth to different parts of the screen. As a result, he concluded that the use of telop could be optimized primarily by matching the *syntax* of oral and written language.

The use of the television eye camera to inform on-screen composition and shot framing was further directed toward the needs of the Olympics. Live broadcasts of sports in particular would benefit from heuristics that could inform the split-second decisions of directors regarding the camera feed to which they should cut at any given moment. With judo debuting as an Olympic sport in 1964, NHK turned its attention to studying the visual fixation points of judo spectators. Watanabe tested subjects who had judo training and those who did not, in order to understand where the eyes of initiated and uninitiated viewers might be drawn. Based on the results, he drew up lists of comparative visual fixation points and their movement over time relative to the types of holds and throws common in matches, as seen in Figure 4.

In this way, television was redeployed as an experimental apparatus to understand television viewers themselves. Specifically, in the form of the television eye camera, television became a device for transforming the voluntary and involuntary physiological behaviours of the human gaze into productive

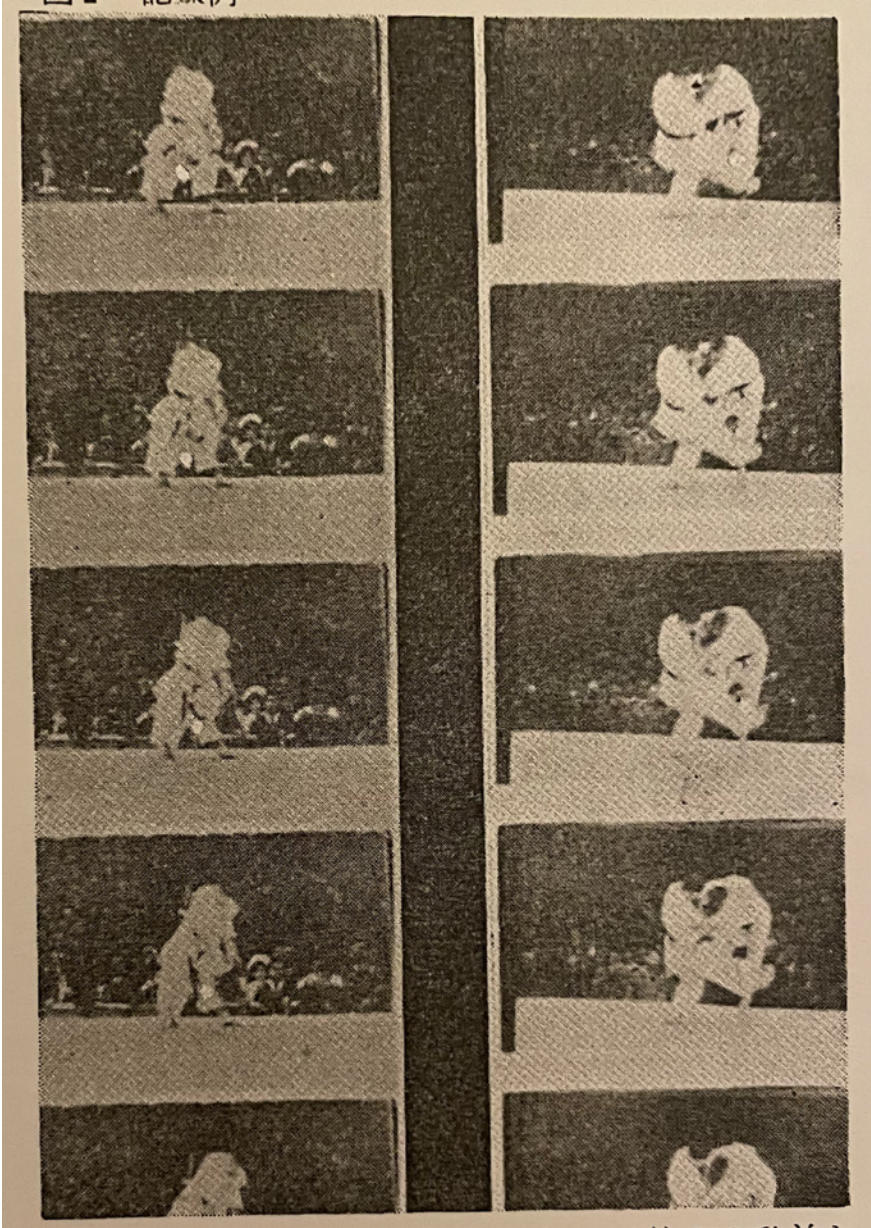
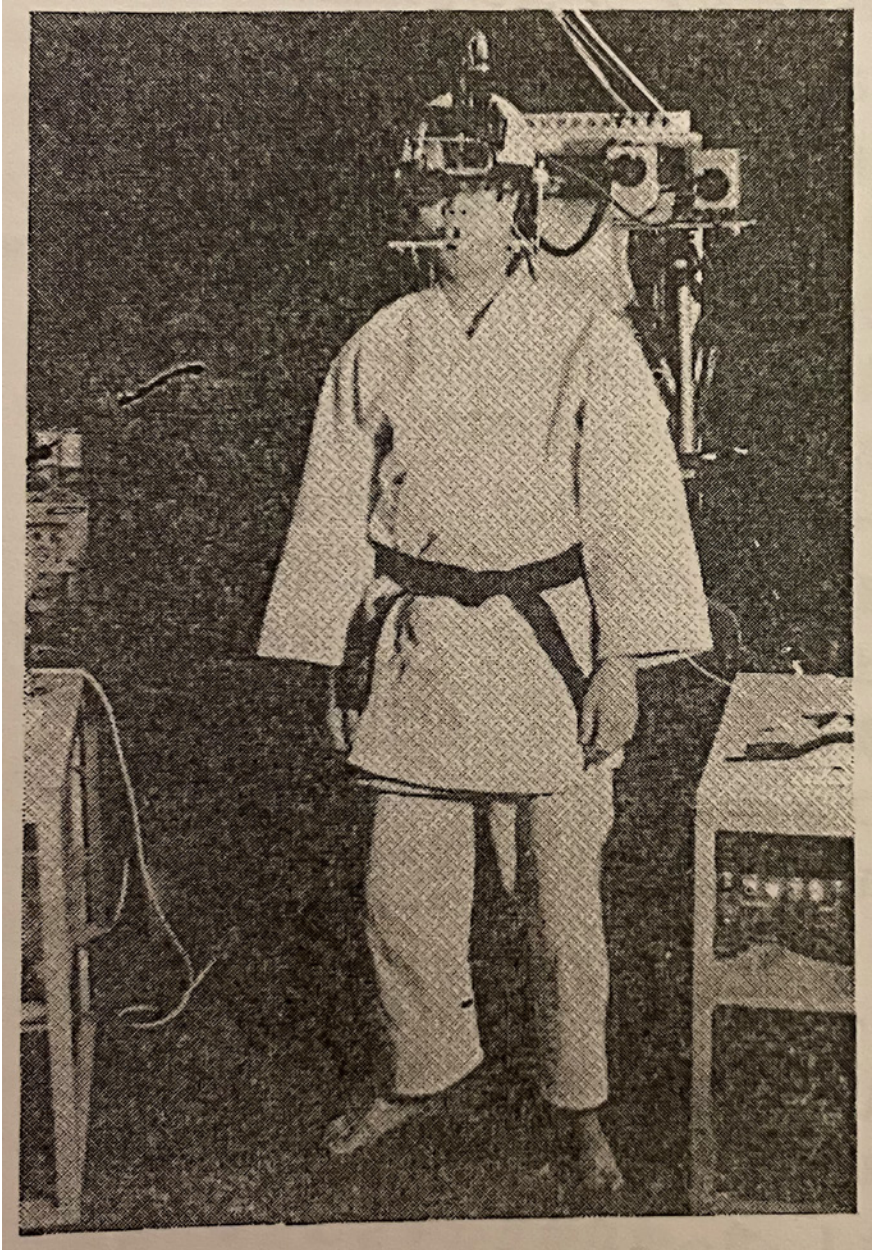


FIGURE 4 Visual fixation points of trained judo practitioners when watching televised matches, and a mobile version of the television eye camera attached to a judo practitioner.

SOURCE: MATSUMOTO AND WATANABE (1969: 120)



data. The most immediate surplus value emerging from this transformation of the televisual consumer into a data producer was a new vantage on strategies of production, direction, and editing across expanded genres of broadcast programming. But at the very same time, voices at Giken were also beginning to call for the mobilization of viewer data toward a new technology of vision itself.

2 From Audiences to the Eye-Brain Complex

Although NHK's initial interest in the television eye camera centred on the relationship between visual fixation and broadcasting content, Watanabe had larger ambitions. In his first published report on the television eye camera, Watanabe (1963: 571) concluded as follows: 'Investigating visual fixation points can be used not only for the applied needs of television, but for research into pattern recognition, ergonomics, and automata'. This, indeed, was not only his own view, but reflected a more expansive research agenda that his senior colleague, Hiwatashi Kenji, had been attempting to foster at Giken. Although by training a radio specialist, Hiwatashi's had turned his interests increasingly to visual phenomena after joining NHK, particularly through his participation in a project on instant slow-motion replay that would later debut at the 1964 Olympics. Above all, his core fascination settled on the relation between electrical systems and the nervous physiology of human perception (Satō et al. 2009). In November 1962, through connections at his alma mater, Hiwatashi invited Tōhoku University physiologist Motokawa Kōichi (1903–1971) to act as a guest lecturer at Giken for a year. Motokawa's inaugural address, 'The Audiovisual World', stressed that the true crux of vision lay not in understanding the function of the eye but, rather, the neural eye-brain complex. 'The retina', according to Motokawa (1963: 22), 'is essentially one part of the brain, and because the retina's structure and functions are similarly complex as those of the brain, both must be researched together'. Put differently, if, as Jonathan Crary (1992: 129–131) has suggested, the new visual technologies of the nineteenth century were 'contiguous instruments' of the eye, supplementing and enhancing its capacities, then under Motokawa's formulation, future visual technologies might be better thought of as supplementing and enhancing the capacities of a larger eye-brain complex tied together by networks of neurons.

Hiwatashi's efforts met with a fortuitous change in NHK's fundamental self-understanding. Riding on the success of the Tokyo Olympics, NHK leaders expressed the opinion that televisual technologies in Japan had by now attained, if not surpassed, Western standards. Henceforth, NHK's goal should not merely be immediate 'applied' research aimed at 'catching up' with foreign

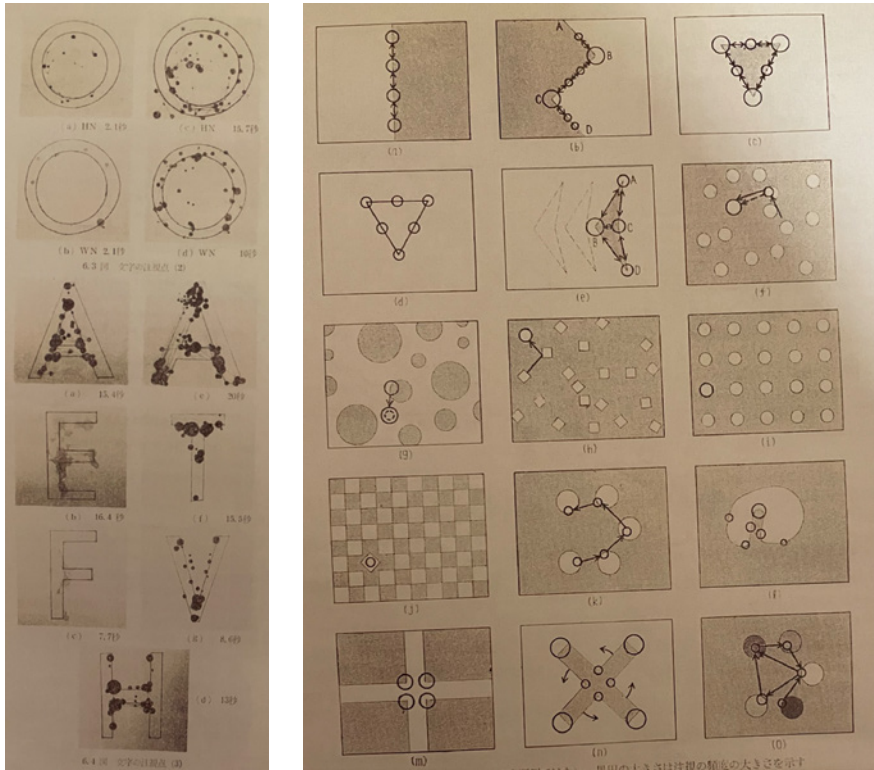


FIGURE 5 Accumulated visual fixation points for various capital letters and shapes
 SOURCE: HIWATASHI AND WATANABE (1965: 15)

technologies, but long-term ‘basic research’ (*kiso kagaku*) that would blaze a trail toward the future possibilities of television (Itō 2024: 190). In short, at a moment when the post-war developmental state and its ‘economic miracle’ had begun to be vaunted as a viable – if not desirable – alternative to liberal capitalism, televisual technologies emerged as a representative site for the inversion of Japan’s long-held sentiment of temporal inferiority vis-à-vis the West. Japanese television was no longer late. Soon, it would shape the world’s future.

The manifestation of NHK’s temporal reorientation, in January 1965, was the creation of a new institute – the BSRL – that would house a new kind of research unit: VAISU. As NHK’s own announcement made clear, the unit sought ‘the elucidation of the superior visual and auditory information-processing mechanisms unique to human begins in anticipation of the rapid progress of future broadcast technologies’ (Nihon Hōsō Kyōkai 1964: 3). In defining vision as a mechanism of ‘information processing’, VAISU’s mission statement equated the future of broadcast television with the design of machines that

would emulate human eye-brain behaviour. Therefore, by 1965 already, the ‘screen-brain apparatus’ described by Thomas Lamarre (2018: 24–25, 33–108) had materialized as one prominent axis of NHK’s research agenda. Through VAISU, neuroscience became internal to the very constitution of televisual technologies themselves, the goal being to endow the latter with their own information-processing ‘brain’.

Fittingly, VAISU began with an interdisciplinary roster. Hiwatashi and Watanabe were immediately transferred to the unit, with the former as lead researcher, and they were soon joined by Fukada Yoshirō, a physiologist; Ōgushi Kengo and Sakai Hisao, researchers in psychoacoustics; Nagata Shōjirō, a cognitive neuroscientist; Kira Kenji and Kutsuzawa Junnosuke, computer graphics researchers; and Uesaka Yoshinori, a computer scientist. During their first year of activity, Watanabe’s television eye camera became the focal point of research. Under Hiwatashi, the device was repurposed to study the automation of visual cognition. Specifically, in a series of collaborative projects with Watanabe from 1965–1966, Hiwatashi argued that the television eye camera revealed the operation of visual cognition as an ‘automatic control system’ regulated by feedback loops. The argument started with a claim about the physiology of visual information processing. Although together the two eyes furnished humans with an overall field of peripheral vision of ca. 60 degrees at top and bottom, left and right, only a small part of the eye – the fovea centralis at the back of the retina – possessed high visual acuity, acting as the eye’s nerve centre. This fovea comprised only approximately one degree of the field of vision, and the more an object deviated from it, the poorer the perception of that object. By the time one reached 30 degrees of deviation, visual acuity dropped below 50%. Recognition of objects was thus a function of feedback between peripheral vision and the fovea. When peripheral cells of the eye registered differences in contrast, colour, and motion, they triggered a reflex that would cause the eye to position the fovea centralis on an object, enabling accurate recognition of that object. The various visual fixation points and movements recorded by the television eye camera thus also served as a record of the steps taken in an overall cognitive process of vision that was self-regulating and involuntary, and thus potentially susceptible to machine automation (Hiwatashi & Watanabe 1965, 1966).

To demonstrate this, Hiwatashi and Watanabe redeployed the television eye camera in experiments that focused not on visual fixation relative to NHK programs, but fixation in simple letters of the roman alphabet and geometric shapes (see Figure 5). The results indicated that the eye fixated only on certain regions of an object – primarily corners and edges – and furthermore did so repeatedly, ignoring other parts of the object. Visual cognition, in other words, could be conceptualized as a process that pooled together a minimal set of

stimuli or identifying features, reasoning from this reduced set of features to the likely identity of the whole. For instance, in the more familiar case of the upper-case letter T, the eye fixated continually on the corners of the intersection between the letter's horizontal and vertical lines, with little attention to other components of the letter. It was, in short, only through a few select features that the brain recognized T (Hiwatashi & Watanabe 1965).

The more one understood this 'control system' wherein eye-brain feedback powered an automatic process of 'feature extraction', the more one might be able to artificially simulate it. In this manner, Hiwatashi (1966a: 77) began to argue that 'image transmission and processing' technologies such as television were really a stepping stone towards AI, proposing that NHK become the hub for a new 'bionics of vision' (*shikaku no baionikusu*) – that is, the engineering of new visual technologies modelled after the operation of biological systems. Introducing the work of the VAISU to anglophone audiences in the pages of the *New Scientist*, Hiwatashi (1966b) dubbed the group's goal a 'quest for a seeing machine'. The initial basis for this 'seeing machine' would be data drawn from television eye camera studies, which could be used to design computers and cameras capable of predicting an audience's visual fixation points. 'Electronic simulation of the visual system', Hiwatashi wrote, 'encourages hopes that computers can acquire similar faculties, and that television cameras will be able to search out the interest in a scene just as a human observer would do' (232).

This articulation of a bionics of vision occurred from 1965–1966. And in this way, the stage was set for the arrival of a young engineer, fresh from his doctorate at Kyoto University's Department of Engineering: Fukushima Kunihiko.

3 The Neocognitron

Artificial intelligence and machine learning were not yet at the front of Fukushima's mind when he began employment at NHK. Hired in 1958 while still pursuing a doctorate at Kyoto University, Fukushima's early research instead centred on signal compression. Compression had become a topic of ever greater critical importance given the massive bandwidth consumed by television's rapid national expansion. By mid-1961, Japan had over 110 public and private TV stations covering 82 percent of the archipelago, with larger stations broadcasting seven or more channels of content (Adam 1961: 14; Bush 1961: 21). Standard compression techniques utilized statistical analysis to identify redundancies, i.e., repeated patterns, in electrical signals, targeting these redundancies for elimination. The first half of Fukushima's doctoral dissertation (1966) explored different approaches to bandwidth reduction on this

basis, albeit coming to the dismal conclusion that television signals – unlike those in telephony – contained insufficient statistical redundancy to be susceptible to meaningful compression.

The second half of the dissertation then pivoted. The traditional approach to compression, Fukushima noted, focused on the informational content of signals from the perspective of sending and receiving devices. This, however, was a flawed assumption. The true receiver at the end of a televisual signal was not the television apparatus itself, Fukushima argued, but rather a different device – a device with a different kind of ‘camera’, ‘transmission lines’, and information processing unit. In more familiar terms, one called this camera an ‘eye’, the transmission lines ‘nerves’, and the processing unit a ‘brain’. Compression, put simply, was not a matter of redundancy within the signal, but redundancy relative to the human eye–brain complex (Fukushima 1966, Abstract). More efficient transmission of TV signals was thus primarily a matter of eye-brain physiology and psychology, and only secondarily a matter of electrical engineering.

In this sense, Fukushima’s ideas stood squarely in line with what Jonathan Sterne (2012: 19, 32–60), in reference to audio compression, has called ‘perceptual technics’, i.e., ‘the application of perceptual research for the purposes of economizing signals’. More narrowly, Fukushima was proposing for television a kind of ‘lossy compression’ that would eliminate parts of signals which lay outside the probable threshold of human visual perception. For NHK, this kind of lossy compression worked alongside attempts to understand its viewers, functioning as part of a two-pronged plan of economization. On the one hand, early uses of Watanabe’s television eye camera extracted data from users, transforming the act of consuming television into a generator of surplus value for TV production. On the other hand, the same data indicated that they could employ a physiological model of vision to *reduce* the amount of data sent to viewers and thus minimize bandwidth costs.

Although Fukushima made the case for this approach in his dissertation, he reluctantly concluded that neuroscientific research at the time was not yet advanced enough to yield models for workable visual compression algorithms. Nevertheless, it is clear that as early as 1964, Fukushima had begun to link lossy compression and questions of AI. In March of that year, Giken organized a roundtable, for which Watanabe Akira served as moderator, to address the theme of ‘Pattern Recognition and Broadcasting’. Present at the discussion was Fukushima, who was struck in particular by analogies between the problems faced by compression engineering and the puzzle of feature extraction within pattern recognition. The way in which feedback loops between eye and brain directed the eye to examine only minimal features was essentially a means to reduce the amount of information that the brain had to process in order to

recognize an object. In short, eye-brain feedback was in fact an algorithm of lossy compression. Currently, Fukushima explained (Watanabe et al. 1964b: 157),

it is impossible to significantly compress the bandwidth of television signals, because whereas there is feedback between the human eye and brain, there is no feedback across the entirety of television systems. Thus, television systems are forced to transmit information that is in fact unnecessary for the [human] visual system.

Were it only possible to ‘endow a device with this kind of feedback function, the device would become vastly more efficient’ (Ibid.). Furthermore, an algorithm that could compress television signals efficiently by extracting only necessary features might in turn serve as the basis for a pattern recognition machine. Fukushima (Watanabe et al. 1964a: 97) began to perceive lossy compression as sharing identical goals with AI research, telling the roundtable participants:

Once feature extraction is possible, the problem of pattern recognition seems basically solved. In other words, pattern recognition research all comes down to the task of finding which features are best to extract, and how best to extract them.

Hiwatashi and Watanabe’s use of the television eye camera offered hope that the mechanisms of lossy compression in human vision might actually be clarified, offering visual data that could serve as a preliminary model of the feature extraction process in humans. The results of the television eye camera thus became the basis for Fukushima’s transition from compression to computer vision. Leaving his group on compression, Fukushima joined VAISU and began pursuing this new project.

The first stage of work from 1966–1971 involved a scramble to catch up with the state of computer vision research, leading to the creation in 1971 of the Cognitron, Fukushima’s ‘deep’ version of Frank Rosenblatt’s 1957 Perceptron. Deepness, in this context, referred to the existence of two or more layers in a neural network between input and output layers. From a historical perspective, Fukushima’s Cognitron was nothing novel. Although at the time only one other lab in Japan – at the University of Tokyo – was developing a similar deep neural network, the underlying design itself had already been actualized in the USSR. However, tests with the Cognitron did demonstrate to Fukushima the fundamental barrier to genuine feature detection and extraction in machines: the problem of dimensionality. Deep neural networks, including the Cognitron, flattened images into a single line of data by converting their pixel values into a one-dimensional series. Each pixel value in this one-dimensional line of input

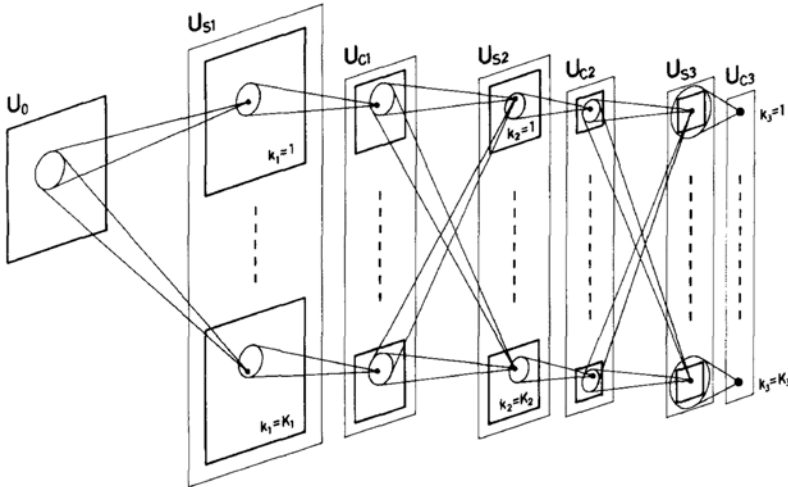


FIGURE 6 Example of the operation of convolutional layers in Fukushima's initial Neocognitron
SOURCE: FUKUSHIMA (1980: 198)

data was then compared with the value of stored images of pre-identified objects. Far from selective features, this approach required processing the entirety of the image again and again from each layer in the network to the next (Dobson 2023).

Yet according to the television eye camera, the human gaze worked differently. It worked through an automatic, self-regulated feedback mechanism that ignored vast swathes of the object, searching only for certain features. Once a potential feature was found, the brain directed the eye to return to the areas around those features, fixing the gaze on specific two-dimensional regions, particularly areas where edges intersected to create angles or curves (see Figure 5). The advantage of this approach was not only a reduction in the amount of information that needed to be processed – the initial attraction which had drawn Fukushima's interest from compression to pattern recognition. It was also what enabled humans to recognize objects despite changes in size, shape, position, and lighting conditions, all of which could be set aside so long as certain basic feature criteria were met. Hidden in this model of vision was an epistemic claim. Visual cognition is possible precisely because of what people do not see, or, to be exact, because people see only in a highly selective manner. Hence, an artificial *intelligence* capable of seeing would ideally be a machine capable of *ignoring* what it sees in the right ways. Intelligence is thus learned ignorance.

The hint as to how this learned ignorance might be implemented in a machine was provided by Motokawa Kōichi. As mentioned above, Motokawa

had earlier visited NHK over the course of 1962–1963 as an invited lecturer. After VAISU's founding, however, he was brought back in a permanent advisory capacity to shape the unit's programme of physiological research. Through Motokawa, Fukushima became aware of studies on the visual cortex of cats by David Hubel and Torstein Wiesel. These studies, performed at Harvard's Neurophysiology Laboratory, mapped a hierarchical structure of the visual cortex in which 'simple cells' detected discrete image features and then passed these features to 'complex cells' that grouped the features together (Hubel & Wiesel 1962). Important here was the insight that different features activated different pathways through a structured chain of simple and complex cells. This suggested to Fukushima two points. First, a deep neural network might be constructed through successive layers that *filtered* for potential features then *pooled* the extracted results together into a two-dimensional array (see Figure 6), mimicking the functions of simple and complex cells, respectively. Second, the identification of objects should not be based not on a comparison of whether the individual pixels of an input image were roughly identical to a stored image, but, rather, whether the *pathway of connections* through successive filters and pools was roughly similar to stored precedents. This approach thus added an ontological claim to the epistemic claim of learned ignorance. A given object should not be fundamentally defined by the sum of its size, shape, colour, and other properties. Instead, an object should ultimately be defined by the specific neural structure of the feedback mechanism that it activates – the unique pathway that it constructs between the eye and the brain.

Once equipped with successive operations of filtering and pooling, a deep neural network would become, in mathematical terms, *convolutional*. The end result of Fukushima's search for convolution, announced first in Japanese in 1979, then in English in 1980, was the Neocognitron – the world's first deep CNN. As seen in Figure 6, the Neocognitron comprised six layers (*sō*) organized into three successive 'levels' (*dan*). In the figure, the letter A serves as visual input at U_0 , the input layer. This input then feeds into the first level, comprising two layers, U_{S1} and U_{C1} . The U_{S1} layer filters the input image for simple features such as intersections and endpoints, shown in the smaller bold circles in the image. These simple features are then passed on to layer U_{C1} , which pools them together, as illustrated by the larger circles in thin outline. At the first level, five features in the letter A detected by U_{S1} are pooled into three groups by U_{C1} . The pooled groups are then passed to the next level (comprising the layers U_{S2} and U_{C2}), where the process is repeated again at a higher level of abstraction. U_{S2} filters the pooled data for intersections and endpoints, and then U_{C2} pools the results together – this time into one, rather than three pools. The end result of this procedure of filtering and pooling, as seen at the final level (U_{S3} and

U_{C3} layers), is a reduced version – one might also say *compressed* version – of the original input based on the reconstruction of select features. The pathway through the network taken to produce this reduced version is then stored as the letter A. By repeating this process, the neural network learns to ‘see’ the letter A. If, in future iterations, an input activates a sufficiently similar pathway of filtering and pooling – a similar set of connections between layers – then it will be classified as A (Fukushima 1979, 1980). With the Neocognitron, a type of ‘seeing machine’ had, at least in theory, been achieved.

4 Conclusion

Television was at the origins of this seeing machine: as an institutional home for research, as an experimental apparatus, and as a means to understand the human subject. Beginning in the 1960s, the rapid expansion of television in Japan prompted concerns about how to grasp and maximize this new political economy. Watanabe’s television eye camera addressed the centrality of attention as a commodity in this political economy at a physiological level, extracting data on visual fixation from consumers to shape broadcast production. Data from that same television eye camera, however, also suggested that the act of seeing itself comprised a highly efficient, self-regulating information processing system. Understanding that information processing system became the impetus for the more radical project of a ‘bionics of vision’, which wed the development of broadcast technologies to the artificial emulation of human physiology and psychology. At first, this, too, served an eminent economizing purpose: the reduction of bandwidth, and thus television transmission costs, through lossy compression. Yet joining NHK to work on televisual signal compression, Fukushima instead found in the bionics of vision a thread that led to pattern recognition machines: feature extraction. If a machine could be engineered to structurally mimic the feedback loops through which humans minimized their informational load, then it could also ‘learn’ by associating each class of object with a unique neural pathway of feature extraction and pooling.

On the one hand, then, the invention of deep CNNs was eminently rooted in the specificities of a reorientation in NHK’s research. This reorientation was born from the sense that Japanese broadcasting technologies had ‘caught up’ temporally with the West. It was Japan now that would lead the world in defining television’s future, and this future, for NHK, lay in the ‘bionics of vision’. The neural network model at the centre of contemporary computer vision was thus created in response to a set of local needs seemingly distant from today’s world of AI, and different from genealogies of Cold War cybernetics in the

United States. It is thus necessary to recover the local needs and institutions which shaped a plurality of divergent, yet partially overlapping, cultures of AI research, lest histories of AI merely recount a certain arc of American globalization in the second half of the twentieth century.

On the other hand, NHK's 'bionics of vision' also reveals facets of the genealogy of computer vision that extend beyond the Japanese context. As Iida Yutaka (2016: 19, 339–348) has argued, work on television remains dominated by a certain 'broadcast history' model of research that interests itself in questions of mass communication. Based on this, a discourse has emerged proclaiming the 'end of television', 'post-television', and, in Japanese, 'departure from television' (*terebi banare*) (Katz & Scannell 2009; Tse 2024). Yet in fact, as the key technology alongside fax that enabled the conversion of images into electrical signals for information processing, television arguably ushered in the birth of the digital image, and thus the computational processing of visual data. For this reason, the Institute of Image Information and Television Engineers in Japan (Eizō Jōhō Media Gakkai 2009: i) has continued to maintain that it was not computing, but television, that marked the 'greatest development in the 150 years of the history of images since cinema'. Just as studies of digital audio have sought to decentre the role of the computer, tracing histories back to early twentieth-century telephony (Mills & Tresch 2011: 8, 10; Sterne 2012), so too do we require work that can excavate television – across physiology, psychology, neuroscience, and engineering – as a structuring force in the genesis of modes of computer vision, including visual pattern recognition AI.

In short, Lamarre's (2018: 4) call for a tracing of 'genealogical transitions between one assembling of television and another' can in fact be positioned as the central axis along which we approach the broader project of historicizing how images came to be digitally processed, recognized, and classified. Writing in 1984, Jonathan Crary pointed to the 'increasing adjacency of television to telecommunications and computers' (1984: 284), a web tied together by the use of 'information, structured by automated data processing [...] [as] a new kind of raw material' (286).⁴ The story of Fukushima, however, reveals the opposite trajectory: it was television that gave rise to a new model of eye-brain information processing, only after which the Neocognitron became possible. Television was not growing adjacent to computing. Rather, television was a paradigmatic form of electronic visuality (Galili 2020) to which computing itself was growing adjacent. Developments at NHK perhaps reveal this history more prominently. But they are surely not alone.

4 I am indebted to Ōkubo Ryō for this reference.

Acknowledgements

This article owes a significant debt to members of the Historical Media Epistemology group at Kyoto University and the conferences ‘Techniques of the Shichōsha’, held in June 2023, and ‘Terebijon Ākaibuzu o saisōzō suru’, held in December 2023. Special thanks are due to Okazawa Yasuhiro and Maejima Masahiro. The author is also grateful for funding from the Suntory Foundation and the Great Britain Sasakawa Foundation.

References

- Adam, Kenneth (1961), ‘Broadcasting in Japan’. *The Listener*, 66(1694), 14–17.
- Azar, Mitra, Cox, Geoff, & Impett, Leonardo (2021), ‘Introduction: Ways of Machine Seeing’. *AI & Society: Knowledge, Culture and Communication*, 36(4), 1093–1104.
- Boehringer, Carl H. (1951, August 24), [Letter to Department of State]. 994.50, Other Internal Affairs, Communications, Transportation, Science, Japan, Television, Facsimile Transmission, March 9, 1950–February 16, 1954, p. 60. National Archives, Washington, DC.
- Bush, Lewis (1961), ‘Television in Japan’. *The Listener*, 66(1694), 21–23.
- Crary, Jonathan (1984), ‘Eclipse of the Spectacle’. In: Wallis, Brian (Ed.), *Art after Modernism: Rethinking Representation*. New York: New Museum of Contemporary Art, pp. 283–294.
- Crary, Jonathan (1992), *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge, MA: MIT Press.
- Crawford, Kate, & Paglen, Trevor (2019), ‘The Politics of Images in Machine Learning Training Sets’. *Excavating AI*. Retrieved from www.excavating.ai.
- Dobson, James E. (2023), *The Birth of Computer Vision*. Minneapolis: University of Minnesota Press.
- Edwards, Paul (1997), *The Closed World: Computers and the Politics of Discourse in Cold War America*. Cambridge, MA: MIT Press.
- Eizō Jōhō Media Gakkai (Eds.) (2009), *Gazō to shikaku jōhō kagaku (Image and Visual Information Science)*. Tokyo: Koronasha.
- Erickson, Paul, Klein, Judy L., Daston, Lorraine, Lemov, Rebecca, Sturm, Thomas, & Gordin, Michael D. (2013), *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. Chicago: University of Chicago Press.
- Franklin Institute (2020), ‘Kunihiko Fukushima’. Retrieved from <https://www.fi.edu/en/laureates/kunihiko-fukushima/>.
- Fukushima, Kunihiko (1966), ‘Terebijon shingō no densō taiiki asshuku no kenkyū’ (Research on the Transmission Bandwidth Compression of Television Signals). PhD diss., Kyoto University.

- Fukushima, Kunihiko (1979), 'Ichizure ni eikyō sarenai patān ninshiki kikō no shinkei kairo moderu: Neocogunitoron' (A Neural Network Model for a Pattern Recognition Mechanism Unaffected by Shifts in Position: Neocognitron). *Denshi jōhō tsūshin gakkai ronbunshi A*, J62-A(10), 658–665.
- Fukushima, Kunihiko (1980), 'Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position'. *Biological Cybernetics*, 36, 193–202.
- Furuhata, Yuriko (2019), 'Tange Lab and Biopolitics: From the Geopolitics of the Living Sphere to the Nervous System of the Nation'. In: Inoue, Mayumo & Choe, Steve (Eds.), *Beyond Imperial Aesthetics: Theories of Art and Politics in East Asia*. Hong Kong: Hong Kong University Press, pp. 219–242.
- Galili, Doron (2020), *Seeing by Electricity: The Emergence of Television, 1878–1939*. Durham, NC: Duke University Press.
- Galison, Peter (1994), 'The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision'. *Critical Inquiry*, 21(1), 228–266.
- Halpern, Orit (2015), *Beautiful Data: A History of Vision and Reason since 1945*. Durham, NC: Duke University Press.
- Halpern, Orit, & Mitchell, Robert (2023), *The Smartness Mandate*. Cambridge, MA: MIT Press.
- Hiwatashi, Kenji (1965), 'Shikaku kikō to sono moderu' (The Visual Mechanism and Its Model). *NHK Giken geppō*, 8(7), 407–411.
- Hiwatashi, Kenji (1966a), 'Sōsetsu: shikaku kenkyū ni tsuite' (Introduction: On Vision Research). *NHK Gijutsu ken'yū*, 18(2), 75–78.
- Hiwatashi, Kenji (1966b), 'The Quest for a 'Seeing' Machine'. *New Scientist*, 32(519), 232–235.
- Hiwatashi, Kenji, & Watanabe, Akira (1965), 'Gazō to chūshiten no bunpu' (Images and the Distribution of Visual Fixation Points). *NHK Gijutsu ken'yū*, 17(1), 4–20.
- Hiwatashi, Kenji, & Watanabe, Akira (1966), 'Gankyū undō no seigyō kikō' (The Control Mechanism for Eye Movements). *Rinshō kagaku*, 2(10), 1403–1414.
- Holthusen, Henry F. (1951, November 15). [Letter to Alexis Johnson]. 994.50, Other Internal Affairs, Communications, Transportation, Science, Japan, Television, Facsimile Transmission, July 1, 1950–November 20, 1951, pp. 2–13. National Archives, Washington, DC.
- Hubel, D., & Wiesel, T.N. (1962), 'Receptive Fields, Binocular Interaction and Functional Architecture in Cat's Visual Cortex'. *Journal of Physiology*, 160, 106–154.
- Humphrey, David (2022), 'Sensing the Human: Biometric Surveillance in the Japanese Technology Industry'. *Media, Culture & Society*, 44(1), 72–87.
- Iida, Yutaka (2016), *Terebi ga misemono datta koro: shoki terebijon no kōkogaku (When TV Was Spectacle: The Archaeology of Early Television)*. Tokyo: Seiyūsha.
- Ikuta, Masaki (1960), 'Television in Japan'. *International Communication Gazette*, 6(1), 43–50.

- Itō, Takayuki (2024), 'Terebi no hattatsu to kiso kenkyū' (The Development of TV and Basic Research). In: NHK Hōsō Bunka Kenkyūjo (Ed.), *Gijutsu no hattatsu to hōsō media. (Technological Development and Broadcast Media)*. Tokyo: NHK shuppan, pp. 185–202.
- Jones, Matthew (2018), 'How We Became Instrumentalists (Again): Data Positivism since World War II'. *Historical Studies in the Natural Sciences*, 48(5), 673–684.
- Jones, Matthew (2023), 'AI in History'. *American Historical Review*, 128(3), 1360–1367.
- Katz, Elihu, & Scannell, Paddy (Eds.) (2009), 'The End of Television? Its Impact on the World So Far'. *Annals of the American Academy of Political and Social Science*, 625 [special issue].
- Kittler, Friedrich (2010), *Optical Media: Berlin Lectures 1999*. Translated by Anthony Enns. London: Polity.
- Kline, Ronald (2015), *The Cybernetics Moment, or Why We Call Our Age the Information Age*. Baltimore: Johns Hopkins University Press.
- Krizhevsky, Alex, Sutskever, Ilya, & Hinton, Geoffrey E. (2012), 'ImageNet Classification with Deep Convolutional Neural Networks'. In: Pereira, F., Burges, C.J., Bottou, L., & Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems*. Red Hook, NY: Curran Associates, pp. 1097–1105.
- Lamarre, Thomas (2018), *The Anime Ecology: A Genealogy of Television, Animation, and Game Media*. Minneapolis: University of Minnesota Press.
- Liu, Xiao (2019), *Information Fantasies: Precarious Mediation in Postsocialist China*. Minneapolis: University of Minnesota Press.
- Loukissas, Yanni (2019), *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.
- Maslow, Sebastian (2015), 'A Blueprint for a Strong Japan? Abe Shinzō and Japan's Evolving Security System'. *Asian Survey*, 55(4), 739–765.
- Matsumoto, Yoshizō, & Watanabe, Akira (1969), 'Jūdō tanrensha no chūshiten ni kansuru kenkyū' (Research on the Visual Fixation Points of Judo Practitioners). *Kōdōkan jūdō kagaku kenkyūkai kiyō*, (3), 103–120.
- Mills, Mara, & Tresch, John (2011), 'Introduction: Audio/Visual'. *Grey Room*, 43, 6–15.
- Mitsuya, Keiko (2014), 'Tanjō kara 60-nen o heta terebi shichō' (TV Viewship Sixty Years After The Birth of Broadcast Television in Japan). *NHK Hōsō bunka kenkyūjo nenpō*, 58, 7–44.
- Motokawa, Kōichi (1963), 'Shichōkaku no sekai' (The Audiovisual World). *NHK Giken geppō*, 6(1), 22–27.
- Murphy, Sheila (2011), *How Television Invented New Media*. New Brunswick, NJ: Rutgers University Press.
- Nihon Hōsō Kyōkai Sōgō Gijutsu Kenkyūjo, & Hōsō Kagaku Gijutsu Kenkyūjo (Eds.) (1964), *Kenkyū nenpō Shōwa 39 nendo. (Annual Research 1964)* Tokyo: Nihon hōsō kyōkai sōgō gijutsu kenkyūjo.

- Okazawa, Yasuhiro (2023), 'Techniques of the Shichōsha: On the Technoscientific Formation of Cultural Subjects'. *Repre*, 49, <https://www.repre.org/repre/vol49/topics/04/>. [in Japanese].
- Satō, Katsuaki et al. (2009), 'Ōraru hisutorī Hiwatashi Kenji meiyō kaiin' (Oral History: Emeritus Member Hiwatashi Kenji). *Eizō jōhō media gakkaiishi*, 63(7), 892–895.
- Steinberg, Marc, Zhang, Lin, & Mukherjee, Rahul (2024), 'Platform Capitalisms and Platform Cultures'. *International Journal of Cultural Studies*. <https://doi.org/10.1177/13678779231223544>.
- Sterne, Jonathan (2012), *MP3: The Meaning of a Format*. Durham, NC: Duke University Press.
- Striphas, Ted (2015), 'Algorithmic Culture'. *European Journal of Cultural Studies*, 18(4–5), 395–412.
- 'Television in Japan'. (1953, February 2). *The Times* (London), 6.
- Tse, Yu-Kei (2024), "'Terebi banare": Historicising Internet-Distributed Television and the "Departure from Television" in Japan'. *International Journal of Cultural Studies*, 27(1), 99–118.
- Virilio, Paul (1989), *War and Cinema: The Logistics of Perception*. Translated by Patrick Camiller. London: Verso.
- Watanabe, Akira (1963), 'Terebi gamen no chūshiten no ugoki' (The Movement of Visual Fixation Points on the TV Screen). *NHK Giken geppō*, 6(11), 571–575.
- Watanabe, Akira (1964), 'Terebi gazō no chūshiten' (Visual Fixation Points of TV Images). *Terebijon*, 18(10), 10–11.
- Watanabe, Akira (1965), 'Shichōsha wa terebi no gamen no doko o miru ka' (Where Do Viewers Look on TV Screens?). *Hōsō bunka*, 20(5), 24–25.
- Watanabe, Akira, Kuroki, Sōichirō, Hiwatashi, Kenji, Uesaka, Yoshinori, Fukushima, Kunihiko, Ishikawa, Hiroshi, Itō, Gen, Murasaki, Keisuke, Terayama, Yoshirō, Fujimura, Yasushi, & Ōta, Tatsuji (1964a), 'Patān ninshiki to hōsō (1)' (Pattern Recognition and Broadcasting 1). *NHK Giken geppō*, 7(2), 91–99.
- Watanabe, Akira, Kuroki, Sōichirō, Hiwatashi, Kenji, Uesaka, Yoshinori, Fukushima, Kunihiko, Ishikawa, Hiroshi, Itō, Gen, Murasaki, Keisuke, Terayama, Yoshirō, Fujimura, Yasushi, & Ōta, Tatsuji (1964b), 'Patān ninshiki to hōsō (2)' (Pattern Recognition and Broadcasting 2). *NHK Giken geppō*, 7(3), 153–161.
- Wiener, Norbert (1954), *The Human Use of Human Beings: Cybernetics and Society* (2nd ed.). New York: Doubleday.
- Yoshimi, Shun'ya (2014), 'From Street Corner to Living Room: Domestication of TV Culture and National Time/Narrative'. *Mechademia: Second Arc*, 9, 126–142.